

Bayesian Inference for Stochastic Volatility Models

by

Zhongxian Men

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, 2012

© Zhongxian Men 2012

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

Stochastic volatility (SV) models provide a natural framework for a representation of time series for financial asset returns. As a result, they have become increasingly popular in the finance literature, although they have also been applied in other fields such as signal processing, telecommunications, engineering, biology, and other areas.

In working with the SV models, an important issue arises as how to estimate their parameters efficiently and to assess how well they fit real data. In the literature, commonly used estimation methods for the SV models include general methods of moments, simulated maximum likelihood methods, quasi Maximum likelihood method, and Markov Chain Monte Carlo (MCMC) methods. Among these approaches, MCMC methods are most flexible in dealing with complicated structure of the models. However, due to the difficulty in the selection of the proposal distribution for Metropolis-Hastings methods, in general they are not easy to implement and in some cases we may also encounter convergence problems in the implementation stage. In the light of these concerns, we propose in this thesis new estimation methods for univariate and multivariate SV models. In the simulation of latent states of the heavy-tailed SV models, we recommend the slice sampler algorithm as the main tool to sample the proposal distribution when the Metropolis-Hastings method is applied. For the SV models without heavy tails, a simple Metropolis-Hastings method is developed for simulating the latent states. Since the slice sampler can adapt to the analytical structure of the underlying density, it is more efficient. A sample point can be obtained from the target distribution with a few iterations of the sampler, whereas in the original Metropolis-Hastings method many sampled values often need to be discarded.

In the analysis of multivariate time series, multivariate SV models with more general specifications have been proposed to capture the correlations between the innovations of the asset returns and those of the latent volatility processes. Due to some

restrictions on the variance-covariance matrix of the innovation vectors, the estimation of the multivariate SV (MSV) model is challenging. To tackle this issue, for a very general setting of a MSV model we propose a straightforward MCMC method in which a Metropolis-Hastings method is employed to sample the constrained variance-covariance matrix, where the proposal distribution is an inverse Wishart distribution. Again, the log volatilities of the asset returns can then be simulated via a single-move slice sampler.

Recently, factor SV models have been proposed to extract hidden market changes. Geweke and Zhou (1996) propose a factor SV model based on factor analysis to measure pricing errors in the context of the arbitrage pricing theory by letting the factors follow the univariate standard normal distribution. Some modification of this model have been proposed, among others, by Pitt and Shephard (1999a) and Jacquier et al. (1999). The main feature of the factor SV models is that the factors follow a univariate SV process, where the loading matrix is a lower triangular matrix with unit entries on the main diagonal. Although the factor SV models have been successful in practice, it has been recognized that the order of the component may affect the sample likelihood and the selection of the factors. Therefore, in applications, the component order has to be considered carefully. For instance, the factor SV model should be fitted to several permuted data to check whether the ordering affects the estimation results. In the thesis, a new factor SV model is proposed. Instead of setting the loading matrix to be lower triangular, we set it to be column-orthogonal and assume that each column has unit length. Our method removes the permutation problem, since when the order is changed then the model does not need to be refitted. Since a strong assumption is imposed on the loading matrix, the estimation seems even harder than the previous factor models. For example, we have to sample columns of the loading matrix while keeping them to be orthonormal. To tackle this issue, we use the Metropolis-Hastings method to sample the loading matrix one column at a time, while the orthonormality

between the columns is maintained using the technique proposed by Hoff (2007). A von Mises-Fisher distribution is sampled and the generated vector is accepted through the Metropolis-Hastings algorithm.

Simulation studies and applications to real data are conducted to examine our inference methods and test the fit of our model. Empirical evidence illustrates that our slice sampler within MCMC methods works well in terms of parameter estimation and volatility forecast. Examples using financial asset return data are provided to demonstrate that the proposed factor SV model is able to characterize the hidden market factors that mainly govern the financial time series. The Kolmogorov-Smirnov tests conducted on the estimated models indicate that the models do a reasonable job in terms of describing real data.

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisors Professor Adam Kolkiewicz and Professor Don McLeish for their insight, guidance, constant encouragement, and for their support. During my tenure as a doctoral student, I have gained a lot of knowledge from them.

I am also very grateful for the assistance and valuable advice I received from my committee members: Professor Tony Wirjanto and Professor David Saunders, and for all the help they have given me during a past couple of years.

Many thanks to my friends, Longyang Wu, Zhijian Chen, Zhaoxia Ren, Hua Shen, Yanqiao Zhang, Zhiyue Huang, Jianfa Cong, Ker-Ai Lee, and Hui Zhao for their help and encouragement during my time at the University of Waterloo. I also want to take this opportunity to thank Professors Wendy Huang, Deli Li, and Tianxuan Miao at Lakehead University for their continual encouragement. Thanks also to the staff, Mary Lou Dufton, Mary Flatt, Marg Feeney, Lucy Simpson, Karen Richardson, and Melissa Cambridge.

I am indebted to my family, my parents, my sisters and brother, for their support and encouragement throughout my life. Last, but not least, I am deeply grateful for my wife, Yanhong Song, and my son, Yusong Men. Without their understanding, love, and patience, I would not have been able to finish this thesis.

Contents

List of Figures	x
List of Tables	xv
1 Introduction	1
1.1 Stochastic Volatility Models	1
1.2 Univariate SV Models	3
1.3 Literature Review	7
1.3.1 Estimation of the Univariate SV Models	7
1.3.2 Basic Multivariate SV Models and Parameter Estimation	10
1.3.3 Factor Multivariate SV Models and Parameter Estimation	14
1.3.4 Dynamic Correlation MSV Models and Parameter Estimation	16
1.4 Contributions and the Presentation of the Thesis	17
2 Slice Sampler within MCMC Methods for Univariate SV Models	20
2.1 Introduction	20
2.2 Slice Sampler Within MCMC Algorithms for Univariate SV Models	24
2.2.1 Estimation of the ASV Model	24
2.2.2 Estimation of a Heavy-tailed ASV (ASV-t) Model	31
2.2.3 Particle Filter	34

2.2.4	Diagnostics	37
2.3	Simulation Studies and Methodology Comparison Based on Real Data	38
2.3.1	Simulation Studies for the ASV Model	38
2.3.2	Comparison Between Three Single-move Estimation Methods	44
2.3.3	Comparison in Sampling the Latent States with and without the Slice Sampler	46
2.4	Empirical Illustrations	49
2.4.1	Data Analysis of the ASV and ASV-t Models	50
2.4.2	Estimation of Several Competing Models	51
2.5	Conclusion	58
2.6	Appendix	59
3	Efficient Bayesian Estimation of a Multivariate Stochastic Volatility Model with Cross Leverage	67
3.1	Introduction	67
3.2	Model and Estimation	70
3.2.1	Model	70
3.2.2	Estimation: an MCMC Algorithm	72
3.2.3	Simulation of Σ	79
3.3	Simulation Studies	82
3.4	Application	90
3.4.1	Stock Returns from Different Industries	91
3.4.2	Stock Returns from the Finance Sector	94
3.5	Conclusion	97
3.6	Appendix	98
4	Factor Stochastic Volatility with Orthogonal Loadings	105
4.1	Introduction	105

4.2	Factor SV Models: a Brief Review	107
4.3	Proposed Model, Identification, Estimation and Assessment	111
4.3.1	Model and Its Identification	111
4.3.2	Estimation: an MCMC Algorithm	116
4.3.3	Model Selection and Assessment	121
4.4	Empirical Exploration and Comparative Analysis	123
4.4.1	Forecast Assessment for the two-factor Idiosyncratic PPCAF Model	124
4.5	Factor Structure of Returns of International Stock Indices	132
4.5.1	Model Fit and Data Analysis	132
4.5.2	Forecasting Analysis	141
4.6	Conclusion and Remarks	142
4.7	Appendix	144
5	Avenues for Future Work	151
5.1	Summary of Contributions	151
5.2	Estimation of More General Univariate SV Models	152
5.3	Multivariate SV Models Under the General PPCA Framework	153
	Bibliography	154

List of Figures

1.1	Time series of prices, log returns and absolute log returns of the IBM stock.	2
2.1	Time series of the first 5000 sampled points from the full conditionals of parameters in the ASV model based on the generated return data.	40
2.2	Histograms and dynamics of the samples from the full conditionals of parameters in the ASV model based on the generated return data.	40
2.3	Comparison between the theoretical CDF of the standard normal distribution and empirical CDF of the residuals after fitting the ASV model to the generated returns.	42
2.4	Comparison between the absolute returns and the true volatilities with the estimated and one-step ahead forecasted volatilities under the ASV model based on the generated asset return data.	43
2.5	Ratio comparison between true and estimated and forecasted volatilities based on the generated asset return data.	43
2.6	The top panel shows the scatter plot of $u(t)$ while the bottom the histogram of $u(t)$	44
2.7	Comparison between the theoretical uniform CDF and the empirical CDF of the PITs from generated return data.	44
2.8	Acceptance rate of the latent states of the heavy-tailed SV models fitted to the exchange rate data.	49

2.9	Comparison between the theoretical and empirical CDFs of the observation errors of the ASV (left) and ASV-t (right) models based on the IBM return data.	51
2.10	Comparison between the theoretical and empirical CDFs of the PITs of the ASV (left) and ASV-t (right) models upon the IBM returns.	52
2.11	The estimated $\sqrt{\lambda_t}$ s and volatilities obtained from the fit of the ASV-d model to the IBM return data.	54
2.12	Comparison of absolute asset returns vs. forecasted volatilities from the competing models.	57
3.1	Histograms and time series of samples from the full conditionals of the location parameters based on the generated return data.	83
3.2	Histograms and time series of samples from the full conditionals of the persistence parameters based on the generated return data.	84
3.3	Volatility comparison for the fourth component of the generated time series of returns.	86
3.4	Differences between the true and estimated volatilities from the MSV model based on the generated return data.	87
3.5	Scatter plot (top) and histogram (bottom) of the PITs for the generated return data.	89
3.6	Comparison between the empirical CDF of the PITs under the MSV model based on the generated returns and the theoretic CDF of the uniform distribution over the interval $[0, 1]$	90
3.7	Histograms and dynamics of samples of the location parameters of the MSV model for the four industry stock return data.	93
3.8	Scatter (top) and histogram (bottom) plots of the PITs for the four-dimensional industry return data.	93

3.9	Comparison between the CDF of the PITs for the four-dimensional industry return data and the theoretical CDF of the uniform distribution over the interval $[0, 1]$	94
3.10	Scatter (top) and histogram (bottom) plots of the PITs for the five-bank return data.	96
3.11	Comparison between the CDF of the PITs for the five-bank return data and the theoretical CDF of the uniform distribution over the interval $[0, 1]$	97
4.1	Comparison of the absolute first factor time series and the corresponding true volatilities with the estimated and the in-sample forecasted volatilities based on a simulated data set from the two-factor idiosyncratic PPCAF model.	128
4.2	Comparison of the absolute second factor time series and the corresponding simulated volatilities with the estimated and the in-sample forecasted volatilities based on a simulated data set from the two-factor idiosyncratic PPCAF model.	129
4.3	Comparison of the absolute simulated third time series with the in-sample and out-of-sample forecasts based on a simulated data set from the two-factor idiosyncratic PPCAF model.	130
4.4	RMSE comparison between true and out-of-sample forecasted volatilities through the two-factor idiosyncratic PPCAF models (the dotted line) and univariate SV models (the solid line) based on the last four block data.	131
4.5	Comparison between true and out-of-sample forecasted volatilities of the first component from the two-factor idiosyncratic PPCAF and univariate SV models in the last four blocks.	131

4.6	Time series comparison of the estimated factors from the indices return data using the four-factor idiosyncratic PPCAF model.	133
4.7	Check the independence between the first factor and additive innovations.	137
4.8	Factor effect comparison.	140
4.9	Comparison of the estimated first factor time series with the corresponding MCMC estimated and forecasts volatilities based on the international return data set.	141
4.10	Comparison between the absolute returns of NASDAQ with the one-step ahead forecasted volatilities.	142
4.11	Comparison of volatilities explained by the first three dominant factors for the returns of TSX from the PPCAF model with idiosyncratic errors.	145
4.12	Comparison of volatilities explained by the first three dominant factors for the returns of S&P500 from the PPCAF model with idiosyncratic errors.	146
4.13	Comparison of volatilities explained by the first three dominant factors for the returns of NASDAQ from the PPCAF model with idiosyncratic errors.	146
4.14	Comparison of volatilities explained by the first three dominant factors for the returns of DJI from the PPCAF model with idiosyncratic errors. .	147
4.15	Comparison of volatilities explained by the first three dominant factors for the return of FTSE from the PPCAF model with idiosyncratic errors.	147
4.16	Comparison of volatilities explained by the first three dominant factors for the returns of DAX from the PPCAF model with idiosyncratic errors.	148
4.17	Comparison of volatilities explained by the first three dominant factors for the returns of HS from the PPCAF model with idiosyncratic errors. .	148
4.18	Comparison of volatilities explained by the first three dominant factors for the returns of Nikkei from the PPCAF model with idiosyncratic errors.	149

4.19 Comparison of the estimated second factor time series with the corresponding MCMC estimated and forecasts volatilities based on the international return data set. 149

4.20 Comparison of the estimated third factor time series with the corresponding MCMC estimated and forecasts volatilities based on the international return data set. 150

List of Tables

2.1	MCMC algorithm for the ASV model.	26
2.2	True and estimated parameters of the ASV model based on the simulated return data.	40
2.3	True and estimated parameters of the BSV model via the slice sampler within MCMC method.	45
2.4	True and estimated parameters of the BSV model through the JPR approach.	46
2.5	True and estimated parameters of the BSV model via the method in Kim et al. (1998).	46
2.6	Estimated parameters of the SV models under various MCMC methods for daily observations of weekday close exchange rates for the U.K. Sterling/U.S. Dollar exchange rate from 1/10/81 to 28/6/85.	47
2.7	Estimates of parameters obtained from daily returns of the IBM stock through the ASV, ASV-t and competing models. The standard errors are in the parentheses.	55
2.8	The AIC and BIC values for the ASV and ASV-t and competing models based on the IBM return data.	56
3.1	MCMC algorithm for the MSV model.	74
3.2	The MCMC sampler for sampling \mathbf{h}_t from iteration n to $n + 1$	76

3.3	The variance-covariance matrix (in boldface) and correlations (in italics) used for data generation.	83
3.4	Bayesian estimates of the variance-variance (in boldface) and correlations (in italics) from the generated return data.	85
3.5	Comparison between true and estimated parameters from the ASV and MSV models based on generated data.	85
3.6	RMSE for the ASV and MSV models.	87
3.7	Bayesian estimates of the unconditional variance-covariance (in boldface) and correlation matrices (in italics) from the return time series of four industries.	92
3.8	Comparison of estimated parameters between ASV and MSV models.	92
3.9	Bayesian estimates of the unconditional variance-covariance (in boldface) and correlation matrices (in italics) from the return time series of the five banks.	95
3.10	Comparison of estimated parameters between ASV and MSV models based on the five-bank return data.	96
4.1	MCMC algorithm for the PPCAF model.	117
4.2	Model selection using the AIC and BIC criteria.	125
4.3	Comparison between true and estimated parameters of the latent AR(1) processes.	125
4.4	Comparison between true and estimated parameters of the measurement equations.	126
4.5	RMSE comparison between true and forecasted volatilities on the simulated eight-dimensional return data from an idiosyncratic two-factor PPCAF model. All values have been multiplied by 1000.	128
4.6	Model selection using the AIC and BIC criteria.	133

4.7 Estimated parameters of the three latent AR(1) processes from the three-factor PPCAF model with idiosyncratic observation errors. 134

4.8 Percentage of the variance of each series explained by each factor in analysis of the international indices return time series from the three-factor idiosyncratic PPCAF model. 138

Chapter 1

Introduction

1.1 Stochastic Volatility Models

While the time series of asset returns are observable, volatilities of asset returns are unobservable. In addition, one of the stylized facts about financial asset returns is that volatilities of asset returns are time varying and clustered over time. To discuss this, let us consider the IBM stock prices downloaded from the web site finance.yahoo.com with 1730 observations. The top trajectory in Figure 1.1 is the time series of daily closing prices of the IBM stock from January 3, 2003 to November 13, 2009. The time series in the middle is the log returns defined as $r_t = \ln(p_t) - \ln(p_{t-1})$, where p_t is the closing price at day t , and the bottom plot is the dynamics of the absolute log returns. From these graphs, we can see that the volatilities of asset returns are time-dependent.

There are two commonly used types of models to characterize the time-varying volatility of asset returns. They are generalized autoregressive conditional heteroskedasticity (GARCH) models and stochastic volatility (SV) models. These models attempt to describe volatility as a random process. The GARCH models, proposed by Bollerslev (1986), are extensions of the autoregressive conditional heteroscedasticity (ARCH)

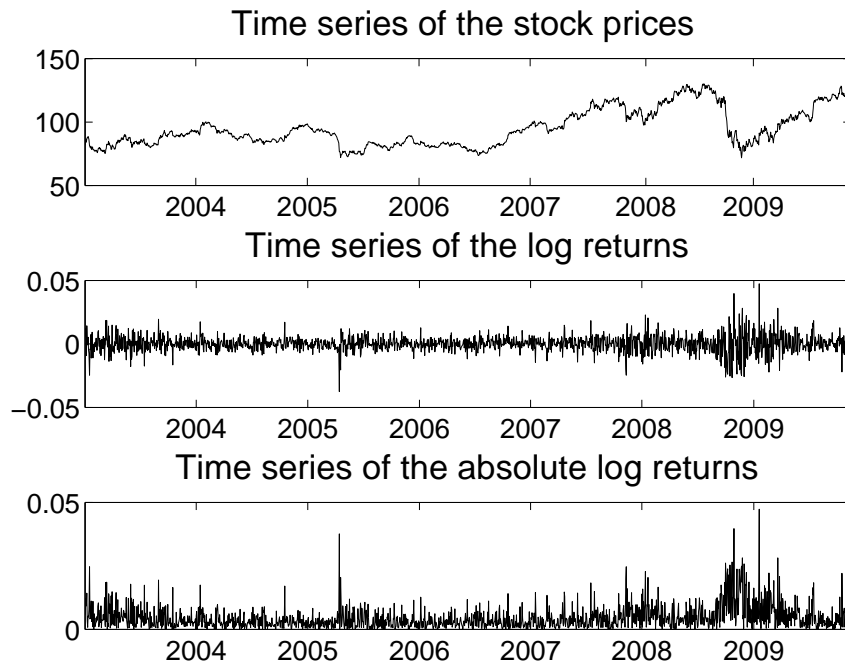


Figure 1.1: Time series of prices, log returns and absolute log returns of the IBM stock.

model by Engle (1982) to allow current volatility to depend not only on past returns but also on past volatilities, and this has also been extended to various directions. Unlike in GARCH related models, in SV models (see Taylor (1986)) the log volatility of asset returns is modeled as a latent first-order autoregressive (AR(1)) process. SV models are attractive because they are closer to the theoretical models often specified in financial theory to represent the behaviour of financial prices, which are generalizations of the Black-Scholes option pricing formula to allow volatility clustering in asset returns (see, for instance, Hull and White (1987)). Comparing with GARCH models, SV models can better capture the main empirical properties observed in daily series of financial asset returns. For instance, the persistence of volatility implied by the GARCH(1,1) models is usually higher than that implied by the SV models, and also the SV models with normally distributed innovations tend to fit better than GARCH models even with heavy-tailed distributed innovations (see, for example, Broto and Ruiz (2004) and

Carnero et al. (2003)). This thesis works with univariate and multivariate SV models and focuses on developing efficient estimation methods for the models as well as proposing novel factor SV models.

1.2 Univariate SV Models

As discussed in the last section, our focus is on the stochastic volatility models in which volatility is an unobserved random process. Specifically, the log volatility follows a hidden Markov process. Define by y_t the observation of asset returns at time t ($t \leq T$), a generic univariate SV model can be formulated as follows:

$$y_t = \sigma_t \epsilon_t, \quad (1.1a)$$

$$\alpha(\sigma_{t+1}^2, \delta) = \mu + \phi(\alpha(\sigma_t^2, \delta) - \mu) + \sqrt{\sigma} \eta_{t+1}, \quad (1.1b)$$

where the innovation vectors $(\epsilon_t, \eta_{t+1})'$ are independently and identically distributed (i.i.d.) according to a joint probability density function $f(\epsilon_t, \eta_{t+1})$, and σ_t is a time-varying scalar representing the standard volatility of y_t . In the latent AR(1) process, $\alpha(\sigma_t^2, \delta)$, a function of σ_t^2 and other parameter δ , may vary according to the behaviour of volatility. As usual, we set $|\phi| < 1$ to ensure covariance or weak stationarity of the process. The latent AR(1) process indicates some volatility persistency through the function $\alpha(\sigma_t^2, \delta)$. In the literature, a common representation of the function $\alpha(\sigma_t^2, \delta)$ is given by a Box-Cox transformation of σ_t^2 . This transformation, applied by Yu et al. (2006) and Zhang and King (2008) in the specification of model (1.1), is defined as

$$\alpha(\sigma_t^2, \delta) = \begin{cases} (\sigma_t^{2\delta} - 1)/\delta & \text{if } \delta \neq 0; \\ \ln(\sigma_t^2) & \text{if } \delta = 0. \end{cases}$$

Let $h_t = \alpha(\sigma_t^2, \delta)$ denote the Box-Cox transformation. Then (1.1) can be equivalently represented as

$$y_t = \sqrt{q(h_t, \delta)} \epsilon_t, \quad (1.2a)$$

$$h_{t+1} = \mu + \phi(h_t - \mu) + \sqrt{\sigma} \eta_{t+1}, \quad (1.2b)$$

where

$$q(h_t, \delta) = \begin{cases} (1 + \delta h_t)^{1/\delta} & \text{if } \delta \neq 0; \\ \exp(h_t) & \text{if } \delta = 0. \end{cases}$$

Yu et al. (2006) estimate their proposed SV model without leverage effect using daily returns of the dollar/pound exchange rate for the period from January 1, 1986 to December 31, 1998. The authors find that the estimate of δ has a positive sign. Zhang and King (2008) fit the daily returns of the Australian All Ordinaries stock index and obtain a negative estimate of δ .

Most of the univariate and multivariate SV models in the literature are special cases of (1.2) with $\delta = 0$. In this case, the Box-Cox transformed model becomes

$$y_t = \exp(h_t/2) \epsilon_t, \quad (1.3a)$$

$$h_{t+1} = \mu + \phi(h_t - \mu) + \sqrt{\sigma} \eta_{t+1}, \quad (1.3b)$$

where $h_t, t = 1, \dots, T$, are the log volatilities of asset returns that follow an AR(1) process. The conditional mean of y_t is $E(y_t|h_t) = 0$ and its conditional variance is $Var(y_t|h_t) = \exp(h_t)$. The variance σ of the log volatility process (or the transition process) (1.3b), measures the uncertainty about the future volatility. Models like (1.3) are also referred to as non-linear state-space or hidden Markov models since the measurement equation (1.3a) depends non-linearly on h_t which is unobservable. This model is studied in Taylor (1986) under the assumptions: $\epsilon_t \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1)$, $\eta_{t+1} \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1)$ and

$corr(\epsilon_s, \eta_{t+1}) = 0$ for all s and t . As discussed by the author, even this simple SV model is capable of producing excess kurtosis in the marginal distribution of data in line with the empirical stylized facts of financial asset returns (for a discussion, see Broto and Ruiz (2004)). In the thesis we will exclusively work with this special case of the Box-Cox transformation.

By specifying different distributions for the innovation ϵ_t in (1.3a), the univariate SV model (1.3) can be extended to capture many different stylized facts of financial asset returns. The first extension of model (1.3) is an asymmetric stochastic volatility model (ASV) introduced in Harvey and Shephard (1996), Jacquier et al. (2004) and Omori et al. (2007), where the two innovations have bivariate normal distribution with correlation $\rho = corr(\epsilon_t, \eta_{t+1})$. In practice, the correlation coefficient is often found to be negative and ρ is interpreted as the leverage effect between the asset returns and the latent log volatility process. In other words, if an underlying asset experiences a positive (negative) return, then the volatility at the next observation time will tend to decrease (increase).

As argued by Yu (2005), the conditional expectation of the log variance of y_{t+1} , based on the ASV model, is

$$E(h_{t+1}|y_t) = \mu(1 - \phi) + \frac{\mu\phi}{1 + \phi} + \rho\sigma \exp\left(-\frac{\sigma^2}{4(1 - \phi^2)^2} + \frac{\mu\sigma}{(1 - \phi^2)}\right)y_t,$$

which results in a leverage effect between the two equations whenever $\rho < 0$. On the other hand, for the model of Jacquier et al. (2004)

$$h_t = \mu + \phi(h_{t-1} - \mu) + \sqrt{\sigma}\eta_t,$$

and $\rho = \text{corr}(\epsilon_t, \eta_t)$, Asai et al. (2006) give the result

$$\frac{\partial h_{t+1}}{\partial y_t} = \frac{\rho\sqrt{\sigma} \exp(h_{t+1})/\exp(h_t)}{1 + 0.5\rho\sqrt{\sigma}\epsilon_{t+1}}, \quad (1.4)$$

from which it is unclear whether the correlation ρ can be interpreted as a leverage effect since the partial derivative of (1.4) also depends on the sign of ϵ_{t+1} through the denominator. More specifically, the sign of (1.4) depends on the sign of the denominator. As a consequence, the leverage effect is not guaranteed to exist. Thus, the authors conclude that the ASV model formulated through (1.3) has a clear leverage effect, while the leverage effect defined in Jacquier et al. (2004) is not correct. They use a single-move algorithm which has slow convergent, highly dependent consecutive states and inefficient mixing. From now on, the ASV model based on the equations in (1.3) will be referred to as the leverage effect SV model.

The second extension of model (1.3) is to allow the innovation ϵ_t of the observation equation (1.3a) to follow a Student- t distribution with unknown v degrees of freedom, retaining the normality assumption on the latent noise η_t . In this context, the measurement equation (1.3a) is able to accommodate the financial asset returns with thick tails. This heavy-tailed SV model, called the SV- t model in short, is studied in Harvey and Shephard (1996), Shephard and Pitt (1997), Chib et al. (1998) and Kim et al. (1998). They proposed multi-move algorithm that sample latent volatility vector in one block to improve simulation efficiency. It should be mentioned that if the leverage effect is introduced in SV- t models, the Student- t distribution is often decomposed into a mixture of a standard normal and square root of an inverse Gamma distribution from which the correlation is permitted between the two standard normal variables.

The third extension is to impose more general distributions on the observation errors. Since these distributions of the two innovations are not known, Xu (2006) approximates the joint density of $(\epsilon_t, \eta_{t+1})'$ by a mixture of two bivariate normal distributions.

Abraham et al. (2006) propose a SV model under the assumption that the volatility of asset returns follows an autoregressive process with a Gamma innovation. Also, Barndorff-Nielsen (1997, 1998) propose univariate SV models by assuming that the asset returns follow a stochastic process whose innovation is a product of normal inverse Gaussian (NIG) and standard normal variables.

1.3 Literature Review

1.3.1 Estimation of the Univariate SV Models

Since the work of Taylor (1986), univariate SV models have had some success in modelling volatilities of financial asset returns. A variety of methods have been developed to estimate the parameters and log volatilities of the model. Four of these are widely used in practice and cited in the literature. They are method of moments (MM), maximum likelihood (ML) method including the simulated maximum likelihood (SML) method, and Markov Chain Monte Carlo (MCMC) method. Taylor (1986) uses the MM method to estimate parameters in the uncorrelated SV model. Later, the generalized method of moments (GMM) was proposed by Duffie and Singleton (1993) and Melino and Turnbull (1990) under very general conditions about the error distributions. These approaches are based on the convergence of selected sample moments to their unconditional expected values. An alternative, proposed by Duffie and Singleton (1993), is the simulated method of moments (SMM) that replaces the analytic moments by the moments of a simulated process. However, as discussed in Broto and Ruiz (2004), these estimation methods have poor finite sample properties, and their efficiency is suboptimal relative to likelihood-based approaches. Harvey and Shephard (1996) develop a so-called quasi-maximum likelihood (QML) method. The idea of the QML is to first linearize the observation equation by taking logarithm of the squared observation

equation (1.3a), that is

$$\log(y_t^2) = \alpha + h_t + \xi_t, \quad (1.5a)$$

$$h_{t+1} = \mu + \phi(h_t - \mu) + \sqrt{\sigma}\eta_{t+1}, \quad (1.5b)$$

where $\alpha = E(\log(\epsilon_t^2))$, $\xi_t = \log(\epsilon_t^2) - E(\log(\epsilon_t^2))$.

Model (1.5) is in a linear state space form with equation (1.5a) having a non-Gaussian innovation. Under the assumption that ϵ_t is Gaussian, $\log(\epsilon_t^2)$ has the log Gamma distribution with mean -1.27 and variance $\frac{\pi^2}{2}$ (see Abramowitz and Stegun (1970) for details). In order to use Kalman filter, Harvey and Shephard (1996) assume that $\xi_t \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \frac{\pi^2}{2})$ and obtain the MLE of parameters. Because the linearized measurement equation does not actually have a normal innovation, the estimation method is called a quasi-maximum likelihood method and has been extensively applied by researchers to estimate Multivariate SV models (see Chib et al. (2009), Lopes and Polson (2010) and references therein).

Due to the difficulty in obtaining a closed form expression for the likelihood function of the observed data, approximate maximum likelihood methods have also been proposed to estimate parameters in the univariate SV models. For instance, Fridman and Harris (1998) propose a direct ML method from which the likelihood is calculated by using a recursive numerical integration procedure of Kitagawa (1987) for non-Gaussian SV models. Xu et al. (2011) establish an empirical characteristic function approach to capture the leverage between the two equations. Kawakatsu (2007) obtains the MLE through a technique similar to that of Xu et al. (2011). Instead of approximating the joint density, the author approximates the marginal densities of individual innovations of the SV model by a finite combination of weighted univariate normal densities. To evaluate the likelihood of the observed data, the Gauss-Hermite quadrature numerical integration was employed.

Recently, there have been some advances in the likelihood-based technique of estimating the SV models primarily due to the contribution of simulation methods such as importance sampling and MCMC. Notably, the SML approach for estimating the SV models is studied in Danielsson and Richard (1993) and Danielsson (1994). The main advantage of importance sampling is that it is computationally less demanding relative to MCMC methods, which are more time consuming and sometimes fail to converge. Furthermore, the accuracy of the SML methods can be improved by increasing the sample size of the simulations of the latent states. However, MCMC methods have the advantage of allowing a large dimensional problem to be split into smaller dimensional tasks. It should be noted that most of the MCMC methods involve the use of Gibbs procedures where the posteriors are simulated repeatedly, and the parameters of the model are estimated by sample means.

MCMC methods, which were also proposed independently by Shephard (1993) and Jacquier et al. (1994) to estimate the SV models, have been widely employed in the estimation of univariate SV models. In their MCMC algorithms, the authors first calculate the posterior distributions of the model's parameters and latent states, and then these posteriors are sampled cyclically in which the latent states are sampled through the Metropolis-Hastings (MH) algorithm. For the correlated SV models, Jacquier et al. (2004) estimate the parameters of the model by reparameterization through the Cholesky decomposition. The authors even consider a heavy-tailed SV model with leverage effect, where a Student- t distribution is introduced to the measurement equation. Although, as discussed in the previous section, their model does not lead to a proper interpretation of the leverage effect, the estimation methodology is commonly applied to estimate both univariate and multivariate SV models. In fact Broto and Ruiz (2004) comment that the MCMC methods are the most efficient approaches for estimating univariate SV models. The biggest advantage of using the MCMC methods in estimating the SV models, compared to other estimation methods, is that the log

volatilities can be estimated as a by-product of the estimation process. In the past few years, many MCMC methods have been developed to estimate various versions of the univariate SV model under different assumptions on the two innovations driving the mean and volatility equations. Recently, Omori et al. (2007) fit a ASV model by approximating the joint density of the two innovations with a ten-component mixture of bivariate normal distributions. In their approach, the authors first take the logarithm of the squared measurement equation of the ASV model. Then the innovation distribution of the transformed measurement equation is approximated by a mixture of ten univariate normal distributions, which is similar to those in Kim et al. (1998) and Chib et al. (2002), where the authors use a seven-component mixture for the density approximation to fit a univariate SV model without leverage effect. The joint distribution of the two innovations of the transformed ASV model is then approximated by a ten-component mixture of bivariate normal distributions. By introducing the sign random variables based on the observations, an MCMC algorithm is derived for the estimation of parameters of the model.

1.3.2 Basic Multivariate SV Models and Parameter Estimation

Due to the success of univariate SV models in capturing many stylized facts about the financial asset returns, multivariate SV (MSV) models have attracted considerable attention in modeling a group of financial asset returns. There are mainly three types of MSV models: basic MSV models (BMSV), factor SV (FSV) models, and dynamic correlation (DMSV) models. In this section we discuss the BMSV models and their parameter estimation. The other two types of MSV models will be discussed in the next two subsections.

A natural extension of the univariate SV model to the multivariate case is to simply increase the number of univariate SV models while specifying correlations between the

innovations. The motivation behind this extension is that multivariate asset returns evolve in a complex way. It is often found that the trajectories of asset returns share common features. For instance, asset prices may drop or jump almost at the same time. To model this common feature, a series of independent univariate SV models will not be ideal since they only characterize the individual time series of returns and the corresponding log volatilities. To overcome this drawback, BMSV models have been formulated. For a review of the MSV models, see Asai et al. (2006), Chib et al. (2009) or Lopes and Polson (2010).

Let $\mathbf{y}_t = (y_{1t}, \dots, y_{mt})'$, $t \leq T$, be a vector of m time series of asset returns, where m is a positive integer representing the dimension of \mathbf{y}_t . The more general BMSV model is defined as

$$\mathbf{y}_t = \mathbf{H}_t^{\frac{1}{2}} \boldsymbol{\epsilon}_t, \quad t = 1, \dots, T, \quad (1.6)$$

$$\mathbf{h}_{t+1} = \boldsymbol{\mu} + \boldsymbol{\Phi}(\mathbf{h}_t - \boldsymbol{\mu}) + \boldsymbol{\eta}_{t+1}, \quad t = 1, \dots, T - 1, \quad (1.7)$$

$$\mathbf{h}_1 \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}_0), \quad (1.8)$$

where

$$\mathbf{H}_t^{\frac{1}{2}} = \text{diag}(\exp(h_{1,t}/2), \dots, \exp(h_{m,t}/2)), \quad (1.9)$$

$$\boldsymbol{\Phi} = \text{diag}(\phi_1, \dots, \phi_m), \quad (1.10)$$

$$\begin{pmatrix} \boldsymbol{\epsilon}_t \\ \boldsymbol{\eta}_{t+1} \end{pmatrix} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}), \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{\epsilon\epsilon} & \boldsymbol{\Sigma}_{\epsilon\eta} \\ \boldsymbol{\Sigma}_{\eta\epsilon} & \boldsymbol{\Sigma}_{\eta\eta} \end{pmatrix}, \quad (1.11)$$

and $\boldsymbol{\mu} = (\mu_1, \dots, \mu_m)'$ is the location vector and $\boldsymbol{\Phi} = \text{diag}\{\phi_1, \dots, \phi_m\}$ is the persistence parameters' diagonal matrix. In order for the AR(1) processes to be stationary, all persistence parameters are assumed to satisfy the weak or covariance stationarity condition as $|\phi_i| < 1, i = 1, \dots, m$. Further, if the (i, j) element of $\boldsymbol{\Sigma}_0$ is also assumed to equal

the (i, j) element of $\Sigma_{\eta\eta}$ divided by $1 - \phi_i\phi_j$, then it can be verified that Σ_0 satisfies the following stationary condition:

$$\Sigma_0 = \Phi\Sigma_0\Phi + \Sigma_{\eta\eta}, \quad (1.12)$$

The cross covariance matrix $\Sigma_{\epsilon\eta}$ between the two innovations $\epsilon_t = (\epsilon_{1t}, \dots, \epsilon_{mt})'$ and $\eta_t = (\eta_{1t}, \dots, \eta_{mt})'$ are allowed to be a non-zero matrix, so that the model is capable of capturing the cross correlation between the two innovation vectors. For the BMSV model to be identifiable, Σ_ϵ is defined as a correlation matrix. The latent vector AR(1) process of $\mathbf{h}_t = (h_{1t}, \dots, h_{mt})'$ models the unobserved log volatilities of asset returns. In a more general setting, the Student- t distribution can be assumed for the innovations ϵ_t as in Harvey et al. (1994), Yu and Meyer (2006) and Jacquier et al. (2004).

The BMSV models are studied by Harvey et al. (1994), Danielsson (1998), Smith and Pitts (2006) and Chan et al. (2006) under different distributional assumptions on the innovations and the cross correlations. Specifically, Harvey et al. (1994) use the QML method to estimate the BMSV model where the cross covariance matrix $\Sigma_{\epsilon\eta}$ is a zero matrix. In order to obtain the MLE through a Kalman filter, a linearization of the observation equations is required. As mentioned by the authors, the QML method can not be extended to estimate the leverage BMSV model. So et al. (1997) follow a similar idea of Harvey et al. (1994) but consider a situation where the off-diagonal elements of the persistence matrix Φ may not be zero. To estimate parameters of the model, the authors derive a computationally efficient expectation-maximization (EM) algorithm. Extended from the univariate case of the SML, Danielsson (1998) applies the SML to estimate the parameters of model defined above. Smith and Pitts (2006) propose a bivariate MSV model with an intervention factor contained in the latent equation which represents the intervention by banks. An MCMC sampling scheme is derived for parameter estimation. So and Kwok (2006) consider a MSV model with

the specification of the measurement equation being the same as (1.6) but the latent dynamics following an autoregressive fractionally integrated moving average process. The model is called the ARFIMA(p, d, q) model and is estimated by the QML estimation method.

For the asymmetric MSV model with $\Sigma_{\epsilon\eta}$ having potentially non-zero entries, Asai and McAleer (2006) permit leverage effect to enter only between innovations of the observation equations and their corresponding log volatility processes. That is, they require

$$\Sigma_{\epsilon\eta} = \text{diag}(\lambda_1\sigma_{1,\eta}, \dots, \lambda_m\sigma_{m,\eta}),$$

where $\sigma_\eta = (\sigma_{1,\eta}^2, \dots, \sigma_{m,\eta}^2)'$ is the variance vector of η_{t+1} . If the components in the vector $\lambda = (\lambda_1, \dots, \lambda_p)'$ are negative, then $\Sigma_{\epsilon\eta}$ can capture these partially specified leverage effect. It is obvious that this is a direct extension of the formulation of leverage effect for a univariate SV model studied for instance in Harvey and Shephard (1996).

A more general BMSV model is proposed by Chan et al. (2006), where the matrix $\Sigma_{\epsilon\eta}$ could be a covariance matrix with any structure. This model permits non-zero correlations within and between the two innovation vectors. The authors develop an MCMC algorithm to generate estimates of parameters of their model and the latent states. To sample the variance-covariance matrix of innovations, the authors employ a methodology proposed in Wong et al. (2003). In their algorithm, the correlation matrix \mathbf{R} calculated from Σ is not sampled directly, but is parameterized as

$$\mathbf{R}^{-1} = \mathbf{T}\mathbf{G}\mathbf{T}, \quad \mathbf{T} = \text{diag}(\sqrt{G^{11}}, \dots, \sqrt{G^{2m \times 2m}}),$$

where \mathbf{G} is a correlation matrix and G^{ii} is defined as the main diagonal components of the inverse of \mathbf{G} . With this reparameterization, the authors sample component wise

the off-diagonal elements of G by using the MH algorithm. Through specified prior distributions, each off-diagonal component is allowed to be zero. Obviously, the simulation of the correlation matrix is complex and likely to be time-consuming for high-dimensional financial returns.

1.3.3 Factor Multivariate SV Models and Parameter Estimation

In recent years, factor-based stochastic volatility (FSV) models have also been used in the analysis of multivariate financial asset returns. The motivation of the FSV model is to detect the hidden factors that partially drive the underlying multivariate time series of asset returns. Geweke and Zhou (1996) propose a factor model to measure the pricing errors of the arbitrage pricing theory, where the time series of observed returns is a linear regression of the latent factors with observation errors. In their model, all factors are assumed to be independent and follow a standard normal distribution. It also assumed that the observation errors are idiosyncratic and each follows a univariate normal distribution. These observation errors are independent of factors. In order for the proposed factor model to be identifiable, the loading matrix is set to be a lower triangular matrix with positive entries on the main diagonal, and the entries below the main diagonal are free parameters. Subsequent to the model studied by Geweke and Zhou (1996), various FSV models have been proposed in the literature, such as Jacquier et al. (1999), Pitt and Shephard (1999a), Liesenfeld and Richard (2003) and Chib et al. (2006), among others. A common structure of the FSV models is an extension of the model studied by Geweke and Zhou (1996), such that the loading matrix has ones on the diagonal. The factors in the FSV models follow standard univariate SV processes and the observation error vector has a multivariate normal distribution with zero mean. Conditioned on the factors, the observed returns have independent multivariate normal distributions. Jacquier et al. (1999) consider a FSV model by assuming that the

observation errors are independently and identically distributed according to a multivariate normal distribution with zero mean and a general constant variance-covariance matrix. Liesenfeld and Richard (2003) study a model similar to that of Jacquier et al. (1999) but their model is fitted with only one factor. The time-varying correlations are mainly captured by the dynamics of unobservable factor(s). Pitt and Shephard (1999a) generalize the above FSV models even further by assuming that the additive errors are uncorrelated and follow univariate latent SV processes from which the time-varying correlation of the considered time series of asset returns is characterized by both the factors and stochastic observation errors. An even more general extension is given by Chib et al. (2006) where fat-tailed Student- t distributions and jumps are assumed for the observation equations. Lopes and Carvalho (2007) have considered a general model which includes the models studied by Pitt and Shephard (1999a), and Aguilar and West (2000), and extend it in two directions by (i) letting the loading matrix to be time-varying and (ii) allowing Markov switching in the log volatility of common factors. Han (2006) modifies the model of Pitt and Shephard (1999a) and Chib et al. (2006) by allowing the factors to be Markovian and follow first-order autoregressive processes. The key point of the FSV models is that not only the conditional variance-covariance of the asset returns changes with time but also the conditional correlation depends on time.

Since the likelihood function of observed multivariate returns for FSV models does not have a closed form expression in general, MCMC methods in Bayesian framework have been proposed as a preferred approach. Jacquier et al. (1999) and Pitt and Shephard (1999a) for instance propose MCMC methods, where the log volatilities (or the state random variables) are augmented as parameters and sampled one at a time or within blocks from their posterior distributions. Liesenfeld and Richard (2003) show how the MLE can be obtained using importance sampling. Chib et al. (2006) derive an MCMC based method to fit their complex FSV model. Lastly, Han (2006) fits the

proposed model by adapting the approach of Chib et al. (2006) and uses the model for asset allocation.

1.3.4 Dynamic Correlation MSV Models and Parameter Estimation

In the BMSV models, the conditional correlation of financial asset returns is a constant matrix which seems somewhat inconsistent with a model in which variances are dynamically changing. Dynamic conditional correlation (DCC) model allows the conditional correlations among the asset returns to be time-dependent. Yu and Meyer (2006) propose a bivariate SV model, where the Fisher transformation of the correlation of the two innovations follows a stationary AR(1) process. The WinBUGS program is used for the estimation of the model. The authors find that the models that allow for time-varying coefficients generally fit the data better. Tsay (2005) considers a DCC model based on a Cholesky decomposition of the conditional correlation matrix. After the decomposition, the author assumes that the components on the main diagonal of the lower triangular matrix follow univariate SV processes. Since the decomposition is performed at each discrete observation time, the free parameters in the lower triangular matrix are also evolving with time. Jungbacker and Koopman (2006) consider a similar model assuming that these free parameters are time-invariant. A Monte Carlo likelihood method is developed and the model is fitted to daily exchange rate returns.

Another type of DCC model is defined through a Wishart process. Philipov and Glickman (2006a, 2006b) propose a MSV model by assuming that the conditional covariance follows an inverse Wishart distribution where the scalar matrix depends on the past covariance matrix. Asai and McAleer (2007) propose two similar models, where the correlation matrix is represented by a singular value decomposition. In these models the orthogonal matrices are time-dependent. The settings of the two models ensures that the random variance-covariance matrices are positive definite. Recently,

more DCC models via Wishart processes have been used to address the dynamic correlation by *Gourieroux et al. (2004)* and *Gourieroux (2006)*.

1.4 Contributions and the Presentation of the Thesis

It is common to use univariate and multivariate SV models to model the evolution of time series of financial asset returns over time. Efficient estimation of SV models has been extensively studied in the past two decades. Many methods have been proposed to estimate the models according to specific assumptions made about the innovations of SV models. MCMC is a general and more efficient approach to estimate both the univariate and multivariate SV models. Under the Bayesian framework, *Jacquier et al. (1994, 2004)* propose MCMC algorithms to fit univariate SV models, which avoid the difficulty of evaluating the likelihood function analytically. Most MCMC approaches are based on the MH technique for simulation under various proposal densities. Some of them are not efficient and have convergence problems due to the low acceptance rate produced by the proposal distributions. Accordingly, this thesis first considers how to estimate parameters and log volatilities efficiently for univariate SV models and a generalized BMSV model. MCMC based simulation strategies for latent states are developed via the slice sampler introduced in *Neal (2003)*. For the general MSV model defined in *Chan et al. (2006)*, an MH algorithm is derived for the simulation of the variance-covariance matrix of the two observation errors. The inverse Wishart distribution is selected as the proposal when we implement the algorithm for estimating the model. To model latent market factors, a factor SV model is proposed under a probabilistic principal component analysis (PPCA) framework to determine the market factors that govern the multivariate process and model the time-varying correlation of asset returns.

Our first contribution is to consider fitting univariate SV models by using a slice

sampler within an MH algorithm for the simulation of latent state. It is well known that the proposal distribution of the MH algorithm is important for the performance of the simulation procedure. Chib and Greenberg (1995) comment that choosing a good proposal density is like searching for a proverbial needle in a haystack. Usually, a proposal density can be found by approximation of the underlying full conditional (see Jacquier et al. 1994, 2004) or by choosing a standard normal density (see Zhang and King 2008). As is noticed in the literature, the crucial aspect of MCMC for fitting SV models is the sampling quality of the full conditionals of the augmented parameters, which are the log volatilities. Our contribution is to develop MH methods, where the proposal distributions are either simple distributions that can be sampled directly, or unknown distributions that can be simulated by the slice sampler. Since the analytic structure of the underlying density can be adapted by the slice sampler, the resulting algorithm is more efficient. The efficiency of the slice sampler is studied by Roberts and Rosenthal (1999) and Mira and Tierney (2002). The actual use of the slice sampler depends on the specified innovations of SV models and will be presented in the following chapters.

Our second contribution is the estimation of the MSV model under a general correlation structure studied in Chan et al. (2006). In this model, correlations are not only permitted among the innovations of observation and latent equations, but also permitted between the two innovations. We develop an efficient MCMC method for this more general MSV model, where the whole variance-covariance matrix is simulated simultaneously in each iteration through a parameter-extension MH algorithm. The proposal is an inverse Wishart density. For the simulation of the latent states, we propose a slice sampler within an MH method which moves one state at a time. Comparing with the MCMC method in Chan et al. (2006), where the states are simulated within blocks using a multivariate normal proposal, our solution is relatively more straightforward. For instance, in Chan et al. (2006), the proposal is a Gaussian approx-

imation to the underlying posterior distribution whose mode is obtained through the Newton-Raphson method which is time-consuming. Our method is more applicable for high-dimensional data because of the simulation of the variance matrix of innovations.

The last contribution is to propose a factor SV model based on the PPCA framework. Stimulated by the principal component analysis (PCA) and the PPCA (see, for example, Tipping and Bishop 1999), our factor SV model estimates the unobserved factors that drive the processes of financial asset returns. To do this, the observed vectors of asset returns are projected onto a subspace spanned by several constant orthogonal unit vectors, where the projection errors are either isotropic or idiosyncratic. The columns of the loading matrix are closely related to the eigenvectors of the observational time series, and the weights along these orthonormal directions are called factors. In performing this projection, we believe that asset returns are mainly determined by several hidden market factors. In addition to this, we further assume that each factor follows a univariate SV process which is also unobserved. Upon the formulation of this factor SV model, not only the conditional variance-covariance matrix, but also the correlations of the asset returns are allowed to be time-varying. To estimate this hidden Markov model, an MCMC algorithm is developed. Assuming a uniform prior distribution on the unit hypersphere, the posterior distribution of each column, given that other columns having been sampled, is assumed to follow either a von Mises-Fisher distribution in the isotropic case or an unknown distribution for the idiosyncratic type. So the loading matrix can be sampled one column at a time using the method introduced in Hoff (2007) by directly simulating a von Mises-Fisher distribution or using MH algorithm through a von Mises-Fisher proposal.

Chapter 5 concludes the thesis and discusses avenues for future research.

Chapter 2

Slice Sampler within MCMC Methods for Univariate SV Models

2.1 Introduction

Stochastic volatility (SV) models have enjoyed some success in the analysis of time series for financial asset returns. Define by y_t the observation of excess returns at time t ($t \leq T$), the basic univariate SV (BSV) model, proposed by Taylor (1986), can be formulated as

$$y_t = \exp(h_t/2)\epsilon_t, \quad t = 1, \dots, T, \quad (2.1)$$

$$h_{t+1} = \mu + \phi(h_t - \mu) + \sqrt{\sigma}\eta_{t+1}, \quad t = 1, \dots, T - 1, \quad (2.2)$$

$$h_1 \sim \mathcal{N}(\mu, \sigma/(1 - \phi^2)), \quad (2.3)$$

where $\epsilon_t \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1)$, $\eta_{t+1} \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1)$, $\text{corr}(\epsilon_s, \eta_{t+1}) = 0$ for all s and t , and $\mathcal{N}(a, b)$ is a univariate normal distribution with mean a and variance b . The latent random variables $h_t, t = 1, \dots, T$, are the log volatilities of y_t assumed to follow a first-order autoregressive

(AR(1)) process. In Taylor's model, extensions have been proposed to the distribution of ϵ_t and the definition of the correlation between the two equations. For instance, Harvey and Shephard (1996) propose that the two innovations have a correlated bivariate normal distribution. In practice, as the correlation coefficient is often found to be negative, this model is called the asymmetric stochastic volatility model (ASV) or the SV model with leverage effect. In order to characterize the heavy tail property of the marginal distribution of asset returns, a Student- t distribution with unknown degrees of freedom is assumed for the innovations ϵ_t . The heavy-tailed SV models are introduced in Harvey and Shephard (1996) and Jacquier et al. (2004) and adapted by other researchers, such as Shimada and Tsukuda (2005), Yu et al. (2006), Chib et al. (2006) and Zhang and King (2008).

In the estimation of univariate SV models, many methods have been proposed in the literature, such as the quasi-maximum likelihood (QML) method by Harvey and Shephard (1996), the numerical integration method in Kawakatsu (2007), the simulated maximum likelihood (SML) methods by Danielsson (1994), Shephard and Pitt (1997) and Sandmann and Koopman (1998). Recently, Bayesian inference approaches that use Markov Chain Monte Carlo (MCMC) methods have been proposed in Jacquier et al. (1994, 2004), Kim et al. (1998), Omori et al. (2007) and Zhang and King (2008) for the SV models with or without leverage effect. The greatest advantage of the MCMC methodology is that a large dimensional problem can be divided into lower dimensional simulation tasks, in which the log volatilities are estimated simultaneously. Broto and Ruiz (2004) claim that MCMC approaches are more efficient among other estimation methods, such as the QML and the general method of moment by Melino and Turnbull (1990). As usual, in the MCMC methods for SV models, the posterior densities of the parameters and the augmented parameters are either simple density functions or proportional to some positive functions. These distributions generally cannot be sampled directly and therefore simulation is carried out by using Metropolis-Hastings

algorithm. It is well known that the performance of the Metropolis-Hastings algorithm depends on the selection of proposal density. Choosing an appropriate proposal density is difficult and different proposal densities yield different acceptance rates. Due to a high correlation of latent states, the simulation of the log volatilities is critical, which is discussed in Jacquier et al. (2004). There are two main methods for simulating the latent states. One is called the single-move simulation method developed in Jacquier et al. (1994, 2004), Yu et al. (2006), Zhang and King (2008) and Kim et al. (1998), in which the states are simulated one at a time. Another method, known as a block sampling, is introduced by Shephard and Pitt (1997) and has been employed in Pitt and Shephard (1999a) and Chib et al. (2006). In a block sampling algorithm, the latent states are divided into random blocks and the blocks are sampled via the Metropolis-Hastings algorithm. The proposal distribution is either a multivariate Gaussian as in Shephard and Pitt (1997) or a multivariate Student- t distribution as in Chib et al. (2006), where the modes of these proposal distributions can be found by the Newton-Raphson method. As discussed in Shephard and Pitt (1997), the Newton-Raphson algorithm may converge slowly. Given slow convergence of Newton-Raphson algorithm, the authors suggest using a pseudo dominating MCMC and only two or five iterations are needed for calculating the modes of the proposal distribution. Thus, the block sampling is relatively more time-consuming.

In this chapter, we mainly focus on the estimation of univariate SV models, where the measurement equation has either standard normal or Student- t observation errors. The non-zero correlation between the innovations of asset returns and the latent AR(1) process is permitted. The focus of this exercise is to consider simulation-based inference for the parameters of the model and its log volatilities. Our contributions are two-fold. First, we develop single-move MCMC algorithms for the heavy-tailed SV models in simulating latent states based on the Metropolis-Hastings algorithm, where an observation from the proposal distribution is simulated by the slice sampler intro-

duced in Edwards and Sokal (1988) and Neal (2003). For the SV models without heavy tails, single-move simulation methods based on the Metropolis-Hastings method are developed, where the proposal distribution is either a univariate normal distribution or a product of several positive functions that can be sampled by the slice sampler. The second contribution relates to the heavy-tailed SV models as well. compared with the method, where a mixture decomposition of the Student- t distribution (for examples, see Jacquier et al. (2004) and Zhang and King (2008)) is required at each observation time, our method uses the Student- t distribution directly. We do not need to estimate the extra parameters from the mixture decomposition. This methodology is studied via simulation in terms of parameter and volatility estimation and statistical test of model assumptions. Using the slice sampler within MCMC approach, the estimation of the heavy-tailed SV models is relatively straightforward.

The rest of the chapter is organized as follows. In Section 2.2, we propose a slice sampler within the MCMC approach for the heavy-tailed SV model and a simple Metropolis-Hastings method for the ASV model in the simulation of the latent states. To assess the overall model fit, in addition to checking the realized innovations of the observation equation, we test the so-called probability integral transforms (PITs) calculated from the density forecast introduced by Diebold et al. (1998). Since it is difficult to obtain the analytical conditional densities of observed data, we employ the auxiliary particle filter in Pitt and Shephard (1999b) to evaluate the likelihoods. Section 2.3 presents simulation studies for the ASV model. A comparison of our approaches with that of Jacquier et al. (1994) and Kim et al. (1998) is made on the same simulated return data. We illustrate by means of asset return data that the SV models without heavy tail can be fitted without the slice sampler but through a Metropolis-Hastings method. This is not the case for heavy-tailed SV models, where the slice sampler is necessary to generate a stationary time series from the full conditionals of parameters. In Section 2.4, we apply our estimation methods to the ASV and heavy-tailed ASV models and

several other competing SV models to a data set of the IBM stock returns and use the AIC in Akaike (1987) and the BIC in Schwarz (1978) for model comparison. Conclusions and remarks can be found in the last section.

2.2 Slice Sampler Within MCMC Algorithms for Univariate SV Models

In this section, the slice sampler within the MCMC approach for the ASV and heavy-tailed ASV models are discussed in detail. To assess the goodness of fit of the models, beside checking on the normality assumption of the realized observation errors, we also test the PITs produced by the estimated model. In order to evaluate the sample likelihood, calculate the PITs from the estimated model and perform the one-step ahead volatility forecast, the auxiliary particle filter is introduced and employed.

2.2.1 Estimation of the ASV Model

As discussed earlier, the ASV model allows for non-zero correlation between the two Gaussian noises of the BSV model formulated by equations (2.1) to (2.3). Define by $\Theta = (\mu, \phi, \rho, \sigma)$ the parameter of the ASV model, where $\rho = \text{corr}(\epsilon_t, \eta_{t+1})$ is the correlation coefficient, $\mathbf{h} = (h_1, \dots, h_T)'$ the vector of latent random variables (or the latent states) and $\mathbf{y} = (y_1, \dots, y_T)'$ the vector of observations. In the implementation of the MCMC algorithm, \mathbf{h} is augmented as a vector of parameters and will be estimated simultaneously. Let $f(\cdot)$ denote a generic density function. Applying Bayes' Theorem the joint conditional distribution of Θ and \mathbf{h} is

$$f(\Theta, \mathbf{h}|\mathbf{y}) \propto f(\mathbf{y}|\mathbf{h}, \Theta)f(\mathbf{h}|\Theta)f(\Theta), \quad (2.4)$$

where $f(y|\mathbf{h}, \Theta)$ is the likelihood of y given (Θ, \mathbf{h}) , $f(\mathbf{h}|\Theta)$ is the density of \mathbf{h} and $f(\Theta)$ is the prior density of Θ . Bayesian inference of (Θ, \mathbf{h}) is based upon the posterior distribution $f(\Theta, \mathbf{h}|y)$. Since the two errors ϵ_t and η_{t+1} are correlated, the ASV model can be written in an equivalent form (see Jacquier et al. (2004), Zhang and King (2008) and Omori et al. (2007)) as

$$y_t = \exp(h_t/2)\epsilon_t, \quad (2.5a)$$

$$h_{t+1} = \mu + \phi(h_t - \mu) + \rho\sqrt{\sigma} \exp(-h_t/2)y_t + \sqrt{\sigma(1 - \rho^2)}v_{t+1}, \quad (2.5b)$$

where $\epsilon_t \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1)$, $v_{t+1} \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1)$ and $\text{cov}(\epsilon_t, v_{t+1}) = 0$. The log volatility h_{t+1} is defined by the shock v_{t+1} , the correlation ρ and the return y_t . Following the re-parametrization of ρ and σ in Jacquier et al. (2004) and Zhang and King (2008), we denote $\psi = \sqrt{\sigma}\rho$ and $\tau = \sigma(1 - \rho^2)$ and obtain

$$y_t = \exp(h_t/2)\epsilon_t, \quad (2.6a)$$

$$h_{t+1} = \mu + \phi(h_t - \mu) + \psi \exp(-h_t/2)y_t + \sqrt{\tau}v_{t+1}. \quad (2.6b)$$

Instead of sampling ρ and σ directly, we now sample ψ and τ and then transform back by $\sigma = \psi^2 + \tau$ and $\rho = \psi/\sqrt{\sigma}$. With this transformation, we redefine $\Theta = (\mu, \phi, \psi, \tau)$.

The ASV model is completed by specifying informative prior distributions for the parameters. We assume that all parameters in Θ are mutually independent. The prior distributions of μ and ψ are chosen to be $\mu \sim \mathcal{N}(0, 10)$ and $\psi \sim \mathcal{N}(0, 10)$, respectively. These prior distributions result in reasonably flat densities over their support regions. To impose stationary condition on the latent process, the prior distribution of ϕ follows a univariate normal distribution $\phi \sim \mathcal{N}(0, 10)$ truncated in the interval $(-1, 1)$. The prior distribution of σ is an inverse Gamma distribution given by $\mathcal{IG}(2.5, 0.025)$, which was also used in Pitt and Shephard (1999a). It is seen that all prior distributions

except for ϕ are conjugate, which is convenient for the calculation of the posterior distributions or the full conditionals (note: the full conditional of a parameter is defined as the conditional distribution given that other parameters in the model have been previously sampled). The outline of the MCMC algorithm is listed in Table 2.1 followed by detailed explanations.

Table 2.1: MCMC algorithm for the ASV model.

Step 0. Initialize \mathbf{h} , μ , ϕ , ψ and τ .

Step 1. Sample $h_t, t = 1, \dots, T$.

Step 2. Sample ϕ .

Step 3. Sample ψ , μ and τ .

Step 4. Go to Step 1.

Step 0. Initialize \mathbf{h} , μ , ϕ , ψ and τ . To start the MCMC algorithm, the parameters of the latent Markov process are set as $\mu = -5.5$, $\phi = -0.5$, $\psi = -0.08$ and $\tau = 0.04$, respectively. The initial values of \mathbf{h} are generated from the distribution $\mathcal{N}(-5.5, 0.06)$, which is the equilibrium distribution of the latent states.

Step 1. Sample \mathbf{h} . The simulation is conducted via a single-move Metropolis-Hastings algorithm. The full conditionals of the latent random variables are expressed as follows:

$$\begin{aligned} f(h_1|\mathbf{y}, h_2, \Theta) &\propto f(y_1|h_1)f(h_1|\Theta)f(h_1|h_2, y_1, \Theta), \\ f(h_t|\mathbf{y}, h_{t-1}, h_{t+1}, \Theta) &\propto f(y_t|h_t)f(h_t|h_{t-1}, y_{t-1}, \Theta)f(h_t|h_{t+1}, y_t, \Theta), \\ f(h_T|\mathbf{y}, h_{T-1}, \Theta) &\propto f(y_T|h_T)f(h_T|h_{T-1}, y_{T-1}, \Theta), \end{aligned}$$

where $f(y_t|h_t), t = 1, \dots, T$, are the conditional densities of the asset returns at discrete

time points and $f(h_1|\Theta)$ is the density of the latent log volatility h_1 . $f(h_t|h_{t-1}, y_{t-1}, \Theta)$ and $f(h_t|h_{t+1}, y_t, \Theta)$ are the conditional densities of h_t given h_{t-1} and of h_t given h_{t+1} by the latent equation (2.6b), respectively. Since y_T is the last observation, the posterior distribution of h_T depends only on y_T, y_{T-1} and h_{T-1} .

Here we only present the simulation algorithm for the full conditionals of $h_t, t = 2, \dots, T - 1$.

$$\begin{aligned}
& f(h_t|\mathbf{y}, h_{t-1}, h_{t+1}, \Theta) \\
&= c_{1t} f(y_t|h_t) f(h_t|h_{t-1}, y_{t-1}, \Theta) f(h_t|h_{t+1}, y_t, \Theta) \\
&= c_{2t} \exp\left\{\frac{-h_t}{2}\right\} \exp\left\{-\frac{y_t^2}{2} \exp(-h_t)\right\} \\
&\quad \times \exp\left\{-\frac{[(h_t - \mu) - \phi(h_{t-1} - \mu) - \psi \exp(-h_{t-1}/2)y_{t-1}]^2}{2\tau}\right\} \\
&\quad \times \exp\left\{-\frac{[(h_{t+1} - \mu) - \phi(h_t - \mu) - \psi \exp(-h_t/2)y_t]^2}{2\tau}\right\} \\
&= c_{2t} \exp\left\{-\frac{[(h_{t+1} - \mu) - \phi(h_t - \mu) - \psi \exp(-h_t/2)y_t]^2}{2\tau}\right\} \times g(h_t), \tag{2.7}
\end{aligned}$$

where c_{1t} and c_{2t} are the two normalizing constants, and

$$\begin{aligned}
g(h_t) &= \exp\left\{\frac{-h_t}{2}\right\} \exp\left\{-\frac{y_t^2}{2} \exp(-h_t)\right\} \\
&\quad \times \exp\left\{-\frac{[(h_t - \mu) - \phi(h_{t-1} - \mu) - \psi \exp(-h_{t-1}/2)y_{t-1}]^2}{2\tau}\right\}. \tag{2.8}
\end{aligned}$$

The full conditionals of h_1 and h_T are given in the Appendix. The full conditional of h_t in (2.7) is not a simple distribution and therefore can not be simulated directly. It is noticed that $g(h_t)$ is a product of three positive functions, which can be sampled by the slice sampler introduced in Edwards and Sokal (1988) and Neal (2003). Thus, the Metropolis-Hastings method can be applied to simulate the full conditional of h_t with

proposal density proportional to $g(h_t)$.

Algorithm for the slice sampler

It is easy to verify that the proposal density in (2.8) can be expressed by

$$g(h_t) \propto \exp \left\{ \frac{-y_t^2}{2} \exp(-h_t) \right\} \exp \left\{ -\frac{(h_t - \mu_t)^2}{2\tau} \right\}, \quad (2.9)$$

where $\mu_t = \mu - \frac{\tau}{2} + \phi(h_{t-1} - \mu) + \psi \exp(-h_{t-1}/2)y_{t-1}$.

1. Initialize $h_t^{(0)}$ and set $n = 0$.

2. Draw u_1 uniformly from the interval $\left(0, \exp \left\{ \frac{-y_t^2}{2} \exp(-h_t^{(n)}) \right\}\right)$. Then we define an interval for h_t through the inequality

$$u_1 \leq \exp \left\{ \frac{-y_t^2}{2} \exp(-h_t) \right\},$$

which, under the condition $y_t \neq 0$, is equivalent to

$$h_t \geq -\log \left(\frac{-2 \log(u_1)}{y_t^2} \right). \quad (2.10)$$

3. Draw u_2 uniformly from the interval $\left(0, \exp \left\{ -\frac{(h_t^{(n)} - \mu_t^{(n)})^2}{2\tau} \right\}\right)$. Then we define an interval for h_t through the inequality

$$u_2 < \exp \left\{ -\frac{(h_t - \mu_t^{(n)})^2}{2\tau} \right\},$$

which leads to

$$\mu_t^{(n)} - \sqrt{-2\tau \log(u_2)} \leq h_t \leq \mu_t^{(n)} + \sqrt{-2\tau \log(u_2)}. \quad (2.11)$$

4. If $y_t \neq 0$, draw $h_t^{(n+1)}$ uniformly from the interval determined by the inequalities

(2.10) and (2.11):

$$h_t^{(n+1)} \sim \mathcal{U}\left(\max\left\{-\log\left(\frac{-2\log(u_1)}{y_t^2}\right), \mu_t^{(n)} - \sqrt{-2\tau\log(u_2)}\right\}, \mu_t^{(n)} + \sqrt{-2\tau\log(u_2)}\right).$$

Otherwise,

$$h_t^{(n+1)} \sim \mathcal{U}\left(\mu_t^{(n)} - \sqrt{-2\tau\log(u_2)}, \mu_t^{(n)} + \sqrt{-2\tau\log(u_2)}\right).$$

5. Stop if a stopping criterion is met; otherwise, set $n = n + 1$ and repeat from 2.

In the slice sampling procedure, two auxiliary uniform random variables u_1 and u_2 over the interval $(0, 1)$ were introduced. Once $h_t^{(n+1)}$ is sampled, the two values simulated from the auxiliary variables are ignored. To start the slice sampling procedure, the sampled value of h_t from the last MCMC step is set as the initial value. Since the distributions of the full conditionals of h_t in each MCMC step are similar, this initial value is a good starting point. Because the slice sampler adapts to the analytical density function of the underlying random variable, it is more efficient. For instance, the target distribution can be sampled directly with the help of auxiliary random variable(s). Under some sufficient conditions, Roberts and Rosenthal (1999) show that the slice algorithm is robust and has extremely robust geometric ergodicity properties. Mira and Tierney (2002) prove that the slice sampler has a smaller second-largest eigenvalue, which ensures faster convergence to the underlying distribution. Typically the proposed slice sampler is iterated for five times when sampling the full conditional of h_t in our MCMC algorithm.

It is noticed that the second part of the proposal density $g(h_t)$ defined in (2.9) is a univariate normal density. Then the proposal distribution could be a univariate normal distribution. That is, the slice sampler for simulating the states may not be needed. Our

experience shows that using the slice sampler, the acceptance rates of the Metropolis-Hastings method are higher but the running speed of the algorithm is slower. We will compare this in Section 2.3.3.

Step 2. Sample ϕ . Given a truncated prior distribution $\phi \sim \mathcal{N}(\alpha_\phi, \beta_\phi^2)$, the full conditional of ϕ is

$$\begin{aligned} f(\phi|\mathbf{y}, \mu, \psi, \tau) &\propto p(h_1|\Theta) \prod_{t=1}^{T-1} p(h_{t+1}|h_t, \Theta, y_t) \exp\left\{-\frac{(\phi - \alpha_\phi)^2}{2\beta_\phi^2}\right\} \\ &\propto \mathcal{N}\left(\phi; \frac{d}{c}, \frac{1}{c}\right) (1 - \phi^2)^{\frac{1}{2}}, |\phi| < 1, \end{aligned}$$

where

$$\begin{aligned} c &= \frac{-(h_1 - \mu)^2 + \sum_{t=1}^{T-1} (h_t - \mu)^2}{\tau} + \frac{1}{\beta_\phi^2}, \\ d &= \frac{\sum_{t=1}^{T-1} (h_t - \mu)(h_{t+1} - \mu - \psi \exp(-h_t/2)y_t)}{\tau} + \frac{\alpha_\phi}{\beta_\phi^2}. \end{aligned}$$

The full conditional is detailed in the Appendix. It is proportional to the product of a univariate normal distribution and a positive function, and so can be sampled by the slice sampler or acceptance-rejection method.

Step 3. Sample parameters μ, ψ and τ . Since the priors for these parameters are conjugate, the full conditionals are normal and inverse Gamma distributions, respectively. Those full conditionals are presented in the Appendix, and can be easily simulated.

We have proposed an MCMC method for fitting the ASV model, where the latent states are sampled one at a time. In the literature, the single-move simulation approaches for sampling the latent states are popular. See for instance Jacquier et al. (1994, 2004), Yu et al. (2006), and Zhang and King (2008), among others. In spite of high autocorrelation between the latent states, our experience shows that for the ASV model and the heavy-tailed ASV model (to be introduced in the next subsection) our

simulation methods work very well. As clarified in Mira and Tierney (2002), the slice sampler converges faster than the independent Metropolis-Hastings method. The advantage of the slice sampler is that after a few iterations, a sampled point may be obtained from the target distribution. In applications of the slice sampling, we obtained very similar results even when only one iteration of the slice sampler algorithm was used. This is unlike the Metropolis-Hastings algorithm, where many generated points have to be discarded.

2.2.2 Estimation of a Heavy-tailed ASV (ASV-t) Model

Assuming the innovations ϵ_t in the ASV model follow a Student- t distribution with v degrees of freedom, a heavy-tailed ASV model is specified as

$$y_t = \exp(h_t/2)\epsilon_t, \quad \epsilon_t \sim t(v), \quad (2.12)$$

$$h_{t+1} = \mu + \phi(h_t - \mu) + \sqrt{\sigma}u_{t+1}, \quad u_{t+1} \sim \mathcal{N}(0, 1). \quad (2.13)$$

In the literature, the Student- t distribution is often decomposed as $\epsilon_t = \sqrt{\lambda_t}e_t$, where $\lambda_t \sim \mathcal{IG}(v/2, v/2)$, an inverse Gamma distribution, and $e_t \sim \mathcal{N}(0, 1)$ as, for instance, in Harvey and Shephard (1996), Kim et al. (1998), Chib et al. (2002), Jacquier et al. (2004), Shimada and Tsukuda (2005), Chib et al. (2006) and Zhang and King (2008). The correlation between the two equations is then defined as $\rho = \text{corr}(e_t, u_{t+1})$, which is the correlation of the two standard normal distributions.

In the above heavy-tailed SV models, the inverse Gamma random variables λ_t have to be estimated during the MCMC estimation procedure. Instead of a mixture decomposition, we introduce a type of heavy-tailed SV model that uses a Student- t distribu-

tion directly. The model is defined as follows

$$y_t = \exp(h_t/2)\epsilon_t, \quad \epsilon_t \sim t(v), \quad (2.14)$$

$$h_{t+1} = \mu + \phi(h_t - \mu) + \psi\epsilon_t + \sqrt{\tau}u_{t+1}, \quad u_{t+1} \sim \mathcal{N}(0, 1), \quad (2.15)$$

where ϵ_t and u_{t+1} are i.i.d innovations and ϵ_t and u_{t+1} are mutually independent. Because of the presence of ϵ_t in (2.15), the latent process is effectively correlated with the asset return process. The innovation of the latent process is driven by a mixture of a Student- t distribution and a univariate normal distribution. We expect that in applications the estimate of ψ will be negative, indicating that the observed asset return process and the latent AR(1) process are negatively correlated. Based on this consideration, the proposed model will be called the ASV-t model hereafter. It is noticed that this model can be rewritten equivalently as

$$y_t = \exp(h_t/2)\epsilon_t, \quad \epsilon_t \sim t(v), \quad (2.16)$$

$$h_{t+1} = \mu + \phi(h_t - \mu) + \psi y_t \exp(-h_t/2) + \sqrt{\tau}u_{t+1}, \quad u_{t+1} \sim \mathcal{N}(0, 1), \quad (2.17)$$

which is similar to the ASV model with a Cholesky decomposition except here we assume that the innovation of the measurement equation has a Student- t distribution. In our ASV-t model, there is only one extra parameter v to be estimated when compared with the ASV model, while in the mixture case we have to estimate v together with T augmented parameters λ_t . Again, we propose a slice sampler within the MCMC algorithm to fit the ASV-t model. The algorithms for simulating $(\mu, \phi, \psi, \tau)'$ are the same as those for the ASV model. Therefore, below we only describe methods to simulate the latent states h_t and v .

- Sample the latent states $h_t, t = 1, \dots, T - 1$. The simulation of h_1 and h_T are similar.

The full conditionals of $h_t, t = 2, \dots, T - 1$, is

$$\begin{aligned}
& f(h_t | \mathbf{y}, h_{t-1}, h_{t+1}, \Theta) \\
&= c_{1t} f(y_t | h_t) f(h_t | h_{t-1}, y_{t-1}, \Theta) f(h_t | h_{t+1}, y_t, \Theta) \\
&= c_{2t} e^{-h_t} \left(1 + \frac{y_t^2 e^{-h_t}}{v} \right)^{-\frac{v+1}{2}} \\
&\quad \times \exp \left\{ -\frac{[(h_t - \mu) - \phi(h_{t-1} - \mu) - \psi \exp(-h_{t-1}/2)y_{t-1}]^2}{2\tau} \right\} \\
&\quad \times \exp \left\{ -\frac{[(h_{t+1} - \mu) - \phi(h_t - \mu) - \psi \exp(-h_t/2)y_t]^2}{2\tau} \right\} \\
&= c_{2t} \exp \left\{ -\frac{[(h_{t+1} - \mu) - \phi(h_t - \mu) - \psi \exp(-h_t/2)y_t]^2}{2\tau} \right\} \times g(h_t), \tag{2.18}
\end{aligned}$$

where c_{1t} and c_{2t} are the two normalizing constants, and

$$\begin{aligned}
g(h_t) &= e^{-h_t} \left(1 + \frac{y_t^2 e^{-h_t}}{v} \right)^{-\frac{v+1}{2}} \\
&\quad \times \exp \left\{ -\frac{[(h_t - \mu) - \phi(h_{t-1} - \mu) - \psi \exp(-h_{t-1}/2)y_{t-1}]^2}{2\tau} \right\}. \tag{2.19}
\end{aligned}$$

Similar to the simulation algorithm for the ASV model, the full conditional of h_t in (2.18) can not be sampled from a known distribution. We use a Metropolis-Hastings method with proposal distribution proportional to $g(h_t)$, where $g(h_t)$ can be sampled via the slice sampler. The slice sampler for this proposal distribution can be found in the Appendix.

We notice that a proposal distribution for the full conditional of h_t in (2.18) can also be written as

$$g(h_t) = c_{2t} e^{-h_t} \left(1 + \frac{y_t^2 e^{-h_t}}{v} \right)^{-\frac{v+1}{2}}. \tag{2.20}$$

To sample this proposal distribution, we first sample a Student- t distribution with v degrees of freedom. By variable transformation, a sample point from the proposal distribution is obtained, which is accepted with a probability calculated through the Metropolis-Hastings step. In Section 2.3.3, we will compare the use of (2.19) and (2.20) and show (2.20) has lower acceptance rate for proper mixing.

- Sample v . The full conditional of v is given by

$$\begin{aligned} f(v|\mathbf{y}, \mathbf{h}, \mu, \phi, \sigma^2) &\propto f(\mathbf{y}|\mathbf{h}, v)f(v) \\ &= f(v) \prod_{t=1}^T \frac{v^{v/2}\Gamma((v+1)/2)}{\Gamma(v/2)\Gamma(1/2)} (v + y_t^2 \exp(-h_t))^{-(v+1)/2}, \end{aligned} \quad (2.21)$$

where $f(v)$ is a prior density of v . In the literature, there are several ways to specify this prior distribution. For instance, Jacquier et al. (2004) propose a discrete prior distribution $\mathcal{U}[3, 40]$ from which the full conditional can be sampled directly from a multinomial distribution. Geweke (1993) suggests $\alpha \exp(-\alpha v)$ with $\alpha = 0.2$ as an alternative, while Zhang and King (2008) choose a normal distribution $v \sim \mathcal{N}(20, 25)$. Bauwens and Lubrano (1998) use a Cauchy prior proportional to $1/(1 + v^2)$. In our procedure we use a normal prior proposed in Zhang and King (2008). Since this full conditional is an unknown distribution, we use a random-walk Metropolis-Hastings algorithm, in which the proposal density is a standard Gaussian density and the acceptance probability is computed using equation (2.21).

2.2.3 Particle Filter

To perform model comparison, we use the AIC and BIC, which require that we evaluate the sample likelihood approximately. For the ASV and ASV-t models, the likelihoods are very difficult to compute due to their non-linear structure. We employ the auxiliary particle filter in Pitt and Shephard (1999b) to perform this task, which is a

recursive efficient algorithm to approximate the filter and one-step ahead predictive distributions. The likelihood of the specific SV model via a successive conditional decomposition is

$$f(\mathbf{y}|\Theta) = f(y_1|\Theta) \prod_{t=2}^T f(y_t|\mathcal{F}_{t-1}, \Theta), \quad (2.22)$$

where $\mathcal{F}_t = (y_1, \dots, y_t)$ is the information known at time t . The conditional density of y_{t+1} given Θ and \mathcal{F}_t has the following expression

$$\begin{aligned} f(y_{t+1}|\mathcal{F}_t, \Theta) &= \int f(y_{t+1}|h_{t+1}, \Theta) dF(h_{t+1}|\mathcal{F}_t, \Theta) \\ &= \int f(y_{t+1}|h_{t+1}, \Theta) f(h_{t+1}|h_t, \Theta) dF(h_t|\mathcal{F}_t, \Theta). \end{aligned} \quad (2.23)$$

In general, it is impossible to have an analytical solution of (2.23); instead numerical methods such as the auxiliary particle filter will have to be employed. Suppose that we have a particle sample $\{h_t^{(i)}, k = 1, \dots, N\}$ of h_t from a filtered distribution of $(h_t|\mathcal{F}_t, \Theta)$ with weights $\{\pi_i, k = 1, \dots, N\}$ such that $\sum_{i=1}^N \pi_i = 1$. Based on this sample, the one-step ahead predictive density of h_{t+1} is

$$f(h_{t+1}|\mathcal{F}_t, \Theta) \approx \sum_{i=1}^N \pi_i f(h_{t+1}|h_t^{(i)}, \Theta). \quad (2.24)$$

Then the one-step ahead prediction distribution of h_{t+1} can be sampled and the conditional density (2.23) can be evaluated numerically by

$$f(y_{t+1}|\mathcal{F}_t, \Theta) \approx \sum_{i=1}^N \pi_i f(y_{t+1}|h_{t+1}^{(i)}, \Theta), \quad (2.25)$$

where $h_{t+1}^{(i)}$ are particles from the prediction distribution of $(h_{t+1}|\mathcal{F}_t, \Theta)$. The assumption for the above evaluations (2.24) and (2.25) to be valid is that the prediction density of h_{t+1} is at least approximately known. This assumption is satisfied since from the latent AR(1) process, h_{t+1} has a conditional normal distribution $h_{t+1} \sim \mathcal{N}(\mu + \phi(h_t - \mu) + \psi \exp(-h_{t-1}/2)y_{t-1}, \tau)$, which can also be used for volatility forecast.

Now the question is how to sample $(h_{t+1}|\mathcal{F}_{t+1}, \Theta)$ given that we have a particle sample from the filter distribution of $(h_t|\mathcal{F}_t, \Theta)$. We present the algorithm for the ASV and ASV-t models based on the procedure in Chib et al. (2006).

Step 1. Given a sample $\{h_t^{(i)}, i = 1, \dots, N\}$ from $(h_t|\mathcal{F}_t, \Theta)$, we calculate the expectation $\hat{h}_{t+1}^{*(i)} = E(h_{t+1}|h_t^{(i)})$ and

$$\pi_i = f(y_{t+1}|\hat{h}_{t+1}^{*(i)}, \Theta), i = 1, \dots, N. \quad (2.26)$$

Sample N times with replacement the integers of $1, \dots, N$ with probability $\hat{\pi}_i = \pi_i / \sum_{i=1}^N \pi_i$. Denote the number of occurrences of the indices $1, \dots, n$ in one sample by n_1, \dots, n_N and associate these with particles $\{\mathbf{h}_t^{(n_1)}, \dots, \mathbf{h}_t^{(n_N)}\}$.

Step 2. For each value n_i from Step 1, sample the values $\{h_{t+1}^{*(1)}, \dots, h_{t+1}^{*(N)}\}$ from

$$h_{t+1}^{*(i)} = \mu + \phi(h_t^{(n_i)} - \mu) + \psi \exp(-h_t^{(n_i)}/2)y_t + \sqrt{\tau}v_{t+1}, i = 1, \dots, N, \quad (2.27)$$

where $v_{t+1} \sim \mathcal{N}(0, 1)$.

Step 3. Calculate the weights of the values $\{h_{t+1}^{*(1)}, \dots, h_{t+1}^{*(N)}\}$ as

$$\pi_i^* = \frac{f(y_{t+1}|h_{t+1}^{*(i)}, \Theta)}{f(y_{t+1}|\hat{h}_{t+1}^{*(i)}, \Theta)}, i = 1, \dots, N, \quad (2.28)$$

and resample the values $\{h_{t+1}^{*(1)}, \dots, h_{t+1}^{*(N)}\}$ N times with replacement using these weights to obtain a fair sample $\{h_{t+1}^{(1)}, \dots, h_{t+1}^{(N)}\}$ with weights $1/N$ from the filter distribution of

$(h_{t+1}|\mathcal{F}_{t+1}, \Theta)$.

In our experience, $N = 3000$ is sufficient for our simulation studies and the real stock return data that we use to illustrate our estimation methods.

2.2.4 Diagnostics

There are many techniques to check the overall fit of a specific ASV or ASV-t model. One method is a Kolmogorov-Smirnov test that can be used to examine whether the realized observation errors come from the assumed distribution. Another method is to analyze the PITs proposed in Diebold et al. (1998). We discuss this method below.

Suppose that $\{f(y_t|\mathcal{F}_{t-1})\}_{t=1}^T$ is a sequence of conditional densities that guides the time series of y_t . Let $\{p(y_t|\mathcal{F}_{t-1})\}_{t=1}^T$ be the corresponding sequence of one-step ahead density forecasts. The PIT of y_t is defined as

$$u(t) = \int_{-\infty}^{y_t} p(z|\mathcal{F}_{t-1})dz. \quad (2.29)$$

Under the hypothesis that the sequence $\{p(y_t|\mathcal{F}_{t-1})\}_{t=1}^T$ coincides with $\{f(y_t|\mathcal{F}_{t-1})\}_{t=1}^T$, the sequence $\{u(t)\}_{t=1}^T$ is independent and identically distributed (i.i.d.) Uniform $[0, 1]$. In our univariate ASV and ASV-t models, the PITs can be calculated by the formulas below.

$$u(t) \approx \frac{1}{N} \sum_{i=1}^N \int_{-\infty}^{y_t} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2} e^{-h_t^{(i)}} - \frac{h_t^{(i)}}{2}\right) dz, \quad (2.30)$$

in the ASV model, and

$$u(t) \approx \frac{1}{N} \sum_{i=1}^N \int_{-\infty}^{y_t} \frac{\Gamma(\frac{v+1}{2})}{\Gamma(\frac{v}{2})} \frac{1}{\sqrt{v\pi}} \left(1 + \frac{z^2 \exp(-h_t^{(i)})}{v}\right)^{-\frac{v+1}{2}} dz, \quad (2.31)$$

in the ASV-t model. In the computation of $u(t)$, $h_t^{(i)}$ are particles from the corresponding

predictive distribution of h_t with weights $1/N$.

It is noticed that the PITs can only be calculated from the fitted SV models indexed by the estimated parameters, since the true values of the parameters are not known. As remarked in Diebold et al. (1998), when density forecast analysis is applied in real applications, the parameter estimation uncertainty is often ignored. Clements and Smith (2000) discuss this issue and suggest using the parameter estimates as the population values for generating the forecasts. In this and the following chapters, when conducting simulation studies and analyzing real data, we follow these ideas to assess the goodness-of-fit for univariate and multivariate SV models. We use the Kolmogorov-Smirnov test to assess the normality assumption of the realized observation errors and the uniform distribution $\mathcal{U}[0, 1]$ for PITs produced by the fitted SV models. Because our Kolmogorov-Smirnov test is based on the fitted model, the assessment of the overall model fit should not rely only on the test results. Instead of using this test alone, we also compare theoretical cumulative distribution functions (CDFs) with the empirical CDFs of the realized observation errors and the PITs.

2.3 Simulation Studies and Methodology Comparison Based on Real Data

2.3.1 Simulation Studies for the ASV Model

In this subsection, we conducted simulation studies for the ASV model. Here we only present the results when the proposed slice sampler within the MCMC method was applied to an artificially generated data set of asset returns. To generate asset returns,

we make a transformation to the innovation ϵ_t of the ASV model and obtain

$$y_t = \exp(h_t/2)(\eta_{t+1} + \sqrt{1 - \rho^2}v_t), \quad (2.32)$$

$$h_{t+1} = \mu + \phi(h_t - \mu) + \sqrt{\sigma}\eta_{t+1}, \quad (2.33)$$

where v_t and η_{t+1} are independent and i.i.d. with $v_t \sim \mathcal{N}(0, 1)$ and $\eta_{t+1} \sim \mathcal{N}(0, 1)$. For given Θ , the following equations will be used to generate \mathbf{h} and \mathbf{y} .

$$h_{t+1} \sim \mathcal{N}(\mu + \phi(h_t - \mu), \sigma), \quad (2.34)$$

$$y_t \sim \mathcal{N}(\exp(h_t/2)\rho(h_{t+1} - \mu - \phi(h_t - \mu))/\sqrt{\sigma}, \exp(h_t)(1 - \rho^2)), \quad (2.35)$$

where $h_1 \sim \mathcal{N}(\mu, \sigma/(1 - \phi^2))$ and $y_T \sim \mathcal{N}(0, \exp(h_T))$.

The parameters used to generate the asset returns were set in the second column of Table 2.2. We generated 2000 observations from the ASV model. The prior distribution of v is $v \sim \mathcal{N}(20, 25)$, which is the same as in Zhang and King (2008). Our proposed slice sampler within MCMC algorithm was iterated 50,000 iterations and the first 10,000 sampled points were discarded as the burn-in prior to conducting Bayesian inference. Figure 2.1 graphs the time series of the first 5000 sampled points from the full conditionals of parameters. These time series appear to converge in distribution quickly, certainly in less than 10,000 iterations, indicating that a short burn-in period is necessary before convergence. As we mentioned when we derived the MCMC algorithm, all the slice samplers for the latent states were iterated for five iterations. In Figure 2.2 we plot the histograms and sample paths of parameters after the burn-in, while Table 2.2 includes summaries in terms of standard errors and Bayesian highest probability density (HPD) confidence intervals for the parameters. The estimated parameters are close to their true values with relatively small standard errors.

To assess the overall model fit, we first check the normality assumption of the mea-

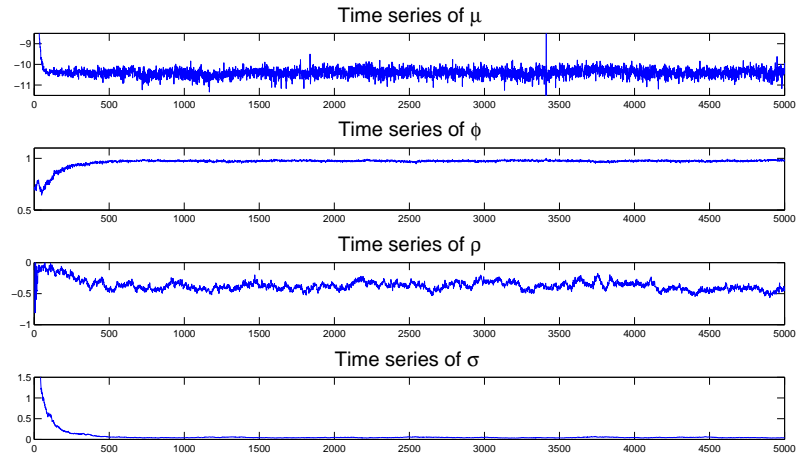


Figure 2.1: Time series of the first 5000 sampled points from the full conditionals of parameters in the ASV model based on the generated return data.

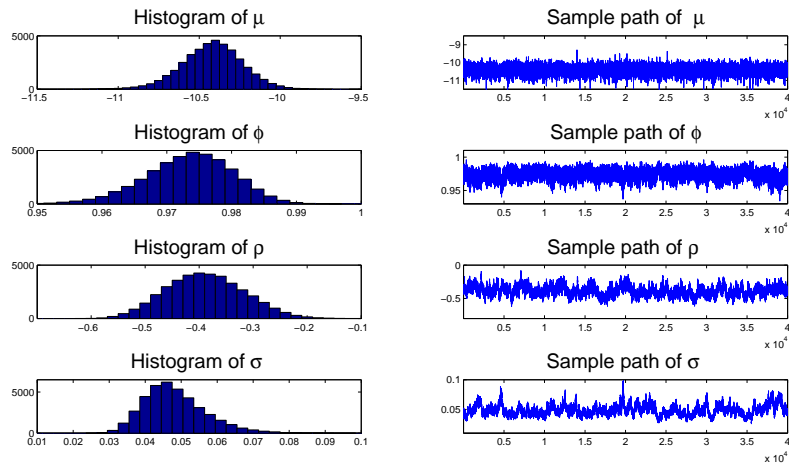


Figure 2.2: Histograms and dynamics of the samples from the full conditionals of parameters in the ASV model based on the generated return data.

Table 2.2: True and estimated parameters of the ASV model based on the simulated return data.

Parameter	True	Est.	Std.	HPD CI(95%)
μ	-10.45	-10.42	0.14	(-10.70, -10.15)
ϕ	0.98	0.98	0.01	(0.97, 0.99)
ρ	-0.41	-0.39	0.06	(-0.52, -0.27)
σ	0.036	0.048	0.009	(0.033, 0.066)

surement equation. The realized observation errors can be calculated using the following formula

$$\hat{\epsilon}_t = y_t \exp(-\hat{h}_t/2), t = 1, \dots, T, \quad (2.36)$$

where \hat{h}_t are the estimated log volatilities. Intuitively, we plot and compare the theoretical CDF of the standard normal distribution with the empirical CDF of the realized innovations in Figure 2.3. The two CDFs appear to be close, which means that the realized innovations are close to the standard normal distribution of the actual innovations. For a statistical test of the normality assumption, we perform a Kolmogorov-Smirnov test. Although the realized innovations came from the fitted ASV model, the applicability of the Kolmogorov-Smirnov test may still be justified (e.g., Diebold et al. (1998) and Clements and Smith (2000)). The computed Kolmogorov-Smirnov statistic is 0.0239 with the critical value 0.0303 at the 5% significance level. So we do not reject the null hypothesis at the 5% significance level that the realized innovations follow a standard normal distribution.

We compare the absolute values of the simulated asset returns y , the simulated volatilities $\exp(-h_t/2), t = 1, \dots, T$, with the estimated and forecasted volatilities depicted in Figure 2.4. The Bayesian estimated volatilities can be calculated by the following formula

$$\hat{V}_t = \frac{1}{N-n} \sum_{k=n+1}^N \exp(h_t^{(k)}/2), \quad (2.37)$$

where $h_t^{(k)}$ is the k -th simulated value of h_t and n is the size of the burn-in. The estimated and forecasted volatilities appear to visually resemble the simulated volatilities. An alternative comparison is to check the ratios between the true and estimated volatilities with the ratios between the true and forecasted volatilities graphed in Figure 2.5

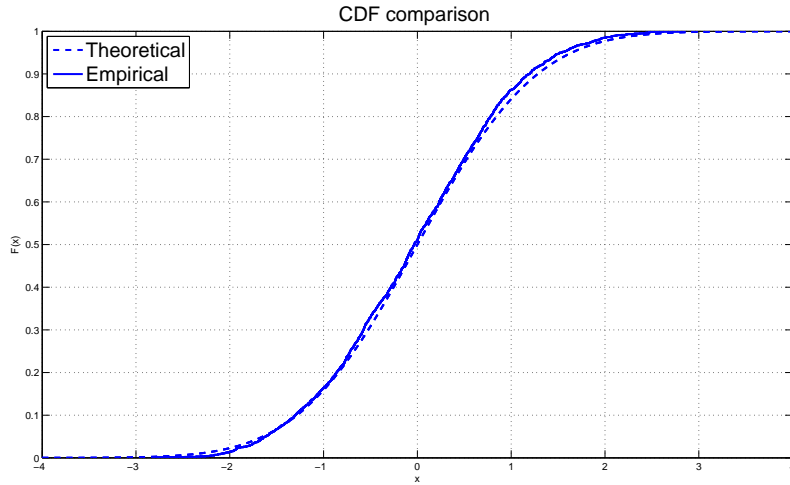


Figure 2.3: Comparison between the theoretical CDF of the standard normal distribution and empirical CDF of the residuals after fitting the ASV model to the generated returns.

with root mean square errors (RMSE). These ratios vary about the value of 1, indicating that the estimated and forecasted volatilities “agree” with the true volatility.

The overall model fit can also be assessed through the analysis of the PITs obtained from the formula (2.30). The uniform distribution of $u(t)$ on the interval $[0,1]$ is visualized in Figure 2.6 by scatter and histogram plots. The Kolmogorov-Smirnov test statistic is 0.0204 with the critical value 0.0303 at the 5% significance level, upon which we do not reject the null hypothesis at the 5% significant level that the PITs are Uniform $[0,1]$. The empirical CDF of the PITs is depicted together with a theoretical CDF of the Uniform $[0,1]$ in Figure 2.7. From the above comparisons and the Kolmogorov-Smirnov tests, we may conclude that the proposed slice sampler within MCMC approach for the ASV model fits the simulated asset return data reasonably well. The comparison with other MCMC algorithms will be seen in the next subsection.

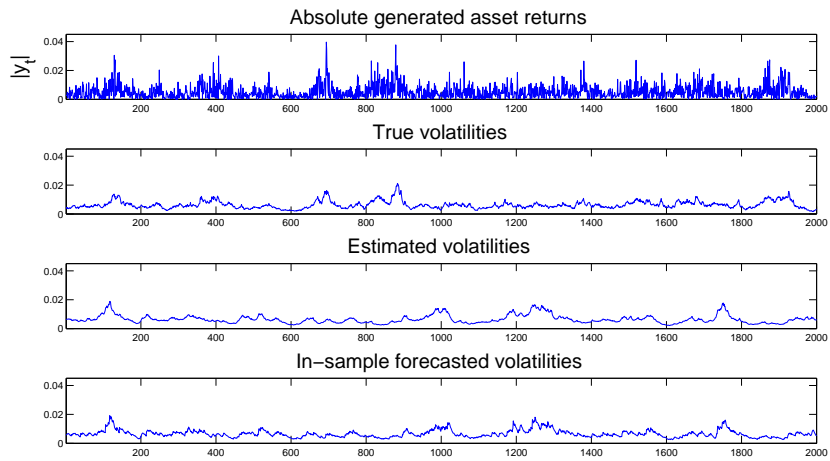


Figure 2.4: Comparison between the absolute returns and the true volatilities with the estimated and one-step ahead forecasted volatilities under the ASV model based on the generated asset return data.

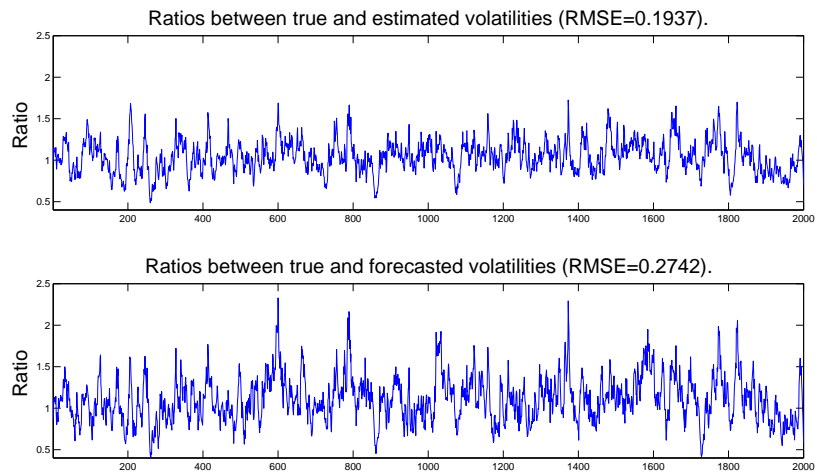


Figure 2.5: Ratio comparison between true and estimated and forecasted volatilities based on the generated asset return data.

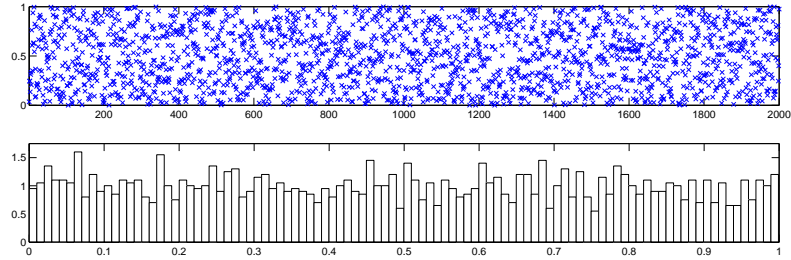


Figure 2.6: The top panel shows the scatter plot of $u(t)$ while the bottom the histogram of $u(t)$.

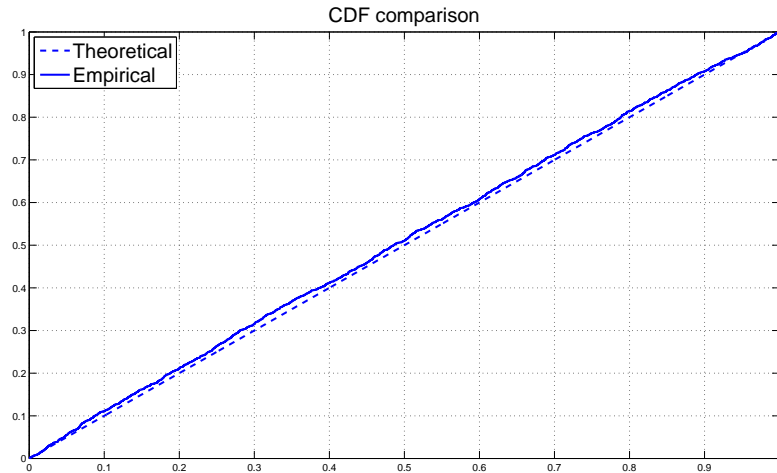


Figure 2.7: Comparison between the theoretical uniform CDF and the empirical CDF of the PITs from generated return data.

2.3.2 Comparison Between Three Single-move Estimation Methods

We compare our approach with two other single-move methods used in the literature by simulations. This comparison was conducted under the BSV model, where the correlation between the two innovations in the ASV model is zero. The first method, proposed in Jacquier et al. (2004) (called JPR hereafter), is a Metropolis-Hastings within the MCMC method, which uses a non-Gaussian proposal density based on an approximation of the full conditional of h_t . The other is derived in Kim et al. (1998), where a

full conditional of h_t is simulated via a Metropolis-Hastings method with a Gaussian proposal obtained through a Taylor expansion. We generated 100 data sets each with 2000 observations from the BSV model indexed with the same given parameters. Each data set was fitted by our slice sampler within the MCMC method and the approaches in Jacquier et al. (2004) and Kim et al. (1998), respectively. The estimated parameters are summarized in Tables 2.3 to 2.5. It is found that the estimated parameters from our approach and those by Kim et al. (1998) have a shorter confidence interval, while the estimated volatility and location parameters of the latent AR(1) process based on the JPR are not contained in the corresponding confidence intervals. This illustrates that at least for the BSV model, the JPR may not be efficient. The reason behind this phenomena is possibly due to the inefficient sampling of the log volatilities caused by either a high autocorrelation of latent states, which is discussed in Kim et al. (1998), or inappropriate proposal distributions. It is seen that the confidence intervals from the method in Kim et al. (1998) are even shorter than those from our method. This does not mean that our method is not comparable to their method. It is noticed that for the ASV and ASV-t models, the single-move acceptance-rejection simulation method of Kim et al. (1998) is not applicable, where a block sampling algorithm has to be used for the simulation of the latent states. As discussed in Section 2.1, to deal with these leveraged SV models, the block sampling algorithm is employed.

Table 2.3: True and estimated parameters of the BSV model via the slice sampler within MCMC method.

Parameter	True	Est.	Std.	HPD CI(95%)
μ	-10.45	-10.39	0.24	(-10.80, -9.98)
ϕ	0.98	0.98	0.01	(0.96, 0.99)
σ	0.036	0.036	0.009	(0.018, 0.054)

Table 2.4: True and estimated parameters of the BSV model through the JPR approach.

Parameter	True	Est.	Std.	HPD CI(95%)
μ	-10.45	-11.14	0.23	(-11.58, -10.69)
ϕ	0.98	0.97	0.01	(0.96, 0.99)
σ	0.036	0.055	0.009	(0.037, 0.073)

Table 2.5: True and estimated parameters of the BSV model via the method in Kim et al. (1998).

Parameter	True	Est.	Std.	HPD CI(95%)
μ	-10.45	-10.41	0.20	(-10.78, -10.04)
ϕ	0.98	0.98	0.01	(0.96, 0.99)
σ	0.036	0.035	0.008	(0.019, 0.048)

2.3.3 Comparison in Sampling the Latent States with and without the Slice Sampler

We noticed earlier that, in the algorithms for the ASV and ASV-t models, the slice sampler for simulating the latent states may not be necessary. In the estimation algorithm for the ASV model, the proposal distribution of the full conditional of h_t could be

$$g(h_t) = \frac{1}{\sqrt{2\pi\tau}} \exp \left\{ -\frac{(h_t - \mu_t)^2}{2\tau} \right\}. \quad (2.38)$$

This leads to a Metropolis-Hastings simulation method for the full conditional of h_t with a univariate normal distribution as the proposal distribution.

To check whether the slice sampler can actually affect the fit of the SV model, we consider a data set of daily pound/dollar exchange rate from 01/10/1981 to 28/06/1985 using the BSV and ASV models. The sample size is 945. This data set was also analyzed in Harvey et al. (1994), Shephard and Pitt (1997), Kim et al. (1998), Meyer and Yu (2000), Skaug and Yu (2007) and Huang and Yu (2008). In Kim et al. (1998), the BSV model was fitted by an MCMC method. The latent states were simulated one at a time

by using an acceptance-rejection method, where the dominating density is a univariate normal density obtained through a Taylor's expansion. We name this method the BSV-K method. Huang and Yu (2008) consider the BSV model but using an importance sampler based on Laplace approximation, which is called the BSV-Y method. The BSV model was also fitted by applying our MCMC methods with and without the slice sampler, which are named as the BSV-S and BSV-N methods, respectively. Similarly, ASV-S and ASV-N are the estimation methods for the ASV model with and without the slice sampler simulation, respectively.

When the data was fitted, each algorithm was iterated 50,000 times on the computing server of the University of Waterloo, and the last 40,000 sampled points were used for Bayesian inference. As we mentioned earlier, in all of the estimation procedures, if the slice sampler was used, it was iterated for five times. The estimated parameters of the SV models are included in Table 2.6, where the last row contains the CPU time in seconds. Since the BSV-Y method is not an MCMC algorithm, the running speed is not given here.

Table 2.6: Estimated parameters of the SV models under various MCMC methods for daily observations of weekday close exchange rates for the U.K. Sterling/U.S. Dollar exchange rate from 1/10/81 to 28/6/85.

Parameter	BSV-K	BSV-Y	BSV-S	BSV-N	ASV-S	ASV-N
μ	-0.8724	-0.9128	-0.8523	-0.8615	-0.8460	-0.9061
ϕ	0.9797	0.9734	0.9804	0.9821	0.9856	0.9825
σ	0.1479	0.1687	0.1485	0.1418	0.1429	0.1503
ρ					-0.1222	-0.0955
Duration	56		314	54	335	92

It is not surprising that the estimated parameters are roughly the same across the SV models and the estimation methods. But the running speeds are quite different. For the BSV model, the BSV-N method is about 6 time faster than the BSV-S method, while the ASV-N model is about 4 time faster than the BSV-S model. Hence, we can conclude

that with the slice sampler, the running time is much longer than the approaches using a simple Metropolis-Hastings method although with the former, we can have higher acceptance rate.

For the ASV-t model a slice sampler within the MCMC approach has been proposed for simulating latent states. Simulation studies illustrated that this method works well on parameter estimation. As mentioned earlier, the sampling algorithm for latent states may be further simplified. The proposal density for the full conditional of h_t can also be expressed as

$$g(h_t) = c_{2t} e^{-h_t} \left(1 + \frac{y_t^2 e^{-h_t}}{v} \right)^{-\frac{v+1}{2}}. \quad (2.39)$$

If we can simulate this proposal distribution easily, the proposed MCMC method for the heavy-tailed SV model would be faster. We know that $y_t \exp(-h_t/2)$ follows a student- t distribution $t(v)$. To obtain a sample point of h_t from the proposal distribution $g(h_t)$, the Student- t distribution is first sampled, then the sampled value is transformed to a candidate value. This value is accepted with a probability calculated from a Metropolis-Hastings method using (2.14). We found that the acceptance rates for this simulation method based on the ASV-t and heavy-tailed BSV (BSV-t) models, plotted in Figure 2.8, are less than 10%, have very small acceptance probabilities to accept a point generated from the proposal distribution. The reason behind these low acceptance rates probably came from the proposal distribution since only the parameter v was used. A direct consequence of this is that the simulated time series from the full conditionals of latent states may not mix well. This may cause the samples from full conditionals of latent states to be unreliable, which will affect the estimates of the static parameters in the latent log volatility process. By checking the sampled time series from the full conditional of v , we found that the time series are not ergodic showing a slow downside trend. Consequently we did not use the simple Metropolis-Hastings

method for heavy-tailed SV models.

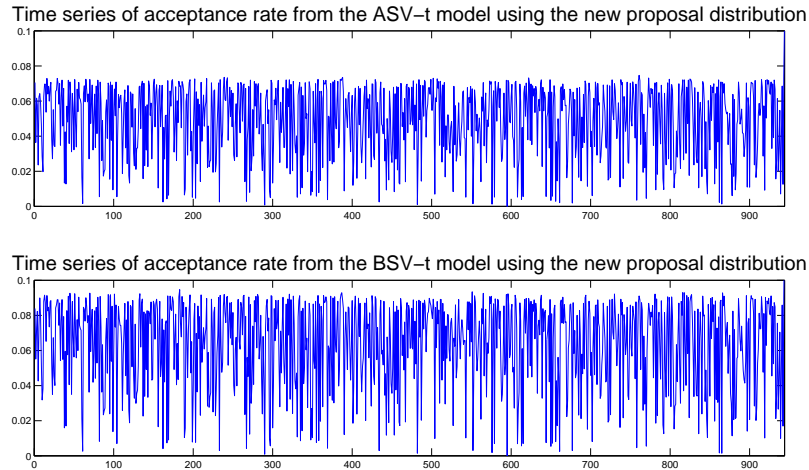


Figure 2.8: Acceptance rate of the latent states of the heavy-tailed SV models fitted to the exchange rate data.

Simulation studies and comparisons with the two competing methods demonstrate that our estimation methods is consistent in terms of parameter estimation. The simple Metropolis-Hastings method can be applied to the latent states of SV models without heavy tails, and the estimation speed can increase substantially. The normality assumption of the realized observation errors and the uniform distribution of the PITs from the ASV model are all supported by the results of the Kolmogorov-Smirnov tests.

2.4 Empirical Illustrations

In this section, we present an application of our MCMC methods to the ASV and ASV-t models using daily returns of the IBM stock. The historical data was downloaded from the web site finance.yahoo.com yielding 1730 observations from Jan. 3, 2003 to Nov. 13, 2009. We fit the ASV and ASV-t models to the first 1500 observations and the other 230 observations were used for an out-of-sample forecast assessment.

2.4.1 Data Analysis of the ASV and ASV-t Models

We fitted the ASV and ASV-t models to the asset return data. For each model, the estimation algorithm was iterated 50,000 times. The parameters and log volatilities were estimated by means of a sample average based on the last 40,000 observations.

After fitting the ASV and ASV-t models, we checked the innovation assumptions of the measurement equations of the two models. For the ASV model, we assessed the normality assumption of the realized errors using the Kolmogorov-Smirnov test. The statistic of the Kolmogorov-Smirnov test is 0.0203 with the the critical value 0.035 at the 5% significance level. So we can not reject the null hypothesis at the 5% significance level that the realized innovations of the ASV model follow a standard normal distribution. In the test of Student- t distribution for the observation errors of the ASV-t model, the degrees of freedom is unknown; so we use its estimate $v = 15.98$. The Kolmogorov-Smirnov test statistic is 0.0148 with the critical value 0.035 at the 5% significance level. Thus, we do not reject the null hypothesis that the realized innovations follow a Student- t distribution with 15.98 degrees of freedom. Figure 2.9 compares the empirical CDFs of the realized errors and the related theoretical CDFs, which is consistent with both the normality and Student- t assumptions.

To assess the overall fit of the ASV and ASV-t models to the return data, we test the PITs. In Figure 2.10, we plot the theoretical CDF of Uniform $[0, 1]$ and empirical CDFs of the PITs. The Kolmogorov-Smirnov test statistic for the PITs from the ASV model is 0.0207 with the critical value 0.035 at the 5% significance level. Thus, we do not reject the null hypothesis that the estimated ASV model agrees with the IBM return data. Similarly, for the ASV-t model, the Kolmogorov-Smirnov test statistic for the PITs is 0.0183 with the the critical value 0.035 at the 5% significance level, indicating that the PITs are i.i.d. Uniform $[0, 1]$. Both of the Kolmogorov-Smirnov tests show that the ASV and ASV-t models are good candidates for the IBM daily return data. The estimated

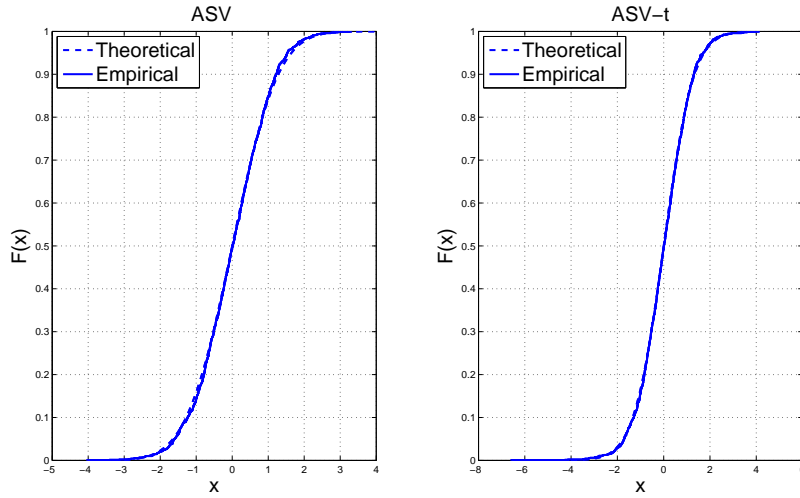


Figure 2.9: Comparison between the theoretical and empirical CDFs of the observation errors of the ASV (left) and ASV-t (right) models based on the IBM return data.

parameters are presented in the second and third columns of Table 2.7.

2.4.2 Estimation of Several Competing Models

In order to obtain a good model for the IBM return data, we compare the results from several competing models estimated from the first 1500 observations of the IBM stock.

A heavy-tailed asymmetric SV model with leverage

Let us recall that for the heavy-tailed ASV model presented in Jacquier et al. (2004), it is not obvious how to obtain and interpret the leverage effect. The revised version of this model is given by

$$y_t = \exp(h_t/2)\epsilon_t, \quad \epsilon_t \sim t(v), \quad (2.40)$$

$$h_{t+1} = \mu + \phi(h_t - \mu) + \sqrt{\sigma}\eta_{t+1}, \quad \eta_{t+1} \sim \mathcal{N}(0, 1), \quad (2.41)$$

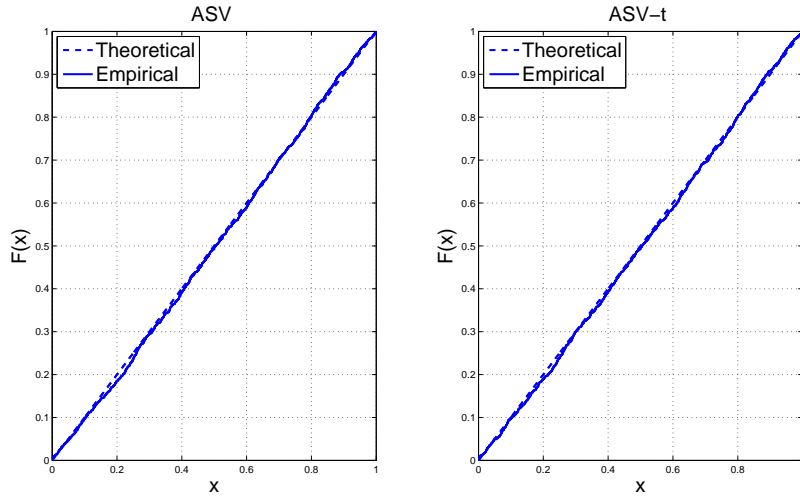


Figure 2.10: Comparison between the theoretical and empirical CDFs of the PITs of the ASV (left) and ASV-t (right) models upon the IBM returns.

where the measurement error ϵ_t is decomposed as $\epsilon_t = \sqrt{\lambda_t}e_t$, λ_t follows an inverse Gamma distribution $\lambda_t \sim \mathcal{IG}(v/2, v/2)$, and $e_t \sim \mathcal{N}(0, 1)$, which is correlated with η_{t+1} , such that $\rho = \text{corr}(e_t, \eta_{t+1})$. We call this model a ASV-d model, where the letter d stands for the decomposition of Student- t distribution. This mixture decomposition is used by Harvey et al. (1994), Jacquier (2004) and can be found in some latter papers, such as in Hautsch and Ou (2008), Shimada and Tsukuda (2005), and Zhang and King (2008). The reason for introducing inverse Gamma variables λ_t in the ASV-d model is that the existing estimation methods for the ASV model can be easily modified and applied, which will be seen later. Under the mixture specification, the volatilities of asset returns are not only described by h_t but also $\sqrt{\lambda_t}$. That is, the conditional variance of y_t is $\lambda_t \exp(h_t)$. In terms of interpreting volatilities of asset returns, we consider the two values together. It is also noticed that under the mixture decomposition, the correlation between the two equations is determined by the correlation between the two standard normal noises. Moreover, The latent process partially describes the log volatilities of y .

To estimate this model, we notice that once $\lambda_t, t = 1, \dots, T$, have been sampled, $y_t^* = y_t/\sqrt{\lambda_t}, t = 1, \dots, T$, can be modeled by the ASV model. So the parameters of the AR(1) process can be estimated in the same way as in the ASV model. The simulation of v proceeds in the same way as that in the ASV-t model by using formula (2.21). To sample λ_t , the density of λ_t , given v , is

$$f(\lambda_t|v) = \frac{(v/2)^{(v/2)}}{\Gamma(v/2)} \left(\frac{1}{\lambda_t}\right)^{v/2+1} \exp\left\{-\frac{v/2}{\lambda_t}\right\}. \quad (2.42)$$

The full conditional of λ_t , given y_t, h_t and v , is

$$\begin{aligned} f(\lambda_t|y_t, h_t, v) &\propto \left\{\frac{1}{\lambda_t}\right\}^{(1/2)} \exp\left\{-\frac{y_t^2 \exp(-h_t/2)}{2\lambda_t}\right\} \frac{(v/2)^{(v/2)}}{\Gamma(v/2)} \left(\frac{1}{\lambda_t}\right)^{v/2+1} \exp\left\{-\frac{v/2}{\lambda_t}\right\} \\ &\propto \left\{\frac{1}{\lambda_t}\right\}^{(v+1)/2+1} \exp\left\{-\frac{y_t^2 \exp(-h_t/2) + v}{2\lambda_t}\right\}. \end{aligned} \quad (2.43)$$

Hence we can sample λ_t directly from the inverse Gamma distribution

$$\lambda_t|y_t, h_t, v \sim IG((v+1)/2, (y_t^2 \exp(-h_t/2) + v)/2). \quad (2.44)$$

Using the same prior densities of parameters as those used for the ASV and ASV-t models, we applied the sampling algorithm to the ASV-d model for the first 1500 observations of the IBM returns. A summary of the posterior means of the parameters is listed in the fourth column of Table 2.7. If the λ_t 's were all equal to one, the innovation would be a standard normal distribution. If it has an inverse gamma distribution, the innovation has a Student- t distribution. Figure 2.11 graphs the time series of the estimated volatilities and the time series of $\sqrt{\lambda_t}$, respectively. The mean of $\sqrt{\lambda_t}$ is 1.032 with standard deviation 0.16. This indicates that the innovations ϵ_t follow a Student- t distribution. The AIC and BIC values of the ASV-t and ASV-d models presented in Table 2.8 are very close, which says that these two models are ranked first when they

are estimated on asset return data. It is found that the estimated degrees of freedom ν in both of the ASV-t and ASV-d models are approximately the same. In practice, we would prefer the ASV-t model since it is convenient for obtaining volatility forecasts. Specifically, to perform an out-of-sample forecast, the ASV-t model is not required to predict and filter λ_t during the auxiliary particle filter procedure; instead it only requires ν and the conditional Student- t distribution of y_t . We can directly sample the Student- t distribution without any mixture decomposition.

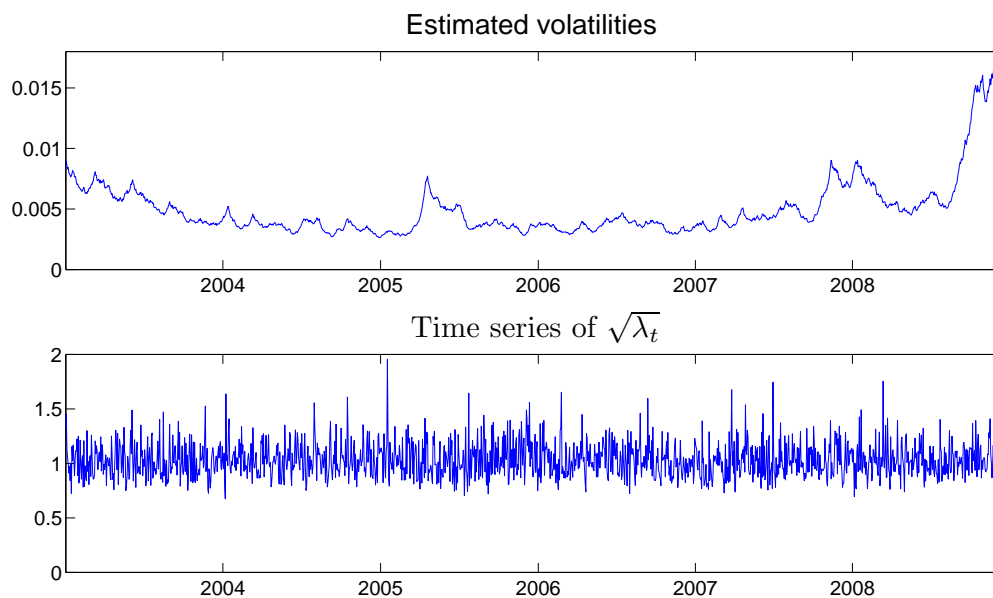


Figure 2.11: The estimated $\sqrt{\lambda_t}$ s and volatilities obtained from the fit of the ASV-d model to the IBM return data.

The BSV model

The BSV model was defined by equations (2.1) to (2.3), where a non-zero correlation between the two innovations are not permitted. Using the same prior distributions of parameters as those in ASV model and fixing $\rho = 0$, the IBM return data was fitted and the parameter estimates are included in Table 2.7.

Heavy-tailed basic SV (BSV-t) model

Similar to the BSV model, if the correlation coefficient in the ASV-t model is assumed to be zero, then we obtain a heavy-tailed basic SV model (BSV-t) expressed as:

$$y_t = \exp(h_t/2)\epsilon_t, \quad \epsilon_t \sim t(v), \quad (2.45)$$

$$h_{t+1} = \mu + \phi(h_t - \mu) + \sqrt{\sigma}\eta_{t+1}, \quad \eta_{t+1} \sim \mathcal{N}(0, 1), \quad (2.46)$$

where the correlation $\text{corr}(\epsilon_t, \eta_{t+1}) = 0$. We fitted this model to the asset return data with prior distributions being the same as those in the ASV-t model and by pre-setting $\rho = 0$. The estimated parameters are presented in the last column of Table 2.7.

Table 2.7: Estimates of parameters obtained from daily returns of the IBM stock through the ASV, ASV-t and competing models. The standard errors are in the parentheses.

Parameter	ASV	ASV-t	ASV-d	BSV	BSV-t
μ	-10.49 (0.39)	-10.18 (1.24)	-10.45 (0.72)	-10.47 (0.50)	-10.17 (1.32)
ϕ	0.98 (0.01)	0.99 (0.01)	0.99 (0.01)	0.98 (0.01)	0.99 (0.01)
ρ	-0.42 (0.09)		-0.34 (0.10)		
ψ		-0.06 (0.01)			
σ	0.035 (0.013)	0.011 (0.005)	0.021 (0.009)	0.029 (0.011)	0.012 (0.006)
v		15.98 (4.49)	16.10 (5.59)		15.98 (4.36)

According to the estimation results in Table 2.7, we found that the estimated variances of the latent AR(1) processes are different across various specifications of the SV model. Using the Student- t distribution, the estimates of volatility parameter σ are much smaller than those from other SV models, which is related to the AIC and

BIC values on the next page. Moreover, the estimated correlations in the ASV and ASV-d models are significantly negative, indicating that the leverage effect between the returns and the volatilities does exist. The estimate of ψ in the ASV-t model is negative, indicating this model can also capture a negative correlation between the two processes. Table 2.8 presents the values of AIC and BIC of the fitted SV models. The last column is obtained when the first 1500 observations of the IBM returns were fitted by a GARCH(1,1) model with i.i.d. univariate standard normal innovations. Both of the AIC and BIC attained their smallest values for the ASV-t model. Thus, the ASV-t model seems to have the best fit among the six models for the IBM daily return data.

Table 2.8: The AIC and BIC values for the ASV and ASV-t and competing models based on the IBM return data.

Model	ASV	ASV-t	ASV-d	BSV	BSV-t	GARCH(1,1)
AIC	-11538.4	-11560.4	-11557.3	-11522.7	-11539.9	-11449.0
BIC	-11517.1	-11533.8	-11530.7	-11501.5	-11513.3	-11428.0

Figure 2.12 compares the absolute IBM returns with the forecasted volatilities from the ASV, ASV-t and the other competing models. The time series before and after the dotted vertical line at $t = 1500$ are the in-sample and out-of-sample one-step ahead forecasted volatilities. Although we can not observe the true volatilities, we may be confident that the ASV-t and ASV-d model will perform better in terms of volatility forecast.

To further assess whether the ASV-t model fits the IBM return data well, we repeatedly generated asset returns and fitted the ASV-t model to see how often the fitted model agreed with the generated data. We generated 100 data sets of asset returns from this model indexed by the estimated parameters. Each data set contains 1500 observations which also has the same size as the IBM return data used to fit the ASV-t model. For each generated data set, we fitted the ASV-t model and computed the Kolmogorov-Smirnov statistic of the PITs. We found that there were 45 times that the

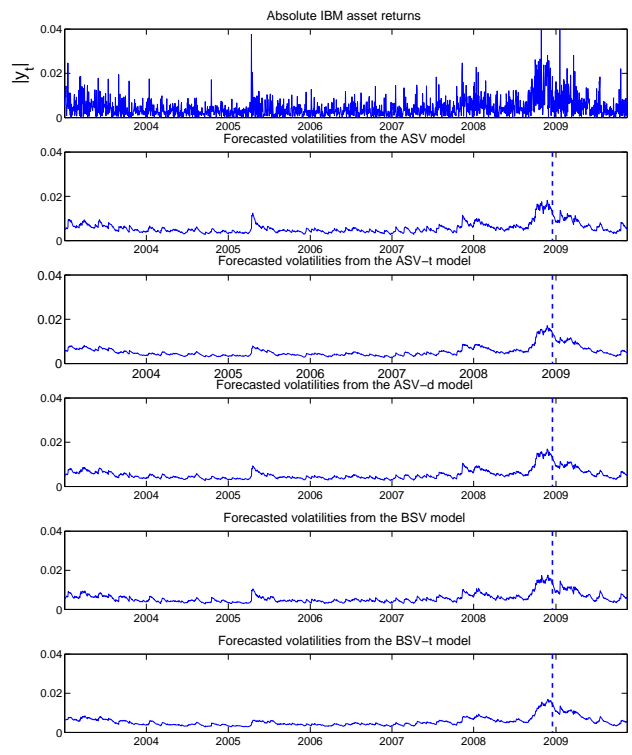


Figure 2.12: Comparison of absolute asset returns vs. forecasted volatilities from the competing models.

calculated statistics were greater than 0.0183, which was the value of the Kolmogorov-Smirnov statistic for the Uniform $[0, 1]$ test of PITs from the ASV-t model to the first 1500 observations of the IBM returns. An empirical estimate of the significance probability is 0.55. Therefore, there is no evidence to suggest that the ASV-t model fails to fit the real data.

2.5 Conclusion

In this chapter, the slice sampler within the MCMC methods were developed for various specifications of the univariate SV models including the ASV and ASV-t models. For heavy-tailed SV models, the Student- t distribution can be used directly without any decomposition of inverse Gamma and standard normal distributions. Simulation studies and empirical applications showed that our proposed approaches work well in terms of parameter estimation. We employed the AIC and BIC criteria for model selection, where the sample likelihoods were evaluated via the auxiliary particle filter. The overall model fit was assessed by the Kolmogorov-Smirnov test of the PITs realized from the estimated models. Since the ASV-t model seems to fit the IBM asset return data better than the other competing SV models, volatility forecasts conducted via the auxiliary particle filter based on this model resemble the time series of the absolute asset returns.

2.6 Appendix

Full Conditionals of Parameters and Latent States in the ASV Model

- Full conditional of μ , given a normal prior $\mathcal{N}(\alpha_\mu, \beta_\mu^2)$, is expressed as

$$\begin{aligned}
& f(\mu|\mathbf{y}, \phi, \psi, \tau) \\
& \propto f(\mathbf{h}|\Theta, \mathbf{y}) \exp\left\{-\frac{(\mu - \alpha_\mu)^2}{2\beta_\mu^2}\right\} \\
& \propto f(h_1|\Theta) \prod_{t=1}^{T-1} f(h_{t+1}|h_t, \Theta, y_t) \exp\left\{-\frac{(\mu - \alpha_\mu)^2}{2\beta_\mu^2}\right\} \\
& \propto \exp\left\{-\frac{\sum_{t=1}^{T-1} [(h_{t+1} - \mu) - \phi(h_t - \mu) - \psi \exp(-h_t/2)y_t]^2}{2\tau}\right\} \\
& \quad \times \exp\left\{-\frac{(1 - \phi^2)(h_1 - \mu)^2}{2\sigma}\right\} \exp\left\{-\frac{(\mu - \alpha_\mu)^2}{2\beta_\mu^2}\right\} \\
& = \exp\left(-\frac{1}{2}\left[\mu^2\left\{\frac{1 - \phi^2}{\sigma} + \frac{(T - 1)(1 - \phi)^2}{\tau} + \frac{1}{\beta_\mu^2}\right\}\right.\right. \\
& \quad \left.\left.- 2\mu\left\{\frac{h_1(1 - \phi^2)}{\sigma} + \frac{(1 - \phi) \sum_{t=1}^{T-1} (h_{t+1} - \phi h_t - \psi \exp(-h_t/2)y_t)}{\tau} + \frac{\alpha_\mu}{\beta_\mu^2}\right\}\right]\right) \\
& \propto \mathcal{N}\left(\mu; \frac{b}{a}, \frac{1}{a}\right).
\end{aligned}$$

where

$$\begin{aligned}
a &= \frac{1 - \phi^2}{\sigma} + \frac{(T - 1)(1 - \phi)^2}{\tau} + \frac{1}{\beta_\mu^2}, \\
b &= \frac{h_1(1 - \phi^2)}{\sigma} + \frac{(1 - \phi) \sum_{t=1}^{T-1} (h_{t+1} - \phi h_t - \psi \exp(-h_t/2)y_t)}{\tau} + \frac{\alpha_\mu}{\beta_\mu^2}.
\end{aligned}$$

- Full conditional of ϕ , given a normal prior $\mathcal{N}(\alpha_\phi, \beta_\phi^2)$, is given by

$$\begin{aligned}
& f(\phi|\mathbf{y}, \mu, \psi, \tau) \\
& \propto f(\mathbf{h}|\Theta, \mathbf{y}) \exp\left\{-\frac{(\phi - \alpha_\phi)^2}{2\beta_\phi^2}\right\} \\
& \propto f(h_1|\Theta) \prod_{t=1}^{T-1} f(h_{t+1}|h_t, \Theta, y_t) \exp\left\{-\frac{(\phi - \alpha_\phi)^2}{2\beta_\phi^2}\right\} \\
& \propto \exp\left\{-\frac{\sum_{t=1}^{T-1} [(h_{t+1} - \mu) - \phi(h_t - \mu) - \psi \exp(-h_t/2)y_t]^2}{2\tau}\right\} \\
& \quad \times \exp\left\{-\frac{(1 - \phi^2)(h_1 - \mu)^2}{2\sigma}\right\} \exp\left\{-\frac{(\phi - \alpha_\phi)^2}{2\beta_\phi^2}\right\} (1 - \phi^2)^{\frac{1}{2}} \\
& = \exp\left(-\frac{1}{2}\left[\phi^2\left\{\frac{-(h_1 - \mu)^2}{\sigma} + \frac{\sum_{t=1}^{T-1}(h_t - \mu)^2}{\tau} + \frac{1}{\beta_\phi^2}\right\}\right.\right. \\
& \quad \left.\left.- 2\phi\left\{\frac{\sum_{t=1}^{T-1}(h_t - \mu)(h_{t+1} - \mu - \psi \exp(-h_t/2)y_t)}{\tau} + \frac{\alpha_\phi}{\beta_\phi^2}\right\}\right]\right) (1 - \phi^2)^{\frac{1}{2}} \\
& \propto \mathcal{N}\left(\phi; \frac{d}{c}, \frac{1}{c}\right) (1 - \phi^2)^{\frac{1}{2}},
\end{aligned}$$

where

$$\begin{aligned}
c &= \frac{-(h_1 - \mu)^2}{\sigma} + \frac{\sum_{t=1}^{T-1}(h_t - \mu)^2}{\tau} + \frac{1}{\beta_\phi^2}, \\
d &= \frac{\sum_{t=1}^{T-1}(h_t - \mu)(h_{t+1} - \mu - \psi \exp(-h_t/2)y_t)}{\tau} + \frac{\alpha_\phi}{\beta_\phi^2}.
\end{aligned}$$

- Full conditional of ψ , given a normal prior density $\mathcal{N}(\alpha_\psi, \beta_\psi^2)$, is given by

$$\begin{aligned}
& f(\psi|\mathbf{y}, \mu, \phi, \tau) \\
& \propto f(\mathbf{h}|\Theta, \mathbf{y}) \exp \left\{ -\frac{(\psi - \alpha_\psi)^2}{2\beta_\psi^2} \right\} \\
& \propto f(h_1|\Theta) \prod_{t=1}^{T-1} f(h_{t+1}|h_t, \Theta, y_t) \exp \left\{ -\frac{(\psi - \alpha_\psi)^2}{2\beta_\psi^2} \right\} \\
& \propto \exp \left\{ -\frac{\sum_{t=1}^{T-1} [(h_{t+1} - \mu) - \phi(h_t - \mu) - \psi \exp(-h_t/2)y_t]^2}{2\tau} \right\} \\
& \quad \times \exp \left\{ -\frac{(\psi - \alpha_\psi)^2}{2\beta_\psi^2} \right\} \\
& = \exp \left(-\frac{1}{2} \left[\psi^2 \left\{ \frac{\sum_{t=1}^{T-1} [\exp(-h_t/2)y_t]^2}{\tau} + \frac{1}{\beta_\psi^2} \right\} \right. \right. \\
& \quad \left. \left. - 2\psi \left\{ \frac{\sum_{t=1}^{T-1} [\exp(-h_t/2)y_t \{ (h_{t+1} - \mu) - \phi(h_t - \mu) \}]}{\tau} + \frac{\alpha_\psi}{\beta_\psi^2} \right\} \right] \right) \\
& \propto \mathcal{N} \left(\psi; \frac{b_1}{a_1}, \frac{1}{a_1} \right),
\end{aligned}$$

where

$$\begin{aligned}
a_1 &= \frac{\sum_{t=1}^{T-1} [\exp(-h_t/2)y_t]^2}{\tau} + \frac{1}{\beta_\psi^2}, \\
b_1 &= \frac{\sum_{t=1}^{T-1} [\exp(-h_t/2)y_t \{ (h_{t+1} - \mu) - \phi(h_t - \mu) \}]}{\tau} + \frac{\alpha_\psi}{\beta_\psi^2}.
\end{aligned}$$

- Full conditional of τ , given an inverse Gamma prior density $f(\alpha_\tau, \beta_\tau)$, is

$$\begin{aligned}
& f(\tau|\mathbf{y}, \mu, \psi, \phi) \\
& \propto f(\mathbf{h}|\Theta, \mathbf{y})f(\alpha_\tau, \beta_\tau) \\
& \propto f(h_1|\Theta) \prod_{t=1}^{T-1} f(h_{t+1}|h_t, \Theta, y_t)f(\alpha_\tau, \beta_\tau) \\
& \propto \exp \left\{ -\frac{\sum_{t=1}^{T-1} [(h_{t+1} - \mu) - \phi(h_t - \mu) - \psi \exp(-h_t/2)y_t]^2}{2\tau} \right\} \\
& \quad \times \exp \left\{ -\frac{(1 - \phi^2)(h_1 - \mu)^2}{2\tau} \right\} \times \left(\frac{1}{\tau} \right)^{\frac{T}{2}} \frac{(\beta_\tau)^{\alpha_\tau} e^{-\beta_\tau/\sigma_\tau}}{\Gamma(\alpha_\tau)(\sigma_\tau)^{\alpha_\tau+1}} \\
& \propto \exp \left(\frac{\beta_\tau + \frac{1}{2}(h_1 - \mu)^2(1 - \phi^2) + \frac{1}{2} \sum_{t=1}^{T-1} [(h_{t+1} - \mu) - \phi(h_t - \mu) - \psi \exp(-h_t/2)y_t]^2}{\tau} \right) \\
& \quad \times \left(\frac{1}{\tau} \right)^{(\alpha_\tau + \frac{T}{2})+1} \\
& \propto \mathcal{IG}(\tau; a, b),
\end{aligned}$$

which is an inverse Gamma distribution with parameters,

$$\begin{aligned}
a &= \alpha_\tau + \frac{T}{2}, \\
b &= \beta_\tau + \frac{1}{2}(h_1 - \mu)^2(1 - \phi^2) + \frac{1}{2} \sum_{t=1}^{T-1} [(h_{t+1} - \mu) - \phi(h_t - \mu) - \psi \exp(-h_t/2)y_t]^2.
\end{aligned}$$

- Full conditional of h_1 can be specified as

$$\begin{aligned}
& P(h_1|y_{1:T}, h_2, \Theta) \\
&= c_1 f(y_1|h_1) f(h_1|\Theta) f(h_1|h_2, y_1, \Theta) \\
&= c_2 \exp\left\{\frac{-h_1}{2}\right\} \exp\left\{-\frac{y_1^2}{2} \exp(-h_1)\right\} \exp\left\{-\frac{(1-\phi^2)(h_1-\mu)^2}{2\sigma}\right\} \\
&\quad \times \exp\left\{-\frac{[(h_2-\mu) - \phi(h_1-\mu) - \psi \exp(-h_1/2)y_1]^2}{2\tau}\right\} \\
&< c_2 \exp\left\{\frac{-h_1}{2}\right\} \exp\left\{-\frac{y_1^2}{2} \exp(-h_1)\right\} \exp\left\{-\frac{(1-\phi^2)(h_1-\mu)^2}{2\sigma}\right\},
\end{aligned}$$

where c_1 and c_2 are the two normalizing constants.

- Full conditional of h_T is given by

$$\begin{aligned}
& f(h_T|y_{1:T}, h_{T-1}, \Theta) \\
&\propto f(y_T|h_T) f(h_T|h_{T-1}, y_{T-1}, \Theta) \\
&\propto \exp\left\{\frac{-h_T}{2}\right\} \exp\left\{-\frac{y_T^2}{2} \exp(-h_T)\right\} \\
&\quad \times \exp\left\{-\frac{[(h_T-\mu) - \phi(h_{T-1}-\mu) - \psi \exp(-h_{T-1}/2)y_{T-1}]^2}{2\tau}\right\}.
\end{aligned}$$

Full Conditionals of Latent States of the ASV-t Model

- Full conditional of h_1 can be expressed as

$$\begin{aligned}
 & f(h_1|y_{1:T}, h_2, \Theta) \\
 &= c_1 f(y_1|h_1) f(h_1|\Theta) f(h_1|h_2, y_1, \Theta) \\
 &= c_2 e^{-h_1/2} \left(1 + \frac{y_1^2 e^{-h_1}}{v}\right)^{-\frac{v+1}{2}} \exp\left\{-\frac{(1-\phi^2)(h_1-\mu)^2}{2\sigma}\right\} \\
 &\quad \times \exp\left\{-\frac{[(h_2-\mu) - \phi(h_1-\mu) - \psi \exp(-h_1/2)y_t]^2}{2\tau}\right\} \\
 &< c_2 e^{-h_1/2} \left(1 + \frac{y_1^2 e^{-h_1}}{v}\right)^{-\frac{v+1}{2}} \exp\left\{-\frac{(1-\phi^2)(h_1-\mu)^2}{2\sigma}\right\},
 \end{aligned}$$

where c_1 and c_2 are the two normalizing constants.

- Full conditional of h_T .

$$\begin{aligned}
 & f(h_T|y_{1:T}, h_{T-1}, \Theta) \\
 &\propto f(y_T|h_T) f(h_T|h_{T-1}, y_{T-1}, \Theta) \\
 &\propto \frac{\Gamma(\frac{v+1}{2})}{\Gamma(\frac{v}{2})} \frac{1}{\sqrt{\pi v}} e^{-h_T/2} \left(1 + \frac{y_T^2 e^{-h_T}}{v}\right)^{-\frac{v+1}{2}} \\
 &\quad \times \exp\left\{-\frac{[(h_T-\mu) - \phi(h_{T-1}-\mu) - \psi \exp(-h_{T-1}/2)y_{T-1}]^2}{2\tau}\right\} \\
 &\propto e^{-h_T/2} \left(1 + \frac{y_T^2 e^{-h_T}}{v}\right)^{-\frac{v+1}{2}} \\
 &\quad \times \exp\left\{-\frac{[(h_T-\mu) - \phi(h_{T-1}-\mu) - \psi \exp(-h_{T-1}/2)y_{T-1}]^2}{2\tau}\right\}.
 \end{aligned}$$

Slice Sampler for the ASV-t Model

Based on the full conditionals of h_t , $t = 1, \dots, T$, we only give the slice sampler for sampling h_t . It can be verified that the full conditional of h_t , obtained in Section 2.2, is

dominated by $f(h_t)$ as follows.

$$\begin{aligned} f(h_t) &\propto g_1(h_t)g_2(h_t) \\ &\propto \left\{ 1 + \frac{y_t^2 \exp(-h_t)}{v} \right\}^{-\frac{v+1}{2}} \exp \left\{ -\frac{(h_t - \mu_t)^2}{2\tau} \right\}, \end{aligned}$$

where $\mu_t = \mu - \frac{\tau}{2} + \phi(h_{t-1} - \mu) + \psi \exp(-h_{t-1}/2)y_{t-1}$.

Step 0. Given $h_t^{(k)}$, the draw at the iteration k .

Step 1. Draw $u_1 \sim \mathcal{U}(0, 1)$. Let $u_2 = u_1 * \left\{ 1 + \frac{y_t^2 \exp(-h_t^{(k)})}{v} \right\}^{-\frac{v+1}{2}}$ and

$$u_2 \leq \left\{ 1 + \frac{y_t^2 \exp(-h_t)}{v} \right\}^{-\frac{v+1}{2}}.$$

If $y_t \neq 0$, then we have

$$\begin{aligned} \exp(-h_t) &\leq v \left(\left(\frac{1}{u_2} \right)^{\frac{2}{v+1}} - 1 \right) / y_t^2 \\ h_t &\geq -\log \left[v \left(\left(\frac{1}{u_2} \right)^{\frac{2}{v+1}} - 1 \right) / y_t^2 \right]. \end{aligned} \tag{A.1}$$

Step 2. Draw $u_3 \sim \mathcal{U}(0, 1)$.

Let $u_4 = u_3 * \exp \left\{ -\frac{(h_t^{(k)} - \mu_t^{(k)})^2}{2\tau} \right\}$ and

$$\begin{aligned} u_4 &\leq \exp \left\{ -\frac{(h_t - \mu_t^{(k)})^2}{2\tau} \right\} \\ \log(u_4) &\leq -\frac{(h_t - \mu_t^{(k)})^2}{2\tau} \\ (h_t - \mu_t^{(k)})^2 &\leq -2\tau \log(u_4). \end{aligned}$$

Then we have

$$\mu_t^{(k)} - \sqrt{-2\tau \log(u_4)} \leq h_t \leq \mu_t^{(k)} + \sqrt{-2\tau \log(u_4)}. \quad (\text{A.2})$$

Step 3. If $y_t \neq 0$, draw $h_t^{(k+1)}$ uniformly from the interval determined by the inequalities (A.1) and (A.2),

$$h_t \sim \mathcal{U}\left(\max\left\{-\log\left[v\left(\left(\frac{1}{u_2}\right)^{\frac{2}{v+1}} - 1\right)/y_t^2\right], \mu_t^{(k)} - \sqrt{-2\tau \log(u_4)}\right\}, \mu_t^{(k)} + \sqrt{-2\tau \log(u_4)}\right),$$

otherwise,

$$h_t \sim \mathcal{U}\left(\mu_t^{(k)} - \sqrt{-2\tau \log(u_4)}, \mu_t^{(k)} + \sqrt{-2\tau \log(u_4)}\right).$$

Chapter 3

Efficient Bayesian Estimation of a Multivariate Stochastic Volatility Model with Cross Leverage

3.1 Introduction

In equity markets, leverage effects are often observed between asset returns and latent volatilities. As discussed by Yu (2005) and also studied in Jacquier et al. (2004) and Omori et al. (2007), there is strong evidence for a leverage effect and leverage is a particularly important feature of asset returns in the equity market. Univariate stochastic volatility (SV) models have been successful in capturing this property of asset returns. Current research work in this area includes studies by Jacquier et al. (2004), Omori et al. (2007), Zhang and King (2008), Kawakatsu (2007) and Harvey and Shephard (1996). Since likelihood functions are difficult to evaluate for SV models, Bayesian inference methods such as the Markov Chain Monte Carlo (MCMC) methods have been proposed instead (e.g. Shephard and Pitt (1997), Jacquier et al. (2004) and Omori et al.

(2007)).

It is known that the individual asset return's time series evolve not only on their own but are also related to other financial time series in the market. To study such complex phenomena, univariate SV models are not sufficient since they are unable to model the correlations of multivariate asset returns. Based on this observation, univariate SV models have been extended to multivariate SV (MSV) settings to capture the properties of multivariate financial asset returns. Analogous to the univariate SV models, non-zero cross correlations between the innovations of multivariate returns and those of the latent Markov processes are permitted. Our aim in this chapter is to develop a novel MCMC method to estimate a more general specification of the MSV model.

The leverage MSV model was first mentioned by Danielsson (1998) in his empirical study of foreign exchange rates and stock indices, but the author did not fit the model to real data. The reason that the leverage effect MSV model has not been extensively studied is the difficulty arising in the estimation of conditional variances of the two innovations; for instance, the covariance matrix of the observation errors in this model is a correlation matrix. Chan et al. (2006) propose a more general leverage MSV model, where non-zero correlations are permitted across the innovations of asset returns and those of the volatility dynamics. In estimating this model, the authors focus not on the simulation of the variance-covariance matrix directly but the correlation matrix instead. An MCMC algorithm is proposed from which an inverted correlation matrix is simulated one entry at a time through the Metropolis-Hastings (MH) algorithm using a parsimonious reparameterization proposed in Wong et al. (2003). In the implementation, Chan et al. (2006) perform an element selection on the off-diagonal elements of the inverted correlation matrix. By imposing a prior, the off-diagonal elements of the inverse of the correlation matrix are allowed to be identically zero. The technique is also adapted by Pitt et al. (2006) in their fit of Gaussian copula regression models.

In order to improve the simulation efficiency, Chan et al. (2006) simulate the latent random variables by blocks with the MH algorithm as in Smith and Pitts (2006). The proposal is a Gaussian density obtained by a quadratic approximation to the full conditional around the mode. The mode is found by using an additional Newton-Raphson method. Because of the extra numerical algorithm required in this case, the realized simulation method is relatively more time-consuming. To the best of our knowledge, the study by Chan et al. (2006) represents the first attempt to fit such a general MSV model.

In this chapter, we consider the same model as in Chan et al. (2006) but develop a straightforward MCMC approach for parameter and log volatility estimation. The main difficulty with the MSV model is the sampling of the latent states and the restricted conditional variance-covariance matrix of the model. Our contributions beyond Chan et al. (2006) are as follows. First, we provide a framework in the simulation of the constrained variance-covariance matrix under which the required computation altogether is reduced by simulating the variance-covariance matrix simultaneously rather than element-by-element. Second, for the latent states, we use an MH method, where the proposal distribution is sampled by the slice sampler proposed by Edwards and Sokal (1988) and Neal (2003). Because the slice sampler can adapt to the analytical structure of the target density, it is expected to be more efficient when compared with other sampling methods, such as the one given in Chan et al. (2006). Third, our method is relatively more straightforward in dealing with high-dimensional data than the method of Chan et al. (2006).

The rest of the chapter is organized as follows. Section 3.2 gives the definition of a MSV model with cross variance-covariance structure. An MCMC algorithm for estimating the parameters and log volatilities is presented in Section 3.3. Simulation studies and applications to asset return data are given in Section 3.4 and Section 3.5, respectively. Concluding remarks can be found in Section 3.6.

3.2 Model and Estimation

3.2.1 Model

To simplify the exposition, we assume that the considered financial asset returns are demeaned. Before formulating the MSV model, we first reproduce the specification of the univariate SV model with a leverage effect. Define by y_t the return at time t , $t \leq T$. The univariate model is specified as

$$y_t = \exp(h_t/2)\epsilon_t, \quad t = 1, \dots, T, \quad (3.1)$$

$$h_{t+1} = \mu + \phi(h_t - \mu) + \sqrt{\sigma}\eta_{t+1}, \quad t = 1, \dots, T - 1, \quad (3.2)$$

where $h_1 \sim \mathcal{N}(\mu, \sigma/(1 - \phi^2))$ and

$$\begin{pmatrix} \epsilon_t \\ \eta_{t+1} \end{pmatrix} \sim \mathcal{N}(\mathbf{0}, \Sigma), \quad \Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \quad (3.3)$$

where ρ is the correlation coefficient between the two innovations, $\mathcal{N}(a, b)$ denotes a univariate normal distribution with mean a and variance b , and $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$ is a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and variance-covariance matrix Σ . The log volatility of y_t follows a first-order autoregressive (AR(1)) process. We assume that the persistence parameter in the AR(1) process, ϕ , satisfies the weak stationarity restriction $|\phi| < 1$. In various applications of this model to financial time series, the correlation ρ is often found to be negative, which can be interpreted as a leverage effect between asset returns and latent volatilities. The resulting model is called the univariate asymmetric stochastic volatility (ASV) model.

Now we extend the ASV model to a multivariate setting following the definition in Chan et al. (2006). Let $\mathbf{y}_t = (y_{1,t}, \dots, y_{m,t})'$ denote an m -dimensional vector of asset

returns at time t , and $\mathbf{h}_t = (h_{1,t}, \dots, h_{m,t})'$ be the $m \times 1$ vector of log volatility of \mathbf{y}_t . We assume that \mathbf{h}_t follows a vector AR(1) process with mean $\boldsymbol{\mu} = (\mu_1, \dots, \mu_m)'$ and persistence parameter vector $\boldsymbol{\phi} = (\phi_1, \dots, \phi_m)'$ that satisfies the weak stationarity condition $|\phi_i| < 1$ for $i = 1, \dots, m$. The MSV model is given by

$$\mathbf{y}_t = \mathbf{H}_t^{\frac{1}{2}} \boldsymbol{\epsilon}_t, \quad t = 1, \dots, T, \quad (3.4)$$

$$\mathbf{h}_{t+1} = \boldsymbol{\mu} + \boldsymbol{\Phi}(\mathbf{h}_t - \boldsymbol{\mu}) + \boldsymbol{\eta}_{t+1}, \quad t = 1, \dots, T - 1, \quad (3.5)$$

$$\mathbf{h}_1 \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}_0), \quad (3.6)$$

where

$$\mathbf{H}_t^{\frac{1}{2}} = \text{diag}(\exp(h_{1,t}/2), \dots, \exp(h_{m,t}/2)), \quad (3.7)$$

$$\boldsymbol{\Phi} = \text{diag}(\phi_1, \dots, \phi_m), \quad (3.8)$$

$$\begin{pmatrix} \boldsymbol{\epsilon}_t \\ \boldsymbol{\eta}_{t+1} \end{pmatrix} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}), \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{\epsilon\epsilon} & \boldsymbol{\Sigma}_{\epsilon\eta} \\ \boldsymbol{\Sigma}_{\eta\epsilon} & \boldsymbol{\Sigma}_{\eta\eta} \end{pmatrix}. \quad (3.9)$$

The (i, j) element of $\boldsymbol{\Sigma}_0$ equals the (i, j) element of $\boldsymbol{\Sigma}_{\eta\eta}$ divided by $1 - \phi_i\phi_j$. Then it can be verified that $\boldsymbol{\Sigma}_0$ satisfies the weak stationarity condition

$$\boldsymbol{\Sigma}_0 = \boldsymbol{\Phi}\boldsymbol{\Sigma}_0\boldsymbol{\Phi} + \boldsymbol{\Sigma}_{\eta\eta}. \quad (3.10)$$

Assume that the marginal variances of components in $\boldsymbol{\epsilon}_t$ are $\text{var}(\epsilon_{k,t}) = 1, k = 1, \dots, m$, so that $\boldsymbol{\Sigma}_{\epsilon\epsilon}$ is a correlation matrix. The covariance matrix between $\boldsymbol{\epsilon}_t$ and $\boldsymbol{\eta}_{t+1}$ is $\boldsymbol{\Sigma}_{\epsilon\eta}$, and $\boldsymbol{\Sigma}_{\eta\eta}$ is the variance matrix of $\boldsymbol{\eta}_{t+1}$. As usual, we denote $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_T)$, $\mathbf{h} = (\mathbf{h}_1, \dots, \mathbf{h}_T)$, and the collection of the parameters to be estimated by $\boldsymbol{\Theta} = (\boldsymbol{\mu}, \boldsymbol{\phi}, \boldsymbol{\Sigma})$. We permit non-zero correlations within and between $\boldsymbol{\epsilon}_t$ and $\boldsymbol{\eta}_{t+1}$. In the literature, this more general setup of cross correlation only appears in the study by Chan et al. (2006).

3.2.2 Estimation: an MCMC Algorithm

In this section we develop an MCMC algorithm for our proposed MSV model. For convenience in presenting the sampling scheme, we introduce some additional notation. The innovations in (3.4) and (3.5) are assumed to be jointly represented as a column vector \mathbf{Z}_t with length $2m$:

$$\mathbf{Z}_t = \left((\mathbf{H}_t^{-\frac{1}{2}} \mathbf{y}_t)', (\mathbf{h}_{t+1} - \boldsymbol{\mu} - \boldsymbol{\Phi}(\mathbf{h}_t - \boldsymbol{\mu}))' \right)'. \quad (3.11)$$

Then $\mathbf{Z} = (\mathbf{Z}_1, \dots, \mathbf{Z}_{T-1})$ represents an i.i.d. sample from the multivariate normal distribution $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$. This \mathbf{Z} will be used when we specify the proposal distribution in the MH algorithm for the simulation of $\boldsymbol{\Sigma}$.

Our proposed algorithm uses a Cholesky decomposition of $\boldsymbol{\Sigma}$, which enables us to find posterior distributions of latent variables. At each MCMC iteration, when $\boldsymbol{\Sigma}$ is simulated and accepted, we perform a Cholesky decomposition of the form $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}'$, where \mathbf{A} is a lower triangular matrix with strictly positive entries on the main diagonal and \mathbf{A}' is a transpose of \mathbf{A} . We represent \mathbf{A} as

$$\mathbf{A} = \left(\begin{array}{c|c} \mathbf{C} & \mathbf{0} \\ \mathbf{D} & \mathbf{E} \end{array} \right), \quad (3.12)$$

where \mathbf{C} and \mathbf{E} are two lower triangular $m \times m$ matrices. Note the MSV model defined by equations (3.4) and (3.5) can be rewritten as

$$\mathbf{y}_t = \mathbf{H}_t^{\frac{1}{2}} \mathbf{C} \mathbf{e}_t, \quad \mathbf{e}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{m \times m}), \quad (3.13)$$

$$\mathbf{h}_{t+1} = \boldsymbol{\mu} + \boldsymbol{\Phi}(\mathbf{h}_t - \boldsymbol{\mu}) + \mathbf{D} \mathbf{e}_t + \mathbf{E} \mathbf{v}_{t+1}, \quad \mathbf{v}_{t+1} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{m \times m}). \quad (3.14)$$

If we denote $\mathbf{X}_t = (\mathbf{e}_t', \mathbf{v}_{t+1}')'$, then $\mathbf{X}_t, t = 1, \dots, T-1$, are i.i.d. multivariate standard

normal random vectors, such that $\mathbf{X}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{2m \times 2m})$. Then (3.13) and (3.14) can be represented by

$$\mathbf{y}_t = \mathbf{H}_t^{\frac{1}{2}} \mathbf{C} \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \mathbf{I}_{\mathbf{m} \times \mathbf{m}}), \quad (3.15)$$

$$\mathbf{h}_{t+1} = \boldsymbol{\mu} + \boldsymbol{\Phi}(\mathbf{h}_t - \boldsymbol{\mu}) + \mathbf{D}\mathbf{C}^{-1} \mathbf{H}_t^{-\frac{1}{2}} \mathbf{y}_t + \mathbf{E}\mathbf{v}_{t+1}, \quad \mathbf{v}_{t+1} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{\mathbf{m} \times \mathbf{m}}). \quad (3.16)$$

It is seen that the log volatility \mathbf{h}_{t+1} includes a contribution from the previous observation \mathbf{y}_t . The transition equation (3.16) is very convenient for the calculation of the posteriors of latent states. Like the asymmetric univariate SV models, the leverage effect is introduced by $\mathbf{D}\mathbf{C}^{-1}$ in the coefficient of \mathbf{y}_t in (3.16).

As usual, for the use of MCMC sampler, all prior distributions of the parameters in Θ have to be specified in advance. We assume that the prior distributions of μ_i follow univariate normal distributions, where the location hyperparameters are estimated by fitting a univariate ASV model to the corresponding component of the considered multivariate time series, and the variance hyperparameters are all set to 10, so these prior distributions are reasonably flat over their supports. For the prior distributions of ϕ_i , we set univariate norm distributions with restrictions $|\phi_i| < 1, i = 1, \dots, m$. The prior distribution of Σ is $\Sigma \sim \mathcal{IW}(\mathbf{R}_0, n_0)$, which is an inverse Wishart distribution with the probability density function $f_0(\Sigma)$ given by

$$f_0(\Sigma) \propto |\Sigma|^{-\frac{n_0+m+1}{2}} \exp \left\{ -\frac{1}{2} \text{tr}(\mathbf{R}_0 \Sigma^{-1}) \right\}, \quad (3.17)$$

where n_0 is the number of degrees of freedom, \mathbf{R}_0 is a positive-definite matrix called the scale matrix and $\text{tr}(\cdot)$ represents the trace of the argument matrix.

The MCMC algorithm for the MSV model is outlined in Table 3.1 followed by a detailed explanation.

The initial values for the states and latent parameters of the MSV model are set

Table 3.1: MCMC algorithm for the MSV model.

Step 1. Sample $h_{k,t}, k = 1, \dots, m, t = 1, \dots, T$.

Step 2. Sample μ .

Step 3. Sample ϕ .

Step 4. Sample Σ .

Step 5. Repeat Step1 to Step 5.

equal to the estimates obtained from fitting the univariate ASV model to each univariate time series. The variance-covariance matrix Σ is initialized to be a diagonal matrix, where the first m components on the diagonal are set to one and the last m components are the estimates of the asymptotic volatilities of the latent AR(1) processes from the estimated univariate models. For the convenience of presenting the algorithm, by $f(\cdot)$ we denote a generic density function.

Step 1. Sample $\mathbf{h}_t, t = 1, \dots, T$. The full conditional of $\mathbf{h}_t, t = 2, \dots, T - 1$, given that other parameters have been sampled, satisfies

$$\begin{aligned}
 & f(\mathbf{h}_t | \Theta, \mathbf{y}, \mathbf{h}_{-t}) \\
 &= \beta_{1t} f(\mathbf{y}_t | \Theta, \mathbf{h}_t) f(\mathbf{h}_t | \Theta, \mathbf{y}_{t-1}, \mathbf{h}_{t-1}) f(\mathbf{h}_t | \Theta, \mathbf{y}_t, \mathbf{h}_{t+1}) \\
 &= \beta_{2t} \exp \left\{ -\frac{\sum_{i=1}^m h_{i,t}}{2} \right\} \exp \left\{ -\frac{(\mathbf{H}_t^{-\frac{1}{2}} \mathbf{y}_t)' (\mathbf{C}\mathbf{C}')^{-1} (\mathbf{H}_t^{-\frac{1}{2}} \mathbf{y}_t)}{2} \right\} \\
 &\quad \times \exp \left\{ -\frac{(\mathbf{h}_t - \boldsymbol{\mu}_t)' (\mathbf{E}\mathbf{E}')^{-1} (\mathbf{h}_t - \boldsymbol{\mu}_t)}{2} \right\} \\
 &\quad \times \exp \left\{ -\frac{(\mathbf{h}_{t+1} - \boldsymbol{\mu}_{t+1})' (\mathbf{E}\mathbf{E}')^{-1} (\mathbf{h}_{t+1} - \boldsymbol{\mu}_{t+1})}{2} \right\}
 \end{aligned}$$

$$\begin{aligned}
&< \beta_{2t} \exp \left\{ -\frac{\sum_{i=1}^m h_{i,t}}{2} \right\} \exp \left\{ -\frac{(\mathbf{H}_t^{-\frac{1}{2}} \mathbf{y}_t)' (\mathbf{C}\mathbf{C}')^{-1} (\mathbf{H}_t^{-\frac{1}{2}} \mathbf{y}_t)}{2} \right\} \\
&\times \exp \left\{ -\frac{(\mathbf{h}_t - \boldsymbol{\mu}_t)' (\mathbf{E}\mathbf{E}')^{-1} (\mathbf{h}_t - \boldsymbol{\mu}_t)}{2} \right\}, \tag{3.18}
\end{aligned}$$

where β_{1t} and β_{2t} are the two normalizing constants, \mathbf{h}_{-t} denotes all elements of $\mathbf{h} = (\mathbf{h}_1, \dots, \mathbf{h}_T)$ excluding \mathbf{h}_t , and

$$\boldsymbol{\mu}_t = \boldsymbol{\mu} + \boldsymbol{\Phi}(\mathbf{h}_{t-1} - \boldsymbol{\mu}) + \mathbf{D}\mathbf{C}^{-1} \mathbf{H}_{t-1}^{-\frac{1}{2}} \mathbf{y}_{t-1}. \tag{3.19}$$

The full conditionals of \mathbf{h}_1 and \mathbf{h}_T are given in the Appendix. It follows from (3.18) that the full conditional of \mathbf{h}_t is bounded above by a product of three positive and integrable functions. According to the structure of the right-hand side of (3.18), we could simulate \mathbf{h}_t using the vector slice sampler. That is, the dominating function in (3.18) can be simulated through a slice sampler, and then an MH step is needed to sample the left-hand side, which is the full conditional of \mathbf{h}_t . However, as discussed in Chapter 8 of Robert and Casella (2004), the direct multi-dimensional slice sampler may not be efficient. If we inspect the inequality (3.18) carefully, we find that the full conditional of $h_{k,t}$ is bounded above by a product of several univariate positive functions. Then a univariate slice sampler with the MH method can be applied to sample $h_{k,t}$. Thus, the MH algorithm proposed by Jacquier et al. (1994, 1999) or the block simulation approach in Smith and Pitts (2006) are not needed here. The procedure of sampling h_t is summarized in Table 3.2, which splits the multivariate simulation problem into a series of univariate simulation tasks. That is, $h_{k,t}, k = 1, \dots, m$, are simulated cyclicly by the proposed MH method with a univariate slice sampler. It is observed that the sampling of each latent state may also be conducted by the single-move simple MH method introduced in Chapter 2 for the univariate SV models without heavy tails. In this chapter we use the slicing sampling as an example to illustrate that the simulation can be conducted in a simpler way relative to the method used in Chan et al. (2006).

Table 3.2: The MCMC sampler for sampling \mathbf{h}_t from iteration n to $n + 1$.

1. $h_{1,t}^{(n+1)} \sim f(h_{1,t}|h_{1,t-1}^{(n)}, h_{1,t+1}^{(n)}, h_{2,t}^{(n)}, \dots, h_{m,t}^{(n)}, \Theta^{(n)}, \mathbf{y}),$
2. $h_{2,t}^{(n+1)} \sim f(h_{2,t}|h_{2,t-1}^{(n)}, h_{2,t+1}^{(n)}, h_{1,t}^{(n+1)}, h_{3,t}^{(n)}, \dots, h_{m,t}^{(n)}, \Theta^{(n)}, \mathbf{y}),$
- ⋮
- ⋮
- ⋮
- $m.$ $h_{m,t}^{(n+1)} \sim f(h_{m,t}|h_{m,t-1}^{(n)}, h_{m,t+1}^{(n)}, h_{1,t}^{(n+1)}, \dots, h_{m-1,t}^{(n+1)}, \Theta^{(n)}, \mathbf{y}).$

We only give the full conditional of $h_{i,t}$, $i = 1, \dots, m, t = 2, \dots, T - 1$. Let $\mathbf{B} = (\mathbf{C}\mathbf{C}')^{-1}$ and $\mathbf{G} = (\mathbf{E}\mathbf{E}')^{-1}$ be two positive definite matrices, where \mathbf{C} and \mathbf{E} were defined in (3.12). In the following derivation of the full conditional, we assume $y_{i,t} \neq 0$, otherwise the derivation requires a minor adjustment. Analogous to the formula (3.18), it is easy to see that the full conditional of $h_{i,t}$ satisfies the following:

$$\begin{aligned}
& f(h_{i,t}|h_{-i,t}) \\
&= \eta_{1t} \exp \left\{ -\frac{h_{i,t}}{2} \right\} \exp \left\{ -\frac{b_{i,i}y_{i,t}^2 e^{-h_{i,t}} + 2y_{i,t}e^{-\frac{h_{i,t}}{2}} \sum_{j=1, j \neq i}^m (b_{i,j} y_{j,t} e^{-\frac{h_{j,t}}{2}})}{2} \right\} \\
&\quad \times \exp \left\{ -\frac{g_{i,i} (h_{i,t} - \mu_{i,t})^2 + 2(h_{i,t} - \mu_{i,t}) \sum_{j=1, j \neq i}^m (g_{i,j} (h_{j,t} - \mu_{j,t}))}{2} \right\} \\
&\quad \times \exp \left\{ -\frac{g_{i,i} (h_{i,t+1} - \mu_{i,t+1})^2 + 2(h_{i,t+1} - \mu_{i,t+1}) \sum_{j=1, j \neq i}^m (g_{i,j} (h_{j,t+1} - \mu_{j,t+1}))}{2} \right\} \\
&= \eta_{2t} \exp \left\{ -\frac{h_{i,t}}{2} \right\} \exp \left\{ -\frac{\left(e^{-\frac{h_{i,t}}{2}} + \frac{\sum_{j=1, j \neq i}^m (b_{i,j} y_{j,t} e^{-\frac{h_{j,t}}{2}})}{b_{i,i}y_{i,t}} \right)^2}{2 \frac{1}{b_{i,i}y_{i,t}^2}} \right\} \\
&\quad \times \exp \left\{ -\frac{\left((h_{i,t} - \mu_{i,t}) + \frac{\sum_{j=1, j \neq i}^m (g_{i,j} (h_{j,t} - \mu_{j,t}))}{g_{i,i}} \right)^2}{2 \frac{1}{g_{i,i}}} \right\} \\
&\quad \times \exp \left\{ -\frac{g_{i,i} (h_{i,t+1} - \mu_{i,t+1})^2 + 2(h_{i,t+1} - \mu_{i,t+1}) \sum_{j=1, j \neq i}^m (g_{i,j} (h_{j,t+1} - \mu_{j,t+1}))}{2} \right\}
\end{aligned}$$

$$\begin{aligned}
&< \eta_{2t} \exp \left\{ -\frac{h_{i,t}}{2} \right\} \exp \left\{ -\frac{\left(e^{-\frac{h_{i,t}}{2}} + \frac{\sum_{j=1, j \neq i}^m (b_{i,j} y_{j,t} e^{-\frac{h_{j,t}}{2}})}{b_{i,i} y_{i,t}} \right)^2}{2 \frac{1}{b_{i,i} y_{i,t}^2}} \right\} \\
&\times \exp \left\{ -\frac{\left((h_{i,t} - \mu_{i,t}) + \frac{\sum_{j=1, j \neq i}^m (g_{i,j} (h_{j,t} - \mu_{j,t}))}{g_{i,i}} \right)^2}{2 \frac{1}{g_{i,i}}} \right\} \\
&= \eta_{2t} \exp \left\{ -\frac{(h_{i,t} - \hat{\mu}_{i,t})^2}{2 \frac{1}{g_{i,i}}} \right\} \exp \left\{ -\frac{\left(e^{-\frac{h_{i,t}}{2}} + \frac{\sum_{j=1, j \neq i}^m (b_{i,j} y_{j,t} e^{-\frac{h_{j,t}}{2}})}{b_{i,i} y_{i,t}} \right)^2}{2 \frac{1}{b_{i,i} y_{i,t}^2}} \right\}, \quad (3.20)
\end{aligned}$$

where $\hat{\mu}_{i,t} = \mu_{i,t} - \frac{1}{2g_{i,i}} - \frac{\sum_{j=1, j \neq i}^m (g_{i,j} (h_{j,t} - \mu_{j,t}))}{g_{i,i}}$, and η_{1t} and η_{2t} are the two normalizing constants. The dominating distribution is a mixture of univariate normal of $h_{i,t}$ and a truncated normal of $\exp(h_{i,t}/2)$. So the full conditional of $h_{i,t}$ can be sampled by the MH method, where the proposal distribution is the dominating distribution that can be simulated using the slice sampler. The slice sampling of the proposal distribution can be found in the Appendix.

Step 2. Sample $\boldsymbol{\mu}$.

The full conditional of $\boldsymbol{\mu}$ is given in the Appendix, which is a multivariate normal distribution. The simulation of this parameter can be performed using most of the statistical software packages.

Step 3. Sample $\boldsymbol{\phi}$.

Denote $\mathbf{r}_t = \mathbf{h}_{t+1} - \boldsymbol{\mu} - \boldsymbol{\Phi}(\mathbf{h}_{t-1} - \boldsymbol{\mu}) - \mathbf{D}\mathbf{C}^{-1}\mathbf{H}_{t-1}^{-\frac{1}{2}}\mathbf{y}_{t-1}$ and $\mathbf{r} = \{\mathbf{r}_1, \dots, \mathbf{r}_{T-1}\}$, then the full conditional of $\boldsymbol{\phi}$ is

$$\begin{aligned}
f(\boldsymbol{\phi} | \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{y}, \mathbf{h}) &\propto f(\mathbf{h}_1 | \boldsymbol{\Sigma}_0) \prod_{t=2}^{T-1} f(\mathbf{r}_t | \mathbf{E}) f(\boldsymbol{\phi}) \\
&\propto f(\mathbf{h}_1 | \boldsymbol{\Sigma}_0) \prod_{t=2}^{T-1} \exp\left(-\frac{1}{2} \mathbf{r}_t' (\mathbf{E}\mathbf{E}')^{-1} \mathbf{r}_t\right) f(\boldsymbol{\phi}), \quad (3.21)
\end{aligned}$$

where $f(\mathbf{h}_1 | \boldsymbol{\Sigma}_0)$ is the conditional density of \mathbf{h}_1 and $f(\boldsymbol{\phi})$ is the prior density of $\boldsymbol{\phi}$, re-

spectively. Because of the weak stationarity condition (3.10), the full conditional of ϕ is not a standard distribution function and its simulation is not routine. But, it is noticed that $\prod_{t=2}^{T-1} \exp(-\frac{1}{2}\mathbf{r}_t'\mathbf{E}\mathbf{E}'\mathbf{r}_t)f(\phi)$ is proportional to a truncated multivariate normal density. Hence, the full conditional (3.21) can be sampled by the MH algorithm with the proposal density proportional to $\prod_{t=2}^{T-1} \exp(-\frac{1}{2}\mathbf{r}_t'\mathbf{E}\mathbf{E}'\mathbf{r}_t)f(\phi)$ truncated in the interval $(-1, 1)$. To sample the multivariate truncated proposal, for each k we calculate the full conditional of ϕ_k , which is a truncated univariate normal distribution and can be sampled easily. The proposal density of $(\phi|\mathbf{r})$ is given in the Appendix.

Step 4. Sample Σ . The full conditional of Σ , given the prior density $f_0(\Sigma)$ in (3.17), is given by

$$f(\Sigma|\mathbf{y}, \mathbf{h}, \boldsymbol{\mu}, \phi) \propto f(\Sigma) \prod_{t=1}^{T-1} f(\mathbf{Z}_t|\mathbf{0}, \Sigma) f(\mathbf{y}_T|\mathbf{h}_T, \Sigma_{\epsilon\epsilon}) f(h_1|\boldsymbol{\mu}, \Sigma_0),$$

$$\propto |\Sigma|^{-\frac{n+m+1}{2}} \exp\left\{-\frac{1}{2}\text{tr}(\mathbf{R}\Sigma^{-1})\right\} \times g_1(\Sigma), \quad (3.22)$$

$$g_1(\Sigma) = |\Sigma_0|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\mathbf{h}'_1\Sigma_0^{-1}\mathbf{h}_1\right\}$$

$$\times |\Sigma_{\epsilon\epsilon}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\mathbf{y}'_T \exp\left(-\frac{\mathbf{h}_T}{2}\right)\Sigma_{\epsilon\epsilon}^{-1} \exp\left(-\frac{\mathbf{h}_T}{2}\right)\mathbf{y}_T\right\}, \quad (3.23)$$

where $\mathbf{R} = \mathbf{R}_0 + \Sigma_h$, $\Sigma_h = \mathbf{Z}\mathbf{Z}'$, $n = n_0 + T - 1$, and $f(\mathbf{Z}_t|\mathbf{0}, \Sigma)$ is the multivariate normal density function of \mathbf{Z}_t with mean $\mathbf{0}$ vector and variance-covariance matrix Σ , respectively. Since this full conditional is not a simple distribution, it can not be simulated directly. Moreover, draws from this full conditional may not satisfy the condition that the upper-left $m \times m$ square matrix is a correlation matrix. We discuss simulation from this full conditional in the next subsection.

3.2.3 Simulation of Σ

We now focus on a method for sampling Σ and discuss how to set up its prior distribution. In the literature, the Bayesian inference of special matrices such as the correlation matrix is an active research area. Liu (2001), Zhang et al. (2006), Liu and Daniels (2006) and Liu (2008) propose parameter-extended methods to simulate the correlation matrices in multivariate probit and multivariate regression analysis models. The main idea is simple, first a variance matrix is drawn from an inverse Wishart distribution, and then, using a decomposition of this matrix, it is transformed into a diagonal matrix and a correlation matrix. The latter is accepted through the MH algorithm. Motivated by that approach, we propose a parameter-extended Metropolis-Hastings (PX-MH) algorithm for the simulation of our constrained covariance matrix.

It can be seen from (3.22) that the first part of the full conditional is the kernel of an inverse Wishart distribution $\mathcal{IW}(\Sigma|\mathbf{R}, n)$. We choose this distribution as the proposal distribution for the MH algorithm when sampling the full conditional of Σ . The probability density function of this proposal distribution is

$$f_{iw}(\mathbf{W}|\mathbf{Z}) = c^{-1}|\mathbf{R}|^{-\frac{m}{2}}|\mathbf{W}|^{-\frac{T+m}{2}} \exp\left\{-\frac{\text{tr}(\mathbf{R}\mathbf{W}^{-1})}{2}\right\}, \quad (3.24)$$

where c is the normalizing constant and $|\mathbf{D}|$ is the determinant of matrix \mathbf{D} . The reason that we use \mathbf{W} rather than Σ is that a draw from this distribution does not have to meet the constraint that the upper-left $m \times m$ submatrix is a correlation matrix. To obtain the required matrix Σ we perform the following transformation:

$$\Sigma = \mathbf{D}^{\frac{1}{2}}\mathbf{W}\mathbf{D}^{\frac{1}{2}}, \quad (3.25)$$

where $\mathbf{D} = \text{diag}(w_{11}, \dots, w_{mm}, 1, \dots, 1)$ and $w_{ii}, i = 1, \dots, m$, are the main diagonal ele-

ments of \mathbf{W} . The Jacobian of the transformation $\mathbf{W} \rightarrow (\boldsymbol{\Sigma}, \mathbf{D})$ is

$$\mathbf{J}_{\mathbf{W} \rightarrow \boldsymbol{\Sigma}, \mathbf{D}} = |\mathbf{D}|^{\frac{3m-1}{2}}. \quad (3.26)$$

By transforming the variables, we can find that the joint distribution of $(\boldsymbol{\Sigma}, \mathbf{D})$ is of the form

$$\begin{aligned} p(\boldsymbol{\Sigma}, \mathbf{D} | \mathbf{y}, \mathbf{h}, \boldsymbol{\mu}, \boldsymbol{\phi}) &= c^{-1} |\mathbf{D}|^{\frac{3m-1}{2}} |\mathbf{R}|^{-\frac{m}{2}} |\mathbf{D}|^{-\frac{T+m}{2}} \exp \left\{ -\frac{\text{tr}(\mathbf{R} \mathbf{D}^{\frac{1}{2}} \boldsymbol{\Sigma}^{-1} \mathbf{D}^{\frac{1}{2}})}{2} \right\} \\ &= c^{-1} |\mathbf{D}|^{-\frac{T-2m+1}{2}} |\mathbf{R}|^{-\frac{m}{2}} \exp \left\{ -\frac{\text{tr}(\mathbf{R} \mathbf{D}^{\frac{1}{2}} \boldsymbol{\Sigma}^{-1} \mathbf{D}^{\frac{1}{2}})}{2} \right\}. \end{aligned} \quad (3.27)$$

It would appear that the marginal density of $\boldsymbol{\Sigma}$ can be obtained by integrating out \mathbf{D} from the above joint density. As discussed in Zhang et al. (2006), this is virtually impossible unless $\boldsymbol{\Sigma}$ is a nonsingular diagonal matrix. In our specified MSV model, $\boldsymbol{\Sigma}$ can be a more general variance-covariance matrix including cross correlations. In other words, we are not able to find an analytical representation of the posterior distribution of $\boldsymbol{\Sigma}$ through the above joint distribution. So the simulation of $\boldsymbol{\Sigma}$ is not to be done directly.

Although the closed form for the marginal density of $\boldsymbol{\Sigma}$ is unavailable in general, it is noticed that under the transformation (3.25), the joint density of $(\boldsymbol{\Sigma}, \mathbf{D})$ has an explicit representation. Instead of accepting $\boldsymbol{\Sigma}$ alone, we will accept $\boldsymbol{\Sigma}$ and \mathbf{D} together. At this point, the auxiliary parameter matrix \mathbf{D} makes the sampling of $\boldsymbol{\Sigma}$ quite easy. Based on this argument, the sampling of the full conditional of $\boldsymbol{\Sigma}$ is replaced by the simulation of the joint full conditional of $(\boldsymbol{\Sigma}, \mathbf{D})$, and $\boldsymbol{\Sigma}$ then is accepted by the MH algorithm.

Like the usual PX-MH algorithms, for our model, we apply this strategy by focusing on the joint posterior distribution $f(\boldsymbol{\Sigma}, \boldsymbol{\mu}, \boldsymbol{\phi}, \mathbf{D} | \mathbf{y})$. The full conditionals of $\boldsymbol{\mu}$ and $\boldsymbol{\phi}$

are the same as before. The full conditional distribution for Σ , similar to the full conditional (3.22), is replaced by the joint posterior distribution as follows:

$$f(\Sigma, \mathbf{D}|\mathbf{Z}) \propto |\Sigma|^{-\frac{n+m+1}{2}} \exp\left\{-\frac{1}{2}\text{tr}(\mathbf{R}^{-1}\Sigma)\right\} \times g_1(\Sigma), \quad (3.28)$$

where $g_1(\Sigma)$ was defined in (3.23). The MH algorithm for sampling the constrained matrix is the following.

PX-MH Algorithm

Set initial values of $(\Sigma^{(0)}, \mathbf{D}^{(0)})$ through the setting $\mathbf{W}^{(0)} = \mathbf{D}^{(0)\frac{1}{2}}\Sigma^{(0)}\mathbf{D}^{(0)\frac{1}{2}}$ to an initial covariance matrix, and let $n = 1$. Calculate $\Sigma_{\mathbf{h}}^{(n)}$ from the definition (3.22).

1. Generate (Σ^*, \mathbf{D}^*) by generating $\mathbf{W}^* = \mathbf{D}^{*\frac{1}{2}}\Sigma^*\mathbf{D}^{*\frac{1}{2}}$ from the inverse Wishart distribution $\mathcal{IW}(\mathbf{W}|\mathbf{R}, n)$.
2. Calculate Σ_0^* according the stationary condition (3.10)

$$\Sigma_0^* = \Phi\Sigma_0^*\Phi + \Sigma_{\eta\eta}^*. \quad (3.29)$$

3. Calculate the acceptance probability

$$\beta = \min\left\{1, \frac{g_1(\Sigma^*)}{g_1(\Sigma^{(n)})}\right\}. \quad (3.30)$$

4. Generate α uniformly from the interval $[0,1]$. If $\alpha \leq \beta$ then $\Sigma^{(n+1)} = \Sigma^*$ otherwise $\Sigma^{(n+1)} = \Sigma^{(n)}$, keep the generated matrix from the last iteration.

To put the above MH algorithm in use, the prior distribution of (Σ, \mathbf{D}) has to be specified in advance. In our proposed algorithm, instead of setting a prior distribution for Σ explicitly, we set up a prior distribution for \mathbf{W} , and transform it to a joint prior

distribution of (Σ, \mathbf{D}) . In the algorithm we construct an inverse Wishart prior for \mathbf{W} , which is a diagonal matrix where the first m elements on the main diagonal equal 1 and the other m elements are estimated marginal variances of the latent AR(1) processes. To ensure that this prior distribution has high prior variance the degrees of freedom is set to $2m + 2$.

3.3 Simulation Studies

This section summarizes results of our simulation studies, designed to illustrate the ability of our method to recover the parameters and correlations among innovations. A four-dimensional correlated time series was generated from a MSV model. Table 3.3 includes the variance-covariance matrix in boldface and the correlations in italics for data generation, and the true location and persistence parameters are provided in Table 3.5 in boldface. Those parameter values are close to the estimates from asset return data, which will be seen in the next section. There are 2000 observations generated from this specified MSV model. The MCMC algorithm was run 50,000 iterations with the first 10,000 iterations discarded as the burn-in and the following 40,000 iterations used for inference. In order for the prior distribution of the unconstrained variance matrix to have a higher variability, its degrees of freedom was set to 10.

Histograms and time series of samples from the full conditionals of selected parameters are given in Figures 3.1 and 3.2, respectively. Those plots show that the sample paths of the location and persistence parameters are well mixed.

Table 3.4 provides the estimated variance-covariance matrix in the lower triangular matrix and the true cross correlations in the upper triangular matrix. Compared with the values in Table 3.3, the estimated covariance matrix and the correlations are similar to their true counterparts. Table 3.5 includes the estimated location, persistence and volatility parameters of the latent AR(1) processes, where the numbers in the paren-

Table 3.3: The variance-covariance matrix (in boldface) and correlations (in italics) used for data generation.

1.00	<i>0.40</i>	<i>0.43</i>	<i>0.47</i>	-0.23	<i>-0.07</i>	<i>-0.13</i>	<i>-0.19</i>
0.39	1.00	<i>0.32</i>	<i>0.42</i>	<i>-0.17</i>	0.07	<i>-0.10</i>	<i>-0.19</i>
0.43	0.32	1.00	<i>0.44</i>	<i>-0.17</i>	<i>0.00</i>	-0.10	<i>-0.19</i>
0.47	0.42	0.44	1.00	<i>-0.20</i>	<i>0.00</i>	<i>-0.13</i>	-0.30
-0.07	-0.06	-0.05	-0.06	0.09	<i>0.47</i>	<i>0.78</i>	<i>0.76</i>
-0.01	0.01	0.00	0.00	0.02	0.02	<i>0.71</i>	<i>0.53</i>
-0.04	-0.03	-0.03	-0.04	0.07	0.03	0.09	<i>0.63</i>
-0.05	-0.05	-0.05	-0.08	0.06	0.02	0.05	0.07

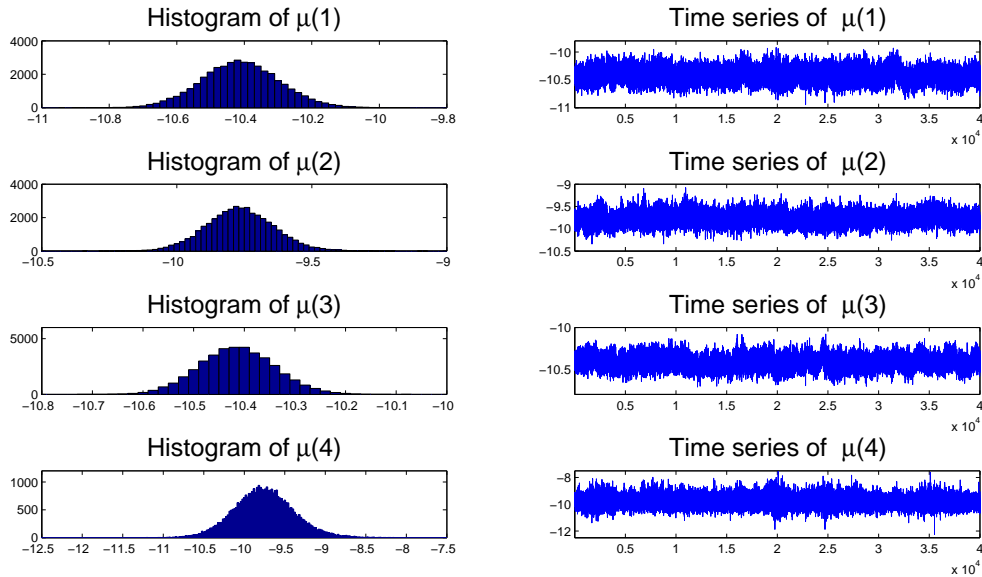


Figure 3.1: Histograms and time series of samples from the full conditionals of the location parameters based on the generated return data.

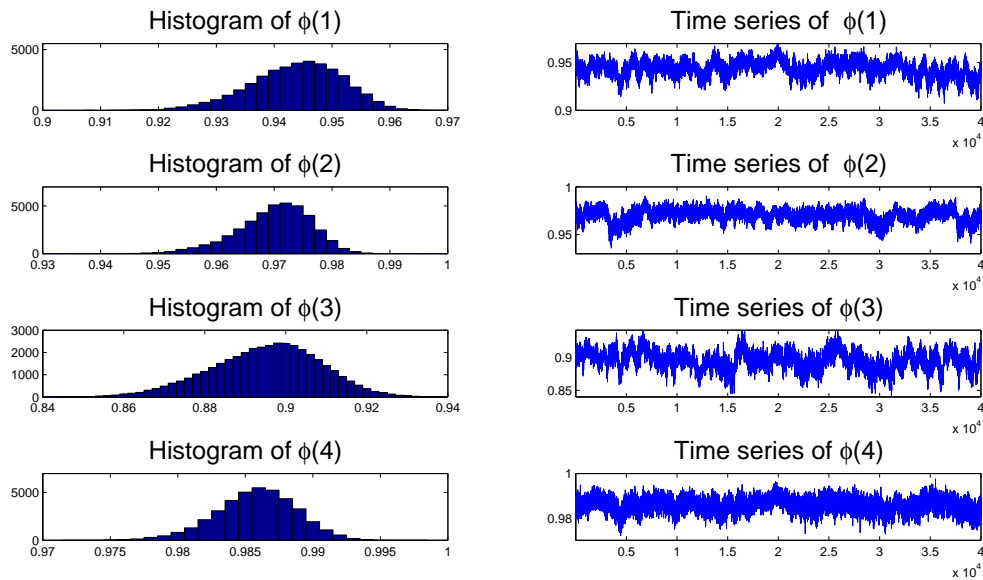


Figure 3.2: Histograms and time series of samples from the full conditionals of the persistence parameters based on the generated return data.

theses are standard errors. The columns titled 'ASV' provide the estimated parameters and standard errors when each component time series was fitted by the univariate ASV model. Although the estimated parameters of the latent AR(1) processes from both of the ASV and MSV models are close to their true values, the MSV model has several advantages. First, it not only models the leverage effect between the asset returns and the corresponding volatilities but also the cross correlations. The correlation matrix captures the complex dependence among the dynamics of the asset returns and latent volatilities. Second, for forecasting volatility, the MSV model provides more information than univariate SV models. Third, in portfolio selection, the MSV model gives us a more complete picture about the relationships among individual components.

Figure 3.3 depicts the absolute simulated returns, the simulated volatilities and the estimated volatilities for the fourth component of the generated returns. We can see that these estimated volatilities resemble the true volatilities.

Table 3.4: Bayesian estimates of the variance-variance (in boldface) and correlations (in italics) from the generated return data.

1.00	<i>0.40</i>	<i>0.39</i>	<i>0.45</i>	<i>-0.18</i>	<i>-0.06</i>	<i>-0.13</i>	<i>-0.12</i>
0.40	1.00	<i>0.34</i>	<i>0.42</i>	<i>-0.19</i>	<i>0.14</i>	<i>-0.09</i>	<i>-0.14</i>
0.45	0.34	1.00	<i>0.47</i>	<i>-0.19</i>	<i>0.02</i>	<i>-0.05</i>	<i>-0.11</i>
0.46	0.42	0.47	1.00	<i>-0.22</i>	<i>0.14</i>	<i>-0.12</i>	<i>-0.26</i>
-0.05	-0.06	-0.05	-0.06	0.09	<i>0.65</i>	<i>0.81</i>	<i>0.85</i>
-0.01	0.02	0.00	0.02	0.03	0.03	<i>0.70</i>	<i>0.61</i>
-0.04	-0.03	-0.02	-0.04	0.08	0.04	0.11	<i>0.77</i>
-0.03	-0.03	-0.03	-0.06	0.06	0.03	0.06	0.06

Table 3.5: Comparison between true and estimated parameters from the ASV and MSV models based on generated data.

Series	μ_i			ϕ_i			σ_i		
	True	ASV	MSV	True	ASV	MSV	True	ASV	MSV
1	-10.56	-10.48 (0.13)	-10.41 (0.12)	0.93	0.95 (0.01)	0.94 (0.01)	0.09	0.08 (0.02)	0.09 (0.01)
2	-10.01	-10.40 (0.12)	-9.81 (0.13)	0.97	0.97 (0.01)	0.97 (0.01)	0.02	0.03 (0.01)	0.03 (0.01)
3	-10.56	-10.40 (0.02)	-10.44 (0.08)	0.90	0.90 (0.02)	0.90 (0.03)	0.09	0.11 (0.02)	0.11 (0.02)
4	-9.95	-9.86 (0.32)	-9.75 (0.38)	0.98	0.98 (0.01)	0.99 (0.003)	0.07	0.07 (0.01)	0.06 (0.01)

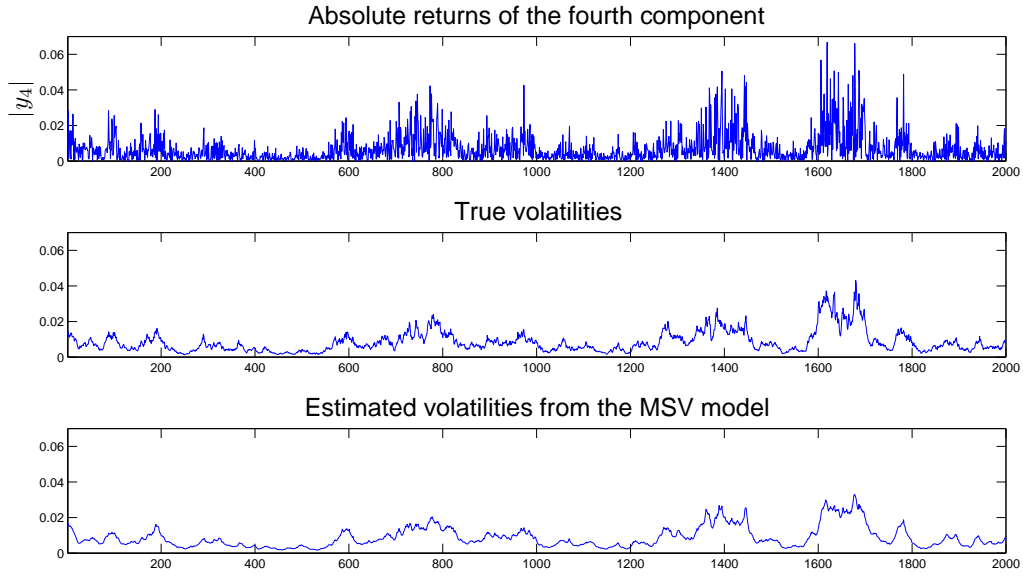


Figure 3.3: Volatility comparison for the fourth component of the generated time series of returns.

To assess goodness-of-fit of the MSV model we can compare the root mean square errors (RMSE) of the true (simulated) and estimated volatilities from the MSV and ASV models. The RMSEs can be calculated using the following formula

$$v_k = \sqrt{\frac{1}{T} \sum_{t=1}^T (\exp(h_{k,t}) - \exp(\hat{h}_{k,t}))^2}, \quad k = 1, \dots, 4, \quad (3.31)$$

where $h_{k,t}$ and $\hat{h}_{k,t}$, $k = 1, \dots, 4, t = 1, \dots, T$, are the true and estimated log volatilities, respectively. Table 3.6 summarizes the RMSEs from both of the MSV and ASV models. We can see that the RMSEs from the MSV model are smaller than the ones for the ASV models, suggesting that the MSV model fits the simulated data better. We repeated the simulations several times and obtained similar results.

We may check the differences between the true and estimated volatilities from the

Table 3.6: RMSE for the ASV and MSV models.

Component	ASV	MSV
1	0.0014	0.0013
2	0.0013	0.0012
3	0.0014	0.0013
4	0.0023	0.0013

MSV model. Figure 3.4 graphs these differences for the four components of the generated return time series. It can be seen that the time series of differences are around zero and there is no explicit pattern, indicating reasonable agreement between estimated and true volatilities.

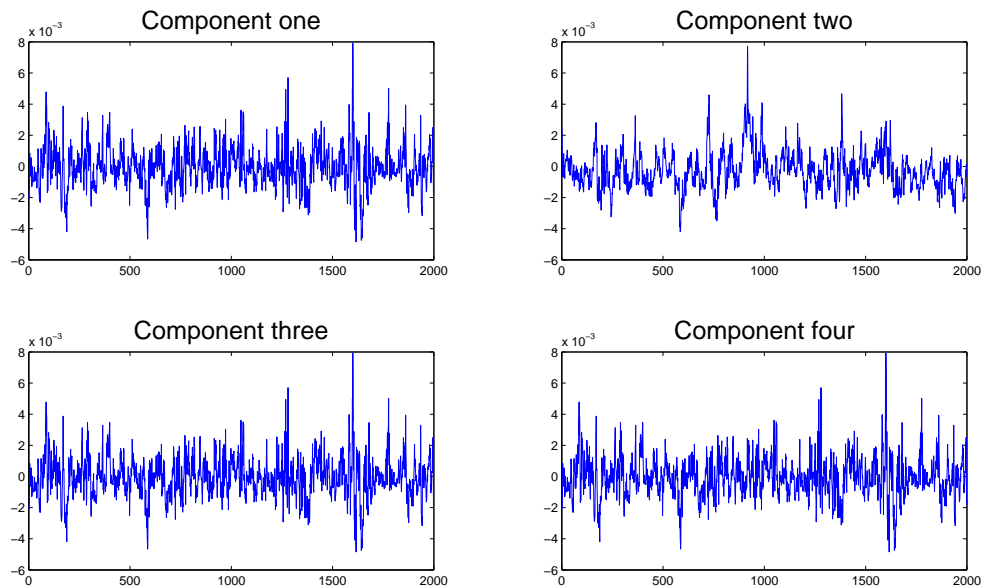


Figure 3.4: Differences between the true and estimated volatilities from the MSV model based on the generated return data.

An alternative method to assess the overall fit of the MSV model can be conducted by means of the probability integral transform (PIT) test proposed in Diebold et al. (1998, 1999), which was also employed in Chapter 2. Suppose that $\{f_t(y_t|\mathcal{F}_{t-1})\}_{t=1}^T$

is a sequence of densities that governs the time series \mathbf{y}_t , and $\{p_t(\mathbf{y}_t|\mathcal{F}_{t-1})\}_{t=1}^T$ is the corresponding sequence of the one-step ahead density forecast of \mathbf{y}_t , where \mathcal{F}_t is the information known up to time t . Then the PIT of \mathbf{y}_t is defined as the following

$$u_t = \int_{A_t} p_t(\mathbf{x}|\mathcal{F}_{t-1})d\mathbf{x}, \quad (3.32)$$

where $A_t = \{\mathbf{x}; -\infty < \mathbf{x} < \mathbf{y}_t\}$. For the univariate SV models y_t is observed scalar time series and p_t is a univariate normal density function. Under the null hypothesis that the sequence $\{p_t(y_t|\mathcal{F}_{t-1})\}_{t=1}^T$ coincides with $\{f_t(y_t|\mathcal{F}_{t-1})\}_{t=1}^T$, the sequence $\{u_t\}_{t=1}^T$ is i.i.d. according to the uniform distribution over the interval $[0, 1]$. In our MSV model, \mathbf{y}_t is the observed vector of returns at time t and its one-step predictive distribution is a multivariate normal distribution. Since p_t in (3.32) is a multivariate normal density, the PITs u_t are no longer uniformly distributed on the interval $[0,1]$. We can still use the method in Diebold et al. (1998, 1999) by successive conditioning. Considering the PIT of \mathbf{y}_t , the conditional density of \mathbf{y}_t can be factored into conditionals as

$$p_t(\mathbf{y}_t|\mathcal{F}_{t-1}) = p(y_{m,t}|y_{1:m-1,t}, \mathcal{F}_{t-1}) \dots p(y_{2,t}|y_{1,t}, \mathcal{F}_{t-1})p(y_{1,t}|\mathcal{F}_{t-1}), \quad (3.33)$$

where $y_{1:k,t} = (y_{1,t}, \dots, y_{k,t})$, $2 \leq k \leq m - 1$, is the partial information of \mathbf{y}_t .

Define $\mathbf{u}_t = (u_{1,t}, \dots, u_{m,t})'$. The PIT of \mathbf{y}_t based on the factored conditional density is given by

$$u_{k,t} = \int_{-\infty}^{y_{k,t}} p(x|y_{1:k-1,t}, \mathcal{F}_{t-1})dx, \quad k = 2, \dots, m - 1, \quad (3.34)$$

and

$$u_{1,t} = \int_{-\infty}^{y_{1,t}} p(x|\mathcal{F}_{t-1})dx, \quad (3.35)$$

where $y_{i,t}, i = 1, \dots, k$, are the components taking scalar real values.

We can also factor the joint prediction density into the conditionals starting from $y_{k,t}$, where k could be any integer number between 1 to m . If all PITs $u_{k,t}$ are stacked together as a sample, according to Diebold et al. (1998, 1999), these PITs are i.i.d. uniform over the interval $[0,1]$. To calculate the PITs (3.34) and (3.35) we use the auxiliary particle filter in Pitt and Shephard (1999b), which is also employed in Chib et al. (2006) for density forecast. The detailed procedure is provided in the Appendix.

Figure 3.5 presents the scatter plot and the histogram of the PITs, and Figure 3.6 is the comparison between the empirical CDF of PITs and the theoretical CDF of the uniform distribution over the support $[0,1]$. As discussed in Chapter 2, we can use the Kolmogorov-Smornov test even when the population values are estimated. The KS statistic is 0.0067 which is less than 0.0152, the critical value at the 5% significance level. Since the p -value for this test is 0.8647, we do not reject the null hypothesis that the PITs are i.i.d. uniform on $[0, 1]$ and conclude that the MSV model agrees with the generated return data.

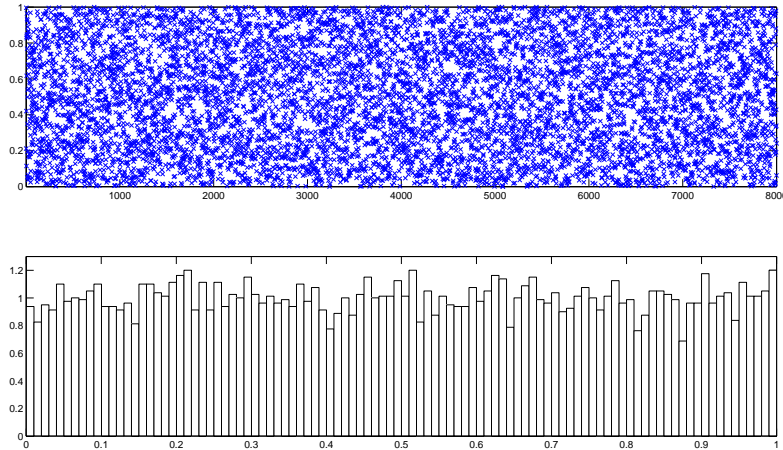


Figure 3.5: Scatter plot (top) and histogram (bottom) of the PITs for the generated return data.

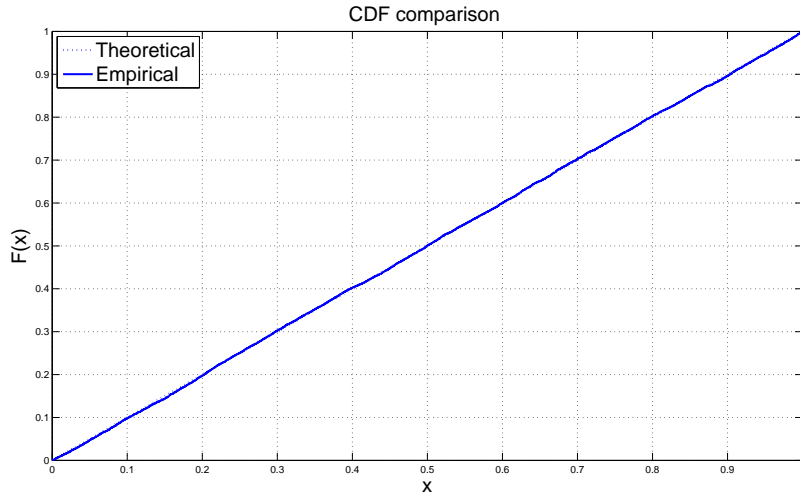


Figure 3.6: Comparison between the empirical CDF of the PITs under the MSV model based on the generated returns and the theoretic CDF of the uniform distribution over the interval $[0, 1]$.

The study of the proposed approach for the general MSV model on the simulated return data demonstrates that our variance-covariance sampling scheme via the inverse Wishart distribution works well in terms of parameter estimation. With a linear transformation to the variance-covariance matrix simulated from an inverse Wishart proposal distribution, the transformed constraint matrix is accepted with a certain probability. Also the slice sampler within the MCMC method is easier to implement for the simulation of the latent random variables, which avoids the commonly used complicated MH algorithms. The correlation structure among innovations of the observation equations and that of the latent processes are then easily estimated.

3.4 Application

In this section we provide two examples in analyzing financial return data from the stock market. The first example is based on four-dimensional return data from different industries and the second is five-dimensional return data from the Canadian banks. The proposed MCMC method was run 50,000 iterations in each of the two applications,

where the first 10,000 sampled points were burn-in and 40,000 iterations recorded for the inference of parameters and log volatilities.

3.4.1 Stock Returns from Different Industries

The data is a collection of time series of stock daily returns of IBM, Toyota, Walmart, and Citibank whose historical prices were downloaded from the web site finance.yahoo.com from Jan. 3, 2003 to Nov. 13, 2009 yielding 1730 observations for each stock. For comparison purpose, each individual time series of asset returns was also fitted by the univariate ASV model. The estimated variance-covariance matrix in boldface and correlations in italics are presented in Table 3.7. It is observed from the estimated correlation matrix that the innovations of the measurement equations have high positive correlations. The same evidence can also be found for innovations of the latent processes. Moreover, each measurement equation is also negatively correlated with the corresponding volatility process except for the returns of the Toyota stock, where the returns and the latent volatilities are positively correlated. The interesting thing here is that each return time series also has a negative correlation with other latent innovations of the AR(1) processes except for the process of Toyota returns, which is positively correlated with the latent innovation of Citibank and has no correlation to the AR(1) process of Walmart.

Table 3.8 summarizes estimated location, persistence and variance parameters of the latent AR(1) processes with standard errors in the parentheses from the MSV and ASV models. It is found that the volatility parameters σ_i of the latent processes of IBM and Walmart stocks are higher than the other two stocks. This can be explained by the corresponding smaller persistence parameters. Figure 3.7 graphs the histograms and dynamics of samples from the full conditionals of location parameters from the MSV model. These graphs show that the sampled time series from the corresponding

Table 3.7: Bayesian estimates of the unconditional variance-covariance (in boldface) and correlation matrices (in italics) from the return time series of four industries.

1.00	<i>0.39</i>	<i>0.43</i>	<i>0.47</i>	-0.24	-0.04	-0.13	-0.17
0.39	1.00	<i>0.32</i>	<i>0.42</i>	-0.18	<i>0.06</i>	-0.11	-0.18
0.43	0.32	1.00	<i>0.44</i>	-0.15	<i>0.00</i>	-0.09	-0.20
0.47	0.42	0.44	1.00	-0.21	<i>0.03</i>	-0.14	-0.30
-0.07	-0.06	-0.05	-0.06	0.09	<i>0.51</i>	<i>0.74</i>	<i>0.72</i>
-0.01	0.01	0.00	0.01	0.02	0.02	<i>0.69</i>	<i>0.46</i>
-0.04	-0.03	-0.03	-0.04	0.07	0.03	0.09	<i>0.61</i>
-0.05	-0.05	-0.05	-0.08	0.06	0.02	0.05	0.07

Table 3.8: Comparison of estimated parameters between ASV and MSV models.

	μ_i		ϕ_i		σ_i	
	ASV	MSV	ASV	MSV	ASV	MSV
Stock						
IBM	-10.47 (0.24)	-10.56 (0.11)	0.98 (0.01)	0.93 (0.01)	0.04 (0.01)	0.09 (0.02)
Toyota	-10.01 (0.13)	-10.01 (0.15)	0.99 (0.02)	0.97 (0.02)	0.03 (0.02)	0.02 (0.01)
Walmart	-10.53 (0.19)	-10.56 (0.08)	0.96 (0.02)	0.90 (0.01)	0.04 (0.02)	0.09 (0.04)
Citibank	-10.12 (0.74)	-9.95 (0.32)	0.99 (0.002)	0.98 (0.01)	0.04 (0.01)	0.07 (0.02)

full conditionals are mixed well, and the location parameters can be estimated by the Bayesian means of samples from their full conditionals.

Figure 3.8 includes the scatter plot and the histogram for the PITs, and Figure 3.9 includes the CDF plot of PITs together with theoretical CDF plot of the uniform distribution over the interval $[0, 1]$. The KS test statistic for the PITs is 0.0114 with a critical value of 0.0163 at the 5% significance level. Therefore, at the 5% level, we do not reject the null hypothesis that the PITs follow the uniform distribution over the interval $[0, 1]$ and conclude that the four-dimensional MSV model fits the four-dimensional return

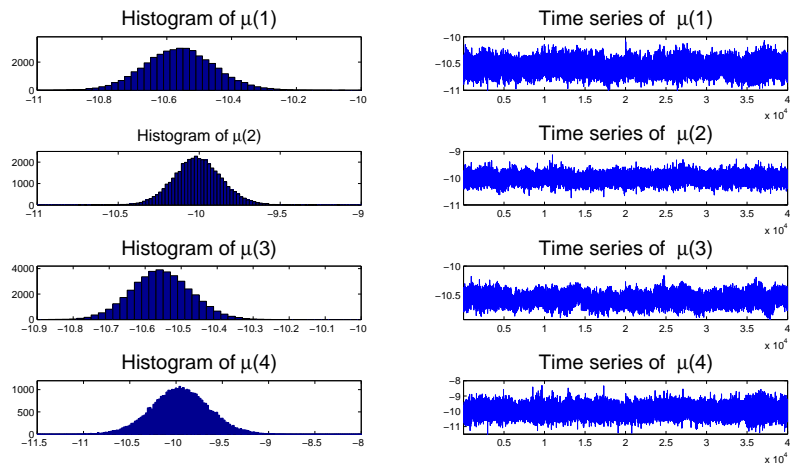


Figure 3.7: Histograms and dynamics of samples of the location parameters of the MSV model for the four industry stock return data.

data well.

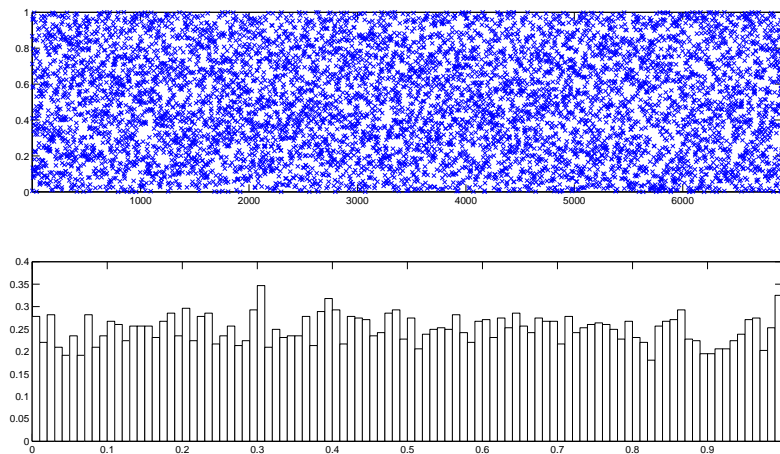


Figure 3.8: Scatter (top) and histogram (bottom) plots of the PITs for the four-dimensional industry return data.

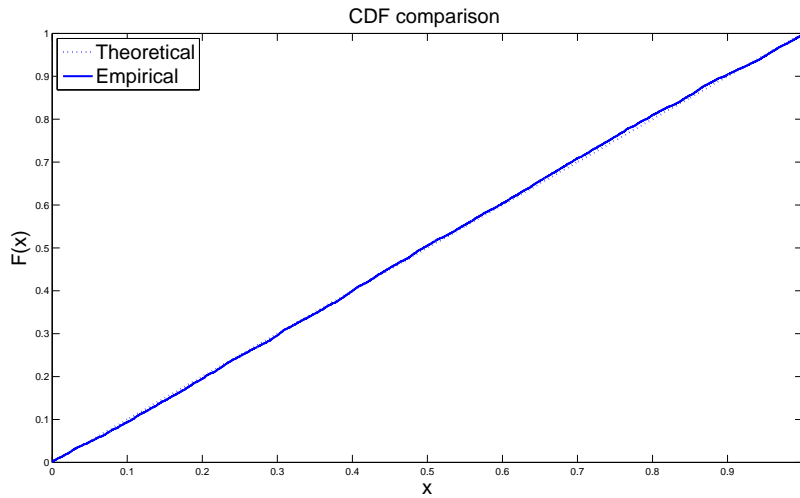


Figure 3.9: Comparison between the CDF of the PITs for the four-dimensional industry return data and the theoretical CDF of the uniform distribution over the interval $[0, 1]$.

3.4.2 Stock Returns from the Finance Sector

The data set is a collection of five time series of stock returns from the Canadian banks: Royal Bank of Canada (RY), Toronto-Dominion Bank (TD), Bank of Montreal (BMO), Bank of Nova Scotia (BNS) and Canadian Imperial Bank of Commerce (CM). The historical daily closing prices were downloaded from the web site finance.yahoo.com from Jan. 3, 2003 to Nov. 13, 2009 yielding 1730 observations for each stock. Again, each return time series was also fitted by the ASV models. The estimated variance-covariance matrix and correlation matrix are presented together in Table 3.9. All innovations of the asset returns are positively correlated and the positive correlation can also be found in those of the latent AR(1) processes. The cross correlations are all negative, which illustrates that each asset return also has a leverage effect with other latent volatility processes. This is reasonable since these stocks belong to the same industry in the same country, and common factors such as political issues, financial policy adjustments, and government intervention can cause similar changes to stock

prices. Investors will adjust their portfolios based on such information causing similar changes to stock prices and returns across the five banks. The estimated positive correlations between the observation innovations and the latent process are higher than the absolute values of the cross correlations, which indicates that the leverage effect seems to play a relatively minor role for the dynamics of the multivariate asset return time series.

Table 3.9: Bayesian estimates of the unconditional variance-covariance (in boldface) and correlation matrices (in italics) from the return time series of the five banks.

1.00	<i>0.69</i>	<i>0.66</i>	<i>0.70</i>	<i>0.64</i>	<i>-0.20</i>	<i>-0.17</i>	<i>-0.20</i>	<i>-0.19</i>	<i>-0.11</i>
0.69	1.00	<i>0.68</i>	<i>0.69</i>	<i>0.66</i>	<i>-0.21</i>	<i>-0.19</i>	<i>-0.20</i>	<i>-0.17</i>	<i>-0.19</i>
0.66	0.68	1.00	<i>0.66</i>	<i>0.64</i>	<i>-0.25</i>	<i>-0.20</i>	<i>-0.20</i>	<i>-0.17</i>	<i>-0.10</i>
0.70	0.69	0.66	1.00	<i>0.66</i>	<i>-0.24</i>	<i>-0.18</i>	<i>-0.23</i>	<i>-0.28</i>	<i>-0.18</i>
0.64	0.66	0.64	0.66	1.00	<i>-0.23</i>	<i>-0.20</i>	<i>-0.19</i>	<i>-0.24</i>	<i>-0.26</i>
-0.07	-0.08	-0.09	-0.09	-0.08	0.13	<i>0.93</i>	<i>0.89</i>	<i>0.87</i>	<i>0.83</i>
-0.05	-0.06	-0.06	-0.06	-0.06	0.11	0.10	<i>0.92</i>	<i>0.85</i>	<i>0.87</i>
-0.06	-0.06	-0.06	-0.07	-0.06	0.10	0.09	0.10	<i>0.89</i>	<i>0.90</i>
-0.04	-0.03	-0.03	-0.05	-0.05	0.06	0.05	0.06	0.04	<i>0.88</i>
-0.03	-0.05	-0.03	-0.05	-0.07	0.09	0.08	0.08	0.05	0.08

Table 3.10 summarizes estimated location, persistence and variance parameters of the latent AR(1) processes with standard errors in the parentheses. The estimated persistence parameters from the MSV model are smaller than those from the ASV models. This can be explained by the smaller estimated volatilities of the latent processes of the univariate SV models. It seems that the fitted univariate SV models are more persistent than the MSV model.

Figure 3.10 provides the scatter plot and the histogram for the PITs, and Figure 3.11 draws the CDF of PITs together with theoretical CDF of the uniform distribution over the interval $[0, 1]$. The KS test statistic for the PITs is 0.0172 with a critical value of 0.0175 at the 1% significance level. Therefore we do not reject the null hypothesis at the 1% significance level that the PITs follow the uniform distribution over the interval

Table 3.10: Comparison of estimated parameters between ASV and MSV models based on the five-bank return data.

	μ_i		ϕ_i		σ_i	
	ASV	MSV	ASV	MSV	ASV	MSV
Stock						
RY	-10.57 (0.26)	-10.66 (0.13)	0.97 (0.01)	0.93 (0.006)	0.07 (0.02)	0.13 (0.03)
TD	-10.27 (0.52)	-10.59 (0.13)	0.99 (0.004)	0.95 (0.008)	0.02 (0.01)	0.10 (0.03)
BMO	-10.59 (0.36)	-10.60 (0.15)	0.98 (0.01)	0.95 (0.006)	0.05 (0.01)	0.10 (0.02)
BNS	-10.26 (0.75)	-10.63 (0.15)	0.99 (0.002)	0.97 (0.007)	0.014 (0.004)	0.04 (0.02)
CM	-10.32 (0.38)	-10.43 (0.15)	0.99 (0.005)	0.96 (0.008)	0.04 (0.01)	0.08 (0.03)

$[0, 1]$ and that the MSV model fits the five-dimensional return data well.

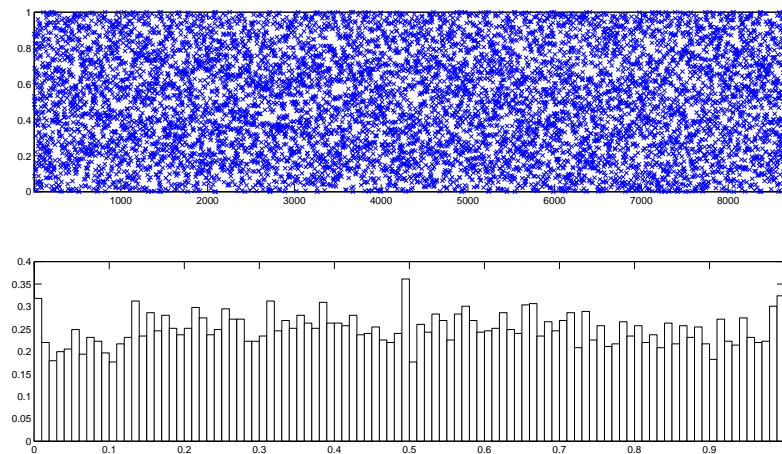


Figure 3.10: Scatter (top) and histogram (bottom) plots of the PITs for the five-bank return data.

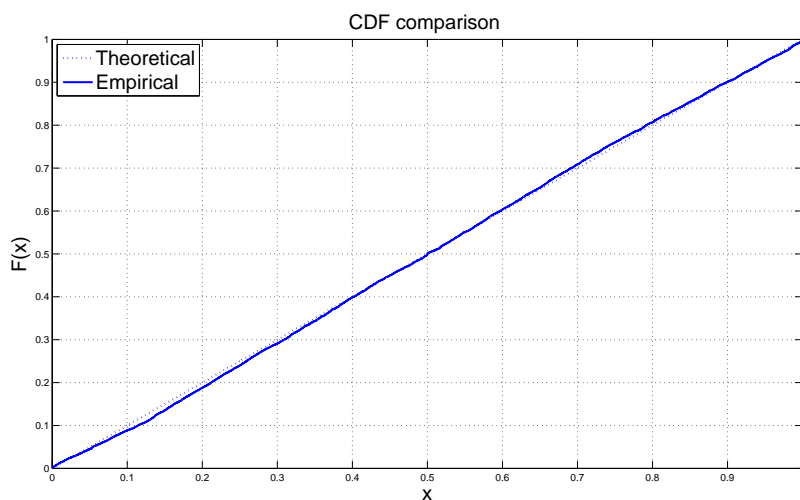


Figure 3.11: Comparison between the CDF of the PITs for the five-bank return data and the theoretical CDF of the uniform distribution over the interval $[0, 1]$.

3.5 Conclusion

In this chapter, we have developed an MCMC approach for a more general multivariate stochastic volatility model. Non-zero correlations are permitted between the innovations of measurement and latent processes. Instead of simulating the constrained variance-covariance matrix one entry at a time, we use the MH algorithm to sample the whole matrix in each iteration. In sampling of log volatilities, we propose a slice sampler in the simulation of full conditionals. The derived slice algorithms are easier to implement compared to other commonly used sampling schemes, where some extra auxiliary numerical tools have to be employed. Moreover, compared with the method in Chan et al. (2006), our proposed MCMC algorithm for the multivariate stochastic volatility model can handle higher dimension data more easily because of the straightforward simulation of the covariance matrix.

3.6 Appendix

Full Conditionals of the Parameters in the MSV Model

- Full conditional of $\boldsymbol{\mu}$.

From the equation (3.19) we have the following, for $t = 1, \dots, T - 1$,

$$\begin{aligned}\mathbf{h}_{t+1} &= \boldsymbol{\mu} + \Phi(\mathbf{h}_t - \boldsymbol{\mu}) + \mathbf{DC}^{-1}\mathbf{A}_t^{-\frac{1}{2}}\mathbf{y}_t + \mathbf{E}\mathbf{v}_{t+1} \\ (\mathbf{I} - \Phi)\boldsymbol{\mu} &= \mathbf{h}_{t+1} - \boldsymbol{\mu} - \mathbf{DC}^{-1}\mathbf{A}_t^{-\frac{1}{2}}\mathbf{y}_t - \mathbf{E}\mathbf{v}_{t+1}\end{aligned}$$

Let us define

$$\begin{aligned}\boldsymbol{\mu}_t &= (\mathbf{I} - \Phi)^{-1}(\mathbf{h}_{t+1} - \boldsymbol{\mu} - \mathbf{DC}^{-1}\mathbf{A}_t^{-\frac{1}{2}}\mathbf{y}_t), \\ \boldsymbol{\Sigma}_t &= (\mathbf{I} - \Phi)^{-1}\mathbf{E}\mathbf{E}'(\mathbf{I} - \Phi)^{-1}.\end{aligned}$$

Then the conditional distribution of $\boldsymbol{\mu}$, for $t = 1, \dots, T - 1$, is

$$\boldsymbol{\mu} \sim \mathcal{N}(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t).$$

The conditional density of $\boldsymbol{\mu}$ is

$$f(\boldsymbol{\mu}|\mathbf{y}, \mathbf{h}, \Phi, \boldsymbol{\Sigma}) \sim \mathcal{N}(\boldsymbol{\mu}_*, \boldsymbol{\Sigma}\boldsymbol{\mu}_*),$$

where

$$\begin{aligned}\boldsymbol{\Sigma}\boldsymbol{\mu}_* &= (\mathbf{I} - \Phi)^{-1}(\mathbf{E}\mathbf{E}')(\mathbf{I} - \Phi)^{-1}/(T - 1), \\ \boldsymbol{\mu}_* &= (\mathbf{I} - \Phi)^{-1}\sum_{t=1}^{T-1}(\mathbf{h}_{t+1} - \boldsymbol{\mu} - \mathbf{DC}^{-1}\mathbf{A}_t^{-\frac{1}{2}}\mathbf{y}_t)/(T - 1).\end{aligned}$$

Following the above conditional we can calculate the full conditional of $\boldsymbol{\mu}$ given its conjugate prior.

- Proposal density for the full conditional of $\boldsymbol{\phi}$ defined in (3.21).

From the equation (3.19) we have the following, for $t = 1, \dots, T - 1$,

$$\mathbf{h}_{t+1} = \boldsymbol{\mu} + \boldsymbol{\Phi}(\mathbf{h}_t - \boldsymbol{\mu}) + \mathbf{D}\mathbf{C}^{-1}\mathbf{A}_t^{-\frac{1}{2}}\mathbf{y}_t + \mathbf{E}\mathbf{v}_{t+1}.$$

Then we have

$$\begin{aligned} \mathbf{h}_{t+1} - \boldsymbol{\mu} - \mathbf{D}\mathbf{C}^{-1}\mathbf{A}_t^{-\frac{1}{2}}\mathbf{y}_t &= \text{diag}(\mathbf{h}_t - \boldsymbol{\mu})\boldsymbol{\phi} + \mathbf{E}\mathbf{v}_{t+1}, \\ \text{diag}^{-1}(\mathbf{h}_t - \boldsymbol{\mu}) * (\mathbf{h}_{t+1} - \boldsymbol{\mu} - \mathbf{D}\mathbf{C}^{-1}\mathbf{A}_t^{-\frac{1}{2}}\mathbf{y}_t) &= \boldsymbol{\phi} + \text{diag}^{-1}(\mathbf{h}_t - \boldsymbol{\mu})\mathbf{E}\mathbf{v}_{t+1}. \end{aligned}$$

Let us define

$$\begin{aligned} \boldsymbol{\phi}_t &= \text{diag}^{-1}(\mathbf{h}_t - \boldsymbol{\mu}) * (\mathbf{h}_{t+1} - \boldsymbol{\mu} - \mathbf{D}\mathbf{C}^{-1}\mathbf{A}_t^{-\frac{1}{2}}\mathbf{y}_t), \\ \boldsymbol{\Psi}_t &= \text{diag}^{-1}(\mathbf{h}_t - \boldsymbol{\mu})\mathbf{E}\mathbf{E}'\text{diag}^{-1}(\mathbf{h}_t - \boldsymbol{\mu}). \end{aligned}$$

Then the distribution of $\boldsymbol{\phi}$, for $t = 1, \dots, T - 1$, is

$$\boldsymbol{\phi} \sim N(\boldsymbol{\phi}_t, \boldsymbol{\Psi}_t).$$

Thus the conditional density of $\boldsymbol{\phi}$, given \mathbf{h} , is

$$f(\boldsymbol{\phi}|\mathbf{y}, \mathbf{h}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) \sim \mathcal{N}(\boldsymbol{\phi}^*, \boldsymbol{\Sigma}_{\boldsymbol{\phi}^*}),$$

where

$$\Sigma_{\phi^*}^{-1} = \left[\sum_{t=1}^{T-1} \text{diag}(\mathbf{h}_t - \boldsymbol{\mu}) \left(\mathbf{E}\mathbf{E}' \right)^{-1} \text{diag}(\mathbf{h}_t - \boldsymbol{\mu}) \right],$$

and

$$\begin{aligned} \phi^* &= \left[\sum_{t=1}^{T-1} \text{diag}(\mathbf{h}_t - \boldsymbol{\mu}) \left(\mathbf{E}\mathbf{E}' \right)^{-1} \text{diag}(\mathbf{h}_t - \boldsymbol{\mu}) \right]^{-1} \\ &\times \left[\sum_{t=1}^{T-1} \text{diag}(\mathbf{h}_t - \boldsymbol{\mu}) \left(\mathbf{E}\mathbf{E}' \right)^{-1} \text{diag}(\mathbf{h}_t - \boldsymbol{\mu}) \phi_t \right]. \end{aligned}$$

Following the above conditional we can calculate the proposal density of ϕ given its conjugate prior. This sampling scheme is also employed in Aguilar and West (2000).

- Full conditional of \mathbf{h}_1 is given by

$$\begin{aligned} &f(\mathbf{h}_1 | \mathbf{y}, \mathbf{h}_2, \boldsymbol{\Theta}) \\ &= \eta_1 f(\mathbf{y}_1 | \mathbf{h}_1, \boldsymbol{\Theta}) f(\mathbf{h}_1 | \boldsymbol{\Theta}) f(\mathbf{h}_1 | \mathbf{h}_2, \boldsymbol{\Theta}) \\ &= \eta_2 \exp \left\{ -\frac{\sum_{i=1}^m h_{i,1}}{2} \right\} \exp \left\{ -\frac{\mathbf{y}'_1 \text{diag}(e^{-\frac{h_1}{2}}) (\mathbf{C}\mathbf{C}')^{-1} \text{diag}(e^{-\frac{h_1}{2}}) \mathbf{y}_1}{2} \right\} \\ &\quad \exp \left\{ -\frac{(\mathbf{h}_1 - \boldsymbol{\mu})' (\boldsymbol{\Sigma}_0)^{-1} (\mathbf{h}_1 - \boldsymbol{\mu})}{2} \right\} \exp \left\{ -\frac{(\mathbf{h}_2 - \boldsymbol{\mu}_2)' (\mathbf{E}\mathbf{E}')^{-1} (\mathbf{h}_2 - \boldsymbol{\mu}_2)}{2} \right\} \\ &< \eta_2 \exp \left\{ -\frac{\sum_{i=1}^m h_{i,1}}{2} \right\} \exp \left\{ -\frac{\mathbf{y}'_1 \text{diag}(e^{-\frac{h_1}{2}}) (\mathbf{C}\mathbf{C}')^{-1} \text{diag}(e^{-\frac{h_1}{2}}) \mathbf{y}_1}{2} \right\} \\ &\quad \exp \left\{ -\frac{(\mathbf{h}_1 - \boldsymbol{\mu})' (\boldsymbol{\Sigma}_0)^{-1} (\mathbf{h}_1 - \boldsymbol{\mu})}{2} \right\}, \end{aligned}$$

where η_1 and η_2 are the two normalizing constants, $\boldsymbol{\Sigma}_0$ and $\boldsymbol{\mu}_2$ were defined in (3.10) and (3.19), respectively.

- Full conditional of \mathbf{h}_T is

$$\begin{aligned}
& f(\mathbf{h}_T | \mathbf{y}, \mathbf{h}_{-T}, \Theta) \\
& \propto p(\mathbf{y}_T | \mathbf{h}_T, \Theta) p(\mathbf{h}_T | \Theta, \mathbf{h}_{T-1}) \\
& \propto \exp \left\{ -\frac{\sum_{i=1}^m h_{i,T}}{2} \right\} \exp \left\{ -\frac{\mathbf{y}'_T \text{diag}(e^{-\frac{\mathbf{h}_T}{2}}) (\mathbf{C}\mathbf{C}')^{-1} \text{diag}(e^{-\frac{\mathbf{h}_T}{2}}) \mathbf{y}_T}{2} \right\} \\
& \propto \exp \left\{ -\frac{(\mathbf{h}_T - \boldsymbol{\mu}_T)' (\mathbf{E}\mathbf{E}')^{-1} (\mathbf{h}_T - \boldsymbol{\mu}_T)}{2} \right\},
\end{aligned}$$

where $\boldsymbol{\mu}_T$ was defined in (3.19).

Slice Sampler for Simulating $h_{i,t}$, $t = 1, \dots, T - 1$.

The dominant distribution for the full conditional of $h_{i,t}$ derived in Section 3.2.2 is

$$p(h_{i,t}) = \eta_{2t} \exp \left\{ -\frac{(h_{i,t} - \hat{\mu}_{i,t})^2}{2 \frac{1}{g_{i,i}}} \right\} \exp \left\{ -\frac{\left(e^{-\frac{h_{i,t}}{2}} + \frac{\sum_{j=1, j \neq i}^m (b_{i,j} y_{j,t} e^{-\frac{h_{j,t}}{2}})}{b_{i,i} y_{i,t}} \right)^2}{2 \frac{1}{b_{i,i} y_{i,t}^2}} \right\},$$

where $\hat{\mu}_{i,t} = \mu_{i,t} - \frac{1}{2g_{i,i}} - \frac{\sum_{j=1, j \neq i}^m (g_{i,j} (h_{j,t} - \hat{\mu}_{j,t})^2)}{g_{i,i}}$, and η_{1t} and η_{2t} are the two normalizing constants. Suppose that we have a sampled point $h_{i,t}^{(n)}$ from the n th iteration. Assume $y_{i,t} \neq 0$, otherwise, the algorithm can be adjusted easily.

Step 1. Draw u_1 uniformly from the interval $(0, 1)$.

Let $u_1 = u_1 * \exp \left\{ -\frac{(h_{i,t}^{(n)} - \hat{\mu}_{i,t}^{(n)})^2}{2\frac{1}{g_{i,i}}} \right\}$ and

$$u_3 \leq \exp \left\{ -\frac{(h_{i,t} - \hat{\mu}_{i,t}^{(n)})^2}{2\frac{1}{g_{i,i}}} \right\}$$

$$\frac{2 \log(u_3)}{g_{i,i}} \leq -(h_{i,t} - \hat{\mu}_{i,t}^{(n)})^2.$$

Then we have

$$-\sqrt{-\frac{2 \log(u_3)}{g_{i,i}}} + \hat{\mu}_{i,t}^{(n)} \leq h_{i,t} \leq \sqrt{-\frac{2 \log(u_3)}{g_{i,i}}} + \hat{\mu}_{i,t}^{(n)}. \quad (\text{A.1})$$

Step 2. Draw u_2 uniformly from the interval $(0, 1)$.

Let $u_2 = u_2 * \exp \left\{ -\frac{\left(e^{-\frac{h_{i,t}^{(n)}}{2}} + \frac{\sum_{j=1, j \neq i}^m (b_{i,j} y_{j,t} e^{-\frac{h_{j,t}^{(n)}}{2}})}{b_{i,i} y_{i,t}} \right)^2}{2\frac{1}{b_{i,i} y_{i,t}^2}} \right\}$

and

$$u_2 \leq \exp \left\{ -\frac{\left(e^{-\frac{h_{i,t}}{2}} + \frac{\sum_{j=1, j \neq i}^m (b_{i,j} y_{j,t} e^{-\frac{h_{j,t}}{2}})}{b_{i,i} y_{i,t}} \right)^2}{2\frac{1}{b_{i,i} y_{i,t}^2}} \right\}$$

$$\log(u_2) \leq -\frac{\left(e^{-\frac{h_{i,t}}{2}} + \frac{\sum_{j=1, j \neq i}^m (b_{i,j} y_{j,t} e^{-\frac{h_{j,t}}{2}})}{b_{i,i} y_{i,t}} \right)^2}{2\frac{1}{b_{i,i} y_{i,t}^2}}$$

$$\left(e^{-\frac{h_{i,t}}{2}} + \frac{\sum_{j=1, j \neq i}^m (b_{i,j} y_{j,t} e^{-\frac{h_{j,t}}{2}})}{b_{i,i} y_{i,t}} \right)^2 \leq -\frac{2 \log(u_2)}{b_{i,i} y_{i,t}^2}.$$

Define $\beta_{h_{i,t}} = \frac{\sum_{j=1, j \neq i}^m (b_{i,j} y_{j,t} e^{-\frac{h_{i,t}}{2}})}{b_{i,i} y_{i,t}}$. Then we have

$$\begin{aligned}
-\sqrt{-\frac{2 \log(u_2)}{b_{i,i} y_{i,t}^2}} - \beta_{h_{i,t}} &\leq e^{-\frac{h_{i,t}}{2}} \leq \sqrt{-\frac{2 \log(u_2)}{b_{i,i} y_{i,t}^2}} - \beta_{h_{i,t}} \\
\log\left(-\sqrt{-\frac{2 \log(u_2)}{b_{i,i} y_{i,t}^2}} - \beta_{h_{i,t}}\right) &\leq \frac{-h_{i,t}}{2} \leq \log\left(\sqrt{-\frac{2 \log(u_2)}{b_{i,i} y_{i,t}^2}} - \beta_{h_{i,t}}\right) \\
-2 \log\left(\sqrt{-\frac{2 \log(u_2)}{b_{i,i} y_{i,t}^2}} - \beta_{h_{i,t}}\right) &\leq h_{i,t} \leq -2 \log\left(-\sqrt{-\frac{2 \log(u_2)}{b_{i,i} y_{i,t}^2}} - \beta_{h_{i,t}}\right).
\end{aligned} \tag{A.2}$$

Step 3. Draw $h_{i,t}^{(n+1)}$ uniformly from the joint interval determined by the inequalities (A.1) and (A.2).

Auxiliary particle filter algorithm for the MSV model

The algorithm for our MSV model is based on the procedure in Chib et al. (2006).

Step 1. Given a sample $\{\mathbf{h}_t^{(i)}, i = 1, \dots, N\}$ from $(\mathbf{h}_t | \mathbf{y}_{1:t}, \Theta)$, we calculate the expectation $\hat{\mathbf{h}}_{t+1}^{*(i)} = E(\mathbf{h}_{t+1} | \mathbf{h}_t^{(i)})$ and

$$\pi_i = p(\mathbf{y}_{t+1} | \hat{\mathbf{h}}_{t+1}^{*(i)}, \Theta), i = 1, \dots, N. \tag{A.3}$$

Sample N times with replacement the integers of $1, \dots, N$ with probability $\hat{\pi}_i = \pi_i / \sum_{i=1}^N \pi_i$. Define the sampled indexes n_1, \dots, n_N and associate these with particles $\{\mathbf{h}_t^{(n_1)}, \dots, \mathbf{h}_t^{(n_N)}\}$.

Step 2. For each values of n_i from Step 1, sample the values $\{\mathbf{h}_{t+1}^{*(1)}, \dots, \mathbf{h}_{t+1}^{*(N)}\}$ from

$$h_{k,t+1}^{*(i)} = \mu_k + \phi_k(h_{k,t}^{(n_i)} - \mu_k) + \sqrt{\sigma_k} u_{k,t+1}, \tag{A.4}$$

where $u_{k,t+1} \sim \mathcal{N}(0, 1)$, $k = 1, \dots, q$, $i = 1, \dots, N$.

Step 3. Calculate the weights of the values $\{\mathbf{h}_{t+1}^{*(1)}, \dots, \mathbf{h}_{t+1}^{*(N)}\}$ as

$$\pi_i^* = \frac{p(\mathbf{y}_{t+1}|\mathbf{h}_{t+1}^{*(i)}, \Theta)}{p(\mathbf{y}_{t+1}|\hat{\mathbf{h}}_{t+1}^{*(i)}, \Theta)}, i = 1, \dots, N, \quad (\text{A.5})$$

and resample the values $\{\mathbf{h}_{t+1}^{*(1)}, \dots, \mathbf{h}_{t+1}^{*(N)}\}$ N times with replacement using these weights to obtain a random sample $\{\mathbf{h}_{t+1}^{(1)}, \dots, \mathbf{h}_{t+1}^{(N)}\}$ from the filter distribution of $(\mathbf{h}_{t+1}|\mathbf{y}_{1:t+1}, \Theta)$.

Chapter 4

Factor Stochastic Volatility with Orthogonal Loadings

4.1 Introduction

In recent years, factor-based stochastic volatility models have been used in the analysis of multivariate financial time series of asset returns. This is largely motivated by the observation that the underlying multivariate time series are governed by latent common factors. For instance Geweke and Zhou (1996) propose a factor model to measure the pricing errors in the context of the arbitrage pricing theory, where the time series of the observed returns is a linear regression of latent factors with idiosyncratic observation errors. Lopes and West (2004) employ the same model to detect the factor structure in international exchange rates. Following Geweke and Zhou (1996), the mean factor stochastic volatility (FSV) models were proposed by Jacquier et al. (1999), Pitt and Shephard (1999a), Chib et al. (2006), Aguilar and West (2000), Lopes and Carvalho (2007), among others. From the models of Geweke and Zhou (1996) and Lopes and West (2004), we note that the common structure of the FSV models is that

the latent factors follow standard univariate SV processes, and the observation errors either have a multivariate normal distribution with a zero mean vector and a constant variance-covariance matrix or follow uncorrelated univariate SV processes. Because the likelihood function of observed multivariate returns does not always have a closed form, to estimate FSV models simulation based approaches have been applied using Bayesian framework. Sampling the latent states of the factor stochastic volatility models is the most difficult part. Jacquier et al. (2004), Pitt and Shephard (1999a), Chib et al. (2006), Aguilar and West (2000) and Lopes and Carvalho (2007) propose Markov Chain Monte Carlo (MCMC) methods, where the log volatilities (or the state random variables) are augmented as parameters and sampled one at a time or within blocks from their full conditionals.

Although FSV models have been successful in financial time series analysis, the structure of the loading matrix may make the sample likelihood critically depend on the order of the multivariate time series. Thus the analysis results may be affected. This is undesirable if components are observed simultaneously. Lopes and West (2004) show that interchanging the order of two individual time series does not affect the estimates of the corresponding rows of the loading matrix but influences the likelihood of data and hence the number of factors. Aguilar and West (2000) discuss the same issue and suggest the use of a specific ordering to define and interpret the factor effects. Lopes and Carvalho (2007) point out that the number of factors can even be affected when both common factors and the idiosyncratic errors follow univariate SV processes.

Motivated by FSV models, we propose a multivariate stochastic volatility model based on standard factor analysis. Instead of requiring that the factor loading matrix is a lower triangular matrix, we assume that its columns are orthogonal with unit length. Similar to the FSV model, all factors follow univariate standard SV processes, but the observation errors are assumed to be independent and follow a multivariate normal distribution with zero mean and a constant diagonal variance matrix. An MCMC al-

gorithm is used to estimate the loading matrix, the latent factors, their log volatilities, and the parameters of latent Markov processes. The loading matrix is sampled one column at a time either directly from a von Mises-Fisher full conditional or through the Metropolis-Hastings algorithm, where the proposal distribution is a von Mises-Fisher distribution. In the simulation of latent states, we use the single-move Metropolis-Hastings algorithms derived in Chapter 2. One merit of the proposed factor model is that it avoids problems concerning the order of components. Consequently, our results are invariant under a permutation of the components.

The rest of the chapter is organized as follows. Section 4.2 briefly reviews the standard factor analysis model and the FSV models. In Section 4.3, we present a factor SV model, discuss the model identifiability, the MCMC algorithm and how to perform model selection and assessment. Section 4.4 includes some empirical results from applying our model and estimation methods to simulated asset return data. Section 4.5 presents the results from applying our model to international index return data. We conclude this chapter with a short discussion in the last section.

4.2 Factor SV Models: a Brief Review

In this section, we briefly review the standard factor analysis and FSV models, and discuss model identification.

Standard normal factor analysis (see Press (1985)) is the basis of FSV models. Define by $\mathbf{y}_t = (y_{1,t}, \dots, y_{m,t})'$, $t \leq T$, a random sample from a zero-mean multivariate normal distribution with a non-singular variance-covariance matrix. Assuming q ($q < m$) unobserved factors $\mathbf{f}_t = (f_{1,t}, \dots, f_{q,t})'$, we have

$$\mathbf{y}_t = \mathbf{D}\mathbf{f}_t + \boldsymbol{\epsilon}_t, \tag{4.1}$$

where (i) $\mathbf{D} = [D_1, \dots, D_q]$ is a $m \times q$ matrix of unknown parameters, (ii) $\epsilon_t, t = 1, 2, \dots$, are independent m -multivariate normal random variables with $\epsilon_t \sim \mathcal{N}(\mathbf{0}, \Psi)$ and $\Psi = \text{diag}(\psi_1, \dots, \psi_m)$ is a constant variance-covariance matrix, which may have idiosyncratic entries on the main diagonal, (iii) \mathbf{f}_t are independent with $\mathbf{f}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_q)$, where \mathbf{I}_q is the $q \times q$ identity matrix and (iv) ϵ_t are independent of the latent factors \mathbf{f}_s for all t and s .

There is a fundamental rotation problem in factor analysis which leads to multiplicities of solutions. Under the factor analysis model, the variance-covariance structure of the data \mathbf{y}_t is $V(\mathbf{y}_t | \mathbf{D}, \Psi) = \mathbf{D}\mathbf{D}' + \Psi$. Suppose that \mathbf{R} is any $q \times q$ orthogonal matrix, such that $\mathbf{R}'\mathbf{R} = \mathbf{I}_q$. Then equation (4.1) can be re-written as

$$\mathbf{y}_t = \mathbf{D}\mathbf{R}'\mathbf{R}\mathbf{f}_t + \epsilon_t. \quad (4.2)$$

The variables \mathbf{y}_t have the same variance-covariance matrix but a different loading matrix $\mathbf{D}\mathbf{R}'$ and the factors $\mathbf{R}\mathbf{f}_t$, where the latter satisfy $\mathbf{R}\mathbf{f}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_q)$. Hence, model (4.1) is unidentifiable.

There are several ways to eliminate the ambiguity due to rotation by imposing constraints on the loading matrix \mathbf{D} . A detailed discussion about this issue can be found in Anderson and Rubin (1956), where the authors provide several methods to guarantee that the factor model (4.1) has a unique solution. The first approach requires that \mathbf{D} is column-orthogonal and satisfies $\mathbf{D}'\Psi^{-1}\mathbf{D} = \mathbf{C}$, where $\mathbf{C} = \text{diag}(c_1, \dots, c_q)$ and $c_1 > c_2 > \dots > c_q$, which can also be found in Press 1982, Chapter 10. The second approach imposes restrictions on the loading matrix by letting it to be a lower triangular matrix with positive elements on the main diagonal. That is, $W_{k,k} > 0, k = 1, \dots, q$, $W_{i,j} = 0, 1 < i < j \leq q, j = 2, \dots, q$, and all other entries are free parameters. This specification of the loading matrix was used in formulating factor SV models in Geweke and Zhou (1996) and Lopes and West (2004). The third approach assumes that the loading matrix in model (4.1) is column-orthonormal. We adopt this approach in the

factor SV model that we propose in the next section. Conditional on the latent common factors, the observed variables \mathbf{y}_t are independent with $(\mathbf{y}_t|\mathbf{D}, \mathbf{f}_t) \sim \mathcal{N}(\mathbf{D}\mathbf{f}_t, \Psi)$. That is, the dependence structure among the m -dimensional time series are mainly explained by the latent factors, while the idiosyncratic errors measure the residual variability in each of the component time series after the contribution by the latent factors.

There is an important issue arising in the models of both Geweke and Zhou (1996) and Lopes and West (2004): how many factors that the model can have without encountering identifiability problems. To answer this question, it is noticed that the \mathbf{f}_t are independent and estimation is largely based on the variance-covariance matrix. Therefore there are at most $m(m+1)/2$ distinct parameters that can be included. For the factor model not to be over-parameterized, the number of parameters that index the model must satisfy the inequalities $q < m$ and

$$m(m+1)/2 - (mq - q(q-1)/2 + m) \geq 0, \quad (4.3)$$

where $mq - q(q-1)/2$ is the total number of free parameters in the loading matrix, and m is the number of the idiosyncratic error parameters. For instance, with $m = 5$ the factor model can have at most 2 factors, with $m = 6$ or 7 we have $q \leq 3$, with $m = 8$ the upper boundary for q is 4. For high-dimensional data where m is large, the number of factors is usually not a constraint since we try to have only a small number of factors.

In the context of SV models, the standard FSV models are considered by Pitt and Shephard (1999a), Aguilar and West (2000), Chib et al. (2006) and Lopes and Carvalho (2007). These papers extend the model in Geweke and Zhou (1996) and Lopes and West (2004), so that both the factors and idiosyncratic errors follow standard univariate SV

models. In the more general specification, in equation (4.1) we assume

$$\begin{pmatrix} \mathbf{f}_t \\ \boldsymbol{\epsilon}_t \end{pmatrix} | \mathbf{A}_t, \mathbf{B}_t \sim \mathcal{N} \left\{ \mathbf{0}, \begin{pmatrix} \mathbf{A}_t & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_t \end{pmatrix} \right\}.$$

That is, \mathbf{f}_t and $\boldsymbol{\epsilon}_t$ are conditionally independent Gaussian random vectors with zero mean and time-dependent variances. The time-varying variance matrices \mathbf{A}_t and \mathbf{B}_t are assumed to be diagonal and depend on the unobserved random variables $\mathbf{h}_t = \text{diag}(h_{1t}, \dots, h_{q+m,t})$ as in the following representation

$$\mathbf{A}_t = \text{diag}(\exp(h_{1,t}), \dots, \exp(h_{q,t})), \quad (4.4)$$

$$\mathbf{B}_t = \text{diag}(\exp(h_{q+1,t}), \dots, \exp(h_{q+m,t})), \quad (4.5)$$

$$h_{k,t+1} = \mu_k + \phi(h_{k,t} - \mu_k) + \sqrt{\sigma_k} u_{k,t+1}, \quad k = 1, \dots, q + m. \quad (4.6)$$

To ensure that the FSV model is identifiable, the loading matrix \mathbf{D} is specified as a lower triangular matrix but the main diagonal entries are set to 1. Pitt and Shephard (1999a) consider this model by setting $u_{k,t}$ to be independent and identically distributed according to the univariate standard normal distribution $u_{k,t} \sim \mathcal{N}(0, 1)$, whereas Aguilar and West (2000) propose that the innovation vectors of the latent volatility processes of factors $u_{A_t} = (u_{1,t}, \dots, u_{q,t})'$ are independent and multivariate Gaussian with zero mean and a constant non-diagonal variance-covariance matrix. Lopes and Carvalho (2007) consider a FSV model which includes the model of Pitt and Shephard (1999a) and Aguilar and West (2000), and extends it in two directions by (i) letting the matrix of factor loadings \mathbf{D} to be time-varying and (ii) allowing Markov switching in the common factor volatilities. For a systematic review of the FSV models, see Chib et al. (2009) and a more recent one by Lopes and Polson (2010).

As in the model of Geweke and Zhou (1996) and Lopes and West (2004), there

is still a question about the maximum number of factors for FSV models. Pitt and Shephard (1999a) fit their model with two factors in the analysis of returns of exchange rates. Although each latent univariate SV process contains three parameters, Lopes and West (2000) and Lopes and Carvalho (2007) determine the maximum number of factors using the inequality (4.3) as if the factors \mathbf{f}_t were independent, and fit their models with a number of factors at the upper boundary.

4.3 Proposed Model, Identification, Estimation and Assessment

4.3.1 Model and Its Identification

Let $\mathbf{y}_t = (y_{1,t}, \dots, y_{m,t})'$ denote the vector of measurements from m objects at time t ($t \leq T$). In this chapter, \mathbf{y}_t is a demeaned vector. Assuming q ($q < m$) unobserved factors $\mathbf{f}_t = (f_{1,t}, \dots, f_{q,t})'$, \mathbf{y}_t can be characterized by the following hierarchical factor model specification:

$$\mathbf{y}_t = \mathbf{W}\mathbf{f}_t + \Psi^{\frac{1}{2}}\boldsymbol{\epsilon}_t, \quad t = 1, \dots, T, \quad (4.7)$$

$$\mathbf{W} = [W_1, \dots, W_q], \quad (4.8)$$

$$f_{k,t} = \exp(h_{k,t}/2)v_{k,t}, \quad k = 1, \dots, q, \quad (4.9)$$

$$h_{k,t+1} = \mu_k + \phi_k(h_{k,t} - \mu_k) + \sqrt{\sigma_k} u_{k,t+1}, \quad t = 1, \dots, T - 1, \quad (4.10)$$

where \mathbf{W} is an unknown $m \times q$ ($q < m$) column-orthogonal loading matrix such that $\mathbf{W}'\mathbf{W} = \mathbf{I}_q$, and $\Psi^{\frac{1}{2}} = \text{diag}(\sqrt{\psi_1}, \dots, \sqrt{\psi_m})$ is a constant diagonal variance matrix which may have idiosyncratic diagonal entries. The vector \mathbf{f}_t is defined, as in FSV models, so that each component follows a univariate SV process. Denote by $\mathbf{h}_t = (h_{1,t}, \dots, h_{q,t})'$

the vector of log volatilities of the latent factors which follows a univariate first order vector AR(1) process. To ensure weak stationarity of the latent AR(1) processes, we assume $|\phi_k| < 1$ for $k = 1, \dots, q$. Let $\mathbf{v}_t = (v_{1,t}, \dots, v_{q,t})'$ and $\mathbf{u}_t = (u_{1,t}, \dots, u_{q,t})'$ be the innovations of the factor processes and that of the corresponding volatility processes, respectively. Assume

$$\begin{pmatrix} \boldsymbol{\epsilon}_t \\ \mathbf{v}_t \\ \mathbf{u}_t \end{pmatrix} \sim \mathcal{N} \left\{ \mathbf{0}, \begin{pmatrix} \mathbf{I}_m & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_q & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_q \end{pmatrix} \right\},$$

which implies that \mathbf{f}_t and $\boldsymbol{\epsilon}_t$ are independent. Generally, q is much smaller than m , so the latent variables \mathbf{f}_t and the loading matrix \mathbf{W} offer a parsimonious representation for multivariate random variables \mathbf{y}_t . Conditioned on the latent factors \mathbf{f}_t , the observations \mathbf{y}_t are allowed to be independent and follow multivariate normal distributions, such that $(\mathbf{y}_t | \mathbf{W}, \mathbf{f}_t) \sim \mathcal{N}(\mathbf{W}\mathbf{f}_t, \boldsymbol{\Psi})$. The conditional dependence structure of \mathbf{y}_t is characterized by the latent factors \mathbf{f}_t . In other words, both the conditional variance and correlation of \mathbf{y}_t are allowed to be time-dependent.

Our model is an extension of the model based on the probabilistic principal component analysis (PPCA) proposed by Tipping and Bishop (1999), where the columns of the loading matrix are orthogonal, the latent factors are orthonormal and the observation errors are isotropic. Due to the similarity of our setting of the loading matrix to that in Tipping and Bishop (1999), we call our factor SV model the PPCAF model. Tipping and Bishop (1999) prove that the subspace defined by the maximum likelihood estimates (MLE) of the columns of \mathbf{W} is the principal subspace of the observed data. This is also true even when the sample covariance is not equal to the model covariance. However, this result does not hold for general factor analysis with idiosyncratic observation errors. Following a similar argument, we can reach the conclusion that for

isotropic observation errors the subspace spanned by the columns of \mathbf{W} is the principal subspace.

In a general situation when the model has idiosyncratic errors, the problem is more difficult to solve. There are two issues that need to be considered. The first one is whether the proposed factor model has a solution and the other is under what condition(s) our model has a unique solution. Anderson and Rubin (1956) provide a necessary and sufficient condition to address the first issue, namely, $\hat{\Sigma}$ is a variance-covariance matrix of a factor SV model with q factors if there exists a diagonal matrix Ψ with nonnegative elements such that $\hat{\Sigma} - \Psi$ is positive semidefinite of rank q . Our PPCAF model satisfies this sufficient and necessary condition. For the second issue, Anderson and Rubin (1956) give a sufficient condition by requesting that $\mathbf{W}'\mathbf{W}$ is diagonal, which was used in our PPCAF model and also stated in Alexander (1994) and Hardle and Simar (2007). As discussed in Anderson and Rubin (1956), this condition and the other two reviewed in the previous section are more or less arbitrary ways to eliminate the indeterminacy of rotation. There are no theoretical considerations that would help us choose any of these representations. In practice, if the estimate of Ψ from our PPCAF model is near isotropic, then the estimated columns of \mathbf{W} are similar to the leading eigenvectors of the sample variance-covariance matrix of the observed data. Simulation studies and applications to financial asset return data appear to support this conclusion.

The PPCAF model specification is completed by specifying prior distributions of the parameters. Define by $\Theta = (W_k, \mu_k, \phi_k, \sigma_k, \psi_k, k = 1, \dots, q)$ the collection of parameters of the idiosyncratic PPCAF model (for the isotropic case all ψ_k are set to equal each other). The prior distributions of parameters in Θ are assumed to be mutually independent. For the columns of the loading matrix, since we have not much information, a uniform distribution on the unit hypersphere is set as prior distribution, which makes us easier to obtain the corresponding posterior distributions. As discussed in

Hoff (2007), a noninformative prior distribution for each column of the loading matrix is nature since it is the unique probability measure on the unit hypersphere that is invariant under left- and right-orthogonal transformations. For parameters in the latent AR(1) processes, we set $\mu_k \sim \mathcal{N}(-1, 9)$ and $\sigma_k \sim \mathcal{IG}(5, 0.05)$, respectively, as in Pitt and Shephard (1999a), where $\mathcal{IG}(\cdot)$ represents a generic inverse Gamma distribution. Further, we assume that each ϕ_k has a uniform distribution over the interval $(-1, 1)$, such that $\phi_k \sim \text{unif}(-1, 1)$. Since we do not have much information about these prior distributions, all hyperparameters are set to make the prior distributions flatter over their support.

Based upon our PPCAF model, the m -multivariate time series \mathbf{y}_t under the common \mathcal{R}^m coordinate system is transformed into a q -dimensional factor time series \mathbf{f}_t through a new orthogonal coordinate system in \mathcal{R}^m defined by the columns of \mathbf{W} . We expect the lower-dimensional time series \mathbf{f}_t to provide a parsimonious fit to the original data \mathbf{y}_t .

There is one important feature of the idiosyncratic PPCAF model that requires some discussion. The idiosyncratic standard errors are the residuals when the observed multivariate time series is projected onto the subspace spanned by the columns of \mathbf{W} . The components of these residuals can be used to measure the extent of factor contributions. A larger idiosyncratic error indicates that the corresponding component of time series is less affected by the factors, while a smaller idiosyncratic error indicates that the corresponding return time series is more influenced by the factors.

As in the basic factor analysis and FSV models, we give an upper boundary for q to prevent the proposed PPCAF model from being over parameterized. The argument for this is similar to that of Geweke and Zhou (1996) and Lopes and Carvalho (2007). For a q -factor m -variate PPCAF model, we have $(m - 1)q$ free parameters for the loading matrix since each column has unit length such as $\|W_k\| = 1$, where $\|\mathbf{x}\|$ stands for a Euclidian norm of vector \mathbf{x} , and q parameters from the standard deviances of latent

factors. Also we have one parameter for the isotropic errors or m parameters for the idiosyncratic constant errors. Again, we have a total of $m(m + 1)/2$ free parameters provided by the sample variance-covariance matrix. So the upper bound of q should satisfy the inequality constraint that $m(m + 1)/2 - mq - 1 \geq 0$ in the isotropic case, or $m(m + 1)/2 - (m + 1)q \geq 0$ in the idiosyncratic case. For instance, in the idiosyncratic PPCAF model, with $m = 5$, we can have at most 2 factors, while, with $m = 8$, the upper boundary of q is 3.

As discussed in Pitt and Shephard (1999a), the unconditional variance-covariance matrix estimated from the model shows the relative importance of the factors for each of the return series. This matrix can be used in comparison with the corresponding sample variance-covariance matrix of the observed data. The Bayesian mean of the unconditional variance from our model is

$$\Sigma = E\left\{\mathbf{W}\text{diag}(\sigma_{f_1}^2, \dots, \sigma_{f_q}^2)\mathbf{W}' + \Psi\right\}, \quad (4.11)$$

where the operator $E(\cdot)$ is the expectation with respect to the posterior density and

$$\begin{aligned} \sigma_{f_k}^2 &= E(f_{k,t}^2) - (E(f_{k,t}))^2 \\ &= E_{h_{k,t}}\{E[f_{k,t}^2|h_{k,t}]\} - (E_{h_{k,t}}[E(f_{k,t}|h_{k,t})])^2 \\ &= E_{h_{k,t}}\{\exp(h_{k,t})\} \\ &= \exp\left\{\mu_k + \frac{\sigma_k}{2(1 - \phi_k^2)}\right\}, \quad k = 1, \dots, q, \end{aligned} \quad (4.12)$$

where the last equality is due to the fact that $h_{k,t}$ follows a normal distribution, $h_{kt} \sim \mathcal{N}(\mu_k, \sigma_k/(1 - \phi_k^2))$. Thus Σ can be unbiasedly estimated by using our MCMC output.

Our model is similar to the principal component analysis because of the orthogonal loading matrix. As discussed in Ahn and Oh (2003) and Roweis (1997), direct use of PCA via eigenvalue decomposition of the sample covariance matrix of the data is

often unsuitable for high-dimensional data due to its high computational complexity. To deal with this situation, our PPCAF model provides an alternative estimate of the sample variance-covariance matrix. We want the dominant eigenvectors from Σ to be close to the leading eigenvectors of the sample variance-covariance matrix. Therefore, the dimension of the observation data can be reduced significantly as in the context of PCA. In the simulation studies and the application to asset return data, we compare the estimated unconditional sample variance-covariance matrix from our PPCAF model with the sample variance-covariance matrix of the observed data and find that the two matrices are very similar. This suggests that our proposed PPCAF model is capable of capturing the dependency structure of multivariate financial time series. We also found that the columns of the loading matrix are very close to the principal sample eigenvectors. This indicates that our model can extract the main factors of the data that govern the multivariate time series.

4.3.2 Estimation: an MCMC Algorithm

There are two difficult issues involved in applying the MCMC sampler to the problem at hand. One is how to sample the column-orthogonal loading matrix and the other is how to sample the full conditionals of latent states efficiently. Considering the first issue, it is known (see, for example, Hoff (2007)) that the full conditional of each column, given the other columns, follows a von Mises-Fisher distribution if the uniform prior distribution on the unit hypersphere is assumed and the PPCAF model has isotropic errors. The simulation of these full conditionals can be done by following the procedure in Hoff (2007) and Dobigeon and Tournaret (2010), which allows us to preserve the orthonormality among columns. In the case of idiosyncratic errors, the full conditionals of these columns no longer follow von Mises-Fisher distributions. We simulate these columns by using a Metropolis-Hastings algorithm with von Mises-Fisher pro-

positional distributions.

Regarding the second issue, once the factors $f_{k,t}$ have been sampled, the sampling of $h_{k,t}$ and the parameters in the latent AR(1) processes can be done by employing some of the commonly used MCMC methods for univariate SV models. Our approach for this simulation is the slice sampler within the MCMC method introduced in Chapter 2. The MCMC algorithm for the PPCAF model is outlined in Table 4.1 followed by detailed procedures.

Table 4.1: MCMC algorithm for the PPCAF model.

Step 1. Sampling $W_k, k = 1, \dots, q$, in the isotropic case.

Step 1'. Sampling $W_k, k = 1, \dots, q$, in the idiosyncratic case.

Step 2. Sampling $f_{k,t}, k = 1, \dots, q, t = 1, \dots, T$.

Step 3. Sampling $h_{k,t}, k = 1, \dots, q, t = 1, \dots, T$.

Step 4. Sampling $\phi_k, \mu_k, \sigma_k, k = 1, \dots, q$.

Step 5. Go to Step 1.

Step 1. Sampling $W_k, k = 1, \dots, q$, in the isotropic case. Define by \mathbf{W}_{-k} the loading matrix where the k -th column has been removed.

The likelihood of $\mathbf{y}_t, t = 1, \dots, T$, given that Θ and \mathbf{f} have been sampled, is

$$p(\mathbf{y}_t | \Theta, \mathbf{f}) = (2\pi\psi)^{-\frac{m}{2}} \exp\left(-\frac{1}{2\psi} \|\mathbf{y}_t - \sum_{i=1}^q W_i f_{i,t}\|^2\right). \quad (4.13)$$

By Bayes' Theorem, from (4.13) we can show that the full conditional of W_k , given the

uniform prior distribution, is

$$\begin{aligned}
p(W_k | \mathbf{y}, \mathbf{W}_{-k}, \mathbf{f}, \psi) &\propto \prod_{t=1}^T p(W_k | \mathbf{y}_{k,t}^*, f_{k,t}, \psi) \\
&= (2\pi\psi)^{-\frac{m}{2}} \exp\left(-\frac{1}{2\psi} \sum_{t=1}^T \|\mathbf{y}_{k,t}^* - W_k f_{k,t}\|^2\right) \\
&\propto \exp\left(\frac{1}{\psi} W_k' \sum_{t=1}^T f_{k,t} \mathbf{y}_{k,t}^*\right), \tag{4.14}
\end{aligned}$$

where $\mathbf{y}_{k,t}^* = \mathbf{y}_t - \sum_{i=1, i \neq k}^q W_i f_{i,t}$ is defined as the residual from the projection of \mathbf{y}_t onto the subspace spanned by the columns of \mathbf{W}_{-k} . The question now is how to sample this full conditional while maintaining orthogonality with other columns. Define by \mathbf{N}_k a basis for the null space of \mathbf{W}_{-k} . As argued in Hoff (2007), conditional on \mathbf{W}_{-k} , there exists a \mathbf{v}_k which is uniform on the unit $(q-1)$ hypersphere and such that $W_k = \mathbf{N}_k \mathbf{v}_k$. Thus from (4.14), the full conditional of \mathbf{v}_k is

$$p(\mathbf{v}_k | \mathbf{y}, \mathbf{W}_{-k}, \mathbf{f}, \psi) \propto \exp\left(\frac{1}{\psi} \mathbf{v}_k' \mathbf{N}_k' \sum_{t=1}^T f_{k,t} \mathbf{y}_{k,t}^*\right), \tag{4.15}$$

which is a von Mises-Fisher distribution with parameter $\frac{1}{\psi} \mathbf{N}_k' \sum_{t=1}^T f_{k,t} \mathbf{y}_{k,t}^*$. The details for simulating this distribution can be found in Hoff (2007) and Dobigeon and Tourneret (2010). Then the full conditional of W_k is simulated by drawing \mathbf{v}_t from the von Mises-Fisher distribution (4.15) and setting $W_k = \mathbf{N}_k \mathbf{v}_k$.

Step 1'. Sampling $W_k, k = 1, \dots, q$, in the idiosyncratic case.

The likelihood of $\mathbf{y}_t, t = 1, \dots, T$, given that Θ and \mathbf{f} have been sampled, is

$$\begin{aligned}
&p(\mathbf{y}_t | \Theta, \mathbf{f}) \\
&= (2\pi)^{-\frac{m}{2}} |\Psi|^{-\frac{1}{2}} \exp\left(-\frac{(\mathbf{y}_t - \sum_{i=1}^q W_i f_{i,t})' \Psi^{-1} (\mathbf{y}_t - \sum_{i=1}^q W_i f_{i,t})}{2}\right), \tag{4.16}
\end{aligned}$$

where $|\Psi|$ is the determinant of the square matrix Ψ . According to Bayes' Theorem, the full conditional of W_k is given by

$$\begin{aligned}
& p(W_k | \mathbf{y}, \mathbf{W}_{-k}, \mathbf{f}, \Psi) \\
& \propto (2\pi)^{-\frac{mT}{2}} |\Psi|^{-\frac{T}{2}} \prod_{t=1}^T \exp \left(- \frac{(\mathbf{y}_t - \sum_{i=1}^q W_i f_{i,t})' \Psi^{-1} (\mathbf{y}_t - \sum_{i=1}^q W_i f_{i,t})}{2} \right) \\
& = (2\pi)^{-\frac{mT}{2}} |\Psi|^{-\frac{T}{2}} \prod_{t=1}^T \exp \left(- \frac{(\mathbf{y}_{k,t}^* - W_k f_{k,t})' \Psi^{-1} (\mathbf{y}_{k,t}^* - W_k f_{k,t})}{2} \right), \tag{4.17}
\end{aligned}$$

where $\mathbf{y}_{k,t}^*$ is defined as before. This full conditional is no longer a von Mises-Fisher distribution and unlike the isotropic case, can not be simulated directly. We use instead a Metropolis-Hastings algorithm. Define by \mathbf{B} an $m \times m$ diagonal matrix where the entries on the diagonal are equal to the average of the idiosyncratic variance, given by $b = \sum_{k=1}^m \psi_k / m$. By replacing Ψ in formula (4.17) with this \mathbf{B} , the proposal density is chosen as

$$f(X | \mathbf{y}, \mathbf{W}_{-k}, \mathbf{f}, b) \propto \exp \left(\frac{1}{b} X' \sum_{t=1}^T f_{k,t} \mathbf{y}_{k,t}^* \right). \tag{4.18}$$

Comparing with the distribution (4.14), this proposal distribution is a von Mises-Fisher distribution. To sample X , which is orthonormal to \mathbf{W}_{-k} , we first sample the following von Mises-Fisher distribution

$$f(\mathbf{v}_k | \mathbf{y}, \mathbf{W}_{-k}, \mathbf{f}, b) \propto \exp \left(\frac{1}{b} \mathbf{v}_k' \mathbf{N}_k' \sum_{t=1}^T f_{k,t} \mathbf{y}_{k,t}^* \right), \tag{4.19}$$

and then set $X = \mathbf{N}_k \mathbf{v}_k$, where \mathbf{N}_k is a basis for the null space of \mathbf{W}_{-k} , and X will be accepted from the Metropolis-Hastings algorithm. The whole procedure for sampling from the full conditional of W_k can be summarized as follows:

Suppose that we have a simulated value $W_k^{(n)}$ from the n -th iteration of the MCMC

algorithm. Then at iteration $n + 1$, we perform the following steps:

1. Generate a uniform random number α from the interval $[0,1]$.
2. Sampling \mathbf{v}_k from the von Mises-Fisher distribution (4.19) and set $X = \mathbf{N}_k \mathbf{v}_k$.
3. Define

$$\beta = \min \left\{ 1, \frac{p(X|\mathbf{y}, \mathbf{W}_{-k}, \mathbf{f}, \Psi)}{p(W_k^{(n)}|\mathbf{y}, \mathbf{W}_{-k}, \mathbf{f}, \Psi)} \times \frac{\exp \left(\frac{1}{b} (W_k^{(n)})' \sum_{t=1}^T f_{k,t} \mathbf{y}_{k,t}^* \right)}{\exp \left(\frac{1}{b} X' \sum_{t=1}^T f_{k,t} \mathbf{y}_{k,t}^* \right)} \right\}.$$

4. If $\alpha \leq \beta$ then $W_k^{(n+1)} = X$ otherwise $W_k^{(n+1)} = W_k^{(n)}$, keep the generated vector from the last iteration.

Step 2. Sampling $f_{k,t}$, $k = 1, \dots, q$, $t = 1, \dots, T$.

If the PPCAF mode has isotropic errors, the full conditional of $f_{k,t}$ is

$$\begin{aligned} p(f_{k,t}|\Theta, \mathbf{h}) &\propto p(f_{k,t}|h_{k,t})p(f_{k,t}|\mathbf{y}_{k,t}^*, W_k, \psi) \\ &= p(f_{k,t}|h_{k,t})(2\pi\psi)^{-m/2} \exp \left(-\frac{1}{2\psi} (f_{k,t} - W_k' \mathbf{y}_{k,t}^*)^2 \right), \end{aligned}$$

which is a univariate normal distribution and can be easily sampled. The full conditional of $f_{k,t}$ in the idiosyncratic case is also easy to calculate and it will not be given here.

Step 3. Sampling $h_{k,t}$, $k = 1, \dots, q$, $t = 1, \dots, T$.

Once the q factor time series \mathbf{f}_t has been sampled, the simulation of $h_{k,t}$ can be done by the slice sampler within the MCMC method. The detailed procedure can be found in Chapter 2.

Step 4. Sampling ϕ_k, μ_k, σ_k , $k = 1, \dots, q$.

For given priors, the full conditionals of these parameters can be easily sampled by using the method proposed in the previous chapters.

4.3.3 Model Selection and Assessment

In financial time series, the goal of our PPCAF model is to detect the latent factors (within the upper boundary) that govern the underlying time series. This is a model selection problem. In the literature, traditional model selection criteria based on the likelihood include the AIC in Akaike (1987) and the BIC in Schwarz (1978), which are the functions of parameters of the model. For hierarchical models, the log likelihood involves high-dimensional integrals and is virtually impossible to obtain its analytical form. In order to use AIC and BIC, the likelihood of data has to be given approximately. In the past decade, the auxiliary particle filter (APF) proposed in Pitt and Shephard (1999b) has been extensively employed as an efficient tool for hidden Markov models, especially the SV models, in the calculation of likelihood. Typically it is assumed that the underlying model is only indexed by static (non time-varying) parameters and the sample likelihood can be obtained by APF recursively. Once the likelihood is available, we can apply AIC and BIC to assess the model fit and determine the number of factors.

From the definition of the PPCAF model, the observations $\mathbf{y}_t, t = 1, \dots, T$, can be defined sequentially. Specifically, at each time t the observation equation defines the observation density

$$p(\mathbf{y}_t | \mathbf{h}_t, \Theta), \quad (4.20)$$

and the latent AR(1) Markov process defines the transition density

$$p(\mathbf{h}_t | \mathbf{h}_{t-1}, \Theta). \quad (4.21)$$

Thus, the log-likelihood of the data can be written as

$$\log p(\mathbf{y}_1, \dots, \mathbf{y}_T | \Theta) = \sum_{t=1}^{T-1} \log p(\mathbf{y}_{t+1} | \mathbf{y}_{1:t}, \Theta) + \log p(\mathbf{y}_1 | \Theta),$$

where $\mathbf{y}_{1:t} = (\mathbf{y}_1, \dots, \mathbf{y}_t)$ is the information known at time t .

At time $t + 1$ we have

$$p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t}, \Theta) = \int p(\mathbf{y}_{t+1}, \mathbf{h}_{t+1}|\mathbf{y}_{1:t}, \Theta)d\mathbf{h}_{t+1} \quad (4.22)$$

$$= \int p(\mathbf{y}_{t+1}|\mathbf{h}_{t+1}, \mathbf{y}_{1:t}, \Theta)p(\mathbf{h}_{t+1}|\mathbf{y}_{1:t}, \Theta)d\mathbf{h}_{t+1} \quad (4.23)$$

$$= \int p(\mathbf{y}_{t+1}|\mathbf{h}_{t+1}, \Theta)p(\mathbf{h}_{t+1}|\mathbf{y}_{1:t}, \Theta)d\mathbf{h}_{t+1}. \quad (4.24)$$

Since

$$p(\mathbf{h}_{t+1}|\mathbf{y}_{1:t}, \Theta) = \int p(\mathbf{h}_{t+1}, \mathbf{h}_t|\mathbf{y}_{1:t}, \Theta)d\mathbf{h}_t \quad (4.25)$$

$$= \int p(\mathbf{h}_{t+1}|\mathbf{h}_t, \mathbf{y}_{1:t}, \Theta)p(\mathbf{h}_t|\mathbf{y}_{1:t}, \Theta)d\mathbf{h}_t \quad (4.26)$$

$$= \int p(\mathbf{h}_{t+1}|\mathbf{h}_t, \Theta)p(\mathbf{h}_t|\mathbf{y}_{1:t}, \Theta)d\mathbf{h}_t, \quad (4.27)$$

the conditional density of \mathbf{y}_{t+1} can be calculated if we know the one-step ahead prediction distribution of \mathbf{h}_{t+1} , where the latter depends on the filtered density of \mathbf{h}_t . The challenge here is that we do not know the filter distribution of \mathbf{h}_t explicitly. Our solution is to approximate this filter density (or posterior distribution) with an efficient sequential Monte Carlo particle filtering procedure, the APF in Pitt and Shephard (1999b), which is also applied in Chib et al. (2006), Omori et al. (2007) and among others for FSV models. Suppose that we approximate a sample $\{\mathbf{h}_t^{(i)}, i = 1, \dots, N\}$ of \mathbf{h}_t from the filtered distribution of $(\mathbf{h}_t|\mathbf{y}_{1:t}, \Theta)$ with the weights $\{\pi_i, i = 1, \dots, N\}$ such that $\sum_{i=1}^N \pi_i = 1$. Given this sample, the one-step ahead predictive density of \mathbf{h}_{t+1} is

$$p(\mathbf{h}_{t+1}|\mathbf{y}_{1:t}, \Theta) \approx \frac{1}{N} \sum_{i=1}^N p(\mathbf{h}_{t+1}|\mathbf{h}_t^{(i)}, \Theta). \quad (4.28)$$

Under this particle approximation, the one-step ahead conditional density (4.22) can

be calculated numerically. The APF used here is similar to that in Chib et al. (2006), and is given in the Appendix. In our experience $N = 10,000$ particles are sufficient for our simulation studies and the index return data that we use to illustrate our model and the estimation method.

4.4 Empirical Exploration and Comparative Analysis

We conducted simulation studies for the one-factor and two-factor PPCAF models, where the observation errors are either isotropic or idiosyncratic. In total, there are four cases to be examined. The following procedure was used to generate multivariate time series of asset returns.

$$h_{k,t+1} \sim \mathcal{N}(\mu_k + \phi_k(h_{k,t} - \mu_k), \sigma_k), \quad t = 1, \dots, T - 1, \quad (4.29)$$

$$\mathbf{y}_t \sim \mathcal{N}_m(\mathbf{0}, \mathbf{W} \text{diag}(\exp(h_{1,t}), \dots, \exp(h_{q,t})) \mathbf{W}' + \mathbf{\Psi}), \quad t = 1, \dots, T, \quad (4.30)$$

where $h_{k,1} \sim \mathcal{N}(\mu_k, \sigma_k / (1 - \phi_k^2))$, $k = 1, \dots, q$, are the initial distributions of \mathbf{h}_1 .

We considered an eight-dimensional problem in each of the four cases based on generated return data from a PPCAF model indexed by true parameters presented in Tables 4.3 and 4.4. As discussed earlier, there are 36 free parameters in total provided by the sample variance-covariance matrix. The generated return data can be fitted by our proposed PPCAF model with up to four factors in the isotropic type and three factors in the idiosyncratic type. For each data set, we fit the specific type of PPCAF model with the number of factors from 1 to 3 and employed AIC and BIC for model selection. There are 1500 observations in each of the following examples, from which the first 1000 observations were analyzed and the other 500 observations were used for the assessment of out-of-sample one-step ahead forecast. To estimate the PPCAF model, the derived MCMC algorithm was iterated for 50,000 iterations and the first

10,000 sampled values were discarded as the burn-in. All parameters and augmented parameters were then estimated by sample averages of the full conditionals. The estimates of the parameters were used as the true values when the APF was applied to evaluate the likelihoods of data and conduct the in-sample and out-of sample volatility forecast. We chose 10,000 particles when the filter and the prediction densities of latent random variables were approximated recursively.

It should be mentioned, for each of the four types of PPCAF models, we conducted several experiments using simulated asset return data to see how often that AIC and BIC correctly identified the true number of factors. It was found in all replications that AIC and BIC always identified the PPCAF model used to generate the asset returns. We just take one example from each type of the four PPCAF models and present the analysis results. Table 4.2 summarizes the AIC and BIC values from the specified PPCAF models where the boldface values are the smallest. Table 4.3 compares the true and estimated parameters of the latent Markov processes, where $\theta_{f_k} = (\mu_k, \phi_k, \sigma_k)$, $k = 1, 2$, are the vectors of parameters for the latent AR(1) processes. The comparison between the true and estimated loading matrix is provided in Table 4.4. It is clear that the estimated parameters are very close to the true values used to generate the asset returns.

These examples provide some evidence that our proposed PPCAF model and corresponding estimation method work well, and the AIC and BIC criteria can correctly identify the number of factors contained in the observation processes.

4.4.1 Forecast Assessment for the two-factor Idiosyncratic PPCAF Model

In terms of forecasting, we forecast in-sample and out-of-sample volatilities for a two-factor idiosyncratic PPCAF model using 2000 simulated observations. Since MCMC estimation methods are generally time consuming, we used the selected two-factor

Table 4.2: Model selection using the AIC and BIC criteria.

Type	True q	Criterion	$q = 1$	$q = 2$	$q = 3$
Isotropic	1	AIC	-49808.7	-49785.2	-49762.2
		BIC	-49749.8	-49672.3	-49595.3
	2	AIC	-48917.1	-51442.7	-51419.8
		BIC	-48863.1	-51339.6	-51267.6
Idiosyncratic	1	AIC	-50761.3	-50743	-50747.6
		BIC	-50668.1	-50595.7	-50546.4
	2	AIC	-50858.8	-52820.6	-52802.4
		BIC	-50765.6	-52673.4	-52601.2

Table 4.3: Comparison between true and estimated parameters of the latent AR(1) processes.

Type	Factor q	Parameter	True	Estimate
Isotropic	1	θ_{f_1}	(-8.11, 0.99, 0.03)	(-8.07, 0.99, 0.04)
	2	θ_{f_1}	(-7.89, 0.99, 0.02)	(-7.56, 0.99, 0.03)
		θ_{f_2}	(-9.09, 0.98, 0.06)	(-8.99, 0.98, 0.05)
Idiosyncratic	1	θ_{f_1}	(-8.45, 0.99, 0.03)	(-8.03, 0.99, 0.03)
	2	θ_{f_1}	(-7.89, 0.99, 0.02)	(-7.05, 0.99, 0.03)
		θ_{f_2}	(-9.34, 0.99, 0.06)	(-8.33, 0.99, 0.06)

PPCAF model fitted from the first 1000 observations, when forecasting in-sample and out-of-sample one-step ahead volatilities. The multi-step-ahead forecast can be done in the same fashion.

First, we examine the two-factor PPCAF idiosyncratic model in terms of volatility forecast. The idiosyncratic PPCAF model was fitted with $q = 0, 1, 2, 3$ factors to the simulated data. That is, we also considered the situation with $q = 0$, where the data

Table 4.4: Comparison between true and estimated parameters of the measurement equations.

Type	q	Par.	True	Estimate
Iso.	1	\mathbf{W}'	0.33, 0.40, 0.41, 0.36, 0.34, 0.40, 0.31, 0.24	0.32, 0.40, 0.42, 0.37, 0.34, 0.41, 0.30, 0.24
		ψ	0.0097	0.0097
	2	\mathbf{W}'	0.33, 0.40, 0.42, 0.37, 0.34, 0.41, 0.29, 0.23 -0.11, -0.27, -0.30, -0.24, 0.07, 0.03, 0.61, 0.62	0.33, 0.40, 0.42, 0.36, 0.34, 0.41, 0.29, 0.23 -0.10, -0.28, -0.31, -0.24, 0.08, 0.03, 0.64, 0.58
		ψ	0.0073	0.0073
Idi.	1	\mathbf{W}'	0.33, 0.40, 0.41, 0.36, 0.34, 0.41, 0.31, 0.24	0.32, 0.38, 0.41, 0.35, 0.33, 0.40, 0.33, 0.29
		Ψ	0.0077, 0.0070, 0.0084, 0.0066, 0.0081, 0.0091, 0.0135, 0.0136	0.0081, 0.0078, 0.0089, 0.0074, 0.0084, 0.0093, 0.0128, 0.0127
	2	\mathbf{W}'	0.33, 0.40, 0.42, 0.37, 0.34, 0.41, 0.29, 0.23 -0.11, -0.27, -0.30, -0.24, 0.07, 0.03, 0.62, 0.62	0.32, 0.39, 0.40, 0.34, 0.35, 0.42, 0.32, 0.27 -0.12, -0.28, -0.34, -0.25, 0.02, -0.01, 0.60, 0.61
		Ψ	0.0073, 0.0048, 0.0062, 0.0048, 0.0078, 0.0087, 0.0089, 0.0091	0.0073, 0.0057, 0.0067, 0.0056, 0.0077, 0.0084, 0.0084, 0.0086

In order to save space, we use Par. for parameter, Iso. for isotropic and Idi. for idiosyncratic, respectively.

does not depend on any factors, and the situation with $q = 3$ where we added an extra factor for comparison purpose. The standard deviations of univariate time series of asset returns are compared with the root mean square errors (RMSEs) obtained from the idiosyncratic PPCAF model with different number of factors in Table 4.5, where the RMSEs were calculated based on the differences between the true and forecasted volatilities. The model used for forecasting was estimated using only the first 1000 observations. The "In" columns in Table 4.5 show the in-sample forecast RMSEs for the first 1000 observations, while the columns "Out" show the RMSEs for the second half of the data. It is obvious that the RMSEs with $q > 0$ are smaller than the RMSEs with $q = 0$. This result indicates that by introducing factors into the SV model the RMSEs can be highly reduced. We also observe that between $q = 2$ and $q = 3$ the forecast RMSEs do not change much. It is evident that the correct number of factors has been detected. By using the AIC and BIC, we can also reach the same conclusion in this experiment. That means, the third factor is not significant in terms of forecasting volatility. Figures 4.1 and 4.2 compare the generated time series of the two factors and their true volatilities with the corresponding estimated and in-sample forecasted volatilities. It is easy to see the forecasted volatilities of factors resemble the true volatilities.

Next, we compare the forecasted volatilities between the two-factor PPCAF and univariate SV models. The first 1000 observations in each component were also fitted by the univariate SV model introduced in Chapter 2, and the estimated univariate SV models were employed to perform the in-sample and out-of-sample volatility forecast. Similarly we compare the absolute value of the generated returns and the true volatilities with the in-sample and out-of-sample forecasted volatilities. In Figure 4.3, we present a comparison for the third component of the multivariate time series of asset returns based on the two-factor PPCAF and univariate SV models, where the asset returns before and after the vertical dotted line at $t = 1000$ are the in-sample and out-of-sample volatility forecasts, respectively. The RMSEs from the PPCAF model is 0.0022

Table 4.5: RMSE comparison between true and forecasted volatilities on the simulated eight-dimensional return data from an idiosyncratic two-factor PPCAF model. All values have been multiplied by 1000.

Component	$q = 0$		$q = 1$		$q = 2$		$q = 3$	
	In	Out	In	Out	In	Out	In	Out
1	10.4	10.0	1.4	1.8	1.4	1.7	1.3	1.8
2	10.2	10.8	2.4	2.6	1.8	2.4	1.8	2.4
3	11.6	12.2	2.5	2.6	1.9	2.4	1.9	2.4
4	9.8	10.2	2.1	2.4	1.7	2.1	1.7	2.2
5	10.3	11.1	1.2	1.7	1.1	1.7	1.1	1.7
6	11.7	13.1	1.4	2.1	1.4	2.1	1.4	2.2
7	14.5	12.9	5.3	3.5	2.8	2.2	2.8	2.2
8	14.3	12.6	5.3	3.5	2.7	2.0	2.8	2.0

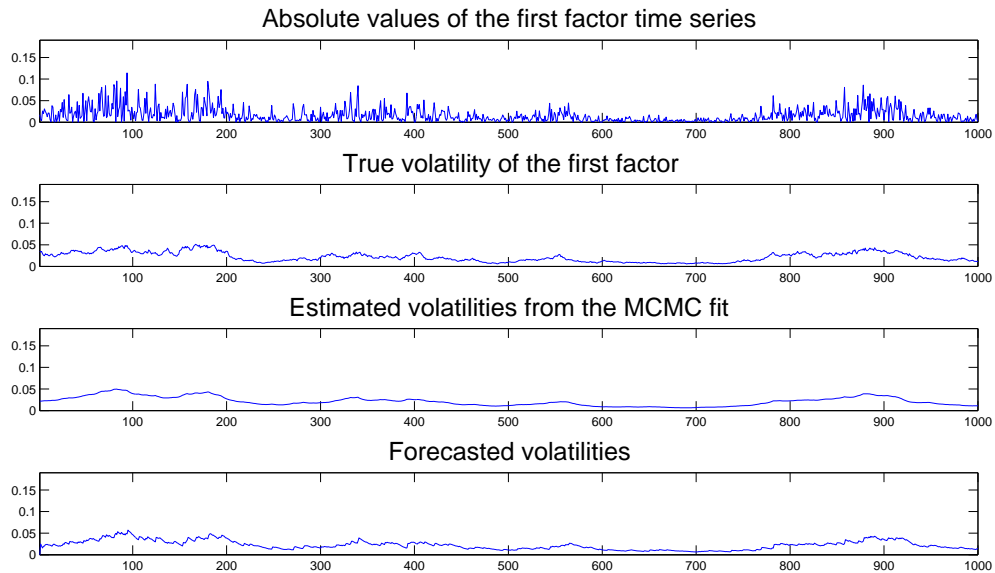


Figure 4.1: Comparison of the absolute first factor time series and the corresponding true volatilities with the estimated and the in-sample forecasted volatilities based on a simulated data set from the two-factor idiosyncratic PPCAF model.

which is smaller than 0.0026 from the univariate model. Thus, the proposed PPCAF

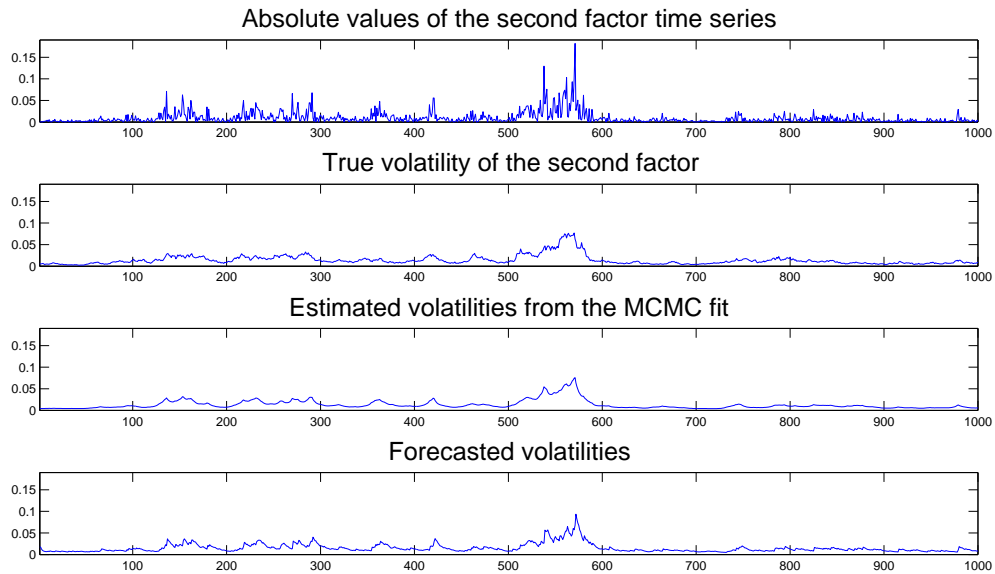


Figure 4.2: Comparison of the absolute second factor time series and the corresponding simulated volatilities with the estimated and the in-sample forecasted volatilities based on a simulated data set from the two-factor idiosyncratic PPCAF model.

model is able to forecast the volatilities of multivariate time series of asset returns.

A more sophisticated way to check the forecast ability is called the block assessment. To do this, we first generate multivariate time series of asset returns with N observations from a PPCAF model, and then divide the data into j blocks, each of the blocks consists of N/j observations. Within each block, the N/j observations were fitted by the same type of PPCAF model, and the resulting model is used to perform the in-sample volatility forecast within the block and the out-of-sample volatility forecast for the successive block. In the first block, the in-sample forecast is conducted and for other four blocks both out-of-sample and in-sample forecasts are performed. In this example, we studied a two-factor idiosyncratic PPCAF model with settings $m = 8$, $N = 10,000$ and $j = 5$. Figure 4.4 compares out-of-sample forecast RMSEs based on the last four block data from the two-factor idiosyncratic PPCAF and univariate SV

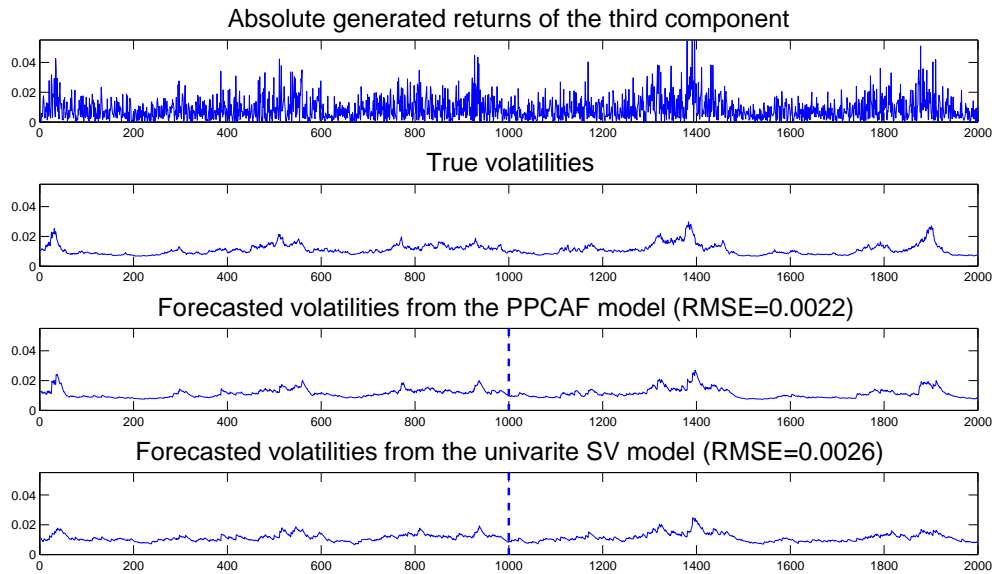


Figure 4.3: Comparison of the absolute simulated third time series with the in-sample and out-of-sample forecasts based on a simulated data set from the two-factor idiosyncratic PPCAF model.

models. It shows that the RMSEs of the out-of-sample forecast from the two-factor idiosyncratic PPCAF are smaller than those from the univariate SV models. Figure 4.5 compares the time series of true volatility and the out-of-sample forecasted volatilities of the first component in the last four blocks.

Again, the forecast errors from the PPCAF model are smaller than those from the univariate SV models. The reason for this is that the PPCAF model takes advantage of the cross correlations that drive the multivariate time series of asset returns, while the univariate SV model only captures the correlation between the observed time series and the corresponding latent volatility process. In the analysis of asset return data, although the true volatilities are unobservable, we expect the PPCAF model to perform better than the univariate SV model in terms of forecasting volatility.

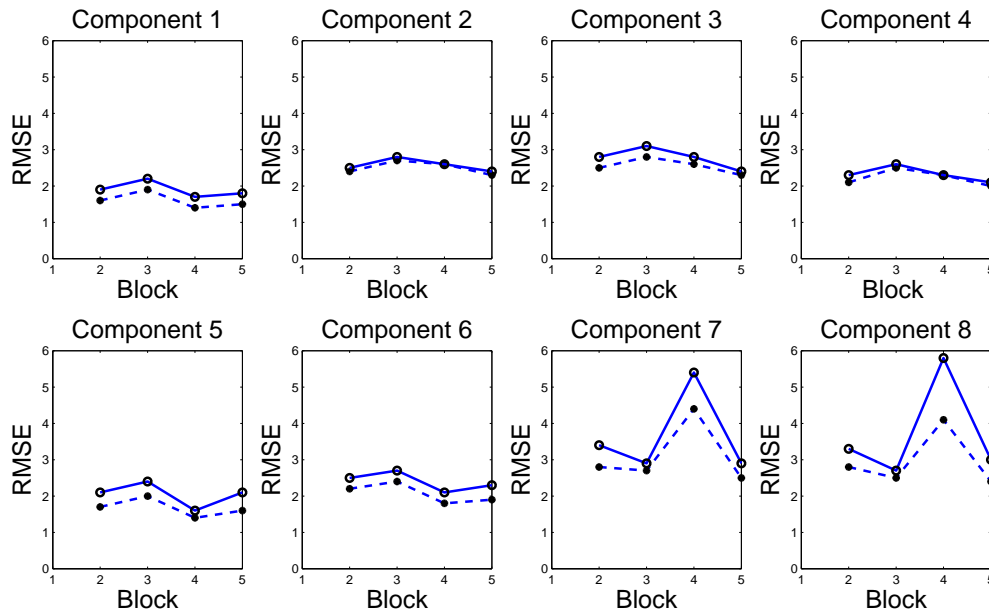


Figure 4.4: RMSE comparison between true and out-of-sample forecasted volatilities through the two-factor idiosyncratic PPCAF models (the dotted line) and univariate SV models (the solid line) based on the last four block data.

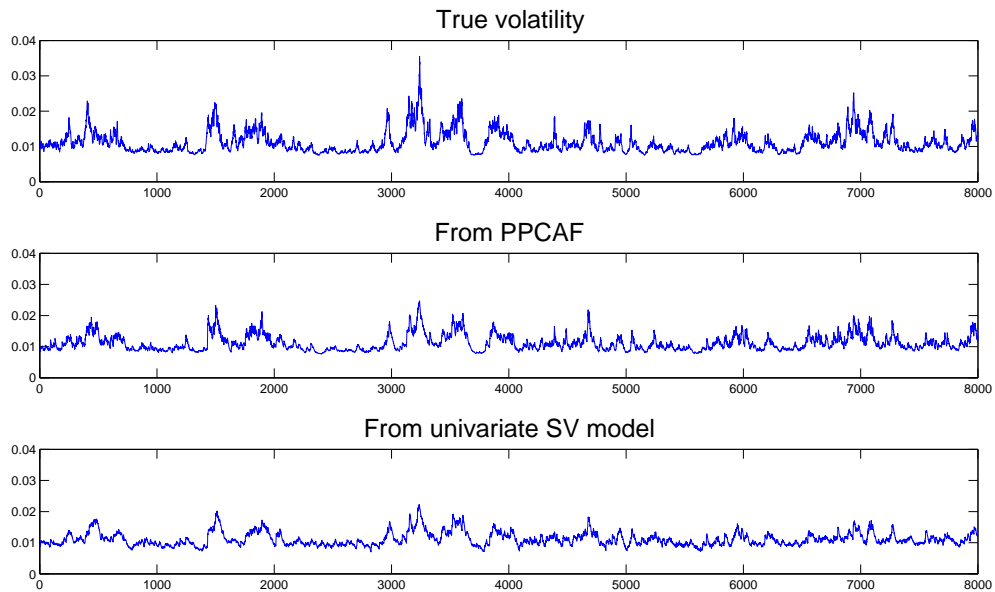


Figure 4.5: Comparison between true and out-of-sample forecasted volatilities of the first component from the two-factor idiosyncratic PPCAF and univariate SV models in the last four blocks.

4.5 Factor Structure of Returns of International Stock Indices

We considered a return data set of eight international stock indices. The AIC and BIC were the main tools to determine the number of factors within the upper boundary. Since the volatilities of the return time series are unavailable, we compared the in-sample and out-of-sample one-step ahead forecasted volatilities with the absolute asset returns.

4.5.1 Model Fit and Data Analysis

The time series considered here are the returns of stock indices from eight main stock markets: the Toronto Stock Exchange (TSX), the Standard & Poor's 500 (S&P500), the NASDAQ Stock Market (NASDAQ), the Dow Jones Industrial Average (DJI), the London Stock Exchange FTSE 100 Index (FTSE), the Frankfurt Stock Exchange index (DAX), the Hang Seng Index (HS) and the Nikkei Stock Average (Nikkei) from the Tokyo Stock Exchange. The data was downloaded from the web site finance.yahoo.com, for the period from January 6, 2003 to July 8, 2011 yielding 2204 observations for each index. Because this is an eight-dimensional data set, the PPCAF model can be fit with up to $q = 4$ factors in the isotropic case and $q = 3$ for the idiosyncratic model. As in the simulation studies, we fit the data based on the first 2000 observations and the last 204 observation were used for the assessment of out-of-sample one-step ahead volatility forecast.

To determine how well the PPCAF model fits the data, we used the AIC and BIC. To do this, the return data was fitted by the isotropic PPCAF models with up to four factors and the idiosyncratic PPCAF models with up to three factors, respectively. Table 4.6 presents the AIC and BIC values for the fitted PPCAF models. Both AIC and

BIC attained their smallest values at the upper boundaries for the two types of models. By looking at these values, we found that the three-factor idiosyncratic PPCAF model has the smallest AIC and BIC values and, therefore, this model is a preferred candidate for the asset return data.

Table 4.6: Model selection using the AIC and BIC criteria.

Type	Criterion	$q = 1$	$q = 2$	$q = 3$	$q = 4$
Isotropic	AIC	-98918.6	-104068	-106585	-108431
	BIC	-98851.4	-103940	-106395	-108179
Idiosyncratic	AIC	-102076	-106690	-109659	
	BIC	-101969	-106522	-109429	

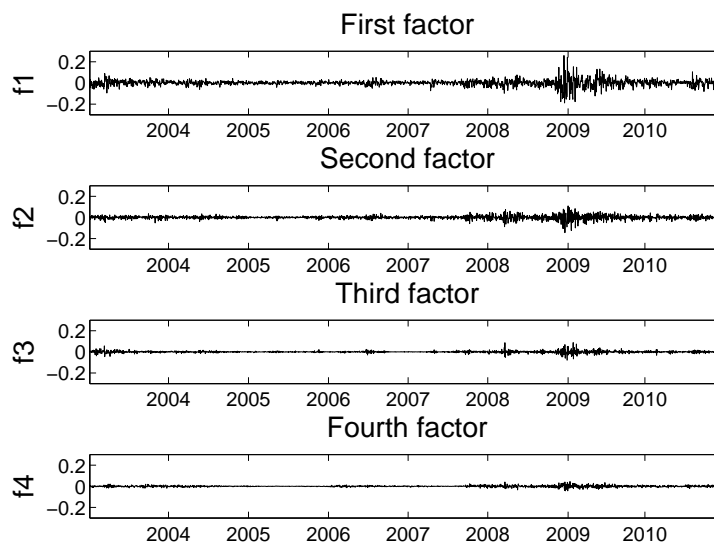


Figure 4.6: Time series comparison of the estimated factors from the indices return data using the four-factor idiosyncratic PPCAF model.

Figure 4.6 plots the time series of the estimated four factors from the idiosyncratic four-factor PPCAF model, where the magnitudes of the third and fourth factor time series are much smaller than those of the first two factor time series. Accordingly, this

provides additional evidence that a two or three factor PPCAF model is an adequate representation for the international indices return data.

Table 4.7 includes summaries of standard errors and Bayesian highest probability density (HPD) confidence intervals for the parameters. It is clear that the 95% HPD intervals contain the estimated values.

Table 4.7: Estimated parameters of the three latent AR(1) processes from the three-factor PPCAF model with idiosyncratic observation errors.

Parameter	Estimate	Std.	HPD CI (95%)
μ_1	-7.78	0.80	(-8.78, -6.74)
ϕ_1	0.99	0.01	(0.98, 0.99)
σ_1	0.02	0.01	(0.01, 0.03)
μ_2	-9.04	0.72	(-10.19, -7.89)
ϕ_2	0.99	0.01	(0.98, 0.99)
σ_2	0.04	0.01	(0.02, 0.06)
μ_3	-10.33	0.39	(-11.12, -9.61)
ϕ_3	0.97	0.01	(0.96, 0.99)
σ_3	0.13	0.04	(0.06, 0.23)

For model assessment, as in Pitt and Shephard (1999a), we can compare the sample variance-covariance matrix of the observed data given as Σ in (4.31) with the Bayesian mean $\hat{\Sigma}$ of the unconditional variance-covariance matrix in (4.32) from the three-factor idiosyncratic PPCAF model using the formulas (4.11) and (4.12). The two matrices (having been multiplied by 1000) do not differ much, indicating that the variation of

the index returns is mostly accounted for by the first three leading factors.

$$\Sigma = \begin{pmatrix} 0.1469 & 0.1207 & 0.1232 & 0.1056 & 0.0872 & 0.0946 & 0.0633 & 0.0454 \\ 0.1207 & 0.1764 & 0.1820 & 0.1600 & 0.0926 & 0.1182 & 0.0516 & 0.0293 \\ 0.1232 & 0.1820 & 0.2096 & 0.1630 & 0.0936 & 0.1250 & 0.0529 & 0.0305 \\ 0.1056 & 0.1600 & 0.1630 & 0.1500 & 0.0848 & 0.1090 & 0.0477 & 0.0279 \\ 0.0872 & 0.0926 & 0.0936 & 0.0848 & 0.1606 & 0.1556 & 0.0825 & 0.0717 \\ 0.0946 & 0.1182 & 0.1250 & 0.1090 & 0.1556 & 0.2161 & 0.0884 & 0.0769 \\ 0.0633 & 0.0516 & 0.0529 & 0.0477 & 0.0825 & 0.0884 & 0.2666 & 0.1595 \\ 0.0454 & 0.0293 & 0.0305 & 0.0279 & 0.0717 & 0.0769 & 0.1595 & 0.2338 \end{pmatrix}, \quad (4.31)$$

$$\hat{\Sigma} = \begin{pmatrix} 0.1382 & 0.1101 & 0.1160 & 0.1000 & 0.0781 & 0.0959 & 0.0585 & 0.0409 \\ 0.1101 & 0.1616 & 0.1550 & 0.1329 & 0.0886 & 0.1117 & 0.0476 & 0.0252 \\ 0.1160 & 0.1550 & 0.1930 & 0.1405 & 0.0921 & 0.1164 & 0.0485 & 0.0247 \\ 0.1000 & 0.1329 & 0.1405 & 0.1380 & 0.0812 & 0.1022 & 0.0441 & 0.0238 \\ 0.0781 & 0.0886 & 0.0921 & 0.0812 & 0.1615 & 0.1544 & 0.0641 & 0.0663 \\ 0.0959 & 0.1117 & 0.1164 & 0.1022 & 0.1544 & 0.2176 & 0.0682 & 0.0700 \\ 0.0585 & 0.0476 & 0.0485 & 0.0441 & 0.0641 & 0.0682 & 0.2604 & 0.1678 \\ 0.0409 & 0.0252 & 0.0247 & 0.0238 & 0.0663 & 0.0700 & 0.1678 & 0.2274 \end{pmatrix}. \quad (4.32)$$

The leading three eigenvalues (multiplied by 10000) of the sample variance-covariance matrix (4.31) are 8.712, 3.568 and 1.412 with corresponding eigenvectors $\hat{\mathbf{E}}'$ listed in (4.33). The estimated loading matrix $\hat{\mathbf{E}}'_1$ is provided in (4.34) with the corresponding factor variances (multiplied by 10000) equal to 7.709, 2.928 and 1.157. For a comparison purpose, the estimated eigenvalues from the isotropic PPCAF model are 7.619, 2.966 and 1.299 (multiplied by 10000) with corresponding eigenvectors provided as $\hat{\mathbf{E}}'_2$ in (4.35). It is observed that the eigenvalues and the corresponding eigenvectors from

the sample variance-covariance matrix (4.31) are close to the columns of the estimated loading matrices (4.34) and (4.35), and the leading eigenvalues of the sample variance-covariance matrix are closely related to the variances of extracted factors. Also the corresponding eigenvectors and the columns of the loading matrices appear to point to similar directions.

$$\hat{\mathbf{E}}' = \begin{pmatrix} 0.326 & 0.397 & 0.419 & 0.361 & 0.339 & 0.408 & 0.302 & 0.240 \\ -0.110 & -0.279 & -0.305 & -0.249 & 0.066 & 0.022 & 0.615 & 0.611 \\ 0.143 & 0.230 & 0.268 & 0.199 & -0.543 & -0.637 & 0.325 & 0.100 \end{pmatrix}, \quad (4.33)$$

$$\hat{\mathbf{E}}_1' = \begin{pmatrix} 0.330 & 0.380 & 0.406 & 0.348 & 0.352 & 0.425 & 0.305 & 0.252 \\ -0.118 & -0.247 & -0.279 & -0.222 & 0.009 & -0.060 & 0.635 & 0.627 \\ 0.190 & 0.248 & 0.304 & 0.223 & -0.527 & -0.611 & 0.326 & 0.049 \end{pmatrix}, \quad (4.34)$$

$$\hat{\mathbf{E}}_2' = \begin{pmatrix} 0.323 & 0.389 & 0.408 & 0.354 & 0.354 & 0.422 & 0.300 & 0.243 \\ -0.111 & -0.279 & -0.304 & -0.249 & 0.055 & -0.010 & 0.621 & 0.605 \\ 0.143 & 0.247 & 0.268 & 0.219 & -0.548 & -0.607 & 0.354 & 0.061 \end{pmatrix}. \quad (4.35)$$

Lastly, we check whether the factors appear to be independent of the innovations by plotting the realized observation errors against each of the estimated factors, where the observation errors can be calculated by the following formula

$$\hat{R}_{i,t} = \mathbf{y}_t - \sum_{k=1}^q \hat{W}_k \hat{f}_{k,t}, \quad t = 1, \dots, T; i = 1, \dots, q.$$

Here we only give the scatter plots of these errors versus the estimated first factor in Figure 4.7. There is no systematic pattern discernible among these scatter plots and the assumption of independence between the first factor and observation errors is acceptable.

The estimated idiosyncratic standard errors, listed in (4.36), are quite different show-

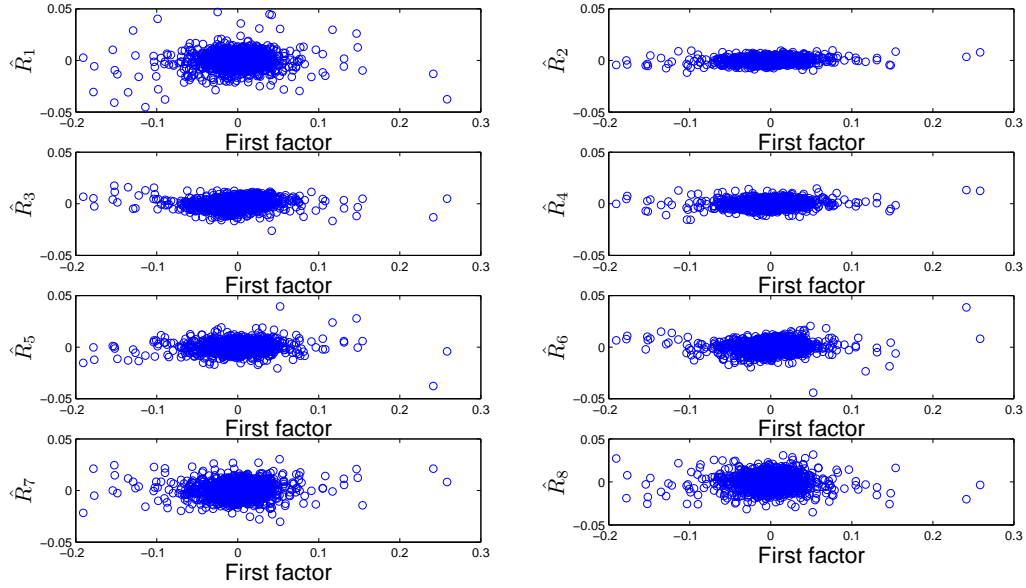


Figure 4.7: Check the independence between the first factor and additive innovations.

ing that the projection errors are asset dependent.

$$\hat{\Psi} = \text{diag} \left(0.0070 \quad 0.0038 \quad 0.0053 \quad 0.0041 \quad 0.0053 \quad 0.0058 \quad 0.0080 \quad 0.0088 \right). \quad (4.36)$$

An important alternative model assessment is to check the percentages of the marginal variances explained by individual factors. For each return series $i = 1, \dots, m$, the percentage $d_{i,k}$ of the variance explained by each factor $k, k = 1, \dots, q$, is simply given by

$$d_{i,k} = \frac{W(i,k)^2 \sigma_{f_k}^2}{\sum_{l=1}^q W(i,l)^2 \sigma_{f_l}^2 + \psi_i^2} \times 100.$$

Table 4.8 summarizes the values of these quantities with W and ψ_i and $\sigma_{f_k}, k = 1, \dots, q$, estimated at their posterior means. This table will be referred to later when we conclude the discussion of the empirical analysis.

We can draw various conclusions based on the estimated loading matrix (4.34), the

Table 4.8: Percentage of the variance of each series explained by each factor in analysis of the international indices return time series from the three-factor idiosyncratic PPCAF model.

Index	Factor 1	Factor 2	Factor 3	Total
TSX	58.53	1.97	1.91	62.42
S&P 500	73.90	11.88	5.00	90.78
NASDAQ	68.11	11.77	5.02	84.90
DJI	71.77	11.02	4.56	87.35
FTSE	60.00	0.16	21.57	81.63
DAX	63.90	0.05	19.63	83.58
HS	24.23	47.53	4.32	76.08
Nikkei	17.99	47.71	0.04	65.74

percentage Table 4.8 and the estimated idiosyncratic standard errors in (4.36) for the indices returns. The first factor represents a dominating common market factor that influences the stock market indices. All eight markets have similar sensitivity to this factor since the components of the first loading vector in (4.34) are positive and nearly equal. These market indices will go up or down when this factor changes. The second factor has an interesting interpretation. Looking at the second loading vector in (4.34), the entries for the North American market are (-0.118 -0.247 -0.279 -0.222), which are all negative and far from zero, while the entries for the European market are (0.009 -0.060), which are close to zero and the entries (0.635 0.627) for the Asian market are positive and large. This factor is negatively correlated with the North American market and strongly positively correlated with the Asian market. In other words, the North American market will have a minor negative response to the second factor and the Asian market has a strong positive response. The third panel of Figures 4.11 to 4.18 shows volatility contributed by this second factor to the eight indices. For the third factor, all components in the third column of the loading matrix (4.34) are positive except for the European market where the components (-0.527 -0.611) are negative. As a consequence, this factor will cause the European market to move in the opposite

direction of the other markets. Since the elements of the European markets in this loading vector have the largest absolute values, this market depends strongly on the third factor, which is seen in the fourth panel of Figures 4.15 and 4.16. Table 4.8 is also informative. First, the weights on factor 1 are relatively high, indicating that the first factor explains most of the variation except in the Asian indices. Second, the Canadian market puts little weight on the second and third factors and all three only explain about 62% of the volatility. The United States and Asian markets largely depend on the two leading factors shown in Figures 4.12 to 4.14 and 4.17 to 4.18. Finally, the European markets are largely driven by the first and third factors. In summary, we might say that the first factor is the dominant factor representing the global market changes, while the second factor is an Asian factor and the third factor represents the European market. The last column in Table 4.8 includes the ratio of the variances explained by these three factors to the total marginal variance. We notice that the Canadian market and the Asian market are less influenced by these three factors, whereas the others account for more than 80% of their volatility. If we look at the estimated standard idiosyncratic errors in (4.36), the first and the last two entries are larger, confirming this fact.

The factor effects can also be visualized in Figure 4.8, which plots the cumulative percentages explained by the three extracted factors. The positive signs in the parts that represent the factor effects from the first two factors indicate that the corresponding loadings are positive, while the two positive signs in the third factor areas indicate that the corresponding loadings are opposite to that of the other markets. It can be seen that the second factor has almost no impact on the two European markets, and the third factor does not contribute much to the Japanese market.

It is also reasonable to check the correlations between the eight international markets. The estimated correlation matrix $\hat{\mathbf{R}}$ is given in (4.37). We see that the eight index returns are positively correlated. It is noticed that the four indices of the North American market are highly correlated, the correlation coefficient between the two Eu-

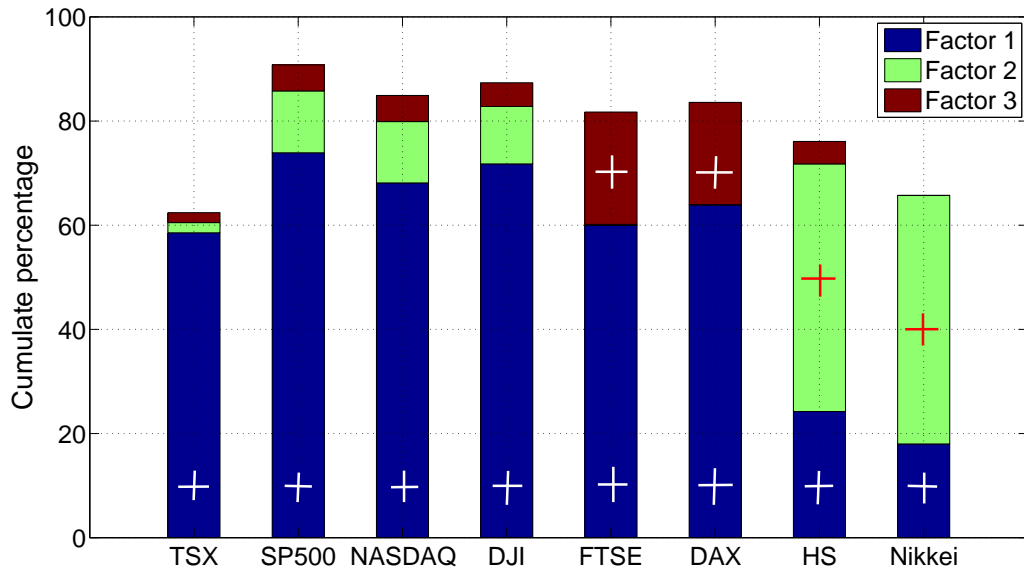


Figure 4.8: Factor effect comparison.

ropean indices is 0.8236 indicating a high correlation and the two Asian markets are also highly correlated with correlation coefficient 0.6894.

$$\hat{\mathbf{R}} = \begin{pmatrix} 1.0000 & 0.7370 & 0.7105 & 0.7243 & 0.5229 & 0.5533 & 0.3086 & 0.2305 \\ 0.7370 & 1.0000 & 0.8778 & 0.8905 & 0.5486 & 0.5956 & 0.2320 & 0.1313 \\ 0.7105 & 0.8778 & 1.0000 & 0.8609 & 0.5218 & 0.5679 & 0.2163 & 0.1178 \\ 0.7243 & 0.8905 & 0.8609 & 1.0000 & 0.5440 & 0.5897 & 0.2325 & 0.1345 \\ 0.5229 & 0.5486 & 0.5218 & 0.5440 & 1.0000 & 0.8236 & 0.3124 & 0.3462 \\ 0.5533 & 0.5956 & 0.5679 & 0.5897 & 0.8236 & 1.0000 & 0.2864 & 0.3147 \\ 0.3086 & 0.2320 & 0.2163 & 0.2325 & 0.3124 & 0.2864 & 1.0000 & 0.6894 \\ 0.2305 & 0.1313 & 0.1178 & 0.1345 & 0.3462 & 0.3147 & 0.6894 & 1.0000 \end{pmatrix}, \quad (4.37)$$

4.5.2 Forecasting Analysis

As in the simulation studies, we performed in-sample and out-of-sample one-step ahead volatility forecasts for the fitted model. All Bayesian estimated parameters from the first 2000 observations were taken as the true parameters. Figures 4.8 plots the dynamics of the estimated first factor with the corresponding estimated and forecasted volatilities, whereas in Figures 4.9, the absolute returns of NASDAQ is compared with the in-sample and out-sample forecasted volatilities separated by the vertical dotted line at $t = 2000$. It is demonstrated that the magnitude of the estimated volatilities of the first factor and that of the asset returns of NASDAQ stock are similar to those of the corresponding forecasted volatilities. As illustrated in the simulation studies, although we can not observe the true volatilities of asset returns, we are confident that the one-step-ahead forecasted volatilities can resemble future volatilities of the asset returns.

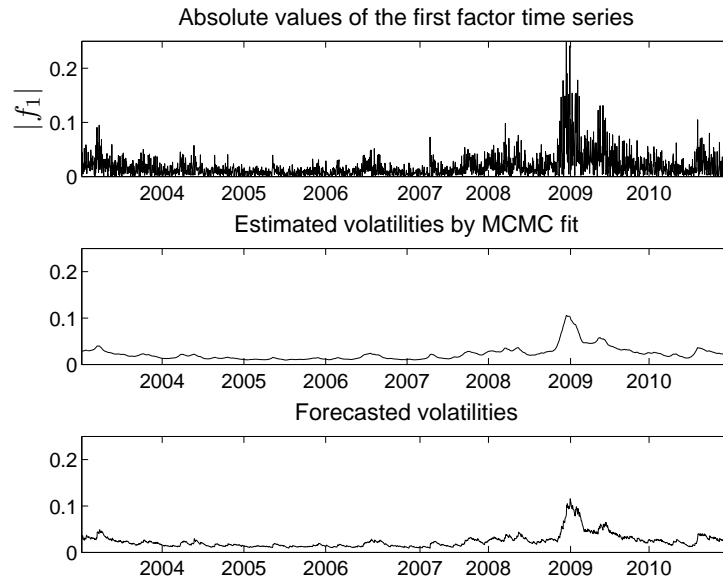


Figure 4.9: Comparison of the estimated first factor time series with the corresponding MCMC estimated and forecasts volatilities based on the international return data set.

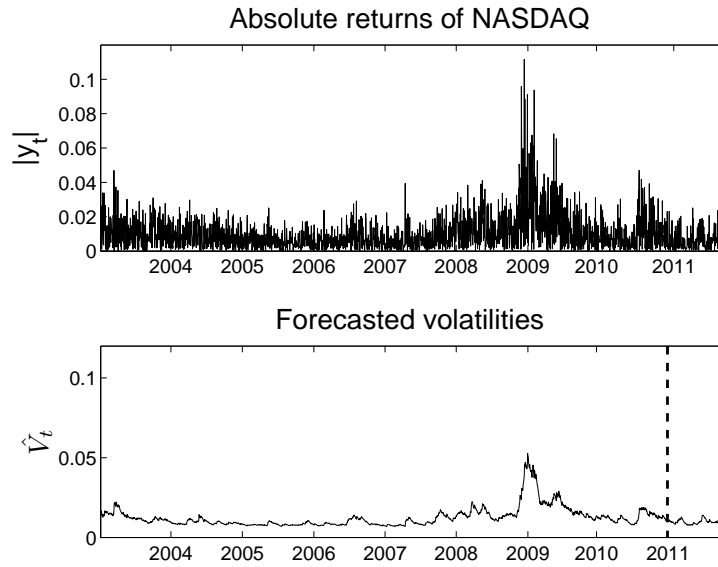


Figure 4.10: Comparison between the absolute returns of NASDAQ with the one-step ahead forecasted volatilities.

The application to the international index returns illustrates that our model and the corresponding estimation method are able to capture the market dynamics, and the extracted three factors characterize the economic changes that affect the world-wide stock markets. In addition to this, with the help of the APF, the PPCAF model can perform one-step ahead volatility forecast reasonably well.

4.6 Conclusion and Remarks

In this chapter, we proposed a factor stochastic volatility model with an orthonormal loading matrix. The aim of the model is to extract the key factors that govern the movement of underlying time series of financial asset returns. In our PPCAF model, the factors contribute to the underlying asset returns in orthonormal directions. An MCMC method has been developed for the Bayesian estimation of parameters. The complexity of sampling of the loading matrix was resolved by simulation via the von Mises-Fisher distribution while the orthogonality of the columns is preserved in the process. The slice sampler within the MCMC method derived in Chapter 2 was applied

to simulate the latent states. Since the slice sampler adapts the analytical expressions of the target densities, this procedure may make the simulation of full conditionals faster than other sampling schemes introduced in the literature, where complex Metropolis-Hastings algorithms would otherwise be needed. Our model not only can remove the rotation ambiguity of the estimated loading matrix, but also permits factors to vary stochastically. The PPCAF model can explain most of the properties of the observed return data. Simulation studies confirm that the true factors can be extracted correctly and the assumptions for the PPCAF models are satisfied. In an application to financial return data, the latent factors that drive the market evolution can be characterized, and the fitted model was shown to satisfy the assumed conditions.

4.7 Appendix

The APF algorithm for the PPCAF model

The algorithm for our PPCAF model based on the procedure in Chib et al. (2006).

Step 1. Given a sample $\{\mathbf{h}_t^{(i)}, i = 1, \dots, N\}$ from $(\mathbf{h}_t | \mathbf{y}_{1:t}, \Theta)$, we calculate the expectation $\hat{\mathbf{h}}_{t+1}^{*(i)} = E(\mathbf{h}_{t+1} | \mathbf{h}_t^{(i)})$ and

$$\pi_i = p(\mathbf{y}_{t+1} | \hat{\mathbf{h}}_{t+1}^{*(i)}, \Theta), i = 1, \dots, N. \quad (4.38)$$

Sample N times with replacement the integers of $1, \dots, N$ with probability $\hat{\pi}_i = \pi_i / \sum_{i=1}^N \pi_i$. Define the sampled indexes n_1, \dots, n_N and associate these with particles $\{\mathbf{h}_t^{(n_1)}, \dots, \mathbf{h}_t^{(n_N)}\}$.

Step 2. For each values of n_i from Step 1, sample the values $\{\mathbf{h}_{t+1}^{*(1)}, \dots, \mathbf{h}_{t+1}^{*(N)}\}$ from

$$h_{k,t+1}^{*(i)} = \mu_k + \phi_k(h_{k,t}^{(n_i)} - \mu_k) + \sqrt{\sigma_k} u_{k,t+1}, \quad u_{k,t+1} \sim \mathcal{N}(0, 1), k = 1, \dots, q, i = 1, \dots, N. \quad (4.39)$$

Step 3. Calculate the weights of the values $\{\mathbf{h}_{t+1}^{*(1)}, \dots, \mathbf{h}_{t+1}^{*(N)}\}$ as

$$\pi_i^* = \frac{p(\mathbf{y}_{t+1} | \mathbf{h}_{t+1}^{*(i)}, \Theta)}{p(\mathbf{y}_{t+1} | \hat{\mathbf{h}}_{t+1}^{*(i)}, \Theta)}, i = 1, \dots, N, \quad (4.40)$$

and resample the values $\{\mathbf{h}_{t+1}^{*(1)}, \dots, \mathbf{h}_{t+1}^{*(N)}\}$ N times with replacement using these weights to obtain a fair sample $\{\mathbf{h}_{t+1}^{(1)}, \dots, \mathbf{h}_{t+1}^{(N)}\}$ from the filter distribution of $(\mathbf{h}_{t+1} | \mathbf{y}_{1:t+1}, \Theta)$.

Figures

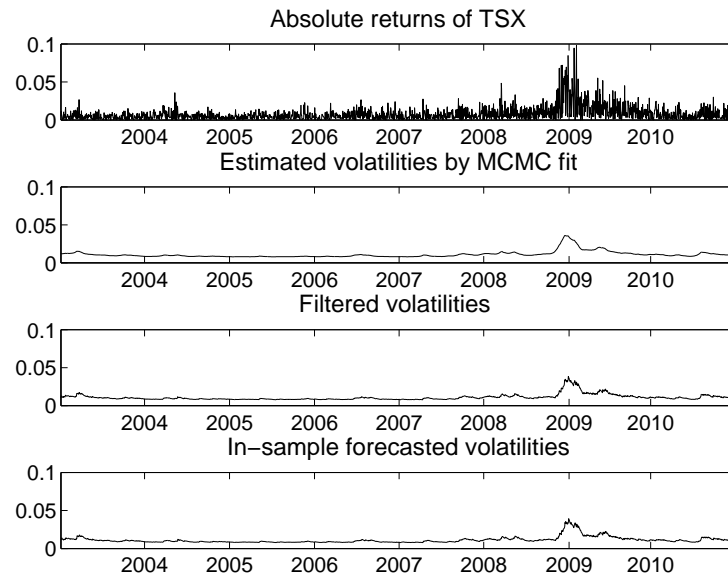


Figure 4.11: Comparison of volatilities explained by the first three dominant factors for the returns of TSX from the PPCAF model with idiosyncratic errors.

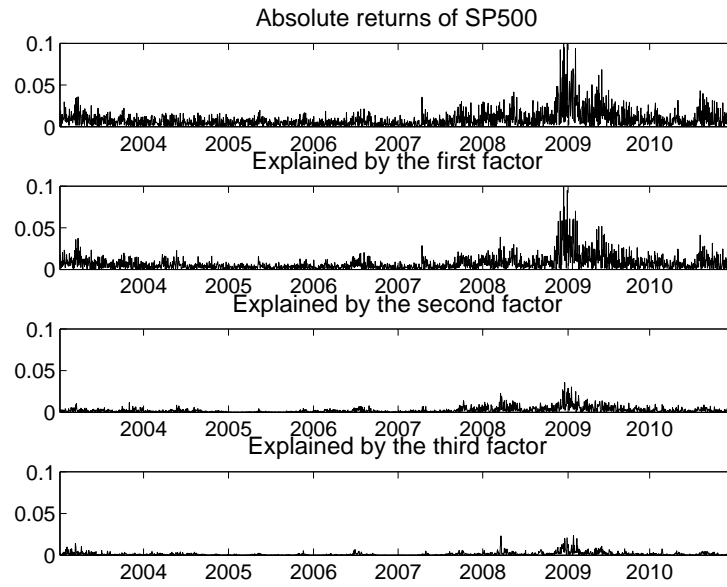


Figure 4.12: Comparison of volatilities explained by the first three dominant factors for the returns of S&P500 from the PPCAF model with idiosyncratic errors.

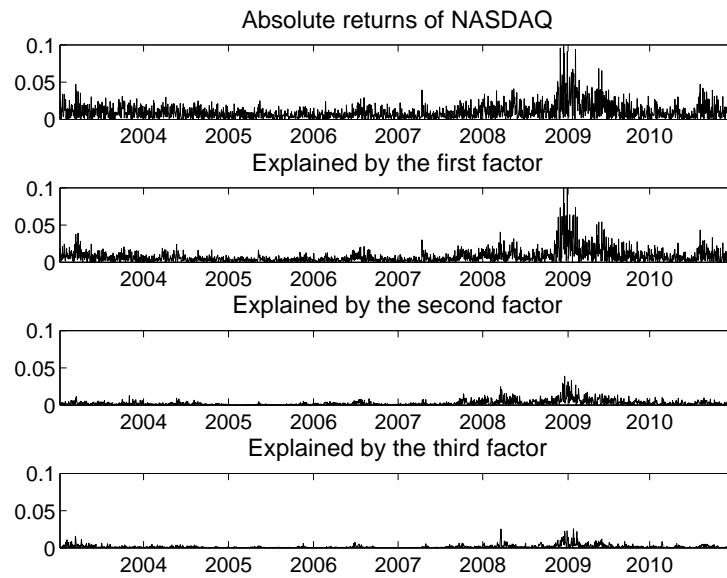


Figure 4.13: Comparison of volatilities explained by the first three dominant factors for the returns of NASDAQ from the PPCAF model with idiosyncratic errors.

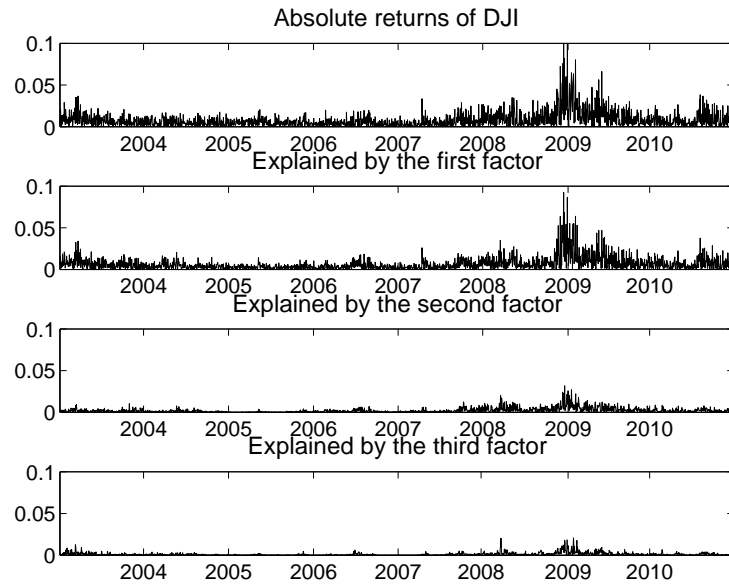


Figure 4.14: Comparison of volatilities explained by the first three dominant factors for the returns of DJI from the PPCAF model with idiosyncratic errors.

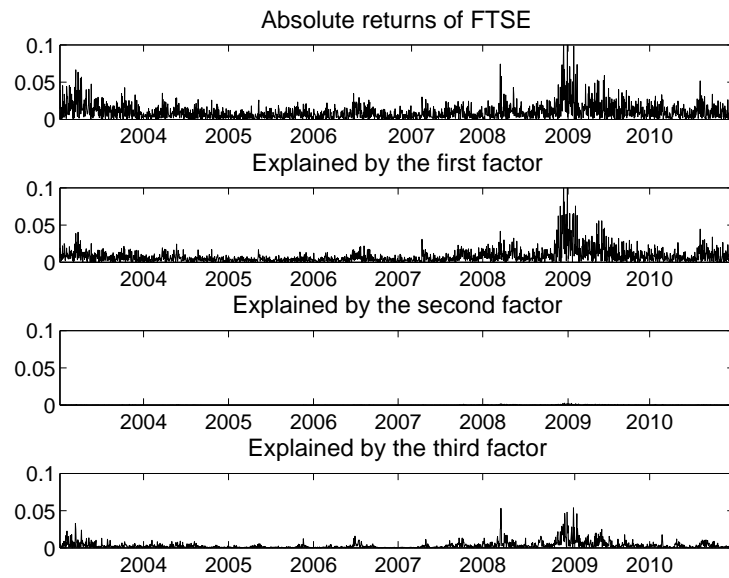


Figure 4.15: Comparison of volatilities explained by the first three dominant factors for the return of FTSE from the PPCAF model with idiosyncratic errors.

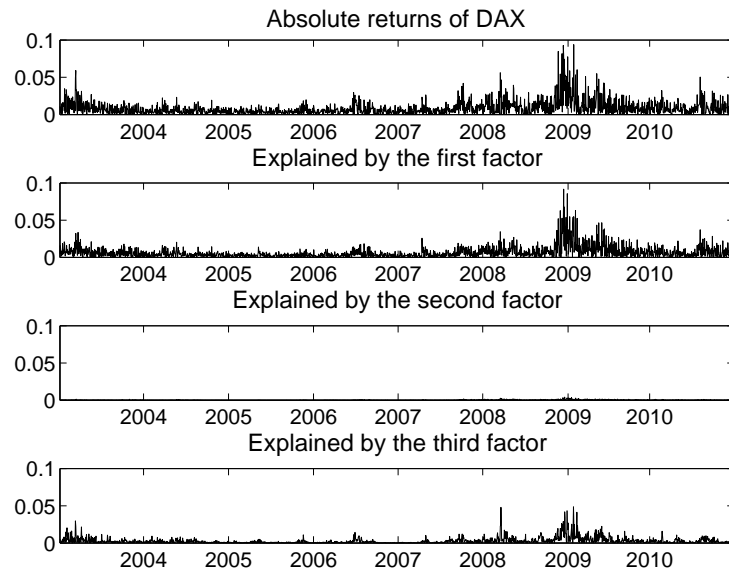


Figure 4.16: Comparison of volatilities explained by the first three dominant factors for the returns of DAX from the PPCAF model with idiosyncratic errors.

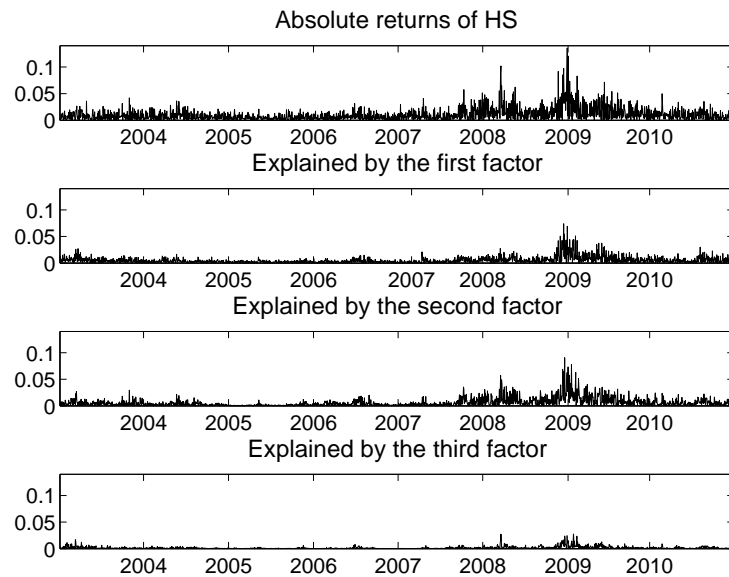


Figure 4.17: Comparison of volatilities explained by the first three dominant factors for the returns of HS from the PPCAF model with idiosyncratic errors.

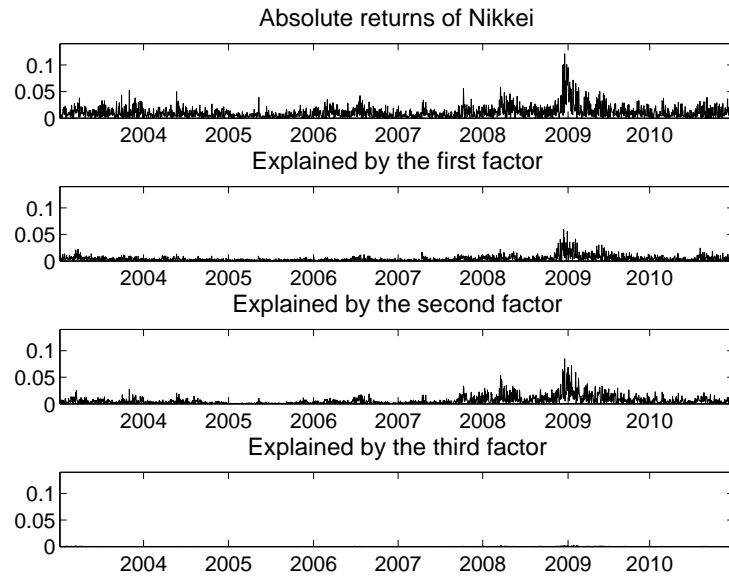


Figure 4.18: Comparison of volatilities explained by the first three dominant factors for the returns of Nikkei from the PPCAF model with idiosyncratic errors.

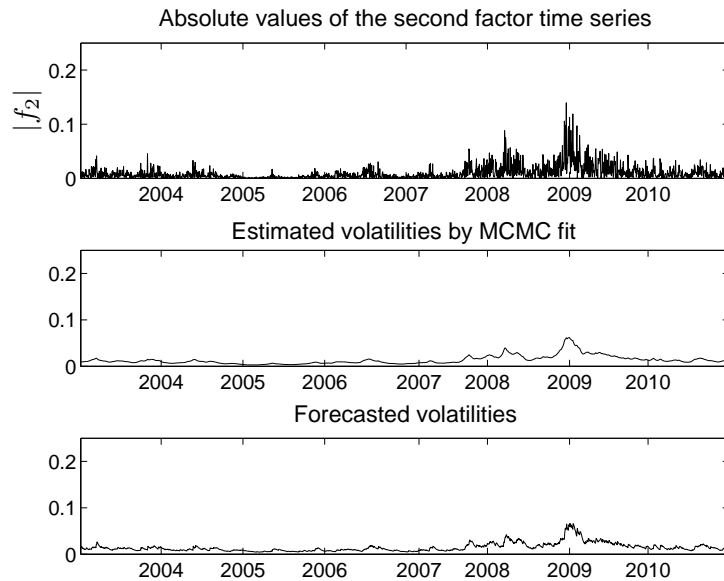


Figure 4.19: Comparison of the estimated second factor time series with the corresponding MCMC estimated and forecasts volatilities based on the international return data set.

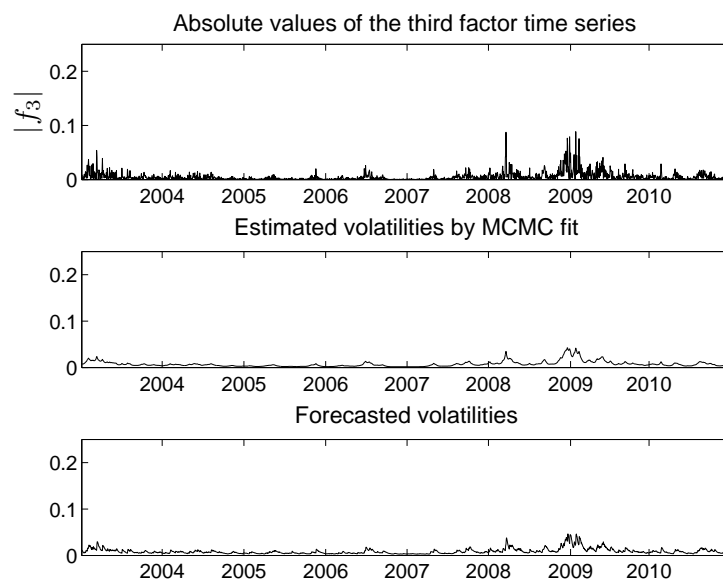


Figure 4.20: Comparison of the estimated third factor time series with the corresponding MCMC estimated and forecasts volatilities based on the international return data set.

Chapter 5

Avenues for Future Work

We first summarize what has been accomplished in this thesis. Then we discuss various possible extensions of our proposed models and methods.

5.1 Summary of Contributions

In this thesis, we have studied univariate and multivariate SV models, by focussing on the estimation of parameters and log volatilities based on different distributional assumptions of the observation innovations, and the leverage specification between asset returns and log volatilities. A novel factor SV model was proposed to extract market factors and characterize the time-varying correlation.

In Chapter 2, the slice sampler within the MCMC algorithms were developed for various univariate SV models. The innovation of measurement equation has either univariate normal or Student- t distribution, and the correlation is permitted between the observation equation and the latent process. If a Student- t distribution is introduced, it does not need a mixture decomposition. For the SV models without heavy tails, we used a Metropolis-Hastings method to sample the latent states, where the proposal distribution is either a univariate normal distribution that can be sampled

directly or an unknown distribution that can be simulated by the slice sampler. In the proposed Metropolis-Hastings simulation methods for the latent states of heavy-tailed SV models, the proposal distributions were simulated through the slice sampler, so that the popular Metropolis-Hastings algorithms used in the literature for SV models were avoided. To assess the goodness-of-fit, in addition to the test of the normality assumption of the measurement equation, we examined the probability integral transforms produced by the estimated model.

In Chapter 3, a multivariate SV model with more general cross correlation was considered. One main difficulty in fitting this model under the Bayesian framework is how to sample the variance-covariance matrix. Our solution to this problem is a Metropolis-Hastings method whose proposal was given by an inverse Wishart distribution. A slice sampler within the MCMC algorithm was then proposed to estimate the augmented parameters and the log volatilities of the model.

In Chapter 4, we proposed a novel factor SV (FSV) model within the framework of PCA and PPCA. The new model is motivated by the factor SV models and is combined with the PPCA. The main purpose of this model is to determine how many factors can actually affect the behavior of asset returns. Although our study focussed only on the cases under either the isotropic or idiosyncratic observation errors, our empirical application provides some indication that the idiosyncratic FSV model is of practical relevance.

5.2 Estimation of More General Univariate SV Models

The goal of the univariate SV model includes modeling the time-varying volatility of asset returns and the correlation between the two random processes. In the literature, the heavy tail property is modeled by introducing the Student- t distribution to the asset returns equation. This assumption may not adequately fit some daily return

time series. Accordingly, we can impose alternative distributions such as the asymmetric Laplace distribution or the normal inverse Gaussian (NIG) distribution, where the properties of the latter are detailed in McLeish (2005). Mathematical correlated errors of these distributions may be used. MCMC methods may be modified and applied to these more general models.

An alternative approach for modeling correlation between asset returns and latent volatilities is to introduce copulas to the innovations. The advantage of doing this is that the correlation explained by copulas is independent of the marginal distributions. Thus copula specification gives us more freedom to look at the correlations between the two innovations of the univariate SV models.

5.3 Multivariate SV Models Under the General PPCA Framework

Recently, FSV models have been proposed to analyze latent market factors that affect the dynamics of asset returns. Under the FSV model, the time-varying correlations of the asset returns can be captured by the dynamics of factors. Our PPCAF model is parsimoniously specified and yet flexible enough to capture many empirical regularities. One extension of our PPCAF model, analogous to the model in Han (2006), is to let the factors to be stochastic but follow univariate AR(1) processes. In other words, the factors are Markovian, whose shocks are assumed to have univariate SV processes. The Markov factor model formulation is suitable and has been used to study many financial time series, such as the dividend yield and Treasury-bill yield where the observations have strong serial dependence. After minor modifications, MCMC can be efficiently used to study the dynamics of these time series.

Bibliography

Abraham, B., Balakrishna, N. and Sivakumar, R. (2006). Gamma Stochastic Volatility Models. *Journal of Forecasting*, 25:153-171.

Abramowitz, M. and Stegun, N. C. (1970). Handbook of Mathematical Functions. New York: Dover Publications Inc.

Aguilar, O. and West, M. (2000). Bayesian dynamic factor models and portfolio allocation. *Journal of Business and Economic Statistics*, 18:338-357.

Ahn, J. H. and Oh, J. H. (2003). A Constrained EM algorithm for principal component analysis. *Neural Computation*, 15:57-65.

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, 52:317-332.

Alexander, B. (1994). Statistical factor analysis and related methods: Theory and applications. Wiley.

Andersen, T. W. and Rubin H. (1956). Statistical inference in factor analysis. *Proceedings of the third Berkeley symposium on mathematical statistics and probability*, 5:111-150.

Asai, M., and McAleer, M. (2005). Dynamic asymmetric leverage in stochastic volatility models. *Econometric Reviews*, 24:317-332.

Asai, M. and McAleer, M. (2006). Asymmetric multivariate stochastic volatility. *Econometric Reviews*, 25(2-3):453-473.

- Asai, M., McAleer, M. and Yu, J. (2006). Multivariate Stochastic Volatility: A Review. *Econometric Reviews*, 25(2-3):145-175.
- Barndorff-Nielsen, O. E. (1997). Normal inverse Gaussian distributions and stochastic volatility modeling. *Scandinavian Journal of Statistics*, 24(1):1-13.
- Barndorff-Nielsen, O. E. (1998). Process of normal inverse Gaussian type. *Finance Stochastics*. 2:41-68.
- Bauwens, L. and Lubrano, M. (1998). Bayesian inference on GARCH models using the Gibbs sampler. *Econometrics Journal*, 1:C23-C26.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31:307-327.
- Broto, c. and Ruiz, E. (2004). Estimation methods for stochastic volatility models: a survey. *Journal of Economic Surveys*, 18:613-649.
- Carnero, A., Pena, D. and Ruiz, E. (2003). Persistence and kurtosis in autoregressive stochastic volatility models. *Journal of financial econometrics*, 2:319-342.
- Chan, D., Kohn, R. and Kirby, C. (2006). Multivariate stochastic volatility models with correlated errors. *Econometric Reviews*, 25:245-274.
- Chib, S., Nardari, F. and Shephard, N. (2002). Markov chain Monte Carlo methods for generalized stochastic volatility models. *Journal of Econometrics*, 108:281-316.
- Chib, S., Nardarib, F. and Shephard, N. (2006). Analysis of high dimensional multivariate stochastic volatility models. *Journal of Econometrics*, 134:341-371.
- Chib, S., Omori, Y. and Asai, M. (2009). Multivariate stochastic volatility. *Handbook of Financial Time Series*, Springer-Verlag Berlin Heidelberg.

- Chib, S. and Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician*, 49(4):327-335.
- Clements, M. and Smith, J (2000). Evaluating the forecast densities of linear and non-linear models: applications to output growth and unemployment. *Journal of Forecasting*, 19:255-276.
- Danielsson, J. (1994). Stochastic volatility in asset prices: estimation with simulated maximum likelihood. *Journal of Econometrics*, 64:375-400.
- Danielsson, J. (1998). Multivariate stochastic volatility models: Estimation and a comparison with VGARCH models. *Journal of Empirical Finance*, 5:155-173.
- Danielsson, J. and Richard, J.F. (1993). Accelerated Gaussian importance sampler with application to dynamic latent variable models. *Journal of Applied Econometrics*, 8:153-173.
- Diebold, F. X., Guther, T. A. and Tay, A. S. (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review*, 39:863-883.
- Diebold, F. X., Guther, T. A. and Tay, A. S. (1999). Multivariate Density Forecast Evaluation and Calibration in Financial Risk Management: High-Frequency Returns on Foreign Exchange. *The Review of Economics and Statistics*, 81:661-673.
- Dobigeon N. and Tourneret J. (2010). Bayesian orthogonal component analysis for sparse representation. *IEEE Transactions on Signal Processing*, 58(5):2675-2685.
- Duffie, D. and Singleton, K. J. (1993). Simulated moments estimation of Markov models of asset prices. *Econometrica*, 61:929-952.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4):987-1007.

- Edwards, R. G. and Sokal, A. D. (1988). Generalization of the Fortuin-Kasteleyn-Swendsen-Wang representation and Monte Carlo algorithm. *Phys. Rev. D*, 38(6):2009-2012.
- Fridman, M. and Harris, L. (1998). A maximum likelihood approach for non-Gaussian stochastic volatility models. *Journal of Business & Economic Statistics*, 16(3): 284-291.
- Geweke, J. (1993). Bayesian treatment of the independent Student-t linear model. *Journal of Applied Econometrics*, 8:S19-S40.
- Geweke, J. and G. Zhou. (1996). Measuring the Pricing Error of the Arbitrage Pricing Theory, *The Review of Financial Studies*, 9(2):557-587.
- Gourieroux, C., Jasiak, J. and Sufana, R. (2004). The Wishart autoregressive process of multivariate stochastic volatility. Discussion paper: University of Toronto.
- Gourieroux, C. (2006). Continuous time Wishart process for stochastic risk. *Econometric Reviews*, 25:177-217.
- Han, Y. (2006). The economics value of volatility modelling: asset allocation with a high dimensional dynamic latent factor multivariate stochastic volatility model. *Review of Financial Studies*, 19:237-271.
- Hardle, W. and Simar, L (2007). Applied stochastic analysis. Springer-Verlag Berlin Heidelberg.
- Harvey, A. C., Ruiz, E. and Shephard, N. (1994). Multivariate stochastic variance models. *Review of Economic Studies*, 61:247-264.
- Harvey, A. C. and Shephard, N. (1996). Estimation of an asymmetric stochastic volatility model for asset returns. *Journal of Business & Economic Statistics*, 14:429-434.

- Hautsch, N. and Ou, Y. (2008). Discrete-time stochastic volatility models and MCMC-based statistical inference. Working paper, Humboldt-University zu Berlin, CASE, CFS, QPL.
- Hoff P. (2007). Model averaging and dimension selection for the singular value decomposition. *Journal of the American Statistical Association*, 102(478):674-685.
- Huang, S. and Yu, J. (2008). An efficient method for maximum likelihood estimation of a stochastic volatility models. *Statistics and Its Interface*, 1:289-296.
- Hull, J. and White, A. (1987). The pricing of options on assets with stochastic volatilities. *Journal of Finance*, 42(2):281-300.
- Jacquier, E., Polson, N. G. and Rossi, P. E. (1994). Bayesian analysis of stochastic volatility models (with discussion). *Journal of Business & Economic Statistics*, 12(4):371-417.
- Jacquier, E., Polson, N. G. and Rossi, P. E. (1999). Stochastic volatility: univariate and multivariate extensions. *CIRANO Working paper*, 99s-26, Montreal.
- Jacquier, E., Polson, N. G. and Rossi, P. E. (2004). Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *Journal of Econometrics*, 122:185-212.
- Jungbacker, B. and Koopman, S. J. (2006). Monte Carlo likelihood estimation for three multivariate stochastic volatility models. *Econometric Reviews*, 25:385-408.
- Kawakatsu, H. (2007). Numerical integration-based Gaussian mixture filters for maximum likelihood estimation of asymmetric stochastic volatility models. *Econometrics Journal*, 10:342-358.

- Kim, S., Shephard, N. and Chib, S. (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models. *Review of Economic Studies*, 65:361-393.
- Kitagawa, G. (1987). Non-Gaussian state-space modeling of nonstationary time series. *Journal of the American Statistical Association*, 82:1032-1063.
- Liesenfeld, R. and Richard, J. F. (2003). Univariate and Multivariate Stochastic Volatility Models: Estimation and Diagnostics. *Journal of Empirical Finance*, 10:505-531.
- Liesenfeld, R. and Richard, J. F. (2006). Classical and Bayesian analysis of univariate and multivariate stochastic volatility models. *Econometric Reviews*, 25:335-360.
- Liu, C. (2001). Discussion on "The Art of Data Augmentation" by van Dyk and Meng *Journal of Computational and Graphical Statistics*, 10:75-81.
- Liu, X. F. (2008). Parameter expansion for sampling a correlation matrix: an efficient GPX-RPMH algorithm. *Journal of Statistical Computation and Simulation*, 78:1065-1076.
- Liu, X. F. and Daniels, M. J. (2006). A new algorithm for simulating a correlation matrix based on parameter expansion and re-parameterization. *Journal of Computational and Graphical Statistics*, 15:897-914.
- Lopes, H. F. and Carvalho, C. M. (2007). Factor stochastic volatility with time varying loadings and Markov switching regimes. *Journal of Statistical Planning and Inference*, 137:3082-3091.
- Lopes H. and Polson N. (2010). Bayesian inference for stochastic volatility modeling, in Bocker, K. (Ed.) *Rethinking Risk Measurement and Reporting: Uncertainty, Bayesian Analysis and Expert Judgement*, 2010, 515-551.
- Lopes, H. F. and West, M. (2004). Bayesian model assessment in factor analysis. *Statistica Sinica*, 14:41-67.

McLeish D. L. (2005). Monte Carlo Methods and Finance. John Wiley & Sons, Inc.

Melino, A. and Turnbull, S. M. (1990). Pricing foreign currency options with stochastic volatility. *Journal of Econometrics*, 45:239-265.

Meyer, R. and Yu, J. (2000). BUGS for a Bayesian analysis of stochastic volatility models. *Econometrics Journal*, 3:198-215.

Mira, A. and Tierney, L. (2002). Efficiency and convergence properties of slice samplers. *Scandinavian Journal of Statistics*, 29:1:12.

Neal, R. N. (2003). Slice Sampling. *The Annals of Statistics*, 31(3): 705-767.

Omori, Y., Chib, S., Shephard, N. and Nakajima, J. (2007). Stochastic volatility with leverage Fast and efficient likelihood inference. *Journal of Econometrics*, 140:425-449

Philipov, A. and M. E. Glickman (2006a). Factor multivariate stochastic volatility via Wishart processes. *Econometric Reviews*, 25:311-334.

Philipov, A. and M. E. Glickman (2006b). Multivariate stochastic volatility via Wishart processes. *Journal of Business and Economic Statistics*, 24:313-328.

Pitt, M. K., D. Chan, and R. Kohn (2006). Efficient Bayesian inference for Gaussian copula regression models. *Biometrika*, 93:537-554.

Pitt, M., Shephard, N. (1999a). Time varying covariances: a factor stochastic volatility approach. In: Bernardo, J. M., Berger, J. O., David, A. P., Smith, A. F. M., eds. *Bayesian Statistics*, Volume 6, pp:547-570. Oxford University Press.

Pitt, M., Shephard, N. (1999b). Filtering via simulation: auxiliary particle filter. *Journal of the American Statistical Association*, 94:590-599.

- Press, S. J. (1985). *Applied Multivariate Analysis: Using Bayesian and Frequentist Methods of Inference*, California: Krieger.
Springer-Verlag Berlin Heidelberg 2003, 2007
- Robert C. P. and Casella G. (2004). *Monte Carlo Statistical Methods*. Springer Verlag, New York, second edition, 2001.
- Roberts G. O. and Rosenthal, J. S. (1999). Convergence of slice sampler Markov chains. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 61(3):643-660.
- Roweis, S. R. (1997). EM algorithms for PCA and SPCA. In M. Kearns, M. Jordan, & S. Solla (Eds.), *Advances in neural information processing systems*, 10:626-632. Cambridge, MA: MIT Press.
- Sandmann, G. and Koopman, S. J. (1998). Estimation of stochastic volatility models via Monte Carlo maximum likelihood. *Journal of Econometrics*, 87:271-301.
- Schwarz G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6(2):461-464.
- Shephard, N. G. (1993). Fitting non-linear time series models, with application to stochastic variance models. *Journal of Applied Econometrics*, 8:135-152.
- Shephard, N. and Pitt, M. K. (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika*, 84:653-667.
- Shimada, J. and Tsukuda, Y. (2005). Estimation of stochastic volatility models: an approximation to the nonlinear state space representation. *Communications in Statistics - Simulation and Computation*, 34:429-450.
- Skaug, H. and Yu, J. (2007). Automated likelihood based inference for stochastic volatility models, Working Paper, Singapore Management University.

- Smith, M. and Pitts, A. (2006). Foreign exchange intervention by the Bank of Japan: Bayesian analysis using a bivariate stochastic volatility model. *Econometric Reviews*, 25:425-451.
- So, M. K. P. and Kwok, W. Y. (2006). A multivariate long memory stochastic volatility model. *Physica A*, 362:450-464.
- So, M. K. P., Li, W. K. and Lam, K. (1997). Multivariate modelling of the autoregressive random variance process. *Journal of Time Series Analysis*, 18:429-446.
- Taylor, S. J. (1986). Modeling financial time series. Chichester, Uk:Wiley.
- Tipping, M. E. and Bishop, C. M. (1999). Probabilistic principal component analysis. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 61:611-622.
- Tsay, R. S. (2002). Analysis of Financial Time Series. Wiley.
- Tsay, R. S. (2005). Analysis of Financial Time Series: Financial Econometrics (2 ed.). New York: Wiley.
- Wong, F., Carter, C. K. and Kohn, R. (2003). Efficient estimation of covariance selection models. *Biometrika*, 90(4):809-830.
- Xu, D. (2006). Stochastic Volatility Model under Discrete Mixtures of Normal Specification. Department of Economics, University of Western Ontario, London, Canada. Working paper.
- Xu, D., Knight, J. L. and Wirjanto T. S. (2011). Asymmetric stochastic conditional duration model: a mixture-of-normal approach. *Journal of Econometrics*, vol 9, No. 3:469-488.
- Yu, J. (2005). On leverage in a stochastic volatility model. *Journal of Econometrics*, 127:165-178.

- Yu, J. and Meyer, R. (2006). Multivariate stochastic volatility models: Bayesian estimation and model comparison. *Econometric Reviews*, 51:2218-2231.
- Yu, J., Yang, Z. and Zhang, X. (2006). A class of nonlinear stochastic volatility models and its implications for pricing currency options. *Computational Statistics & Data Analysis*, 25:361-384.
- Zhang, X., Boscardin, W. J. and Belin, T. R. (2004). Sampling correlation matrices in Bayesian models with correlated latent variables. *Journal of Computational and Graphical Statistics* 15, 880-896.
- Zhang, X. and King, L. (2008). Box-Cox stochastic volatility models with heavy-tails and correlated errors. *Journal of Empirical Finance*, 15:549-566.