Bayesian Analysis of Stochastic Volatility Models: Modelling and Applications

Joanna Jia Jia Wang

A thesis submitted in fulfillment of
the requirements for the degree of
Doctor of Philosophy

School of Mathematics and Statistics
University of Sydney

January 2012

Abstract

It has long been recognised that the return volatility of financial assets tends to vary over time with persistence. One way of modelling this feature is allowing the conditional variance to be a function of previous observations and past variances, which leads to the ARCH-type models developed by Engle (1982). An alternative to the ARCH framework is a model which specifies the variance to follow a latent stochastic process, known as the stochastic volatility (SV) model. A fundamental theme of this thesis is the extension of current SV model specifications in a Bayesian framework: firstly to increase model flexibility in capturing prominent empirical characteristics of financial time series data and secondly widen the range of applications of SV models.

Firstly, we consider a heavy-tailed SV model with leverage effect, modelled by a (negative) correlation between the innovations of returns and latent volatilities. Choy *et al.* (2008) investigated this model by assigning a bivariate Student-t distribution as a scale mixture of bivariate normal distributions. We propose a novel alternative formulation by first deriving a conditional Student-t distribution for the returns and a marginal Student-t distribution for the volatilities. This is followed by an employment of a scale mixture of normals representation for these two Student-t distributions. Our approach separates the sources of outliers by distinguishing between outliers generated by the return process or by the volatility process, and hence providing an improvement over the approach of Choy *et al.* (2008).

With motivation from the finding that the SV model with a conditional normal distribution does not adequately account for the leptokurtic distribution of the returns, we study the use of different heavy-tailed distributions in SV models and investigate the effects of choosing different conditional distributions for the returns on both the estimates of model parameters and latent volatilities. More specifically, we propose the use of a generalised-t (GT) distribution, which incorporates commonly used distributions in the SV literature, including the Student-t and exponential power distributions, into a unified distributional framework, expressed as a scale mixture of uniform distributions. The SV model with GT distribution is

further extended to accommodate the asymmetric behaviour between the returns and volatilities, adopting an alternative specification to the bivariate SV model in which the leverage effect is modelled through a constant correlation coefficient of the bivariate distribution. In this alternative specification, both the sign and the magnitude of the previous return is included in the volatility equation (Asai and McAleer (2005a); So *et al.* (2002)). Asai and McAleer (2005a) showed that a specific case of their proposed asymmetric SV model with normal error distribution leads to time varying correlation and we generalise this result to the case of the GT distribution. This time varying feature provides a more flexible way to model the dynamic of asymmetric effect compared to the constant correlation between the return and volatility processes.

In view of another prominent characteristic, we extend the SV model to take into account the skewness observed in financial asset returns by replacing the normality assumption of the returns with the asymmetric Laplace (AL) distribution. In addition, past observations and exogenous variables are incorporated into the model to examine possible effects on the returns. When the AL distribution is used as the error distribution for the returns, the model can be interpreted from a different perspective: a quantile regression model where the heteroscedasticity is controlled by a SV model for the volatility. This study is a stochastic extension of Chen *et al.* (2009) who assumed a GARCH specification to account for heteroscedasticity in a quantile regression problem.

Furthermore, we consider a multivariate extension of the SV model to explain the cross correlation between multivariate time series. In response to the lack of literature in the modelling of electricity price and price volatility of the Australian wholesale spot electricity market using SV models, we explore the applicability of multivariate SV models in this field. Different specifications regarding both the correlation structure and distributional assumption of the multivariate SV models are adopted and compared using empirical data.

For statistical inference, these extended models are implemented using Bayesian Markov chain Monte Carlo method with Gibbs sampler algorithm. A series of simulation experiments and empirical studies are conducted to demonstrate the properties and applicability of the proposed models. Extended models are shown to provide satisfactory performance in both simulation studies and real data analyses, and they have made substantial contributions to enrich the SV model literature.

Declaration

I declare that this thesis represents my own work, except where due acknowledgement is made, and that it has not been previously included in a thesis, dissertation or report submitted to this University or to any other institution for a degree, diploma or other qualifications.

Signed _____

Joanna Jia Jia Wang

Acknowledgements

First and foremost, I would like to sincerely thank my primary supervisor, Dr. Jennifer S.K. Chan for being such a wonderful supervisor. I owe my deepest gratitude to her, for providing enormous support, guidance and encouragement throughout my entire PhD studies. Dr. Jennifer Chan has made my research experience extremely rewarding and has inspired me to pursue an academic career.

I would also like to express gratitude to my associate supervisor, Dr. Boris Choy, who has provided much assistance and encouragement. I wish to thank him for introducing me to financial and econometrical concepts. His enthusiasm in research and his innovative ideas always motivate me.

It is a great pleasure for me to thank all staff of the School of Mathematics and Statistics for their teaching and assistance, especially Professor Neville Weber, who simulated my interest in statistics in the first place. Special thanks to Professor Eugene Seneta, Dr. John Ormerod and Associate Professor Shelton Peiris, who often engage me in insightful conversation and help me to broaden my perspective on future career development. I thank my fellow postgraduates and ex-students of the department: Anna, Connie, Emi, Erwin, Jenny, Justin, Thomas, Xiang and Yinny, for their friendship and support.

Lastly, I am indebted to my family for their continual care and understanding, especially during the final stage of writing-up this thesis. I wish to thank them for their constant support and encouragement which have been invaluable in helping me to complete this thesis.

CONTENTS

List of Figures

List of Tables

CHAPTER 1

Introduction

1.1. Background

Volatility is an important phenomenon in financial markets and it is simply a measure for variability of the price of a financial instrument over time. It is used to quantify the risk of a financial instrument over the specified time period. A thorough understanding of volatility enables realistic pricing of options, accurate risk assessment, precise hedging of contracts and efficient asset allocation in a portfolio. In their famous paper, Black and Scholes (1973) derived a pricing formula for a standard European call option based on the assumption that stock returns are normal and volatility is constant. However, there is an increasing amount of evidence that this assumption is not consistent with empirical data. Even Black and Scholes have themselves, commented: '...there is evidence of non-stationarity in the variance. More work must be done to predict variances using the information available.' (Black and Scholes, 1972). A direct consequence of the violation of this assumption is biases in pricing options on assets. Therefore, building econometric models to account for time varying volatility has been the subject of theoretical and empirical investigation by both researchers and practitioners since late last century. Before we introduce various volatility models, we will discuss some stylised facts about the volatility.

A number of stylised facts about the volatility of financial asset prices have emerged over the years and have been confirmed by empirical studies. They include:

- *Volatility clustering*: Mandelbrot (1963) and Fama (1965) both noted that large changes in the price of an asset tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes. This clustering of large moves or small moves was one of the first documented features of the volatility process. A quantitative manifestation of volatility clustering is that while returns themselves are not autocorrelated, absolute or squared returns exhibit significant, positive and slowly decaying autocorrelations or persistence. Volatility persistence

implies that the shocks of the volatility today will exert persistent influence on the expectation of volatility over a long period of time in the future.

- *Mean-reverting*: Volatility mean reversion refers to the fact that there is a long-run level to which volatility will eventually return. As can be inferred from volatility clustering, a period of high volatility will eventually lead to lower volatility and similarly, a period of low volatility will be followed by a rise in volatility.
- *Leverage effect*: Black (1976), Christie (1982), Nelson (1991) and Engle and Ng (1993) found evidence that volatility is negatively correlated with equity returns. The usual claim is that in response to bad news, the price of a stock decreases, causing an increase in the debt-to-equity ratio which makes the firm riskier and hence increases future expected volatility.

A widely used class of models for the conditional volatility is the autoregressive conditional heteroscedastic (ARCH) models introduced by Engle (1982) and the generalised ARCH (GARCH) by Bollerslev (1986). The primary interest of ARCH/GARCH models is in modelling the changes in volatility. These models are able to characterise the stylised features of the volatility mentioned above. Since the introduction of ARCH models in 1982, there has been an enormous body of research extending the original ARCH/GARCH model specification designed to improve upon the model's ability to capture and reflect empirically observed features of real life data. These include, nonlinear GARCH introduced by Engle and Ng (1993); exponential GARCH by Nelson (1991); Glosten-Jagannathan-Runkle GARCH introduced by Glosten *et al.* (1993); threshold GARCH by Zakoian (1994), among many others. Bollerslev *et al.* (1994) provided a thorough review on most of the earlier literature on GARCH models. The distinguishing and essential feature of ARCH type models is that the conditional variance of the returns is explicitly specified as a deterministic function of past conditional variances and past values of the return itself. This feature is in clear contrast with the stochastic volatility (SV) model, which is the fundamental theme of this thesis.

The SV models have gradually emerged as a successful alternative to the class of ARCH models in accounting for the time-varying and persistent volatility of financial returns. They are motivated by the mixture-of-distributions hypothesis postulated by Clark (1973). In Clark's approach, asset returns follow a mixture of normal distributions with a mixing process depending on the unobservable flow of price-relevant information. The work of Clark

was then followed by influential articles by Tauchen and Pitts (1983) and Gallant *et al.* (1991) who noted that if the unobserved information flows are positively autocorrelated, then the resulting return process with time-varying and autocorrelated conditional variance reveals volatility clustering, which is a typical feature of financial return series. The mixture-of-distributions hypothesis leads to the idea that asset return volatility follows its own stochastic process with unobservable innovations. This is in contrast with the ARCH models, where the conditional variance given the available information is a deterministic function of past observations and innovations. In the field of mathematical finance and financial economics, SV models are typically treated in a continuous-time framework which is employed extensively for derivative pricing and portfolio optimisation. Hull and White (1987) generalised the Black and Scholes (1973) option pricing formula based on SV and they allowed the volatility process to follow a diffusion process. However, in empirical applications, discrete-time SV models are equally important as real data is observed at discrete points in time. In the framework of discrete-time, Taylor (1982, 1986) first published a paper on the direct volatility clustering SV model, in which the SV model can be thought of as an Euler discretisation of the diffusion process.

Despite the theoretical appeal, SV models are less popular in financial practice as compared to the ARCH type models. This is due to the ease in the evaluation of the ARCH likelihood function on one hand and the high level of difficulty involved in the efficient estimation of the SV models on the other. There are studies that compare the relative performance of the SV and GARCH models. Geweke (1994a) compared a GARCH (1,1) model against a basic SV model using the US/Canadian exchange rate and he concluded that the SV model accommodates volatility changes better than a GARCH model, especially after large shocks. Kim *et al.* (1998) also showed the superior performance of the SV models in terms of in-sample fitting. On the other hand, Gerlach and Tuyl (2006) compared Markov switching GARCH and SV models and they found that the simple GARCH model with Student-t innovations performs the best in terms of marginal likelihood. It should be pointed out that the aim of this thesis is not to investigate whether GARCH or SV models provide a better approximation to the data-generating process for financial data. Our focus is entirely on SV models, and we contribute to the SV model literature by developing novel modelling strategies and broadening the area in which SV models can be applied.

The standard discrete-time SV model as introduced by Taylor (1982) is given by

$$\begin{aligned} y_t &= \exp(h_t/2)\epsilon_t, \\ h_{t+1} &= \mu + \phi(h_t - \mu) + \tau\eta_t, \quad \eta_t \sim \text{ iid } N(0,1), \end{aligned} \tag{1.1}$$

where y_t is the asset return on day t, h_t is the log-volatility which is assumed to follow a Gaussian AR(1) process with persistence parameter ϕ and τ is the standard deviation of volatility innovations. We assume $|\phi| < 1$ for stationary SV model and the error processes ϵ_t and η_t are mutually independent. The error of the return is typically simplified to an independent and identically distributed (iid) process with $E(\epsilon_t) = 0, \text{var}(\epsilon_t) = 1$. In the case where ϵ_t is assumed to be Gaussian, we often refer to (1.1) as the log-normal SV model. The unconditional distribution of h_t is given by

$$h_t \sim N\left(\mu, \frac{\tau^2}{1-\phi^2}\right).$$

Assuming $|\phi| < 1$ and $E(y_t^4) < \infty$, the unconditional moments of the returns $E(y_t^2)$ and $E(y_t^4)$ are given by

$$\begin{aligned} E(y_t^2) &= E(\exp(h_t)\epsilon_t^2) \\ &= E(\exp(h_t))E(\epsilon_t^2) \\ &= E(\exp(h_t)) \\ &= \exp\left(\mu + \frac{\tau^2}{2(1-\phi^2)}\right) \\ E(y_t^4) &= E(\exp(2h_t))E(\epsilon_t^4) \\ &= \exp\left(2\mu + \frac{2\tau^2}{1-\phi^2}\right) E(\epsilon_t^4). \end{aligned}$$

Hence the kurtosis of the unconditional distribution of the returns is

$$\begin{aligned} \kappa &= \frac{E(y_t^4)}{[E(y_t^2)]^2} \\ &= \exp\left(\frac{\tau^2}{1-\phi^2}\right) E(\epsilon_t^4), \end{aligned}$$

which consists of the kurtosis of the standardised errors $E(\epsilon_t^4)$ and the variation in the volatility process. When the normal distribution is assumed for the conditional distribution of the

returns, $E(\epsilon_t^4) = 3$ and the kurtosis of the unconditional returns is $3\exp\left(\frac{\tau^2}{1-\phi^2}\right) > 3$. Therefore, the unconditional distribution of returns in the SV model (1.1) is leptokurtic even when the conditional distribution is normal, as long as the volatility varies with time.

Although the standard SV model is able to capture volatility clustering exhibited in financial time series, Liesenfeld and Jung (2000) showed that the kurtosis implied by the model is often too small to match the sample kurtosis observed. Hence, the leptokurtosis of returns cannot be fully explained by changes in volatility. Many studies considered a heavy-tailed distribution for y_t and demonstrated a significant improvement of the SV model with heavy-tailed distribution over the normal distribution. For instance, Liesenfeld and Jung (2000) employed a Student-t and generalised error distribution (GED) and they found that the fit of the SV model with a Student-t distribution is generally better than a GED. They also demonstrated that the choice of the conditional distribution for the returns has a significant and systematic effect on the parameter estimates of the volatility process. For example, they found that the estimates of the volatility persistence parameter are systematically higher and the estimates of the variance of volatility errors are systematically lower under the SV model with Student-t distribution than the SV model with normal distribution, respectively. Asai (2009) compared a SV model with a mixture-of-normal distribution with Student-t and GED and showed that the mixture-of-normal distribution gives a better fit to the Yen/Dollar exchange rate than other distributions based on Bayes factor. It should be pointed out that another name for the GED is the exponential power (EP) distribution and both names are used in the literature. Other studies that use Student-t distribution include Harvey *et al.* (1994), Meyer and Yu (2000), Chib *et al.* (2002), Jacquier *et al.* (2004), Omori (2007), Choy *et al.* (2008), Ishihara and Omori (2011), among many others. Steel (1998) assumed a skewed EP distribution (introduced in Fernández *et al.* (1995)) for $\log(\epsilon_t^2)$ while Cappuccio *et al.* (2004) assumed an alternative skew EP distribution with density obtained via Azzalini's method (Azzalini, 1985) for the return innovations. Lastly, Nakajima and Omori (2011) used a generalised hyperbolic skew Student-t error distribution in order to account for asymmetric heavy-tailness of the returns simultaneously.

When employing heavy-tailed distributions in SV modelling, an increasing popular approach is to use the scale mixture form of distributions. Choy and Smith (1997) utilised scale mixture of normal (SMN) distributions in a Bayesian random effects model and showed that

the SMN distribution can protect statistical inference from the distorting effects of outliers and hence, robustifies statistical analyses. In the context of SV models, Choy and Chan (2000) studied the SV model with an EP distribution for both the returns and volatilities via a scale mixture of uniform (SMU) distributions; Wang *et al.* (2011a) and Abanto-Valle *et al.* (2010) studied the SV model by expressing the Student-t density function into a SMN representation while Choy *et al.* (2008) used a SMU form for the Student-t density function. Nakajima and Omori (2011) expressed the generalised hyperbolic skew Student-t distribution as a normal variance-mean mixture of the generalised inverse Gaussian distribution. A major advantage of using such mixture distributions is that the conditional posterior distributions of model parameters can sometimes be reduced to tractable forms, conditional on the mixing parameters. This is in contrast to the case when the direct likelihood function of the heavy-tailed distribution is used in the model. Moreover, Choy and Smith (1997) showed that potential outliers can be identified using these mixing parameters and this is also explored in Choy and Chan (2008).

In addition to adopting a more flexible error distribution for the asset returns, the basic SV model specification can also be extended to enhance the flexibility of the model to capture important regularities observed in real financial time series. For example, Chib (2002) extended the basic SV model to include covariates; Harvey *et al.* (1994) and Shephard (1996) introduced two independent autoregressive processes in the observation equation; Chib (2002), Berg *et al.* (2004) and Nakajima and Omori (2009) studied the SV models with jumps. The use of exogenous thresholds for regime switching have been shown to be important in volatility modelling, particularly in GARCH models. For example, Li and Li (1996) attempted to model the asymmetries in both returns and volatility by using a double threshold ARCH model. Chen *et al.* (2003) modelled a double threshold GARCH model with US market returns as the threshold variable while Gerlach et al (2006) employed a double threshold GARCH model with trading volume as the threshold variable and observed higher volatility following increases in trading volume. So *et al.* (2002) proposed a threshold SV model, which is analogous to the threshold ARCH model of Li and Li (1996). Chen *et al.* (2008) generalised the model by So *et al.* (2002) by including exogenous variables in the mean equation and using a non-zero threshold variable. In the situation where the volatility responds asymmetrically towards positive and negative returns, a SV model with leverage

(SVL) is used where the leverage effect is measured by the correlation between the innovations of returns and volatilities. Meyer and Yu (2000) and Omori *et al.* (2007) investigated the SVL model by modelling the pair (ϵ_t, η_t) with a bivariate normal distribution while Choy *et al.* (2008) assumed a bivariate Student-t distribution. In comparison, Asai (2008) modelled ϵ_t and η_t using the Student-t and normal distributions respectively. Asai and McAleer (2005a) proposed alternative asymmetric SV model which nests the SVL model. The univariate SV model is also generalised to the multivariate SV models, although the literature on multivariate SV is relatively limited. See for example, Harvey *et al.* (1994), Yu and Meyer (2006), Chib *et al.* (2006) and Asai *et al.* (2006). In the next section, we will give an overview of the different inference methods that have been proposed in the literature for the estimation of SV models.

1.2. Estimation techniques for SV models

Despite the theoretical appeal, the empirical application of SV models has been limited due to the difficulties involved in the evaluation of the likelihood function. In order to derive the exact likelihood function, the vector of latent volatilities needs to be integrated out of the joint probability distribution. Consider the standard SV model in (1.1) and denote the vector of observed data by $\boldsymbol{y} = (y_1, \ldots, y_n)$ and the vector of log-volatilities by $\boldsymbol{h} = (h_1, \ldots, h_n)$. Then the likelihood is given by

$$f(\boldsymbol{y}|\mu, \phi, \tau^2) = \int f(\boldsymbol{y}|\boldsymbol{h}, \mu, \phi, \tau^2) f(\boldsymbol{h}|\mu, \phi, \tau^2) d\boldsymbol{h}, \tag{1.2}$$

which has dimension equal to the sample size. Hence, classical maximum likelihood estimation of the parameters of SV models is difficult as numerical methods must be used to evaluate the likelihood (1.2) which involves high dimensional integral.

Important issues that need to be considered when choosing a particular estimation method include statistical efficiency, computational efficiency, estimation of the latent volatilities and to the best of our interest, the applicability for flexible modelling. The simplest estimator for the SV model is the method of moments (MM), which was extended by Melino and Turnbull (1990) to the generalised method of moments (GMM). These methods are based on the convergence of the sample moments to their expected values. However, it is well known that the MM estimators have poor finite sample properties and may be inefficient relative

to a likelihood-based method of inference. Despite the limitations, a number of studies, for example, Andersen (1994) and Fleming *et al.* (1998) employed the GMM estimator due to its simplicity.

Harvey *et al.* (1994) proposed a quasi-maximum-likelihood estimator (QML) based on linearising the SV model by taking the logarithms of the squares of the observations y_t in (1.1). Approximating $\log(\epsilon_t^2)$ with a Gaussian distribution, the Kalman filter can be applied to obtain the QML estimates. The QML method is computationally convenient and very flexible, however, it is shown through the Monte Carlo experiments that the QML method is less efficient than the Bayesian MCMC technique and other likelihood based methods because the QML method does not rely on the exact likelihood of $\log(\epsilon_t^2)$. For further details, see Jacquier *et al.* (1994).

Much of the research in the SV literature in recent years is on the development and the use of Markov chain Monte Carlo (MCMC) methods. The first MCMC estimation procedure in the context of basic univariate SV model was proposed by Jacquier *et al.* (1994), where their method is commonly known as JPR. The JPR method exploits the latent volatility structure and treats it as a vector of unknown parameters to be estimated. Hence they augment the model parameters with time series of volatilities and draw directly from the joint posterior distribution. More specifically, instead of focusing on the usual posterior density $f(\mu, \phi, \tau^2|\boldsymbol{y})$, their method focuses on the density $f(\boldsymbol{h}, \mu, \phi, \tau^2|\boldsymbol{y})$. Their algorithm combines the data augmentation idea of Tanner and Wong (1987) and the MCMC procedure to produce an algorithm which alternates back and forth between sampling from $f(\boldsymbol{h}|\mu, \phi, \tau^2, \boldsymbol{y})$ and $f(\mu, \phi, \tau^2|\boldsymbol{h}, \boldsymbol{y})$. A Metropolis accept/reject independence chain is also used in the algorithm for the sampling of the latent volatilities. The advantage of the JPR procedure is that it allows inference on the model parameters as well as obtaining smoothed estimates of the unobserved volatilities simultaneously. A technical advantage of the Bayesian MCMC method over classical inferential techniques is that MCMC does not require numerical optimisation in general. This advantage becomes particularly important when the number of parameters that need to be estimated in the model is large. For example, when a multivariate SV model is used in empirical applications. Andersen *et al.* (1999) compared various estimation methods for the SV model and found that MCMC is the most efficient estimation tool. Jacquier *et al.* (1994) also showed that MCMC method outperforms both

QML and GMM through an extensive Monte Carlo study. However, the JPR approach was criticised for being computationally inefficient for the sampling of the volatility by Shephard and Kim (1994) as the JPR method proposes to update the latent volatility, one at a time. This single-move sampler typically requires a large number of iterations before the Markov chain converges. Shephard and Pitt (1997) later proposed a more efficient multi-move algorithm which samples a block of volatility states at a time and demonstrated that the multi-move sampler outperforms the single-move method. Kim *et al.* (1998) developed a mixture sampler MCMC algorithm for the SV model and it was used and extended by Chib *et al.* (2002), Omori *et al.* (2007) and Nakajima and Omori (2009) for the estimation of various extensions of the basic SV model. Wong (2002) however, suggested to use the single-move JPR algorithm due to its simpler implementation procedure in practice.

Another estimation procedure that is used widely in the SV literature is the simulated maximum likelihood (SML) estimator based on importance sampling. The first SML method is the accelerated Gaussian importance sampling approach, developed by Danielsson and Richard (1993). This method has limited application to more flexible SV models as it only applies to models with a latent Gaussian process. The second method is the efficient importance sampling procedure, as proposed by Liesenfeld and Richard (2003, 2006). The last SML method is known as the Monte Carlo likelihood method which decompose the likelihood into a Gaussian part and a remainder function. Several studies adopted this method to estimate various SV models, for example, Sandmann and Koopman (1998), Jungbacker and Koopman (2006) and Asai and McAleer (2005b). Danielsson (1994) compared the MM, QML, JPR and SML methods and concludes that SML nearly outperforms the MM and QML estimators and has similar performance to JPR. However, the SML procedure has several drawbacks as mentioned in Broto and Ruiz (2004) which includes, for example, difficulty in measuring the accuracy of the proposed approximation.

Other estimation methods such as the indirect inference method (Gourieroux *et al.*, 1993) and the efficient method of moments (Gallant and Tauchen, 1996) have been proposed and used in practice. For a detailed survey regarding these estimation methods and their main advantages and limitations for univariate SV models, see Broto and Ruiz (2004).

Taking into account the statistical efficiency, convenience of implementation, popularity in the literature and the flexibility of modelling various extensions of the SV model, we will

use the Bayesian MCMC method for our estimation of various univariate and multivariate SV models. Next, we will give a brief overview of Bayesian inference methodology and the MCMC algorithm for the standard SV model in (1.1).

1.3. Bayesian inference and MCMC

Bayesian statistics is concerned with generating the posterior distribution of the unknown parameters incorporating both the observed data information and some prior belief through prior distributions for these parameters. The Bayes' theorem sets the foundation of Bayesian statistics, which states that the posterior distribution of the parameter $\boldsymbol{\theta}$ given the data $\boldsymbol{y}$ is proportional to the product of the data likelihood $f(\boldsymbol{y}|\boldsymbol{\theta})$ and the prior density $f(\boldsymbol{\theta})$, that is,

$$f(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{f(\boldsymbol{y}|\boldsymbol{\theta})f(\boldsymbol{\theta})}{\int f(\boldsymbol{y}|\boldsymbol{\theta})f(\boldsymbol{\theta})d\boldsymbol{\theta}} \propto f(\boldsymbol{y}|\boldsymbol{\theta})f(\boldsymbol{\theta}).$$

In the absence of specific prior information on $\boldsymbol{\theta}$, non-informative priors are adopted. For a large sample size, Bayes estimators are asymptotically equivalent to maximum likelihood estimators under appropriate regularity conditions. (See Gelman *et al.* (2004) for further details.)

A major limitation towards the implementation of Bayesian approaches in early years is that the posterior distributions are analytically tractable for only a small number of simple models. In most cases, the posterior distribution does not have analytic expression and requires the evaluation of high-dimensional integrals. Simulation-based MCMC methods became widely known among statisticians since the early 1990 and it has now become the routine tool for Bayesian computation for a wide range of complicated statistical models. The principle of MCMC is the sampling from one or more Markov chains after convergence to a stationary distribution that is exactly the posterior distribution. More specifically, the MCMC algorithm consists of constructing an irreducible and aperiodic Markov chain, whose equilibrium distribution is the desired posterior distribution $f(\boldsymbol{\theta}|\boldsymbol{y})$. After selecting an initial value $\boldsymbol{\theta}^0$, we generate dependent samples $\boldsymbol{\theta}^1, \ldots, \boldsymbol{\theta}^t, \ldots$ until the equilibrium is reached. Outputs from the simulated chain are used for posterior analysis such as obtaining summaries of the posterior distribution. A variety of MCMC methods have been proposed to sample from posterior densities and here we summarise two most commonly used MCMC techniques.

1.3.1. Metropolis-Hastings Algorithm. The Metropolis-Hastings (MH) algorithm is the baseline for MCMC method and it was first developed by Metropolis *et al.* (1953) and later generalised by Hastings (1970). Suppose we wish to generate a sample from a target distribution, which can be taken as the posterior distribution $f(\boldsymbol{\theta}|\boldsymbol{y})$ in Bayesian framework. The MH algorithm can be described by the following iterative steps:

(1) Set any initial value $\boldsymbol{\theta}^{(0)}$.

(2) Given the current value $\boldsymbol{\theta} = \boldsymbol{\theta}^{(k)}$, sample a candidate parameter value $\boldsymbol{\theta}^{'}$ from a proposal distribution $q(\boldsymbol{\theta}^{'}|\boldsymbol{\theta})$.

(3) Accept $\boldsymbol{\theta}^{'}$ as the new state $\boldsymbol{\theta}^{(k+1)}$ with probability α, where

$$\alpha = \min\left(1, \frac{f(\boldsymbol{\theta}^{'}|\boldsymbol{y})q(\boldsymbol{\theta}|\boldsymbol{\theta}^{'})}{f(\boldsymbol{\theta}|\boldsymbol{y})q(\boldsymbol{\theta}^{'}|\boldsymbol{\theta})}\right).$$

With probability $1-\alpha$, we set $\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}$.

In particular, when the proposal distribution is symmetric, that is, $q(\boldsymbol{\theta}|\boldsymbol{\theta}^{'}) = q(\boldsymbol{\theta}^{'}|\boldsymbol{\theta})$ and so $\alpha = \min(1, f(\boldsymbol{\theta}^{'}|\boldsymbol{y})/f(\boldsymbol{\theta}|\boldsymbol{y}))$, we reduce the MH algorithm to the Metropolis algorithm (Metropolis *et al.*, 1953). Depending on the form and the complexity of the sampling problem, different choices of proposal distribution (which lead to different schemes) can be made. The two general approaches for the practical construction of a MH proposal are random walk and independent chain sampling. A discussion of the choices of proposal distribution is given by Tierney (1994).

1.3.2. Gibbs Sampler. The Gibbs sampler was introduced by Geman and Geman (1984) in the context of Bayesian image processing. Gelfand and Smith (1990) and Gelfand *et al.* (1990) made the wider Bayesian community aware of the Gibbs sampler and highlighted the wide applicability of the method to a range of statistical models. Gibbs sampler is a special case of the MH algorithm where the acceptance probability $\alpha = 1$ and hence the proposed value is accepted in all iterations. It is a useful and convenient method particularly when the posterior distribution is of high dimension because it allows random values to be drawn from univariate full conditional posterior distributions. Thus, rather than generating a single d-dimensional vector using the joint distribution, we simulate d random variables sequentially from the d univariate full conditional distributions.

Suppose that $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$. Given a particular state of the chain $\boldsymbol{\theta}^{(k)} = (\theta_1^{(k)}, \ldots, \theta_d^{(k)})$, the Gibbs sampler generates the new parameter values by successively sampling

$$\begin{array}{lll}
\theta_1^{(k+1)} & \text{from} & f(\theta_1|\theta_2^{(k)}, \theta_3^{(k)}, \ldots, \theta_d^{(k)}, \boldsymbol{y}), \\
\theta_2^{(k+1)} & \text{from} & f(\theta_2|\theta_1^{(k+1)}, \theta_3^{(k)}, \ldots, \theta_d^{(k)}, \boldsymbol{y}), \\
\vdots & \vdots & \vdots \\
\theta_j^{(k+1)} & \text{from} & f(\theta_j|\theta_1^{(k+1)}, \theta_2^{(k+1)}, \ldots, \theta_{j-1}^{(k+1)}, \theta_{j+1}^{(k)}, \ldots, \theta_d^{(k)}, \boldsymbol{y}), \\
\vdots & \vdots & \vdots \\
\theta_d^{(k+1)} & \text{from} & f(\theta_d|\theta_1^{(k+1)}, \theta_2^{(k+1)}, \ldots, \theta_{d-1}^{(k+1)}, \boldsymbol{y}),
\end{array}$$

The distribution $f(\theta_j|\theta_1, \theta_2, \ldots, \theta_{j-1}, \theta_{j+1}, \ldots, \theta_d, \boldsymbol{y})$ is known as the full conditional distribution of θ_j. The full conditional densities are obtained from the joint density $f(\boldsymbol{\theta}, \boldsymbol{y})$. In some cases, they can be reduce to standard densities from which sampling is straightforward. However, for the SV models we are dealing with in this thesis, these full conditional distributions are often non-standard, and random variate generation techniques such as the MH method, adaptive rejection sampling technique and slice sampling may be required. We will show in later chapters, the use of scale mixture representations of some distributions such as the Student-t distribution, can effectively simplify the full conditional distributions to mostly standard forms. More detailed description of the Gibbs sampler can be found in Smith and Roberts (1993).

1.3.3. Estimation of the Standard SV Model using the MCMC Algorithm. In this section, we illustrate the estimation of the standard SV model defined in (1.1) using the MCMC algorithm. Bayesian inference for the model parameters (μ, ϕ, τ^2) together with the volatility states $\boldsymbol{h} = (h_1, \ldots, h_n)$ in the SV model is based on the joint posterior distribution $f(\mu, \phi, \tau^2, \boldsymbol{h}|\boldsymbol{y})$, which is proportional to the product of the likelihood function $f(\boldsymbol{y}|\mu, \phi, \tau^2, \boldsymbol{h})$, the conditional distribution of the volatilities $f(\boldsymbol{h}|\mu, \phi, \tau^2)$ and the joint prior $f(\mu, \phi, \tau^2, \boldsymbol{h})$. The joint prior distribution of all unobservables, including the model parameters and latent volatilities, is given by

$$f(\mu, \phi, \tau^2, \boldsymbol{h}) = f(\mu)f(\phi)f(\tau^2)f(h_1|\mu, \tau^2)\prod_{t=1}^{n-1} f(h_{t+1}|h_t, \mu, \phi, \tau^2)$$

by assuming prior independence of the model parameters and successive conditioning of the latent volatilities. The likelihood of observed data given the latent volatilities is specified by:

$$f(\boldsymbol{y}|\mu,\phi,\tau^2,\boldsymbol{h}) = \prod_{t=1}^{n} f(y_t|h_t).$$

The Bayesian estimation procedure requires the specification of the prior distribution for each model parameter. We assign the following priors to the parameters:

$$\begin{aligned} f(\mu) &= N(a_\mu, b_\mu), \\ f(\phi^*) &= Be(a_\phi, b_\phi), \\ f(\tau^2) &= IG(a_\tau, b_\tau), \end{aligned}$$

where $\phi^* = \frac{\phi+1}{2}$, $Be(a,b)$ is the beta distribution with density

$$f(x|a,b) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1},$$

where $B(\cdot,\cdot)$ is the beta function and $IG(\alpha,\beta)$ is the inverse gamma distribution with shape and scale parameters α and β respectively and has a pdf given by

$$IG(\lambda|a,b) = \frac{b^a}{\Gamma(a)} \lambda^{-(a+1)} e^{-b/\lambda}, \quad \lambda, a, b > 0.$$

The Gibbs sampler requires successive sampling from the univariate full conditional posterior distributions for the model parameters and they are derived as follows

$$\begin{aligned} f(\mu|\boldsymbol{y},\boldsymbol{h},\phi,\tau^2) &\propto f(\boldsymbol{h}|\mu,\phi,\tau^2) f(\mu), \\ f(\phi|\boldsymbol{y},\boldsymbol{h},\mu,\tau^2) &\propto f(\boldsymbol{h}|\mu,\phi,\tau^2) f(\phi), \\ f(\tau^2|\boldsymbol{y},\boldsymbol{h},\mu,\phi) &\propto f(\boldsymbol{h}|\mu,\phi,\tau^2) f(\tau^2). \end{aligned}$$

The likelihood function $f(\boldsymbol{y}|\boldsymbol{h},\mu,\phi,\tau^2)$ is omitted in these full conditional distributions since the volatility $\boldsymbol{h}$ contains all information about the states parameters and the likelihood function can be taken as a constant with respect to these parameters.

By successive conditioning, the conditional distribution of the log-volatilities is given by

$$f(\boldsymbol{h}|\mu,\phi,\tau^2) = f(h_1|\mu,\phi,\tau^2)\prod_{t=1}^{n-1} f(h_{t+1}|h_t,\mu,\phi,\tau^2). \tag{1.3}$$

We assign a normal $N(a_\mu, b_\mu)$ prior to μ and an inverse gamma $IG(a_\tau, b_\tau)$ prior to τ^2 since they are both conjugate priors. For ϕ, we assign a Beta(a_ϕ, b_ϕ) prior to $\phi^* = (\phi+1)/2$. In our work, we select $a_\phi = 20$ and $b_\phi = 1.5$, implying a prior mean of 0.86. As discussed in Kim *et al.* (2008), a flat prior is also attractive because it leads to an analytically tractable full conditional density. However, this prior can cause problems when the data are close to being non-stationary (Phillips, 1991). Furthermore, if $\phi = 1$, then μ becomes unidentified from the data. The Beta prior that has support on the interval $(-1, 1)$ avoids this problem as well.

Combining (1.3) with each of the prior distribution and eliminating the constant terms, we obtain the following full conditional distributions for the volatility parameters.

- Full conditional distribution for μ

$$\mu|\boldsymbol{h},\phi,\tau^2,\boldsymbol{y} \sim N\left(\sigma_\mu^2\left(\frac{(1-\phi^2)h_1 + (1-\phi)\sum_{t=1}^{n-1}(h_{t+1}-\phi h_t)}{\tau^2} + \frac{a_\mu}{b_\mu}\right), \sigma_\mu^2\right)$$

  where

$$\sigma_\mu^2 = \left(\frac{1-\phi^2+(n-1)(1-\phi)^2}{\tau^2} + \frac{1}{b_\mu}\right)^{-1}.$$

- Full conditional distribution for τ^2

$$\tau^2|\boldsymbol{h},\mu,\phi,\boldsymbol{y} \sim IG\left(a_\tau + \frac{n}{2}, b_\tau + \frac{(1-\phi^2)(h_1-\mu)^2 + \sum_{t=1}^{n-1}\left(h_{t+1}-\mu-\phi(h_t-\mu)\right)^2}{2}\right)$$

- Full conditional distribution for ϕ

$$\phi|\boldsymbol{h},\mu,\tau^2,\boldsymbol{y} \propto N\left(h_1|\mu, \frac{\tau^2}{1-\phi^2}\right)\prod_{t=1}^{n-1} N\left(h_{t+1}|\mu+\phi(h_t-\mu),\tau^2\right)\left(\frac{\phi+1}{2}\right)^{a_\phi-1}\left(\frac{1-\phi}{2}\right)^{b_\phi-1}$$

It is straightforward to sample from the full conditional posteriors of μ and τ^2 since they are standard distribution and the method of Chib and Greenberg (1994) (based on the MH

algorithm) can be used to sample from the full conditional distribution of ϕ. Note that

$$\log f(\boldsymbol{h}|\mu,\phi,\tau^2) \propto -\frac{(1-\phi^2)(h_1-\mu)^2}{2\tau^2} + \frac{1}{2}\log(1-\phi^2) + \frac{\sum_{t=1}^{n-1}[(h_{t+1}-\mu)-\phi(h_t-\mu)]^2}{2\tau^2}.$$

The method of Chib and Greenberg (1994) involves using a normal proposal distribution $N(\hat{\phi}, V_\phi)$, where

$$\begin{aligned}\hat{\phi} &= \frac{\sum_{t=1}^{n-1}(h_{t+1}-\mu)(h_t-\mu)}{\sum_{t=1}^{n-1}(h_t-\mu)^2},\\ V_\phi &= \frac{\tau^2}{\sum_{t=1}^{n-1}(h_t-\mu)^2}.\end{aligned}$$

A proposal value ψ^* is draw from $N(\hat{\phi}, V_\phi)$, given the current value $\phi^{(i-1)}$ at the $(i-1)$-st iteration. Then, provided ϕ^* is in the stationary region, accept ϕ^* as $\phi^{(i)}$ with probably $\exp(g(\phi^*) - g(\phi^{(i-1)}))$ where

$$g(\phi) = \log f(\phi) - \frac{(1-\phi^2)(h_1-\mu)}{2\tau^2} + \frac{1}{2}\log(1-\phi^2).$$

If the proposal is rejected, then set $\phi^{(i)}$ to equal $\phi^{(i-1)}$.

The most difficult part in the estimation of SV models is to sample effectively from the full conditional posterior of the latent volatility h_t. By Bayes' theorem, full conditional distribution for h_t is

$$\begin{aligned}f(h_t|\boldsymbol{y}, \boldsymbol{h}_{\setminus t}, \mu, \phi, \tau^2) &\propto f(y_t|h_t) f(h_t|\boldsymbol{h}_{\setminus t}, \mu, \phi, \tau^2),\\ &= \frac{1}{\sqrt{2\pi}\exp(h_t/2)} \exp\left(-\frac{y_t^2}{2\exp(h_t)}\right) f(h_t|\boldsymbol{h}_{\setminus t}, \mu, \phi, \tau^2),\end{aligned}$$

where $\boldsymbol{h}_{\setminus t} = (h_1, \ldots, h_{t-1}, h_{t+1}, \ldots, h_n)$.

Exploiting the Markovian structure of the SV model, it can be shown

$$\begin{aligned}f(h_t|\boldsymbol{h}_{\setminus t}, \mu, \phi, \tau^2) &= f(h_t|h_{t-1}, h_{t+1}, \mu, \phi, \tau^2)\\ &= f(h_t|h_{t-1}, \mu, \phi, \tau^2) f(h_{t+1}|h_t, \mu, \phi, \tau^2)\\ &= N(\mu_{h_t}, \sigma_{h_t}^2)\end{aligned}$$

where

$$\mu_{h_t} = \begin{cases} \mu(1-\phi)+\phi(h_2-\mu), & t=1, \\ \mu+\frac{\phi[(h_{t-1}-\mu)+(h_{t+1}-\mu)]}{1+\phi^2}, & t=2,\ldots,n-1, \\ \mu+\phi(h_{n-1}-\mu), & t=n; \end{cases}$$

$$\sigma^2_{h_t} = \begin{cases} \tau^2, & t=1,n, \\ \frac{\tau^2}{1+\phi^2}, & t=2,\ldots,n-1. \end{cases}$$

Kim *et al.* (1998) proposed a method for sampling from $f(h_t|\boldsymbol{y}, \boldsymbol{h}_{\setminus t}, \mu, \phi, \tau^2)$ by exploiting the fact that $\exp(-h_t)$ is a convex function and can be bound by a linear function in h_t. An acceptance-rejection procedure (Ripley, 1987) can be then implemented to sample h_t. Other MCMC methods have been proposed, for example, by Jacquier *et al.* (1994), Geweke (1994b) and Shephard and Kim (1994). However, note that these methods use a single-move sampler for the volatility.

To summarise, a single-move Gibbs sampler for the standard SV model as suggested by Jacquier *et al.* (1994) is given as follows:

(1) Initialise $\boldsymbol{h}^{(0)}$, $\mu^{(0)}$, $\phi^{(0)}$ and $\tau^{2(0)}$.

(2) For $k = 1, ..., T$:

(a) For $t = 1, \ldots, n$,
Sample $h_t^{(k)}$ from $f(h_t|\boldsymbol{y}, h_1^{(k)}, \ldots, h_{t-1}^{(k)}, h_{t+1}^{(k-1)}, \ldots, h_n^{(k-1)}, \tau^{2(k-1)}, \mu^{(k-1)}, \phi^{(k-1)})$.

(b) Sample $\tau^{2(k)}$ from $f(\tau^2|\boldsymbol{y}, \boldsymbol{h}^{(k)}, \phi^{(k-1)}, \mu^{(k-1)})$.

(c) Sample $\phi^{(k)}$ from $f(\phi|\boldsymbol{y}, \boldsymbol{h}^{(k)}, \tau^{2(k)}, \mu^{(k-1)})$.

(d) Sample $\mu^{(k)}$ from $f(\mu|\boldsymbol{y}, \boldsymbol{h}^{(k)}, \phi^{(k)}, \tau^{2(k)})$.

In practice, we will use the following Bayesian software to implement the above Gibbs sampling algorithm for SV models.

1.4. `WinBUGS`

Due to the advancement of computational power, the emergence of the statistical software `WinBUGS` enables a more straightforward and feasible implementation of the Bayesian

methodology. WinBUGS is an interactive Windows program for Bayesian analysis of complex statistical models using MCMC technique, where the acronym BUGS stands for **B**ayesian inference **U**sing **G**ibbs **S**ampling (Lunn *et al.*, 2000). WinBUGS is a stand-along program and can be run from other software packages such as R, Matlab, Stata and SAS. There are two ways to specify a model in WinBUGS. The first way is to use a text-based model description that is similar to the popular S language used by Splus and R statistical packages. Alternatively the model can be constructed graphically by drawing directed acyclic graphs. The latter method is convenient and user-friendly for practitioners who are not familiar with programming and it has the advantage of the conceiving the structure of the model in terms of graphical paths and links between data and variables. The text-based method describes the model by specifying different components of the model. The first is the stochastic component of the model which specifies the distributions of the random variables of the model. The random variables are typically model parameters described by prior distributions and the response data by the sampling distribution. Secondly the logical components of the model are the variables which are functions or transformations of other model parameters. Lastly the constants are fixed values specified in the data section, which usually include the explanatory variables.

When estimating statistical models in WinBUGS, the user needs to firstly specify a distribution for the observed data and prior distributions for the model parameters. The observed sample values add information to the prior distribution to result in a posterior distribution of parameters. WinBUGS adopts the Gibbs sampler to simulate realisations of a Markov chain whose limiting distribution is the posterior distribution of the parameters. This consists of sampling each parameter in turn from their full conditional posterior distributions. WinBUGS has an expert system for choosing the best sampling method to draw from these univariate distributions. The system first checks for conjugacy, in which case the full conditional distribution reduces analytically to a well-known distribution and direct sampling using standard algorithms can be applied. If conjugacy is not detected, but instead the density is log-concave, then the adaptive rejection sampling (Gilks and Wild, 1992) algorithm is employed. If the target density is not log-concave, then a MH algorithm will be used. In the latest version of WinBUGS (version 1.4.3), efficient MH implementation using slice sampling (Neal, 1997) is adopted for target distribution with restricted range and current

point Metropolis method for distribution with unrestricted range. `WinBUGS` uses adaptive algorithms that tune the proposal density to achieve optimal acceptance rates. For example, when updating non-conjugate continuous full conditional distribution that is neither available in closed form nor log-concave, a MH sampler is used. The proposal distribution is a normal distribution centred at the current point with the variance adjusted every 100 iterations over the first 4,000 iterations. This is to obtain an acceptance rate of between 20% to 40%. The self tuning variance is included in a MH updater option, which is coupled to each node requiring Metropolis updates. Thus a distinct and independently adapted proposal variance is used for each such node such that as oppose to using a globally adapted variance (which may perform poorly for some nodes), the sampling of each full conditional distribution is 'optimised'.

We run the MCMC algorithm for T iterations and we discard the first B iterations from the sample in order to exclude parameters sampled from non-stationary distribution. From the remaining samples after the burn-in period, we keep every Lth iteration to reduce the autocorrelation in the sample. The resulting sample of $\frac{T-B}{L}$ realisations are used for MCMC output analysis. The output analysis includes, firstly, the analysis of the sample used for the description of the posterior distribution and inference about the parameters and secondly, the monitoring of the MCMC convergence. Firstly, a summary of the descriptive statistics, including the posterior mean, standard deviation, Monte Carlo error and selected quantiles is given for the sample. The posterior mean is usually taken to be the estimate of the parameter while the posterior median is sometimes adopted when the posterior distributions are highly skewed. The test for the significance of parameters of interest is based on a credible interval in Bayesian analysis. For the analyses performed in this thesis, the 95% credible interval constructed from the 2.5% and 97.5% posterior quantiles will be used. We say that a parameter is significantly different from zero if the credible interval does not contain zero. Secondly, the convergence of the sampled values can be monitored by examining the history and autocorrelation (acf) plots. When the sample values do not exhibit strong periodicity and tendency, we can assume convergence of parameters. On the other hand, independence of the posterior sample can be concluded when we observe a sharp cut-off in the acf plot. Other convergence diagnostics using formal statistical tests have been developed and they are discussed in detail in later chapters when they are applied.

Since 1998, WinBUGS has become popular among researchers in a wide variety of scientific disciplines. The popularity of this all-purpose Bayesian software is due to its ability to perform sampling-based computations for a variety of statistical models, including but not limited to, random effects models, generalised linear models and proportion hazards models. In financial context, WinBUGS is employed in studies such as Meyer and Yu (2000), Finlay and Seneta (2008) and Fung and Seneta (2010a). In particular, Meyer and Yu (2000) explored the implementation of a Bayesian analysis of univariate SV models using WinBUGS and Yu and Meyer (2006) compared multivariate SV models via WinBUGS. As discussed in Meyer and Yu (2000), the main strength of WinBUGS is that any changes in the model, such as choosing different prior distributions for the parameters, adopting different error distributions for the returns and assigning different structures in the return and/or volatility equations, can be accomplished with ease. This makes the WinBUGS software particularly useful when it comes to exploring new models. Moreover, we will use the mixing representation of distributions in all of our analyses. Thus another major advantage of WinBUGS is that it allows us to use such representation to specify the model. On the other hand, the major weakness of the software is slow convergence to stationary distribution and inefficiency in simulation using a single move Gibbs sampling algorithm, particularly when sampling the latent volatilities. However, more efficient samplers such as the multi-move sampler (Shephard and Pitt, 1997) require specialised code in low-level programming language such as C++ and any modifications to the model necessitate major reprogramming. Therefore, Meyer and Yu (2000) commented that the gain in efficiency in using a more sophisticated sampling scheme is 'largely outweighed by the ease of implementation in BUGS and the feasibility of running large chains, nowadays, on fast computers.'

The MCMC algorithms for different proposed models in this thesis are implemented using WinBUGS because of both its ease of model implementation and more importantly, the allowance for straightforward comparison among a large number of competitive SV models. Several other packages including R2WinBUGS and BOA (Bayesian Output Analysis, Smith (2005)) in R will also be used as the former is particularly useful in performing simulation study when WinBUGS needs to be called repeatedly and the latter is useful for convergence diagnostic testing.

1.5. Model selection criteria for comparing SV models

As we aim to compare a wide range of SV models, we need efficient model selection criteria that are directly applicable to the SV models. The well-known Akaike information criterion (AIC) (Akaike, 1973) and the Bayesian information criterion (BIC) (Schwarz, 1978) cannot be used for comparing the SV models. The reason for this is that both criteria require the specification of the number of parameters in each model. The number of parameters is not well-defined in the SV model, as we augment the parameter space by the latent volatilities and the existence of the Markovian dependence structure between the volatilities, they cannot be considered as additional free parameters.

Bayes factor have been viewed as the only correct way to carry out Bayesian model comparison for many years. Kim *et al.* (1998) and Chib *et al.* (2002) demonstrated how to compute Bayes factors for the SV models using the marginal likelihood approach of Chib (1995). They also evaluated the marginal likelihood at the posterior mean using particle filtering method (Kitagawa, 1996; Pitt and Shephard, 1999). One major drawback of Bayes factors is that in order to compute this for comparing any two models, it requires the marginal likelihoods and, hence, a marginalisation over the parameter vectors in each model. However, if the dimension of the parameter space is large, then these extremely high-dimensional integration problems pose a formidable computational challenge. This is particularly the case for SV models, where the number of unknown parameters is large and typically exceeding the number of observations due to the latent volatilities. Han and Carlin (2001) commented that computing Bayes factor 'requires substantial time and effort (both human and computer) for a rather modest payoff, namely a collection of posterior model probability estimates...' For our study, we need a more computationally realistic alternative to Bayes factors.

Spiegehalter *et al.* (2002) proposed the Deviance Information Criterion (DIC) which is a Bayesian generalisation of the AIC. It is particularly useful in Bayesian model selection problems where the posterior distributions have been obtained using MCMC simulation. Similar to the AIC and BIC, the DIC combines a measure of model fit with a measure of complexity. DIC is applicable to a wide range of statistical models, with successful applications in the field of medical statistics (Zhu and Carlin, 2000) and financial time series (Berg *et al.*, 2004). In particular, Berg *et al.* (2004) demonstrated the practical performance of DIC as a model selection criterion for comparing various SV models. More specifically, by

using a simulation study and through empirical analysis, they showed that the DIC generally gives the same model ranking as Chib's marginal likelihood. They further concluded that the performance of the DIC for model comparison purposes is stable, as the Monte Carlo error is fairly low and the DIC is robust to the change of prior distributions. However, some authors believe the DIC to be undesirable and details can be found in the discussion section of Spiegelhalter *et al.* (2002).

Following the original suggestion of Dempster (1974), computation for DIC is based on posterior distribution of the deviance, defined by

$$D(\boldsymbol{\theta}) = -2 \ln f(\boldsymbol{y}|\boldsymbol{\theta}),$$

where $\boldsymbol{\theta}$ denotes the parameter vector, $\boldsymbol{y}$ is observed data and $f(\boldsymbol{y}|\boldsymbol{\theta})$ is the likelihood function. The posterior mean of the deviance, defined as

$$\bar{D} = E_{\boldsymbol{\theta}|\boldsymbol{y}}[D(\boldsymbol{\theta})],$$

is a Bayesian measure of model adequacy, which attains smaller value for more appropriate models. The second component of the DIC is the effective number of parameters, computed as the the difference between the posterior mean of the deviance and the deviance evaluated at the posterior mean of the parameters $\bar{\boldsymbol{\theta}}$, that is,

$$p_D = \bar{D} - D(\bar{\boldsymbol{\theta}}).$$

This quantity measures the complexity of the model. By combining the two components above, the DIC is calculated as

$$DIC = \bar{D} + p_D.$$

The model with a smaller value for DIC is preferred.

Using the concept of effective number of parameters, DIC is applicable to complex hierarchical models, including SV models, where the number of unknown parameters is not well-defined and exceeds the number of observations. This is in clear contrast to AIC and BIC where the number of parameters must be specified. Moreover, in contrast with the computationally expensive Bayes factor, DIC can be easily computed from any MCMC output. An estimate of $\bar{D}$ can be calculated by monitoring $D(\boldsymbol{\theta})$ and then average the simulated

value of $D(\boldsymbol{\theta})$. The effective number of parameters p_D can be calculated from first evaluating $D(\boldsymbol{\theta})$ at the posterior mean of the simulated values of $\boldsymbol{\theta}$ and then subtract this value from the estimate of $\bar{D}$. The `WinBUGS` software contains a DIC module that automatically calculates value for DIC, hence the computation is almost trivial. For all of our studies in the later chapters, we will employ the DIC as an efficient model selection criterion for comparing various SV models. DIC is also used in the study by Yu and Meyer (2006) and Wang *et al.* (2011a,b).

1.6. Objective and structure of the thesis

The objective of this thesis is to develop and explore more flexible SV models for various applications, through modifying the model structure as well as adopting more flexible error distributions. We focus on improving the ability of the SV models in capturing stylised features of financial time series data, including conditional skewness and leptokurtosis of the returns and general asymmetric behaviour between the return and underlying volatility processes. We examine the effect of adopting different model specifications and various error distributions for the returns on the estimation of model parameters and latent volatilities. Throughout the thesis, symmetric and skewed, univariate and multivariate distributions will be expressed using scale mixture of uniform or normal distributions whenever possible, and we show such representations are particularly useful in simplifying the conditional posterior distributions of the parameters used for the Gibbs sampler and in providing diagnostic tools for the detection of possible outliers.

This thesis is set out as follows. In Chapter 2, we consider a heavy-tailed SV model with leverage and propose an alternative formulation to the one employed by Choy *et al.* (2008). Since Black (1976) revealed that an increase in volatility is associated with a drop in equity return, studies have been conducted to model this leverage effect by allowing a correlation between the innovations of the returns and volatilities, for example, Meyer and Yu (2000), Omori *et al.* (2007) and Choy *et al.* (2008), among many others. Choy *et al.* (2008), in particular, replaced the bivariate normal distribution for the SV model with a bivariate Student-t distribution, in order to account for the leptokurtosis of the returns and to improve model robustness. They further expressed the bivariate Student-t distribution as a bivariate scale mixture of normal distributions such that conditional on the mixing parameters, the returns and volatilities have a joint normal distribution. This not only simplifies the Gibbs sampling

algorithm but also the mixing parameters as the by-products of the scale mixture representation are used to identify outliers, as was done in Choy *et al.* (2008). However, the major drawback of this method is that once outliers are identified, it is not possible to determine whether they are outlying returns or outlying volatilities, since the mixing parameter is assigned to a pair of return and volatility at time t. We show that by simply interchanging the order in which we derive the conditional distribution of the returns given the future volatility and employing the scale mixture form for the Student-t distribution, we are able to obtain separate mixing parameters for the return and volatility processes, hence we separate the sources of outliers. A shorter version of the chapter forms the paper published as Wang *et al.* (2011a).

In Chapter 3, we employ a more flexible distribution, known as the generalised-t (GT) distribution for the returns. Since it has long been recognised that the leptokurtosis of asset returns cannot fully be explained by time-varying volatility, many studies focus on adopting heavy-tailed distributions such as the Student-t and EP distributions for the conditional returns. However, the general approach is to fit different heavy-tailed distributions individually to the data. Their empirical performance is then assessed by some model selection criterion such as the Bayes factor and DIC. We propose to use the GT distribution, not only because the shape parameters of the distribution cover a wide range of density shapes, but also because it nests several commonly used distributions including the normal, Student-t and EP distributions for SV models. Hence we incorporate different distributions into a unified distributional framework, expressed as scale mixture of uniform distributions. The first part of Chapter 3 is due to be published as Wang *et al.* (2011b). In the second part, we extend the SV model with GT distribution to take into account the general asymmetric behaviour between the returns and volatilities, following the formulation in Asai and McAleer (2005a). Hence, our work can be considered an extension of Asai and McAleer (2005a), with a more flexible error distribution for the returns.

Furthermore, the symmetric heavy-tailed distribution is appropriate in capturing leptokurtic feature of the returns, but is unable to model the empirical skewness observed in the asset return data. Hence the focus of Chapter 4 is the extension of SV model to skewed distribution for the returns. We will adopt the asymmetric Laplace (AL) distribution to simultaneously capture leptokurtosis and skewness in the returns. We will also include lagged

returns and exogenous variables in the return equation. By exploiting the link between quantile regression and the AL distribution, we are also able to examine causal effects between dynamic variables across different quantile levels. In our study, heteroscedasticity is accounted for by assuming a stochastic process for volatility, which is a stochastic extension of Chen *et al.* (2009) who controlled for the volatility dynamics by a GARCH specification in a quantile regression model. Firstly, we allow the shape parameter of the AL distribution to be random and perform the usual estimation of the SV model, with return innovations follow an AL distribution. Afterwards, we consider the quantile regression model with SV dynamics, where we investigate causal relations between the covariates and returns across different quantile levels.

The SV model has been extended to allow for different prominent features in financial time series including leverage effect, general asymmetric effect between returns and volatilities and their leptokurtic and skewed distributions in univariate framework, we consider in Chapter 5, a multivariate extension of SV model to describe the dynamic of cross-correlation between several financial time series. In particular, we employ different multivariate SV models to investigate the daily price volatilities in Australian electricity markets. Compared to the conventional financial markets, the use of sophisticated econometric models is limited in energy markets including electricity markets. Some work has been done to examine the applicability of a range of GARCH specifications to model price volatilities in the Australian national electricity markets, however, the modelling of electricity data using SV model is an area that has not been widely explored. The only current study using SV models focuses on a single regional electricity market of Australia, hence the correlation between the regional markets has not been considered. We propose to employ multivariate SV models which incorporate the correlation between the electricity prices as well as price volatilities from different regional markets. In addition, pronounced characteristics including weekday/weekend effects, skewness and leptokurtosis of electricity price data are accounted for by incorporating appropriate covariates and assuming skewed heavy-tailed distributions. The contents of Chapter 5 are being submitted for publication.

Lastly in Chapter 6, we summarise this research with some concluding remarks and implications for future developments.

CHAPTER 2

SV models with Leverage and Heavy-tailed Distribution

2.1. Overview

The leverage effect, revealed by Black (1976), is used to describe the phenomenon of a negative correlation between return and volatility. Choy *et al.* (2008) considered the SV model with leverage where the returns and volatilities follow a bivariate Student-t distribution to allow for the negative correlation between them. Using the scale mixture of uniforms and the scale mixture of normals representations, they demonstrated that the Gibbs sampler for the SV model can be simplified and a means for outlier diagnostics is warranted.

The focus of this chapter is on the SV model with leverage (SVL) and the Student-t error distribution. We develop an alternative formulation of the bivariate Student-t SV model to Choy *et al.* (2008) and we show through simulation and empirical studies, that our approach provides an improvement over Choy's approach.

2.2. Background

In financial markets, there is voluminous evidence that financial time series exhibit asymmetric reactions to positive and negative returns. As with the asymmetric GARCH models, several studies have considered an asymmetric variance in the SV model framework. Early studies by Black (1976) and Nelson (1991) revealed that a drop in equity return was associated with an increase in volatility. This asymmetric return-volatility relationship was attributed to changes in financial leverage, or debt-to-equity ratios, hence this phenomenon is commonly known as the financial leverage effect. The other explanation for the volatility asymmetry is known as the volatility feedback effect. The theory states that 'if volatility is priced, an anticipated increase in volatility would raise the required rate of return, in turn necessitating an immediate stock price decline in order to allow for higher future returns.' (Bollerslev *et al.*, 2006). The fundamental difference between leverage and volatility feedback effects is that the leverage effect explains why a negative return leads to higher subsequent volatility as it hinges on the reverse causal relationship while the volatility feedback

effect explains how an increase in volatility may result in negative returns as the causality relationship runs from volatility to prices. Many studies compared the magnitude of these two effects empirically. Earlier studies by Nelson (1991), Glosten *et al.* (1993), and Engle and Ng (1993), have found that volatility increases more following negative returns than positive returns and that the relationship between expected returns and volatility is insignificant. Later studies by Bekaert and Wu (2000) and Wu (2001) found that the volatility feedback effect dominates the leverage effect empirically.

We will be focusing on the leverage effect, which is modelled by introducing a direct correlation between the returns and volatilities. However, two alternative specifications of the incorporation of this correlation co-exist in the literature. The first approach was pursued by Harvey and Shephard (1996). Their discrete time SV model is the Euler approximation to the continuous time asymmetric SV model used commonly in option pricing literature. The leverage effect is modelled as the correlation between the current return and future volatility. Using stock data, Harvey and Shephard (1996) found strong evidence of leverage effect. An alternative approach was taken by Jacquier *et al.* (2004) where the leverage effect is modelled by allowing a contemporaneous dependence between returns and volatilities. Yu (2005) noted this subtle but important difference and he formally compared these two alternative specifications on both theoretical and empirical grounds. Firstly, the contemporaneous model by Jacquier *et al.* (2004) does not comply with the efficient market hypothesis as the model violates the martingale property. Secondly, Yu (2005) showed that while it is straightforward to interpret the leverage effect in the model by Harvey and Shephard (1996), the interpretation of the leverage effect is unclear in the model used by Jacquier *et al.* (2004). Lastly, Yu (2005) compared the empirical performance of the two alternative specifications using daily return data of two market indices. His results demonstrated the superiority of Harvey and Shephard's model over the model by Jacquier *et al.* (2004). The specification of Harvey and Shephard (1996) is subsequently employed by Meyer and Yu (2000) and Choy *et al.* (2008) and we will describe each of their methods and how their approaches motivated the development of our study on the SVL model in the next section.

2.3. SVL model specification

Before we describe the model specification, we need the following definition.

2.3.1. Student-t distribution as a SMN. A wide class of continuous and symmetric distribution can be constructed as scale mixture of normal (SMN) distributions. Some examples include the well-known contaminated normals, the Student-t distribution, logistic and Laplace distribution. Andrews and Mallows (1974) presented the class of SMN distributions. Let Z be a standard normal random variable and λ follows some distribution on the positive real line $\mathbb{R}^+$ with a continuous or discrete density $\pi(\lambda)$ which is independent of Z. Then we refer to the distribution of the univariate random variable $X = Z\lambda$ as a SMN. The probability density function (pdf) of X with location parameter μ and scale parameter σ can be expressed in the following mixture form

$$f(x|\mu,\sigma) = \int_0^\infty N(x|\mu,\kappa(\lambda)\sigma^2)\pi(\lambda)d\lambda, \tag{2.1}$$

where $N(x|.,.)$ is a normal pdf and $\kappa(.)$ is a positive function. The parameter λ is called the mixing parameter and $\pi(.)$ is known as the mixing density of this SMN representation.

From (2.1), a SMN random variable X with location μ and scale σ can be expressed hierarchically as

$$\begin{aligned} X|\lambda &\sim N(\mu,\kappa(\lambda)\sigma^2), \\ \lambda &\sim \pi(\lambda). \end{aligned}$$

The univariate SMN can be extended to the multivariate case. Let $\boldsymbol{X}$ be a p-dimensional vector of continuous random variables with location vector $\boldsymbol{\mu}$ and scale matrix $\boldsymbol{\Sigma}$. If the pdf of $\boldsymbol{X}$ can be expressed into the following mixture representation

$$f(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \int_0^\infty N_p(\boldsymbol{x}|\boldsymbol{\mu},\kappa(\lambda)\boldsymbol{\Sigma})\,\pi(\lambda)\,d\lambda,$$

where $N_p(\boldsymbol{x}|\cdot,\cdot)$ is a p-dimensional multivariate normal pdf, then we say that the pdf of $\boldsymbol{X}$ has a multivariate SMN representation.

For a multivariate Student-t distribution with location vector $\boldsymbol{\mu}$, scale matrix $\boldsymbol{\Sigma}$ and degrees of freedom ν, it has a SMN representation with $\kappa(\lambda) = \lambda^{-1}$ and $\pi(\lambda)$ is the pdf of the gamma $Ga\left(\frac{\nu}{2},\frac{\nu}{2}\right)$ distribution given by

$$f(x|a,b) = \frac{b^a}{\Gamma(a)}x^{a-1}e^{-bx}, \quad x > 0.$$

Therefore, the pdf of the multivariate Student-t distribution can be rewritten as

$$f(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)=\int_0^\infty N\left(\boldsymbol{x}|\boldsymbol{\mu},\lambda^{-1}\boldsymbol{\Sigma}\right)Ga\left(\lambda|\frac{\nu}{2},\frac{\nu}{2}\right)\,d\lambda, \tag{2.2}$$

or we can express the multivariate Student-t distribution hierarchically as:

$$\begin{aligned}\boldsymbol{X}|\boldsymbol{\mu},\boldsymbol{\Sigma},\nu,\lambda &\sim N\left(\boldsymbol{\mu},\lambda^{-1}\boldsymbol{\Sigma}\right),\\ \lambda|\nu &\sim Ga\left(\frac{\nu}{2},\frac{\nu}{2}\right).\end{aligned}$$

This hierarchical structure of SMN representation has two main advantages. Firstly, when the posterior distributions of the parameters lack the conjugate structure, it is sometimes difficult to sample from these distributions and various sampling methods are required. Representing a heavy-tailed distribution such as the Student-t distribution in a scale mixture form results in a simpler set of full conditional distributions and alleviates the computational burden of the Gibbs sampler in the MCMC algorithm. Secondly, the scale mixture form of the density contains a mixing parameter which can be used to identify extreme values in outlier diagnosis (Choy and Smith, 1997). An outlier is associated with a large value of the mixing parameter in (2.2) which inflates the variance of the corresponding normal distribution to accommodate the outlier. Therefore, the extremeness of observations is closely related to the magnitude of the mixing parameters. In practice, we identify outliers by comparing the posterior means of the mixing parameters of all observations.

2.3.2. SVL model. Adopting the specification by Harvey and Shephard (1996), the univariate SVL model is given by

$$\begin{aligned}y_t|h_t &= \exp(h_t/2)\epsilon_t,\\ h_{t+1}|h_t &= \mu+\phi(h_t-\mu)+\tau\eta_t,\\ \begin{pmatrix}\epsilon_t\\ \eta_t\end{pmatrix} &\overset{iid}{\sim} N\left(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}\right),\end{aligned} \tag{2.3}$$

where $\rho=\text{cor}(\epsilon_t,\eta_t)$ which is used to measure the leverage effect and $h_1\sim N\left(\mu,\tau^2/(1-\phi^2)\right)$. Meyer and Yu (2000) analysed this SVL model and their approach is to express the bivariate normal density as the product of the conditional density of $y_t|h_{t+1},h_t$ and the marginal density of $h_{t+1}|h_t$. To facilitate an efficient posterior inference using the `WinBUGS` software,

Meyer and Yu (2000) showed that this SVL model can alternatively be specified by

$$\begin{aligned} y_t|h_{t+1},h_t &\sim N\left(\frac{\rho}{\tau}\exp\left(\frac{h_t}{2}\right)[h_{t+1}-\mu-\phi(h_t-\mu)],\exp(h_t)(1-\rho^2)\right), \\ h_{t+1}|h_t &\sim N\left(\mu+\phi(h_t-\mu),\tau^2\right), \\ h_1 &\sim N\left(\mu,\frac{\tau^2}{1-\phi^2}\right). \end{aligned} \tag{2.4}$$

To see the equivalence of (2.3) and (2.4), we need to derive the conditional distribution of $y_t|h_{t+1},h_t$. The joint distribution of $y_t,h_{t+1}|h_t$ is a bivariate normal distribution with the following density

$$\begin{aligned} f(y_t,h_{t+1}|h_t) &= \frac{1}{2\pi\exp(\frac{h_t}{2})\tau\sqrt{1-\rho^2}}\exp\Big(-\frac{1}{2\exp(h_t)\tau^2(1-\rho^2)}\Big(y_t^2\tau^2 \\ &\quad -2\rho y_t\tau\exp\left(\frac{h_t}{2}\right)[h_{t+1}-\mu-\phi(h_t-\mu)]+\exp(h_t)[h_{t+1}-\mu-\phi(h_t-\mu)]^2\Big)\Big) \end{aligned}$$

The marginal density of $h_{t+1}|h_t$ is

$$f(h_{t+1}|h_t)=\frac{1}{\sqrt{2\pi}\tau}\exp\left(-\frac{1}{2\tau^2}[h_{t+1}-\mu-\phi(h_t-\mu)]^2\right)$$

The conditional distribution of $y_t|h_{t+1},h_t$ is obtained by dividing the joint density by the marginal density. After algebraic simplification, we get

$$y_t|h_{t+1},h_t\sim N\left(\frac{\rho}{\tau}\exp\left(\frac{h_t}{2}\right)[h_{t+1}-\mu-\phi(h_t-\mu)],\exp(h_t)(1-\rho^2)\right).$$

To adequately account for the leptokurtic feature of financial return series, a bivariate Student-t distribution with ν degrees of freedom can replace the bivariate normal distribution and (2.3) becomes

$$\begin{pmatrix}\epsilon_t\\ \eta_t\end{pmatrix} \overset{iid}{\sim} t_\nu\left(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}1&\rho\\ \rho&1\end{pmatrix}\right).$$

Equivalently, model can be represented by

$$\begin{pmatrix}y_t|h_t\\ h_{t+1}|h_t\end{pmatrix} \overset{iid}{\sim} t_\nu\left(\begin{pmatrix}0\\ \mu+\phi(h_t-\mu)\end{pmatrix},\begin{pmatrix}\exp(h_t)&\tau\rho\exp(h_t/2)\\ \tau\rho\exp(h_t/2)&\tau^2\end{pmatrix}\right). \tag{2.5}$$

Choy *et al.* (2008) studied the robustness property of this t SVL model by first expressing the bivariate Student-t error distribution into the following bivariate SMN representation

$$\begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} \overset{iid}{\sim} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \frac{1}{\lambda_t}\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right),$$

or equivalently,

$$\begin{aligned} \begin{pmatrix} y_t|h_t \\ h_{t+1}|h_t \end{pmatrix} &\sim N\left(\begin{pmatrix} 0 \\ \mu+\phi(h_t-\mu) \end{pmatrix}, \frac{1}{\lambda_t}\begin{pmatrix} \exp(h_t) & \tau\rho\exp(h_t/2) \\ \tau\rho\exp(h_t/2) & \tau^2 \end{pmatrix}\right), \\ \lambda_t &\overset{iid}{\sim} Ga\left(\frac{\nu}{2},\frac{\nu}{2}\right), \end{aligned} \tag{2.6}$$

for $t = 1, \ldots, n$ and then adopted the formulation of Meyer and Yu (2000). Note that by conditioning on the mixing parameter λ_t, the joint distribution of (y_t, h_{t+1}) is now bivariate normal. Under Choy's approach, the t SVL model is reformulated as

$$\begin{aligned} y_t|h_{t+1},h_t &\sim N\left(\frac{\rho}{\tau}\exp\left(\frac{h_t}{2}\right)[h_{t+1}-\mu-\phi(h_t-\mu)], \frac{\exp(h_t)(1-\rho^2)}{\lambda_t}\right), \\ h_{t+1}|h_t &\sim N\left(\mu+\phi(h_t-\mu), \frac{\tau^2}{\lambda_t}\right), \\ h_1 &\sim N\left(\mu, \frac{\tau^2}{\lambda_1(1-\phi^2)}\right). \end{aligned} \tag{2.7}$$

Choy *et al.* (2008) showed that the Student-t distribution can provide a robust inference in financial applications and the mixing parameter λ_t, associated with the pairs (y_t, h_{t+1}) can be used to identify outlying pairs. However, one shortfall of their approach is that once an outlier is identified, it is unclear whether it is outlying in return, volatility or both.

In order to improve upon Choy's approach, we propose an alternative formulation for the t SVL model by first deriving a conditional Student-t distribution for the returns and a marginal Student-t distribution for the volatilities from the joint distribution of (y_t, h_{t+1}). More specifically, the joint density for the bivariate Student-t distribution of $y_t, h_{t+1}|h_t$ is given by

$$f(y_t, h_{t+1}|h_t) = \frac{\Gamma(\frac{\nu}{2}+1)}{\Gamma(\frac{\nu}{2})\nu\pi\tau\exp(\frac{h_t}{2})\sqrt{1-\rho^2}}\left[1+\frac{1}{\nu}\frac{M_t^2-2\rho M_t N_t+N_t^2}{D_t}\right]^{-(\frac{\nu}{2}+1)},$$

where

$$\begin{aligned} D_t &= \tau^2(1-\rho^2)\exp(h_t), \\ M_t &= y_t\tau, \\ N_t &= \exp\left(\frac{h_t}{2}\right)\{h_{t+1}-[\mu+\phi(h_t-\mu)]\} \end{aligned}$$

and the pdf for the marginal distribution of $h_{t+1}|h_t$ is

$$f(h_{t+1}|h_t) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\nu\pi}\tau}\left[1+\frac{1}{\nu\tau^2}\{h_{t+1}-[\mu+\phi(h_t-\mu)]\}^2\right]^{-(\frac{\nu+1}{2})}.$$

We now show that the bivariate SMN form (2.6) of Choy results in the same joint density for $y_t, h_{t+1}|h_t$ as stated above.

Let

$$\begin{aligned} \boldsymbol{X}_t &= (y_t|h_t, h_{t+1}|h_t)^T, \\ \boldsymbol{\mu}_t &= (0, \mu+\phi(h_t-\mu))^T, \\ \boldsymbol{\Sigma} &= \begin{pmatrix} \exp(h_t) & \tau\rho\exp(h_t/2) \\ \tau\rho\exp(h_t/2) & \tau^2 \end{pmatrix}. \end{aligned}$$

Now using (2.2), the pdf of $\boldsymbol{X}_t$ is given by

$$\begin{aligned} &\int_0^\infty N\left(\boldsymbol{X}_t|\boldsymbol{\mu}_t, \frac{\boldsymbol{\Sigma}}{\lambda_t}\right) Ga\left(\lambda_t|\frac{\nu}{2}, \frac{\nu}{2}\right) d\lambda_t \\ = &\int_0^\infty \frac{\lambda_t}{2\pi\exp(\frac{h_t}{2})\tau\sqrt{1-\rho^2}} \exp\left\{-\frac{\lambda_t(M_t^2-2\rho M_tN_t+N_t^2)}{2\exp(h_t)\tau^2(1-\rho^2)}\right\} \frac{\frac{\nu}{2}^{\frac{\nu}{2}}\lambda_t^{\frac{\nu}{2}-1}}{\Gamma\left(\frac{\nu}{2}\right)} \exp\left(-\frac{\nu}{2}\lambda_t\right) d\lambda_t \\ = &\frac{1}{2\pi\exp(\frac{h_t}{2})\tau\sqrt{1-\rho^2}} \frac{\frac{\nu}{2}^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)} \int_0^\infty \lambda_t^{\frac{\nu}{2}} \exp\left\{-\lambda_t\left(\frac{M_t^2-2\rho M_tN_t+N_t^2}{2\exp(h_t)\tau^2(1-\rho^2)}+\frac{\nu}{2}\right)\right\} d\lambda_t \\ = &\frac{\frac{\nu}{2}^{\frac{\nu}{2}}}{2\pi\exp(\frac{h_t}{2})\tau\sqrt{1-\rho^2}} \frac{\Gamma\left(\frac{\nu}{2}+1\right)}{\Gamma\left(\frac{\nu}{2}\right)} \left(\frac{M_t^2-2\rho M_tN_t+N_t^2}{2\exp(h_t)\tau^2(1-\rho^2)}+\frac{\nu}{2}\right)^{-(\frac{\nu}{2}+1)} \\ &\times\int_0^\infty \lambda_t^{\frac{\nu}{2}} \exp\left\{-\lambda_t\left(\frac{M_t^2-2\rho M_tN_t+N_t^2}{2\exp(h_t)\tau^2(1-\rho^2)}+\frac{\nu}{2}\right)\right\} \left(\frac{M_t^2-2\rho M_tN_t+N_t^2}{2\exp(h_t)\tau^2(1-\rho^2)}+\frac{\nu}{2}\right)^{(\frac{\nu}{2}+1)} \\ &\qquad\qquad \times\frac{1}{\Gamma\left(\frac{\nu}{2}+1\right)} d\lambda_t \\ = &\frac{(\frac{\nu}{2})^{\frac{\nu}{2}}}{2\pi\exp(\frac{h_t}{2})\tau\sqrt{1-\rho^2}} \frac{\Gamma\left(\frac{\nu}{2}+1\right)}{\Gamma\left(\frac{\nu}{2}\right)} \left[\frac{\nu}{2}\left(1+\frac{M_t^2-2\rho M_tN_t+N_t^2}{\nu\exp(h_t)\tau^2(1-\rho^2)}\right)\right]^{-(\frac{\nu}{2}+1)} \end{aligned}$$

$$= \frac{\Gamma\left(\frac{\nu}{2}+1\right)}{\Gamma\left(\frac{\nu}{2}\right)\nu\pi\tau\exp(\frac{h_t}{2})\sqrt{1-\rho^2}}\left(1+\frac{M_t^2-2\rho M_tN_t+N_t^2}{\nu\exp(h_t)\tau^2(1-\rho^2)}\right)^{-(\frac{\nu}{2}+1)}$$

which is desired bivariate Student-t density.

Adopting the idea of Meyer and Yu (2000), we express the conditional density of $y_t|h_{t+1},h_t$ as the quotient of bivariate Student-t distribution and the marginal density of $h_{t+1}|h_t$. That is,

$$\begin{aligned}
& f(y_t|h_{t+1},h_t) \\
= & \frac{\frac{\Gamma(\frac{\nu}{2}+1)}{\Gamma(\frac{\nu}{2})\nu\pi\tau\exp(\frac{1}{2}h_t)\sqrt{1-\rho^2}}\left(1+\frac{1}{\nu}\frac{M_t^2-2\rho M_tN_t+N_t^2}{D_t}\right)^{-(\frac{\nu}{2}+1)}}{\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})(\nu\pi)^{\frac{1}{2}}\tau}\left(1+\frac{1}{\nu\tau^2}\left\{h_{t+1}-[\mu+\phi(h_t-\mu)]\right\}^2\right)^{-(\frac{\nu+1}{2})}} \\
= & \frac{\Gamma(\frac{\nu}{2}+1)}{\Gamma(\frac{\nu+1}{2})[\nu\pi(1-\rho^2)]^{\frac{1}{2}}\exp(\frac{h_t}{2})\left[1+\frac{1}{\nu\tau^2}\left\{h_{t+1}-[\mu+\phi(h_t-\mu)]\right\}^2\right]^{\frac{1}{2}}} \\
& \times\left[\frac{1+\frac{1}{\nu}\frac{M_t^2-2\rho M_tN_t+N_t^2}{D_t}}{1+\frac{\left\{h_{t+1}-[\mu+\phi(h_t-\mu)]\right\}^2}{\nu\tau^2}}\right]^{-(\frac{\nu}{2}+1)}
\end{aligned}$$

The expression $\dfrac{1+\frac{1}{\nu}\frac{M_t^2-2\rho M_tN_t+N_t^2}{D_t}}{1+\frac{\{h_{t+1}-[\mu+\phi(h_t-\mu)]\}^2}{\nu\tau^2}}$ is simplified to

$$1+\frac{\left(y_t-\frac{\rho}{\tau}\left\{h_{t+1}-[\mu+\phi(h_t-\mu)]\right\}\exp\left(\frac{h_t}{2}\right)\right)^2}{\nu(1-\rho^2)\exp(h_t)\left(1+\frac{\left\{h_{t+1}-[\mu+\phi(h_t-\mu)]\right\}^2}{\nu\tau^2}\right)}.$$

Hence the conditional density is given by

$$
\begin{aligned}
& f(y_t|h_{t+1},h_t) \\
= & \frac{\Gamma(\frac{\nu+1}{2}+\frac{1}{2})}{\Gamma(\frac{\nu+1}{2})[(\nu+1)\pi]^{\frac{1}{2}}\left\{\frac{\nu}{\nu+1}(1-\rho^2)\exp(h_t)\left[1+\frac{1}{\nu\tau^2}\left\{h_{t+1}-[\mu+\phi(h_t-\mu)]\right\}^2\right]\right\}^{\frac{1}{2}}} \times \\
& \left[1+\frac{\left[y_t-\frac{\rho}{\tau}\exp\left(\frac{h_t}{2}\right)\left\{h_{t+1}-[\mu+\phi(h_t-\mu)]\right\}\right]^2}{(\nu+1)(\frac{\nu}{\nu+1})(1-\rho^2)\exp\left(h_t\right)\left[1+\frac{1}{\nu\tau^2}\left\{h_{t+1}-[\mu+\phi(h_t-\mu)]\right\}^2\right]}\right]^{-(\frac{(\nu+1)+1}{2})}.
\end{aligned}
$$

Therefore, we show that the t SVL model in (2.5) can be specified by

$$
\begin{aligned}
y_t|h_{t+1},h_t &\sim t_{\nu+1}\Big(\frac{\rho}{\tau}\exp\left(\frac{h_t}{2}\right)\left\{h_{t+1}-[\mu+\phi(h_t-\mu)]\right\}, \\
&\qquad \left(\frac{\nu}{\nu+1}\right)(1-\rho^2)\exp(h_t)\Big[1+\frac{1}{\nu\tau^2}\left\{h_{t+1}-[\mu+\phi(h_t-\mu)]\right\}^2\Big]\Big), \\
h_{t+1}|h_t &\sim t_\nu\left(\mu+\phi(h_t-\mu),\tau^2\right), \\
h_1 &\sim t_\nu\left(\mu,\frac{\tau^2}{1-\phi^2}\right). \qquad (2.8)
\end{aligned}
$$

Now, our approach is then to express the Student-t distributions in (2.8) as SMN distributions and the model becomes

$$
\begin{aligned}
y_t|h_{t+1},h_t &\sim N\Big(\frac{\rho}{\tau}\exp\left(\frac{h_t}{2}\right)\left\{h_{t+1}-[\mu+\phi(h_t-\mu)]\right\}, \\
&\qquad \frac{1}{\lambda_{y_t}}\left(\frac{\nu}{\nu+1}\right)(1-\rho^2)\exp(h_t)\Big[1+\frac{1}{\nu\tau^2}\left\{h_{t+1}-[\mu+\phi(h_t-\mu)]\right\}^2\Big]\Big), \\
h_{t+1}|h_t &\sim N\left(\mu+\phi(h_t-\mu),\frac{\tau^2}{\lambda_{h_{t+1}}}\right), \\
h_1 &\sim N\left(\mu,\frac{\tau^2}{\lambda_{h_1}(1-\phi^2)}\right), \\
\lambda_{y_t} &\sim Ga\left(\frac{\nu+1}{2},\frac{\nu+1}{2}\right), \\
\lambda_{h_t} &\sim Ga\left(\frac{\nu}{2},\frac{\nu}{2}\right). \qquad (2.9)
\end{aligned}
$$

Comparing (2.4), (2.7) and (2.9), which correspond to the approaches taken by Meyer and Yu (2000), Choy *et al.* (2008) and us respectively, we see that Choy's approach is essentially taking the conditional density of $y_t|h_{t+1},h_t$ in Meyer and Yu (2000) and conditioning on the

mixing parameter λ_t (note this mixing parameter is common for a pair of (y_t, h_{t+1})) to arrive at the t SVL model. The main difference between our approach and Choy's approach lies in the order of conditioning on the mixing parameter and obtaining the conditional density of $y_t|h_{t+1}$. The consequence is that the density of $y_t|h_{t+1}$ has a different expression for the variance and more importantly, by using two mixing parameters, λ_{y_t} and λ_{h_t}, our approach is now capable of identifying outliers in the return and volatility processes separately.

2.4. Bayesian inference for the proposed t SVL model

2.4.1. MCMC algorithm. The `WinBUGS` software described in Chapter 1 is used to implement the MCMC algorithm for the t SVL model described in (2.9). To complete the Bayesian paradigm, we assign the following prior distributions to the model parameters:

$$\begin{aligned}
\mu &\sim N(a_\mu, b_\mu), \\
\tau^2 &\sim IG(a_\tau, b_\tau), \\
\phi^* &\sim Be(a_\phi, b_\phi), \\
\rho &\sim U(-1, 1), \\
\nu &\sim IG(a_\nu, b_\nu),
\end{aligned}$$

where $\phi^* = \frac{\phi+1}{2}$, $U(a, b)$ is the uniform distribution and $IG(a, b)$ is the inverse gamma distribution with density given in Chapter 1. The joint prior distribution of all unobservables, including the model parameters and latent volatilities, is given by

$$f(\mu, \phi, \tau^2, \rho, h_1, \ldots, h_n, h_{n+1}|y_1, \ldots, y_n) = f(\mu)f(\phi)f(\tau^2)f(\rho)f(h_1)\prod_{t=1}^{n} f(h_{t+1}|h_t)$$

by assuming prior independence of the model parameters and successive conditioning of the latent volatilities. The likelihood of observed data given the latent volatilities is specified by:

$$f(y_1, \ldots, y_n|\mu, \phi, \tau^2, \rho, h_1, \ldots, h_n, h_{n+1}) = \prod_{t=1}^{n} f(y_t|h_{t+1}, h_t).$$

By Bayes' theorem, the posterior distribution of the unobservables is proportional to the joint distribution of data likelihood times the prior distribution. So for the t SVL model, the joint

density is

$$f(\mu,\phi,\tau^2,\rho,h_1,\ldots,h_n,h_{n+1}) \propto f(\mu)f(\phi)f(\tau^2)f(\rho)f(h_1) \times\prod_{t=1}^{n} f(h_{t+1}|h_t)\prod_{t=1}^{n} f(y_t|h_{t+1},h_t).$$

The Gibbs sampler generates a sequence of samples from the joint posterior distribution by sampling from the univariate full conditional distribution for each variable iteratively, conditional on the current values of other variables.

Let $\boldsymbol{y}=(y_1,\ldots,y_n)$ be the vector of log-returns, $\boldsymbol{h}=(h_1,\ldots,h_n)$ be the vector of log-volatilities, $\boldsymbol{\lambda}_y=(\lambda_{y_1},\ldots,\lambda_{y_n})$ the mixing parameters for returns and $\boldsymbol{\lambda}_h=(\lambda_{h_1},\ldots,\lambda_{h_n})$ the mixing parameters for log-volatilities. Also define $\boldsymbol{h}_{\setminus t}=(h_1,\ldots,h_{t-1},h_{t+1},\ldots,h_n)$, $\boldsymbol{\lambda}_{y,\setminus t}=(\lambda_{y_1},\ldots,\lambda_{y_{t-1}},\lambda_{y_{t+1}},\ldots,\lambda_{y_n})$ and $\boldsymbol{\lambda}_{h,\setminus t}=(\lambda_{h_1},\ldots,\lambda_{h_{t-1}},\lambda_{h_{t+1}},\ldots,\lambda_{h_n})$. Further define

$$\begin{aligned}\mu_{y_t} &= \frac{\rho}{\tau}\exp\left(\frac{h_t}{2}\right)\left\{h_{t+1}-[\mu+\phi(h_t-\mu)]\right\},\\ \sigma^2_{y_t} &= \left(\frac{\nu}{\nu+1}\right)(1-\rho^2)\exp(h_t)\Big[1+\frac{1}{\nu\tau^2}\left\{h_{t+1}-[\mu+\phi(h_t-\mu)]\right\}^2\Big].\end{aligned}$$

We derive the following set of univariate full conditional distributions for the volatility and model parameters.

- Full conditional density for h_t,

$$
\begin{aligned}
f(h_1|y_1,h_2,\mu,\phi,\tau,\rho,\nu) &\propto N(h_1|\gamma_1[\mu(\lambda_{h_1}+\lambda_{h_2}\phi^2)-\lambda_{h_1}\phi^2\mu+\phi\lambda_{h_2}(h_2-\mu)],\gamma_1\tau^2) \\
&\quad \times N\left(y_1|\mu_{y_1},\frac{\sigma^2_{y_1}}{\lambda_{y_1}}\right) \\
&\propto \left[1+\frac{1}{\nu\tau^2}\{h_2-[\mu+\phi(h_1-\mu)]\}^2\right]^{-1/2} \\
&\quad \times \exp\left\{-\frac{1}{2\gamma_1\tau^2}(h_1-\gamma_1\mu_{h_1})^2-\frac{\lambda_{y_1}}{2\sigma^2_{y_1}}(y_1-\mu_{y_1})^2-\frac{h_1}{2}\right\}, \qquad t=1 \\
f(h_t|\boldsymbol{y},\boldsymbol{h}_{\setminus t},\mu,\phi,\tau,\rho,\nu) &\propto N\big(h_t|\gamma_t\{\mu(\lambda_{h_t}+\lambda_{h_{t+1}}\phi^2)+\phi[\lambda_{h_t}(h_{t-1}-\mu)+\lambda_{h_{t+1}}(h_{t+1}-\mu)]\}, \\
&\quad \gamma_t\tau^2\big) \times N\left(y_t|\mu_{y_t},\frac{\sigma^2_{y_t}}{\lambda_{y_t}}\right) N\left(y_{t-1}|\mu_{y_{t-1}},\frac{\sigma^2_{y_{t-1}}}{\lambda_{y_{t-1}}}\right) \\
&\propto \left[1+\frac{1}{\nu\tau^2}\{h_{t+1}-[\mu+\phi(h_t-\mu)]\}^2\right]^{-1/2} \\
&\quad \times \left[1+\frac{1}{\nu\tau^2}\{h_t-[\mu+\phi(h_{t-1}-\mu)]\}^2\right]^{-1/2} \\
&\quad \times \exp\Big\{-\frac{1}{2\gamma_t\tau^2}(h_t-\gamma_t\mu_{h_t})^2-\frac{h_t}{2}-\frac{h_{t-1}}{2} \\
&\quad -\frac{\lambda_{y_t}}{2\sigma^2_{y_t}}(y_t-\mu_{y_t})^2-\frac{\lambda_{y_{t-1}}}{2\sigma^2_{y_{t-1}}}(y_{t-1}-\mu_{y_{t-1}})^2\Big\}, \qquad t=2,\ldots,n,
\end{aligned}
$$

where

$$
\begin{aligned}
\mu_{h_1} &= \mu(\lambda_{h_1}+\lambda_{h_2}\phi^2)-\lambda_{h_1}\phi^2\mu+\phi\lambda_{h_2}(h_2-\mu), \\
\mu_{h_t} &= \mu(\lambda_{h_t}+\lambda_{h_{t+1}}\phi^2)+\phi[\lambda_{h_t}(h_{t-1}-\mu)+\lambda_{h_{t+1}}(h_{t+1}-\mu)], \\
\gamma_1 &= (\lambda_{h_1}(1-\phi^2)+\lambda_{h_2}\phi^2)^{-1}, \\
\gamma_t &= (\lambda_{h_t}+\lambda_{h_{t+1}}\phi^2)^{-1}.
\end{aligned}
$$

- Full conditional density for λ_{y_t}

$$
\begin{aligned}
\lambda_{y_t}|\boldsymbol{h},\boldsymbol{y},\boldsymbol{\lambda}_{y,\setminus t},\mu,\phi,\tau^2,\nu \;\sim\; & Ga\Big(\frac{\nu}{2}+1,\frac{\nu+1}{2}+ \\
& \frac{\{y_t-\frac{\rho}{\tau}\exp\left(\frac{h_t}{2}\right)\{h_{t+1}-[\mu+\phi(h_t-\mu)]\}\}^2}{2\left\{\frac{\nu}{\nu+1}(1-\rho^2)\exp(h_t)\left[1+\frac{1}{\nu\tau^2}\{h_{t+1}-[\mu+\phi(h_t-\mu)]\}^2\right]\right\}}\Big)
\end{aligned}
$$

- Full conditional density for λ_{h_t}

$$\begin{aligned}
\lambda_{h_1}|h_1, \boldsymbol{\lambda}_{y_t}, \boldsymbol{y}, \boldsymbol{\lambda}_{h,\setminus t} &\sim Ga\left(\frac{\nu+1}{2}, \frac{1-\phi^2}{2\tau^2}(h_1-\mu)^2+\frac{\nu}{2}\right), \qquad t=1 \\
\lambda_{h_t}|\boldsymbol{h}, \boldsymbol{\lambda}_{y_t}, \boldsymbol{y}, \boldsymbol{\lambda}_{h,\setminus t}, \mu, \phi, \tau^2, \nu &\sim Ga\left(\frac{\nu+1}{2}, \frac{1}{2\tau^2}\{h_{t+1}-[\mu+\phi(h_t-\mu)]\}^2+\frac{\nu}{2}\right), t=2,\ldots,n
\end{aligned}$$

- Full conditional density for μ,

$$\begin{aligned}
f(\mu|\boldsymbol{h}, \boldsymbol{y}, \boldsymbol{\lambda}_y, \boldsymbol{\lambda}_h, \phi, \tau^2, \rho, \nu) &\propto N(\mu|\gamma_\mu m_\mu, \gamma_\mu)\prod_{t=1}^{n} N(y_t|\mu_{y_t}, \frac{\sigma^2_{y_t}}{\lambda_{y_t}}) \\
&\propto \exp\left\{-\frac{1}{2\gamma_\mu}(\mu-\gamma_\mu m_\mu)^2-\frac{1}{2}\sum_{t=1}^{n}\frac{\lambda_{y_t}}{\sigma^2_{y_t}}(y_t-\mu_{y_t})^2\right\} \\
&\quad \times\prod_{t=1}^{n}\left[1+\frac{1}{\nu\tau^2}\{h_{t+1}-[\mu+\phi(h_t-\mu)]\}^2\right]^{-1/2}
\end{aligned}$$

where

$$\begin{aligned}
\gamma_\mu &= \left(\frac{\lambda_{h_1}(1-\phi^2)+(1-\phi)^2\sum_{t=1}^{n}\lambda_{h_{t+1}}}{\tau^2}+\frac{1}{b_\mu}\right)^{-1}, \\
m_\mu &= \frac{\lambda_{h_1}(1-\phi^2)h_1+(1-\phi)\sum_{t=1}^{n}\lambda_{h_{t+1}}(h_{t+1}-\phi h_t)}{\tau^2}+\frac{a_\mu}{b_\mu}.
\end{aligned}$$

- Full conditional density for τ^2,

$$\begin{aligned}
f(\tau^2|\boldsymbol{h}, \boldsymbol{y}, \boldsymbol{\lambda}_y, \boldsymbol{\lambda}_h, \mu, \phi, \rho, \nu) &\propto IG\Big(a_\tau+\tau^2|\frac{n+1}{2}, \\
&\quad b_\tau+\frac{\lambda_{h_1}(1-\phi^2)(h_1-\mu)^2+\sum_{t=1}^{n}\lambda_{h_{t+1}}\{h_{t+1}-[\mu+\phi(h_t-\mu)]\}^2}{2}\Big) \\
&\quad \times\prod_{t=1}^{n} N\left(y_t|\mu_{y_t}, \frac{\sigma^2_{y_t}}{\lambda_{y_t}}\right) \\
&\propto \left(\frac{1}{\tau^2}\right)^{\frac{n+3}{2}+a_\tau}\exp\Big\{-\frac{1}{\tau^2}\Big[b_\tau+\frac{\lambda_{h_1}(1-\phi^2)}{2}(h_1-\mu)^2 \\
&\quad +\frac{1}{2}\sum_{t=1}^{n}\lambda_{h_{t+1}}\{h_{t+1}-[\mu+\phi(h-\mu)]\}^2\Big]-\frac{1}{2}\sum_{t=1}^{n}\frac{\lambda_{y_t}}{\sigma^2_{y_t}}(y_t-\mu_{y_t})^2\Big\} \\
&\quad \times\prod_{t=1}^{n}\left[1+\frac{1}{\nu\tau^2}\{h_{t+1}-[\mu+\phi(h_t-\mu)]\}^2\right]^{-1/2}
\end{aligned}$$

- Full conditional density for ϕ,

$$\begin{aligned} f(\phi|\boldsymbol{h},\boldsymbol{y},\boldsymbol{\lambda}_y,\boldsymbol{\lambda}_h,\mu,\tau^2,\rho,\nu) &\propto N\left(h_1|\mu,\frac{\tau^2}{\lambda_{h_1}(1-\phi^2)}\right)\prod_{t=1}^{n}N\left(y_t|\mu_{y_t},\frac{\sigma^2_{y_t}}{\lambda_{y_t}}\right)N\left(h_{t+1}|\mu_{h_t},\frac{\tau^2}{\lambda_{h_{t+1}}}\right)\\ &\quad\times Be(\phi^*|a_\phi,b_\phi)\\ &\propto \sqrt{1-\phi^2}\left(\frac{\phi+1}{2}\right)^{a_\phi-1}\left(\frac{1-\phi}{2}\right)^{b_\phi-1}\\ &\quad\times\exp\left\{-\frac{\lambda_{h_1}(1-\phi^2)}{2\tau^2}(h_1-\mu)^2-\frac{1}{2}\sum_{t=1}^{n}\left(\frac{\lambda_{y_t}}{\sigma^2_{y_t}}(y_t-\mu_{y_t})^2\right.\right.\\ &\quad\left.\left.-\frac{\lambda_{h_{t+1}}}{\tau^2}(h_{t+1}-\mu_{h_{t+1}})^2\right)\right\}\\ &\quad\times\prod_{t=1}^{n}\left[1+\frac{1}{\nu\tau^2}\{h_{t+1}-[\mu+\phi(h_t-\mu)]\}^2\right]^{-1/2} \end{aligned}$$

- Full conditional density for ρ,

$$\begin{aligned} f(\rho|\boldsymbol{h},\boldsymbol{y},\boldsymbol{\lambda}_y,\boldsymbol{\lambda}_h,\mu,\tau^2,\phi,\nu) &\propto \prod_{t=1}^{n}N\left(y_t|\mu_{y_t},\frac{\sigma^2_{y_t}}{\lambda_{y_t}}\right)I(-1,1)\\ &\propto (1-\rho^2)^{-n/2}\exp\left\{-\frac{1}{2}\sum_{t=1}^{n}\frac{\lambda_{y_t}}{\sigma^2_{y_t}}(y_t-\mu_{y_t})^2\right\}I(-1,1) \end{aligned}$$

- Full conditional density for ν,

$$\begin{aligned} f(\nu|\boldsymbol{h},\boldsymbol{y},\boldsymbol{\lambda}_y,\boldsymbol{\lambda}_h,\mu,\tau^2,\phi,\rho) &\propto Ga\left(\lambda_{h_1}|\frac{\nu}{2},\frac{\nu}{2}\right)\prod_{t=1}^{n}N\left(y_t|\mu_{y_t},\frac{\sigma^2_{y_t}}{\lambda_{y_t}}\right)\\ &\quad\times Ga\left(\lambda_{y_t}|\frac{\nu+1}{2},\frac{\nu+1}{2}\right)Ga\left(\lambda_{h_{t+1}}|\frac{\nu}{2},\frac{\nu}{2}\right)IG(\nu|a_\nu,b_\nu)\\ &\propto \frac{1}{[\Gamma(\frac{\nu}{2})]^{n+1}}\frac{1}{[\Gamma(\frac{\nu+1}{2})]^{n}}\frac{\nu^{\frac{1}{2}(\nu n+\nu-n)}(\nu+1)^{n(\frac{\nu}{2}+1)}}{2^{n\nu+\frac{1}{2}(\nu+n)}}\lambda_{h_1}^{\frac{\nu}{2}-1}\\ &\quad\times\exp\left\{-\frac{b_\nu}{\nu}-\frac{1}{2}\sum_{t=1}^{n}\frac{\lambda_{y_t}}{\sigma_t^2}(y_t-\nu_{y_t})^2-\frac{\nu+1}{2}\sum_{t=1}^{n}\lambda_{y_t}-\frac{\nu}{2}\sum_{t=0}^{n}\lambda_{h_{t+1}}\right\}\\ &\quad\times\prod_{t=1}^{n}\lambda_{h_{t+1}}^{\frac{\nu}{2}-1}\lambda_{y_t}^{\frac{\nu-1}{2}}\prod_{t=1}^{n}\left[1+\frac{1}{\nu\tau^2}\{h_{t+1}-[\mu+\phi(h_t-\mu)]\}^2\right]^{-1/2} \end{aligned}$$

The MCMC algorithms are implemented using the user-friendly software `WinBUGS` which contains an expert system for choosing the best sampling method to sample from these univariate full conditional densities.

In the simulation and empirical studies presented in Section 2.5 and 2.6 respectively, a vague prior is assigned to μ, a uniform prior $U(-1, 1)$ is assigned to ρ, a non-informative prior is assigned to τ^2, a beta $Be(20, 1.5)$ prior distribution is assigned to ϕ^* following Kim *et al.* (1998) and a non-informative prior with restricted range $(1, 40)$ is assigned to ν. In the Gibbs sampling scheme, a single Markov chain is run for 300,000 iterations. We discard the initial 50,000 iterations as the burn-in period to ensure convergence. Simulated values from the Gibbs sampler after the burn-in period are taken from every 100th iteration to avoid high autocorrelation and to mimic a random sample of size 2,500 from the joint posterior distribution for posterior inference. The convergence of the Markov chain is assessed using Geweke's convergence test and Heidelberger and Welch stationary and halfwidth tests described below. This procedure can be performed using the `BOA` (Bayesian Output Analysis) package in the `R` software. Smith (2005) gave a detailed description of the use and the methodology upon which the package is based.

2.4.2. Model selection criterion. For the empirical study in Section 2.6, model comparison is based on the deviance information criterion (DIC), discussed in Section 1.5. Berg *et al.* (2004) gave an account on the use of DIC in SV modelling. Since the DIC value of a model can be automatically computed using `WinBUGS`, it is adopted as a user-friendly model comparison criterion. Other model selection criteria can also be used, for example, Kim *et al.* (1998) calculated Bayes factors using the approach of Chib (1995). However, the calculation requires the marginal likelihood and it becomes computationally demanding when the number of unknown parameters in the model is large.

2.4.3. Convergence diagnostics. Posterior summaries of model parameters are computed from MCMC chains, hence it is important to check that the chains have converged and provide representative samples from the joint posterior distribution. Several diagnostic test have been developed to verify the convergence of MCMC output. In the empirical studies, we will use the following convergence diagnostics.

2.4.3.1. *Geweke convergence diagnostic.* Geweke (1992) suggested assessing the convergence of the MCMC output by comparing the sample mean in an early segment of the

chain with the mean in a later segment. Let $\theta^{(i)}$ be the ith draw of a parameter in a sample of n draws and let $\bar{\theta}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1}\theta^{(i)}$ and $\bar{\theta}_2 = \frac{1}{n_2}\sum_{i=1}^{n_2}\theta^{(i)}$ represent the means of draws in an early segment and a later segment of the chain respectively. Based on this, Geweke (1992) proposed the following convergence diagnostics statistic

$$z = \frac{\bar{\theta}_1 - \bar{\theta}_2}{\sqrt{\hat{s}_1^2/n_1 + \hat{s}_2^2/n_2}},$$

where the variance estimate $\hat{s}$ is calculated as the spectral density at frequency zero to account for serial correlation in the sample output. If the sequence of $\theta^{(i)}$ is stationary, then the limiting distribution of this statistic is a standard normal. Thus a p-value can be computed as evidence against the two sequences being from a stationary distribution. Following Geweke's suggestion, we set $n_1 = 0.1n$ and $n_2 = 0.5n$.

2.4.3.2. *Heidelberger and Welch convergence diagnostic.* The Heidelberger and Welch test (Heidelberger and Welch, 1983) consists of two stages: the stationary test and a half width test. The null hypothesis that the sampled values come from a stationary distribution is based on Brownian bridge theory and the test statistic used is the Cramer-von-Mises test statistic. The entire chain is first used to construct the statistic. If the test is passed, then we can conclude the whole chain is stationary. If the test is failed, then the first $0.1n$ of samples are discarded and the test is reapplied to the remaining chain. This procedure is repeated until either the null hypothesis is accepted or 50% of the samples have been discarded, in which case the chain is taken to be non-stationary. The second stage is the half-width test, which is performed by computing a $(1-\alpha)100\%$ confidence interval for the mean, using the portion of the chain which passed the stationary test in the first stage. This test is passed if the ratio between the half-width of the confidence interval and the estimate of the mean is lower than a specified level of accuracy ϵ, otherwise the length of the sample is not long enough to accurately estimate the mean with $(1-\alpha)$ confidence under tolerance ϵ.

2.5. A Simulation study

To investigate the performance of our proposed t SVL model in (2.9) (Wang's approach), in particular, its ability to provide accurate estimates, we conduct a simulation study in which 100 independent data sets are simulated from the t SVL model with true parameter values $\mu = -10, \tau = 0.2, \phi = 0.8, \rho = 0.8$ and $\nu = 5$. Our simulation study begins with an investigation on the minimal sample sizes needed to obtain decent estimation results. The

t SVL model using our approach is fitted to each data set with sample sizes $n = 100$, 200 and 500 using the `WinBUGS` software the `R2WinBUGS` package. Our results in Table 2.1 indicate that the parameter estimates for ν and ρ incur large percentage error and large mean square errors (defined below) for sample sizes 100 and 200. Based on the results for sample size of 500 observations, we conclude that this is a minimal sample size needed to obtain decent estimation results. Choy's approach (2.7) is also applied to the data sets with sample size of 500. To compare the performance of the two approaches, the following goodness-of-fit measures are used: the percentage error (PE), mean square error (MSE) and posterior coverage (PC), defined respectively by

$$\begin{aligned} PE &= \frac{\widehat{\theta} - \theta}{|\theta|} \times 100\%, \\ MSE &= \frac{1}{N}\sum_{j=1}^{N}(\widehat{\theta}_j - \theta)^2, \\ PC &= \frac{1}{N}\sum_{j=1}^{N} I[\theta \in (\widehat{\theta}_{j,0.025}, \widehat{\theta}_{j,0.975})], \end{aligned}$$

where $\widehat{\theta}$ is the average of the posterior means $\widehat{\theta}_j$, $j = 1, ..., 100$ for each parameter θ and $(\widehat{\theta}_{j,0.025}, \widehat{\theta}_{j,0.975})$ is the 95% credible interval of θ in data set j.

Table 2.2 presents the posterior summaries, PE, MSE and PC for the five parameters under Choy's and our approaches. Firstly, the point estimates under the two approaches are very close to each other and are also very close to their true values. This indicates that the two approaches for estimating the t SVL model through the MCMC algorithm give unbiased and precise estimates as both MSE and standard deviation of most parameters are reasonably small. Generally speaking, point estimates from Choy's approach are marginally closer to the true values than those from our approach but they have larger posterior standard errors and hence wider credible intervals. Secondly, our approach gives a smaller MSE and a better posterior coverage while Choy's method gives a smaller PE. Finally, the estimate for ν, which is the degrees of freedom in the Student-t distribution, has relatively larger standard error, PE and MSE than other parameter estimates using both approaches. This indicates the higher level of difficulty in estimating this particular parameter, which reflects the level of heavy-tailedness in a Student-t distribution due to the insensitivity of the distribution to ν, particularly when it approaches the normal distribution. Overall, the simulation experiment

shows that the performance of our proposed approach for the t SVL model is satisfactory and is comparable with Choy's approach.

TABLE 2.1. Simulation results based on 100 replicates on n = 100, 200 observations using Wang's approach

$n = 100$							
	True	Estimate	sd	95 % CI	PE (%)	MSE	PC (%)
ν	5	9.5710	5.6584	(2.9509, 23.7464)	91.42	15.55	95
μ	-10	-9.9519	0.3501	(-10.6020, -9.2751)	0.48	0.0729	95
ϕ	0.8	0.8361	0.1018	(0.5891, 0.9754)	4.51	0.0052	98
ρ	0.8	0.4427	0.3354	(-0.3036, 0.9199)	-44.66	0.0766	80
τ	0.2	0.2212	0.1294	(0.0647, 0.5446)	10.60	0.0129	96
$n = 200$							
	True	Estimate	sd	95 % CI	PE (%)	MSE	PC (%)
ν	5	7.0001	4.0165	(3.3582, 18.2058)	40.00	15.5283	86
μ	-10	-9.9737	0.1877	(-10.3224, -9.6177)	0.26	0.0398	90
ϕ	0.8	0.8351	0.0935	(0.5937, 0.9507)	4.39	0.0047	97
ρ	0.8	0.2557	0.3354	(-0.0191, 0.9072)	25.80	0.0722	93
τ	0.2	0.1877	0.0946	(0.0841, 0.4404)	-6.13	0.0093	89

TABLE 2.2. Simulation results based on 100 replicates on n = 500 observations.

Par.	Approach	True	Estimate	sd	95% CI	PE (%)	MSE	PC (%)
ν	Wang's	5	5.7937	2.4708	(3.7092, 13.1320)	15.87	2.3335	98
	Choy's	5	5.7159	3.2069	(3.3416, 15.2972)	14.32	7.1341	92
μ	Wang's	-10	-9.9758	0.1133	(-10.1860, -9.7506)	0.24	0.0111	96
	Choy's	-10	-9.9999	0.1600	(-10.2836, -9.6827)	0.00	0.0684	96
ϕ	Wang's	0.8	0.8107	0.0747	(0.6205, 0.9057)	1.35	0.0039	99
	Choy's	0.8	0.8104	0.0831	(0.6064, 0.9209)	1.30	0.0035	98
ρ	Wang's	0.8	0.7136	0.1538	(0.3421, 0.9044)	-10.80	0.0169	95
	Choy's	0.8	0.7146	0.1810	(0.2977, 0.9279)	-10.67	0.0554	83
τ	Wang's	0.2	0.2085	0.0735	(0.1209, 0.3988)	4.24	0.0045	97
	Choy's	0.2	0.1913	0.0833	(0.0885, 0.4018)	-4.34	0.0048	96

2.6. Empirical studies

In this section, we analyse two sets of financial time series data. The data are the mean-corrected log-returns y_t of the daily exchange rate and the daily stock market closing index given by

$$y_t = \ln\left(\frac{P_t}{P_{t-1}}\right) - \frac{1}{n}\sum_{t=1}^{n} \ln\left(\frac{P_t}{P_{t-1}}\right), \qquad t = 1, \ldots, n,$$

where P_t is the exchange rate or the stock market closing index at time t.

2.6.1. Exchange rate data.

2.6.1.1. *Data.* The data used in this section are the mean-corrected daily exchange rate log-returns of the Australian dollar (AUD) against ten other currencies. These currencies are the American Dollar (USD), Euro (EUR), Great British Pound (GBP), Swiss Franc (CHF), Canadian Dollar (CAD), New Zealand Dollar (NZD), Japanese Yen (JPY), Chinese Yuan (CNY), Hong Kong Dollar (HKD) and Singapore Dollar (SGD). The exchange rate data cover the 2-year period from January 2006 to December 2007, consisting of 510 observations. These ten currencies are chosen as they are the most common currencies in foreign exchange in Australia. Although two years is a short time span for daily data, it is employed in order to make direct comparison with the results in Choy *et al.* (2008) as the data were analysed using Choy's approach in Choy *et al.* (2008).

Table 2.3 presents the summary statistics of the data. As shown in the table, the daily log-returns of the AUD to the ten currencies are all mildly negatively skewed and have excess kurtosis. Between the ten currencies, NZD has the smallest kurtosis (3.77) and JPY has the largest kurtosis (13.08). Figure 2.1 shows the time series plots for the ten exchange rate return data. All data sets, except for NZD, display large fluctuation in the returns towards the end of 2007. To demonstrate the superiority of the use of heavy-tailed Student-t distribution in SV modelling, we fit a normal SVL (N SVL) model using Meyer and Yu's approach (2.4) and a t SVL model using our approach (2.9). The estimation results using Choy's approach (2.7) are presented in the same table and we compare all three types of models/approaches.

TABLE 2.3. Summary statistics for exchange rate return data and stock market index return data.

	sd	Max	Min	Skewness	Kurtosis
Currency					
USD	0.00703	0.0258	-0.0400	-0.92	7.00
CAD	0.00576	0.0154	-0.0341	-0.83	6.03
EUR	0.00578	0.0185	-0.0398	-1.12	7.61
GBP	0.00576	0.0159	-0.0273	-0.63	5.40
CHF	0.00695	0.0223	-0.0381	-1.09	7.08
NZD	0.00498	0.0162	-0.0205	-0.04	3.77
JPY	0.00951	0.0381	-0.0643	-1.69	13.08
CNY	0.00685	0.0241	-0.0398	-0.95	6.96
HKD	0.00702	0.0254	-0.0386	-0.91	6.89
SGD	0.00585	0.0199	-0.0358	-0.04	8.44
Market index					
AORD	0.01267	0.0537	-0.0854	-0.65	9.24
CAC	0.01499	0.1061	-0.0946	0.10	14.30
DAX	0.01428	0.1079	-0.0745	0.27	15.68
FTSE	0.01392	0.0939	-0.0926	-0.11	13.72
HSI	0.01908	0.1341	-0.1358	0.06	13.44
NIKKEI	0.01752	0.1326	-0.1208	-0.54	13.74
S&P 500	0.01462	0.1099	-0.0944	-0.33	16.65

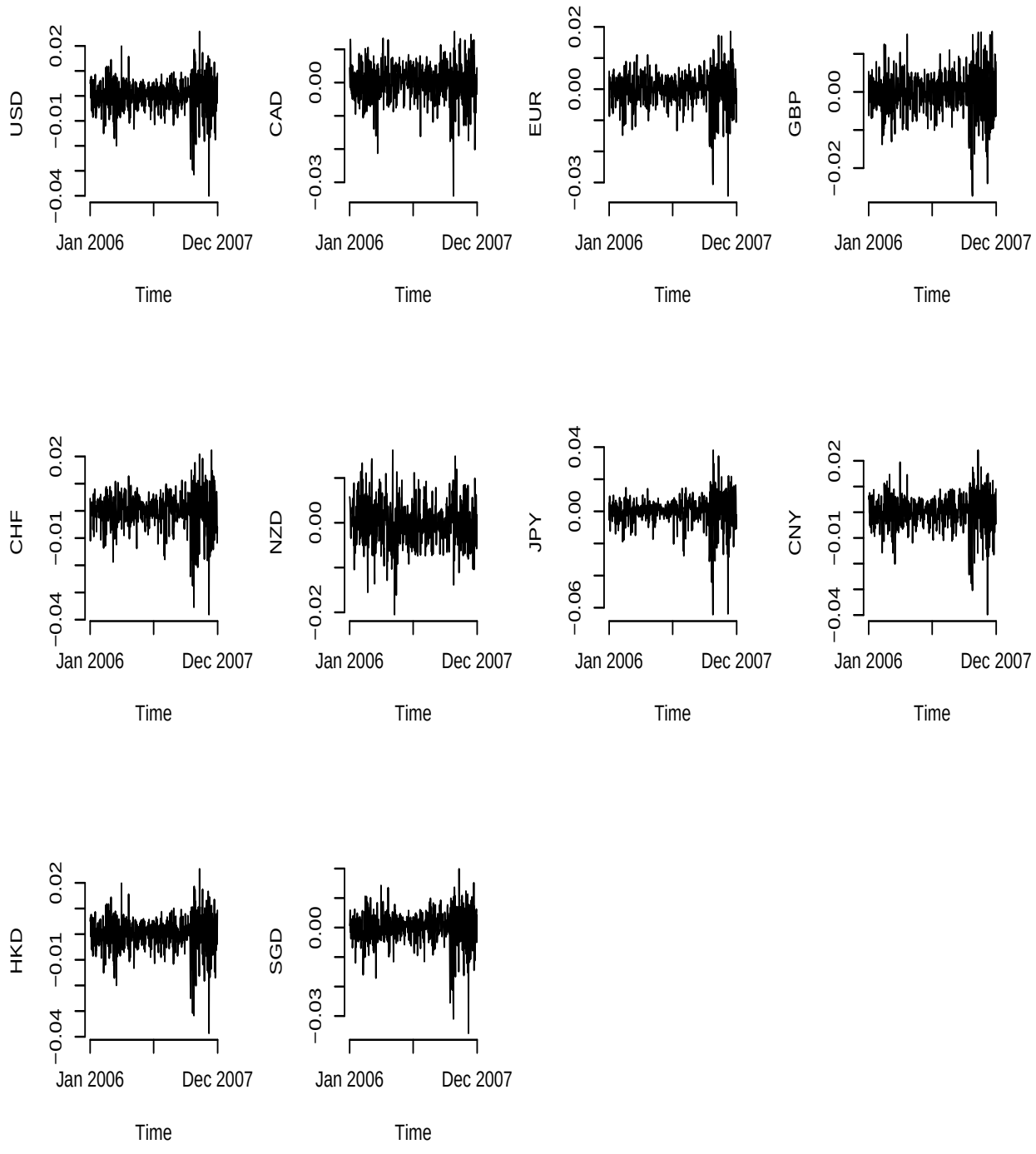


FIGURE 2.1. Time series plots of ten exchange rate return data.

2.6.1.2. *Results.* Table 2.4 displays the posterior means and standard errors of the model parameters in the N SVL and t SVL model using our approach. We focus first on the comparison between N SVL and t SVL models. The estimate of the correlation coefficient, ρ, which is a measure of the leverage effect, ranges from -0.2413 (GBP) to 0.1999 (NZD) for N SVL model and from -0.3754 (CHF) to 0.2734 (NZD) for t SVL model. For both models and all ten currencies, the estimates of ρ are insignificant at the 5% level. This result is not surprising as it is discovered in many literatures that the exchange rate data do not exhibit strong leverage effect. The posterior estimate of the degrees of freedom ν varies between 6.47 (JPY) and 17.98 (GBP). The fact that AUD/JPY has the smallest estimated degrees of freedom is reflected in the summary statistics in which we observe the largest kurtosis for JPY out of the ten currencies. The relatively small estimated degrees of freedom shows the necessity of replacing the conditional normal distribution of the returns by a heavy-tailed distribution, in this case the Student-t distribution, to capture the leptokurtic feature of daily returns more adequately. In fact for model comparison, the smaller DIC values shown in Table 2.4 for the t SVL model indicate a substantial improvement of the t SVL model over the N SVL model for all ten currencies after allowing for model complexity. Hence in the context of SVL models in this analysis, a bivariate Student-t distribution is more appropriate than the bivariate normal distribution in modelling the random error (ϵ_t, η_t). In all ten currencies, the estimates of μ are smaller under t SVL than N SVL model, indicating that the estimated unconditional means of the volatility process are smaller under the t SVL specification.

The estimates of the persistence parameter ϕ are very high and lie between 0.9646 (CAD) and 0.9816 (CNY) for N SVL model, and between 0.9499 (CAD) and 0.9842 (CNY) for t SVL model. This result is in accordance with the well-known stylised feature of asset returns, that volatility is highly persistent. We also notice that the estimates of the persistence parameter under the t SVL model are in general greater than those under the N SVL model (with the exception of CAD and NZD). A general interpretation of this effect is as follows. The t SVL model attributes a larger proportion of extreme returns to the shocks of the mean equation instead of the underlying volatility process than the N SVL model. Hence t SVL model is more likely to capture extreme data as Student-t distributed shocks in the mean equation than the N SVL model. As a result, these shocks are interpreted as non-persistent

since mean shocks in the SV models are assumed to have no influence in the future. On the other hand, the N SVL model captures these non-persistent shocks as shocks in the latent volatility process, which results in a decrease in the estimated persistence parameter relative to the t SVL model.

Relating to the above result regarding the persistence parameter is the finding that the estimates of the standard deviation, τ, of the log-volatility are smaller under the t SVL models than those under the N SVL model in all ten data sets. This is due to the different treatment of the t SVL and N SVL models on the highly leptokurtic return distribution. The t SVL model uses a heavier-tailed error distribution which effectively captures the data in the extreme tail areas. The N SVL model on the other hand, accounts for the high kurtosis by inflating the variation in the latent volatility process, resulting in a higher estimated variance parameter for the volatility. This is confirmed by the empirical result that the unconditional variance of the underlying volatility process, $\tau^2/(1-\phi^2)$, calculated from the posterior parameter estimates, is larger under the N SVL model than under the t SVL model. This is shown in the last column of Table 2.4. The smaller unconditional variance of the volatility process under the t SVL model may also explain the smaller Monte Carlo sampling errors of most parameter estimates under t SVL model relative to the N SVL specification. With the smaller unconditional variance for the latent volatility, simulated values of the latent volatility under the t SVL model have reduced variation, which in turn results in a decrease in the Monte Carlo sampling error for the estimates of associated parameters under the t SVL model compared to the N SVL model.

Next, our estimation results are compared to those obtained by Choy's approach. Our approach and Choy's approach indeed produce similar parameter estimates although the latter uses a slightly different prior distribution for the degrees of freedom ν and the numbers of iterations and posterior samples are much smaller. One point to note is that Choy *et al.* (2008) found significant leverage effect indicated by the significant estimates of ρ for currencies EUR and CHF, which is at odds with our results and the general belief that exchange rate data do not show profound leverage effect.

To illustrate the difference of the estimated latent volatilities under the t SVL and N SVL specifications and to perform outlier diagnostics, we choose the AUD/JPY data for the following reasons. Firstly, the estimated degrees of freedom for this data set is the smallest

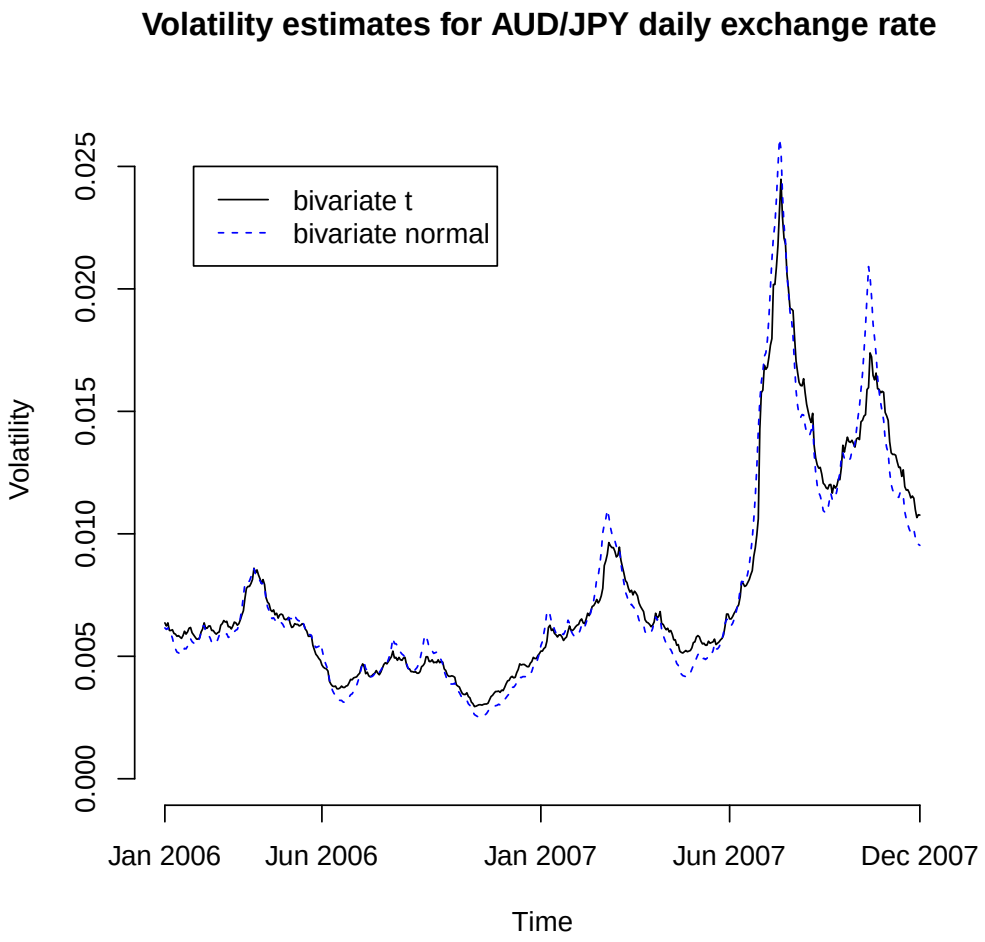


FIGURE 2.2. Posterior means of return volatility under t SVL and N SVL models for the AUD/JPY exchange rate return data.

among the ten exchange rates. The Student-t distribution with a smaller degrees of freedom has a stronger robustness property and a stronger potential to identify outliers. Secondly, it facilitates comparison with Choy *et al.* (2008) to demonstrate the improvement of our approach in outlier diagnostics.

For the AUD/JPY data, the posterior means of volatilities $\exp(h_t/2)$ under the t SVL and N SVL models are displayed in Figure 2.2. Note that for the t SVL model, we do not have $var(y_t|h_t) = \exp(h_t)$ since the Student-t distribution we used is not standardised. In order to compare return volatilities across different models, we need to adjust for the variance of the Student-t distribution. That is, we plot $\exp(h_t/2)\sqrt{\frac{\nu}{\nu-2}}$ where ν is the posterior mean of the degrees of freedom for the t SVL model. As shown in the graph, the posterior means of volatilities from the t SVL model exhibit smoother movement than those from the N SVL

model. The differences are clearer at extreme returns such as the returns at the great financial crisis towards the end of the time series. The volatilities associated with these extreme returns are greater under the N SVL model than under the t SVl model, since the t model partially captures these extreme return observations as large shocks in the mean process.

TABLE 2.4. Parameter estimates (standard errors in parentheses) of the N SVL and t SVL models for ten exchange rate return data.

Currency	Model	DIC	μ	ϕ	ρ	τ	ν	$\tau^2/(1-\phi^2)$
USD	N SVL	-3719.4	-10.24	0.9815	0.1555	0.1489	-	0.605
			(0.59)	(0.0118)	(0.1786)	(0.0349)		
	t SVL	-3727.1	-10.34	0.9827	0.0996	0.1268	16.66	0.469
			(0.56)	(0.0102)	(0.1947)	(0.0270)	(5.67)	
	Choy t SVL	-3732.5	-10.44	0.9822	0.15	0.1295	16.1	
			(0.32)	(0.0108)	(0.21)	(0.0273)	(5.7)	
CAD	N SVL	-3835.4	-10.43	0.9646	-0.0789	0.1244	-	0.223
			(0.36)	(0.0305)	(0.2191)	(0.0443)		
	t SVL	-3879.3	-10.61	0.9499	-0.1734	0.1069	12.42	0.117
			(0.26)	(0.0450)	(0.2527)	(0.0372)	(5.59)	
	Choy t SVL	-3925.8	-10.70	0.9551	-0.28	0.0988	10.3	
			(0.19)	(0.0330)	(0.27)	(0.0306)	(4.9)	
EUR	N SVL	-3919.9	-10.60	0.9746	-0.2189	0.1654	-	0.545
			(0.48)	(0.0134)	(0.1720)	(0.0348)		
	t SVL	-4088.1	-10.77	0.9766	-0.3657	0.1359	13.24	0.399
			(0.41)	(0.0127)	(0.1943)	(0.0314)	(5.73)	
	Choy t SVL	-3979.5	-10.86	0.9765	-0.38*	0.1309	11.0	
			(0.27)	(0.0121)	(0.19)	(0.0305)	(5.5)	
GBP	N SVL	-3905.2	-10.58	0.9797	-0.2413	0.1450	-	0.523
			(0.46)	(0.0116)	(0.1870)	(0.0305)		
	t SVL	-3957.2	-10.71	0.9836	-0.2906	0.1211	17.98	0.451
			(0.54)	(0.0101)	(0.2195)	(0.0280)	(5.59)	
	Choy t SVL	-3933.0	-10.85	0.9843	-0.32	0.1241	17.0	
			(0.33)	(0.0100)	(0.21)	(0.0270)	(6.0)	
CHF	N SVL	-3750.1	-10.25	0.9734	-0.2337	0.1795	-	0.614
			(0.44)	(0.0141)	(0.1557)	(0.0422)		
	t SVL	-3782.0	-10.50	0.9754	-0.3754	0.1434	11.26	0.423
			(0.43)	(0.0132)	(0.1870)	(0.0341)	(4.89)	
	Choy t SVL	-3798.2	-10.57	0.9771	-0.34*	0.1443	11.6	
			(0.30)	(0.0121)	(0.19)	(0.0347)	(5.5)	

Currency	Model	DIC	μ	ϕ	ρ	τ	ν	$\tau^2/(1-\phi^2)$
NZD	N SVL	-3937.2	-10.72	0.9660	0.1999	0.1179	-	0.208
			(0.28)	(0.0208)	(0.1754)	(0.0295)		
	t SVL	-4001.6	-10.82	0.9653	0.2734	0.1118	17.91	0.183
			(0.25)	(0.0198)	(0.2054)	(0.0272)	(5.71)	
	Choy t SVL	-3964.6	-10.85	0.9682	0.25	0.1087	17.5	
			(0.20)	(0.0191)	(0.20)	(0.0266)	(5.8)	
JPY	N SVL	-3582.6	-9.95	0.9762	-0.0961	0.2316	-	1.140
			(0.62)	(0.0124)	(0.1503)	(0.0460)		
	t SVL	-3643.3	-10.27	0.9823	-0.2561	0.1459	6.38	0.607
			(0.62)	(0.0095)	(0.1873)	(0.0340)	(2.39)	
	Choy t SVL	-3639.3	-10.39	0.9830	-0.23	0.1408	6.0	
			(0.38)	(0.0091)	(0.18)	(0.0354)	(2.7)	
CNY	N SVL	-3735.3	-10.30	0.9816	0.1486	0.1516	-	0.630
			(0.57)	(0.0115)	(0.1864)	(0.0339)		
	t SVL	-3762.9	-10.39	0.9842	0.1417	0.1249	16.81	0.498
			(0.60)	(0.0091)	(0.2015)	(0.0279)	(5.86)	
	Choy t SVL	-3762.7	-10.51	0.9846	0.19	0.1243	14.8	
			(0.32)	(0.0093)	(0.20)	(0.0281)	(5.8)	
HKD	N SVL	-3719.8	-10.24	0.9798	0.1124	0.1592	-	0.634
			(0.52)	(0.0114)	(0.1685)	(0.0290)		
	t SVL	-3751.2	-10.36	0.9831	0.1061	0.1272	16.50	0.483
			(0.61)	(0.0108)	(0.1968)	(0.0298)	(6.00)	
	Choy t SVL	-3742.7	-10.45	0.9828	0.17	0.1226	15.0	
			(0.33)	(0.0102)	(0.20)	(0.0286)	(6.1)	
SGD	N SVL	-3912.3	-10.66	0.9713	0.0220	0.1895	-	0.635
			(0.49)	(0.0185)	(0.1696)	(0.0539)		
	t SVL	-3944.0	-10.87	0.9802	-0.0533	0.1262	11.10	0.406
			(0.52)	(0.0120)	(0.2186)	(0.0291)	(4.46)	
	Choy t SVL	-3960.9	-11.09	0.9806	-0.08	0.1242	10.4	
			(0.37)	(0.0125)	(0.24)	(0.0317)	(4.6)	

A asterisk * indicates that ρ is significant at the 5% level.

For outlier detection, Choy *et al.* (2008) found six pairs of possible outliers which are (y_t, h_{t+1}) for day t, t=173, 467, 104, 254, 294 and 394. Since Choy's approach only involves a single mixing parameter λ_t, it is unclear whether these pairs are outlying in the

return, volatility or both. Alternatively, our approach uses two mixing parameters, λ_{y_t} and λ_{h_t}, to identify outlying returns and unobserved outlying volatilities separately. From Table 2.5, our approach identifies that y_{173} (8 Sep 2006), y_{476} (12 Nov 2007), y_{104} (1 Jun 2006), y_{254} (5 Jan 2007), y_{294} (5 Mar 2007) and h_{395} (27 July 2007) are outlying, indicating that, the first five outliers found by Choy's approach are outlying in return only and the sixth outlier is outlying in volatility only. Hence this shows the greater flexibility in our approach in identifying the sources of the potential outliers as compared to Choy's approach. Figure 2.3 displays the posterior means of λ_{y_t} and λ_{h_t}. This graph shows that the first five outliers have the most outlying returns while the sixth outlier has the largest log-volatility.

TABLE 2.5. Bayes estimates of the mixing parameters for possible outliers: AUD/JPY data and S&P 500 data.

Day t	Wang's		Choy's
	λ_{y_t}	$\lambda_{h_{t+1}}$	λ_t
	AUD/JPY		
Day 173 (8 Sep 2006)	2.387	1.189	4.055
Day 467 (12 Nov 2007)	2.213	1.353	4.574
Day 104 (1 Jun 2006)	2.151	1.178	3.693
Day 254 (5 Jan 2007)	2.098	1.239	4.181
Day 294 (5 Mar 2007)	1.940	1.200	3.397
Day 394 (27 July 2007)	1.371	1.583	2.105
	S&P 500		
Day 542 (27 Feb 2007)	1.299	2.610	3.057
Day 943 (29 Sep 2008)	1.048	1.751	2.263

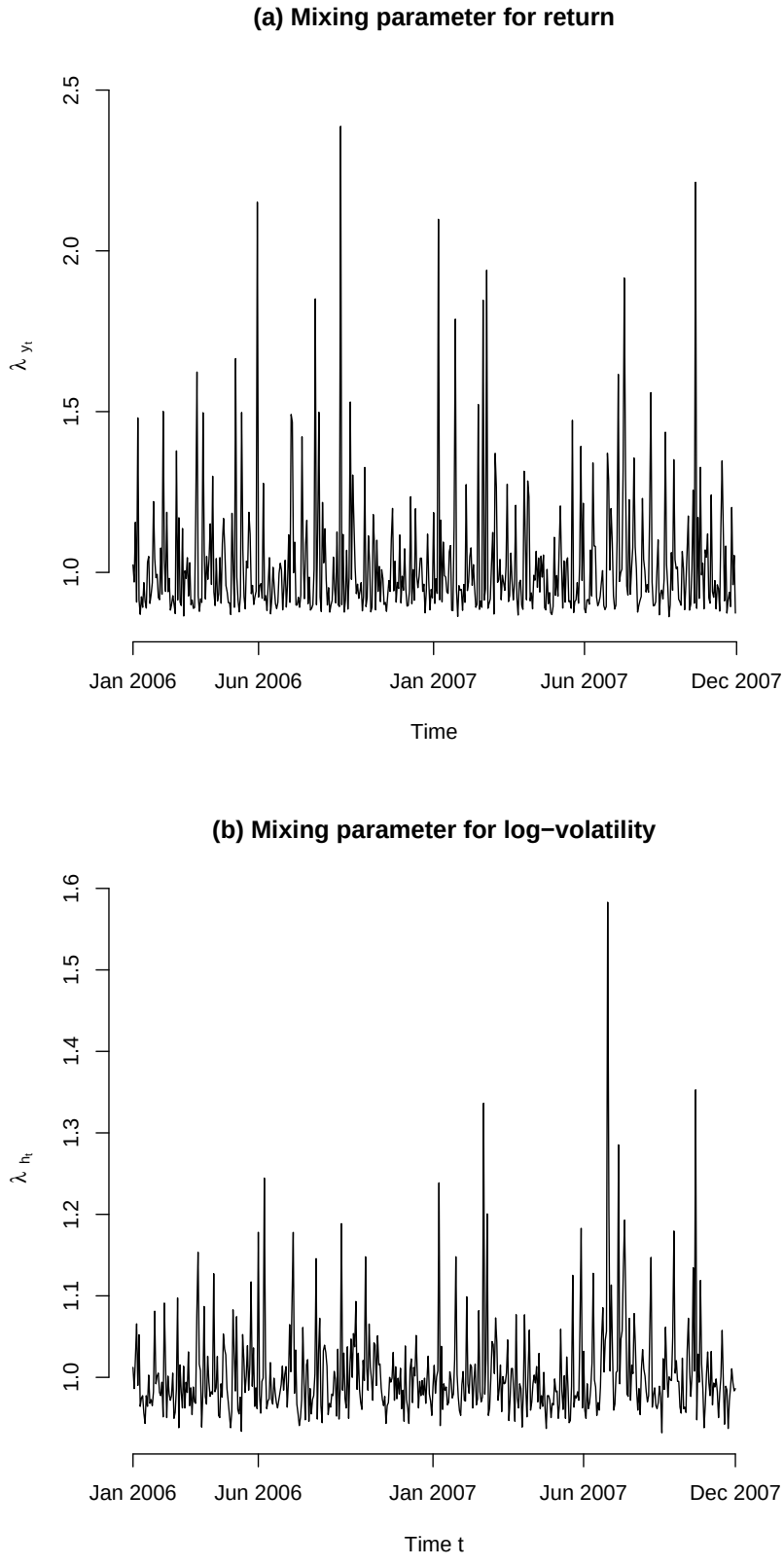


FIGURE 2.3. Outlier diagnostics using (a) mixing parameter, λ_{y_t}, for return and (b) mixing parameter λ_{y_t}, for volatility in our proposed approach for the AUD/JPY exchange rate return data. Large values are associated with possible outliers.

Lastly, we present results for the convergence tests for the Markov chains in the MCMC algorithm. The convergence is assessed using Geweke's convergence test and Heidelberger and Welch stationary and halfwidth tests described in Section 2.4.3. The p-values of Geweke's convergence test for all parameters are greater than the 5% level, indicating that the Geweke diagnostic does not provide evidence of non-convergence for the Markov chains in the MCMC algorithm. For the AUD/JPY exchange rate data, the p-values of the Geweke's convergence test for the N SVL and t SVL models are reported in the first and second rows of Table 2.6.

TABLE 2.6. p-values for Geweke's convergence test for selected model parameters: AUD/JPY data S&P 500 data.

Dataset	μ	ϕ	ρ	τ	ν	h_{100}	λ_{100}	$\lambda_{y_{100}}$	$\lambda_{h_{100}}$
AUD/JPY N SVL	0.092	0.504	0.068	0.092	-	0.200	-	-	-
AUD/JPY t SVL	0.531	0.840	0.549	0.478	0.181	0.524	-	0.240	0.425
S&P 500 Wang	0.476	0.887	0.072	0.360	0.867	0.375	-	0.410	0.568
S&P 500 Choy	0.868	0.568	0.522	0.159	0.886	0.177	0.529	-	-

The results of Heidelberger and Welch stationarity and halfwidth tests for the AUD/JPY data set is given below. The results indicate that all iterations are retained for posterior inference and none are discarded as burn-in sequence. Also, for all model parameters, including the unobserved volatilities and mixing parameters, there is no significant evidence of non-stationarity in the retained iterations, hence all parameters passed the Heidelberger and Welch stationarity test. Furthermore, the results from the halfwidths test for the means of all parameters are less than the specified level of 10%. Hence the MCMC outputs from the t SVL and N SVL models also pass the halfwidth test.

Heidelberger and Welch stationarity and interval halfwidth tests

Halfwidth test accuracy = 0.1

Model: N SVL					
	Stationarity Test	Keep	Discard	Cramer-von-Mises statistic	Halfwidth Test
μ	passed	2500	0	0.12620381	passed
ϕ	passed	2500	0	0.09334390	passed
ρ	passed	2500	0	0.14684249	passed
τ	passed	2500	0	0.23102779	passed
h_{100}	passed	2500	0	0.05370562	passed
	Halfwidth				
μ	0.0170186491				
ϕ	0.0008232576				
ρ	0.0100620667				
τ	0.0048114873				
h_{100}	0.0256341736				

Model: t SVL					
	Stationarity Test	Keep	Discard	Cramer-von-Mises statistic	Halfwidth Test
ν	passed	2500	0	0.17528169	passed
$\lambda_{y_{100}}$	passed	2500	0	0.36551358	passed
$\lambda_{h_{100}}$	passed	2500	0	0.10232375	passed
μ	passed	2500	0	0.18183392	passed
ϕ	passed	2500	0	0.05790712	passed
ρ	passed	2500	0	0.15843654	passed
τ	passed	2500	0	0.14699065	passed
h_{100}	passed	2500	0	0.06163448	passed
	Halfwidth				
ν	0.054141350				
$\lambda_{y_{100}}$	0.028555804				
$\lambda_{h_{100}}$	0.037530867				
μ	0.051664816				
ϕ	0.001145524				
ρ	0.023341044				
τ	0.005171355				
h_{100}	0.053296478				

2.6.2. Stock market index data.

2.6.2.1. *Data.* In the analysis of stock market data, many studies reveal that a negative return generally results in a larger volatility in the future than a positive return of the same magnitude. In this empirical study, we analyse the mean-corrected daily log-returns of seven international stock market indices, namely the Standard & Poors 500 (S&P 500), Financial Times Stock Exchange 100 (FTSE), Deutscher Aktien (DAX), Cotation Assiste en Continu 40 (CAC), All Ordinaries (AORD), Nikkei 225 (Nikkei) and Hang Seng Index (HSI) indices, covering the period from January 2005 to December 2008. The summary statistics in Table 2.3 show that the daily log-returns of all indices have excess kurtosis and have generally larger kurtosis as compared to the exchange rate data. The time series plots of seven market index return data are displayed in Figure 2.4. In all markets, we observe period of very high volatility in the returns in the year of 2008. This is most likely due to the late-2000s financial crisis that began in the United States in late 2007. The financial crisis consisted of a significant decline in the global economic activity, including downturns in stock markets around the world. This results in a severe global economic recession in 2008.

For each stock market index, the dataset is fitted by a t SVL model using both Choy's and our approaches and we focus on the comparison between the two approaches. We do not fit the N SVL model to the market index data as the superiority of the t SVL model has already been demonstrated in the previous empirical study when analysing data with excess kurtosis.

2.6.2.2. *Results.* Table 2.7 reports the posterior means, standard errors and the DIC of model parameters under the two approaches. The estimates of the model parameters under Choy's and our approach are very similar. The percentage difference between the two approaches is less than 5% for the parameters μ and τ and less than 1% for ϕ. For ρ, the percentage difference is less than 4% except for the AORD and HSI data which are 6.7% and 13.3%, respectively. The estimates for the persistence parameter ϕ are greater than 0.96 for all market indices, under both approaches. This indicates that the volatility for the market index is highly persistent. Comparing the estimates of τ, t SVL model using our approach generally gives smaller estimates, except for HSI and S&P 500. This shows under our approach, the estimated latent volatility has slightly less variation than that under Choy's

approach. For the degrees of freedom ν, the differences between the estimates measured in statistical standard errors from our approach are within 0.76.

Asai (2008) also analysed the daily S&P 500 index return using a SVL model with Student-t distributed innovations for the returns and normal innovations for the volatilities. Comparing our result with Asai (2008), we notice that the two studies give a quite consistent estimate of the persistence of volatility ϕ. However, we obtain higher estimate for the degrees of freedom ν than the one found in Asai (2008). Note that no direct comparison can be made because Asai (2008) considered a longer time series which had no overlap with the time series used in our study. Since the estimates of ρ for the seven indices are significantly negative, our finding is in agreement with the fact that the leverage effect is more profound in stock market index return data (Harvey and Shephard, 1996). The DIC evaluated for the t SVL model under Choy's and our approaches are quite close, in most cases within a difference of 2.5% except for the S&P 500 index which is 3.1% and the Nikkei 225 index which is 5.8%. In five out of the seven market indices, the DICs under our approach are smaller than those under Choy's approach, indicating a better fit of the model for those data sets.

For the S&P 500 index data, the smoothed estimates of the return volatilities, $\exp(h_t/2)\sqrt{\frac{\nu}{\nu-2}}$, under Choy's and our approaches are shown in Figure 2.5. We observe that the estimated volatilities from the two approaches are comparable with no major difference between them because the differences in parameter estimates are small. The plot of the estimated volatilities in Figure 2.5 shows that the market was very volatile in September and October 2008 because of the US financial tsunami.

Regarding outlier diagnostics, Figure 2.6 plots the posterior means of the mixing parameter λ_t of Choy's approach and λ_{y_t} and λ_{h_t} under our approach. By comparing the magnitude of the mixing parameters, Choy's method successfully identifies two most significant outliers, Day 542 and Day 943. On February 27, 2007 (Day 542), the S&P 500 index tumbled 3.5%, its worst one-day percentage loss since March 2003, due to the worries of the global economic growth. On September 29, 2008 (Day 943), the US Congress unexpectedly rejected the President's financial rescue plan and the S&P 500 index fell 107 points or 8.8%. Moreover, our approach further reveals that the data on February 27, 2007 is outlying in both return and volatility while the data on September 29, 2008 is outlying in volatility only (see Table 2.5).

For the S&P 500 index data, the last two row of Table 2.6 presents the result of Geweke's convergence test for t SVL model under Choy's and our approaches. Since all tests are insignificant, we conclude the constructed Markov chain converges for all parameters. The Heidelberger and Welch stationarity and halfwidth diagnostics of the MCMC output for S&P 500 data using our approach is given below. The results from BOA package indicate that all parameters pass the stationarity and halfwidth tests at the 10% level.

Heidelberger and Welch stationarity and interval halfwidth tests

Halfwidth test accuracy = 0.1

Model: t SVL using Wang's approach					
	Stationarity Test	Keep	Discard	Cramer-von-Mises statistic	Halfwidth Test
ν	passed	2500	0	2.47236999	passed
$\lambda_{y_{100}}$	passed	2500	0	0.07809373	passed
$\lambda_{h_{100}}$	passed	2500	0	0.09451190	passed
μ	passed	2500	0	0.21362299	passed
ϕ	passed	2500	0	0.10274592	passed
ρ	passed	2500	0	0.33109281	passed
τ	passed	2500	0	0.13847219	passed
h_{100}	passed	2500	0	0.08142537	passed
	Halfwidth				
ν	0.9143959117				
$\lambda_{y_{100}}$	0.0238671326				
$\lambda_{h_{100}}$	0.0266779280				
μ	0.0341524006				
ϕ	0.0008881347				
ρ	0.0185198309				
τ	0.0046199442				
h_{100}	0.0456702867				

TABLE 2.7. Parameter estimates (standard errors in parentheses) of the t SVL model for the stock market index data under Wang's and Choy's approaches.

Market Index	Approach	DIC	μ	ϕ	ρ	τ	ν
AORD	Wang's	-7169.5	-9.253	0.9663	-0.8315*	0.2018	21.94
			(0.130)	(0.0076)	(0.0647)	(0.0298)	(4.70)
	Choy's	-7097.0	-9.421	0.9627	-0.7795*	0.2119	18.38
			(0.149)	(0.0085)	(0.0613)	(0.0299)	(6.06)
CAC	Wang's	-7421.3	-9.090	0.9689	-0.9042*	0.2123	19.37
			(0.127)	(0.0068)	(0.0448)	(0.0284)	(5.33)
	Choy's	-7607.9	-9.211	0.9675	-0.8905*	0.2145	16.85
			(0.123)	(0.0062)	(0.0454)	(0.0243)	(5.62)
DAX	Wang's	-7038.6	-9.193	0.9665	-0.7988*	0.1903	15.06
			(0.151)	(0.0087)	(0.0752)	(0.0310)	(5.45)
	Choy's	-6942.5	-9.293	0.9635	-0.7941*	0.1975	13.91
			(0.146)	(0.0076)	(0.0661)	(0.0240)	(5.68)
FTSE	Wang's	-7242.6	-9.309	0.9808	-0.8215*	0.1833	20.24
			(0.215)	(0.0046)	(0.0623)	(0.0212)	(5.20)
	Choy's	-7220.6	-9.517	0.9802	-0.8210*	0.1848	17.02
			(0.190)	(0.0046)	(0.0748)	(0.0208)	(5.75)
HSI	Wang's	-5935.5	-8.648	0.9874	-0.4531*	0.1549	16.63
			(0.578)	(0.0056)	(0.1087)	(0.0272)	(5.75)
	Choy's	-5893.6	-9.078	0.9879	-0.3998*	0.1486	15.09
			(0.379)	(0.0058)	(0.1179)	(0.0271)	(5.67)
N225	Wang's	-6394.4	-8.770	0.9744	-0.6701*	0.1948	19.80
			(0.226)	(0.0077)	(0.0666)	(0.0291)	(5.41)
	Choy's	-6044.9	-8.947	0.9721	-0.6502*	0.1980	18.77
			(0.206)	(0.0083)	(0.0759)	(0.0292)	(5.58)
S&P 500	Wang's	-7331.3	-9.316	0.9822	-0.8467*	0.1682	11.84
			(0.231)	(0.0048)	(0.0654)	(0.0223)	(4.08)
	Choy's	-7561.3	-9.527	0.9819	-0.8751*	0.1637	10.16
			(0.191)	(0.0049)	(0.0572)	(0.0234)	(3.92)

A asterisk * indicates that ρ is significant at the 5% level.

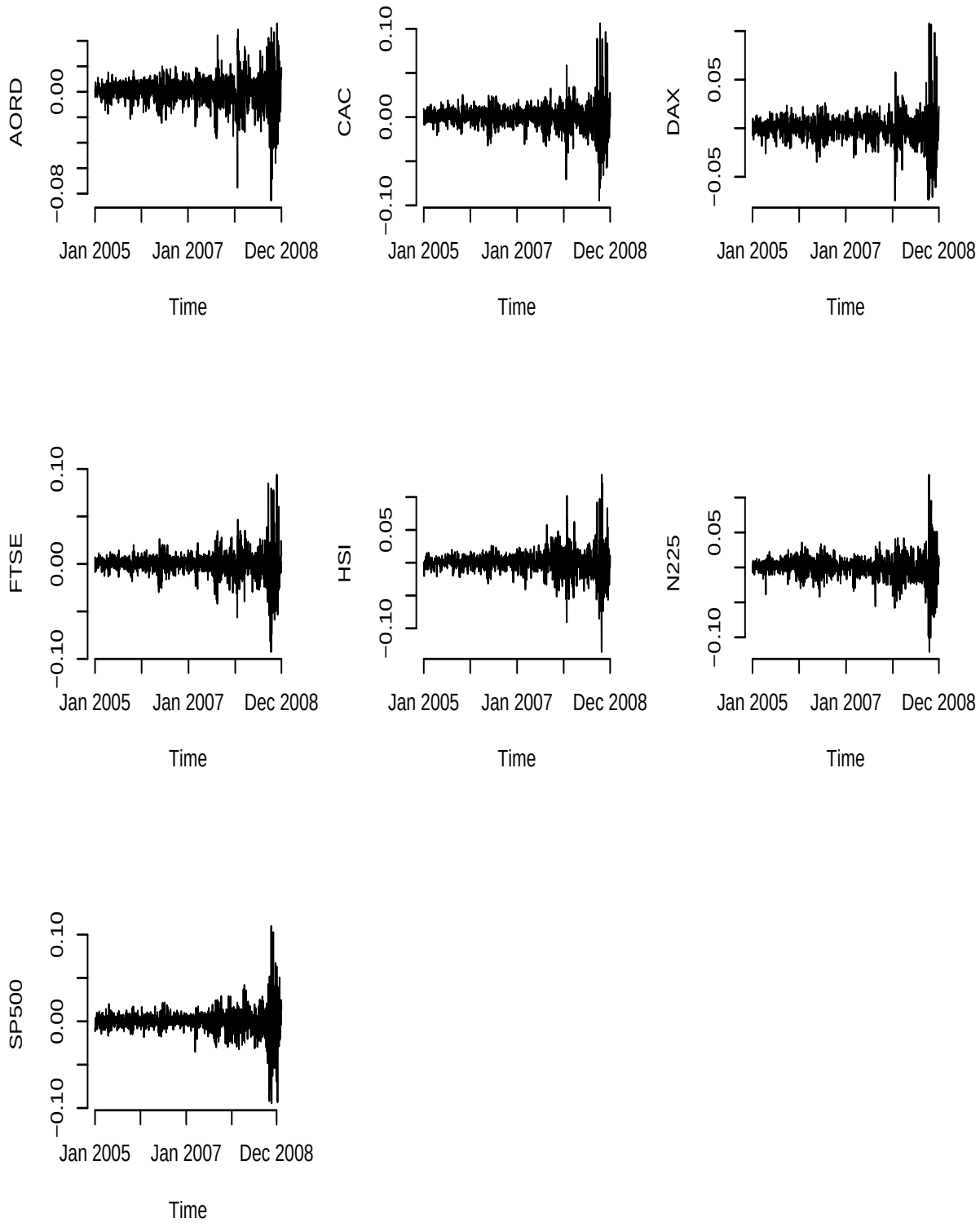


FIGURE 2.4. Time series plots of seven stock market index return data.

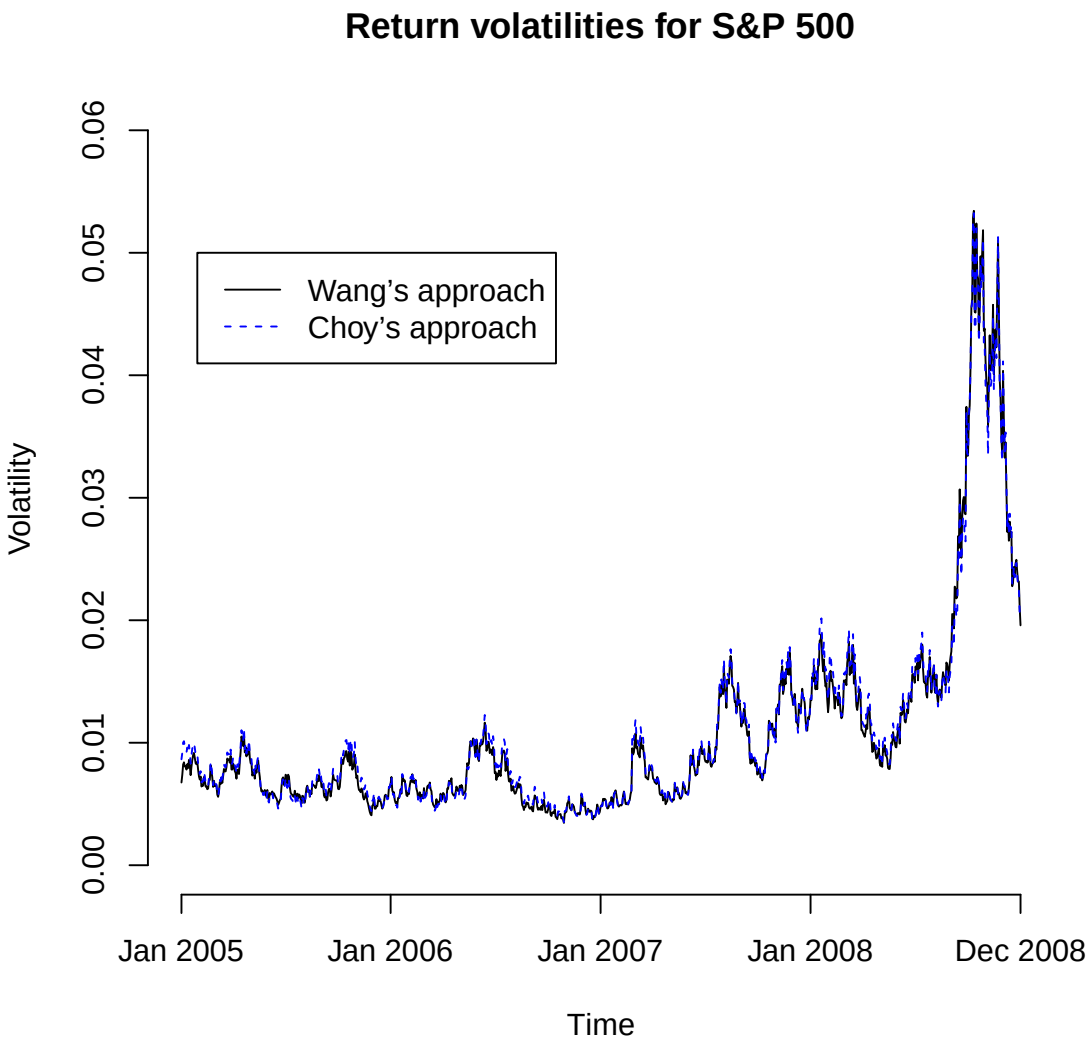


FIGURE 2.5. Comparison of Choy's and our approaches in the estimation of return volatilities under a t SVL model for the S&P 500 index return data.

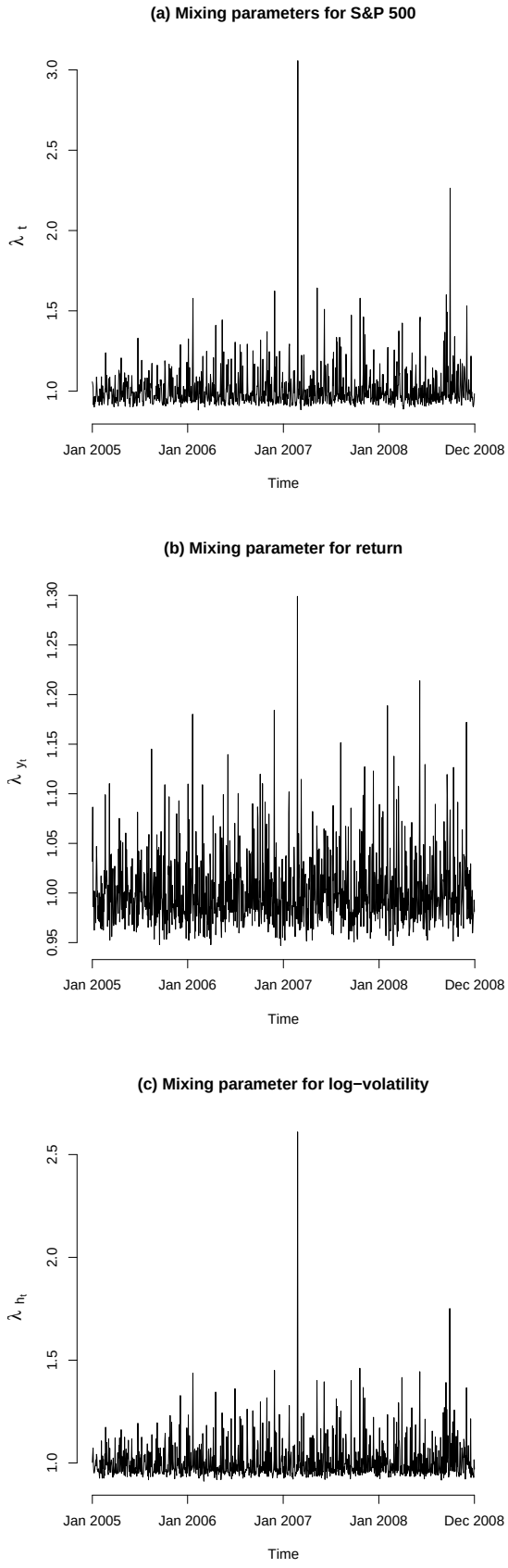


FIGURE 2.6. Outlier diagnostics using (a) λ_t in Choy's approach and (b) λ_{y_t} and (c) λ_{h_t} in our approach for the S&P 500 index return data. Large values are associated with possible outliers.

2.7. Discussion

The stochastic volatility model with leverage (SVL) is an extension of the basic SV model to account for the leverage effect observed in some financial time series. Following the formulation by Harvey and Shephard (1996), the leverage effect is modelled by a bivariate distribution for the innovations of the return at time t and the log-volatility at time $t+1$. Meyer and Yu (2000) analysed the SVL model with normal error distribution using `WinBUGS`. To facilitate implementation of the model, they expressed the joint distribution of (y_t, h_{t+1}) as the product of the conditional distribution $y_t|h_{t+1}, h_t$ and the marginal distribution of $h_{t+1}|h_t$. Choy *et al.* (2008) adopted this idea and extended the normal SVL model to t SVL model for robustness consideration. The key contribution of Choy *et al.* (2008) is that they expressed the Student-t distribution as a scale mixture of normals (SMN). By doing so, not only the full conditional distributions for the model parameters can be simplified to mostly standard distributions, but the mixing parameters, as the by-product of the SMN form, can be used as proxies for outlier diagnostics.

In this chapter, we propose an alternative formulation for the t SVL model of Choy *et al.* (2008). Although both Choy's and our approaches express the Student-t distribution as a SMN and that we also use the mixing parameters in the SMN representation to serve the purpose of identifying possible outliers, the SMN procedure is done at different stages of the model development process. Both Choy's and our approaches, being the extension of the formulation of the normal SVL model analysed by Meyer and Yu (2000), allow Bayesian analysis of the t SVL model to be straightforwardly implemented using the `WinBUGS` software. A simulation study confirms that both Choy's and our approaches give very close and accurate results but our approach achieves a smaller standard error in the parameter estimates, a smaller MSE and a higher posterior coverage than Choy's approach. In the empirical study of exchange rate return data, we confirm that a bivariate Student-t distribution is more appropriate than a bivariate normal distribution in modelling the random error terms in the return and volatility equations. Furthermore, we confirm that the exchange rate returns of AUD to ten other currencies do not exhibit significant leverage effect. In the empirical study of stock market return data, all seven stock market indices show a very high degree of persistence in volatility and a significant leverage effect. In identifying possible outliers in

the return and volatility equations, Choy's approach can only identify outlying pairs of return and volatility but our approach can further identify these pairs to be outlying in return or volatility or both. As to conclude, our proposed approach is a new and preferable approach for analysing Bayesian t SVL models.

Furthermore, it is interesting to investigate other heavy-tailed distributions in the application of SVL models and compare the model performance with the t SVL model. For example, the exponential power (EP) distribution which can accommodate both leptokurtic and platykurtic shapes is a good alternative to consider. Gómez (1998) proposed the following multivariate EP distribution with density

$$f(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \beta) = k|\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left[(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]^{\beta}\right\},$$

where $\boldsymbol{\mu}$ is the location vector, $\boldsymbol{\Sigma}$ is a $(n \times n)$ positive definite symmetric matrix, $\beta \in (0, \infty)$ is the shape parameter and

$$k = \frac{n\Gamma(\frac{n}{2})}{\pi^{\frac{n}{2}}\Gamma(1+\frac{n}{2\beta})2^{1+\frac{n}{2\beta}}}.$$

To use our method, we need to calculate the conditional distribution of the bivariate EP distribution. However, the conditional distribution of EP is an elliptically contoured distribution, and is no longer an EP distribution. Hence to use our method, we need to use special methods such as 'onetrick' or 'zerotrick' in `WinBUGS` which allow us to specify the exact form of the density for the conditional distribution. Alternatively, since the EP distribution can be expressed as a scale mixture of uniform (SMU) distributions (Fung and Seneta, 2008), we can express the bivariate EP as a bivariate SMU, and proceed with the calculation using Choy's approach. This work is currently under progress and has been deferred to future research.

CHAPTER 3

SV Model with Generalised t Distribution

3.1. Background

In Chapter 2, we used a heavy-tailed Student-t distribution as the error distribution for the returns and showed it outperforms the Gaussian error distribution. In fact in financial time series modelling, not limited to SV models, the Student-t distribution and the exponential power (EP) distributions are commonly used since it has been long recognised that risky-asset returns are leptokurtic. This has motivated us to search for more flexible error distributions for the returns. More specifically, we aim to employ a general distribution which nests other commonly used distributions, such that instead of fitting different distributions to the data and compare their relative performance, we fit a single distribution and we allow the data to choose the distribution to adopt through different parameter values. It is known that the commonly used Student-t and the EP distributions belong to a broader family of distributions — the generalised t (GT) distribution. In this Chapter, we explore the use of the GT distribution in SV models and compare the empirical performance of its nested distributions.

The GT distribution was first introduced by McDonald and Newey (1988) as an error distribution in the context of regression analysis. Using the GT distribution, they developed a robust partially adaptive estimation procedure which includes least squares, least absolute deviations and some other estimators. Arslan and Genç (2003) considered the GT distribution as an alternative robust model for modelling the errors in a location-scale estimation problem. They showed that the maximum likelihood estimators for the location and scale parameters of a GT distribution with known shape parameters provide alternative robust estimators for the location and scale parameters of a dataset.

Since the GT distribution is able to adjust for the leptokurtosis of the nonnormal asset returns, it has been used by researchers as an alternative robust modelling distribution in financial econometrics. McDonald (1989) considered the autoregressive moving average (ARMA) time series model based on the GT distribution and developed partially adaptive estimates of the model. McDonald and Nelson (1989) applied the GT distribution using

partially adaptive technique for the estimation of the market model. Applications of the GT distribution to the US stock index returns were given in Bollerslev *et al.* (1994).

Although the GT distribution provides a flexible alternative to the commonly used distributions, its implementation in modelling is complicated. For classical likelihood inference, McDonald and Newey (1988) and Arslan and Genç (2003) used a two-stage estimation procedure where the shape parameters are firstly estimated from the data and then they fix the shape parameters at those estimates throughout the estimation of other parameters such as the location and scale parameters of the distribution. A useful result regarding the property of the GT distribution is given in McDonald and Newey (1988) and Arslan and Genç (2003), where they showed that the GT distribution can be redefined as a scale mixture of the EP and the generalised gamma (GG) distributions. Choy and Chan (2008) further proposed a scale mixture of uniforms (SMU) representation for the GT distribution since the EP distribution can be expressed as a SMU (Walker and Gutiérrez-Peña, 1999). A major advantage of the SMU representation is that in Bayesian estimation through MCMC, some of the full conditional distributions can be reduced to standard forms hence facilitating an efficient Gibbs sampling algorithm for parameter estimation. The same idea applies to SMN which is explored in Chapter 2. Chan *et al.* (2008) successfully applied the GT distribution with SMU in predicting loss reserves for the insurance companies using fully Bayesian approach with the MCMC algorithm.

The GT distribution nests several well known distributions, including the normal, uniform, Laplace, Student-t and EP distributions. These distributions are popular in many financial applications. To name a few, Bollerslev (1987) and Ballie and Bollerslev (1989) assumed conditional Student-t error distribution in ARCH and GARCH models for the analysis of foreign exchange rates and financial market indices; Nelson (1991) used the EP distribution to model daily market returns and So *et al.* (2008) compared the Student-t and EP distributions in terms of their model fitting and forecasting capability in eight international financial market indices.

The objective of this study is to explore and demonstrate the use of GT distribution as the error distribution in a SV model. We will adopt the GT distribution via the SMU representation and demonstrate how this procedure produces a system of mostly standard full

conditional posterior distributions for the SV model parameters in a fully Bayesian estimation procedure using Gibbs sampler. The mixing parameters that arise from the SMU form can be used as global diagnostic tools for the identification of possible outliers.

The basic SV model falls short of allowing for the leverage effect. The notion of a leverage effect is explored in Chapter 2 when we use a bivariate Student-t distribution for the SV modelling of various exchange rate returns and market index returns. It should be clarified, that although the leverage effect implies asymmetry, not all asymmetric effect can be classified as having a leverage effect. The leverage effect is a specific type of asymmetry which specifically refers to the (negative) correlation between the innovations of the returns and volatilities. Asai and McAleer (2011) defines asymmetry as the 'differential impacts of positive and negative shocks on the volatility.' Apart from the leverage effect, other types of the asymmetry for SV models have been proposed and explored in the literature. So *et al.* (2002) proposed a threshold SV model, where the model is divided into two distinct regimes in response to bad news (lag-one return is negative) and good news (lag-one return is positive), given by the following:

$$\begin{aligned} y_t|h_t &= \psi_{0_{s_t}} + \psi_{1_{s_t}} y_{t-1} + \exp\left(\frac{h_t}{2}\right)\epsilon_t, \quad \epsilon_t \sim N(0,1) \\ h_{t+1}|h_t &= \alpha_{s_{t+1}} + \phi_{s_{t+1}} h_t + \tau\eta_t, \quad \eta_t \sim N(0,1), \end{aligned}$$

where s_t is a set of Bernoulli random variables defined by

$$s_t = \begin{cases} 0, & \text{if } y_t < 0 \\ 1, & \text{if } y_t \geq 0. \end{cases}$$

The model parameters in a threshold SV model switch between the two regimes corresponding to the rise and fall of the lag-one return and hence capturing both mean and volatility asymmetries. The threshold SV model proposed is analogous to the threshold ARCH model of Li and Li (1996). Chen *et al.* (2008) generalised the model by So *et al.* (2002) to include exogenous variables and non-zero threshold variable. Asai and McAleer (2005a) pointed out that So *et al.* (2002) only used the sign of the previous return and neglected its magnitude and consequently, they proposed a more general asymmetric SV model that accommodates

both the sign and the magnitude of the previous return as follows.

$$\begin{aligned} y_t|h_t &= \exp\left(\frac{h_t}{2}\right)\epsilon_t, \quad \epsilon_t \sim N(0,1) \\ h_{t+1}|h_t &= \mu + \phi(h_t - \mu) + \gamma_1 y_t + \gamma_2|y_t| + \tau\eta_t, \quad \eta_t \sim N(0,1) \\ E(\epsilon_t\eta_t) &= \rho \end{aligned}$$

The above 'dynamic asymmetric leverage' SV model, directly incorporates the sign and the magnitude of the previous return separately in the volatility equation and in addition, the usual leverage effect is captured through the negative correlation between the innovations of the returns and volatilities. This general asymmetric model nests a few types of asymmetric effect and each is explored through Monte Carlo experiments and their relative performance is compared through an empirical study. Asai and McAleer (2011) suggested a new asymmetric SV model, as a stochastic extension of the EGARCH (Nelson, 1991), where the unobserved standardised errors of the returns are included in the volatility process. They further investigated the difference among several asymmetric effects by classifying them into five categories. Asai and Unite (2010) also employed a general asymmetric SV model, which is an extension of the model in Asai and McAleer (2005), for the analysis of range data of financial returns.

In this second part of this chapter, we extend the basic SV model with GT error distribution to include asymmetric relationship between the returns and volatilities. In particular, we will use the general asymmetric SV model specification proposed in Asai and McAleer (2005a) and compare its nested models with the model of So *et al.* (2002). We will further show that a specific case of the asymmetric model in Asai and McAleer (2005a) leads to time-varying correlation and this model has not been estimated before in empirical applications. Apart from comparing different asymmetric models under the same distribution, we will also compare the performance of different nested distributions of the GT distribution under the same model.

The remainder of this chapter is organised as follows. Firstly, Section 3.2 presents the GT distribution and discusses some properties of this flexible distribution. The SMU representation of the GT distribution is also given, which plays a crucial role in deriving the full conditional densities of the model parameters. The SV model with GT error distribution for both returns and volatilities is specified next in Section 3.3. Property of the SV model

such as the unconditional kurtosis of the returns is derived. Implementation of the GT SV model using MCMC algorithm is presented in the same section by giving the full derivation of the full conditional distributions of the unknown parameters. After that, the properties and practicability of the GT SV model will be demonstrated using two exchange rate return data. Section 3.4 describes the characteristics of the data used and Section 3.5 presents and discusses important empirical findings. The second part of this chapter starts with Section 3.6, which briefly talks about the different types of asymmetric SV models. We specify our asymmetric GT SV model in Section 3.7 and illustrate our models in an empirical study in Section 3.8. Finally, concluding remarks are given in Section 3.9.

3.2. The generalised t (GT) distribution

The GT distribution was proposed by McDonald and Newey (1988) to model econometric data. The GT density is given by

$$GT(x|\mu,\sigma,p,q) = \frac{p}{2q^{\frac{1}{p}}\sigma B\left(\frac{1}{p},q\right)}\left(1+\frac{1}{q}\left|\frac{x-\mu}{\sigma}\right|^{p}\right)^{-\left(q+\frac{1}{p}\right)}, \quad x\in\mathbb{R} \tag{3.1}$$

where $B(.)$ is the beta function, $\mu\in\mathbb{R}$ is a location parameter, $\sigma>0$ is a scale parameter and $p>0$ and $q>0$ are the two shape parameters. The GT distribution is able to accommodate both leptokurtic and platykurtic distributions by varying the values of p and q which control the thickness of the tails. Small p and q values are associated with heavy-tailed distributions and large p and q values give distributions with lighter tails relative to normal. In our study, we use an alternative parameterisation $\gamma^{-1}=pq$ so that the pdf becomes

$$GT(x|\mu,\sigma,p,\gamma) = \frac{p^{1+\frac{1}{p}}\gamma^{\frac{1}{p}}}{2\sigma B\left(\frac{1}{p},\frac{1}{p\gamma}\right)}\left(1+p\gamma\left|\frac{x-\mu}{\sigma}\right|^{p}\right)^{\frac{1}{p}\left(1+\frac{1}{\gamma}\right)}.$$

The advantage of using this reparametrisation is that the condition for the existence of moments depends solely on γ instead of on both p and q. When $\mu=0$ and $\sigma=1$, the even moments of the GT distribution are given by

$$E(X^{2r}) = \frac{\Gamma\left(\frac{1}{p}(2r+1)\right)\Gamma\left(\frac{1}{p}\left(\frac{1}{\gamma}-2r\right)\right)}{(p\gamma)^{\frac{2\gamma}{p}}\Gamma\left(\frac{1}{p}\right)\Gamma\left(\frac{1}{p\gamma}\right)}, \quad r=1,2,\ldots,$$

provided that $\gamma^{-1} > 2r$. The variance is given by

$$\text{var}(X) = \frac{\Gamma\left(\frac{3}{p}\right)\Gamma\left(\frac{1}{p}\left(\frac{1}{\gamma}-2\right)\right)}{(p\gamma)^{\frac{2}{p}}\Gamma\left(\frac{1}{p}\right)\Gamma\left(\frac{1}{p\gamma}\right)}, \tag{3.2}$$

when $\gamma^{-1} > 4$ and the kurtosis is given by

$$\begin{aligned}\kappa &= \frac{E(X^4)}{(E(X^2))^2} \\ &= \frac{\Gamma\left(\frac{5}{p}\right)\Gamma\left(\frac{1}{p}\left(\frac{1}{\gamma}-4\right)\right)\Gamma\left(\frac{1}{p}\right)\Gamma\left(\frac{1}{p\gamma}\right)}{\left[\Gamma\left(\frac{3}{p}\right)\Gamma\left(\frac{1}{p}\left(\frac{1}{\gamma}-2\right)\right)\right]^2}.\end{aligned}$$

The GT family nests a number of well-known distributions: the Laplace ($p = 1, \gamma \to 0$), EP with shape parameter p ($\gamma \to 0$), Cauchy ($p = 2, \gamma = 1$), Student-t with γ^{-1} degrees of freedom ($p = 2$), normal ($p = 2, \gamma \to 0$) and uniform ($p \to \infty$) distributions. In addition, when $p \leq 1$, the GT distribution is cuspidate (that is, the distribution has a sharp peak) and we shall avoid this in real applications. Figure 3.1 shows the density of the standardised GT distribution with different shape parameters p and q. Special cases of the GT density, including the Laplace, Student-t, EP, normal and uniform distributions, are also presented in the same plot. Note that the third graph with $q = 100$ represents the EP distribution with different shape parameter p. When $p = 1$, we obtain the Laplace distribution and when $p = 2$, we obtain the normal distribution. Hence the Laplace and normal distributions are both special cases of the EP distribution.

3.2.1. The SMU representation of the GT distribution. The class of scale mixture of uniform (SMU) distributions was proposed by Walker and Gutiérrez-Peña (1999). Let X be a continuous random variable with location parameter μ and scale parameter σ. The pdf of X is said to have a SMU representation if it can be expressed as

$$f(x|\mu,\sigma) = \int_0^\infty U(x|\mu-\kappa(u)\sigma, \mu+\kappa(u)\sigma)\pi(u)du, \tag{3.3}$$

where $U(x|a,b)$ is the uniform density function with support $[a,b]$, $\kappa(.)$ is a positive function and $\pi(.)$ is a density function defined on $\mathbb{R}^+$. We refer to u and $\pi(.)$ as the mixing parameter and mixing density, respectively.

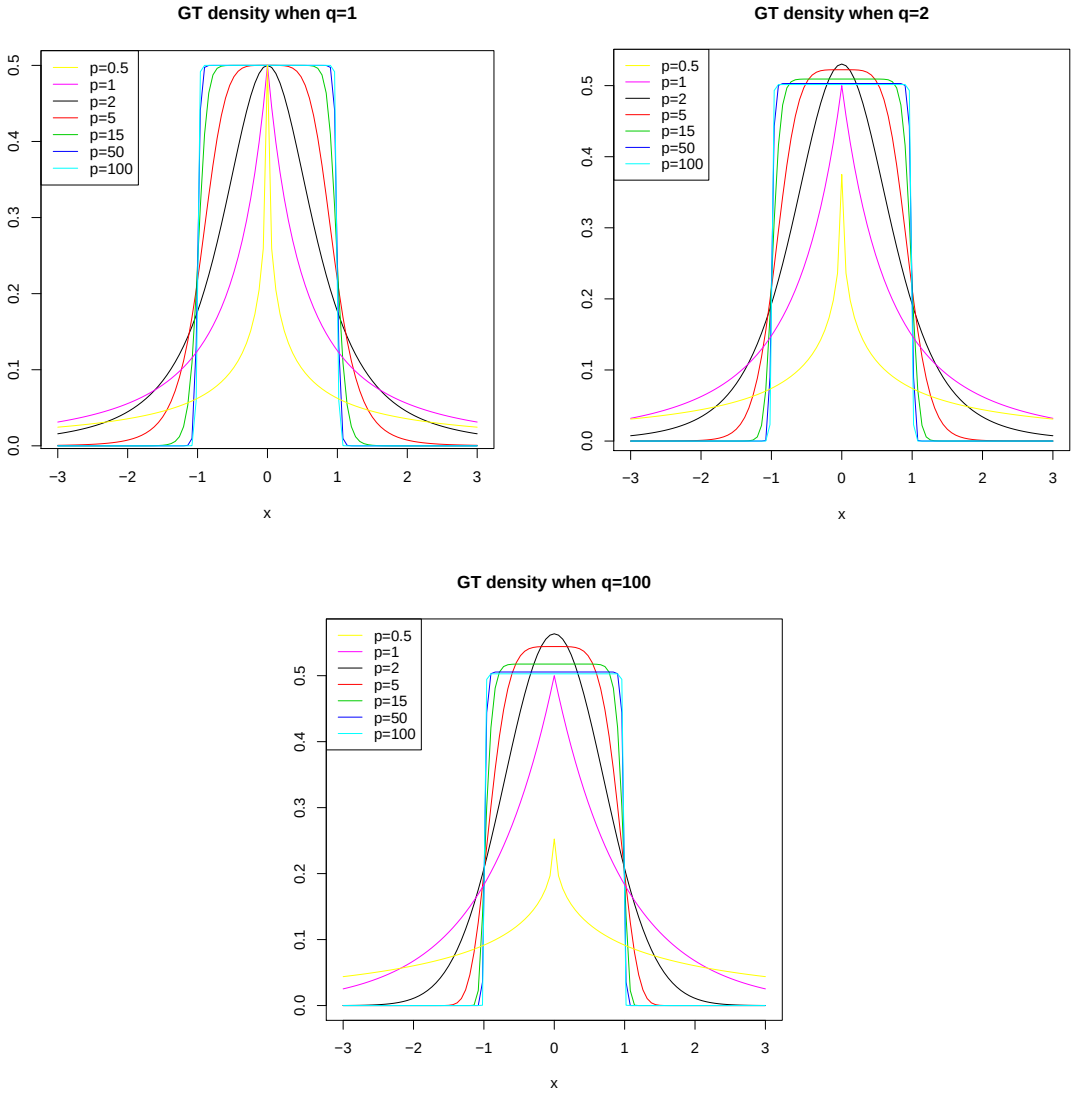


FIGURE 3.1. The density plot of the GT distribution with various values of p and q.

Choy and Chan (2008) provided the following SMU representation for the GT pdf in (3.1)

$$\int_0^\infty \int_\delta^\infty U(x|\mu - q^{\frac{1}{p}} s^{-\frac{1}{2}} u^{\frac{1}{p}} \sigma, \mu + q^{\frac{1}{p}} s^{-\frac{1}{2}} u^{\frac{1}{p}} \sigma) Ga\left(u|1+\frac{1}{p},1\right) GG\left(s|q,1,\frac{p}{2}\right) duds \quad (3.4)$$

where $Ga(.|a,b)$ is the pdf of the gamma distribution with mean a/b, $GG(.|a,b,c)$ is the generalised gamma (GG) distribution and

$$\delta = \frac{1}{q}\left|\frac{x-\mu}{\sigma}\right|^p s^{p/2}.$$

The pdf of the GG distribution with scale parameter $b > 0$ and shape parameters $a > 0$ and $c > 0$ is given by

$$GG(s|a,b,c) = \frac{cb^{ac}}{\Gamma(a)} s^{ac-1} \exp(-b^c s^c), \quad s > 0. \tag{3.5}$$

We provide the following proof.

Proof.

$$\begin{aligned}
& \int_0^\infty \int_\delta^\infty U(x|\mu - q^{\frac{1}{p}} s^{-\frac{1}{2}} u^{\frac{1}{p}} \sigma, \mu + q^{\frac{1}{p}} s^{-\frac{1}{2}} u^{\frac{1}{p}} \sigma) Ga\left(u|1+\frac{1}{p},1\right) GG\left(s|q,1,\frac{p}{2}\right) duds \\
= & \int_0^\infty \int_\delta^\infty \frac{1}{2q^{1/p} s^{-1/2} u^{1/p} \sigma} \frac{u^{1/p} \exp(-u)}{\Gamma(1+1/p)} \frac{p}{2\Gamma(q)} s^{pq/2-1} \exp(-s^{p/2}) duds \\
= & \frac{p}{4q^{1/p}\sigma\Gamma(1+1/p)\Gamma(q)} \int_0^\infty \int_\delta^\infty s^{pq/2-1/2} \exp(-s^{p/2}) \exp(-u) duds \\
= & \frac{p}{4q^{1/p}\sigma\Gamma(1+1/p)\Gamma(q)} \int_0^\infty s^{pq/2-1/2} \exp(-s^{p/2}) \exp\left(-\frac{1}{q}\left|\frac{x-\mu}{\sigma}\right|^p s^{p/2}\right) ds \\
= & \frac{p}{4q^{1/p}\sigma\Gamma(1+1/p)\Gamma(q)} \int_0^\infty s^{pq/2-1/2} \exp\left(-s^{p/2}\left(1+\left(-\frac{1}{q}\left|\frac{x-\mu}{\sigma}\right|^p s^{p/2}\right)\right)\right) ds \\
= & \frac{p}{4q^{1/p}\sigma\Gamma(1+1/p)\Gamma(q)} \frac{\Gamma(q+1/p)}{p/2} \left(\left(1+\left(-\frac{1}{q}\left|\frac{x-\mu}{\sigma}\right|^p s^{p/2}\right)\right)^{2/p}\right)^{pq/2+1/2} \\
= & \frac{\Gamma(q+1/p)}{2q^{1/p}\sigma\Gamma(1+1/p)\Gamma(q)} \left(1+\frac{1}{q}\left|\frac{x-\mu}{\sigma}\right|^p\right)^{-(q+1/p)} \\
= & \frac{p}{2q^{1/p}\sigma B(1/p,q)} \left(1+\frac{1}{q}\left|\frac{x-\mu}{\sigma}\right|^p\right)^{-(q+1/p)}
\end{aligned}$$

□

Using the reparameterisation $\gamma^{-1} = pq$, the SMU representation in (3.4) becomes

$$\int_0^\infty \int_\delta^\infty U(x|\mu - p^{-\frac{1}{p}} \gamma^{-\frac{1}{p}} \omega\sigma, \mu + p^{-\frac{1}{p}} \gamma^{-\frac{1}{p}} \omega\sigma) Ga(u|1+p^{-1},1) GG(s|p^{-1}\gamma^{-1},1,p/2) duds \tag{3.6}$$

where

$$\delta = \frac{1}{p\gamma}\left|\frac{x-\mu}{\sigma}\right|^p s^{p/2} \quad \text{and} \quad \omega = u^{1/p} s^{-1/2},$$

which is a combined mixing parameter used for outlier diagnostic.

3.3. SV model with GT distribution

We add flexibility to the basic SV model (1.1) by incorporating an exogenous variable z_t for the returns. In addition, we also consider the GT distribution for the log-volatilities such that the model is able to capture the leptokurtic feature in both the returns and log-volatilities series. We shall call the SV model with GT error distribution in both the return and log-volatility processes the GT-GT SV model and it is defined as follows:

$$\begin{aligned} y_t|h_t &\sim GT\left(\psi_0+\psi_1 z_{t-1}, \exp(h_t/2), p_y, \gamma_y\right), \qquad t=1,2,...,n, \\ h_{t+1}|h_t &\sim GT\left(\mu+\phi(h_t-\mu), \tau, p_h, \gamma_h\right), \qquad t=1,2,...,n-1, \\ h_1 &\sim GT\left(\mu, \frac{\tau}{\sqrt{1-\phi^2}}, p_h, \gamma_h\right) \end{aligned} \tag{3.7}$$

where z_{t-1} is a lag one exogenous variable, $\gamma_y^{-1} = p_y q_y$, $\gamma_h^{-1} = p_h q_h$ and (p_y, q_y) and (p_h, q_h) are the shape parameters of the GT distributions in the return and log-volatility series, respectively.

Define $y_t^* = y_t - (\psi_0 + \psi_1 z_{t-1})$. Then the marginal moments of y_t^* are

$$\begin{aligned} E(y_t^{*^2}) &= \exp\left(\mu + \frac{\tau^2}{2(1-\phi^2)}\right) \frac{\Gamma\left(\frac{3}{p_y}\right)\Gamma\left(\frac{1}{p_y}\left(\frac{1}{\gamma_y}-2\right)\right)}{(p_y\gamma_y)^{2/p_y}\Gamma\left(\frac{1}{p_y}\right)\Gamma\left(\frac{1}{p_y\gamma_y}\right)}, \\ E(y_t^{*^4}) &= \exp\left(2\mu + \frac{2\tau^2}{1-\phi^2}\right) \frac{\Gamma\left(\frac{5}{p_y}\right)\Gamma\left(\frac{1}{p_y}\left(\frac{1}{\gamma_y}-4\right)\right)}{(p_y\gamma_y)^{4/p_y}\Gamma\left(\frac{1}{p_y}\right)\Gamma\left(\frac{1}{p_y\gamma_y}\right)}. \end{aligned}$$

Hence the unconditional kurtosis of y_t^* is given by

$$\kappa = \exp\left(\frac{\tau^2}{1-\phi^2}\right) \frac{\Gamma\left(\frac{5}{p_y}\right)\Gamma\left(\frac{1}{p_y}\left(\frac{1}{\gamma_y}-4\right)\right)\Gamma\left(\frac{1}{p_y}\right)\Gamma\left(\frac{1}{p_y\gamma_y}\right)}{\left[\Gamma\left(\frac{3}{p_y}\right)\Gamma\left(\frac{1}{p_y}\left(\frac{1}{\gamma_y}-2\right)\right)\right]^2} \tag{3.8}$$

which depends on both p_y and γ_y.

By incorporating the SMU form for the GT distribution given in (3.6), this GT-GT SV model can be formulated alternatively as

$$\begin{aligned} y_t|h_t, p_y, \gamma_y, u_{y_t}, s_{y_t}, \psi_0, \psi_1 \sim U\Big(&\psi_0+\psi_1 z_{t-1} - \left(p_y^{-1}\gamma_y^{-1}\right)^{1/p_y}\omega_{y_t}\exp(h_t/2), \\ &\psi_0+\psi_1 z_{t-1} + \left(p_y^{-1}\gamma_y^{-1}\right)^{1/p_y}\omega_{y_t}\exp(h_t/2)\Big) \end{aligned}$$

with volatility equation:

$$\begin{aligned} h_t|h_{t-1},\mu,\phi,\tau,p_h,\gamma_h,u_{h_t},s_{h_t} &\sim U\Big(\mu+\phi(h_{t-1}-\mu)-\left(p_h^{-1}\gamma_h^{-1}\right)^{1/p_h}\omega_{h_t}\tau, \\ &\qquad \mu+\phi(h_{t-1}-\mu)+\left(p_h^{-1}\gamma_h^{-1}\right)^{1/p_h}\omega_{h_t}\tau\Big) \\ h_1|\mu,\phi,\tau,p_h,\gamma_h,u_{h1},s_{h1} &\sim U\Big(\mu-\left(p_h^{-1}\gamma_h^{-1}\right)^{1/p_h}\omega_{h1}\tau(1-\phi^2)^{-1/2}, \\ &\qquad \mu+\left(p_h^{-1}\gamma_h^{-1}\right)^{1/p_h}\omega_{h1}\tau(1-\phi^2)^{-1/2}\Big) \end{aligned}$$

where

$$\begin{aligned} \omega_{y_t} &= (u_{y_t})^{1/p_y}(s_{y_t})^{-1/2}, \\ \omega_{h_t} &= (u_{h_t})^{1/p_h}(s_{h_t})^{-1/2}, \\ u_{y_t} &\sim Ga(1+p_y^{-1},1), \\ s_{y_t} &\sim GG\left((p_y\gamma_y)^{-1},1,p_y/2\right), \\ u_{h_t} &\sim Ga(1+p_h^{-1},1), \\ s_{h_t} &\sim GG\left((p_h\gamma_h)^{-1},1,p_h/2\right). \end{aligned}$$

3.3.1. The MCMC algorithm for GT-GT SV model. As in the previous chapter, we derive the full conditional distributions for the model parameters and latent volatilities and we show that most of these full conditional distributions are of standard forms due to the use

of the SMU representation. We assign the following priors to the model parameters:

$$\begin{aligned}
\psi_0 &\sim N(a_{\psi_0}, b_{\psi_0}),\\
\psi_1 &\sim U(a_{\psi_1}, b_{\psi_1}),\\
\mu &\sim N(a_\mu, b_\mu),\\
\tau^2 &\sim IG(a_\tau, b_\tau),\\
\phi^* &\sim Be(\alpha, \beta),\\
p_y &\sim U(a_{p_y}, b_{p_y}),\\
\gamma_y &\sim U(a_{\gamma_y}, b_{\gamma_y}),\\
p_h &\sim U(a_{p_h}, b_{p_h}),\\
\gamma_h &\sim U(a_{\gamma_h}, b_{\gamma_h}),
\end{aligned} \tag{3.9}$$

where $\phi^* = \frac{\phi+1}{2}$ as before. Normal conjugate priors are assigned to mean and volatility intercept parameters ψ_0 and μ. A conjugate inverse gamma prior is assigned to τ^2 and a Beta prior is assigned to ϕ for the reason stated in Chapter 1. A uniform prior with restricted range (-1,1) is assigned to the autoregressive coefficient ψ_1 and uniform priors with suitable ranges are assigned to the shape parameters of the GT distribution. In real applications, we choose values of the hyperparameters for the shape parameters p_y, γ_y, p_h, γ_h that are able to cover a wide range of GT distributions and at the same time, prevent the GT distribution from being too cuspidate.

Let $\boldsymbol{y} = (y_1, \ldots, y_n)$, $\boldsymbol{y}^* = (y_1^*, \ldots, y_n^*)$ where $y_t^* = y_t - (\psi_0 + \psi_1 z_{t-1})$, $\boldsymbol{z} = (z_1, \ldots, z_n)$, $\boldsymbol{h} = (h_1, \ldots, h_n)$ be the vector of log-volatilities, $\boldsymbol{u}_y = (u_{y_1}, \ldots, u_{y_n})$, $\boldsymbol{s}_y = (s_{y_1}, \ldots, s_{y_n})$ be the mixing parameters for returns and $\boldsymbol{u}_h = (u_{h_1}, \ldots, u_{h_n})$, $\boldsymbol{s}_h = (s_{h_1}, \ldots, s_{h_n})$ be the mixing parameters for log-volatilities. Also define

$\boldsymbol{h}_{\setminus t} = (h_1, \ldots, h_{t-1}, h_{t+1}, \ldots, h_n)$, $\boldsymbol{u}_{y,\setminus t} = (u_{y_1}, \ldots, u_{y_{t-1}}, u_{y_{t+1}}, \ldots, u_{y_n})$,
$\boldsymbol{s}_{y,\setminus t} = (s_{y_1}, \ldots, s_{y_{t-1}}, s_{y_{t+1}}, \ldots, s_{y_n})$, $\boldsymbol{u}_{h,\setminus t} = (u_{h_1}, \ldots, u_{h_{t-1}}, u_{h_{t+1}}, \ldots, u_{h_n})$ and
$\boldsymbol{s}_{h,\setminus t} = (s_{h_1}, \ldots, s_{h_{t-1}}, s_{h_{t+1}}, \ldots, s_{h_n})$.

The system of full conditional distributions are given as follows.

- Full conditional distribution for h_t, $t = 1, \ldots, n$:

$$h_1|\boldsymbol{h}_{\setminus t}, \boldsymbol{u}_h, \boldsymbol{s}_h, \boldsymbol{u}_y, \boldsymbol{s}_y, \mu, \phi, \tau^2, p_h, \gamma_h, p_y, \gamma_y, \boldsymbol{y}^* \sim U(h_{1_a}, h_{1_b}),$$

where

$$\begin{aligned} h_{1_a} &= \max\left(\frac{\mu - (p_h^{-1}\gamma_h^{-1}u_{h_1})^{1/p_h}s_{h_1}^{-1/2}\tau}{\sqrt{1-\phi^2}}, \frac{h_1 - \mu(1-\phi) - (p_h^{-1}\gamma_h^{-1}u_{h_2})^{1/p_h}s_{h_2}^{-1/2}\tau}{\phi}\right), \\ h_{1_b} &= \min\left(\frac{\mu + (p_h^{-1}\gamma_h^{-1}u_{h_1})^{1/p_h}s_{h_1}^{-1/2}\tau}{\sqrt{1-\phi^2}}, \frac{h_1 - \mu(1-\phi) + (p_h^{-1}\gamma_h^{-1}u_{h_2})^{1/p_h}s_{h_2}^{-1/2}\tau}{\phi}\right). \end{aligned}$$

$$h_t|\boldsymbol{h}_{\setminus t}, \boldsymbol{u}_h, \boldsymbol{s}_h, \boldsymbol{u}_y, \boldsymbol{s}_y, \mu, \phi, \tau^2, p_h, \gamma_h, p_y, \gamma_y, \boldsymbol{y}^* \sim Exp\left(\frac{1}{2}\right) I\left(\max(h_{t_a}, h_{t_b}, h_{t_c}), \min(h_{t_d}, h_{t_e})\right), \quad t = 2, \ldots, n,$$

where

$$\begin{aligned} h_{t_a} &= \ln\left(\frac{(y_t^*)^2 s_{y_t}}{(p_y^{-1}\gamma_y^{-1}u_{y_t})^{2/p_y}}\right), \\ h_{t_b} &= \mu + \phi(h_{t-1} - \mu) - (p_h^{-1}\gamma_h^{-1}u_{h_t})^{1/p_h}s_{h_t}^{-1/2}\tau, \\ h_{t_c} &= \frac{h_{t+1} - \mu(1-\phi) - (p_h^{-1}\gamma_h^{-1}u_{h_{t+1}})^{1/p_h}s_{h_{t+1}}^{-1/2}\tau}{\phi}, \\ h_{t_d} &= \mu + \phi(h_{t-1} - \mu) + (p_h^{-1}\gamma_h^{-1}u_{h_t})^{1/p_h}s_{h_t}^{-1/2}\tau, \\ h_{t_e} &= \frac{h_{t+1} - \mu(1-\phi) + (p_h^{-1}\gamma_h^{-1}u_{h_{t+1}})^{1/p_h}s_{h_{t+1}}^{-1/2}\tau}{\phi}. \end{aligned}$$

- Full conditional distribution for u_{y_t}, $t = 1, 2, \ldots, n$:

$$u_{y_t}|\boldsymbol{h}, \boldsymbol{u}_{y,\setminus t}, \boldsymbol{u}_h, \boldsymbol{s}_h, \boldsymbol{s}_y, \mu, \phi, \tau^2, p_h, \gamma_h, p_y, \gamma_y, \boldsymbol{y}^* \sim Exp(1) I\left(u_{y_t} > p_y\gamma_y\left(|y_t^*|s_{y_t}^{1/2}e^{-h_t/2}\right)^{p_y}\right)$$

- Full conditional distribution for s_{y_t}, $t = 1, 2, \ldots, n$:

$$\begin{aligned} s_{y_t}|\boldsymbol{h}, \boldsymbol{s}_{y,\setminus t}, \boldsymbol{u}_h, \boldsymbol{s}_h, \boldsymbol{u}_y, \mu, \phi, \tau^2, p_h, \gamma_h, p_y, \gamma_y, \boldsymbol{y}^* &\sim GG\left(\frac{1}{p_y}\left(\frac{1}{\gamma_y} + 1\right), 1, \frac{p_y}{2}\right) \\ &\quad \times I\left(s_{y_t} < \frac{(p_y^{-1}\gamma_y^{-1}u_{y_t})^{2/p_y}e^{h_t}}{(y_t^*)^2}\right) \end{aligned}$$

- Full conditional distribution for u_{h_t}, $t = 1, \ldots, n$:

$$u_{h_1}|\boldsymbol{h}, \boldsymbol{u}_{h,\setminus t}, \boldsymbol{s}_h, \boldsymbol{u}_y, \boldsymbol{s}_y, \mu, \phi, \tau^2, p_h, \gamma_h, p_y, \gamma_y, \boldsymbol{y}^* \sim Exp(1) I\Big(u_{h_1} > |h_1 - \mu|^{p_h} p_h \gamma_h (1-\phi^2)^{p_h/2} s_{h_1}^{p_h/2} \tau^{-p_h}\Big)$$

$$\begin{aligned} u_{h_t}|\boldsymbol{h}, \boldsymbol{u}_{h,\setminus t}, \boldsymbol{s}_h, \boldsymbol{u}_y, \boldsymbol{s}_y, \mu, \phi, \tau^2, p_h, \gamma_h, p_y, \gamma_y, \boldsymbol{y}^* &\sim Exp(1) I\Big(u_{h_t} > \\ &\quad |h_t - \mu - \phi(h_{t-1} - \mu)|^{p_h} p_h \gamma_h s_{h_t}^{p_h/2} \tau^{-p_h}\Big) \end{aligned}$$

- Full conditional distribution for s_{h_t}, $t = 1, \ldots, n$:

$$\begin{aligned} s_{h_1}|\boldsymbol{h}, \boldsymbol{s}_{h,\setminus t}, \boldsymbol{u}_h, \boldsymbol{u}_y, \boldsymbol{s}_y, \mu, \phi, \tau^2, p_h, \gamma_h, p_y, \gamma_y, \boldsymbol{y}^* \sim \; & GG\left(\frac{1}{p_h}\left(\frac{1}{\gamma_h} + 1\right), 1, \frac{p_h}{2}\right) \\ & \times I\left(s_{h_1} < \frac{(p_h^{-1}\gamma_h^{-1} u_{h_1})^{2/p_h} \tau^2}{(1-\phi^2)|h_1 - \mu|^2}\right) \end{aligned}$$

$$\begin{aligned} s_{h_t}|\boldsymbol{h}, \boldsymbol{s}_{h,\setminus t}, \boldsymbol{u}_h, \boldsymbol{u}_y, \boldsymbol{s}_y, \mu, \phi, \tau^2, p_h, \gamma_h, p_y, \gamma_y, \boldsymbol{y}^* \sim \; & GG\left(\frac{1}{p_h}\left(\frac{1}{\gamma_h} + 1\right), 1, \frac{p_h}{2}\right) \\ & \times I\left(s_{h_t} < \frac{(p_h^{-1}\gamma_h^{-1} u_{h_t})^{2/p_h} \tau^2}{|h_t - \mu - \phi(h_{t-1} - \mu)|^2}\right) \end{aligned}$$

- Full conditional distribution for ψ_0:

$$\psi_0|\boldsymbol{h}, \boldsymbol{u}_h, \boldsymbol{s}_h, \boldsymbol{u}_y, \boldsymbol{s}_y, \mu, \phi, \tau^2, p_h, \gamma_h, p_y, \gamma_y, \psi_1, \boldsymbol{z}, \boldsymbol{y} \sim N(a_{\psi_0}, b_{\psi_0}) I(a^*_{\psi_0} < \psi_0 < b^*_{\psi_0})$$

where

$$\begin{aligned} a^*_{\psi_0} &= \max\left(y_t - \psi_1 z_{t-1} - \left(p_y^{-1}\gamma_y^{-1} u_{y_t}\right)^{1/p_y} s_{y_t}^{-1/2} e^{h_t/2}\right), \quad t = 1, \ldots, n, \\ b^*_{\psi_0} &= \min\left(y_t - \psi_1 z_{t-1} + \left(p_y^{-1}\gamma_y^{-1} u_{y_t}\right)^{1/p_y} s_{y_t}^{-1/2} e^{h_t/2}\right), \quad t = 1, \ldots, n. \end{aligned}$$

- Full conditional distribution for ψ_1:

$$\psi_1|\boldsymbol{h}, \boldsymbol{u}_h, \boldsymbol{s}_h, \boldsymbol{u}_y, \boldsymbol{s}_y, \mu, \phi, \tau^2, p_h, \gamma_h, p_y, \gamma_y, \psi_0, \boldsymbol{z}, \boldsymbol{y} \sim U(-1, 1) \times \prod_{t=1}^{n} I(a^*_{\psi_1} < \psi_1 z_{t-1} < b^*_{\psi_1})$$

where

$$\begin{aligned} a^*_{\psi_1} &= y_t - \psi_0 - \left(p_y^{-1}\gamma_y^{-1} u_{y_t}\right)^{1/p_y} s_{y_t}^{-1/2} e^{h_t/2}, \\ b^*_{\psi_1} &= y_t - \psi_0 + \left(p_y^{-1}\gamma_y^{-1} u_{y_t}\right)^{1/p_y} s_{y_t}^{-1/2} e^{h_t/2}. \end{aligned}$$

- Full conditional distribution for μ:

$$\mu|\boldsymbol{h}, \boldsymbol{u}_h, \boldsymbol{s}_h, \boldsymbol{u}_y, \boldsymbol{s}_y, \phi, \tau^2, p_h, \gamma_h, p_y, \gamma_y, \boldsymbol{y}^* \sim N(a_\mu, b_\mu) I(\mu_a < \mu < \mu_b),$$

where

$$\begin{aligned}
\mu_a &= \max\left(\frac{h_1 - (p_h^{-1}\gamma_h^{-1}u_{h_1})^{1/p_h} s_{h_1}^{-1/2}\tau}{\sqrt{1-\phi^2}}, \frac{h_t - \phi h_{t-1} - (p_h^{-1}\gamma_h^{-1}u_{h_t})^{1/p_h} s_{h_t}^{-1/2}\tau}{1-\phi} \right), t = 1, \ldots, n \\
\mu_b &= \min\left(\frac{h_1 + (p_h^{-1}\gamma_h^{-1}u_{h_1})^{1/p_h} s_{h_1}^{-1/2}\tau}{\sqrt{1-\phi^2}}, \frac{h_t - \phi h_{t-1} + (p_h^{-1}\gamma_h^{-1}u_{h_t})^{1/p_h} s_{h_t}^{-1/2}\tau}{1-\phi} \right), t = 1, \ldots, n
\end{aligned}$$

- Full conditional distribution for $\phi^* = (\phi + 1)/2$:

$$\begin{aligned}
\phi^*|\boldsymbol{h}, \boldsymbol{u}_h, \boldsymbol{s}_h, \boldsymbol{u}_y, \boldsymbol{s}_y, \mu, \tau^2, p_h, \gamma_h, p_y, \gamma_y, \boldsymbol{y}^* \sim\ & Be\left(\alpha + \frac{1}{2}, \beta + \frac{1}{2}\right) I(\phi_a^* < \phi^* < \phi_b^*) \\
& \times I\left(\phi^* > \frac{1}{2}\left(1 + \sqrt{1 - \frac{(p_h^{-1}\gamma_h^{-1}u_{h_1})^{1/p_h} s_{h_1}^{-1/2}\tau^2}{|h_1 - \mu|^2}}\right)\right)
\end{aligned}$$

where

$$\begin{aligned}
\phi_a^* &= \max\left(\frac{h_t - \mu - (p_h^{-1}\gamma_h^{-1}u_{h_1})^{1/p_h} s_{h_1}^{-1/2}\tau + 1}{2(h_t - \mu)} \right), \quad t = 1, 2, \ldots, n, \\
\phi_b^* &= \min\left(\frac{h_t - \mu + (p_h^{-1}\gamma_h^{-1}u_{h_1})^{1/p_h} s_{h_1}^{-1/2}\tau + 1}{2(h_t - \mu)} \right), \quad t = 1, 2, \ldots, n.
\end{aligned}$$

- Full conditional distribution for τ^2:

$$\tau^2|\boldsymbol{h}, \boldsymbol{u}_h, \boldsymbol{s}_h, \boldsymbol{u}_y, \boldsymbol{s}_y, \mu, \phi, p_h, \gamma_h, p_y, \gamma_y, \boldsymbol{y}^* \sim IG\left(a_\tau + \frac{n+1}{2}, b_\tau\right) I\left(\tau^2 > \max(\tau_a^2, \tau_b^2)\right)$$

where

$$\begin{aligned}
\tau_a^2 &= \frac{(h_1 - \mu)^2 s_{h_1}(1-\phi^2)}{(p_h^{-1}\gamma_h^{-1}u_{h_1})^{2/p_h}}, \\
\tau_b^2 &= \frac{(h_t - \mu - \phi(h_{t-1} - \mu))^2 s_{h_t}}{(p_h^{-1}\gamma_h^{-1}u_{h_t})^{2/p_h}}, \quad t = 2, \ldots, n.
\end{aligned}$$

- Full conditional distribution for p_y:

$$p_y|\boldsymbol{h},\boldsymbol{u}_y,\boldsymbol{s}_y,\boldsymbol{u}_h,\boldsymbol{s}_h,\mu,\phi,\tau^2,p_h,\gamma_h,\gamma_y,\boldsymbol{y}^* \propto \left(\frac{p_y^{1/p_y+1}\gamma_y^{1/p_y}}{\Gamma(1+1/p_y)\Gamma(1/(p_y\gamma_y))}\right)^n \exp\left(-\sum_{t=1}^n s_{y_t}^{p_y/2}\right) \times \prod_{t=1}^n I\left(\left(\frac{u_{y_t}}{p_y\gamma_y}\right)^{1/p_y} > |y_t^*|s_{y_t}^{1/2}e^{-h_t/2}\right) \times I(a_{p_y},b_{p_y})$$

- Full conditional distribution for p_h:

$$p_h|\boldsymbol{h},\boldsymbol{u}_y,\boldsymbol{s}_y,\boldsymbol{u}_h,\boldsymbol{s}_h,\mu,\phi,\tau^2,\gamma_h,p_y,\gamma_y,\boldsymbol{y}^* \propto \left(\frac{p_h^{1/p_h+1}\gamma_h^{1/p_h}}{\Gamma(1+1/p_h)\Gamma(1/(p_h\gamma_h))}\right)^n \exp\left(-\sum_{t=1}^n s_{h_t}^{p_h/2}\right) \times I\left(\left(\frac{u_{h_1}}{p_h\gamma_h}\right)^{1/p_h} > |h_1-\mu|(1-\phi^2)^{1/2}\frac{s_{h_1}^{1/2}}{\tau}\right) \times \prod_{t=2}^n I\left(\left(\frac{u_{h_t}}{p_h\gamma_h}\right)^{1/p_h} > |h_t-\mu-\phi(h_{t-1}-\mu)|\frac{s_{h_t}^{1/2}}{\tau}\right) \times I(a_{p_h},b_{p_h})$$

- Full conditional distribution for γ_y:

$$\gamma_y|\boldsymbol{h},\boldsymbol{u}_y,\boldsymbol{s}_y,\boldsymbol{u}_h,\boldsymbol{s}_h,\mu,\phi,\tau^2,p_h,\gamma_h,p_y,\boldsymbol{y}^* \propto \left(\frac{\gamma_y^{1/p_y}}{\Gamma(1/(p_y\gamma_y))}\right)^n \left(\prod_{t=1}^n s_{y_t}^{\frac{1}{2\gamma_y}}\right) \times I\left(\gamma_y<\gamma_y^*\right) \times I(a_{\gamma_y},b_{\gamma_y})$$

where

$$\gamma_y^* = \min\left(\frac{u_{y_t}}{p_y}\left(\frac{e^{h_t/2}}{|y_t^*|s_{y_t}^{1/2}}\right)^{p_y}\right), \quad t=1,\ldots,n.$$

- Full conditional distribution for γ_h:

$$\gamma_h|\boldsymbol{h},\boldsymbol{u}_y,\boldsymbol{s}_y,\boldsymbol{u}_h,\boldsymbol{s}_h,\mu,\phi,\tau^2,p_h,p_y,\gamma_y,\boldsymbol{y}^* \propto \left(\frac{\gamma_h^{1/p_h}}{\Gamma(1/(p_h\gamma_h))}\right)^n \left(\prod_{t=1}^n s_{h_t}^{\frac{1}{2\gamma_h}}\right) I(\gamma_h<\min(\gamma_a,\gamma_b)) \times I(a_{\gamma_h},b_{\gamma_h})$$

where

$$\gamma_a = \frac{u_{h_1}}{p_h}\left(\frac{\tau}{|h_1-\mu|s_{h_1}^{1/2}\sqrt{1-\phi^2}}\right)^{p_h}$$

$$\gamma_b = \frac{u_{h_t}}{p_h}\left(\frac{\tau}{|h_t-\mu-\phi(h_{t-1}-\mu)|s_{h_t}^{1/2}}\right)^{p_h}, \quad t=2,\ldots,n.$$

From the above full conditional distributions, we can see that that the full conditional distribution for h_1 is uniform distribution and the full conditional distributions for ψ_0, ψ_1 and μ are truncated normal distributions. The full conditional distributions for h_t, u_{y_t} and u_{h_t}, $t = 1, 2, \ldots, n$ are truncated exponential distributions and those for s_{y_t} and s_{h_t} truncated GG distributions. The full conditional distributions for ϕ and τ^2 are truncated beta and truncated inverse gamma distributions, respectively. Simulation from the truncated exponential distribution can be done using the inversion method. To simulate from the truncated normal and truncated gamma distributions, the algorithms proposed by Robert (1995) and Philippe (1997) can be used, respectively. To simulate from the truncated GG distribution, we can make use of the fact that if $X \sim Ga(\alpha, \beta^\gamma)$, then the pdf of $Y = X^{1/\gamma}$ is given by

$$f_Y(y) = \frac{1}{\Gamma(\alpha)}\beta^{\alpha\gamma}\exp(-\beta^\gamma y^\gamma)y^{\alpha\gamma-1}$$

where is $GG(\alpha, \beta, \gamma)$. Hence suppose we want to simulate a random variate from truncated GG distribution $Y \sim GG(\alpha, \beta, \gamma)I(Y \leq y^*)$, we can alternatively simulate from truncated gamma distribution, $X \sim G(\alpha, \beta)I(X \leq (y^*)^\gamma)$ and $X^{1/\gamma}$ is a random variate from the desired truncated GG distribution. An algorithm for simulating from a right truncated gamma distribution is given by Philippe (1997). The full conditional distributions for the shape parameters are non-standard, hence sampling method such as the Metropolis-Hastings algorithm can be used to sample random variates from these distributions.

3.4. An empirical study

In this section, we analyse two sets of exchange rate data, they are the Australian dollar/Euro (AUD/EUR) and AUD/Japanese Yen (AUD/JPY) exchange rates. The AUD/US dollar (AUD/USD) exchange rate is used as the covariate z_{t-1} in equation (3.7). Exchange rate data are chosen in particular to demonstrate the GT-GT SV model in (3.7) because it is shown in Chapter 2 that exchange rate data typically do not exhibit a significant leverage

effect. The sampling period is from 6/1/2003 to 30/6/2009, yielding 1626 observations. The returns are defined as $y_t = (\log P_t - \log P_{t-1}) \times 100\%$ where P_t is the closing exchange rate on day t.

Table 3.1 summarises the descriptive statistics for the two return series as well as the AUD/USD covariate including the mean, standard deviation, skewness and kurtosis. Note that the kurtoses are all greater than 3, indicating the well-known leptokurtic characteristic of financial return data. The Jarque-Bera p-values are close to 0, suggesting significant departure from normality for all three data sets. Time series plots and histograms of the AUD/EUR and AUD/JPY exchange rate data, as well as the covariate AUD/USD exchange rate are shown in Figure 3.2.

TABLE 3.1. Summary statistics for AUD/EUR, AUD/JPY and AUD/USD exchange rates.

	AUD/EUR	AUD/JPY	AUD/USD
Mean	0.0016	0.0038	0.0098
Median	0.0219	0.0377	0.0336
Standard deviation	0.3017	0.5132	0.4088
Skewness	-0.6132	-0.6918	-0.4514
Kurtosis	9.3945	13.4849	9.5287
J-B statistic	6081.29	12449.53	6208.08
J-B p-value	<0.001	<0.001	<0.001

3.5. Empirical results

We fit the GT-GT SV model and the t-t SV model, which is a GT-GT SV model with parameter values $p_y = 2$ and $p_h = 2$, respectively to our data. For model comparison, we use the DIC defined earlier as it has been successfully applied to the family of SV models (Berg, *et al.*, 2004) and is less computationally demanding than other model comparison measures such as the Bayes factor. DIC can be straightforwardly output from `WinBUGS`.

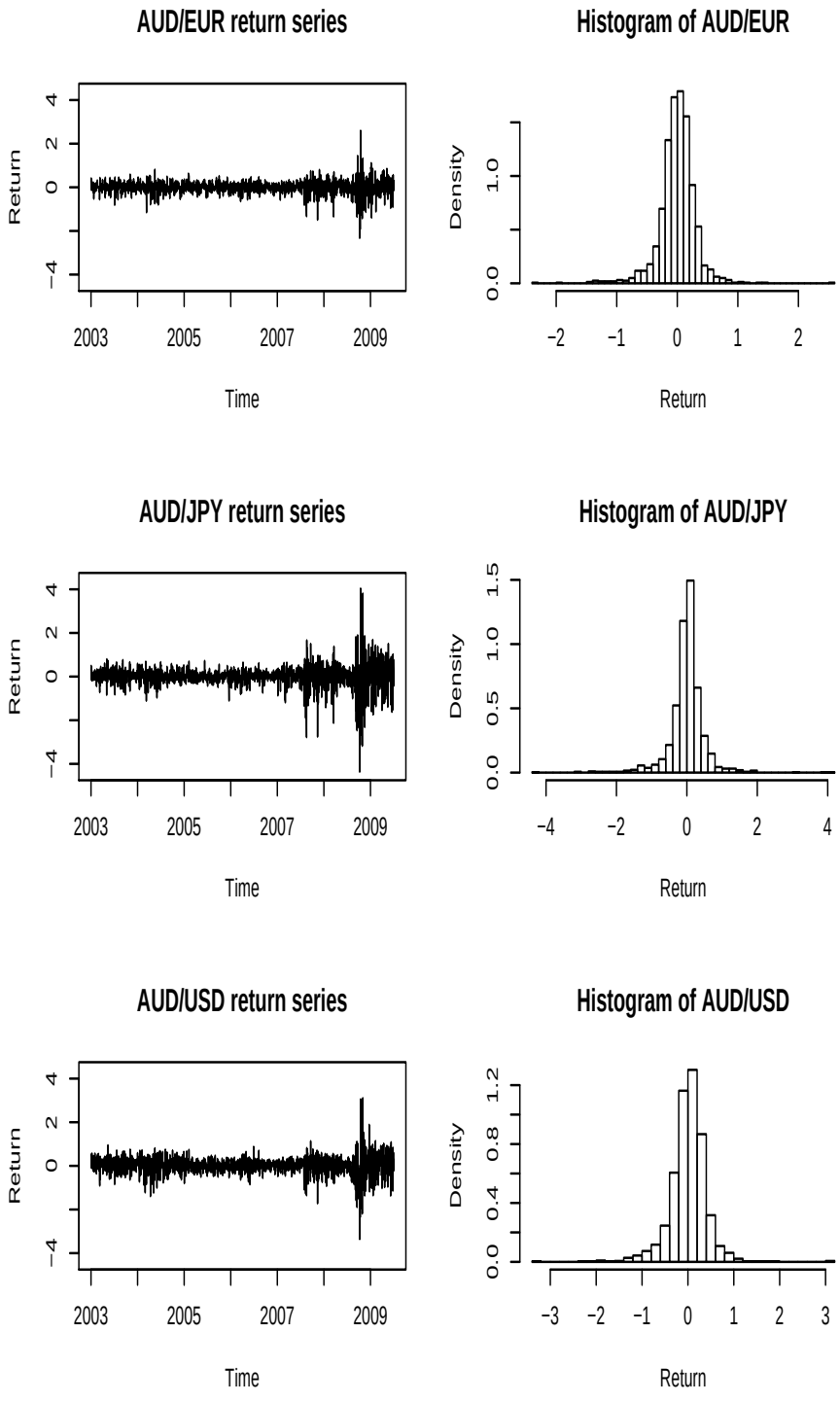


FIGURE 3.2. Time series plots and histograms for AUD/EUR, AUD/JPY and AUD/USD exchange rate data.

To implement the GT-GT model in the `WinBUGS` package, we set the hyperparameter values of the prior distribution in (3.9) as:

$$\begin{aligned}
\psi_0 &\sim N(0, 100), \\
\psi_1 &\sim U(-1, 1), \\
\mu &\sim N(-3, 100), \\
\tau^2 &\sim IG(0.001, 0.001), \\
\phi^* &\sim Be(20, 1.5).
\end{aligned}$$

Instead of using the two-stage estimation procedure as in McDonald and Newey (1988), we estimate the shape parameters p_y, γ_y, p_h, γ_h simultaneously with other model parameters from their full conditional posterior distribution. We assign uniform prior distributions with suitable range to these parameters as follows:

$$\begin{aligned}
p_y &\sim U(0.5, 4), \\
\gamma_y &\sim U(0.01, 1.5), \\
p_h &\sim U(0.5, 4), \\
\gamma_h &\sim U(0.01, 1.5).
\end{aligned}$$

The reason for choosing these particular sets of hyperparameters for the shape parameters is that these values cover a wide range of GT distributions. Moreover, the prior distribution $\gamma_h \sim U(0.01, 1.5)$ restricts p and q to satisfy the constraint $0.67 < pq < 100$ which prevents the GT distribution from being too cuspidate. As shown in Section 3.3.1, the full conditional distributions for these parameters are not standard and we adopt the Metropolis-Hastings technique to sample random variates from these distributions. We conduct a single Markov chain for 300,000 iterations. The initial 100,000 iterations are discarded as the burn-in period to ensure convergence. Simulated values from the Gibbs sampler after the burn-in period are taken from every 50th iteration to reduce the high autocorrelation between successive values. The resulting sample of size 4,000 is used for posterior inference. Table 3.2 presents the posterior summaries and the DIC values for these two data sets. Although we sample the

shape parameter using the reparametrisation $\gamma = 1/(pq)$, we choose to report the estimate of q, as it leads to easier interpretations.

For the AUD/EUR exchange rate data, we observe that the unconditional mean of the log-volatilities μ is smaller under the t-t SV model with a smaller standard deviation. On the other hand, the GT-GT SV model gives lower estimate of the volatility persistence parameter ϕ than the corresponding t-t SV model. The estimates of the scale parameter of the log-volatility distribution, τ are very close under the GT-GT and the t-t SV models, although the variance of the GT distribution depends on p_h, q_h and τ. The estimates of the intercept ψ_0 in the return equation are close to 0 under both distributions, and both the 95% credible intervals include 0, indicating that this parameter is not significant. For the covariate parameter ψ_1, it is estimated to be 0.450 and 0.451 under the GT-GT and t-t SV models, respectively and both 95% credible intervals do not cover 0. The significance of the estimated ψ_1 shows the positive and significant impact of the AUD/USD exchange rate on the AUD/EUR exchange rate over the sampling period. Substituting the shape parameters estimates p_y, q_h and volatility parameters ϕ and τ into (3.8), we obtain the unconditional kurtosis of the returns to be 5.206 and 3.923 for the GT-GT and t-t SV models, respectively. This indicates that the GT-GT SV model is able to account for the heavy-tailedness of the returns to a greater extent than the t-t SV model. The Student-t distribution is obtained by setting the GT distribution with $p = 2$ and has $2q$ degrees of freedom. Hence with the estimated q_y and q_h parameter values, we obtain Student-t distribution with 13.17 and 7.63 degrees of freedom for the returns and log-volatilities, respectively. Figure 3.3 displays the time series plot of the smoothed estimates of the unobserved volatilities under the GT-GT and t-t SV models. Since the GT distribution we used is not standardised to have variance 1, we adjust the volatility estimates by the variance of the GT distribution given by (3.2). As can be observed from the plot, the estimated volatilities are very similar under the two distributions although a close inspection of the plot reveals that the volatilities are generally smaller under the GT-GT SV model.

TABLE 3.2. Posterior means, standard errors, lower and upper limits of 95% Bayes interval.

GT	AUD/EUR			AUD/JPY		
	Mean	sd	95% CI	Mean	sd	95% CI
μ	-2.575	0.1084	(-2.912, -2.406)	-2.345	0.1495	(-3.053, -2.857)
ϕ	0.917	0.0214	(0.881, 0.956)	0.938	0.0210	(0.894, 0.970)
τ	0.294	0.0051	(0.289, 0.308)	0.290	0.0040	(0.278, 0.305)
ψ_0	0.001	0.0042	(-0.007, 0.010)	0.022	0.0049	(0.010, 0.030)
ψ_1	0.450	0.0170	(0.422, 0.488)	0.581	0.0360	(0.523, 0.630)
p_y	3.032	0.4427	(2.471, 3.957)	2.906	0.3611	(2.484, 3.778)
p_h	2.228	0.2045	(1.867, 2.588)	1.443	0.1671	(1.195, 1.767)
q_y	3.065	0.5630	(2.401, 3.904)	1.888	0.3802	(1.076, 2.411)
q_h	2.647	0.1803	(2.332, 2.968)	3.635	0.4147	(2.943, 4.345)
DIC	-1740.670			-1613.310		
Student-t						
μ	-2.718	0.0262	(-2.927, -2.517)	-2.390	0.1990	(-2.795, -2.009)
ϕ	0.9315	0.0201	(0.876, 0.959)	0.967	0.0080	(0.952, 0.981)
τ	0.295	0.0066	(0.289, 0.315)	0.293	0.0050	(0.289, 0.308)
ψ_0	0.005	0.0047	(-0.001, 0.009)	0.016	0.0021	(0.012, 0.020)
ψ_1	0.451	0.0162	(0.422, 0.485)	0.582	0.0440	(0.492, 0.632)
q_y	6.585	0.6620	(5.773, 7.254)	4.163	0.5316	(3.141, 5.422)
q_h	3.813	0.7030	(3.101, 4.156)	3.145	0.4656	(2.549, 4.230)
DIC	-2020.880			-1299.570		

Looking at the results for AUD/JPY exchange rate return in Table 3.2, we found similar results to those for AUD/EUR exchange rate returns. Firstly, the estimated unconditional mean of the log-volatilities μ is again slightly smaller under the t-t SV model and both μ estimates (under the GT and Student-t distributions) are larger than those for AUD/EUR data. This shows that the marginal mean volatility is higher for the AUD/JPY return data. The estimated volatility persistence ϕ is also higher for AUD/JPY as compared to the AUD/EUR

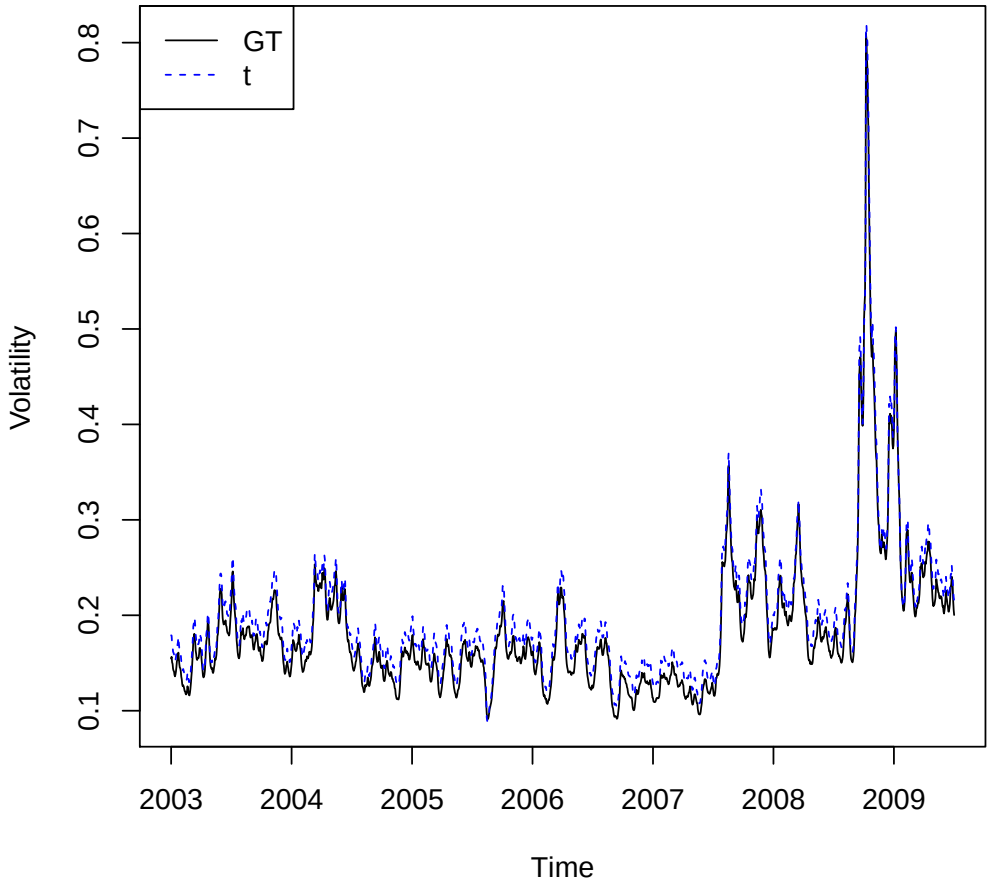


FIGURE 3.3. Comparison of estimated volatilities under SV model with GT-GT and t-t distribution for AUD/EUR exchange rate return.

data with the estimate relatively larger under the t-t SV model. Unlike the insignificant estimates of ψ_0 for AUD/EUR, the estimates of ψ_0 are positive and significant for AUD/JPY under both GT-GT and t-t SV models as both 95% credible intervals do not include 0. This may due to the higher level of overall mean for AUD/JPY exchange rate return reported in Table 3.1. Significant and positive estimate of ψ_1 is also obtained for AUD/JPY return data, suggesting the AUD/USD exchange rate also has a positive influence on the exchange rate of AUD/JPY. In particular, with a larger magnitude of the estimated ψ_1 for AUD/JPY, the effect of AUD/USD exchange rate is stronger for AUD/JPY as compared with AUD/EUR data. For both the AUD/EUR and AUD/JPY exchange rate returns, the kurtosis exists under both GT and Student-t distributions as the product of the posterior means of $p_y q_y$ and $p_h q_h$

for returns and log-volatilities are both greater than 4. The unconditional kurtosis of the AUD/JPY exchange rate return is similar for GT-GT and t-t SV models, with values 9.717 and 9.027 respectively. These values suggest that the AUD/JPY exchange rate return data is more heavy-tailed than the AUD/EUR series, after taking into account the time-varying volatilities. This is also reflected in the estimated degrees of freedom, which is 8.33 for returns and 6.20 for log-volatilities, when a Student-t distributions is fitted, and they are both smaller than the corresponding values for AUD/EUR data. Figure 3.4 shows the plot of the smoothed estimates of volatilities under both GT-GT and t-t SV models for the AUD/JPY series. Similar to those obtained for AUD/EUR data, volatility estimates are very close under the two distributions, with the GT distribution producing slightly smaller estimated volatilities, particularly during the great financial crisis period from 2007 to 2009.

Figure 3.5 depicts the estimated volatilities for the AUD/EUR and AUD/JPY series under the GT-GT SV model. It is obvious from the plot that the volatilities of AUD/JPY are

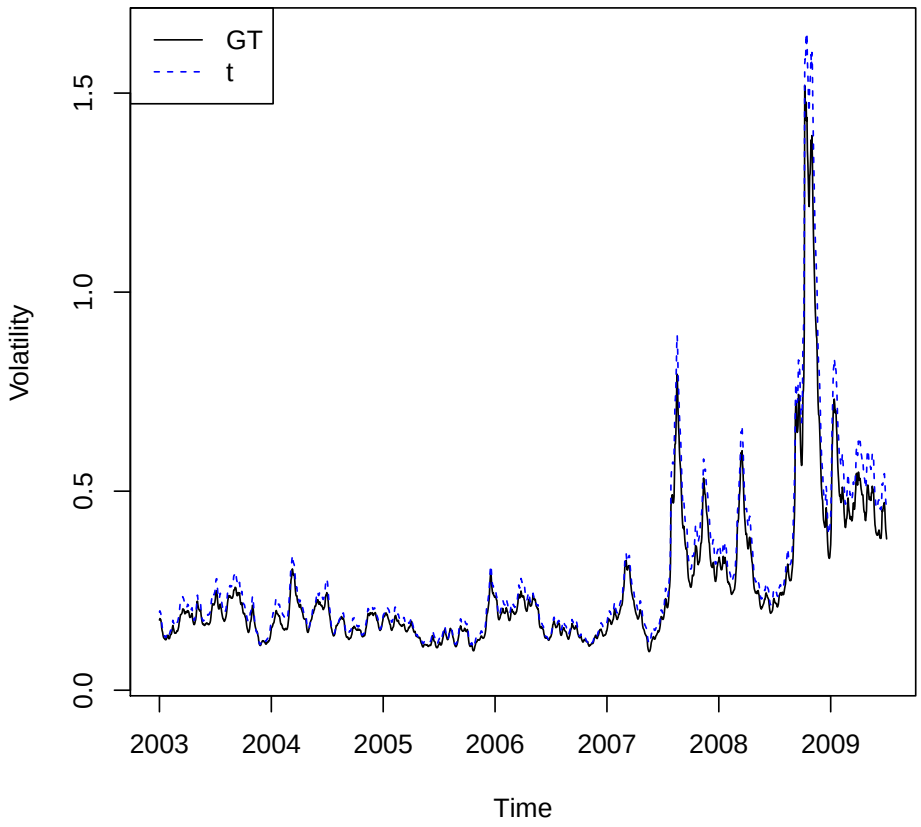


FIGURE 3.4. Comparison of estimated volatilities under SV model with GT-GT and t-t distribution for AUD/JPY exchange rate return.

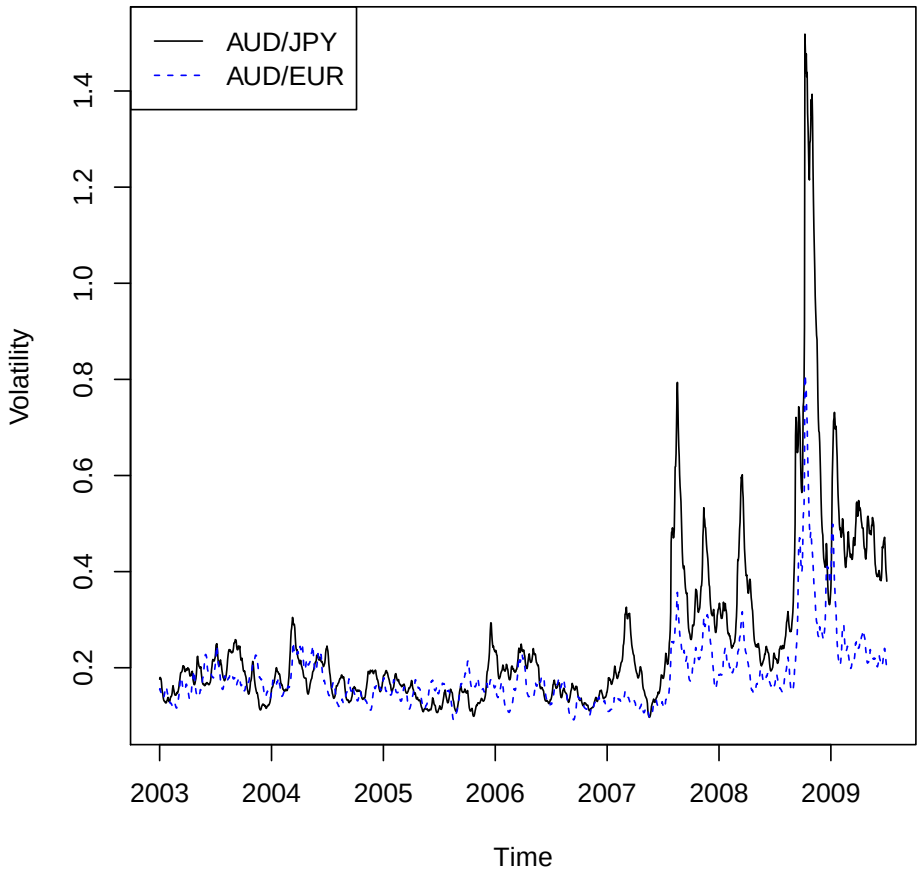


FIGURE 3.5. Comparison of estimated volatilities between AUD/EUR and AUD/JPY series under SV model with GT-GT distribution.

generally higher than that of AUD/EUR. This finding is not out of our expectation as the time series plot in Figure 3.2 also shows a greater degree of fluctuation in the returns of AUD/JPY and hence higher volatility is expected. This result is also in accordance with the result we obtained in Chapter 2 where the degrees of freedom, estimated from a SV model with leverage, is the smallest for the AUD/JPY pair of exchange rate return among the ten currency pairs analysed. In fact, it has been recognised that the AUD/JPY pair is one of the highly volatile pair in foreign exchange market. One contributing factor is the interest rate differential between the two countries (currencies). In the beginning of 2007, the Reserve Bank of Australia had interest rate at 6.25% while the Bank of Japan had interest at 0.5%. This large interest rate differential may help to explain the observed volatility peak for AUD/JPY series in the beginning of 2007. On the other hand, the interest rate in the Euro area (set by the

European Central Bank) in 2007 was 3.75 to 4%. The interest rate differential between Australia and Europe is smaller than that between Australia and Japan. For both exchange rate pairs, increased volatility is observed from the second half of 2007 with the highest volatility at around the end of 2008.

The DIC values for comparing the GT-GT and t-t SV models for the two data sets are reported in Table 3.2. For the AUD/EUR data, the DIC for GT-GT and t-t SV models are -1741 and -2021, respectively, suggesting that the t-t SV model is preferred. It is worth noticing, that the shape parameter of the GT distribution for log-volatilities p_h is estimated to be 2.228 and more importantly, the 95% credible interval includes the value 2. This suggests that the GT distribution employed for the log-volatilities can be reduced to the Student-t distribution. On the other hand, the GT-GT SV model performs better than the t-t SV model for the AUD/JPY data as the DIC values are -1613 and -1300 respectively. Moreover, the 95% credible intervals of p_y and p_h do not support the reduction from the GT distribution to the Student-t distribution.

By substituting smoothed estimates of the log-volatilities and other model parameter estimates for the returns, we calculate the standardised returns for the AUD/EUR and AUD/JPY data sets. Figure 3.6 and 3.7 plot the standardised returns for the AUD/EUR and AUD/JPY data respectively. The two return series can be fitted well by both the GT and Student-t distributions. An investigation of the standardised returns reveals that they are slightly negatively skewed. In fact, the skewness of the standardised returns of AUD/EUR is found to be -0.167 under the GT distribution and -0.185 under the Student-t distribution while the skewness of the standardised returns of AUD/JPY is calculated to be -0.169 and -0.208 respectively for the GT and Student-t distributions. The symmetric GT and Student-t distributions focus on leptokurtic distributions, but it is also worthwhile to consider fitting skewed distributions, which may result in a better model fit.

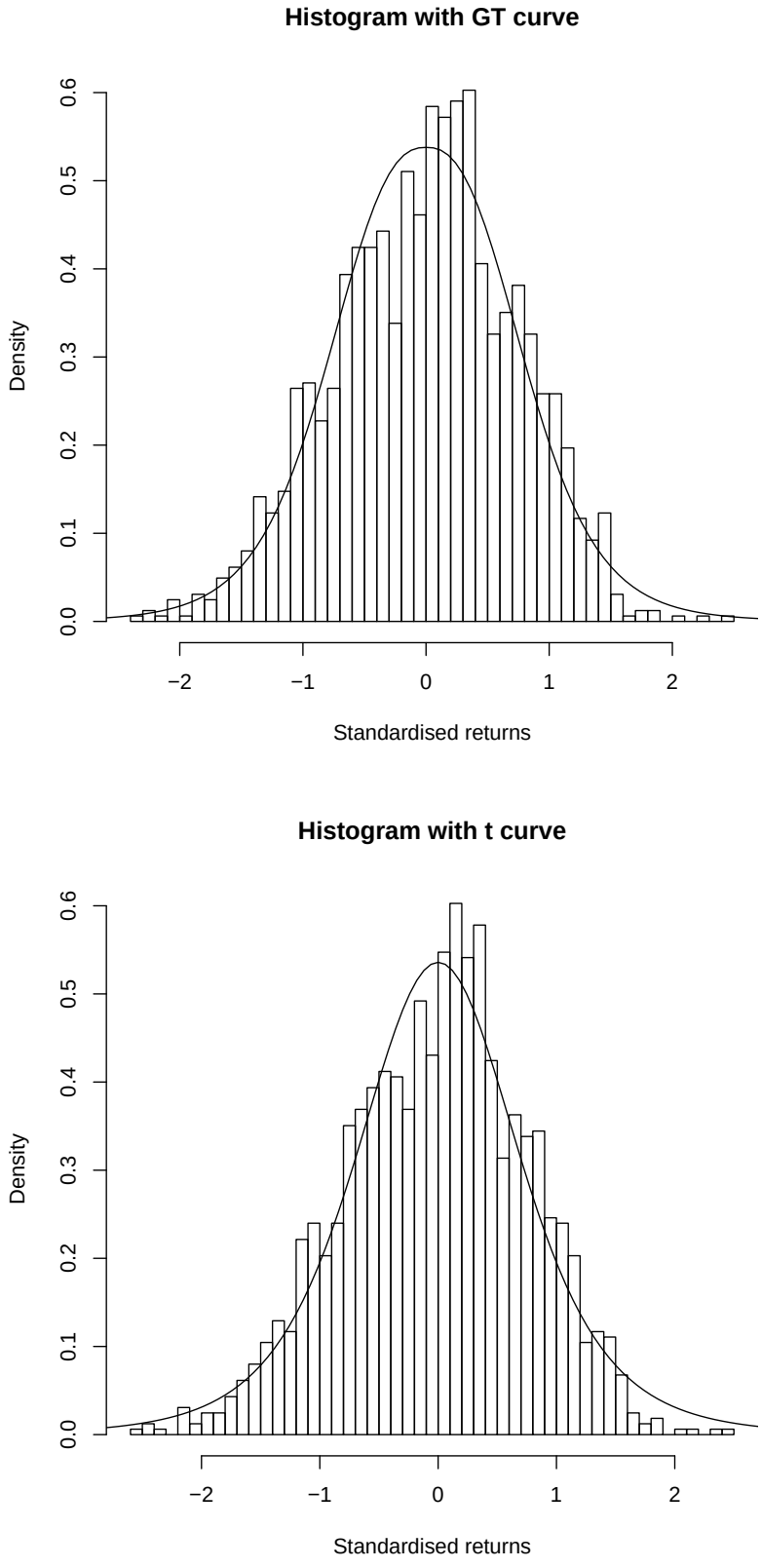


FIGURE 3.6. Comparison of standardised returns of GT and Student-t distributions for AUD/EUR data.

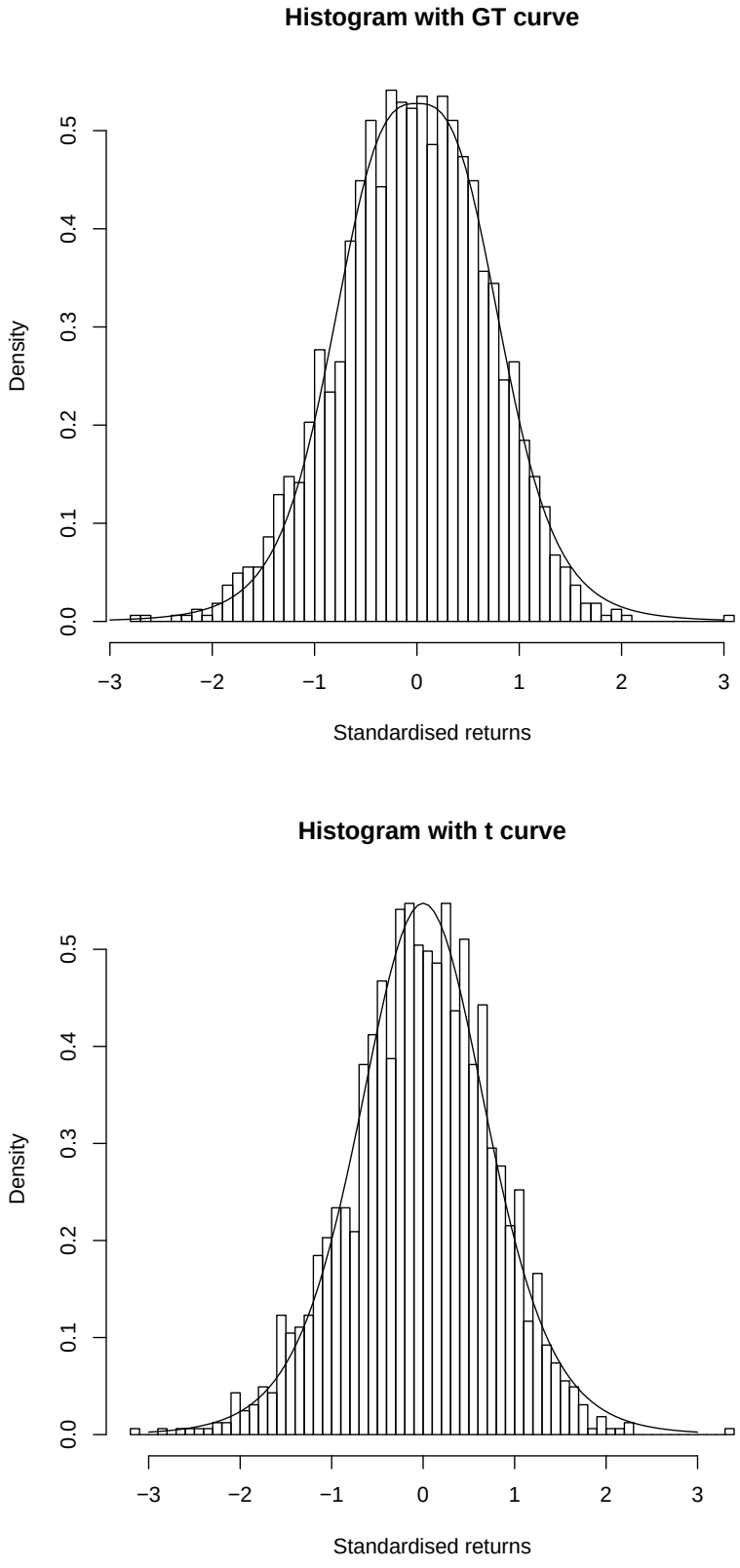


FIGURE 3.7. Comparison of standardised returns of GT and Student-t distributions for AUD/JPY data.

3.5.1. Outlier diagnostics. It is well-known that heavy-tailed distributions can provide a robust inference by accommodating for outliers. Choy and Smith (1997) and Choy and Chan (2008) proposed to perform outlier diagnostics using the latent mixing parameters of the normal and uniform scale mixture distributions. A normal scale mixture distribution accommodates an outlier by inflating the variance component of the corresponding normal distribution and a uniform scale mixture distribution do so by widening the support of the uniform distribution. In our analysis, we compare the posterior means of the mixing parameters for each observation and each volatility to identify possible outliers. For the GT distribution, we use ω_{y_t} and ω_{h_t} as proxies to identify outliers and large values of ω_{y_t} and ω_{h_t} indicate possible outlying returns and outlying log-volatilities, respectively.

For the AUD/EUR data, we detect 5 most outlying returns as y_{299} (12/03/2004), y_{1017} (24/01/2007), y_{1303} (17/03/2008), y_{1217} (12/11/2007) and y_{524} (7/02/2005) in descending order of ω_{y_t}. The most outlying log-volatilities are identified to be h_{1443} (7/10/2008), h_{1428} (15/09/2008), h_{1439} (30/09/2008), h_{1429} (16/09/2008) and h_{1442} (3/10/2008). Table 3.3 presents the posterior means for five largest mixing parameters ω_{y_t} and ω_{h_t} under the GT and Student-t distribution. The same set of outlying returns is identified under the two distributions. However, two observations (h_{1428} and h_{1442}) which are identified to be among the five most outlying log-volatilities under the GT distribution, are not identified as being among the five most outlying log-volatilities under the Student-t distribution. Figure 3.8 displays the posterior means of ω_{y_t} and ω_{h_t} for AUD/EUR data under the GT and Student-t distributions. It appears that for the outlying returns, the magnitude of the mixing parameters is greater under the Student-t distribution. This result can be explained by our earlier finding that the GT distribution accommodates outliers better by having a larger kurtosis. In order to accommodate the same outlying effect, the Student-t distribution would require larger values for the mixing parameters ω_{y_t} to widen the support of the uniform distribution in the SMU form.

For the AUD/JPY pair, five most outlying returns are detected to be y_{611} (14/06/2005), y_{1004} (5/01/2007), y_{1217} (12/11/2007), y_{180} (22/09/2003) and y_{93} (20/05/2003). Among these five outlying returns, y_{611}, y_{1004}, y_{1217} and y_{93} are also identified as outlying returns under the Student-t distribution although in a slightly different order and instead of y_{180}, the Student-t distribution identifies y_{1261} (16/01/2008) as one of the five most outlying returns with a mixing parameter value of 2.256. As for the outlying log-volatilities, they are identified

to be h_{1443} (7/10/2008), h_{1144} (27/07/2007), h_{1422} (5/09/2008), h_{1439}(30/09/2008) and h_{1503} (2/01/2009) under the GT distribution. The first four outlying volatilities are also identified as outlying when the Student-t distribution is fitted to the model. However, h_{1442} (3/10/2008) which is detected as an outlying volatility under the Student-t distribution, is not detected as outlying under the GT distribution. Table 3.4 shows the posterior means of five largest mixing parameters under GT and Student-t distributions for the AUD/JPY data and Figure 3.9 displays the posterior means of ω_{y_t} and ω_{h_t} for AUD/JPY data under the GT and Student-t distributions. The values of the mixing parameters for the outlying returns are actually quite close under the GT and Student-t distributions, since similar kurtosis, about 9, is estimated for the returns under both distributions. It is interesting to observe that some outlying returns and/or outlying volatilities are common to both AUD/EUR and AUD/JPY exchange rate returns. For example, y_{1217} (12/11/2007) is identified as outlying return for both AUD/EUR and AUD/JPY under the GT and Student-t distributions and the outlying h_{1443} (7/10/2008) and h_{1439} (30/09/2008) are also common to both AUD/EUR and AUD/JPY pairs. This may be due to the fact that both exchange rate pairs are equally affected by significant market movements at the time, for example, in the global financial crisis in 2008.

Although the GT distribution can offer flexible leptokurtic error distribution for the return and volatility processes, the SV model considered here, does not allow the leverage effect, which is the negative correlation between the returns and volatilities commonly observed in financial asset return series. The leverage effect is not a crucial issue in the above study, since we use the exchange rate data in the empirical analysis which typically do not exhibit significant leverage effect. In the next section, we will extend the SV model in (3.7) to allow for asymmetric effects in the volatility and we will compare different specifications of asymmetry through an empirical study.

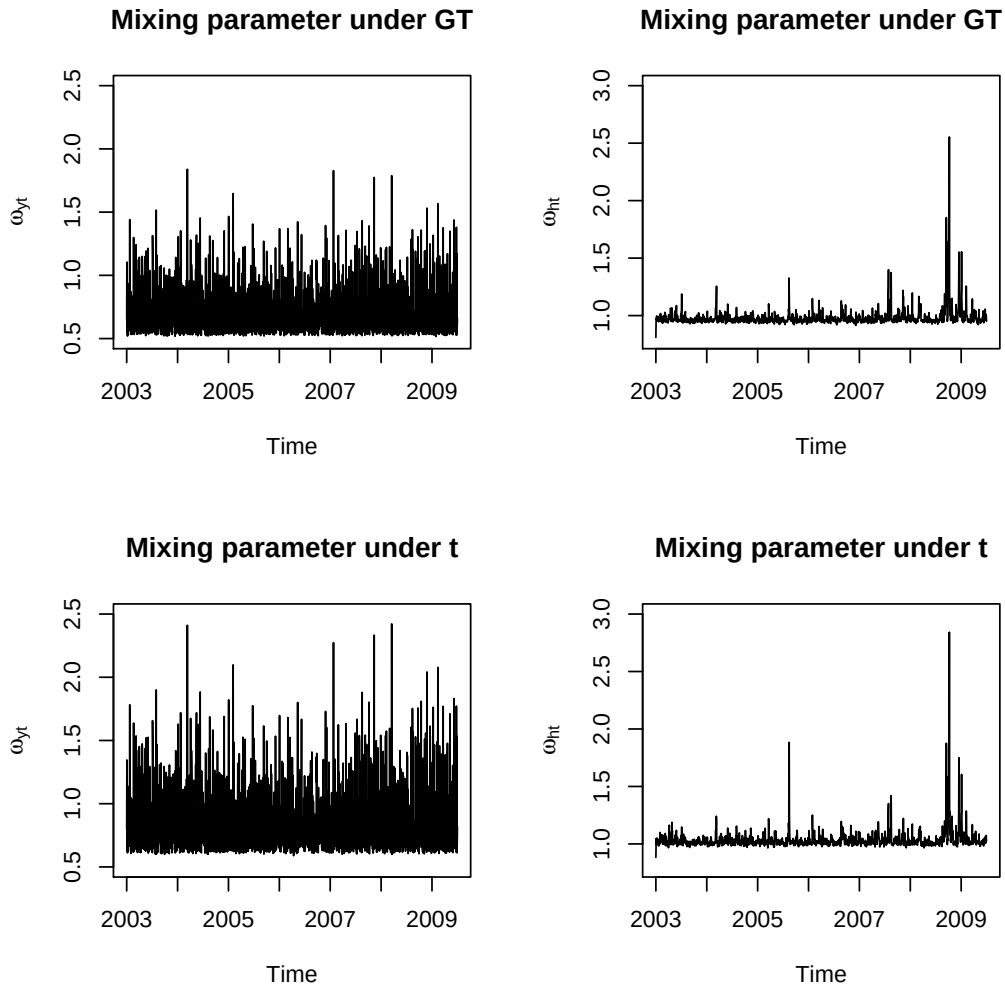


FIGURE 3.8. Plots of the posterior means of ω_{y_t} and ω_{h_t} for AUD/EUR exchange rate under GT and t distribution.

TABLE 3.3. Bayes estimates of the combined mixing parameters ω_t for outliers: AUD/EUR data.

Outlying Return (ω_{y_t})			Outlying Volatility (ω_{h_t})		
Day (t)	GT	t	Day (t)	GT	t
12/03/2004 (299)	1.839	2.411	7/10/2008 (1443)	2.554	2.842
24/01/2007 (1017)	1.828	2.274	15/09/2008 (1428)	1.852	1.876
17/03/2008 (1303)	1.788	2.422	30/09/2008 (1439)	1.649	1.876
12/11/2007 (1217)	1.774	2.333	16/09/2008 (1429)	1.620	1.825
7/02/2005 (524)	1.647	2.089	3/10/2008 (1442)	1.608	1.695

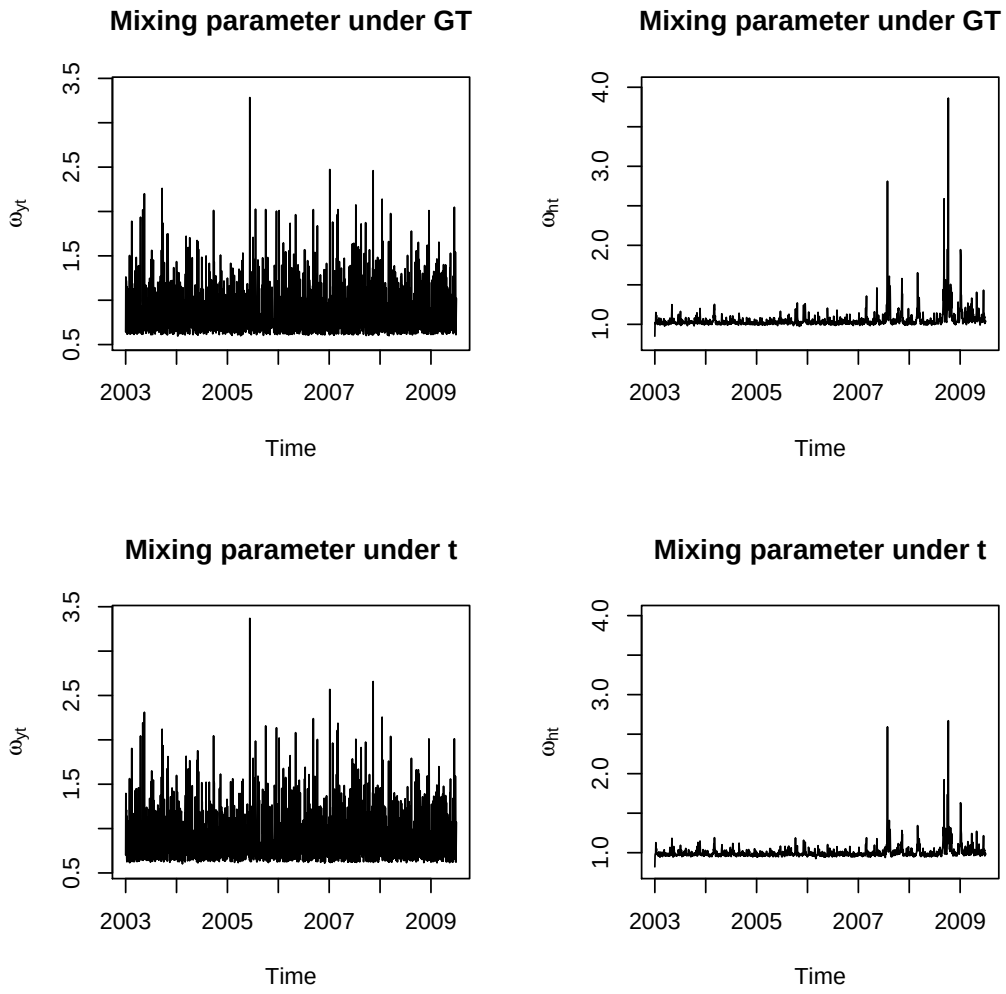


FIGURE 3.9. Plots of the posterior means of ω_{y_t} and ω_{h_t} for AUD/JPY exchange rate under GT and t distribution.

TABLE 3.4. Bayes estimates of the combined mixing parameters ω_t for outliers: AUD/JPY data.

Outlying Return (ω_{y_t})			Outlying Volatility (ω_{h_t})		
Day (t)	GT	t	Day (t)	GT	t
14/06/2005 (611)	3.284	3.369	7/10/2008 (1443)	3.863	2.670
5/01/2007 (1004)	2.473	2.568	27/07/2007 (1144)	2.811	2.592
12/11/2007 (1217)	2.460	2.657	5/09/2008 (1422)	2.589	1.923
22/09/2003 (180)	2.260	2.120	30/09/2008 (1439)	1.951	1.738
20/05/2003 (93)	2.200	2.311	2/01/2009 (1503)	1.945	1.632

3.6. Asymmetric SV models

The concept of the 'leverage' effect, in which negative shocks to return increase the future volatility to a greater extent than positive shocks of similar magnitudes, is explored in Chapter 2. In that chapter, we employed a bivariate Student-t distribution via SMN to model this negative correlation. It is not easy, however, if we extend the GT-GT SV model in (3.7) to incorporate the leverage effect due to the complexity of the multivariate GT distribution. This direct (negative) correlation between the innovations of the returns and volatilities, which is referred to as the 'dynamic leverage' in Asai and McAleer (2005a), is a specific type of asymmetric effects. Other types of asymmetry have been proposed, for example by So *et al.* (2002) and Asai and McAleer (2005a). So *et al.* (2002) proposed a threshold SV model, in which the constant term and the autoregressive parameters in both the return and volatility equations vary according to the sign of the previous day return. However, the magnitude of the previous return is neglected in their model. The formulation in So *et al.* (2002) is extended by Chen *et al.* (2008). Asai and McAleer (2005a) developed the 'dynamic asymmetric leverage' SV model. In their model, apart from allowing for the direct negative correlation between the return and volatility innovations, the sign as well as the magnitude of the previous return is explicitly included in the volatility process. However, as Gaussian error distribution is used for the returns, the model of Asai and McAleer (2005a) lacks the flexibility to accommodate the leptokurtic characteristic of the returns. We will employ the asymmetric SV model, which is a special case of the dynamic asymmetric leverage model of Asai and McAleer (2005a) when the direct correlation between the innovations of the returns and volatilities is zero. In addition, we will extend the Gaussian specification of the return errors to the case of the GT distribution. The threshold SV model of So *et al.* (2002) will also be considered, as is done in Asai and McAleer (2005a).

3.7. Asymmetric GT SV model specifications

For simplicity, we will drop the exogenous variable in the return equation and assume that the volatilities in the SV model follow a Gaussian process. Hence the SV model in (3.7)

becomes

$$\begin{aligned} y_t &= \exp\left(\frac{h_t}{2}\right)\epsilon_t, \quad t = 1, \ldots, n \\ \epsilon_t &\sim GT(0, 1, p, q), \\ h_{t+1}|h_t &= \mu + \phi(h_t - \mu) + \tau\eta_t, \\ h_1 &\sim N\left(\mu, \frac{\tau^2}{1-\phi^2}\right), \\ \eta_t &\sim N(0, 1). \end{aligned}$$

We refer to this SV model without leverage or asymmetric effects as the GT-SV model. In the asymmetric SV model (GT-ASV), we allow the volatilities to be affected by both the sign and the magnitude of the previous return as follows:

$$h_{t+1}|h_t, y_t = \mu + \phi(h_t - \mu) + \gamma_1 y_t + \gamma_2|y_t| + \tau\eta_t. \tag{3.10}$$

When $\gamma_1 = 0$ and $\gamma_2 > 0$, the asymmetric model reduces to the 'alternative asymmetric leverage' SV model in Asai and McAleer (2005a), where the positive and negative shocks increase the future volatility by the same magnitude. Asai and McAleer (2011) proposed an asymmetric model similar to (3.10) where the unobserved innovations ϵ_t and $|\epsilon_t|$ are included in the volatility equation instead of the observed return y_t and $|y_t|$ to capture standardised size effects. Their proposed model is in fact a stochastic version of the EGARCH model of Nelson (1991). We include the information contained in the previous returns rather than previous shocks as it is noted in Asai and McAleer (2005b) that is it not computationally straightforward to incorporate the previous normalised shocks in the framework of SV models. In order to have the 'leverage' effect, we need $\gamma_1 < 0$, which implies one-unit of positive shock would have a smaller effect on the volatility than one-unit of negative shock. According to Black (1976), this effect may also imply $\gamma_1 < 0$ and $\gamma_1 < \gamma_2 < -\gamma_1$.

It is argued in Asai and McAleer (2005a) that the sign of γ_2 has no effects on the asymmetry and the effect of its magnitude can be absorbed by the values of γ_1. In addition, their empirical results demonstrate that some estimates of γ_2 are insignificant. Hence we consider, as our second asymmetric SV model, which is a special case of (3.10) with $\gamma_2 = 0$ as

$$h_{t+1}|h_t, y_t = \mu + \phi(h_t - \mu) + \gamma_1 y_t + \tau\eta_t. \tag{3.11}$$

We refer to this asymmetric model as GT-ASV2. Although this model is not estimated in the empirical study, Asai and McAleer (2005a) discussed some interesting properties of this model. In particular, they have shown that when $\gamma_1 y_t$ is incorporated into the volatility equation, not only it induces correlation between the returns and volatilities, but also this correlation structure is time-varying. We will extend their result to our case where the returns follow a GT distribution. Consider (3.11) and define $\xi_t = \gamma_1 \exp(h_t/2)\epsilon_t + \tau\eta_t$ to be the general innovation of the volatility. Then given h_t, the expected value and variance of ξ_t are

$$\begin{aligned} E(\xi_t) &= 0, \\ \text{var}(\xi_t) &= \gamma_1^2 \exp(h_t)\text{var}(\epsilon_t) + \tau^2 \text{var}(\eta_t) \\ &= \gamma_1^2 \exp(h_t) \frac{\Gamma\left(\frac{3}{p}\right)\Gamma\left(\frac{1}{p}\left(\frac{1}{\gamma}-2\right)\right)}{(p\gamma)^{\frac{2}{p}}\Gamma\left(\frac{1}{p}\right)\Gamma\left(\frac{1}{p\gamma}\right)} + \tau^2, \end{aligned}$$

using the variance of the GT distribution given in (3.2) and recall $\gamma = 1/(pq)$. Therefore, the correlation between ϵ_t and ξ_t, given h_t, is given by

$$\begin{aligned} \rho_t := \text{cor}(\epsilon_t, \xi_t) &= \frac{E(\epsilon_t \xi_t)}{\sqrt{\text{var}(\epsilon_t)\text{var}(\xi_t)}} \\ &= \frac{E(\gamma_1 \exp(h_t/2)\epsilon_t^2)}{\sqrt{\text{var}(\epsilon_t)\text{var}(\xi_t)}} \\ &= \frac{\gamma_1 \exp(h_t/2)\sqrt{\text{var}(\epsilon_t)}}{\sqrt{\text{var}(\xi_t)}} \\ &= \frac{\gamma_1 \exp(h_t/2)\frac{\sqrt{\Gamma\left(\frac{3}{p}\right)\Gamma\left(\frac{1}{p}\left(\frac{1}{\gamma}-2\right)\right)}}{(p\gamma)^{\frac{1}{p}}\sqrt{\Gamma\left(\frac{1}{p}\right)\Gamma\left(\frac{1}{p\gamma}\right)}}}{\sqrt{\gamma_1^2 \exp(h_t)\frac{\Gamma\left(\frac{3}{p}\right)\Gamma\left(\frac{1}{p}\left(\frac{1}{\gamma}-2\right)\right)}{(p\gamma)^{\frac{2}{p}}\Gamma\left(\frac{1}{p}\right)\Gamma\left(\frac{1}{p\gamma}\right)} + \tau^2}} \\ &= \frac{\gamma_1\sqrt{\Gamma\left(\frac{3}{p}\right)\Gamma\left(\frac{1}{p}\left(\frac{1}{\gamma}-2\right)\right)}}{\sqrt{\gamma_1^2\Gamma\left(\frac{3}{p}\right)\Gamma\left(\frac{1}{p}\left(\frac{1}{\gamma}-2\right)\right) + \exp(-h_t)\tau^2\Gamma\left(\frac{1}{p}\right)\Gamma\left(\frac{1}{p\gamma}\right)(p\gamma)^{\frac{2}{p}}}} \end{aligned} \tag{3.12}$$

When $\gamma_1 < 0$, we have $\rho_t < 0$, implying the correlation between ϵ_t and ξ_t is always negative for negative γ_1. Furthermore, this correlation is time-varying through its dependence on the

time-varying volatility h_t. In particular, when the volatility increases, the negative correlation between the returns and volatilities becomes stronger. The time-varying structure of the correlation inferred by the GT-ASV2 model provides a more flexible way in describing the correlation between the returns and volatilities than the leverage SV model analysed in Chapter 2, where the correlation is specified as a constant.

So *et al.* (2002) proposed a new SV model to capture the asymmetry in the mean and volatility simultaneously by incorporating threshold non-linear structures in the mean and volatility equations. We will adopt this model specification but with two modifications. Firstly, we consider asymmetry only in the volatility equation for the purpose of a direct comparison with previously mentioned asymmetric SV models and secondly, we extend the normality assumption of the return innovations in So *et al.* (2002) to the case of a GT distribution. Our threshold SV model, hereafter, GT-TSV is given by

$$h_{t+1} = \mu_{s_t} + \phi_{s_t} h_t + \tau \eta_t,$$

where s_t a set of indicator variables defined by

$$s_t = \begin{cases} 0, & \text{if } y_t < 0 \\ 1, & \text{if } y_t \geq 0. \end{cases}$$

In the symmetric case, $\mu_0 = \mu_1$ and $\phi_0 = \phi_1$. If $\phi_0 = \phi_1$, then $\mu_0 > \mu_1$ implies that the volatility is higher when the past return is negative than when it is positive. The parameter ϕ_{s_t} measures the effect of the past volatility on the current volatility and if $\phi_0 > \phi_1$, then the past volatility will have a greater influence on the present volatility after a fall in price than after a rise in price. It is hypothesised that it would take longer for the bad news contained in the past variance to be absorbed by the market, hence the volatility is expected to be more persistent when $y_t < 0$. In order to implement the GT-TSV model, we parameterise μ_{s_t} and ϕ_{s_t} as follows

$$\begin{aligned} \mu_{s_t} &= \mu + c s_t \\ \phi_{s_t} &= \phi + d s_t \end{aligned}$$

In general, $\mu_0 = \mu$, $\mu_1 = \mu + c$, $\phi_0 = \phi$ and $\phi_1 = \phi + d$. When $c = d = 0$, we obtain the special case which corresponds to the symmetric model.

3.8. An empirical study

Asai and McAleer (2005a) used the Monte Carlo likelihood method for the estimation of asymmetric SV models and So *et al.* (2002) employed the Bayesian MCMC method for the estimation of TSV model. We will use the MCMC method to facilitate comparison between various asymmetric and threshold SV models and to obtain parameter estimates and smoothed volatilities. We use the same set of priors for the model parameters as in the previous section and in addition, we assign a normal prior to $c \sim N(0, 10)$ and normal prior to d with restricted range $I(|\phi + d| < 1)$. When the volatility process is assumed to be Gaussian, many full conditional distributions are reduced to standard forms, for example, the full conditional distributions for volatility h_t, μ and τ^2 are now truncated normal, normal and inverse gamma, respectively.

We apply the GT-SV, GT-ASV, GT-ASV2 and GT-TSV models to the daily log-returns of Standard & Poors 500 (S&P 500) Index of the USA from January 2005 to December 2008. We have used this data in Chapter 2 for the analysis of the SV model with leverage (SVL). We will however, multiply the returns by 100% to obtain the percentage returns. The S&P 500 return data were also used in the study by Asai and McAleer (2005a) and So *et al.* (2002), although the sampling period in Asai and McAleer (2005a) was from January 1986 to April 2000 and that of So *et al.* (2002) was from 1981 to 1986. For each model, we fit the GT, EP ($\gamma = 0$) and Student-t ($p = 2$) distributions and compare the estimates of the model parameters and smoothed volatilities. We iterate the MCMC algorithm 200,000 times and discard the first 20,000 as burn-in period. Parameters are then sub-sampled from every 30th iteration to reduce the autocorrelation in the samples. A final sample of size 6,000 is used for posterior analysis.

3.8.1. Results. Table 3.5 reports the posterior mean, posterior standard deviation and 95% posterior credible interval for the model parameters under the four models using the GT distribution for the returns. The estimate of the marginal mean of the volatilities μ is the highest under the GT-SV model as it does not incorporate any asymmetry structure based on the past return in the volatility process. Similar μ estimate is obtained between the GT-ASV and GT-ASV2 models, although the GT-ASV model reports higher standard deviation. Under the GT-TSV model, $\mu_0 = 0.1382$ for $y_t < 0$ and $\mu_1 = -0.1003$ for $y_t \geq 0$. This result supports the earlier belief that the volatility tends to be higher when

the previous day return is negative which also can be seen from the significantly negative estimate of c in the GT-TSV model. This result is in contrast to So *et al.* (2002) who found the estimate of c insignificant. Chen *et al.* (2008), however, found this effect significant using the Hang Seng Index and the Nikkei 225. All four models give high estimates close to 1 for the volatility persistence ϕ with lowest value 0.9799 under the GT-ASV2 model. In addition, $\phi_0 = 0.9868$ and $\phi_1 = 0.9815$ for the GT-TSV model which may suggest that the volatility resulting from bad news ($y_t < 0$) is more persistent than those from good news ($y_t \geq 0$). However, this effect may not be significant as the estimate of d is insignificant since the 95% credible interval includes zero. Our result is consistent with the finding of So *et al.* (2002) who also employed the S&P 500 data, although a different sampling period was used. The significant negative estimate of γ_1 indicates the 'leverage' effect, in agreement with the finding in Asai and McAleer (2005a). However, our result indicates that the estimate of γ_2 is close to zero and insignificant and the effect of the magnitude of the past return is absorbed into γ_1, resulting in a more negative estimate of γ_1 in the GT-ASV2 model. On the contrary, Asai and McAleer (2005a) found significant γ_2 for the S&P 500 return, which may due to the different sampling period used, but they found γ_2 insignificant for the TOPIX return. The estimate of the shape parameter p of the GT distribution suggests that the GT distribution can be reduced to the Student-t distribution only in the GT-ASV2 model as the 95% credible interval covers 2.

The significance of the asymmetry parameters γ_1 and c in the GT-ASV, GT-ASV2 and GT-TSV models leads to a clear rejection of the GT-SV model which does not capture the asymmetric structure in the volatility process. Same result is obtained based on the DIC, favouring the asymmetric SV models. Among the asymmetric models, GT-ASV2 is slightly better than GT-ASV and GT-TSV model is worse than both GT-ASV and GT-ASV2. Figure 3.10 shows the comparison of the smoothed estimates of the volatilities, given by $\exp(h_t/2)\sqrt{\frac{\Gamma\left(\frac{3}{p}\right)\Gamma\left(\frac{1}{p}\left(\frac{1}{\gamma}-2\right)\right)}{(p\gamma)^{\frac{2}{p}}\Gamma\left(\frac{1}{p}\right)\Gamma\left(\frac{1}{p\gamma}\right)}}$, under the four models using the GT distribution. The smoothed estimates of latent volatilities are very close under the GT-ASV and GT-ASV2 models, with values overlapping most of the time. The estimated volatilities under the GT-TSV model are reasonably close to that of GT-ASV and GT-ASV2 models during market tranquil period. However, when the market is highly volatile due to the financial crisis, estimated volatilities under the GT-TSV model are lower than those under the GT-ASV and GT-ASV2 models.

The estimated volatilities using GT-SV model show a more smooth movement across time as compared to those under the asymmetric models, indicating that the fine dynamics in the volatility process is not well captured by this model. Figure 3.11 depicts the time-varying correlation between the returns and volatilities under the GT-ASV2 model. Instead of using a single coefficient to describe the correlation between the returns and volatilities as in the SVL model, the time-varying correlation ρ_t describes how this correlation varies across time according to the changes in volatility. It can be seen from the plot, that when the market is more volatile, the leverage effect becomes stronger as the correlation is more negative.

Table 3.6 and 3.7 presents the parameter estimates for the four models using the Student-t distribution (GT distribution with $p = 2$) and the EP distribution (GT distribution with γ=0.01), respectively. In Table 3.6, the results generally lead to similar implications to those for the GT distribution in Table 3.5. There is significant evidence of asymmetries in the GT-ASV, GT-ASV2 and GT-TSV models, which can be inferred from significantly negative estimates of γ_1 and c. Similar to the asymmetric models with GT distribution, GT-ASV and GT-TSV models with Student-t distribution also lead to insignificant estimates of γ_2 and d. The estimated degrees of freedom ($2q$) are 7.52, 8.91, 8.08 and 12.06 for GT-SV, GT-ASV, GT-ASV2 and GT-TSV models respectively. This demonstrates that it is appropriate to employ an error distribution with heavier tails than a normal distribution for the returns to capture the empirical leptokurtosis. Comparing the models based on DIC, the GT-SV model without asymmetry is clearly rejected in favour of the other three asymmetric models. Although γ_2 is insignificant, the full GT-ASV model is still preferred than the GT-ASV2 model and both GT-ASV and GT-ASV2 models are superior to the GT-TSV model.

The results for using the EP distribution presented in Table 3.7 are similar to those obtained from using GT and Student-t distributions. The estimate of γ_1 is significantly negative and μ_0 is significantly different from μ_1. This demonstrates the existence of asymmetric effect in the volatility. The estimate of d is negative which may suggest volatility persistence asymmetry, nonetheless, it is insignificant. The shape parameter p of the EP distribution is estimated to be less than 2 for all four models, and the 95% credible intervals do not include 2. This indicates the return distribution is heavier tailed than the normal distribution, which is expected as the standard property of market return data. DIC again suggests that the GT-ASV model is the best model and GT-TSV model is the worst among the asymmetric models.

In fact, Asai and McAleer (2005a) also drew the same conclusion regarding the inferiority of the GT-TSV model to the dynamic leverage and asymmetric leverage SV models. The inferiority of the GT-TSV model may due to the fact that only the sign of the previous return is incorporated to determine the asymmetry level, while the magnitude which can provide essential information is neglected in the model. The behaviour of the smoothed volatility estimates under the four models when using the Student-t and EP distributions is essentially the same as the GT distribution, as demonstrated in Figures 3.12 and 3.13 respectively.

For each of the GT-SV, GT-ASV, GT-ASV2 and GT-TSV models, we compare the performance of the GT, Student-t and EP distributions. Generally speaking, GT distribution tends to give smaller estimates of μ than the Student-t distribution, which in turn produces smaller values than the EP distribution. Furthermore, the estimates of the standard deviation of the volatility τ are also consistently smaller under the GT distribution across different models. This shows the latent volatilities estimated under the GT distribution have the least variability as compared to the Student-t and EP distributions. On the basis of DIC, GT distribution performs better than both Student-t and EP distributions while Student-t and EP have similar performance as their DIC values are mostly quite close. Figure 3.14 compares the estimated volatilities under the GT, Student-t and EP distributions for each of the four models. It can be seen from the plots, that the model which leads to the greatest difference in the volatility estimates among different distributions is the GT-SV model which does not allow for volatility asymmetry. It can be clearly seen from the volatility plot under the GT-SV model that the GT distribution gives the highest estimates of the volatilities, followed by Student-t and lastly EP which gives the lowest estimates. However, we do not observe this obvious difference in the volatilities when asymmetric SV models are used. Estimated volatilities are indistinguishable across different distributions hence the lines used for plotting volatility estimates under different distributions overlap with each other. This suggests that when we account for volatility asymmetry in the SV model, the difference between estimated volatilities from fitting different error distributions, is minimised or partially compensated by using a more flexible asymmetric model that provides a better description of the dynamics of the volatility process.

TABLE 3.5. Posterior mean, standard deviation (in square brackets) and 95% Bayes interval (in parentheses) obtained from fitting GT distribution.

	GT-SV	GT-ASV	GT-ASV2	GT-TSV
μ	0.3072	0.2750	0.2553	0.1382
	[0.1574]	[0.4622]	[0.1820]	[0.0302]
	(0.0291, 0.5832)	(-0.4783, 1.3350)	(-0.0013, 0.6122)	(0.0799, 0.1982)
ϕ	0.9942	0.9817	0.9799	0.9868
	[0.0035]	[0.0064]	[0.0045]	[0.0085]
	(0.9859, 0.9993)	(0.9667, 0.9917)	(0.9703, 0.9880)	(0.9671, 0.9987)
τ	0.1277	0.0911	0.0897	0.1320
	[0.0224]	[0.0178]	[0.0174]	[0.0213]
	(0.0886, 0.1730)	(0.0671, 0.1352)	(0.0668, 0.1201)	(0.0952, 0.1769)
p	1.706	1.839	1.819	2.391
	[0.1914]	[0.1097]	[0.3181]	[0.2476]
	(1.2646, 1.9975)	(1.6060, 1.9830)	(1.3850, 2.3160)	(2.0100, 2.9110)
q	3.860	5.555	3.815	2.566
	[0.2977]	[1.0750]	[0.5028]	[0.6201]
	(3.2630, 4.3820)	(4.1110, 8.0690)	(3.0060, 4.8410)	(1.5260, 3.7710)
γ_1	-	-0.1319	-0.1445	-
		[0.0221]	[0.0182]	-
	-	(-0.1781, -0.0932)	(-0.1791, -0.1082)	-
γ_2	-	0.0004	-	-
		[0.0083]	-	-
	-	(-0.0145,0.0183)	-	-
c	-	-	-	-0.2385
		-	-	[0.0531]
	-	-	-	(-0.3439, -0.1355)
d	-	-	-	-0.0053
		-	-	[0.0162]
	-	-	-	(-0.0352, 0.0264)
DIC	2228.98	2192.46	2191.11	2221.00

TABLE 3.6. Posterior mean, standard deviation (in square brackets) and 95% Bayes interval (in parentheses) obtained from fitting GT(p=2) distribution.

	GT-SV	GT-ASV	GT-ASV2	GT-TSV
μ	0.4252 [0.1632] (0.1532, 0.6884)	0.3076 [0.4366] (-0.3660, 1.3170)	0.2756 [0.1968] (-0.0923, 0.6882)	0.1334 [0.0303] (0.0768, 0.1949)
ϕ	0.9942 [0.0036] (0.9859, 0.9993)	0.9800 [0.0069] (0.9642, 0.9910)	0.9817 [0.0049] (0.9712, 0.9913)	0.9868 [0.0084] (0.9672, 0.9987)
τ	0.1281 [0.0208] (0.0921, 0.1754)	0.0935 [0.0167] (0.0638, 0.1307)	0.0949 [0.0184] (0.0632, 0.1340)	0.1326 [0.0202] (0.0973, 0.1757)
q	3.760 [0.3451] (3.1940, 4.6530)	4.456 [1.0530] (3.0980, 6.2270)	4.038 [0.8751] (2.6850, 6.2300)	6.029 [1.7190] (3.5590, 9.0230)
γ_1	- - -	-0.1411 [0.0215] (-0.1833, -0.0976)	-0.1368 [0.0265] (-0.1952, -0.0843)	- - -
γ_2	- - -	0.0012 [0.0092] (-0.0145, 0.0213)	- - -	- - -
c	- - -	- - -	- - -	-0.2298 [0.0533] (-0.3378, -0.1299)
d	- - -	- - -	- - -	-0.0048 [0.0158] (-0.0342, 0.0264)
DIC	2256.20	2206.37	2214.51	2231.85

TABLE 3.7. Posterior mean, standard deviation (in square brackets) and 95% Bayes interval (in parentheses) obtained from fitting GT($\gamma = 0.01$) distribution.

	GT-SV	GT-ASV	GT-ASV2	GT-TSV
μ	0.4902	0.4469	0.3865	0.1300
	[0.1781]	[0.5091]	[0.2154]	[0.0323]
	(0.2033, 0.7002)	(-0.3699, 1.6200)	(-0.0201, 0.8331)	(0.0671, 0.1932)
ϕ	0.9936	0.9831	0.9829	0.9869
	[0.0038]	[0.0057]	[0.0039]	[0.0086]
	(0.9845, 0.9992)	(0.9706, 0.9926)	(0.9750, 0.9903)	(0.9662, 0.9987)
τ	0.1342	0.0966	0.09842	0.1330
	[0.0209]	[0.0141]	[0.0144]	[0.0195]
	(0.1016, 0.1810)	(0.0638, 0.1307)	(0.0725, 0.1293)	(0.0997, 0.1771)
p	1.796	1.914	1.822	1.744
	[0.0401]	[0.0509]	[0.0911]	[0.0590]
	(1.7360, 1.8880)	(1.8090, 1.9830)	(1.6940, 1.9950)	(1.6740, 1.8680)
γ_1	-	-0.1239	-0.1197	-
	-	[0.0203]	[0.0195]	-
	-	(-0.1629, -0.0859)	(-0.1588, -0.0842)	-
γ_2	-	-0.0003	-	-
	-	[0.0082]	-	-
	-	(-0.0145, 0.0175)	-	-
c	-	-	-	-0.2238
	-	-	-	[0.0568]
	-	-	-	(-0.3340, -0.1121)
d	-	-	-	-0.0051
	-	-	-	[0.0159]
				(-0.0340, 0.0274)
	2254.80	2205.00	2222.17	2230.93

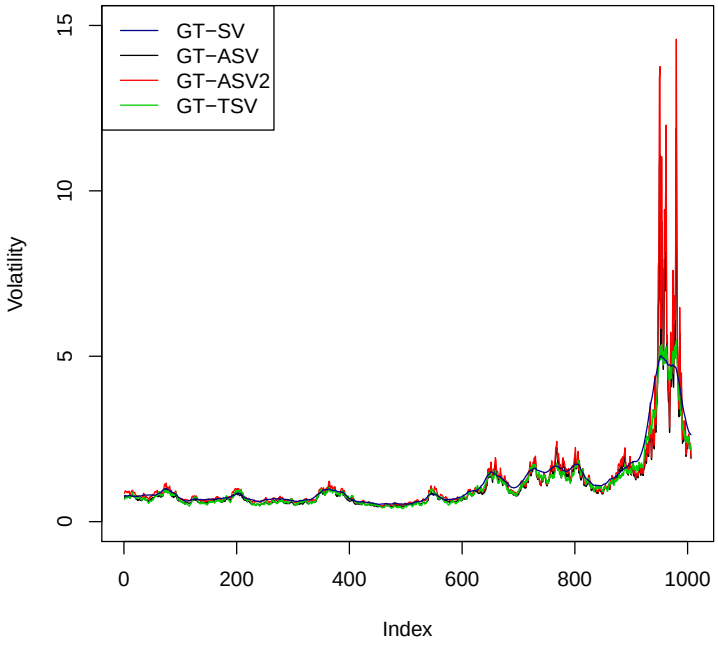


FIGURE 3.10. Comparison of the estimated volatilities estimated from GT-SV, GT-ASV, GT-ASV2 and GT-TSV models under the GT distribution.

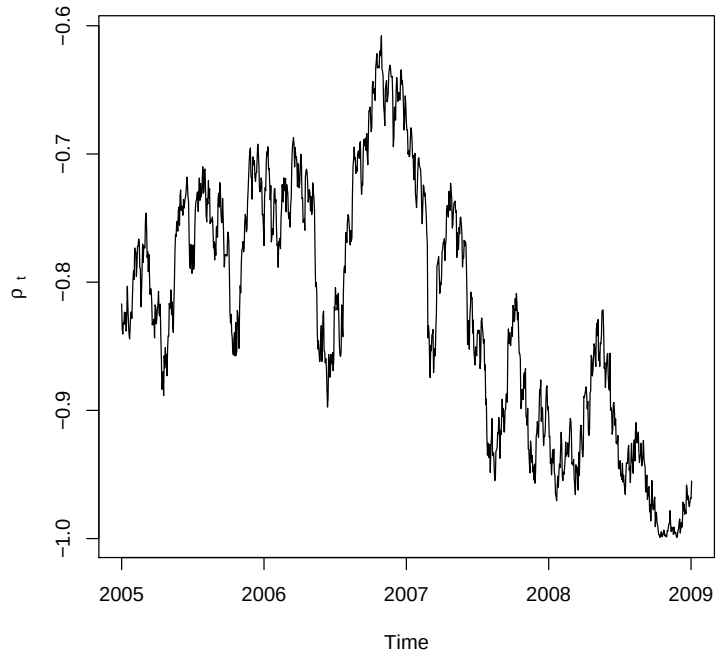


FIGURE 3.11. Plot of the time-varying correlation between returns and log-volatilities in GT-ASV2.

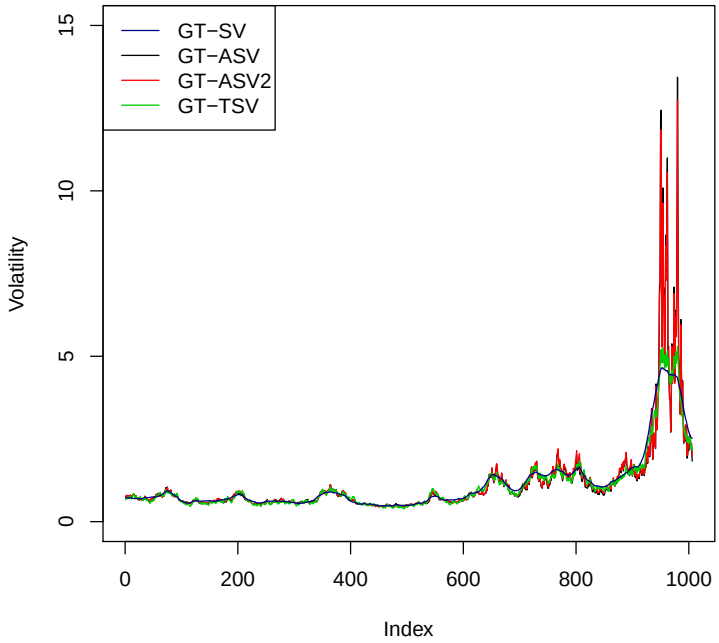


FIGURE 3.12. Comparison of the estimated volatilities estimated from GT-SV, GT-ASV, GT-ASV2 and GT-TSV models under the Student-t distribution.

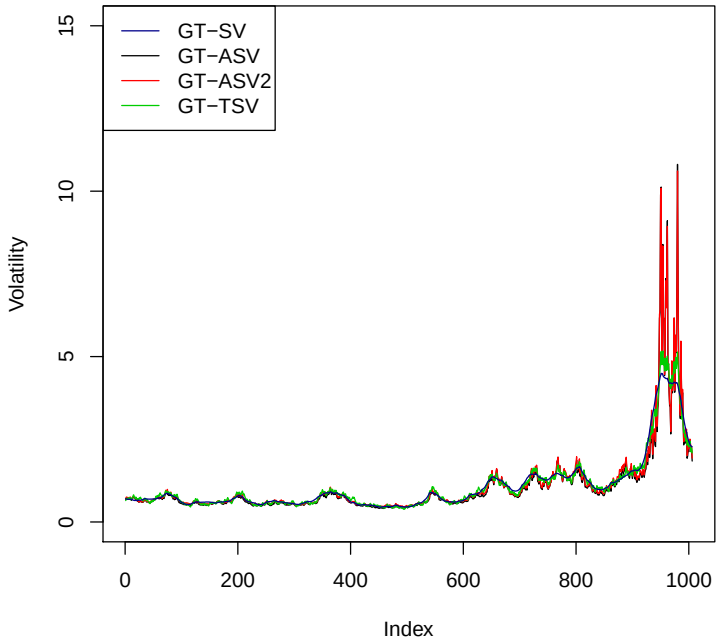


FIGURE 3.13. Comparison of the estimated volatilities estimated from GT-SV, GT-ASV, GT-ASV2 and GT-TSV models under the EP distribution.

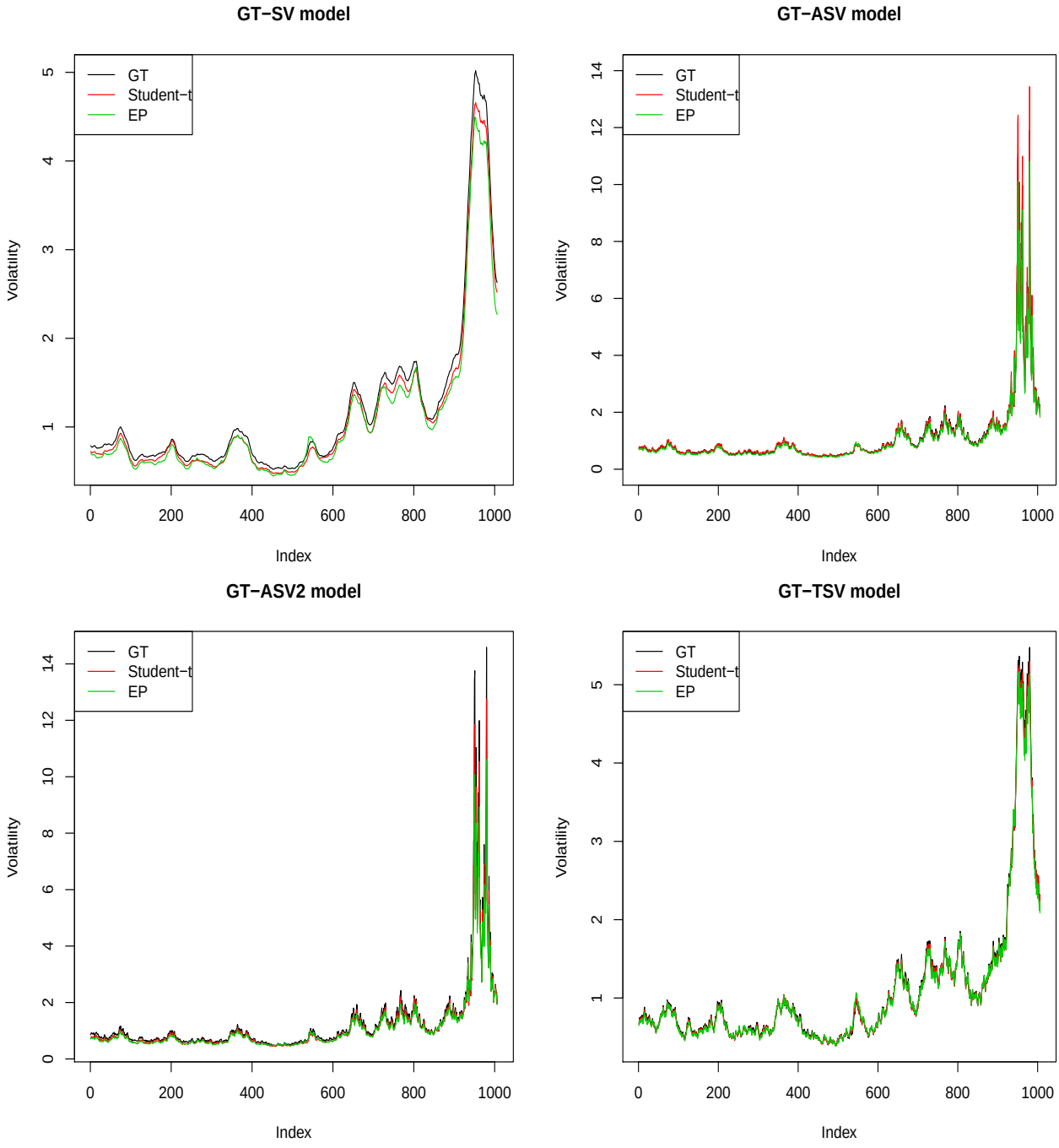


FIGURE 3.14. Comparison of the estimated volatilities using GT, Student-t and EP distribution for each of GT-SV, GT-ASV, GT-ASV2 and GT-TSV models.

3.9. Concluding remarks

In this chapter, we propose to use the GT distribution in the SV model with asymmetry and covariate effect. The extension from the conventional normal error distribution to the GT distribution allows for greater flexibility in modelling the error distribution of the returns and log-volatilities since the GT distribution nests several heavy-tailed distributions including the popular Student-t distribution. We show that by using the SMU representation, we obtain a set of standard full conditional distributions for the Gibbs sampler and the mixing parameters, as the by-product of scale mixture form, are examined for the identification of potential outliers. We implement the GT-GT SV model and its extension to asymmetric specifications using the Bayesian software `WinBUGS` and obtain smoothed estimates of the latent volatilities.

In the empirical study, we fit the SV model with GT and Student-t distributions, the latter is a special case of the GT distribution with $p = 2$, to AUD/EUR and AUD/JPY exchange rate return data, using the AUD/USD exchange rate return as the covariate. Based on the DIC, we find that the GT-GT SV model provides a better fit than the t-t SV model for the AUD/JPY data while AUD/EUR exchange rate returns favour the t-t SV model. Outlier diagnostic is performed using the mixing parameters and we successfully identify the outlying returns and log-volatilities for the two exchange rate series. Moreover, comparing the estimated volatilities, we found that the volatility of AUD/JPY is generally higher than AUD/EUR pair, which is expected as the AUD/JPY exchange rate is known to be one of the highly volatile pair in the foreign exchange market.

Although the GT distribution provides a flexible alternative distribution to the Student-t distribution in modelling the innovations of the returns and volatilities, the SV model considered falls short of allowing for other dynamics in financial time series and can be enriched by including the leverage and threshold effects in the volatility model. In the second part of our study, we extend the basic SV model with GT distribution to the case of asymmetric SV models, based on the model specifications proposed by Asai and McAleer (2005a) and So *et al.* (2002). We employ the popular S&P 500 return data to estimate different asymmetric SV models with GT, Student-t and EP distributions. The empirical results indicate that under all distributions, the asymmetric effects are found to be statistically significant and the GT-SV model which ignores the asymmetric effects, is the worst model as expected. Among

the asymmetric models, we find that the GT-ASV2 model performs the best under the GT distribution while GT-ASV is the preferred model when we use Student-t and EP distributions. Moreover, the incorporation of volatility asymmetry into the SV model can partially compensate for the difference between estimated volatilities as a result of using different error distributions. For the same model with different distributions, we find the model with the GT distribution produces the best results in terms of the DIC.

In summary, the asymmetric SV model proposed in Asai and McAleer (2005a) allows the effects of both the sign and magnitude of the previous day return to be explicitly included in the volatility equation and it does not have the complication of using a multivariate distribution as in the case of the SVL model. However, it is not entirely impossible to consider the SVL model with a bivariate GT distribution. Arslan (2004) proposed a family of multivariate GT distribution as the scale mixture of multivariate EP distributions and the inverse generalised gamma distribution. Fung and Seneta (2008) further proposed to express the multivariate EP distribution as a scale mixture of multivariate uniform and gamma distributions. Although the multivariate GT density is rather complicated and it is unclear how to implement the multivariate scale mixture form of the GT distribution, it would be interesting to compare the performance of the GT SVL model with the GT asymmetric SV models investigated in this chapter.

CHAPTER 4

SV Model with Asymmetric Laplace Distribution

4.1. Background

The successful application of the GT distribution in SV models has motivated us to consider other types of heavy-tailed distributions. Although the GT distribution is able to account for excess kurtosis by varying the values of the shape parameters p and q, it falls short of allowing for skewness. There is abundant empirical evidence that apart from the leptokurtosis in daily asset returns, there is also mild negative skewness. Several studies, which include Harvey and Siddique (2000), Ait-Sahalia and Brandt (2001) and Chen (2001), highlighted the importance of including skewness in financial applications, especially in financial decision making. Hence we are interested in adopting a distribution that is able to accommodate heavy-tails as well as skewness of financial returns for the SV models. A candidate distribution for this purpose is the asymmetric Laplace (AL) distribution.

The AL distribution was first introduced in Hinkley and Revankar (1977) and it has been applied to model financial data, particularly in the estimation and forecasting of Value-at-Risk (VaR). A common approach to estimate the conditional volatility of short horizon asset returns is based on the JP Morgan's RiskMetrics model, in which the exponentially weighted moving average (EWMA) estimator of the conditional variance defines the next period's variance as a weighted average of this period's variance and squared return. Guermat and Harris (2001) argued that the EWMA estimator is only optimal when the conditional returns are normal, and there is considerable evidence that the conditional distribution of short horizon asset returns is leptokurtic. Thus, they proposed an alternative EWMA estimator based on the Laplace distribution, that is robust to the non-normality of the conditional returns. Through empirical studies, they demonstrated that this robust estimator in terms of the absolute returns, provides an improvement over the original EWMA estimator. Lu *et al.* (2010) extended the work of Guermat and Harris (2001) by proposing a skewed robust EWMA estimator for VaR forecasting model, based on an AL distribution. The use of AL distribution

takes into account both skewness and heavy tails in the distribution of financial asset returns. As an alternative to the EWMA forecasting technique, Chen *et al.* (2011) adopted the well-known Glosten-Jagannathan-Runkle (GJR) GARCH model for the volatility process and assumed an AL distribution as the conditional distribution of the financial return series when estimating and forecasting VaR. Apart from the flexible excess kurtosis and skewness properties of the AL distribution, one other reason for considering this distribution is that it plays an important role in the inference for a quantile regression problem.

Regression models are used to quantify the relationship between a response variable and some explanatory variables and they have been one of the most important statistical methods in applied research. A standard linear regression model estimates the conditional expectation of the response variable using the least-squares minimisation method. However, standard regression sometimes does not offer a complete picture of the response distribution as it only fits the average relationship between the response variable and the regressors. Quantile regression has emerged as a comprehensive approach to the statistical analysis of linear and nonlinear regression models. It estimates the p-th quantile of the response variable given the regressors and thus explores the complete relationship between the response and explanatory variables. Koenker and Bassett (1978) developed the general quantile regression problem and suggested a general technique for estimating a family of conditional quantile functions. The general linear quantile regression problem is defined as follows

$$y_t = \boldsymbol{X}_t^{'}\boldsymbol{\beta} + \epsilon_t,$$

where y_t is the response at time t, $\boldsymbol{X}$ is a vector of regressors, $\boldsymbol{\beta}$ is a vector of unknown coefficients and ϵ_t is an error term with an unspecified distribution. The regression quantile at level $p \in (0, 1)$ is the solution to the minimisation problem

$$\min_{\boldsymbol{\beta}} \sum_t \rho_p(y_t - \boldsymbol{X}_t^{'}\boldsymbol{\beta}). \tag{4.1}$$

The function ρ is a loss or 'check' function which is defined as

$$\rho_p(u) = u(p - I(u < 0)).$$

This special class of loss function has robust properties in contrast to the quadratic loss function used for mean regression estimation. For example, if the observations are highly

skewed or have outliers, then the conditional median regression is more robust than the conditional mean regression. Furthermore, quantile regression admits separate assessment of relationships between the response and explanatory variables, in particular, it provides a method for the examination of the influence of variables on the location, scale and shape of the response distribution at different quantiles. More importantly, quantile plots can be exploited as a useful descriptive tool for the detection of heteroscedasticity, in which case the regression quantiles are not all parallel.

Quantile regression has a wide range of applications to medical reference charts (Cole and Green, 1992), survival analysis (Koenker and Geling, 2001), labour economics (Buchinsky, 1995) and environmental modelling such as pollution concentration (Pandey and Nguygen, 1999). The application closest to our interest is in financial research, particularly in the modelling of VaR as the VaR estimation is closely related to the extreme quantile estimation through estimating the tail of financial returns. Bassett and Chen (2001) provided a general discussion of using quantile regression for return-based analysis.

Yu *et al.* (2003) gave an overview of the different estimation methods and algorithms for quantile regression, including parametric quantile regression estimation through the Box-Cox transformation, nonparametric estimation through kernel-smoothing technique and semi-parametric estimation using the penalised likelihood approach. The inference method we favour and adopt is Bayesian quantile regression for its two advantages over classical inference methods. Firstly, Bayesian approach leads to exact inference of the parameters by providing us with the entire joint posterior distribution of the parameter of interest as opposed to asymptotic inference of the classical approach. Secondly, it allows parameter uncertainty to be taken into account in model estimation and forecasts.

Yu and Moyeed (2001) developed a fully Bayesian approach for quantile regression by forming the likelihood function based on the AL distribution as the minimisation of the loss function (4.1) is equivalent to the maximisation of a likelihood function formed by iid AL densities. This likelihood function allows the construction of the posterior density from which the regression parameter estimates can be obtained. Considering the observations $\boldsymbol{y} = (y_1, \ldots, y_n)$, the posterior distribution of the unknown regression parameters $\boldsymbol{\beta}$, $f(\boldsymbol{\beta}|\boldsymbol{y})$, is given by

$$f(\boldsymbol{\beta}|\boldsymbol{y}) \propto L(\boldsymbol{y}|\boldsymbol{\beta})\pi(\boldsymbol{\beta}),$$

where $\pi(\boldsymbol{\beta})$ is the prior distribution of $\boldsymbol{\beta}$ and $L(\boldsymbol{y}|\boldsymbol{\beta})$ is the likelihood function based on the AL densities defined in (4.2). As pointed out in Yu and Moyeed (2001), because a conjugate prior distribution is not available for the quantile regression, we can use virtually any prior distribution. In practice, we choose mainly diffuse priors in the absence of substantial prior information, such that the data information dominates in the inference of the parameters. As the likelihood is not one of the standard distributions, computational MCMC method is required for the posterior inference of the parameters.

Recently, Chen *et al.* (2009) investigated the testing of Granger non-causality in a quantile regression setting such that quantiles for which causality effect is significant can be identified. They argued that if heteroscedastic effect is not explicitly modelled in quantile regression, then the effects of the explanatory variables may be biased and incorrect conclusion regarding non-causality may be obtained. Hence, their analysis involved robust estimation of the quantile regression effects after allowing for heteroscedasticity. The work of Chen *et al.* (2009) is an extension of Koenker and Zhao (1996), as a GARCH model is employed for the volatility process rather than a simple ARCH representation used in Koenker and Zhao (1996).

Our study is motivated by the fact that the SV model offers a flexible alternative volatility modelling to GARCH type specifications and we investigate the applicability of quantile regression where the dynamic volatility is controlled by a SV model. Firstly, we extend the Bayesian analysis of the SV model by assuming an AL distribution for the asset return distribution, which has not been considered before in the literature. Furthermore, unlike previous studies where the direct AL density is used in model estimation, we express the AL distribution using the scale mixture of uniform (SMU) distributions which allows us to perform Bayesian inference with MCMC and Gibbs sampling methods with ease. Secondly, we investigate the applicability of controlling for heteroscedasticity using a SV model in a quantile regression setting.

4.2. AL distribution as a SMU

Koenker and Machado (1999) and Yu and Moyeed (2001) presented a new skewed distribution in the context of quantile regression. This three parameter distribution, asymmetric

Laplace (AL) distribution, denoted by $X \sim AL(\mu, \sigma, p)$ has the following pdf:

$$f(x|\mu, \sigma, p) = \frac{p(1-p)}{\sigma} \exp\left(-\frac{(x-\mu)}{\sigma}[p - I(x \leq \mu)]\right), \quad x \in \mathbb{R}, \tag{4.2}$$

where $\mu \in \mathbb{R}$ is the location parameter, $\sigma > 0$ is the scale parameter and $0 < p < 1$ is the skewness parameter. Here the shape parameter is defined such that $p = P(X < \mu)$ and μ is the mode.

Figure 4.1 shows the density plot of the AL distribution with different p values. The Laplace distribution, also known as the double exponential distribution, is a special case of the AL distribution when $p = \frac{1}{2}$, which is a symmetric distribution with density

$$f(x|\mu, \sigma, \frac{1}{2}) = \frac{1}{4\sigma} \exp\left(-\frac{|x-\mu|}{2\sigma}\right).$$

The mean and variance of $X \sim AL(\mu, \sigma, p)$ are given by

$$\begin{aligned} E(X) &= \mu + \frac{\sigma(1-2p)}{p(1-p)}, \\ \text{var}(X) &= \frac{\sigma^2(1-2p+2p^2)}{(1-p)^2p^2}. \end{aligned}$$

The skewness and kurtosis are given by

$$\begin{aligned} \text{skewness} &= \frac{2[(1-p)^3 - p^3]}{((1-p)^2 + p^2)^{3/2}}, \\ \text{kurtosis} &= \frac{9p^4 + 6p^2(1-p)^2 + 9(1-p)^4}{(1-2p+2p^2)^2}. \end{aligned}$$

The derivation of these four moments is given in Appendix 4.1. Hence both skewness and kurtosis depend solely on the shape parameter p and are independent of the scale parameter σ. Figure 4.2 shows the behaviour of skewness and kurtosis as p varies. The AL distribution is skewed to the right when $p < \frac{1}{2}$ and skewed to the left when $p > \frac{1}{2}$. The kurtosis has minimum at $p = 1/2$, where the excess kurtosis is 3, indicating that the AL distribution has a relatively peaked density as compared to the normal distribution.

If the location and scale parameters are set to be 0 and 1 respectively, that is, $X \sim AL(0, 1, p)$, then X has density

$$f(x|0, 1, p) = p(1-p)\exp(-x(p - I(x \leq 0)))$$

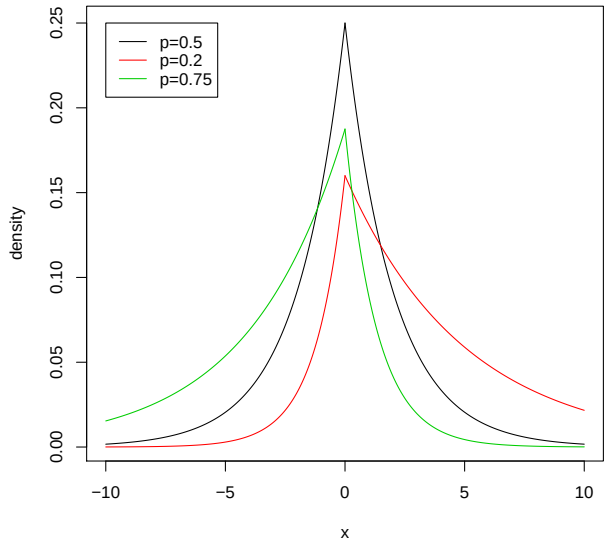


FIGURE 4.1. AL density different p values with $\mu = 0$ and $\sigma = 1$.

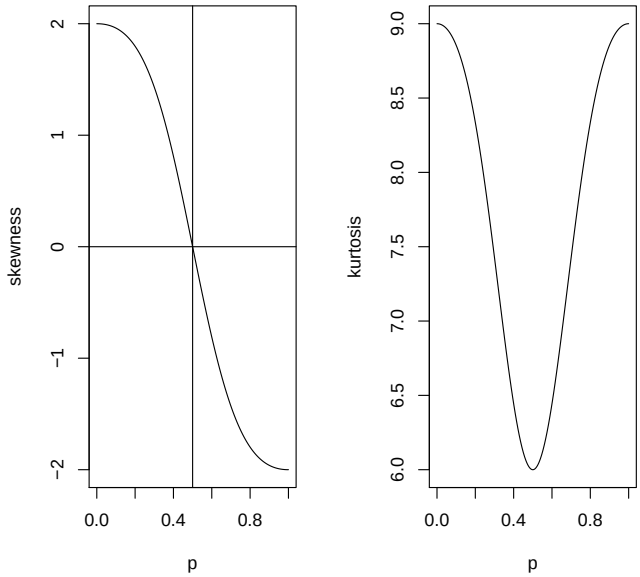


FIGURE 4.2. Skewness and kurtosis of AL distribution.

and

$$\begin{aligned} E(X) &= \frac{1-2p}{p(1-p)}, \\ \text{var}(X) &= \frac{1-2p+2p^2}{(1-p)^2p^2}. \end{aligned}$$

Then $X^* = \frac{X-E(X)}{\sqrt{\text{var}(X)}}$ has a standard AL distribution with mean 0, variance 1 and shape parameter p. The density of X^* using density transformation is thus

$$f(x^*|p) = c_p \exp\left(\frac{c_p x^*}{p - I(x^* \geq 0)}\right), \tag{4.3}$$

where $c_p = \sqrt{1 - 2p + 2p^2}$. For the above standard AL density, the mode is 0 and the shape parameter p is defined such that $p = P(X^* < 0)$. The skewness and kurtosis formulas still hold since they are invariant under linear transformation.

Choy and Chan (2008) provided the SMU representation of the exponential power (EP) distribution, which has the symmetric Laplace distribution as a special case. Wichitaksorn *et al.* (2011) proposed the following SMU form for the standard AL density in (4.3):

$$f(x^*|p) = \int_0^\infty U\left(x^*| -\frac{p}{c_p}u, \frac{(1-p)}{c_p}u\right) Ga(u|2,1)du. \tag{4.4}$$

Proof.

$$\begin{aligned} f(x^*|p) &= \int_0^\infty U\left(x^*| -\frac{p}{c_p}u, \frac{(1-p)}{c_p}u\right) Ga(u|2,1)du \\ &= \int_0^\infty \frac{c_p}{u} I\left(-\frac{p}{c_p}u < x^* < \frac{(1-p)}{c_p}u\right) \frac{1}{\Gamma(2)} u\exp(-u)du \\ &= c_p \int_0^\infty I\left(-\frac{p}{c_p}u < x^* < \frac{(1-p)}{c_p}u\right) \exp(-u)du \end{aligned}$$

Consider two cases: 1) $-\frac{p}{c_p}u < x^*$, or $u > -\frac{c_p x^*}{p}$ and 2) $x^* < \frac{(1-p)}{c_p}u$ or $u > \frac{c_p x^*}{1-p}$. Note that in case 1, $-\frac{c_p x^*}{p} \geq 0$, hence this condition is equivalent to $x^* \leq 0$. Similarly, case 2 corresponds to the condition $x^* > 0$. In follows that in case 1,

$$\int_0^\infty I\left(u > -\frac{c_p x^*}{p}\right) \exp(-u)du = \exp\left(\frac{c_p x^*}{p}\right).$$

For case 2,

$$\int_0^\infty I\left(u > \frac{c_p x^*}{1-p}\right) \exp(-u)du = \exp\left(-\frac{c_p x^*}{1-p}\right) = \exp\left(\frac{c_p x^*}{p-1}\right)$$

Hence by combine case 1 and 2, we obtain

$$\int_0^\infty U\left(x^*| -\frac{p}{c_p}u, \frac{(1-p)}{c_p}u\right) Ga(u|2,1)du = c_p \exp\left(\frac{c_p x^*}{p - I(x^* \geq 0)}\right)$$

□

4.3. SV model with AL distribution and quantile regression SV model

4.3.1. SV model with AL distribution. Many studies have extended the basic SV model in (1.1) with Gaussian error to the case of heavy-tails for the return distribution, motivated by the empirical leptokurtic feature observed in financial return series. The most common heavy-tail distribution employed is the Student-t distribution, found in, for example, Chib *et al.* (2002), Asai (2008) and Wang *et al.* (2011a). The extension to skewed distribution for the returns in SV models is much more limited, although apart from the leptokurtosis, financial returns are also recognised as slightly negatively skewed. Studies which employ skewed distribution for the returns include Steel (1998) which allowed a skewed EP distribution, introduced in Fernández, Osiewalski and Steel (1995), for the logarithm of the square of the return shocks, Cappuccio *et al.* (2004) which assumed an alternative skewed EP distribution with density constructed via Azzalini's method (Azzalini, 1985) for the return shocks and Nakajima and Omori (2011) extended their analysis to the case of generalised hyperbolic skew Student-t distribution. In this section, we extend the normality assumption of the return distribution to the AL distribution, which is able to account for leptokurtosis and skewness in the return distribution simultaneously. Lu *et al.* (2010) and Chen *et al.* (2010) also used the AL distribution in the estimation and forecasting of VaR.

Our SV model for the return series y_t is given by

$$y_t = \psi_0 + \psi_1 y_{t-1} + \psi_2 z_{t-1} + \exp\left(\frac{h_t}{2}\right)\epsilon_t, \tag{4.5}$$

where ϵ_t follows a standard AL distribution (defined in (4.3)) with mean 0, variance 1 and shape parameter p. The conditional volatility is given by $\exp(h_t)$. The log-volatility follows a AR(1) Gaussian process:

$$h_t|h_{t-1} = \mu + \phi(h_{t-1} - \mu) + \tau\eta_t, \tag{4.6}$$

where ϕ is the persistence parameter, τ is the standard deviation of h_t and $\eta_t \sim N(0,1)$. Moreover, y_{t-1} is the lagged return from local market and the regressor z_{t-1} is the lagged return from other market which may influence the local market return at time t. Furthermore, given y_{t-1} and z_{t-1}, the unconditional variance of y_t is given by

$$\text{var}(y_t) = \exp\left(\frac{h_t}{2}\right)\frac{(1-2p+2p^2)}{p^2(1-p)^2}.$$

One common problem when using non-standard distribution for the returns is that the full conditional distributions of the parameters do not attain closed forms. Hence sampling techniques such as the Metropolis-Hastings algorithm is required for the sampling from these full conditional distributions. Other studies overcome this problem by exploiting the SMN or SMU representation of these non-standard distributions. For example, Asai (2008), Nakajima and Omori (2009), Abanto-Valle *et al.* (2010) and Wang *et al.* (2011a) (Chapter 2) expressed the Student-t distribution as a SMN for the return distribution and Wang *et al.* (2011b) expressed the generalised-t distribution as a SMU (Chapter 3) and showed the full conditional distributions attain mostly standard forms. However, the above mentioned studies applied the scale mixture forms only to symmetric distributions. We make the first attempt to estimate the SV model with AL error by employing the SMU representation of the AL distribution. Following (4.4), we can express the model defined in (4.5) and (4.6) hierarchically as

$$\begin{aligned} y_t|\psi_0,\psi_1,\psi_2,y_{t-1},z_{t-1},h_t,p &\sim U\left(\mu_{y_t}-\frac{pe^{h_t/2}u_t}{c_p},\mu_{y_t}+\frac{pe^{h_t/2}u_t}{c_p}\right), \\ h_t|h_{t-1},\mu,\phi,\tau &\sim N(\mu+\phi(h_{t-1}-\mu),\tau^2), \quad t=2,\ldots,n \\ h_1|\mu,\phi,\tau &\sim N\left(\mu,\frac{\tau^2}{1-\phi^2}\right), \\ u_t &\sim Ga(2,1), \end{aligned}$$

where $\mu_{y_t}=\psi_0+\psi_1 y_{t-1}+\psi_2 z_{t-1}$.

The estimation of this SV model with AL error distribution is straightforward using Bayesian method with MCMC and Gibbs sampling algorithms. The details are given in the next section. This model assumes a random shape parameter p of the AL distribution and the estimated p gives an indication of the magnitude and direction of skewness in the return distribution. In the next subsection, we discuss the implication of the model when p is assumed to be fixed at different levels.

4.3.2. Quantile regression with SV model. Quantile regression with SV model allows the study of causal effects of some covariates across quantiles, when the dynamic volatility is controlled for. In particular, Granger causality which was developed by Granger (1969) and has been widely used in economics since the 1960s, gives a statistical concept of causality. It says if a series X 'Granger-causes' another series Y, then past values of X provide

statistically significant information that helps to predict Y above and beyond the information contained in the past values of Y alone. Accordingly, the Granger causality test is a statistical test for determining whether one series is useful in forecasting another series and its mathematical formulation is based on linear regression modelling of stochastic processes. Granger non-causality on the other hand, occurs when the conditional distribution of y_t given the past information of X and Y up to time $t-1$ is the same as the conditional distribution of y_t given the past information of Y alone. Granger non-causality test at different quantiles has been used in financial studies when the causal effects differ across quantiles and the results of the test enable identification of the quantiles where causality is significant. Chuang *et al.* (2007) investigated the causal relations between stock returns and volumes by testing Granger non-causality based on quantile regression. They concluded that causal effects at tail quantiles may be much different from those at middle quantiles and hence the conclusion on non-causality based on a test on the mean relation may be misleading. However, heteroscedasticity is not considered in Chuang *et al.* (2007). Chen *et al.* (2009) tested for Granger non-causality across different quantile levels and the results were used to infer dynamic quantile linkages for various international stock markets. More importantly, heteroscedasticity was taken into consideration through the use of GARCH-type modelling of the volatilities.

We take a similar approach in our study, however with a different formulation of the volatility process by the SV models according to (4.6). The quantile effects are usually estimated according to (4.1). However, in the presence of heteroscedasticity, an extension of the usual criterion function in (4.1) is required as to take into account the estimation of the volatility parameters. This is done in Chen *et al.* (2009) who argued that without including an extra term involving the volatility in the usual criterion function, the minimisation can be achieved by setting the volatility parameters as infinite. In order to cope with this problem in our case, we propose the following quantile criterion function

$$\min_{\boldsymbol{\psi}(p),\boldsymbol{\alpha}(p)} \sum_{t=2}^{n} \frac{(y_t - \psi_0 - \psi_1 y_{t-1} - \psi_2 z_{t-1})[p - I(y_t < \psi_0 + \psi_1 y_{t-1} - \psi_2 z_{t-1})]}{\exp\left(\frac{h_t}{2}\right)} - \frac{h_t}{2}, \quad (4.7)$$

where the vector of parameters in the return equation at the p-th quantile is $\boldsymbol{\psi}(p) = (\psi_0(p), \psi_1(p), \psi_2(p))$ and the vector of parameters in the volatility equation is $\boldsymbol{\alpha}(p) = (\mu(p), \phi(p), \tau(p))$. These

quantile causality parameters and volatility parameters are quantile dependent but we will write $\psi_0(p)$ as ψ_0 and similarly for other parameters for simplicity.

Yu and Moyeed (2001) explored the idea of Bayesian inference for quantile regression and illustrated that the parameters in the quantile regression model can be estimated using the likelihood function of the AL distribution defined in (4.2). More specifically, if we assume that ϵ_t in the return equation (4.5) are iid AL distributed random variables with the density given in (4.3), the likelihood function is given by

$$f(\boldsymbol{Y}|\boldsymbol{h},\boldsymbol{\psi},\boldsymbol{\alpha}) = \sqrt{1-2p+2p^2}\left(\prod_{t=2}^{n}\exp\left(-\frac{h_t}{2}\right)\right) \times \exp\left\{\sum_{t=2}^{n}\frac{\sqrt{1-2p+2p^2}(y_t-\psi_0 y_{t-1}-\psi_2 z_{t-1})}{\exp\left(\frac{h_t}{2}\right)\left(p-I(y_t\geq\psi_0+\psi_1 y_{t-1}+\psi_2 z_{t-1})\right)}\right\}, \quad (4.8)$$

then the maximum likelihood estimates obtained from the above model is equivalent to the quantile estimators from (4.7). Note that in this 'quantile SV' model, we do not assume a specific distribution directly for ϵ_t and the use of the AL distribution is only for the purpose of facilitating the estimation of quantile effects in a quantile regression model. Hence the shape parameter of the AL distribution, p, will be fixed at specific quantile levels and the MCMC method is applied separately and independently for each value of p. Moreover, we assume independence between the parameter values at different quantile levels. There are other work in quantile regression that do not assume independence of parameters over different quantile levels. See for example, Gourieroux and Jasiak (2008).

Based on the likelihood function (4.8), we adopt the Bayesian methodology for the advantages mentioned in the background section of this chapter. By specifying prior distributions on the unknown model parameters, we obtain the full conditional posterior distribution for each parameter which is non-standard distribution. We extend the MCMC methods in Yu and Moyeed (2001) where heteroscedasticity was not accounted for and Chen *et al.* (2009) who assumed GARCH-type volatility, to cover the SV case we consider here.

From the computational perspective, the only difference between the the SV model with AL distribution and the quantile SV model is that whether p is fixed at specific levels or needs to be estimated by sampling from the full conditional distribution. However, these two cases are different in nature and hence they should be treated separately when applied in empirical studies.

4.4. Model estimation

4.4.1. Prior specification. As we adopt the Bayesian estimation approach, specification of prior distributions on the unknown model parameters is required. We assign the following prior distributions to the model parameters, which are mostly non-informative:

$$\begin{aligned}
\psi_0 &\sim N(\mu_{\psi_0}, \sigma^2_{\psi_0}),\\
\psi_1 &\sim U(-1, 1),\\
\psi_2 &\sim N(\mu_{\psi_2}, \sigma^2_{\psi_2}),\\
\mu &\sim N(a_\mu, b_\mu),\\
\phi^* &\sim Be(20, 1.5), \quad \text{where } \phi^* = \frac{\phi+1}{2},\\
\tau^2 &\sim IG(a_\tau, b_\tau).
\end{aligned}$$

Furthermore, to fit the quantile SV model, we fix the shape parameter of the AL distribution at specified quantiles, that is, $p = (0.01, 0.05, 0.10, 0.25, 0.50, 0.75, 0.90, 0.95, 0.99)$. For the SV model with AL error, we assign an $U(0, 1)$ as the prior distribution for p.

4.4.2. Full conditional densities. The system of full conditional densities derived in Wichitaksorn *et al.* (2011) is included in Appendix 4.2. The full conditional densities for the log-volatility and its parameters are mostly standard distributions. More specifically, the full conditional density of the log-volatility h_t is truncated normal, the full conditional distribution of μ is normal and the full conditional distribution of τ^2 is inverse gamma. The full conditional distribution for ϕ in the volatility equation is non-standard. The parameters in the return equation have standard full conditional distributions, in particular, the full conditional distribution for ψ_0 and ψ_2 is truncated normal and that of ψ_1 is uniform distribution. Moreover, the mixing parameter u_t has truncated exponential full conditional distribution.

These full conditional densities have similar forms to those obtained from the GT-GT SV model in Chapter 3, especially for the parameters in the return equation. This is because for both the AL distribution and the GT distribution, we express them using the SMU representation, and as a result, the full conditional densities are truncated distributions. The full conditional distributions for the parameters in the volatility equation are the same as

those from the basic SV model in Chapter 1, as in both cases, log-volatilities follow a AR(1) Gaussian process.

To check for the convergence of model parameters, we use the following convergence diagnostics.

4.4.3. Convergence diagnostics.

4.4.3.1. *Brooks-Gelman-Rubin diagnostic.* Gelman and Rubin (1992) proposed a method for monitoring the convergence of iterative simulations by constructing a statistic based on comparing the between chain variance to within chain variance of multiple chains. To begin with, we run $m \geq 2$ sequences of simulations, each of length n after discarding the first n iterations of the sequence. Consider a single random variable, θ, which has mean μ and variance σ^2 under the target distribution. Let θ_{ij} denote the ith iteration value of θ in chain j. Then we calculate the between-m-sequence variance B/n and the within-sequence variance W, defined as

$$\begin{aligned} B/n &= \frac{1}{m-1}\sum_{j=1}^{m}(\bar{\theta}_{\cdot j} - \bar{\theta}_{\cdot\cdot})^2, \quad \text{where } \bar{\theta}_{\cdot j} = \frac{1}{n}\sum_{i=1}^{n}\theta_{ij}, \quad \bar{\theta}_{\cdot\cdot} = \frac{1}{m}\sum_{j=1}^{m}\bar{\theta}_{\cdot j} \\ W &= \frac{1}{m(n-1)}\sum_{j=1}^{m}\sum_{i=1}^{n}(\theta_{ij} - \bar{\theta}_{\cdot j})^2 \end{aligned}$$

An estimate of the target variance σ^2 is a weighted average of B and W, that is

$$\hat{\sigma}^2 = \left(1 - \frac{1}{n}\right)W + \frac{B}{n}.$$

This is an unbiased estimator for σ^2 if the initial values were drawn from the target distribution or when $n \to \infty$. After accounting for the sampling variability of the estimator $\hat{\mu} = \bar{\theta}_{\cdot\cdot}$ for the target mean μ, the pooled posterior variance estimate is given by

$$\hat{V} = \hat{\sigma}^2 + \frac{B}{mn}.$$

The potential scale reduction factor used as a convergence diagnostic, is constructed by comparing the pooled chain variance to within-chain variance expressed as a variance ratio. It is estimated by

$$\sqrt{\hat{R}} = \sqrt{\frac{\hat{V}}{W}} = \sqrt{\frac{m+1}{m}\left(\frac{n-1}{n} + \frac{B}{nW}\right) - \frac{n-1}{mn}}.$$

If $\hat{R}$ is high, then either $\hat{V}$ can further decreased by running more simulations, since $\hat{V}$ overestimates σ^2 due to over-dispersed starting values before convergence, or W will be increased in further simulations, since it underestimates the variance because it has not fully explored the target distribution. Once convergence is reached, the variation between the chains and variation within the chains should coincide, hence $\hat{R}$ should be approximately equal to 1.

Brooks and Gelman (1998) corrected the computation of Gelman and Rubin (1992) potential scale reduction factor and pointed out that monitoring convergence by examining only the ratio $\hat{R}$ ignores some information in the simulation and would consequently result in misleading conclusion about the convergence. In particular, they pointed out that apart from the condition that $\hat{R}$ should approach 1 at convergence, the variance between the parallel chains as well as the variance within the chains should both stabilise as a function of n. Furthermore, Gelman and Rubin (1992) extended the method to other measures of mixing besides using estimates of variances. More specifically, from the combined m chains, we calculate the $100(1-\alpha)\%$ interval which gives us a total-sequence interval length estimate. On the other hand, we calculate the empirical $100(1-\alpha)\%$ interval length estimates for each of the m individual chain. The diagnostic $\hat{R}$ is then interpreted as the ratio of the length of total-sequence interval to the mean length of the within-sequence intervals. This measure also has the property of approaching 1 as the chain converges.

In `WinBUGS`, **bgr diag** function calculates the Gelman-Rubin statistic, as modified by Brooks and Gelman (1998) and is presented in a graphical form. The width of the 80% credible interval for the parameter is taken as a measure of posterior variability. The width of the 80% credible interval constructed from the pooled runs is plotted in green. The average width of 80% credible intervals constructed from each run is plotted in blue. Finally, when convergence is approached, both pooled and within-interval widths should stabilise and their ratio (plotted in red) should converge to 1.

4.5. A simulation study

In this section, we conduct a simulation study in order to assess the performance of our models estimated using Bayesian MCMC method. The simulation study consists of two parts. In the first part, we focus on the estimation of the quantile SV model. It is noted by

Chen *et al.* (2009) that for any given set of true parameter values that change with quantile level p, it is unclear how to simulate data that would have the true implied quantile function over p. Thus we need to test our quantile SV models under a specified error distribution. Data was simulated from the following model:

$$\begin{aligned} y_t &= \psi_0 + \psi_1 y_{t-1} + \psi_2 z_{t-1} + \exp\left(\frac{h_t}{2}\right)\epsilon_t, \quad \epsilon_t \sim D(0,1) \\ h_t|h_{t-1} &= \mu + \phi(h_t - \mu) + \tau\eta_t, \quad \eta_t \sim N(0,1). \end{aligned}$$

We consider three error distributions: i) a standard Gaussian $\epsilon_t \sim N(0,1)$, ii) a Student-t distribution with 8 degrees of freedom $\epsilon_t \sim t_8(0,1)$, standardised to have zero mean and unit variance and iii) a standard Laplace distribution $\epsilon_t \sim AL(p = 0.5)$. The true parameter values were chosen as $\psi_0 = 0.1$, $\psi_1 = -0.1$, $\psi_2 = 0.4$, $\mu = 0.2$, $\phi = 0.9$ and $\tau = 0.1$. We simulate 100 datasets of size 1,000 from the above model using the true parameter values, repeated over each specified error distribution.

In the second part, we estimate the SV model with AL distribution for the returns and hence the shape parameter p is estimated from our simulated data. We set the true values of p to range from 0.05 to 0.95 as it is highly unlikely to observe financial returns that are extremely skewed, that is, the ones with 0.01 or 0.99 skewness parameters. We set the true values of other model parameters to resemble the ones that would be obtained from the estimation of real data and hence each set is different across the p values. For each quantile level p, we generate random samples of size 1,000 observations with 100 replications. To generate returns from the AL distribution, we make use of the SMU representation of the AL distribution, that is, we simulate mixing parameters from $Ga(2,1)$ and substitute them into the uniform density defined in (4.4) from which the return series are then generated.

For both part of the simulation study, we run the MCMC algorithm for 70,000 iterations and discard the initial 20,000 as burn-in. As an illustration, we present the Brooks, Gelman and Rubin convergence plots (described in Section 4.4) for a random data set simulated from the $AL(p = 0.5)$ distribution in Figures 4.3 and 4.4. From the plots, we conclude that the model parameters have converged for both quantiles ($p = 0.5$ and $p = 0.5$). Results for $p = 0.95$ are omitted as they are similar to those for $p = 0.05$.

Table 4.1 presents the simulation results for the quantile SV model. Summaries including the posterior means, posterior standard deviations and posterior 95% credible intervals are

presented for each of the fixed value of p. Note that since true values can only be calculated for ψ_1, ψ_2 and ϕ, which are constant for each p, only these estimates can be assessed using PE and PC (both defined in Chapter 2). In order to be consistent with the rest of the parameters, we do not report the PE and PC. Overall, the posterior means of ψ_1 and ψ_2 are always close to their true values with small standard deviations for all three error distributions, over quantiles. At $p = 0.5$, parameters are closest to their true values and the PCs are the highest for ψ_1, ψ_2 and ϕ, as expected. The parameters ψ_0, μ and τ vary with p, since the location and variance parameters are quantile-dependent. Comparing between error distributions, we find that the estimates of ϕ and τ are relatively more stable across quantiles for the normal distribution while they vary widely with quantile for data generated from Student-t and $AL(p = 0.5)$ distributions. Furthermore, the estimates of ϕ are lower and the estimates of τ are higher for Student-t and $AL(p = 0.5)$ distributions than normal distribution for extreme quantiles. The reason for this effect is that the $AL(p = 0.5)$ and Student-t distributions have heavier-tails than the normal distribution. Thus it is more likely to generate extreme observations under those distributions which effectively lowers the persistence of the volatility process and raise the volatility variability at extreme quantiles.

Table 4.2 reports the estimation results for SV model with AL error distribution. The posterior means of the parameters are close to their true values and the posterior coverages are reasonably high for all parameters except for τ when $p = 0.05$ where the PC is only 77%. More importantly, the relative small PE for p indicates that the shape parameter of the AL distribution can be accurately estimated using the MCMC method.

In both part of the simulation study, our results show that the estimate of τ, which is the standard deviation of the latent volatility process, has the largest PE among all parameters for most values of p. This indicates the higher level of difficulty in estimating this parameter as compared to other parameters. Furthermore, the PCs of τ are the largest at extreme quantiles, that is, when $p = 0.01, 0.95$. This shows the increased difficulty in estimating this parameter at extreme quantiles and when the AL distribution is extremely skewed. Overall, the simulation study results demonstrate satisfactory performance of the MCMC method in both scenarios.

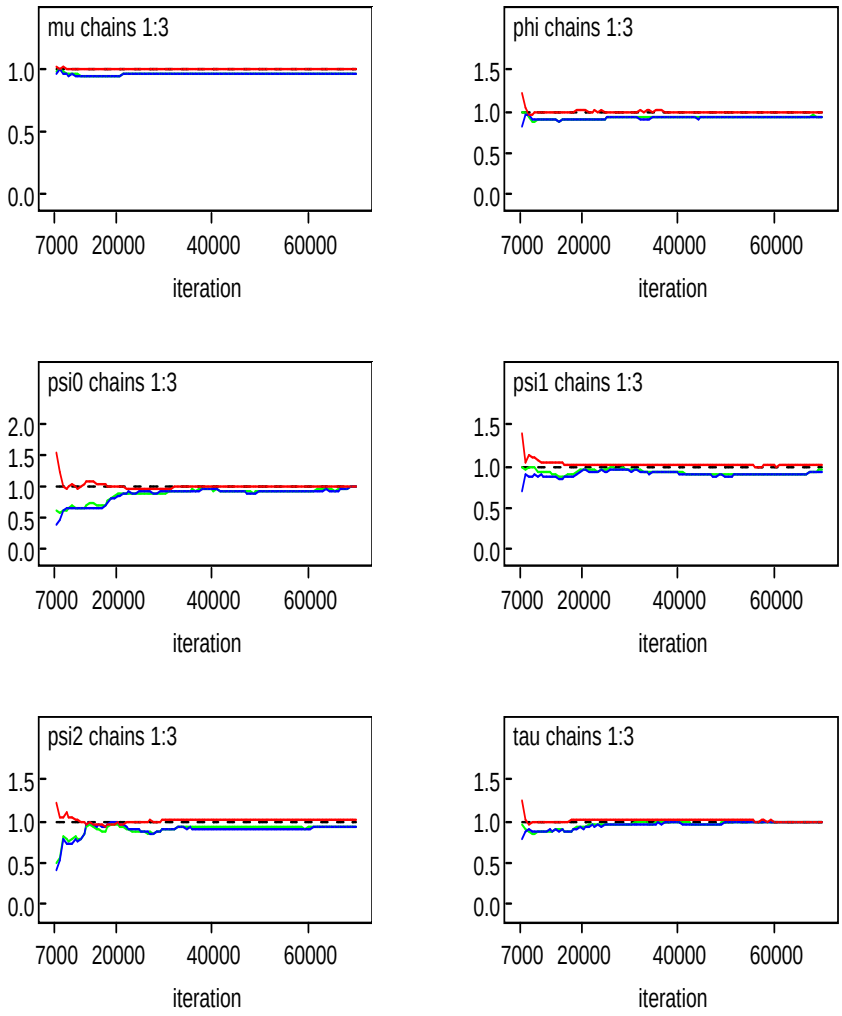


FIGURE 4.3. Brooks, Gelman, Rubin convergence plot for a random data set simulated from $AL(p = 0.5)$ distribution, estimated at $p = 0.05$. The parameters 'mu', 'phi', 'psi0', 'psi1', 'psi2' and 'tau' correspond to μ, ϕ, ψ_0, ψ_1, ψ_2 and τ, respectively.

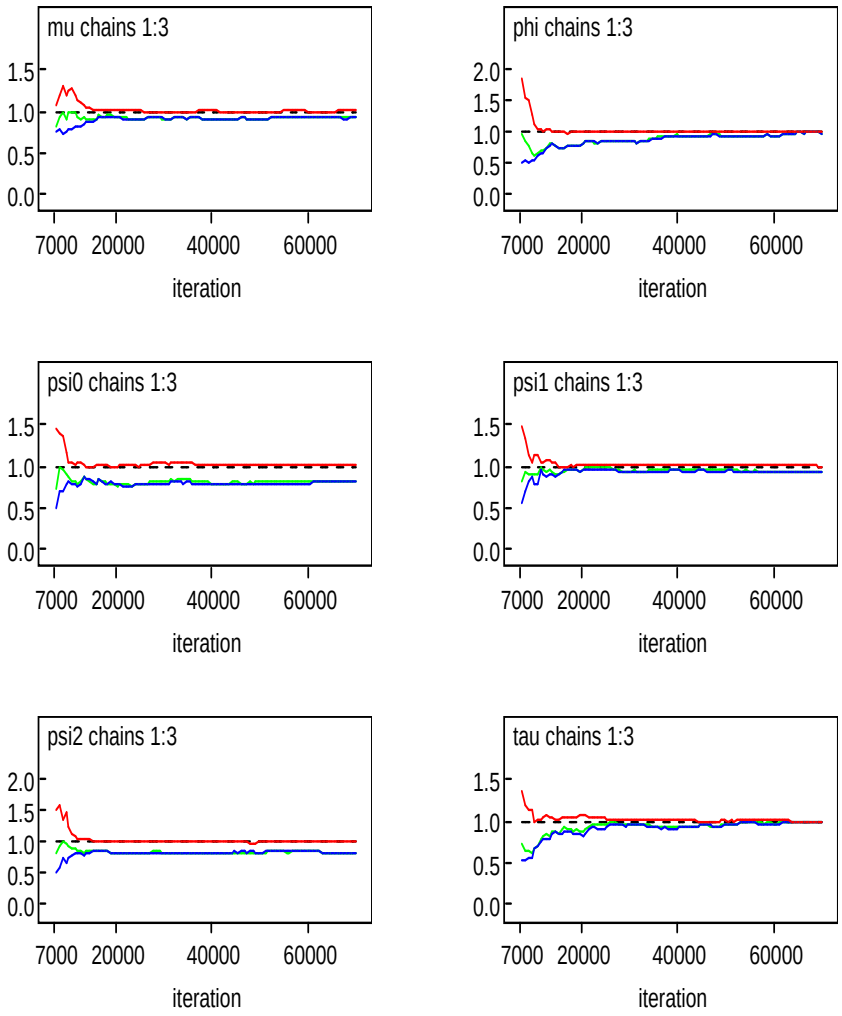


FIGURE 4.4. Brooks, Gelman, Rubin convergence plot for a random data set simulated from $AL(p = 0.5)$ distribution, estimated at $p = 0.5$. The parameters 'mu', 'phi', 'psi0', 'psi1', 'psi2' and 'tau' correspond to μ, ϕ, ψ_0, ψ_1, ψ_2 and τ, respectively.

TABLE 4.1. Simulation results for quantile SV models

Normal						
$p = 0.05$	ψ_0	ψ_1	ψ_2	μ	ϕ	τ
True Value	0.1000	-0.1000	0.4000	0.2000	0.9000	0.1000
Mean	-1.6974	-0.1028	0.4059	1.6552	0.8365	0.1430
S.D.	0.0375	0.0288	0.0350	0.0795	0.0934	0.0497
2.5%	-1.7716	-0.1603	0.3377	1.4995	0.6017	0.0728
97.5%	-1.6250	-0.0472	0.4750	1.8108	0.9557	0.2635
$p = 0.5$	ψ_0	ψ_1	ψ_2	μ	ϕ	τ
True Value	0.1000	-0.1000	0.4000	0.2000	0.9000	0.1000
Mean	0.0979	-0.0985	0.4052	0.4544	0.7846	0.0894
S.D.	0.0341	0.0290	0.0344	0.0672	0.1150	0.0222
2.5%	0.0318	-0.1546	0.3377	0.3241	0.5054	0.0566
97.5%	0.1651	-0.0418	0.4721	0.5866	0.9431	0.1431
$p = 0.95$	ψ_0	ψ_1	ψ_2	μ	ϕ	τ
True Value	0.1000	-0.1000	0.4000	0.2000	0.9000	0.1000
Mean	1.9353	-0.1037	0.4019	1.6855	0.8397	0.1316
S.D.	0.0385	0.0301	0.0361	0.0771	0.0943	0.0466
2.5%	1.8591	-0.1619	0.3313	1.5350	0.6004	0.0682
97.5%	2.0105	-0.0437	0.4731	1.8362	0.9590	0.2486
Student-t with 8 degrees of freedom						
$p = 0.05$	ψ_0	ψ_1	ψ_2	μ	ϕ	τ
True Value	0.1000	-0.1000	0.4000	0.2000	0.9000	0.1000
Mean	-1.5652	-0.1063	0.3911	1.6532	0.6817	0.4598
S.D.	0.0451	0.0311	0.0369	0.0884	0.1171	0.1022
2.5%	-1.6568	-0.1678	0.3178	1.4814	0.4158	0.2727
97.5%	-1.4811	-0.0459	0.4623	1.8283	0.8663	0.8663
$p = 0.5$	ψ_0	ψ_1	ψ_2	μ	ϕ	τ
True Value	0.1000	-0.1000	0.4000	0.2000	0.9000	0.1000
Mean	0.0956	-0.1015	0.3934	0.3725	0.8019	0.0937
S.D.	0.0315	0.0271	0.0323	0.0692	0.1106	0.0243
2.5%	0.0339	-0.1542	0.3303	0.2376	0.5283	0.0581
97.5%	0.1576	-0.0479	0.4556	0.5090	0.9481	0.1522

$p = 0.95$	ψ_0	ψ_1	ψ_2	μ	ϕ	τ
True Value	0.1000	-0.1000	0.4000	0.2000	0.9000	0.1000
Mean	1.7549	-0.0988	0.4040	1.6446	0.6852	0.4548
S.D.	0.0446	0.0316	0.0363	0.0881	0.1167	0.0988
2.50%	1.6710	-0.1603	0.3328	1.4733	0.4170	0.2805
97.50%	1.8437	-0.0373	0.4742	1.8193	0.8651	0.6591
$ALD(p = 0.5)$						
$p = 0.05$	ψ_0	ψ_1	ψ_2	μ	ϕ	τ
True Value	0.1000	-0.1000	0.4000	0.2000	0.9000	0.1000
Mean	-1.3529	-0.1048	0.3962	1.5939	0.4415	0.8797
S.D.	0.0500	0.0340	0.0392	0.0902	0.1276	0.1074
2.5%	-1.4516	-0.1718	0.3216	1.4182	0.1744	0.6683
97.5%	-1.2569	-0.0388	0.4735	1.7723	0.6701	1.0869
$p = 0.5$	ψ_0	ψ_1	ψ_2	μ	ϕ	τ
True Value	0.1000	-0.1000	0.4000	0.2000	0.9000	0.1000
Mean	0.0987	-0.1010	0.4003	0.2010	0.8424	0.1123
S.D.	0.0250	0.0217	0.0255	0.0765	0.0942	0.0364
2.5%	0.0502	-0.1437	0.3508	0.0512	0.6029	0.0632
97.5%	0.1480	-0.0585	0.4507	0.3503	0.9597	0.2042
$p = 0.95$	ψ_0	ψ_1	ψ_2	μ	ϕ	τ
True Value	0.1000	-0.1000	0.4000	0.2000	0.9000	0.1000
Mean	1.5618	-0.1032	0.4053	1.5943	0.4509	0.8491
S.D.	0.0487	0.0317	0.0382	0.0892	0.1267	0.1059
2.5%	1.4710	-0.1662	0.3324	1.4201	0.1848	0.6411
97.5%	1.6585	-0.0409	0.4800	1.7706	0.6757	1.0529

TABLE 4.2. Simulation results for SV model with AL distribution

$p = 0.05$	ψ_0	ψ_1	ψ_2	μ	ϕ	τ	$\hat{p}$
True Value	-1.5000	0.0250	0.4000	-4.0000	0.9000	0.3000	0.0500
Mean	-1.4970	0.0265	0.3999	-3.9800	0.9190	0.2216	0.0553
S.D.	0.0104	0.0057	0.0011	0.1318	0.0377	0.0641	0.0158
2.5%	-1.5103	0.0184	0.3978	-4.2370	0.8270	0.1199	0.0379
97.5%	-1.4686	0.0413	0.4021	-3.7197	0.9720	0.3684	0.1007
PE(%)	0.2000	6.0000	-0.0250	0.5000	2.1111	-26.1333	10.6000
PC(%)	90.0	91.0	92.0	95.0	94.0	77.0	97.0
$p = 0.10$	ψ_0	ψ_1	ψ_2	μ	ϕ	τ	$\hat{p}$
True Value	-1.1000	0.0240	0.4200	-3.3000	0.9400	0.1500	0.1000
Mean	-1.1003	0.0240	0.4200	-3.2879	0.9106	0.1329	0.1005
S.D.	0.0053	0.0045	0.0021	0.1085	0.0570	0.0424	0.0103
2.5%	-1.1104	0.0152	0.4159	-3.5000	0.7616	0.0730	0.0817
97.5%	-1.0894	0.0329	0.4240	-3.0805	0.9765	0.2367	0.1224
PE(%)	-0.0273	0.0000	0.0000	0.3666	-3.1277	-11.4000	0.5000
PC(%)	94.0	94.0	89.0	93.0	98.0	100.0	96.0
$p = 0.25$	ψ_0	ψ_1	ψ_2	μ	ϕ	τ	$\hat{p}$
True Value	-0.5000	0.0260	0.4900	-2.2500	0.9500	0.1250	0.2500
Mean	-0.4992	0.0270	0.4900	-2.2403	0.9061	0.1252	0.2526
S.D.	0.0087	0.0099	0.0057	0.1077	0.0601	0.0387	0.0140
2.5%	-0.5164	0.0076	0.4789	-2.4485	0.7483	0.0705	0.2259
97.5%	-0.4823	0.0466	0.5012	-2.0366	0.9762	0.2200	0.2806
PE(%)	0.1600	3.8462	0.0000	0.4311	-4.6211	0.1600	1.0400
PC(%)	91.0	91.0	94.0	93.0	97.0	100.0	91.0
$p = 0.50$	ψ_0	ψ_1	ψ_2	μ	ϕ	τ	$\hat{p}$
True Value	0.0400	0.0200	0.4600	-1.8000	0.9200	0.1100	0.5000
Mean	0.0386	0.0199	0.4597	-1.7921	0.8562	0.1132	0.4988
S.D.	0.0128	0.0153	0.0093	0.0829	0.0891	0.0356	0.0157
2.5%	0.0134	-0.0104	0.4415	-1.9512	0.6249	0.0639	0.4678
97.5%	0.0637	0.0496	0.4779	-1.6345	0.9642	0.2009	0.5292
PE(%)	-3.500	-0.5000	-0.0652	0.4389	-6.9348	2.9091	-0.2400
PC(%)	96.0	95.0	92.0	91.0	98.0	100.0	98.0

$p = 0.75$	ψ_0	ψ_1	ψ_2	μ	ϕ	τ	$\hat{p}$
True Value	0.7700	0.0140	0.4000	-2.2400	0.8500	0.1070	0.7500
Mean	0.7697	0.0142	0.4002	-2.2410	0.8303	0.1097	0.7494
S.D.	0.0103	0.0113	0.0058	0.0754	0.0985	0.0349	0.0135
2.5%	0.7495	-0.0078	0.3889	-2.3890	0.5801	0.0620	0.7226
97.5%	0.7900	0.0366	0.4114	-2.0942	0.9555	0.1964	0.7757
PE(%)	-0.0390	0.0000	0.0500	-0.0446	-2.3176	2.5234	-0.0800
PC(%)	89.0	93.0	93.0	95.0	100.0	100.0	93.0
$p = 0.90$	ψ_0	ψ_1	ψ_2	μ	ϕ	τ	$\hat{p}$
True Value	1.3000	0.0200	0.3780	-3.3600	0.8800	0.1300	0.9000
Mean	1.3000	0.0201	0.3779	-3.3575	0.8419	0.1118	0.8998
S.D.	0.0066	0.0049	0.0021	0.0774	0.0954	0.0358	0.0099
2.5%	1.2873	0.0101	0.3738	-3.5090	0.5991	0.0634	0.8796
97.5%	1.3135	0.0295	0.3819	-3.2079	0.9606	0.2020	0.9185
PE(%)	0.0000	0.5000	-0.0265	0.0744	-4.3295	-14.0000	-0.0222
PC(%)	94.0	94.0	95.0	95.0	100.0	100.0	93.0
$p = 0.95$	ψ_0	ψ_1	ψ_2	μ	ϕ	τ	$\hat{p}$
True Value	1.7000	0.0120	0.3900	-4.3700	0.8800	0.1700	0.9500
Mean	1.7178	0.0019	0.3899	-4.3113	0.8560	0.1177	0.9439
S.D.	0.0379	0.0213	0.0013	0.1444	0.0903	0.0397	0.0177
2.5%	1.6913	-0.0633	0.3873	-4.5152	0.6204	0.0641	0.8939
97.5%	1.8334	0.0168	0.3925	-3.9318	0.9643	0.2176	0.9639
PE(%)	1.0471	-84.1667	-0.0256	1.3432	-2.7273	-30.7647	-0.6421
PC(%)	94.0	96.0	97.0	98.0	99.0	87.0	97.0

4.6. An empirical study

In this section, we apply the SV model with AL error and the quantile SV model to five stock market indices from different countries. For the quantile SV model, we apply the MCMC algorithm separately for each of $p = (0.01, 0.05, 0.10, 0.25, 0.50, 0.75, 0.90, 0.95)$. We run the MCMC for 80,000 iterations, discarding the first 40,000 iterations as burn-in and retain 40,000 iterations for posterior inference to each dataset. Trace plots of our MCMC iterates are extensively examined to check for convergence.

4.6.1. Data. The data used in our empirical study are daily percentage log returns calculated from different stock markets in the period 2000-2010. These include: the All Ordinaries of Australia (AORD), the Nikkei 225 (N225) of Japan, Hong Kong's Hang Seng Index (HSI), the Stock Exchange of Thailand (SET) and the UK FTSE. The US S&P 500 Index is used as an exogenous variable in the return equation.

A summary of descriptive statistics for all six markets, including the US market, is presented in Table 4.3. Both the means and medians are approximately zero for all markets. All returns series exhibit excess kurtosis and mild negative skewness, which are empirical regularities of financial asset return data. Among these markets, Thailand has the largest negative skewness as well as excess kurtosis. The Jarque-Bera (J-B) statistic is also reported for each market. All return series fail the J-B test at the 1% level, rejecting the null hypothesis of no skewness and zero excess kurtosis. Furthermore, we calculate the proportion of days that have returns which are lower than yesterday's return, referred to as % Negative in Table 4.3. For all markets, the % Negative is just above 50, indicating that the marginal distribution of all return series is only mildly skewed. Time series plot and histogram of each market are shown in Figure 4.5. The large return fluctuations in all markets in 2008-2009 is due to the global financial crisis, which caused global economies to go into recession. Following the bankruptcy of Lehman Brothers in September, 2008, the financial markets began a period of extreme volatility, until June 2009, which was declared to be the official end of the great recession in the U.S. by the National Bureau of Economic Research.

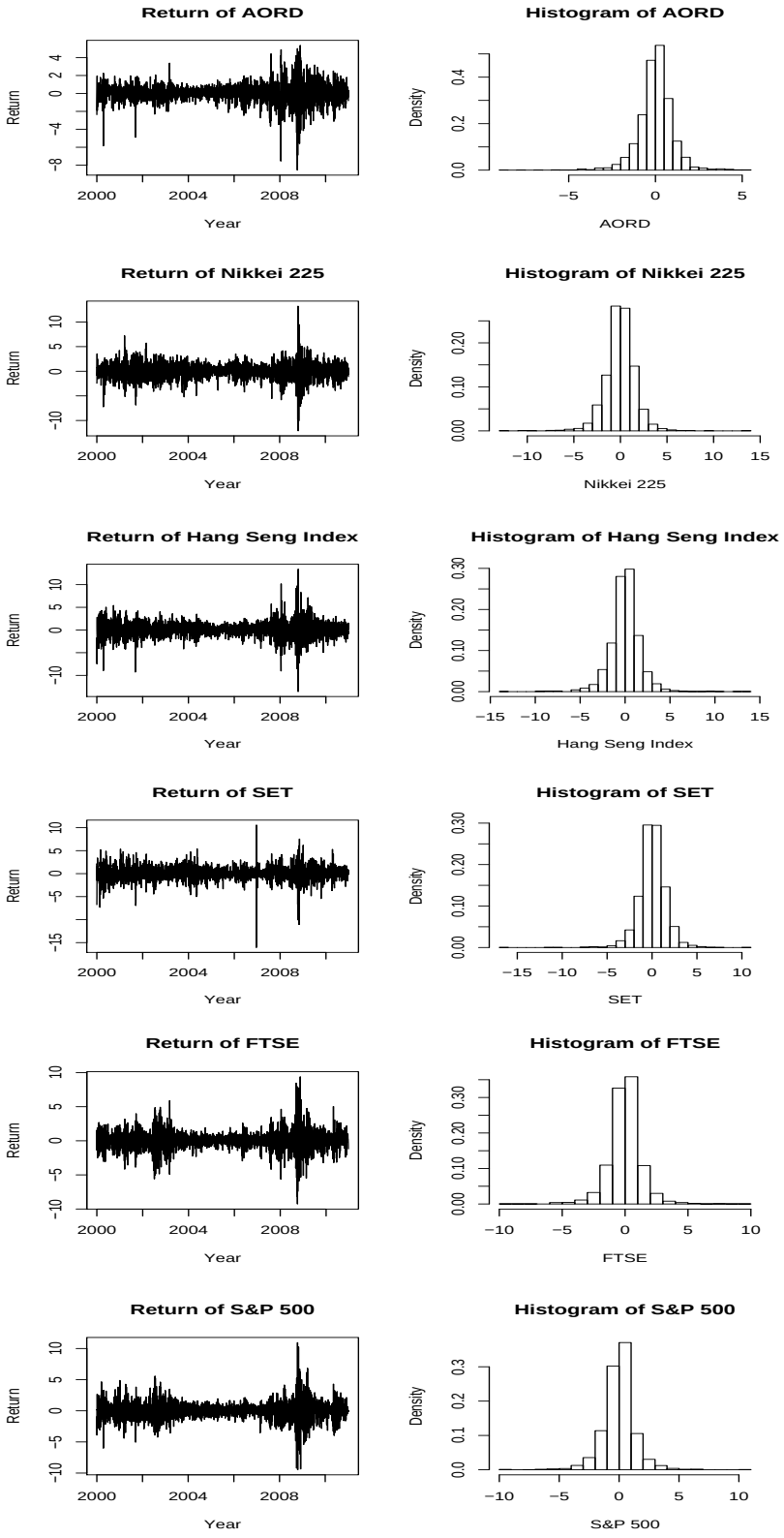


FIGURE 4.5. Time series plots and histograms of log daily average spot price for the six markets.

TABLE 4.3. Summary statistics of different market index returns in 2000-2010.

	U.S.	Australia	Japan	HK	Thailand	U.K.
Mean	-0.0056	0.0158	-0.0230	0.0104	0.0283	-0.0044
Median	0.0523	0.0544	-0.0008	0.0493	0.0388	0.0387
Standard deviation	1.3789	1.0191	1.6131	1.6706	1.5179	1.3202
Skewness	-0.1127	-0.6655	-0.2946	-0.0462	-0.7456	-0.1012
Ex. Kurtosis	4.5253	4.1618	3.0891	4.6800	5.9410	3.0312
Min	-9.4695	-8.5536	-12.1110	-13.5820	-16.0633	-9.2646
Max	10.9572	5.3601	13.2346	13.4068	10.5770	9.3842
1st Quart	-0.6276	-0.4435	-0.8420	-0.7564	-0.7458	-0.6245
3rd Quart	0.6206	0.5436	0.8572	0.8326	0.8608	0.6598
% Negative	52.49	51.15	51.00	51.88	51.51	50.49
Obs.	2766	2772	2698	2714	2691	2777
J-B statistic	2365.93	2205.14	1111.75	2477.75	4206.79	1067.88
J-B p-value	<0.001	<0.001	< 0.001	< 0.001	< 0.001	< 0.001

4.7. Empirical results

4.7.1. SV model with AL error. In this section, we fit a SV model with AL error (defined in Section 4.5) to the five markets. A SV model with Laplace (AL distribution with $p = 0.5$) error is also fitted as a comparison. Figure 4.6 presents the Brooks, Gelman and Rubin convergence plots for the SV model with AL error fitted to the AORD dataset. By examining the plots, we are confident to conclude that all model parameters have converged as the Gelman-Rubin statistic (plotted in red) converged to 1 and both pooled (plotted in green) and within-interval (plotted in blue) widths stablised to the same value after the 40,000 burn-in period.

To classify whether estimates of ψ_0, ψ_1, ψ_2 and μ are significant, we check the 2.5th and 97.5th sample quantiles of the MCMC iterates and an estimate is classified as significant if this 95% credible interval does not include 0. Table 4.4 presents the posterior mean and standard errors of the parameters for the five markets and significant estimates are shown in bold. Firstly, the intercept parameter ψ_0 in the return equation is mostly positive and significant, except for HSI under both AL and Laplace error distributions and Nikkei 225 under Laplace

distribution. The effects of lagged returns in each local market are small and significant, as indicated by the negative and significant AR(1) coefficient ψ_1, except for SET under both error distributions. This effect is slightly stronger under the SV model with Laplace error (except for SET) as the absolute magnitude of ψ_1 is slightly bigger. The positive and significant ψ_2 estimates for all markets under both distributions indicate that the lagged US returns significantly and positively affect the returns in other markets considered and in particular, US returns have the strongest effect on the Japanese and Hong Kong markets and weakest on the Thai market. This result is not surprising as the Thai stock market is heavily influenced by its own political instability. It is analysed in Aggarwal *et al.* (1999) that in emerging markets, there were local events along with the stock market crash that caused increased volatility. In the case of Thailand, the volatility increased substantially during political unrest in 1988 and the military coup in 1991. The estimates of the unconditional mean of the volatilities μ are quite similar for the SV model with AL and Laplace error distributions, with slightly smaller estimates for the Laplace error (except for FTSE). Moreover, the estimate of μ is the smallest for AORD and largest for SET under both error distributions, although it is significantly different from zero only for Nikkei 225 and SET indices.

Figure 4.7a) compares the smoothed estimates of log-volatilities under SV models with AL and Laplace error distributions for the five markets. It seems from the plots, that there is no essential difference between the estimated volatilities using either distribution. However, when we compare the estimated volatilities across markets under the same distribution, there is a large degree of variability between the series which is displayed in Figure 4.7b). It can be seen that the general trend of volatility for all five markets decreases until around 2006 and then gradually increase, reaching a peak near the end of 2008. The only exception is the Australian market in which the increasing trend starts earlier in 2004. It is also observed that the Thai market is one of the markets with higher volatilities and variation in the volatilities and the Australian market has the lowest volatilities and variation in the volatilities. The reason for the high volatilities observed in the Thai market is not hard to understand: high volatility is one of the distinguishing features of emerging markets. Bekaert and Harvey (1997) examined the emerging equity market characteristics in relation to developed markets and they argued that the volatility of emerging markets, especially segmented markets, is strongly influenced by local factors, unlike the fully integrated markets where the volatility is

heavily influenced by world factors. This is the case for the Thai market, with the instability in political situation of the country causing the SET composite index to be volatile. Chen *et al.* (2009) found that the HSI is more volatile than Nikkei 225 and the HSI tail distribution is fatter than that of the Nikkei 225 during January 1996 to January 2007. Our result is in contrast to their finding: the volatility of Nikkei 255 is generally larger than that of HSI before 2007. The discrepancy may due to the different sampling period used. However, the opposite is observed after 2007, particularly during the financial crisis period, the volatility of HSI is substantially higher than that of Nikkei 225 until 2010.

In all of the markets considered, estimates of ϕ are highly positive and close to 1, which suggest a strong AR(1) structure and high persistence of the volatility process. This is expected for stock market daily return data. Among the five markets, persistence is estimated to be the highest for AORD and lowest for SET, which is also the case when Laplace error is used. Comparing across the AL and Laplace errors, we note that the estimates of ϕ are almost the same for the two error distributions, with numbers differ only at the fourth decimal place. The estimated standard deviation of the log-volatility τ ranges from 0.080 (AORD) to 0.136 (SET) under SV model with AL error. The estimates of ϕ and τ suggest that the Thai market not only possesses the highest mean volatility of the returns but the variability of the latent volatility process is also the greatest whereas the Australian market has the lowest variability of the volatilities as shown in Figure 4.7b). It is not surprising to observe that the estimated τ is very similar under the SV model with AL and Laplace errors. As mentioned in Chapter 2, a SV model accounts for high kurtosis of empirical data by using an error distribution that is more heavy-tailed than the normal distribution. A SV model with normally distributed error distribution on the other hand, accommodates the high kurtosis by amplifying the variation of the underlying volatility process, thus a higher estimate of τ is obtained. However, the two distributions we employ here have very similar kurtoses (as the estimated p is very close to 0.5). Therefore, we expect the distributions to account for high kurtosis of the data to a very similar extent, resulting in very similar estimates of τ. Lastly, the estimated skewness parameter p is slightly above 0.5 for AORD, Nikkei 225, SET and FTSE. This indicates that for these markets, the 11-years return data from 2000-2010 is slightly left skewed. More importantly, the estimates of p are significant for AORD, Nikkei 225 and FTSE as the 95% credible intervals do not include 0.5. This demonstrates the importance and necessity for

the explicit modelling of (negative) skewness when analysing financial return data. For HSI and SET, the estimates of p are 0.494 and 0.512 respectively, both of which have 95% credible intervals including 0.5. Hence we may argue that p is essentially 0.5 and the return distributions of HSI and SET over the 11-years are roughly symmetric. Figure 4.8 plots the standardised residuals for the AORD market return under Laplace and AL distributions. Both distributions fit the data reasonably well in the tail areas but the peaks of the distributions are higher than those for the data, particularly under the Laplace distribution.

For model comparison, we report the DIC (described in Chapter 1) for SV model with AL and Laplace error distributions for each market in Table 4.4. For AORD, Nikkei 225 and FTSE, the SV model with AL error distribution attains smaller DIC values, implying that this model provides a better fit to the data than the SV model with Laplace error. The opposite can be said about the HSI and SET markets, with smaller DIC values for the SV model with Laplace distribution. This result is not out of our expectation, as the SV model with AL error distribution offers a more flexible way to account for the empirical skewness observed in the return data, hence the empirical performance of the model is expected to be superior than the SV model with Laplace error which falls short of allowing for skewness. On the other hand, since we do not acquire significant skewness estimates for the HSI and SET indices, the SV model with Laplace distribution should perform equally well as the one with AL distribution. Moreover, the Laplace distribution has one less parameter (p) than the AL distribution, as a result, the DIC chooses the SV model with Laplace distribution as the preferred model for these two markets because the model complexity is taken into account in the computation of DIC.

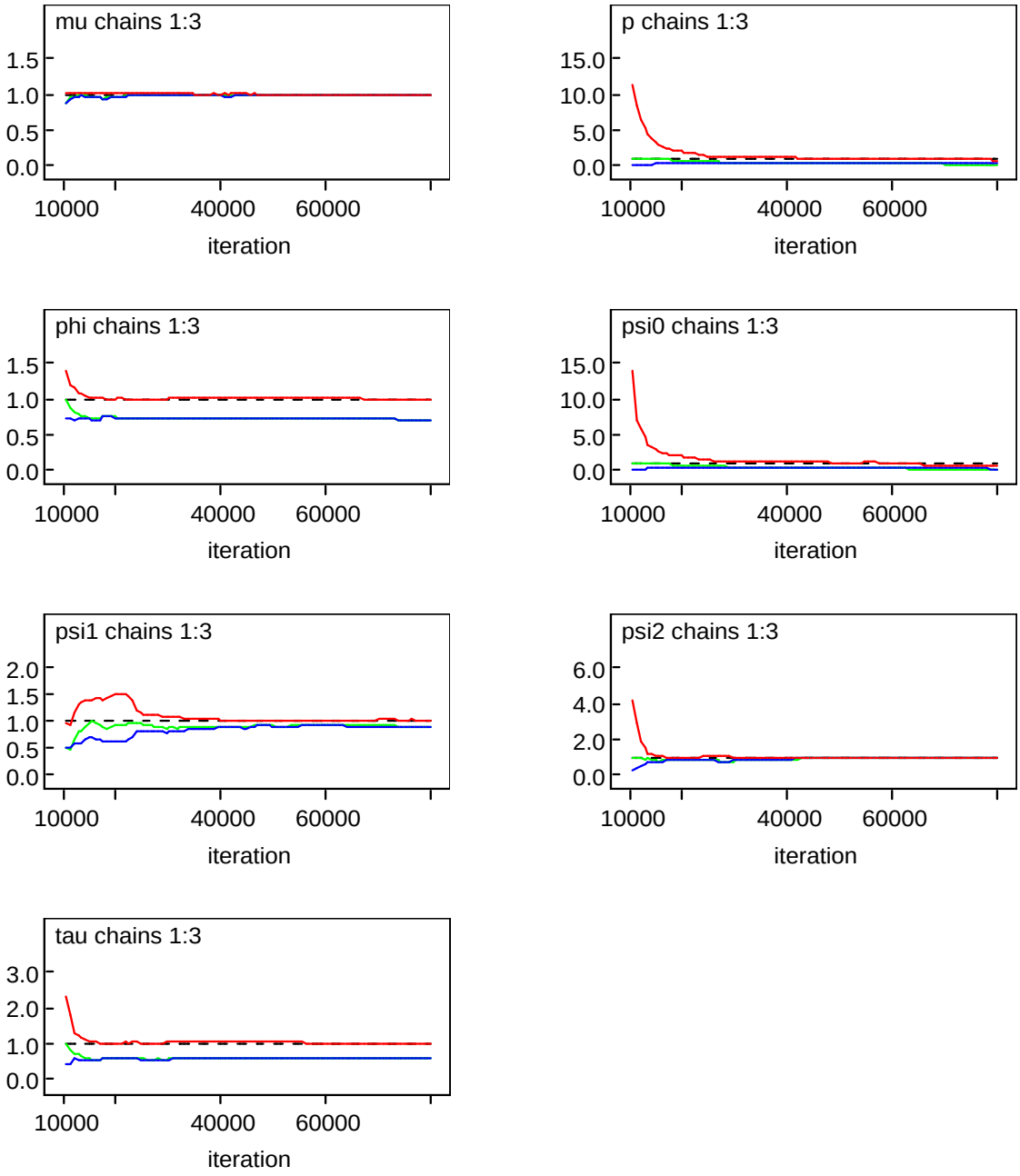


FIGURE 4.6. Brooks, Gelman, Rubin convergence plot for the SV model with AL error for AORD. The parameters 'mu', 'phi', 'psi0', 'psi1', 'psi2' and 'tau' correspond to μ, ϕ, ψ_0, ψ_1, ψ_2 and τ, respectively.

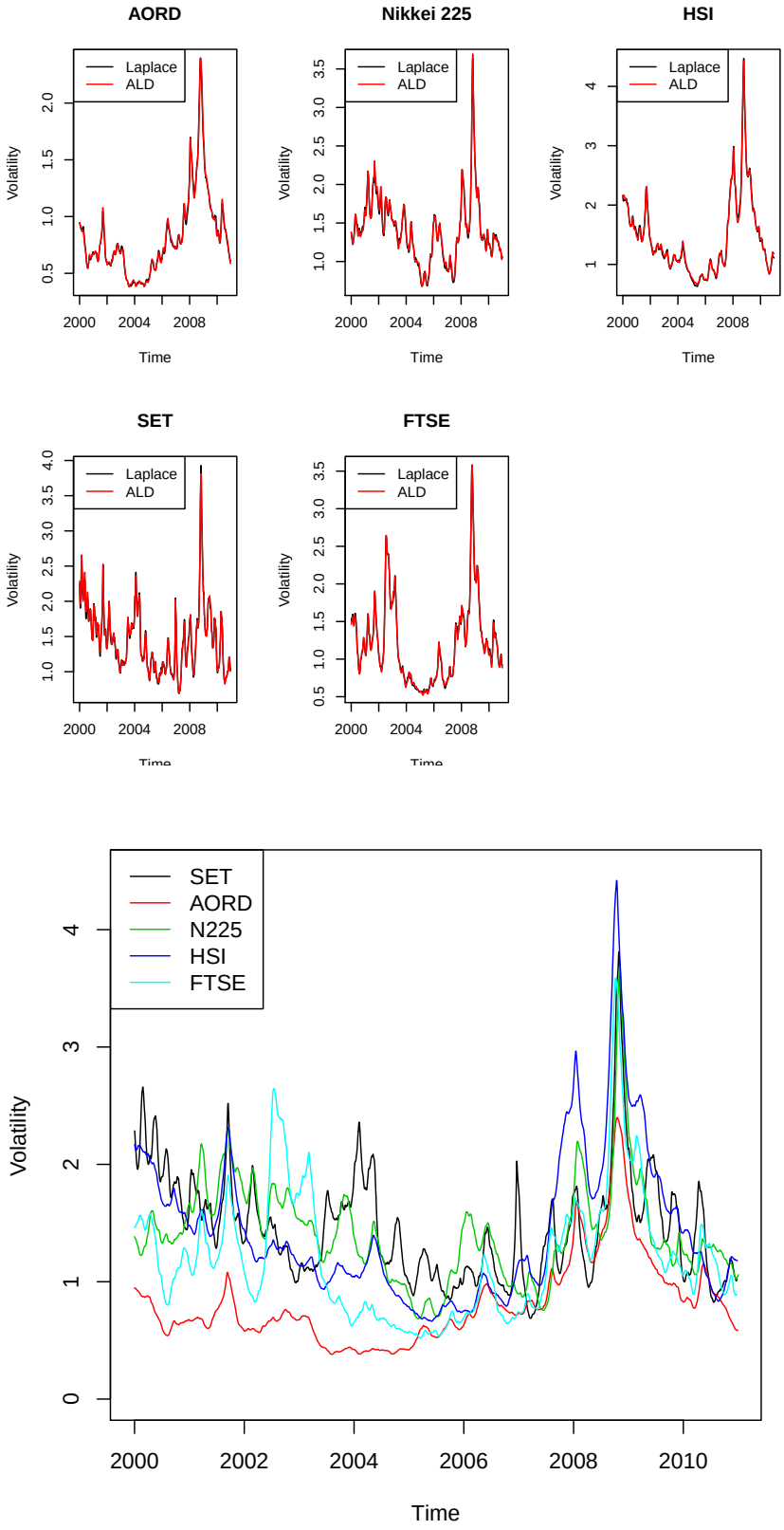


FIGURE 4.7. a) Top: Estimated volatilities under SV model with AL error and Laplace error for each of the five markets; b) Bottom: estimated volatilities across five markets under SV model with AL distribution.

TABLE 4.4. Estimation results of SV Model for the five markets.

Market	ψ_0	ψ_1	ψ_2	μ	ϕ	τ	$\hat{p}$	DIC
AL Error								
AORD	**0.107**	**−0.098**	**0.352**	-0.595	0.996	0.080	**0.537**	2049
	(0.012)	(0.015)	(0.012)	(0.490)	(0.002)	(0.012)	(0.010)	
Nikkei 225	**0.085**	**−0.080**	**0.542**	**0.590**	0.989	0.099	**0.525**	5139
	(0.025)	(0.018)	(0.017)	(0.213)	(0.004)	(0.014)	(0.009)	
HSI	0.014	**−0.045**	**0.507**	0.600	0.995	0.083	0.494	5714
	(0.025)	(0.019)	(0.018)	(0.491)	(0.002)	(0.011)	(0.010)	
SET	**0.095**	-0.005	**0.234**	**0.671**	0.982	0.136	0.512	6258
	(0.042)	(0.025)	(0.026)	(0.161)	(0.006)	(0.017)	(0.013)	
FTSE	**0.095**	**−0.225**	**0.314**	0.175	0.993	0.105	**0.527**	5018
	(0.020)	(0.026)	(0.019)	(0.388)	(0.003)	(0.013)	(0.009)	
Laplace Error								
AORD	**0.055**	**−0.092**	**0.353**	-0.599	0.996	0.080	n.a.	2800
	(0.010)	(0.014)	(0.010)	(0.506)	(0.002)	(0.011)	n.a.	
Nikkei 225	0.026	**−0.081**	**0.558**	**0.577**	0.989	0.097	n.a.	6154
	(0.017)	(0.016)	(0.024)	(0.223)	(0.004)	(0.014)	n.a.	
HSI	**0.044**	**−0.052**	**0.497**	0.576	0.995	0.084	n.a.	5006
	(0.023)	(0.018)	(0.017)	(0.529)	(0.002)	(0.012)	n.a.	
SET	**0.051**	-0.006	**0.256**	**0.669**	0.981	0.139	n.a.	6153
	(0.021)	(0.018)	(0.016)	(0.157)	(0.006)	(0.019)	n.a.	
FTSE	**0.043**	**−0.240**	**0.322**	0.194	0.993	0.104	n.a.	5065
	(0.012)	(0.025)	(0.017)	(0.391)	(0.003)	(0.014)	n.a.	

4.7.2. Quantile SV model. In this section, we present the estimation results for the quantile SV model, where the shape parameter p of the AL distribution is specified at each level. Table 4.5 reports the posterior means and standard errors of the parameters for the five markets and the significant estimates are shown in bold.

To have a better picture of how the parameter estimates vary across quantiles, we plot the estimates and their 95% credible intervals against p, for each market as shown in Figures 4.9-4.13. Firstly, the mean intercept of the return equation ψ_0 increases steadily with increasing quantile p. Clearly for all markets, the estimate of ψ_0 varies from significantly negative for low quantiles to significantly positive for quantiles larger than 0.5. The variation of ψ_0 across quantiles is the largest for HSI and the smallest for AORD. Secondly, the AR(1) coefficient ψ_1 displays a similar pattern for AORD, Nikkei 225 and HSI. For these three

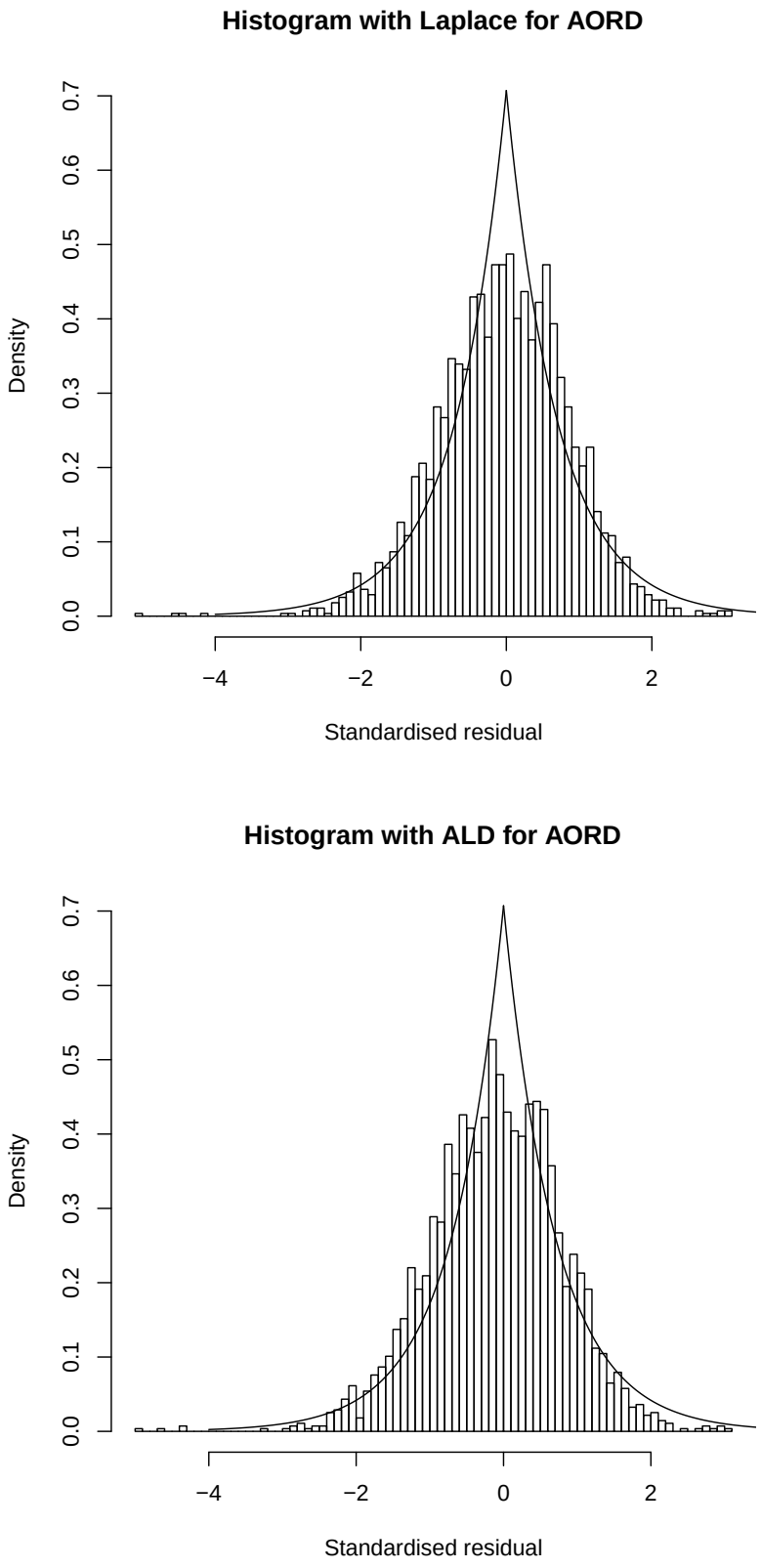


FIGURE 4.8. Comparison of standardised residuals of Laplace and AL distributions for AORD data.

markets, estimated ψ_1 are negative for most of the quantiles, and they are significant for $p \geq 0.5$. This indicates that for these markets, past returns in the local market significantly affect today's market return in a negative relationship and this relationship is only significant at higher quantile levels. For SET, ψ_1 is positive for $p < 0.5$ and significantly negative for $p > 0.5$. FTSE is the only market under investigation which has significantly negative ψ_1 estimates at all quantiles.

Thirdly, the estimates of the effect of lagged S&P 500 returns, ψ_2, are positive and significant across different quantiles for each market. This indicates that the previous day's US return has a significant and positive effect on these markets and in particular, the effects are the strongest in the Japanese and Hong Kong markets and the weakest in the Thai market. Examining the behaviour of ψ_2 over varying quantiles, we observe that there is no large variation in the estimates, hence the effects are quite consistent at different quantiles although a slight curvature in the trend of the estimates can be seen for AORD, suggesting that the US effect is slightly smaller for $p \in (0.25, 0.50, 0.75)$. Furthermore, the effect of lagged US return is stronger than local market lagged return for all markets, which can be inferred by the larger magnitude of ψ_2 estimate as compared with ψ_1.

Next, we observe that the unconditional means of the volatilities are the highest for very low and very high quantiles and they are relatively lower for middle quantiles. This is demonstrated by the obvious U-shape in the trend of the estimates of μ in all markets. In particular, the estimates of μ are significant and positive at all quantiles for SET and Nikkei 225. AORD and FTSE, on the other hand, have mean volatility estimates insignificantly different from zero for $p \in (0.25, 0.50, 0.75)$ and HSI has insignificant estimate of μ only for $p = 0.50$. Generally speaking, the sampling variation will increase as the value of p approaches 0 or 1 as those estimates which are further away from the 50th percentile of the distribution usually cannot be estimated as precisely. Hence we expect to observe narrow credible intervals for the middle quantiles and large credible intervals for extremely large or small quantiles. However, the pattern illustrated in Figures 4.9-4.13 for the estimate of μ is opposite to our expectation. In fact, the specific changes of the sampling variation across quantiles are dependent on the model, data distribution, sample size and number of parameters. The full conditional posterior distribution of μ (given in Appendix 4.2) depends on the estimated values of τ and ϕ, which are also quantile dependent. For extreme quantiles, we find that the

estimate of ϕ is generally lower and the estimate of τ is higher than those for the middle quantiles. As a result, the variance of the full conditional distribution of μ, σ_μ^2, is smaller for extreme quantiles, which may cause the 95% credible interval for μ to be the smallest for the extreme quantiles in all markets.

Looking at the estimates of volatility persistence, ϕ, a n-shape pattern is observed for Nikkei 225 and FTSE. This indicates that for these two markets, the estimated persistence is relatively lower for very low and high quantiles and it is highly positive and close to 1 for the middle quantiles. An inspection of the plots for the estimates of ϕ for AORD, SET and HSI reveals that the persistence drop rapidly from the 0.95 quantile to the 0.99 quantile. This is especially the case for SET, where the persistence at $p = 0.99$ is estimated to be only 0.556. Moreover for AORD, SET and HSI, the standard errors of the estimates of ϕ are much larger at $p = 0.99$, as compared to other quantiles, particularly for HSI. As a result, the 95% credible intervals for ϕ are much larger at $p = 0.99$, which explains the 'long tail' behaviour of the persistence at extremely high quantiles for these three markets. This means that the volatility at high quantiles is less driven by previous volatility but more by other factors.

Lastly, the plots of the estimates of τ, the standard deviation of the volatility, resemble the mirror images of the plots of ϕ estimates about the horizontal axis. In particular, for Nikkei 225 and FTSE, estimates of τ are the largest at very low and high quantiles and smallest at middle quantiles. For AORD, SET and HSI, estimates of τ increase rapidly for extremely high quantile, especially for SET where the estimated τ is as large as 1.094 at $p = 0.99$. Similar to the estimates of ϕ, the standard errors of τ are the greatest at extremely high quantile, resulting in the observed 'long tails' at $p = 0.99$ due to the large credible interval. This result is consistent with what we found in the simulation study: the standard deviation of the volatility is relatively more difficult to estimate at extremely high quantiles.

Furthermore, note that the parameter estimates at $p = 0.50$ are the same as those estimated from SV model with Laplace distribution as the quantile SV model estimated at the 50th quantile is essentially equivalent to fitting a symmetric Laplace error distribution for the returns. This result is discussed in detail with comparison to SV model with AL error distribution in the previous subsection.

TABLE 4.5. Estimation Results of Quantile SV Model for the five markets.

AORD (Australia)						
Quantile (p)	ψ_0	ψ_1	ψ_2	μ	ϕ	τ
0.01	$\mathbf{-1.573}$	-0.008	$\mathbf{0.418}$	$\mathbf{1.597}$	0.945	0.444
	(0.024)	(0.012)	(0.009)	(0.164)	(0.010)	(0.033)
0.05	$\mathbf{-0.937}$	-0.023	$\mathbf{0.433}$	$\mathbf{0.938}$	0.964	0.323
	(0.019)	(0.019)	(0.014)	(0.182)	(0.009)	(0.037)
0.10	$\mathbf{-0.714}$	-0.017	$\mathbf{0.426}$	$\mathbf{0.541}$	0.985	0.175
	(0.014)	(0.018)	(0.010)	(0.255)	(0.005)	(0.027)
0.25	$\mathbf{-0.313}$	$\mathbf{-0.062}$	$\mathbf{0.386}$	-0.138	0.994	0.099
	(0.015)	(0.019)	(0.010)	(0.424)	(0.003)	(0.013)
0.50	$\mathbf{0.055}$	$\mathbf{-0.092}$	$\mathbf{0.353}$	-0.599	0.996	0.080
	(0.010)	(0.014)	(0.010)	(0.506)	(0.002)	(0.011)
0.75	$\mathbf{0.406}$	$\mathbf{-0.134}$	$\mathbf{0.361}$	-0.255	0.995	0.085
	(0.013)	(0.013)	(0.011)	(0.488)	(0.002)	(0.012)
0.90	$\mathbf{0.768}$	$\mathbf{-0.166}$	$\mathbf{0.411}$	0.379	0.989	0.141
	(0.011)	(0.011)	(0.009)	(0.294)	(0.004)	(0.021)
0.95	$\mathbf{0.969}$	$\mathbf{-0.130}$	$\mathbf{0.402}$	$\mathbf{0.759}$	0.972	0.268
	(0.013)	(0.015)	(0.011)	(0.196)	(0.008)	(0.034)
0.99	$\mathbf{1.328}$	$\mathbf{-0.156}$	$\mathbf{0.429}$	$\mathbf{1.297}$	0.863	0.766
	(0.044)	(0.0101)	(0.019)	(0.123)	(0.033)	(0.099)
Nikkei 225 (Japan)						
Quantile (p)	ψ_0	ψ_1	ψ_2	μ	ϕ	τ
0.01	$\mathbf{-2.894}$	-0.012	$\mathbf{0.661}$	$\mathbf{2.593}$	0.874	0.486
	(0.030)	(0.028)	(0.011)	(0.086)	(0.086)	(0.045)
0.05	$\mathbf{-1.985}$	0.010	$\mathbf{0.556}$	$\mathbf{2.070}$	0.958	0.233
	(0.032)	(0.022)	(0.020)	(0.118)	(0.011)	(0.029)
0.10	$\mathbf{-1.454}$	$\mathbf{-0.044}$	$\mathbf{0.593}$	$\mathbf{1.683}$	0.972	0.173
	(0.032)	(0.019)	(0.020)	(0.133)	(0.008)	(0.022)
0.25	$\mathbf{-0.695}$	$\mathbf{-0.037}$	$\mathbf{0.586}$	$\mathbf{1.016}$	0.986	0.111
	(0.020)	(0.016)	(0.015)	(0.189)	(0.005)	(0.015)
0.50	$\mathbf{0.026}$	$\mathbf{-0.081}$	$\mathbf{0.558}$	$\mathbf{0.577}$	0.989	0.097
	(0.017)	(0.016)	(0.024)	(0.223)	(0.004)	(0.014)
0.75	$\mathbf{0.701}$	$\mathbf{-0.125}$	$\mathbf{0.554}$	$\mathbf{0.972}$	0.987	0.107
	(0.015)	(0.018)	(0.019)	(0.199)	(0.005)	(0.016)
0.90	$\mathbf{1.436}$	$\mathbf{-0.130}$	$\mathbf{0.542}$	$\mathbf{1.616}$	0.979	0.152
	(0.027)	(0.015)	(0.022)	(0.157)	(0.007)	(0.023)
0.95	$\mathbf{1.859}$	$\mathbf{-0.129}$	$\mathbf{0.532}$	$\mathbf{1.958}$	0.966	0.225
	(0.023)	(0.016)	(0.020)	(0.141)	(0.009)	(0.027)
0.99	$\mathbf{2.518}$	$\mathbf{-0.162}$	$\mathbf{0.554}$	$\mathbf{2.480}$	0.918	0.495
	(0.083)	(0.026)	(0.024)	(0.129)	(0.027)	(0.088)

HSI (Hong Kong)						
Quantile (p)	ψ_0	ψ_1	ψ_2	μ	ϕ	τ
0.01	$\mathbf{-2.956}$	0.030	$\mathbf{0.410}$	$\mathbf{2.819}$	0.947	0.408
	(0.072)	(0.025)	(0.031)	(0.159)	(0.011)	(0.036)
0.05	$\mathbf{-1.807}$	-0.034	$\mathbf{0.518}$	$\mathbf{2.171}$	0.980	0.228
	(0.021)	(0.021)	(0.018)	(0.240)	(0.006)	(0.027)
0.10	$\mathbf{-1.307}$	0.015	$\mathbf{0.533}$	$\mathbf{1.762}$	0.989	0.153
	(0.031)	(0.017)	(0.017)	(0.325)	(0.004)	(0.021)
0.25	$\mathbf{-0.555}$	-0.009	$\mathbf{0.510}$	$\mathbf{1.030}$	0.994	0.102
	(0.028)	(0.017)	(0.020)	(0.485)	(0.003)	(0.012)
0.50	0.044	$\mathbf{-0.052}$	$\mathbf{0.497}$	0.576	0.995	0.084
	(0.023)	(0.018)	(0.017)	(0.529)	(0.002)	(0.012)
0.75	$\mathbf{0.690}$	$\mathbf{-0.097}$	$\mathbf{0.514}$	$\mathbf{0.988}$	0.995	0.091
	(0.023)	(0.016)	(0.021)	(0.488)	(0.002)	(0.013)
0.90	$\mathbf{1.356}$	$\mathbf{-0.087}$	$\mathbf{0.576}$	$\mathbf{1.622}$	0.990	0.138
	(0.026)	(0.018)	(0.018)	(0.334)	(0.004)	(0.017)
0.95	$\mathbf{1.748}$	$\mathbf{-0.105}$	$\mathbf{0.588}$	$\mathbf{2.003}$	0.982	0.216
	(0.021)	(0.017)	(0.023)	(0.262)	(0.005)	(0.025)
0.99	$\mathbf{2.291}$	$\mathbf{-0.155}$	$\mathbf{0.588}$	$\mathbf{2.522}$	0.838	0.866
	(0.173)	(0.027)	(0.037)	(0.133)	(0.121)	(0.309)
SET (Thailand)						
Quantile (p)	ψ_0	ψ_1	ψ_2	μ	ϕ	τ
0.01	$\mathbf{-2.939}$	$\mathbf{0.079}$	$\mathbf{0.260}$	$\mathbf{2.718}$	0.863	0.580
	(0.015)	(0.010)	(0.011)	(0.092)	(0.025)	(0.043)
0.05	$\mathbf{-1.711}$	0.039	$\mathbf{0.231}$	$\mathbf{2.066}$	0.883	0.534
	(0.028)	(0.024)	(0.016)	(0.100)	(0.024)	(0.052)
0.10	$\mathbf{-1.327}$	$\mathbf{0.051}$	$\mathbf{0.236}$	$\mathbf{1.730}$	0.943	0.317
	(0.026)	(0.022)	(0.020)	(0.119)	(0.014)	(0.038)
0.25	$\mathbf{-0.628}$	0.035	$\mathbf{0.263}$	$\mathbf{1.083}$	0.976	0.172
	(0.027)	(0.019)	(0.016)	(0.155)	(0.007)	(0.023)
0.50	$\mathbf{0.050}$	-0.006	$\mathbf{0.256}$	$\mathbf{0.669}$	0.981	0.139
	(0.021)	(0.018)	(0.016)	(0.157)	(0.006)	(0.019)
0.75	$\mathbf{0.744}$	$\mathbf{-0.047}$	$\mathbf{0.239}$	$\mathbf{1.039}$	0.976	0.157
	(0.025)	(0.019)	(0.019)	(0.143)	(0.007)	(0.020)
0.90	$\mathbf{1.495}$	$\mathbf{-0.044}$	$\mathbf{0.255}$	$\mathbf{1.682}$	0.952	0.238
	(0.021)	(0.021)	(0.020)	(0.109)	(0.014)	(0.035)
0.95	$\mathbf{1.980}$	$\mathbf{-0.054}$	$\mathbf{0.310}$	$\mathbf{2.018}$	0.893	0.388
	(0.023)	(0.017)	(0.014)	(0.084)	(0.025)	(0.047)
0.99	$\mathbf{2.456}$	$\mathbf{-0.068}$	$\mathbf{0.264}$	$\mathbf{2.449}$	0.556	1.094
	(0.100)	(0.019)	(0.017)	(0.066)	(0.064)	(0.100)

FTSE (U.K.)						
Quantile (p)	ψ_0	ψ_1	ψ_2	μ	ϕ	τ
0.01	$\mathbf{-2.393}$	$\mathbf{-0.268}$	$\mathbf{0.393}$	$\mathbf{2.451}$	0.942	0.453
	(0.024)	(0.024)	(0.022)	(0.159)	(0.012)	(0.039)
0.05	$\mathbf{-1.463}$	$\mathbf{-0.180}$	$\mathbf{0.356}$	$\mathbf{1.802}$	0.972	0.274
	(0.025)	(0.016)	(0.020)	(0.202)	(0.007)	(0.029)
0.10	$\mathbf{-1.000}$	$\mathbf{-0.166}$	$\mathbf{0.333}$	$\mathbf{1.350}$	0.983	0.207
	(0.022)	(0.020)	(0.018)	(0.257)	(0.005)	(0.023)
0.25	$\mathbf{-0.477}$	$\mathbf{-0.217}$	$\mathbf{0.336}$	0.629	0.991	0.127
	(0.017)	(0.022)	(0.019)	(0.354)	(0.003)	(0.014)
0.50	$\mathbf{0.043}$	$\mathbf{-0.240}$	$\mathbf{0.322}$	0.194	0.993	0.104
	(0.012)	(0.025)	(0.017)	(0.391)	(0.003)	(0.013)
0.75	$\mathbf{0.553}$	$\mathbf{-0.233}$	$\mathbf{0.286}$	0.548	0.993	0.114
	(0.015)	(0.019)	(0.015)	(0.400)	(0.003)	(0.013)
0.90	$\mathbf{1.032}$	$\mathbf{-0.298}$	$\mathbf{0.319}$	$\mathbf{1.230}$	0.989	0.163
	(0.022)	(0.017)	(0.019)	(0.334)	(0.004)	(0.018)
0.95	$\mathbf{1.472}$	$\mathbf{-0.314}$	$\mathbf{0.363}$	$\mathbf{1.659}$	0.983	0.206
	(0.024)	(0.016)	(0.016)	(0.260)	(0.005)	(0.023)
0.99	$\mathbf{2.135}$	$\mathbf{-0.296}$	$\mathbf{0.357}$	$\mathbf{2.257}$	0.943	0.485
	(0.015)	(0.012)	(0.021)	(0.175)	(0.013)	(0.044)

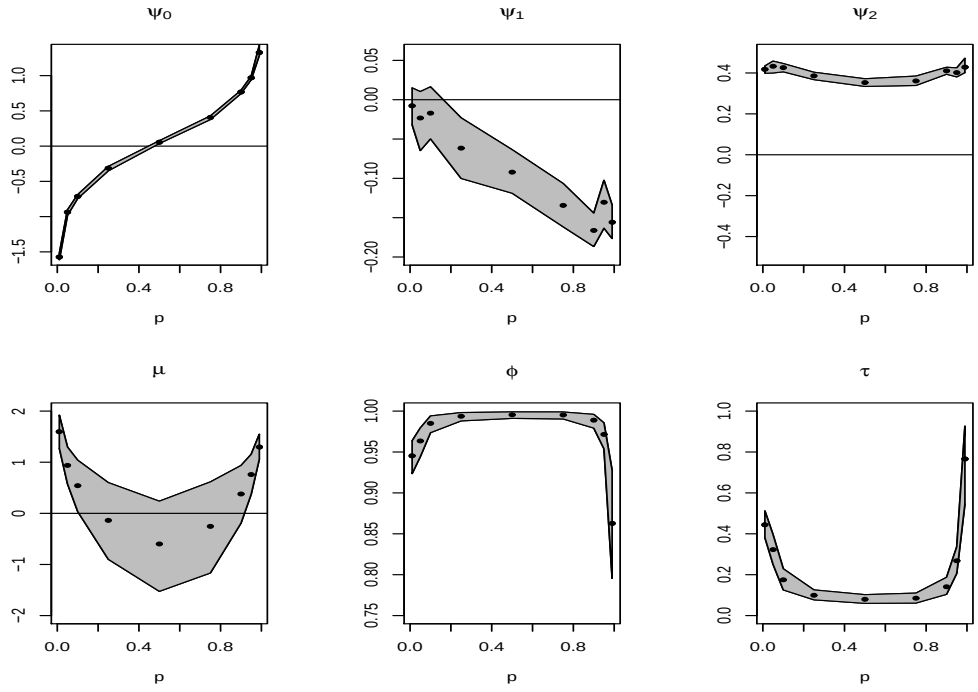


FIGURE 4.9. Parameter estimates plotted against quantile levels for AORD. The shadows represent the 95% credible interval for each parameter.

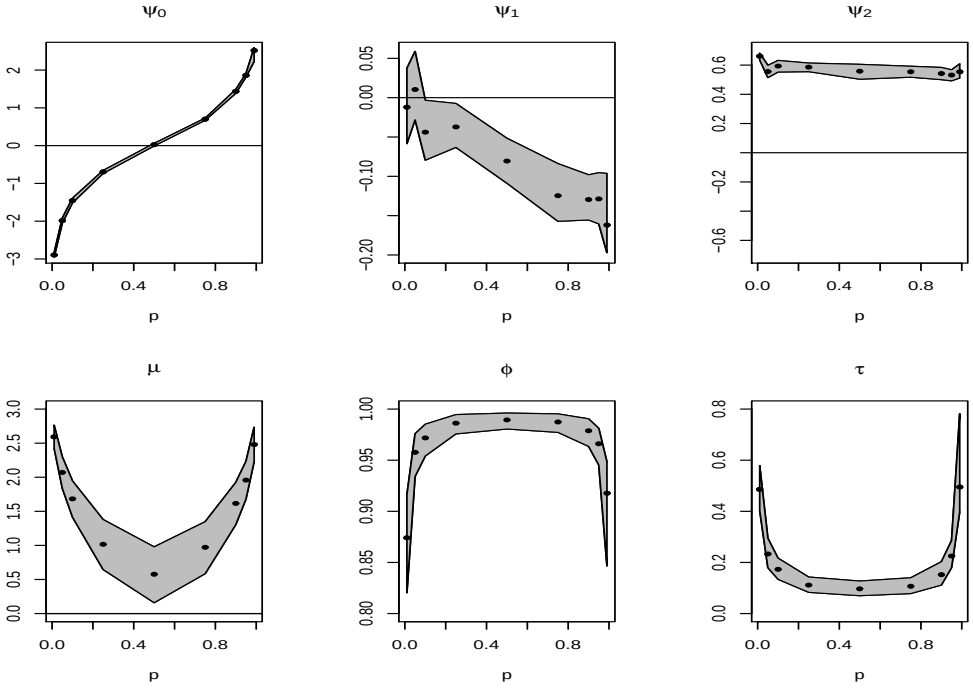


FIGURE 4.10. Parameter estimates plotted against quantile levels for Nikkei 225. The shadows represent the 95% credible interval for each parameter.

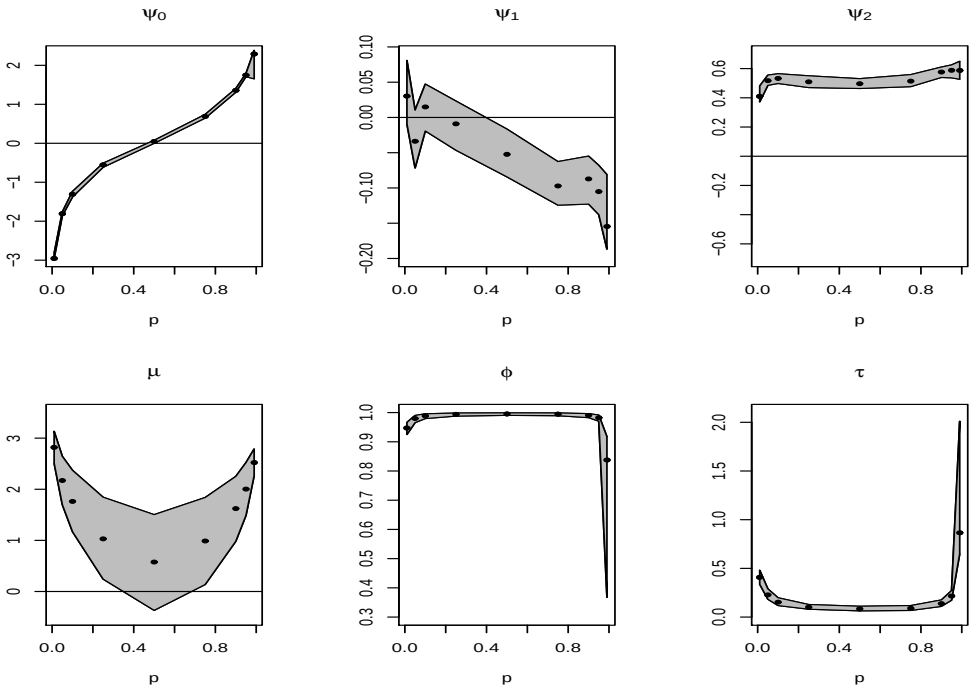


FIGURE 4.11. Parameter estimates plotted against quantile levels for HSI. The shadows represent the 95% credible interval for each parameter.

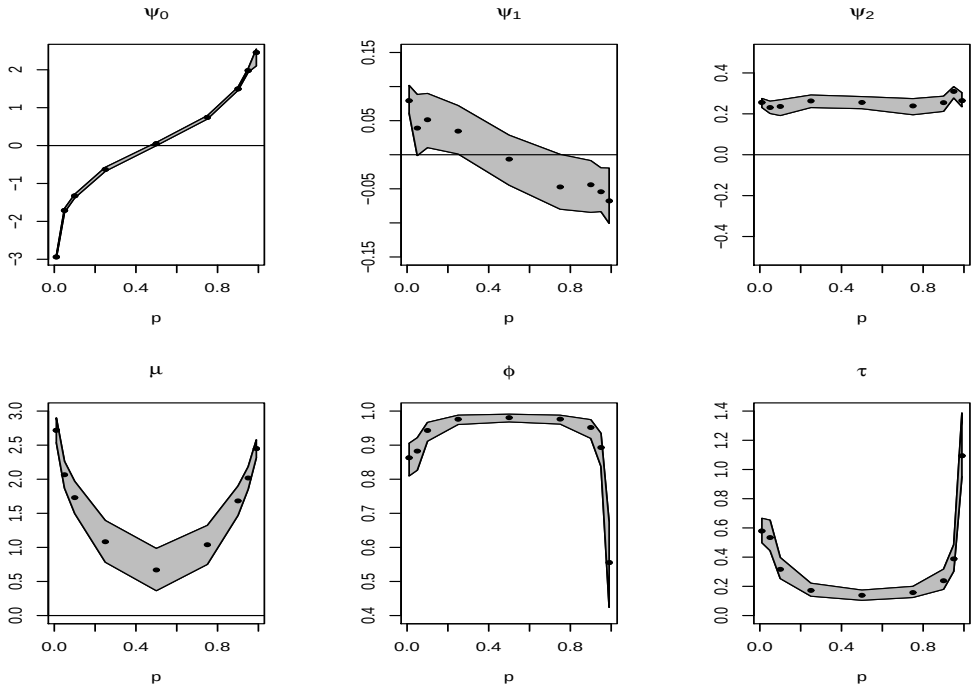


FIGURE 4.12. Parameter estimates plotted against quantile levels for SET. The shadows represent the 95% credible interval for each parameter.

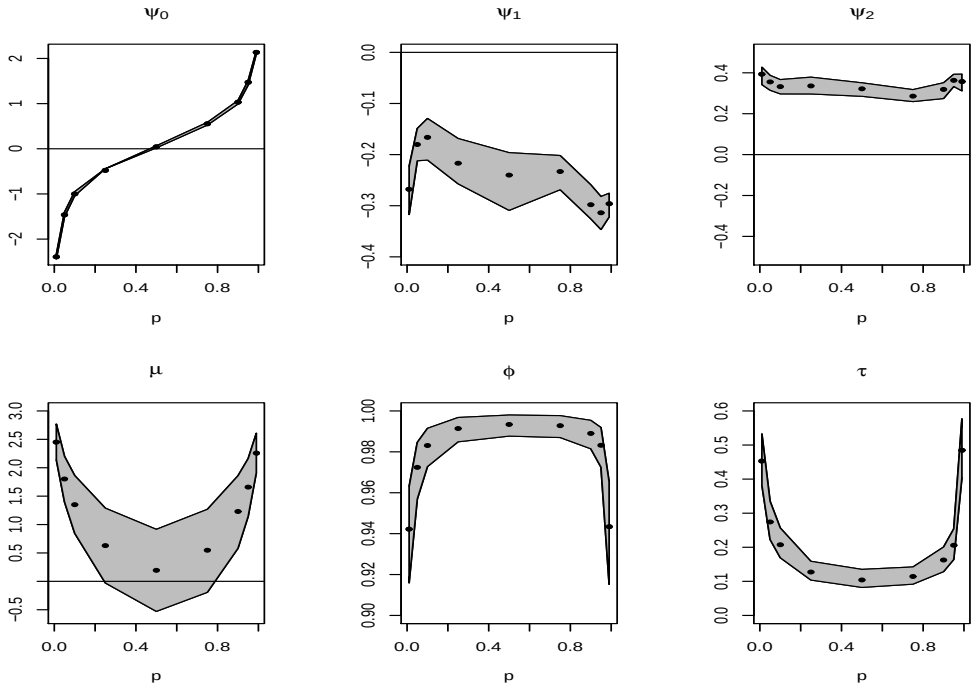


FIGURE 4.13. Parameter estimates plotted against quantile levels for FTSE. The shadows represent the 95% credible interval for each parameter.

4.8. Conclusion

In this chapter, we investigate the SV model with AL distribution for the return innovations and we show that this error assumption allows the model to have two separate implications. Firstly, the use of this asymmetric distribution is motivated by our findings in Chapter 3, where the standardised returns after modelled by a SV model with GT distribution, exhibit negative skewness. The estimate of the shape parameter p of the AL distribution gives indication of the direction and the magnitude of skewness. Secondly, if we fix the shape parameter at specific levels, then our model becomes a 'quantile SV' in the sense that we account for heteroscedasticity by assuming SV in a quantile regression setting. Our work is an extension of Chen *et al.* (2009) who modelled heteroscedasticity by a GARCH model. A quantile regression model allows us to investigate the varying effects of the covariate on the response variable across quantiles as well as changes in the underlying volatility process which are also quantile dependent.

As with studies in previous chapters, we estimate the SV model with AL distribution using Bayesian MCMC method via `WinBUGS`. The AL distribution is expressed as a SMU since the full conditional posterior distributions for the model parameters are shown to have mostly standard forms. A simulation experiment with 100 replications of a sample size 1000 is conducted separately for the case when p is random and when p is fixed at different levels. The sample means of the Bayesian estimators are shown to be very close to the true values, hence indicating small finite sample bias in both cases.

In the empirical study, five market indices are analysed using both the SV model with AL error distribution and the quantile SV model. We find significant negative skewness for the log-returns of AORD, Nikkei 225 and FTSE. For these markets, SV model with AL distribution that accommodates skewness is superior to the SV model with symmetric Laplace distribution with respect to DIC. On the other hand, for HSI and SET indices where the estimated skewness is not significant, the model with Laplace distribution is preferred. Among the five markets, highest volatility is estimated for SET and lowest for AORD, which is consistent with previous findings of high volatility in emerging markets. Concerning the results for the quantile SV model, we find all parameters except for ψ_2 are sensitive to the quantile level. The effect of S&P 500 returns is significant and positive for all markets, with strongest impact on the Hong Kong and Japanese markets and weakest on the Thai market.

Other findings regarding the systematic changes of the volatility parameters are generally in agreement with those in Chen *et al.* (2009). Overall, we demonstrate the practicability of controlling for heteroscedasticity in a quantile regression setting by using a SV model for the volatility and at the same time we extend the basic SV model to adopt the AL error distribution that jointly captures skewness and leptokurtosis of financial return data. One important thing to note is that the only outcome from the analysis on quantile SV model is to assess Granger-causality and dynamics in the return and volatility equations at different quantiles. Since neither the return nor the volatility intercept parameters can be estimated directly from a heteroskedastic quantile model, quantiles themselves cannot be calculated directly in this case.

Since the use of AL distribution is an extension of the symmetric heavy-tailed distribution, we can certainly consider other types of skewed distributions that are more general than the AL distribution, for example, the skewed generalised Student-t distribution (Theodossiou, 1998). However, we consider the AL in particular for two reasons. Firstly, the AL density has a SMU representation which enables easier implementation in `WinBUGS`. At the same time, because the SMU representation is used, the full conditional posterior distributions of the model parameters can be reduced to standard forms, allowing the sampling from these distributions required in the Gibbs sampling algorithm more efficient. This may not be achieved by using other skewed distributions as they may not have convenient scale mixture forms. Secondly, the AL distribution was chosen because it allows the model to have a different interpretation: a quantile regression model with SV type variance. The use of AL distribution allows Bayesian estimation of this model because the minimisation of the loss function in a quantile regression problem is equivalent to the maximisation of the likelihood formed by i.i.d. AL distributed random variables. Other types of asymmetric distributions may not allow such linkage.

Appendix 4.1

According to Yu and Zhang (2005), the kth moment of $X \sim AL(\mu, \sigma, p)$ is given by

$$E(x-\mu)^k = k!\sigma^k p(1-p)\left(\frac{1}{p^{k+1}} + \frac{(-1)^k}{(1-p)^{k+1}}\right).$$

Note that μ is the location parameter, not the expected value of X. Putting $k = 1$, we easily get the expected value of X as

$$E(X) = \mu + \frac{\sigma(1-2p)}{p(1-p)}.$$

The variance of X is

$$\begin{aligned}
\text{var}(X) &= E(X^2) - (E(X))^2 \\
&= E(X-\mu)^2 + 2\mu E(X) - \mu^2 - (E(X))^2 \\
&= \frac{2\sigma^2(1-3p+3p^2)}{p^2(1-p)^2} + 2\mu\left(\mu + \frac{\sigma(1-2p)}{p(1-p)}\right) - \mu^2 - \left(\mu + \frac{\sigma(1-2p)}{p(1-p)}\right)^2 \\
&= \frac{\sigma^2(1-2p+2p^2)}{p^2(1-p)^2}
\end{aligned}$$

The skewness is given by

$$\begin{aligned}
\text{skewness} &= \frac{E(X-E(X))^3}{(\text{var}(X))^{3/2}} \\
&= \frac{E\left(X-\mu-\frac{\sigma(1-2p)}{p(1-p)}\right)^3}{\left(\frac{\sigma^2(1-2p+2p^2)}{p^2(1-p)^2}\right)^{3/2}} \\
&= \frac{E\left((X-\mu)^3 - 3(X-\mu)^2\frac{\sigma(1-2p)}{p(1-p)} + 3(X-\mu)\frac{\sigma^2(1-2p)^2}{p^2(1-p)^2} - \frac{\sigma^3(1-2p)^3}{p^3(1-p)^3}\right)}{\left(\frac{\sigma^2(1-2p+2p^2)}{p^2(1-p)^2}\right)^{3/2}} \\
&= \frac{6(1-4p+6p^2-4p^3) - 6(1-2p)(1-3p+3p^2) + 3(1-2p)^2(1-2p) - (1-2p)^3}{(1-2p+2p^2)^{3/2}} \\
&= \frac{2-6p+6p^2-4p^3}{(1-2p+2p^2)^{3/2}} \\
&= \frac{2((1-p)^3-p^3)}{(1-2p+2p^2)^{3/2}}
\end{aligned}$$

The kurtosis is given by

$$\begin{aligned}
\text{kurtosis} &= \frac{E(X-E(X))^4}{(\text{var}(X))^2} \\
&= \frac{E\left(X-\mu-\frac{\sigma(1-2p)}{p(1-p)}\right)^4}{\left(\frac{\sigma^2(1-2p+2p^2)}{p^2(1-p)^2}\right)^2} \\
&= \frac{p^4(1-p)^4}{\sigma^4(1-2p+2p^2)^2}\Big(E\Big((X-\mu)^4-4(X-\mu)^3\frac{\sigma(1-2p)}{p(1-p)}+6(X-\mu)^2\frac{\sigma^2(1-2p)^2}{p^2(1-p)^2} \\
&\quad -4(X-\mu)\frac{\sigma^3(1-2p)^3}{p^3(1-p)^3}+\frac{\sigma^4(1-2p)^4}{p^4(1-p)^4}\Big)\Big) \\
&= \frac{1}{(1-2p+2p^2)^2}\Big(24((1-p)^5+p^5)-24(1-2p)((1-p)^4-p^4) \\
&\quad +12(1-2p)^2((1-p)^3+p^3)-4(1-2p)^3((1-p)^2-p^2)+(1-2p)^4\Big) \\
&= \frac{9p^4+6p^2(1-p)^2+9(1-p)^4}{(1-2p+2p^2)^2}
\end{aligned}$$

Note that the skewness formula given in Yu and Zhang (2005) is incorrect. As a result, the skewness plot shown in Yu and Zhang (2005) is the reverse of the plot in Figure 4.2, that is, positive skewness and negative skewness correspond to $p > 1/2$ and $p < 1/2$, respectively.

Appendix 4.2

Using the SMU form of the ALD, given in (4.4), the list of full conditionals for the parameters in the SV model with AL distributed error is given as follows:

- Full conditional density for h_t,

  $h_t|\boldsymbol{h}_{-t}, \boldsymbol{y}, \psi_0, \psi_1, \psi_2, \boldsymbol{u}, p, \mu, \phi, \tau^2 \sim N(\mu_{h_t}, \sigma^2_{h_t})I(a_{h_t} < h_t < b_{h_t})$

$$\text{where } \mu_{h_t} = \begin{cases} \mu+\phi(h_2-\mu)-\frac{\tau^2}{2} & \text{if} \quad t=1, \\ \mu+\frac{\phi[(h_{t-1}-\mu)+(h_{t+1}-\mu)]-\tau^2/2}{1+\phi^2} & \text{if} \quad t=2,...,n-1, \quad \text{and} \\ \mu+\phi(h_{n-1}-\mu)-\frac{\tau^2}{2} & \text{if} \quad t=n, \end{cases}$$

$$\sigma^2_{h_t} = \begin{cases} \tau^2 & \text{if} \quad t=1,n \quad \text{and} \\ \frac{\tau^2}{1+\phi^2} & \text{if} \quad t=2,...,n-1, \end{cases}$$

$$
\begin{aligned}
a_{h_t} &= \ln(y_t - \mu_{y_t})^2 + \ln c_p^2 - \ln(1-p)^2 - \ln u_t^2, \\
b_{h_t} &= \ln(y_t - \mu_{y_t})^2 + \ln c_p^2 - \ln p^2 - \ln u_t^2,
\end{aligned}
$$

- Full conditional density for ψ_0,

$$
\begin{aligned}
\psi_0|\boldsymbol{h}, \boldsymbol{y}, \psi_1, \psi_2, \boldsymbol{u}, p, \mu, \phi, \tau^2 &\propto N(\mu_{\psi_0}, \sigma^2_{\psi_0}) I(-10, 10) \\
&\quad \times \prod_{t=1}^{n} \left(y_t | \mu_{y_t} - \frac{e^{h_t/2} p u_t}{c_p}, \mu_{y_t} + \frac{e^{h_t/2}(1-p)u_t}{c_p} \right) \\
&\propto N(\mu_{\psi_0}, \sigma^2_{\psi_0}) I(\max(-10, a_{\psi_0,t}) < \psi_0 < \min(10, b_{\psi_0,t}))
\end{aligned}
$$

where

$$
\begin{aligned}
a_{\psi_0,t} &= y_t - \psi_1 y_{t-1} - \psi_2 z_{t-1} - \frac{e^{h_t/2}(1-p)u_t}{c_p}, \\
b_{\psi_0,t} &= y_t - \psi_1 y_{t-1} - \psi_2 z_{t-1} + \frac{e^{h_t/2} p u_t}{c_p}.
\end{aligned}
$$

- Full conditional density for ψ_1,

$$
\begin{aligned}
\psi_1|\boldsymbol{h}, \boldsymbol{y}, \psi_0, \psi_2, \boldsymbol{u}, p, \mu, \phi, \tau^2 &\propto U(\psi_1|-1, 1) \times \prod_{t=1}^{n} U\left(y_t | \mu_y - \frac{e^{h_t/2} p u_t}{c_p}, \mu_y + \frac{e^{h_t/2}(1-p)u_t}{c_p} \right) \\
&\propto U\left(\max(-1, a_{\psi_1,t}), \min(1, b_{\psi_1,t})\right),
\end{aligned}
$$

where

$$
\begin{aligned}
a_{\psi_1,t} &= \frac{1}{y_{t-1}} \left(y_t - \psi_0 - \psi_2 z_{t-1} - \text{sign}(y_{t-1}) \frac{e^{h_t/2}(1-p^*)u_t}{c_p} \right), \\
b_{\psi_1,t} &= \frac{1}{y_{t-1}} \left(y_t - \psi_0 - \psi_2 z_{t-1} + \text{sign}(y_{t-1}) \frac{e^{h_t/2} p^* u_t}{c_p} \right), \\
p^* &= pI(y_{t-1} > 0) + (1-p)(1 - I(y_{t-1} > 0)).
\end{aligned}
$$

- Full conditional density for ψ_2,

$$
\begin{aligned}
\psi_2|\boldsymbol{h}, \boldsymbol{y}, \psi_0, \psi_1, \boldsymbol{u}, p, \mu, \phi, \tau^2 &\propto N(\mu_{\psi_2}, \sigma^2_{\psi_2}) I(-10, 10) \\
&\quad \times \prod_{t=1}^{n} U\left(y_t | \mu_{y_t} - \frac{e^{h_t/2} p u_t}{c_p}, \mu_{y_t} + \frac{e^{h_t/2}(1-p)u_t}{c_p} \right) \\
&\propto N(\mu_{\psi_2}, \sigma^2_{\psi_2}) I\left(\max(-10, a_{\psi 2,t}), \min(10, b_{\psi_2,t})\right),
\end{aligned}
$$

where

$$\begin{aligned} a_{\psi_2,t} &= \frac{1}{z_{t-1}}\left(y_t - \psi_0 - \psi_1 y_{t-1} - \text{sign}(z_{t-1})\frac{e^{h_t/2}(1-p^{'})u_t}{c_p}\right), \\ b_{\psi_2,t} &= \frac{1}{z_{t-1}}\left(y_t - \psi_0 - \psi_1 y_{t-1} + \text{sign}(z_{t-1})\frac{e^{h_t/2}p^{'}u_t}{c_p}\right), \\ p^{'} &= pI(z_{t-1}>0) + (1-p)(1-I(z_{t-1}>0)). \end{aligned}$$

- Full conditional density for u_t,

$$\begin{aligned} u_t|\boldsymbol{h},\boldsymbol{y},\psi_0,\psi_1,\psi_2,p,\mu,\phi,\tau^2 &\propto Ga(u_t|2,1)\times U\left(y_t|\mu_{y_t} - \frac{e^{h_t/2}pu_t}{c_p}, \mu_{y_t} + \frac{e^{h_t/2}(1-p)u_t}{c_p}\right) \\ &\propto \exp(u_t|1)I\left(u_t > \max\left(-\frac{c_p(y_t-\mu_{y_t})e^{-h_t/2}}{p}, \frac{c_p(y_t-\mu_{y_t})e^{-h_t/2}}{1-p}\right)\right) \end{aligned}$$

- Full conditional density for p,

$$\begin{aligned} p|\boldsymbol{h},\boldsymbol{y},\psi_0,\psi_1,\psi_2,\boldsymbol{u},\mu,\phi,\tau^2 &\propto U(0,1)\times\prod_{t=1}^{n} U\left(y_t|\mu_{y_t} - \frac{e^{h_t/2}pu_t}{c_p}, \mu_{y_t} + \frac{e^{h_t/2}(1-p)u_t}{c_p}\right) \\ &\propto U(\max a_{p,t}, \min b_{p,t}) \end{aligned}$$

where

$$a_{p,t} = \begin{cases} 0, & y_t > \mu_{y_t}, w_t > 1, \\ \frac{-1+\sqrt{w_t-1}}{w_t-2}, & y_t < \mu_{y_t}, w_t > 1, w_t \neq 2, \\ 1/2, & y_t < \mu_{y_t}, w_t = 2; \end{cases}$$

$$b_{p,t} = \begin{cases} \frac{w_t-1-\sqrt{w_t-1}}{w_t-2}, & y_t > \mu_{y_t}, w_t > 1, w_t \neq 2, \\ 1/2, & y_t > \mu_{y_t}, w_t = 2, \\ 1, & y_t < \mu_{y_t}, w_t > 1, \end{cases}$$

$$w_t = \frac{e^{h_t/2}u_t^2}{(y_t-\mu_{y_t})^2}.$$

- Full conditional density for μ,

$$\begin{aligned} \mu|\boldsymbol{h},\boldsymbol{y},\psi_0,\psi_1,\psi_2,\boldsymbol{u},p,\phi,\tau^2 &\propto N(a_\mu,b_\mu)N\left(h_1|\mu,\frac{\tau^2}{1-\phi^2}\right) \\ &\quad \prod_{t=2}^{n} N(h_t|\mu+\phi(h_{t-1}-\mu),\tau^2) \\ &\sim N(m_\mu,\sigma_\mu^2) \end{aligned}$$

where

$$\begin{aligned} m_\mu &= \sigma_\mu^2 \left(\frac{a_\mu}{b_\mu} + \frac{(1-\phi^2)}{\tau^2} h_1 + \frac{(1-\phi)}{\tau^2} \sum_{t=2}^{n} (h_t - \phi h_{t-1}) \right) \\ \sigma_\mu^2 &= \left(\frac{1}{b_\mu} + \frac{1}{\tau^2} [(n-1)(1-\phi)^2 + (1-\phi^2)] \right)^{-1} \end{aligned}$$

- Full conditional density for τ^2,

  $\tau^2 | \boldsymbol{h}, \boldsymbol{y}, \psi_0, \psi_1, \psi_2, \boldsymbol{u}, p, \mu, \phi \sim IG(a, b)$

  where

$$\begin{aligned} a &= a_\tau + \frac{n}{2} \\ b &= b_\tau + \frac{(1-\phi^2)(h_1-\mu)^2 + \sum_{t=2}^{n} \{h_t - [\mu + \phi(h_{t-1} - \mu)]\}^2}{2} \end{aligned}$$

- Full conditional density for ϕ,

$$\begin{aligned} \phi | \boldsymbol{h}, \boldsymbol{y}, \psi_0, \psi_1, \psi_2, \boldsymbol{u}, p, \mu, \tau^2 \;\propto\; & Be\left(\frac{\phi+1}{2} | 20, 1.5 \right) \times N\left(h_1 | \mu, \frac{\tau^2}{1-\phi^2} \right) \\ & \times \prod_{t=2}^{n} N(h_t | \mu + \phi(h_{t-1} - \mu), \tau^2). \end{aligned}$$

CHAPTER 5

Bivariate SV models with Application to Electricity Prices

So far we have been concentrating on univariate SV models. In this chapter, we extend the univariate SV models to multivariate SV models, for the modelling of price and price volatilities of the Australian wholesale electricity market. The use of SV models for electricity prices is still quite limited, as GARCH models have been widely adopted in this area. Our study using electricity price data is motivated by the fact that the operation of the electricity market is similar to that of the financial market and the data share many similar characteristics. We make a significant contribution in developing multivariate SV models that incorporate heavy-tailness, skewness, dynamic correlation for the prices as well as bilateral Granger causality to explain the dependence between different regional market volatilities in the context of the Australian electricity market.

5.1. Background

As mentioned in previous chapters, univariate SV models offer flexible alternatives to ARCH-type models in characterising the volatility inherent in financial time series data. Hence the univariate SV models have received considerable attention in the literature, regardless of whether it is from the modelling, distributional or estimation perspectives. A survey on univariate SV literature is given in Shephard (1996, 2005) and various estimation methods for univariate SV models are reviewed in Broto and Ruiz (2004).

Although there have been many successful applications of a wide variety of univariate GARCH and SV models, there are both theoretical and practical reasons why univariate volatility models should be extended to the multivariate case. This is especially the case of financial decision making, for example, derivative pricing, asset allocation, portfolio management and risk management, that clearly needs the knowledge of correlation structures. As a result, a wide range of multivariate GARCH models has been developed, analysed and applied in financial context. A comprehensive literature review on multivariate GARCH models can be found in McAleer (2005) and Bauwens *et al.* (2006).

Compared to the multivariate GARCH model literature, the number of studies on multivariate SV model is much fewer. The first multivariate SV model proposed was due to Harvey *et al.* (1994) and this basic multivariate SV model is analogous to the constant conditional correlation multivariate GARCH model proposed by Bollerslev (1990) when there are no volatility spillovers from other assets. This first basic SV model is given by

$$\begin{aligned} y_t &= H_t^{1/2}\epsilon_t, \\ H_t^{1/2} &= \text{diag}\{\exp(h_{1t}/2),\ldots,\exp(h_{mt}/2)\} = \text{diag}\{\exp(\boldsymbol{h}_t/2)\} \\ \boldsymbol{h}_{t+1} &= \boldsymbol{\mu} + \boldsymbol{\phi}\circ\boldsymbol{h}_t + \boldsymbol{\eta}_t, \\ \begin{pmatrix} \boldsymbol{\epsilon}_t \\ \boldsymbol{\eta}_t \end{pmatrix} &\sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \boldsymbol{P}_\epsilon & O \\ O & \boldsymbol{\Sigma}_\eta \end{pmatrix}\right), \end{aligned}$$

where $\boldsymbol{h} = (h_{1t},\ldots,h_{mt})'$ is a $m\times 1$ vector of log-volatilities, $\boldsymbol{\mu}$ and $\boldsymbol{\phi}$ are $m\times 1$ parameter vectors, the operator $\circ$ denotes the Hadamard (or element-by-element) product, $\boldsymbol{\Sigma}_\eta = \{\sigma_{\eta,ij}\}$ is a positive definite covariance matrix and $\boldsymbol{P}_\epsilon = \{\rho_{ij}\}$ is the correlation matrix with $\rho_{ii} = 1$ and $|\rho_{ij}| < 1$ for any $i \neq j$, $i,j = 1,\ldots,m$.

Since then, the literature on multivariate SV models has developed significantly over the years. Important contributions include Danielsson (1998), Liesenfeld and Richard (2003) and Chib *et al.* (2006), among many others. A detailed review of the literature on the modelling, estimation and evaluation of discrete time multivariate SV models is given in Asai *et al.* (2006).

While the multivariate SV models have some theoretical appeal and statistical attractions, the model estimation is a difficult task. Apart from the inherent problem of estimating multivariate models, for example the parameter space is usually of high dimensionality, the multivariate SV models also have the problem of having a likelihood function depending upon high-dimensional integrals. This is the same problem mentioned when univariate SV models are discussed in previous chapters. Moreover, as mentioned in Yu and Meyer (2006), 'the computation of model comparison criteria becomes extensive and demanding' which is a direct consequence of the difficulties associated with parameter estimation. However, model comparison criteria that are computationally inexpensive are required so that a range of models of various types can be easily and quickly compared in empirical applications.

A variety of estimation methods for univariate SV models is briefly discussed in Chapter 1. These include some less efficient methods such as the GMM and QML methods. More efficient methods include the SML, numerical maximum likelihood and MCMC methods. However, only a small subset of these methods has been applied in the estimation of multivariate SV models in the literature. Bayesian MCMC method, in particular, has been extensively adopted to estimate multivariate SV models as it has the technical advantage over classical inferential techniques that there is no need to use numerical optimisation. This advantage has practical importance when the number of parameters is large, as in the case of multivariate SV models. Yu and Meyer (2006) presented some existing multivariate SV model specifications and proposed several extensions. More importantly, the authors demonstrated how different multivariate SV models can be estimated and compared routinely with MCMC via `WinBUGS`.

The majority of existing research in the volatility model literature illustrates the model specifications and estimation techniques using financial asset return data. Much less work using these advanced econometric techniques has been undertaken in energy markets, for example, electricity markets. We aim to investigate the application of SV models in this less studied market by first giving a review of the existing studies below.

One prominent characteristic of electricity markets is the high level of price volatility. Devising a model that describes the dynamics of price volatility is crucial since high volatility causes uncertainty and instability in generators' revenues, suppliers' costs, investment decisions, risk management and pricing derivatives and hedge contracts. Existing models that attempt to describe the volatilities observed in the electricity markets as well as accounting for volatility clustering include the ARCH and GARCH models. Thomas and Mitchell (2005) examined the applicability of four different GARCH model specifications to describe the intra-day price volatilities for five regional markets in the Australian National Electricity market over a six-year sample period. The approach used in their study involves removing seasonal effects and outlier effects prior to estimating the conditional mean and variance equations in the GARCH estimation process, which is different to previous GARCH-based studies on Australian electricity prices where seasonality and outliers effects are incorporated into the conditional mean equation.

Higgs and Worthington (2005) investigated the half-hourly price volatility processes in four major Australian electricity markets using five different ARCH model specifications, namely, the GARCH, Risk Metrics, normal asymmetric power ARCH (APARCH), Student-t APARCH and skew Student-t APARCH models. Based on some model selection criteria such as the log-likelihood and AIC, their results demonstrated that the skew Student-t APARCH model, which takes account of right skewed and heavy-tailed characteristics of the data, performs the best in three of the four markets analysed. The statistical significance of the skewness parameter in the skew Student-t distribution and the superior performance of the model with skew Student-t distribution have motivated the use of different types of skew distributions in our study. The skew Student-t distribution and its multivariate extension is also used in Panagiotelis and Smith (2008) where a first order vector autoregressive model is applied to hourly average electricity prices of NSW. In their study, model estimation is performed using Bayesian method with MCMC algorithm and empirical results indicate that better model fit is achieved with skew Student-t disturbance in peak hours using the continuous ranked probability score as model assessment. Surveys of GARCH models used in electricity markets outside Australia can be found in Hadsell *et al.* (2004) and Solibakke (2002).

The studies mentioned above for the Australian electricity market concentrated on analysing each regional market separately. It is important however, to investigate the inter-relationships between regional electricity markets as 'a fuller understanding of the pricing relationships between these markets enables the benefits of interconnection to be assessed as a step towards the fuller integration of the regional electricity markets into a national electricity market.' (Higgs, 2009). Each region in the Australian National Electricity market is connected to another state's system through interconnection transmission lines but there are limitations on the physical transfer capacity of these interconnectors. Details of the interconnection is given in Section 4.2. Worthington *et al.* (2005) applied a multivariate GARCH model to examine the inter-relationships among five Australian spot electricity markets. More specifically, the model is applied to identify the source and magnitude of volatility spillovers using daily spot prices from December 1998 to June 2001. Their empirical results indicate that only Queensland and the Snowy Mountains Hydroelectric Scheme markets exhibit significant own price spillover. This would suggest that in general, lagged price from either each local market or

from other markets may not be useful in forecasting the Australian spot electricity prices, despite the presence of a national electricity market. However, when looking at volatility spillover effects, they found significant own-volatility and cross market volatility spillovers for nearly all markets, indicating shocks in those markets exert significant influence on price volatility due to the presence of strong ARCH and GARCH effects.

Higgs (2009) extended the work of Worthington *et al.* (2005) by employing three conditional correlation multivariate GARCH models with Student-t distribution to analyse the inter-relationships of electricity prices and price volatilities in four Australian spot electricity markets. The conditional correlation multivariate GARCH models include constant correlation and dynamic correlation specifications. Their results indicate that Tse and Tsui's (2002) dynamic conditional correlation model produces the best results for all four markets as the dynamic correlation process is able to accommodate the time-varying correlation volatility spillover effects across different regional markets. Furthermore, they found the highest correlations are evident between NSW and QLD; NSW and VIC; and SA and VIC due to their geographical proximity and large number of interconnectors between the markets. The lowest correlation is found between SA and QLD as there is no direct interconnector between these two regions. The success of the application of dynamic conditional correlation multivariate GARCH models to capture the volatility spillover effects in electricity markets has motivated us to employ a model which incorporates dynamic correlation in our study.

While much investigations have been done on the electricity markets using various GARCH model specifications, the literature on the use of SV models in electricity markets is still quite scarce, despite the fact that the SV models offer flexible and powerful alternatives to the GARCH models in modelling time varying volatility for 'conventional' financial markets. Elliott *et al.* (2007) proposed a continuous time SV model with a hidden Markov chain to describe the behaviour of energy prices, including crude oil prices and electricity markets of Canada and Nordic countries. Montero *et al.* (2010) applied a threshold asymmetric SV model to describe the volatility behaviour of returns of other than electricity energy products. Montero *et al.* (2011) made the first attempt to apply a univariate threshold asymmetric SV model to the Spanish electricity spot market. The asymmetric GARCH model and the exponential GARCH model are also estimated to provide comparison and their results clearly favour the SV model specification over the popular GARCH-type models. Smith (2010)

extended the basic SV model to allow for periodic autoregression and applied the model to intraday electricity spot price in NSW. Exogenous variables including electricity demand and dummy variable for day type effects are included in the observation and log-volatility equations to capture some of the empirical features in the data. Bayesian methods using MCMC scheme is adopted for model estimation. To the best of our knowledge, this is the only study so far that employs the SV models for the analysis of electricity prices in the context of Australian electricity market. And yet in his study, only one regional market is analysed.

The main aim of this research is to investigate and apply the multivariate SV models, which are more commonly used in financial markets, to capture the effects of cross-correlation between prices and price volatilities of the four major regional electricity markets in the Australian National Electricity market. We employ some of the multivariate SV model specifications presented in Yu and Meyer (2006) and propose new specifications which take account of the empirical features of electricity data including leptokurtic and right-skewed characteristics by using two types of skew Student-t distributions. The two particular types of skew Student-t distributions are chosen as they conveniently have similar scale mixture of normals (SMN) representation and under such representation, we demonstrate the full conditional distributions for some of the model parameters can be reduced to standard distributions, hence parameter estimation using Gibbs sampling can be simplified. We demonstrate the ease of model implementation using the MCMC software `WinBUGS`, which is used in our previous studies.

The rest of the chapter is organised as follows. Section 5.2 describes in detail the Australian electricity market. Section 5.3 presents some details about the two types of skew Student-t distributions to be used. The specifications of various proposed bivariate SV models are given in Section 5.4. Section 5.5 discusses the methodology employed and describes model estimation and issues related to implementation of the models in `WinBUGS`. A list of full conditional distributions of the model parameters is also given in the same section. Section 5.6 describes the data employed in the empirical study and some summary statistics along with some empirical features of the data are also presented. Empirical results and detailed discussion are given in Section 5.7. Finally, a concluding remark with possible future extensions is given in Section 5.8.

5.2. The Australian National Electricity Market (NEM)

The Australian National Electricity Market (NEM) was established in 1998 and it encompasses five state-based regional markets, namely New South Wales (NSW) (including the Australian Capital Territory (ACT)), Victoria (VIC), Queensland (QLD), South Australia (SA) and Tasmania. The non-state based Snowny Mountains Hydroelectric Scheme was part of the NEM and it was abolished in 2008. The remaining state of Western Australia and Northern Territory are not members of the NEM due to interconnection and transmission infeasibility across geographically distant regions. Of these members of the NEM, NSW has the largest generation capacity and highest electricity demand.

The NEM is managed by the Australian Energy Market Operator (AEMO), which was established in July, 2009. The AEMO is responsible for electricity market and system operation, whose electricity functions are previously undertaken by the National Electricity Market Management Company (NEMMCO). The AEMO operates the wholesale market for electricity as a spot market where electricity supply and demand requirements are balanced through a centrally coordinated dispatch process. Generators submit bids (or stacks) to the AEMO prior to generation specifying the amount of electricity they are prepared to supply at a certain price. From all the bids submitted, AEMO determines which generators to dispatch based on the principle of balancing electricity supply and demand requirements in the most cost-efficient way. More specifically, bids are stacked in ascending price order in each five minute dispatch period. AEMO then matches forecast of expected electricity demand in the next five minutes against the industry supply curve submitted by all generators. On a least-cost generation basis, generators are progressively scheduled to dispatch until forecast demand is fully met. The dispatch price for each five minute interval is the price offered by the most expensive generator to be dispatched to meet total demand. The spot price of electricity for the half-hour trading interval is the time-weighted average of the six dispatch prices. All generators receive this price for dispatching electricity into the pool and it is the price market customers pay for electricity consumption in that half hour period. Each of the NEM's state regions has a separate spot price for each trading interval.

The NEM consists of five state-based electrical regions where each state developed its own generation and transmission network. Between adjacent NEM regions there are high-voltage interconnector transmission lines that link one state to another state's system. These

interconnectors transmit electricity into a region when demands are higher than local generators can provide, or when the price of electricity in an adjacent region is low enough to displace the local supply. However, the scheduling of generators to meet demand by AEMO using interconnectors is constrained by the physical transfer capacity of the interconnectors. For example, electricity can be transmitted from a region with very low prices to an adjacent region with high prices across interconnectors up to the maximum technical capacity of the interconnector. Once the limit is reached, AEMO's system will dispatch the lowest priced local generators to meet the outstanding consumer demand. There are currently two interconnectors between QLD and NSW. QLD can export 207 megawatt (MW) of electricity to NSW through the Terranora interconnector and the flow limit is capped at 1078 MW for the QLD and NSW Interconnector (QNI). NSW, on the other hand, can export up to 520 MW to QLD through QNI. VIC can export 460 MW to SA through the Heywood interconnector and 220 MW through the Murraylink interconnector. The two interconnectors together can import 460 MW to VIC from SA. With the abolition of Snowy Mountains Hydroelectric Scheme, the VIC-NSW interconnector commenced operation from 1 July 2008 where NSW can export up to 1900 MW to VIC. TAS, as the last state to join the NEM, is connected to VIC by the Basslink interconnector, which was fully activated in 2006. The Basslink interconnector enables TAS to import 480 MW from and export 600 MW to VIC. At present, QLD is only directly linked to NSW and there is no direct interconnector between SA and NSW. An illustration of the inter-regional connections via interconnectors is shown in Figure 5.1. With increasing transmission capacity and greater number of regulated interconnectors between separate regions, the NEM is developing towards a nationally integrated and efficient electricity market.

5.3. Skew Student-t distributions as SMN

In this section, we give details about the two types of skew Student-t distributions that we employ in our study. In particular, we present the SMN representation for each of the distributions. Firstly, we need to following definition.

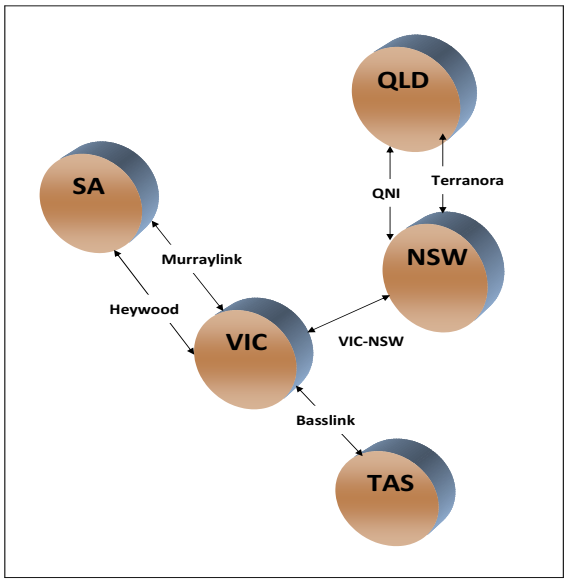


FIGURE 5.1. Interconnectors in the NEM.

Definition 5.3.1. *If a random variable Y follows a Generalised Inverse Gaussian (GIG) distribution with parameters p, a and b, denoted by $GIG(p, a, b)$, then it has density give by*

$$f_{GIG}(y|p,a,b) = \frac{\left(\frac{b}{a}\right)^p}{2K_p(ab)} y^{p-1} \exp\left(-\frac{1}{2}(a^2 y^{-1} + b^2 y)\right), \quad \text{for } y > 0 \tag{5.1}$$

where $p \in \mathbb{R}$, $a > 0$, $b > 0$ and K_p is the modified Bessel function of the third kind with index p.

The GIG distribution nests the gamma distribution (when $p > 0$, $a \to 0$) and the inverse gamma distribution (when $p < 0$, $b \to 0$).

5.3.1. The SMN representations of two skew Student-t distributions. The skew Student-t distribution introduced in Demarta and McNeil (2005), hereafter referred to as ST_1, has the density

$$f_{ST_1}(\boldsymbol{x}) = c \frac{K_{\frac{\nu+n}{2}}\left(\sqrt{(\nu + (\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}))\boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}}\right) \exp((\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma})}{\left(\sqrt{(\nu + (\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}))\gamma'\boldsymbol{\Sigma}^{-1}\gamma}\right)^{-\frac{\nu+n}{2}} \left(1 + \frac{(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}{\nu}\right)^{\frac{\nu+n}{n}}}, \tag{5.2}$$

where $\boldsymbol{\mu} \in \mathbb{R}^n$ is the location vector, $\boldsymbol{\Sigma}$ is an $n \times n$ symmetric positive definite matrix, $\boldsymbol{\gamma}$ controls the asymmetry of the distribution, ν is the shape parameter and K_τ is the modified

Bessel function of the third kind with index p. The normalising constant is given by

$$c = \frac{2^{1-\frac{\nu+n}{2}}}{\Gamma\left(\frac{\nu}{2}\right)(\pi\nu)^{\frac{n}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}}$$

The density (5.2) has a normal mean-variance mixture representation defined as

$$f_{ST_1}(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\gamma},\nu) = \int_0^\infty N_n\left(\boldsymbol{\mu}+\boldsymbol{\gamma}\lambda^{-1},\lambda^{-1}\boldsymbol{\Sigma}\right) Ga\left(\lambda|\frac{\nu}{2},\frac{\nu}{2}\right) d\lambda, \tag{5.3}$$

where $N_n(,)$ is the standard n-multivariate normal distribution and λ is the mixing parameter. We shall show this for the univariate case.

Proof. Using (5.3), the pdf of X is

$$\begin{aligned}
&\int_0^\infty N\left(x|\mu+\frac{\gamma}{\lambda},\frac{\sigma^2}{\lambda}\right) Ga\left(\lambda|\frac{\nu}{2},\frac{\nu}{2}\right) d\lambda \\
=\ & \int_0^\infty \frac{\sqrt{\pi}}{\sqrt{2\pi}\sigma}\exp\left(-\frac{\lambda}{2\sigma^2}\left(x-\left(\mu+\frac{\gamma}{\lambda}\right)\right)^2\right)\frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)}\lambda^{\frac{\nu}{2}-1}\exp\left(-\frac{\nu}{2}\lambda\right)d\lambda \\
=\ & \frac{1}{\sqrt{2\pi}\sigma}\frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)}\int_0^\infty \lambda^{\frac{\nu}{2}-\frac{1}{2}}\exp\left(-\frac{\lambda}{2\sigma^2}\left(x-\left(\mu+\frac{\gamma}{\lambda}\right)\right)^2-\frac{\nu}{2}\lambda\right)d\lambda \\
=\ & \frac{1}{\sqrt{2\pi}\sigma}\frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)}\exp\left(\frac{\gamma}{\sigma^2}(x-\mu)\right)\int_0^\infty \lambda^{\frac{\nu}{2}-\frac{1}{2}}\exp\left(-\frac{1}{2}\left(\lambda\left(\frac{(x-\mu)^2}{\sigma^2}+\nu\right)+\frac{\gamma^2}{\sigma^2}\lambda^{-1}\right)\right)d\lambda \\
=\ & \frac{1}{\sqrt{2\pi}\sigma}\frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)}\exp\left(\frac{\gamma}{\sigma^2}(x-\mu)\right)2K_{\frac{\nu+1}{2}}\left(\sqrt{\nu+\frac{(x-\mu)^2}{\sigma^2}}\frac{|\gamma|}{\sigma}\right)\left(\frac{\frac{|\gamma|}{\sigma}}{\sqrt{\nu+\frac{(x-\mu)^2}{\sigma^2}}}\right)^{\frac{\nu+1}{2}} \\
=\ & \frac{\sqrt{2}}{\sqrt{\pi}\sigma}\frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)}\exp\left(\frac{\gamma}{\sigma^2}(x-\mu)\right)K_{\frac{\nu+1}{2}}\left(\sqrt{\nu+\frac{(x-\mu)^2}{\sigma^2}}\frac{|\gamma|}{\sigma}\right)\left(\frac{1}{\sqrt{\nu+\frac{(x-\mu)^2}{\sigma^2}}\frac{|\gamma|}{\sigma}}\right)^{-\frac{\nu+1}{2}}\left(\frac{1}{\nu+\frac{(x-\mu)^2}{\sigma^2}}\right)^{\frac{\nu+1}{2}} \\
=\ & \frac{\sqrt{2}}{\sqrt{\pi}\sigma}\frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}\nu^{-\frac{\nu+1}{2}}}{\Gamma\left(\frac{\nu}{2}\right)}\exp\left(\frac{\gamma}{\sigma^2}(x-\mu)\right)K_{\frac{\nu+1}{2}}\left(\sqrt{\nu+\frac{(x-\mu)^2}{\sigma^2}}\frac{|\gamma|}{\sigma}\right)\left(\frac{1}{\sqrt{\nu+\frac{(x-\mu)^2}{\sigma^2}}\frac{|\gamma|}{\sigma}}\right)^{-\frac{\nu+1}{2}}\left(\frac{1}{1+\frac{(x-\mu)^2}{\nu\sigma^2}}\right)^{\frac{\nu+1}{2}} \\
=\ & \frac{2^{\frac{1}{2}-\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\nu}\sigma}\exp\left(\frac{\gamma}{\sigma^2}(x-\mu)\right)K_{\frac{\nu+1}{2}}\left(\sqrt{\nu+\frac{(x-\mu)^2}{\sigma^2}}\frac{|\gamma|}{\sigma}\right)\left(\frac{1}{\sqrt{\nu+\frac{(x-\mu)^2}{\sigma^2}}\frac{|\gamma|}{\sigma}}\right)^{-\frac{\nu+1}{2}}\left(\frac{1}{1+\frac{(x-\mu)^2}{\nu\sigma^2}}\right)^{\frac{\nu+1}{2}},
\end{aligned}$$

since we are integrating over the $GIG(p,a,b)$ density in (5.1) with $p=\frac{\nu+1}{2}$, $a^2=\frac{\gamma^2}{\sigma^2}$ and $b^2=\nu+\frac{(x-\mu)^2}{\sigma^2}$. □

Moments of this skew Student-t distribution can be calculated using the normal mixture representation of the distribution and are given by

$$\begin{aligned} E(\boldsymbol{X}) &= \boldsymbol{\mu} + \frac{\nu}{\nu-2}\boldsymbol{\gamma}, \\ \text{cov}(\boldsymbol{X}) &= \frac{\nu}{\nu-2}\boldsymbol{\Sigma} + \frac{2\nu^2}{(\nu-2)^2(\nu-4)}\boldsymbol{\gamma}\boldsymbol{\gamma}^{'}. \end{aligned}$$

The covariance matrix is only finite when $\nu > 4$ which is in contrast with the symmetric Student-t distribution where only $\nu > 2$ is required.

The second skew Student-t distribution we consider was first introduced in Branco and Dey (2001) in multivariate form, hereafter referred to as ST_2. This type of skew Student-t distribution is defined as resulting from variance-mixing of the skew normal with a gamma random variable, that is,

$$\boldsymbol{X} = \lambda^{-\frac{1}{2}}\boldsymbol{Z}^*, \tag{5.4}$$

where $\lambda \sim Ga\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$. The multivariate skew normal, $\boldsymbol{Z}^* \sim SN_n(\boldsymbol{\gamma}, \boldsymbol{\Sigma})$, is constructed by the method in Azzalini and Dalla Valle (1996) and the normal mean mixture representation. The pdf of $\boldsymbol{Z}^*$ is given by

$$f(\boldsymbol{z}^*) = 2\phi_n(\boldsymbol{z}^*|\boldsymbol{0}, \boldsymbol{\Sigma})\Phi(\boldsymbol{\gamma}^{'}\boldsymbol{z}^*), \tag{5.5}$$

where ϕ_n is the density of the n-dimensional normal density with mean $\boldsymbol{0}$ and covariance matrix $\boldsymbol{\Sigma}$ and Φ is the distribution function of the normal distribution. By combining (5.4) and the normal mean mixing representation of the multivariate skew normal distribution, we have the following normal mean-variance mixture form of the ST_2 distribution defined as

$$\begin{aligned} \boldsymbol{X}|(\lambda, V) &\sim N_n(\boldsymbol{\mu} + \boldsymbol{\gamma}\lambda^{-1/2}V, \lambda^{-1}\boldsymbol{\Sigma}) \\ \lambda &\sim Ga\left(\frac{\nu}{2}, \frac{\nu}{2}\right) \end{aligned} \tag{5.6}$$

where $V \sim$ half normal with pdf

$$f(v) = \begin{cases} \sqrt{\frac{2}{\pi}}e^{-\frac{v^2}{2}}, & \text{if } v > 0, \\ 0, & \text{otherwise.} \end{cases}$$

That is $V = |W|$, with $W \sim N(0, 1)$.

Using (5.4) and (5.5), the density of X defined in (5.6) is given by

$$f_{ST_2}(\boldsymbol{x}) = 2t_\nu(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\Sigma})F_{t_{\nu+n}}\left(\boldsymbol{\gamma}'\boldsymbol{x}\sqrt{\frac{\nu+n}{\nu+(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}}\right), \tag{5.7}$$

where t_ν is the pdf of a n-variate Student-t distribution with ν degrees of freedom and location and scale parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, respectively. $F_{t_{\nu+n}}$ is the distribution function of the Student-t distribution with $\nu+n$ degrees of freedom.

Proof. (Univariate case) The univariate skew normal rv Z^* has density

$$\frac{2}{\sigma}\phi(z^*|\mu,\sigma^2)\Phi\left(\gamma\frac{z^*-\mu}{\sigma}\right)$$

Let $X = Z^*Y^{-\frac{1}{2}}$, where $Y \sim Ga\left(\frac{\nu}{2},\frac{\nu}{2}\right)$. Then the density of X is

$$\begin{aligned}
f(x) &= \frac{2}{\sigma}\int_0^\infty \phi\left(x|\mu,\frac{\sigma^2}{y}\right)\Phi\left(\gamma\frac{x-\mu}{\sigma}\sqrt{y}\right)\frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)}y^{\frac{\nu}{2}-1}\exp\left(-\frac{\nu}{2}y\right)dy \\
&= \frac{2}{\sigma}\frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)}\int_0^\infty \frac{\sqrt{y}}{\sqrt{2\pi}\sigma}\exp\left(-\frac{y}{2\sigma^2}(x-\mu)^2\right)\Phi\left(\gamma\frac{x-\mu}{\sigma}\sqrt{y}\right)y^{\frac{\nu}{2}-1}dy \\
&= \frac{2}{\sigma}\frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)}\frac{1}{\sqrt{2\pi}\sigma}\int_0^\infty y^{\frac{\nu}{2}-\frac{1}{2}}\exp\left(-\frac{1}{2}y\left(\nu+\frac{(x-\mu)^2}{\sigma^2}\right)\right)\Phi\left(\gamma\frac{x-\mu}{\sigma}\sqrt{y}\right)dy \\
\text{let } s &= \left(\nu+\frac{(x-\mu)^2}{\sigma^2}\right)y; \\
&= \frac{2}{\sigma}\frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)}\frac{1}{\sqrt{2\pi}\sigma}\int_0^\infty \left(\frac{s}{\nu+\frac{(x-\mu)^2}{\sigma^2}}\right)^{\frac{\nu}{2}-\frac{1}{2}}\exp\left(-\frac{1}{2}s\right)\Phi\left(\gamma\frac{x-\mu}{\sigma}\sqrt{\frac{s}{\nu+\frac{(x-\mu)^2}{\sigma^2}}}\right) \\
&\qquad\qquad \times\left(\nu+\frac{(x-\mu)^2}{\sigma^2}\right)^{-1}ds \\
&= \frac{2}{\sigma}\frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)}\frac{1}{\sqrt{2\pi}\sigma}\left(\nu+\frac{(x-\mu)^2}{\sigma^2}\right)^{-\frac{\nu+1}{2}}\int_0^\infty s^{\frac{\nu+1}{2}-1}\exp\left(-\frac{1}{2}s\right)\Phi\left(\gamma\frac{x-\mu}{\sigma}\sqrt{\frac{s}{\nu+\frac{(x-\mu)^2}{\sigma^2}}}\right)ds \\
&= \frac{2}{\sigma}\frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)}\frac{1}{\sqrt{2\pi}\sigma}\left(\nu+\frac{(x-\mu)^2}{\sigma^2}\right)^{-\frac{\nu+1}{2}}\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\left(\frac{1}{2}\right)^{\frac{\nu+1}{2}}}\int_0^\infty \frac{\left(\frac{1}{2}\right)^{\frac{\nu+1}{2}}}{\Gamma\left(\frac{\nu+1}{2}\right)}s^{\frac{\nu+1}{2}-1}\exp\left(-\frac{1}{2}s\right) \\
&\qquad\qquad \times\Phi\left(\gamma\frac{x-\mu}{\sigma}\sqrt{\frac{s}{\nu+\frac{(x-\mu)^2}{\sigma^2}}}\right)ds \\
&= \frac{2}{\sigma}\frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu+1}{2}\right)}\frac{1}{\sigma\sqrt{2\pi}}\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\left(\frac{1}{2}\right)^{\frac{\nu+1}{2}}}\left(\nu+\left(\frac{x-\mu}{\sigma}\right)^2\right)^{-\frac{\nu+1}{2}}E_S\left(\Phi\left(\gamma\frac{x-\mu}{\sigma}\sqrt{\frac{s}{\nu+\frac{(x-\mu)^2}{\sigma^2}}}\right)\right)
\end{aligned}$$

where $S \sim Ga\left(\frac{\nu+1}{2},\frac{1}{2}\right)$

$$= \frac{2}{\sigma}\frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\sqrt{2\pi}\sigma}2^{\frac{\nu+1}{2}}\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}\nu^{-\frac{\nu+1}{2}}\left(1+\frac{(x-\mu)^2}{\nu\sigma^2}\right)^{-\frac{\nu+1}{2}}F_{t_{\nu+1}}\left(\gamma\frac{x-\mu}{\sigma}\sqrt{\frac{\nu+1}{\nu+\frac{(x-\mu)^2}{\sigma^2}}}\right)$$

according to Gupta, Chang and Huang (2002), Lemma 3 (given in Appendix 5.1.)

$$= \frac{2}{\sigma}\frac{1}{\sqrt{\pi\nu}\sigma}\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{(x-\mu)^2}{\nu\sigma^2}\right)^{-\frac{\nu+1}{2}}F_{t_{\nu+1}}\left(\gamma\frac{x-\mu}{\sigma}\sqrt{\frac{\nu+1}{\nu+\frac{(x-\mu)^2}{\sigma^2}}}\right)$$

$$= \frac{2}{\sigma}t_\nu\left(\frac{x-\mu}{\sigma}\right)F_{t_{\nu+1}}\left(\gamma\frac{x-\mu}{\sigma}\sqrt{\frac{\nu+1}{\nu+\frac{(x-\mu)^2}{\sigma^2}}}\right)$$

□

A comparison of the densities of these two type of skew Student-t distributions is shown in Figure 5.2. Both distributions are standardised to have mean 0 and unit variance. For both distributions positive γ values indicate the densities are skewed to the right and negative values of γ indicate the densities are skewed to the left, as shown in the graph. For the same skewness parameter value and the degrees of freedom, ST_1 has higher peak than ST_2 but ST_2 has a heavier-tail than ST_1 which may affect their performance in empirical applications. For more details on these two skew Student-t distributions, see Fung and Seneta (2010b).

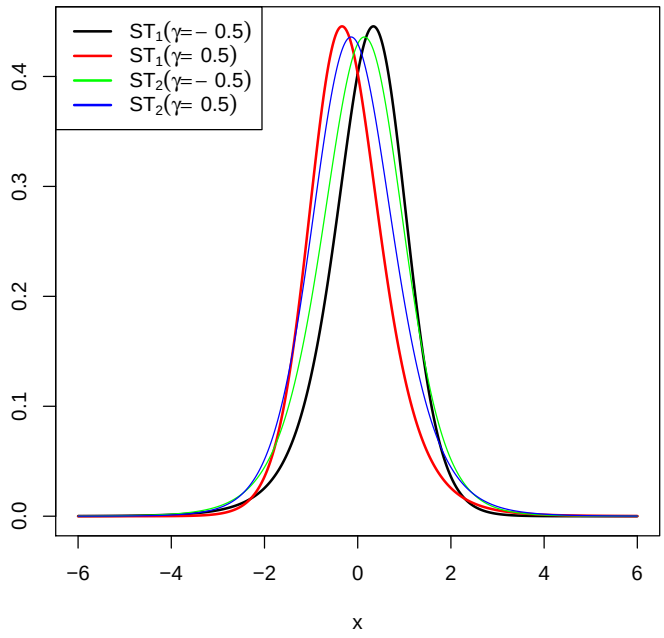


FIGURE 5.2. Density plots for ST_1 ST_2 with 8 degrees of freedom, standardised to have 0 mean and unit variance

5.4. Model specifications

To investigate the interconnection between the regional electricity markets, we concentrate on the bivariate models. Extending the simplest SV model in (1.1) with covariate effects for the observed prices, we define the model as follows

$$\begin{aligned} \boldsymbol{y}_t &= \boldsymbol{\mu}_y + \text{diag}(\xi_1, \xi_2)\boldsymbol{W}_t + \boldsymbol{\Sigma}_t\boldsymbol{\epsilon}_t \\ \boldsymbol{h}_{t+1} &= \boldsymbol{\mu} + \boldsymbol{\Phi}(\boldsymbol{h}_t - \boldsymbol{\mu}) + \text{diag}(\tau_1, \tau_2)\boldsymbol{\eta}_t \end{aligned} \tag{5.8}$$

where $\boldsymbol{y}_t = (y_{1t}, y_{2t})$ is the natural logarithms of daily prices of two markets respectively and $\boldsymbol{h}_t = (h_{1t}, h_{2t})$ are the corresponding log-volatilities. Also, $\boldsymbol{\mu}_y = (\mu_{y1}, \mu_{y2})$, $\boldsymbol{\mu} = (\mu_1, \mu_2)$, $\boldsymbol{\Sigma}_t = \text{diag}(\exp(\boldsymbol{h}_t)/2)$ and $\boldsymbol{\Phi} = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix}$. The vector $\boldsymbol{W}_t = (W_{1t}, W_{2t})$ consists of indicator variables for the day-of-the-week effect taking values of 1 if t is a weekend or public holiday and 0 otherwise. The day-of-the-week effect is documented in previous studies, for example, Higgs and Worthington (2008) found that the electricity prices are lower for weekends and public holidays than for weekday in four Australian electricity markets. In this basic bivariate SV model, no correlations across the observed price series or the volatilities are allowed as the two error processes $\boldsymbol{\epsilon}_t$ and $\boldsymbol{\eta}_t$ are both assumed to be independent and follow bivariate normal distributions. The diagonal elements of the matrix Φ, ϕ_{11} and ϕ_{22}, define the volatility persistence of each market. If the off-diagonal elements, ϕ_{12} and ϕ_{21}, are non-zero, then we have bilateral Granger causality existing between the two market volatilities.

To account for correlation between the two price series as well as empirical leptokurtic feature of the data, we consider as the first model, the *constant correlation t-SV model with Granger causality* (cc-t-G) as an extension of (5.8) in which we allow the innovations $\boldsymbol{\epsilon}_t$ to follow a bivariate Student-t distribution. That is,

$$\begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix} \sim t_\nu \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_\epsilon \\ \rho_\epsilon & 1 \end{pmatrix} \right).$$

Liesenfeld and Jung (2000), Chib *et al.* (2002) and Asai (2008) considered univariate SV models with Student-t distributed returns. Meyer and Yu (2000) and Wang *et al.* (2011a) studied the Student-t SV model by expressing the density of the Student-t distribution into a

SMN representation. The advantage of using the SMN representation is that the full conditional distributions for the parameters can be reduced to simpler forms. Hence the SV models with heavy-tailed distributions can be implemented using the MCMC algorithm without substantially increase the computing time. We use the multivariate SMN form of the bivariate Student-t density as shown in (2.2) such that conditioning on the mixing parameter λ_t, $\boldsymbol{\epsilon}_t$ has a bivariate normal distribution.

This cc-t-G SV model is capable of capturing several important characteristics. Firstly, a heavy-tailed Student-t distribution is assumed for the prices hence excess kurtosis is allowed. The Student-t distribution can also provide robust inference by accommodating outlier effects as compared to the normal distribution. Secondly, the correlation between the regional electricity prices is measured by the parameter ρ_ϵ and its magnitude indicates the degree of dependence. Lastly, not only the price series are allowed to be correlated, the dependence between the volatilities is modelled by bilateral Granger-causality. If the coefficients ϕ_{12} and ϕ_{21} are both significant, then we conclude the volatility of the first market price, h_1, Granger-causes the volatility of the second price, h_2 and vice versa.

An alternative way to model the dependence between the volatilities rather than using Granger causality is to assume a bivariate normal distribution. This leads to our second model, the cc-t-bN model. In this model, we have

$$\begin{aligned} \boldsymbol{h}_{t+1} &= \boldsymbol{\mu} + \text{diag}(\phi_{11}, \phi_{22})(\boldsymbol{h}_t - \boldsymbol{\mu}) + \text{diag}(\tau_1, \tau_2)\boldsymbol{\eta}_t, \\ \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix} &\sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_\eta \\ \rho_\eta & 1 \end{pmatrix} \right). \end{aligned}$$

Yu and Meyer (2006) compared these two specifications empirically by using exchange rate data and showed that the model with bivariate normal distribution for the volatilities outperformed the one with Granger causality when using DIC as a criterion for model selection. However, in their study, only one-way Granger causality was employed. We are interested in looking at the relative performance of these two specifications when a bilateral Granger causality is used in the context of modelling electricity prices.

The fixed correlation models presented above fall short of allowing for possible time-varying behaviour of the correlations. Hence a natural extension would be to extend fixed correlations to dynamic correlations that are time-varying. Based on the cc-t-G model, this

proposed model has the flexible formulation given by

$$\begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix} \sim t_\nu \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_{\epsilon t} \\ \rho_{\epsilon t} & 1 \end{pmatrix} \right).$$

We use Fisher transformation on $\rho_{\epsilon t}$ to ensure that it is bounded by -1 and 1 and for the modelling of time varying correlation $\rho_{\epsilon t}$, we follow Tsay (2002) and Yu and Meyer (2006) to assign an AR(1) stochastic process to the transformed correlation q_t as follows

$$\begin{aligned} \rho_{\epsilon t} &= \frac{\exp(q_t) - 1}{\exp(q_t) + 1}, \\ q_{t+1} &= \psi_0 + \psi_1(q_t - \psi_0) + \tau_\rho \upsilon_t, \quad \upsilon_t \overset{iid}{\sim} N(0,1). \end{aligned}$$

We refer to this model as the dc-t-G model. It is a more general model than the dynamic correlation SV model in Yu and Meyer (2006) who assumed no dependence between the volatilities and the model in Johansson (2009) who assumed normal distribution for the observed data.

The symmetric Student-t distribution is able to account for the observed excess kurtosis in the data, but it falls short of accounting for the observed skewness. Therefore, for the fourth and fifth proposed model, we further extend the cc-t-G model to have skew Student-t distributed innovations for the price series. The first type of skew Student-t distribution (ST_1) we adopt was first introduced in Demarta and McNeil (2005). The density of a multivariate ST_1 distribution and its scale mixture representation are given by (5.2) and (5.3) respectively. For the purpose of comparing different types of skew heavy-tailed distribution, we also consider the skew Student-t distribution, ST_2, proposed by Azzalini and Capitanio (2003). The multivariate version of ST_2 was originally introduced by Branco and Dey (2001) and the density and the variance-mixing form are given by (5.7) and (5.6) respectively. These two skew Student-t distributions are chosen because they have similar normal mixing representation which facilitates model implementation in `WinBUGS`. Moreover, these distributions have been employed and compared in Fung and Seneta (2010a) in financial applications and their statistical properties, in particular, the tail behaviour, are studied in Fung and Seneta (2010b). We refer to the cc-t-G model with symmetric Student-t distribution replaced by ST_1 and ST_2 distributions as cc-ST_1-G model and cc-ST_2-G model respectively. We will show in the empirical study, that the model with ST_2 distributed innovations outperforms the

one with ST_1 error distribution. Therefore, combining dynamic correlation specification and ST_2 error distribution, we propose the dc-ST_2-G SV model which is shown below in full

$$\begin{aligned}
\boldsymbol{y}_t &= \boldsymbol{\mu}_y + \text{diag}(\xi_1, \xi_2)\boldsymbol{W}_t + \boldsymbol{\Sigma}_t\boldsymbol{\epsilon}_t, \\
\boldsymbol{h}_{t+1} &= \boldsymbol{\mu} + \boldsymbol{\Phi}(\boldsymbol{h}_t - \boldsymbol{\mu}) + \text{diag}(\tau_1, \tau_2)\boldsymbol{\eta}_t, \\
\begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix} &\sim ST_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_{\epsilon t} \\ \rho_{\epsilon t} & 1 \end{pmatrix}\right),
\end{aligned}$$

with all the parameters defined earlier. This new model, as the last proposed model, allows for time-varying correlation between the observed prices as well as volatility spillover from one regional market to the other. Furthermore, empirical regularities including skewness and excess kurtosis of the prices are accounted for by employing a heavy-tailed skew distribution.

5.5. Model estimation

5.5.1. MCMC algorithm. For the models presented in the previous section, we use the Bayesian methodology to sample model parameters and latent volatilities from their posterior distributions using MCMC method with Gibbs sampling algorithm. We assign the following prior distribution for each of the model parameters. In our studies, we assign mostly non-informative or diffuse prior distributions. A beta prior $Be(a_\phi, b_\phi)$ is assigned to the persistence parameters ψ_1, $\phi^*_{11} = \frac{\phi_{11}+1}{2}$ and $\phi^*_{22} = \frac{\phi_{22}+1}{2}$. A conjugate normal prior with large variance is assigned to Granger causality coefficients ϕ_{12} and ϕ_{21}, intercept coefficients ψ_0, μ_{y1}, μ_{y2}, μ_{h1} and μ_{h2} and weekend and public holiday effect parameters ξ_1, ξ_2. A non-informative gamma prior distribution with restricted range is assigned to the degrees of freedom parameter ν and a conjugate inverse gamma prior distribution is assigned to variance parameters τ_1^2, τ_2^2 and τ_ρ^2. Lastly, a uniform prior is assigned to the correlation coefficient ρ_ε and ρ_η and skewness parameters γ_1 and γ_2.

Define $\boldsymbol{\nu}_{\boldsymbol{y}_t} = (\nu_{y_{1t}}, \nu_{y_{2t}})$, where

$$\begin{aligned}
\nu_{y_{1t}} &= \mu_{y1} + \xi_1 W_{1t}, \\
\nu_{y_{2t}} &= \mu_{y2} + \xi_2 W_{2t}.
\end{aligned}$$

The system of full conditional distributions is shown below.

- Full conditional density for h_{1t},

$$
\begin{aligned}
&f(h_{11}|y_{11}, y_{21}, h_{21}, h_{21}, h_{22}, \mu_1, \mu_2, \phi_{11}, \phi_{12}, \phi_{21}, \phi_{22}, \tau_1, \tau_2, \lambda_1, \rho_\epsilon, \boldsymbol{\nu_{y_1}}) \\
&\propto N\left(h_{11}|\mu_1, \frac{\tau_1^2}{1-\phi_{11}^2}\right) \times N(h_{12}|\mu_1 + \phi_{11}(h_{11}-\mu_1) + \phi_{12}(h_{21}-\mu_2), \tau_1^2) \\
&\quad \times N(h_{22}|\mu_2 + \phi_{21}(h_{11}-\mu_1) + \phi_{22}(h_{21}-\mu_2), \tau_2^2) \times N_2(y_{11}, y_{21}|\boldsymbol{\nu_{y_1}}, \boldsymbol{\Sigma}_t) \\
&\propto \exp\Bigg\{ -\frac{1}{2}\Bigg\{ h_{11}^2\left(\frac{1}{\tau_1^2} + \frac{\phi_{21}^2}{\tau_2^2}\right) - 2h_{11}\Bigg[-\frac{1}{2} - (h_{21}-\mu_2)\left(\frac{\phi_{11}\phi_{12}}{\tau_1^2} + \frac{\phi_{21}\phi_{22}}{\tau_2^2}\right) \\
&\quad + \frac{\phi_{11}(h_{12}-\mu_1)}{\tau_1^2} + \frac{\phi_{21}(h_{22}-\mu_2)}{\tau_2^2} + \mu_1\left(\frac{1}{\tau_1^2} + \frac{\phi_{21}^2}{\tau_2^2}\right)\Bigg] \\
&\quad - \frac{\lambda_1}{1-\rho_\epsilon}\left(\frac{(y_{11}-\nu_{y_{11}})}{\exp(h_{11})}\left((y_{11}-\nu_{y_{11}}) - 2\rho_\epsilon(y_{21}-\nu_{y_{21}})\exp\left(\frac{h_{11}-h_{21}}{2}\right)\right)\right)\Bigg\}\Bigg\}, \quad t=1 \\
&f(h_{1t}|y_{1t}, y_{2t}, h_{2t}, \mu_1, \mu_2, \phi_{11}, \phi_{12}, \phi_{21}, \phi_{22}, \tau_1, \tau_2, \lambda_t, \rho_\epsilon, \boldsymbol{\nu_{y_t}}) \\
&\propto N(h_{1_{t+1}}|\mu_1 + \phi_{11}(h_{1t}-\mu_1) + \phi_{12}(h_{2t}-\mu_2), \tau_1^2) \times N(h_{1t}|\mu_1 + \phi_{11}(h_{1_{t_1}}-\mu_1) + \phi_{12}(h_{2_{t-1}}-\mu_2), \tau_1^2) \\
&\quad \times N(h_{2_{t+1}}|\mu_2 + \phi_{21}(h_{1t}-\mu_1) + \phi_{22}(h_{2t}-\mu_2), \tau_2^2) \times N_2(y_{1t}, y_{2t}|\boldsymbol{\nu_{y_t}}, \boldsymbol{\Sigma}_t) \\
&\propto \exp\Bigg\{ -\frac{1}{2}\Bigg\{ h_{1t}^2\left(\frac{1+\phi_{11}^2}{\tau_1^2} + \frac{\phi_{21}^2}{\tau_2^2}\right) - 2h_{1t}\Bigg[-\frac{1}{2} - (h_{2t}-\mu_2)\left(\frac{\phi_{11}\phi_{12}}{\tau_1^2} + \frac{\phi_{21}\phi_{22}}{\tau_2^2}\right) \\
&\quad + \frac{\phi_{11}(h_{1t+1}-\mu_1+h_{1t-1}-\mu_1)}{\tau_1^2} + \frac{\phi_{21}(h_{2t+1}-\mu_2)}{\tau_2^2} + \frac{\phi_{12}(h_{2t-1}-\mu_2)}{\tau_2^2} + \mu_1\left(\frac{1+\phi_{11}^2}{\tau_1^2} + \frac{\phi_{21}^2}{\tau_2^2}\right)\Bigg] \\
&\quad - \frac{\lambda_t}{1-\rho_\epsilon}\left(\frac{(y_{1t}-\nu_{y_{1t}})}{\exp(h_{1t})}\left((y_{1t}-\nu_{y_{1t}}) - 2\rho_\epsilon(y_{2t}-\nu_{y_{2t}})\exp\left(\frac{h_{1t}-h_{2t}}{2}\right)\right)\right)\Bigg\}\Bigg\}, \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad t = 2, \ldots, n
\end{aligned}
$$

- Full conditional density for h_{2t},

$$
\begin{aligned}
&f(h_{21}|y_{21}, y_{22}, h_{11}, h_{12}, h_{22}, \mu_1, \mu_2, \phi_{11}, \phi_{12}, \phi_{21}, \phi_{22}, \tau_1, \tau_2, \lambda_1, \rho_\epsilon, \boldsymbol{\nu_{y}}_1) \\
&\propto N\left(h_{21}|\mu_2, \frac{\tau_2^2}{1-\phi_{22}^2}\right) \times N(h_{22}|\mu_2 + \phi_{21}(h_{11}-\mu_1) + \phi_{22}(h_{21}-\mu_2), \tau_2^2) \\
&\quad \times N(h_{12}|\mu_1 + \phi_{11}(h_{11}-\mu_1) + \phi_{12}(h_{21}-\mu_2), \tau_1^2) \times N_2(y_{11}, y_{21}|\boldsymbol{\nu_{y}}_1, \boldsymbol{\Sigma}_t) \\
&\propto \exp\Bigg\{ -\frac{1}{2}\Bigg\{ h_{21}^2\left(\frac{1}{\tau_2^2} + \frac{\phi_{12}^2}{\tau_1^2}\right) - 2h_{21}\Bigg[-\frac{1}{2} - (h_{11}-\mu_1)\left(\frac{\phi_{22}\phi_{21}}{\tau_2^2} + \frac{\phi_{12}\phi_{11}}{\tau_1^2}\right) \\
&\quad + \frac{\phi_{22}(h_{22}-\mu_2)}{\tau_2^2} + \frac{\phi_{12}(h_{12}-\mu_1)}{\tau_1^2} + \mu_2\left(\frac{1}{\tau_2^2} + \frac{\phi_{12}^2}{\tau_1^2}\right)\Bigg] \\
&\quad - \frac{\lambda_1}{1-\rho_\epsilon}\left(\frac{(y_{21}-\nu_{y_{21}})}{\exp(h_{21})}\left((y_{21}-\nu_{y_{21}}) - 2\rho_\epsilon(y_{11}-\nu_{y_{11}})\exp\left(\frac{h_{21}-h_{11}}{2}\right)\right)\right)\Bigg\}\Bigg\}, \quad t=1
\end{aligned}
$$

$$
\begin{aligned}
&f(h_{2t}|y_{1t}, y_{2t}, h_{1t}, \mu_1, \mu_2, \phi_{11}, \phi_{12}, \phi_{21}, \phi_{22}, \tau_1, \tau_2, \lambda_t, \rho_\epsilon, \boldsymbol{\nu_{y}}_t) \\
&\propto N(h_{2_{t+1}}|\mu_2 + \phi_{21}(h_{1t}-\mu_1) + \phi_{22}(h_{2t}-\mu_2), \tau_2^2) \times N(h_{2t}|\mu_2 + \phi_{21}(h_{1_{t_1}}-\mu_1) + \phi_{22}(h_{2_{t-1}}-\mu_2), \tau_2^2) \\
&\quad \times N(h_{1_{t+1}}|\mu_1 + \phi_{11}(h_{1t}-\mu_1) + \phi_{12}(h_{2t}-\mu_2), \tau_1^2) \times N_2(y_{1t}, y_{2t}|\boldsymbol{\nu_{y}}_t, \boldsymbol{\Sigma}_t) \\
&\propto \exp\Bigg\{ -\frac{1}{2}\Bigg\{ h_{2t}^2\left(\frac{1+\phi_{22}^2}{\tau_2^2} + \frac{\phi_{12}^2}{\tau_1^2}\right) - 2h_{2t}\Bigg[-\frac{1}{2} - (h_{1t}-\mu_1)\left(\frac{\phi_{22}\phi_{21}}{\tau_2^2} + \frac{\phi_{12}\phi_{11}}{\tau_1^2}\right) \\
&\quad + \frac{\phi_{22}(h_{2t+1}-\mu_2+h_{2t-1}-\mu_2)}{\tau_2^2} + \frac{\phi_{12}(h_{1t+1}-\mu_1)}{\tau_1^2} + \frac{\phi_{21}(h_{1t-1}-\mu_1)}{\tau_1^2} + \mu_2\left(\frac{1+\phi_{22}^2}{\tau_2^2} + \frac{\phi_{12}^2}{\tau_1^2}\right)\Bigg] \\
&\quad - \frac{\lambda_t}{1-\rho_\epsilon}\left(\frac{(y_{2t}-\nu_{y_{2t}})}{\exp(h_{2t})}\left((y_{2t}-\nu_{y_{2t}}) - 2\rho_\epsilon(y_{1t}-\nu_{y_{1t}})\exp\left(\frac{h_{2t}-h_{1t}}{2}\right)\right)\right)\Bigg\}\Bigg\}, \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad t=2,\ldots,n
\end{aligned}
$$

- Full conditional density for μ_1,

$$
\begin{aligned}
&f(\mu_1|h_{1t}, h_{2t}, \mu_2, \phi_{11}, \phi_{12}, \phi_{21}, \phi_{22}, \tau_1, \tau_2) \\
&\propto N\left(h_{11}|\mu_1, \frac{\tau_1^2}{1-\phi_{11}^2}\right)\prod_{t=1}^{n-1} N(h_{1t+1}|\mu_1 + \phi_{11}(h_{1t}-\mu_1) + \phi_{12}(h_{2t}-\mu_2), \tau_1^2) \times N(\mu_1|a_{\mu 1}, b_{\mu 1}) \\
&\sim N\left(\frac{b_{\mu 1}(1-\phi_{11}^2)h_{11} + (1-\phi_{11})\prod_{t=1}^{n-1}[h_{1t+1} - \phi_{11}h_{1t} - \phi_{12}(h_{2t}-\mu_2)] + a_{\mu 1}\tau_1^2}{\gamma_{\mu 1}}, \frac{b_{\mu 1}\tau_1^2}{\gamma_{\mu 1}}\right)
\end{aligned}
$$

where

$$\gamma_{\mu 1} = b_{\mu 1}(1-\phi_{11}^2 + (n-1)(1-\phi_{11})^2) + \tau_1^2$$

- Full conditional density for μ_2,

$$
\begin{aligned}
&f(\mu_2|h_{1t}, h_{2t}, \mu_1, \phi_{11}, \phi_{12}, \phi_{21}, \phi_{22}, \tau_1, \tau_2) \\
&\propto N\left(h_{21}|\mu_1, \frac{\tau_2^2}{1-\phi_{22}^2}\right) \prod_{t=1}^{n-1} N(h_{2t+1}|\mu_2 + \phi_{21}(h_{1t} - \mu_1) + \phi_{22}(h_{2t} - \mu_2), \tau_2^2) \times N(\mu_2|a_{\mu 2}, b_{\mu 2}) \\
&\sim N\left(\frac{b_{\mu 2}(1-\phi_{22}^2)h_{21} + (1-\phi_{22})\prod_{t=1}^{n-1}[h_{2t+1} - \phi_{22}h_{2t} - \phi_{21}(h_{1t} - \mu_1)] + a_{\mu 2}\tau_2^2}{\gamma_{\mu 2}}, \frac{b_{\mu 2}\tau_2^2}{\gamma_{\mu 2}}\right)
\end{aligned}
$$

where

$$\gamma_{\mu 2} = b_{\mu 2}(1 - \phi_{22}^2 + (n-1)(1-\phi_{22})^2) + \tau_2^2$$

- Full conditional density for τ_1^2,

$$
\begin{aligned}
&f(\tau_1^2|h_{1t}, h_{2t}, \mu_1, \mu_2, \phi_{11}, \phi_{12}, \phi_{21}, \phi_{22}, \tau_2) \\
&\propto N\left(h_{11}|\mu_1, \frac{\tau_1^2}{1-\phi_{11}}\right) \prod_{t=1}^{n-1} N(h_{1t+1}|\mu_1 + \phi_{11}(h_{1t} - \mu_1) + \phi_{12}(h_{2t} - \mu_2), \tau_1^2) \times IG(\alpha_{\tau 1}, \beta_{\tau 1}) \\
&\sim IG\left(\alpha_{\tau 1} + \frac{n}{2}, \beta_{\tau 1} + \frac{(1-\phi_{11}^2)(h_{11} - \mu_1)^2 + \sum_{t=1}^{n-1}[h_{1t+1} - \mu_1 - \phi_{11}(h_{1t} - \mu_1) - \phi_{12}(h_{2t} - \mu_2)]^2}{2}\right)
\end{aligned}
$$

- Full conditional density for τ_2^2,

$$
\begin{aligned}
&f(\tau_2^2|h_{1t}, h_{2t}, \mu_1, \mu_2, \phi_{11}, \phi_{12}, \phi_{21}, \phi_{22}, \tau_1) \\
&\propto N\left(h_{21}|\mu_2, \frac{\tau_2^2}{1-\phi_{22}}\right) \prod_{t=1}^{n-1} N(h_{2t+1}|\mu_2 + \phi_{21}(h_{1t} - \mu_1) + \phi_{22}(h_{2t} - \mu_2), \tau_2^2) \times IG(\alpha_{\tau 2}, \beta_{\tau 2}) \\
&\sim IG\left(\alpha_{\tau 2} + \frac{n}{2}, \beta_{\tau 2} + \frac{(1-\phi_{22}^2)(h_{21} - \mu_2)^2 + \sum_{t=1}^{n-1}[h_{2t+1} - \mu_2 - \phi_{21}(h_{1t} - \mu_1) - \phi_{22}(h_{2t} - \mu_2)]^2}{2}\right)
\end{aligned}
$$

- Full conditional density for ϕ_{11},

$$
\begin{aligned}
&f(\phi_{11}|h_{1t}, h_{2t}, \mu_1, \mu_2, \phi_{12}, \phi_{21}, \phi_{22}, \tau_1) \\
&\propto N\left(h_{11}|\mu_1, \frac{\tau_1^2}{1-\phi_{11}^2}\right) \prod_{t=1}^{n-1} N(h_{1_{t+1}}|\mu_1 + \phi_{11}(h_{1t} - \mu_1) + \phi_{12}(h_{2t} - \mu_2), \tau_1^2) \times Be(a_{\phi_{11}}, b_{\phi_{11}}) \\
&\propto \sqrt{1-\phi_{11}^2} \left(\frac{\phi_{11}+1}{2}\right)^{a_{\phi_{11}}-1} \left(\frac{1-\phi_{11}}{2}\right)^{b_{\phi_{11}}-1} \\
&\quad \times \exp\left(-\frac{1}{2\tau_1^2}\left[\phi_{11}^2\left(-(h_{11}-\mu_1)^2 + \sum_{t=1}^{n-1}(h_{1t}-\mu_1)^2\right) - 2\phi_{11}\sum_{t=1}^{n-1}(h_{1t}-\mu_1)(h_{1t+1}-\mu_1-\phi_{12}(h_{2t}-\mu_2))\right]\right)
\end{aligned}
$$

- Full conditional density for ϕ_{22},

$$
\begin{aligned}
&f(\phi_{22}|h_{1t}, h_{2t}, \mu_1, \mu_2, \phi_{11}, \phi_{12}, \phi_{21}, \tau_2) \\
&\propto N\left(h_{21}|\mu_2, \frac{\tau_2^2}{1-\phi_{22}^2}\right) \prod_{t=1}^{n-1} N(h_{2_{t+1}}|\mu_2 + \phi_{21}(h_{1t} - \mu_1) + \phi_{22}(h_{2t} - \mu_2), \tau_2^2) \times Be(a_{\phi_{22}}, b_{\phi_{22}}) \\
&\propto \sqrt{1-\phi_{22}^2} \left(\frac{\phi_{22}+1}{2}\right)^{a_{\phi_{11}}-1} \left(\frac{1-\phi_{22}}{2}\right)^{b_{\phi_{22}}-1} \\
&\quad \times \exp\left(-\frac{1}{2\tau_2^2}\left[\phi_{22}^2\left(-(h_{21}-\mu_2)^2 + \sum_{t=1}^{n-1}(h_{2t}-\mu_2)^2\right) - 2\phi_{22}\sum_{t=1}^{n-1}(h_{2t}-\mu_2)(h_{2t+1}-\mu_2-\phi_{21}(h_{1t}-\mu_1))\right]\right)
\end{aligned}
$$

- Full conditional density for ϕ_{12},

$$
\begin{aligned}
&f(\phi_{12}|h_{1t}, h_{2t}, \mu_1, \mu_2, \phi_{11}, \phi_{21}, \phi_{22}, \tau_1) \\
&\propto \prod_{t=1}^{n-1} N(h_{1t+1}|\mu_1 + \phi_{11}(h_{1t} - \mu_1) + \phi_{12}(h_{2t} - \mu_2), \tau_1^2) \times N(\phi_{12}|a_{\phi_{12}}, b_{\phi_{12}}) \\
&\sim N\left(\frac{b_{\phi_{12}}\sum_{t=1}^{n-1}(h_{2t}-\mu_2)[h_{1t+1}-\mu_1-\phi_{11}(h_{1t}-\mu_1)] + a_{\phi_{12}}\tau_1^2}{\gamma_{\phi_{12}}}, \frac{\tau_1^2 b_{\phi_{12}}}{\gamma_{\phi_{12}}}\right)
\end{aligned}
$$

where

$$\gamma_{\phi_{12}} = b_{\phi_{12}}\sum_{t=1}^{n-1}(h_{2t}-\mu_2)^2 + \tau_1^2$$

- Full conditional density for ϕ_{21},

$$
\begin{aligned}
&f(\phi_{21}|h_{1t}, h_{2t}, \mu_1, \mu_2, \phi_{11}, \phi_{12}, \phi_{22}, \tau_2) \\
&\propto \prod_{t=1}^{n-1} N(h_{2t+1}|\mu_2 + \phi_{21}(h_{1t} - \mu_1) + \phi_{22}(h_{2t} - \mu_2), \tau_2^2) \times N(\phi_{21}|a_{\phi_{21}}, b_{\phi_{21}}) \\
&\sim N\left(\frac{b_{\phi_{21}} \sum_{t=1}^{n-1} (h_{1t} - \mu_1)[h_{2t+1} - \mu_2 - \phi_{22}(h_{2t} - \mu_2)] + a_{\phi_{21}}\tau_2^2}{\gamma_{\phi_{21}}}, \frac{\tau_2^2 b_{\phi_{21}}}{\gamma_{\phi_{12}}} \right)
\end{aligned}
$$

where

$$\gamma_{\phi_{21}} = b_{\phi_{21}} \sum_{t=1}^{n-1} (h_{1t} - \mu_1)^2 + \tau_2^2$$

- Full conditional density for ρ_ϵ,

$$
\begin{aligned}
&f(\rho_\epsilon|y_{1t}, y_{2t}, h_{1t}, h_{2t}, \lambda_t) \\
&\propto \prod_{t=1}^{n} N_2(y_{1t}, y_{2t}|\boldsymbol{\nu}_{\boldsymbol{y}_t}, \boldsymbol{\Sigma}_t) I(-1, 1) \\
&\propto (1 - \rho_\epsilon^2)^{-\frac{n}{2}} \exp\left\{ -\frac{\lambda_t}{2(1-\rho_\epsilon)} \sum_{t=1}^{n} \frac{(y_{1t} - \nu_{y_{1t}})^2}{\exp(h_{1t})} - \frac{2\rho_\epsilon (y_{1t} - \nu_{y_{1t}})(y_{2t} - \nu_{y_{2t}})}{\exp\left(\frac{h_{1t}+h_{2t}}{2}\right)} + \frac{(y_{2t} - \nu_{y_{2t}})^2}{\exp(h_{2t})} \right\} \\
&\quad \times I(-1, 1)
\end{aligned}
$$

- Full conditional density for λ_t,

$$
\begin{aligned}
&f(\lambda_t|y_{1t}, y_{2t}, h_{1t}, h_{2t}, \rho_\epsilon, \nu, \nu_{y_{1t}}, \nu_{y_{2t}}) \\
&\propto N_2(y_{1t}, y_{2t}|\boldsymbol{\nu}_{\boldsymbol{y}_t}, \boldsymbol{\Sigma}_t) Ga\left(\lambda_t|\frac{\nu}{2}, \frac{\nu}{2}\right) \\
&\sim Ga\left(\frac{\nu}{2} + 1, \frac{\nu}{2} + \frac{1}{2\sqrt{1-\rho_\epsilon^2}} \left(\frac{(y_{1t} - \nu_{y_{1t}})^2}{\exp(h_{1t})} - \frac{2\rho_\epsilon (y_{1t} - \nu_{y_{1t}})(y_{2t} - \nu_{y_{2t}})}{\exp\left(\frac{h_{1t}+h_{2t}}{2}\right)} + \frac{(y_{2t} - \nu_{y_{2t}})^2}{\exp(h_{2t})} \right) \right)
\end{aligned}
$$

- Full conditional density for μ_{y1},

$$
\begin{aligned}
&f(\mu_{y1}|y_{1t}, y_{2t}, h_{1t}, h_{2t}, \rho_\epsilon, \lambda_t, \mu_{y2}, \xi_1, \xi_2) \\
&\propto \prod_{t=1}^{n} N_2(y_{1t}, y_{2t}|\boldsymbol{\nu}_{\boldsymbol{y}_t}, \boldsymbol{\Sigma}_t) N(\mu_{y1}|a_{\mu_{y1}}, b_{\mu_{y1}}) \\
&\sim N\left(\left(\frac{1}{\sqrt{1-\rho_\epsilon^2}} \sum_{t=1}^{n} \lambda_t \left(\frac{y_{1t} - \xi_1 W_{1t}}{\exp(h_{1t})} - \frac{\rho_\epsilon (y_{2t} - \mu_{y2} - \xi_2 W_{2t})}{\exp\left(\frac{h_{1t}+h_{2t}}{2}\right)} \right) + \frac{a_{\mu_{y1}}}{b_{\mu_{y1}}} \right) \gamma_{\mu_{y1}}^{-1}, \gamma_{\mu_{y1}}^{-1} \right)
\end{aligned}
$$

where

$$\gamma_{\mu_{y1}} = \frac{1}{\sqrt{1-\rho_\epsilon^2}} \sum_{t=1}^{n} \frac{\lambda_t}{\exp(h_{1t})} + \frac{1}{b_{\mu_{y1}}}$$

- Full conditional density for μ_{y2},

$$\begin{aligned} &f(\mu_{y2}|y_{1t}, y_{2t}, h_{1t}, h_{2t}, \rho_\epsilon, \lambda_t, \mu_{y1}, \xi_1, \xi_2) \\ &\propto \prod_{t=1}^{n} N_2(y_{1t}, y_{2t}|\boldsymbol{\nu_{y_t}}, \boldsymbol{\Sigma}_t) N(\mu_{y2}|a_{\mu_{y2}}, b_{\mu_{y2}}) \\ &\sim N\left(\left(\frac{1}{\sqrt{1-\rho_\epsilon^2}} \sum_{t=1}^{n} \lambda_t \left(\frac{y_{2t} - \xi_2 W_{2t}}{\exp(h_{2t})} - \frac{\rho_\epsilon (y_{1t} - \mu_{y1} - \xi_1 W_{1t})}{\exp\left(\frac{h_{1t}+h_{2t}}{2}\right)}\right) + \frac{a_{\mu_{y2}}}{b_{\mu_{y2}}}\right) \gamma_{\mu_{y2}}^{-1}, \gamma_{\mu_{y2}}^{-1}\right) \end{aligned}$$

where

$$\gamma_{\mu_{y2}} = \frac{1}{\sqrt{1-\rho_\epsilon^2}} \sum_{t=1}^{n} \frac{\lambda_t}{\exp(h_{2t})} + \frac{1}{b_{\mu_{y2}}}$$

- Full conditional density for ξ_1,

$$\begin{aligned} &f(\xi_1|y_{1t}, y_{2t}, h_{1t}, h_{2t}, \rho_\epsilon, \lambda_t, \mu_{y1}, \mu_{y2}, \xi_2) \\ &\propto \prod_{t=1}^{n} N_2(y_{1t}, y_{2t}|\boldsymbol{\nu_{y_t}}, \boldsymbol{\Sigma}_t) N(\xi_1|a_{\xi_1}, b_{\xi_1}) \\ &\sim N\left(\frac{1}{\sqrt{1-\rho_\epsilon^2}} \sum_{t=1}^{n} \lambda_t \left[\frac{W_{1t}(y_{1t} - \nu_{y_{1t}})}{\exp(h_{1t})} - \frac{\rho W_{1t}(y_{2t} - \mu_{y2} - \xi_2 W_{2t})}{\exp\left(\frac{h_{1t}+h_{2t}}{2}\right)}\right] + \frac{a_{\xi_1}}{b_{\xi_1}} \gamma_{\xi_1}^{-1}, \gamma_{\xi_1}^{-1}\right) \end{aligned}$$

where

$$\gamma_{\xi_1} = \frac{1}{\sqrt{1-\rho_\epsilon^2}} \sum_{t=1}^{n} \frac{\lambda_t}{\exp(h_{1t})} W_{1t}^2 + \frac{1}{b_{\xi_1}}$$

- Full conditional density for ξ_2,

$$\begin{aligned} &f(\xi_2|y_{1t}, y_{2t}, h_{1t}, h_{2t}, \rho_\epsilon, \lambda_t, \mu_{y1}, \mu_{y2}, \xi_1) \\ &\propto \prod_{t=1}^{n} N_2(y_{1t}, y_{2t}|\boldsymbol{\nu_{y_t}}, \boldsymbol{\Sigma}_t) N(\xi_2|a_{\xi_2}, b_{\xi_2}) \\ &\sim N\left(\frac{1}{\sqrt{1-\rho_\epsilon^2}} \sum_{t=1}^{n} \lambda_t \left[\frac{W_{2t}(y_{2t} - \nu_{y_{2t}})}{\exp(h_{2t})} - \frac{\rho W_{2t}(y_{1t} - \mu_{y1} - \xi_1 W_{1t})}{\exp\left(\frac{h_{1t}+h_{2t}}{2}\right)}\right] + \frac{a_{\xi_2}}{b_{\xi_2}} \gamma_{\xi_2}^{-1}, \gamma_{\xi_2}^{-1}\right) \end{aligned}$$

where

$$\gamma_{\xi_2} = \frac{1}{\sqrt{1-\rho_\epsilon^2}} \sum_{t=1}^{n} \frac{\lambda_t}{\exp(h_{2t})} W_{2t}^2 + \frac{1}{b_{\xi_2}}$$

- Full conditional density for ν,

$$f(\nu|\lambda_t) \propto \prod_{t=1}^{n} Ga\left(\lambda_t|\frac{\nu}{2},\frac{\nu}{2}\right) \times Ga(\nu|a_\nu,b_\nu) \propto \frac{\nu^{a_\nu-1+\frac{n\nu}{2}}}{2^{\frac{n\nu}{2}}(\Gamma(\frac{\nu}{2}))^n} \exp\left(-\frac{\nu}{2}\sum_{t=1}^{n}\lambda_t - b_\nu\nu\right)\prod_{t=1}^{n}\lambda_t^{\frac{\nu}{2}}$$

The above full conditional distributions are mostly standard. These include normal distributions for μ_1, μ_2, ϕ_{12}, ϕ_{21}, μ_{y1}, μ_{y2}, ξ_1 and ξ_2; inverse gamma distribution for τ_1^2 and τ_2^2 and gamma distribution for λ_t. The full conditional densities for h_{1t}, h_{2t}, ϕ_{11}, ϕ_{22}, ρ_ϵ and ν are nonstandard, hence some sampling methods such as adaptive rejection sampling for log-concave distribution will be used to sample from these nonstandard distributions.

We run three parallel Markov chain for 80,000 iterations. The initial 20,000 iterations are discarded as the burn-in period. We take simulated values every 20th iterations after the burn-in period to avoid high-autocorrelations. A final sample of size 3,000 is used for posterior inference. Trace plots and autocorrelation plots of our MCMC iterates are examined carefully, to ensure convergence is achieved in each case.

5.5.2. Technical issues with `WinBUGS`. For some distributions such as normal and Student-t distributions, the precision parameter, rather than the variance parameter, needs to be specified in `WinBUGS`. Hence in the multivariate case, we need to specify the precision matrix, which is the inverse of the covariance matrix. The covariance matrix of (y_{1t}, y_{2t}) is given by

$$\Sigma_{y,t} = \begin{pmatrix} \frac{1}{\lambda_t}\exp(h_{1t}) & \frac{\rho_\epsilon}{\lambda_t}\exp\left(\frac{h_{1t}+h_{2t}}{2}\right) \\ \frac{\rho_\epsilon}{\lambda_t}\exp\left(\frac{h_{1t}+h_{2t}}{2}\right) & \frac{1}{\lambda_t}\exp(h_{2t}) \end{pmatrix}$$

And its inverse is given by

$$\begin{aligned}\Sigma_{y,t}^{-1} &= \frac{\lambda_t^2}{\exp(h_{1t}+h_{2t})(1-\rho_\epsilon^2)} \begin{pmatrix} \frac{1}{\lambda_t}\exp(h_{2t}) & -\frac{\rho_\epsilon}{\lambda_t}\exp\left(\frac{h_{1t}+h_{2t}}{2}\right) \\ -\frac{\rho_\epsilon}{\lambda_t}\exp\left(\frac{h_{1t}+h_{2t}}{2}\right) & \frac{1}{\lambda_t}\exp(h_{1t}) \end{pmatrix} \\ &= \frac{\lambda_t}{\exp(h_{1t}+h_{2t})(1-\rho_\epsilon^2)} \begin{pmatrix} \exp(h_{2t}) & -\rho_\epsilon\exp\left(\frac{h_{1t}+h_{2t}}{2}\right) \\ -\rho_\epsilon\exp\left(\frac{h_{1t}+h_{2t}}{2}\right) & \exp(h_{1t}) \end{pmatrix}\end{aligned}$$

Therefore in `WinBUGS`, the distribution of the data (y_{1t}, y_{2t}) is specified as

```
for (i in 1:N) {
```

```
det[i]<-exp(h[i,1]+h[i,2])*(1-rhoep*rhoep)/lam[i]
yisigma2[i,1,1]<-exp(h[i,2])/det[i]
yisigma2[i,2,2]<-exp(h[i,1])/det[i]
yisigma2[i,1,2]<--rhoep*exp(0.5*h[i,1]+0.5*h[i,2])/det[i]
yisigma2[i,2,1]<-yisigma2[i,1,2]
lam[i]~dgamma(nu/2,nu/2)
y[i,1:2]~dmnorm(muy[i,],yisigma2[i,,])
muy[i,1]<-mu_y1+xi1*W1[i]
muy[i,2]<-mu_y2+xi2*W2[i]
```

Secondly, when implementing the ST_2 distribution via its mixing form, direct specification for the half normal variable V using the absolute value of standard normal distribution definition incurs some syntax error in WinBUGS. Hence, we use the generalised gamma distribution defined in (3.5) with parameter values $a = 1/2$, $b = 1/\sqrt{2}$ and $c = 2$ which reduces to the density of a half normal random variable.

Lastly, we change the default sampling method for log-concave parameters from adaptive rejection sampling to slice sampling, in order to avoid possible traps in WinBUGS. The trap messages referring to 'DFreeARS' indicate numerical problems with the derivative-free adaptive rejection algorithm used for log-concave distributions. In WinBUGS (1.4.3), it is possible to change the sampling methods. For our models, we encountered problems with WinBUGS's adaptive rejection sampler (DFreeARS), and hence we replace the method 'UpdaterDFreeARS' for 'log-concave' by 'UpdaterSlice'. This is suggested in the WinBUGS website

http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml#problems

5.6. Empirical data

The data employed in this study consist of daily spot electricity prices over the 3-year period from January 1, 2007 to December 31, 2009 for each of the four main regional markets of NSW, QLD, SA and VIC. The data were obtained from the AEMO website, originally as half-hourly data. The prices are in Australian dollars per megawatt hour ($/MW h). The natural logarithms of the daily averages are used to construct time series of observations. We take the log transformation of the prices in order to reduce the extreme spike effect. Higgs

and Worthington (2005), Worthington *et al.* (2005) and Higgs (2009) also used daily spot prices in their analyses of electricity markets. Although using daily prices may lead to the loss of some information impounded in the more frequent trading interval data, it is nonetheless important for example, in financial contracts. A summary of descriptive statistics for the four electricity markets is presented in Table 5.1. Sample mean, median, standard deviation, skewness, kurtosis, Jarque-Bera (J-B) statistic, J-B p-values and Augmented DickeyŰFuller (ADF) test and p-values are reported. The average log daily spot electricity prices range from \$3.5877/MW h (QLD) to \$3.6698/MW h (SA). In particular, we are interested in analysing the pairs: NSW and VIC, NSW and QLD and SA and VIC with 1095, 1096 and 1094 observations respectively. These pairs are chosen since it was shown in Higgs (2009) that these pairs have the strongest correlations. All four markets exhibit positive skewness and kurtosis larger than three, indicating that a leptokurtic distribution is more appropriate than the normal distribution in modelling the prices. The J-B statistic is calculated for the test of the joint hypothesis of no skewness and zero excess kurtosis. The p-values are all significant at 1% level of significance, rejecting the null hypothesis for all markets. The p-values of the ADF test are less than 0.01. Hence the ADF statistics reject the null hypothesis of non-stationarity or unit root at the 0.01 level of significance. This indicates that the spot electricity price series in the four markets are stationary.

Figure 5.3 shows the time series plots and histograms of the log daily spot price for each of the four electricity markets. The high prices in the middle of May, 2009 for NSW, QLD and VIC are likely due to the record high mean temperatures set over eastern Australia. The temperature anomalies for May were accompanied by a major heatwave in tropical Australia, affecting southern QLD and northern NSW. In March 2008, another heatwave struck Adelaide which was the longest heatwave recorded for any Australian capital city. This caused prominent price spikes in SA in the early 2008. The price spikes in NSW, VIC and SA in the early 2009 were due to the Southern Australian heatwave and is considered as one of the most extreme weather event in the region. Ten month later, a second heatwave, known as the late 2009 Southern Australian heatwave, struck the same region, causing significant temperature fluctuations. During heatwaves, electricity usage was significantly increased due to increased demand from air conditioning. The electricity prices in adjacent regions are also

TABLE 5.1. Summary statistics of the logarithms of daily spot prices for four Australian electricity markets.

	NSW	VIC	QLD	SA
Mean	3.6257	3.6230	3.5877	3.6698
Median	3.5063	3.5636	3.4556	3.5718
Standard deviation	0.5603	0.5166	0.5937	0.6337
Skewness	2.4398	1.7907	1.7121	2.7856
Kurtosis	9.1843	8.0413	8.4087	12.5870
J-B statistic	2831	1745	1871	5604
J-B p-value	0.0000	0.0000	0.0000	0.0000
ADF statistics	-5.2641	-6.2464	-6.4645	-7.0954
ADF p-value	<0.01	<0.01	<0.01	<0.01

affected because of the violation of constraints on the interconnector flows, with the severe weather conditions affecting the power system.

5.6.1. Prior distributions. For the six models to be fitted, we classify model parameters into three categories: parameters in the mean equations, these include μ_{y1}, μ_{y2}, ξ_1, ξ_2, ρ_ϵ and ν; parameters in the volatility equations, these include μ_1, μ_2, ϕ_{11}, ϕ_{12}, ϕ_{21}, ϕ_{22}, ρ_η, τ_1^2 and τ_2^2 and the parameters in the dynamic correlation models ψ_0, ψ_1 and τ_ρ. We assign the following prior distributions:

$$
\begin{aligned}
\mu_{y1} &\sim N(0,100), & \phi_{12} &\sim N(0,20),\\
\mu_{y2} &\sim N(0,100), & \phi_{21} &\sim N(0,20),\\
\xi_1 &\sim N(0,100), & \phi_{22}^* &\sim Be(20,1.5), \text{ where } \phi_{22}^* = \tfrac{\phi_{22}+1}{2},\\
\xi_2 &\sim N(0,100), & \rho_\eta &\sim U(-1,1),\\
\rho_\epsilon &\sim U(-1,1), & \tau_1^2 &\sim IG(2.5,0.025),\\
\nu &\sim Ga(0.001,0.001), & \tau_2^2 &\sim IG(2.5,0.025),\\
\mu_1 &\sim N(0.3,100), & \psi_0 &\sim N(0,100),\\
\mu_2 &\sim N(0.3,100), & \psi_1^* &\sim Be(20,1.5), \text{ where } \psi_1^* = \tfrac{\psi_1+1}{2},\\
\phi_{11}^* &\sim Be(20,1.5), \text{ where } \phi_{11}^* = \tfrac{\phi_{11}+1}{2}, & \tau_\rho &\sim IG(2.5,0.025).
\end{aligned}
$$

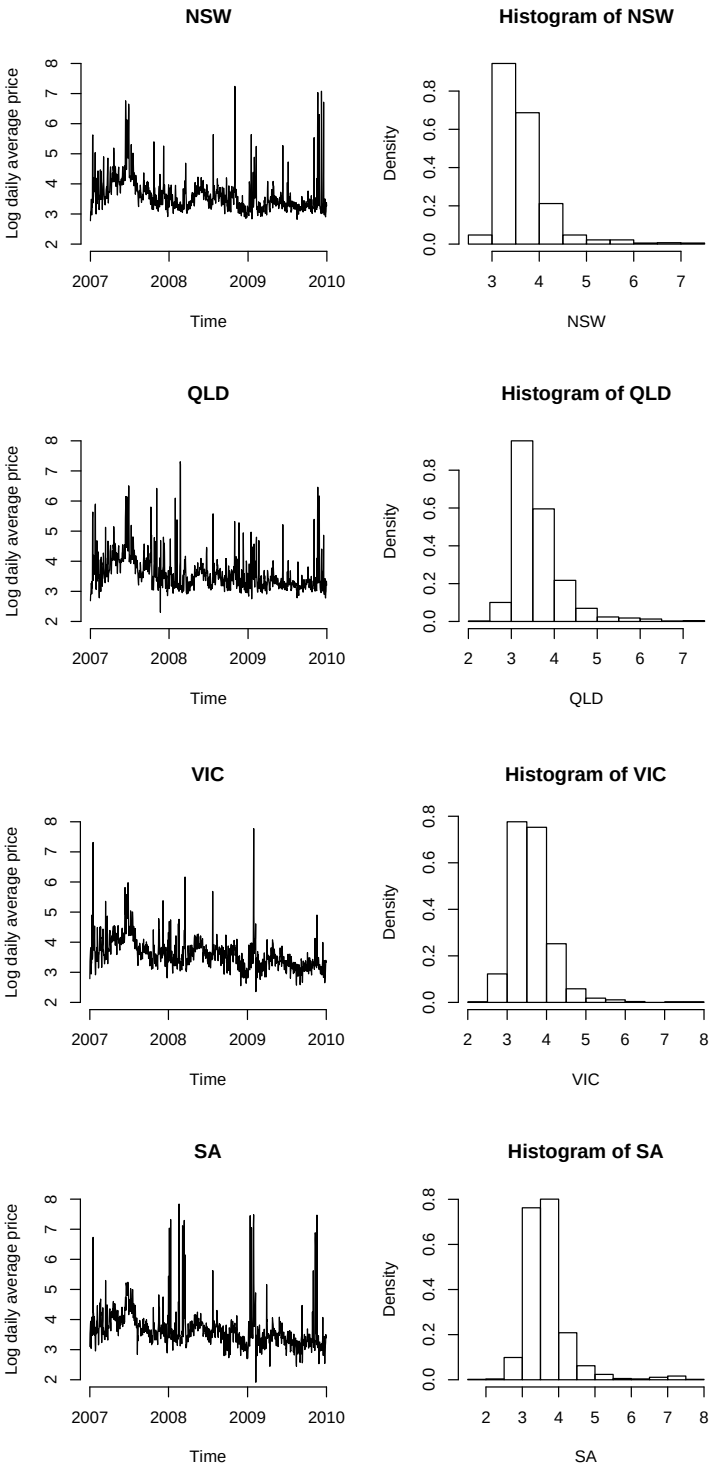


FIGURE 5.3. Time series plots and histograms of log daily average spot price for NSW, VIC, QLD and SA

The hyper-parameter values mainly follow those used by Yu and Meyer (2006). Most of them convey ignorance about the parameter values.

5.7. Empirical results

We fit the cc-t-G, cc-t-bN, dc-t-G, cc-ST_1-G, cc-ST_2-G and dc-ST_2-G models to the NSW/VIC, NSW/QLD and SA/VIC pairs from the Australian electricity markets. The history plots and Brooks, Gelman and Rubin (bgr) convergence plots (described in Chapter 4) of the parameters in the cc-t-G model are presented in Figure 5.4 and 5.5 for the NSW/VIC data for illustrative purposes. Note that although we use 20,000 iterations as burn-in, the history plots and bgr the convergence plots are shown from 10,000 iterations (except for the history plots of the volatilities which are shown from 20,000, due to insufficient RAM in `WinBUGS`) to illustrate that taking 10,000 iterations is not sufficient enough for the burn-in. From the history plots of the model parameters, we observe that the three chains initiated from different starting values are intertwined with each other, providing support for convergence. The bgr plots show that the Gelman-Rubin statistic converge to 1 and both pooled and within-interval widths stablise to the same value for all parameters. Hence we can safely conclude that all parameters have converged in the model. Convergence of parameters in other models for other two dataset is also checked, although the plots are omitted.

Table 5.2 shows the posterior means and standard errors of the parameters in the 6 models for the NSW/VIC pair. Significant estimates of parameters ϕ_{12}, ϕ_{21}, ξ_1, ξ_2, γ_1 and γ_2 are shown in bold. The constant in the conditional mean equation for NSW (μ_{y1}) ranges from 3.195 (cc-ST_1-G model) to 3.549 (dc-t-G model) and μ_{y2} for VIC ranges from 3.235 (cc-ST_1-G model) to 3.594 (cc-t-G model). The estimates of μ_{y1} and μ_{y2} are smaller under the skew Student-t models than the models with symmetric Student-t distribution because the skew distributions capture the right-skewness due to large observations better than the symmetric distribution. The mean estimates in symmetric distribution are usually inflated by a few large observations. The correlation between the NSW and VIC is very high, which is shown by the large estimates of the correlation coefficient ρ_ϵ. In particular, the cc-ST_1-G and cc-ST_2-G models give ρ_ϵ estimates close to 1.

The coefficients ξ_1 and ξ_2 for the indicator variables W_{1t} and W_{2t} indicating the weekend and holiday effect, are negative and significant in all 6 models, showing that the electricity prices on weekends and public holidays are lower than the weekday prices, as expected. In

particular, this effect is stronger in VIC than NSW, as the absolute magnitude of ξ_2 is larger than that of ξ_1, for all models. The estimates of the skewness parameters, γ_1 and γ_2, are all positive and significant for the two skew distributions considered, indicating that the NSW and VIC markets are significantly positively skewed. This shows that for these markets, there is greater chance of obtaining higher prices than lower prices. Higgs and Worthington (2005) employed a different form of the skew Student-t distribution for GARCH type model in their analysis of the Australian electricity market. They also found significant skewness in each of NSW, QLD, VIC and SA regional markets, when analysed individually.

The persistence ϕ_{11} and ϕ_{22} of the log-volatilities for the two individual markets are reasonably high, ranging from 0.9385 (cc-ST_2-G) to 0.9659 (cc-t-bN) for NSW and 0.9228 (cc-ST_1-G) to 0.9655 (cc-t-bN) for VIC. More importantly, the parameters ϕ_{12} and ϕ_{21} are significant across all models as the 95% credible intervals do not include zero, indicating that the volatility of the NSW price series Granger causes the volatility of VIC price and vice versa. The strong correlation between the volatilities is further confirmed by the significant ρ_η in the cc-t-bN model where the log-volatilities are modelled by a bivariate normal distribution. The marginal means of the log-volatilities, μ_{h1} and μ_{h2}, for NSW and VIC respectively, are around -2 for the models with symmetric Student-t error distribution and the estimates are smaller when skew Student-t distributions are employed to capture the right-skewness for the prices. This is illustrated in Figure 5.6 which displays the plots of log-volatilities for NSW and VIC across the 6 models. It shows that the estimated log-volatilities fluctuate around -2 to -3; models with symmetric Student-t distribution (cc-t-G, cc-t-bN and dc-t-G) have mean volatility level around -2 whereas models with skew Student-t distributions (cc-ST_1-G, cc-ST_2-G, dc-ST_2-G) have lower log-volatility estimates which are around -3. A similar pattern can be observed in the volatility plot of the VIC market: models with skew error distributions tend to produce lower estimated volatilities. An explanation for this phenomenon is that the models which employ skew Student-t error distributions effectively capture the data in the extreme tail areas and hence the variability of the data can be accommodated to a greater extent by these models, as compared to the models with symmetric Student-t distributions. As a result, the contribution coming from the latent volatilities is less which leads to smaller estimated values of the volatilities. Figure 5.8 depicts the comparison between the log-volatilities of the NSW and VIC markets plotted separately for each of

the 6 models. In all models, the volatilities of the two markets are highly correlated, which can be explained by the significant Granger causality coefficients ϕ_{12}, ϕ_{21} and the significant correlation coefficient in the cc-t-bN model.

Lastly, when performing model comparison based on the DIC, the proposed dc-ST_2-G model ranks the highest among the six models. The cc-t-G model has a smaller DIC than the cc-t-bN model, indicating that bilateral Granger causality in this case is a more appropriate way in capturing the dependence between volatilities of the two markets than a bivariate normal distribution. This is likely due to fact that Granger causality parameters give indication of the direction of causality whereas this cannot be obtained by a single correlation parameter in a bivariate distribution. On the contrary, Yu and Meyer (2006) found that the bivariate normal distribution for the volatilities performs empirically better than the Granger causality model when the models are applied to weekly log-returns of Australian/US dollar and New Zealand/US dollar exchange rates. However, it is worth noting that only one-way Granger causality was employed in their study. Among the models with symmetric Student-t error distribution, model dc-t-G ranks the highest, since it captures the time-varying dynamic of the correlation between the prices of two markets.

The DICs for models with skew Student-t error distributions are smaller than the models with symmetric Student-t distribution, demonstrating that models with skew distribution allow for greater flexibility in describing the characteristics of the data. Moreover, the cc-ST_2-G model performs better than the cc-ST_1-G model which can be explained by the tail dependence properties of the two skew Student-t distributions. The tail dependence property of a distribution gives idea on the likelihood for the distribution to generate extreme events jointly. Fung and Seneta (2010b) proved that depending on the signs of the skewness parameters, the ST_1 distribution possesses either zero or one asymptotic tail dependence (Theorem 5.8.1. in Appendix 5.3). The ST_2 distribution on the other hand, possesses different degrees of asymptotic tail dependence that is nontrivial for all values of the skewness parameters (Theorem 5.8.2. in Appendix 5.3). Hence from the modelling perspective, the ST_2 distribution is a more appropriate extension of the symmetric Student-t distribution. Therefore, the dc-ST_2-G model which incorporates the time-varying correlation between the price series, the more appropriate Granger causality for the volatilities and the empirically better ST_2 error distribution, produces the best results across all the models.

Table 5.3 reports the parameter estimates for the NSW/QLD pair of markets. These results share many similarities with the results of the NSW/VIC pair. For the NSW/QLD pair, μ_{y2} estimates are smaller than μ_{y1}, showing that the average log daily price of QLD is slightly lower than the price of NSW. As with the NSW/VIC pair, estimated log-volatilities and the marginal means of log-volatilities under the models with skew Student-t error distributions are lower than the models using the symmetric Student-t distribution. The persistence ϕ_{22} of the volatility of QLD ranges from 0.8587 (cc-ST_1-G) to 0.9757 (dc-ST_2-G). In the cc-t-G and dc-t-G models, all ϕ_{12} and ϕ_{21} are positive and significant, indicating there exists a bilateral Granger causality between the volatilities of the NSW and QLD price series. However, under the cc-ST_1-G, cc-ST_2-G and dc-ST_2-G models, these parameters are not significant (except for ϕ_{12} in cc-ST_1-G model). The correlation coefficient ρ_η between the volatilities in the cc-t-bN model is positive and significant, showing that there is a positive dependence between the volatilities. The correlation between the NSW and QLD price series ρ_ε ranges from 0.9633 (cc-ST_2-G model) to 0.9846 (cc-t-bN model). Comparing to the NSW/VIC pair, ρ_ε for NSW/QLD is higher under models with symmetric Student-t distribution and lower under the models with skew Student-t error distributions. The weekend and holiday effect coefficient ξ_2 for QLD is negative and significant under all models, but the magnitude is smaller than those of NSW and VIC. The skewness γ_2 for QLD is positive and significant under all models with skew distributions, indicating that for the QLD market, the price is also positively skewed.

Comparing the DIC of the 6 models, once again the models with skew Student-t distributions perform better than the models with symmetric distribution, except for the cc-ST_1-G model, which performs worse than the dc-t-G model. For the same reason explained for the NSW/VIC pair, the cc-t-G mode has a smaller DIC than the cc-t-bN model. The dc-t-G model which allows for dynamic correlation performs better than both cc-t-G and cc-t-bN models. Yet again, dc-ST_2-G model performs the best among all the models considered.

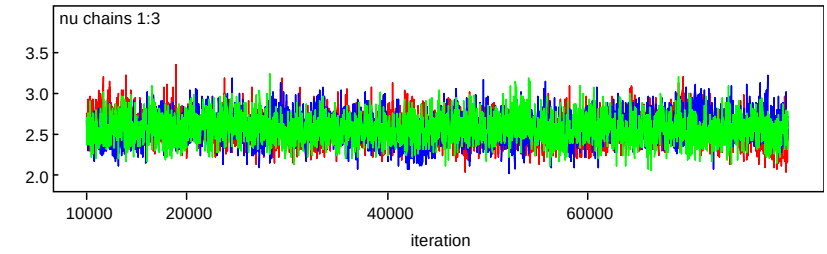
The log-volatilities of QLD under the six models is shown in Figure 5.7. In this plot, the difference between volatilities estimated under symmetric Student-t models and skew Student-t models is more drastic. This is particularly the case for the cc-ST_2-G and dc-ST_2-G models which employ the ST_2 as the error distribution, as the estimated volatilities are consistently lower under these two models. Figure 5.9 shows the comparison between

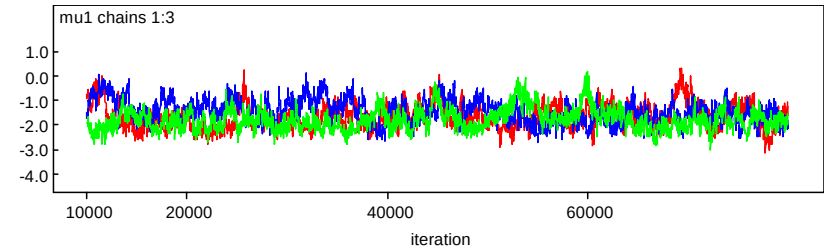
the estimated volatilities of NSW and QLD under the six models. Overall, the volatilities of the regional markets show close alignment, indicating strong correlation between them. However, for the cc-ST_2-G and dc-ST_2-G models, we observe some discrepancies between the volatilities, which may be explained by the insignificant estimates of ϕ_{12} and ϕ_{21} in these two models.

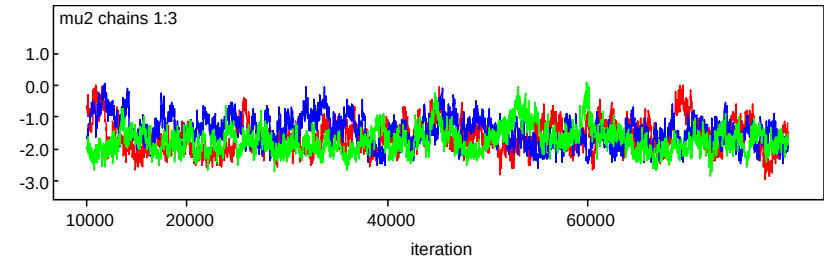
Looking at the results for SA/VIC in Table 5.4, we observe the same pattern that the models with skew Student-t error distributions tend to produce smaller estimates of μ_{y1} and μ_{y2} than the symmetric Student-t models. The average log daily price for SA is higher than NSW, QLD and VIC in most models. This agrees with the summary statistics, that the SA market has the highest mean and median daily electricity price. The persistence ϕ_{11} of the volatility of SA ranges from 0.8555 (cc-t-G model) to 0.9530 (cc-t-bN model), generally smaller than those of the other 3 markets. The estimates of the Granger causality parameter ϕ_{12} are significant for all models, showing that the volatility of VIC Granger causes the volatility of SA. The parameter ϕ_{21} is significant in all models, except for the cc-t-G and cc-ST_2-G models, where the volatility of SA does not Granger cause the volatility of VIC. The correlation between the volatilities of SA and VIC is estimated to be 0.9393, which is larger than both the correlation between NSW/VIC and NSW/QLD. The correlation between the price series ρ_ϵ is also relatively high, ranges from 0.9740 (cc-ST_1-G model) to 0.9919 (cc-ST_2-G model). The weekend and public holiday effect ξ_1 for SA is negative and significant under all models, and the effect is stronger than both NSW and QLD, but weaker than VIC. The estimated log-volatilities for SA is depicted in Figure 5.7. Comparing the log-volatilities estimated under different models for SA with those for other states, the difference in the log-volatilities is smaller across different models, especially after year 2009, where the log-volatilities mostly overlap each other. From 2007 to 2009, the difference in the log-volatilities is larger, where we observe smaller estimates under the cc-ST_2-G and dc-ST_2-G models relative to other models. A comparison of estimated log-volatilities of SA and VIC under the six models is depicted in Figure 5.10, where the strong correlation between the volatilities can be easily observed. Note that for these three data sets, we only compare the estimated log-volatilities across different models, rather than the actual variance (or the standard deviation) of the returns. This is because the variance of the skew Student-t type 1 is only finite when $\nu > 4$ and for the SA/VIC data, the estimated ν is 2.476. Hence the

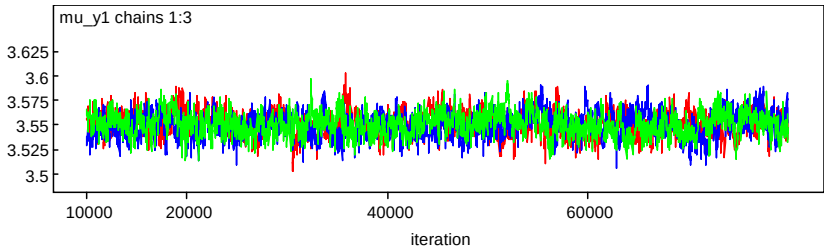
variance of the returns in this case in infinite. To be consistent across different models and data sets, we only compare the log-volatilities h_t.

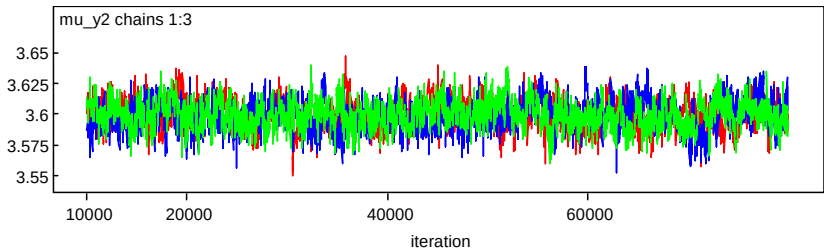
Figure 5.11 displays the dynamic correlations $\rho_{\epsilon t}$ from the dc-t-G and dc-ST_2-G models for the three pairs of markets. These three markets are well-interconnected markets in the NEM: NSW and VIC is linked by a regulated interconnector, NSW and QLD are linked by the Terranora interconnector and the QNI, SA and VIC are linked by the Murrylink and the Heywood interconnectors. Overall there is a mean-reverting pattern in each correlation series. The high correlations close to 1 dispersed throughout the sampling period is mostly a result of concurrent jumps occurring in both regions. In the middle of 2007, price jumps occur in both NSW and VIC and as a result, we estimate very high correlation which is close 1 around this time. In March 2008, an exceptional and prolonged heatwave affected most of southern Australia, including SA and part of VIC. These areas, especially in SA, recorded consecutive extremely high temperatures and as a result, the demand of electricity in these areas was greatly increased. However, the temperature was not as high in NSW and hence the NSW demand was not affected as much by the heat wave. This explains the significant drop of correlation between NSW and VIC and between SA and VIC in the early 2008. All regions in the NEW experienced significant price fluctuations in the beginning of 2009, due to the 'early 2009 southeastern Australia heatwave', striking mainly SA and VIC. During this period, the transmission system of VIC was greatly affected by bushfires and extremely high temperatures were recorded in Melbourne (VIC), Adelaide (SA) and Bankstown (NSW). QLD on the other hand, was much less affected by the heatwave, hence explained the negative correlation between NSW and QLD in the early 2009. The drop of correlation between all three pairs in the end of 2009 is most likely due to the 'late 2009 southeastern Australia heat wave', which was a second struck ten month after the early 2009 heatwave. Comparing the correlations estimated from the dc-t-G and the dc-ST_2-G models across different pairs, it is revealed that correlations under the dc-ST_2-G model are lower than those under the dc-t-G model. This is the result of using skew error distribution in the dc-ST_2-G mode which effectively filters out the extreme observations in each of the bivariate observed price series. In the case of low correlation between two regional markets, for example, when only one of the two markets encounters price jumps, the dc-ST_2-G model is more capable of detecting negative correlation between the prices of two markets as compared to the dc-t-G model.

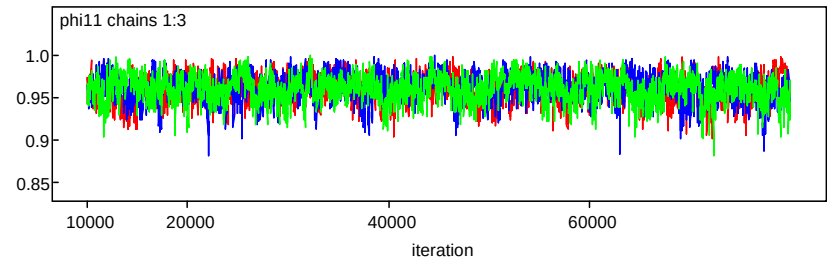

nu chains 1:3
3.5
3.0
2.5
2.0
10000
20000
40000
60000
iteration

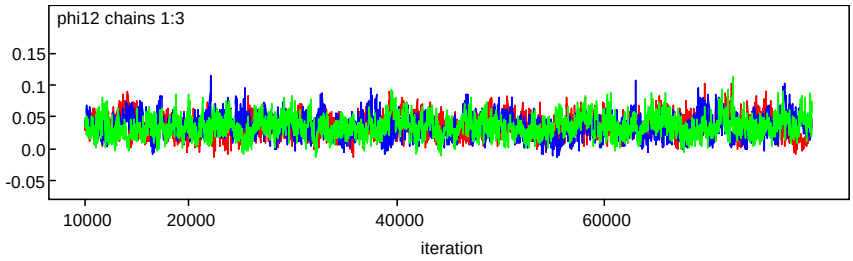

mu1 chains 1:3
1.0
0.0
-1.0
-2.0
-3.0
-4.0
10000
20000
40000
60000
iteration

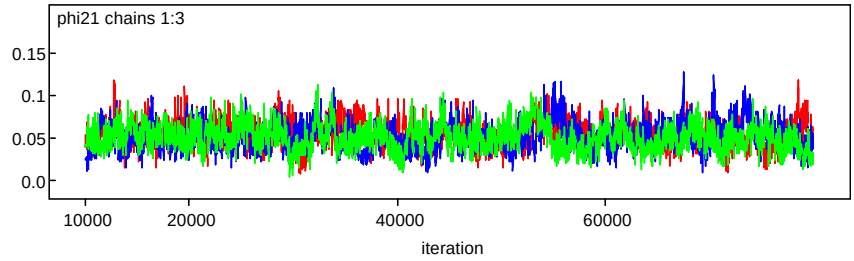

mu2 chains 1:3
1.0
0.0
-1.0
-2.0
-3.0
10000
20000
40000
60000
iteration

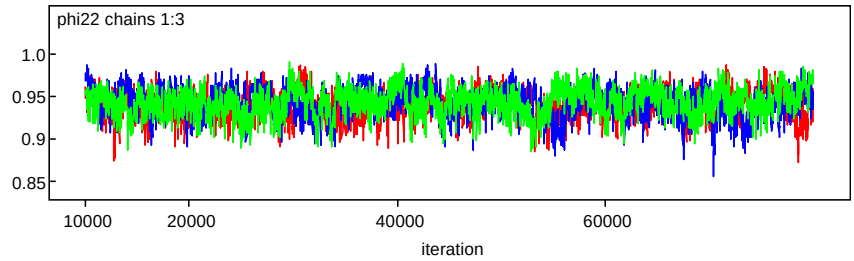

mu_y1 chains 1:3
3.625
3.6
3.575
3.55
3.525
3.5
10000
20000
40000
60000
iteration

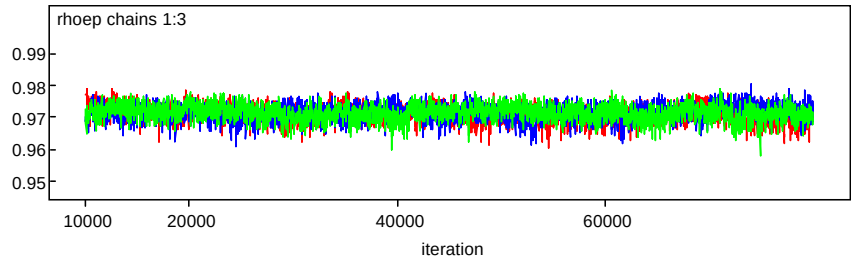

mu_y2 chains 1:3
3.65
3.625
3.6
3.575
3.55
10000
20000
40000
60000
iteration

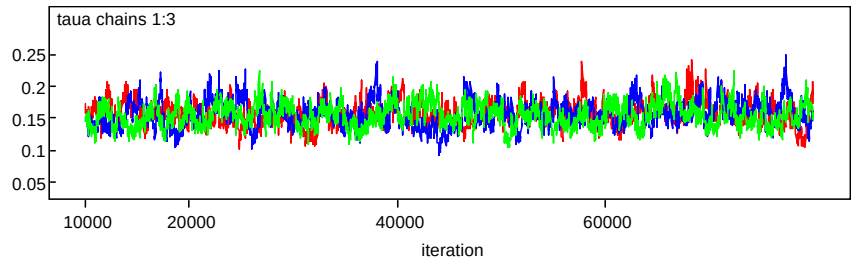

phi11 chains 1:3
1.0
0.95
0.9
0.85
10000
20000
40000
60000
iteration

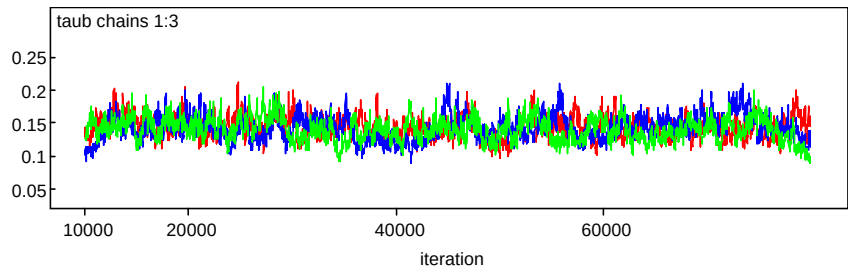

phi12 chains 1:3
0.15
0.1
0.05
0.0
-0.05
10000
20000
40000
60000
iteration


phi21 chains 1:3
0.15
0.1
0.05
0.0
10000
20000
40000
60000
iteration


phi22 chains 1:3
1.0
0.95
0.9
0.85
10000
20000
40000
60000
iteration


rhoep chains 1:3
0.99
0.98
0.97
0.96
0.95
10000
20000
40000
60000
iteration


taua chains 1:3
0.25
0.2
0.15
0.1
0.05
10000
20000
40000
60000
iteration


taub chains 1:3
0.25
0.2
0.15
0.1
0.05
10000
20000
40000
60000
iteration

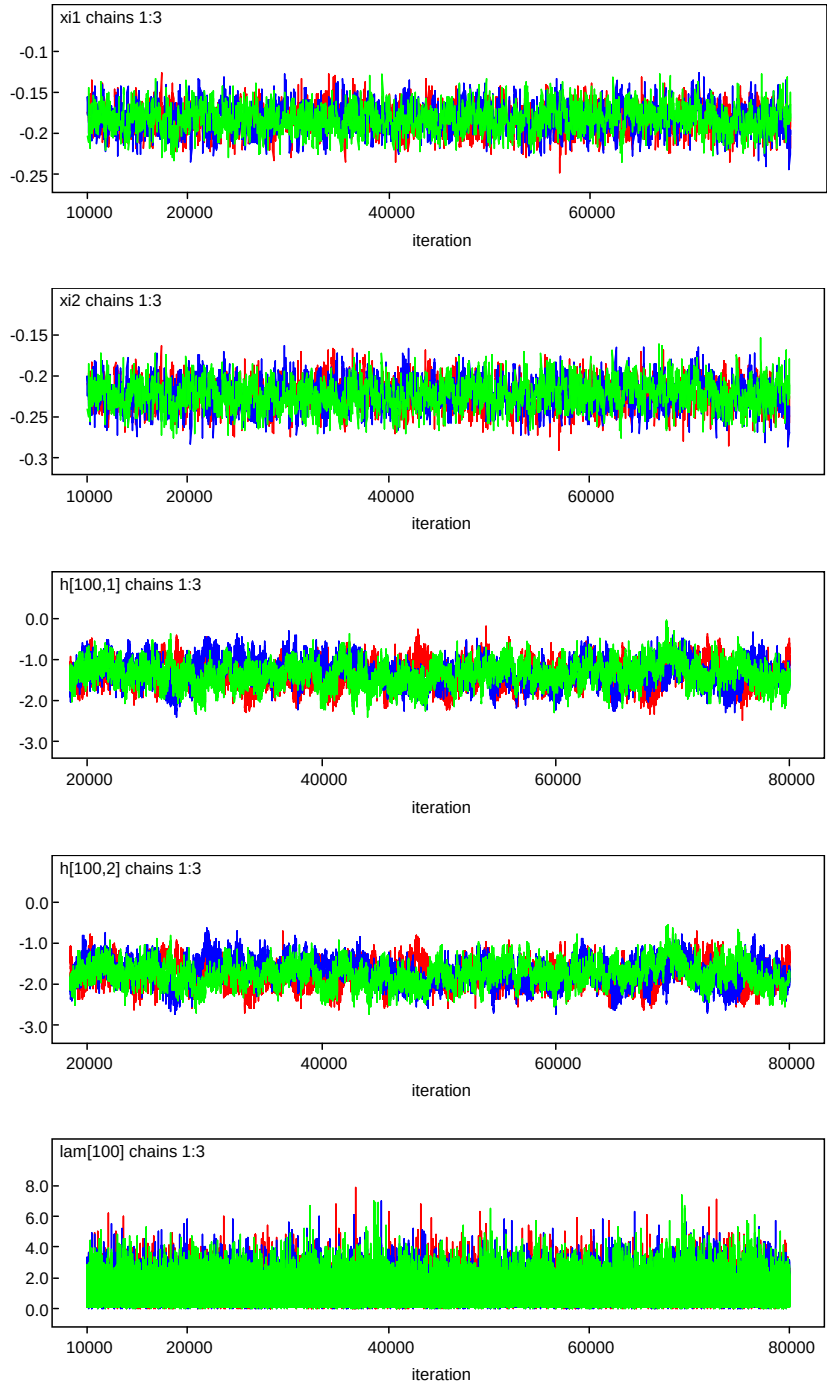


FIGURE 5.4. History plots of parameters of cc-t-G model for NSW/VIC data. Parameters presented are in the order: ν, μ_1, μ_2, μ_{y1}, μ_{y2}, ϕ_{11}, ϕ_{12}, ϕ_{21}, ϕ_{22}, ρ_ϵ, τ_a, τ_b, ξ_1, ξ_2, $h_{1,100}$, $h_{2,100}$ and λ_{100}.

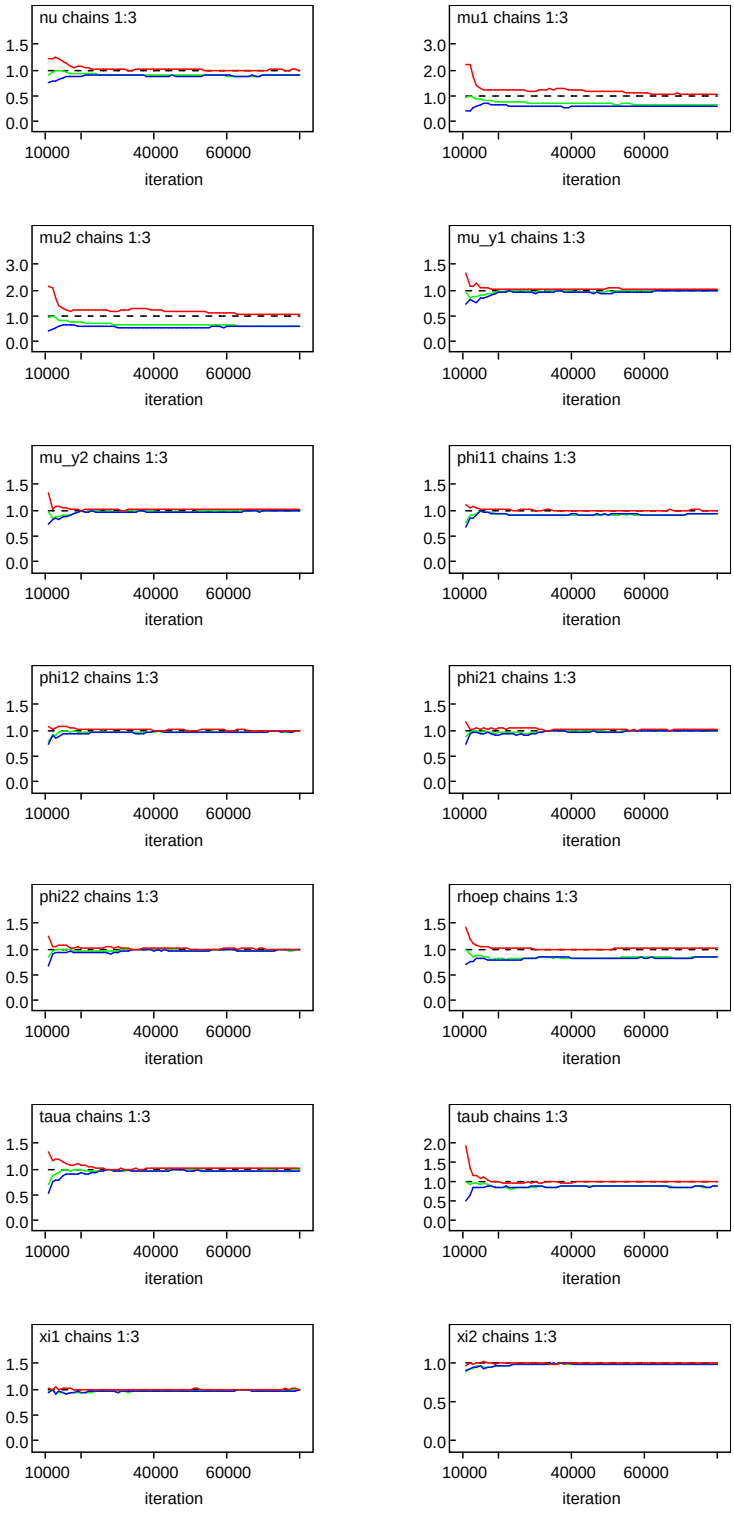


FIGURE 5.5. Brooks, Gelman, Rubin convergence plot. The Gelman-Rubin statistic is plotted in red. Parameters presented are in the order: ν, μ_1, μ_2, μ_{y1}, μ_{y2}, ϕ_{11}, ϕ_{12}, ϕ_{21}, ϕ_{22}, ρ_ϵ, τ_a, τ_b, ξ_1 and ξ_2.

TABLE 5.2. Parameter estimates (standard errors in parentheses) for NSW/VIC data.

	cc-t-G	cc-t-bN	dc-t-G	cc-ST_1-G	cc-ST_2-G	dc-ST_2-G
μ_{y1}	3.545 (0.0127)	3.538 (0.0134)	3.549 (0.0143)	3.195 (0.0183)	3.300 (0.0094)	3.279 (0.0061)
μ_{y2}	3.594 (0.0126)	3.588 (0.0134)	3.588 (0.0126)	3.235 (0.0196)	3.347 (0.0092)	3.333 (0.0082)
μ_{h1}	-1.454 (0.4744)	-2.166 (0.3033)	-2.035 (0.3091)	-2.528 (0.3827)	-3.026 (0.3546)	-3.233 (0.4149)
μ_{h2}	-1.437 (0.4345)	-2.112 (0.3077)	-1.958 (0.2950)	-2.490 (0.4437)	-2.842 (0.3220)	-2.854 (0.3759)
ρ_ε	0.9719 (0.0027)	0.9723 (0.0063)	- -	0.9955 (0.0014)	0.9931 (0.0012)	- -
ρ_η	- -	0.7336 (0.0948)	- -	- -	- -	- -
ϕ_{11}	0.9592 (0.0166)	0.9659 (0.0088)	0.9448 (0.0179)	0.9415 (0.0173)	0.9385 (0.0174)	0.9544 (0.0127)
ϕ_{22}	0.9447 (0.0165)	0.9655 (0.0089)	0.9383 (0.0190)	0.9228 (0.0170)	0.9410 (0.0163)	0.9611 (0.0138)
ϕ_{12}	**0.0383** (0.0178)	- -	**0.0513** (0.0187)	**0.0269** (0.0122)	**0.0370** (0.0182)	**0.0381** (0.0186)
ϕ_{21}	**0.0485** (0.0152)	- -	**0.0519** (0.0184)	**0.0654** (0.0225)	**0.0294** (0.0141)	**0.0283** (0.0126)
ξ_1	**−0.1784** (0.0168)	**−0.1631** (0.0391)	**−0.1802** (0.0168)	**−0.1933** (0.0121)	**−0.1999** (0.0057)	**−0.1643** (0.0093)
ξ_2	**−0.2188** (0.0176)	**−0.2041** (0.0418)	**−0.2237** (0.0175)	**−0.2364** (0.0126)	**−0.2461** (0.0050)	**−0.1980** (0.0123)
γ_1	- -	- -	- -	**0.3295** (0.0208)	**0.3185** (0.0135)	**0.3445** (0.0051)
γ_2	- -	- -	- -	**0.3320** (0.0217)	**0.3164** (0.0117)	**0.3195** (0.0048)
ν	2.581 (0.1725)	2.683 (0.1822)	4.674 (0.5541)	5.672 (0.4771)	3.110 (0.2358)	3.776 (0.6221)
ψ_0	- -	- -	4.963 (0.2941)	- -	- -	5.815 (1.0210)
ψ_1	- -	- -	0.8341 (0.0362)	- -	- -	0.8582 (0.0232)
τ_ρ	- -	- -	0.402 (0.0393)	- -	- -	0.344 (0.0303)
τ_1	0.1562 (0.0198)	0.1986 (0.0316)	0.1454 (0.0188)	0.2938 (0.0397)	0.3782 (0.0495)	0.2014 (0.0216)
τ_2	0.1390 (0.0188)	0.2048 (0.0345)	0.1496 (0.0192)	0.4339 (0.0364)	0.3305 (0.0505)	0.1624 (0.0238)
DIC	-612.97	-501.10	-1698.46	-4410.73	-4625.00	-7079.49

TABLE 5.3. Parameter estimates (standard errors in parentheses) for NSW/QLD data.

	cc-t-G	cc-t-bN	dc-t-G	cc-ST_1-G	cc-ST_2-G	dc-ST_2-G
μ_{y1}	3.487	3.539	3.521	3.116	3.222	3.252
	(0.0159)	(0.0172)	(0.0136)	(0.0775)	(0.0304)	(0.0106)
μ_{y2}	3.410	3.460	3.441	3.028	3.135	3.168
	(0.0158)	(0.0170)	(0.0137)	(0.0745)	(0.0333)	(0.0076)
μ_{h1}	-2.010	-1.854	-1.732	-3.514	-3.611	-3.241
	(0.3423)	(0.2689)	(0.3353)	(0.7756)	(0.4016)	(0.3230)
μ_{h2}	-2.000	-1.848	-1.704	-3.963	-4.004	-3.267
	(0.3626)	(0.2734)	(0.3612)	(1.0010)	(0.4686)	(0.3709)
ρ_ε	0.9766	0.9846	-	0.9764	0.9633	-
	(0.0025)	(0.0036)	-	(0.0273)	(0.0230)	-
ρ_η	-	0.8989	-	-	-	-
	-	(0.0377)	-	-	-	-
ϕ_{11}	0.9495	0.9727	0.9446	0.8927	0.8922	0.9547
	(0.0152)	(0.0078)	(0.0180)	(0.0335)	(0.0732)	(0.0394)
ϕ_{22}	0.9655	0.9661	0.9592	0.8587	0.8739	0.9757
	(0.0142)	(0.0085)	(0.0153)	(0.0320)	(0.0484)	(0.0329)
ϕ_{12}	**0.0452**	-	**0.0488**	**0.0729**	0.0474	0.0354
	(0.0149)	-	(0.0169)	(0.0298)	(0.0627)	(0.0785)
ϕ_{21}	**0.0311**	-	**0.0417**	0.0266	-0.0172	0.0117
	(0.0150)	-	(0.0168)	(0.0362)	(0.0413)	(0.0226)
ξ_1	**−0.1567**	**−0.1443**	**−0.1413**	**−0.1472**	**−0.1540**	**−0.1601**
	(0.0190)	(0.0208)	(0.0222)	(0.0157)	(0.0144)	(0.0088)
ξ_2	**−0.1525**	**−0.1392**	**−0.1360**	**−0.1445**	**−0.1489**	**−0.1588**
	(0.0189)	(0.0210)	(0.0219)	(0.0152)	(0.0138)	(0.0071)
γ_1	-	-	-	**0.3855**	**0.4282**	**0.3405**
	-	-	-	(0.0529)	(0.0236)	(0.0215)
γ_2	-	-	-	**0.3957**	**0.4399**	**0.3473**
	-	-		(0.0523)	(0.0219)	(0.0199)
ν	1.862	2.069	3.545	4.354	2.417	2.804
	(0.1086)	(0.1294)	(0.2996)	(0.5665)	(0.1649)	(0.3388)
ψ_0	-	-	5.367	-	-	5.366
	-	-	(0.3117)	-	-	(1.0830)
ψ_1	-	-	0.8141	-	-	0.8301
	-	-	(0.0328)	-	-	(0.0469)
τ_ρ	-	-	0.727	-	-	0.443
	-	-	(0.1425)	-	-	(0.0728)
τ_1	0.1160	0.1682	0.1011	0.5569	0.3765	0.1144
	(0.0171)	(0.0399)	(0.0106)	(0.2080)	(0.0744)	(0.0269)
τ_2	0.1184	0.1837	0.0933	0.6437	0.1042	0.1187
	(0.0183)	(0.0424)	(0.0095)	(0.2241)	(0.0318)	(0.0173)
DIC	-556.34	-321.80	-1904.49	-581.90	-6446.56	-8022.76

TABLE 5.4. Parameter estimates (standard errors in parentheses) for SA/VIC data.

	cc-t-G	cc-t-bN	dc-t-G	cc-ST_1-G	cc-ST_2-G	dc-ST_2-G
μ_{y1}	3.621 (0.0130)	3.600 (0.0189)	3.617 (0.0104)	3.582 (0.0137)	3.364 (0.0125)	3.393 (0.0059)
μ_{y2}	3.613 (0.0143)	3.588 (0.0208)	3.604 (0.0110)	3.583 (0.0145)	3.351 (0.0143)	3.376 (0.0059)
μ_{h1}	-1.893 (0.3827)	-2.213 (0.3024)	-2.245 (0.3266)	-1.875 (0.4325)	-3.473 (0.3611)	-2.786 (0.4138)
μ_{h2}	-1.791 (0.3679)	-2.108 (0.2922)	-2.136 (0.3137)	-1.749 (0.4253)	-3.003 (0.3273)	-2.587 (0.3597)
ρ_ε	0.9746 (0.0023)	0.9751 (0.0054)	- -	0.9740 (0.0022)	0.9919 (0.0030)	
ρ_η	- -	0.9393 (0.0335)	- -	- -	- -	
ϕ_{11}	0.8555 (0.0377)	0.9530 (0.0095)	0.8923 (0.0391)	0.9102 (0.0317)	0.8676 (0.0287)	0.9081 (0.0284)
ϕ_{22}	0.9344 (0.0298)	0.9477 (0.0106)	0.8801 (0.0316)	0.9063 (0.0307)	0.9475 (0.0223)	0.9330 (0.0236)
ϕ_{12}	**0.1456** (0.0393)	- -	**0.1031** (0.0407)	**0.0873** (0.0326)	**0.1296** (0.0352)	**0.0955** (0.0334)
ϕ_{21}	0.0562 (0.0292)	- -	**0.1099** (0.0306)	**0.0861** (0.0303)	0.0205 (0.0178)	**0.0512** (0.0208)
ξ_1	**−0.2087** (0.0204)	**−0.1999** (0.0307)	**−0.2185** (0.0159)	**−0.2094** (0.0193)	**−0.2224** (0.0157)	**−0.1964** (0.0171)
ξ_2	**−0.2324** (0.0219)	**−0.2240** (0.0322)	**−0.2403** (0.0171)	**−0.2350** (0.0207)	**−0.2441** (0.0166)	**−0.2172** (0.0177)
γ_1	- -	- -	- -	**0.3198** (0.0259)	**0.3466** (0.0184)	**0.2849** (0.0076)
γ_2	- -	- -	- -	**0.3118** (0.0254)	**0.3497** (0.0193)	**0.2895** (0.0066)
ν	2.321 (0.1515)	2.704 (0.2067)	3.625 (0.3306)	2.476 (0.1608)	2.780 (0.2313)	3.921 (0.3755)
ψ_0	- -	- -	5.034 (0.267)	- -	- -	5.724 (0.654)
ψ_1	- -	- -	0.8065 (0.0363)	- -	- -	0.8415 (0.0417)
τ_ρ	- -	- -	0.368 (0.0271)	- -	- -	0.476 (0.0346)
τ_1	0.1464 (0.0204)	0.2564 (0.0574)	0.1611 (0.0220)	0.1317 (0.0172)	0.4672 (0.0630)	0.2503 (0.0438)
τ_2	0.1310 (0.0190)	0.2517 (0.0455)	0.1344 (0.0155)	0.1337 (0.0183)	0.3036 (0.0545)	0.1697 (0.0245)
DIC	-635.01	-594.10	-1618.49	-733.43	-4430.16	-4775.74

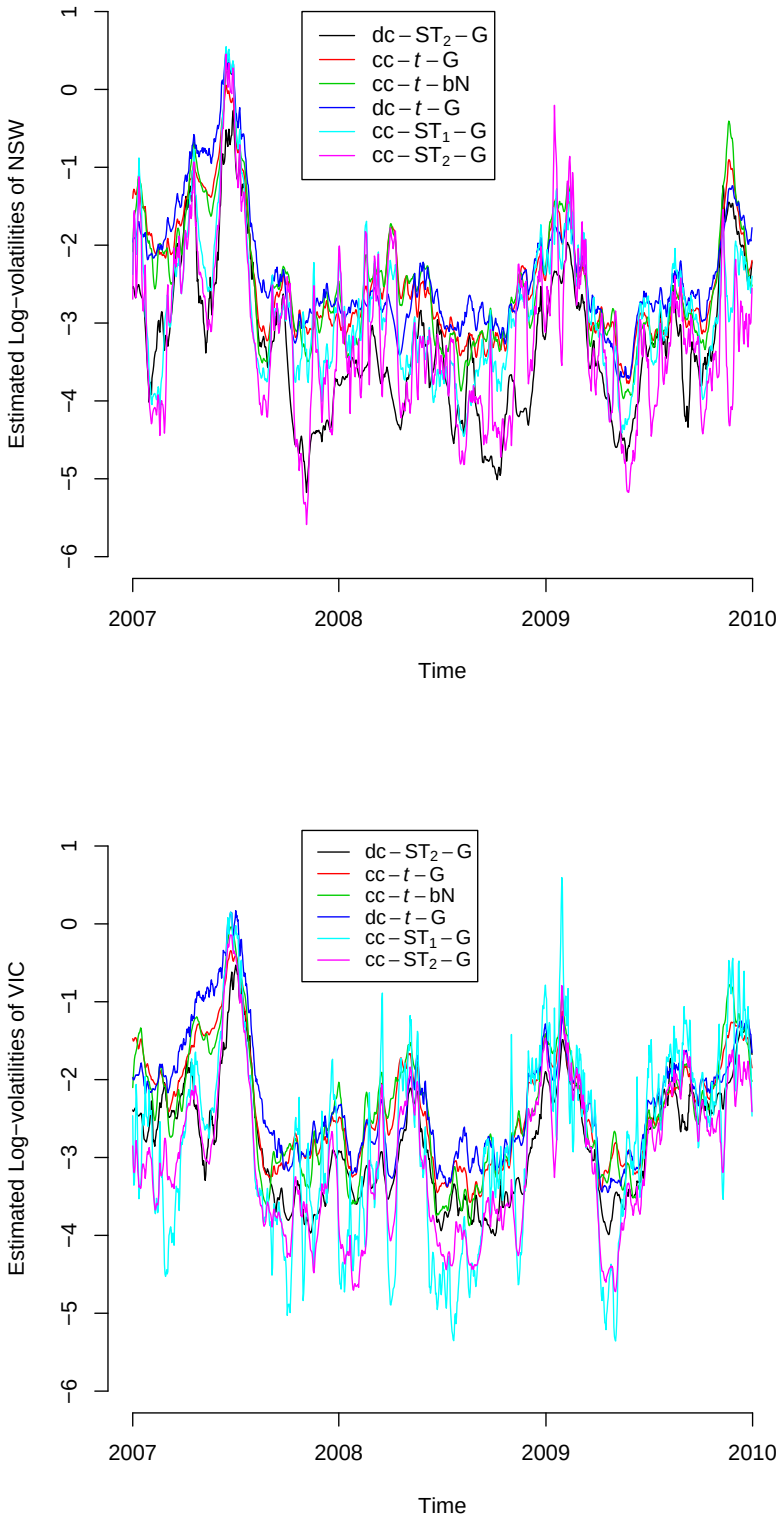


FIGURE 5.6. Comparison of volatilities across 6 models for top: NSW and bottom: VIC.

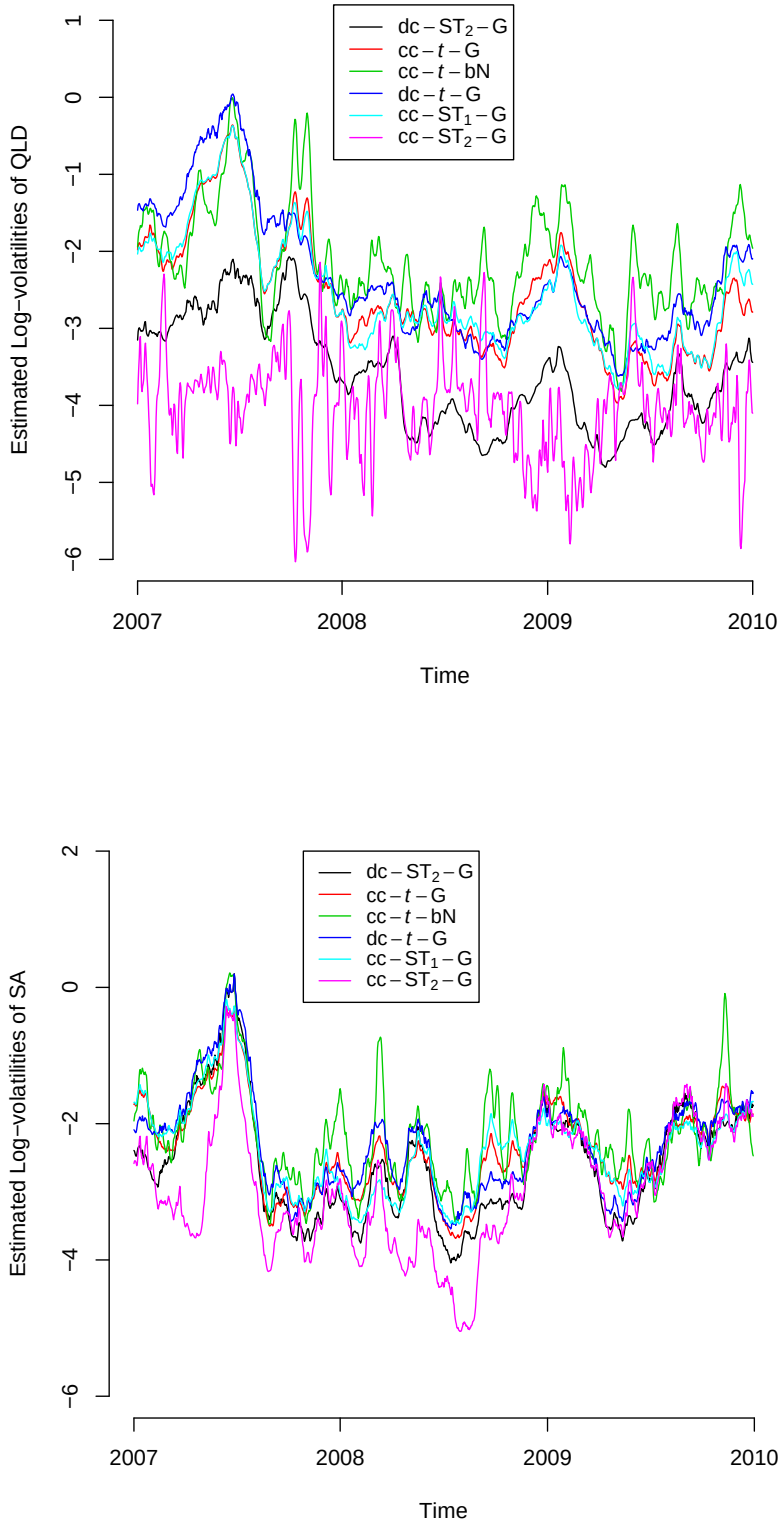


FIGURE 5.7. Comparison of volatilities across 6 models for top: QLD and bottom: SA.

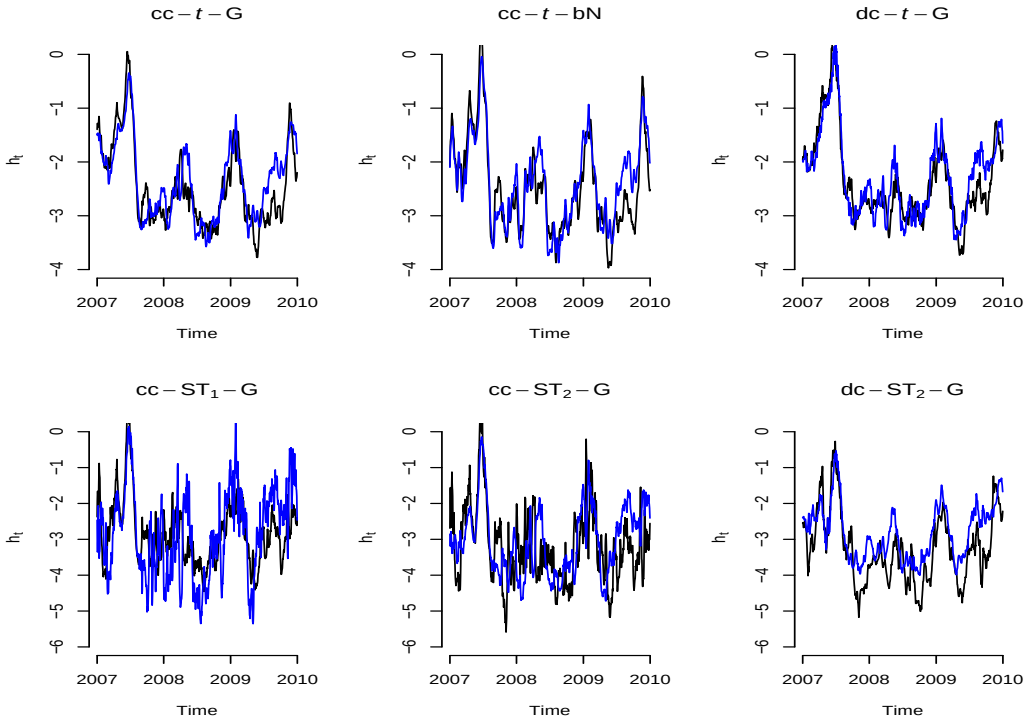


FIGURE 5.8. Comparison of volatilities of NSW (black) and VIC (blue) under the 6 models.

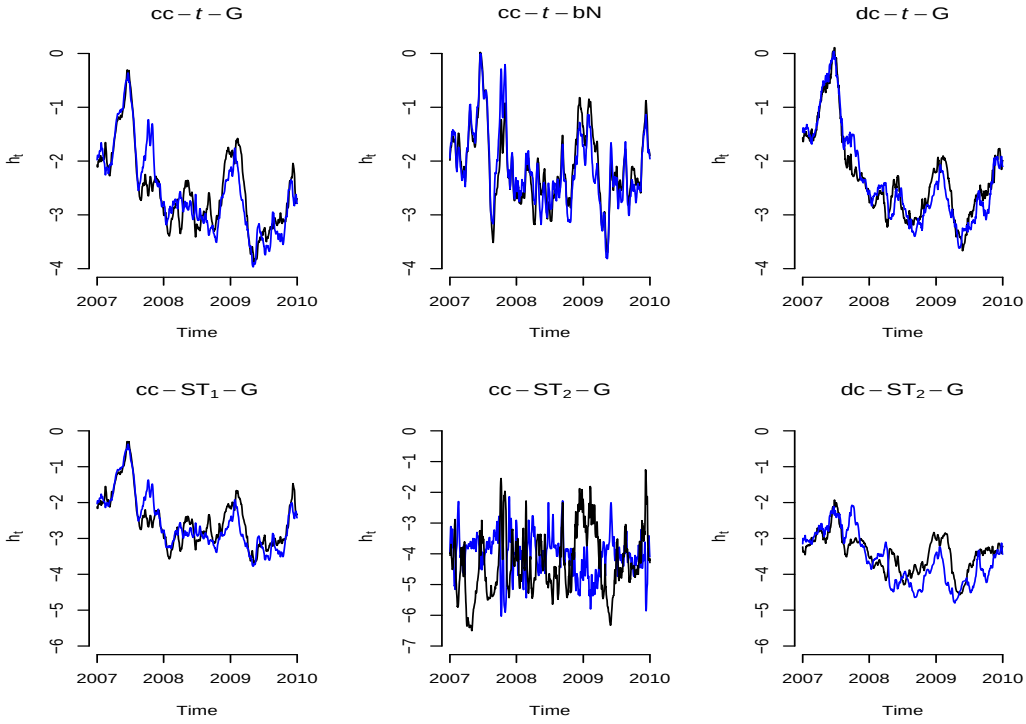


FIGURE 5.9. Comparison of volatilities of NSW (black) and QLD (blue) under the 6 models.

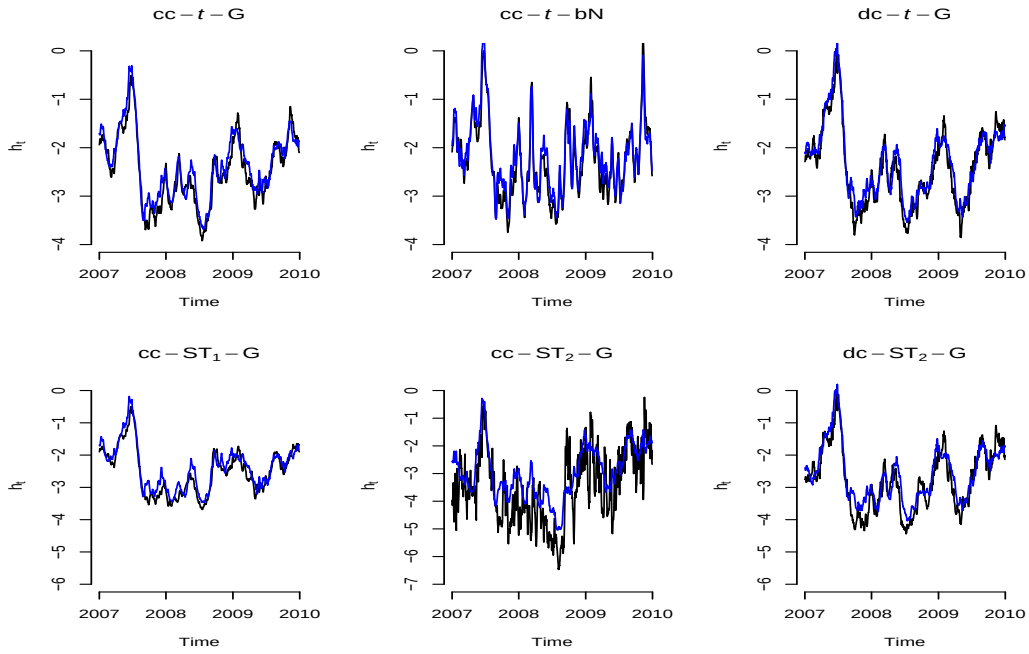


FIGURE 5.10. Comparison of volatilities of SA (black) and VIC (blue) under the 6 models.

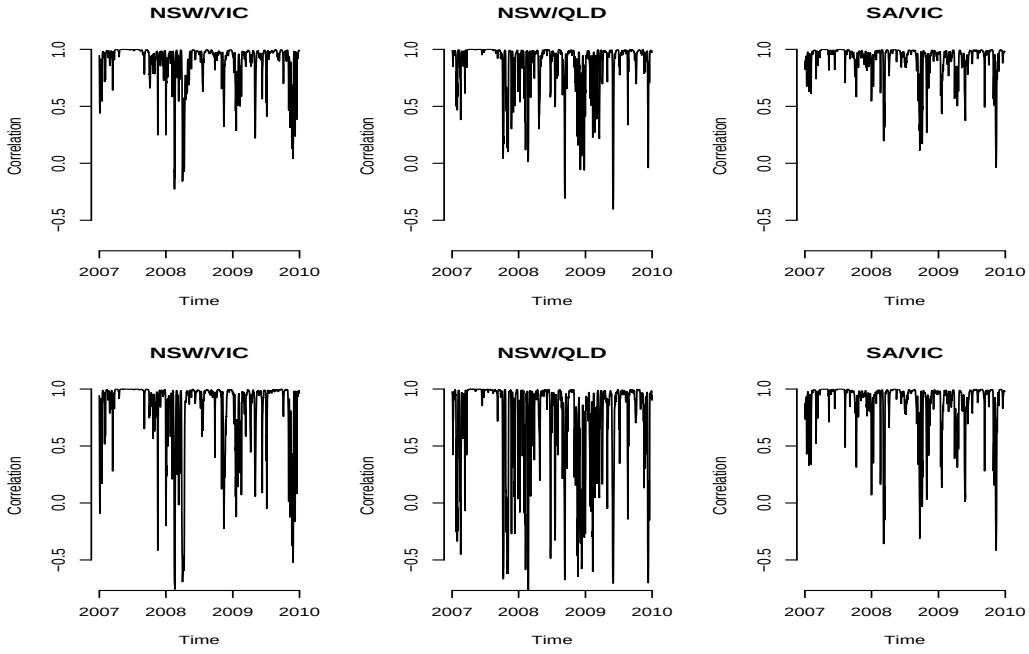


FIGURE 5.11. Dynamic correlations between price series for NSW/VIC, NSW/QLD and SA/VIC under the dc-t-G model (top row) and the dc-ST_2-G model (bottom row).

5.8. Conclusion

In this chapter, we propose some new bivariate SV models to describe the price volatility processes in three pairs (NSW/VIC, NSW/QLD and SA/VIC) of regional markets in Australia's NEM which were identified as the most correlated pairs in a previous study by Higgs (2009). We propose new specifications of the bivariate SV models including bilateral Granger causality for the volatilities as well as bivariate heavy-tailed distributions for the observations which include symmetric and skew Student-t distributions. We apply 6 different specifications of the bivariate SV model to the logarithms of average daily spot electricity prices for the period 1 January 2007 to 31 December 2009 and we compare our proposed models to the existing bivariate SV models. Model estimation is performed using Bayesian MCMC method and implemented in the `WinBUGS` software.

Our results indicate that the prices on weekdays are higher, the correlations between the prices and their volatilities are significant and the persistence of volatilities are strong. When modelling the correlation between volatilities, bilateral Granger causality performs empirically better than a bivariate normal distribution. Allowing for dynamic correlation between the price series also improves the fit of the model, making it the best model among the three models with the symmetric Student-t distribution. In extending the model to skew Student-t distributions, the skewness parameters are all significant showing that it is necessary to allow skewness in modelling, apart from assuming heavy-tailed distributions. Our results further indicate that for all three pairs of markets, models with skew error distributions outperform those with symmetric distributions and among the two skew distributions, skew Student-t type 2 produces the best results in terms of the DIC. The model which accounts for heavy-tailedness, skewness, asymptotic tail dependence and dynamic correlation is the best model out of all the models. Close examination of the volatilities show that estimated volatilities under the models with skew Student-t distributions are smaller than those with symmetric Student-t distributions. Moreover, the estimated dynamic correlations are also lower under the skew Student-t distribution.

Note that in our SV model specifications, we do not include a jump component to account for the observed price jumps. The reason for this is that we attribute extreme observations to the tail behaviour of non-Gaussian distributions, rather than to the superimposition of a jump component on a Gaussian process. Chib *et al.* (2002) considered the heavy-tailed Student-t

SV model and SV model with a jump component in the observation equation. Their results indicate that the excess kurtosis in the data seems to be better characterised by a heavy-tailed distribution as opposed to a Gaussian model with a jump component. They also found that once jumps are permitted, a less heavy-tailed error distribution suffices for capturing the excess kurtosis in the data. Since we employ the heavy-tailed Student-t distribution as the error distribution in the SV models to capture excess kurtosis, we do not include a jump component in the model.

Overall, we demonstrate the applicability of bivariate SV models with various specifications for the modelling of spot electricity price in the Australian market and we show that model implementation can be done in the Bayesian software `WinBUGS`. Further research include extending the bivariate models to multivariate models, where all markets are compared simultaneously. Moreover, the skewness parameters can be made time-varying, as the skewness of the prices may vary during a day (Panagiotelis and Smith, 2008).

Appendix 5.1

Lemma 5.8.1. *(Gupta, Chang and Huang (2002)). Let Z have a standard normal distribution with distribution function $\Phi(z)$ and Y has Chi squared distribution with ν degrees of freedom and is independent of Z. Then*

$$E_Z(\Phi(a\sqrt{Z})) = F_{t_\nu}(a\sqrt{\nu}),$$

where F_{t_ν} is the distribution function of t distribution with ν degrees of freedom.

Note that the Chi squared distribution with ν degrees of freedom is a special case of the gamma distribution, hence $Y \sim Ga\left(\frac{\nu}{2}, \frac{1}{2}\right)$.

Appendix 5.2

Rejection sampling

Suppose we want to sample points independently from a density $f(x)$. Rejection sampling may be performed by using $g(x)$ instead of $f(x)$, where $g(x) = cf(x)$ for some value of c.

To sample n points independently from $f(x)$, define an envelope function $g_u(x)$ such that $g_u(x) \geq g(x)\ \forall x \in D$, where D denotes the domain of $f(x)$. Also, optionally define a squeezing function $g_1(x)$ such that $g_1(x) \leq g(x)\ \forall x \in D$. Then the rejection sampling algorithm is given by the following:

Sample x^* from $g_u(x)$ and sample a value w independently from the $U(0,1)$ distribution.

- Perform the squeezing test: if $w \leq g_1(x^*)/g_u(x^*)$, then accept x^*. Otherwise,
- Perform the rejection test: if $x \leq g(x^*)/g_u(x^*)$, then accept x^*; otherwise reject x^*.
- Repeat until n points have been accepted.

In practice, finding a suitable envelope function $g_u(x)$ can be difficult and often involves finding the supremum of $g(x)$ in D.

Adaptive rejection sampling

Adaptive rejection sampling uses the assumption of log-concavity of $f(x)$, hence avoid the need to locate the supremum of $g(x)$ in D. After each rejection in an adaptive rejection sampling scheme, the probability of needing to evaluate $g(x)$ is reduced by updating the envelope and squeezing functions to incorporate the most recently acquired information about $g(x)$.

Assume $g(x)$ is continuous and differentiable everywhere in D and that $h(x) = \ln g(x)$ is concave everywhere in D.

(1) Initialisation step:

Let $T_k = x_i, i = 1, \ldots, k$ where $x_1 \leq x_2 \leq \ldots \leq x_k$. Define u_k to be a piecewise linear upper hull formed from the tangents to $h(x)$ at the abscissae in T_k. For $j = 1, \ldots, k-1$, the tangents at x_j and x_{j+1} intersect at

$$z_j = \frac{h(x_{j+1}) - h(x_j) - x_{j+1}h'(x_{j+1}) + x_j h'(x_j)}{h'(x_j) - h'(x_{j+1})}$$

Thus for $x \in [z_{j-1}, z_j]$, $j = 1, \ldots, k$, we define

$$u_k(x) = h(x_j) + (x - x_j)h'(x_j)$$

where z_0 is the lower bound of D and z_k is the upper bound of D.

Also define

$$s_k(x) = \frac{\exp u_k(x)}{\int_D \exp u_k(x')dx'}.$$

Lastly, let $l_k(x)$ to be a piecewise linear lower hull formed from the chords between adjacent abscissae in T_k. Thus for $x \in [x_j, x_{j+1}]$,

$$l_k(x) = \frac{(x_{j+1} - x)h(x_j) + (x - x_j)h(x_{j+1})}{x_{j+1} - x_j}$$

for $j = 1, \ldots, k-1$. For $x < x_1$ or $x > x_k$, define $l_k(x) = -\infty$.

The initialisation step involves initialise the abscissae in T_k, then calculate each of the functions $u_k(x)$, $s_k(x)$ and $l_k(x)$.

(2) Sampling step:

Sample a value x^* from $s_k(x)$ and sample a value w independently from the $U(0, 1)$ distribution.

- Perform the squeezing test: if $x \leq \exp(l_k(x^*) - u_k(x^*))$, then accept x^*. Otherwise,
- Evaluate $h(x^*)$ and $h'(x^*)$ and perform the rejection test: if $w \leq \exp(h(x^*) - u_k(x^*))$, then accept x^*; otherwise reject x^*.

(3) Updating step:

If $h(x^*)$ $h'(x^*)$ were evaluated at the sampling step, include x^* in T_k to form T_{k+1}; relabel the elements of T_{k+1} in ascending order; construct the functions $u_{k+1}(x)$, $s_{k+1}(x)$ and $l_{k+1}(x)$ on the basis of T_{k+1} then increment k. Return to the sampling step if n points have not yet been accepted.

Slice sampling

Suppose we want to sample from a density $f(x)$. This is equivalent to sampling uniformly from the area under the graph $f(x)$:

$$\mathcal{A} = \{(x, u) : 0 \leq u \leq f(x)\}.$$

The slice sampler uses data augmentation to introduce an auxiliary variable U, which is uniformly distributed on the interval $[0, f(x)]$. Hence, the joint density function of (X, U) is

$$f(x, u) = f(x)f(u|x) \propto 1_{(x,u)\in\mathcal{A}},$$

which can be sampled using the Gibbs sampler as follows:

(1) Draw $u_{t+1}|x_t \sim U[0, f(x_t)]$.

(2) Draw $x_{t+1}|u_t$ uniformly from the region $\{x : f(x) \geq u_{t+1}\}$.

Appendix 5.3

Tail dependence for the bivariate skew Student-t distributions

Consider a random bivariate vector $\boldsymbol{X} = (X_1, X_2)'$ with marginal quantile functions F_1^{-1} and F_2^{-1}. The coefficient of asymptotic lower tail dependence of $\boldsymbol{X}$ is defined as

$$\lambda_L = \lim_{q \to 0^+} P(X_2 \leq F_2^{-1}(q) | X_1 \leq F_1^{-1}(q))$$

The vector $\boldsymbol{X}$ is said to have asymptotic lower tail dependence if λ_L exists and is positive. On the other hand, if $\lambda_L = 0$, then $\boldsymbol{X}$ is said to be asymptotically independent in the lower tail. This quantity measures the degree of dependence in the lower tails of a bivariate distribution. Thus it provides insights into the tendency for the bivariate distribution to generate joint extreme events.

For each of the two types of skew Student-t distributions presented in Section 5.3, Fung and Seneta (2010a) provided theorems that state the lower tail dependence coefficients, as follow.

Theorem 5.8.1. *(Fung and Senta (2010a)). For $\boldsymbol{X} = (X_1, X_2)$ defined in (5.3) with $n = 2$, $\boldsymbol{mu}$ and $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)' \in \mathbb{R}^2$ and $\nu > 0$, the lower tail dependence coefficient is given by*

(1) If $\gamma_1 = \gamma_2 = 0$, then

$$\lambda_L = 2P\left(t_{\nu+1} \leq -\sqrt{\frac{(\nu+1)(1-\rho)}{1+\rho}}\right);$$

(2) If γ_1 and $\gamma_2 > 0$, then $\lambda_L = 0$;

(3) If γ_1 and $\gamma_2 < 0$, then $\lambda_L = 1$;

(4) $\gamma_1 > 0$ and $\gamma_2 < 0$, then $\lambda_L = 0$;

(5) $\gamma_1 = 0$ and $\gamma_2 > 0$, then $\lambda_L = 0$;

(6) $\gamma_1 = 0$ and $\gamma_2 < 0$, then

$$\lambda_L = \int_0^1 \left(1 - \Phi\left(\left(\frac{2^{\nu/2}\Gamma(\frac{\nu+1}{2})}{2\sqrt{\pi}}\right)^{1/\nu} u^{1/\nu}\right)\right) du,$$

where $\Phi(.)$ is the standard normal distribution function, $t_{\nu+1}$ is the Student-t distribution with $\nu + 1$ degrees of freedom and $|\rho| = |\frac{\Sigma_{12}}{\sqrt{\Sigma_{11}\Sigma_{22}}}| < 1$.

Note that when both skewness parameters are positive or negative, this leads to trivial values —0 or 1— of the lower tail dependence coefficient, respectively.

Theorem 5.8.2. *(Fung and Senta (2010a)). For $\boldsymbol{X}$ defined in (5.6) with $n = 2$, the lower tail dependence coefficient is given by*

$$\begin{aligned}
\lambda_L &= \int_{-\infty}^{-\left\{\left(\frac{F_{t_{\nu+1}}(-\lambda_2\sqrt{\nu+1})}{F_{t_{\nu+1}}(-\lambda_1\sqrt{\nu+1})}\right)^{1/\nu}-\rho\right\}\sqrt{\frac{\nu+1}{1-\rho^2}}} f_{t_{\nu+1}}(z) \\
&\quad \times \left\{ \frac{F_{t_{\nu+2}}\left(\left(\alpha_2\sqrt{\frac{(1-\rho^2)}{\nu+1}}z-(\alpha_1+\rho\alpha_2)\right)\sqrt{\frac{\nu+2}{1+\frac{z^2}{\nu+1}}}\right)}{F_{t_{\nu+1}}(-\lambda_1\sqrt{\nu+1})} \right\} dz \\
&\quad + \int_{-\infty}^{-\left\{\left(\frac{F_{t_{\nu+1}}(-\lambda_1\sqrt{\nu+1})}{F_{t_{\nu+1}}(-\lambda_2\sqrt{\nu+1})}\right)^{1/\nu}-\rho\right\}\sqrt{\frac{\nu+1}{1-\rho^2}}} f_{t_{\nu+1}}(z) \\
&\quad \times \left\{ \frac{F_{t_{\nu+2}}\left(\left(\alpha_1\sqrt{\frac{(1-\rho^2)}{\nu+1}}z-(\alpha_2+\rho\alpha_1)\right)\sqrt{\frac{\nu+2}{1+\frac{z^2}{\nu+1}}}\right)}{F_{t_{\nu+1}}(-\lambda_2\sqrt{\nu+1})} \right\} dz,
\end{aligned}$$

where $f_{t_\nu}(.)$ and $F_{t_\nu}(.)$ are the pdf and cdf of the univariate symmetric Student-t distribution with ν degrees of freedom and

$$\begin{aligned}
\boldsymbol{\alpha}^* &= (\alpha_1^*, \alpha_2^*)' = \frac{\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}}{\sqrt{1+\boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}}}; \quad \Psi = \begin{pmatrix} \Psi_{11} & \Psi_{12} \\ \Psi_{21} & \Psi_{22} \end{pmatrix} = \boldsymbol{\Sigma} + \boldsymbol{\gamma}\boldsymbol{\gamma}'; \\
\boldsymbol{\alpha}' &= (\alpha_1, \alpha_2) = (\alpha_1^*\sqrt{\Psi_{11}}, \alpha_2^*\sqrt{\Psi_{22}}); \quad \rho = \frac{\Psi_{12}}{\sqrt{\Psi_{11}\Psi_{22}}}; \\
\lambda_1 &= \frac{\alpha_1+\rho\alpha_2}{\sqrt{1+\alpha_2^2(1-\rho^2)}} \quad \textit{and} \quad \lambda_2 = \frac{\alpha_2+\rho\alpha_1}{\sqrt{1+\alpha_1^2(1-\rho^2)}}.
\end{aligned}$$

Hence this skew Student-t distribution does not have trivial values of the asymptotic lower tail dependence. As a result, it is a more appropriate skew extension to the multivariate symmetric Student-t distribution in terms of tail dependence than skew Student-t type 1.

CHAPTER 6

Summary

6.1. Overview

Time varying volatility is a characteristic of many financial time series and it has been an area of extensive research for many years. Apart from the non-constant evolution over time, other stylised features of the volatility include volatility clustering, mean reversion and leverage effect. The autoregressive conditional heteroscedastic (ARCH) (Engle, 1982) model and the generalised ARCH (GARCH) (Bollerslev, 1986) model and its extensions have been proposed and employed in real applications to account for those features of the volatility. A natural alternative to the popular ARCH framework is a stochastic volatility (SV) model in which the conditional mean and variance are driven by separate stochastic processes. In this thesis, we bring together some popular SV model specifications in the literature and extend the model further in different directions to enhance the empirical performance of the SV models in various applications.

The first extension of the basic SV model we consider is the SV model with leverage. Leverage is used to describe the phenomenon that a drop in equity return is associated with a rise in volatility and is modelled by a direct correlation between the returns and volatilities. Meyer and Yu (2000) studied a SV model with leverage by assuming a bivariate normal distribution and then, expressed the bivariate distribution into the conditional distribution of the return given future volatility and the marginal distribution of the volatility. This simplifies the implementation in `WinBUGS` as no specification of multivariate distribution is required. Choy *et al.* (2008) extended the bivariate normal SV model analysed in Meyer and Yu (2000) to a bivariate Student-t distribution in order to capture the leptokurtosis in the conditional returns. A bivariate scale mixture of normals (SMN) is firstly used such that conditioning on the mixing parameters, they retrieve the joint distribution of returns and volatilities as a bivariate normal. Hence Choy *et al.* (2008) proceeded according to the method in Meyer and Yu (2000) and implemented this heavy-tailed SV model with leverage. The approach used by Choy *et al.* (2008) is appealing not only because heavy-tailed distribution is employed for

robustness consideration, but the mixing parameters are used as diagnostic tools for the detection of outlying returns or volatilities. However, the major drawback of Choy's approach is that it is unable to identify whether an outlier is an outlying return or outlying volatility. Outlying in return and outlying in volatility are two distinct characteristics of a financial time series. Our empirical studies also confirm that an observation can be solely outlying in return or volatility or both. A model which can distinguish these two features helps practitioners in more accurate pricing of assets and management of risk. This explains the rationale behind Chapter 2 where we proposed an alternative approach, by deriving the distribution of the return conditional on the future volatility and expressing the Student-t distributions for the conditional returns and marginal volatilities via two SMN forms. By doing so, we can separately identify outlying returns and outlying volatilities by having two mixing parameters. The results from the simulation experiment showed the competitive performance of our approach by obtaining parameter estimates which are close to the true values. When compared to Choy's approach, our estimates have smaller mean square errors and better posterior coverages. In the empirical study, we applied our method to a set of ten exchange rate returns which was previously analysed in Choy *et al.* (2008). Our results confirmed the empirical phenomenon of insignificant leverage effect in exchange rate returns. More importantly, when performing outlier diagnostics for the AUD/JPY exchange rate pair, we found that among those (pairs of) outliers identified in Choy *et al.* (2008), the majority of them were outlying returns and that only one observation was an outlying volatility. Furthermore, we applied both our approach and Choy's approach to seven international stock market indices and as expected, both methods found significant leverage effects in all markets. In five out of the seven markets, the DICs computed for the SV model using our approach are smaller than those using Choy's approach, indicating a better fit of the model for those markets. Outlier diagnostics were performed using the S&P 500 returns, Choy's approach successfully identified two most significant outliers and we further revealed that one outlier was outlying in both return and volatility while the other outlier was outlying in volatility only.

The popularity and success of using heavy-tailed distributions such as the Student-t and the exponential power (EP) distributions in SV modelling has motivated us to investigate other flexible error distributions for the returns. In particular, we are interested in employing a general distribution that nests those commonly used distributions such that the data

automatically adopt the most appropriate distribution through different parameter values. In addition, it is also desirable that this flexible distribution has a scale mixture form to facilitate model implementation via Gibbs sampler and to enable outlier detection. In Chapter 3, we adopted the generalised-t (GT) distribution as the more flexible error distribution for the returns. The GT distribution has two shape parameters which define a very general family of density functions, nesting the uniform, normal, Student-t, EP and Laplace distributions. Moreover, it was shown that the GT distribution can be expressed as a scale mixture of the EP distributions with a generalised gamma mixing distribution and further expressed as a scale mixture of uniform (SMU) distributions (Choy and Chan, 2008). We derived the set of full conditional distributions for the parameters and obtained mostly (truncated) standard distributions. We applied the SV model with GT distribution to the AUD/EUR and AUD/JPY exchange rate returns and found that for the AUD/JPY pair, the SV model with GT distribution outperformed the Student-t distribution based on DIC. The smoothed estimates of the log-volatilities are quite similar under the two distributions while the parameter estimates implied a larger unconditional kurtosis for the returns under the GT distribution. This indicates that the GT distribution provides a more flexible tail thickness than the Student-t distribution. Furthermore, our empirical results showed that the AUD/JPY is more heavy-tailed than the AUD/EUR exchange rate pair and that they are both positively and significantly influenced by the AUD/USD exchange rate returns. Outlier diagnostics were performed using the combined mixing parameters and we successfully identified the sets of outlying returns and volatilities under both GT and Student-t distributions for AUD/EUR and AUD/JPY exchange rates. We further revealed that mixing parameters under the GT distribution have a smaller magnitude than those under the Student-t distribution and that some outlying returns and volatilities were common to both AUD/EUR and AUD/JPY pairs. In the second part of Chapter 3, we extended the GT SV model to incorporate the general asymmetric behaviour between returns and volatilities, by adopting the asymmetric SV model specifications of Asai and McAleer (2005a) and So *et al.* (2002). Various asymmetric SV models under each of the GT, Student-t and EP distributions were fitted to the S&P 500 return data. The results showed the presence of significant asymmetric effects under all distributions and under all models. We also found that the GT SV model has the best model fit and that the asymmetric SV model specification of Asai and McAleer (2005a) performed empirically better

than that of So *et al.* (2002). An analysis of the standardised returns disclosed that they are mildly negatively skewed, which motivated us to employ skewed distribution for the returns in Chapter 4.

We investigated the SV model with the asymmetric Laplace (AL) distribution, which is able to account for both leptokurtosis and skewness of the financial returns, in Chapter 4. When the AL distribution is used as the error distribution in a SV model, the model can be interpreted as a quantile regression model with heteroscedasticity in the response controlled by a SV model. The use of AL distribution facilitates the estimation of causal effects across varying quantile levels in a quantile regression problem. We applied both the SV model with AL error distribution which has a random shape parameter and the quantile SV model with the shape parameter fixed at different quantile levels to five market indices in an empirical study. Statistically significant negative skewness was found for AORD, Nikkei 225 and FTSE indices, thus the SV model with AL distribution is superior to its symmetric counterpart (Laplace distribution) with respect to DIC. In the quantile SV model, we observed that the systematic changes of parameters are generally in agreement with those in Chen *et al.* (2009). Hence in this chapter, we demonstrated the practicability of controlling for heteroscedasticity by employing a SV model in a quantile regression framework and at the same time, we extended the error distribution to AL distribution to capture the skewness and leptokurtosis of financial returns simultaneously.

Compared to the abundant literature on univariate SV models, studies on multivariate SV models are relatively limited. In Chapter 5, we explored the application of multivariate SV models in modelling electricity price volatilities, an area where the application of SV models is scarce. We proposed some new specifications for the bivariate SV model including bivariate heavy-tailed skewed Student-t distributions for the prices as well as bilateral Granger causality for the latent volatilities. We applied six bivariate SV model specifications to the daily spot electricity prices of three pairs of Australian regional electricity markets. Proposed models were shown to capture prominent characteristics of electricity markets including leptokurtosis and positive skewness of the prices and time-varying correlation between the price volatilities. The weekday/weekend effect was estimated to be negative and significant, agreeing with our hypothesis that the electricity prices are lower on weekends than weekdays. Furthermore, we demonstrated the empirical superiority of the skew Student-t

distribution introduced by Branco and Dey (2001) over the alternative skew Student-t introduced in Demarta and McNeil (2005), although they both can be expressed as a SMN.

For the estimation of various SV models, we adopted the Bayesian Markov chain Monte Carlo (MCMC) approach as it was shown to be one of the most efficient estimation tools for SV models (Andersen *et al.*, 1999). The Bayesian software `WinBUGS` was used for all simulation and empirical studies in this thesis, as it allows straightforward implementation and modifications of the SV models. Model comparison criterion, DIC, and smoothed estimates of the volatilities, are readily obtained by `WinBUGS`. Despite the ease of implementation using `WinBUGS`, we derived a whole set of full conditional posterior distributions for all parameters in all the proposed models for the implementation in other environments for expert users.

6.2. Further research

Although some substantial development of the SV models have been made in this thesis, there are still unattended areas that are worthy of further studies.

6.2.1. Extension of the proposed models. Some extensions of our proposed SV models were discussed in individual chapters, and they are summarised below:

- In Chapter 2, we modelled the leverage effect in financial returns by assigning a bivariate Student-t distribution followed by expressing the conditional distribution of the return and marginal distribution for the volatility in SMN forms. Other bivariate heavy-tailed distributions might have also been used, for example, the EP and GT distributions, despite their implementation being more involved. As both univariate EP and GT distributions can be expressed as a SMU, a useful extension given by Fung and Seneta (2008) characterises multivariate EP distribution as multivariate SMU. Moreover, with heavy-tailed distributions incorporated into the model, it is useful to compare the leverage effect modelled by a direct correlation between the returns and volatilities and the general asymmetry effect modelled by including past return information in the volatility process as was done in Chapter 3.
- In Chapter 4, the AL distribution was used to account for both leptokurtosis and skewness of returns. An alternative skew distribution is the skewed GT distribution, developed by Theodossiou (1998). This skew GT distribution nests the symmetric

GT, and all the nested distributions of the GT, making it a promising error distribution to be used in SV models.

- In Chapter 5, we applied bivariate SV models to three pairs of the Australian regional electricity markets. A natural extension is to employ the multivariate SV models, where all regional markets can be estimated and compared simultaneously. Furthermore, a more flexible modelling of skewness of electricity price is to allow time-varying skewness, as it was shown that the skewness of prices may vary during a day (Panagiotelis and Smith, 2008).

6.2.2. More efficient sampling methods. Although we choose to adopt the more efficient Bayesian MCMC method as our estimation tool for the SV models, `WinBUGS` uses single-move Gibbs sampler which generates a single state variable at a time given the rest of the state variables and other parameters. The single move sampler converges slowly and produces posterior samples that are highly autocorrelated. Therefore, we need to run a long Markov chain and obtain a large number of samples to conduct statistical inference. Other methods have been proposed to reduce the sample autocorrelation effectively. For example, Kim *et al.* (1998) proposed the mixture sampler which firstly transforms the model into a linear state-space model and approximates the error distribution by a mixture of normal distributions. The drawback of this method is that its use is limited to the models that can be transformed into a linear state-space form. Another method is the block sampler (or the multimove sampler) proposed by Shephard and Pitt (1997), which samples a block of volatility state variables at a time. Currently, the multimove sampler is not available in `WinBUGS`. Therefore, to implement more sophisticated sampling methods such as the multimove sampler, specialised codes need to be written in, for example, Ox programming language (Doornik, 1996).

6.2.3. Comparison with GARCH models. As the main objective of this thesis is to develop more flexible SV models, no comparison has been made with the popular GARCH models. Asai (2009) showed that there are cases where the SV model with the normal distribution is less effective in capturing leptokurtosis than the GARCH model with heavy-tailed distributions. Therefore, we can compare the empirical performance of SV and GARCH models, taking into account different features of the financial data, such as leptokurtosis, skewness and leverage effect. Comparison of the SV and GARCH models can also take place

in a quantile regression model, where heteroscedasticity is controlled by different volatility processes. In this thesis, we are only concerned with different SV model specifications and the resulting parameter and latent volatility estimates. However, the SV model can be used in option pricing and risk management, for example, to evaluate and forecast Value-at-Risk and Expected shortfall, and for the comparison of empirical performance with the GARCH models, in order to obtain useful insights into the appropriateness of the two classes of model in different context.

Bibliography

[1] Abanto-Valle, C.A., Bandyopadhyay, D., Lachos, V.H. and Enriquez, I. (2010). Robust Bayesian analysis of heavy-tailed stochastic volatility models using scale mixture of normal distributions. *Computational Statistics & Data Analysis* **54**, 2883-2898.

[2] Aggarwal, R. Inclan, C. and Leal, R. (1999). Volatility in emerging stock markets. *The Journal of Financial and Quantitative Analysis* **34**, 33-55.

[3] Ait-Sahalia, Y. and Brandt, M.W. (2001). Variable selection for portfolio choice. *Journal of Finance* **56**, 1297-1355.

[4] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *Proceedings of the Second International Symposium on Information Theory*, eds. Petrov, B. N. and Csaki, F., Budapest: Akademiai Kiado, pp. 267-281.

[5] Andersen, T.G. (1994). Stochastic autoregressive volatility: a framework for volatility modelling. *Mathematical Finance* **4**, 75-102.

[6] Andersen, T.G., Chung, H. and Sorensen, B. (1999). Efficient method of moments estimation of a stochastic volatility model: a Monte Carlo study. *Journal of Econometrics* **91**, 61-87.

[7] Andrews, D.F. and Mallows, C.L. (1974). Scale mixtures of normal distributions. *Journal of the Royal Statistical Society, Series B* **36**, 99-102.

[8] Arslan, O. (2004). Family of multivariate generalized t distributions. *Journal of Multivariate Analysis* **89**, 329-337.

[9] Arslan, O. and Genç, A.I. (2003). Robust location and scale estimation based on the univariate generalized t (GT) distribution. *Communications in Statistics - Theory and Methods* **32**, 1505-1525.

[10] Asai, M. (2008). Autoregressive stochastic volatility models with heavy-tailed distributions: a comparison with multifactor volatility models. *Journal of Empirical Finance* **15**, 332-341.

[11] Asai, M. (2009). Bayesian analysis of stochastic volatility models with mixture-of-normal distributions. *Mathematics and Computers in Simulation* **79**, 2579-2596.

[12] Asai, M and McAleer, M. (2005a). Dynamic asymmetric leverage in stochastic volatility models. *Econometric Reviews* **24**, 317-332.

[13] Asai, M. and McAleer, M. (2005b). Asymmetric multivariate stochastic volatility. *Econometric Reviews* **25**, 453-473.

[14] Asai, M and McAleer, M. (2011). Alternative asymmetric stochastic volatility models. *Econometric Reviews* **30**, 548-564.

[15] Asai, M and Unite, A. (2010). General asymmetric stochastic volatility models using range data: estimation and empirical evidence from emerging equity markets. *Applied Financial Economics* **20**, 1041-1049.

[16] Asai, M., McAleer, M. and Yu, J. (2006). Multivariate stochastic volatility models: a review. *Econometric Reviews* **25**, 145-175.

[17] Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics* **12**, 171-178.

[18] Azzalini, A. and Capitanio A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew-t distribution. *Journal of the Royal Statistical Society, Series B* **65**, 367-389.

[19] Azzalini, A. and Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika* **83**, 715-726.

[20] Ballie, R.T. and Bollerslev, T. (1989). The message in daily exchange rates: a conditional-variance tale. *Journal of Business & Economic Statistics* **7**, 297-305.

[21] Bassett, G.W. and Chen, H. (2001). Portfolio style: return-based attribution using quantile regression. *Empirical Economics* **26**, 293-305.

[22] Bauwens, L., Laurent, S. and Rombouts, J.K.V. (2006). Multivariate GARCH models: a survey. *Journal of Applied Econometrics* **21**, 79-109.

[23] Bekaert, G. and Harvey, C.R. (1997). Emerging equity market volatility. *Journal of Financial Economics* **43**, 29-77.

[24] Bekaert, G. and Wu, G. (2000). Asymmetric volatility and risk in equity markets. *The Review of Financial Studies* **13**, 1-42.

[25] Berg, A., Meyer, R. and Yu, J. (2004). Deviance information criterion for comparing stochastic volatility models. *Journal of Business & Economic Statistics* **22**, 107-120.

[26] Black, F. (1976). Studies of stock market volatility changes. *Proceedings of the American Statistical Association, Business and Economic Statistical Section*, pp. 177-181.

[27] Black, F. and Scholes, M. (1972). The valuation of options contracts and a test of market efficiency. *Journal of Finance* **27**, 399-418.

[28] Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy* **81**, 637-654.

[29] Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics* **31**, 307-327.

[30] Bollerslev, T. (1987). A conditional heteroskedastic time series model for speculative prices and rates of return. *The Review of Economics and Statistics* **69**, 542-547.

[31] Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH approach. *Review of Economics and Statistics* **72**, 498-505.

[32] Bollerslev, T., Engle, R.F. and Nelson, D.B. (1994). *ARCH Models*. In: Engle R.F. and McFadden D.L., eds. *Handbook of Econometrics*, Volume IV, Amsterdam, North-Holland, pp. 2959-3038.

[33] Bollerslev, T., Litvinova, J. and Tauchen, G. (2006). Leverage and volatility feedback effects in high-Frequency data. *Journal of Financial Econometrics* **4**, 353-384.

[34] Branco, M.D. and Dey, D.K. (2001). A general class of multivariate skew-elliptical distributions. *Journal of Multivariate Analysis* **79**, 99-113.

[35] Brooks, S. and Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* **7**, 434-456.

[36] Broto, C. and Ruiz, E. (2004). Estimation methods for stochastic volatility models: a survey. *Journal of Economic Surveys* **18**, 613-649.

[37] Buchinsky, M. (1995). Quantile regression, Box-Cox transformation model, and the U.S. wage structure, 1963-1987. *Journal of Econometrics* **65**, 109-154.

[38] Cappuccio, N., Lubian, D. and Raggi, D. (2004). MCMC Bayesian esimation of a skew-GED stochastic volatility model. *Studies in Nonlinear Dynamics & Econometrics* **8**, Article 6.

[39] Chan, J.S.K., Choy, S.T. and Makov, U.E. (2008). Robust Bayesian analysis of loss reserves data using the generalised-t distribution. *Astin Bulletin* **38**, 207-230.

[40] Chen, Y.T. (2001). Testing conditional symmetry with an application to stock returns. Working paper, Institute for Social Science and Philosophy, Academia Sinica.

[41] Chen, C.W.S., Chiang, T.C. and So, M.K.P. (2003). Asymmetrical reaction to US stock-return news: evidence from major stock markets based on a double-threshold model. *The Journal of Economics and Business* Special issue on globalization in the new millennium: evidence on financial and economic integration, **55**, 487-502.

[42] Chen, C.W.S., Gerlach, R and Wei, D.C.M. (2009). Bayesian causal effects in quantiles: accounting for heteroscedasticity. *Computational Statistics and Data Analysis* **53**, 1993-2007.

[43] Chen, C.W.S., Liu, F.C. and So, M.K.P. (2008). Heavy-tailed-distributed threshold stochastic volatility models in financial time series. *The Australian & New Zealand Journal of Statistics* **50**, 29-51.

[44] Chen, Q., Gerlach, R. and Lu, Z. (2011). Bayesian Value-at-Risk and expected shortfall forecasting via the asymmetric Laplace distribution. *Computational Statistics and Data Analysis*, in press.

[45] Chib, S. (1995). Marginal likelihood from Gibbs output. *Journal of American Statistical Association* **90**, 1313-1321.

[46] Chib, S. and Greenberg, E. (1994). Bayes inference for regression models with ARMA (p, q) errors. *Journal of Econometrics* **64**, 183-206.

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

[48] Chib, S., Nardari, F. and Shephard, N. (2006). Analysis of high dimensional multivariate stochastic volatility models. *Journal of Econometrics* **134**, 341-371.

[49] Choy, S.T.B. and Smith, A.F.M. (1997). Hierarchical models with scale mixtures of normal distribution. *TEST* **6**, 205-221.

[50] Choy, S.T.B. and Chan, C.M. (2000). Bayesian estimation of stochastic volatility model via scale mixtures distributions. *Statistics and Finance: An Interface* **1**, 186-204.

[51] Choy, S.T.B. and Chan, J.S.K. (2008). Scale mixture representaions in statistical modelling. *The Australian & New Zealand Journal of Statistics* **50**, 135-146.

[52] Choy, S.T.B., Wan, W.Y. and Chan, C.M. (2008). Bayesian Student-t stochastic volatility models via scale mixtures. *Advances in Econometrics*, Special issue on Bayesian Econometrics Methods **23**, 595-618.

[53] Christie, A.A. (1982). The stochastic behavior of common stock variances. *Journal of Financial Economics* **10**, 407-432.

[54] Chuang, C.C., Kuan, C.M. and Lin, H. (2007). Causality in quantiles and dynamic stock return-volume relations. In: Workshop on Forecasting and Risk Management held on Dec. http://ccer.org.cn/download/80201.pdf.

[55] Clark, P.K. (1973). A subordinated stochastic process model with finite variance for speculative prices. *Econometrica* **41**, 135-156.

[56] Cole,T.J. and Green, P.J. (1992). Smoothing reference centile curves: the LMS method and penalized likelihood. *Statistics in Medicine* **11**, 1305-1319.

[57] Danielsson, J. (1994). Stochastic volatility in asset prices: estimation with simulated maximum likelihood. *Journal of Econometrics* **64**, 375-400.

[58] Danielsson, J. (1998). Multivariate stochastic volatility models: estimation and comparison with VGARCH models. *Journal of Empirical Finance* **5**, 155-173.

[59] Danielsson, J. and Richard, J.-F. (1993). Accelerated gaussian importance sampler with application to dynamic latent variable models. *Journal of Applied Econometrics* **8**, 153-173.

[60] Demarta, S. and McNeil, A.J. (2005). The t copula and related copulas. *International Statistical Review* **73**, 111-129.

[61] Dempster, A. P. (1974). The direct use of likelihood for significance testing. *Proceedings of Conference on Foundational Questions in Statistical Inference,* University of Aarhus, pp. 335-352.

[62] Doornik, J.A. (1996). *Ox: Object oriented matrix programming, 1.10.* Chapman & Hall, London.

[63] Elliott, R.J., Lin, T. and Miao, H. (2007). A hidden markov stochastic volatility model for energy prices. The commodities and finance centre forum for commodity modelling.

[64] Engle, R.F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of the United Kingdom inflation. *Econometrica* **50**, 987-1007.

[65] Engle, R. and Ng, V. (1993). Measuring and testing the impact of news in volatility. *Journal of Finance* **43**, 1749-1778.

[66] Fama, E.F. (1965). The behaviour of stock market prices. *Journal of Business* **38**, 34-105.

[67] Fernández, C., Osiewalski, J. and Steel, M.F.J. (1995). Modelling and inference with v-spherical distributions. *Journal of the American Statistical Association* **90**, 1331-1340.

[68] Finlay, R. and Seneta, E. (2008). Stationary-increment Variance-Gamma and t models: Simulation and parameter estimation. *International Statistical Review* **76**, 167-186.

[69] Fleming, J., Kirby, C. and Ostdiek, B. (1998). Information and volatility linkages in the stock, bond and money markets. *Journal of Financial Econometrics* **49**, 111-137.

[70] Fung T and Seneta E. (2008). A characterisation of scale mixtures of the uniform distribution. *Statistics and Probability Letters* **78**, 2883-2888.

[71] Fung, T. and Seneta, E. (2010a). Modelling and estimation for bivariate financial returns. *International Statistical Review* **78**, 117-133.

[72] Fung, T. and Seneta, E. (2010b). Tail dependence for two skew t distributions. *Statistics and Probability Letters* **80**, 784-791.

[73] Gallant, A.R., Hsieh, D. and Tauchen, G. (1991). On fitting a recalcitrant series: the Pound/Dollar exchange rates, 1974-83. In: Barnett, W.A., Powell,J. and Tauchen, G., eds. *Nonparametric and Semiparametric Methods in Econometrics and Statistics, Proceedings of the Fifth International Symposium in Economic Theory and Econometrics*, Cambridge, U.K., Cambridge University Press.

[74] Gallant, A.R. and Tauchen, G. (1996). Which moments to match? *Econometric Theory* **12**, 657-681.

[75] Gelfand, A.E., Hills, S.E., Racine-Poon, A. and Smith, A.F.M. (1990). Illustration of Bayesian inference in normal models using Gibbs sampling. *Journal of the American Statistical Association* **85**, 972-985.

[76] Gelfand, A.E. and Smith, A.F.M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85**, 398-409.

[77] Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (2004). *Bayesian Data Analysis 2nd Ed*. Chapman & Hall, New York.

[78] Gelman, A. and Rubin, D. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science* **7**, 457-511.

[79] Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distribution and Bayesian restoration of images. *IEE Transaction on Pattern Analysis and Machine Intelligence* **6**, 721-741.

[80] Gerlach, R., Chen, C.W.S., Lin, D.S.Y. and Huang, M-H. (2006). Asymmetric responses of international stock markets to trading volume. *Physica A-Statistical Mechanics And Its Applications* **360**, 422-444.

[81] Gerlach, R. and Tuyl, F. (2006). MCMC methods for comparing stochastic volatility and GARCH models. *International Journal of Forecasting* **22**, 91-107.

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

[83] Geweke, J. (1994a). Bayesian comparison of econometric models. Working Papers 532, Research Department, Federal Reserve Bank of Minneapolis.

[84] Geweke, J. (1994b). Comment on Bayesian analysis of stochastic volatility models. *Journal of Business and Economic Statistics* **12**, 397-399.

[85] Gilks, W.R. and Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Applied Statistics* **41**, 337-348.

[86] Glosten, L.R., Jagannanthan, R. and Runkle, D. (1993). Relationship between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance* **48**, 1779-1802.

[87] Gómez, E., Gómez-Villegas, M.A. and Marín, J.M. (1998). A multivariate generalization of the power exponential family of distributions. *Communications in Statistics - Theory and Methods* **27**, 589-600.

bibitemgour Gourieroux, C. and Jasiak, J. (2008). Dynamic quantile models. *Journal of Econometrics* **147**, 198-205.

[88] Gourieroux, C., Monfort, A. and Renault, E. (1993). Indirect inference. *Journal of Applied Econometrics* **8**, 85-118.

[89] Granger, C.W.J. (1969). Investigating causal relations by econometric models and cross spectral methods. *Econometrica* **37**, 424-438.

[90] Guermat, C. and Harris, R.D.F. (2001). Robust conditional variance estimation and value-at-risk. *Journal of Risk* **4**, 25-41.

[91] Hadsell, L., Marathe, A. and Shawky, H.A. (2004). Estimating the volatility of wholesale electricity spot prices in the US. *The Energy Journal* **25**, 23-40.

[92] Han, C. and Carlin, B. P. (2001). Markov chain Monte Carlo methods for computing Bayes factors: a comparative review. *Journal of the American Statistical Association* **96**, 1122-1132.

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

[94] Harvey, A.C. and Shephard, N. (1996). The estimation of an asymmetric stochastic volalitity model for asset returns. *Journal of Business and Economic Statistics* **14**, 429-34.

[95] Harvey, C.P. and Siddique, A. (2000). Conditional skewness in asset pricing tests. *Journal of Finance* **55**, 1263-1295.

[96] Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97-109.

[97] Heidelberger, P. and Welch, P. (1983). Simulation run length control in the presence of an initial transient. *Operations Research* **31**, 1109-1144.

[98] Higgs, H. (2009). Modelling price and volatility inter-relationships in the Australian wholesale spot electricity markets. *Energy Economics* **31**, 748-756.

[99] Higgs, H. and Worthington, A. (2005). Systematic features of high-frenquency volatility in the Australian electricity market: intraday patterns, information arrival and calender effects. *The Energy Journal* **26**, 1-20.

[100] Higgs, H. and Worthington, A. (2008). Stochastic price modelling of high volatility, mean-reverting, spike-prone commodities: the Australian wholesale electricity market. *Energy Economics* **30**, 3172-3195.

[101] Hinkley, D.V. and Revankar, N.S. (1977). Estimation of the Pareto law from underreported data. *Journal of Econometrics* **5**, 1-11.

[102] Hull, J. and White, A. (1987). The pricing of options on assets with stochastic volatilities. *Journal of Finance* **42**, 281-300.

[103] Ishihara, T. and Omori, Y. (2011). Efficient Bayesian estimation of a multivariate stochastic volatility model with cross leverage and heavy-tailed errors. *Computational Statistics and Data Analysis*, in press.

[104] Jacquier, E., Polson, N.G. and Rossi, P.E. (1994). Bayesian analysis of stochastic volatility models. *Journal of Business and Economic Statistics* **12**, 371-389.

[105] 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.

[106] Johansson, A.C. (2009). Stochastic volatility and time-varying country risk in emerging markets. *The European Journal of Finance* **15**, 337-363.

[107] Jungbacker, B. and Koopman, S.J. (2006). Monte Carlo likelihood estimation for three multivariate stochastic volatility models. *Econometric Reviews* **25**, 385-408.

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

[109] Kitagawa, G. (1996). Monte Carlo filter and smoother for gaussian nonlinear state space models. *Journal of Computational and Graphical Statistics* **5**, 1-25.

[110] Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica* **46**, 33-50.

[111] Koenker, R. and Geling, R. (2001). Reappraising medfly longevity: a quantile regression survival analysis. *Journal of the American Statistical Association* **96**, 458-468.

[112] Koenker, R. and Machado, J. (1999). Goodness of fit and related inference processes for quantile regression. *Journal of American Statistical Association* **94**, 1296-1309.

[113] Koenker, R. and Zhao, Q. (1996). Conditional quantile estimation and inference for ARCH models. *Econometric Theory* **12**, 793-813.

[114] Li, C.W. and Li, W.K. (1996). On a double-threshold autoregressive heteroscedastic time series model. *Journal in Applied Econometrics* **11**, 253-274.

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

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

[117] Liesenfeld, R. and Richard, J.-F. (2006). Classical and bayesian analysis of univariate and multivariate stochastic volatility models. *Econometric Reviews* **25**, 335-360.

[118] Lu, Z., Huang, H. and Gerlach, R. (2010). Estimating Value at Risk: from JP Morgan's standard-EWMA to skewed-EWMA forecasting. Working paper, The University of Sydney.

[119] Lunn, D.J., Thomas, A., Best, N. and Spiegelhalter, D. (2000). WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing* **10**, 325-337.

[120] Mandelbrot, B. (1963). The variation of certain speculative prices. *Journal of Business* **36**, 394-419.

[121] McAleer, M. (2005). Automated inference and learning in modeling financial volatility. *Econometric Theory* **21**, 232-261.

[122] McDonald, J.B. (1989). Partially adaptive estimation of ARMA time series models. *International Journal of Forecasting* **5**, 217-230.

[123] McDonald, J.B. and Nelson, R.D. (1989). Alternative beta estimation for the market model using partially adaptive techiniques. *Communications in Statistics - Theory and Methods* **18**, 4039-4058

[124] McDonald, J.B. and Newy, W.K. (1988). Partially adaptive estimation of regression models via the generalized t distribution. *Economic Theory* **4**, 428-457.

[125] Melino, A. and Turnbull, S.M. (1990). Pricing foreign currency options with stochastic volatility. *Journal of Econometrics* **45**, 239-265.

[126] Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. and Teller, E. (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics* **21**, 1087-1092.

[127] Meyer, R. and Yu, J. (2000). BUGS for a Bayesian analysis of stochastic volatility models. *Econometrics Journal* **3**, 198-215.

[128] Montero, J.M., Fernández-avilés, G. and García, M.C. (2010). Estimation of asymmetric stochastic volatility models: application to daily average price of energy products. *International Statistical Review* **78**, 330-347.

[129] Montero, J.M. García, M.C. and Fernández-avilés, G. (2011). On the leverage effect in the Spanish electricity spot market. *Proceedings of the 6th IASME/WSEAS international conference on Energy and Environment.*

[130] Nakajima, J. and Omori, Y. (2009). Leverage, heavy-tails and correlated jumps in stochastic volatility models. *Computational Statistics and Data Analysis* **53**, 2335-2353.

[131] Nakajima, J. and Omori, Y. (2011). Stochastic volatility model with leverage and asymmetrically heavy-tailed error using GH skew Student's t-distribution. *Computational Statistics and Data Analysis*, in press.

[132] Neal, R.M. (1997). Markov chain Monte Carlo methods based on 'slice' the density function. Technical Report No. 9722. Department of Statistics, University of Toronto.

[133] Nelson, D.B. (1991). Conditional heteroskedasticity in asset returns: a new approach. *Econometrics* **59**, 347-370.

[134] Omori, Y., Chib, S., Shephard, N. and Nakajima, J. (2007). Stochastic volatility with leverage: fast and efficient likelihood inference. *Journal of Econometrics* **140**, 425-449.

[135] Panagiotelis, A. and Smith, M. (2008). Bayesian density forecasting of intraday electricity prices using multivariate skew t distributions. *International Journal of Forecasting* **24**, 710-727.

[136] Pandey, G.R. and Nguyen, V.T. (1999). A comparative study of regression based methods in regional flood frequency analysis. *Journal of Hydrology* **225**, 92-101.

[137] Philippe, A. (1997). Simulation of right and left truncated gamma distributions by mixtures. *Statistics and Computing* **7**, 173-181.

[138] Phillips, P.C.B. (1991). To criticise the critics: an objective Bayesian analysis of stochastic trends. *Journal of Applied Econometrics* **6**, 333-364.

[139] Pitt, M. and Shephard, N. (1999). Filtering via simulation: auxiliary particle filter. *Journal of the American Statistical Association* **94**, 590-599.

[140] Ripley, B.D. (1987). *Stochastic Simulation*. Wiley, New York.

[141] Robert, C.P. (1995). Simulation of truncated normal variables. *Statistics and Computing* **5**, 121-125.

[142] Sandmann, G. and Koopman, S.J. (1998). Estimation of stochastic volatility models via Monte Carlo maximum likelihood. Journal of *Econometrics* **87**, 271-301.

[143] Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics* **6**, 461-464.

[144] Shephard, N. (1996). Statistical aspects of ARCH and stochastic volatility. In Cox, D.R., Barnodorff-Nielson, O.E. and Hinkley, D.V., eds, *Time Series Models in Econometrics, Finance and Other Fields*, Chapman & Hall, London, pp. 1-67.

[145] Shephard, N. (2005). *Stochastic Volatility: Selected Readings*. Oxford, Oxford University Press.

[146] Shephard, N.G. and Kim, S. (1994). Comments on 'Bayesian analysis of stochastic volatility models' by Jacquier, Polson and Rossi. *Journal of Business and Economic Statistics* **11**, 406-410.

[147] Shephard, N.G. and Pitt, M.K. (1997). Likelihood analysis of non-gaussian measurement time series. *Biometrika* **84**, 653-667.

[148] Smith, B. (2005). *Bayesian Output Analysis program (BOA) Version 1.1.5*. The University of Iowa.

[149] Smith, M.S. (2010). Bayesian inference for a periodic stochastic volatility model of intraday electricity prices. In: Kneib, T. and Tutz, G., eds. *Statistical Modelling and Regression Structures: Festschrift in Honour of Ludwig Fahrmeir*, Springer, Berlin, pp. 353-376.

[150] Smith, A.F.M. and Robert, G.O. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte-Carlo methods (with discussion). *Journal of the Royal Statistical Society, Series B* **55**, 3-23.

[151] So, M.K.P., Chen C.W.S., Lee J.Y. and Chang Y.P. (2008). An empirical evaluation of fat-tailed distributions in modeling financial time series. *Mathematics and Computers in Simulation* **77**, 96-108.

[152] So, M.K.P., Li, W.K. and Lam, K. (2002). A threshold stochastic volatility model. *Journal of Forecasting* **21**, 473-500.

[153] Solibakke, P. (2002). Efficient esimated mean and volatility characteristics for the Nordic spot electricity spot electricity power market. *International Journal of Business* **7**, 17-35.

[154] Spiegehalter, D., 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**, 583-639.

[155] Steel, M.F.J. (1998). Bayesian analysis of stochastic volatility models with flexible tails. *Econometric Reviews* **17**, 109-143.

[156] Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation (with discussion). *Journal of the American Statistical Association* **82**, 528-550.

[157] Tauchen, G. and Pitts, M. (1983). The price variability-volume relationship on speculative markets. *Econometrica* **51**, 485-505.

[158] Taylor, S.J. (1982). Financial returns modelled by the product of two stochastic processes—a study of daily sugar prices 1961-79. In Anderson, O.D., ed, *Time Series Analysis: Theory and Practice* 1, Amsterdam, North-Holland, pp. 203-226.

[159] Taylor, S.J. (1986). *Modelling Financial Time Series*. John Wiley, Chichester.

[160] Theodossiou, P. (1998). Financial data and the skewed generalised t distribution. *Management Science* **44**, 1650-1661.

[161] Thomas, S. and Mitchell, H. (2005). GARCH modelling of high-frequency volatility in Australia's national electricity market. Discussion paper. Melbourne Centre for Financial Studies.

[162] Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion). *The Annals of Statistics* **22**, 1701-1762.

[163] Tsay, R.S. (2002). *Analysis of financial time series.* John Wiley, New York.

[164] Tse, Y. and Tusi, A. (2002). A multivariate GARCH model with time-varying correlations. *Journal of Business and Economic Statistics* **17**, 351-362.

[165] Walker, S.G. and Gutiérrez-Peña, E. (1999). Robustifying Bayesian procedures. In: Bernardo, J.M., Dawid, A.P. and Smith, A.F.M., eds. *Bayesian Statistics* **6**, Oxford, New York, pp. 685-710.

[166] Wang, J.J.J., Chan, J.S.K. and Choy, S.T.B. (2011a). Stochastic volatility models with leverage and heavy-tailed distributions: A Bayesian approach using scale mixtures. *Computational Statistics and Data Analysis* **55**, 852-862.

[167] Wang, J.J.J., Choy, S.T.B. and Chan, J.S.K. (2011b). Modelling stochastic volatility using generalised t distribution. *Journal of Statistical Computation and Simulation*, in press.

[168] Wichitaksorn, N, Wang, J.J.J., Choy, S.T.B. and Chan, J.S.K. (2011). A new approach for quantile regression and stochastic volatility model with asymmetric laplace distribution. Working paper, The University of Sydney.

[169] Wong, C.K.J. (2002). Latent factor models of high frequency financial time series. Unpublished Ph.D. Thesis. University of Oxford.

[170] Worthington, A.C., Kay-Spratley, A. and Higgs, H. (2005). Transmission of prices and price volatility in Australia electricity spot markets: a multivariate GARCH analysis. *Energy Economics* **2**, 337-350.

[171] Wu, G. (2001). The determinats of asymmetric volatility. *Review of Financial Studies* **14**, 837-859.

[172] Yu, J. (2005). On leverage in a stochastic volatility model. *Journal of Econometrics* **127**, 165-178.

[173] Yu, J. and Meyer, R. (2006). Multivariate stochastic volatility models: Bayesian estimation and model comparison. *Econometric Reviews* **25**, 361-384.

[174] Yu, K., Lu, Z. and Stander, J. (2003). Quantile regression: applications and current research area. *The Statistician* **52**, 331-350.

[175] Yu, K. and Moyeed, R.A. (2001). Bayesian quantile regression. *Statistics and Probability Letters* **54**, 437-447.

[176] Yu, K. and Zhang, J. (2005). A three parameter asymmetric laplace distribution and its extension. *Communications in Statistics - Theory and Methods* **34**, 1867-1879.

[177] Zakoian, J.-M. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and Control* **18**, 931-955.

[178] Zhu, L. and Carlin, B. P. (2000). Comparing hierarchical models for spatio-temporally misaligned data using the deviance information criterion. *Statistics in Medicine* **19**, 2265-2278.