Stochastic Volatility Models and Simulated Maximum Likelihood Estimation

by

Ji Eun Choi

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Statistics

Waterloo, Ontario, Canada, 2011

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Financial time series studies indicate that the lognormal assumption for the return of an underlying security is often violated in practice. This is due to the presence of time-varying volatility in the return series. The most common departures are due to a fat left-tail of the return distribution, volatility clustering or persistence, and asymmetry of the volatility. To account for these characteristics of time-varying volatility, many volatility models have been proposed and studied in the financial time series literature. Two main conditional-variance model specifications are the autoregressive conditional heteroscedasticity (ARCH) and the stochastic volatility (SV) models.

The SV model, proposed by Taylor (1986), is a useful alternative to the ARCH family (Engle (1982)). It incorporates time-dependency of the volatility through a latent process, which is an autoregressive model of order 1 (AR(1)), and successfully accounts for the stylized facts of the return series implied by the characteristics of time-varying volatility. In this thesis, we review both ARCH and SV models but focus on the SV model and its variations. We consider two modified SV models. One is an autoregressive process with stochastic volatility errors (AR–SV) and the other is the Markov regime switching stochastic volatility (MSSV) model. The AR–SV model consists of two AR processes. The conditional mean process is an AR(p) model , and the conditional variance process is an AR(1) model. One notable advantage of the AR–SV model is that it better captures volatility persistence by considering the AR structure in the conditional mean process. The MSSV model consists of the SV model and a discrete Markov process. In this model, the volatility can switch from a low level to a high level at random points in time, and this feature better captures the volatility movement. We study the moment properties and the likelihood functions associated with these models.

In spite of the simple structure of the SV models, it is not easy to estimate parameters by conventional estimation methods such as maximum likelihood estimation (MLE) or the Bayesian method because of the presence of the latent log-variance process. Of the various estimation methods proposed in the SV model literature, we consider the simulated maximum likelihood (SML) method with the efficient importance sampling (EIS) technique, one of the most efficient estimation methods for SV models. In particular, the EIS technique is applied in the SML to reduce the MC sampling error. It increases the accuracy of the estimates by determining an importance function with a conditional density function of the latent log variance at time t given the latent log variance and the return at time $t-1$.

Initially we perform an empirical study to compare the estimation of the SV model using

the SML method with EIS and the Markov chain Monte Carlo (MCMC) method with Gibbs sampling. We conclude that SML has a slight edge over MCMC. We then introduce the SML approach in the AR–SV models and study the performance of the estimation method through simulation studies and real-data analysis. In the analysis, we use the AIC and BIC criteria to determine the order of the AR process and perform model diagnostics for the goodness of fit. In addition, we introduce the MSSV models and extend the SML approach with EIS to estimate this new model. Simulation studies and empirical studies with several return series indicate that this model is reasonable when there is a possibility of volatility switching at random time points. Based on our analysis, the modified SV, AR–SV, and MSSV models capture the stylized facts of financial return series reasonably well, and the SML estimation method with the EIS technique works very well in the models and the cases considered.

Acknowledgements

Foremost, I would like to express my sincere gratitude to my supervisor Prof. Bovas Abraham for the continuous support of my Ph.D study and research, for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis.

I also would like to thank my co-supervisor Prof. Ken Seng Tan for his support, patience, and encouragement. His guidance especially helped me to enlarge my view more on actuarial science. Besides my supervisors, I would like to thank my internal thesis committee: Prof. Adam Kolkiewicz and Prof. Yulia Gel and my external committee: Prof. Nalini Ravishanker and Prof. Ranjini Jha for their encouragement and insightful comments.

My sincere thanks also goes to Ken Mungan and Peter Sun for offering me opportunities on diverse exciting projects and giving me continuous encouragement, insights, and support on my career in the financial risk management practice in Milliman. Working experience with them in Milliman took me into a different level.

Last but not the least, I would like to thank my parents, Woo Young Choi and Myoung Shok Mo, my sister, Il Jung Choi, my sister-in-law, Hyung Tae Kim, and my little brother, Hun Seok Choi, for their unconditional love, support, and encouragement throughout my life. Without them, it means nothing to accomplish my Ph.D degree.

This is dedicated to my parents,
Woo Young Choi and Myoung Shok Mo,
who I love the most.

Contents

List of Tables

List of Figures

Chapter 1

Introduction: Volatility Models and Estimation Methods

Studies on financial time series reveal that changes in volatility occur over time for many classes of assets such as stocks, currencies, and commodities. There are two types of models for time-dependent variances: observation-driven and parameter-driven models. The most well-known example of an observation-driven model is the autoregressive conditional heteroscedasticity (ARCH) model, introduced by Engle (1982). In this model, the conditional variance is taken as a function of the squares of previous observations. An alternative to ARCH is the stochastic volatility (SV) model introduced by Taylor (1982, 1986), in which the conditional variance follows a latent stochastic process.

The SV model can successfully capture the empirical properties of the financial return series: (1) the unconditional distribution of returns is leptokurtic or its kurtosis is larger than that of a normal distribution, (2) volatility clustering is often exhibited, that is, a large movement is followed by a large movement and a small change is followed by a small change, and (3) the squared returns exhibit serial autocorrelations whereas there is little or no serial correlation in the return series itself. The SV model is also related to a continuous-time diffusion process that is widely used in option pricing (*see* Hull and White (1987)).

Maximum likelihood estimation (MLE) is often used to estimate the parameters in ARCH models, but it cannot be directly applied to parameter estimation in the SV model because of the existence of a latent process. To efficiently estimate parameters and volatility in the

SV model, various estimation methods have been proposed in the literature.

In this chapter, we describe ARCH and SV models in Sections 1.1 and 1.2. We introduce some of the estimation methods for SV models and discuss their advantages and disadvantages in Section 1.3.

1.1 ARCH Models

The basic idea of ARCH models is that the return of any underlying security y_t is serially uncorrelated but dependent, and the dependence of y_t is described through the conditional variance which is taken as a simple quadratic function of its lagged values. Throughout Chapter 1, y_t is referred to as the mean-corrected return of a security.

Let p_t be a price of an underlying asset at time t; its mean-corrected return is expressed as

$$y_t = 100 \left[ln \left(\frac{p_t}{p_{t-1}} \right) - \frac{\sum_{t=1}^{T} ln \left(\frac{p_t}{p_{t-1}} \right)}{T} \right].$$

Typically, the ARCH(q) model is defined by

$$\begin{aligned} y_t &= \sqrt{\lambda_t} u_t, \\ \lambda_t &= \alpha_0 + \sum_{i=1}^{q} \alpha_i y_{t-i}^2, \end{aligned} \tag{1.1}$$

where $\{u_t\}$ is a sequence of identically and independently distributed (iid) random variables with mean zero and variance one. It is often assumed to follow the standard normal or a heavy-tailed distribution such as a Student-t distribution. $\alpha_0 > 0$, $\alpha_i \geq 0$ $(i = 1, ..., q)$, and $0 < \sum_{i=1}^{q} \alpha_i < 1$ is a necessary and sufficient condition for a weakly stationary process $\{y_t\}$. The order of q determines the volatility persistence, which increases with the value of q. Equation (1.1) implies that large values of squared past observations lead to a large variance λ_t of the return series y_t, and a small movement in squared past returns leads to a small movement in λ_t. This indicates that volatility clustering can be captured by the ARCH structure.

To study further properties of the ARCH process, consider the ARCH(1) model, given

by

$$\begin{aligned} y_t &= \sqrt{\lambda_t} u_t, \qquad u_t \sim iid\ \mathrm{N}(0,1), \\ \lambda_t &= \alpha_0 + \alpha_1 y_{t-1}^2, \end{aligned}$$

where $\alpha_0 > 0$ and $0 < \alpha_1 < 1$. Then, the unconditional moments for y_t are

$$\begin{aligned} E[y_t^2] &= E[\lambda_t u_t^2] = \frac{\alpha_0}{1-\alpha_1}, \\ E[y_t^4] &= E[\lambda_t^2 u_t^4] = 3\left(\frac{\alpha_0^2 + 2\alpha_0\alpha_1 E[y_t^2]}{1-3\alpha_1^2}\right). \end{aligned}$$

Thus, the kurtosis of y_t is given by

$$\kappa = \frac{E[y_t^4]}{E[y_t^2]^2} = 3\left(\frac{1-\alpha_1^2}{1-3\alpha_1^2}\right). \tag{1.2}$$

The value of the kurtosis in (1.2) is greater than three if $\alpha_1^2 < 1/3$. This implies that the ARCH(1) model can have a heavier tail than the normal distribution.

In spite of its nice properties, the ARCH model has some limitations:

1. Regardless of the sign of a shock, the effect on the volatility is the same because it depends on the square of past shocks. In practice, this assumption is frequently violated, particularly by stock returns: the volatility increases more after negative shocks than after positive shocks. This is the so-called *leverage effect*, first introduced by Black (1976).

2. When there are other sources of variation in financial time series such as exogenous factors in the economy, the model does not fit the data well.

3. The model is likely to overestimate the volatility because it responds slowly to large isolated shocks in the return series.

To overcome these weaknesses, there are many extensions of the ARCH model. The most important extended model, proposed by Bollerslev (1986), is the generalized ARCH (GARCH) model. Other extensions are an integrated GARCH (IGARCH) model (Bollerslev and Engle (1993)) and a fractional integrated ARCH model (Ding, Granger, and Engle (1993)), which account for the characteristic that volatility tends to change quite slowly,

with the effect of shocks taking a reasonably long time to decay. To account for the asymmetry of the volatility to shocks, the exponential GARCH (EGARCH) and threshold ARCH (TARCH) models were proposed by Nelson (1991) and Zakoian (1994), respectively. Engle, Lilien, and Robins (1987) proposed the ARCH-in-mean model to incorporate a risk/reward trade-off in the ARCH model. These extended models and their advantages and disadvantages are described below.

GARCH model

The GARCH model merely includes a moving average part in the ARCH model specification, and can be written as an autoregressive moving average (ARMA) model in the squares. The GARCH(p, q) model (Bollerslev (1986)) is defined by

$$\begin{aligned} y_t &= \sqrt{\lambda_t} u_t, \\ \lambda_t &= \alpha_0 + \sum_{i=1}^{q} \alpha_i y_{t-i}^2 + \sum_{j=1}^{p} \beta_j \lambda_{t-j}, \end{aligned} \tag{1.3}$$

where $\{u_t\}$ is a sequence of *iid* random variables with mean zero and unit variance. For stationarity, it is assumed that $\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$, and $\sum_{l=1}^{max(p,q)}(\alpha_l + \beta_l) < 1$. The constraint on $(\alpha_l + \beta_l)$ implies that the unconditional variance of $\{y_t\}$ is finite, but its conditional variance λ_t varies over time.

For simplicity, let us consider the GARCH model of order (1,1), denoted GARCH(1,1), with a normal innovation u_t, given by

$$\begin{aligned} y_t &= \sqrt{\lambda_t} u_t, \\ \lambda_t &= \alpha_0 + \alpha_1 y_{t-1}^2 + \beta_1 \lambda_{t-1}, \end{aligned} \tag{1.4}$$

where $\alpha > 0$, $\alpha_1, \beta_1 \geq 0$, and $\alpha_1 + \beta_1 < 1$. From Eq. (1.4), it can be easily seen that a large y_{t-1}^2 or λ_{t-1} gives rise to a large λ_t, and a small y_{t-1}^2 or λ_{t-1} leads to a small λ_t. This is volatility clustering.

Another nice property of the GARCH(1,1) model is obtained by letting $\eta_t = y_t^2 - \lambda_t$, which is a martingale difference (MD). That is, $E[|\eta_t|] < \infty$ and $E[\eta_t | \mathcal{F}_{t-1}] = 0$, which implies $E[\eta_t] = 0$ and $Cov(\eta_t, \eta_{t-j}) = 0$ for $j \geq 1$. However, $\{\eta_t\}$ is not an *iid* sequence in

the GARCH(1,1) model. Since $\eta_{t-1} = y_{t-1}^2 - \lambda_{t-1}$, we can rewrite Eq. (1.4) as

$$\begin{aligned} y_t^2 &= \alpha_0 + (\alpha_1 + \beta_1) y_{t-1}^2 - \beta_1 \eta_{t-1} + \eta_t, \\ \Phi(B) y_t^2 &= \alpha_0 + \beta(B) \eta_t, \end{aligned}$$

where $\Phi(B) = \{1 - (\alpha_1 + \beta_1)B\}$, $\beta(B) = \{1 - \beta_1 B\}$, and B is a backshift operator. This is an ARMA(1,1) process for the squared returns. In general, a GARCH(p, q) model can be written as an ARMA(p, q) process for y_t^2.

Using the unconditional mean of an ARMA(1,1) model, the first moment of the squared y_t is given by

$$E[y_t^2] = \frac{\alpha_0}{1 - (\alpha_1 + \beta_1)}, \tag{1.5}$$

provided $(\alpha_1 + \beta_1) < 1$. Equation (1.5) provides a simple parametric function to describe the volatility evolution because $E[\lambda_t] = E[y_t^2]$. For any GARCH (p, q) model, we can find a parametric form of $E[y_t^2]$.

The kurtosis of the return y_t of the GARCH(1,1) model is

$$\frac{E[y_t^4]}{(E[y_t^2])^2} = \frac{3[1 - (\alpha_1 + \beta_1)^2]}{1 - (\alpha_1 + \beta_1)^2 - 2\alpha_1^2}. \tag{1.6}$$

If $(\alpha_1 + \beta_1)^2 - 2\alpha_1^2 < 1$, then the value of the kurtosis in (1.6) is greater than three, which indicates that the tail of the GARCH(1,1) model is heavier than that of a normal distribution.

Forecasts of volatility in the GARCH(1,1) model can be obtained using methods similar to those of the ARMA process. Assume that the forecast origin is time h. That is, y_h and λ_h are known at time h. For the 1-*step* ahead forecast function, we have

$$\lambda_{h+1} = \alpha_0 + \alpha_1 y_h^2 + \beta_1 \lambda_h,$$

and thus the 1-*step* ahead forecast is

$$\lambda_h(1) = \alpha_0 + \alpha_1 y_h^2 + \beta_1 \lambda_h.$$

For the 2-*step* ahead forecast, using $y_t^2 = \lambda_t u_t^2$, Eq. (1.4) can be written as

$$\lambda_{t+1} = \alpha_0 + (\alpha_1 + \beta_1)\lambda_t + \alpha_1 \lambda_t (u_t^2 - 1).$$

When $t = h + 1$, this equation becomes

$$\lambda_{h+2} = \alpha_0 + (\alpha_1 + \beta_1)\lambda_{h+1} + \alpha_1\lambda_{h+1}(u_{h+1}^2 - 1).$$

Thus, the 2-*step* ahead forecast, originated at time h, is given by

$$\lambda_h(2) = \alpha_0 + (\alpha_1 + \beta_1)\lambda_h(1)$$

because $E[u_{h+1}^2 - 1|\mathcal{F}_h] = 0$. In general, the l-*step* ahead volatility forecast can be given, for $l > 1$, by

$$\begin{aligned}\lambda_h(l) &= \alpha_0 + (\alpha_1 + \beta_1)\lambda_t(l-1) \\ &= \frac{\alpha_0[1 - (\alpha_1 + \beta_1)^{l-1}]}{1 - \alpha_1 - \beta_1} + (\alpha_1 + \beta_1)^{l-1}\lambda_h(1).\end{aligned}$$

Therefore, provided $\alpha_1 + \beta_1 < 1$,

$$\lambda_h(l) \ \to \ \frac{\alpha_0}{1 - \alpha_1 - \beta_1} \quad \text{as} \quad l \to \infty.$$

This implies that the multistep-ahead volatility forecasts of a GARCH(1,1) model converge to the unconditional variance of y_t as the time horizon increases, assuming the existence of the variance of y_t. Nevertheless, the model has the same weaknesses as the ARCH model. For example, the *leverage effect* is still not taken into account and the tail behavior of the GARCH model is still short for the high-frequency data, even with the Student-t innovations (Engle and Bollerslev (1986) and Nelson (1991)).

Integrated GARCH

The integrated GARCH (IGARCH) model is a unit-root GARCH model. If the AR polynomial in (1.3) has a unit root, the GARCH model becomes the IGARCH model. In the IGARCH model, the impact of past squared shocks ($\eta_t = y_{t-i}^2 - \lambda_{t-i}$) on the squared return y_t^2 is persistent, which is similar to the behavior of an ARIMA model. IGARCH(p, q) (Bollerslev and Engle (1993)) is defined as

$$\begin{aligned}y_t &= \sqrt{\lambda_t}u_t, \\ \lambda_t &= \alpha_0 + \sum_{i=1}^{q} \alpha_i y_{t-i}^2 + \sum_{j=1}^{p} \beta_j \lambda_{t-j},\end{aligned}$$

where u_t is a Gaussian white noise and for stationarity we assume that $\alpha_i \geq 0$, $\beta_j \geq 0$ for $i = 1, ..., q$ and $j = 1, ..., p$ and $\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j = 1$.

Exponential GARCH

The exponential GARCH (EGARCH) model was proposed by Nelson (1991). In this model, a weight innovation is considered to allow for asymmetric effects between positive and negative shocks in return series.

The EGARCH(p, q) (Nelson (1991)) model is defined by

$$\begin{aligned} y_t &= \sqrt{\lambda_t} u_t, \\ \log \lambda_t &= \alpha_0 + \sum_{i=1}^{q} \alpha_i \log \lambda_{t-i} + \sum_{j=1}^{p} \beta_j g(u_{t-j}), \end{aligned} \tag{1.7}$$

where

$$g(x) = \omega x + \delta(|x| - E|x|) \tag{1.8}$$

and α_i and β_j are constant for $i = 0, 1, ..., q$ and $j = 1, ..., p$. The function $g(\cdot)$ in (1.8) is piecewise linear and contains both the size and sign effects of the shocks on volatility. The term ωx determines the sign effect and the term $\delta(|x| - E|x|)$ the size effect. Typically, the parameter ω is positive and δ is negative. Moreover, since u_t is iid, so is $g(u_t)$. Since u_t and $u_t - E|u_t|$ are zero-mean iid sequences, it can be seen that $E[g(u_t)] = 0$.

Similarly to the GARCH model, the unconditional mean of $\log \lambda_t$ is $\frac{\alpha_0}{1-\sum \alpha_i}$. However, there are some differences: first, the coefficients are not required to be positive because we take the log of the variance. Second, we account for the asymmetric effect by adding a function $g(\cdot)$ in the process $\log\{\lambda_t\}$. The asymmetry of information is potentially useful because it allows the variance to respond more rapidly to drops in the market than to corresponding rises. This is important for many equities (Schwert (1989), Sentana (1995), Campbell and Hentschel (1992)).

Consider the simplest EGARCH(1,1) model where $\beta_1 = 1$, given by

$$\begin{aligned} y_t &= \sqrt{\lambda_t} u_t, \\ (1 - \alpha_1 B) \log \lambda_t &= \alpha_0 + g(u_{t-1}), \end{aligned}$$

where B is a backshift operator and u_t is a Gaussian white noise so that $E[|u_t|] = \sqrt{2/\pi}$. A function $g(\cdot)$ is defined via (1.8) and has a zero mean and a constant variance so that $\log \lambda_t$ is a stationary autoregression provided $|\alpha_1| < 1$. Setting $\alpha_* = \alpha_0 - \delta\sqrt{2/\pi}$, the model becomes

$$(1 - \alpha_1 B) \log \lambda_t = \begin{cases} \alpha_* + (\omega + \delta) u_{t-1} & \text{if } u_{t-1} \geq 0, \\ \\ \alpha_* + (\omega - \delta) u_{t-1} & \text{if } u_{t-1} < 0. \end{cases}$$

This is a nonlinear function of λ_t so that it suffices to say that for the EGARCH(1,1) model with $\beta_1 = 1$, the conditional variance λ_t evolves in a nonlinear manner depending on the sign of y_{t-1}. In particular, the above equation can be rewritten as

$$\lambda_t = \lambda_{t-1}^{\alpha_1} \exp(\alpha_*) * \begin{cases} \exp\left[(\omega + \delta) \frac{y_{t-1}}{\sqrt{\lambda_{t-1}}}\right] & \text{if } y_{t-1} \geq 0, \\ \\ \exp\left[(\omega - \delta) \frac{y_{t-1}}{\sqrt{\lambda_{t-1}}}\right] & \text{if } y_{t-1} < 0. \end{cases}$$

The coefficients $\omega + \delta$ and $\omega - \delta$ measures the asymmetric effects on positive and negative shocks, respectively. Furthermore, in contrast to ARCH, this model allows the autocorrelation function of the squared returns to be negative so that EGARCH provides oscillations in the autocorrelation function of the squares.

Threshold ARCH (TARCH) model

The idea of the TARCH model is to divide the distribution of the innovations into two disjoint intervals and then approximate a piecewise linear function for the conditional standard deviation (Zakoian (1994)) and the conditional variance (Glosten, Jagannathan, and Runkle (1993)). That is, the model distinguishes the effect of negative and positive movements of the return on the volatility. If a threshold is zero, the TARCH(q) model in Glosten *et al.* (1993) is of the form

$$\lambda_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i^+ \, y_{t-i}^2 \, I(y_t > 0) + \sum_{i=1}^{q} \alpha_i^- \, y_{t-i}^2 \, I(y_t \leq 0),$$

where α_i^+ is a coefficient when $I(y_t > 0)$ and α_i^- is a coefficient when $I(y_t \leq 0)$.

ARCH-in-mean model

In economic theory, an investor should have a reward for taking risk. This is known as the risk/reward trade-off. To reflect this, Engle *et al.* (1987) proposed the ARCH-in-mean (ARCH-M) model. This model provides an explicit link between the expected return of a risky asset and the level of volatility. The ARCH-M model (Engle *et al.* (1987)) is given by

$$\begin{aligned} y_t &= g(\lambda_t^2, \theta) + u_t \lambda_t, \\ \lambda_t^2 &= \alpha_0 + \alpha_1 \{y_{t-1} - g(\lambda_t^2, \theta)\}^2, \end{aligned}$$

where $g(\lambda_t^2, \theta)$ is typically linear, for example, $g(\lambda_t^2, \theta) = \alpha_0 + \alpha_1 \lambda_t^2$. The statistical properties have been studied by Hong (1991).

1.2 Stochastic Volatility Model

An alternative to ARCH models is to allow λ_t to depend on some unobserved structure. The most popular parameter-driven model is the stochastic volatility (SV) model, introduced by Taylor (1986). Its standard form can be written

$$y_t = \exp\{\lambda_t/2\} u_t, \tag{1.9}$$

$$\lambda_t = \alpha + \beta \lambda_{t-1} + \gamma v_t, \tag{1.10}$$

$$\lambda_1 \sim N\left(\frac{\alpha}{1-\beta}, \frac{\gamma^2}{1-\beta^2}\right), \tag{1.11}$$

for $t = 1, ..., T$. The error processes u_t and v_t are assumed to be *iid* with zero mean and unit variance, and v_t is normally distributed. Both processes are serially and mutually uncorrelated at all lags, and uncorrelated with λ_1. The latent process λ_t is a stationary AR(1) process with a persistent parameter β if $|\beta| < 1$ and represents the log-variance of the observed returns y_t. Its physical interpretation is the random and uneven flow of new information, which is difficult to model directly for financial markets. $\gamma > 0$ is the volatility of the log-variance process λ_t. Conditional on λ_t, the return series y_t is simply a heteroscedastic Gaussian noise.

Typically, the process u_t is a Gaussian white noise, but it has been empirically shown that the heavy-tailed assumption of the process, such as the Student-t distribution (Ruiz (1994), Harvey, Ruiz, and Shephard (1994), Sandmann and Koopman (1998), and Chib,

Nardari, and Shephard (1998)), and a generalized error distribution can capture the empirical regularities of financial data (Liesenfeld and Jung (2000)). In Chapter 2 we discuss the statistical properties of the SV models: a model with normal error (SV-normal), a model with Student-t distributed error (SV-t), and with a model with generalized-error distributed error (SV-GED).

1.3 Estimation Methods

In the ARCH models discussed earlier, the parameter estimation is straightforward. Maximum likelihood estimation (MLE) is commonly used. However, this is difficult for the SV model because of the latent process λ_t. We will now discuss several estimation methods used in the parameter estimation for the SV model.

For the SV model in (1.9) and (1.10), the likelihood associated with known observations $Y_T = \{y_t\}_{t=1}^T$ and latent variables $\Lambda_T = \{\lambda_t\}_{t=1}^T$ is given by

$$L(\theta; Y_T) = \int_{R^T} f(Y_T, \Lambda_T|\theta)\, d\Lambda_T, \tag{1.12}$$

where θ denotes the vector of parameters α, β, and γ to be estimated, and $f(\cdot)$ is a probability density. Despite the simplicity of the model specification, an analytical solution of the integral in (1.12) is not available, and simulations or numerical methods must be used to estimate the integral.

There are numerous ways of performing this estimation. Some approaches use a method of moments such as the generalized method of moments (GMM)(Hansen (1982); Melino and Turnbull (1990)), popular in econometrics. Others approximate the likelihood using, for example, quasi maximum likelihood (QML) (Harvey *et al.* (1994)). However, both GMM and QML have some drawbacks. For example, the selection of the moment conditions is not trivial in GMM, and QML is computationally intensive. More elaborate methods have also been proposed, such as the simulated maximum likelihood (SML) method (Danielsson and Richard (1993)), the efficient method of moments (EMM) (Gallant, Hsieh, and Tauchen (1997); Andersen, Chung, and Sorensen (1999)), and Markov chain Monte Carlo (MCMC) (Jacquier, Polson, and Rossi (1994); Kim, Shephard, and Chib (1998)), which strongly dominates GMM, QML, and EMM. SML is a good alternative to MCMC in the sense that it is as efficient as MCMC and its implementation is fairly straightforward. Danielsson (1994) showed that the finite sample properties of the ML estimator based on

importance sampling are identical to those of the Bayesian posterior means evaluated by MCMC (Jacquier *et al.* (1994)). Liesenfeld and Jung (2003) introduced the efficient importance sampling (EIS) technique in the context of SML and later developed (*see* Liesenfeld and Richard (2008)) an improved MCMC method based on the Metropolis Hastings procedure using the EIS (MH-MCMC-EIS). Ozturk (2009) implemented the HM-MCMC-EIS method for EGARCH and SV models. Since our focus is more on a stand-alone method such as GMM, EMM, SML, or MCMC. We will discuss those stand-alone methods in the next subsections.

1.3.1 Generalized Method of Moments

When the method of moments (MM) is used to estimate parameters in the SV model, many possible moments could be used. In contrast to MM, GMM is designed to consider all the moment restrictions in a reasonably sensible way, for example by weighting the moment conditions and exploiting the convergence of selected sample moments to their unconditional expected values.

In general, the GMM procedure can be described as follows. Let $Y_T = \{y_1, ..., y_T\}$ be a financial time series and let $m_t(\theta) = (m_{1t}(\theta), ..., m_{Qt}(\theta))$ denote the vector of the sample realizations of the moments at time t. The constant Q can be larger than the dimension of θ, which is a parameter vector to be estimated. Then, the corresponding sample moments $M_{iT}(\theta)$ are given by

$$M_{iT}(\theta) = \frac{1}{T-j} \sum_{t=j+1}^{T} m_{it}(\theta), \qquad i = 1, ..., Q,$$

where j is the maximum lag between variables defining the sample moments. The vector of the analytical moments is denoted $A(\theta) = (A_1(\theta)), ..., A_Q(\theta))$. Then, the GMM estimator $\hat{\theta}$ minimizes the distance between $M_T(\theta) = (M_{iT}(\theta), ..., M_{QT}(\theta))$ and $A(\theta)$:

$$\hat{\theta} = \text{argmin}_\theta[(M_T(\theta) - A(\theta))' \, \Sigma^{-1} \, (M_T(\theta) - A(\theta))],$$

where the matrix Σ is an arbitrary positive definite weighting matrix. Hansen (1982) showed that the GMM estimator $\hat{\theta}$ is consistent and asymptotically normal, *i.e.*, $\sqrt{T}(\hat{\theta} - \theta_0) \sim N(0, \Omega)$, where θ_0 is the true parameter vector and Ω is an asymptotic covariance matrix.

Given the properties of $\hat{\theta}$, the optimal choice of the weighting matrix Σ is the one that minimizes the asymptotic covariance matrix Ω. It can be given by the covariance matrix of the standardized moment conditions:

$$\Sigma = \lim_{T\to\infty} E\left[\sum_{t,\tau=1}^{T}(m_t(\theta_0) - A(\theta_0))(m_\tau(\theta_0) - A(\theta_0))'/T\right].$$

This matrix can be estimated by a kernel estimator for the density of the vector of sample moments.

For simplicity of the moment conditions, let us consider the SV model obtained by setting $\lambda_t = \log\sigma_t^2$, given by

$$\begin{aligned} y_t &= \sigma_t u_t, \\ \log\sigma_t^2 &= \alpha + \beta\log\sigma_{t-1}^2 + \gamma v_t. \end{aligned}$$

The error u_t is assumed to be standard normal. The other assumptions given in the standard SV model in (1.9) and (1.10) also hold for this model. For $k = 1, ..., K$, the first few moments are

$$\begin{aligned} E[y_t^2] = E[\sigma_t^2], \qquad & E[y_t^4] = 3E[\sigma_t^4], \\ E[|y_t|] = \sqrt{2/\pi}E[\sigma_t], \qquad & E[|y_t^3|] = 2\sqrt{2/\pi}E[\sigma_t^3], \\ E[|y_t\ y_{t-k}|] = \frac{2}{\pi}E[\sigma_t\ \sigma_{t-k}], \qquad & E[y_t^2\ y_{t-k}^2] = E[\sigma_t^2\ \sigma_{t-k}^2], \end{aligned} \tag{1.13}$$

where, for any positive constants r and s and any positive integer k,

$$\begin{aligned} E[\sigma_t^r] &= \exp[r\mu/2 + r^2\sigma^2/8], \\ E[\sigma_t^r\sigma_{t-k}^s] &= E[\sigma_t^r]E[\sigma_t^s]\exp[rs\beta^k\sigma^2/4]. \end{aligned}$$

Note that $\mu = \frac{\alpha}{1-\beta}$ and $\sigma^2 = \frac{\gamma^2}{1-\beta^2}$ are the unconditional mean and variance of the log-variance process λ_t. For moment conditions, only lower-order moments are important because of the erratic behavior of higher-order moments and the requirement of asymptotic normality of $\hat{\theta}$. The variances of the higher-order moments are usually large because of the presence of the fat tail in the return series. The asymptotic normality of the GMM estimator $\hat{\theta}$ requires finite variances of the moment conditions and good estimates of these quantities in finite samples.

It should be noted that sample moments of the return series are likely to be heavily

correlated and serially dependent; this influences the choice of an appropriate weighting matrix for the SV model. If this factor is ignored, suitable efficiency cannot be obtained. For example, when Andersen and Sorensen (1996) took an identity matrix as a weighting matrix, they found that the GMM behaved badly and sometimes did not converge. Hence, some preliminary scaling of the moments through a weighting matrix is required for meaningful inference via GMM.

There are some drawbacks to this method:

1. GMM can only be applied when λ_t is stationary. That is, as the persistent parameter β gets close to unity, GMM works poorly for many high-frequency financial time series.

2. The parameter estimates are not invariant, *i.e.*, if the model is reparameterized as $\tau = f(\theta)$, then $\hat{\tau} \neq f(\hat{\theta})$.

3. Under the normal assumption for u_t, the autocorrelation function of y_t in the SV model at lag k is given by

$$\rho(k) \approx \frac{\exp\{\sigma^2\} - 1}{3\exp\{\sigma^2\} - 1}\beta^k,$$

   where $\sigma^2 = \frac{\gamma^2}{1-\beta^2}$. The derivation of $\rho(k)$ is given in Chapter 2. This indicates that if σ^2 is small, typically the case in practice, the autocorrelation $\rho(k)$ is small but positive for many k. Consequently, the number of moment equations to be considered must be high to capture the low but persistent ACF of the squares, which increases the computational burden.

4. GMM itself cannot provide estimates of the volatility so another estimation technique is required for volatility estimation.

5. Conventional tests are not available after fitting the model.

1.3.2 Efficient Method of Moments

EMM seeks to improve the efficiency of GMM, while maintaining its general flexibility. Instead of using the exact moment conditions, EMM takes as a moment condition a score function from a pseudo (or auxiliary) model rather than the true model. Hence, the key

issue of EMM is to carefully select moment conditions with the likelihood of an auxiliary model, depending on the characteristics of the observations. It then leads to an efficient estimation via the standard GMM procedure.

EMM is particularly useful when the ML approach is not feasible or is computationally intensive, as is the case in the SV model. The idea underlying EMM is to match the efficiency of the ML estimation with the flexibility of the GMM procedure. Thus, the efficiency of EMM is likely to fall between that of GMM and likelihood-based inference.

We first introduce some notation. The process that generates the data Y is referred to as a structural model. In the SV model, (1.9) and (1.10) define the structural model. The parameter vector to be estimated in the structural model, such as θ in the SV model, is called a structural parameter vector. The vector of parameters in an auxiliary model is called an auxiliary parameter vector, denoted η. Then, the EMM procedure can be summarized in the following steps:

Step 1: Choice of an auxiliary model and its estimation

The auxiliary model should approximate the structural (true data-generating) model as closely as possible and allow the maximum likelihood to be applicable. Let the auxiliary model be specified as $f(y_t|Y_{t-1}, \eta)$, where $Y_{t-1} = \{y_{t-1}, ..., y_1\}$ and η is the auxiliary or pseudo parameter vector whose dimension is greater than or equal to that of θ. Then, the estimated η is obtained by the quasi maximum likelihood (QML) estimation method, and the QML estimator $\hat{\eta}$ satisfies

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\partial}{\partial\eta}\log f(y_t|Y_{t-1},\hat{\eta}) = \frac{1}{T}\sum_{t=1}^{T}S(y_t,\hat{\eta}) = 0, \tag{1.14}$$

where $S(y_t, \hat{\eta}) = \frac{\partial}{\partial\eta}\log f(y_t|Y_{t-1}, \hat{\eta})$ is the score generator corresponding to the auxiliary model. The advantage of QML is that even if the auxiliary model is misspecified, under some regularity conditions, the QML estimator $\hat{\eta}$ is consistent (White (1994)).

An auxiliary model could be a (full) parametric model or a semi-parametric model (SNP) (Gallant and Nychka (1987)). Although the choice of the auxiliary model is important for the estimation performance, there is no systematic approach to this. Thus, we need to parametrize a potential auxiliary model in several different ways and investigate the resulting efficiency. Because of the conditional heteroscedasticity of the SV model, ARCH models are popular auxiliary models for a fully parameterized model, while the SNP mod-

els vary (*see* Anderson *et al.* (1999)).

Step 2: Setting up of the moment conditions and computation of the estimated moments

For an initial value of a structural parameter θ, we generate the new series $Y_{new}^{(1)}(\theta), ..., Y_{new}^{(N)}(\theta)$ for a large $N > T$ using the structural model. Let the sample moments be

$$m(\theta, \hat{\eta}) = \frac{1}{N} \sum_{i=1}^{N} S(Y_{new}^{(i)}(\theta), \hat{\eta}),$$

where $Y_{new}^{(i)}(\theta) = \{y_{new,1}^{(i)}(\theta), y_{new,2}^{(i)}(\theta), ..., y_{new,T}^{(i)}(\theta)\}$. Then, the estimated moment evaluated at $\hat{\eta}$ is calculated by

$$\hat{m}(\theta, \hat{\eta}) = \frac{1}{N} \sum_{i=1}^{N} \prod_{t=1}^{T} S(y_{new,t}^{(i)}(\theta) | y_{new,1}^{(i)}(\theta), ..., y_{new,t-1}^{(i)}(\theta), \hat{\eta}),$$

where $\hat{\eta}$ is an auxiliary parameter estimator.

Step 3: Computation of the parameter estimates

The estimator of θ is a solution of the following minimization problem:

$$\arg\min_{\theta} [\hat{m}(\theta, \hat{\eta})' \, \hat{V}^{-1} \, \hat{m}(\theta, \hat{\eta})],$$

where $\hat{V}$ denotes a consistent estimator of the asymptotic covariance matrix V of the sample score vector. If an appropriate auxiliary model is selected, the likelihood inference implies that the score vector consists of a (near) MD sequence, and an estimator of the information matrix $\hat{V}$ is given by

$$\hat{V} = \frac{1}{T} \sum_{t=1}^{T} S(y_t, \hat{\eta}) S(y_t, \hat{\eta})'.$$

1.3.3 Simulated Maximum Likelihood

The simulated maximum likelihood (SML) method was introduced by Danielsson and Richard (1993); it depends on Monte Carlo (MC) integration to evaluate the likelihood

in (1.12). After estimating the likelihood by MC simulation, SML seeks values of the parameter vector θ that maximize the estimated likelihood.

Danielsson and Richard (1993) first implemented SML to estimate the SV model. It works poorly because the latent variable λ_t is drawn from the density without any information on the return series y_t, and many iterations are required to achieve convergence. To resolve this inefficiency issue, one of the variance reduction methods for MC simulation, called importance sampling (IS) (Ripley (1987)), has been applied by several researchers. The two variants used are an efficient importance sampling (EIS) technique (Liesenfeld and Richard (2003)) and an alternative approach, proposed by Shephard and Pitt (1997), Durbin and Koopman (1997, 2000), and Sandmann and Koopman (1998). These methods are briefly discussed below along with a crude MC simulation.

Crude Monte Carlo Simulation

The likelihood in (1.12) can be rewritten as

$$L(\theta; Y_T) = \int_{R^T} f(Y_T, \Lambda_T|\theta)\, d\Lambda_T = \int_{R^T} f(Y_T|\Lambda_T, \theta) f(\Lambda_T|\theta)\, d\Lambda_T. \tag{1.15}$$

This is an expectation of the conditional density $f(Y_T|\Lambda_T)$ under the distribution of Λ_T. Throughout this thesis, $f(\cdot)$ denotes a probability density function. In the crude MC simulation, we first generate N replications $\Lambda_T^{(i)}$ from the unconditional density $f(\Lambda_T|\theta)$, which is given by

$$f(\Lambda_T|\theta) = \prod_{t=1}^{T} f(\lambda_t|\lambda_{t-1}, \theta). \tag{1.16}$$

Then, the estimated likelihood can be obtained by

$$\hat{L}(\theta; Y_T) = \frac{1}{N} \sum_{i=1}^{N} f(Y_T|\Lambda_T^{(i)}),$$

where the $\Lambda_T^{(i)}$ are drawn from (1.16). However, in the context of the SV model, the replications $\Lambda_T^{(i)}$ generated from $f(\Lambda_T|\theta)$ in (1.16) do not resemble the process λ_t under which the observed returns Y_T were obtained. Thus, a large number N is required to achieve a certain accuracy with the crude MC simulation, and this is quite burdensome.

To overcome this problem, IS can be applied (Ripley (1987)). The idea underlying IS

is to take a "good" probability density function, called the importance function (IF), instead of the original probability density function. The IF is good in the sense that its distribution is known and it behaves similarly to the original density function.

In general, an IF can be any function and need not be a density function. If the IF is not a density function, the integrand must be multiplied by a quantity that makes the integral equal to one. One way to choose a good IF is the efficient importance sampling (EIS) method described below.

Efficient Importance Sampling

The EIS technique (Liesenfeld and Richard (2003)) chooses in the following way an IF that can provide a good approximation of $f(Y_T, \Lambda_T|\theta)$.

Consider the following factorization of the likelihood function in (1.15) using an IF $g(\Lambda_T|Y_T, \theta)$:

$$\begin{aligned} L(\theta; Y_T) &= \int_{R^T} f(Y_T|\Lambda_T, \theta) f(\Lambda_T|\theta)\, d\Lambda_T \qquad (1.17) \\ &= \int_{R^T} \frac{f(Y_T|\Lambda_T, \theta) f(\Lambda_T|\theta)}{g(\Lambda_T|Y_T, \theta)} g(\Lambda_T|Y_T, \theta)\, d\Lambda_T \\ &= \int_{R^T} h(\Lambda_T, Y_T|\theta) g(\Lambda_T|Y_T, \theta)\, d\Lambda_T \qquad (1.18) \end{aligned}$$

where $h(\Lambda_T, Y_T|\theta) = f(Y_T|\Lambda_T, \theta) f(\Lambda_T|\theta)/g(\Lambda_T|Y_T, \theta)$ is a remainder function (RF). To find an IF $g(\Lambda_T|Y_T, \theta)$ that depends on both λ_T and Y_T, EIS considers a particular factorization:

$$g(\Lambda_T|Y_T, \theta) = \prod_{t=1}^{T} f(\lambda_t|\Lambda_{t-1}, Y_{t-1}, \theta, e_t), \qquad (1.19)$$

where e_t is an auxiliary parameter that depends on the observations Y_T. Then, we find a function $k(\Lambda_t; e_t)$ such that

$$g(\Lambda_T|Y_T, \theta) = \prod_{t=1}^{T} C(\lambda_{t-1}, e_t) k(\Lambda_t; e_t), \qquad (1.20)$$

where $C(\lambda_{t-1}, e_t) = \frac{1}{\int k(\Lambda_t; e_t) d\lambda_t}$ is an integral constant that makes $k(\Lambda_t; e_t)$ a density function. For ease of notation, we remove θ from the functional notations of C and k.

Moreover, the choice of $C(\lambda_{t-1}, e_t)$ plays an important role in matching $g(\Lambda_T|Y_T, \theta)$ to the joint density $f(Y_T, \Lambda_T|\theta) = \prod_{t=1}^{T} f(y_t, \lambda_t|\Lambda_{t-1}, Y_{t-1}, \theta)$ as closely as possible. Thus, the estimator of e_t is obtained by solving the least-squares problem:

$$\hat{e}_t = \arg\min_{e_t} \sum_{i=1}^{N} \Big(\log[f(y_t, \lambda_t^{(i)}|\Lambda_{t-1}^{(i)}, Y_{t-1}, \theta)/C(\lambda_t^{(i)}; \hat{e}_{t+1})] - c_t - \log k(\Lambda_t^{(i)}; e_t) \Big)^2,$$

where $C(\lambda_T; e_{T+1}) \equiv 1$ and $\lambda_t^{(i)}$ is drawn from the conditional density $f(\lambda_t|\lambda_{t-1}, \theta)$ for all t. The unknown constant c_t has to be estimated jointly with e_t for all t. This least-squares problem can be solved recursively backward for $t = T, ..., 1$.

Once e_t is obtained and $g(\Lambda_T|Y_T, \theta)$ is determined, the likelihood can be estimated as described before. The ML estimator $\hat{\theta}$ is then obtained by maximizing the estimated likelihood. The details of the EIS implementation will be given in Chapter 2.

Alternative Importance Sampling

This method, originally proposed by Durbin and Koopman (1997, 2000) and Sandmann and Koopman (1998), is based on a linear Gaussian state-space approximation to the original model. In this approach, the observation equation, which approximates (1.9), is given by

$$y_t = \lambda_t + a_t + b_t w_t, \tag{1.21}$$

and Eq. (1.10) is the transition equation of the state variable λ_t. The error process $\{w_t\}$ is *iid* N(0,1) and uncorrelated with the error $\{v_t\}$ in (1.10). The conditional density $g^*(\Lambda_T|Y_T, \theta)$ of the approximating model in (1.21) serves as the IF. Throughout this section, the density marked with an asterisk (*) denotes the density function of the approximating model in (1.21).

The location parameter a_t and scale parameter b_t must be selected to ensure that the observation equation (1.21) is a good approximation to the original equation (1.9). One way to choose "good" parameters a_t and b_t is to equate the first and second derivatives of the logarithm of the conditional density of y_t with respect to λ_t for both the original (1.9) and observation (1.21) equations.

Since the approximating marginal density $f^*(\Lambda_T|\theta)$ is the same as $f(\Lambda_T|\theta)$ in the original model, we have the relationship $g^*(\Lambda_T|Y_T,\theta) = f^*(Y_T|\Lambda_T,\theta)f(\Lambda_T|\theta)/f^*(Y_T|\theta)$. Thus, the likelihood in (1.15) can be written as

$$\begin{aligned} L(\theta;Y_T) &= \int \frac{f(Y_T|\Lambda_T,\theta)f(\Lambda_T|\theta)}{g^*(\Lambda_T|Y_T,\theta)} g^*(\Lambda_T|Y_T,\theta)\, d\Lambda_T \\ &= f^*(\Lambda_T|\theta)\int \frac{f(Y_T|\Lambda_T,\theta)f(\Lambda_T|\theta)}{f^*(\Lambda_T|Y_T,\theta)f(\Lambda_T|\theta)} g^*(\Lambda_T|Y_T,\theta)\, d\Lambda_T \\ &= f^*(\Lambda_T|\theta)\int \frac{f(Y_T|\Lambda_T,\theta)}{f^*(\Lambda_T|Y_T,\theta)} g^*(\Lambda_T|Y_T,\theta)\, d\Lambda_T. \end{aligned}$$

This likelihood can be estimated as

$$\hat{L}(\theta;Y_T) = f^*(\Lambda_T|\theta)\frac{1}{N}\sum_{i=1}^{N}\frac{f(Y_T|\Lambda_T^{(i)},\theta)}{f^*(Y_T|\Lambda_T^{(i)},\theta)}, \tag{1.22}$$

where the $\Lambda_T^{(i)}$ are generated from the IF $g^*(\Lambda_T|Y_T,\theta)$ by a simulation smoother.

In summary, the algorithm of the alternative IS proceeds as follows:

1. Generate a trial Λ_T from (1.10) and (1.11).
2. Find a_t and b_t by equalizing the derivatives of the log-density of the approximating model (1.21) and the original model (1.9).
3. Estimate Λ_T from the state space model in (1.21) and (1.10) by the Kalman smoother.
4. Repeat Steps 2 and 3 until either a_t, b_t, or Λ_T converges.
5. Compute the estimated likelihood in (1.22).

Volatility Estimation

For a given $\hat{\theta}$ and the data Y_T, an estimator of the volatility is the conditional expectation:

$$\hat{\sigma} = E[\sigma_t|Y_T,\hat{\theta}] = \frac{\int \sigma_t f(Y_T,\Lambda_T|\hat{\theta})\, d\Lambda_T}{\int f(Y_T,\Lambda_T|\hat{\theta})\, d\Lambda_T},$$

where $\sigma_t = \exp\{\lambda_t/2\}$. Applying the idea of importance sampling, the joint density $f(Y_T, \Lambda_T|\hat{\theta})$ can be factorized into an IF and RF so that the conditional expectation becomes

$$\hat{\sigma} = E[\sigma_t|Y_T, \hat{\theta}] = \frac{E_g[\sigma_t h(\Lambda_T, Y_T|\hat{\theta})]}{E_g[h(\Lambda_T, Y_T|\hat{\theta})]}. \tag{1.23}$$

To calculate this expectation, the expectations in the numerator and denominator can be estimated by the corresponding MC sample mean, where the IF $g(\Lambda_T|Y_T, \theta)$ is determined by the EIS algorithm associated with the evaluation of the likelihood function discussed in Section 2.3.1. Therefore, the MC estimates of $E[\sigma_t|Y_T, \hat{\theta}]$ are obtained as a byproduct of the likelihood evaluation at the SML estimates of θ based on the EIS sampler.

1.3.4 Markov Chain Monte Carlo

The MCMC method is widely used in applied Bayesian statistical analysis. In the SV context, the MCMC method is useful for the parameter estimation. However, the difficulty of directly computing the likelihood makes it impossible to sample from the posterior distribution $f(\theta|Y_T)$. Instead, the MCMC approach employs $f(\theta, \Lambda_T|Y_T)$ in sampling for θ, where Λ_T is a latent log variance process. We will briefly introduce the MCMC method based on Gibbs sampling (Kim *et al.* (1998)) below. For volatility estimation, a particle filter is usually employed with MCMC estimation in the SV model. We will describe the implementation of a particle filter for volatility estimation in Chapter 2.

In the SV model, both Λ_T and θ are to be sampled from their posterior densities. For given initial values (Λ_T^0, θ_0), the Gibbs sampling algorithm to sample (Λ_T, θ) is given as follows (Kim *et al.* (1998)):

1. Draw λ_t from $\lambda_t|\Lambda_{-t}, Y_T, \theta$ for $t = 1, ..., T$.
2. Draw γ^2 from $\gamma^2|Y_T, \Lambda_T, \mu, \beta$.
3. Draw β from $\beta|Y_T, \Lambda_T, \mu, \gamma^2$.
4. Draw μ from $\mu|Y_T, \Lambda_T, \beta, \gamma^2$, where $\mu = \frac{\alpha}{1-\beta}$.

Here, Λ_{-t} denotes a vector of λ_t excluding the t^{th} component, that is, $(\lambda_1, ..., \lambda_{t-1}, \lambda_{t+1}, ..., \lambda_T)$. Note that a cycle through Steps 1 to 4 is called a sweep. To achieve a certain accuracy, this

algorithm requires a sufficiently large number of sweeps, usually several thousand. Each step needs assumptions and approximations for the prior distribution of the parameters to be estimated. More details of the implementation will be given in Chapter 2.

1.4 Organization and Contributions of the Thesis

This thesis is organized as follows. In Chapter 2, we derive moment properties of the SV model and perform empirical analysis with the historical data from six equity returns. We perform parameter estimation with the SML and MCMC methods and volatility estimation with a particle filter. In the following two chapters, we propose other stochastic volatility models that are based on the SV model but can specify more stylized facts of the return series. In Chapter 3, we introduce the autoregressive with stochastic volatility errors (AR–SV) model. Moment properties and the exact likelihood of the AR–SV model are derived. Empirical analysis of the AR–SV model with historical data from some indices is performed to demonstrate the benefits of the model. Chapter 4 introduces Markov switching stochastic volatility (MSSV) models. We study the statistical properties of the MSSV model and perform empirical analysis with actual data. In Chapter 5, we summarize our work in Chapters 1 to 4, and discuss the vector autoregressive model with SV errors. We also present directions for future research.

New research contributions in this thesis are as listed below:

1. For the Stochastic Volatility Model,
   (a) We conduct an empirical study with six actual return series to investigate the performance of the SML with the EIS technique for the SV models with normal and heavy-tail error assumptions,
   (b) We compare the performance of SML and MCMC with the same data set.
2. For an Autoregressive Model with a Stochastic Volatility error,
   (a) We derive moment properties such as kurtosis, ACF of the series, ACF of the squares, and cross correlation between the series and the squares. We also obtain the exact likelihood function,

(b) We conduct a simulation study to investigate the performance of the SML with the EIS technique for AR–SV model and a sensitivity analysis to study the robustness of the SML method for different tail behavior,

(c) We also conduct empirical studies with four actual return series. We implement the SML with the EIS technique for the AR–SV models with normal and t error assumptions, determine the order of the models, and perform model diagnostics.

3. For a Markov Switching Stochastic Volatility Model,

(a) We derive the exact likelihood function,

(b) We conduct a simulation study to investigate the performance of the SML with the EIS technique,

(c) We also conduct empirical studies with four actual return series. We implement the SML with the EIS technique for the MSSV models with normal and t error assumptions.

Chapter 2

Statistical and Empirical Analysis of Stochastic Volatility Models

In spite of the simple structure of the SV model, its estimation is not trivial because of the existence of the latent process. Various estimation methods for this model have been developed. We introduced in Chapter 1 several estimation methods that are commonly used in the literature such as the generalized method of moments (Hansen (1982)), the efficient method of moments (Gallant *et al.* (1997)), the Markov chain Monte Carlo (MCMC) (Jacquier *et al.* (1994) and Kim *et al.* (1998)), and the simulated maximum likelihood (SML) (Danielsson and Richard (1993)). In particular, Danielsson (1994) showed in an empirical study using the S&P 500 index that SML was as efficient as MCMC for finite samples. Liesenfeld and Richard (2008) proposed an improved MCMC method based on the Metropolis Hastings procedure using EIS technique (MH-MCMC-EIS) and Ozturk (2009) implemented the HM-MCMC-EIS method for EGARCH and SV models. However, our focus in this thesis is more on a stand-alone method such as SML or MCMC rather than the combined method such as MG-MCMC-EIS.

The SML method is a ML estimation approach, except that the likelihood is not directly computed. Instead, the likelihood is estimated by the Monte Carlo (MC) method. That is, SML seeks the parameter estimates that maximize the estimated likelihood calculated by MC simulation. To improve the accuracy of the likelihood estimation, a variance reduction method such as importance sampling (IS) is typically considered. Liesenfeld and Richard (2003) introduced efficient importance sampling (EIS), and Shephard and Pitt (1997) and Durbin and Koopman (1997, 2000) introduced another IS technique using the Kalman

filter. Liesenfeld and Richard (2003) showed that EIS could find the MC estimates of the likelihood efficiently. Koopman, Lucas, and Scharth (2011) combined those importance sampling techniques and called it numerically accelerated importance sampling (NAIS) and showed that NAIS worked as efficient as Liesenfeld and Richard's EIS method via a simulation study. In our empirical analysis of the SV model, we implement SML with EIS and the MCMC method for several financial return series, and study the performance of each method. Asymptotic properties of the SML estimator with importance sampling were studied by Koopman, Shephard, and Creal (2009), but those of the SML with the EIS has not been studied yet.

The organization of this chapter is as follows. The standard SV model is considered in Section 1 and Section 2 describes certain statistical properties of the SV model with two different assumptions for the error process u_t. One assumes a normal distribution and the other assumes a heavy-tailed distribution. Section 3 summarizes the SML and MCMC methods for parameter estimation and specifies a particle filter for volatility estimation. In Section 4, the empirical results from the actual data are discussed with model diagnostics. Both normal and heavy-tailed error assumptions for the SV model are considered. Some concluding remarks are given in Section 5.

2.1 Model

The standard SV model, proposed by Taylor (1986), is defined as

$$
\begin{aligned}
y_t &= \exp\{\lambda_t/2\}u_t, && (2.1)\\
\lambda_t &= \alpha + \beta\lambda_{t-1} + \gamma v_t, && (2.2)\\
\lambda_1 &\sim N\left(\frac{\alpha}{1-\beta}, \frac{\gamma^2}{1-\beta^2}\right) && (2.3)
\end{aligned}
$$

for $t = 1, ..., T$, where T is the number of observations, y_t is a financial return series and the error processes u_t and v_t are uncorrelated white noises with zero mean and unit variance. The v_t are assumed to be normally distributed, but the assumption for u_t may be a normal or heavy-tailed distribution. Both processes u_t and v_t are uncorrelated with λ_1. It is worthwhile to note that the errors u_t and v_t can be mutually correlated or more specifically negatively correlated, similar to a common assumption for continuous-time models in the finance literature. A negative correlation between those error processes can allow

a volatility asymmetry. The coefficient $\gamma > 0$ is the conditional standard deviation of λ_t given λ_{t-1}. It is assumed that the absolute value of the persistent parameter β is less than 1, *i.e.*, $|\beta| < 1$, so that the log variance λ_t is a stationary AR(1) process. Thus, λ_t linearly depends on λ_{t-1} (*see* Eq. (2.2)). A small log volatility at time $t-1$ implies a small λ_t, and a large λ_{t-1} leads to a large λ_t. This is volatility clustering, one of the empirical stylized facts of a financial return series.

Typically, the error process u_t is assumed to be a Gaussian white noise, but sometimes the assumption of a heavy-tailed distribution for u_t such as a Student-t distribution (Ruiz (1994), Harvey *et al.* (1994), Sandmann and Koopman (1998), and Chib *et al.* (1998)) is made to capture the heavy tail of the return distribution better than a normal innovation. Another heavy-tailed distribution often used is the generalized error distribution (GED) (Liesenfeld and Jung (2000)). The SV model with a normal error (SV-normal) and the SV model with a Student-t distributed error (SV-t) are considered in this chapter and their moment properties are discussed below.

2.2 Moment Properties

In this section, we consider the kurtosis of the series and the autocorrelation function (ACF) of the squared series under two assumptions for the error process u_t. One is a normal distribution and the other is a heavy-tailed distribution (Liesenfeld and Jung (2000)).

Let us consider the case where the error process u_t follows a normal distribution with mean zero and variance one. For $|\beta| < 1$, the unobserved log variance process λ_t is stationary so that it is normally distributed with unconditional mean $\mu = \frac{\alpha}{1-\beta}$ and variance $\sigma^2 = \frac{\gamma^2}{1-\beta^2}$. The properties of a lognormal distribution imply that all the odd moments of y_t are zero and all the even moments exist if $\{\lambda_t\}$ is stationary. Thus, the second and fourth moments of y_t are given by

$$E[y_t^2] = E[e^{\lambda_t}] = \exp\{\mu + \sigma^2/2\}, \tag{2.4}$$

$$E[y_t^4] = E[e^{2\lambda_t} u_t^4] = \exp\{2(\mu + \sigma^2)\} E[u_t^4]. \tag{2.5}$$

Hence, the kurtosis of the returns is given as

$$\kappa = E[y_t^4]/(E[y_t^2])^2 = \exp\{\sigma^2\} E[u_t^4]. \tag{2.6}$$

The above expression for the kurtosis has two components: the kurtosis due to the term $E[u_t^4]$ that represents the kurtosis of the error u_t, and the kurtosis due to the variation

in the log-variance process λ_t. Under the normal assumption for an error process u_t, the kurtosis $E[u_t^4]$ is equal to three, which leads to an unconditional kurtosis κ greater than three. If we take u_t to be a Student-t distribution with unit variance and ω degrees of freedom, then $E[u_t^4] = 3(\omega-2)/(\omega-4)$ provided $\omega > 4$, which is greater than three. This is consistent with the leptokurtosis of the empirical distribution of financial return series.

The ACF of the squared returns can be obtained provided $E[u_t^4] < \infty$ and $|\beta| < 1$. Setting $\sigma_t^2 = \exp\{\lambda_t\}$, the autocovariance of y_t^2 is expressed (Jacquier *et al.* (1994)) as

$$Cov(y_t^2, y_{t-k}^2) = Cov(\sigma_t^2, \sigma_{t-k}^2) = (\exp\{\sigma^2\beta^k\} - 1)(E[\sigma_t^2])^2 \tag{2.7}$$

for any integer $k > 0$. Moreover, from (2.4) and (2.5), the variance of the squared returns y_t^2 is

$$\begin{aligned} Var[y_t^2] &= E[(\sigma_t^2)^2 u_t^4] - (E[\sigma_t^2 u_t^2])^2 \\ &= E[(\sigma_t^2)^2]E[u_t^4] - (E[\sigma_t^2])^2 \\ &= (E[\sigma_t^2])^2 \left[\left(\frac{Var[\sigma_t^2]}{(E[\sigma_t^2])^2} + 1 \right) E[u_t^4] - 1 \right] \\ &= (E[\sigma_t^2])^2 (\exp\{\sigma^2\} E[u_t^4] - 1). \end{aligned} \tag{2.8}$$

With Eqs. (2.7) and (2.8), the ACF of the squared return at lag k, denoted $\rho(k)$ for any positive integer k, is

$$\rho(k) = \frac{\exp\{\sigma^2\beta^k\} - 1}{\exp\{\sigma^2\} E[u_t^4] - 1} \tag{2.9}$$

$$\approx \frac{\exp\{\sigma^2\} - 1}{\exp\{\sigma^2\} E[u_t^4] - 1} \, \beta^k. \tag{2.10}$$

From (2.10), it can be seen that $\rho(k)$ goes to zero as k increases.

Expressing the ACF in (2.10) as a function of its unconditional kurtosis, we can see in the SV model the properties of high kurtosis and low, but slowly decaying, autocorrelations of the squared returns. The theoretical form of the ACF of the squares, proposed by Teräsvirta (1996), is given by

$$\rho(k) = \frac{(\kappa / E[u_t^4])^{\beta^k} - 1}{\kappa - 1}. \tag{2.11}$$

Under the assumption of a finite fourth moment of u_t and $|\beta| < 1$, this function allows us to compare the theoretical combination of $\rho(k)$ and κ with the empirical $\rho(k)/\kappa$ combination

for a given lag k. Thus, in the SV-normal model, the first-order autocorrelation in (2.11) can be written

$$\rho(1) = \frac{(\kappa/3)^{\beta} - 1}{\kappa - 1}. \tag{2.12}$$

Note that $\rho(1)$ is similar to the first-lag autocorrelation of an ARMA(1,1) model. In addition, the ACF of GARCH(1,1) has the same properties as that of ARMA(1,1) and hence the SV model behaves similarly to the GARCH(1,1) model.

Using the relationship between the ACF of the squared series and the kurtosis of the series in (2.12), the relationship between β and κ can be described as follows. Assuming $\beta \geq 0.9$, which is typical for a daily return, if $\rho(1) \leq 0.15$, which is often used as an empirical first-order autocorrelation for a daily return, the predicted kurtosis is less than 11.2. On the other hand, for $\kappa \geq 7$, also a typical value of the empirical kurtosis, the predicted $\rho(1)$ is greater than 0.191. Thus, the SV model with a normal innovation is unlikely to simultaneously capture both the low but slowly decaying ACF of y_t^2 and the leptokurtosis of the return distribution. This can be seen more clearly in Fig. 2.1. The graph on the left-hand side shows the function in (2.12) for different values of β, 0.9, 0.95, and 0.99, with the empirical $\rho(1)/\kappa$ combination of six return series under a normal error assumption. The data will be described in Section 2.4.1. Obviously, two of the empirical values are quite far from the theoretical curves calculated by the formula in (2.12), while the others are near or on the curves.

Suppose that the error u_t follows a heavy-tailed distribution such that $E[u_t^4] > 3$. We will consider a scaled Student-t distribution and the generalized error distribution (GED) (Box and Tiao (1973) and Liesenfeld and Jung (2000)), typically used for a heavy-tailed distribution.

First, let us consider a Student-t distribution with zero mean and unit variance for u_t. Its density function is given by

$$f(x) = \pi(\omega - 2)^{-1/2} \frac{\Gamma\left(\frac{\omega+1}{2}\right)}{\Gamma\left(\frac{\omega}{2}\right)} \left[1 + \frac{x^2}{\omega - 2}\right]^{-(\omega+1)/2}, \tag{2.13}$$

where $\Gamma(\cdot)$ is a gamma function and the degree of freedom $\omega > 2$. Since its variance is one, the kurtosis of u_t becomes $E[u_t^4] = 3(\omega - 2)/(\omega - 4)$, provided $\omega > 4$, which is greater than three if ω is finite. Notice that the unconditional fourth moment $E[u_t^4]$ does not exist if

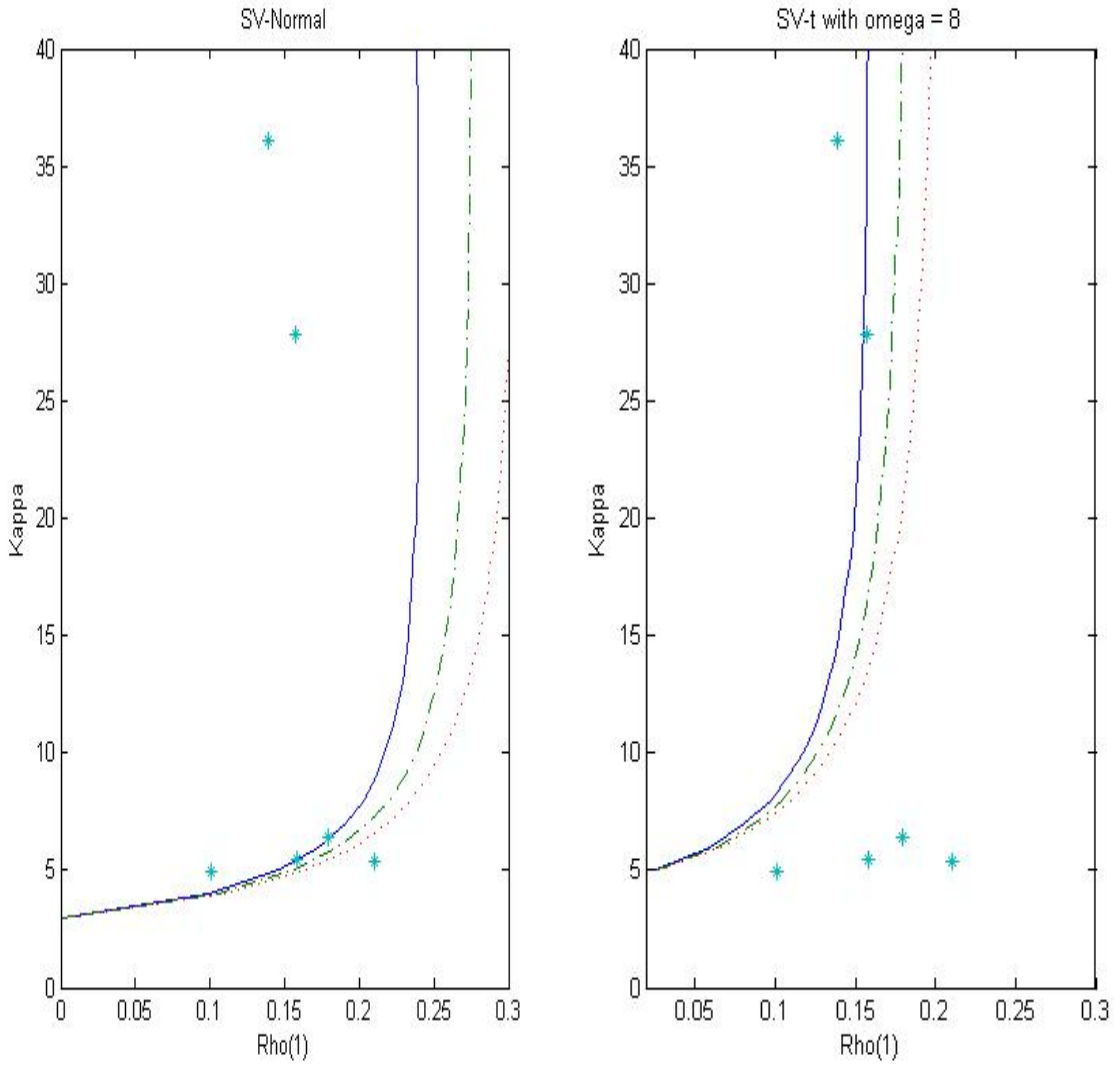

Figure 2.1: $\rho(1)/\kappa$ combination for the data and its theoretical value in SV with normal (left) and Student-t (right) error assumptions for different values of β. Asteroid(*) represents an empirical combination and lines represent a set of theoretical combinations. Solid line for $\beta = 0.9$, solid-dotted line for $\beta = 0.95$, and dotted line for $\beta = 0.99$.

$\omega \leq 4$. Plugging this into (2.11), the first-order autocorrelation of the squared returns is

$$\rho(1) = \frac{\left(\frac{\kappa(\omega-4)}{3(\omega-2)}\right)^{\beta} - 1}{\kappa - 1}. \tag{2.14}$$

Another possible heavy-tailed distribution for u_t is a GED. The density of a GED random variable with zero mean and unit variance is given by

$$f(x) = \frac{\nu \ \exp[-1/2|x/\xi|^{\nu}]}{\xi \ \Gamma(1/\nu) 2^{1+1/\nu}}, \qquad 0 < \nu < \infty \tag{2.15}$$

where

$$\xi = \sqrt{\left(\frac{1}{2}\right)^{2/\nu} \frac{\Gamma(1/\nu)}{\Gamma(3/\nu)}}.$$

It should be noted that the GED is identical to a normal distribution when $\nu = 2$ and for $\nu < 2$, its kurtosis is $E[u_t^4] = \Gamma(1/\nu)\Gamma(5/\nu)/[\Gamma(3/\nu)]^2$, which is greater than three (Johnson and Kotz (1970)).

The right panel of Fig. 2.1 shows the theoretical first-order autocorrelation/kurtosis combination curves and the empirical values from our data in the SV model with the Student-t error assumption. We assume $\omega = 8$ degrees of freedom. This is equivalent to the case where $\nu \approx 1.26$ in the GED distribution. Both lead to $E[u_t^4] = 4.5$. The two values that deviated from the theoretical curve under a normal assumption (left panel) are now quite close to the theoretical values. On the other hand, the four values that were close to the curve under a normal assumption are now quite far from the theoretical values. Therefore, the assumption of a leptokurtic distribution for the error process u_t helps the model to be more compatible with a low but persistent ACF of the squared returns and to have a high kurtosis of the returns.

2.3 Estimation Methods

Despite the simplicity of the model, it is not easy to estimate the SV model because of the existence of the latent process λ_t. There are various ways of estimating this model. Some approaches use the method of moments, and others approximate the likelihood. In this chapter, two of the most popular and efficient estimation methods, SML and MCMC, are considered. We will introduce a particle filter for the volatility estimation.

2.3.1 Simulated Maximum Likelihood

For the SV model in (2.1) and (2.2), the likelihood associated with the known observations $Y_T = \{y_t\}_{t=1}^T$ and latent variables $\Lambda_T = \{\lambda_t\}_{t=1}^T$ is given by

$$L(\theta; Y_T) = \int_{R^T} f(Y_T, \Lambda_T|\theta)\, d\Lambda_T, \tag{2.16}$$

where θ denotes the vector of parameters α, β, and γ to be estimated, and $f(\cdot)$ is a probability density. The simulated maximum likelihood (SML) method was introduced by Danielsson and Richard (1993); it depends on Monte Carlo (MC) integration to evaluate the likelihood in (2.16). After estimating the likelihood by MC simulation, SML seeks values of the parameter vector θ that maximize the estimated likelihood. To efficiently estimate the likelihood in (2.16), we will use the efficient importance sampling (EIS) proposed by Liesenfeld and Richard (2003). The EIS implementation is discussed below.

Efficient Importance Sampling

The EIS technique (Liesenfeld and Richard (2003)) chooses in the following way an IF that can provide a good approximation of $f(Y_T, \Lambda_T|\theta)$.

Consider the following factorization of the likelihood function in (2.16) using an IF $g(\Lambda_T|Y_T, \theta)$:

$$\begin{aligned} L(\theta; Y_T) &= \int_{R^T} f(Y_T|\Lambda_T, \theta) f(\Lambda_T|\theta)\, d\Lambda_T \qquad (2.17) \\ &= \int_{R^T} \frac{f(Y_T|\Lambda_T, \theta) f(\Lambda_T|\theta)}{g(\Lambda_T|Y_T, \theta)} g(\Lambda_T|Y_T, \theta)\, d\Lambda_T \\ &= \int_{R^T} h(\Lambda_T, Y_T|\theta) g(\Lambda_T|Y_T, \theta)\, d\Lambda_T \qquad (2.18) \end{aligned}$$

where $h(\Lambda_T, Y_T|\theta) = f(Y_T|\Lambda_T, \theta) f(\Lambda_T|\theta)/g(\Lambda_T|Y_T, \theta)$ is a remainder function (RF). To find an IF $g(\Lambda_T|Y_T, \theta)$ that depends on both λ_T and Y_T, EIS considers a particular factorization:

$$g(\Lambda_T|Y_T, \theta) = \prod_{t=1}^{T} f(\lambda_t|\Lambda_{t-1}, Y_{t-1}, \theta, e_t), \tag{2.19}$$

where e_t is an auxiliary parameter that depends on the observations Y_T. Then, we find a function $k(\Lambda_t; e_t)$ such that

$$g(\Lambda_T|Y_T, \theta) = \prod_{t=1}^{T} C(\lambda_{t-1}, e_t) k(\Lambda_t; e_t), \tag{2.20}$$

where $C(\lambda_{t-1}, e_t) = \frac{1}{\int k(\Lambda_t; e_t) d\lambda_t}$ is an integral constant that makes $k(\Lambda_t; e_t)$ a density function. For ease of notation, we remove θ from the functional notations of C and k.

Estimation of e_t

For the SV model, the natural choice of $k(\Lambda_t; e_t)$ could be proportional to a Gaussian density for λ_t given λ_{t-1}. Thus, we can express the function k as a product of two functions:

$$k(\Lambda_t; e_t) = \xi(\lambda_t, e_t)\ f(\lambda_t|\lambda_{t-1}, \theta). \tag{2.21}$$

Since $f(\lambda_t|\lambda_{t-1}, \theta)$ is a Gaussian density, an auxiliary vector e_t can be estimated via regression coefficients from a linear regression if $\xi(\lambda_t, e_t)$ is chosen as a Gaussian density kernel $\exp\{e_{1,t}\lambda_t + e_{2,t}\lambda_t^2\}$. That is, $\log \xi(\lambda_t, e_t)$ approximates $\log f(y_t|\lambda_t, \theta) - \log C(\lambda_t, e_{t+1})$ as in Liesenfeld and Richard (2003).

The regression equation is given by

$$\log f(y_t|\lambda_t, \theta) - \log C(\lambda_t, e_{t+1}) = constant + e_{1,t}\lambda_t + e_{2,t}\lambda_t^2 + a_t, \tag{2.22}$$

for $t = 1, ..., T$, where a_t is a regression error term. Assuming $C(\lambda_T; e_{T+1}) = 1$, this regression can be done backward.

The conditional density of λ_t is given by

$$f(\lambda_t|\lambda_{t-1}, \theta) = \frac{1}{\sqrt{2\pi\gamma^2}} \exp\left[-\frac{(\lambda_t - \alpha - \beta\lambda_{t-1})^2}{2\gamma^2}\right]. \tag{2.23}$$

Substituting this density into (2.21), the function $f(\lambda_t|\Lambda_{t-1}, Y_{t-1}, \theta, e_t)$ in (2.19) is given by

$$f(\lambda_t|\lambda_{t-1}, y_{t-1}, \theta, e_t) = \frac{1}{\sqrt{2\pi V_t}} \exp\left[-\frac{(\lambda_t - M_t)^2}{2V_t}\right],$$

where

$$M_t = \left(\frac{\alpha + \beta\lambda_{t-1}}{\gamma^2} + e_{1,t}\right) V_t,\ V_t = \frac{\gamma^2}{1 - 2e_{2,t}\gamma^2},$$

and $\log C(\lambda_t, e_{t+1})$ is chosen as

$$\log C(\lambda_t, e_{t+1}) = -\frac{M_t^2}{2V_t} + \frac{(\alpha + \beta\lambda_{t-1})^2}{2\gamma^2} - \frac{1}{2}\log\frac{V_t}{\gamma^2}.$$

Since $C(\lambda_t, e_{t+1})$ depends on $e_{1,t+1}$ and $e_{2,t+1}$, the regression coefficients are computed recursively, from $t = T, T-1, ..., 1$, assuming $C(\lambda_T, e_{T+1}) = 1$.

Once e_t is estimated, we have the IF g, given by

$$g(\Lambda_T|Y_T, \theta) = \prod_{t=1}^{T} C(\lambda_{t-1}, \hat{e}_t) \exp\{\hat{e}_{1,t}\lambda_t + \hat{e}_{2,t}\lambda_t^2\} f(\lambda_t|\lambda_{t-1}, \theta). \tag{2.24}$$

Using this IF, the corresponding likelihood is estimated by

$$\hat{L}(\theta; Y_T) = \frac{1}{N} \sum_{i=1}^{N} \prod_{t=1}^{T} \frac{f(y_t|\lambda_t^{(i)}, \theta)}{C(\lambda_{t-1}, \hat{e}_t) \exp\{\hat{e}_{1,t}\lambda_t^{(i)} + \hat{e}_{2,t}\lambda_t^{(i)2}\}}, \tag{2.25}$$

where $\lambda_t^{(i)}$ is drawn from $g(\Lambda_T^{(i)}|Y_T, \theta)$ in (2.24). Note that the conditional distribution of y_t given λ_t is normal with mean zero and variance $\exp\{\lambda_t\}$.

In summary, the EIS algorithm to evaluate the likelihood for the SV model in (2.16), given an initial parameter vector θ, is as follows:

1. Generate the $\Lambda_T^{(i)}$ according to (2.2).
2. Assuming $C(\lambda_T, e_{T+1}) = 1$, obtain e_t by performing T regressions as in (2.22), working backward from $t = T$ to $t = 1$.
3. Generate the new $\Lambda_T^{(i)}$ from the importance density $g(\Lambda_T|Y_T, \theta)$ in (2.24).
4. Repeat Steps 2 and 3 until either e_t or the new λ_t converges for all t.
5. Compute the estimated likelihood in (2.25).

2.3.2 Markov Chain Monte Carlo

The MCMC method is also useful for obtaining the estimate of a parameter vector θ in the SV model. It employs $f(\theta, \Lambda_T|Y_T)$ in sampling for θ, where λ_t is a latent log-variance process. We will consider the MCMC method based on Gibbs sampling proposed by Kim *et al.* (1998).

Gibbs sampling algorithm

In the MCMC with Gibbs sampling, both Λ_T and θ are to be sampled from their posterior densities. For ease of computation, we will generate μ, which is the unconditional mean of λ_t, by Gibbs sampling and convert it to α using the following relationship:

$$\mu = \frac{\alpha}{1-\beta}. \tag{2.26}$$

Throughout this section, the parameter vector θ is (μ, β, γ).

For given initial values $(\Lambda_T^{(0)}, \theta_0)$, the algorithm to sample (Λ_T, θ) is given as follows (Kim *et al.* (1998)):

1. Draw λ_t from $\lambda_t|\Lambda_{-t}, Y_T, \theta$ for $t = 1, ..., T$.
2. Draw γ^2 from $\gamma^2|Y_T, \Lambda_T, \mu, \beta$.
3. Draw β from $\beta|Y_T, \Lambda_T, \mu, \gamma^2$.
4. Draw μ from $\mu|Y_T, \Lambda_T, \beta, \gamma^2$.

Here, Λ_{-t} denotes a vector of λ_t excluding the t^{th} component, that is, $(\lambda_1, ..., \lambda_{t-1}, \lambda_{t+1}, ..., \lambda_T)$. Note that a cycle through Steps 1 to 4 is called a sweep. To achieve a certain accuracy, this algorithm requires a sufficiently large number of sweeps, usually several thousand. The details of each step are given below.

Step 1. Draw λ_t

The most difficult part of this process is to effectively sample λ_t from $\lambda_t|\Lambda_{-t}, Y_T, \theta$ because this step should be carried out T times at each sweep. To improve the computational efficiency, an acceptance/rejection procedure can be applied in sampling λ_t.

The conditional density $f(\lambda_t|\Lambda_{-t}, Y_T, \theta)$ can be expressed as

$$\begin{aligned} f(\lambda_t|\Lambda_{-t}, Y_T, \theta) &= \frac{f(Y_T|\lambda_t, \Lambda_{-t}, \theta) f(\lambda_t|\Lambda_{-t}, \theta)}{f(Y_T|\Lambda_{-t}, \theta)} \\ &\propto f(y_t|\lambda_t, \theta) f(\lambda_t|\Lambda_{-t}, \theta) \end{aligned} \tag{2.27}$$

for $t = 1, ..., T$. Note that, in the SV model,

$$f(\lambda_t|\Lambda_{-t}, \theta) \sim N(m_t, v^2), \tag{2.28}$$

where

$$m_t = \mu + \frac{\beta[(\lambda_{t-1} - \mu) + (\lambda_{t+1} - \mu)]}{1 + \beta^2} \qquad \text{and} \qquad v^2 = \frac{\gamma^2}{1 + \beta^2},$$

where μ is given in (2.26).

To establish the acceptance/rejection criterion, let the log density of y_t given λ_t and θ be a constant plus the logarithm of the joint density $\log f^*(y_t, \lambda_t, \theta)$. Then $\log f^*(\cdot)$ can be approximated by a linear function of λ_t using the first-order Taylor expansion, originated at m_t. That is,

$$\begin{aligned} \log f^*(y_t, \lambda_t, \theta) &= -\frac{1}{2}\lambda_t - \frac{y_t^2}{2}\exp\{-\lambda_t\} \\ &\leq -\frac{1}{2}\lambda_t - \frac{y_t^2}{2}[(1 + m_t)\exp\{-m_t\} - \lambda_t \exp\{-m_t\}] \\ &\doteq \log g^*(y_t, \lambda_t, m_t, \theta). \end{aligned} \tag{2.29}$$

Substituting (2.29) into (2.27), the acceptance/rejection probability is induced from the given inequality:

$$\begin{aligned} f(y_t|\lambda_t, \theta) f(\lambda_t|\Lambda_{-t}, \theta) &\leq f^*(y_t, \lambda_t, \theta) f(\lambda_t|\Lambda_{-t}, \theta) \\ &\leq g^*(y_t, \lambda_t, m_t, \theta) f(\lambda_t|\Lambda_{-t}, \theta). \end{aligned}$$

Notice that the product of $g^*(\cdot)$ and $f(\cdot)$ on the right-hand side of the above inequality is proportional to a normal density with variance v^2 and mean $m_t^* = m_t + \frac{v^2}{2}[y_t^2 \exp\{-m_t\} - 1]$.

The acceptance/rejection procedure, therefore, can be performed as follows: First, draw a proposed λ_t from $N(m_t^*, v^2)$. Second, accept the value with probability f^*/g^*; if rejected, go to the first step.

Step 2. Draw γ^2 and β

Sampling γ^2 is relatively intuitive and straightforward. Assume an *inverse* Gamma (IG)

distribution with parameters $a/2$ and $b/2$ as a conjugate prior for γ^2. In Bayesian analysis, an *inverse* Gamma distribution is commonly used as a conjugate prior for the variance of a normal distribution. In the same spirit, it is chosen here because a log-variance λ_t is normally distributed assuming a normal distribution for v_t and $|\beta| < 1$. Setting $a = 5$ and $b = 0.01 \times a$, the posterior becomes

$$\begin{aligned} \gamma^2 \quad & |Y_T, \Lambda_T, \mu, \beta \\ \sim \quad & IG\left(\frac{T+a}{2}, \frac{b + (\lambda_1 - \mu)^2(1-\beta^2) + \sum_{t=1}^{T-1}((\lambda_{t+1} - \mu) - \beta(\lambda_t - \mu))^2}{2}\right), \end{aligned}$$

where μ is given in (2.26).

Setting $\beta = 2\beta^* - 1$, where $\beta^* \sim Beta(\beta_1, \beta_2)$, a prior density of β can be written as

$$f(\beta) \propto \left(\frac{1+\beta}{2}\right)^{\beta_1 - 1} \left(\frac{1-\beta}{2}\right)^{\beta_2 - 1} \tag{2.30}$$

for $-1 < \beta < 1$. Other priors can be chosen for β (*see* Phillips (1991), Schotman and Van Dijk (1991), Chib and Greenberg (1994), and Marriott and Smith (1992)). However, this prior is chosen to ensure stationarity. Similarly to Step 1, we can establish an acceptance/rejection criterion for β, and the algorithm for drawing β is as follows:

1. Initialize β and set $\beta^{(0)}$.
2. Sample $\beta_{prop} \sim N(\hat{\beta}, v_\beta^2)$, where $\hat{\beta} = \frac{\sum_{t=1}^{T}(\lambda_{t+1}-\mu)(\lambda_t - \mu)}{\sum_{t=1}^{T-1}(\lambda_t - \mu)^2}$ and $v_\beta^2 = \frac{\gamma^2}{\sum_{t=1}^{T-1}(\lambda_t - \mu)^2}$.
3. Accept the proposed value as $\beta^{(1)}$ with probability $\exp[g(\beta_{prop} - g(\beta^{(0)}))]$; if rejected, set $\beta^{(1)}$ as $\beta^{(0)}$ and go to Step 2. A function $g(\beta)$, proposed by Chib and Greenberg (1995), is given by
$$g(\beta) = \log f(\beta) - \frac{(\lambda_1 - \mu)^2(1-\beta^2)}{2\gamma^2} + \frac{1}{2}\log(1-\beta^2).$$
4. Repeat Steps 2 and 3 until convergence.

Step 4. Draw μ and Determine α

For μ, we take an informative prior $N(0, 10)$. This prior is chosen for the ease of likelihood calculations (*see* Kim *et al.*(1998)). Then the posterior becomes a normal distribution with mean

$$\hat{\mu} = v_\mu^2 \left[\frac{1-\beta^2}{\gamma^2}\lambda_1 + \frac{1-\beta}{\gamma^2}\sum_{t=1}^{T-1}(\lambda_{t+1} - \beta\lambda_t) \right]$$

and variance

$$v_\mu^2 = \frac{\gamma^2}{(T-1)(1-\beta)^2 + (1-\beta^2)}.$$

Therefore, α can be obtained by the relationship in (2.26).

2.3.3 Volatility Estimation

In our empirical analysis, we found that a particle filter (Pitt and Shephard (1997)) produced better volatility estimates than a volatility estimation implied by SML. Given this, we employ a particle filter to estimate the log-variance λ_t or volatility $\sigma_t = \exp\{\lambda_t/2\}$ for a given estimated parameter $\hat{\theta}$ obtained by SML or MCMC. The idea is for a given $\hat{\theta}$ to draw M new samples λ_t^{new} from $\lambda_t|Y_t, \hat{\theta}$, given $\lambda_{t-1}^{new,1}, ..., \lambda_{t-1}^{new,M}$ from $\lambda_{t-1}|Y_{t-1}, \hat{\theta}$, where $Y_t = (y_1, ..., y_t)$.

According to Bayes' theorem, the conditional density $f(\lambda_t|Y_t, \hat{\theta})$ is proportional to $f(y_t|\lambda_t, \hat{\theta})f(\lambda_t|Y_{t-1}, \hat{\theta})$:

$$f(\lambda_t|Y_t, \hat{\theta}) \propto f(y_t|\lambda_t, \hat{\theta})f(\lambda_t|Y_{t-1}, \hat{\theta}),$$

where

$$f(\lambda_t|Y_{t-1}, \hat{\theta}) = \int f(\lambda_t|\lambda_{t-1}, \hat{\theta})f(\lambda_{t-1}|Y_{t-1}, \hat{\theta})d\lambda_{t-1}. \tag{2.31}$$

Then, the integral in (2.31) can be estimated by

$$f(\lambda_t|Y_{t-1}, \hat{\theta}) \approx \frac{1}{M}\sum_{j=1}^{M} f(\lambda_t|\lambda_{t-1}^{(j)}, \hat{\theta}),$$

where $\lambda_{t-1}^{(j)}$ is drawn from $f(\lambda_{t-1}|Y_{t-1},\hat{\theta})$ for $j = 1, ..., M$. This leads to

$$f(\lambda_t|Y_t,\hat{\theta}) \propto f(y_t|\lambda_t,\hat{\theta})\frac{1}{M}\sum_{j=1}^{M} f(\lambda_t|\lambda_{t-1}^{(j)},\hat{\theta}). \tag{2.32}$$

To efficiently sample λ_t, we consider an acceptance/rejection procedure, similar to that in Step 1 in Section 2.3.2. Let $\lambda_{t|t-1} = \hat{\alpha} + \hat{\beta}(M^{-1}\sum \lambda_{t-1}^{(j)})$ and $\log f(y_t|\lambda_t,\hat{\theta}) = \textit{constant} + \log f^*(y_t,\lambda_t,\hat{\theta})$. By the first-order Taylor approximation, we obtain the following acceptance probability:

$$\begin{aligned}
\log f^*(y_t,\lambda_t,\hat{\theta}) &= -\frac{1}{2}\lambda_t - \frac{y_t^2}{2}\exp\{-\lambda_t\} \\
&\leq -\frac{1}{2}\lambda_t - \frac{y_t^2}{2}[(1+\lambda_{t|t-1})\exp\{-\lambda_{t|t-1}\} - \lambda_t\exp\{-\lambda_{t|t-1}\}] \\
&\doteq \log g^*(\lambda_t,\lambda_{t|t-1},\hat{\theta}).
\end{aligned} \tag{2.33}$$

It can also be shown via some algebra that

$$g^*(\lambda_t,\lambda_{t|t-1},\hat{\theta})f(\lambda_t|\lambda_{t-1}^{(j)},\hat{\theta}) \propto \pi_j f_N(\lambda_t|\lambda_{t|t-1}^{(j)},\hat{\gamma}^2), \tag{2.34}$$

where $f_N(\lambda_t|\lambda_{t|t-1}^{(j)},\hat{\gamma}^2)$ denotes a normal density with mean $\lambda_{t|t-1}^{(j)}$, and variance $\hat{\gamma}^2$,

$$\pi_j = \exp\left[\frac{1}{2\hat{\gamma}^2}\{(\hat{\alpha}+\hat{\beta}\lambda_{t|t-1}^{(j)2})^2 - \lambda_{t|t-1}^{(j)2}\}\right],$$

and

$$\lambda_{t|t-1}^{(j)} = \hat{\alpha} + \hat{\beta}\lambda_{t-1}^{(j)} + \frac{\hat{\gamma}^2}{2}[y_t^2\exp\{-\lambda_{t|t-1}\} - 1].$$

Hence, because of the relationship in (2.34), the right-hand side of (2.32) is bounded as

$$\begin{aligned}
f(y_t|\lambda_t,\hat{\theta})\frac{1}{M}\sum_{j=1}^{M} f(\lambda_t|\lambda_{t-1}^{(j)},\hat{\theta}) &\leq g^*(\lambda_t,\lambda_{t-1},\hat{\theta})\frac{1}{M}\sum_{j=1}^{M} f(\lambda_t|\lambda_{t-1}^{(j)},\hat{\theta}) \\
&\propto \frac{1}{M}\sum_{j=1}^{M}\pi_j f_N(\lambda_t|\lambda_{t|t-1}^{(j)},\hat{\gamma}^2).
\end{aligned}$$

In summary, the algorithm for drawing λ_t is given as follows: First, draw a proposal λ_t^{prop} for the mixture density $\sum_{j=1}^{M} \pi_j^* f_N(\lambda_t|\lambda_{t|t-1}^{(j)}, \hat{\gamma}^2)$, where $\pi_j^* = \pi_j / \sum \pi_j$, and then accept λ_t^{prop} with probability $f^*(y_t, \lambda_t, \hat{\theta})/g^*(\lambda_t, \lambda_{t|t-1}, \hat{\theta})$. If rejected, go to the first step. It should be noted that the volatility estimate supplied by the particle filter is in fact a one-step ahead prediction (*see* Larsson (2005) and Pederzoli (2006)).

2.4 Empirical Analysis

In this section, we perform an empirical analysis with six actual return series. Parameter estimation results for the SV model obtained by SML with the EIS and MCMC methods and volatility estimation results obtained by a particle filter are discussed. Moreover, model diagnostics based on the residual analysis are performed. All the code is written in MATLAB 7.3.0.

2.4.1 Data

In our empirical analysis, we study the following financial time series: the weekly spot exchange rates of the US Dollar/Japanese YEN ($/YEN) from October 1, 1981 to June 28, 1985 (Harvey *et al.* (1994)), the S&P500 index, which are the daily closing prices from January 2, 1980 to December 12, 1987 (Jacquier *et al.* (1994)), the stock prices of IBM from January 1, 1973 to December 31, 1991 (Anderson (1994)), the stock prices of Honda from January 1, 1996 to August 29, 2003, the Nikkei 225 index from December 2, 1992 to January 15, 2001, and the stock prices of CIBC from January 1, 1996 to August 29, 2003. All the stock prices are the daily closing prices. The total number of observations and the summary statistics for each return series are given in Table 2.1.

Note that each return series y_t is mean corrected as

$$y_t = \log(p_t/p_{t-1}) - \frac{\sum_{t=1}^{T} \log(p_t/p_{t-1})}{T} \quad \text{for } t = 1, ..., T,$$

where p_t is the price of an underlying asset at time t and $Y_T = (y_1, ..., y_T)$. The sample kurtosis of the return for each series is greater than three (*see* Table 2.1), which indicates that the distribution of the return is leptokurtic. In particular, the kurtoses of S&P500

Table 2.1: Summary statistics of data

Statistics	\$/YEN	S&P500	IBM	HONDA	NIKKEI 225	CIBC
Sample size	945	2023	4693	1557	2000	1700
Mean	0.0069	0.0445	0.0151	-0.0500	0.0258	0.0385
Std. dev.	0.5928	1.1399	1.4566	2.1384	1.4731	1.6171
Kurtosis	4.9632	36.1123	27.8268	6.4482	5.4816	5.3648
$LB_y(30)$	25.0678	45.9152	50.3135	36.8192	39.7589	70.5274
	(0.7217)	(0.0317)	(0.0115)	(0.1825)	(0.1096)	(0.0000)
$LB_{y^2}(30)$	274.8873	177.7401	208.7415	284.0462	539.3733	386.2076
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)

and IBM are large. This is due to the extremely unusual values in the series during the Black Monday period.

In addition to the summary statistics in Table 2.1, we also show the autocorrelations of the returns in Fig. 2.2 and the autocorrelation of the squared returns in Fig. 2.3. All the returns, except possibly the S&P500 index and CIBC, show little autocorrelation. The ACFs of the squares however are significant. For all the series the squares of the ACFs start at low levels and decay slowly, implying strong persistence. This is typical empirical behavior of return series. This is also confirmed by the Ljung-Box statistics shown in the bottom panels in Table 2.1. $LB_y(30) = T(T+2)\sum_{j=1}^{30} \frac{\hat{\rho}_j^2}{T-j}$ denotes the Ljung-Box statistic calculated with the first thirty sample autocorrelations of y_t while $LB_y^2(30)$ is that of the squares. The quantities in parentheses are the p-values. Diebold (1988) has cautioned about the use of Ljung-Box statistics in checking for serial correlation in time series with conditional heteroscedasticity. However, several authors in this area (*see* for example Liesenfeld and Richard (2003)) still use it for checking serial correlation. Thus, we list these values here as an indicator and not as a real test statistic.

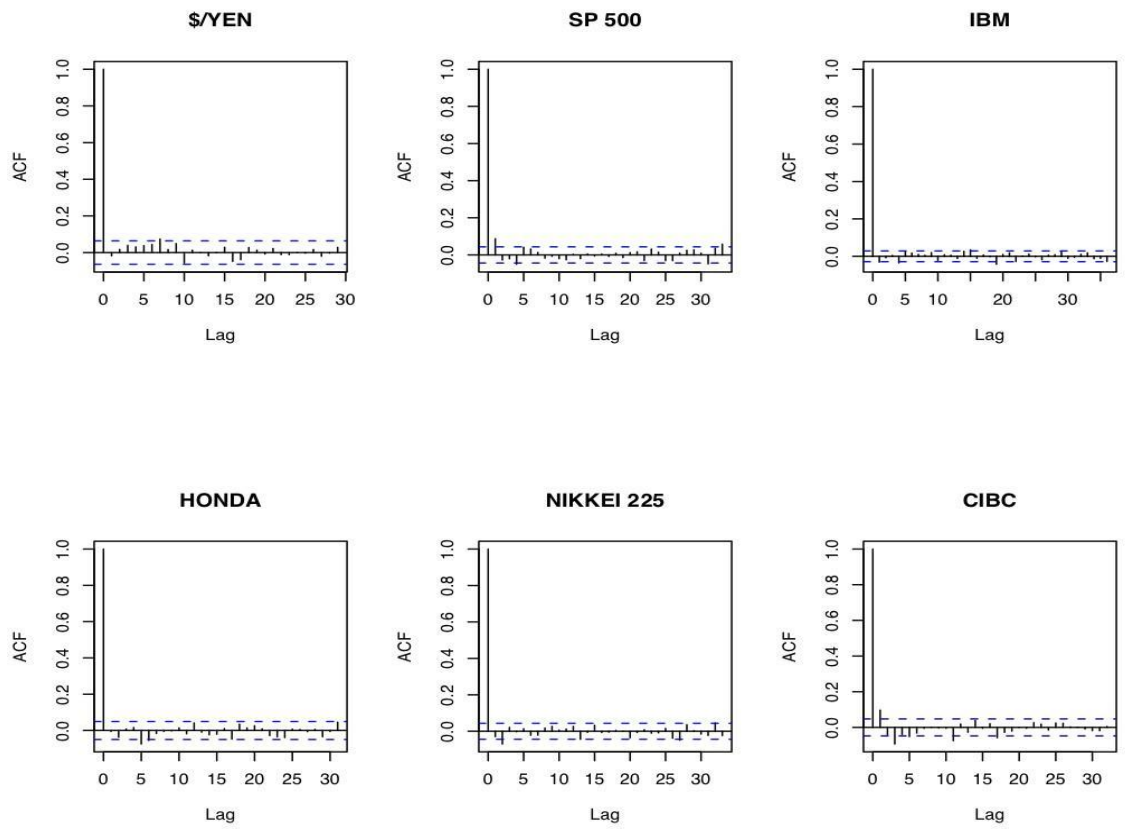

Figure 2.2: Autocorrelation functions of $/YEN, S&P500, IBM, Honda, Nikkei 225, and CIBC

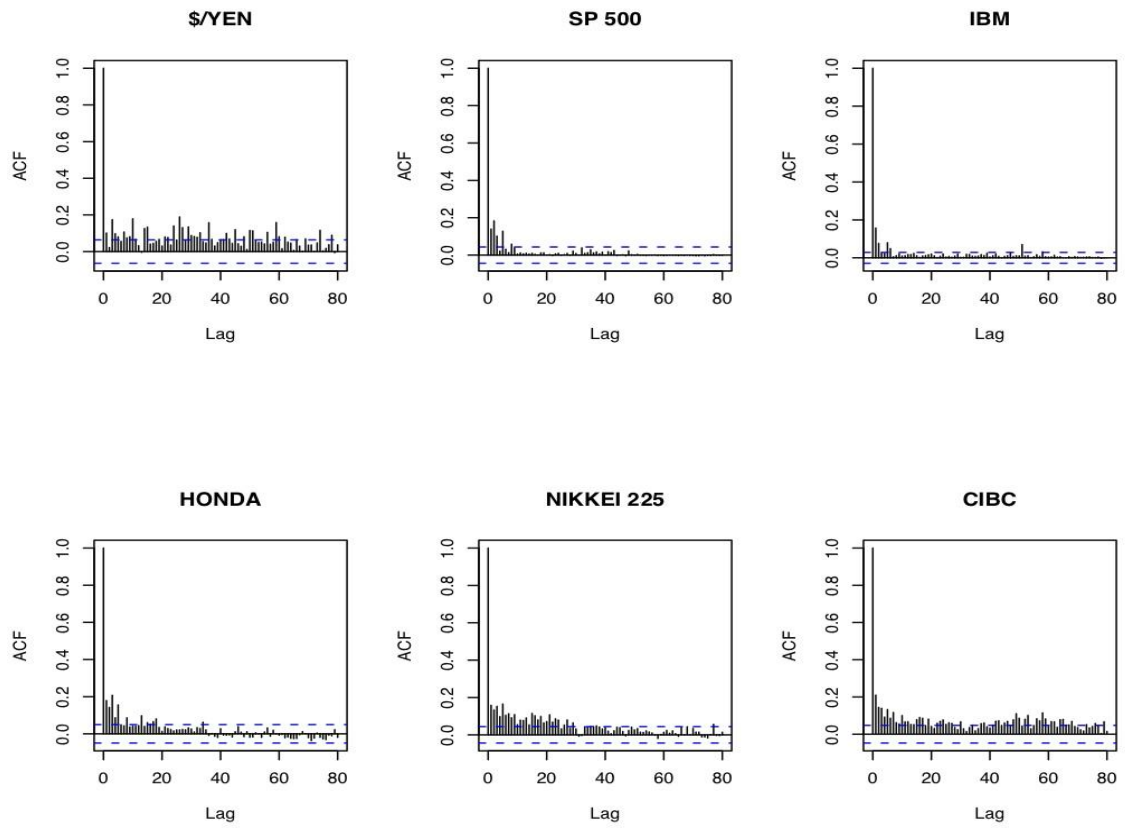

Figure 2.3: Autocorrelation functions of the squares of $/YEN, S&P500, IBM, Honda, Nikkei 225, and CIBC

Table 2.2: SML estimation of SV-normal model

Parameter	$/YEN	S&P500	IBM	HONDA	NIKKEI 225	CIBC
α	0.0001	-0.0342	0.0020	0.0001	-0.0121	-0.0092
MC std. error	0.0008	0.0019	0.0105	0.0014	0.0010	0.0005
Std. error	0.0222	0.0159	0.0122	0.0463	0.0388	0.0428
β	0.9719	0.9728	0.9660	0.9702	0.9648	0.9707
MC std. error	0.0005	0.0008	0.0035	0.0010	0.0013	0.0004
Std. error	0.0197	0.0498	0.0325	0.0208	0.0197	0.0608
γ	0.1510	0.1982	0.2140	0.1937	0.1636	0.1746
MC std. error	0.0029	0.0016	0.0014	0.0008	0.0007	0.0013
Std. error	0.0582	0.0344	0.0329	0.0479	0.0577	0.0491
$\hat{\sigma}$	0.4115	0.7321	0.6851	0.6390	0.3870	0.5280
$\hat{\kappa}$	4.5257	6.2382	5.9520	5.6840	4.4176	5.0864
CPU time*	2052.6	2012.4	2968.1	2362.4	2491.7	2504.9

*in seconds

2.4.2 Estimation Results: SML

In the implementation of SML, the EIS method is employed with four iterations and an MC simulation sample size of fifty. Our initial experimentation with several values of M, the number of MC simulations, indicated that the MC simulations converges approximately when $M = 50$. Both SV-normal and SV-t models are considered.

The SV-normal estimation results are summarized in Table 2.2 along with the standard errors and the MC sampling standard errors of the parameter estimates. The standard error is calculated from fifty repetitions with the same random numbers, while the MC standard error is computed with different random numbers. Both the MC standard errors and the standard errors of the point estimators are significantly small, which indicates that the SML estimates are quite precise. The estimated β's are highly significant in all cases in the sense that all the estimated values of β are above 0.95. This result is consistent with the property of high persistence in the log-variance process. To see the leptokurtosis of the return series implied by the models, the estimated kurtosis is calculated. Given the estimated parameter $\hat{\theta}$, the estimated kurtosis is calculated by

$$\hat{\kappa} = \exp\{\hat{\sigma}^2\} E[u_t^4],$$

Table 2.3: SML estimation of SV-t distribution with $\omega = 8$

Parameter	\$/YEN	S&P500	IBM	HONDA	NIKKEI 225	CIBC
α	-0.0099	0.0001	0.0051	-0.0173	0.0048	-0.0102
MC std. error	0.0007	0.0004	0.0010	0.0004	0.0006	0.0004
Std. error	0.0174	0.0227	0.0109	0.0337	0.0307	0.0189
β	0.9901	0.9899	0.9904	0.9873	0.9853	0.9882
MC std. error	0.0006	0.0004	0.0027	0.0007	0.0001	0.0003
Std. error	0.0195	0.0351	0.0206	0.0188	0.0187	0.0599
γ	0.0698	0.1794	0.1078	0.0971	0.0898	0.0899
MC std. error	0.0006	0.0012	0.0006	0.0004	0.0006	0.0003
Std. error	0.0442	0.0103	0.0211	0.0399	0.0401	0.0328
$\hat{\sigma}$	0.2473	1.6014	0.6082	0.3736	0.2763	0.3445
$\hat{\kappa}$	5.7625	22.3192	8.2668	6.5381	5.9322	6.3507
CPU time*	2679.4	2449.5	2971.5	2207.7	2641.3	2601.1

*in seconds

where $\hat{\sigma}^2 = \hat{\gamma}^2/(1-\hat{\beta}^2)$. The values of $\hat{\kappa}$ in Table 2.2 are somewhat close to the sample values given in Table 2.1, except for the S&P500 and IBM series. In those two cases, the sample kurtosis is extraordinarily large resulting from some unusual values in the series. It is hence reasonable to consider a distribution with heavier tails for u_t. We choose a Student-t distribution with $\omega = 8$ degrees of freedom for the error u_t.

Table 2.3 shows parameter estimation results for the SV model with a Student-t distribution with $\omega = 8$ degrees of freedom. The MC standard errors for all the series are smaller than those from the SV-normal model. The decline in the MC standard errors may be expected. The estimates of the parameter γ in the SV-t specification are lower than those of the SV-normal model. This induces a reduction in the unconditional variance of λ_t, which is $\sigma^2 = \gamma^2/(1-\beta^2)$, so that the variation of the simulated latent variable λ_t is smaller. Consequently, the MC sampling errors of the estimated parameters decline under the SV-t model. Moreover, the estimated values of the kurtosis are greater than those under the SV-normal model. In particular, the estimated kurtosis of the S&P500 series gets even closer to the sample kurtosis of the series.

In Figs. 2.4 and 2.5, the MC estimates of the sequence of volatilities $\hat{\sigma}_t = \exp\{\hat{\lambda}_t/2\}$ of all the series, resulting from the SV-normal (dashed line) and the SV-t (solid line) models,

are presented along with the absolute value of the return series. Both SV models seem to capture large shocks in the return series more accurately than small shocks. For instance, this can be seen for the S&P500 and IBM series at the October 1987 crash. Moreover, the jumps in the volatility process associated with these extreme values are much greater under the SV-normal than under the SV-t model. This is because the SV-t model has a larger persistence than the SV-normal model. In most cases, the estimated volatilities from the SV-t model exhibit smoother movements than those from the SV-normal model.

2.4.3 Model Diagnostics

The estimates of the parameters in both the SV-normal and SV-t models are reasonable for most of the return series. However, the SV-t model performs slightly better than the SV-normal model in capturing the regularities of the financial return series mentioned earlier. To validate the fitted model, we performed model diagnostics for the series Nikkei 225, S&P500, and IBM. Under the normal assumption for u_t, consider the standardized residual at time t from Eq. (2.1):

$$\hat{u}_t \equiv \frac{y_t}{\hat{\sigma}_t}, \qquad t = 1, ..., T, \tag{2.35}$$

where y_t is the observed return and $\hat{\sigma}_t = \exp\{\hat{\lambda}_t/2\}$, which is the estimated volatility displayed in Figs. 2.4 and 2.5. If the fitted model is adequate for the data, $\hat{u}_t$ follows a standard normal distribution under the standard normal assumption and a Student-t distribution with zero mean and unit variance under the scaled Student-t distributional assumption.

Table 2.4: $LB_u(30)$: Ljung-Box Statistics for standardized residual $\hat{u}_t$ along with p-value

Assumption	NIKKEI 225	S&P500	IBM
Normal	31.8700	42.6106	42.9798
	(0.3736)	(0.0634)	(0.0588)
Student-t	35.1249	42.5384	42.5903
	(0.2381)	(0.0525)	(0.0636)

Figure 2.6 displays the diagnostic plots of $\hat{u}_t$ for the Nikkei 225, S&P500, and IBM series

under the normal assumption. The residual plots in the first row show that the behavior of the residuals is quite similar to that of the actual series. In particular, large jumps still appear for the S&P500 and IBM series. For S&P500, the ACF plot of $\hat{u}_t$ shows a slightly large spike at lag 1. However, its value is small (< 0.05), so it can be ignored. For Nikkei 225 and IBM, there are no significant autocorrelations in the standardized residuals. As a reference, the Ljung-Box statistics are given in Table 2.4. All three p-values are greater than 0.05. Thus, the standardized residuals $\hat{u}_t$ tend to be uncorrelated for the three series except maybe for the S&P500 series. To check normality, QQ plots of $\hat{u}_t$ are given in the last row. There are large deviations in both the left and right tails of the QQ plots for all the data. This indicates a violation of the normality assumption for $\hat{u}_t$.

Figure 2.7 exhibits the diagnostic plots of $\hat{u}_t$ when u_t has a Student-t distribution with 8 degrees of freedom. In comparison with the residual plots under the SV-normal model, the variation of the standardized residuals for the series seems to be constant. For the ACF plots of $\hat{u}_t$, patterns similar to those under the SV-normal model are observed. Thus, the residuals are uncorrelated for the series. The QQ plots in the last row display the sample quantiles versus the quantiles of the Student-t distribution with 8 degrees of freedom. Although there are small deviations from the straight lines, the sizes of the deviations are smaller than in the normal case. Therefore, we conclude that the heavy-tailed distributional assumption produces a better fit to the Nikkei 225, S&P500, and IBM series.

2.4.4 Estimation Results: MCMC

The initial values of the parameters are $(\Lambda, \alpha, \beta, \gamma) = (\overrightarrow{0}, 0.01, 0.95, 0.14)$, where $\overrightarrow{0}$ is a $1 \times T$ zero vector. The number of iterations for the log variance and parameters are 500 and 1000, respectively. We repeat Steps 1 to 4, described in Section 2.3.2, 100,000 times so that we have 100,000 sweeps. The first 5,000 iterations are taken as a burn-in period.

The parameter estimation results under the SV-normal model are displayed in Table 2.5. The standard errors of the parameters are small, and thus the estimated values are statistically reasonable. For the persistent parameter β, the estimated values are slightly larger than those under the SV-normal model via SML, but they are still reasonably close. This implies that SML performs as efficiently as MCMC, which is consistent with Danielsson's finding (1994). However, the computing time (in seconds) of the MCMC algorithm is longer than that of SML: approximately thirty times longer for our data set. This may be the result of the large number of iterations: 100,000 sweeps and at each sweep 500 repetitions for volatility and 1,000 repetitions for a

Table 2.5: MCMC estimation of SV-normal model

Parameter	$/YEN	S&P500	IBM	HONDA	NIKKEI 225	CIBC
α	0.0002	0.0000	0.0001	0.0000	0.0000	-0.0001
MC std. error	0.0054	0.0006	0.0003	0.0014	0.0014	0.0025
Std. error	0.0132	0.0014	0.0482	0.0468	0.0468	0.0181
β	0.9791	0.9803	0.9752	0.9806	0.9806	0.9811
MC std. error	0.0004	0.0010	0.0008	0.0006	0.0006	0.0016
Std. error	0.0171	0.0328	0.0493	0.0092	0.0092	0.0328
γ	0.1602	0.1892	0.1749	0.1349	0.1349	0.1613
MC std. error	0.0009	0.0012	0.0014	0.0007	0.0007	0.0041
Std. error	0.0479	0.0524	0.0421	0.0329	0.0329	0.0116
$\hat{\sigma}$	0.6205	0.9176	0.6245	0.4736	0.4736	0.6949
$\hat{\kappa}$	5.5793	7.5097	5.6018	4.8174	4.8174	6.0103
CPU time*	6279.4	6237.3	7368.9	6307.4	6307.4	6049.3

parameter. In general, 1,000,000 or more sweeps and more iterations are required to minimize the dependence of the samples. However, this inefficiency of the MCMC method can be reduced by the improved MCMC algorithm (Kim *et al.* (1998)). The computing time could also be reduced by the use of a different computer language such as C/C++.

In the filtering for volatilities, $M = 1000$, which is the number of samples for λ_t for $t = 1, ..., T$. Figures 2.8 and 2.9 display the estimated volatility of the six series with the MCMC parameter estimates. Similarly to the SML volatility estimation shown in Figs. 2.4 and 2.5, sudden large jumps of the return series are clearly captured by a particle filter for most of the return series. Overall, the filtered volatilities correspond fairly well with the movement of the returns.

2.5 Concluding Remarks

We have implemented SML to estimate the standard SV model in the context of six financial return series. Consistent with the well-known empirical regularities of return series, the six series are leptokurtic and have no serial correlations in the series, but serial correlations exist in the squares. Volatility clustering is also observed. For the parameter

estimation, the EIS algorithm was employed for the SV-normal and SV-t models. The kurtosis implied by the SV-t model is generally much closer than that of the SV-normal model to the sample kurtosis. The model diagnostics indicate that the SV-t specification is more appropriate than the SV-normal for most of our data.

In particular, the S&P500 and IBM series have extremely large sample kurtoses. This is because of the existence of jumps in the series during the Black Monday period. To model these extreme values more precisely, one could consider a model with a jump component. On the other hand, if the unusual values are excluded, the tail behaviors of these series are similar to those of the other series.

To compare the performances of SML and MCMC, we also implemented the MCMC method with the same data set. The estimation results obtained by SML and MCMC are similar and statistically precise in the sense of a small standard error of the parameter estimates. Despite the efficiency of SML and MCMC, both methods have to be carefully implemented. For SML, the common random numbers (CRN) have been used to accelerate the convergence of EIS. For MCMC, a more efficient algorithm should be applied to effectively draw a latent variable Λ_T. In our empirical study, SML has an edge on MCMC in terms of computing time.

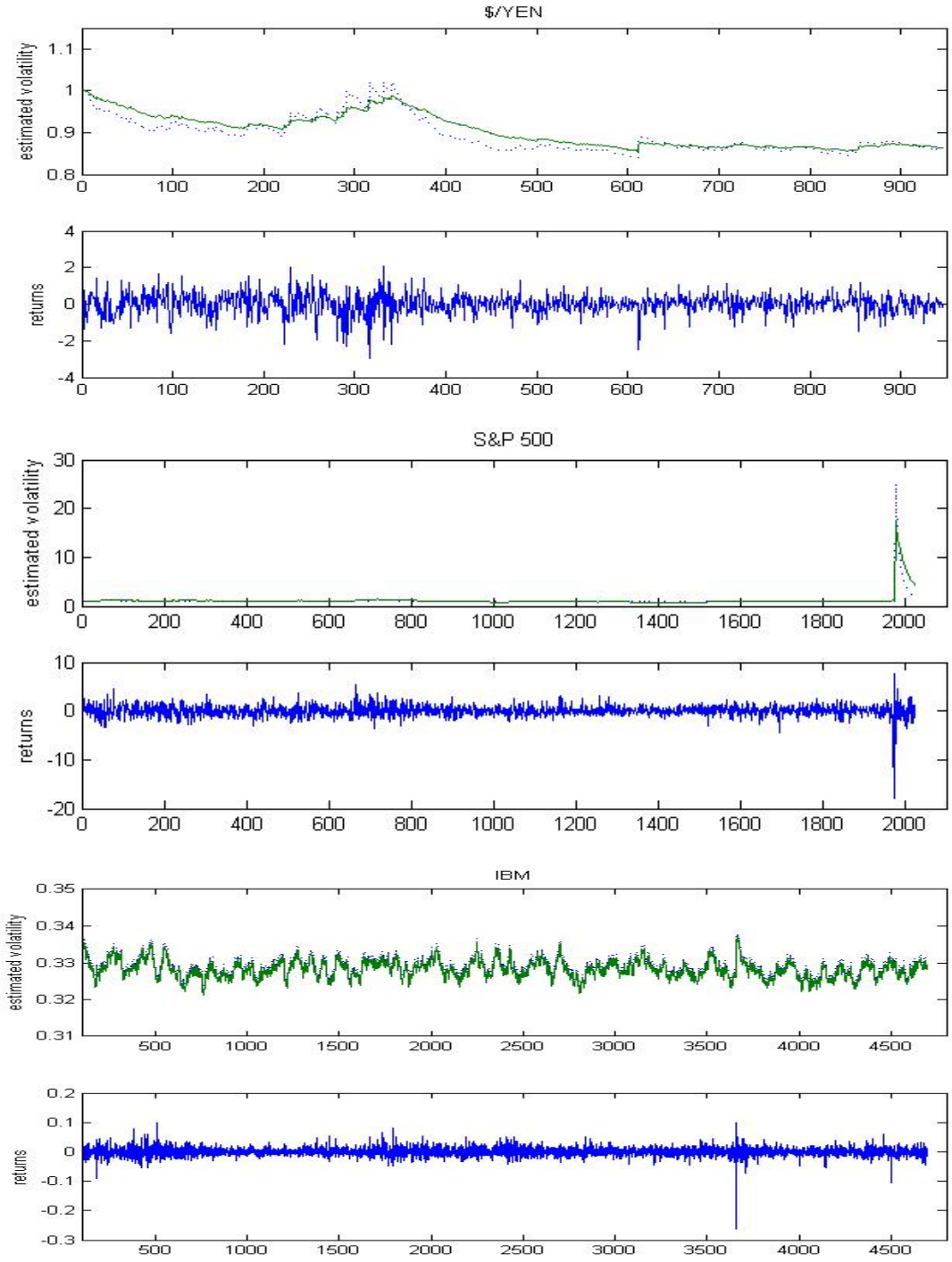

Figure 2.4: SML: Top panels: return series; Bottom panels: estimated volatilities for $/YEN, S&P500, and IBM

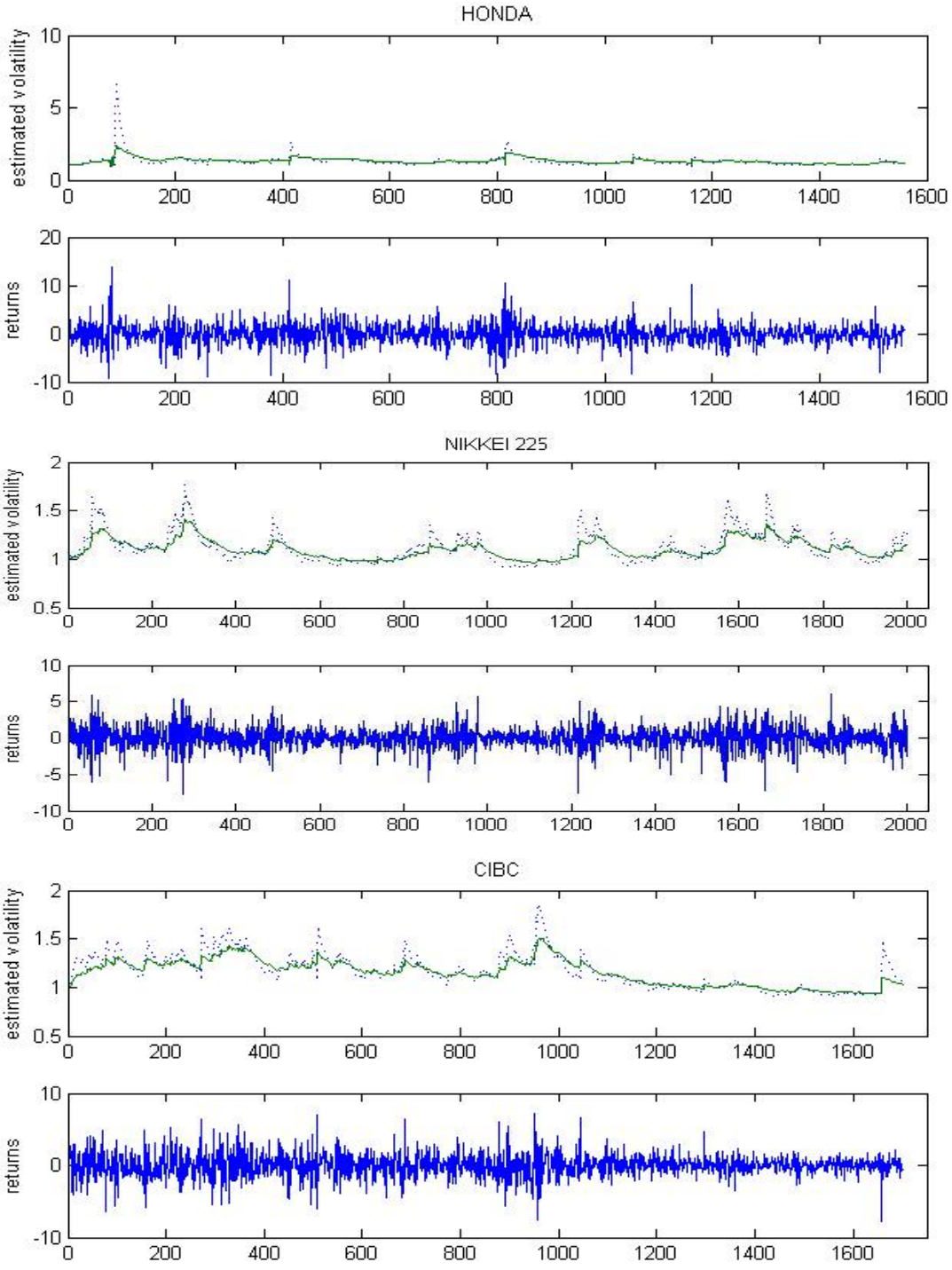

Figure 2.5: SML: Top panels: return series; Bottom panels: estimated volatilities for Honda, Nikkei 225, and CIBC

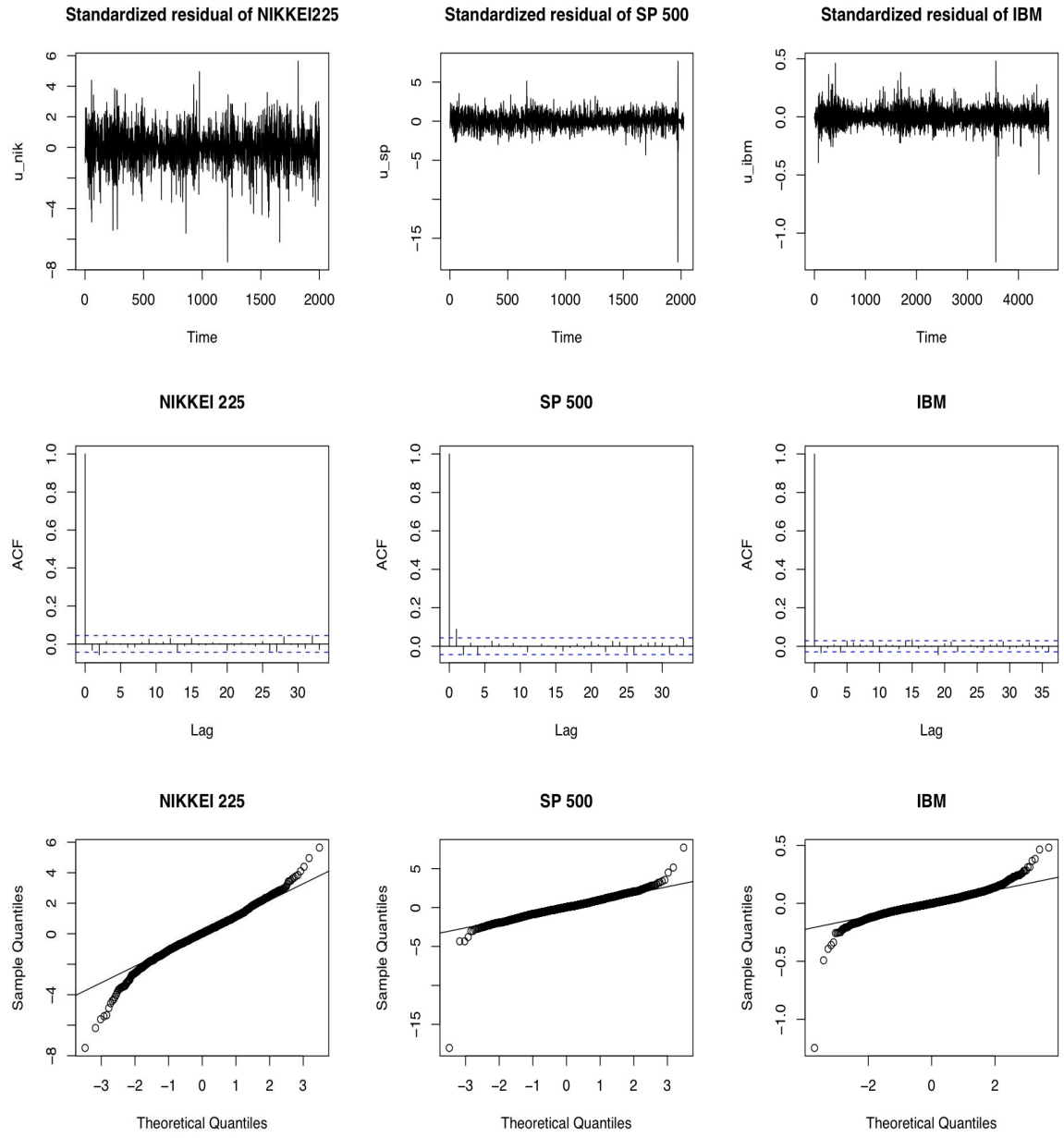

Figure 2.6: SV-normal model via SML: the residual plots are in the first row, the ACFs of $\hat{u}_t$ are in the second row, and the QQ plots are in the last row.

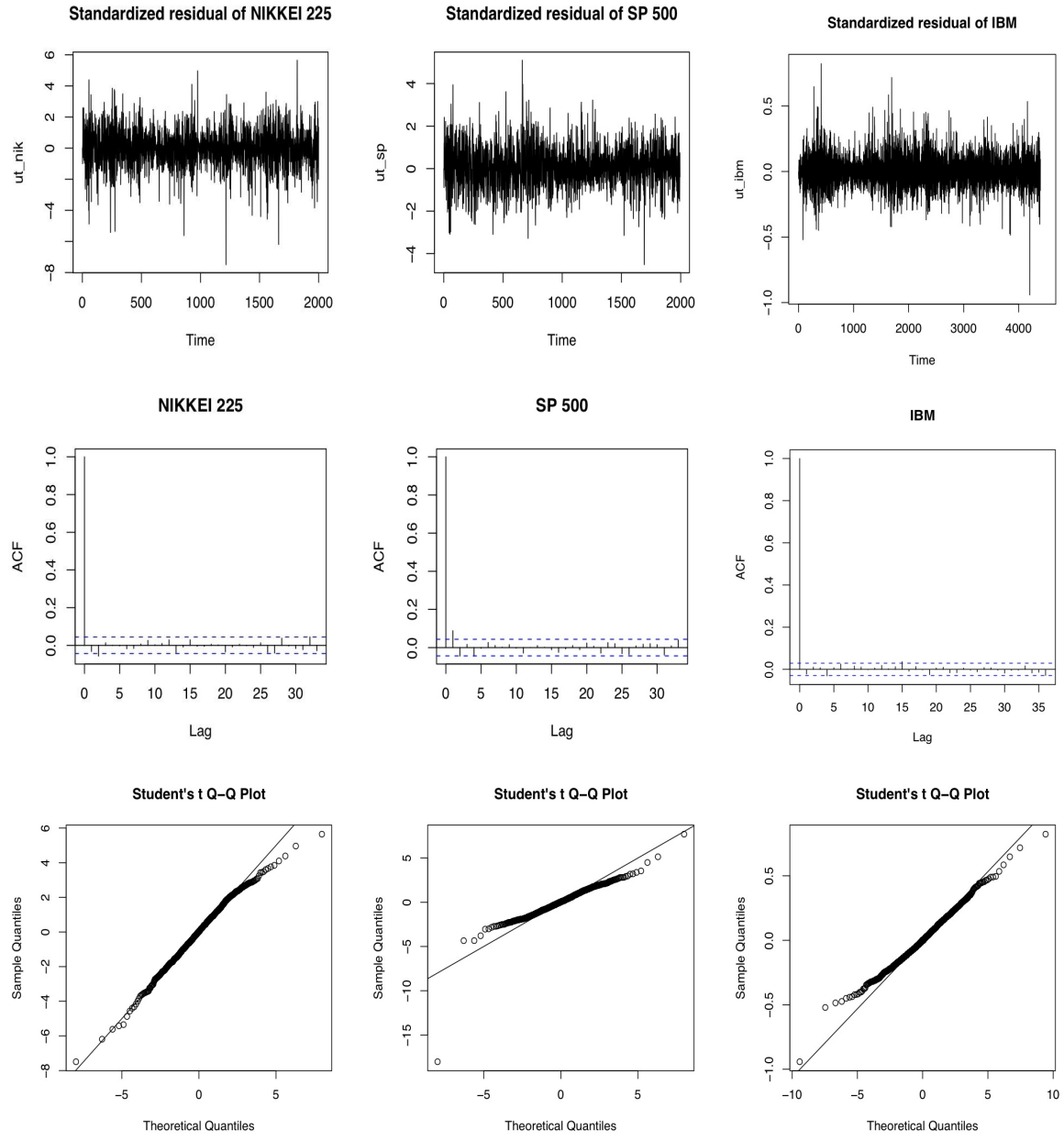

Figure 2.7: SV-t model via SML: the residual plots are in the first row, the ACFs of $\hat{u}_t$ are in the second row, and the QQ plots are in the last row.

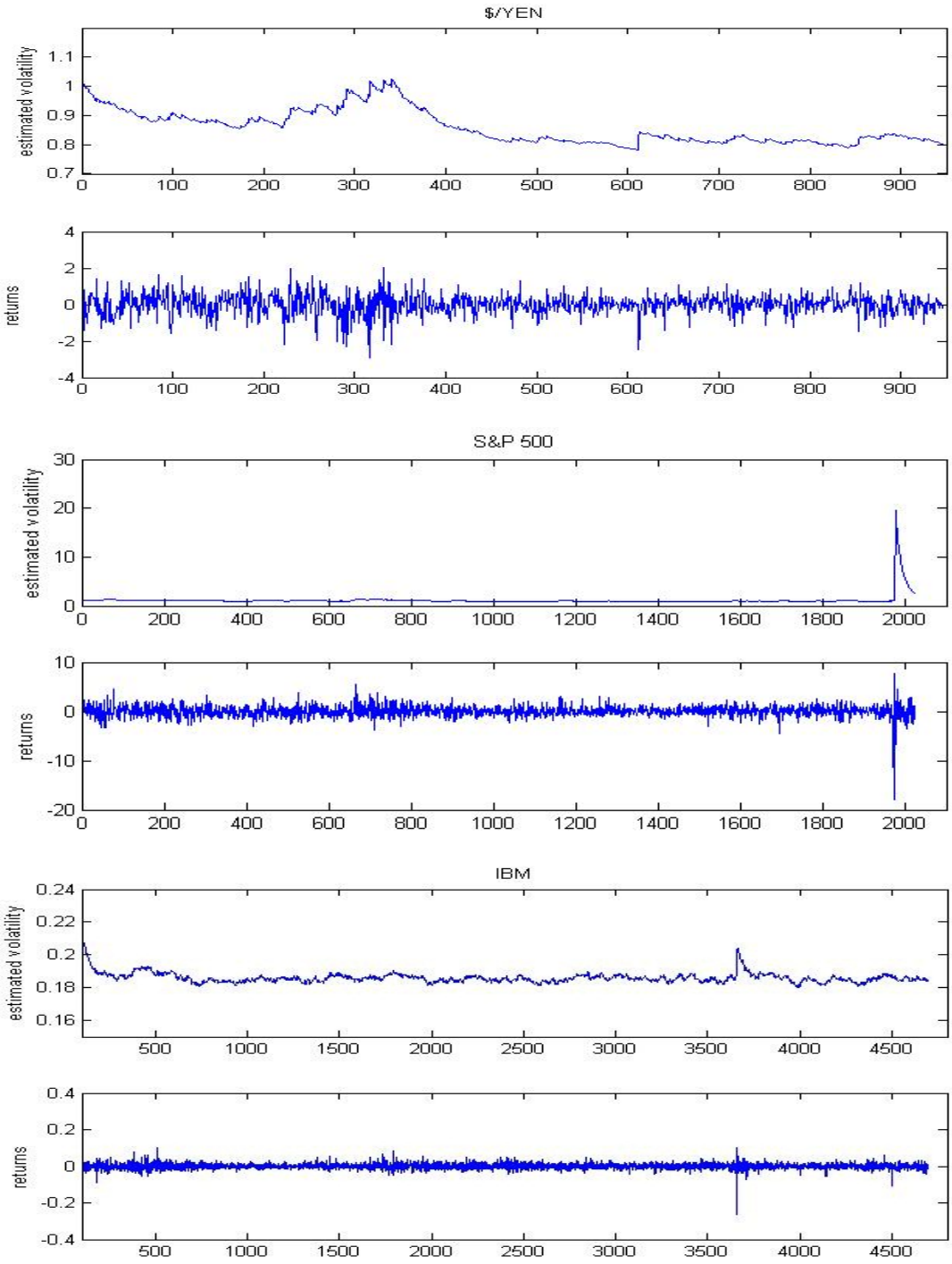

Figure 2.8: MCMC: Top panels: return series; Bottom panels: estimated volatilities for $/YEN, S&P500, and IBM

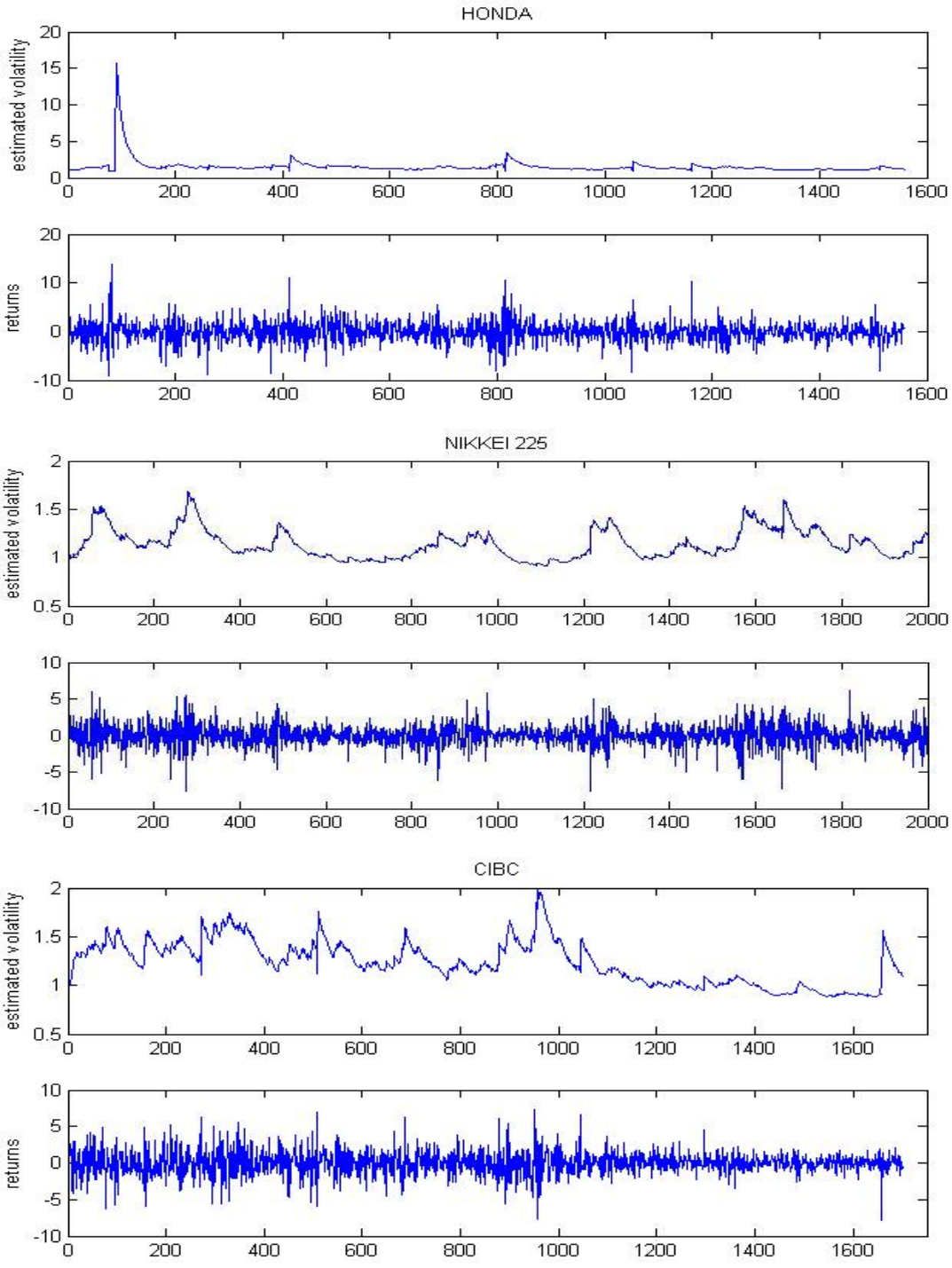

Figure 2.9: MCMC: Top panels: return series; Bottom panels: estimated volatilities for HONDA, Nikkei 225, and CIBC

Chapter 3

Autoregressive Models with Stochastic Volatility Errors

ARCH and SV models are designed to analyze and model the conditional variance (volatility), but the specification of the conditional mean is still important. If the conditional mean is not appropriately specified, it is impossible to find a consistent estimator of the variance or volatility process. For instance, consider the standard SV model. If the conditional mean of the data is captured by an adequate model, the residuals of the model satisfy the assumption for the error u_t, which provides a conditional variance for the series y_t. This can be considered to be the joint estimation of two models: that for the conditional mean and that for the conditional variance.

Several extended models have been developed and their moment properties have been studied. Anderson and Bollerslev (1998) and Tsay (2002) considered the ARMA–GARCH(1,1) model to fit the US dollar and DM exchange rate and the S&P 500 index, respectively. The autocorrelation function (ACF) of the squared series in the ARMA–GARCH model was derived by Karanasos (1999) and He and Teräsvirta (1999). Karanasos (2001) showed the ACF of the series in the ARMA–GARCH-in-mean model. For other combination models, Demos (2002) derived the ACF of a model combining the EGARCH and SV models, and Karanasos and Kim (2003) studied the moments of the ARMA–EGARCH model. Modified ARMA-GARCH models are also studied by Goldman and Tsurumi (2005), Yoo (2010), and Pati and Rajib (2010).

In this chapter, we consider a combination model in which the conditional mean is

modeled by an AR(p) process and the conditional variance is modeled by the SV model. We refer to this as the AR(p)–SV model. Murphy and Izzeldin (2005) in an unpublished manuscript examined the size and power properties of some long memory tests using bootstrap with observations generated from few models including an autoregressive model of order 1 with a Stochastic Volatility error model. Koopman and Bos (2004) considered state space models with stochastic error variances and some special cases of these models could be linked to ARMA models. Bos and Shephard (2006) discussed similar models and implemented MCMC. Pellegrini (2009) in an unpublished thesis considered unobserved component local trend models in which the errors had the traditional SV structure which they referred to as ARSV(1) structure. The unobserved component local trend models that they used are equivalent to ARIMA(0,1,1) or ARIMA(0,2,2) models. We focus in this chapter on the moment properties, exact likelihood, and SML estimation using EIS of AR(p)–SV model.

In spite of the simple structure of the AR(p)–SV model, its estimation is not trivial because of the existence of the latent log-variance process, as in the SV model. To efficiently estimate the model, we first consider the estimation methods used for the SV model, such as the generalized method of moments (Hansen (1982)), the efficient method of moments (Gallant *et al.* (1997)), the Markov chain Monte Carlo (MCMC) (Jacquier *et al.* (1994) and Kim *et al.* (1998)), and the simulated maximum likelihood (SML) (Danielsson and Richard (1993)). Based on our empirical study in Chapter 2, both methods are powerful and efficient, but SML has an edge over MCMC. Thus, we use the SML approach to estimate the AR(p)–SV model. We perform a simulation study to investigate the performance of the SML with the EIS and a sensitivity analysis to examine the robustness of our estimation method.

To choose the order of the AR process in the AR–SV model, we consider model selection criteria such as the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). This is similar to Karanasos and Kim's (2003) approach in the selection of the ARMA–EGARCH model. Residual analysis is performed with the fitted model.

This chapter is organized as follows. The standard AR(p)–SV model is introduced in Section 1, and the moment properties of the model are studied in Section 2. Section 3 describes the SML method with EIS for estimating the AR(p)–SV model and a particle filter for volatility estimation. We summarize the order determination and diagnostic checking in Section 4. Section 5 gives a simulation study and a sensitivity analysis, and Section 6 presents an empirical study with actual data.

3.1 AR(p)–SV model

The AR(p)–SV model is designed to capture the conditional mean with a standard AR process and model the conditional variance with the SV model. For any financial time series y_t, a standard AR(p)–SV model is defined as

$$\Phi(B)y_t = \epsilon_t, \tag{3.1}$$

$$\epsilon_t = \exp\{\lambda_t/2\}u_t, \tag{3.2}$$

$$\lambda_t = \alpha + \beta\lambda_{t-1} + \gamma v_t, \tag{3.3}$$

where

$$\Phi(B) = \sum_{i=0}^{p} \phi_i B^i$$

for $t = 1, ..., T$. $\phi_0 = 1$ and B is a backshift operator such as $By_t = y_{t-1}$. The error processes u_t and v_t are uncorrelated white noises with zero mean and unit variance. v_t is assumed to be normally distributed, but the assumption for u_t can be a normal or heavy-tailed distribution. The processes u_t and v_t are also uncorrelated with λ_1, where $\lambda_1 \sim N(\frac{\alpha}{1-\beta}, \frac{\gamma^2}{1-\beta^2})$. The coefficient $\gamma > 0$ is the conditional standard deviation of λ_t given λ_{t-1}.

For a stationary process y_t, the following assumptions are needed.

Assumption 3.1. *The absolute value of the persistent parameter β is less than 1, i.e., $|\beta| < 1$.*

Assumption 3.2. *All the roots of the autoregressive polynomial $\Phi(B)$ lie outside the unit circle.*

Assumption 3.1 ensures that the latent log-variance process λ_t is a stationary AR(1) process. Throughout this chapter, these two assumptions always hold. Compared with other time-varying volatility models, the AR(p)–SV model has at least two advantages in applications. First, it is designed to account for serial correlation in the series y_t with the AR process in Eq. (3.1), so that the model can be applied to any financial time series and not only the return series. In contrast, the series y_t is assumed to be uncorrelated in a standard SV model. Second, the conditional variance can be captured more explicitly by the SV structure in Eqs. (3.2) and (3.3).

3.2 Moment Properties

The kurtosis and ACF of the AR process are well-known (*for example, see* Box and Jenkins (1976)). For the SV model, most of the important properties have been discussed by Taylor (1986, 1994). Jacquier *et al.* (1994) showed the covariance of the squared return series, and the ACF of the squared returns was expressed as a function of the kurtosis by Teräsvirta (1996).

In this section, we discuss the moment structure of the AR(p)–SV model: the kurtosis, the ACF of the series, the ACF of the squared series, and the cross-correlation. First, we derive the moments of any stationary process, and then we specialize the results to the AR(p)–SV model. Furthermore, we derive the exact likelihood function for this model.

3.2.1 Kurtosis

Kurtosis is a basic statistical tool to measure the thickness of the tail for the distribution. To investigate the tail behavior of the AR(p)–SV model, we study the kurtosis of y_t.

Proposition 3.1. *Suppose that a linear stationary process y_t has a moving-average representation of the form:*

$$y_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}, \tag{3.4}$$

where ϵ_t is an uncorrelated sequence with $E[\epsilon_t] = 0$ and $Var[\epsilon_t] = \sigma_\epsilon^2$. Then, the second and fourth moments of y_t are given by

$$E[y_t^2] \;=\; Var[y_t] \;=\; \sigma_\epsilon^2 \sum_{j=0}^{\infty} \psi_j^2, \tag{3.5}$$

$$E[y_t^4] \;=\; E[\epsilon_t^4] \sum_{j=0}^{\infty} \psi_j^4 + 6[E[\epsilon_t^2]]^2 \sum_{i<j} \sum \psi_i^2 \psi_j^2, \tag{3.6}$$

provided $E[\epsilon_t^4] < \infty$. In addition, the kurtosis of y_t in (3.4) is defined by

$$\kappa = \frac{E[y_t^4]}{E[y_t^2]^2} = \frac{\kappa^{(\epsilon)} \sum_{j=0}^{\infty} \psi_j^4 + 6 \sum \sum_{i<j} \psi_i^2 \psi_j^2}{(\sum_{j=0}^{\infty} \psi_j^2)^2},$$

where $\kappa^{(\epsilon)}$ is the kurtosis of ϵ_t.

Proof. For the moments, note that

$$y_t^2 = \sum_{j=0}^{\infty} \psi_j^2 \epsilon_{t-j}^2 + 2 \sum \sum_{i<j} \psi_i \psi_j \epsilon_{t-i} \epsilon_{t-j}, \tag{3.7}$$

$$y_t^4 = \sum_{j=0}^{\infty} \psi_j^4 \epsilon_{t-j}^4 + 6 \sum \sum_{i<j} \psi_i^2 \psi_j^2 \epsilon_{t-i}^2 \epsilon_{t-j}^2 + 4\, \delta(\epsilon) + 4\, \omega(\epsilon), \tag{3.8}$$

where

$$\omega(\epsilon) = \sum_{i<j<k<l} \sum \sum \sum \psi_i \psi_j \psi_k \psi_l \epsilon_{t-i} \epsilon_{t-j} \epsilon_{t-k} \epsilon_{t-l},$$

$$\delta(\epsilon) = \sum_{i<j} \psi_i^3 \psi_j \epsilon_{t-i}^3 \epsilon_{t-j} + \sum_{i<j} \sum \psi_i \psi_j^3 \epsilon_{t-i} \epsilon_{t-j}^3 + \sum_{i<j<l} \sum \sum \psi_i \psi_j \psi_l^2 \epsilon_{t-i} \epsilon_{t-j} \epsilon_{t-l}^2.$$

Since the ϵ_t are uncorrelated, $E[\epsilon_{t-i}\epsilon_{t-j}] = E[\epsilon_{t-i}^3 \epsilon_{t-j}] = 0$ for all $i \neq j$. Thus, moments $E[y_t^2]$ and $E[y_t^4]$ are easily obtained by taking the expectation on both sides of Eqs. (3.7) and (3.8).

For the kurtosis, substitute Eqs. (3.5) and (3.6) into the definition of kurtosis. QED □

Proposition 3.1 is true for any stationary process expressed as (3.4). Therefore, the moments of y_t for the AR(p)–SV model can be obtained in a similar fashion as follows. Under Assumptions 3.1 and 3.2, Eq. (3.1) has the following representation:

$$y_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}, \tag{3.9}$$

where $\psi_0 = 1$. Other ψ_j's for $j = 1, ..., p$ are obtained by comparing the coefficients in the identity $(1 + \psi_1 B + \psi_2 B^2 + \cdots + \psi_p B^p)(1 - \psi_1 B - \psi_2 B^2 - \cdots - \psi_p B^p) = 1$. In contrast to a conventional stationary process, the error process ϵ_t for the AR(p)–SV model satisfies the SV model specification in (3.2) and (3.3). The following results hold for the SV error process ϵ_t (*see* Eqs. (2.4)–(2.6)).

Provided $E[u_t^4] < \infty$, the second and fourth moments of ϵ_t are given by

$$E[\epsilon_t^2] \;=\; E[e^{\lambda_t}] \;=\; \exp\{\mu + \sigma^2/2\}, \tag{3.10}$$

$$E[\epsilon_t^4] \;=\; E[e^{2\lambda_t} u_t^4] \;=\; \exp\{2(\mu + \sigma^2)\} E[u_t^4], \tag{3.11}$$

where $\mu = \frac{\alpha}{1-\beta}$ is the unconditional mean of λ_t and $\sigma^2 = \frac{\gamma^2}{1-\beta^2}$ is the unconditional variance of λ_t. Moreover, the kurtosis of ϵ_t is defined as

$$\kappa^{(\epsilon)} = \kappa^{(u)} \exp\{\sigma^2\}, \tag{3.12}$$

where $\kappa^{(u)}$ is the kurtosis of the error process u_t.

Depending on the distributional assumption for u_t, the kurtosis $\kappa^{(u)}$ has different expressions. For instance, if u_t is normally distributed, $\kappa^{(u)} = 3$; if u_t follows a Student-t distribution with ω degrees of freedom, then $\kappa^{(u)} = 3(\omega-2)/(\omega-4)$. Under the generalized-error distributional assumption for u_t, the kurtosis is $\kappa^{(u)} = \Gamma(1/\nu)\Gamma(5/\nu)/[\Gamma(3/\nu)]^2$ (Johnson and Kotz (1970)). Using Eqs. (3.10)–(3.12), we can obtain the moments and kurtosis of y_t for the AR(p)–SV model.

Theorem 3.1. *Suppose that a financial time series y_t follows an AR(p)–SV model. Then the second and fourth moments of y_t are given by*

$$E[y_t^2] = \exp\{\mu + \sigma^2/2\} \sum_{j=0}^{\infty} \psi_j^2, \tag{3.13}$$

$$E[y_t^4] = \exp\{2(\mu + \sigma^2)\} E[u_t^4] \sum_{j=0}^{\infty} \psi_j^4 + 6 \exp\{\mu + \sigma^2/2\} \sum \sum_{i<j} \psi_i^2 \psi_j^2, \tag{3.14}$$

provided $E[u_t^4] < \infty$. In addition, the kurtosis of y_t is defined by

$$\kappa = \frac{\kappa^{(u)} \exp\{\sigma^2\} \sum_{j=0}^{\infty} \psi_j^4 + 6 \sum \sum_{i<j} \psi_i^2 \psi_j^2}{(\sum_{j=0}^{\infty} \psi_j^2)^2}, \tag{3.15}$$

where $\kappa^{(u)}$ is the kurtosis of u_t.

Proof. Substitute the results for ϵ_t in (3.12) into κ in *Proposition* 3.1. QED □

For a special case, consider the AR–SV model of order 1 with the normal assumption for u_t. In the AR(1)–SV model, it can easily be seen that $\psi_j = \phi_1^j$ in (3.9). Thus, the kurtosis of y_t is given by

$$\kappa = \left(\frac{1 - \phi_1^2}{1 + \phi_1^2} \right) [3 \exp\{\sigma^2\} + 6\phi_1^2]. \tag{3.16}$$

3.2.2 Autocorrelations of the Series and Squared Series

Next, we consider the behavior of ACFs of the series y_t and the squared series y_t^2. Like any other ARMA process with a heteroscedastic error process, such as the ARMA–GARCH model, the ACF of the series in the AR(p)–SV model has the same expression as a conventional AR(p) process. The autocovariance and ACF of y_t for the AR–SV model are given in the following proposition. The proofs of the proposition and the theorem are standard.

Proposition 3.2. *Suppose a financial time series y_t follows the AR–SV model. Then the autocovariance of y_t at lag k is given by*

$$\gamma_k = cov(y_t, y_{t-k}) = \sum_{j=0}^{\infty} \psi_j \psi_{j+k} \sigma_\epsilon^2,$$

and thus the ACF of y_t at lag k is given by

$$\rho_y(k) = \frac{\sum_{j=0}^{\infty} \psi_j \psi_{j+k}}{\sum_{j=0}^{\infty} \psi_j^2} \tag{3.17}$$

for any $k > 0$.

Proposition 3.2 implies that in the AR(p)–SV model, the autocorrelation of the series is described by the AR part only. In contrast to the ACF of y_t, the autocorrelation of y_t^2 is impacted by the error process ϵ_t, which is governed by the SV structure in the AR(p)–SV model. The following theorem is a special case of Corollary 1 in Palma and Zevallos (2004), where the error ϵ_t has finite kurtosis, but satisfies the SV specification.

Theorem 3.2. *Suppose a financial time series y_t follows the AR(p)–SV model. Provided $E[u_t^4] < \infty$, the ACF of y_t^2 at lag k is given by*

$$\rho_{y^2}(k) = \frac{2}{\kappa - 1} \rho_y^2(k) + \frac{\kappa - 3}{\kappa - 1} \left(\frac{\sum_{j=0}^{\infty} \psi_j^2 \psi_{j+k}^2}{\sum_{j=0}^{\infty} \psi_j^4} \right), \tag{3.18}$$

where κ is the kurtosis of y_t in (3.15).

In addition, we consider the cross-correlation between the levels and the squared series. Such a measure is useful to understand volatility asymmetry. Demos (2002) derived the cross-correlations in the time-varying parameter generalized stochastic volatility model in

mean (TVP-GSV-M), and Karanasos and Kim (2003) derived the cross-correlation between the level and the squared series for the ARMA–EGARCH model. Similarly, we obtain the cross-correlation between the level and the squared series for the AR(p)–SV model.

Theorem 3.3. *Suppose a financial time series y_t follows the AR(p)–SV model. Provided $E[\epsilon_t] < \infty$, the cross-correlation between the level and the squared series is given by*

$$\rho(y_t^2, y_{t-k}) = \frac{E[y_t^2 y_{t-k}]}{\sqrt{V[y_t^2]E[y_t^2]}},$$

where

$$E[y_t^2 y_{t-k}] = \sum_{j=0}^{l+k-1} \sum_{l=0}^{\infty} \psi_j^2 \psi_l \; \rho(\epsilon_t^2, \epsilon_{t-(l+k-j)})(\kappa^{(\epsilon)} - 1)^{\frac{1}{2}} E[\epsilon_t^2]^{\frac{3}{2}}, \tag{3.19}$$

$$V[y_t^2] = (\kappa - 1)E[y_t^2]^2,$$

and

$$E[y_t^2] = \sigma_\epsilon^2 \sum_{j=0}^{\infty} \psi_j^2.$$

Proof. For the stationary AR(p)–SV process y_t given in (3.9), the product of y_t and y_{t-k} is written as

$$y_t^2 \; y_{t-k} = \sum_{j=0}^{\infty} \sum_{l=0}^{\infty} \psi_j^2 \psi_j \epsilon_{t-j}^2 \epsilon_{t-k-l} + 2 \sum \sum_{i<j} \sum_{l=0}^{\infty} \psi_i \psi_j \psi_l \epsilon_{t-i} \epsilon_{t-j} \epsilon_{t-l}.$$

Because there is no correlation with ϵ_t, the expectation of the product is given by

$$E[y_t^2 \; y_{t-k}] = \sum_{j=0}^{\infty} \sum_{l=0}^{\infty} \psi_j^2 \psi_j E[\epsilon_{t-j}^2 \epsilon_{t-k-l}]. \tag{3.20}$$

According to Eqs. (B.3) and (B.4) in Karanasos and Kim (2003), we have

$$\begin{aligned} E[\epsilon_{t-j}^2 \; \epsilon_{t-k-l}] &= \begin{cases} E[\epsilon_{t-j} \; \epsilon_{t-(k+l-j)}], & \text{if } j < l+k, \\ 0, & \text{otherwise} \end{cases} \\ &= \begin{cases} \rho(\epsilon_{t-j}, \epsilon_{t-(k+l-j)})(\kappa^{(\epsilon)} - 1)^{1/2} E[\epsilon_t^2]^{3/2}, & \text{if } j < l+k, \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

Substituting the above equation into (3.20), we obtain the autocovariance of y_t^2 and y_{t-k}, satisfying (3.19). QED □

3.2.3 Exact Likelihood Function

Similarly to the derivation of the exact likelihood function for an AR process, we can derive the likelihood function for the AR(p)–SV model. The likelihood associated with $Y_T = (y_1, ..., y_T)$ in (3.1), (3.2), and (3.3), for $p \leq T$, is defined by

$$L(\xi; Y_T) = \int_{R^T} f(Y_T, \Lambda_T|\xi) d\Lambda_T \tag{3.21}$$

$$= \int_{R^T} f(Y_T|\Lambda_T, \xi) f(\Lambda_T|\xi) d\Lambda_T, \tag{3.22}$$

where T is the number of observations and $\Lambda_T = (\lambda_1, ..., \lambda_T)$ is a vector of the latent variable λ_t. The parameter vector $\xi = (\Phi, \alpha, \beta, \gamma)$ is to be estimated, where $\Phi = (\phi_1, ..., \phi_p)$. The density function $f(\Lambda_T|\xi)$ is a product of $f(\lambda_t|\lambda_{t-1}, \xi)$ for all t and the conditional distribution of λ_t given λ_{t-1} is normal with mean $\alpha+\beta\lambda_{t-1}$ and variance γ^2. The conditional density function $f(Y_T|\Lambda_T, \xi)$ can be factorized as

$$f(Y_T|\Lambda_T, \xi) = f(y_{p+1}, ..., y_T|Y_p, \xi, \Lambda_T) f(y_1, ..., y_p|\xi, \Lambda_T). \tag{3.23}$$

Because the AR(p)–SV model is a combination of the AR(p) process and the SV model, the likelihood function in (3.21) has a similar functional form to that for the AR(p) process. One notable difference between the AR(p) and AR(p)–SV models arises from the error process. The error ϵ_t in the AR(p) process is a Gaussian white noise, while it is a stochastic volatility error for the AR(p)–SV model. More details on the functional forms of the likelihood function are given below.

Suppose that the error process u_t in (3.2) is normally distributed with mean zero and unit variance. In the special case of order 1 ($p = 1$), let $\sigma_t^2 = \exp\{\lambda_t\}$. Then, the conditional density function $f(Y_T|\Lambda_T, \xi)$ is written as

$$f(Y_T|\Lambda_T, \xi) = f(y_2, ..., y_T|y_1, \xi, \Lambda_T) f(y_1|\xi, \Lambda_T),$$

where

$$f(y_1|\xi, \Lambda_T) = \frac{\sqrt{1-\phi_1^2}}{\sigma_1\sqrt{2\pi}} \exp\left[-\frac{y_1^2(1-\phi_1^2)}{2\sigma_1^2}\right]$$

and

$$\begin{aligned} f(y_2, ..., y_T|y_1, \xi, \Lambda_T) &= \prod_{t=2}^{T} f(y_t|y_{t-1}, \lambda_t) \\ &= (2\pi)^{-\frac{T-1}{2}} \left(\prod_{t=2}^{T} \sigma_t \right)^{-1} \exp\left[-\sum_{t=2}^{T} \frac{(y_t - \phi_1 y_{t-1})^2}{2\sigma_t^2} \right]. \end{aligned}$$

Similarly, the conditional density $f(Y_T|\Lambda_T, \xi)$ for the AR(2)–SV model is written as

$$f(Y_T|\Lambda_T, \xi) = f(y_3, ..., y_T|Y_2, \xi, \Lambda_T) f(y_1, y_2|\xi, \Lambda_T),$$

where

$$\begin{aligned} f(y_1, y_2|\xi, \Lambda_T) &= \frac{(1+\phi_2)[(1-\phi_2)^2 - \phi_1^2]}{2\pi} * \\ &\quad \exp\left[\frac{(1-\phi_2^2)\sigma_1^2 y_1^2 - \phi_1(1+\phi_2)\sigma_1\sigma_2 y_1 y_2 (1-\phi_2^2)\sigma_2^2 y_2^2}{2\sigma_1^2\sigma_2^2} \right] \end{aligned}$$

and

$$f(y_3, ..., y_T|Y_2, \xi, \Lambda_T) = (2\pi)^{-\frac{T-2}{2}} \left(\prod_{t=3}^{T} \sigma_t \right)^{-1} \exp\left[-\sum_{t=3}^{T} \frac{(y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2})^2}{2\sigma_t^2} \right].$$

In general, the conditional density $f(y_1, ..., y_p|\xi, \Lambda_T)$ for $p \leq T$ is given by

$$f(y_1, ..., y_p|\xi, \Lambda_T) = (2\pi)^{-\frac{p}{2}} \left(\prod_{t=1}^{p} \sigma_t \right)^{-1} |M_p^{(p,0)}|^{\frac{1}{2}} \exp\left[-\sum_{t=1}^{p} \frac{1}{2\sigma_t^2} Y_p' M_p^{(p,0)} Y_p \right], \qquad (3.24)$$

where $M_p^{(p,0)}$ denotes the covariance matrix of Y_p such that

$$M_p^{(p,0)} = \{m_{ij}^{(p)}\} = (\Gamma^{(p)})^{-1}\sigma_p, \qquad (3.25)$$

where $\sigma_p = (\sigma_1, ..., \sigma_p)'$ and

$$\Gamma^{(p)} = \begin{pmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{p-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{p-2} \\ \vdots & \ddots & \ddots & \vdots \\ \gamma_{p-1} & \gamma_{p-2} & \cdots & \gamma_0 \end{pmatrix}.$$

Note that $\gamma_i = Cov(y_t, y_{t-i})$ for $i = 0, 1, ..., p$; it is different from the coefficient γ in (3.3).

The conditional density $f(y_{p+1}, ..., y_T|Y_p, \xi, \Lambda_T)$ is

$$f(y_{p+1}, ..., y_T|Y_p, \xi, \Lambda_T) = (2\pi)^{-\frac{T-p}{2}} \left(\prod_{t=p+1}^{T} \sigma_t \right)^{-1} \exp\left[- \sum_{t=p+1}^{T} \frac{(y_t - \sum_{i=1}^{p} \phi_i y_{t-i})^2}{2\sigma_t^2} \right]. \tag{3.26}$$

Thus, the conditional density $f(Y_T|\Lambda_T)$ is the product of (3.24) and (3.26).

Suppose that u_t in (3.2) has a scaled Student-t distribution with mean zero and unit variance. When $p = 1$, the conditional density function $f(Y_T|\Lambda_T, \xi)$, for $\omega > 2$, is factorized by

$$f(Y_T|\Lambda_T, \xi) = f(y_2, ..., y_T|y_1, \xi, \Lambda_T) f(y_1|\xi, \Lambda_T),$$

where

$$f(y_1|\xi, \Lambda_T) = \left(\frac{\pi(\omega-2)(1-\phi_1^2)}{\sigma_1^2} \right)^{-\frac{1}{2}} \frac{\Gamma\left(\frac{\omega+1}{2}\right)}{\Gamma\left(\frac{\omega}{2}\right)} \left[1 + \frac{y_1^2(1-\phi_1^2)}{\sigma_1^2(\omega-2)} \right]^{-\frac{(\omega+1)}{2}}$$

and

$$f(y_2, ..., y_T|y_1, \xi, \Lambda_T) = [\pi(\omega-2)]^{-\frac{T-2}{2}} \frac{\Gamma\left(\frac{\omega+T-2}{2}\right)}{\Gamma\left(\frac{\omega}{2}\right)} \prod_{t=2}^{T} \sigma_t^{-1} \left[1 + \frac{(y_t - \phi_1 y_{t-1})^2}{\sigma_t^2(\omega-2)} \right]^{-\frac{(\omega+1)}{2}}.$$

In general, for the order $0 < p < T$, the conditional density function $f(y_1, ..., y_p|\xi, \Lambda_T)$ (Tarami and Pourahmadi (2003)) is written as

$$f(y_1, ..., y_p|\xi, \Lambda_T) = (2\pi)^{-\frac{p}{2}} \left(\frac{\omega-2}{3} \right)^{-\frac{p}{2}} \frac{\Gamma\left(\frac{\omega+p}{2}\right)}{\Gamma\left(\frac{\omega}{2}\right)} |M_p^{(p,0)}|^{-\frac{1}{2}} \left(1 + \frac{Y_p' M_p^{(p,0)} Y_p}{\omega-2} \right)^{-\frac{(\omega+p)}{2}}, \tag{3.27}$$

where $M_p^{(p,0)}$ satisfies the relationship in (3.25). The conditional density $f(y_{p+1}, ..., y_T|Y_p, \xi, \Lambda_T)$ is written as

$$f(y_{p+1}, ..., y_T|Y_p, \xi, \Lambda_T) = [\pi(\omega-2)]^{-\frac{T-p}{2}} \frac{\Gamma\left(\frac{\omega+T-p}{2}\right)}{\Gamma\left(\frac{\omega}{2}\right)} \prod_{t=p+1}^{T} \sigma_t^{-1} \left[1 + \frac{(y_t - \sum_{t=p+1}^{T} \phi_i y_{t-i})^2}{\sigma_t^2(\omega-2)} \right]^{-\frac{(\omega+1)}{2}}. \tag{3.28}$$

Thus, the conditional density $f(Y_T|\Lambda_T)$ is the product of (3.27) and (3.28).

3.3 Estimation Methods

Although the model specification of the AR(p)–SV model is fairly simple, its estimation is not straightforward because of the latent log-variance process λ_t. This is the same obstacle as in the SV model. Thus, one can consider the estimation methods that are employed for the SV model. We use the simulated maximum likelihood (SML) approach (Danielsson and Richard (1993)). Since the AR(p)–SV model has the same structure as the SV model except that the conditional mean equation follows an AR process, the SML approach is applicable.

Consider the likelihood function associated with the observations Y_T for the AR(p)–SV model in (3.21). Since the log-variance process λ_t is not observed in practice, the direct calculation of the integral in (3.21) is not feasible. To efficiently evaluate this integral, an MC integration is used in the SML approach. In contrast to other numerical integrations, MC simulation is independent of the dimension of the integral. Thus, high-dimensional integration, which is usually the case, can be handled effectively. After estimating the likelihood by MC integration, SML seeks parameter estimates that maximize the estimated likelihood function. To reduce the MC sampling variance in estimating the likelihood, an EIS technique (Liesenfeld and Richard (2003)) is employed. Details of the implementation of the EIS method for the AR(p)–SV model are given below.

3.3.1 Parameter Estimation

In the AR(p)–SV model, consider the following factorization for the likelihood in (3.21):

$$L(\xi; Y_T) = \int f(Y_T|\Lambda_T, \xi) f(\Lambda_T|\xi) d\Lambda_T, \tag{3.29}$$

where

$$f(\Lambda_T|\xi) = \prod_{t=1}^{T} f(\lambda_t|\lambda_{t-1}, \xi), \tag{3.30}$$

$$f(Y_T|\Lambda_T, \xi) = f(Y_p|\Lambda_p, \xi) f(y_{p+1}, ..., y_T|Y_p, \lambda_{p+1}, ..., \lambda_T, \xi), \tag{3.31}$$

where $Y_p = (y_1, ..., y_p)$ and $\Lambda_p = (\lambda_1, ..., \lambda_p)$. In a crude MC simulation, the likelihood function in (3.21) is estimated by

$$\hat{L}(\xi; Y_T) = \frac{1}{N} \sum_{i=1}^{N} f(Y_T | \Lambda_T^{(i)}, \xi), \tag{3.32}$$

where N sets of $\Lambda_T^{(i)}$ are generated from the $f(\Lambda_T|\xi)$ in (3.30). Similarly to the difficulty of MC simulation in the standard SV model, the factorization in (3.29) causes a large MC sampling variance because the simulated Λ_T does not resemble that where the observations Y_T are obtained. As discussed in Chapter 2, Liesenfeld and Richard (2003) proposed an EIS method that sought an IF that minimized the MC sampling variance effectively to resolve the inefficiency problem. We will pursue only the EIS technique for the AR(p)–SV model.

To implement the EIS technique, let us consider the following factorization:

$$\begin{aligned} L(\xi; Y_T) &= \int \frac{f(Y_T, \Lambda_T|\xi)}{g(\Lambda_T|Y_T, \xi)} g(\Lambda_T|Y_T, \xi) d\Lambda_T \\ &= \int h(Y_T, \Lambda_T|\xi) g(\Lambda_T|Y_T, \xi) d\Lambda_T, \end{aligned} \tag{3.33}$$

where g and h denote an importance function (IF) and a remainder function (RF), respectively.

The EIS technique chooses in the following way an IF $g(\Lambda_T|Y_T, \xi)$ that can provide a good approximation of $f(Y_T, \Lambda_T|\xi)$ and that depends on Y_T. Consider a particular factorization:

$$g(\Lambda_T|Y_T, \xi) = \prod_{t=1}^{T} f(\lambda_t|\Lambda_{t-1}, Y_{t-1}, \xi, e_t), \tag{3.34}$$

where e_t is an auxiliary vector depending on Y_T, as in the SV model. Then we choose a function $k(\cdot)$ depending on e_t such that

$$g(\Lambda_T|Y_T, \xi) = \prod_{t=1}^{T} C(\lambda_{t-1}, e_t) k(\Lambda_t; e_t), \tag{3.35}$$

where $C(\cdot)$ is an integral constant that makes $k(\Lambda_t; e_t)$ a density function. For ease of notation, we remove ξ from the functional notations of C and k.

The choice of $C(\lambda_{t-1}, e_t)$ plays an important role in matching $g(\Lambda_T|Y_T, \xi)$ to the joint density $f(Y_T, \Lambda_T|\xi) = \prod_{t=1}^{T} f(y_t, \lambda_t|\Lambda_{t-1}, Y_{t-1}, \xi)$ as closely as possible.

For the AR(p)–SV model, a natural choice of $k(\Lambda_t; e_t)$ would be proportional to a Gaussian density for λ_t given λ_{t-1}. Thus, we can express the function k as a product of two functions:

$$k(\Lambda_t; e_t) = l(\lambda_t, e_t)\ f(\lambda_t|\lambda_{t-1}, \xi). \tag{3.36}$$

Since $f(\lambda_t|\lambda_{t-1}, \xi)$ is a Gaussian density, an auxiliary vector e_t can be estimated via regression coefficients from a linear regression if $l(\lambda_t, e_t)$ is chosen as a Gaussian density kernel $\exp\{e_{1,t}\lambda_t + e_{2,t}\lambda_t^2\}$. That is, $\log l(\lambda_t, e_t)$ approximates $\log f_p(y_t|\lambda_t, \xi) - \log C(\lambda_t, e_{t+1})$, where the conditional density $f_p(y_t|\lambda_t, \xi)$ has the following expression:

$$f_p(y_t|\lambda_t) = \begin{cases} f(y_1, ..., y_t|\lambda_t, \xi), & \text{if } t \leq p, \\ f(y_t|Y_p, \lambda_t, \xi), & \text{otherwise.} \end{cases} \tag{3.37}$$

Thus, the regression equation is given by

$$\log f_p(y_t|\lambda_t, \xi) - \log C(\lambda_t, e_{t+1}) = constant + e_{1,t}\lambda_t + e_{2,t}\lambda_t^2 + a_t \tag{3.38}$$

for $t = 1, ..., T$, where a_t is a regression error term. Assuming $C(\Lambda_T; e_{T+1}) = 1$, this regression can be done backward. The details of the functional forms are as follows.

The conditional density of λ_t given λ_{t-1} is given by

$$f(\lambda_t|\lambda_{t-1}, \xi) = \frac{1}{\sqrt{2\pi\gamma^2}} \exp\left[-\frac{(\lambda_t - \alpha - \beta\lambda_{t-1})^2}{2\gamma^2}\right]. \tag{3.39}$$

Substituting this density into (3.36), the function $f(\lambda_t|\Lambda_{t-1}, Y_{t-1}, \xi, e_t)$ in (3.34) is given by

$$f(\lambda_t|\Lambda_{t-1}, Y_{t-1}, \xi, e_t) = \frac{1}{\sqrt{2\pi V_t}} \exp\left[-\frac{(\lambda_t - M_t)^2}{2V_t}\right],$$

where

$$M_t = \left(\frac{\alpha + \beta\lambda_{t-1}}{\gamma^2} + e_{1,t}\right) V_t,\ V_t = \frac{\gamma^2}{1 - 2e_{2,t}\gamma^2},$$

and $\log C(\lambda_t, e_{t+1})$ is chosen as

$$\log C(\lambda_t, e_{t+1}) = -\frac{M_t^2}{2V_t} + \frac{(\alpha + \beta\lambda_{t-1})^2}{2\gamma^2} - \frac{1}{2}\log\frac{V_t}{\gamma^2}.$$

Since $C(\lambda_t, e_{t+1})$ depends on $e_{1,t+1}$ and $e_{2,t+1}$, the regression coefficients are computed recursively, from $t = T, T-1, ..., 1$, assuming $C(\lambda_T, e_{T+1}) = 1$.

Hence, the IF g for the AR(p)–SV model is

$$g(\Lambda_T|Y_T, \xi) = \prod_{t=1}^{T} C(\lambda_{t-1}, \hat{e}_t)\exp\{\hat{e}_{1,t}\lambda_t + \hat{e}_{2,t}\lambda_t^2\} f(\lambda_t|\lambda_{t-1}, \xi). \tag{3.40}$$

Using this IF, the corresponding likelihood is estimated by

$$\hat{L}(\xi; Y) = \frac{1}{N}\sum_{i=1}^{N}\prod_{t=1}^{T}\frac{f_p(y_t|\lambda_t^{(i)})}{C(\lambda_{t-1}, \hat{e}_t)\exp\{\hat{e}_{1,t}\lambda_t^{(i)} + \hat{e}_{2,t}\lambda_t^{(i)2}\}}, \tag{3.41}$$

where

$$f_p(y_t|\lambda_t^{(i)}) = \begin{cases} f(y_1, ..., y_t|\lambda_t^{(i)}, \xi), & \text{if } t \le p, \\ f(y_t|Y_p, \lambda_t^{(i)}, \xi), & \text{otherwise,} \end{cases}$$

where $\lambda_t^{(i)}$ is drawn from $g(\Lambda_T^{(i)}|Y_T, \xi)$ in (3.40). Assuming the standard normal distribution for the error $\{u_t\}$, the functional forms of $f(y_1, ..., y_t|\lambda_t^{(i)}, \xi)$ and $f(y_t|Y_p, \lambda_t^{(i)}, \xi)$ are given in (3.24) and (3.26), respectively. Under the Student-t distributional assumption for u_t, $f(y_1, ..., y_t|\lambda_t^{(i)}, \xi)$ and $f(y_t|Y_p, \lambda_t^{(i)}, \xi)$ are given in (3.27) and (3.28), respectively.

In summary, the EIS algorithm to evaluate the likelihood for the AR(p)–SV model in (3.21), given an initial parameter vector ξ, is as follows:

1. Generate $\Lambda_T^{(i)}$ according to (3.3).
2. Assuming $C(\lambda_T, e_{T+1}) = 1$, obtain e_t by performing T regressions as in (3.38), working backward from $t = T$ to $t = 1$.
3. Generate the new $\Lambda_T^{(i)}$ from the importance density $g(\Lambda_T|Y_T, \xi)$ in (3.40).
4. Repeat Steps 2 and 3 until either e_t or the new λ_t converges for all t.
5. Compute the estimated likelihood in (3.41).

3.3.2 Volatility Estimation

For a given estimated parameter $\hat{\xi} = (\hat{\Phi}, \hat{\alpha}, \hat{\beta}, \hat{\gamma})$ obtained by SML, a particle filter (Pitt and Shephard (1997) and Kim *et al.* (1998)) can be employed to estimate the volatility, $\sigma_t = \exp\{\lambda_t/2\}$. The idea is to draw, for a given $\hat{\xi}$, M new samples λ_t^{new} from $\lambda_t|Y_t, \hat{\xi}$, given $\lambda_{t-1}^{new,1}, ..., \lambda_{t-1}^{new,M}$ from $\lambda_{t-1}|Y_{t-1}, \hat{\xi}$, where $Y_t = (y_1, ..., y_t)$.

According to Bayes' theorem, the conditional density $f(\lambda_t|Y_t, \hat{\xi})$ is proportional to $f_p(y_t|\lambda_t, \hat{\xi})f(\lambda_t|Y_{t-1}, \hat{\xi})$:

$$f(\lambda_t|Y_t, \hat{\xi}) \propto f_p(y_t|\lambda_t, \hat{\xi})f(\lambda_t|Y_{t-1}, \hat{\xi}),$$

where $f_p(y_t|\lambda_t, \hat{\xi})$ is given in (3.37) and

$$f(\lambda_t|Y_{t-1}, \hat{\xi}) = \int f(\lambda_t|\lambda_{t-1}, \hat{\xi})f(\lambda_{t-1}|Y_{t-1}, \hat{\xi})d\lambda_{t-1}. \tag{3.42}$$

Then, the integral in (3.42) can be estimated by

$$f(\lambda_t|Y_{t-1}, \hat{\xi}) \approx \frac{1}{M}\sum_{j=1}^{M} f(\lambda_t|\lambda_{t-1}^{(j)}, \hat{\xi}),$$

where $\lambda_{t-1}^{(j)}$ is drawn from $f(\lambda_{t-1}|Y_{t-1}, \hat{\xi})$ for $j = 1, ..., M$. This leads to

$$f(\lambda_t|Y_t, \hat{\xi}) \propto f_p(y_t|\lambda_t, \hat{\xi})\frac{1}{M}\sum_{j=1}^{M} f(\lambda_t|\lambda_{t-1}^{(j)}, \hat{\xi}). \tag{3.43}$$

To sample λ_t, we adopt the following procedure. Let $\lambda_{t|t-1} = \hat{\alpha} + \hat{\beta}(M^{-1}\sum \lambda_{t-1}^{(j)})$. Note that the logarithmic conditional density of y_t given λ_t can be written as

$$\log f_p(y_t|\lambda_t, \hat{\xi}) = constant - \frac{1}{2}\lambda_t - \frac{y_t^2}{2}\exp\{-\lambda_t\}.$$

Setting $\log f^*(y_t, \lambda_t, \hat{\xi}) = -\frac{1}{2}\lambda_t - \frac{y_t^2}{2}\exp\{-\lambda_t\}$, the expansion of $\log f^*(y_t, \lambda_t, \hat{\xi})$ by a first-order Taylor approximation around $\lambda_{t|t-1}$ leads to:

$$\begin{aligned}\log f^*(y_t, \lambda_t, \hat{\xi}) &\leq -\frac{1}{2}\lambda_t - \frac{y_t^2}{2}[(1 + \lambda_{t|t-1})\exp\{-\lambda_{t|t-1}\} - \lambda_t \exp\{-\lambda_{t|t-1}\}] \\ &= \log g^*(\lambda_t, \lambda_{t|t-1}, \hat{\xi}),\end{aligned} \tag{3.44}$$

where $\log g^*$ is the right-hand side of the inequality. Also, it can be shown via some algebra that

$$g^*(\lambda_t, \lambda_{t|t-1}, \hat{\xi}) f(\lambda_t|\lambda_{t-1}^{(j)}, \hat{\xi}) \propto \pi_j f_N(\lambda_t|\lambda_{t|t-1}^{(j)}, \hat{\gamma}^2), \tag{3.45}$$

where $f_N(\lambda_t|\lambda_{t|t-1}^{(j)}, \hat{\gamma}^2)$ denotes a normal density of λ_t with mean $\lambda_{t|t-1}^{(j)}$, and variance $\hat{\gamma}^2$,

$$\pi_j = \exp\left[\frac{1}{2\hat{\gamma}^2}\{(\hat{\alpha} + \hat{\beta}\lambda_{t|t-1}^{(j)2})^2 - \lambda_{t|t-1}^{(j)2}\}\right],$$

and

$$\lambda_{t|t-1}^{(j)} = \hat{\alpha} + \hat{\beta}\lambda_{t-1}^{(j)} + \frac{\hat{\gamma}^2}{2}[y_t^2 \exp\{-\lambda_{t|t-1}\} - 1].$$

Hence, because of the relationship in (3.45), the right-hand side of (3.43) is bounded as

$$\begin{aligned} f_p(y_t|\lambda_t, \hat{\xi})\frac{1}{M}\sum_{j=1}^{M} f(\lambda_t|\lambda_{t-1}^{(j)}, \hat{\xi}) &\leq g^*(\lambda_t, \lambda_{t|t-1}, \hat{\xi})\frac{1}{M}\sum_{j=1}^{M} f(\lambda_t|\lambda_{t-1}^{(j)}, \hat{\xi}) \\ &\propto \frac{1}{M}\sum_{j=1}^{M} \pi_j f_N(\lambda_t|\lambda_{t|t-1}^{(j)}, \hat{\gamma}^2). \end{aligned}$$

In summary, the algorithm for drawing λ_t is as follows: First, draw a proposal λ_t^{prop} for the mixture density $\sum_{j=1}^{M} \pi_j^* f_N(\lambda_t|\lambda_{t|t-1}^{(j)}, \hat{\gamma}^2)$, where $\pi_j^* = \pi_j / \sum \pi_j$, and then accept λ_t^{prop} with probability $f^*(y_t, \lambda_t, \hat{\xi})/g^*(\lambda_t, \lambda_{t|t-1}, \hat{\xi})$. If rejected, go to the first step. From the estimated log-variance $\hat{\lambda}_t$, we obtain the estimated volatility $\hat{\sigma}_t = \exp\{\hat{\lambda}_t/2\}$.

Under the scaled Student-t distributional assumption for u_t, the conditional density of y_t given λ_t is

$$f(y_t|\lambda_t) = \left(\sqrt{\pi(\omega-2)\exp\{\lambda_t\}}\right)^{-1} \frac{\Gamma((\omega+1)/2)}{\Gamma(\omega/2)} \left(1 + \frac{y_t^2}{\exp\{\lambda_t\}(\omega-2)}\right)^{-(\omega+1)/2},$$

and so its corresponding acceptance/rejection probability f^*/g^* can be described by the following two equations:

$$\begin{aligned} \log f^*(y_t, \lambda_t, \hat{\xi}) &= \frac{\omega}{2}\lambda_t - \frac{(\omega+1)}{2}\log[\exp\{\lambda_t\}(\omega-2) + y_t^2], \\ \log g^*(\lambda_t, \lambda_{t|t-1}, \hat{\xi}) &= \frac{\omega}{2}\lambda_t - \frac{(\omega+1)}{2}\log[(\omega-2)(1+\lambda_{t|t-1})\exp\{\lambda_{t|t-1}\} \\ &+ \lambda_t \exp\{\lambda_{t|t-1}\}) + y_t^2]. \end{aligned}$$

Then, the weight function π_j is the same as $g^*(\lambda_t^{(j)}, \lambda_{t|t-1}^{(j)}, \hat{\xi})$ in (3.44) and the normal mixture density is given in (3.45). The estimation of volatility in the case of the scaled Student-t distribution can be implemented by modifying the algorithm for the normal case.

3.4 Order Selection and Model Diagnostics

To choose the "best" model for the data, we first determine the order of the AR(p)–SV process with some statistical model-selection criteria. Karanasos and Kim (2003) used the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) to select the order of the ARMA process in the ARMA(p, q)–EGARCH model. Kitamura (2011) and Ari and Ünal (2011) used the AIC to determine the order of the ARMA–GARCH model. In the same spirit, we also use these criteria to determine the order of the AR process in the AR(p)–SV model although the properties of AIC and BIC are currently unknown for this model.

We start with a higher order model and reduce the order of the model to fit the data by comparing AIC and BIC; this is referred to as the general-to-specific approach. AIC is defined by $2k - 2 * \text{logLH}$, where k is the number of parameters and logLH denotes the log likelihood. BIC is defined by $-2*\text{logLH}+k*\log(T)$, where T is the number of observations.

After the model parameters are estimated, we perform a residual analysis to check the adequacy of the fitted model. For the residual analysis, we define the standardized residual at time t from (3.1) and (3.2) as

$$\hat{u}_t \equiv \frac{\hat{\Phi}(B)y_t}{\hat{\sigma}_t}, \qquad t = 1, ..., T, \tag{3.46}$$

where y_t is the observed return, $\hat{\Phi}(B)$ is the polynomial with the estimated AR coefficients $\hat{\phi}_i$ for $i = 1, ..., p$, and $\hat{\sigma}_t = \exp\{\hat{\lambda}_t/2\}$, which is the estimated volatility obtained by the particle filter described in Section 3.3.2. If the fitted model is adequate, $\hat{u}_t$ follows a standard normal distribution under the standard normal distributional assumption and a Student-t distribution with zero mean and unit variance under the scaled Student-t distributional assumption.

Table 3.1: Parameter estimates: AR(1)–SV model

	$\hat{\phi}_1$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	κ	$\hat{\kappa}$
True Value	0.5	0.01	0.97	0.2		
Series 500	0.5189	-0.0000	0.9668	0.2069	4.9623	4.2571
Series 1000	0.4982	0.0011	0.9783	0.1982	5.7834	5.4087
Series 2000	0.5107	0.0073	0.9737	0.2124	5.4384	5.1112

3.5 Simulation Study and Sensitivity Analysis

The purpose of the simulation study is to investigate the performance of the SML with the EIS algorithm in the parameter estimation of the AR(p)–SV model, and that of a particle filter for volatility estimation. In particular, we consider the AR(1)–SV model with the parameter vector $(\phi_1, \alpha, \beta, \gamma) = (0.5, 0.01, 0.97, 0.2)$ as a true model. Three different series are generated according to the true model. We will refer to these as Series 500, Series 1000, and Series 2000, with 500, 1000, and 2000 observations respectively. The hypothetical values of the parameters are similar to those found empirically. For parameter estimation, we choose the SML approach and implement the EIS algorithm to estimate the likelihood in (3.21) when $p = 1$.

Figure 3.1 displays the autocorrelation functions of the simulated series and its squares. We observe that the series have large first-lag autocorrelations followed by decaying autocorrelation. In addition, the ACFs of the squared series show persistence. The parameter estimation results are summarized in Table 3.1. All the estimates are close to the true values, which indicates that the SML approach with the EIS algorithm can be used for parameter estimations in the AR(1)–SV model. Although we show the results only for an AR–SV model with order 1, the estimation procedure for a higher-order model is the same as that for AR(1)–SV except that the likelihood functions are more involved and there are more parameters to be estimated.

The empirical kurtosis, referred to as κ in Table 3.1, is directly calculated from the series, while the estimated kurtosis, referred to as $\hat{\kappa}$, is computed with the parameter estimates. The values of the empirical and estimated kurtosis are close to each other for all the simulated series. This indicates that our parameter estimates are fairly accurate.

The estimates of the volatility are obtained by a particle filter. Figure 3.2 shows the absolute values of the returns along with the estimated volatility for each series. We can see that the estimated volatilities are able to capture the movement of the simulated series. In particular, a large movement is captured more readily than a small movement.

For the model diagnostics, the standardized residual of the AR(1)–SV model is defined by

$$\hat{u}_t = \frac{y_t - \hat{\phi}_1 y_{t-1}}{\hat{\sigma}_t},$$

where $\hat{\sigma}_t = \exp\{\hat{\lambda}_t/2\}$ and $\hat{\lambda}_t$ is the estimated log variance obtained by a particle filter for $t = 1, ..., T$. If the AR(1)–SV model fits the data well, this standardized residual $\hat{u}_t$ follows a normal distribution with mean 0 and variance 1. Figure 3.3 displays the diagnostic plots: the residual plot, ACF plot, and QQ plot for each series. The residual plots imply constant variance, and no autocorrelations are observed in any of the ACF plots. Moreover, the QQ plots show that all the points are near or on the straight line, which indicates no violation of the normal assumption for the error process u_t. Thus, it can be concluded that the AR(1)–SV model fits all the simulated series well, as expected. In addition, our simulation study shows that the SML approach with the EIS algorithm is applicable to the estimation of the AR(1)–SV model.

In order to understand the robustness of the SML method with the EIS, we generate two sets of 1000 observations from AR(1)–SV model, assuming that u_t follows a Generalized Lambda Distribution (GLD) (Ramberg, Dudewicz, and Tadikamalla (1979)): one with the parameter vector of $(\lambda_1, \lambda_2, \lambda_3, \lambda_4) = (0, -1, -1/6, -1/6)$ and the other with $(\lambda_1, \lambda_2, \lambda_3, \lambda_4) = (0, -1, -1/2, -1/2)$. We refer to the first series as "Series A" and the second series as "Series B". The sample kurtosis of the series are 4.65 and 10.88, respectively. The true parameter values are the same as the ones used in our simulation study. Figure 3.4 shows time series plots of two series, ACF of the series, and ACF of the squares. For both series, ACF plot of the series displays decaying autocorrelations indicating an AR(1) process. Based on the large kurtosis and the decaying autocorrelation, we entertain an AR(1)–SV model for those series.

To fit AR(1)–SV model to Series A and B, we consider two error assumptions for u_t: one is a standard normal and the other is a scaled Student-t with degree of freedom 8. Table 3.2 summarizes the parameter estimation results along with the true parameter values. Regardless of the error assumptions, parameter estimates are quite close to the true values. Both MC standard errors and standard errors of the estimators are very small.

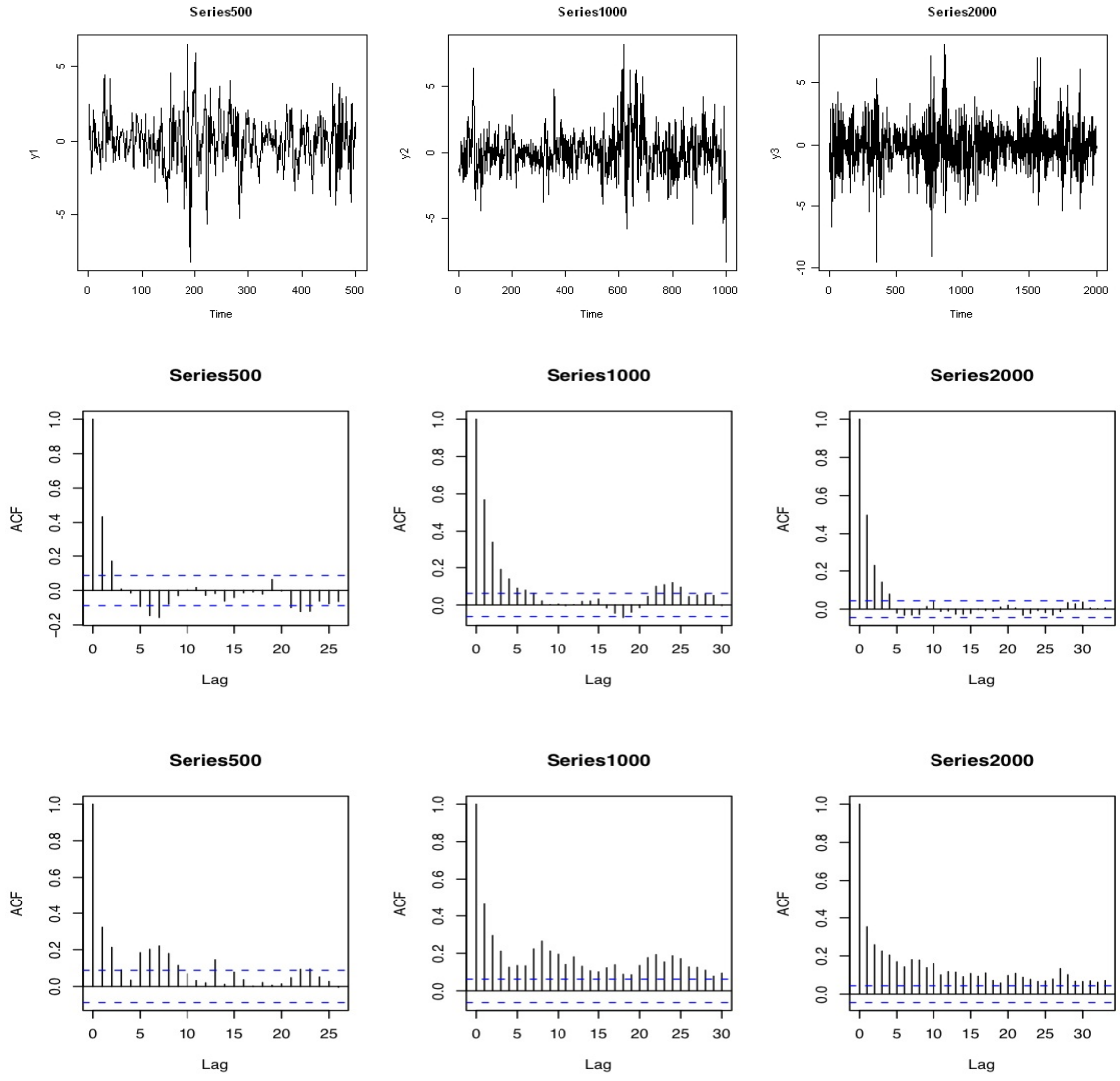

Figure 3.1: The simulated series (top), autocorrelation functions of returns (middle), and squared returns (bottom)

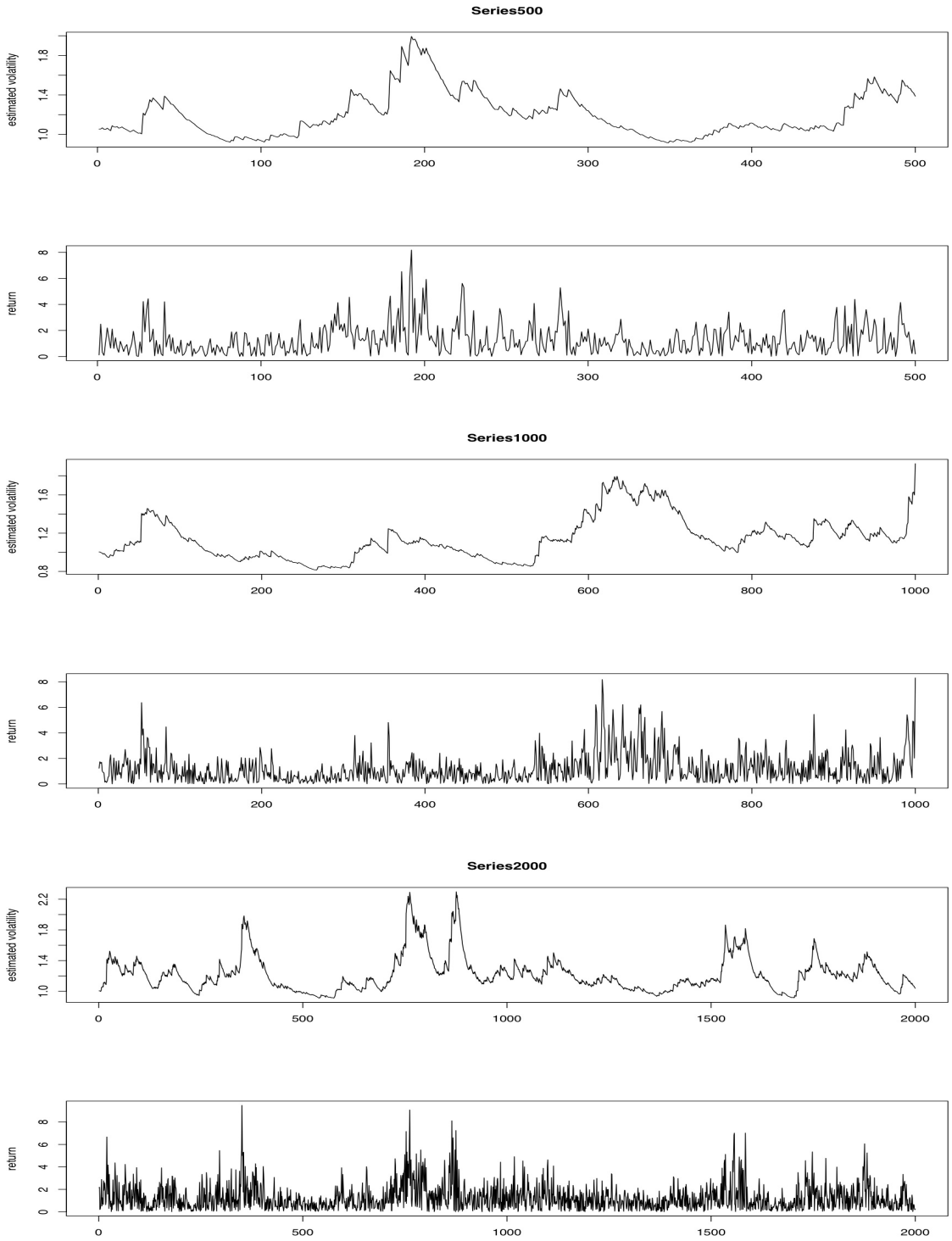

Figure 3.2: Top panels: estimated volatilities; Bottom panels: absolute values of return series.

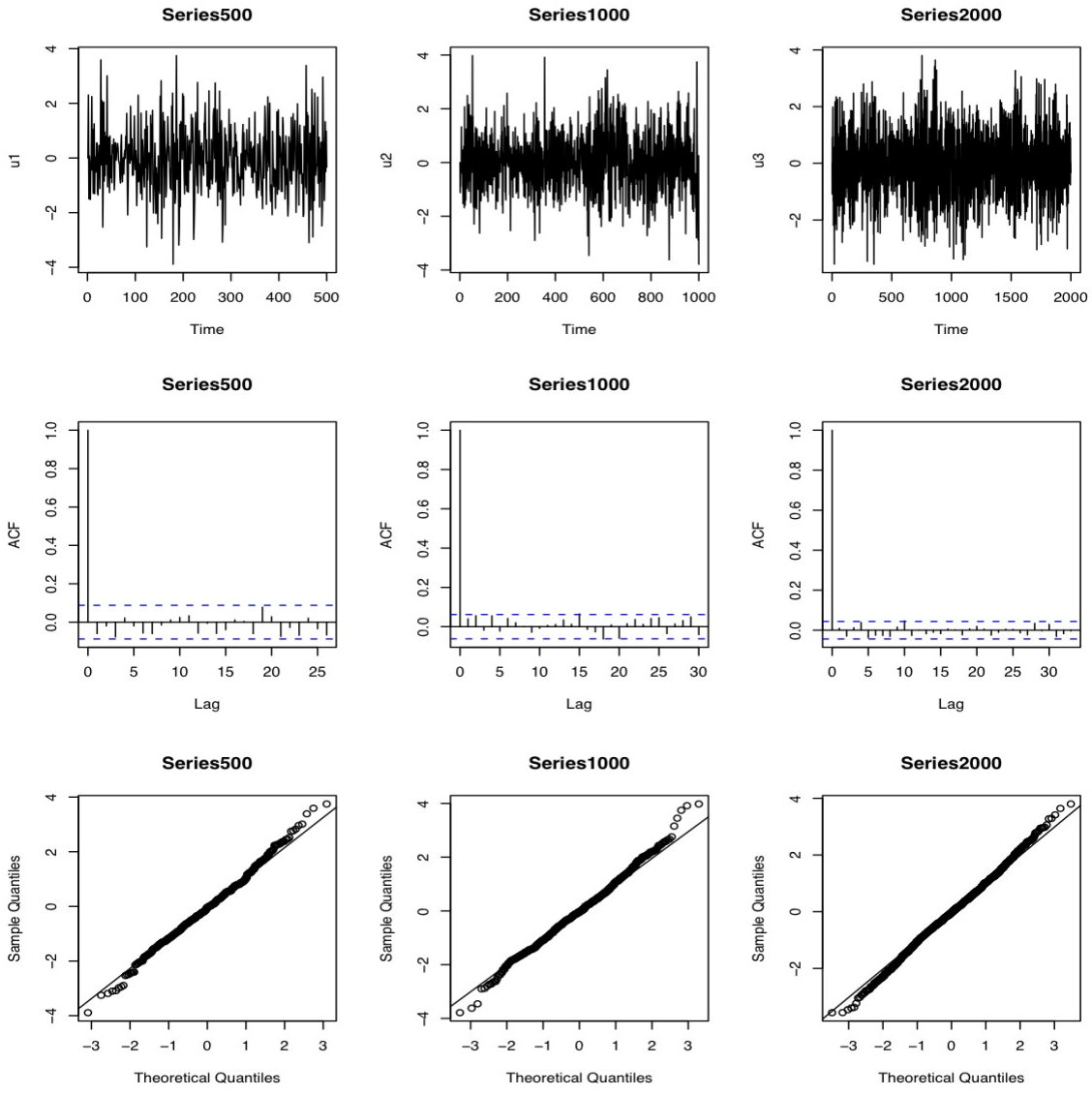

Figure 3.3: The residual plots are in the first row, the ACFs of $\hat{u}_t$ are in the second row, and the QQ plots are in the last row.

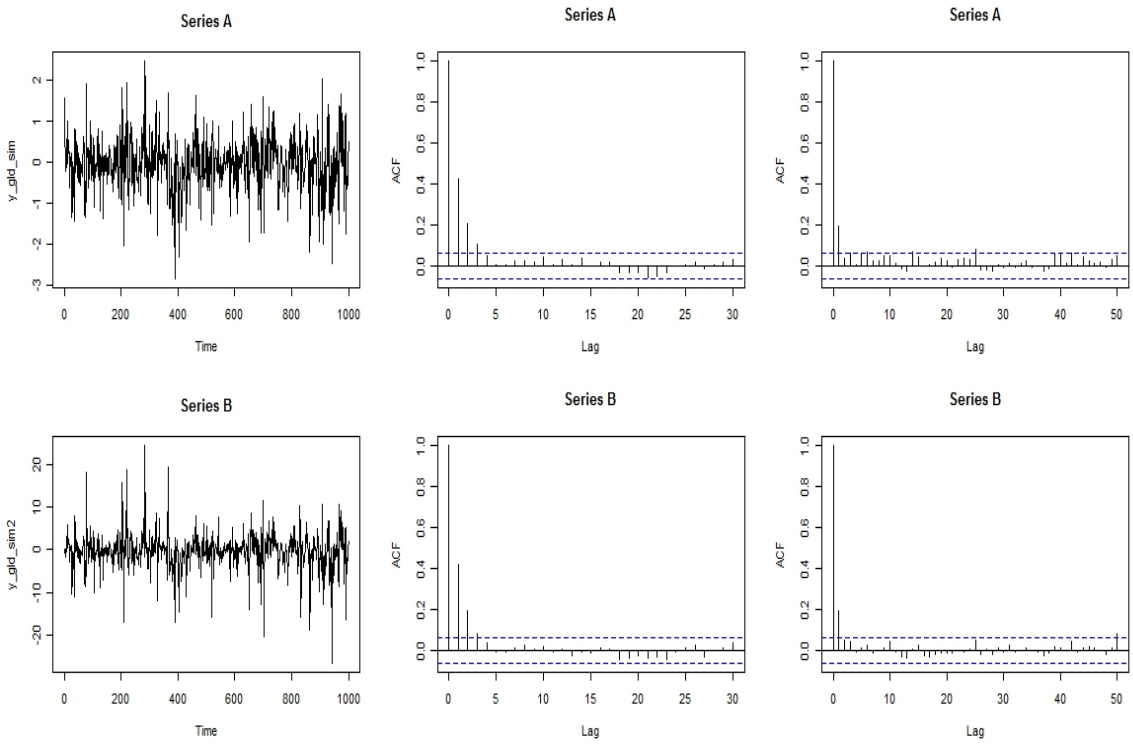

Figure 3.4: Data plots of Series A and B

Table 3.2: Parameter estimates: AR(1)–SV model with N(0,1) and t(0,1) errors

	$\hat{\phi}_1$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$
True Value	0.5	0.01	0.97	0.2
Series A with N(0,1)	0.4717	0.0125	0.9599	0.2207
MC std. error	0.0002	0.0002	0.0001	0.0001
Std. error	0.0108	0.0109	0.0237	0.0197
Series A with t(0,1)	0.4918	0.0140	0.9601	0.1611
MC std. error	0.0001	0.0002	0.0002	0.0001
Std. error	0.0128	0.0101	0.0198	0.0201
Series B with N(0,1)	0.4813	0.0144	0.9719	0.2019
MC std. error	0.0003	0.0004	0.0001	0.0002
Std. error	0.0110	0.0182	0.0215	0.0201
Series B with t(0,1)	0.5191	0.0191	0.9742	0.2004
MC std. error	0.0001	0.0004	0.0003	0.0001
Std. error	0.0181	0.0136	0.0102	0.0186

Figure 3.5 shows the estimated volatility $\sigma_t = \exp\{\lambda_t/2\}$, obtained by a particle filter and the absolute returns for Series A. The dotted line is the estimated volatility under a normal assumption and the solid line is the one under a scaled Student-t assumption. The variation in the estimated volatility under the heavy-tail error assumptions is much smaller than that under a normal error assumption. This can also be seen from the parameter estimation results in Table 3.2. The volatility of the log-variance γ under a scaled Student-t error assumption is less than that under a normal error assumption.

We perform a residual analysis to test the goodness of fit of AR(1)–SV model with two error assumptions to Series A. The standardized residual of the AR(1)–SV model at time t is obtained as in the simulation study. Diagnostic plots are given in Figure 3.6. Graphs on the left are under a normal error assumption and those on the right are under a scaled Student-t error assumption with degree of freedom 8. No serial correlation is observed under a heavy tail error assumption while there is a small autocorrelation at lag 1 under a normal assumption. In the normal QQ plot, there are some deviations at both corners, but deviations at the corners of the Student-t QQ plot are smaller. Based on our observations on the diagnostics plot, we conclude that the AR(1)–SV model with Student-t error fits

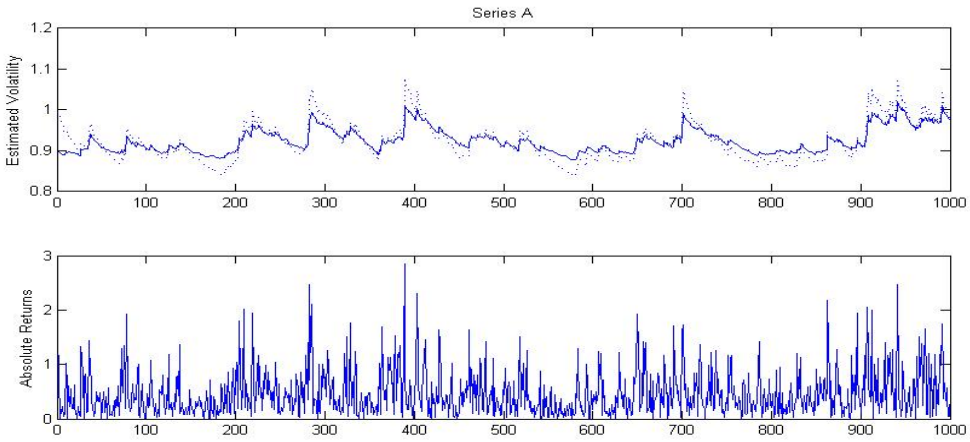

Figure 3.5: Top panels: estimated volatilities; Bottom panels: absolute values of return series

the series reasonably well.

In addition, we note that the estimates of the parameters using the SML with EIS for the AR(1)–SV model with normal and Student-t errors are close to the true values although the true error distribution (GLD) has somewhat different tail behavior. We thus conclude that the SML estimation with EIS technique is robust to certain deviations from normal or t tail behavior.

3.6 Empirical Study

In this section, we carry out an empirical analysis of the AR(p)–SV model using four actual financial time series. The SML approach with the EIS technique is used for the parameter estimation, and a particle filter is used to estimate volatility. Moreover, model selection and model diagnostics are performed. All the code is written in MATLAB 7.3.0.

3.6.1 Data

Four time series are considered in our empirical analysis: the daily closing prices of the Kospi 200 index from 1/4/1989 to 12/29/2000, the daily closing prices of RBC stock from

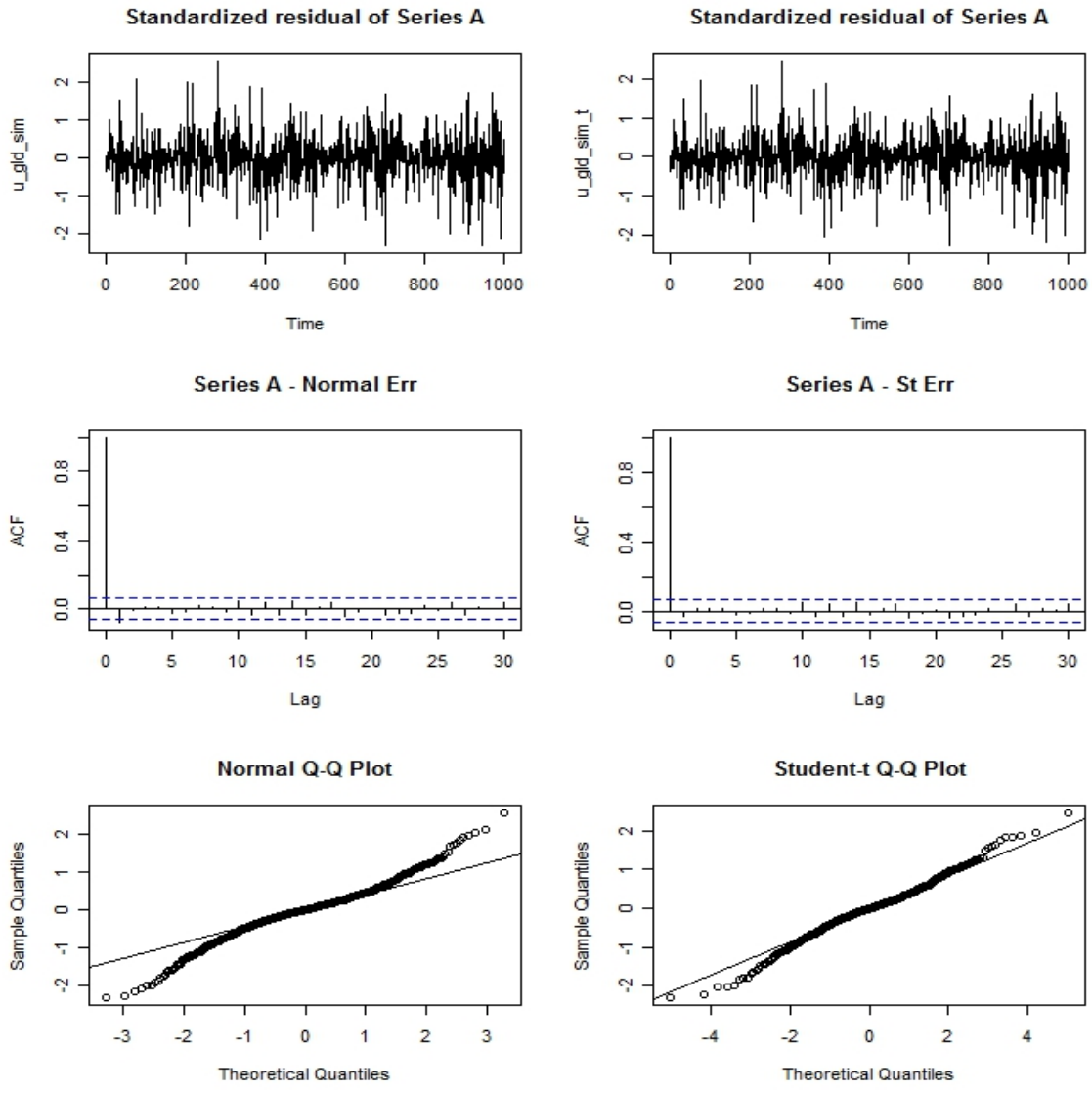

Figure 3.6: Series A: The residual plots are in the first row, the ACFs of $\hat{u}_t$ are in the second row, and the QQ plots are in the last row.

Table 3.3: Summary statistics of data

Statistics	KOSPI 200	RBC	NIKKEI 225	$/YEN
Sample size	3124	3542	2000	2516
Mean	-0.0340	0.0473	0.0258	0.0004
Std. dev.	2.3084	1.3445	1.4731	0.3212
Kurtosis	15.5597	5.7700	5.4816	7.2309
$\rho(1)$	0.156	0.106	-0.03	0.008
$LB^1(30)$	295.6143	81.4052	39.7589	31.3883
	(0.0000)	(0.0000)	(0.1096)	(0.3965)
$LB^2(30)$	4103.5670	916.8041	539.3733	726.2045
	(0.0000)	(0.0000)	(0.0000)	(0.0000)

5/3/1980 to 4/20/1998, the daily closing prices of the Nikkei 225 index from 2/3/1990 to 3/30/2000, and the daily closing exchange rates of US dollar to Japanese Yen ($/YEN) from 10/4/1980 to 12/29/2000. The prices of the series are converted into the mean corrected return, defined by $y_t = 100 * [\log(p_t/p_{t-1}) - \frac{1}{T}\sum \log(p_t/p_{t-1})]$, where p_t is the price of a stock or an exchange rate at time t. The total number of observations and summary statistics for each return series are given in Table 3.3, and the data plots are given in Fig. 3.7.

The kurtosis for each series is greater than three, which implies that the distributions of the series have heavier tails than those of the normal distribution. At the bottom of the table, $LB^1(30)$ denotes the Ljung-Box (LB) statistics of the series for the first thirty lags and $LB^2(30)$ denotes the Ljung-Box statistics of the squared series for the first thirty lags. The values in parentheses are the corresponding p-values of the LB statistics. These indicate that there are non-negligible autocorrelations in the series except for the Nikkei 225 and $/YEN series. This can be seen in Fig. 3.8, where the autocorrelation functions of the series are displayed. In particular, the autocorrelations at lag 1 for the Nikkei 225 and $/YEN series are not as significant as those of the RBC and Kospi 200 series. The values of the first-order autocorrelation, denoted $\rho(1)$, are given in Table 3.3. $\rho(1)$ of Kospi 200 and RBC are greater than 0.1, but $\rho(1)$ of Nikkei 225 and $/YEN are close to zero. Figure 3.9 shows that the serial correlations in the squared series are large.

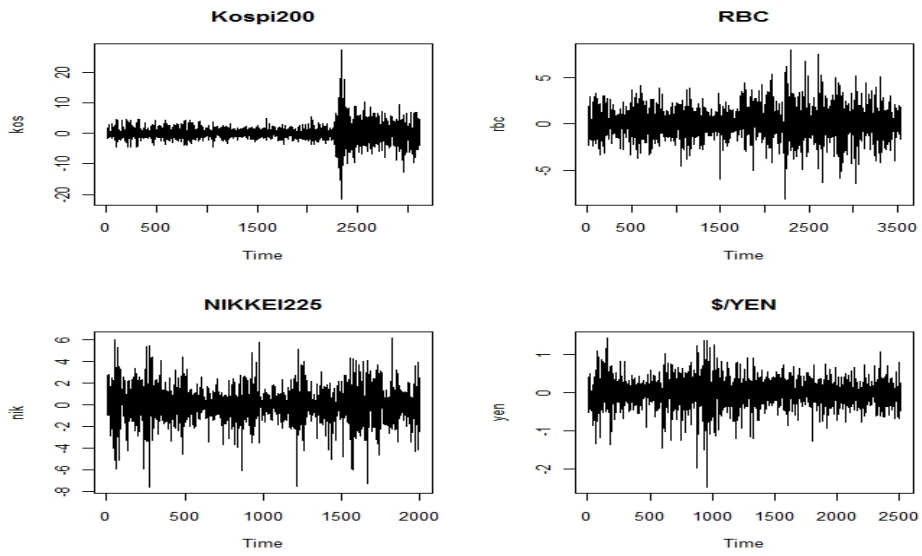

Figure 3.7: Plots of four series: KOSPI 200, RBC, NIKKEI 225, and $/YEN

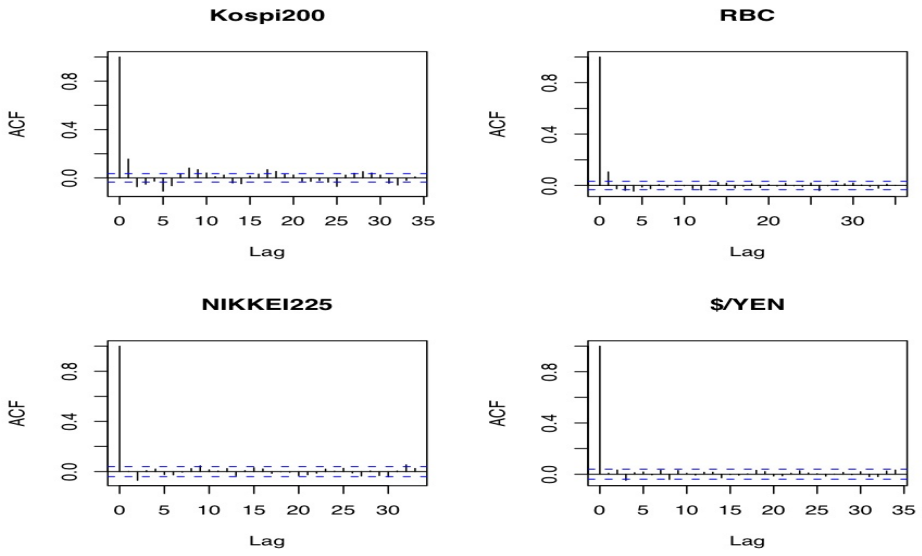

Figure 3.8: Autocorrelation functions of returns

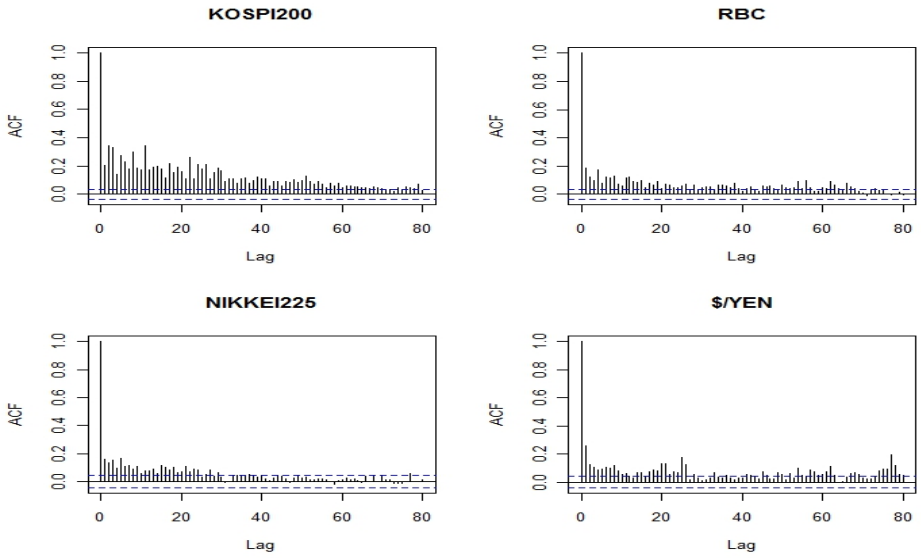

Figure 3.9: Autocorrelation functions of squared returns

3.6.2 Estimation Results

The AIC and BIC values of the data are summarized in Table 3.4. The log-likelihood values in the AIC and BIC calculations are the values of the log likelihood obtained in the parameter estimation by SML. Based on the AIC and BIC values, AR(1)–SV models are the best choice for the Kospi 200 and RBC series because the AIC and BIC values are the smallest. For the Nikkei 225 and \$/YEN series, AR(0)–SV models are the best. This is consistent with our observation from the ACF plots in Fig. 3.8. We see almost no correlations in the Nikkei 225 and \$/YEN series, while there are relatively large correlations at lag 1 in the Kospi 200 and RBC series.

The parameter estimation results for the AR(1)–SV model for all the series are given in Table 3.5. The MC sampling errors and standard errors of the estimates are small. In the AR(1)–SV model, the coefficient ϕ_1 is the first order autocorrelation. The estimates of ϕ_1 for all the series are close to the values of the first-order autocorrelation in Table 3.3. In Fig. 3.8, $\hat{\phi_1}$ is nearly zero for the Nikkei 225 and \$/YEN series. This indicates that the SV model is more appropriate than the AR(1)–SV model for these series. The estimated values of the persistent parameter β are greater than 0.95. The estimated kurtosis of the series is quite close to the sample value of the kurtosis in Table 3.3. This indicates that the AR(1)–SV model is able to capture the heavy tails of the return distribution. Further

Table 3.4: AIC and BIC for model selection

Data	Criteria	AR(0)–SV	AR(1)–SV	AR(2)–SV	AR(3)–SV
KOSPI 200	AIC	5895	5867.2	6298	
	BIC	5901	5891.4	6328	
RBC	AIC	13194	8387	10154	
	BIC	13219	8411.7	10185	
NIKKEI 225	AIC	7910	8169.4	9425	
	BIC	7933	8191.8	9453	
$/YEN	AIC	4942	5021.4	5022	5717.8
	BIC	4965	5044.7	5051	5752.8

model validation via the residual analysis is given in Section 3.6.3.

Given the parameter estimation results in Table 3.5, we estimate the volatility of $\sigma_t = \exp\{\lambda_t/2\}$ by a particle filter. We use 100 simulated λ_t at each time step for $t = 1, ..., T$. Figure 3.10 shows the estimated volatility $\hat{\sigma}_t$ and the corresponding absolute values of the returns. We observe that the AR(1)–SV model captures a large movement more readily than a small movement.

3.6.3 Model Diagnostics

To validate the fitted AR(1)–SV models, we perform residual analysis. Under the normal assumption for u_t, consider the standardized residual at time t from (3.46):

$$\hat{u}_t = \frac{y_t - \hat{\phi}_1 y_{t-1}}{\hat{\sigma}_t}, \qquad t = 1, ..., T,$$

where y_t is the observed return and $\hat{\sigma}_t = \exp\{\hat{\lambda}_t/2\}$, which is the estimated volatility displayed in Fig. 3.10. If the fitted model is adequate for the data, $\hat{u}_t$ follows a standard normal distribution.

Figures 3.11 and 3.12 display the diagnostic plots of $\hat{u}_t$ of the four series. The residual plots are given in the first row. The residual plot of the Kospi 200 series shows

Table 3.5: SML estimation of AR(1)–SV model

Parameter	KOSPI 200	RBC	NIKKEI 225	$/YEN
ϕ_1	0.1562	0.1582	-0.0000	0.0072
MC std. error	0.0001	0.0002	0.0001	0.0001
Std. error	0.0193	0.0255	0.0246	0.0213
α	0.0100	0.2001	0.0089	0.0184
MC std. error	0.0002	0.0002	0.0003	0.0004
Std. error	0.0213	0.0391	0.0401	0.0315
β	0.9856	0.9736	0.9765	0.9889
MC std. error	0.0001	0.0003	0.0002	0.0003
Std. error	0.0103	0.0114	0.0299	0.0201
γ	0.2201	0.1547	0.1847	0.1315
MC std. error	0.0012	0.0009	0.0015	0.0017
Std. error	0.0249	0.0392	0.0551	0.0519
$\hat{\kappa}$	15.6898	4.6599	6.2531	6.5655
CPU time[1]	11019	11948.3	8011.6	8898.7

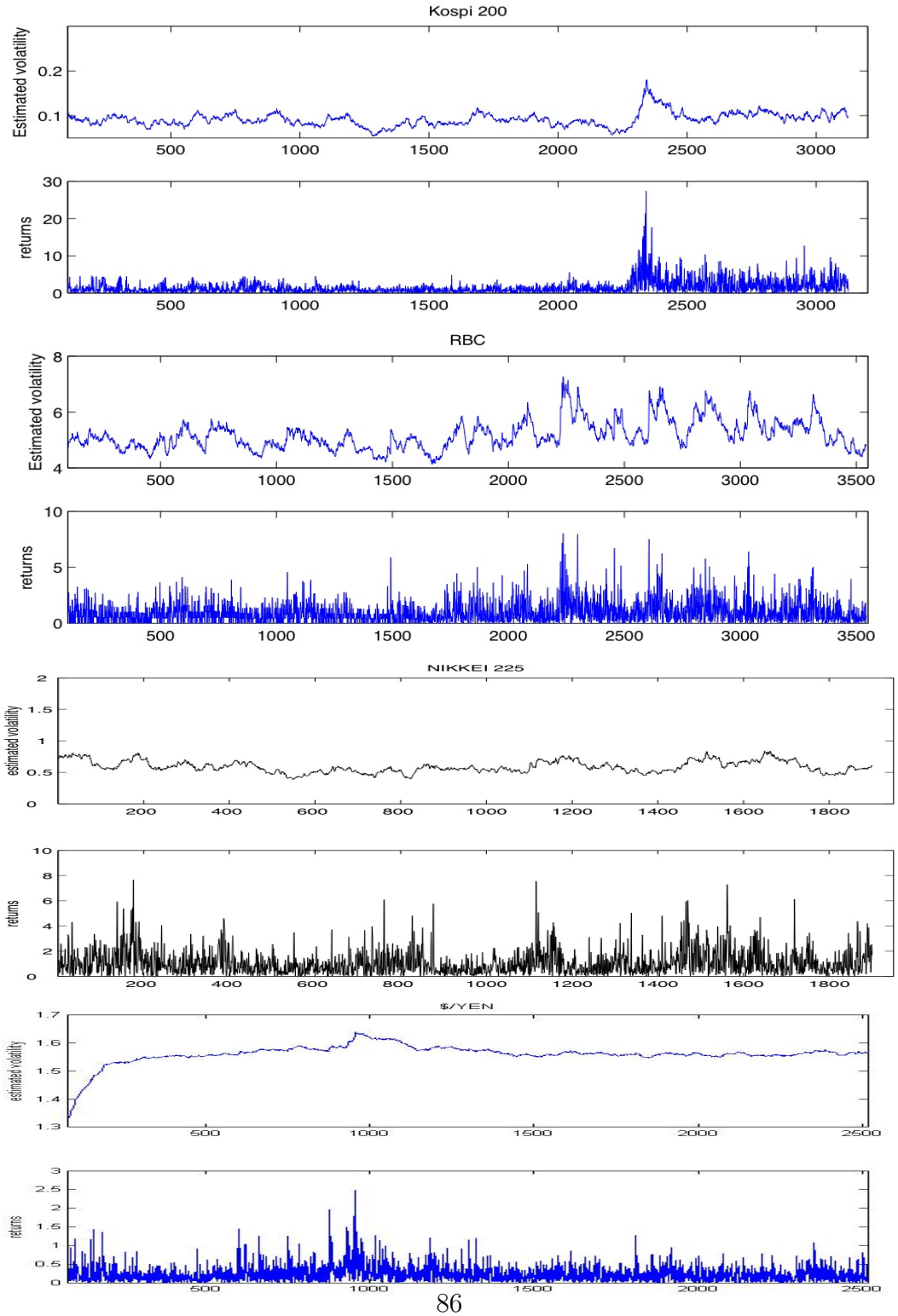

Figure 3.10: Top panels: estimated volatilities; Bottom panels: absolute values of return series

some similarity to the original return series, but the $\hat{u}_t$ are fairly random for the other series. The plots in the second row are the autocorrelation functions of the standardized residuals. There are almost no correlations for the four series. To validate the normality of the error process $\hat{u}_t$, QQ plots are given in the last row. There is considerable deviation at the tails of the QQ plots of the Kospi 200 and RBC series, but the corresponding deviation for the Nikkei 225 and \$/YEN series is less serious. This suggests that an improvement in the error assumption is necessary to capture the heavy tail of the Kospi 200 and RBC series.

We now choose a scaled Student-t distribution with 8 degrees of freedom, denoted $t(0, 1)$, as an alternative assumption for u_t. We do the estimation and diagnostics only for the Kospi 200 and RBC series because the violation of normality for these two series is more serious than that of the others.

The parameter estimation results under the normal and Student-t distributional assumptions are summarized in Table 3.6. The values in parentheses are the MC standard errors of the estimates. For both series, the estimates are close to those obtained under the normal assumption. In particular, the estimated value of the persistence parameter β is slightly larger and the estimate of the conditional standard deviation of the log variance γ is smaller under the Student-t distributional assumption. Based on this estimation result, we expect that the volatility of the estimated volatility under the Student-t error assumption is less than that under the normal assumption because of the smaller unconditional volatility of the log variance.

Figure 3.13 displays the estimated volatilities obtained by a particle filter under the normal and Student-t distributional assumptions. The solid line is the estimated volatility under the Student-t distributional assumption, and the dotted line is the estimated volatility under the normal assumption. The variation of the solid line is smaller, that is, the estimated volatilities under the $t(0, 1)$ assumption are less, as anticipated. QQ plots of the series under the normal and Student-t assumptions are given in Fig. 3.14. The improvement in the goodness of fit of the model can be clearly seen. Hence, for the Kospi 200 and RBC series, the AR(1)–SV model with a Student-t distributional error assumption provides a better fit to the data.

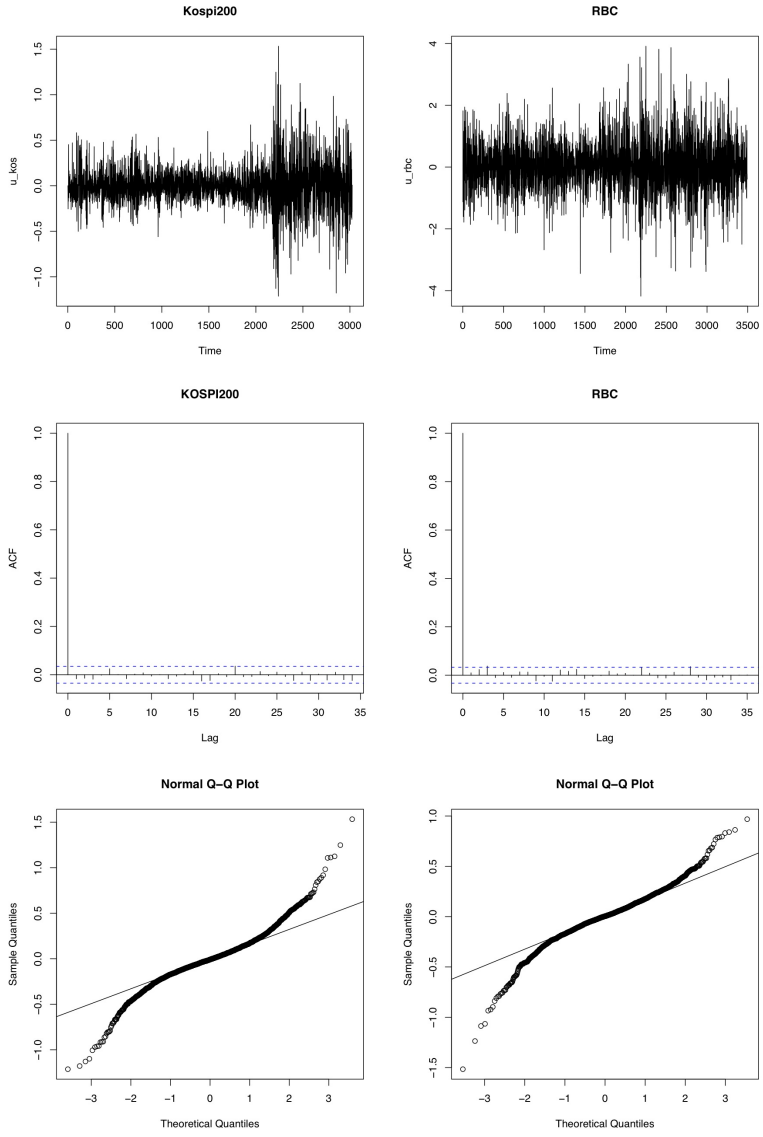

Figure 3.11: KOSPI 200 and RBC: The residual plots are in the first row, the ACFs of $\hat{u}_t$ are in the second row, and the QQ plots are in the last row.

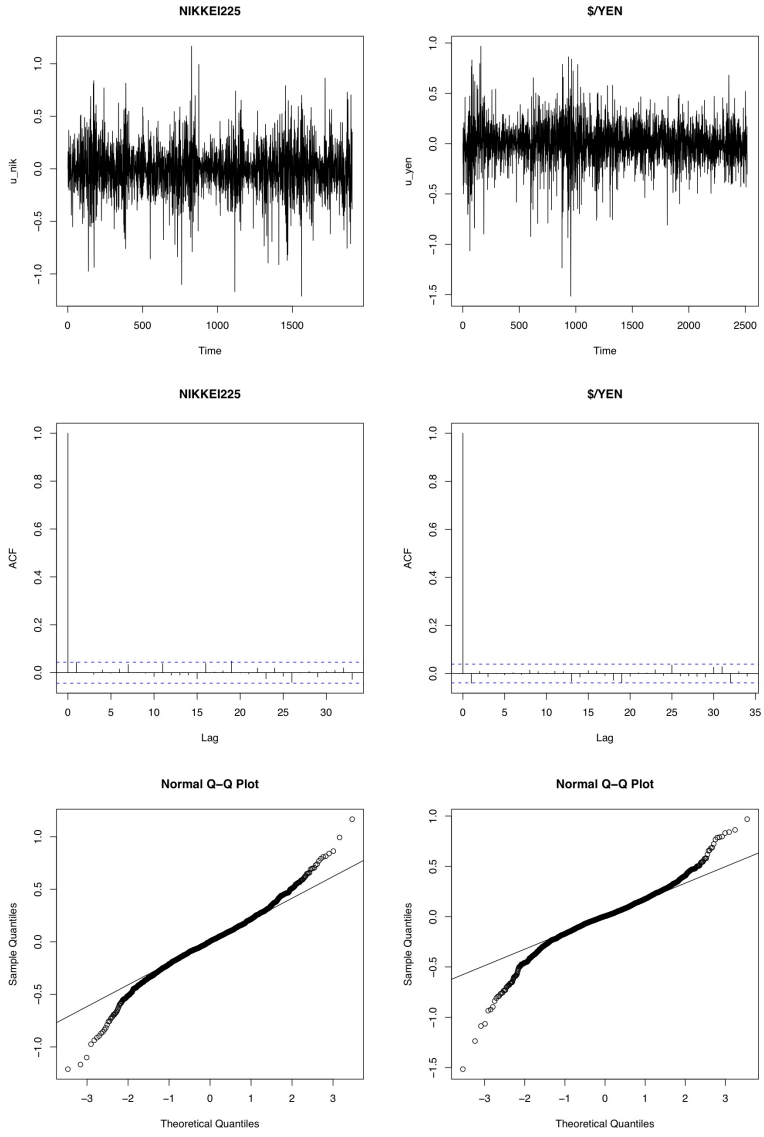

Figure 3.12: NIKKEI 225 and \$/YEN: The residual plots are in the first row, the ACFs of $\hat{u}_t$ are in the second row, and the QQ plots are in the last row.

Table 3.6: SML estimation of AR(1)–SV model for KOSPI 200 and RBC

Data	Assumption	ϕ	α	β	γ	$\hat{\kappa}$
Kospi 200	$u_t \sim$ t(0,1)	0.1579	0.0112	0.9898	0.1912	17.4293
		(0.0001)	(0.0002)	(0.0001)	(0.0010)	
	$u_t \sim$ N(0,1)	0.1562	0.01	0.9856	0.2201	15.6898
RBC	$u_t \sim$ t(0,1)	0.1590	0.1191	0.9785	0.1499	4.9811
		(0.0003)	(0.0004)	(0.0002)	(0.0009)	
	$u_t \sim$ N(0,1)	0.1582	0.2001	0.9736	0.1547	4.6599

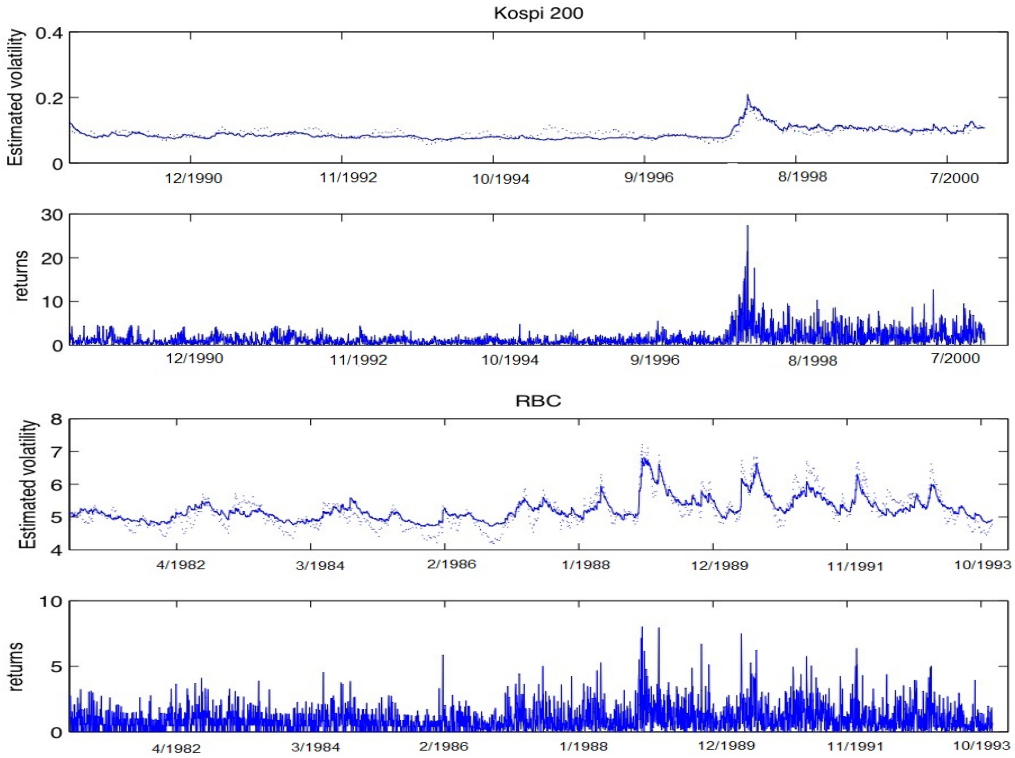

Figure 3.13: Series: Kospi 200 and RBC. Top panels: estimated volatility under normal (dotted line) and Student-t (solid line) distributional assumptions; bottom panels: absolute return

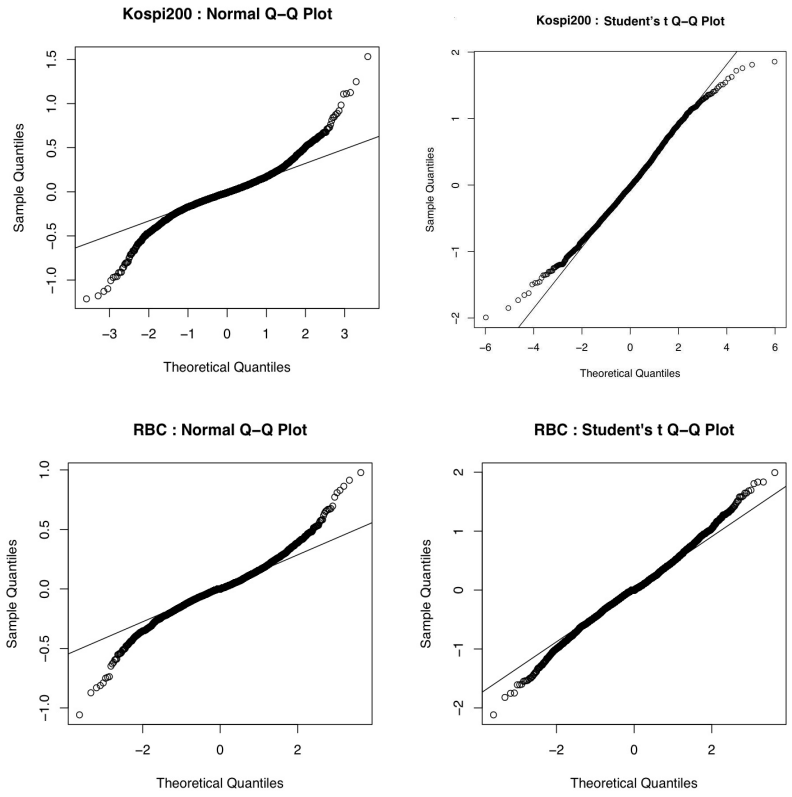

Figure 3.14: First row: QQ plots of Kospi 200 with normal innovation (left) and Student-t innovation (right); Second row: QQ plots of RBC with normal innovation (left) and Student-t innovation (right)

Chapter 4

Markov Switching Stochastic Volatility Models

In the discussion of the stochastic volatility models in previous chapters, we assumed that the parameters were constant over time. In the empirical analysis in Chapter 3, we used a first-order autoregressive model to capture serial correlation in the series and the movements in volatility over time. It seemed, however, in the analysis of the Kospi 200 series that the volatility changed drastically in the latter part of the series. Specifically, the volatility level was low before the Asia crisis in 1997 and suddenly increased afterwards. Such a change motivates us to consider a model in which the volatility may switch to different levels at random points in time.

Several researchers have acknowledged a structural shift in the volatility of financial return series and attempted to incorporate it into existing stochastic volatility models. Lamoureux and Lastrapes (1990) identified the occurrence of switching in financial series and suggested a Markov switching model (Hamilton (1988, 1989)) to model it. Hamilton and Susmel (1994) proposed the Markov switching ARCH (SWARCH) model, which incorporated the ARCH specification and switching in the conditional variance process. Cai (1994) proposed a similar model, called the switching-AR-Markov-ARCH model, to account for discrete shifts in the conditional variance process in the ARCH model. The difference between the SWARCH and switching-AR-Markov-ARCH models is the parameterization of the conditional mean process in the ARCH model. In the GARCH specification, Gray (1996) and Klaassen (2002) first combined the Markov process to characterize different levels of volatility with the GARCH model. Dueker (1997)

considered some fat-tailed distributions for the error in the conditional variance, including the Student-t distribution and GED. Marucucci (2005) extensively studied the volatility forecasting ability of the Markov regime-switching GARCH (MRS-GARCH) model.

As an application of the Markov switching model (Hamilton (1988, 1989)) in the SV specification, So, Lam, and Li (1998) combined the standard SV model with a first-order Markov process, called the Markov switching SV (MSSV) model. In their paper, they considered the MSSV model with three regimes (high-, medium-, and low-volatility states) to fit S&P 500 weekly returns. They used the MCMC method with Gibbs sampling to estimate the parameters and estimated the volatility by the smoothing method (Albert and Chib (1993)). They showed that the persistence in the volatility of S&P 500 returns was explained by low- and medium-volatility states, and the crash on Black Monday in 1987 was captured by the high-volatility state. Kalimipalli and Susmel (2004) proposed a regime-switching stochastic volatility (RSV) model, but the primary purpose of their model was to fit the dynamics of short-term interest rates. Shibata and Watanabe (2005) fitted the MSSV model with two regimes (So *et al.* (1998)) to TOPIX weekly returns and employed the MCMC method for parameter estimation and a particle filter for volatility estimation.

It is worthwhile to note that all the earlier studies assumed a normal distribution for the error in the conditional variance process in the SV specification in regime-switching SV models. To extend So *et al.* (1998)'s work, we consider the MSSV model with not only a normal error but also a heavy-tailed error. We choose a scaled Student-t distribution as a heavy-tailed error assumption. To estimate the parameters of the MSSV model, we consider the SML with EIS technique, and a particle filter is used to estimate the volatility. In our empirical study, we fit the MSSV model with two regimes to four financial series (Kospi 200, S&P 500, Russell 2000, and Nikkei 225), perform model diagnostics by residual analysis, and compare the goodness of fit of the MSSV model with that of the AR–SV model.

This chapter is organized as follows. We introduce the MSSV model with K regimes in Section 1 and derive its likelihood function in Section 2. Section 3 describes the SML method with the EIS technique for parameter estimation and a particle filter for volatility estimation. Section 4 summarizes the results of our empirical analysis with four financial return series.

4.1 Model

The standard stochastic volatility model (Taylor (1986)) for a financial time series y_t is

$$y_t = \exp\{\lambda_t/2\}u_t, \tag{4.1}$$
$$\lambda_t = \alpha + \beta\lambda_{t-1} + \gamma v_t, \tag{4.2}$$

where the assumptions given in Chapter 1 hold. Suppose that s_t is a latent discrete variable with a sample space $\{0, 1, ..., K-1\}$. Assume that s_t is a K-state first-order Markov process with transition probability matrix P given by

$$P = \begin{pmatrix} p_{0,0} & p_{0,1} & \cdots & p_{0,K-1} \\ p_{1,0} & p_{1,1} & \cdots & p_{1,K-1} \\ \vdots & \vdots & \cdots & \vdots \\ p_{K-1,0} & p_{K-1,1} & \cdots & p_{K-1,K-1} \end{pmatrix},$$

where a transition probability from state i to state j is denoted $p_{i,j} = P(s_t = j | s_{t-1} = i)$ and $\sum_{j=0}^{K-1} p_{ij} = 1$ for all $i = 0, 1, ..., K-1$. The latent variable s_t defines a regime that has unique properties. For instance, when $K = 2$, the conditional mean of the log variance is low when $s_t = 0$ and high when $s_t = 1$.

To incorporate discrete shifts in volatility into the SV model, So *et al.* (1998) generalized Eq. (4.2) with a Markov-switching model (Hamilton (1988, 1989)) as

$$\lambda_t(s_t) = \alpha(s_t) + \beta\lambda_{t-1} + \gamma v_t, \tag{4.3}$$

where

$$\alpha(s_t) = \alpha_0 + \sum_{j=1}^{K-1} \alpha_j I_{jt} \tag{4.4}$$

for $\alpha_j > 0$ for all $j = 1, ...K-1$. I_{jt} is an indicator variable at state j at time t and equals one if $s_t \geq j$ or zero otherwise. The log variance $\lambda_t(s_t)$ at state j is denoted $\lambda_t^{(j)}$ for all $j = 0, 1, ...K-1$. β is a persistent parameter such that $|\beta| < 1$ and the volatility of the log variance γ is always nonnegative. Equations (4.1), (4.3), and (4.4) form the K-state Markov switching stochastic volatility model, denoted MSSV(K).

In particular, when $K = 2$, the MSSV model (MSSV(2)) has two states and Eq. (4.3) is written as

$$\lambda_t^{(0)} = \alpha_0 + \beta\lambda_{t-1} + \gamma v_t, \quad \text{if } s_t = 0 \tag{4.5}$$

$$\lambda_t^{(1)} = \alpha_0 + \alpha_1 + \beta\lambda_{t-1} + \gamma v_t, \text{ if } s_t = 1 \tag{4.6}$$

for $\alpha_1 > 0$. Here, s_t is a latent Markov process with a sample space $\{0, 1\}$. We will refer to the regime when $s_t = 0$ as regime 0 and that when $s_t = 1$ as regime 1. The positive coefficient α_1 ensures a high level in volatility for regime 1. Thus, regime 0 is a low-level volatility state and regime 1 is a high-level volatility state. The transition probability matrix P of s_t has the following form:

$$P = \begin{pmatrix} p & 1-p \\ 1-q & q \end{pmatrix}.$$

4.2 Likelihood Function

The likelihood associated with a financial time series $Y_T = (y_1, ..., y_T)$ for the MSSV(K) model is similar to the likelihood for the standard stochastic volatility model. One distinction is the coefficients in the latent log-variance process λ_t; they are constant over time in the standard SV model and vary depending on the state at time t in the MSSV(K) model. The likelihood functions under a normal and a scaled Student-t distributional assumption for the error term u_t in Eq. (4.1) are given in this section.

The likelihood associated with $Y_T = (y_1, ..., y_T)$ in Eqs. (4.1), (4.3), and (4.4) is defined by

$$L(\theta; Y_T) = \prod_{t=1}^{T} f(y_t, \lambda_t, s_t; \theta) \tag{4.7}$$

where T is an index of last time when returns are observed, $\Lambda_T = (\lambda_1, ..., \lambda_T)$ is a vector of the latent variable λ_t, and $S_T = (s_1, ..., s_T)$ is a vector of the latent variable s_t. The parameter vector $\theta = (\vec{\alpha}_K, \beta, \gamma, P)$ is to be estimated, where $\vec{\alpha}_K = (\alpha_0, \alpha_1, ..., \alpha_{K-1})$ such that $\sum_{j=0}^{K-1} \alpha_j = 1$ and P is a transition probability matrix.

For simplicity, let us consider the MSSV model with two regimes, denoted MSSV(2). The joint density function $f(y_t, \lambda_t, s_t; \theta)$ in Eq. (4.7) can be written

$$\begin{aligned} f(y_t, \lambda_t, s_t; \theta) &= f(y_t, \lambda_t | s_t = 0) P(s_t = 0) + f(y_t, \lambda_t | s_t = 1) P(s_t = 1) \\ &= f(y_t, \lambda_t | s_t = 0) P_{0,t} + f(y_t, \lambda_t | s_t = 1) P_{1,t} \\ &= \sum_{i=0}^{1} f(y_t, \lambda_t | s_t = i) P_{i,t}, \end{aligned} \tag{4.8}$$

where $P_{i,t}$ is the marginal distribution, which is the probability that the Markov chain is in regime i at time t for $i = 0, 1$ such that $P_{0,t} + P_{1,t} = 1$. Note that the parameter θ is ignored on the right-hand side for notational convenience, and this convention will be followed hereafter. In Eq. (4.8), the marginal distribution $P_{0,t}$ has the following form:

$$\begin{aligned} P_{0,t} &= P(s_t = 0) \\ &= (1 - q) \left[\frac{f(y_t | \lambda_{t-1}^{(1)})(1 - P_{0,t-1})}{f(y_t | \lambda_{t-1}^{(0)}) P_{0,t-1} + f(y_t | \lambda_{t-1}^{(1)})(1 - P_{0,t-1})} \right] \\ &+ p \left[\frac{f(y_t | \lambda_{t-1}^{(0)}) P_{0,t-1}}{f(y_t | \lambda_{t-1}^{(0)}) P_{0,t-1} + f(y_t | \lambda_{t-1}^{(1)})(1 - P_{0,t-1})} \right], \end{aligned} \tag{4.9}$$

where $f(y_t | \lambda_t^{(i)})$ denotes a conditional density $f(y_t | \lambda_t, s_t = i)$ and $\lambda_t^{(i)}$ denotes a latent variable λ_t in regime i for $i = 0, 1$.

Assuming u_t follows a standard normal distribution, the conditional density $f(y_t | \lambda_t^{(i)})$ for $i = 0, 1$ has the following normal density function:

$$f(y_t | \lambda_t^{(i)}) = \frac{1}{\sqrt{2\pi \exp\{\lambda_t^{(i)}\}}} \exp\left[-\frac{y_t^2}{2 \exp\{\lambda_t^{(i)}\}} \right]. \tag{4.10}$$

Under the scaled Student-t distributional assumption for u_t, it has the following form:

$$f(y_t | \lambda_t^{(i)}) = \left(\sqrt{\pi(\omega - 2) \exp\{\lambda_t^{(i)}\}} \right)^{-1} \frac{\Gamma\left(\frac{\omega+1}{2}\right)}{\Gamma\left(\frac{\omega}{2}\right)} \left(1 + \frac{y_t^2}{\exp\{\lambda_t^{(i)}\}(\omega - 2)} \right)^{-\frac{\omega+1}{2}}. \tag{4.11}$$

The conditional density $f(y_t, \lambda_t | s_t = i)$ in Eq. (4.8) for $i = 0, 1$ can be factorized as

$$\begin{aligned} f(y_t, \lambda_t | s_t = i) &= f(y_t | \lambda_t^{(i)}) f(\lambda_t | \lambda_{t-1}, s_t = i) \\ &= f(y_t | \lambda_t^{(i)}) f(\lambda_t^{(i)} | \lambda_{t-1}), \end{aligned} \tag{4.12}$$

where the conditional density $f(y_t | \lambda_t^{(i)})$ is given in Eq. (4.10) or (4.11) and $f(\lambda_t^{(i)} | \lambda_{t-1})$ for $i = 0, 1$ can be expressed as

$$f(\lambda_t^{(i)} | \lambda_{t-1}) = \sum_{j=0}^{1} f(\lambda_t^{(i)} | \lambda_{t-1}^{(j)}) p_{ji,t-1}, \tag{4.13}$$

where $f(\lambda_t^{(i)} | \lambda_{t-1}^{(j)})$ is a normal density function with mean $\alpha_0 + \alpha_1 i + \beta \lambda_{t-1}^{(j)}$ and variance γ^2 and the probability $p_{ji,t}$ is given by

$$p_{ji,t} = \frac{p_{ji} P_{j,t}}{P_{i,t+1}}, \tag{4.14}$$

where the marginal distribution $P_{j,t}$ for $j = 0, 1$ is given in Eq. (4.9).

4.3 Parameter Estimation

Similarly to the SV model case, the likelihood in Eq. (4.7) cannot be evaluated analytically because of the presence of the latent process λ_t and the latent state process s_t. To numerically evaluate the likelihood and estimate the parameters in the MSSV model, we employ the SML method. In the implementation of the SML method for the MSSV model, the EIS technique is applied to reduce the MC standard error. For simplicity, we will describe the parameter estimation of the MSSV(2) model. The estimation procedure can be generalized to the K-state MSSV model for $K > 2$.

Setting $\theta = (\vec{\alpha}_1, \beta, \gamma, P)$, the likelihood function in Eq. (4.7) can be re-expressed as

$$\begin{aligned} L(\theta; Y_T) &= \prod_{t=1}^{T} \sum_{i=0}^{1} f(y_t | \lambda_t, s_t = i, \theta) f(\lambda_t, s_t = i, \theta) \\ &= \prod_{t=1}^{T} \sum_{i=0}^{1} f(y_t | \lambda_t^{(i)}, \theta) f(\lambda_t^{(i)} | \lambda_{t-1}, \theta) P_{i,t}, \end{aligned} \tag{4.15}$$

where the marginal distribution $P_{i,t}$ for $i = 0, 1$ is given in Eq. (4.9).

In a crude MC simulation, the MLE of a parameter vector θ can be obtained by maximizing the estimated likelihood given by

$$\hat{L}(\theta; Y_T) = \frac{1}{N} \sum_{n=1}^{N} \left(\prod_{t=1}^{T} \sum_{i=0}^{1} f(y_t|\lambda_{t,n}^{(i)}, \theta) f(\lambda_{t,n}^{(i)}|\lambda_{t-1,n}, \theta) P_{i,t} \right),$$

where N samples of $\lambda_t^{(i)}$ are drawn from the conditional density $f(\lambda_t^{(i)}|\lambda_{t-1}, \theta)$ given in Eq. (4.13). Similarly to the SML implementation in the SV model, $\lambda_t^{(i)}$ drawn from the conditional density of $\lambda_t^{(i)}$ given λ_{t-1} will need many iterations to achieve convergence. Instead of increasing the number of random draws $\lambda_t^{(i)}$ for a certain accuracy, we use the EIS algorithm to reduce the MC sampling variance with a smaller number of iterations and to incorporate information from y_t.

To apply the EIS algorithm for estimating θ in the MSSV(2) model, let us consider the following factorization of the likelihood function in Eq. (4.7):

$$\begin{aligned} L(\theta; Y_T) &= \int \int_{R^T} f(Y_T, \Lambda_T, S_T; \theta) d\Lambda_T dS_T \\ &= \int \int_{R^T} \frac{f(Y_T, \Lambda_T, S_T; \theta)}{g(\Lambda_T|S_T, Y_T, \theta)} g(\Lambda_T|S_T, Y_T, \theta) d\Lambda_T dS_T \\ &\doteq \int \int_{R^T} h(Y_T, \Lambda_T, S_T; \theta) g(\Lambda_T|S_T, Y_T, \theta) d\Lambda_T dS_T, \end{aligned} \tag{4.16}$$

where g and h denote an importance function (IF) and a remainder function (RF), respectively. The EIS technique chooses in the following way an IF $g(\Lambda_T|S_T, Y_T, \theta)$ that can provide a good approximation of $f(Y_T, \Lambda_T, S_T; \theta)$.

Consider a particular factorization:

$$g(\Lambda_T|S_T, Y_T, \theta) = \prod_{t=1}^{T} \sum_{i=0}^{1} g(\lambda_t^{(i)}|\lambda_{t-1}, y_t, \theta) P_{i,t} \tag{4.17}$$

$$= \prod_{t=1}^{T} \sum_{i=0}^{1} f(\lambda_t^{(i)}|\lambda_{t-1}, y_t, e_t^{(i)}, \theta) P_{i,t} \tag{4.18}$$

$$= \prod_{t=1}^{T} \sum_{i=0}^{1} C(\lambda_{t-1}, e_t^{(i)}) k(\Lambda_t, e_t^{(i)}, s_t = i) P_{i,t}, \tag{4.19}$$

where $e_t^{(i)}$ for $i = 0, 1$ is an auxiliary parameter vector depending on y_t and C is an integral constant that makes $k(\Lambda_t, e_t^{(i)}, s_t = i)$ a density function, given by

$$C(\lambda_{t-1}, e_t^{(i)}) = \frac{1}{\int k(\Lambda_t, e_t^{(i)}, s_t = i) d\lambda_t}.$$

EIS seeks a functional approximation $k(\Lambda_t, e_t^{(i)}, s_t = i)$ that has a simpler functional form and is more analytically tractable than the original function $f(\lambda_t^{(i)}|\lambda_{t-1}, \theta)$. The choice of $k(\Lambda_t, e_t^{(i)}, s_t = i)$ plays an important role in sampling $\lambda_t^{(i)}$ from the conditional density of $\lambda_t^{(i)}$ given λ_{t-1}, y_t, and $s_t = i$. It should be noted that we do not sample from the conditional density of λ_t given λ_{t-1} and $s_t = i$ only. For ease of notation, we ignore θ in the functional notations of C and k.

Estimation of e_t

For the MSSV(2) model, a natural choice of $k(\Lambda_t, e_t^{(i)}, s_t = i)$ would be proportional to a Gaussian density for λ_t given λ_{t-1} and $s_t = i$. Thus, we can express the function k as a product of two functions:

$$k(\Lambda_t, e_t^{(i)}, s_t = i) = \xi(\lambda_t^{(i)}, e_t^{(i)})\ f(\lambda_t^{(i)}|\lambda_{t-1}, \theta). \tag{4.20}$$

Since $f(\lambda_t^{(i)}|\lambda_{t-1}, \theta)$ is a Gaussian density, an auxiliary vector $e_t^{(i)}$ can be estimated via regression coefficients from a linear regression if $\xi(\lambda_t^{(i)}, e_t^{(i)})$ is chosen as a Gaussian density kernel $\exp\{e_{1,t}^{(i)}\lambda_t^{(i)} + e_{2,t}^{(i)}(\lambda_t^{(i)})^2\}$. The regression equation is given by

$$\log f(y_t|\lambda_t^{(i)}, \theta) - \log C(\lambda_t^{(i)}, e_{t+1}) = constant + e_{1,t}^{(i)}\lambda_t^{(i)} + e_{2,t}^{(i)}(\lambda_t^{(i)})^2 + a_t^{(i)}, \tag{4.21}$$

for $t = 1, ..., T$ and $i = 0, 1$, where $a_t^{(i)}$ is a regression error term for regime i. Assuming $C(\lambda_T^{(i)}, e_{T+1}) = 1$, this regression can be done backward.

The details of the functional forms are as follows. Substituting the conditional density $f(\lambda_t^{(i)}|\lambda_{t-1})$ in Eq. (4.13) into Eq. (4.20), the function $f(\lambda_t^{(i)}|\lambda_{t-1}, y_t, e_t^{(i)}, \theta)$ in (4.18) is given by

$$f(\lambda_t^{(i)}|\lambda_{t-1}, y_t, e_t^{(i)}, \theta) = \frac{1}{\sqrt{2\pi V_t^{(i)}}} \exp\left[-\frac{(\lambda_t^{(i)} - M_t^{(i)})^2}{2V_t^{(i)}}\right], \tag{4.22}$$

where

$$M_t^{(i)} = \left(\frac{\alpha_0 + \alpha_1 i + \beta \lambda_{t-1}}{\gamma^2} + e_{1,t}^{(i)} \right) V_t^{(i)}, \;\; V_t^{(i)} = \frac{\gamma^2}{1 - 2e_{2,t}^{(i)} \gamma^2},$$

and $\log C(\lambda_t^{(i)}, e_{t+1})$ is chosen as

$$\log C(\lambda_t^{(i)}, e_{t+1}) = -\frac{(M_t^{(i)})^2}{2V_t^{(i)}} + \frac{(\alpha_0 + \alpha_1 i + \beta \lambda_{t-1})^2}{2\gamma^2} - \frac{1}{2} \log \frac{V_t^{(i)}}{\gamma^2}.$$

Replacing $\xi(\lambda_t^{(i)}, e_t^{(i)})$ in Eq. (4.20) with the Gaussian density, the IF g for the MSSV(2) model is written as

$$g(\Lambda_T, S_T | Y_T, \theta) = \prod_{t=1}^{T} \sum_{i=0}^{1} C(\lambda_{t-1}, \hat{e}_t^{(i)}) \exp\{\hat{e}_{1,t} \lambda_t^{(i)} + \hat{e}_{2,t} (\lambda_t^{(i)})^2\} f(\lambda_t^{(i)} | \lambda_{t-1}, \theta). \tag{4.23}$$

Using this IF, the corresponding likelihood is estimated by

$$\hat{L}(\theta; Y) = \frac{1}{N} \sum_{n=1}^{N} \prod_{t=1}^{T} \frac{f(y_t | \lambda_{t,n}^{(i)}, \theta)}{C(\lambda_{t-1,n}, \hat{e}_t^{(i)}) \exp\{\hat{e}_{1,t}^{(i)} \lambda_{t,n} + \hat{e}_{2,t}^{(i)} (\lambda_{t,n}^{(i)})^2\}}, \tag{4.24}$$

where $\lambda_{t,n}^{(i)}$ is drawn from $g(\Lambda_T, S_T | Y_T, \theta)$ in (4.23).

The conditional density function $f(y_t | \lambda_{t,n}^{(i)}, \theta)$ in the numerator is given in Eq. (4.10) for the normal error assumption and in Eq. (4.11) for the scaled Student-t error assumption.

In summary, the EIS algorithm to evaluate the likelihood for the MSSV(2) model in (4.16), given an initial parameter vector θ_0 and the transition probabilities p and q, is as follows:

1. Generate ϵ from Uniform[0,1]. Set $s_1 = 0$ if $\epsilon < \pi_0$; otherwise, $s_1 = 1$, where $\pi_0 = (1 - q)/(2 - p - q)$ is the steady-state probability in regime 0 for t = 1, ...T.
2. Generate Λ_T from Eq. (4.13).
3. Assuming $C(\lambda_T, e_{T+1}) = 1$, obtain e_t by performing T regressions as in (4.21), working backward from $t = T$ to $t = 1$.
4. Generate the new Λ_T from the importance density $g(\Lambda_T, S_T | Y_T, \theta)$ in (4.23).

5. Repeat Steps 3 and 4 until either e_t or the new λ_t converges for all t.

6. Compute the estimated likelihood in (4.24).

4.3.1 Volatility Estimation

We employ a particle filter (Pitt and Shephard (1999)) to sample λ_t from the posterior distribution $f(\lambda_t|Y_t, s_t, \hat{\theta})$. The detailed steps of implementing the particle filter along with the functional forms to estimate the log variance λ_t for the MSSV(2) model are given below.

The filtering density $f(\lambda_t|Y_t, s_t)$ is proportional to $f(y_t|\lambda_t, s_t, \hat{\theta})f(\lambda_t|Y_{t-1}, s_t, \hat{\theta})$ by Bayes' theorem. The conditional density function $f(\lambda_t|Y_{t-1}, s_t, \hat{\theta})$ is written as

$$f(\lambda_t|Y_{t-1}, s_t, \hat{\theta}) = f(\lambda_t^{(0)}|Y_{t-1}, \hat{\theta})P_{0,t} + f(\lambda_t^{(1)}|Y_{t-1}, \hat{\theta})(1 - P_{0,t}),$$

where $\lambda_t^{(i)}$ denotes the latent variable λ_t in regime i for $i = 0, 1$, and the conditional density $f(\lambda_t^{(i)}|Y_{t-1}, \hat{\theta})$ can be written as

$$f(\lambda_t^{(i)}|Y_{t-1}, \hat{\theta}) = \int f(\lambda_t^{(i)}|\lambda_{t-1}, \hat{\theta})f(\lambda_{t-1}|Y_{t-1}, \hat{\theta})d\lambda_{t-1}. \tag{4.25}$$

Then, the integral in (4.25) can be estimated by

$$f(\lambda_t^{(i)}|Y_{t-1}, \hat{\theta}) \approx \frac{1}{M}\sum_{m=1}^{M}\left(\sum_{j=0}^{1} f(\lambda_{t,m}^{(i)}|\lambda_{t-1,m}^{(j)}, \hat{\theta})p_{ji,t}\right),$$

where $\lambda_{t-1,m}^{(i)}$ is drawn from $f(\lambda_{t-1}^{(i)}|Y_{t-1}, \hat{\theta})$ for $m = 1, ..., M$ and the probability $p_{ji,t}$ is given in (4.14). This leads to

$$f(\lambda_t^{(i)}|Y_t, \hat{\theta}) \propto f(y_t|\lambda_t^{(i)}, \hat{\theta})\frac{1}{M}\sum_{m=1}^{M}\left(\sum_{j=0}^{1} f(\lambda_{t,m}^{(i)}|\lambda_{t-1,m}^{(j)}, \hat{\theta})p_{ji,t}\right) \tag{4.26}$$

for $i = 0, 1$.

Let $\lambda_{t|t-1}^{(i)} = \hat{\alpha} + \hat{\beta}(M^{-1}\sum_m \sum_j \lambda_{t-1,m}^{(j)} p_{ji,t})$. Note that the conditional density of y_t

given $\lambda_t^{(i)}$ can be written

$$\log f(y_t|\lambda_t^{(i)}, \hat{\theta}) = constant - \frac{1}{2}\lambda_t^{(i)} - \frac{y_t^2}{2}\exp\{-\lambda_t^{(i)}\}.$$

Let us call the right-hand side without the constant $\log f^*(y_t, \lambda_t^{(i)}, \hat{\theta})$. Then, the expansion of $\log f^*(y_t, \lambda_t^{(i)}, \hat{\theta})$ by a first-order Taylor approximation around $\lambda_{t|t-1}^{(i)}$ leads to:

$$\begin{aligned}
\log f^*(y_t, \lambda_t^{(i)}, \hat{\theta}) &\leq -\frac{1}{2}\lambda_t^{(i)} - \frac{y_t^2}{2}[(1 + \lambda_{t|t-1}^{(i)})\exp\{-\lambda_{t|t-1}^{(i)}\} - \lambda_t^{(i)}\exp\{-\lambda_{t|t-1}^{(i)}\}] \\
&= \log g^*(\lambda_t^{(i)}, \lambda_{t|t-1}^{(i)}, \hat{\theta}), \qquad (4.27)
\end{aligned}$$

where $\log g^*$ is the right-hand side of the inequality. Also, it can be shown via some algebra that

$$g^*(\lambda_t^{(i)}, \lambda_{t|t-1}^{(i)}, \hat{\theta}) f(\lambda_t^{(i)}|\lambda_{t-1,m}^{(i)}, \hat{\theta}) \propto \pi_m^{(i)} f_N(\lambda_t^{(i)}|\lambda_{t|t-1}^{*(i)}, \hat{\gamma}^2), \qquad (4.28)$$

where $f_N(\lambda_t^{(i)}|\lambda_{t|t-1}^{*(i)}, \hat{\gamma}^2)$ denotes a normal density of $\lambda_t^{(i)}$ with mean $\lambda_{t|t-1}^{*(i)}$, and variance $\hat{\gamma}^2$,

$$\lambda_{t|t-1}^{*(i)} = \hat{\alpha} + \hat{\beta}\lambda_{t-1}^{(i)} + \frac{\hat{\gamma}^2}{2}[y_t^2\exp\{-\lambda_{t|t-1}^{(i)}\} - 1], \qquad (4.29)$$

$$\pi_m^{(i)} = \sum_{j=0}^{1} \exp\left[\frac{1}{2\hat{\gamma}^2}\{(\hat{\alpha} + \hat{\beta}(\lambda_{t|t-1}^{*(i)})^2)^2 - (\lambda_{t|t-1}^{*(i)})^2\}\right] p_{ji,t}. \qquad (4.30)$$

From the relationship in (4.28), the right-hand side of (4.26) is bounded as

$$\begin{aligned}
f(\lambda_t^{(i)}|Y_t, \hat{\theta}) &\propto f(y_t|\lambda_t^{(i)}, \hat{\theta})\frac{1}{M}\sum_{m=1}^{M}\left(\sum_{j=0}^{1} f(\lambda_{t,m}^{(i)}|\lambda_{t-1,m}^{(j)}, \hat{\theta})p_{ji,t}\right) \\
&\leq g^*(\lambda_t^{(i)}, \lambda_{t|t-1}^{(i)}, \hat{\theta})\frac{1}{M}\sum_{j=1}^{M}\left(\sum_{j=0}^{1} f(\lambda_{t,m}^{(i)}|\lambda_{t-1,m}^{(j)}, \hat{\theta})p_{ji,t}\right) \\
&\propto \frac{1}{M}\sum_{j=1}^{M}\pi_m^{(i)} f_N(\lambda_t^{(i)}|\lambda_{t|t-1}^{*(i)}, \hat{\gamma}^2).
\end{aligned}$$

In summary, the algorithm for drawing λ_t is as follows:

1. Generate a Markov process s_t as Step 1 in the EIS algorithm given in the previous section.

2. Draw a proposal $\lambda_t^{prop(i)}$ for $i = 0, 1$ for the mixture density

$$\sum_{m=1}^{M} \pi_m^{(i)*} f_N(\lambda_t^{(i)} | \lambda_{t|t-1}^{*(i)}, \hat{\gamma}^2), \tag{4.31}$$

   where $\pi_m^{(i)*} = \pi_m^{(i)} / \sum \pi_m^{(i)}$.

3. Accept $\lambda_t^{prop(i)}$ with acceptance/rejection probability

$$f^*(y_t, \lambda_t^{(i)}, \hat{\theta}) / g^*(\lambda_t^{(i)}, \lambda_{t|t-1}^{(i)}, \hat{\theta}).$$

   If rejected, go to Step1.

Hence, the estimated volatility $\hat{\sigma}$ is calculated by $\hat{\sigma} = \exp\{\hat{\lambda}_t/2\}$.

Under the scaled Student-t distributional assumption for u_t, the conditional density of y_t given $\lambda_t^{(i)}$ is

$$f(y_t|\lambda_t^{(i)}) = \left(\sqrt{\pi(\omega-2)\exp\{\lambda_t^{(i)}\}}\right)^{-1} \frac{\Gamma\left(\frac{\omega+1}{2}\right)}{\Gamma\left(\frac{\omega}{2}\right)} \left(1 + \frac{y_t^2}{\exp\{\lambda_t^{(i)}\}(\omega-2)}\right)^{-\frac{\omega+1}{2}}$$

and so its corresponding acceptance/rejection probability f^*/g^* can be described by the following two equations:

$$\begin{aligned}
\log f^*(y_t, \lambda_t^{(i)}, \hat{\theta}) &= \frac{\omega}{2}\lambda_t^{(i)} - \frac{(\omega+1)}{2}\log[\exp\{\lambda_t^{(i)}\}(\omega-2) + y_t^2], \\
\log g^*(\lambda_t^{(i)}, \lambda_{t|t-1}^{(i)}, \hat{\theta}) &= \frac{\omega}{2}\lambda_t^{(i)} t - \frac{(\omega+1)}{2}\log[(\omega-2)((1+\lambda_{t|t-1}^{(i)})\exp\{\lambda_{t|t-1}^{(i)}\} \\
&+ \lambda_t^{(i)}\exp\{\lambda_{t|t-1}^{(i)}\}) + y_t^2].
\end{aligned}$$

Then, the weight function $\pi_m^{(i)*}$ is given in Eq. (4.30) and the mixture density is given in Eq. (4.31). The algorithm for the estimation of volatility in the case of the scaled Student-t distribution can be described as before, and the estimated volatility $\hat{\sigma}$ is calculated by $\hat{\sigma} = \exp\{\hat{\lambda}_t/2\}$.

Table 4.1: Summary statistics of simulated series

Series	Mean	Std. Dev.	Kurtosis
$p = 0.5$ and $q = 0.01$	-0.2780	13.0817	9.2602
$p = 0.8$ and $q = 0.01$	0.0145	4.4325	11.2182

4.4 Simulation Study

We conduct a simulation study to demonstrate parameter estimation using the EIS algorithm and volatility estimation using a particle filter in the MSSV model. Two series each with 1,000 observations were generated using the MSSV(2) model, one with $p = 0.8$ and the other with a smaller $p = 0.5$. The transition probability from regime 0 to regime 1 equals 0.2 when $p = 0.8$ and 0.5 when $p = 0.5$. The parameter q (*see* Table 4.1) was the same in both series. The hypothetical values of the parameters are similar to those found empirically. In particular, $p = 0.5$ and $q = 0.01$ gives a steady-state probability at regime 0 of 66.4% while it is 83.2% when $(p, q) = (0.8, 0.01)$.

Table 4.1 shows the summary statistics for the two simulated series. Both series have large kurtosis. Figure 4.1 displays time series plots of the two series, their autocorrelations, and the autocorrelations of the squared series. It is easy to see that there are two regimes of low and high volatility in the top and bottom plots on the left in Fig. 4.1. Volatility clustering is observed in both series. In addition, the ACF indicates that the simulated series are uncorrelated, but the squared series indicate persistence in their correlations.

The parameter estimation results are summarized in Table 4.2. All the estimates are close to the true values and the standard errors are fairly small. This indicates that the SML approach with the EIS algorithm can be applied to parameter estimation in the MSSV(2) model.

Volatility estimation is done by a particle filter. Figure 4.2 shows the absolute values of the returns along with the estimated volatility for each series. We can see that the estimated volatilities are able to capture the movement of the simulated series. In particular, the change in volatility is successfully captured.

We also performed model diagnostics to see the adequacy of the fit. The standardized

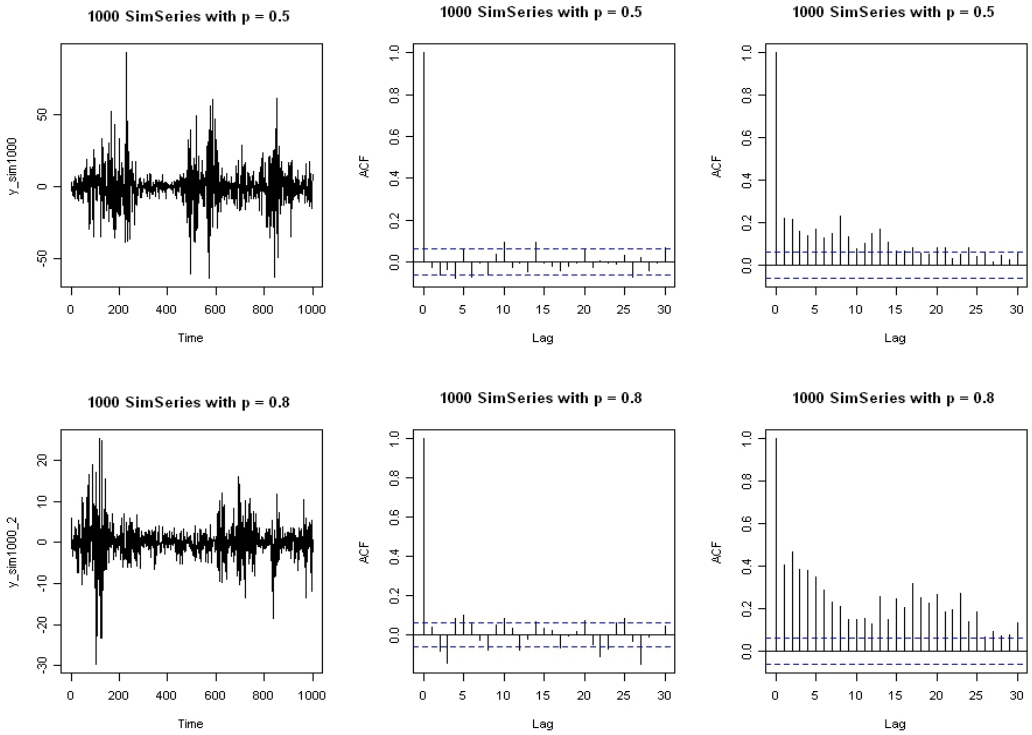

Figure 4.1: Simulated series plots: Returns (left), ACF of returns (center), and ACF of squared returns (right).

Table 4.2: Parameter estimates: MSSV(2) model

$p = 0.5$	α_0	α_1	β	γ	p	q
True Value	0.01	0.3	0.97	0.3	0.5	0.01
Estimated Value	0.0059	0.3014	0.9712	0.2976	0.4998	0.0098
Std. Error	0.0321	0.0129	0.0097	0.0188	0.0107	0.0098
$p = 0.8$	α_0	α_1	β	γ	p	q
True Value	0.01	0.3	0.97	0.3	0.8	0.01
Estimated Value	0.0099	0.3112	0.9643	0.3219	0.8176	0.0091
Std. Error	0.0362	0.0218	0.0038	0.0233	0.0212	0.0074

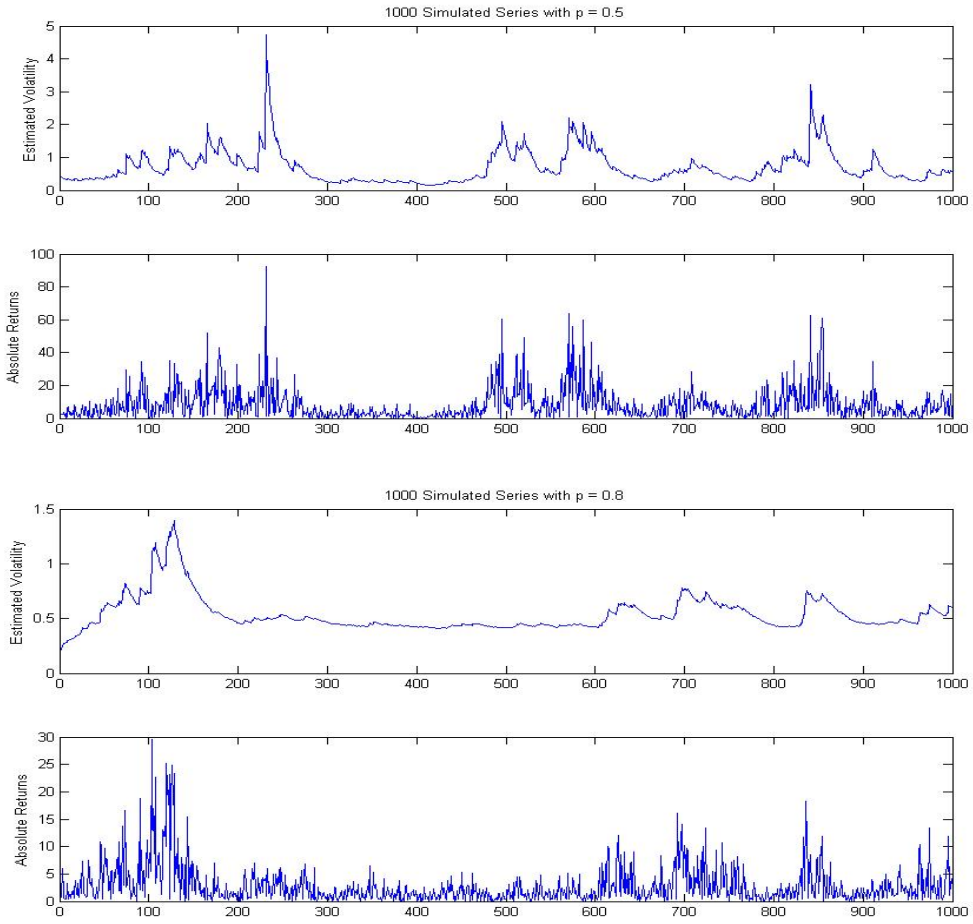

Figure 4.2: Top panels: estimated volatilities; Bottom panels: absolute values of return series.

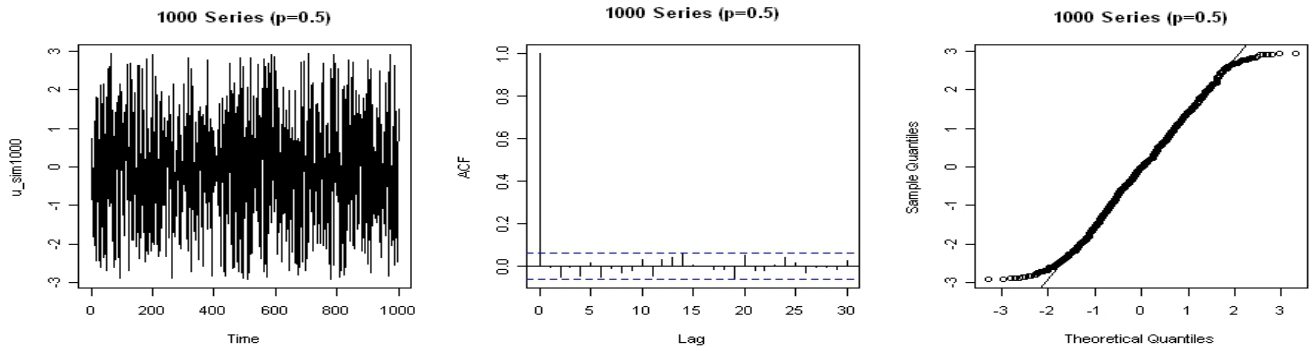

Figure 4.3: Diagnostic plots of simulated series

residual from the MSSV(2) model is defined by

$$\hat{u}_t = \frac{y_t}{\hat{\sigma}_t},$$

where $\hat{\sigma}_t = \exp\{\hat{\lambda}_t/2\}$ and $\hat{\lambda}_t$ is the estimated volatility obtained by a particle filter for $t = 1, ..., T$. If the estimated model is adequate, $\hat{u}_t$ is approximately a normal distribution with mean zero and variance one. Figure 4.3 displays the diagnostic plots for the first series with $p = 0.5$: residual plot, ACF plot, and QQ plot. The residuals vary over a constant band and the autocorrelations are negligible. The QQ plot indicates that the points are close to a straight line although there is a slight deviation at the tails. We conclude that, as expected, the MSSV(2) model fits the data fairly well . In addition, this simulation study shows that the SML approach with the EIS algorithm is reasonable for the estimation of the MSSV(2) model.

4.5 Empirical Analysis

We performed an empirical analysis of the MSSV(2) model using four return series. In each series, the SML approach with the EIS technique is used for the estimation of the parameters, and a particle filter is used to estimate the volatility. Moreover, model selection and model diagnostics are carried out. All the code is written in MATLAB 7.3.0.

Table 4.3: Summary statistics of data

Statistics	KOSPI 200	S&P500	RUSSEL 2000	NIKKEI 225
Sample Size	3124	2681	2681	2000
Mean	-0.0340	-0.0122	0.0072	0.0258
Std. Dev.	2.3084	1.3925	1.6954	1.4731
Kurtosis	15.5597	7.4143	4.0488	5.4816

4.5.1 Data

The following time series are considered for the empirical analysis: the daily closing prices of the Kospi 200 index from 1/4/1989 to 12/29/2000, the daily closing prices of the S&P 500 and Russel 2000 indices from 1/3/2000 to 8/31/2010, and the daily closing prices of the Nikkei 225 index from 2/3/1990 to 3/30/2000. The Kospi 200 and Nikkei 225 series were also used for the empirical analysis of the AR–SV model. The prices of the series are converted to mean-corrected log returns, defined by $y_t = 100 * [\log(p_t/p_{t-1}) - \frac{1}{T}\sum \log(p_t/p_{t-1})]$, where p_t is the price of an index at time t. The total number of observations and some summary statistics for each return series are given in Table 4.3.

The kurtoses of all the series are greater than three, which implies that the distributions of the series have heavier tails than the normal distribution. Figure 4.4 gives the time series of the returns. Figure 4.5 displays the autocorrelation functions of the series, and Fig. 4.6 gives those of the squares. The plots indicate that the volatility of the series shifts at several points throughout the time period. The Kospi 200, S&P 500, and Russel 2000 series have one notable shift in the volatility in the later period. The ACF plots indicate that there are few autocorrelations in the series except for Kospi 200 and large and slowly decaying correlations in the squared series. Given these findings, it seems reasonable to consider MSSV models for these series to capture the discrete switches in the volatility.

4.5.2 Estimation Results

The MSSV(2) model is used for each of the four return series and the SML approach with the EIS algorithm is implemented. The results are in Table 4.4. The MC sampling

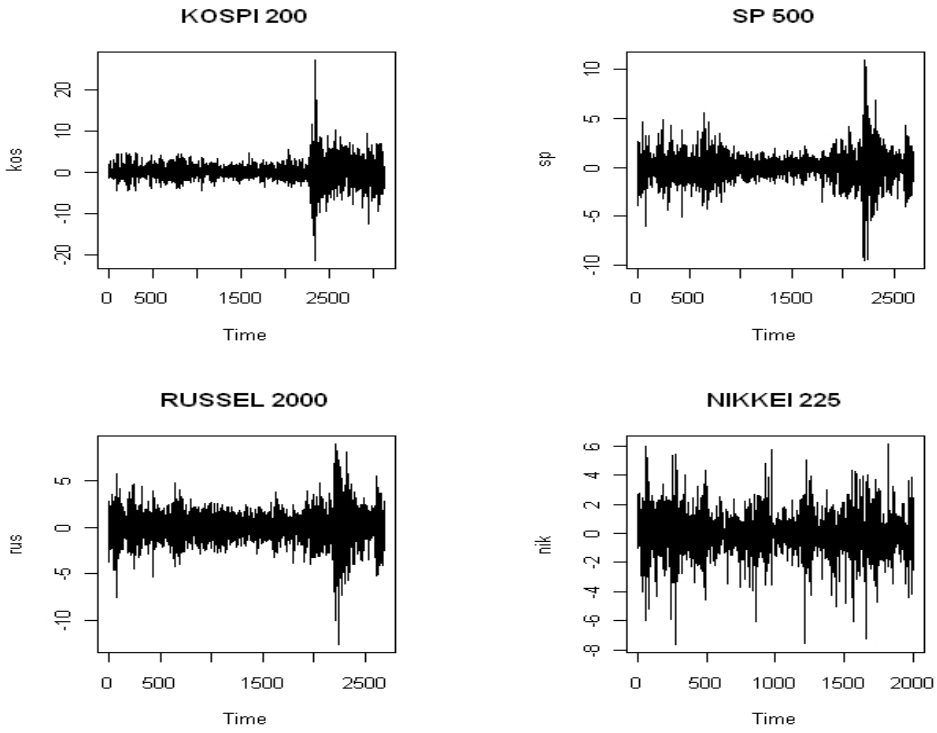

Figure 4.4: Plots of four series: KOSPI 200, S&P 500, RUSSEL 2000, and NIKKEI 225

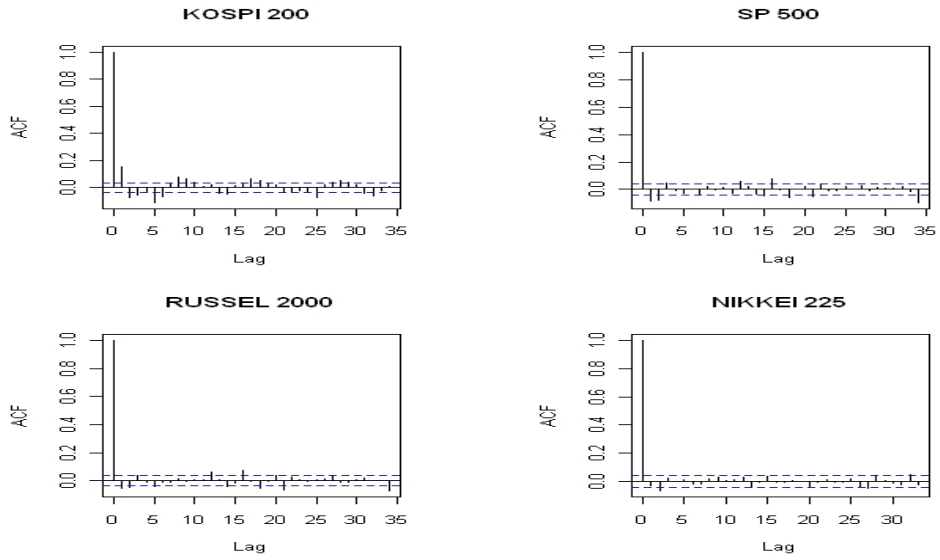

Figure 4.5: Autocorrelation plots of four return series: KOSPI 200, S&P 500, RUSSEL 2000, and NIKKEI 225

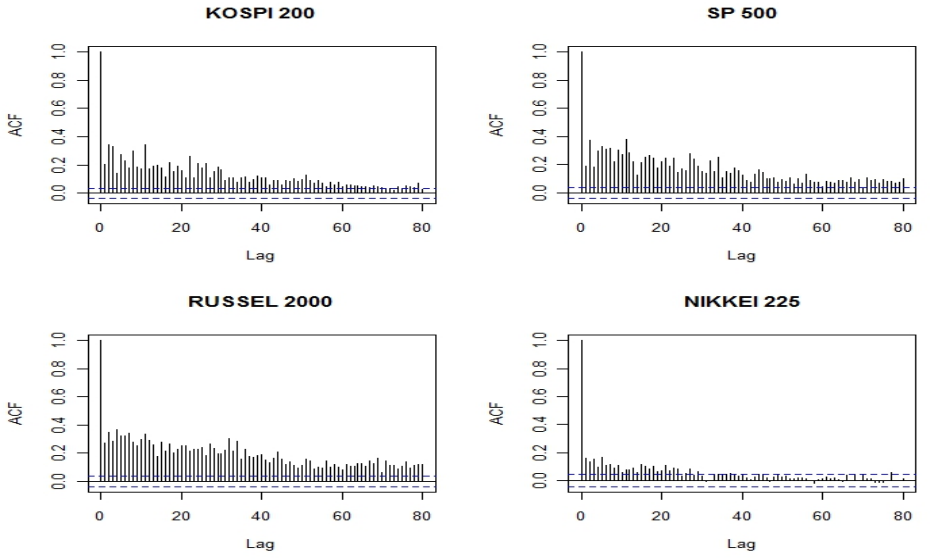

Figure 4.6: Autocorrelation plots of squared returns: KOSPI 200, S&P 500, RUSSEL 2000, and NIKKEI 225

errors and standard errors of the estimates are small. The estimated values of the persistent parameter β are greater than 0.9. This is commonly seen in the empirical analysis of stochastic volatility models. For all the series, the steady-state probabilities at regime 0 calculated by the formula $(1-q)/(2-p-q)$ with the estimated p and q are close to 60%. This means that all of the series are more likely to stay in a normal state, which is regime 0. This is consistent with the characteristics observed from the historical data.

Given the parameter estimation results in Table 4.4, we estimate the volatility $\sigma_t = \exp\{\lambda_t/2\}$ by a particle filter. For this estimation we use fifty simulated λ_t at each time step for $t = 1, ..., T$. Figure 4.7 shows the estimated volatility with the corresponding absolute values of the returns. The plot indicates that the estimated volatility from the MSSV model captures the movement of the return series reasonably well.

4.5.3 Model Diagnostics

We perform a residual analysis to check the adequacy of the MSSV(2) models for the four return series. Under the normal assumption for u_t, consider the standardized residual at time t:

$$\hat{u}_t \equiv \frac{y_t}{\hat{\sigma}_t}, \qquad t = 1, ..., T,$$

where y_t is the observed return and $\hat{\sigma}_t = \exp\{\hat{\lambda}_t/2\}$, which is the estimated volatility displayed in Fig. 4.7. If the fitted model is adequate, $\hat{u}_t$ is approximately standard normal.

Figure 4.8 displays the residual plots of the four series. The standardized residuals do not seem to be abnormal except for the Kospi 200 series. For the Kospi 200 series, there seems to be some change in the volatility. The plots in Fig. 4.9 are the autocorrelation functions of the standardized residuals. The autocorrelations in the residuals are small. In the Kospi 200 series, the first-lag correlation is similar to what was seen in the original series. The model fitting does not seem to have captured this well.

The QQ plots are given in Fig. 4.10 to validate the normality assumption for u_t. For the Russel 2000 series, most of the points are quite close to the straight line. The QQ plots of the S&P500 and Nikkei 225 series indicate that most of the points are close to the straight line although there are some deviations at both ends. For the Kospi 200 series, there is an indication that the normality assumption may not be adequate and an alternative assumption for u_t such as the Student-t distribution may be more appropriate.

Table 4.4: SML estimation of MSSV(2) model

Parameter	KOSPI 200	S&P 500	RUSSEL 2000	NIKKEI 225
α_0	0.1071	0.0108	0.0248	0.0157
MC std. error	0.0002	0.0002	0.0003	0.0002
Std. error	0.0341	0.0381	0.0319	0.0298
α_1	0.0149	0.0157	0.0352	0.0112
MC std. error	0.0001	0.0002	0.0000	0.0001
Std. error	0.0276	0.0301	0.0277	0.0286
β	0.9002	0.9584	0.9821	0.9189
MC std. error	0.0003	0.0001	0.0001	0.0001
Std. error	0.051	0.0489	0.0701	0.0821
γ	0.1078	0.155	0.2147	0.1215
MC std. error	0.0001	0.0002	0.0001	0.0002
Std. error	0.0431	0.0378	0.0403	0.0487
p	0.4218	0.3769	0.5171	0.4714
MC std. error	0.0000	0.0001	0.0001	0.0001
Std. error	0.0712	0.0698	0.0812	0.0714
q	0.1094	0.1328	0.1489	0.1826
MC std. error	0.0001	0.0002	0.0002	0.0003
Std. error	0.0627	0.0612	0.0703	0.0729
Steady-state prob. at regime 0 (%)	60.63	58.19	63.80	60.73
CPU time[1]	13284	12680	12673	11973

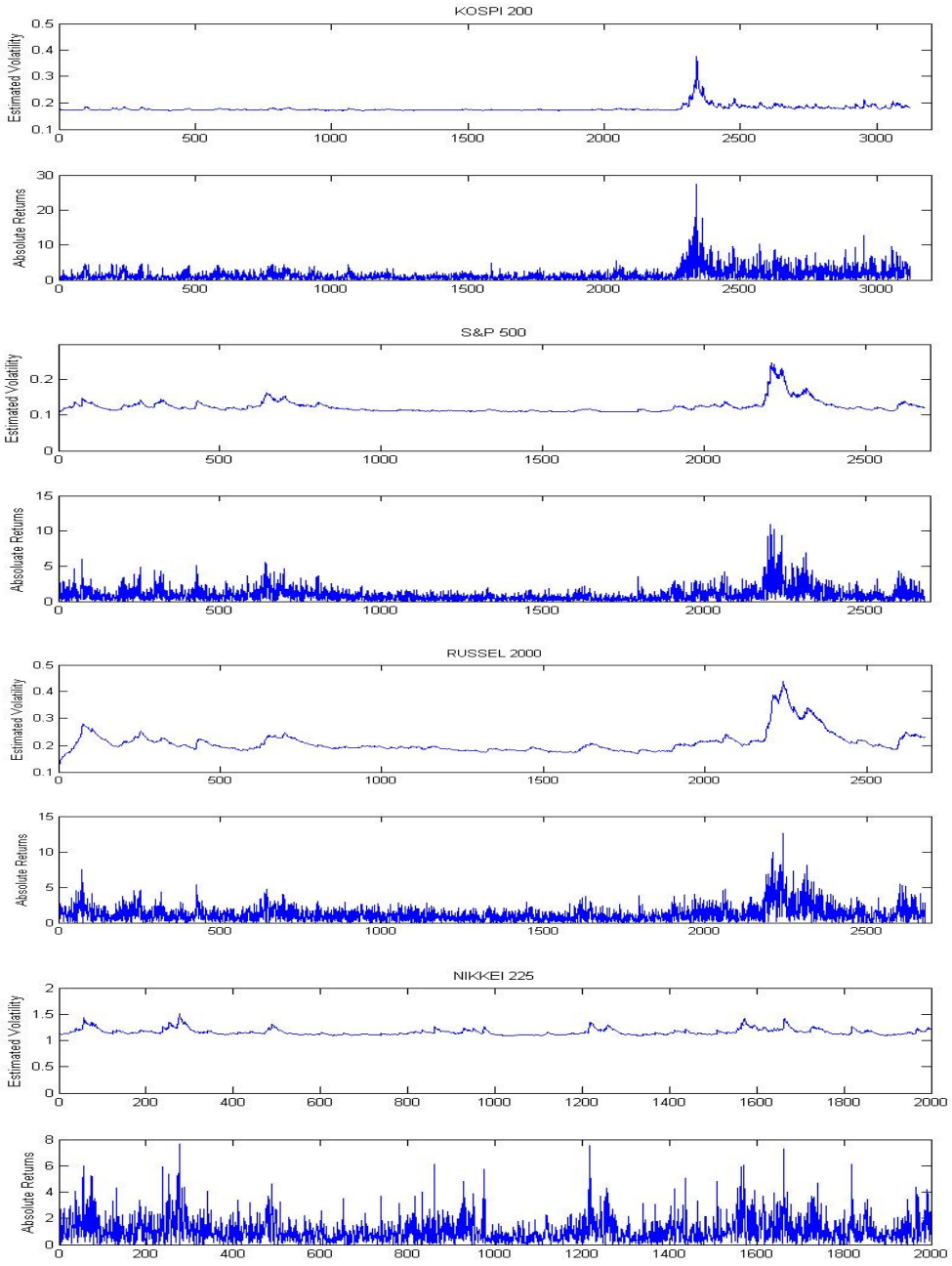

Figure 4.7: Estimated volatility with absolute returns of four series: KOSPI 200, S&P 500, RUSSEL 2000, and NIKKEI 225

We will consider such a specification for the Kospi 200 series in the next subsection. Based on our residual analysis so far, we can conclude that the MSSV(2) model is reasonable for capturing the time-varying volatility in the S&P 500, Russel 2000, and Nikkei 225 series. For the Kospi 200 series, a Student-t assumption for the error u_t may be necessary to improve the goodness of fit of the model.

4.5.4 Further Analysis of Kospi 200 index

In the empirical analysis of the AR–SV model in Chapter 3, we observed a noticeable change in volatility in the Kospi 200 series. Before the crisis in 1997 the volatility was low, while it was high after the crisis. This finding motivated us to consider a stochastic volatility model with regime switching. Further analysis of the Kospi 200 series with the AR(1)–SV model and the MSSV(2) model under normal and Student-t error assumptions is given in this section.

In the empirical analysis earlier, we indicated that the MSSV(2) model with the normal error assumption may have to be improved to better capture the time-dependent volatility. Specifically, the diagnostic tests suggest that the standard normal assumption for the error term u_t is not sufficient to capture the heavy tail of the return distribution of the Kospi 200 series. To investigate this further, we fit the Kospi 200 series with the MSSV(2) model with a scaled Student-t distribution with 8 degrees of freedom.

Figure 4.11 displays the estimated volatility of Kospi 200 from the AR(1)–SV and MSSV(2) models along with the absolute returns. The first two graphs show the estimated volatility from the AR(1)–SV model and the absolute Kospi 200 returns. In the first graph, the dotted line is the estimated volatility under a normal error assumption, and the solid line is the estimated volatility under a Student-t assumption. The last two graphs are the estimated volatility from the MSSV(2) model and the absolute Kospi 200 returns.

Both models seem to capture the movements of the returns fairly well, but the MSSV(2) model has a slight edge in capturing the volatility change. The AR(1)–SV model tends to underestimate the volatility. Regardless of the error assumptions, the average estimated volatility from the AR(1)–SV and MSSV(2) models is about 15% and 18%, respectively. Specifically, for the large jump in volatility in 1997, the estimated volatility of the KOSPI 200 index is only about 20% from that of the AR(1)–SV model, and about 35% from that of the MSSV(2) model. The one-year moving annualized volatility of Kospi

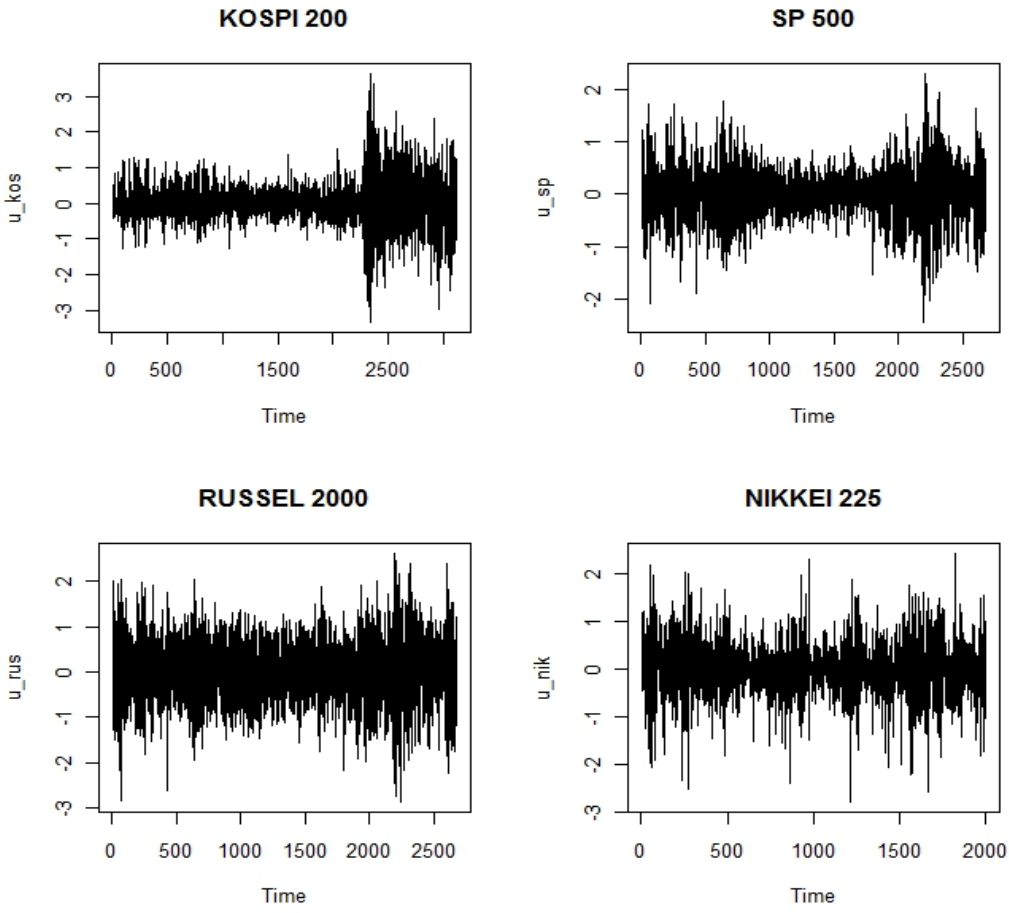

Figure 4.8: Residual plots of four series: KOSPI 200, S&P 500, RUSSEL 2000, and NIKKEI 225

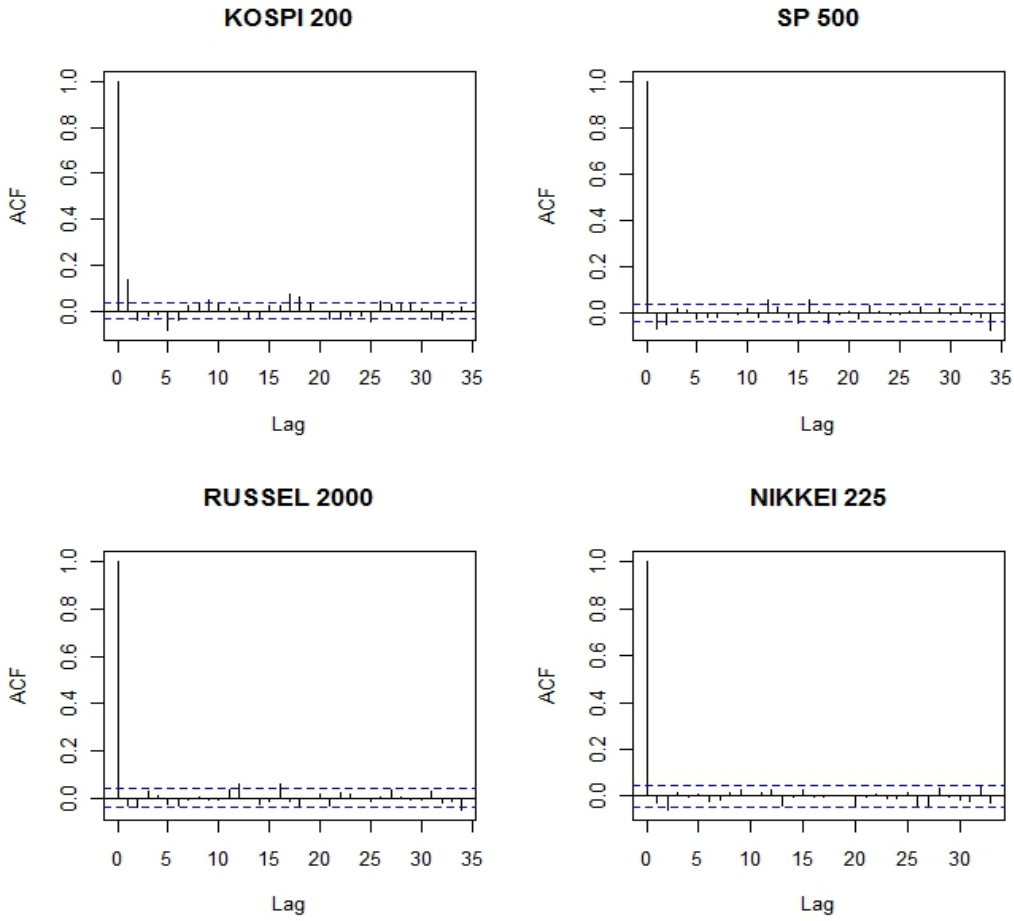

Figure 4.9: Autocorrelation plots of four series: KOSPI 200, S&P 500, RUSSEL 2000, and NIKKEI 225

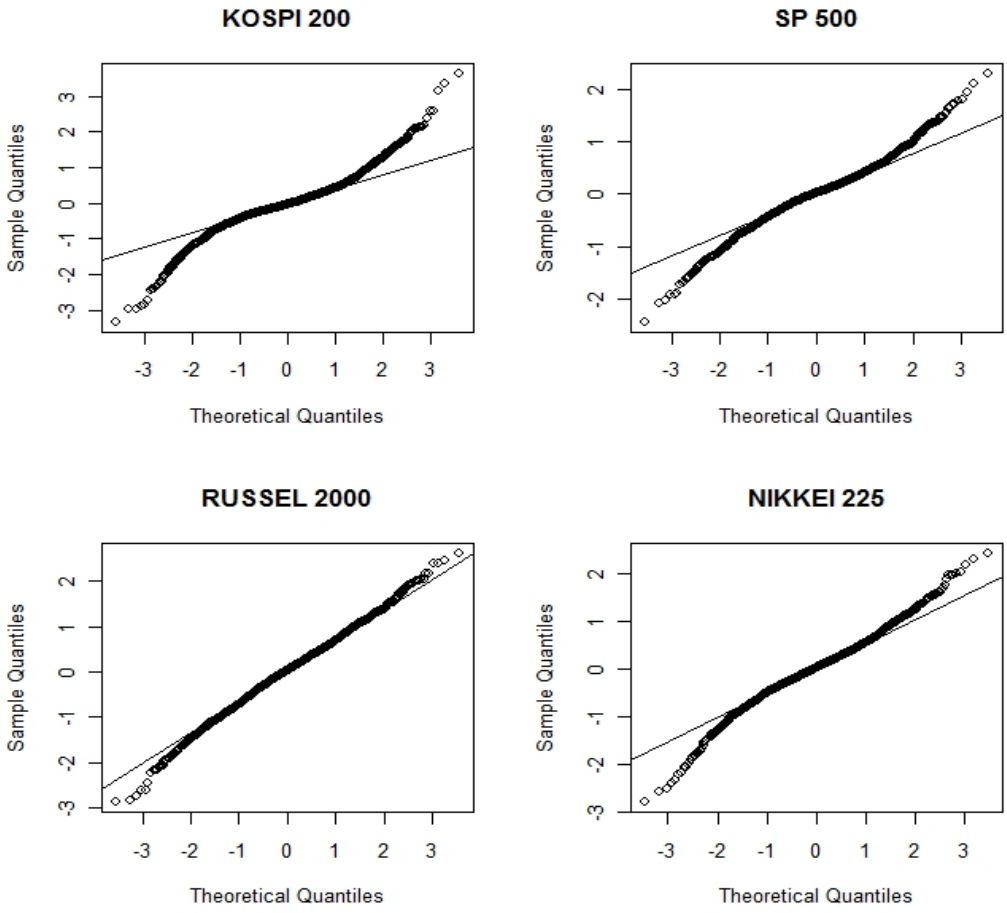

Figure 4.10: QQ plots of four series: KOSPI 200, S&P 500, RUSSEL 2000, and NIKKEI 225

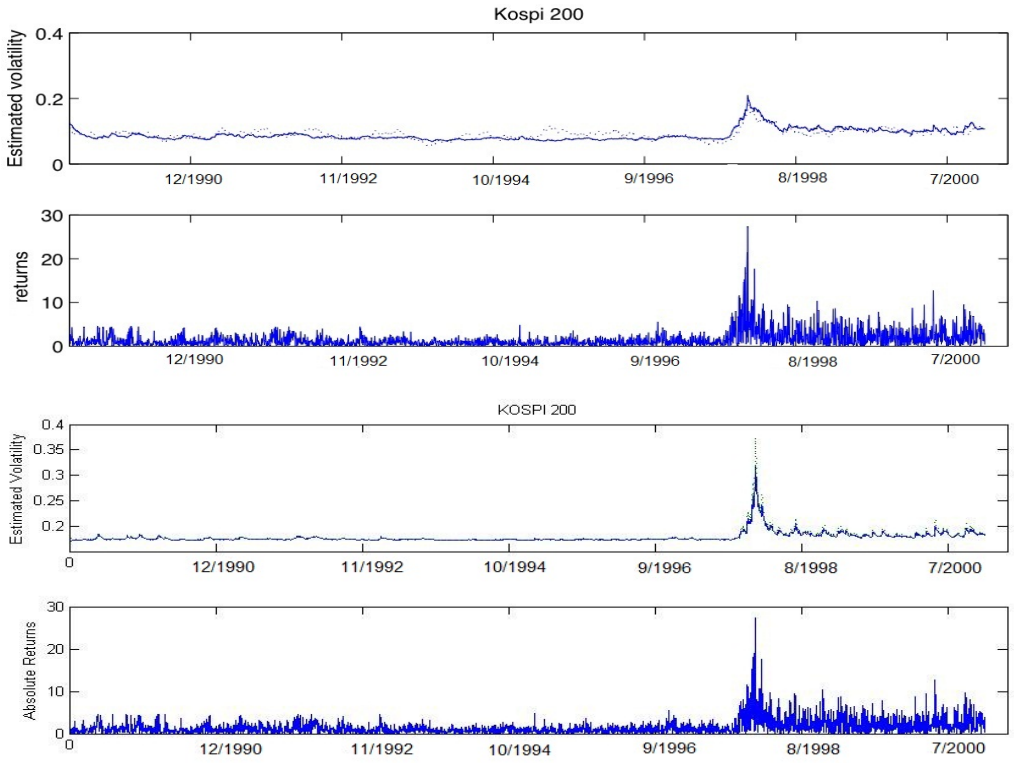

Figure 4.11: Estimated volatility and absolute returns of KOSPI 200: First plot is the estimated volatility from AR(1)–SV model with normal error assumption (dotted line) and Student-t error assumption (solid line). Second and fourth plots are absolute returns of KOSPI 200. Third plot is the estimated volatility from MSSV(2) model with normal error assumption (dotted line) and Student-t error assumption (solid line).

200 during late 1997 and 1998 is about 55%. The estimated annualized volatility of Kospi 200 during the period is 40% for the AR(1)–SV model and 60% for the MSSV(2) model. The estimated annualized volatility is calculated by the estimated daily volatility times the square root of 252. The volatility estimates from the MSSV(2) model are much closer to the actual values of the historical volatility. Given these findings, we conclude that the MSSV(2) model is better than the AR(1)–SV model at capturing the volatility of the Kospi 200 series.

Table 4.5 shows the parameter estimation results under the standard normal and the Student-t with 8 degrees of freedom assumptions for u_t. The persistent parameter β

Table 4.5: SML estimation of MSSV(2) model for KOSPI 200 series: Normal and Student-t error assumptions

Parameter	Normal(0, 1)	Student-t (8)
α_0	0.1071	0.0941
MC std. error	0.0002	0.0001
Std. error	0.0341	0.0129
α_1	0.0149	0.0132
MC std. error	0.0001	0.0001
Std. error	0.0276	0.0218
β	0.9002	0.9213
MC std. error	0.0003	0.0001
Std. error	0.051	0.0318
γ	0.1078	0.0839
MC std. error	0.0001	0.0002
Std. error	0.0431	0.0431
p	0.4218	0.3328
MC std. error	0.0000	0.0001
Std. error	0.0712	0.0522
q	0.1094	0.1182
MC std. error	0.0001	0.0002
Std. error	0.0627	0.031
Steady-state prob. at regime 0 (%)	60.63	56.93

under the Student-t error assumption is slightly larger than that under the standard normal assumption, which is consistent with our expectation. Based on these parameter assumptions, the estimated volatility of the Kospi 200 series is obtained; it is shown in Fig. 4.11.

Figure 4.12 gives diagnostic plots: the plot of standardized residuals, ACF plots, and QQ plots under the error assumptions for the MSSV(2) model. The plots on the left-hand side are under the standard normal error assumption, and those on the right-hand side are under the Student-t error assumption. Under the Student-t error assumption, the residuals are more random and the autocorrelation at lag 1 becomess negligible. Moreover, there is significant improvement in the QQ plot. Based on our diagnostic tests, the MSSV(2)

model under a Student-t with 8 degrees of freedom error assumption fits the Kospi 200 series better than the MSSV(2) model with the standard normal error assumption.

To compare the goodness of fit of the MSSV(2) model with a heavy-tail error and the AR(1)–SV model with a heavy-tail error, we display in Fig. 4.13 diagnostic plots of Kospi 200 in the AR(1)–SV model with a Student-t with 8 degrees of freedom assumption. For both cases, the standard residuals are randomly distributed and no autocorrelations exist. However, the QQ plot of the standardized residuals from the MSSV(2) model shows that most of the points are on the straight line; in the QQ plot corresponding to the AR(1)–SV model there are deviations from the straight line at both ends. From the diagnostic plots, we can conclude that the MSSV(2) model with a Student-t error assumption is more approprite for the Kospi 200 than the AR(1)–SV model with a Student-t error assumption. Furthermore, of the volatility models we considered in this thesis, MSSV(2) is the best for the Kospi 200 series.

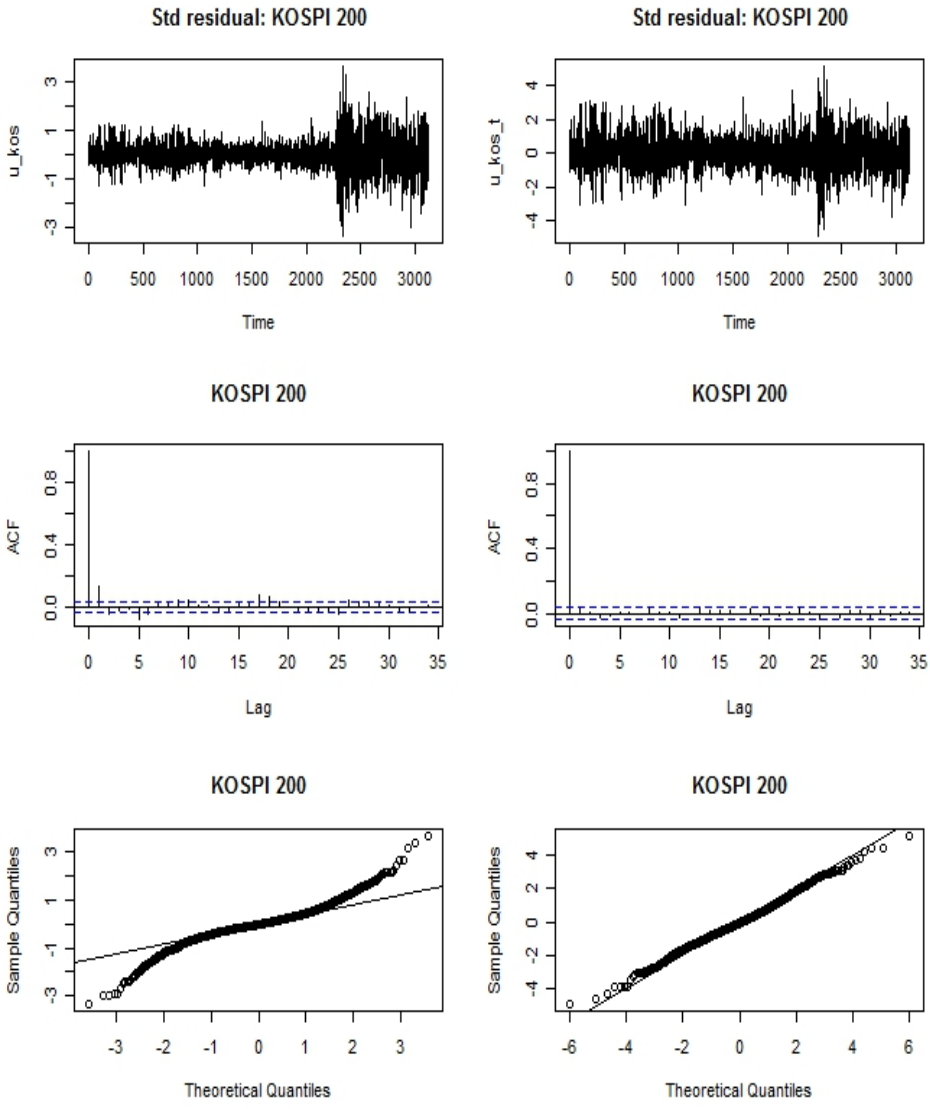

Figure 4.12: Diagnostic plots for KOSPI 200: Standard normal (left) and Student-t (right) distributional assumptions for error term u_t in MSSV(2) model

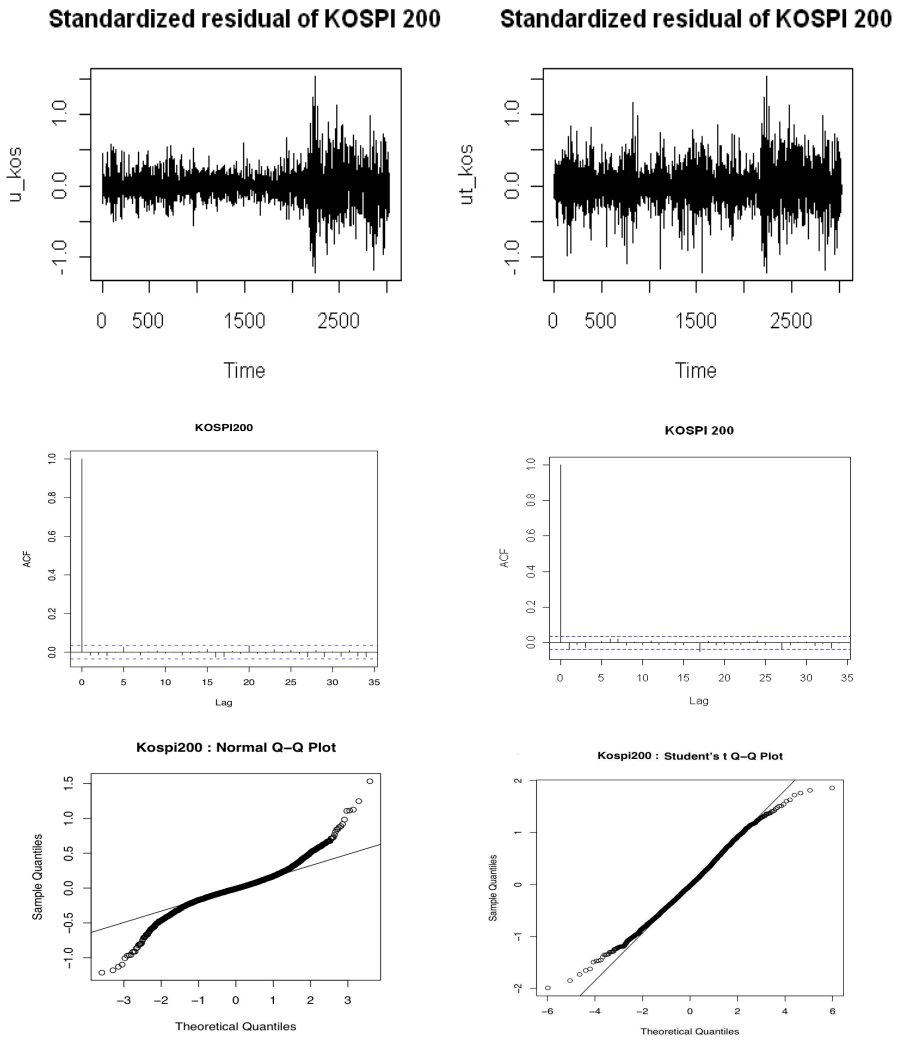

Figure 4.13: Diagnostic plots for KOSPI 200: Standard normal (left) and Student-t (right) distributional assumptions for error term u_t in AR(1)–SV model

Chapter 5

Summary and Future Research

This chapter is organized as follows. We provide a summary of the results in Chapters 1–4 in Section 1. In Section 2, we consider multivariate volatility models as an important topic for future research. In Section 3, we list some additional topics for future research.

5.1 Summary

In Chapter 1, we introduced two types of volatility models: ARCH and SV. In the SV case, we introduced the standard SV model, proposed by Taylor (1986). We also discussed several estimation methods for the SV model including the GMM, EMM, SML, and MCMC methods.

We discussed the statistical properties of the SV model with two error assumptions in Chapter 2. One error assumption is the standard normal distribution and the other is the scaled Student-t distribution. We also described the SML method with the EIS algorithm and the MCMC method for estimating the parameters in the SV model. For the volatility estimation, we considered a particle filter. In the empirical study, we considered the SV model with normal and Student-t errors and studied the performance of the SML method with the EIS technique and the MCMC method with a Gibbs sampler using actual return series.

Our empirical analysis shows that the kurtosis implied by the SV-t model is generally closer to the sample kurtosis of the return series than that for the SV-normal

model. The model diagnostics indicate that the SV-t specification is more appropriate than the SV-normal for most of our data. In particular, the S&P500 and IBM series have extremely large sample kurtoses. This is because of the existence of jumps in the series during the Black Monday period. To model these extreme values more precisely, one can consider a model with a jump component. On the other hand, if the unusual values are excluded, the tail behavior of the series is quite similar to that of the other series.

In the comparison of the performances of SML and MCMC, we found that the estimation results obtained by SML and MCMC are similar and statistically precise in the sense of a small standard error of the parameter estimates, but SML has an edge on MCMC in terms of computing time.

In Chapter 3, we considered the AR(p)–SV model to introduce autocorrelation in the conditional mean process along with the stochastic volatility process. We studied the moment properties and likelihood of the model. We chose the SML method with the EIS technique to estimate the parameters and a particle filter to estimate the volatility. To select the order p of the AR part of the model, we considered AIC and BIC model selection criteria. To demonstrate the estimation of the AR(p)–SV model, we conducted a simulation study with three simulated series. The results showed that the SML method with the EIS algorithm is applicable to the estimation of the AR(1)–SV model.

To examine the robustness of the SML with the EIS for the AR(p)–SV model, we performed a sensitivity analysis with two simulated series where the error u_t follows a Generalized Lambda distribution. We fit the series to the AR(1)-SV models with normal and t errors. Based on the parameter estimation results and model diagnostics, we conclude that the SML method with EIS is robust to certain deviations from normal and t tail behavior.

In the empirical study with four data sets, we implemented the SML method with the EIS technique for the parameter estimation and a particle filter for the volatility estimation. Based on AIC and BIC, the AR(1)–SV model was best for the Kospi 200 and RBC series. This model can capture the persistence and the large kurtosis of the series. However, a normal error assumption for the conditional mean process does not capture the heavy left tail in the return distribution, which is in line with our observation in the SV model. To study the performance of the AR(1)–SV model with a normal and a heavy-tailed error, we performed the parameter estimation and model diagnostics with two error assumptions for the Kospi 200 and RBC series. The results show that the AR(1)–SV model with a scaled Student-t error assumption provides a better fit to the data.

Although the AR(1)–SV model with a scaled Student-t error was the best model for the Kospi 200 series in the empirical study in Chapter 3, we noticed that there was a notable shift in the volatility of the Kospi 200 series in late 1997. To account for this sudden shift in the volatility, in Chapter 4 we considered the regime-switching SV model with K regimes, called the Markov switching SV (MSSV(K)) model. We derived the likelihood function of the MSSV model and used the SML method with the EIS algorithm to estimate the parameters and a particle filter to estimate the volatility. Our simulation study showed that our method was reasonable for the estimation of the MSSV(2) model.

We conducted an empirical study with four series including the Kospi 200 series used in the empirical study in Chapter 3. The MSSV(2) model with a normal error is able to capture the time-varying volatility of the S&P 500, Russel 2000, and Nikkei 225 series, but not that of the Kospi 200 series. We therefore fitted the MSSV(2) model with a scaled Student-t error to the Kospi 200 series, and the goodness of fit substantially improved.

In the comparison of the AR–SV and MSSV models for the Kospi 200 series, we found that both models capture the movement of the Kospi 200 returns well. However, the MSSV(2) model has a slight edge over the AR(1)–SV model in capturing the volatility movement. The AR(1)–SV model tends to underestimate the volatility during the crisis period in 1997, while the estimates from MSSV(2) are quite close to the actual (historical) volatility. Moreover, the MSSV(2) model with a scaled Student-t error assumption fits the data better than the same model with a normal error assumption.

5.2 Vector Autoregressive Models with Stochastic Volatility Errors

The globalization of the economy and the use of the Internet have accelerated the integration of world financial markets during the last decade. Since a price movement of any financial asset in one market can affect another market easily and instantly, the world financial markets are more dependent on each other than ever before. One should therefore consider market phenomena jointly to better understand the dynamic structure of global finance. In a response to the globalization, it is essential to be aware of the interrelationship between financial markets. Accordingly, for investors or financial institutions that hold multiple assets, the dynamic relationship between returns on assets

is a key issue in decision making.

To incorporate the interrelationship between global securities, it would be natural to generalize the univariate volatility models described in Chapters 1 to 4 to the multivariate case. Multivariate volatility models have many important financial applications. They play a crucial role in portfolio selection and asset allocation, and can be used to compute tail risk measures such as the conditional tail expectation (CTE) or the value at risk (VaR) of a financial position consisting of multiple assets.

Many multivariate models have been developed to consider cross-correlations between multiple assets in financial economics. One of the most popular and natural applications is a straightforward extension of the univariate conditional variance models such as the ARCH or SV models. Among the multivariate (G)ARCH models, Bollerslev, Engle, and Wooldridge (1988) proposed the VEC-GARCH model, which is a straightforward generalization of the univariate GARCH model. In this model, the conditional variance and covariance are functions of their lagged conditional variances and covariances, as well as lagged squared returns and cross-products of returns. This model is difficult to estimate because of the high dimension of the parameter space and complicated constraints. To reduce the number of parameters to be estimated in the VEC-GARCH model, Bollerslev (1990) considered a time-invariant multivariate correlation model with the GARCH(p,q) model, called the constant conditional correlation (CCC) GARCH model. An extension to the CCC-GARCH (ECCC-GARCH) model was introduced by Jeantheau (1998). This allows the past squared returns and variances of all series to enter the individual conditional variance equations. Tse and Tsui (1999) proposed the multivariate GARCH model with time-varying correlations (VC-GARCH). The time-varying correlations are modeled as functions of the conditional correlations of the previous period and a set of estimated correlations.

The literature on multivariate stochastic volatility models is rather more limited. Multivariate SV (MSV) models are discussed by Harvey *et al.* (1994), Jacquier, Polson, and Rossi (1995), Kim *et al.* (1998), Pitt and Shephard (1999), Aguilar and West (2000), and Liesenfeld and Richard (2003). The models in these papers are equivalent to a stack of univariate SV models and MCMC is used to estimate the models. Chib, Nardari, and Shephard (2006) proposed a more general version of the MSV model (Harvey *et al.* (1994)) with heavy-tailed errors and jumps and studied the forecasting performance of the MSV model using a particle filter.

In the models stated above, no correlation is assumed for underlying assets, but

time-invariant correlations are assumed for the volatilities of the assets. A natural extension to incorporate cross-correlations between multiple assets is the vector autoregressive process (VAR) for the conditional mean process. Canova (1993), Sims (1993), Stock and Watson (1996), and Cogley and Sargent (2001) consider VARs with drift coefficients and time-invariant variances. Cogley (2003), Cogley and Sargent (2003), Boivin (2001), and Ciccarelli and Rebucci (2003) used VARs with drift coefficients and time-varying variances. Primiceri (2005) considered a VAR model with a time-varying error covariance matrix. Some restrictions were imposed on the structure of the covariance matrix so that it became a special case of the MSV model and then MCMC was used for estimating parameters in the model.

In this section, we review a vector autoregressive (VAR) process and multivariate stochastic volatility (MSV) models and consider the VAR model with the MSV innovation to account for the interrelationship between underlying assets and between the volatilities. To estimate the parameters in the VAR-MSV model, the SML method can be considered, but we will leave this as future research.

5.2.1 Vector Autoregressive Model

Suppose that there are m financial time series $\mathbf{y_t}$. Then, the VAR model of order p is defined as

$$\mathbf{y_t} = \Phi_0 + \sum_{i=1}^{p} \mathbf{\Phi_i y_{t-i}} + \epsilon_t, \tag{5.1}$$

where $\mathbf{y_t} = (y_{1,t}, ..., y_{m,t})'$, $\Phi_0 = (\phi_{1,0}, ..., \phi_{m,0})'$, $\mathbf{\Phi_i}$ is an $m \times m$ matrix, and $\epsilon_t = (\epsilon_{1,t}, ..., \epsilon_{m,t})'$ is a multivariate Gaussian white noise with variance-covariance matrix Σ_ϵ. The coefficient vector Φ_i for all i measures the dynamic dependence of $\mathbf{y_t}$. To obtain the necessary and sufficient condition of (weak) stationarity for this model, we rewrite Eq. (5.1) as

$$\mathbf{\Phi(B) y_t} = \Phi_0 + \epsilon_t,$$

where $\mathbf{\Phi(B)} = 1 - \mathbf{\Phi_1 B} - \cdots - \mathbf{\Phi_p B^p}$ and $\mathbf{B}$ is a backshift operator. $\mathbf{y_t}$ is stationary if and only if the determinant polynomial $|\mathbf{\Phi(B)}| = 0$ has a root outside the unit circle.

Consider the bivariate case [*i.e.*, $m = 2, \mathbf{y_t} = (y_{1,t}, y_{2,t})'$, and $\epsilon_t = (\epsilon_{1,t}, \epsilon_{2,t})'$]. The

bivariate VAR(1) model consists of the following two equations:

$$y_{1,t} = \phi_{1,0} + \phi_{11} y_{1,t-1} + \phi_{12} y_{2,t-1} + \epsilon_{1,t}, \tag{5.2}$$
$$y_{2,t} = \phi_{2,0} + \phi_{21} y_{1,t-1} + \phi_{22} y_{2,t-1} + \epsilon_{2,t} \tag{5.3}$$

for $t = 1, ..., T$. ϕ_{ij} is the $(i,j)^{th}$ component of $\mathbf{\Phi_1}$ and $\phi_{i,0}$ is the i^{th} component of Φ_0 for $i, j = 1, 2$. From Eq. (5.2), ϕ_{12} represents the linear dependence of $y_{1,t}$ on $y_{2,t-1}$ in the presence of $y_{1,t-1}$. That is, ϕ_{12} is the conditional effect of $y_{2,t-1}$ on $y_{1,t}$, given $y_{1,t-1}$. Thus, if $\phi_{12} = 0$, then $y_{1,t}$ depends only on its own past. Similarly, if $\phi_{21} = 0$, then Eq. (5.3) shows that $y_{2,t}$ depends only on $y_{2,t-1}$.

If we consider Eqs. (5.2) and (5.3) jointly, the interpretations of ϕ_{12} and ϕ_{21} are as follows. If $\phi_{12} = 0$ and $\phi_{21} \neq 0$, then everything but $y_{1,t}$ depends on $y_{2,t}$, but not vice versa. If $\phi_{12} = \phi_{21} = 0$, then $y_{1,t}$ and $y_{2,t}$ are not related. If $\phi_{12} \neq 0$ and $\phi_{21} \neq 0$, then there is a feedback relationship between the two series. The concurrent relationship between $y_{1,t}$ and $y_{2,t}$ is shown by the off-diagonal element σ_{12} of the variance-covariance matrix Σ_ϵ. If $\sigma_{12} = 0$, then there is no concurrent linear relationship between the series.

Let $\vec{\mu}$ denote the mean vector of the time series $\mathbf{y_t}$ and $\tilde{\mathbf{y}}_t = \mathbf{y_t} - \vec{\mu}$ denote the mean-corrected series. Then, the bivariate VAR(1) model in (5.2) and (5.3) has the following expression

$$\tilde{\mathbf{y}}_t = \mathbf{\Phi_1} \tilde{\mathbf{y}}_{1,t} + \epsilon_t. \tag{5.4}$$

By repeated substitutions, Eq. (5.4) can be rewritten as

$$\tilde{\mathbf{y}}_t = \epsilon_t + \mathbf{\Phi_1} \epsilon_t + \mathbf{\Phi_1^2} \epsilon_t + \mathbf{\Phi_1^3} \epsilon_t + \cdots . \tag{5.5}$$

From this expression, several characteristics can be found. First, the necessary and sufficient condition for weak stationarity of $\mathbf{y_t}$ is that all eigenvalues of $\mathbf{\Phi_1}$ are less than 1 in modulus provided that the variance-covariance matrix of ϵ_t exists. Second, since ϵ_t is serially uncorrelated, $Cov(\epsilon_t, \mathbf{y_{t-1}}) = \mathbf{0}$, which leads to $Cov(\epsilon_t, \mathbf{y_{t-j}}) = \mathbf{0}$ for $j > 0$. Third, multiplying (5.5) by ϵ_t' and taking the expectation, it can be seen that $Cov(\epsilon_t, \mathbf{y_t}) = \Sigma_\epsilon$ because ϵ_t is uncorrelated.

5.2.2 Multivariate Stochastic Volatility Model

The univariate stochastic volatility (SV) model can be generalized to the multivariate SV (MSV) model in various ways (*see* Asai, McAleer, and Yu (2006) for a survey). In

this section, we introduce an MSV model proposed by Harvey *et al.* (1994) that is straightforward and analytically tractable.

For a multivariate time series $\mathbf{y_t}$, the multivariate stochastic volatility (MSV) model (Harvey *et al.* (1994)) is defined as

$$\begin{aligned} \mathbf{y_t} &= H_t^{1/2}\mathbf{u}_t, & (5.6)\\ \lambda_t &= \alpha + \beta\lambda_{t-1} + \mathbf{v}_t & (5.7) \end{aligned}$$

for $t = 1, ..., T$. $\mathbf{y_t} = (y_{1,t}, ..., y_{m,t})'$ contains the observations at time t, where m is the number of series, $H_t^{1/2} = diag[\exp\{\lambda_{1,t}/2\}, ..., \exp\{\lambda_{m,t}/2\}]$ is the $m \times m$ diagonal matrix, and $\lambda_t = (\lambda_{1,t}, ..., \lambda_{m,t})'$ is the latent log-variance process. α is an $m \times 1$ vector and β is an $m \times m$ matrix. $\mathbf{u_t} = (u_{1,t}, ..., u_{m,t})'$ is a multivariate normal vector with zero mean and correlation matrix $\mathbf{P_u}$, in which the elements on the leading diagonal are unity and the off-diagonal elements are denoted $\rho_{ij}^{(u)}$. Moreover, $\mathbf{v_t}$ is multivariate normal with mean vector zero and variance-covariance matrix Σ_v. Note that the covariance matrices $\mathbf{P_u}$ and Σ_v are positive-definite. If all the eigenvalues of β are less than 1 in modulus, then the latent process λ_t in Eq. (5.7) is a stationary VAR(1) process provided that Σ_v exists.

For the simplest case, consider the bivariate MSV model, defined by

$$\begin{aligned} y_{1,t} &= \exp\{\lambda_{1,t}/2\}u_{1,t}, \\ y_{2,t} &= \exp\{\lambda_{2,t}/2\}u_{2,t}, & (5.8)\\ \lambda_{1,t} = \alpha_1 &+ \beta_{11}\lambda_{1,t-1} + \beta_{12}\lambda_{2,t-1} + v_{1,t}, \\ \lambda_{2,t} = \alpha_2 &+ \beta_{21}\lambda_{1,t-1} + \beta_{22}\lambda_{2,t-1} + v_{2,t}, & (5.9) \end{aligned}$$

$$\begin{pmatrix} \mathbf{u_t} \\ \mathbf{v_t} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \mathbf{P_u} & 0 \\ 0 & \mathbf{\Sigma_v} \end{pmatrix} \right],$$

where the error processes $\mathbf{u_t} = (u_{1,t}, u_{2,t})'$ and $\mathbf{v_t} = (v_{1,t}, v_{2,t})'$ are bivariate normal with the correlation matrix $\mathbf{P_u}$ and the variance-covariance matrix $\mathbf{\Sigma_v}$ given by

$$\mathbf{P_u} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{\Sigma_v} = \begin{pmatrix} \sigma_{11}^{(v)} & \sigma_{12}^{(v)} \\ \sigma_{21}^{(v)} & \sigma_{22}^{(v)} \end{pmatrix},$$

where ρ is a known, constant parameter. The latent log-variance process λ_t follows a VAR(1) model. The coefficient β_{12} describes the relationship between the log variance of series-two $\lambda_{2,t}$ at time $t-1$ and the log variance of series-one $\lambda_{1,t}$ at time t in the presence

of $\lambda_{1,t-1}$. Thus, we can make interpretations for β_{ij} for $i, j = 1, 2$ similar to those in Section 5.2.1.

Although the MSV model in Eqs. (5.6) and (5.7) is natural and reasonable, the estimation cannot easily be accomplished. The main difficulty is that there are too many parameters to be estimated. The number of parameters for the bivariate MSV process is 10, and in general, the number of parameters is $m + 2m^2$. To reduce the number of parameters without losing the nice characteristics of the MSV models, Harvey *et al.* (1994) set β to be an $m \times 1$ vector. That is, each log variance λ_t is specified as a univariate SV model. The form of this MSV model is

$$y_{i,t} = \exp\{\lambda_{i,t}/2\}u_{i,t}, \tag{5.10}$$
$$\lambda_{i,t} = \alpha_i + \beta_i\lambda_{i,t-1} + v_{i,t}, \tag{5.11}$$

where all the assumptions for $\mathbf{u_t}$ and $\mathbf{v_t}$ are as in Eqs. (5.6) and (5.7). Equation (5.11) is a special case of (5.7) in which the off-diagonal elements of β are zeros. In this case, the number of parameters to be estimated is $2m + m^2$. If we assume that the off-diagonal elements of Σ_v are all equal to zero, then the elements of the vector λ_t are independent. Otherwise, the elements of λ_t are not independent.

For more parameter reduction, a multivariate random walk can be considered (Harvey *et al.* (1994)) for λ_t:

$$\lambda_{i,t} = \lambda_{i,t-1} + v_{i,t}. \tag{5.12}$$

The number of parameters in this model is m^2. This MSV model specification is often used in practice, whereas the models in (5.6), (5.7), (5.10), and (5.11) are more difficult to handle.

5.2.3 VAR–MSV Model

In this section, we will introduce a vector autoregressive (VAR) model with the multivariate stochastic volatility (MSV) error. This is basically a vector autoregression with an error for which the covariance matrix has the diagonal elements evolving over time according to the stochastic volatility model specification.

There are at least two advantages of this model. First, by the addition of the

lagged variables of the return series, the concurrent correlations of the series can be explained. Second, by the inclusion of the stochastic volatility evolution for the error process, the variations of several return series can be more explicitly and simultaneously captured. Under suitable assumptions, the VAR(p)–MSV model can be defined as

$$\begin{aligned} \mathbf{y_t} &= \Phi_0 + \sum_{i=1}^{p} \mathbf{\Phi_i y_{t-i}} + \epsilon_t, \\ \epsilon_t &= H_t^{1/2} \mathbf{u_t}, \\ \lambda_t &= \alpha + \beta \lambda_{t-1} + \mathbf{v_t}, \\ \begin{pmatrix} \mathbf{u_t} \\ \mathbf{v_t} \end{pmatrix} &\sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \mathbf{P_u} & 0 \\ 0 & \mathbf{\Sigma_v} \end{pmatrix} \right], \end{aligned} \tag{5.13}$$

where $H_t^{1/2} = diag[\exp\{\lambda_t/2\}]$ for $t = 1, ..., T$. $\mathbf{P_u}$ is a positive definite correlation matrix of $\mathbf{u_t}$ with $\rho_{ii}^{(u)} = 1$ and $|\rho_{ii}^{(u)}| < 1$ for $i \neq j$, and Σ_v is a positive definite covariance matrix of $\mathbf{v_t}$ with the elements $\sigma_{ij}^{(v)}$ for $i, j = 1, ..., m$. The coefficient α is an $m \times 1$ vector and β is an $m \times m$ matrix.

Under this model specification, the number of parameters to be estimated is $m + 3m^2$. To reduce this number without losing the efficiency of the model, we can consider the following VAR(p)–MSV model:

$$\begin{aligned} \mathbf{y_t} &= \Phi_0 + \sum_{i=1}^{p} \mathbf{\Phi_i y_{t-i}} + \epsilon_t, \\ \epsilon_t &= H_t^{1/2} \mathbf{u_t}, \\ \lambda_t &= \lambda_{t-1} + \mathbf{v_t}, \end{aligned} \tag{5.14}$$

where $H_t^{1/2} = diag[\exp\{\lambda_t/2\}]$ for $t = 1, ..., T$. All of the assumptions for (5.13) hold. Note that all of the elements of α are now set to zero and β is an $m \times m$ identity matrix. Thus, λ_t follows a simple random walk in a vector form, in which the number of parameters is only $2m^2$. Compared with the models in (5.13), the number of parameters is relatively small in this model, but the estimation of the covariance matrix is still elaborate.

5.3 Topics for Future Research

In line with the results in Chapters 1–4, possible future research directions are as follows:

1. For the VAR–MSV model,

   (a) Study the moment properties as for the univariate case. Moments and autocorrelations of one series are influenced by the structure of the other series. This can be exploited to study the volatility in one series using that in the other series.

   (b) Obtain the exact likelihood function.

   (c) Estimate the VAR–MSV model via the SML method, and conduct simulations to study how effectively the SML method performs. As an MC variance reduction method, we can consider the EIS technique, as in the univariate case. Empirical analysis can be performed with actual return series.

   (d) Estimate the VAR–MSV model via the MCMC method. Compare the performance of the SML method with the EIS algorithm to that of the MCMC method, which is more popular in the literature.

   (e) Consider the VAR–MSV model with a heavy-tailed error assumption in the MSV process. Study the performance of capturing the fat left tail of the return series with a normal and a heavy-tail distribution such as Student-t in the MSV model.

2. For univariate SV models,

   (a) Study asymptotic properties of the SML estimator with the EIS for the SV, AR–SV, and MSSV models.

   (b) Study the properties of AIC and BIC for AR–SV model.

   (c) Consider other modifications for the SV models such as

      i. Jumps in the conditional mean process with the stochastic volatility error.
      ii. Jumps in the MSSV model.
      iii. Switching in the AR–SV model.

   (d) Apply either SML or MCMC methods to estimate the parameters in the models and conduct empirical studies.

   (e) Consider the generalized error distribution as another heavy-tail error assumption for the SV, AR–SV, and MSSV models.

References

[1] Aguilar, O. and West, M. (2000), Bayesian dynamic factor models and variance matrix discounting for portfolio allocation, *Journal of Business & Economic Statistics*, **18**, 338-357.

[2] Albert, J. H. and Chib, S. (1993), Bayes inference via Gibbs sampling of autoregressive time series subject to Markov and variance shifts, *Journal of Business & Economic Statistics*, **11**, 1-15.

[3] Anderson, T. G. (1994), Stochastic autoregressive volatility: A framework for volatility modelling, *Mathematical Finance*, **4**, 75-102.

[4] Anderson, T. G. and Bollerslev, T. (1998), Answering the skeptics: Yes, standard volatility models do provide accurate forecasts, *International Economic Review*, **39**, 885-905.

[5] Anderson, T. G., Chung, H., and Sorensen, B. E. (1999), Efficient method of moments estimation of a stochastic volatility model: A Monte Carlo study, *Journal of Econometrics*, **91**, 61-87.

[6] Anderson, T. G. and Sorensen, B. E. (1996), GMM estimation of a stochastic volatility model: A Monte Carlo study, *Journal of Business & Economic Statistics*, **14**, 328-352.

[7] Ari, Y. and Ünal, G. (2011), Continuous modeling of foreign exchange rate of USD versus TRY, *International Journal of Advaced Economics and Business Management*, **1**, 102-106.

[8] Asai, M., McAleer, M., and Yu, J. (2006), Multivariate stochastic volatility: A review, *Econometric Reviews*, **25**, 145-175.

[9] Black, F. (1976), Studies in stock price volatility changes, *Proceedings of the 1976 Business Meeting of the Business and Economic Statistics Section*, American Statistical Association, 177-181.

[10] Boivin, J. (2001), The Fed's conduct of monetary policy: Has it changed and does it matter?, *Columbia Business School*, mimeo.

[11] Bollerslev, T. (1986), Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, **31**, 307-327.

[12] Bollerslev, T. (1990), Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model, *Review of Econometrics and Statistics*, **72**, 498-505.

[13] Bollerslev, T. and Engle, R. F. (1993), Common persistence in conditional variances, *Econometrica*, **61**, 167-186.

[14] Bollerslev, T., Engle, R. F., and Wooldridge, J. M. (1988), A capital asset pricing model with time-varying covariances, *The Journal of Political Economy*, **96**, 116-131.

[15] Bos, C. S. and Shephard, N. (2006), Inference for adaptive time series models: Stochastic volatility and conditionally gaussian state space form, *Econometric Reviews*, **25**, 219-244.

[16] Box, G. E. and Jenkins, G. (1976), Time Series Analysis: Forecasting and Control, Holden-Day.

[17] Box, G. E. and Tiao, P. E. (1973), Bayesian Inference in Statistical Analysis, Addison-Wesley, New York.

[18] Cai, J. (1994), A Markov model of switching-regime ARCH, *Journal of Business & Economic Statistics*, **12**, 309-316.

[19] Campell, J. Y. and Hentschel, L. (1992), No news is good news: An asymmetric model of changing volatility in stock returns, *Journal of Financial Economics*, **31**, 281-318.

[20] Canova, F. (1993), Modelling and forecasting exchange rates with a Bayesian time-varying coefficient model, *Journal of Economic Dynamics and Control*, **17**, 233-261.

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

[22] Chib, S., Nardari, F., and Shephard, N. (1998), Markov chain Monte Carlo methods for stochastic volatility models, *Journal of Econometrics*, **108**, 281-316.

[23] Chib, S., Nardari, F., and Shephard, N. (2006), Analysis of high dimensional multivariate stochastic volatility models, *Journal of Econometrics*, **134**, 341-371.

[24] Ciccarelli, M. and Rebucci, A. (2003), Measuring contagion with a Bayesian time-varying coefficient model, *International Monetary Fund*, mimeo.

[25] Cogley, T. (2003), How fast can the new economy grow? A Bayesian analysis of the evolution of trend growth, *University of California, Davis*, mimeo.

[26] Cogley, T. and Sargent, T. J. (2001), Evolving post-world war II U.S. inflation dynamics, *NBER Macroeconomics Annual, MIT Press, Cambridge, Massachusetts*, 331-373.

[27] Cogley, T. and Sargent, T. J. (2003), Drifts and volatilities: Monetary policies and outcomes in the post WWII U.S., *New York University*, mimeo.

[28] Danielsson, J. (1994), Stochastic volatility in asset prices estimation with simulated maximum likelihood, *Journal of Econometrics*, **64**, 375-400.

[29] Danielsson, J. and Richard, J. F. (1993), Accelerated Gaussian importance sampler with application to dynamic latent variable models, *Journal of Applied Econometrics*, **8**, S153-S173.

[30] Demos, A. (2002). Moments and dynamic structure of a time-varying-parameter stochastic volatility in mean model, *The Econometrics Journal*, **5**, 347-357.

[31] Diebold, F. X. (1988), Empirical Modeling of Exchange Rate Dynamics, Springer-Verlag, New York.

[32] Ding, Z., Granger, C. W. J., and Engle, R. F. (1993), A long memory property of stock market returns and a new model, *Journal of Empirical Finance*, **1**, 83-106.

[33] Dueker, M. J. (1997), Markov switching in GARCH processes and mean-reverting stock market volatility, *Journal of Business & Economic Statistics*, **15**, 26-34.

[34] Durbin, J. and Koopman, S. J. (1997), Monte Carlo maximum likelihood estimation of non-Gaussian state space model, *Biometrika*, **84**, 669-684.

[35] Durbin, J. and Koopman, S. J. (2000), A simple and efficient simulation smoother for state space time series analysis, *Biometrika*, **89**, 603-616.

[36] Engle, R. F. (1982), Autoregressive conditional heteroskedasticity with estimates of the variance of the United Kingdom inflation, *Econometrica*, **50**, 987-1007.

[37] Engle, R. F. and Bollerslev, T. (1986), Modelling the persistence of conditional variances, *Econometric Reviews*, **5**, 1-50; 81-87.

[38] Engle, R. F., Lilien, D. M., and Robins, R. P. (1987), Estimating time-varying risk premia in the term structure: The ARCH-M model, *Econometrica*, **55**, 391-407.

[39] Gallant, A. R., Hsieh, D., and Tauchen, G. (1997), Estimation of stochastic volatility models with diagnostics, *Journal of Econometrics*, **8**, S153-S173.

[40] Gallant, A. R. and Nychka, D. W. (1987), Semi nonparametric maximum likelihood estimation, *Econometrica*, **55**, 363-390.

[41] Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993), On the relation between the expected value and the volatility of the nominal excess return on stocks, *The Journal of Finance*, **48**, 1779-1801.

[42] Goldman, E. and Tsurumi, H. (2005), Bayesian Analysis of a Doubly Truncated ARMA-GARCH Model, *Studies in Nonlinear Dynamics & Econometrics*, **9**(2), Article 5.

[43] Gray, S. (1996), Modeling the conditional distribution of interest rates as a regime-switching process, *Journal of Financial Economics*, **42**, 27-62.

[44] Hamilton, J. D. (1988), Rational-experience econometric analysis of changes in regime: An investigation of the term structure of interest rates, *Journal of Economic Dynamics and Control*, **12**, 385-423.

[45] Hamilton, J. D. (1989), A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica*, **57**, 357-384.

[46] Hamilton, J. D. and Susmel, R. (1994), Autoregressive conditional heteroskedasticity and changes in regime, *Journal of Econometrics*, **64**, 307-333.

[47] Hansen, L. P. (1982), Large sample properties of generalized method of moments estimators, *Econometrica*, **8**, S85-S118.

[48] Harvey, A., Ruiz, E., and Shephard, N. (1994), Multivariate stochastic variance models, *Review of Economic Studies*, **61**, 247-264.

[49] He, C. and Teräsvirta, T. (1999), Fourth moment structure of the GARCH(p, q) model, *Econometric Theory*, **15**, 824-846.

[50] Hong, E. P. (1991), The autocorrelation structure for the GARCH-M process, *Economics Letters*, **37**, 129-132.

[51] Hull, J. and White, A. (1987), The pricing of options on assets with stochastic volatilities, *Journal of Finance*, **42**, 281-300.

[52] Jacquier, E., Polson, N. G., and Rossi, P. E. (1994), Bayesian analysis of stochastic volatility models, *Journal of Business & Economic Statistics*, **12**, 371-389.

[53] Jacquier, E., Polson, N. G., and Rossi, P. E. (1995), Models and priors for multivariate stochastic volatility, *CIRANO Working Papers*.

[54] Jeantheau, T. (1998), Strong consistency of estimators for multivariate ARCH models, *Econometric Theory*, **14**, 70-86.

[55] Johnson, N. L. and Kotz, S. (1970), Distributions in Statistics: Continuous Univariate Distributions, *Vol. 1*, John Wiley & Sons, New York.

[56] Kalimipalli, M. and Susmel, R. (2004), Regime-switching stochastic volatility and short-term interest rates, *Journal of Empirical Finance*, **11**, 309-329.

[57] Karanasos, M. (1999), The second moment and the autocovariance function of the squared errors of the GARCH model, *Journal of Econometrics*, **90**, 63-76.

[58] Karanasos, M. (2001), Prediction in ARMA models with GARCH-in-mean effects, *Journal of Time Series Analysis*, **22**, 555-578.

[59] Karanasos, M. and Kim, J. (2003), Moments of the ARMA-EGARCH model, *Econometrics Journal*, **6**, 146-166.

[60] Kim, S., Shephard, N., and Chib, S. (1998), Stochastic volatility: Likelihood inference and comparison with ARCH models, *Reviews of Economic Studies*, **65**, 361-393.

[61] Kitamura, Y. (2011),The impact of order flow on the foreign exchange market: A copula apporach, *Asia-Pacific Financial Markets*, **18**, 1-31.

[62] Klaassen, F. (2002), Improving GARCH volatility forecasts, *Empirical Economics*, **27**, 363-394.

[63] Koopman, S. J. and Bos, C. S. (2004), State space models with a common stochastic variance, *Journal of Business & Economic Statistics*, **22**, 346-357.

[64] Koopman, S. J., Lucas, A., and Scharth, M. (2011), Numerically Accelerated Importance Sampling for Nonlinear Non-Gaussian State Space Models, *Tinbergen Institute Discussion Paper*.

[65] Koopman, S. J., Shephard, N., and Creal, D. (2009), Testing the assumptions behind importance sampling, *Journal of Econometrics*, **149**, 2-11.

[66] Lamoureux, C. G. and Lastrapes, W. D. (1990), Persistence in variance, structural change, and the GARCH model, *Journal of Business & Economic Statistics*, **8**, 225-234.

[67] Larsson, O. (2005), Forecasting volatility and Value at Rik: Stochastic volatility vs GARCH, *Working paper*.

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

[69] Liesenfeld, R. and Richard, J. F. (2003), Univariate and multivariate stochastic volatility models: Estimation and diagnostics, *Journal of Empirical Finance*, **10**, 505-531.

[70] Liesenfeld, R. and Richard, J. F. (2008), Improving MCMC, using efficient importance sampling, *Computational Statistics and Data Analysis*, **53**, 272-288.

[71] Marucucci, J. (2005), Forecasting stock market volatility with regime-switching GARCH models, *Studies in Nonlinear Dynamics & Econometrics*, **9**, Article 6.

[72] Melino, A. and Turnbull, S. M. (1990), Pricing foreign currency options with stochastic volatility, *Journal of Econometrics*, **45**, 239-265.

[73] Murphy, A. and Izzeldin, M. (2005), Bootstrapping Long Memory Tests: Some Monte Carlo Results, *Unpublished paper*.

[74] Nelson, D. B. (1991), Conditional heteroskedasticity in asset pricing: A new approach, *Econometrica*, **61**, 247-264.

[75] Ozturk, S. S. (2009), Applications of Efficient Importance Sampling to Stochastic Volatility Models, *Ph.D. Thesis*, University of Pittsburgh.

[76] Palma, W. and M. Zevallos (2004), Analysis of the correlation structure of square time series, *Journal of Time Series Analysis*, **25**, 529-550.

[77] Pellegrini, S. (2009), Predicción en models de componentes inobservables condicionalmente heteroscedásticos, *Ph.D. Thesis*, Universidad Carlos III De Madrid.

[78] Pederzoli, C. (2006), Stochastic Volatility and GARCH: A comparison based on UK stock data, *The European Journal of Finance*, **12**, 41-59.

[79] Pitt, M. K. and Shephard, N. (1997), Filtering via simulation: An auxiliary variable approach, *Working paper*, Nuffield College, Oxford.

[80] Pitt, M. K. and Shephard, N. (1999), Filtering via simulation: Auxiliary particle filters, *Journal of the American Statistical Association*, **94**, 590-599.

[81] Pati, P. C. and Rajib, P. (2010), Volatility persistence and trading volume in an emerging futures market: Evidence from NSE Nifty stock index futures, *Journal of Risk Finance*, **11**, 296-309.

[82] Primiceri, G. E. (2005), Time varying structural vector autoregressions and monetary policy, *The Review of Economic Studies*, **72**, 821-852.

[83] Ramberg, J. S., Dudewicz, E. J., and Tadikamalla, P. R. (1979), A probability distribution and its uses in fitting data, *Technometrics*, **21**, 201-214.

[84] Ripley, B. D. (1987), Stochastic Simulation, John Wiley, New York.

[85] Ruiz, E. (1994), Quasi-maximum likelihood estimation of stochastic volatility models, *Journal of Econometrics*, **63**, 271-301.

[86] Sandmann, G. and Koopman, S. J. (1998), Estimation of stochastic volatility models via Monte Carlo maximum likelihood, *Journal of Econometrics*, **87**, 271-301.

[87] Schwert, G. W. (1989), Why does stock market volatility change over time?, *Journal of Finance*, **44**, 1115-1153.

[88] Sentana, E. (1995), Quadratic ARCH models, *Review of Economic Studies*, **62**, 639-661.

[89] Shephard, N. and Pitt, M. K. (1997), Likelihood analysis of non-Gaussian measurement time series, *Biometrika*, **84**, 653-667.

[90] Shibata, M. and Watanabe, T. (2005), Bayesian analysis of a Markov switching stochastic volatility model, *Journal of Japan Statistics Society*, **35**, 205-219.

[91] Sims, C. A. (1993), A 9 variable probabilistic macroeconomic forecasting model, in Stock, J. H. and Watson, M. W. (Eds.), Business Cycles, Indicators and Forecasting, NBER Studies in Business Cycles, **28**, 179-214.

[92] So, M. K. P., Lam, K., and Li, W. K. (1998), A stochastic volatility model with Markov switching, *Journal of Business & Economic Statistics*, **16**, 244-253.

[93] Stock, J. M. and Watson, M. W. (1996), Evidence on structural instability in macroeconomic time series relations, *Journal of Business & Economic Statistics*, **14**, 11-30.

[94] Tarami, B. and Pourahmadi, M. (2003), Multi-variate t autoregressions: Innovations, prediction variances and exact likelihood equations, *Journal of Time Series Analysis*, **24**, 739-754.

[95] Taylor, S. J. (1982), Financial returns modeled by the product of two stochastic processes: A study of daily sugar prices, in Anderson, O. D. (Ed.), Time Series Analysis: Theory and Practice 1, North-Holland, Amsterdam, 203-226.

[96] Taylor, S.J. (1986), Modeling Financial Time Series, John Wiley, Chichester.

[97] Taylor, S. J. (1994), Modeling stochastic volatility: A review and comparative study, *Mathematical Finance*, **4**, 183-204.

[98] Teräsvirta, T. (1996), Two stylized facts and the GARCH(1,1) model, *Working paper*, **96**, *Stockholm School of Economics*.

[99] Tsay, R. S. (2002), Analysis of Financial Time Series: Financial Econometrics, John Wiley, New York.

[100] Tse, Y. K. and Tsui, K. C. (1999), A note on diagnosing multivariate conditional heteroscedasticity models, *Journal of Time Series Analysis*, **20**, 679-691.

[101] White, H. (1994), Estimation, Inference and Specification, Cambridge University Press, New York.

[102] Yoo, B. H. (2010), Estimating the Term Premium by a Markov Switching Model with ARMA-GARCH Errors, *Studies in Nonlinear Dynamics & Econometrics*, **14**(2), Article 4.

[103] Zakoian, J. M. (1994), Threshold heteroscedastic models, *Journal of Economic Dynamics and Control*, **18**, 931-955.