# การอนุมานแบบเบย์บนตัวแบบความผันผวนสโตแคสติกของตลาดหุ้น

นางสาวหลิงหลิง โหลว

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรดุษฎีบัณฑิต
สาขาวิชาคณิตศาสตร์ประยุกต์
มหาวิทยาลัยเทคโนโลยีสุรนารี
ปีการศึกษา 2559

# BAYESIAN INFERENCE ON STOCHASTIC VOLATILITY MODELS OF THE STOCK MARKET

**Lingling Luo**

**A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Applied Mathematics**

**Suranaree University of Technology**

**Academic Year 2016**

หลิงหลิง โหลว : การอนุมานแบบเบย์บนตัวแบบความผันผวนสโตแคสติกของตลาดหุ้น (BAYESIAN INFERENCE ON STOCHASTIC VOLATILITY MODELS OF THE STOCK MARKET) อาจารย์ที่ปรึกษา : ศาสตราจารย์ ดร.ไพโรจน์ สัตยธรรม, 95 หน้า.

วิทยานิพนธ์ฉบับนี้ได้เสนอตัวแบบชนิด GARCH E-MSSV E-MSSV-I และ E-MSSV-II เพื่อการประมาณค่าพารามิเตอร์และค่าความผันผวน

ในส่วนแรกของวิทยานิพนธ์ฉบับนี้ ได้มีการใช้ตัวแบบ GARCH EGARCH และ TGARCH พร้อมด้วยนวัตกรรมแบบปกติและแบบสติวเด็นท์ เพื่อการวิเคราะห์ข้อมูลอนุกรมเวลาหยุดนิ่งในตลาดหุ้น ดังเช่น ดัชนี SSE380 ด้วยการพิจารณาจากค่าสูงสุดของ ลอก ไลลี่ฮูด และค่าต่ำสุดของ AIC และ BIC ผลจากการทดลองพบว่า ตัวแบบ GARCH พร้อมด้วยนวัตกรรมแบบสติวเด็นท์จะเป็นตัวแบบที่ดีที่สุด การใช้การจำลองแบบบูทสแตรปแสดงให้เห็นว่าการทดสอบความเชื่อมั่นของตัวแบบซึ่งเป็นตัวแบบ GARCH พร้อมด้วยนวัตกรรมสติวเด็นท์ เป็นตัวแบบที่ดีกว่าตัวแบบอื่น

ส่วนที่สองของวิทยานิพนธ์นี้ศึกษาตัวแบบ E-MSSV ซึ่งเป็นตัวแบบที่เกิดจากการบวกปริมาตรเข้าไปในตัวแบบ MSSV ได้นำตัวแบบนี้มาวิเคราะห์ข้อมูลอนุกรมเวลาไม่นิ่งในตลาดหุ้น ดังเช่น ดัชนีดาวน์โจนส์ (DJI30) ต่อมาได้มีการนำเสนอตัวแบบ E-MSSV-I ซึ่งใช้ประโยชน์จากตัวแบบกราฟระบุทิศทาง และการอนุมานแบบเบย์ เพื่อสร้างสูตรการทำนายการกรอง และการทำฟังก์ชันความน่าจะเป็นให้ราบเรียบ ต่อจากนั้นจะให้วิธีการ ค่าคาดหมาย-สูงสุด ในการประมาณค่าตัวแปรและพารามิเตอร์ ตัวแบบนี้จะประยุกต์ใช้กับตัวแปรสุ่มไม่ต่อเนื่องเท่านั้น

ได้มีการนำเสนอตัวแบบอีกชนิดหนึ่งคือ E-MSSV-II เพื่อใช้กับตัวแปรสุ่มต่อเนื่อง ซึ่งสถานะระบบถูกควบคุมด้วยกระบวนการมาร์คอฟอันดับหนึ่ง ได้มีการใช้ตัวกรองมอนติคาร์โลอันดับ ในการคำนวณค่าพารามิเตอร์และตัวแปรซ่อน วิธีการนี้ได้ถูกทดสอบด้วยอนุกรมเวลาสังเคราะห์ และปรับปรุงด้วยอนุกรมเวลาที่มาจากค่าจริง ดังเช่น DJI30 ผลลัพธ์ที่เกิดขึ้นแสดงให้เห็นว่า ตัวแบบ E-MSSV-II ให้การพยากรณ์ที่แม่นตรงมากกว่าตัวแบบ MSSV ด้วยมาตรวัด MAD MSE และ MPAE

สาขาวิชาคณิตศาสตร์ ลายมือชื่อนักศึกษา Ling Ling Luo

ปีการศึกษา 2559 ลายมือชื่ออาจารย์ที่ปรึกษา

LINGLING LUO : BAYESIAN INFERENCE ON STOCHASTIC VOLATILITY MODELS OF THE STOCK MARKET. THESIS ADVISOR : PROF. PAIROTE SATAYATHAM, Ph.D. 95PP.

VOLATILITY/ GARCH-TYPE MODEL/ MCS TEST/ BAYESIAN INFERENCE/ E-MSSV MODEL/ AUXILIARY PARTICLE FILTER/ EM ALGORITHM

In this thesis, GARCH-type models and the Extended Markov Regime Switching Stochastic Volatility Model, called the E-MSSV model including E-MSSV-I and E-MSSV-II models, are presented to focus on the estimation of parameters and volatilities.

The first part performs GARCH, EGARCH and TGARCH models with Normal innovation and Student's t innovation to analyze stationary time series data, namely the SSE380 index. According to the highest value of Log likelihood, the smallest value of AIC, BIC, the experimental results show that GARCH with Student's t innovation model is the best model. Conducting a bootstrap simulation study shows that the Model Confidence Set test also captures the superior model, which is GARCH with Student's t innovation.

The second part presents a novel approach, called E-MSSV model, based on adding volume to the MSSV model to analyze a non-stationary time series, namely the Dow Jones Industrial Average (DJI30). The E-MSSV-I model which focuses on discrete random variables is proposed by employing advanced probabilistic modeling

methodology called "Directed Graphical Model". Bayesian inference is then used to derive prediction, filtering and smoothing probability distribution function. Then the Expectation-Maximization method is presented to estimate the variables and parameters. This model can only be applied to discrete random variables. Thus, the E-MSSV-II model is introduced to analyze continuous random variables when the regime state is governed by a first-order Markov process. Then the Sequential Monte Carlo filter is presented to evaluate parameters and latent variables. The methodology is tested with a synthetic time series and validated with a real financial time series, namely the DJI30. The results show that the E-MSSV-II model is more accurate at forecasting than the MSSV model, as measured by the MAD, MSE and MAPE loss functions.

School of Mathematics Student's Signature Ling Ling Luo

Academic Year 2016 Advisor's Signature P. Sattayatham

# ACKNOWLEDGEMENTS

I would like to express my deepest appreciation to my advisor, Professor Pairote Sattayatham, for his unyielding patience, numerous discussions and relevant comments.

I am extremely grateful to Doctor Ratthachat Chatpatanasiri for his expert assistance and encouragement in carrying out my thesis work. I would also like to thank the Head of the School of Mathematics, Asst. Prof. Dr. Eckart Schulz, for his professional course guidance.

I am indebted to my husband, Mr. Zhuo Lin, without whose understanding, love, and patience I would not have been able to finish this thesis. I thank all my friends who have contributed by discussing our work and knowledge. Furthermore, I would also like to thank all the staff in the School of Mathematics for providing help. I express my deepest and sincerest thanks to Guizhou University of Finance and Economics for providing the opportunity to further my education.

Lingling Luo

# CONTENTS

**Page**

# CONTENTS (Continued)

**Page**

# CONTENTS (Continued)

**Page**

# CONTENTS (Continued)

**Page**

# LIST OF FIGURES

**Figure** **Page**

# LIST OF TABLES

# LIST OF ABBREVIATIONS

| | |
|---|---|
| ARCH | Autoregressive Conditional Heteroskedasticity |
| GARCH | Generalized Autoregressive Conditional Heteroskedasticity |
| EGARCH | Exponential Generalized Autoregressive Conditional Heteroskedasticity |
| GARCH-N | GARCH model with Normal innovation |
| EGARCH-N | EGARCH model with Normal innovation |
| TGARCH-T | TGARCH model with Student's *t* innovation |
| MCS | Model Confidence Set |
| SSE380 | Shanghai Stock Exchange 380 |
| AIC | Akaike Information Criterion |
| BIC | Bayesian Information Criterion |
| LL | Log Likelihood |
| MSE | Mean Square Error |
| MAD | Mean Absolute Deviation |
| MAPE | Mean Absolute Percent Error |
| SV | Stochastic Volatility |
| MSSV | Markov Regime Switching Stochastic Volatility |
| DAG | Directed Acyclic Graph |
| DGM | Directed Graphical Model |
| ML | Maximum Likelihood |

# LIST OF ABBREVIATIONS (Continued)

| | |
|---|---|
| EM | Expectation Maximization |
| pdf | probability distribution function |
| SMC | Sequential Monte Carlo |
| DJI30 | Dow Jones Industrial Average |

# CHAPTER I

# INTRODUCTION

Volatility, the most extensively used measure of uncertainty, is fundamental in modern finance theory. Quantifying and forecasting volatility are essential to financial asset pricing, portfolio optimization and risk management. Many models of volatility are applied in forecasting stock market movement and evaluating the performance of the stock market. Ability to predict volatility accurately is a crucial job for stock market researchers and practitioners.

In finance, the term volatility is used to denote a measure of the variation of a particular asset. Mathematically it is often defined as standard deviation of asset return (Ramey, 1995; Huang, 2005). Many studies (Day and Lewis, 1992; Liu et al., 1999; Jondeau and Rockinger, 2003) have shown that volatility has wealth statistical properties such as clustering, persistence, long memory and so on. Mandelbrot (1967) and Fama (1965) find that volatility exists in clusters, that is, large changes tend to be followed by large changes and small changes tend to be followed by small changes. Baillie et al. (1996) analyze US Dollar foreign exchange rates and get that volatility possesses long memory and persistence, i.e., there is a long lag linear autocorrelation.

There are two popular classes of widely used models for volatility of the observed time series for capturing the stylized features of volatilities in financial data. One is the class of Autoregressive Conditional Heteroskedasticity (ARCH) models (Engle, 1982)

and their various extensions (Bollerslev, 1986), formulating the serial dependence of volatility and incorporating the past observations into the future volatility. Another one is the class of stochastic volatility (SV) models which have also been well studied in financial econometrics.

After introduction of ARCH and GARCH models (Bollerslev, 1986), many researchers have proposed extensions and alternative specifications on the models such as Exponential GARCH (Nelson, 1991), GARCH-M (Hamilton, 1994), Threshold GARCH (Glosten et al., 1993). Because of the increasingly important demand to explain and to model risk and uncertainty in financial time series, GARCH-type models have been the main tool for volatility forecasting.

In earlier research, Engle (2001) points that the analysis of ARCH and GARCH models and their many extensions provide a statistical stage on which many theories of asset pricing and portfolio analysis can be exhibited and tested. Basel (2005) examines the relative out of sample predictive ability of different GARCH models, with particular emphasis on the predictive content of the asymmetric component. Cathy et al. (2006) introduce a four-regime Double Threshold GARCH (DTGARCH) model, which allows asymmetry in both the conditional mean and variance equations simultaneously by employing two threshold variables, to analyze the stock markets' reactions to different types of information (good/bad news) generated from the domestic markets and the US stock market. Hung (2011) use a fuzzy system method to analyze clustering in GARCH models. Liu and Chiang (2012) employ four GARCH-type models, incorporating the skewed generalized t errors into log returns of Standard and Poor's Depositary Receipts

exhibiting fat-tails, leptokurtosis and skewness to forecast both volatility and value-at-risk. Li et al. (2013) propose the Mixture Memory GARCH volatility model which involves a short memory GARCH and a long memory FIGARCH, using the daily S&P 500 index to illustrate volatility's capabilities.

More recently, Werner et al. (2014) perform a hybrid Neural Networks-GARCH model for volatility forecast in three Latin-American stock exchange indexes from Brazil, Chile and Mexico, and demonstrated that the Artificial Neural Networks models can improve the forecasting performance of the GARCH models when studied in the three Latin-American markets. Peter et al. (2014) introduce a multivariate GARCH model that incorporates realized measures of variances and covariance. Realized measures extract information about the current levels of volatilities and correlations from high-frequency data, which is particularly useful for modeling financial returns during periods of rapid changes in the underlying covariance structure. Mutunga et al. (2015) implement an estimating functions approach combining with the first order EGARCH and GJR-GARCH models to forecast the volatility of two market indices from the USA and Japanese stock markets.

Stochastic Volatility (SV) models which were first introduced by Taylor (1986) concentrate on the time-varying and persistent volatility, as well as on the leptokurtosis in financial return series. Many extensions to the basic SV models have been proposed in the literature (Heston, 1993; Sadorsky, 2005; Vo, 2009). In particular, the Markov Switching Stochastic Volatility models (MSSV) were studied in Diebold (1986) and Mike and So (1998).

In earlier research, Smith (2002) presents the MSSV diffusion model to analyze the short rate volatility and proposes quasi-maximum likelihood estimation techniques to calculate the volatility parameters. Shibata and Watanabe (2005) use the MSSV model to accommodate the shift in the mean of log volatility of the TOPIX index and use the Bayesian Markov Chain Monte Carlo (MCMC) approach to estimate the parameters in the model, which provide evidence that the MSSV model is favored over the standard SV. Carvalho and Hopes (2007) propose a simulation-based algorithm for inference in stochastic volatility models with possible regime switching and develop auxiliary particle filters strategy to sequentially learn about states and parameters of the model in the IBOVESPA stock index. Valle et al. (2010) introduce the modified mixture model with Markov switching volatility specification to analyze the relationship between British Petroleum stock return volatility and trading volume and construct an algorithm based on Markov Chain Monte Carlo simulation methods to estimate all the parameters in the model using a Bayesian approach. Du et al. (2011) assess factors that potentially influence the volatility of crude oil prices and the possible linkage between this volatility and agricultural commodity markets. They apply stochastic volatility models to weekly crude oil, corn, and wheat futures prices from November 1998 to January 2009, and then use Bayesian MCMC methods to estimate the model parameters. Rios and Lopes (2013) explore kernel smoothing and conditional sufficient statistics extensions of the auxiliary particle (Pitt and Shephard, 1999) and bootstrap filters (Gordon et al., 1993) and use simulated data following MSSV models. They show that the LW particle filter degenerates and has the largest Monte Carlo error, while the auxiliary particle filter is

better than it.

More recently, Kastner and Schnatter (2014) represent Bayesian inference for stochastic volatility models. An MCMC method which depends on actual parameter values in terms of sampling efficiency is used to evaluate volatility parameters. The volatility in the latent state equation is small which shows deficiencies for highly persistent latent variable series. Clark and Ravazzolo (2015) compare alternative models of time-varying volatility on the basis of the accuracy of real-time point and density forecasts of key macroeconomic time series for the USA stock. The results show that the AR and VAR specifications with conventional stochastic volatility dominate other volatility specifications, in terms of point forecasting to some degree. Bonfil et al. (2015) present a new method called Support Vector Regression Boltzmann selection for the financial volatility forecasting problem which selects simultaneously the proper kernel and its parameter values. Joshua and Angelia (2016) compare a number of GARCH and SV models using nine series of oil, petroleum product and natural gas prices. The competing models include the standard models of GARCH(1,1) and SV with an AR(1) log-volatility process, as well as more flexible models with jumps, volatility in mean, leverage effects, and *t* distributed and moving average innovations. Using the marginal likelihood, the result shows that the SV model with moving average innovations is the best model for all nine series.

In this thesis, we mainly focus on two parts. Firstly, the Model Confidence Set (MSC) test is introduced to describe the best GARCH-type model which is used to estimate the volatilities. Secondly, Expectation Maximization and Sequential Monte

Carlo (SMC) filter which is based on the Extended Markov Regime Switching Stochastic Volatility Model are proposed to evaluate the parameters.

The remaining parts are organized as follows: In Chapter II, the mathematics preliminaries of GARCH models and Bayesian statistics are introduced. In Chapter III, we seek to identify the superior model in capturing the characteristics of the SSE380 index and use symmetric GARCH and asymmetric GARCH (EGARCH and TGARCH) with normal innovation and student's t innovation models to forecast volatility. Then we use the MCS test based on the bootstrap simulation to choose the best model. In Chapter IV, two novel models, i.e., E-MSSV-I model and E-MSSV-II model, are built to estimate parameters and volatilities. The E-MSSV-I model is proposed to consider the discrete random variables by employing advanced probabilistic modeling methodology called "Directed Graphical model", and then using Bayesian inference to derive filtering, smoothing distribution function and Expectation-Maximization method to jointly estimate the variables and parameters. The E-MSSV-II model is studied to analyze a non-stationary time series, and then a SMC filter is presented to evaluate parameters and latent variables. In Chapter V, the conclusion is presented.

The thesis employs the following symbols: small letters denote random variables or parameter indices; capital letters denote sets or constants; $a_j$ denotes the $j^{th}$ observation; $y_1^t$ denotes the set of $\{y_1,...,y_t\}$; $\{y\}_{-i}$ denotes the set of $\{y\}$ except the $i^{th}$ element; $p(x)$ denotes the probability distribution function of random variable $x$; $p(x,y)$ denotes the joint probability distribution function of random variable $x$ and $y$; $p(x|y)$ denotes the conditional probability distribution of $x$ given $y$; $p(x|y,z)$ denotes the

conditional probability distribution of $x$ given $y$ and $z$, and $\theta$ denotes the set of parameters.

# CHAPTER II

# MATHEMATIC PRELIMINARIES

## 2.1 Autoregressive conditional heteroskedasticity volatility models

While conventional time series and econometric models operate under an assumption of constant variance, the autoregressive conditional heteroskedasticity (ARCH) process introduced in Engle (1982) allows the conditional variance to change over time as a function of past errors leaving the unconditional variance constant. In fact, ARCH models are discrete time models which structure one step ahead forecasting and carry out n-step ahead prediction.

According to Brooks (2002), the following sub-sections define and describe several important concepts in time series analysis. In order to better understand those concepts, we define $\{\varepsilon_t\}$ as a set of random variables, and $\{h_t\}$ as a sequence of variables.

A white noise process: Roughly speaking, a white noise process is one with no discernible structure. A definition of a white noise process is $E(\varepsilon_t) = 0$, $\text{var}(\varepsilon_t) = \sigma^2$. Thus, a white noise process has constant mean and variance, and zero autocovariance, except at lag zero.

If it is assumed that $\varepsilon_t$ is distributed normally, then the sample autocorrelation coefficients are approximately normally distributed.

This result can be used to conduct significance tests for the autocorrelation coefficients by constructing a non-rejection region (like a confidence interval) to

determine whether it is significantly different from zero. For example, a 95% non-rejection region would be given as $\pm 1.96 \times \frac{1}{\sqrt{T}}$, where *T* is the sample size.

Moving average processes: The simplest class of time series model that one could entertain is that of the moving average process. Let $\varepsilon_t$ be a white noise process with $E(\varepsilon_t) = 0$, variance, i.e. $\text{var}(\varepsilon_t) = \sigma^2$ and $\omega$ be constant parameters. Then

$$h_t = \omega + \varepsilon_t + \alpha_1 \varepsilon_{t-1} + ... \alpha_q \varepsilon_{t-q}$$

is a q-th order moving average mode, denoted MA(q). This can be expressed using sigma notation as

$$h_t = \omega + \sum_{j=1}^{q} \alpha_j \varepsilon_{t-j} .$$

A moving average model is simply a linear combination of white noise processes, so that $\sigma_t$ depends on the current and previous values of a white noise disturbance term.

Autoregressive processes**:** An autoregressive model is one where the current value of a variable, $h_t$, depends upon only the values that the variable has taken in previous periods plus an error term. An autoregressive model of order p, denoted as AR(p), can be expressed as

$$h_t = \omega + \beta_1 h_{t-1} + \beta_2 h_{t-2} + ... + \beta_p h_{t-p} + \varepsilon_t ,$$

where $\varepsilon_t$ is a white noise disturbance term, and $\omega$ a constant parameter. A manipulation of the expression will be required to demonstrate the properties of an autoregressive model. Hence the expression can be written more compactly using sigma notation

$$h_t = \omega + \sum_{i=1}^{p} \beta_i h_{t-i} + \varepsilon_t .$$

The autoregressive conditional heteroskedasticity (ARCH) model was introduced by Engle (1982). The ARCH model can model the conditional variance $h_t$ as a function of the lagged $\varepsilon's$. That is, the predictable volatility is dependent on past news. A more detailed model is the p-th order ARCH model, i.e. ARCH(q), which is presented as

$$h_t^2 = \omega + \sum_{j=1}^{q} \alpha_j \varepsilon_{t-j}^2 ,$$

where $\alpha_j$ and $\omega$ are constant parameters. The effect of a return shock $i$ periods ago $(i \le q)$ on current volatility is governed by the parameter $\alpha_i$. That is, in the ARCH(q) model, old news which arrived at the market more than *p* periods ago have no effect at all on current volatility.

A more generalized ARCH model was developed by Bollerslev (1986) for modelling conditional variance. It is denoted GARCH(p,q) where p is the ARCH term specifying the number of autoregressive lags and q is the GARCH term specifying the number of moving average lags.

The GARCH(p, q) model is given by

$$h_t^2 = \omega + \sum_{j=1}^{q} \alpha_j \varepsilon_{t-j}^{\ 2} + \sum_{i=1}^{p} \beta_i h_{t-i}^{\ 2}, \qquad (2.1)$$

with the following restrictions $\omega > 0$, $\alpha_j > 0$, $\beta_i > 0$ and $\sum_{j=1}^{q} \alpha_j + \sum_{i=1}^{p} \beta_i < 1$.

Despite the apparent success of these simple parameterizations, the ARCH and GARCH models cannot capture some important features of the data. The most interesting feature not addressed by these models is the leverage or asymmetric effect discovered by French et al. (1987) and Nelson (1991). One method proposed to capture

such asymmetric effects in Nelson (1991) is the exponential GARCH (EGARCH) model.

The EGARCH is obtained as

$$\log h_t^2 = \omega + \sum_{i=1}^{p} \beta_i \log h_{t-1}^2 + \sum_{j=1}^{q} \alpha_j \left[ \frac{|\varepsilon_{t-j}|}{h_{t-j}} - E\left(\frac{|\varepsilon_{t-j}|}{h_{t-j}}\right) \right] + \sum_{j=1}^{q} \xi_j \left(\frac{\varepsilon_{t-j}}{h_{t-j}}\right). \qquad (2.2)$$

Here the parameter $\alpha_j$ captures the volatility clustering effect and the $\xi_j$ measures the leverage effect. The conditional variance is in logarithmic form, which implies that the model has the following features: Firstly, $h_t^2$ will always be positive regardless of the sign of the parameters, therefore no constraints of non-negativity are needed. Secondly, the asymmetrical effect is not quadratic but exponential, if $\xi_j < 0$, it indicates a leverage effect. The EGARCH model allows good news and bad news to have different impacts on volatility because the level of $\varepsilon_{t-j}/h_{t-j}$ is included with a coefficient $\xi_j$.

The TGARCH model of Glosten et al. (1993) allow for asymmetric effects by augmenting a dichotomous dummy variable into the standard GARCH model. The parameterization of Threshold GARCH (TGARCH) model is obtained as

$$h_t^2 = \omega + \sum_{i=1}^{p} \beta_i h_{t-i}^2 + \sum_{j=1}^{q} \alpha_j \varepsilon_{t-j}^2 + \sum_{j=1}^{q} \xi_i \varepsilon_{t-j}^2 \mathrm{I}_{t-j}, \qquad (2.3)$$

$$\text{where } \mathrm{I}_{t-j} = \begin{cases} 1 & if\ \varepsilon_{t-j} < 0 \\ 0 & if\ \varepsilon_{t-j} \geq 0 \end{cases}.$$

The dummy variable $\mathrm{I}(\cdot)$ stands for the indicator function. If $\varepsilon_{t-j} < 0$, a negative surprise implies the coefficient of $\varepsilon_{t-j}^2$ is $\alpha_j + \xi_j$. Therefore, a leverage effect exists if $\xi_j > 0$. Similar to the standard GARCH model, $\omega > 0$, $\alpha_j, \beta_i \geq 0$ and $\alpha_j + \xi_j \geq 0$ are required.

## 2.2 Stochastic volatility models

Stochastic volatility (SV) models provide a natural alternative to the ARCH family models. The challenge of the SV models is that the volatility is not directly observable. Instead, it is driven by a different unobservable random process (Heston, 1993; Shephard, 2005).

The first SV model was proposed by Taylor (1986). The simplest formulation of the SV model is as follows:

$$y_t = \exp(h_t/2)\varepsilon_t$$

$$h_t = \mu + \phi(h_{t-1} - \mu) + \sigma_\eta \eta_t$$

with $h_0 \sim N(\mu, \sigma_\eta^2/1-\phi^2)$ , $\varepsilon_t$ and $\eta_t$ are independent and identically distributed standard normal random variables, $h_t$ is unobserved log volatility, $y_t$ is the log return of a stock at time $t$, defined as $y_t = \log(p_t/p_{t-1})$ with $p_t$ as the observed stock market price at time $t$. The parameters $\sigma_\eta$ can be thought of as the volatility of $h_t$. $\phi < 1$ is a parameter that measures the persistence of $h_t$, and $\mu$ is the mean of $h_t$.

Many extensions to the basic SV models have been proposed in the literature (Heston, 1993; Sadorsky, 2005; Shibata and Watanabe, 2005; Vo, 2009). In particular, the Markov Switching Stochastic Volatility model (MSSV) was studied in Diebold (1986), Mike and So (1998). The MSSV model is represented as

$$y_t = \exp(h_t/2)\varepsilon_t$$

$$h_t = \mu_{s_t} + \phi h_{t-1} + \sigma_\eta \eta_t \qquad (2.4)$$

$$\mu_{s_t} = \alpha + \beta s_t$$

$$P(s_t = i | s_{t-1} = j) = p_{ij}$$

where $y_t$ is the log return of a stock at time $t$, $h_t$ is unobserved log volatility, $\varepsilon_t \sim i.i.d.N(0,1)$ $\eta_t \sim i.i.d.N(0,1)$, $p_{ij} \geq 0$, $\alpha \in R$, persistence parameter $\phi < 1$, $\beta > 0$ and $\mu_{s_t}$ and $\sigma_\eta^2$ denote the mean and variance of $h_t$ respectively, $i, j \in N$. We note that equation (2.4) can be made more general to be $h_t = \mu_{s_t} + \sum_{i=1}^{t-1} \phi_i h_{t-i} + \sigma_\eta \eta_t$ to better capture data complexity. Nevertheless, many authors (Yu and Zhang, 2011; Pan and Li, 2013; Goutte, 2013) use equation (2.4) for simplicity.

## 2.3 Bayesian inference approaches

An advanced statistical model named "Directed Graphical Model" (DGM) will be employed to model volatility and related variables. Theorems on statistical inference of these variables and on estimation of model's parameters will be derived using Bayesian methods.

### 2.3.1 Elementary Bayesian probability and statistics

This sub-section is a very brief review of the basics of Bayesian probability and statistical theory. More details can be found in Gelman et al. (2003), Bishop (2006) and Murphy (2012). In the Bayesian viewpoint, unlike the conventional viewpoint, probability and statistics are treated as the same subject, i.e. the fundamentals of Bayesian probability and statistics are explained by product, sum and Bayesian rules.

In this thesis, we follow the notations and definitions from Gelman et al. (2003) which is a standard reference on Bayesian statistics. We note that a few notations are different from some textbooks. For example, we use the terms "distribution" and "density" interchangeably. Since the main contribution of this thesis is to develop a

forecasting system which is naturally discrete, we mainly focus on discrete random variables.

Product and Sum Rules: Suppose $x$ and $y$ are random variables, and $p(x, y)$ defines the joint probability distribution function (pdf) of $x$ and $y$, then

$$p(x, y) = p(x|y)p(y) = p(y|x)p(x)$$

is called the product rule. Given a joint probability on two random variables $p(x, y)$ and assuming $y$ is a discrete random variable, we define the marginal probability distribution function as follows:

$$p(x) = \sum_b p(x, y = b) = \sum_b p(x|y = b)p(y = b),$$

here we sum over all possible states $b$. We can define $p(y)$ similarly. This is sometimes called the sum rule. Supposing a set of random variables $\{x_1, ..., x_D\}$, then the product rule can be applied multiple times to yield the chain rule of probability:

$$p(x_1, ..., x_D) = p(x_1)p(x_2|x_1)p(x_3|x_2, x_1)...p(x_D|x_1, ..., x_{D-1}).$$

Bayesian rule: For discrete random variables, combining the definitions of conditional pdf with the product and sum rules yields Bayesian rule, also called Bayesian Theorem. According to the definition, the Bayesian rule of the conditional probability mass function of $x$ given $y$ can be written as:

$$p(x|y) = \frac{p(x, y)}{p(y)} = \frac{p(x)p(y|x)}{\sum_d p(x' = d)p(y|x' = d)}.$$

Philosophy of Bayesian inference: According to Beal (2003), Gelman et al. (2003) and Bishop (2006), a Bayesian approach starts with some prior knowledge or assumptions about the model structure. This initial knowledge is represented in the form of a prior probability distribution over model structures. In the light of observed data,

these are updated to obtain a posterior distribution over models and parameters.

Denote the data set by $Y$, which may be made up of several variables indexed by $t$: $Y = \{y_1, ..., y_t\}$. The Bayesian approach defines a generative model $p(Y|\theta)$ of the data $Y$ through a set of parameters $\theta$, where the data $Y$ is called observed data. Moreover, the prior distribution is defined by $p(\theta)$ which models the prior beliefs on the parameter and the probability posterior distribution is defined by $p(\theta|Y)$ which models the posterior beliefs on the data.

The philosophy of a Bayesian inference is to calculate the posterior distribution over a set of models given a priori knowledge $\theta$ and observed data $Y$. By Bayes' rule, the posterior distribution over parameters having observed data $Y$ is given by:

$$p(\theta|Y) = \frac{p(Y|\theta)\,p(\theta)}{p(Y)}.$$

In some applications, some latent or hidden variables which are unobserved data yet interact through the parameters to generate the data are included in the generative model, here we denote unobserved data by $X$. Given a generating model or a joint likelihood $p(Y, X|\theta)$, the posterior probability distribution function is

$$p(\theta|Y, X) = \frac{p(Y, X|\theta)\,p(\theta)}{p(Y, X)}.$$

We are also interested in calculating other related quantities, such as the predictive density of a new datum $y'$ given observed $Y$, the probability of the data can be written by summing over the possible settings of the hidden states:

$$p(y'|Y, \theta) = \sum_X p(y'|X, Y, \theta)\,p(X|Y, \theta).$$

### 2.3.2 Directed graphical model

A graphical model is a probabilistic model for which a graph expresses the conditional dependence structure between random variables. Graphical models are commonly used in Bayesian probability theory, Bayesian statistics and machine learning. The class of directed graphical models (DGM) is those graphical models in which all the inter-node connections have a direction, indicated visually by an arrowhead. DGM is also widely called a "Bayesian Network".

**2.3.2.1 Difficulty of chain rule in applications** From the chain rule of probability, we can always represent a joint distribution as follows, using any ordering of the variables:

$$p(x_1,...,x_D) = p(x_1)\,p(x_2|x_1)...p(x_D|x_1,...,x_{D-1}),$$

where we have dropped the conditioning on the fixed parameters $\theta$ for brevity. The problem with this expression is that it becomes more and more complicated to represent the conditional distributions $p(x_D|x_1,...,x_{D-1})$ as D gets large.

**2.3.2.2 Conditional independence** The key to efficiently representing large joint distributions is to make some assumptions about conditional independence (CI). We say $x$ and $z$ are conditionally independent given $y$ if and only if their conditional joint distributions can be written as a product of the conditional marginals,

$$x \perp z|y \Leftrightarrow p(x,z|y) = p(x|y)\,p(z|y).$$

CI assumption is widely applied in many situations. For example, given a set of observations $\{x_1,...,x_D\}$, if we assume $x_i$, $i=(1,...,D)$ is discrete, it is appropriate to suppose that the future $x_{D+1}$ is independent of the past given the present, i.e.

$x_{D+1} \perp x_1, ..., x_{D-1} | x_D$ . This CI example is called the first-order Markov assumption widely applied in time-series analysis. Using this assumption, plus the chain rule, we can write the joint distribution as follows:

$$p(x_1, ..., x_D) = p(x_1) \prod_{i=2}^{D} p(x_i | x_{i-1}),$$

this is called the first-order Markov chain. They can be characterized by an initial distribution over states and a state transition matrix.

Although the first-order Markov assumption is useful for defining distributions on one-dimensional sequences, how can we define distributions on two-dimensional images, or three-dimensional videos, or, in general, arbitrary collections of variables? This is where graphical models come in.

A graphical model (GM) (Bishop, 2006) is a way to represent a joint distribution by explicitly making CI assumptions. Several kinds of graphical models depend on whether the graph is directed, or undirected. In this thesis, we are only interested in DGM which is also called a Bayesian Network.

**2.3.2.3 Graph terminology** Here, we give basic definitions on graph theory (Barber, 2012). A graph $G = (V, E)$ consists of a set of nodes or vertices, $V = \{1, ..., v\}$ , and a set of edges, $E = \{(x, y) : x, y \in V\}$ . We can represent the graph by its adjacency matrix, in which we write $G(x, y) = 1$ to denote $(x, y) \in E$ , that is, $x \to y$ is an edge in the graph. If $G(x, y) = 1$ whenever $G(y, x) = 1$, we say the graph is undirected, otherwise it is directed. We usually assume $G(x, x) = 0$ which means there are no self loops. Here are terms we will commonly use:

Parent: For a directed graph, the parents of a node are the set of all nodes that

feed into it:

$$pa(x) \triangleq \{y : G(y,x) = 1\}.$$

Child: For a directed graph, the children of a node are the set of all nodes that feed out of it:

$$ch(x) \triangleq \{y : G(x,y) = 1\}.$$

Descendants: For a directed graph, the descendants are the children, grand-children, etc of a node.

Cycle or loop: For any graph, we define a cycle or loop to be a series of nodes such that we can get back to where we started by following edges, $x_1 - x_2 - ...x_n - x_1, n \geq 2$. If the graph is directed, we may speak of a directed cycle.

DAG : A directed acyclic graph ( DAG) is a directed graph with no directed cycles.

Directed path: In DAG, a directed path is a path (a sequence of nodes) in which the edges are all oriented in the same direction.

Undirected path: In DAG, $G = (V, E)$, an undirected path is an acyclic path of the augmented undirected graph $G' = (V, E')$, where $E'$ is defined as follows:

$\forall x, y \in V$, if $E(x,y) = 1$, we can get $E'(x,y) = 1$ and $E'(y,x) = 1$.

Topological ordering: For a DAG, a topological ordering or total ordering is a numbering of the nodes such that parents have lower numbers than their children. It can be shown that every DAG has at least one topological order.

**2.3.2.4 Directed graphical models** According to Murphy (2012), directed graphical models are represented by a directed acyclic graph (DAG). Each node of the DAG represents a random variable of the model. The nodes are connected by directed

links to describe dependencies.

**Definition 1.** *A Directed Graphical model (DGM) consists of three components. Firstly, a DAG,* $G=(V,E)$. *Secondly, a set of conditional probability distribution function* $p(x|pa(x))$, $x\in V$. *Thirdly, for any* $X,Y,Z\subset V$, *DGM encodes the following CI:* $X\perp Z|Y \Leftrightarrow X$ *is d-separated from* $Z$ *given* $Y$, *where the definition of "d-separation" will be given below.*

**Definition 2.** *Let* $X,Y,Z$ *be three sets of nodes in DAG,* $G=(V,E)$, *where* $Y$ *is the evidence set or the set of observed nodes. We say that* $X$ *and* $Z$ *are d-separated given* $Y$ *if and only if every undirected path P of G between any node* $x\in X$ *and* $z\in Z$ *given* $Y$ *is "blocked".*

*Here, the term "blocked" means that there is an intermediate variable* $y$ *satisfying at least one of the following conditions P with respect to the original edges E:*

*P contains a chain,* $x\to y\to z$ *or* $z\leftarrow y\leftarrow x$, *where* $y\in Y$;

*P contains or fork,* $x\leftarrow y\to z$, *where* $y\in Y$;

*P contains a* ***v-structure***, $x\to y\leftarrow z$, *where y is not in* $Y$ *and nor is any descendant of* $y$.

Example: consider the DGM, as shown in Figure 2.1

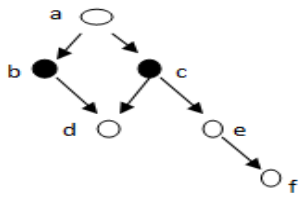


**Figure 2.1** The example of DGM.

Here $Y=\{b,c\}$ is a set of observed nodes. From the graph, Definition 1 and

Definition 2, there are many CI encoded by the given DGM. We only state two examples here.

I. $d \perp a \mid Y$

*Proof.* According to the Definition 2, $X = \{d\}$, $Z = \{a\}$, $Y = \{b, c\}$. From the given graph G, there are two undirected path from *d* to *a*, which are path 1: $d \leftarrow b \leftarrow a$; path 2: $d \leftarrow c \leftarrow a$.

In path 1, $d \leftarrow b \leftarrow a$ is chain path, since $b \in Y$. In path 2, $d \leftarrow c \leftarrow a$ is also a chain path, since $c \in Y$. By Definition 2, every undirected path from d to a is blocked, so $d$ and $a$ d-separated given $Y$. Hence by Definition 1, $d \perp a \mid Y$.

□

II. $d \perp f \mid Y$

*Proof.* According to the Definition 2, $X = \{d\}$, $Z = \{f\}$, $Y = \{b, c\}$. From the given graph G, there are two undirected path from $d$ to $f$, which are path 1: $d \leftarrow c \rightarrow e \rightarrow f$, path 2: $d \leftarrow b \leftarrow a \rightarrow c \rightarrow e \rightarrow f$.

In path 1, $d \leftarrow c \rightarrow e$ is fork path, since $c \in Y$. In path 2, $d \leftarrow b \leftarrow a$ is chain path, $b \in Y$. By Definition 2, every undirected path from $d$ to $f$ is blocked, so $d$ and $f$ are d-separated given $Y$. Hence by Definition 1, $d \perp f \mid Y$.

□

According to the definition of d-separation, a joint probability distribution $p(V)$ satisfy Local Markov Property with respect to a graph $G = (V, E)$.

**Proposition 1.** *(Local Markov Property) Given a DGM, a probability distribution function p over a set of nodes V satisfies the local Markov property with respect to a*

*graph* $G$ *i.e.*

$$\forall x \in V, \quad x \perp (non-desc(x) \setminus pa(x)) \,|\, pa(x)$$

where $non-desc(x)$ are the non-descendants of $x$.

*Proof.* see Cowell et al. (1999).

From the standard chain rule, topological ordering and Local Markov property, setting $p(V)$ the joint distribution of all nodes of a DAG can be expressed by the following simple form:

$$p(V) = \prod_{x \in V} p(x \,|\, pa(x)) . \tag{2.5}$$

The product decomposition in equation (2.5) is called "DGM chain rule" which will be proved shortly. Here, we emphasize that this result is very important as it effectively simplifies the standard chain rule, so the difficulty mentioned in this sub-section is solved. In fact, this result of equation (2.5) is the main reason of employing DGM since in real-world applications it is common to have hundreds or thousands of random variables, so we need a practical method to define their joint distributions. The following proposition formally states this result.

**Proposition 2.** *(DGM chain rule) Given DGM, setting* $p(V)$ *the joint distribution of all nodes of a DAG , then* $p(V)$ *can be expressed by:*

$$p(V) = \prod_{x \in V} p(x \,|\, pa(x))$$

*Proof.* Suppose $\forall x_1, ..., x_D \in V$, and $x_1 \succ x_2 \succ ... \succ x_D$ according to topological ordering.

By the chain rule

$$p(x_1, ..., x_D) = p(x_1)\, p(x_2 \,|\, x_1)\, p(x_3 \,|\, x_2, x_1) ... p(x_D \,|\, x_{D-1}, ..., x_1)$$

By topological ordering, $x_{i-1}, x_{i-2}, ..., x_1$ are non-descendants of $x_i$, so we can

rewrite $p(x_1,...,x_D)$ as

$$p(x_1,...,x_D) = p(x_1)p(x_2|pa(x_2))p(x_3|pa(x_3),A_3)...p(x_D|pa(x_D),A_D)$$

where $A_i \subseteq non-descendants(x_i) \setminus pa(x_i)$.

By Local Markov property, $x_i \perp non-descendants(x_i) \setminus pa(x_i) | pa(x_i)$, hence

$$\begin{aligned} p(x_1,...,x_D) &= p(x_1)p(x_2|pa(x_2))p(x_3|pa(x_3))...p(x_D|pa(x_D)) \\ &= p(x_i)\prod_{i=2}^{D} p(x_i|pa(x_i)) \\ &= \prod_{x\in V} p(x|pa(x)) \end{aligned}$$

□

### 2.3.3 Bayesian parameter estimation

Up until now, when we have spoken of "probabilistic inference", we have assumed that we completely know everything about the mentioned distributions, i.e. its functional form and parameters. Nevertheless, in real-world applications, we are given only the data, not the distributions. It is common to assume that the functional forms are known and to learn the parameters from data. Suppose that, for each observation in $Y = \{y_1,...,y_t\}$, denoting the parameter by $\theta$, in Bayesian statistics, parameter estimation is simply to compute the following posterior distribution

$$p(\theta|Y) \propto p(Y|\theta)p(\theta) \tag{2.6}$$

where $p(\theta)$ is the prior distribution which models the prior beliefs on the parameter before we have seen any data.

However, for most probabilistic models of practical interest, exact inference like equation (6) is intractable, and so we have to resort to some form of approximation. We discuss inference algorithms based on deterministic approximations, which include

methods such as Maximum Likelihood. In particular, we focus on the EM Algorithm to calculate the Maximum Likelihood.

**2.3.3.1 Maximum Likelihood** Recall the definition of the maximum likelihood (ML) estimation problem. We have a density function $p(Y|\theta)$ that is governed by the set of parameters $\theta$. We also have a data set of size $t$, denoted $Y = \{y_1, ..., y_t\}$. We define the ML by reversing the roles of the data $Y$ and the parameters $\theta$ in $p(Y|\theta)$, i.e.

$$L(\theta|Y) \triangleq p(Y|\theta),$$

This function $L(\theta|Y)$ is called the likelihood of the parameters $\theta$ given the data $Y$, or just the likelihood function. The likelihood is thought of as a function of the parameters $\theta$ where the data $Y$ is fixed, i.e. $Y$ is observed. In the maximum likelihood problem, our goal is to find $\theta$ that maximizes $L$. That is, we wish to find $\theta^*$ where

$$\theta^* = \arg\max_{\theta} L(\theta|Y).$$

Often we maximize $\log p(Y|\theta)$ instead because it is analytically easier.

The likelihood function is simplified if we make a standard independent and identically distributed (i.i.d.) assumption : $p(Y|\theta) = \prod_{i=1}^{t} p(y_i|\theta)$.

According to the form of $p(y_i|\theta)$, if it is simply a single Gaussian distribution where $\theta = (\mu, \sigma^2)$, then we can set the derivative of $\log p(Y|\theta)$ to zero, and solve directly for $\mu$ and $\sigma^2$. For many problems, however, it is not possible to find such analytical expressions, and we must resort to more elaborate techniques. The EM algorithm is one important example of such elaborate techniques.

**2.3.3.2 Expectation Maximization algorithm** The expectation maximization

(EM) algorithm (Dempster et al., 1977; Redner and Walker, 1984; Jordan and Jacobs, 1994; Bishop, 1995) is a general method of finding the ML estimate of the parameters of an underlying distribution from a given data set. As mentioned in subsection 2.3.1, in the general setting, data is organized into two different categories, observed data and unobserved data. In some literature, "incomplete data" denotes only the set of observed data, and "complete data" means denotes the set of both observed data and unobserved data.

The EM algorithm is an iterative algorithm, often with closed-form updates at each step. Furthermore, the algorithm automatically enforces the required constraints. Recently, two main applications of EM algorithm have been obtained. The first occurs when the data indeed has missing values, due to problems with or limitations of the observation process. The second occurs when optimizing the likelihood function is analytically intractable. The latter application is more common in machine learning and statistics.

For this discussion, let us suppose that a set of all concerned variables $Z$ can be divided into subsets, $Z = \{X, Y\}$, where $Y$ is the set of variables which we can observe, but $X$ is the set of variables that we cannot observe in the interested application. The set of unobserved variables $X$ is also usually called the set of latent or the set of hidden variables. Therefore, as mentioned earlier, we will refer $Z$ as a complete data set, and $Y$ as an incomplete data set. A joint density function can be obtained as :

$$p(Z|\theta) = p(X, Y|\theta) = p(X|Y, \theta)\, p(Y|\theta)$$

with this probability distribution function, we can define a new likelihood function

$$L(\theta|Z) = L(\theta|X,Y) = p(X,Y|\theta),$$

this is called the complete-data likelihood. Note that this function is in fact a random variable since the missing information $X$ is unknown, random, and presumably governed by an underlying distribution. We think of $L(\theta|X,Y)$ as a function where $Y$ and $\theta$ are constant, and $X$ is a random variable. The original likelihood $L(\theta|Y)$ is referred to as the incomplete-data likelihood function.

The EM algorithm first finds the expected value of the complete-data log-likelihood $\log p(X,Y|\theta)$ with respect to the unknown data $X$ given the observed data $Y$ and the current parameter estimates. We define:

$$Q(\theta,\theta^{(w-1)}) = E[\log(p(X,Y|\theta)|Y,\theta^{(w-1)})]$$

where $\theta^{(w-1)}$ are the current parameter estimates that we have used to evaluate the expectation and $\theta$ are the new parameters that we optimize to increase $Q$.

The evaluation of this expectation is called the E-step of the algorithm. Notice the meaning of the two arguments in the function $Q(\theta,\theta^{(w-1)})$. The first argument $\theta$ corresponds to the parameters that ultimately will be optimized in an attempt to maximize the likelihood. The second argument $\theta^{(w-1)}$ corresponds to the parameters that we use to evaluate the expectation.

The evaluation of this maximization is called the M-step of the EM algorithm and it is to maximize the expectation we computed in the first step. That is, we find:

$$\theta^{(w)} = \arg\max_{\theta} Q(\theta,\theta^{(w-1)}).$$

These two steps are repeated as necessary. Each iteration is guaranteed to

increase the log-likelihood and the algorithm is guaranteed to converge to a local maximum of the likelihood function reference. In the chapter IV, we will exploit the EM algorithm under the Bayesian framework.

# CHAPTER III

# GARCH-TYPE FORECASTING MODELS FOR VOLATILITY OF STOCK MARKET AND MCS TEST

## 3.1 Introduction

A large number of time series based volatility models have been developed since the introduction of the autoregressive conditional heteroskedasticity (ARCH) model of Engle (1982) and generalized ARCH model proposed by Bollerslev (1986).

But which is the best forecasting model? It is difficult to answer this question because asset returns often do not contain sufficient information to identify a single volatility model as "best". Hansen and Lunde (2005) offer some resolution of this quandary, The metric for assessing the forecasts of volatility models is the bootstrap method of superior predictive ability (SPA) test. But to use the SPA test, we need to carefully choose the basic model, as its choice can affect the result. In order to overcome the defects of the SPA test, this thesis uses the model confidence set (MCS) test which is a modified version of the SPA test. The MCS approach has three advantages over tests for SPA.

Firstly, the MCS procedure is independent of any benchmark model, while the SPA tests are not. Secondly, the MCS method characterizes the entire set of models that are not significantly out-performed by other models, while a test for SPA only

provides evidence about the relative performance of a particular model.

Thirdly, the MCS method relies on tests of simple hypotheses. Thus, it avoids the potential problem of SPA tests in which composite hypotheses are examined by Hansen et al. (2005).

In this chapter, we seek to identify the superior model in capturing the characteristics of the SSE380 index. We use the symmetric GARCH and asymmetric GARCH (EGARCH and TGARCH) models with normal innovation and student's t innovation to forecast volatility. Then we use the MCS test based on the bootstrap simulation to choose the best model.

## 3.2 GARCH-type forecasting models for volatility

The volatility of a stock price can be used as an indicator of the uncertainty of stock returns. In a financial market, volatility is measured in terms of standard deviation $\sigma$ or $\sigma^2$. Ser and Clive (2003) compute variance from a set of observations as follows:

$$\sigma^2 = \frac{1}{n-1}\sum_{t=1}^{n}(y_t - \bar{y})^2 ,$$

where $\bar{y}$ and $y_t$ are the mean return and return respectively. Return is defined to be the total gain or loss from an investment over a given period of time. In this part, we compute the daily closing prices as

$$y_t = 10\log(p_t / p_{t-1}),$$

where $p_t$ is stock closing price at time $t$. Then prices are converted into logarithmic returns, $y_t$ denotes the continuously compounded daily returns of the underlying assets at time $t$.

In the part, we assume that the conditional mean equation of stock return is constructed as the constant term[1] plus a residual term,

$$y_t = \mu + \varepsilon_t, \quad \varepsilon_t = h_t z_t$$

where $\{z_t\}$ is a sequence of independent identically distributed random variables with zero-mean and unit variance, $h_t^2$ is the conditional variance of $\varepsilon_t$ derived from mean equation, it is also known as current day's variance or volatility. Larger $h_t^2$ implies higher volatility and higher risk.

### 3.2.1 GARCH (1,1) model

The standard variance model for financial data is GARCH. According to the equation (2.1), the GARCH(1,1) is defined as

$$h_t^2 = \omega + \beta h_{t-1}^2 + \alpha \varepsilon_{t-1}^2,$$

where $\omega > 0$, $\alpha \geq 0$, $\beta \geq 0$, $\alpha + \beta < 1$. There are also some meanings about the parameters. Firstly, $\omega > 0$ means that volatility cannot have a zero or negative mean. Secondly, the positive parameters $\alpha, \beta$ show that the conditional variance forecasts will increase if there is a large fluctuation in returns, the model thus capturing the stylized feature of volatility clustering. Finally, $\alpha + \beta < 1$ indicates the persistence of shocks to volatility will eventually fade away, which depicts another stylized characteristic of volatility, mean reversion.

[1] As discussed by Engle and Patton (2001), the specification of the mean equation is not important for forecasting studies, without significantly degrading the performance of the proposed model. In the part, the results of the model estimations are not presented when the study concentrates on forecasting performance, but the model is available from the author's request.

### 3.2.2 Exponential-GARCH (1,1) model

A more flexible and often cited GARCH extension is Exponential GARCH (EGARCH) (Hamilton, 1994). According to equation (2.2), EGARCH(1,1) can be defined as the following:

$$\log(h_t^2) = \omega + \beta \log(h_{t-1}^2) + \alpha \left[ \frac{|\varepsilon_{t-1}|}{h_{t-1}} - E(\frac{|\varepsilon_{t-1}|}{h_{t-1}}) \right] + r(\frac{\varepsilon_{t-1}}{h_{t-1}}),$$

where $\alpha$ captures the volatility clustering effect, $\omega$ is constant and the $r$ measures the leverage effect.

### 3.2.3 Threshold-GARCH (1,1)

According to Glosten et al. (1993) and from the equation (2.3), we can get TGARCH(1,1) as the following:

$$h_t^2 = \omega + \beta h_{t-1}^2 + \alpha \varepsilon_{t-1}^2 + \gamma \varepsilon_{t-1}^2 \mathrm{I}_{t-1},$$

$$\text{where } \mathrm{I}_{t-1} = \begin{cases} 1 & if\ \varepsilon_{t-1} < 0 \\ 0 & if\ \varepsilon_{t-1} \geq 0 \end{cases}.$$

where $\omega > 0,\ \alpha, \beta \geq 0$ and $\alpha + \gamma \geq 0$.

## 3.3 Experiments with real data

In this part, we analyzes SSE380 index $p_t$, which was first downloaded from HuaChuang securities and then transformed into log returns. The SSE 380 index consists of the 380 stocks with midcap, high growth and good earning records, which aims to comprehensively reflect the performance of the Shanghai new blue chip stocks. So it is useful to analyse SSE 380 index in the Shanghai stock market. The constituents selection space of SSE 380 Index is all the Shanghai Stocks except the following stocks (come

from China Securities Index Co): SSE 180 index constituents, the stocks with negative retained earnings in latest financial report, the stocks over more than five years haven't distributed cash dividends or stock dividends in latest five years.

### 3.3.1 The descriptive statistics of data

The whole sample consists of 1258 daily data spanning from 4 Jan. 2010 to 16 Mar.2015. We select a subsample of size 1000, dated from 4 Jan. 2010 to 24 Feb. 2014, as the training set for the parameters estimation for models and the remaining sample of size 258 daily data, from 25 Feb. 2014 to 16 Mar.2015 is used as the test set or for out of sample forecasting.

Then we need to calculate the log return $y_t = 10\log(p_t / p_{t-1})$. Table 1 summarizes the descriptive statistics of SSE380 index along the whole period.

**Table 3.1** The summary statistic of the SSE380.

| | | | |
|---|---|---|---|
| Sample | 1258 | Kurtosis | 4.4124 |
| Mean | 0.00321 | Skewness | -0.54089 |
| Std. Dev | 0.15331 | JB test | 165.771 |

The Table 3.1 remarks that these facts suggest a highly competitive and volatile market. The skewness is -0.540886<0, the negative skewness indicates that there is a high probability of loss in the market. The value of the Kurtosis is 4.412388>3, it suggests that the market is volatile with high probability of extreme event occurrences. The Jarque-Bera (JB) test is 165.7707 which shows that the returns deviate from the normal distribution significantly and exhibit leptokurtic. Hence the distribution of the index series is not the normal distributed, and it has the feature of asymmetry, zero mean

and left side. Using the Augmented Dickey-Fuller (ADF) Unit Root Tests, the value of ADF test statistic in the log returns of SSE380 is -43.89175 less than 1% level of the critical value -3.435344, it means that the series of $y_t$ is stationary time series.

### 3.3.2 Detecting ARCH effects of data returns

From the Figure 3.1, we can see that the returns appear to fluctuate around a constant level but exhibit volatility clustering. Large changes in the returns tend to cluster together, and small changes tend to cluster together. So the preliminary judgment shows that the series exhibits the conditional heteroscedasticity. Now we use ARCH-LM to detect whether SSE380 returns have ARCH effects.

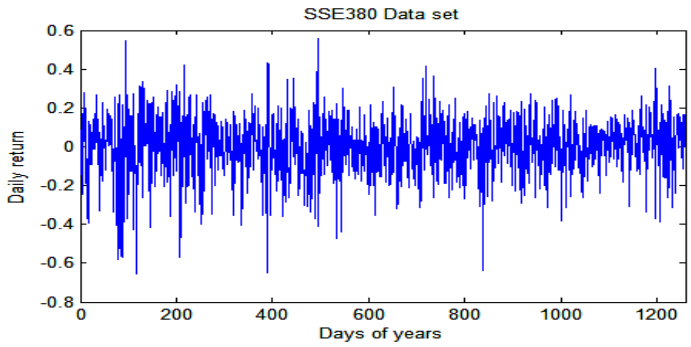


**Figure 3.1** The daily return of the SSE380.

According to the heteroskedasticity test ARCH, the value of F-statistic is 9.810969 and the probability is 0.0018<0.05, $R^2 = 9.750333$, the probability is 0.0018<0.05, and the number of lags is 1, the test of residuals for ARCH(1) rejects the null hypothesis of no conditional heteroskedasticity, so it is clear that SSE380 returns have ARCH effects. Then we can use GARCH-type models to forecast the volatility.

## 3.4 Estimation result of models

We apply the return series to the GARCH, EGARCH and TGARCH models with

Normal innovation and Student's innovation, and then we get their parameters. The estimation results and diagnosis are shown in Table 3.2.

**Table 3.2** The estimation results of models.

| statistics | GARCH-N | EGARCH-N | TGARCH-N |
|---|---|---|---|
| $\mu$ | -0.08998(0.0062) | -0.064876(0.0299) | -0.03359(0.2706) |
| $\omega$ | 0.00156(0.0032) | -0.440637(0.0001) | 0.002476(0.0007) |
| $\alpha$ | 0.07287(0.0001) | 0.152545(0.0001) | -0.006121(0.00) |
| $\beta$ | 0.865212(0.00) | 0.913790(0.00) | 0.126291(0.00) |
| $\gamma$ | —— | -0.077690(0.00) | 0.022767(0.0907) |
| LL | 448.1405 | 452.7252 | 454.1026 |
| AIC | -0.887168 | -0.894345 | -0.897102 |
| BIC | -0.862610 | -0.864875 | -0.885901 |

| statistics | GARCH-T | EGARCH-T | TGARCH-T |
|---|---|---|---|
| $\mu$ | -0.096163(0.0042) | -0.079875(0.0113) | -0.012500(0.6615) |
| $\omega$ | 0.001588 (0.0312) | -0.408222(0.0014) | 0.002820 (0.0026) |
| $\alpha$ | 0.065790(0.0019) | 0.139411(0.0019) | -0.035497(0.1979) |
| $\beta$ | 0.870771(0.00) | -0.063736(0.0084) | 0.829293(0.00) |
| $\gamma$ | —— | -0.026175(0.0718) | 0.161594(0.0002) |
| LL | 458.1541 | 461.1307 | 462.9269 |
| AIC | -0.905213 | -0.909171 | -0.912767 |
| BIC | -0.875743 | -0.874789 | -0.878385 |

Among the parametric models, with normal innovation and student's t innovation, in GARCH-N, the value of Log Likelihood (LL) is 448.1405, Akaike Information Criterion (AIC) is -0.887168 and Bayesian Information Criterion (BIC) is -0.862610, each parameter is significant. In EGARCH-N, the value of LL is 452.7252, AIC is -0.894345 and BIC is -0.864875, each parameter is significant. In TGARCH-N, the parameters $\gamma, \mu$, the respective value of the probability is more than 0.05. In GARCH-T, t the value of LL is 458.1541, AIC is -0.905213 and BIC is -0.4875743, each parameter is significant. In EGARCH-T, the value of $\gamma$ is not significant. In TGARCH-T, the values of $\mu, \alpha$ also are not significantly. Hence according to highest value of LL and smallest value of AIC and BIC, GARCH-T is the best series fit.

In fact, the GARCH-T model has been applied in many financial fields. Dumitru and Cristiana (2010) focus on the US and Romanian stock markets, and find that GARCH-T errors provide a better description for the conditional volatility. Rakesh et al. (2010) apply GARCH-T to examine the asymmetric nature of the US stock market returns. Hence the GARCH-T model can be used to describe properties of volatility in different stock markets and also can be used to forecast the volatility in the future. Lee and Su (2012) use the GARCH-T model to analyse the thirteen stock indices in North America, Europe and Asia to provide data for examining the one-day-ahead VaR forecasting. Furthermore Heitham et al. (2015) examine GARCH with different distributions, as normal, student's t and generalized error distribution (GED), to analyze the Jordanian stock market returns over the period January 2000 – November 2014.

## 3.5 Model confidence set (MCS) test method

When obtaining the predicted values, we can compare with the real proxy variables of a volatility deviation size. However, there is no consensus on the loss functions which are used to measure the prediction error. In the part, we only use two loss functions: mean square error and mean absolute deviation to measure the forecasting error. It is not easy to choose the best model which is always the best under all choices of loss functions or all data samples. Hansen (2003) offers some resolution of this quandary, the metric for assessing the forecasts of volatility models is the Bootstrap method of superior predictive ability (SPA) test. But to use the SPA test, we must need to choose the basic model; it is very vital to choose it which can affect the result. In order to overcome the defects of the SPA test, we use the MCS test which is a modified version of the SPA test.

### 3.5.1 The MCS test procedure

We define a set of models which are denoted by $M_0 = \{1,...,m\}$, the models are indexed by $i = 1,...,m$, and model $i's$ forecasts of $h_t^2$ is denoted by $h_{i,t}^2$, We rank the models according to their expected losses using one of two loss functions: MSE, $L(h_{i,t}^2, h_t^2) = (h_{i,t}^2 - h_t^2)^2$,and MAD, $L(h_{i,t}^2, h_t^2) = \left|h_{i,t}^2 - h_t^2\right|$. The loss differential between models $i$ and $j$ is given by

$$d_{ij,t} = L(h_{i,t}^2, h_t^2) - L(h_{j,t}^2, h_t^2) \quad i, j = 1,...,m, \quad t = 1,...,n.$$

The MCS (Hansen, 2003; Jeff and Chris, 2011) is determined after sequentially trimming the set of candidate models, $M_0$. At each step, we impose the hypothesis

$$H_0 : E(d_{ij,t}) = 0, \text{ for all } \quad i, j \in M \subset M_0.$$

The hypothesis, $H_0$, is a test for Equal Predictive Ability (EPA) over the models in $M$

and if $H_0$ is rejected, the worst performing model is eliminated from $M$. The trimming ends when the first non-rejection occurs. The set of surviving models is the model confidence set $\widehat{M}_{\alpha}^{*}$. By holding the significance level, $\alpha$, fixed at each step of the MCS procedure, we construct a $(1-\alpha)$ confidence set, $\widehat{M}_{\alpha}^{*}$ for the best models in $M_0$. However, the trimming model which is mentioned in the sequential inspection has a drawback. At each step in the test, we need to test the predictive power of any two prediction models and calculate a test statistic. To overcome this drawback, our tests for the EPA employ the rang statistic $T_R$, and the semi-quadratic statistic $T_{SQ}$, given by

$$T_R = \max_{i,j \in M} \frac{\left|\bar{d}_{ij}\right|}{\sqrt{\widehat{\mathrm{var}}(\bar{d}_{ij})}}, \qquad T_{SQ} = \sum_{i<j} \frac{(\bar{d}_{ij})^2}{\widehat{\mathrm{var}}(\bar{d}_{ij})}$$

where the sum is taken over the models in $M$, and $\widehat{\mathrm{var}}(\bar{d}_{ij})$ is an estimate of $\mathrm{var}(\bar{d}_{ij})$, see Dumitru and Cristiana (2010). If the test statistic value of $T_R$ and $T_{SQ}$ are larger, then it means rejecting the null hypothesis. In fact, their distribution is very complicated, and the covariance structure depends on the predictive value of each prediction model. So we use a bootstrap simulation study to find the p-value of the two statistics.

### 3.5.2 The result of MCS test

Table 3.3 shows the MCS test results by using bootstrap simulation at 1000 times. The values in the table represent MCS test p-values. According to Hansen et al. (2011), the part sets a basis p-value which is p=0.1. If the p-value is less than 0.1, then the volatility forecasting model is not good, and the model will be removed in the MCS inspection process. Otherwise, the model will survive.

According to the Table 3.3, when the loss function is the MSE, the p-values of in the GARCH-N and GARCH-T models are more than 0.1. But the p-values of the

other 4 models are less than 0.01. This means that EGARCH-N, EGARCH-T, TGARCH-N and TGARCH-T volatility forecasting models will be removed in the MCS inspection process. Considering the loss function for MAD, we find that only the p-value of the GARCH-T model is more than 0.1. Hence using the loss functions of MSE and MAD, we find that the values of $p_{T_R}$ and $p_{T_{SQ}}$ in the GARCH-T model are more than 0.1. Therefore, the GARCH-T model is the best one.

**Table 3.3** The MCS test results of the models.

| No. | Model | MSE | | MAD | |
|---|---|---|---|---|---|
| | | $p_{T_R}$ | $p_{T_{SQ}}$ | $p_{T_R}$ | $p_{T_{SQ}}$ |
| 1 | GARCH-N | 0.133 | 0.164 | 0.112 | 0.041 |
| 2 | GARCH-T | 0.211 | 0.232 | 0.191 | 0.154 |
| 3 | EGARCH-N | 0.061 | 0.054 | 0.097 | 0.056 |
| 4 | EGARCH-T | 0.054 | 0.044 | 0.063 | 0.041 |
| 5 | TGARCH-N | 0.061 | 0.072 | 0.033 | 0.042 |
| 6 | TGARCH-G | 0.073 | 0.024 | 0.081 | 0.053 |

From the empirical results, the GARCH-T model is the best model for forecasting volatility. Now, it is important to verify the conclusions from the simulation experiment by repeating the analysis for the actual data.

### 3.5.3 The result of prediction

The GARCH-T model is determined to forecast the weekly step ahead volatility in the SSE380. Since the daily return is defined as $y_t = 10\log(p_t / p_{t-1})$, hence

the k−period continuously compounded return is defined as

$$Y_m^k = 10\log(p_m / p_{m-1}) = 10\log(p_{t+k} / p_{t-1}).$$

It follows that $Y_m^k = \sum_{j=1}^{k} y_{t+j}$, in the expression, there are two indexes, one counting the days (t) and the other keeping track of the non-overlapping k periods. In this chapter, we use weekly (i.e. 5 days) return which is denoted by $Y_m^5 = \sum_{j=1}^{5} y_{t+j}$, and then forecast the 5 steps ahead volatility in the out of samples from 25 Feb. 2014 to 16 Mar. 2015 of the SSE380, i.e.258 daily datum. The result is shown in the Figure 3.2.

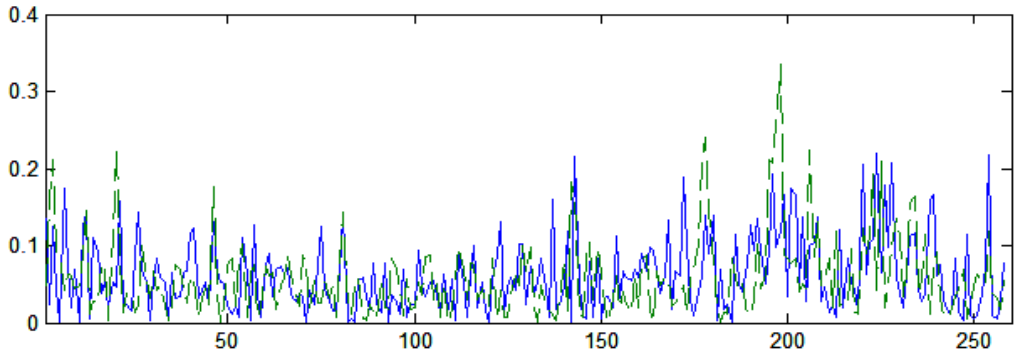


**Figure 3.2** The full line shows weekly step ahead volatility, while the dotted line shows the realized volatility of the out of samples in SSE380.

We obtain that most of the forecasts seem inefficient. Because in the GARCH-T model, the volatility is only a deterministic function of the squares of past returns, while in the real word, many more factors connect with volatility. Then even the GARCH-T model is incomplete. For example, Jeff and Chris (2011) find a positive correlation between trading volume and volatility. Another thing is that in the recent year China's stock market structure had seriously imbalances which caused huge fluctuations in the Shanghai Stock Exchange 380 index. Hence the GARCH-T model is not so good for forecasting, we need to improve it in the future.

# CHAPTER IV

# BAYESIAN INFERENCE FOR AN EXTENDED MARKOV REGIME SWITCHING STOCHASTIC VOLATILITY MODEL

As implied by the title, the mathematical treatment of the models and algorithms in the thesis is Bayesian, which means that all the results are treated as being approximations to certain probability distributions or their parameters. Probability distributions are used both to represent uncertainties in the models and for modeling the physical randomness. The theories of filtering, smoothing, and parameter estimation are formulated in terms of Bayesian inference, and EM algorithms are derived using the same Bayesian notation and formalism.

## 4.1 Introduction

The Stochastic Volatility (SV) (Taylor, 1986) model concentrates on the time-varying and volatility persistence, as well as on the leptokurtosis in financial return series. Many extensions to the basic SV models have been proposed in the literature (Heston, 1993; Sadorsky, 2005; Vo, 2009). In particular, the Markov Switching Stochastic Volatility model (MSSV), was studied in Diebold (1986) and Mike et al. (1998).

But on the basis of Mike and So (1998), there are some disadvantages to the MSSV model as set out in chapter II , when the model is used to forecast volatility.

Firstly, the MSSV model cannot be accurately described as some true values can be missed in an extreme financial event, such as abnormal changes. Secondly, using the MSSV model to estimate the volatility often implies a persistent default assumption. In fact, the traditional risk measurement model generally considers the volatility of financial assets as a single state. However, in some cases, the volatility of financial assets can be caused by the dynamic transformation or structural changes of different risk states. Hence the model may get “down fitting”. Finally, in the existing literature, researchers often use stochastic algorithms to infer about model parameters. But when the number of generations is prespecified, a particular run of genetic programming is not successful.

Fortunately, the dynamics of the relationship between stock return volatility and trading volume has a long history in the finance literature. Karpoff (1987) provides a good survey of this literature, discussing the return–volume relation in various financial markets. Andersen (1996) presents the intuitively appealing mixture of distributions hypothesis (MDH). According to the MDH, return and trading volume are driven by the same underlying latent information flow variable, i.e., price movements. The trading volume changes are caused primarily by the arrival of the volatility process. Grouard et al. (2003) find that persistence decreased when trading volume was used in the conditional variance equation. Alsubaie and Najand (2009) test the effect of trading volume on the persistence of the conditional volatility of returns in the Saudi stock market. Mahajan and Singh (2009) find a positive correlation between trading volume

and volatility, i.e. large change in trading volume means large change in volatility. Hence we can see that volatility not only correlates with stock return data but also correlates with stock volume data.

In order to improve the weakness of the MSSV model, in this section, we propose the Extended Markov Regime Switching Stochastic Volatility Model (E-MSSV) including E-MSSV-I and E-MSSV-II models. Our model is superior to the previous models in two ways. First, avoiding "under fitting", we must increase the model complexity. So the "volume" information is naturally incorporated into a model in order to improve the predictive power of the model. Second, a non-stochastic inference algorithm is derived to guarantee a local maximum of the estimated parameters, to improve the predictive power of the E-MSSV-I model.

## 4.2 Bayesian inference of the E-MSSV-I model

In this section, we consider the discrete random variables in the E-MSSV-I model, details are stated as follows.

### 4.2.1 The directed graphical model of the E-MSSV-I

From the MSSV model and the connection between volatility and volume, recall the terminology in chapter II, we represent the directed graphical model (DGM) of the E-MSSV-I model as:

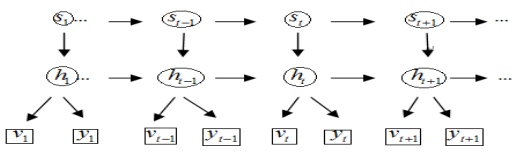

**Figure 4.1** The DGM of the E-MSSV-I.

From the Figure 4.1, the set of random variable is $\{y_t, h_t, s_t, v_t\}$, $t \in N$ and the input set is $\{y_1^T, v_1^T\}$, where $y_1^T = \{y_1, ..., y_T\}$, and the output is the prediction $h_{T+1}$.

the set of parents of $y_t$ is, $pa(y_t) = \{h_t\}$, and

the set of parents of $h_t$ is, $pa(h_t) = \{s_t, h_{t-1}\}, t \geq 2$ $pa(h_1) = \{s_1\}$, and

the set of parents of $s_t$ is, $pa(s_t) = \{s_{t-1}\}, t \geq 2$ $pa(s_1) = \phi$, and

the set of parents of $v_t$ is $pa(v_t) = \{h_t\}$.

The conditional probability distribution function for each child and parent node is defined as follows:

$$p(y_t | h_t = a_i) = (2\pi \exp(a_i))^{-\frac{1}{2}} \exp(-\frac{y_t^2}{2e^{a_i}}) \triangleq g_i(y_t) \tag{4.1}$$

$$p(h_t = a_i | h_{t-1} = a_j, s_t = b_m) \triangleq f_{ijm}, \quad t \geq 2 \tag{4.2}$$

$$p(v_t = c_l | h_t = a_i) \triangleq \gamma_{li}, \quad t \geq 1 \tag{4.3}$$

$$p(s_t = b_m | s_{t-1} = b_n) = p_{mn}, \quad t \geq 2 \tag{4.4}$$

where $i, j \in \{1, ..., I\}$, $m, n \in \{1, ..., M\}$, $l \in \{1, ..., L\}$. This is the E-MSSV-I model.

The initial probability distribution function is

$$p(h_1 = a_i | s_1 = b_m) \triangleq f'_{im} \tag{4.5}$$

$$p(s_1 = b_m) \triangleq p'_m \tag{4.6}$$

where $i \in \{1, ..., I\}$, $m \in \{1, ..., M\}$.

Hence the set of parameters $\theta$ is $\theta = \{p_{mn}\} \cup \{f_{ijm}\} \cup \{\gamma_{li}\} \cup \{g_i\} \cup \{f'_{im}\} \cup \{p'_m\}$.

### 4.2.2 Model inference with known parameters

In order to simplify the technical details behind a Bayesian inference, we start with the assumption that $\theta$ is known.

In this thesis, there are three main inference problems present in Bayesian

inference, including prediction probability distribution function, filtering probability distribution function, and smoothing probability distribution function. Prediction probability distribution function, i.e. $p(h_{T+1}, s_{T+1} \left| y_1^T, v_1^T, \theta \right.)$ , $p(y_{T+1} \left| y_1^T, v_1^T, \theta \right.)$ and $p(v_{T+1} \left| y_1^T, v_1^T, \theta \right.)$, which can be computed with the prediction step of the Bayesian filter is the marginal distributions of the future state $h_{T+1}, s_{T+1}$ and the future value of observed values, with one step after the current time step. Filtering probability distribution function, i.e. $p(h_T, s_T \left| y_1^T, v_1^T, \theta \right.)$ computed by the Bayesian filter is the marginal probability distribution function of the current state $h_T, s_T$ given the current and previous of observed variables $y_1^T$ and $v_1^T$. Smoothing probability distribution function, i.e. $p(h_t, s_t \left| y_1^T, v_1^T, \theta \right.)$ , $t = 1, ..., T-1$ , computed by the Bayesian smoother is the marginal probability distribution function of the state $h_t, s_t$ given a certain interval of observed variables $y_1^T$ and $v_1^T$.

Bayesian smoothing is often considered to be a class of methods within the field of a Bayesian filtering. While the Bayesian filters in their basic form only compute estimates of the current state of the system given the history of observed values, Bayesian smoothers can be used to reconstruct states that happened before the current time.

**4.2.2.1 Prediction probability distribution functions** In the tackle of prediction probability distribution function, $\{y_1, ..., y_T\} \cup \{v_1, ..., v_T\}$ are observed variables, $y_{T+1}, v_{T+1}, h_{T+1}$ and $s_{T+1}$ are unobserved variables. With the preceding notations, the model describing the unobserved variables satisfies the sufficient conditions for existence of a probability distribution function.

When processing data, at each time t, observable variables are used to estimate the current hidden state of the system and the prediction on the state of the system at time t+1. In order to predict the future value of the state of the system, given the information available at time t, we use the Chapman-Kolmogorov equation (Chapman, 1928; Kolmogorov, 1931) to characterize the hidden state evolution and give us the prediction probability distribution function as follows:

**Proposition 3.** *The prediction probability distribution function of* $y_{T+1}$ *given* $\theta$, $y_1^T$ *and* $v_1^T$ *has the following expression*

$$p(y_{T+1}\left|y_1^T, v_1^T, \theta\right.) = \sum_{i=1}^{I}\sum_{m=1}^{M} g_i(y_{T+1})\, p(h_{T+1}=a_i, s_{T+1}=b_m \left| y_1^T, v_1^T, \theta\right.). \tag{4.7}$$

*Proof.* Since

$$p(y_{T+1}\left|y_1^T, v_1^T, \theta\right.) = \sum_{h_{T+1}, s_{T+1}} p(y_{T+1}, h_{T+1}, s_{T+1} \left| y_1^T, v_1^T, \theta\right.), \tag{4.8}$$

according to the Chapman-Kolmogorov equation, the equation (4.8) can be given as

$$p(y_{T+1}\left|y_1^T, v_1^T, \theta\right.) = \sum_{h_{T+1}}\sum_{s_{T+1}} p(y_{T+1}\left|y_1^T, v_1^T, h_{T+1}, s_{T+1}, \theta\right.) p(h_{T+1}, s_{T+1}\left|y_1^T, v_1^T, \theta\right.)$$

with the CI rule of $y_{T+1} \perp s_{T+1}, y_1^T, v_1^T \left| h_{T+1}\right.$, hence

$$p(y_{T+1}\left|y_1^T, v_1^T, \theta\right.) = \sum_{h_{T+1}}\sum_{s_{T+1}} p(y_{T+1}\left|h_{T+1}, \theta\right.) p(h_{T+1}, s_{T+1}\left|y_1^T, v_1^T, \theta\right.).$$

From (4.1), $p(y_{T+1}\left|h_{T+1}=a_i, \theta\right.) \triangleq g_i(y_{T+1}), i \in \{1,...,I\}$, the one step ahead predictive probability distribution function of $y_{T+1}$ can be rewritten as

$$p(y_{T+1}\left|y_1^T, v_1^T, \theta\right.) = \sum_{i=1}^{I}\sum_{m=1}^{M} g_i(y_{T+1})\, p(h_{T+1}=a_i, s_{T+1}=b_m \left| y_1^T, v_1^T, \theta\right.).$$

□

**Proposition 4.** *The prediction probability distribution function of* $v_{T+1}$ *given* $\theta$, $y_1^T$ *and* $v_1^T$ *has the following expression*

$$p(v_{T+1}=c_l \mid y_1^T, v_1^T, \theta)=\sum_{i=1}^{I}\sum_{m=1}^{M}\gamma_{li}\, p(h_{T+1}=a_i, s_{T+1}=b_m \mid y_1^T, v_1^T, \theta) \tag{4.9}$$

*where* $l$ *is constant, satisfying* $v_{T+1}=c_l$, $l=\{1,...,L\}$.

*Proof.* Since

$$p(v_{T+1} \mid y_1^T, v_1^T, \theta)=\sum_{h_{T+1}}\sum_{s_{T+1}} p(v_{T+1}, h_{T+1}, s_{T+1} \mid y_1^T, v_1^T, \theta), \tag{4.10}$$

according to the Chapman-Kolmogorov equation, the equation (4.10) can be given as

$$p(v_{T+1} \mid y_1^T, v_1^T, \theta)=\sum_{h_{T+1}}\sum_{s_{T+1}} p(v_{T+1} \mid y_1^T, v_1^T, h_{T+1}, s_{T+1}, \theta)p(h_{T+1}, s_{T+1} \mid y_1^T, v_1^T, \theta)$$

with the CI rules, $v_{T+1} \perp s_{T+1}, y_1^T, v_1^T \mid h_{T+1}$, hence

$$p(v_{T+1} \mid y_1^T, v_1^T, \theta)=\sum_{h_{T+1}}\sum_{s_{T+1}} p(v_{T+1} \mid h_{T+1}\, \theta)p(h_{T+1}, s_{T+1} \mid y_1^T, v_1^T, \theta).$$

From (4.3) $p(v_{T+1}=c_l \mid h_{T+1}=a_i, \theta)=\gamma_{li}$, $i=1,...,I$, $l=1,...,L$, the one step ahead predictive probability distribution function of $v_{T+1}$ can be rewritten as

$$p(v_{T+1}=c_l \mid y_1^T, v_1^T, \theta)=\sum_{i=1}^{I}\sum_{m=1}^{M}\gamma_{li}\, p(h_{T+1}=a_i, s_{T+1}=b_m \mid y_1^T, v_1^T, \theta).$$

□

**Proposition 5.** *The joint prediction probability distribution function of hidden state* $h_{T+1}, s_{T+1}$ *given* $\theta$, $y_1^T$ *and* $v_1^T$ *is given by*

$$p(h_{T+1}=a_i, s_{T+1}=b_m \mid y_1^T, v_1^T, \theta)=\sum_{j}\sum_{n} f_{ijm}\, p_{mn}\, p(h_T=a_j, s_T=b_n \mid y_1^T, v_1^T, \theta). \tag{4.11}$$

*Proof.* Since

$$p(h_{T+1}, s_{T+1} \mid y_1^T, v_1^T, \theta)=\sum_{h_T}\sum_{s_T} p(h_{T+1}, s_{T+1}, h_T, s_T \mid y_1^T, v_1^T, \theta), \tag{4.12}$$

according to the Chapman-Kolmogorov equation, the equation (4.12) can be given as

$$\begin{aligned}
&p(h_{T+1}, s_{T+1} \mid y_1^T, v_1^T, \theta)\\
&=\sum_{h_T}\sum_{s_T} p(h_{T+1}, s_{T+1} \mid y_1^T, v_1^T, h_T, s_T, \theta)p(h_T, s_T \mid y_1^T, v_1^T, \theta)\\
&=\sum_{h_T, s_T} p(h_{T+1} \mid y_1^T, v_1^T, h_T, s_T, s_{T+1}, \theta)p(s_{T+1} \mid y_1^T, v_1^T, h_T, s_T, \theta)\, p(h_T, s_T \mid y_1^T, v_1^T, \theta)
\end{aligned}$$

since $h_{T+1} \perp y_1^T, v_1^T, s_T \,|\, s_{T+1}, h_T$ and $s_{T+1} \perp h_T, y_1^T, v_1^T \,|\, s_T$ . The equation (4.12) can be rewritten as

$$p(h_{T+1}, s_{T+1} \,|\, y_1^T, v_1^T, \theta) = \sum_{h_T} \sum_{s_T} p(h_{T+1} \,|\, h_T, s_{T+1}, \theta) p(s_{T+1} \,|\, s_T, \theta) p(h_T, s_T \,|\, y_1^T, v_1^T, \theta) .$$

By applying equations (4.2) and (4.4) , we can get

$$p(h_{T+1} = a_i \,|\, h_T = a_j, s_{T+1} = b_m, \theta) \triangleq f_{ijm}, \quad p(s_{T+1} = b_m \,|\, s_T = b_n, \theta) = p_{mn}$$

where $i, j \in \{1, ..., I\}$, $m, n \in \{1, ..., M\}$, so we can get

$$p(h_{T+1} = a_i, s_{T+1} = b_m \,|\, y_1^T, v_1^T, \theta) = \sum_j \sum_n f_{ijm} p_{mn} p(h_T = a_j, s_T = b_n \,|\, y_1^T, v_1^T, \theta) .$$

Here $p(h_T = a_j, s_T = b_n \,|\, y_1^T, v_1^T, \theta)$ is the filtering probability distribution function described in the next subsection.

**4.2.2.2 Filtering probability distribution function** The most important problem in the Extended MSSV model is to estimate latent variables, i.e. $h_t$ and $s_t$, given the data and the model. Filtering is to estimate the joint probability distribution function of latent variables at time $T$ given the data throughout the all $t$, where $t \in \{1, ..., T\}$, i.e. $p(h_T, s_T \,|\, y_1^T, v_1^T, \theta)$.

**Proposition 6.** *Set* $\alpha_{Tim} \triangleq p(h_T = a_i, s_T = b_m \,|\, y_1^T, v_1^T, \theta)$ . *Then the filter posterior probability distribution function* $\alpha_{Tim}$ *is given by*

$$\alpha_{Tim} \propto g_i(y_T) \gamma_{li} \sum_{j=1}^{I} \sum_{n=1}^{M} f_{ijm} p_{mn} \alpha_{(T-1)jn} \tag{4.13}$$

*where*

$g_i(y_T) = p(y_T \,|\, h_T = a_i, \theta)$ , $\gamma_{li} = p(v_T = c_l \,|\, h_T = a_i, \theta)$ , $p_{mn} = p(s_T = b_m \,|\, s_{T-1} = b_n, \theta)$ , $f_{ijm} = p(h_T = a_i \,|\, h_{T-1} = a_j, s_T = b_m, \theta)$ , $i, j \in \{1, ..., I\}$ , $m, n \in \{1, ..., M\}$ *and* $l$ *is constant, satisfying* $v_T = c_l$ .

*When* $t = 1$

$$\alpha_{1im} \propto g_i(y_1)\gamma_{li} f'_{im} p'_m \tag{4.14}$$

*where* $g_i(y_1) = p(y_1 | h_1 = a_i, \theta)$ , $\gamma_{li} = p(v_1 = c_l | h_1 = a_i, \theta)$ , $f'_{im} = p(h_1 = a_i | s_1 = b_m)$ , $p'_m = p(s_1 = b_m)$, $i \in \{1,...,I\}$, $m \in \{1,...,M\}$ *and* $l$ *is constant, satisfying* $v_1 = c_l$.

Now we can use recursion to calculate filtering and continue the process, this is the forward method to solve filtering. Details are shown in Appendix A.

**4.2.2.3 Smoothing probability distribution function** Bayesian smoothing is often considered to be a class of methods within the field of Bayesian filtering. While Bayesian filters in their basic form only compute estimates of the current state of the system given the history of observations, Bayesian smoothers can be used to reconstruct states that happened before the current time. That is, smoothing is to estimate the joint probability distribution function of latent variables at time t given the whole dataset $\{y_t\} \cup \{v_t\}$, i.e. $p(h_t, s_t | y_1^T, v_1^T, \theta)$, where $t \in \{1,...,T-1\}$. We have following results:

**Proposition 7.** *The smoothing probability distribution function* $p(h_t = a_i, s_t = b_m | y_1^T, v_1^T, \theta)$ *is given by*

*when* $1 \le t < T-1$

$$p(h_t = a_i, s_t = b_m | y_1^T, v_1^T, \theta) = \beta_{tim}\alpha_{tim} \tag{4.15}$$

*where* $\beta_{tim} = \dfrac{\sum_{j=1}^{I}\sum_{n=1}^{M} \beta_{(t+1)jn} g_j(y_{t+1}) \gamma_{lj} f_{jin} p_{nm}}{\sum_{i'=1}^{I}\sum_{m'=1}^{M} g_{i'}(y_{t+1})\gamma_{l'i'} \sum_{j'=1}^{I}\sum_{n'=1}^{M} f_{j'i'n'} p_{n'm'} \alpha_{tj'n'}}$, $\alpha_{tim} \triangleq p(h_t = a_i, s_t = b_m | v_1^t, y_1^t, \theta)$

$g_i(y_{t+1}) = p(y_{t+1} | h_{t+1} = a_i, \theta)$, $f_{jim} = p(h_{t+1} = a_j | h_t = a_i, s_{t+1} = b_m, \theta)$,

$\gamma_{li} = p(v_{t+1} = c_l | h_{t+1} = a_i, \theta)$, $p_{nm} = p(s_{t+1} = b_n | s_t = b_m, \theta)$, $i, j \in \{1,...,I\}$,

$m, n \in \{1,...,M\}$ *and* $l$ *is constant, satisfying* $v_1 = c_l$.

*when* $t = T-1$

$$p(h_{T-1} = a_i, s_{T-1} = b_m \,|\, y_1^T, v_1^T, \theta) = \beta_{(T-1)im}\alpha_{(T-1)im} \tag{4.16}$$

*where* $\beta_{(T-1)im} = \dfrac{\sum_{j=1}^{I}\sum_{n=1}^{M} g_j(y_T)\gamma_{lj} f_{jin} p_{nm}}{\sum_{i'=1}^{I}\sum_{m'=1}^{M} g_{i'}(y_T)\gamma_{l'i'}\sum_{j'=1}^{I}\sum_{n'=1}^{M} f_{j'i'n'} p_{n'm'}\alpha_{(T-1)j'n'}}$, $g_j(y_T) = p(y_T \,|\, h_T = a_j, \theta)$,

$\alpha_{(T-1)im} = p(h_{T-1} = a_i, s_{T-1} = b_m \,|\, v_1^t, y_1^t, \theta)$, $f_{jin} = p(h_T = a_j \,|\, h_{T-1} = a_i, s_T = b_n, \theta)$,

$\gamma_{lj} = p(v_T = c_l \,|\, h_T = a_j, \theta)$, $p_{nm} = p(s_T = b_n \,|\, s_{T-1} = b_m, \theta)$, *l is constant, satisfying* $v_1 = c_l$

*and* $i, j \in \{1,...,I\}, m, n \in \{1,...,M\}$.

This is backward method to solve smoothing probability distribution function. Details are presented in Appendix B.

### 4.2.3 Model inference with unknown parameters: EM algorithm

If $\theta$ is unknown, we have only the observed data. We will apply the Expectation-Maximization (EM) algorithm to find the maximum likelihood solutions for models having latent variables. The EM algorithm alternates between an E step and an M step. In the E step, we infer posterior distributions over hidden variables given a current parameter setting, and then we use this posterior distribution to find the expectation of the complete-data log likelihood evaluated for some general parameter value. In the M step, we determine the revised parameter estimate $\theta$ by maximizing the function gathered from the E step.

**4.2.3.1 Model assumptions** From Figure 4.1, we denote the set of all observed data by $\{V, Y\} \triangleq \{v_t, y_t\}_{t=1}^{T}$, and similarly we denote the set of all latent variables by $\{H, S\} \triangleq \{h_t, s_t\}_{t=1}^{T}$. The set of all model parameters is denoted by $\theta$. We shall call $\{V, Y, H, S\}$ the complete data set, and we shall consider the actual observed data $\{V, Y\}$

as incomplete. The joint probability distribution function of every concerned variable is the same as in equations (4.1), (4.2), (4.3) and (4.4). The initial probability distribution function is the same with equation (4.5) and (4.6). The parameters set is

$$\theta = \{p_{mn}\} \cup \{f_{ijm}\} \cup \{\gamma_{li}\} \cup \{g_i(y_t)\} \cup \{f'_{im}\} \cup \{p'_m\}.$$

The parameters space is

$$\Theta = \left\{ \theta \left| \begin{array}{l} 0 \le f'_{im} \le 1, 0 \le p'_m \le 1, 0 \le p_{mn} \le 1, \sum_{i=1}^{I} f'_{im} = 1, \sum_{m=1}^{M} p'_m = 1, \\ \sum_{m}^{M} p_{mn} = 1, 0 \le \gamma_{li} \le 1, \sum_{l=1}^{L} \gamma_{li} = 1, 0 \le f_{ijm} \le 1, \sum_{i=1}^{I} f_{ijm} = 1 \end{array} \right. \right\}$$

We set the initial parameter $\theta^{(0)} \in \Theta$ to

$$\theta^{(0)} \triangleq \{p_{mn}^{(0)}\} \cup \{f_{ijm}^{(0)}\} \cup \{\gamma_{li}^{(0)}\} \cup \{g_i^{(0)}(y_1)\} \cup \{f_{im}'^{(0)}\} \cup \{p_m'^{(0)}\}.$$

**4.2.3.2 The view of EM for the E-MSSV-I** According to the E-MSSV-I model, in the E step, we use the current parameter values $\theta^{(w-1)}$ to find the posterior probability distribution function of the latent variables given by $p(H,S|V,Y,\theta^{(w-1)})$, where $w \ge 1$. Then we use the posterior probability distribution function to find the expectation of the complete-data log likelihood evaluated for some general parameter value $\theta$. This expectation, denoted $Q(\theta,\theta^{(w-1)})$, is represented as

$$Q(\theta,\theta^{(w-1)}) = \sum_{H}\sum_{S} q^{(w)}(H,S) \log p(V,Y,H,S|\theta) \tag{4.17}$$

$$q^{(w)}(H,S) \triangleq p(H,S|V,Y,\theta^{(w-1)}).$$

In the M step, we decide the revised parameter estimate $\theta^{(w)}$ by maximizing this function $\theta^{(w)} = \arg\max_{\theta\in\Theta} Q(\theta,\theta^{(w-1)})$. The EM algorithm can be shown as follows:

**Table 4.1** The EM algorithm.

---

Step 1. Start with an initial setting for the parameters $\theta^{(0)}$.

Step 2. E-step : Evaluate $q^{(w)}(H,S)$.

Step 3. M-step: Evaluate $\theta^{(w)}$ obtained by $\theta^{(w)} = \arg\max_{\theta\in\Theta} Q(\theta,\theta^{(w-1)})$.

where $Q(\theta,\theta^{(w-1)}) = \sum_{H,S} q^{(w)}(H,S)\log p(V,Y,H,S|\theta)$

Step 4. Iterate steps 2 and 3 until convergence. If the convergence criterion is not satisfied, then let $\theta^{(w-1)} \leftarrow \theta^{(w)}$ and return to step 2.

---

**4.2.3.3 Expectation step** Now the important way is that we must simplify $Q(\theta,\theta^{(w-1)})$ by two tricks as follows: First, we must explain the marginalized $q^{(w)}(H,S)$.Second, we need to calculate the factorized $\log p(V,Y,H,S|\theta)$.

**Marginalize:** since $q^{(w)}(H,S) \triangleq p(H,S|V,Y,\theta^{(w-1)})$, it can be simplified in the different situation.

**Factorize:** set $H = \{h_1,...,h_T\} = h_1^T$, $V = \{v_1,...,v_T\} = v_1^T$,

$S = \{s_1,...,s_T\} = s_1^T$, then we can get as follows:

$$\log p(V,Y,H,S|\theta) = \log[p(V,Y|H,S,\theta)p(H,S|\theta)]$$

$$= \log\{[\prod_{t=2}^{T} p(v_t,y_t|v_1^{t-1},y_1^{t-1},h_t,s_t,\theta)]p(v_1,y_1|h_1,s_1,\theta)$$

$$[\prod_{t'=2}^{T} p(h_{t'},s_{t'}|h_1^{t'-1},s_1^{t'-1},\theta)]p(h_1,s_1|\theta)\}$$

$$= \log\{[\prod_{t=2}^{T} p(v_t|v_1^{t-1},y_1^{t-1},y_t,h_t,s_t,\theta)\,p(y_t|v_1^{t-1},y_1^{t-1},h_t,s_t,\theta)]\times$$

$$p(v_1|h_1,s_1,y_1,\theta)\,p(y_1|h_1,s_1,\theta)\times$$

$$[\prod_{t'=2}^{T} p(h_{t'} \left| h_1^{t'-1}, s_1^{t'-1}, s_{t'}, \theta) p(s_{t'} \right| h_1^{t'-1}, s_1^{t'-1}, \theta)] p(h_1 \left| s_1, \theta) p(s_1 \right| \theta)\} .$$

This complexity can be reduced by exploiting the conditional independence structure in the graph. So we can get

$$\log p(V, Y, H, S \left| \theta) = \log\{[\prod_{t=2}^{T} p(v_t \right| h_t, \theta) p(y_t \left| h_t, \theta)] p(v_1 \right| h_1, \theta) p(y_1 \left| h_1, \theta)\}$$

$$[\prod_{t'=2}^{T} p(h_{t'} \left| h_{t'-1}, s_{t'}, \theta) p(s_{t'} \right| s_{t'-1}, \theta)] p(h_1 \left| s_1, \theta) p(s_1 \right| \theta)$$

$$= \log\{[\prod_{t=1}^{T} p(v_t \left| h_t, \theta) p(y_t \right| h_t, \theta)] \times$$

$$[\prod_{t'=2}^{T} p(h_{t'} \left| h_{t'-1}, s_{t'}, \theta) p(s_{t'} \right| s_{t'-1}, \theta)] p(h_1 \left| s_1, \theta) p(s_1 \right| \theta)\}$$

$$= \sum_{t=1}^{T} \log p(v_t \left| h_t, \theta) + \sum_{t=1}^{T} \log p(y_t \right| h_t, \theta) + \sum_{t'=2}^{T} \log p(h_{t'} \left| s_{t'}, h_{t'-1}, \theta) + \right.$$

$$\sum_{t'=2}^{T} \log p(s_{t'} \left| s_{t'-1}, \theta) + \log p(h_1 \right| s_1, \theta) + \log p(s_1 \left| \theta) , \right.$$

hence

$$\log p(V, Y, H, S \left| \theta) = \sum_{t=1}^{T} \log p(v_t \right| h_t, \theta) + \sum_{t=1}^{T} \log p(y_t \left| h_t, \theta) + \right.$$

$$\sum_{t'=2}^{T} \log p(h_{t'} \left| s_{t'}, h_{t'-1}, \theta) + \sum_{t'=2}^{T} \log p(s_{t'} \right| s_{t'-1}, \theta) +$$

$$\log p(h_1 \left| s_1, \theta) + \log p(s_1 \right| \theta) . \qquad (4.18)$$

Substituting (4.18) into (4.17), the expectation can be obtained as

$$Q(\theta, \theta^{(w)}) = \sum_{H} \sum_{S} q^{(w)}(H, S)[\sum_{t=1}^{T} \log p(v_t \left| h_t, \theta) + \sum_{t=1}^{T} \log p(y_t \right| h_t, \theta) + \right.$$

$$\sum_{t'=2}^{T} \log p(h_{t'} \left| s_{t'}, h_{t'-1}, \theta) + \sum_{t'=2}^{T} \log p(s_{t'} \right| s_{t'-1}, \theta) + \log p(h_1 \left| s_1, \theta) + \log p(s_1 \right| \theta)] \quad .$$

Then we can get

$$Q(\theta,\theta^{(w)}) = \sum_H \sum_S q^{(w)}(H,S) \sum_{t=1}^{T} \log p(v_t \mid h_t,\theta) +$$

$$\sum_H \sum_S q^{(w)}(H,S) \sum_{t=1}^{T} \log p(y_t \mid h_t,\theta) +$$

$$\sum_H \sum_S q^{(w)}(H,S) \sum_{t'=2}^{T} \log p(h_{t'} \mid s_{t'}, h_{t'-1},\theta) +$$

$$\sum_H \sum_S q^{(w)}(H,S) \sum_{t'=2}^{T} \log p(s_{t'} \mid s_{t'-1},\theta) +$$

$$\sum_H \sum_S q^{(w)}(H,S) \log p(h_1 \mid s_1,\theta) + \sum_H \sum_S q^{(w)}(H,S) \log p(s_1 \mid \theta). \qquad (4.19)$$

In the equation (4.19), **the first term** can be obtained as follows:

$$\sum_H \sum_S q^{(w)}(H,S) \sum_{t=1}^{T} \log p(v_t \mid h_t,\theta) = \sum_{t=1}^{T} \sum_{H,S} q^{(w)}(H,S) \log p(v_t \mid h_t,\theta)$$

$$= \sum_{t=1}^{T} \sum_{h_t,s_t} \log p(v_t \mid h_t,\theta) \sum_{\{H\}_{-t},\{S\}_{-t}} q^{(w)}(H,S)$$

$$= \sum_{t=1}^{T} \sum_{h_t,s_t} \log p(v_t \mid h_t,\theta) q^{(w)}(h_t,s_t)$$

$$= \sum_{t=1}^{T} \sum_{h_t,s_t} \log p(v_t \mid h_t,\theta)\, p(h_t,s_t \mid v_1^T, y_1^T, \theta^{(w-1)}). \qquad (4.20)$$

When $1 \le t \le T-1$, $q^{(w)}(h_t,s_t) = p(h_t,s_t \mid v_1^T, y_1^T, \theta^{(w-1)})$ is the smoothing probability density function, hence by (4.3), equation (4.20) can be rewritten as

$$\sum_{t=1}^{T-1} \sum_{i=1}^{I} \sum_{m=1}^{M} p(h_t = a_i, s_t = b_m \mid v_1^T, y_1^T, \theta^{(w-1)}) \log p(v_t = c_l \mid h_t = a_i,\theta)$$

$$= \sum_{t=1}^{T-1} \sum_{i=1}^{I} \sum_{m=1}^{M} \beta_{tim}^{(w-1)} \alpha_{tim}^{(w-1)} \log \gamma_{li}. \qquad (4.21)$$

When $t = T$, $q^{(w)}(h_T,s_T) = p(h_T,s_T \mid v_1^T, y_1^T, \theta^{(w-1)})$ is the filtering posterior probability density function, so by (4.3), equation (4.20) can be rewritten as

$$\sum_{i=1}^{I} \sum_{m=1}^{M} p(h_T = a_i, s_T = b_m \mid v_1^T, y_1^T, \theta^{(w-1)}) \log p(v_T = c_l \mid h_T = a_i,\theta)$$

$$= \sum_{i=1}^{I}\sum_{m=1}^{M} \alpha_{Tim}^{(w-1)} \log \gamma_{li} . \tag{4.22}$$

Hence , when $1 \le t \le T$ , (4.20) can be obtained as

$$\sum_{H}\sum_{S} q^{(w)}(H,S) \sum_{t=1}^{T} \log p(v_t = c_l \,|\, h_t = a_i, \theta)$$

$$= \sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{m=1}^{M} p(h_t = a_i, s_t = b_m \,|\, v_1^T, y_1^T, \theta^{(w-1)}) \log p(v_t = c_l \,|\, h_t = a_i, \theta)$$

$$= \sum_{t=1}^{T-1}\sum_{i=1}^{I}\sum_{m=1}^{M} \beta_{tim}^{(w-1)} \alpha_{tim}^{(w-1)} \log \gamma_{li} + \sum_{i=1}^{I}\sum_{m=1}^{M} \alpha_{Tim}^{(w-1)} \log \gamma_{li}$$

$$\triangleq \sum_{t}\sum_{i}\sum_{m} q_{tim}^{(w)} \log \gamma_{li} . \tag{4.23}$$

In equation (4.19), **the second term** can be obtained as follows:

$$\sum_{H}\sum_{S} q^{(w)}(H,S) \sum_{t=1}^{T} \log p(y_t \,|\, h_t, \theta)$$

$$= \sum_{t=1}^{T}\sum_{H}\sum_{S} q^{(w)}(H,S) \log p(y_t \,|\, h_t, \theta)$$

$$= \sum_{t=1}^{T}\sum_{h_t}\sum_{s_t} \log p(y_t \,|\, h_t, \theta) \sum_{\{H\}_{-t},\{S\}_{-t}} q^{(w)}(H,S)$$

$$= \sum_{t=1}^{T}\sum_{h_t}\sum_{s_t} \log p(y_t \,|\, h_t, \theta) q^{(w)}(h_t, s_t)$$

$$= \sum_{t=1}^{T}\sum_{h_t}\sum_{s_t} p(h_t, s_t \,|\, v_1^T, y_1^T, \theta^{(w-1)}) \log p(y_t \,|\, h_t, \theta) . \tag{4.24}$$

According to equation (4.1) and the same analysis process with the first term, when $1 \le t \le T$ , equation (4.24) can be obtained as

$$\sum_{H}\sum_{S} q^{(w)}(H,S) \sum_{t=1}^{T} \log p(y_t \,|\, h_t = a_i, \theta)$$

$$= \sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{m=1}^{M} p(h_t = a_i, s_t = b_m \,|\, v_1^T, y_1^T, \theta^{(w-1)}) \log p(y_t \,|\, h_t = a_i, \theta)$$

$$= \sum_{t=1}^{T-1}\sum_{i=1}^{I}\sum_{m=1}^{M} \beta_{tim}^{(w-1)} \alpha_{tim}^{(w-1)} \log g_i(y_t) + \sum_{i,m} \alpha_{Tim}^{(w-1)} \log g_i(y_T)$$

$$\triangleq \sum_{t}\sum_{i}\sum_{m} q_{tim}^{(w)} \log g_i(y_t) . \tag{4.25}$$

In equation (4.19), **the third term** can be obtained as follows:

$$\sum_{H,S} q^{(w)}(H,S) \sum_{t'=2}^{T} \log p(h_{t'} \left| s_{t'}, h_{t'-1}, \theta \right.)$$

$$= \sum_{t'=2}^{T} \sum_{h_{t'}, s_{t'}, h_{t'-1}, s_{t'-1}} q^{(w)}(h_{t'}, s_{t'}, h_{t'-1}, s_{t'-1}) \log p(h_{t'} \left| s_{t'}, h_{t'-1}, \theta \right.)$$

So when $2 \le t' \le T$,

$$\sum_{H,S} q^{(w)}(H,S) \sum_{t'=2}^{T} \log p(h_{t'} = a_i \left| h_{t'-1} = a_j, s_{t'} = b_m, \theta \right.)$$

$$= \sum_{t'=2}^{T}\sum_{i}\sum_{j}\sum_{m}\sum_{n} q^{(w)}(h_{t'} = a_i, s_{t'} = b_m, h_{t'-1} = a_j, s_{t'-1} = b_n) \log p(h_{t'} = a_i \left| h_{t'-1} = a_j, s_{t'} = b_m, \theta \right.)$$

$$= \sum_{t'=2}^{T-1}\sum_{i}\sum_{j}\sum_{m}\sum_{n} \beta_{t'im}^{(w-1)} \frac{\gamma_{li}^{(w-1)} g_i^{(w-1)}(y_{t'})}{\sum_{i'}\sum_{m'} \alpha_{t'i'm'}^{\prime(w-1)}} f_{ijm}^{(w-1)} p_{mn}^{(w-1)} \alpha_{(t'-1)jn}^{(w-1)} \log f_{ijm} +$$

$$\frac{\gamma_{li}^{(w-1)} g_i^{(w-1)}(y_T)}{\sum_{i'}\sum_{m'} \alpha_{Ti'm'}^{\prime(w-1)}} f_{ijm}^{(w-1)} p_{mn}^{(w-1)} \alpha_{(T-1)jn}^{(w-1)} \log f_{ijm}$$

$$\triangleq \sum_{t'}\sum_{i}\sum_{j}\sum_{m}\sum_{n} q_{t'ijmn}^{(w)} \log f_{ijm} \tag{4.26}$$

Further details are shown in Appendix C.

In equation (4.19), **the fourth term** can be obtained as follows:

$$\sum_{H}\sum_{S} q^{(w)}(H,S) \sum_{t'=2}^{T} \log p(s_{t'} \left| s_{t'-1}, \theta \right.)$$

$$= \sum_{t'=2}^{T} \sum_{h_{t'}, h_{t'-1}, s_t, s_{t-1}} q^{(w)}(h_{t'}, h_{t'-1}, s_{t'}, s_{t'-1}) \log p(s_{t'} \left| s_{t'-1}, \theta \right.) . \tag{4.27}$$

The detailed process is the same as the third term, when $2 \le t' \le T$, so (4.27) can be obtained as

$$\sum_H \sum_S q^{(w)}(H,S) \sum_{t'=2}^{T} \log p(s_{t'} = b_m \,|\, s_{t'-1} = b_n, \theta)$$

$$= \sum_{t'=2}^{T} \sum_i \sum_j \sum_m \sum_n q^{(w)}(h_{t'} = a_i, s_{t'} = b_m, h_{t'-1} = a_j, s_{t'-1} = b_n) \log p(s_{t'} = b_m \,|\, s_{t'-1} = b_n, \theta)$$

$$= \sum_{t'=2}^{T-1} \sum_i \sum_j \sum_m \sum_n \beta_{t'im}^{(w-1)} \frac{\gamma_{li}^{(w-1)} g_i^{(w-1)}(y_{t'})}{\sum_{i'} \sum_{m'} \alpha_{t'i'm'}'^{(w-1)}} f_{ijm}^{(w-1)} p_{mn}^{(w-1)} \alpha_{(t'-1)jn}^{(w-1)} \log p_{mn} +$$

$$\frac{\gamma_{li}^{(w-1)} g_i^{(w-1)}(y_T)}{\sum_{i'} \sum_{m'} \alpha_{Ti'm'}'^{(w-1)}} f_{ijm}^{(w-1)} p_{mn}^{(w-1)} \alpha_{(T-1)jn}^{(w-1)} \log p_{mn}$$

$$\triangleq \sum_{t'} \sum_i \sum_j \sum_m \sum_n q_{t'ijmn}^{(w)} \log p_{mn} \,. \tag{4.28}$$

In equation (4.19), **the fifth term** can be obtained as follows

$$\sum_H \sum_S q^{(w)}(H,S) \log p(h_1 \,|\, s_1, \theta) = \sum_{h_1} \sum_{s_1} \sum_{\{H\}_{-1}} \sum_{\{S\}_{-1}} q^{(w)}(H,S) \log p(h_1 \,|\, s_1, \theta)$$

$$= \sum_{h_1} \sum_{s_1} q^{(w)}(h_1, s_1) \log p(h_1 \,|\, s_1, \theta)$$

$$= \sum_{h_1} \sum_{s_1} p(h_1, s_1 \,|\, v_1^T, y_1^T, \theta^{(w-1)}) \log p(h_1 \,|\, s_1, \theta) \tag{4.29}$$

and by the equation (4.5), (4.29) can be obtained as

$$\sum_H \sum_S q^{(w)}(H,S) \log p(h_1 = a_i \,|\, s_1 = b_m, \theta)$$

$$= \sum_i \sum_m p(h_1 = a_i, s_1 = b_m \,|\, v_1^T, y_1^T, \theta^{(w-1)}) \log p(h_1 = a_i \,|\, s_1 = b_m, \theta)$$

$$= \sum_i \sum_m \beta_{1im}^{(w-1)} \alpha_{1im}^{(w-1)} \log f_{im}'$$

$$\triangleq \sum_i \sum_m q_{1im}^{(w)} \log f_{im}' \,. \tag{4.30}$$

In equation (4.19), **the sixth term** can be obtained as follows:

$$\sum_H \sum_S q^{(w)}(H,S) \log p(s_1 \,|\, \theta) = \sum_{h_1} \sum_{s_1} \sum_{\{H\}_{-1}} \sum_{\{S\}_{-1}} q^{(w)}(H,S) \log p(s_1 \,|\, \theta)$$

$$= \sum_{h_1} \sum_{s_1} q^{(w)}(h_1, s_1) \log p(s_1 \,|\, \theta)$$

$$= \sum_{h_1} \sum_{s_1} p(h_1, s_1 \,|\, v_1^T, y_1^T, \theta^{(w-1)}) \log p(s_1 \,|\, \theta) \tag{4.31}$$

and by the equation (4.6), (4.31) can be obtained as

$$\sum_H \sum_S q^{(w)}(H,S) \log p(s_1 = b_m \,|\, \theta)$$

$$= \sum_i \sum_m p(h_1 = a_i, s_1 = b_m \,|\, v_1^T, y_1^T, \theta^{(w-1)}) \log p(s_1 = b_m \,|\, \theta)$$

$$= \sum_i \sum_m \beta_{1im}^{(w-1)} \alpha_{1im}^{(w-1)} \log p'_m$$

$$\triangleq \sum_i \sum_m q_{1im}^{(w)} \log p'_m . \tag{4.32}$$

According to (4.23), (4.25), (4.26), (4.28), (4.30) and (4.32), $Q(\theta, \theta^{(w-1)})$ can be rewritten as

$$Q(\theta, \theta^{(w)}) = \sum_t \sum_i \sum_m q_{tim}^{(w)} \log \gamma_{li} + \sum_t \sum_i \sum_m q_{tim}^{(w)} \log g_i(y_t) + \sum_{t'} \sum_i \sum_j \sum_m \sum_n q_{t'ijmn}^{(w)} \log f_{ijm}$$
$$+ \sum_{t'} \sum_i \sum_j \sum_m \sum_n q_{t'ijmn}^{(w)} \log p_{mn} + \sum_i \sum_m q_{1im}^{(w)} \log f'_{im} + \sum_i \sum_m q_{1im}^{(w)} \log p'_m .$$

**4.2.3.4 Maximization step** The aim of M-step is to establish the Lagrange function and then derivation of each parameter in the function. The Lagrange function is given as follows:

$$\begin{aligned} H(\theta, \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5) = & \sum_t \sum_i \sum_m q_{tim}^{(w)} \log \gamma_{li} + \sum_t \sum_i \sum_m q_{tim}^{(w)} \log g_i(y_t) + \\ & \sum_{t'} \sum_i \sum_j \sum_m \sum_n q_{t'ijmn}^{(w)} \log f_{ijm} + \sum_{t'} \sum_i \sum_j \sum_m \sum_n q_{t'ijmn}^{(w-1)} \log p_{mn} + \\ & \sum_i \sum_m q_{1im}^{(w)} \log f'_{im} + \sum_i \sum_m q_{1im}^{(w)} \log p'_m + \\ & \lambda_1 (\sum_l \gamma_{li} - 1) + \lambda_2 (\sum_i f_{ijm} - 1) + \lambda_3 (\sum_m p_{mn} - 1) + \\ & \lambda_4 (\sum_i f'_{im} - 1) + \lambda_5 (\sum_m p' - 1) \end{aligned} \tag{4.33}$$

1. In equation (4.33), calculating $\frac{\partial H(\theta, \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5)}{\partial \gamma_{li}}$, and then setting $\frac{\partial H(\theta, \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5)}{\partial \gamma_{li}} = 0$, where $i$ is fixed, it is easy get as

$$\frac{\sum_t \sum_m q_{tim}^{(w)}}{\gamma_{li}} + \lambda_1 = 0 \;\Rightarrow\; \sum_t \sum_m q_{tim}^{(w)} + \gamma_{li} \lambda_1 = 0 . \tag{4.34}$$

Since $\sum_l \gamma_{li} = 1$, then $\sum_l \sum_t \sum_m q_{tim}^{(w)} + \sum_l \gamma_{li} \lambda_1 = 0$, so

$$\lambda_1 = -\sum_l \sum_t \sum_m q_{tim}^{(w)} \tag{4.35}$$

and putting (4.35) into (4.34), then

$$\gamma_{li}^{(w)} = \frac{\sum_t \sum_m q_{tim}^{(w)}}{\sum_l \sum_{t'} \sum_{m'} q_{ti'm'}^{(w)}}. \tag{4.36}$$

By (4.23)

$$\sum_t \sum_m q_{tim}^{(w)} = \sum_{t=1}^{T-1} \sum_{m=1}^{M} \beta_{tim}^{(w-1)} \alpha_{tim}^{(w-1)} \log \gamma_{li} + \sum_{m=1}^{M} \alpha_{Tim}^{(w-1)}.$$

2. In equation (4.33), calculating $\frac{\partial H(\theta, \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5)}{\partial f_{ijm}}$ and then setting $\frac{\partial H(\theta, \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5)}{\partial f_{ijm}} = 0$, where $i, j, m$ are fixed, we obtain as

$$\frac{\sum_{t'} \sum_n q_{t'ijmn}^{(w)}}{f_{ijm}} + \lambda_2 = 0 \Rightarrow \sum_{t'} \sum_n q_{t'ijmn}^{(w)} + \lambda_2 f_{ijm} = 0. \tag{4.37}$$

Since $\sum_i f_{ijm} = 1$, then $\sum_i \sum_{t'} \sum_n q_{t'ijmn}^{(w)} + \lambda_2 \sum_i f_{ijm} = 0$, so

$$\lambda_2 = -\sum_i \sum_{t'} \sum_n q_{t'ijmn}^{(w)} \tag{4.38}$$

and putting (4.38) into (4.37), then

$$f_{ijm}^{(w)} = \frac{\sum_{t'} \sum_n q_{t'ijmn}^{(w)}}{\sum_i \sum_{t''} \sum_{n'} q_{t''ijmn'}^{(w)}} \tag{4.39}$$

by (4.26), where

$$\sum_{t'} \sum_n q_{t'ijmn}^{(w)} = \sum_{t'=2}^{T-1} \sum_{n=1}^{M} \beta_{t'im}^{(w-1)} \frac{\gamma_{li}^{(w-1)} g_i^{(w-1)}(y_{t'})}{\sum_{i'=1}^{I} \sum_{m'=1}^{M} \alpha'^{(w-1)}_{t'i'm'}} f_{ijm}^{(w-1)} p_{mn}^{(w-1)} \alpha_{(t'-1)jn}^{(w-1)} +$$

$$\frac{\gamma_{li}^{(w-1)} g_i^{(w-1)}(y_T)}{\sum_{i'=1}^{I} \sum_{m'=1}^{M} \alpha'^{(w-1)}_{Ti'm'}} f_{ijm}^{(w-1)} p_{mn}^{(w-1)} \alpha_{(T-1)jn}^{(w-1)}.$$

3. In equation (4.33), calculating $\frac{\partial H(\theta, \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5)}{\partial p_{mn}}$ and then setting $\frac{\partial H(\theta, \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5)}{\partial p_{mn}} = 0$, where $m, n$ are fixed, we can obtain as

$$\frac{\sum_{t'}\sum_{i}\sum_{j} q_{t'ijmn}^{(w)}}{p_{mn}} + \lambda_3 = 0 \;\Rightarrow\; \sum_{t'}\sum_{i}\sum_{j} q_{t'ijmn}^{(w)} + \lambda_3 p_{mn} = 0 . \tag{4.40}$$

Since $\sum_{m} p_{mn} = 1$, then $\sum_{m}\sum_{t'}\sum_{i}\sum_{j} q_{t'ijmn}^{(w)} + \sum_{m} \lambda_3 p_{mn} = 0$, so

$$\lambda_3 = -\sum_{m}\sum_{t'}\sum_{i}\sum_{j} q_{t'ijmn}^{(w)} . \tag{4.41}$$

Putting (4.41) into (4.40), then we can get

$$p_{mn}^{(w)} = \frac{\sum_{t'}\sum_{i}\sum_{j} q_{t'ijmn}^{(w)}}{\sum_{m}\sum_{t''}\sum_{i'}\sum_{j'} q_{t''i'j'mn}^{(w)}} \tag{4.42}$$

by (4.28), where

$$\sum_{t'}\sum_{i}\sum_{j} q_{t'ijmn}^{(w)} = \sum_{t'=2}^{T-1}\sum_{i=1}^{I}\sum_{j=1}^{I} \beta_{t'im}^{(w-1)} \frac{\gamma_{li}^{(w-1)} g_i^{(w-1)}(y_{t'})}{\sum_{i'}\sum_{m'} \alpha_{t'i'm'}^{\prime(w-1)}} f_{ijm}^{(w-1)} p_{mn}^{(w-1)} \alpha_{(t'-1)jn}^{(w-1)} +$$

$$\frac{\gamma_{li}^{(w-1)} g_i^{(w-1)}(y_T)}{\sum_{i'}^{I}\sum_{m'}^{M} \alpha_{Ti'm'}^{\prime(w-1)}} f_{ijm}^{(w-1)} p_{mn}^{(w-1)} \alpha_{(T-1)jn}^{(w-1)} .$$

4. In equation (4.33), calculating $\frac{\partial H(\theta, \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5)}{\partial f_{im}'}$ and then setting $\frac{\partial H(\theta, \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5)}{\partial f_{im}'} = 0$, where $i, m$ are fixed, we can obtain as

$$\frac{q_{1im}^{(w)}}{f_{im}'} + \lambda_4 = 0 \;\Rightarrow\; q_{1im}^{(w)} + \lambda_4 f_{im}' = 0 . \tag{4.43}$$

Since $\sum_{i} f_{im}' = 1$, then $\sum_{i} q_{1im}^{(w)} + \sum_{i} \lambda_4 f_{im}' = 0$, so

$$\lambda_4 = -\sum_{i} q_{1im}^{(w)} \tag{4.44}$$

and putting (4.44) into (4.43), then we can get

$$f_{im}'^{(w)} = \frac{q_{1im}^{(w)}}{\sum_{i'} q_{1i'm}^{(w)}} \tag{4.45}$$

by (4.30), where

$$q_{1im}^{(w)} = \beta_{1im}^{(w-1)} \alpha_{1im}^{(w-1)} .$$

5. In equation (4.33), calculating $\frac{\partial H(\theta,\lambda_1,\lambda_2,\lambda_3,\lambda_4,\lambda_5)}{\partial p'_m}$ and then setting $\frac{\partial H(\theta,\lambda_1,\lambda_2,\lambda_3,\lambda_4,\lambda_5)}{\partial p'_m}=0$, where $m$ is fixed, we can obtain as

$$\frac{\sum_i q_{1im}^{(w)}}{p'_m}+\lambda_5=0 \Rightarrow \sum_i q_{1im}^{(w)}+\lambda_5 p'_m=0. \tag{4.46}$$

Since $\sum_m p'_m=1$, then $\sum_m\sum_i q_{1im}^{(w)}+\sum_m \lambda_5 p'_m=0$, so

$$\lambda_5=-\sum_m\sum_i q_{1im}^{(w)} \tag{4.47}$$

and putting (4.47) into (4.46), then we can get

$$p'^{(w)}_m=\frac{\sum_i q_{1im}^{(w)}}{\sum_m\sum_{i'} q_{1i'm}^{(w)}} \tag{4.48}$$

by (4.32), where $\sum_i q_{1im}^{(w)}=\sum_{i=1}^{I}\beta_{1im}^{(w-1)}\alpha_{1im}^{(w-1)}$.

## 4.3 Bayesian inference of the E-MSSV-II model

In this section, we consider stochastic inference algorithm to estimate parameters in the E-MSSV-II model.

### 4.3.1 The description of the E-MSSV-II

In order to avoid over complexity, we start with a simple E-MSSV-II model. Firstly, by analyzing the relationship between return and volume, the new log return can be shown as a function of the volume and the fluctuation range of volume in a stock. Secondly, we define regime variables $s_t$ following a two-state first order Markov process, i.e. high-volatility and low-volatility states, and we can assume that only the mean of volatility shifts depending on the state, i.e. $\mu_{s_t}=\alpha+\beta s_t$. Finally, the transition

probability matrix of first order Markov process is presented as a dimensional matrix. So the E-MSSV-II model can be summarized as follows:

$$y_t' = f(\mathrm{y}_t, v_t) = \exp(h_t/2)\varepsilon_t \tag{4.49}$$

$$h_t = \alpha + \beta s_t + \phi h_{t-1} + \sigma_\eta \eta_t \tag{4.50}$$

$$p = \begin{pmatrix} p_{11} & 1-p_{22} \\ 1-p_{11} & p_{22} \end{pmatrix} \tag{4.51}$$

where $y_t$ explains the return data of a stock, $v_t$ is the state of volume of a stock, $y_t'$ is the function of the volume and log return of price. $h_t$ is the unobserved log volatility of a stock, $s_t \in \{1,2\}$, $\varepsilon_t, \eta_t \sim i.i.d.N(0,1)$, $h_0 \sim N(\alpha + \beta s_1, \sigma_\eta^2/(1-\phi^2))$. The transition probabilities are defined by $p_{11} = p(s_t = 1 | s_{t-1} = 1)$ an $p_{22} = p(s_t = 2 | s_{t-1} = 2)$. $p_{11}$ and $p_{22}$ represent the probabilities of the keeping in the low-volatility regime and high-volatility regime respectively. Hence in the E-MSSV model, the set of parameters is denoted as $\theta = \{\alpha, \beta, \phi, \sigma_\eta^2, p_{11}, p_{22}\}$.

In the E-MSSV-II model, let $y_1'^t \triangleq \{y_1', ..., y_t'\}$, $x_t = \{h_t, s_t\}$ plays the role of the latent state vector. Moreover, in order to simplify the inference process of parameters and latent variables, we make the following independence assumptions: (i) $y_{t+1}'$ is conditionally independent from $y_1'^t$ given $x_{t+1}$; (ii) $x_{t+1}$ is conditionally independent from $x_t$ given $y_1'^t$.

According to Bayesian rule, the conditional probability density function of $x_{t+1}$ is given by

$$p(x_{t+1}, \theta | y_1'^{t+1}) = \frac{p(y_{t+1}' | x_{t+1}, \theta)\, p(x_{t+1}, \theta | y'^t)}{p(y_{t+1}' | y_1'^t)} \tag{4.52}$$

in the equation (4.52), the distribution for $x_{t+1}$ over $y_1'^t$ is given by

$$p(x_{t+1},\theta\left|y_1'^t\right.)=\int p(x_{t+1},\theta\left|x_t\right.)p(x_t\left|y_1'^t\right.)dx_t \tag{4.53}$$

and the distribution for $y_{t+1}$ over $y_1'^t$ is given by

$$p(y_{t+1}\left|y_1'^t\right.)=\int p(y_{t+1}\left|x_{t+1}\right.)p(x_{t+1}\left|y_1'^t\right.)dx_{t+1}.$$

Hence, the posterior distribution for $x_{t+1}$ over $y_1'^{t+1}$ is proportional of the numerator of the right hand side of the equation (4.52), i.e.,

$$p(x_{t+1},\theta\left|y_1'^{t+1}\right.)\propto p(y_{t+1}\left|x_{t+1},\theta\right.)\int p(x_{t+1},\theta\left|x_t\right.)p(x_t\left|y_1'^t\right.)dx_t \tag{4.54}$$

### 4.3.2 Auxiliary particle filter with known parameters

Particle filters (Pitt and Shephard, 1999) are the class of simulation filters that recursively approximate the posterior distribution of $x_t$, i.e., $p(x_t\left|y_1'^t\right.)$ by particles $x_t^{(1)},...,x_t^{(N)}$ with discrete probability mass of $\omega_t^{(1)},...,\omega_t^{(N)}$, and the relation about them will be denoted by

$$\left\{x_t^{(j)},\omega_t^{(j)}\right\}_{j=1}^{N}\sim p(x_t\left|y_1'^t\right.)\approx\sum_{j=1}^{N}\omega_t^{(\mathrm{j})}\delta(x_t-x_t^{(j)}),$$

where $\omega_t^{(j)}$ are suitable weights and $\delta(\cdot)$ is an Dirac $\delta-$function.

In this section, we assume the parameter vector $\theta$ is known, and we will omit the $\theta$ in all probability distributions for ease of understanding. In some literatures, $\omega_t^{(j)}$ are assumed to equal $1/N$. Throughout, $N$ is taken to be very large. If $N\to\infty$, then the particles can be considered to better approximate the density distribution function $p(x_t\left|y_1'^t\right.)$.

Particle filters treat the discrete variables generated by the particles as the true filtering density. This allows us to introduce an approximation to the posterior distribution $p(x_{t+1}\left|y_1'^t\right.)$ in equation (4.53) given by

$$\widetilde{p}(x_{t+1}\left|y_1'^t\right.)=\sum_{j=1}^{N}p(x_{t+1}\left|x_t^{(j)}\right.)\omega_t^{(j)} \tag{4.55}$$

which is called "empirical prediction density". Combining (4.55) with (4.54), by applying the Chapman-Kolmogorov equation it is also possible to obtain an approximation of the posterior distribution of $x_{t+1}$ at time t+1 given by

$$\tilde{p}(x_{t+1}\left|y_1'^{t+1}\right.)=p(y_{t+1}'\left|x_{t+1}\right.)\sum_{j=1}^{N}p(x_{t+1}\left|x_t^{(j)}\right.)\omega_t^{(j)} \tag{4.56}$$

which is called "empirical filtering density" as an approximation to the true filtering density, that is equation (4.55).

In order to complete the filtering process, generically, particle filters sample from this density to produce new particles $x_{t+1}^{(1)},...,x_{t+1}^{(N)}$ with weight $\omega_{t+1}^{(1)},...,\omega_{t+1}^{(N)}$, i.e. $\left\{x_{t+1}^{(j)},\omega_{t+1}^{(j)}\right\}_{j=1}^{N}\sim\widetilde{p}(x_{t+1}\left|y_1'^{t+1}\right.)$. This procedure can then be iterated through the data. Smith and Gelfand (1992) suggest a sampling importance resample filter by using the approximate prior distribution of equation (4.52) as the importance function.

Since Pitt and Shephard (1999) develop the auxiliary particle filter, it looks at the empirical filtering density, i.e. equation 4.56, as a mixture of N distributions and introduces a latent indicator for the mixture components, $p(x_{t+1},k)\propto p(y_{t+1}'\left|x_{t+1}\right.)p(x_{t+1}\left|x_t^{(k)}\right.)\omega_t^{(k)}$, leading to a sequential plan that first samples $k^l$ from $p(y_{t+1}\left|e_{t+1}^{(j)}\right.)\omega_t^{(j)}$, with $e_{t+1}^{(j)}$ representing a guess, such as the mean or some other likely value associated with $p(x_{t+1}\left|x_t^{(j)}\right.)$, and then samples $x_{t+1}^{(l)}$ from $p(x_{t+1}\left|x_t^{(k^l)}\right.)$. Hence the important function is

$$g(x_{t+1},k)\propto p(y_{t+1}'\left|e_{t+1}\right.)p(x_{t+1}\left|x_t^{(k)}\right.)\omega_t^{(k)},$$

these weights $\omega_{t+1}^{(l)}\propto p(y_{t+1}'\left|x_{t+1}^{(l)}\right.)/\,p(y_{t+1}'\left|e_{t+1}^{(k^l)}\right.)$ and $\left\{x_{t+1}^{(l)},\omega_{t+1}^{(l)}\right\}_{l=1}^{N}\sim\widetilde{p}(x_{t+1}\left|y_1'^{t+1}\right.)$.

### 4.3.3 Auxiliary particle filters with unknown parameters

In addition to tracking the unobserved state variables, Liu and West (2001) propose to approximate the posterior distribution $p(\theta|y_1^{'t})$ with a particle set $\{x_t^{(j)}, \theta_t^{(j)}, \omega_t^{(j)}\}$ and to reconstruct the parameter posterior distribution at time t+1 through a Gaussian kernel density estimation.

So when we update $\theta$, the problem can be seen as a Bayesian sequential learning process where the goal is to update the following posterior density:

$$p(x_{t+1}, \theta_{t+1}|y_1^{'t+1}) \propto p(y'_{t+1}|x_{t+1}, \theta_{t+1}) p(x_{t+1}, \theta_{t+1}|y_1^{'t}) p(\theta_{t+1}|y_1^{'t})$$

where $p(\theta_t|y_1^{'t}) \approx \sum_{j=1}^{N} \omega_t^{(j)} N(\theta_{t+1}|m_t^{(j)}, b^2 V_t)$. Combining with West (1993) mixture modeling ideas, we get $m_t^{(j)} = a\theta_t^{(j)} + (1-a)\overline{\theta}_t$, where $a = \frac{3\delta - 1}{2\delta}$, $b^2 = 1-a$, $\delta \in (0,1]$. $\overline{\theta}_t$ and $V_t$ are the mean and variance of the Monte Carlo approximation to $p(\theta_t|y_1^{'t})$, $\overline{\theta}_t = \sum_{j=1}^{N} \omega_t^{(j)} \theta_t^{(j)}$ and $V_t = \sum_{j=1}^{N} \omega_t^{(j)} (\theta_t^{(j)} - \overline{\theta}_t)(\theta_t^{(j)} - \overline{\theta}_t)'$. In order to get the Sequential Monte Carlo (SMC) filter for the E-MSSV-II model, we combine the auxiliary particle filters with the Kernel smoothing approximation.

**Table 4.2** The SMC filter for the E-MSSV-II model.

---

Given an initial set of particles $\{h_t^{(j)}, s_t^{(j)}, \theta_t^{(j)}, \omega_t^{(j)}\}_{j=1}^{N} \sim p(h_t, s_t, \theta|y_1^{'t})$

Step 1: for $j = 1, ..., N$, update

1. $\tilde{s}_{t+1}^{(j)} = \arg\max_{l=1,...,k} p(s_{t+1} = l | s_t = s_t^{(j)})$
2. $e_{t+1}^{(j)} = \mu_{\tilde{s}_{t+1}^{(j)}} + \phi_t^{(j)} h_t^{(j)}$

Step 2: for $l = 1, ..., N$

1. Sample $k^l$ from $\{1, ..., k\}$, with $p(k^{(l)}) \propto p(y'_{t+1}|e_{t+1}^{(l)}, m_t^{(l)}) \omega_t^{(l)}$

---

**Table 4.2** (Continued) The SMC filter for the E-MSSV-II model.

2. Sample $\theta_{t+1}^{(l)}$ from $N(m_t^{(k^l)}, b^2 V_t)$

3. Sample $s_{t+1}^{(l)}$ from $\{1,...,k\}$ with $p(s_{t+1}^{(l)}) = p(s_{t+1}^{(l)} \mid s_t^{(k^l)}))$

4. Sample $h_{t+1}^{(l)}$ from $p(h_{t+1} \mid h_t^{(k^l)}, s_{t+1}^{(l)}, \theta_{t+1}^{(l)})$

Step 3: for $l = 1,..., N$, compute new weights

$$\omega_{t+1}^{(l)} \propto p(y'_{t+1} \mid h_{t+1}^{(l)}) / p(y'_{t+1} \mid e_{t+1}^{(k^l)})$$

Step 4: $\{h_{t+1}^{(j)}, s_{t+1}^{(j)}, \theta_{t+1}^{(j)}, \omega_{t+1}^{(j)}\}_{j=1}^{N} \sim p(h_{t+1}, s_{t+1}, \theta \mid y'^{t+1}_1)$

## 4.4 Application

In this chapter, about the E-MSSV-I model, we use the EM algorithm to calculate the parameters, but it is more difficult to implement the algorithm by computer programming. According to Bishop (2006), if the non-stochastic inference algorithm cannot be used in experiment, we can choose stochastic inference algorithm directly. So in my research, we mainly analyze the E-MSSV-II model in the process of application. A sequential Monte Carlo (SMC) filter is applied in the E-MSSV-II model of one synthetic time series and one real dataset. The simulated examples are based on examples from Raggi (2005) and the real data obtained comes from the Dow Jones Industrial Average Index, i.e. DJI30.

### 4.4.1 Simulation study

To analyze the SMC filter performance, first we simulated 1000 observations from the E-MSSV model with two states in which the parameters are the following:

Volatility process: $\alpha = -2.5$, $\beta = 2$, $\phi = 0.2$, $\sigma_\eta^2 = 0.1$;

Markov process: $p_{11} = 0.99$, $p_{22} = 0.97$.

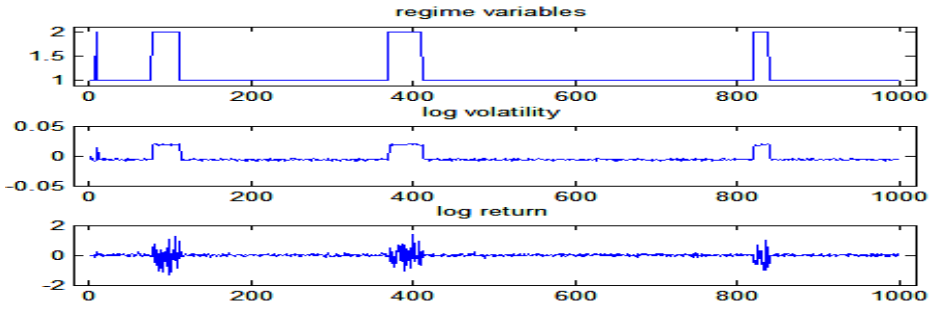


**Figure 4.2** The first graph exhibits the evolution of the true regime variables $s_t$, the second graph presents log-volatility $h_t$, the third graph shows simulated value of log-return $y_t$.

In this part, for verifying consistency of the parameters with the empirical findings, we define the prior distribution of parameter $\theta$ following Eraker et al (2003). We hypothesize the prior distributions as $\alpha \sim N(0,3)$, $\beta \sim Beta(20,1.5)$, $\phi \sim Beta(25,2.5)$, $\sigma_\eta^2 \sim IG(2,0.02)$, where IG indicate the inverse of a Gamma distribution. $p_{11}$ and $p_{22}$ is diagnosed by the transition probability matrix. The values of $a$ and b are determined by a discount factor $\delta$=0.86 which implies $a = 0.9186$ and $b = 0.2853$. For this experiment, basing on SMC filters algorithm, we use 25000 particles to approximate the parameters distribution in E-MSSV model. The results are reported on Figure 4.3.

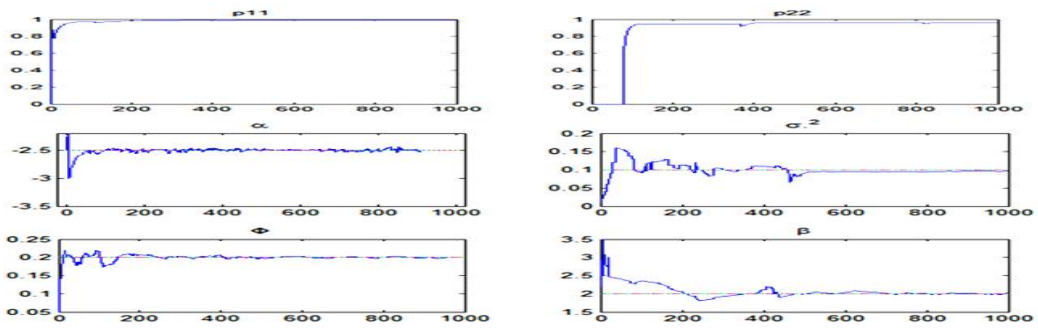


**Figure 4.3** The estimated parameters.

From Figure 4.3, we notice that the filter provides stable estimates for the parameters and the estimates are consistent with the true parameters. So we can use the filter algorithm in the real data.

### 4.4.2 Experiments with real data

We now apply the proposed algorithm to the DJI30 from 03/01/2007 to 09/12/2014 (2000 observations).

The data set has been downloaded from https://www.google.com/finance. $p_t$ be the DJI30 closed price, $y_t = \log(p_t / p_{t-1})$, $v_t$ be as the trading volume (10000 unit)

In the interest of understanding the relation of the closing price and volume, we have drawn Figures from 4.4 to 4.6. From Figure 4.4, we see that in the vicinity of 500, a large amplitude of closing price appears. In fact, the phenomenon also happens to the volume as shown in Figure 4.5. In Figure 4.6, the lower the closing price, the smaller the negative values of the volume. The higher the closing price, the bigger the negative values of the volume. So we can obtain the relationship of volume and closing price, in the part, since $y_t = \log(p_t / p_{t-1})$, we can define

$$y' = \log[(p_t - 5*10^{-3} v_t)/(p_{t-1} - 5*10^{-3} v_{t-1})] \quad .$$

And the new log return is shown in Figure 4.7.

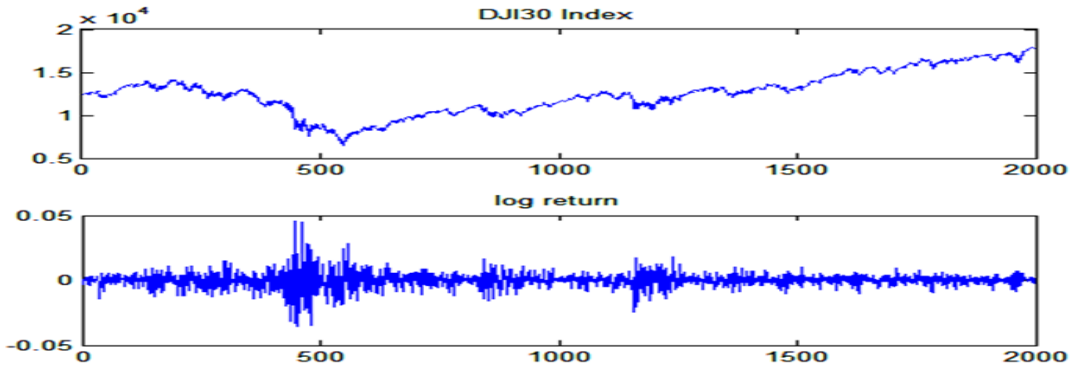


**Figure 4.4** The closed price of the DJI30 and the log return.

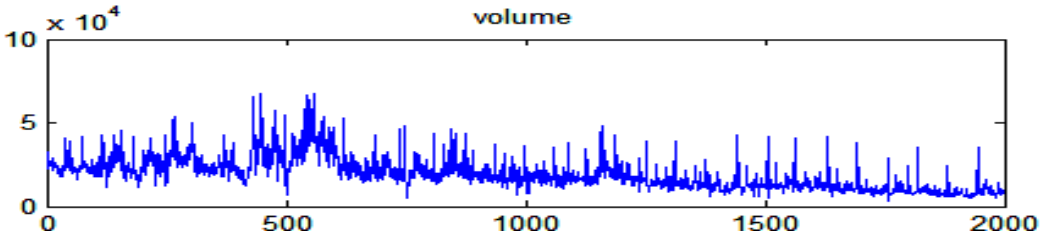


**Figure 4.5** The volumes of the DJI30.

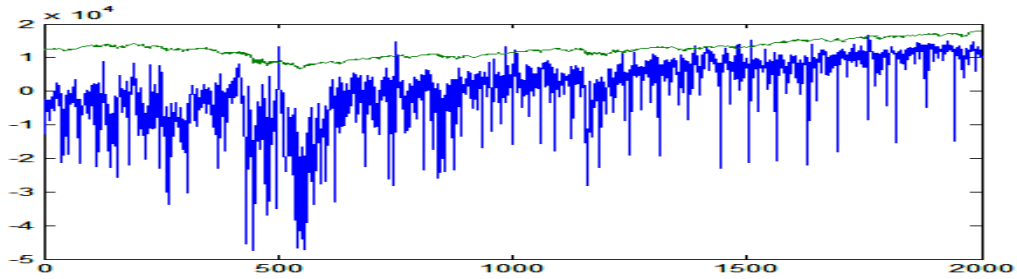


**Figure 4.6** The big line shows the negative values of the volume of the DJI30 and the small line is the closing price of the DJI30.

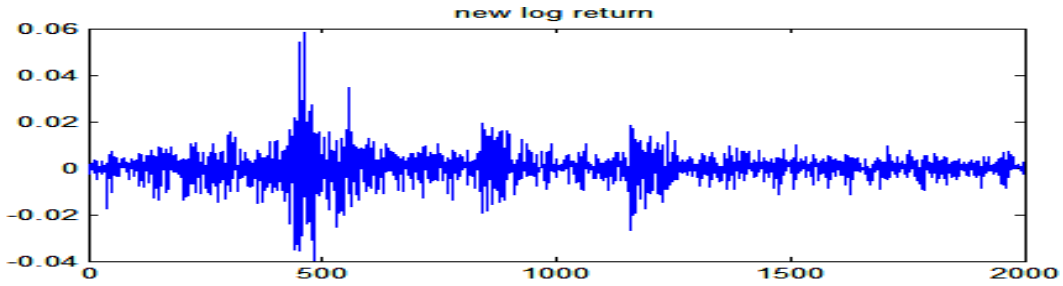


**Figure 4.7** The log return $y_t'$.

According to French et al. (1987), following the SMC filter algorithm and using MATLAB software, we estimate the E-MSSV-II model approximating the distribution through a cloud of 50,000 particles. The results of the parameter estimation in the E-MSSV-II model are summarized in Figure 4.8.

From the Mean Absolute Deviation (MAD), Mean Squared Error (MSE) and Mean Absolute Percentage Error (MAPE), which are shown in Table 4.3, about the E-MSSV-II model, MAD is 0.005679<0.006399, MSE is 0.000068<0.000113 and MAPE is 5.56796<6.47236, for the smaller the prediction error, the better the effect. The

result suggests that the E-MSSV-II model is more accurate in forecasting than the MSSV model. The real volatilities and estimated volatilities are shown in Figure 4.9.

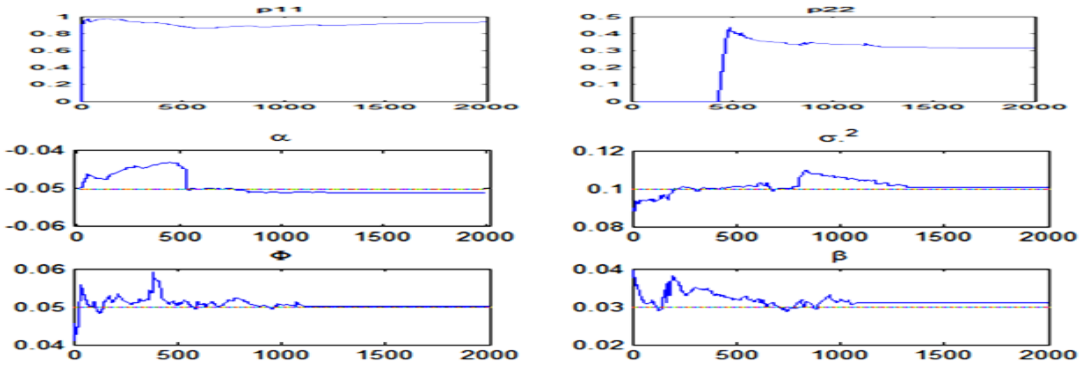


**Figure 4.8** The estimated parameters in the E-MSSV-II model.

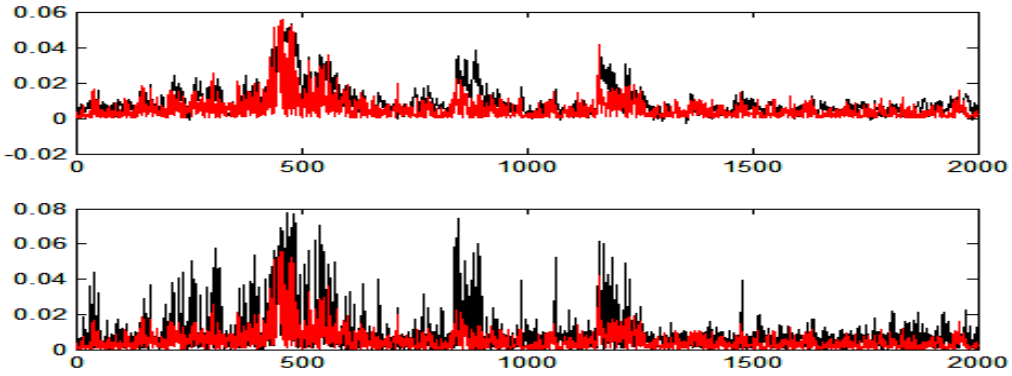

**Figure 4.9** The above picture represents the real volatilities and the volatilities estimated by the E-MSSV-II model. The below picture shows the real volatilities and volatilities estimated by the MSSV model (real volatilities are shown in red).

**Table 4.3** The estimated results by models.

| Model | MAD | MSE | MAPE |
|---|---|---|---|
| MSSV | 0.006399 | 0.000113 | 6.47236 |
| E-MSSV-II | 0.005679 | 0.000068 | 5.56796 |

# CHAPTER V

# CONCLUSION

This thesis mainly focuses on the analysis of volatility in the stock market. There are two classes of widely used models for capturing the stylized feature of volatility.

In chapter III, the study uses SSE380 prices to predict daily volatility changes in the stock market. Firstly, we use descriptive statistics to show that the index series has the feature of asymmetric zero mean and left side, which means it is not normally distributed. Secondly, we consider Augmented Dickey-Fuller Unit Root Tests to find that the series is a stationary time series. And then we use ARCH-Lagrange multiplier to detect SSE380 returns have ARCH effects.

Then the SSE380 index volatility is forecasted with the GARCH, EGARCH and TGARCH models. The volatility models are estimated with normal innovation and Student's t innovation distributions to find the effect of distribution selection on forecasting performance of the models. According to highest value of Log likelihood (LL) and smallest value of AIC and BIC, the results suggest that the GARCH with Student's t innovation model enables more accurate in forecasting than the EGARCH and TGARCH model. Under the evaluation criteria of the less functions of MSE and MAD, using the MCS test, the empirical results also show that the GARCH with Student's t innovation model is the best model.

In chapter IV, we use DJI30 prices to predict daily volatility changes in the stock

market. Firstly, we present two novel approaches, namely the E-MSSV model, including E-MSSV-I and E-MSSV-II model, based on adding volume to DJI30 prices to estimate volatilities. Secondly, in the E-MSSV-I model, Bayesian inference has been used to derive prediction, filtering and smoothing probability distribution function. Then the Expectation-Maximization method is used to estimate the variables and parameters. In the E-MSSV-II model, the Sequential Monte Carlo filter method is used to estimate the parameters. According to the value of MAD, MSE and MAPE, we can see that the E-MSSV-II model is more accurate in forecasting than the MSSV model.

The study can be used as an assistant tool in financial applications, such as describing the risk management and option pricing. Many significant meanings shown in the process of investing: firstly, it can help investors to make rational investment decisions before investing. Secondly, it can improve risk management of institutional and individual investors. Finally, it can assist with the development of relevant policies and help regulatory authorities to improve supervision.

But there are some restrictions on the GARCH-type models and the E-MSSV model. Since GARCH-type models are not novel models, so in the future work, we can structure other models to combine with them. It seems that hybrid models are more useful in extreme event forecasting as the structure of the volatility process becomes more complex. As far the E-MSSV models, in the E-MSSV-I model, the process of the Bayesian inference makes it very difficult to realize the algorithm because there are two discrete random variables in the input. Hence in the future work, we need to find a new way to deal with the case. In the E-MSSV-II model, only one dataset (DJI30) may be not

enough to prove that the new method is really better than existing methods, so in the future work, more datasets should be used in the model.

Moreover, it would also be worthwhile to compare multivariate GARCH-type models and stochastic volatility models in fitting volatility in the stock market.

# REFERENCES

# REFERENCES


Alsubaie, A. and Najand, M. (2009). Trading volume timevarying conditional volatility and asymmetric volatility spillover in the Saudi stock market. **Journal of Multinational Financial Management**. 19: 139–159.

Andersen, T. (1996). Return volatility and trading volume: an information flow interpretation of stochastic volatility. **Journal of Finance**. 51: 69–204.

Baillie R. , Chung C. and Tieslau M. (1996). Analysing inflation by the fractionally integrated ARFIMA-GARCH Model. **Journal of Applied Econometrics**. 74: 23–40.

Barber, D. (2012). **Bayesian Reasoning and Machine Learning**. Cambridge University Press, New York.

Basel, M. A. (2005). Predicting the volatility of the S&P-500 stock index via GARCH models: the role of asymmetries. **International Journal of Forecasting**. 21: 167–183.

Beal, M. J. (2003). **Variational Algorithms for Approximate Bayesian Inference**. PhD thesis, the University of London, UK.

Bishop, C. M. (1995). **Neural Networks for Pattern Recognition**. Oxford University Press, New York.

Bishop, C. M. (2006). **Pattern Recognition and Machine Learning**. Springer Science Press, New York.

Bollerslev, T. (1986). Generalised Autoregressive Conditional Heteroscedasticity. **Journal of Econometrics**. 31: 307–327.

Bonfil, G. S., Solis, J. F. and Rodarte, L. V. (2015). Volatility forecasting using support vector regression and a hybrid genetic algorithm, **Computational Economics**. 45: 111–133.

Brooks, C. (2002). **Introductory Econometrics for Finance**. Cambridge University Press, New York.

Carvalhoand, C. M. and Lopes, H. F. (2007). Simulation-based sequential analysis of Markov switching stochastic volatility models. **Computational Statistics and Data Analysis**. 51: 4526–4542.

Cathy, W. S. and Yang, M. J. (2006). The asymmetric reactions of mean and volatility of stock returns to domestic and international information based on a four-regime double-threshold GARCH model. **Physica**. A366: 401–418.

Chapman, S. (1928). On the Brownian displacements and thermal diffusion of grains suspended in a non uniform fluid. **Proceedings of the Royal Society of London. Series A**. 119: 34–60.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. **Journal of the Royal Statistical Society**. 34: 1–38.

Clark, T. E. and Ravazzolo, F. (2015). Macroeconomic forecasting performance under alternative specifications of time-varying volatility. **Journal of Applied Econometrics**. 30: 551–575.

Diebold, F. X. (1986). Modeling the persistence of conditional variances: A comment. **Econometric Reviews**. 5: 51–56.

Du, X. D., Yu, C. L. and Hayes, D. J. (2011). Speculation and volatility spillover in the crude oil and agricultural commodity markets: A Bayesian analysis. **Energy Economics**. 33: 497–503.

Dumitru, M. and Cristiana, T. (2010). Asymmetric conditional volatility models: empirical estimation and comparison of forecasting accuracy. **Romanian Journal of Economic Forecasting**. 3: 74–92.

Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of variance of UK inflation. **Econometrica**. 50: 987–1008.

Engle, R. F. (2001). GARCH 101: The use of ARCH/GARCH models in applied econometrics. **Journal of Economic Perspectives**. 15: 157–168.

Engle, R. F. and Patton, A. J. (2001). What good is a volatility model? **Quantitative Finance**. 1: 237–245.

Eraker, B. (2001). Mcmc analysis of diffusion models with application to finance. **Journal of Business and Economic Statistics**. 19: 177–197.

Eraker, B., Johannes, M. and Polson, N. (2003). The impact of jumps in volatility and returns. **Journal of Finance**. 58: 1269–1300.

Fama, E. (1965). The behavior of stock market prices. **Journal of Business**. 38: 34–105.

Franses, P. H. and Van, D. (1996). Forecasting stock market volatility using (nonlinear) GARCH models. **Journal of Forecast**. 15: 229–235.

French, K. R., Schwert, G. W. and Stambaugh, R. F. (1987). Expected stock returns and

volatility. **Journal of Financial Economics**. 19: 3–29.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2003). **Bayesian Data Analysis**. CRC Press, New York.

Glosten, L., Jagannathan, R. and Runkle, D. (1993). Relationship between the expected value and volatility of the nominal excess returns on stocks. **The Journal of Finance**. 48: 1779–1801.

Gordon, N. J., Salmond, D. J. and Smith, A. F. (1993). Novel approach to nonlinear/ non-Gaussian Bayesian state estimation. **IEEE Proceedings F - Radar and Signal Processing**. 140: 107–113.

Goutte, S. (2013). Pricing and hedging in stochastic volatility regime switching models. **Journal of Mathematical Finance**. 3: 70–80.

Grouard, M. D., Levy, S. and Lubochinsky, C. (2003). Stock market volatility from empirical data to their interpretation. **France Stability Review**. 2: 57–74.

Hamilton, J. D. (1994). **Time Series Analysis**. Pinceton University Press, New Jersey.

Hansen, P. R. (2003). **Asymptotic Tests of Composite Hypotheses.** Working Paper, Brown University Economics, No. 2003.

Hansen, P. R. and Lunde, A. (2005). A forecast comparison of volatility models: does anything beat a GARCH(1,1). **Journal of Applied Econometrics**. 20: 873–889.

Hansen, P. R. and Lunde, A. (2005). A test for superior predictive ability. **Journal of Business and Economic Statistics**. 23: 365–380.

Hansen, P. R., Luder, A. and James, M. N. (2011). The model confidence set. **Econometrica**. 79: 453–479.

Heitham, A. H, Hashem, A., Timothy, R. and Jacek, N. (2015). Forecasting the Jordanian stock index: modeling asymmetric volatility and distribution effects with in a GARCH framework. **Copernican Journal of Finance and Accounting**. 4: 9–26.

Hentsche, L. (1995). All in the family nesting symmetric andasymmetric GARCH models. **Journal of Financial Economics**. 39: 71–104.

Henton, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. **Review of Financial Studies**. 6: 327–343.

Hung, J. C. (2011). Applying a combined fuzzy systems and GARCH model to adaptively forecast stock market volatility. **Applied Soft Computing**, 11: 3938–3945.

Jeff, F. and Chris, K. (2011). Long memory in volatility and trading volume. **Journal of Banking and Finance**. 35: 1714–1726.

Jordan, M. and Jacobs, R. (1994). Hierarchical mixtures of experts and the EM algorithm. **Neural Computation**. 6: 181–214.

Joshua, C. C. and Angelia L. G. (2016). Modeling energy price dynamics: GARCH versus stochastic volatility. **Energy Economics**. 54: 182-189.

Karpoff, J. M. (1987). The relation between price changes and trading volume: a survey. **The Journal of Financial and Quantitative Analysis**. 22: 109–126.

Kastner, G. and Schnatter, S. F. (2014). Ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC estimation of stochastic volatility models.

**Computational Statistics and Data Analysis**. 76: 408–423.

Kim, S., Shephard, N. and Chib, S. (1998). Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models. **Review of Economic Studies**. 65: 361-393.

Kolmogorov, A. N. (1931). Ueber die analytischen Methoden in der Wahrscheinlichkeitsrechnung. **Mathematische Annalen**. 104: 415–458.

Lee, C. F. and Su, J. B. (2012). Alternative statistical distributions for estimating value at risk: theory and evidence. **Review of Quantitative Finance and Accounting**. 39: 309-331.

Li, M., Li, W. K. and Li, G. (2013). On mixture memory GARCH models. **Journal of Time Series Analysis**. 34: 606–624.

Liu, H. C. and Chiang, S. M. (2012). Forecasting the volatility of S&P depositary receipts using GARCH-type models under intraday range-based and return-based proxy measures, **International Review of Economics and Finance**. 22: 78–91.

Liu, J. and West, M. (2001). **Combined Parameter and State Estimation in Simulation based Filtering**. Springer, Berlin.

Mahajan, S. and Singh, B. (2009). The empirical investigation of relationship between return volume and volatility dynamics in Indian stock markets. **Eurasian Journal of Business and Economics**. 2: 113–137.

Mandelbrot, B. (1967). The valuation of some other speculative prices. **Journal of Business**. 40: 393–413.

Mike, K. P. and So, L. K. (1998). A stochastic volatility model with Markov switching. **Journal of Business and Economic Statistic**. 16: 244–253.

Mincer, J. and Zarnowitz, V. (1969). **The Evaluation of Economic Forecasts**. Columbia University Press, New York.

Murphy, K. P. (2012). **Machine Learning: a Probabilistic Perspective**. MIT Press, Cambridge.

Mutunga, T. N., Islam, A. S. and Orawo, L. A. (2015). Implementation of the estimating functions approach in asset returns volatility forecasting using first order asymmetric GARCH models. **Open Journal of Statistics**. 5: 455–463.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach, **Econometrica**. 59: 347–370.

Pan, Q. and Li, Y. (2013). Testing volatility persistence on Markov switching stochastic volatility models. **Economic Modelling**. 35: 45–50.

Peter, R. H., Asger, L. and Valerl, V. (2014). Realized beta GARCH: a multivariate GARCH model with realized measures of volatility. **Journal of Applied Econometrics**. 29: 774–799.

Pitt, M. and Shephard, N. (1999). Filtering via simulation: Auxiliary particle filters. **Journal of the American Statistical Association**. 94: 590–599.

Raggi, D. (2005). Adaptive MCMC methods for inference on affine stochastic volatility models with jumps. **Econometrics Journal**. 8: 235–250.

Rakesh, K. and Raj, S. D. (2010). Empirical analysis of conditional heteroskedasticity in time series of stock returns and asymmetric effect on volatility. **Global Business Review**. 11: 21–33.

Redner, R. and Walker, H. (1984). Mixture densities maximum likelihood and the em

algorithm. **Journal of Society for Industrial and Applied Mathematics**. 26: 195–235.

Rios, M. P. and Lopes, H. F. (2013). The Extended Liu and West Filter: Parameter Learning in Markov Switching Stochastic Volatility Models. **State-Space Models: Applications in Economics and Finance**. 1: 23–61.

Sadorsky, P. (2005). Stochastic volatility forecasting and risk management. **Applied Financial Economics**. 15: 121–135.

Ser, H. and Clive, W. J. (2003). Forecasting volatility in financial markets: a review. **Journal of Economic Literature**. 5: 478–539.

Shephard, N. (2005). **Stochastic Volatility: Selected Readings**. Oxford University Press, Oxford.

Shibata, M. and Watanabe, T. (2005). Bayesian analysis of a Markov switching stochastic volatility model. **Journal of the Japan Statistical Society**. 35: 205–219.

Smith, A. F. and Gelfand, A. E. (1992). Bayesian statistics without tears: A sampling resampling perspective. **The American Statistician**. 46: 84–88.

Smith, D. R. (2002). Markov switching and stochastic volatility diffusion models of short-term interest rates. **Journal of Business and Economic Statistics**. 20: 183–197.

Taylor, S. J. (1986). **Modelling Financial Time Series**. Wiley Press, New York.

Valle, C. A., Migonand, H. S. and Lopes, H. F. (2010). Bayesian modeling of financial returns: a relationship between volatility and trading volume. **Applied Stochastic Models in Business and Industry**. 26: 172–193.

Vo, M. T. (2009). Regime-switching stochastic volatility: evidence from the crude oil market. **Energy Economics**. 31: 779–788.

Werner, K., Anton F. and Marcel C. M. (2014). Volatility forecast using hybrid neural Network models. **Expert Systems with Applications**. 41: 2437–2442.

West, M. (1993). Approximating posterior distributions by mixture. **Journal of the Royal Statistical Society: Series B**. 55: 409–422.

Yu, C. and Zhang, J. (2011). Bayesian approach to Markov switching stochastic volatility model with jumps. **Communications in Statistics - Simulation and Computation**. 40: 1613–1626.

# APPENDICES

# APPENDIX A

# THE PROOF OF FILTERING PROBABILITY DISTRIBUTION FUNCTION

*Proof of proposition 6.* Since

$$p(h_T = a_i, s_T = b_m \left| y_1^T, v_1^T, \theta \right.) = p(h_T = a_i, s_T = b_m \left| y_1^{T-1}, y_T, v_1^{T-1}, v_T = c_l, \theta \right.)$$

by Bayesian rules , we can get

$$p(h_T = a_i, s_T = b_m \left| y_1^T, v_1^T, \theta \right.)$$

$$= \frac{p(y_T, v_T = c_l \left| h_T = a_i, s_T = b_m, y_1^{T-1}, v_1^{T-1}, \theta \right.) p(h_T = a_i, s_T = b_m \left| y_1^{T-1}, v_1^{T-1}, \theta \right.)}{p(y_T, v_T = c_l \left| y_1^{T-1}, v_1^{T-1}, \theta \right.)}$$

$$\propto p(y_T, v_T = c_l \left| h_T = a_i, s_T = b_m, y_1^{T-1}, v_1^{T-1}, \theta \right.) p(h_T = a_i, s_T = b_m \left| y_1^{T-1}, v_1^{T-1}, \theta \right.)$$

$$= p(y_T \left| h_T = a_i, s_T = b_m, y_1^{T-1}, v_T = c_l, v_1^{T-1}, \theta \right.) p(v_T = c_l \left| h_T = a_i, s_T = b_m, y_1^{T-1}, v_1^{T-1}, \theta \right.) \times$$

$$p(h_T = a_i, s_T = b_m \left| y_1^{T-1}, v_1^{T-1}, \theta \right.).$$

According to the CI rules, i.e. $y_T \perp s_T, y_1^{T-1}, v_T, v_1^{T-1} \left| h_T \right.$ and $v_T \perp s_T, y_1^{T-1}, v_1^{T-1} \left| h_T \right.$, we can obtain

$$p(h_T = a_i, s_T = b_m \left| y_1^T, v_1^T, \theta \right.)$$

$$\propto p(y_T \left| h_T = a_i, \theta \right.) p(v_T = c_l \left| h_T = a_i \, \theta \right.) p(h_T = a_i, s_T = b_m \left| y_1^{T-1}, v_1^{T-1}, \theta \right.).$$

From equations (4.1) and (4.3), we can get

$$p(h_T = a_i, s_T = b_m \left| y_1^T, v_1^T, \theta \right.)$$

$$\propto g_i(y_T) \gamma_{li} p(h_T = a_i, s_T = b_m \left| y_1^{T-1}, v_1^{T-1}, \theta \right.)$$

$$\propto g_i(y_T)\gamma_{li}\sum_j\sum_n p(h_{T-1}=a_j, s_{T-1}=b_n, h_T=a_i, s_T=b_m \,|\, y_1^{T-1}, v_1^{T-1}, \theta)$$

$$\propto g_i(y_T)\gamma_{li}\sum_j\sum_n [\, p(h_T=a_i \,|\, h_{T-1}=a_j, s_T=b_m, s_{T-1}=b_n, y_1^{T-1}, v_1^{T-1}, \theta)\times$$

$$p(s_T=b_m \,|\, h_{T-1}=a_j, s_{T-1}=b_n, y_1^{T-1}, v_1^{T-1}, \theta)\, p(h_{T-1}=a_j, s_{T-1}=b_n \,|\, y_1^{T-1}, v_1^{T-1}, \theta)\,].$$

By using the CI rules, i.e. $h_T \perp s_{T-1}, y_1^{T-1}, v_1^{T-1} \,|\, h_{T-1}, s_T$ and $s_T \perp h_{T-1}, y_1^{T-1}, v_1^{T-1} \,|\, s_{T-1}$, we can get

$$p(h_T=a_i, s_T=b_m \,|\, y_1^T, v_1^T, \theta)$$

$$\propto g_i(y_T)\gamma_{li}\sum_j\sum_n p(h_T=a_i \,|\, h_{T-1}=a_j, s_T=b_m, \theta)\, p(s_T=b_m \,|\, s_{T-1}=b_n, \theta)\times$$

$$p(h_{T-1}=a_j, s_{T-1}=b_n \,|\, y_1^{T-1}, v_1^{T-1}, \theta)$$

According to equation (4.2) and (4.4), we can obtain

$$p(h_T=a_i, s_T=b_m \,|\, y_1^T, v_1^T, \theta)$$

$$\propto g_i(T)\gamma_{li}\sum_j\sum_n f_{ijm}\, p_{mn}\, p(h_{T-1}=a_j, s_{T-1}=b_n \,|\, y_1^{T-1}, v_1^{T-1}, \theta). \qquad \text{(A.1)}$$

In order to calculate the result, we define

$$\alpha_{tim} \triangleq p(h_t=a_i, s_t=b_m \,|\, v_1^t, y_1^t, \theta), 1\le t\le T \qquad \text{(A.2)}$$

$$\alpha'_{Tim} \triangleq g_i(y_T)\gamma_{li}\sum_j\sum_n f_{ijm}\, p_{mn}\alpha_{(T-1)jn}. \qquad \text{(A.3)}$$

By equations (A.3) and (A.2), equation (A.1) can be rewrite as

$$\alpha_{Tim} \propto \alpha'_{Tim} = g_i(y_T)\gamma_{li}\sum_j\sum_n f_{ijm}\, p_{mn}\alpha_{(T-1)jn}.$$

Now we can use recursively to calculate $\alpha_{Tim}$, where $i=1,...,I$ and $m=1,...,M$. Since $\sum_i\sum_m \alpha'_{Tim}=1$, then $\alpha_{Tim}$ can also be written as

$$\alpha_{Tim} = \frac{\alpha'_{Tim}}{\sum_{i'}\sum_{m'}\alpha'_{Ti'm'}} \quad . \qquad \text{(A.4)}$$

We start with $T=1$

$$\alpha_{1im} \triangleq p(h_1 = a_i, s_1 = b_m \,|\, y_1, v_1 = c_l, \theta)$$

$$= \frac{p(y_1, v_1 = c_l \,|\, h_1 = a_i, s_1 = b_m, \theta)\, p(h_1 = a_i, s_1 = b_m \,|\, \theta)}{p(y_1, v_1 = c_l \,|\, \theta)}$$

$$\propto p(y_1, v_1 = c_l \,|\, h_1 = a_i, s_1 = b_m, \theta)\, p(h_1 = a_i, s_1 = b_m \,|\, \theta)$$

$$= p(y_1 \,|\, h_1 = a_i, v_1 = c_l, s_1 = b_m, \theta)\, p(v_1 = c_l \,|\, h_1 = a_i, s_1 = b_m, \theta)$$

$$p(h_1 \,|\, s_1 = b_m, \theta)\, p(s_1 = b_m \,|\, \theta).$$

By the CI rules, that is, $y_1 \perp s_1, v_1 \,|\, h_1$ and $v_1 \perp s_1 \,|\, h_1$, we can get

$$\alpha_{1im} \triangleq p(h_1 = a_i, s_1 = b_m \,|\, y_1, v_1 = c_l, \theta)$$

$$\propto p(y_1 \,|\, h_1 = a_i, \theta)\, p(v_1 = c_l \,|\, h_1 = a_i, \theta)\, p(h_1 = a_i \,|\, s_1 = b_m, \theta)\, p(s_1 = b_m \,|\, \theta).$$

From equations (4.1), (4.3), (4.5) and (4.6), we obtain as

$$\alpha_{1im} \triangleq p(h_1 = a_i, s_1 = b_m \,|\, y_1, v_1, \theta) \propto \alpha'_{1im} = g_i(y_1)\gamma_{li} f'_{im} p'_m.$$

Since $\sum_i \sum_m \alpha'_{1im} = 1$, $\alpha_{1im}$ can also be written as

$$\alpha_{1im} = \frac{\alpha'_{1im}}{\sum_{i'} \sum_{m'} \alpha'_{1i'm'}}$$

where $l$ is constant, satisfying $v_1 = c_l$. Hence proposition 6 has been proved.

□

# APPENDIX B

# THE PROOF OF SMOOTHING PROBABILITY DISTRIBUTION FUNCTION

*Proof of proposition 7.* Since

$$p(h_t = a_i, s_t = b_m \left| y_1^T, v_1^T, \theta \right.)$$

$$= p(h_t = a_i, s_t = b_m \left| y_1^t, v_1^t, y_{t+1}^T, v_{t+1}^T, \theta \right.)$$

$$= \frac{p(y_{t+1}^T, v_{t+1}^T \left| h_t = a_i, s_t = b_m, v_1^t, y_1^t, \theta \right.) p(h_t = a_i, s_t = b_m \left| v_1^t, y_1^t, \theta \right.)}{p(y_{t+1}^T, v_{t+1}^T \left| y_1^t, v_1^t, \theta \right.)}.$$

By using the CI rule, i.e. $y_{t+1}^T, v_{t+1}^T \perp v_1^t, y_1^t \left| h_t, s_t \right.$, we can get

$$p(h_t = a_i, s_t = b_m \left| y_1^T, v_1^T, \theta \right.)$$

$$= \frac{p(y_{t+1}^T, v_{t+1}^T \left| h_t = a_i, s_t = b_m, \theta \right.) p(h_t = a_i, s_t = b_m \left| v_1^t, y_1^t, \theta \right.)}{p(y_{t+1}^T, v_{t+1}^T \left| y_1^t, v_1^t, \theta \right.)}. \quad \text{(B.1)}$$

From equation (B.1) , setting

$$k_{tl} \triangleq p(y_t, v_t = c_l \left| y_1^{t-1}, v_1^{t-1}, \theta \right.), t \geq 2, \quad \text{(B.2)}$$

because we know the value of each variable, so $k_{tl}$ is constant. Here $t = 1$, $k_{1l} = p(y_1, v_1 = c_l \left| \theta \right.)$. by chain rule, The denominator of equation (B.1) can be written as:

$$p(y_{t+1}^T, v_{t+1}^T \left| y_1^t, v_1^t, \theta \right.) = \prod_{t'=t+1}^{T} p(y_{t'}, v_{t'} = c_l \left| v_1^{t'-1}, y_1^{t'-1}, \theta \right.) = \prod_{t'=t+1}^{T} k_{t'l}. \quad \text{(B.3)}$$

From equation (A.2), using Bayesian rules, when $1 \leq t \leq T$, we can get as

$$\alpha_{tim} \triangleq p(h_t = a_i, s_t = b_m \left| v_1^t, y_1^t, \theta \right.)$$

$$= \frac{p(y_t, v_t = c_l \left| h_t = a_i, s_t = b_m, y_1^{t-1}, v_1^{t-1}, \theta \right.) p(h_t = a_i, s_t = b_m \left| y_1^{t-1}, v_1^{t-1}, \theta \right.)}{p(y_t, v_t = c_l \left| y_1^{t-1}, v_1^{t-1}, \theta \right.)}. \tag{B.4}$$

From equations (A.3) and (B.2), (B.4) can be rewritten as

$$\alpha_{tim} = \frac{\alpha'_{tim}}{k_{tl}}. \tag{B.5}$$

Comparing equations (B.5) and (A.4), we can get as

$$k_{tl} = \sum_{i'} \sum_{m'} \alpha'_{ti'm'} \tag{B.6}$$

Here $l$ is constant , satisfying $v_t = c_l$ .

If we calculate the filtering probability density function of $\alpha_{tim}$, we have to solve $\alpha'_{tim}$. Then by equation (B.6), we can get $k_{tl}$ . Therefore $\prod_{t'=t+1}^{T} k_{t'l}$ is known.

Set $\beta'_{tim} \triangleq p(y_{t+1}^T, v_{t+1}^T \left| h_t = a_i, s_t = b_m, \theta \right.)$, from equation (B.1), we define

$$\beta_{tim} \triangleq \frac{p(y_{t+1}^T, v_{t+1}^T \left| h_t = a_i, s_t = b_m, \theta \right.)}{p(y_{t+1}^T, v_{t+1}^T \left| y_1^t, v_1^t, \theta \right.)} = \frac{\beta'_{tim}}{\prod_{t'=t+1}^{T} k_{t'l}}, \quad 1 \le t \le T-1. \tag{B.7}$$

According to equations (A.2) and (B.7), (B.1) can be rewritten as:

$$p(h_t = a_i, s_t = b_m \left| y_1^T, v_1^T, \theta \right.) = \beta_{tim} \alpha_{tim}. \tag{B.8}$$

Now we need to calculate $\beta'_{tim}$. When $t = T-1$,

$$\beta'_{(T-1)im} \triangleq p(y_T, v_T = c_l \left| h_{T-1} = a_i, s_{T-1} = b_m, \theta \right.)$$

$$= \sum_j \sum_n p(y_T, v_T = c_l, h_T = a_j, s_T = b_n \left| h_{T-1} = a_i, s_{T-1} = b_m, \theta \right.)$$

$$= \sum_j \sum_n [\, p(y_T \left| h_{T-1} = a_i, s_{T-1} = b_m, h_T = a_j, s_T = b_n, v_T = c_l, \theta \right.) \times$$

$$p(v_T = c_l \left| h_{T-1} = a_i, s_{T-1} = b_m, h_T = a_j, s_T = b_n, \theta \right.) \times$$

$$p(h_T = a_j \,|h_{T-1} = a_i, s_{T-1} = b_m, s_T = b_n, \theta)\, p(s_T = b_n \,|h_{T-1} = a_i, s_{T-1} = b_m, \theta)].$$

By using CI rules, $y_T \perp h_{T-1}, s_{T-1}, s_T, |h_T$ , $v_T \perp h_{T-1}, s_{T-1}, s_T \,|h_T$ , $h_T \perp s_{T-1} \,|h_{T-1}, s_T$ and $s_T \perp h_{T-1} \,|s_{T-1}$, so we can obtain

$$\beta'_{(T-1)im} \triangleq p(y_T, v_T = c_l \,|h_{T-1} = a_i, s_{T-1} = b_m, \theta)$$

$$= \sum_j \sum_n [\, p(y_T \,|h_T = a_j, \theta) p(v_T = c_l \,|h_T = a_j, \theta)$$

$$p(h_T = a_j \,|h_{T-1} = a_i, s_T = b_n, \theta)\, p(s_T = b_n \,|s_{T-1} = b_m, \theta)].$$

By equations (4.1), (4.2), (4.3) and (4.4), we get

$$\beta'_{(T-1)im} \triangleq p(y_T, v_T = c_l \,|h_{T-1} = a_i, s_{T-1} = b_m, \theta) = \sum_j \sum_n g_j(y_T) \gamma_{lj} f_{jin} p_{nm}$$

where $l$ is constant, satisfying, $v_T = c_l$.

Hence, when $t = T-1$

$$\beta_{(T-1)im} = \frac{\beta'_{(T-1)im}}{k_{Tl}} = \frac{\sum_j \sum_n g_j(y_T) \gamma_{lj} f_{jin} p_{nm}}{k_{Tl}}.$$

When $1 \le t < T-1$,

$$\beta'_{tim} \triangleq p(y_{t+1}^T, v_{t+1}^T \,|h_t = a_i, s_t = b_m, \theta)$$

$$= \sum_j \sum_n p(y_{t+1}^T, v_{t+1}^T, h_{t+1} = a_j, s_{t+1} = b_n \,|h_t = a_i, s_t = b_m, \theta)$$

$$= \sum_j \sum_n [\, p(y_{t+1}^T, v_{t+1}^T \,|h_{t+1} = a_j, s_{t+1} = b_n, h_t = a_i, s_t = b_m, \theta) \times$$

$$p(h_{t+1} = a_j \,|h_t = a_i, s_t = b_m, s_{t+1} = b_n, \theta)\, p(s_{t+1} = b_n \,|h_t = a_i, s_t = b_m, \theta)]$$

since $y_{t+1}^T, v_{t+1}^T \perp s_t, h_t \,|h_{t+1}, s_{t+1}$, $h_{t+1} \perp s_t \,|h_t, s_{t+1}$ and $s_{t+1} \perp h_t \,|s_t$, we can show

$$\beta'_{tim} = \sum_j \sum_n [\, p(y_{t+1}^T, v_{t+1}^T \,|h_{t+1} = a_j, s_{t+1} = b_n, \theta) p(h_{t+1} = a_j \,|h_t = a_i, s_{t+1} = b_n, \theta)$$

$$p(s_{t+1} = b_n \,|s_t = b_m, \theta)].$$

From equation (4.2) and (4.4), we can get

$$\beta'_{tim} = \sum_j \sum_n p(y_{t+1}, y_{t+2}^T, v_{t+1} = c_l, v_{t+2}^T \mid h_{t+1} = a_j, s_{t+1} = b_n, \theta) f_{jin} p_{nm}$$

$$= \sum_j \sum_n [\, p(y_{t+2}^T, v_{t+2}^T \mid h_{t+1} = a_j, s_{t+1} = b_n, y_{t+1}, v_{t+1} = c_l, \theta) \times$$

$$p(y_{t+1}, v_{t+1} = c_l \mid h_{t+1} = a_j, s_{t+1} = b_n, \theta) f_{jin} p_{nm}]$$

$$= \sum_j \sum_n [\, p(y_{t+2}^T, v_{t+2}^T \mid h_{t+1} = a_j, s_{t+1} = b_n, y_{t+1}, v_{t+1} = c_l, \theta) \times$$

$$p(y_{t+1} \mid h_{t+1} = a_j, s_{t+1} = b_n, v_{t+1} = c_l, \theta) p(v_{t+1} = c_l \mid h_{t+1} = a_j, s_{t+1} = b_n, \theta) f_{jin} p_{nm}],$$

since $y_{t+1} \perp v_{t+1}, s_{t+1} \mid h_{t+1}$ , $y_{t+2}^T, v_{t+2}^T \perp v_{t+1}, y_{t+1} \mid h_{t+1}, s_{t+1}$ and $v_{t+1} \perp s_{t+1} \mid h_{t+1}$ , according to equations (4.1) and (4.3) , so we can get as

$$\beta'_{tim} = \sum_j \sum_n p(y_{t+2}^T, v_{t+2}^T \mid h_{t+1} = a_j, s_{t+1} = b_n, \theta) p(y_{t+1} \mid h_{t+1} = a_j, \theta) p(v_{t+1} = c_l \mid h_{t+1} = a_j, \theta) \times$$
$$f_{jin} p_{nm}$$

$$= \sum_j \sum_n p(y_{t+2}^T, v_{t+2}^T \mid h_{t+1} = a_j, s_{t+1} = b_n, \theta) g_j(y_{t+1}) \gamma_{lj} f_{jin} p_{nm}$$

$$= \sum_j \sum_n \beta'_{(t+1)jn} g_j(y_{t+1}) \gamma_{lj} f_{jin} p_{nm}.$$

Hence

$$\beta'_{tim} = \sum_j \sum_n \beta'_{(t+1)jn} g_j(y_{t+1}) \gamma_{lj} f_{jin} p_{nm}. \tag{B.9}$$

This is a backward formula to solve for $\beta'_{tim}$. By equation (B.7) , we can get as

$$\beta'_{tim} = \beta_{tim} \prod_{t'=t+1}^{T} k_{t'l} \tag{B.10}$$

submitting (B.10) into (B.9), equation (B.9) can be rewritten as

$$\beta_{tim} \prod_{t'=t+1}^{T} k_{t'l} = \sum_j \sum_n (\beta_{(t+1)jn} \prod_{t''=t+2}^{T} k_{t''l}) g_j(y_{t+1}) \gamma_{lj} f_{jin} p_{nm}.$$

Hence, when $1 \le t < T-1$

$$\beta_{tim} = \frac{1}{k_{(t+1)l}} \sum_j \sum_n \beta_{(t+1)jn} g_j(y_{t+1}) \gamma_{lj} f_{jin} p_{nm}$$

where $l$ is constant, satisfying $v_{t+1} = c_l$.

Therefore, by equation (B.8), the smoothing probability density function $p(h_t = a_i, s_t = b_m \left| y_1^T, v_1^T, \theta \right.) = \beta_{tim} \alpha_{tim}$ can be obtained as:

When $t = T - 1$

$$p(h_{T-1} = a_i, s_{T-1} = b_m \left| y_1^T, v_1^T, \theta \right.) = \frac{1}{k_{Tl}} [\sum_j \sum_n g_j(y_T) \gamma_{lj} f_{jin} p_{nm}] \alpha_{(T-1)im}$$

where $k_{Tl} = \sum_{i'} \sum_{m'} \alpha'_{Ti'm'} = \sum_{i'} \sum_{m'} g_{i'}(y_T) \gamma_{l'i'} \sum_{j'} \sum_{n'} f_{i'j'm'} p_{m'n'} \alpha_{(T-1)j'n'}$ .

When $1 \le t < T - 1$

$$p(h_t = a_i, s_t = b_m \left| y_1^T, v_1^T, \theta \right.) = \frac{1}{k_{(t+1)l}} [\sum_j \sum_n \beta_{(t+1)jn} g_j(y_{t+1}) \gamma_{lj} f_{jin} p_{nm}] \alpha_{tim}$$

$$k_{(t+1)l} = \sum_{i'} \sum_{m'} \alpha'_{Ti'm'} = \sum_{i'} \sum_{m'} g_{i'}(y_{(t'+1)}) \gamma_{l'i'} \sum_{j'} \sum_{n'} f_{i'j'm'} p_{m'n'} \alpha_{t'j'n'} .$$

□

# APPENDIX C

# CALCULATING THE EXPECTATION OF LOG LIKELIHOOD FUNCTION

*Proof of the third term of equation (4.19).*

$$\sum_{H}\sum_{S} q^{(w)}(H,S)\sum_{t'=2}^{T}\log p(h_{t'}\,|s_{t'},h_{t'-1},\theta)$$

$$=\sum_{t'=2}^{T}\sum_{h_{t'},s_{t'},h_{t'-1},s_{t'-1}} q^{(w)}(h_{t'},s_{t'},h_{t'-1},s_{t'-1})\log p(h_{t'}\,|s_{t'},h_{t'-1},\theta)\,. \tag{C.1}$$

Now we only calculate $q^{(w)}(h_{t'},s_{t'},h_{t'-1},s_{t'-1})$.

When $2\le t'<T$, using Bayesian rules, we obtain as

$$q^{(w)}(h_{t'},s_{t'},h_{t'-1},s_{t'-1})$$

$$= p(h_{t'},s_{t'},h_{t'-1},s_{t'-1}\,|V,Y,\theta^{(w-1)})$$

$$= p(h_{t'},s_{t'},h_{t'-1},s_{t'-1}\,|v_1^T,y_1^T,\theta^{(w-1)})$$

$$= p(h_{t'},s_{t'},h_{t'-1},s_{t'-1}\,|v_1^{t'},y_1^{t'},v_{t'+1}^T,y_{t'+1}^T,\theta^{(w-1)})$$

$$=\frac{p(v_{t'+1}^T,y_{t'+1}^T\,|h_{t'},s_{t'},h_{t'-1},s_{t'-1},v_1^{t'},y_1^{t'},\theta^{(w-1)})}{p(v_{t'+1}^T,y_{t'+1}^T\,|v_1^{t'},y_1^{t'},\theta^{(w-1)})}\,p(h_{t'},s_{t'},h_{t'-1},s_{t'-1}\,|v_1^{t'},y_1^{t'},\theta^{(w-1)})$$

$$=\frac{p(v_{t'+1}^T,y_{t'+1}^T\,|h_{t'},s_{t'},\theta^{(w-1)})}{p(v_{t'+1}^T,y_{t'+1}^T\,|v_1^{t'},y_1^{t'},\theta^{(w-1)})}\,p(h_{t'},s_{t'},h_{t'-1},s_{t'-1}\,|v_1^{t'},y_1^{t'},\theta^{(w-1)})\,. \tag{C.2}$$

By equation (B.7), setting

$$\beta_{t'}^{(w-1)}=\frac{p(v_{t'+1}^T,y_{t'+1}^T\,|h_{t'},s_{t'},\theta^{(w-1)})}{p(v_{t'+1}^T,y_{t'+1}^T\,|v_1^{t'},y_1^{t'},\theta^{(w-1)})}\,. \tag{C.3}$$

Putting equation (C.3) into (C.2), and using Bayesian rules again, then (C.2) can be rewritten as

$$q^{(w)}(h_{t'}, s_{t'}, h_{t'-1}, s_{t'-1})$$

$$= \beta_{t'}^{(w-1)} p(h_{t'}, s_{t'}, h_{t'-1}, s_{t'-1} \left| v_1^{t'}, y_1^{t'}, \theta^{(w-1)} \right.)$$

$$= \beta_{t'}^{(w-1)} p(h_{t'}, s_{t'}, h_{t'-1}, s_{t'-1} \left| v_1^{t'-1}, v_{t'}, y_1^{t'-1}, y_{t'}, \theta^{(w-1)} \right.)$$

$$= \beta_{t'}^{(w-1)} \frac{p(v_{t'}, y_{t'} \left| h_{t'}, s_{t'}, h_{t'-1}, s_{t'-1}, v_1^{t'-1}, y_1^{t'-1}, \theta^{(w-1)} \right.)}{p(v_{t'}, y_{t'} \left| v_1^{t'-1}, y_1^{t'-1}, \theta^{(w-1)} \right.)} p(h_{t'}, s_{t'}, h_{t'-1}, s_{t'-1} \left| v_1^{t'-1}, y_1^{t'-1}, \theta^{(w-1)} \right.).$$

Since $v_{t'} \perp s_{t'}, h_{t'-1}, s_{t'-1}, v_1^{t'-1}, y_1^{t'-1}, y_{t'} \left| h_{t'} \right.$ and $y_{t'} \perp s_{t'}, h_{t'-1}, s_{t'-1}, v_1^{t'-1}, y_1^{t'-1} \left| h_{t'} \right.$

$h_{t'} \perp s_{t'-1}, v_1^{t'-1}, y_1^{t'-1} \left| h_{t'-1}, s_{t'} \right.$ and $s_{t'} \perp h_{t'-1}, v_1^{t'-1}, y_1^{t'-1} \left| s_{t'-1} \right.$, we can get as

$$q^{(w)}(h_{t'}, s_{t'}, h_{t'-1}, s_{t'-1})$$

$$= \beta_{t'}^{(w-1)} \frac{p(v_{t'} \left| h_{t'}, \theta^{(w-1)} \right.) p(y_{t'} \left| h_{t'}, \theta^{(w-1)} \right.)}{p(v_{t'}, y_{t'} \left| v_1^{t'-1}, y_1^{t'-1}, \theta^{(w-1)} \right.)} p(h_{t'} \left| v_1^{t'-1}, y_1^{t'-1}, h_{t'-1}, s_{t'-1}, s_{t'}, \theta^{(w-1)} \right.) \times$$

$$p(s_{t'} \left| v_1^{t'-1}, y_1^{t'-1}, h_{t'-1}, s_{t'-1}, \theta^{(w-1)} \right.) p(h_{t'-1}, s_{t'-1} \left| v_1^{t'-1}, y_1^{t'-1}, \theta^{(w-1)} \right.)$$

$$= \beta_{t'}^{(w-1)} \frac{p(v_{t'} \left| h_{t'}, \theta^{(w-1)} \right.) p(y_{t'} \left| h_{t'}, \theta^{(w-1)} \right.)}{p(v_{t'}, y_{t'} \left| v_1^{t'-1}, y_1^{t''-1}, \theta^{(w-1)} \right.)} p(h_{t'} \left| h_{t'-1}, s_{t'}, \theta^{(w-1)} \right.) p(s_{t'} \left| s_{t'-1}, \theta^{(w-1)} \right.)$$

$$p(h_{t'-1}, s_{t'-1} \left| v_1^{t'-1}, y_1^{t'-1}, \theta^{(w-1)} \right.) \qquad \text{(C.4)}$$

by equations (4.1), (4.2), (4.3) and (4.4), then (C.4) can be rewritten as

$$q^{(w)}(h_{t'} = a_i, s_{t'} = b_m, h_{t'-1} = a_j, s_{t'-1} = b_n)$$

$$= \beta_{t'im}^{(w-1)} \frac{\gamma_{li}^{(w-1)} g_i^{(w-1)}(y_{t'})}{k_{t'l}^{(w-1)}} f_{ijm}^{(w-1)} p_{mn}^{(w-1)} \alpha_{(t'-1)jn}^{(w-1)}. \qquad \text{(C.5)}$$

Hence when $2 \le t' < T$, (C.1) can be obtained as

$$\sum_H \sum_S q^{(w)}(H, S) \sum_{t'=2}^{T} \log p(h_{t'} = a_i \left| h_{t'-1} = a_j, s_{t'} = b_m, \theta \right.)$$

$$= \sum_{t'}\sum_{i}\sum_{j}\sum_{m}\sum_{n} \beta_{t'im}^{(w-1)} \frac{\gamma_{li}^{(w-1)} g_i^{(w-1)}(y_{t'})}{k_{t'l}^{(w-1)}} f_{ijm}^{(w-1)} p_{mn}^{(w-1)} \alpha_{(t'-1)jn}^{(w-1)} \log p(h_{t'} = a_i \,|\, h_{t'-1} = a_j, s_{t'} = b_m)$$

$$= \sum_{t'}\sum_{i}\sum_{j}\sum_{m}\sum_{n} \beta_{t'im}^{(w-1)} \frac{\gamma_{li}^{(w-1)} g_i^{(w-1)}(y_{t'})}{\sum_{i'}\sum_{m'} \alpha_{t'i'm'}^{\prime(w-1)}} f_{ijm}^{(w-1)} p_{mn}^{(w-1)} \alpha_{(t'-1)jn}^{(w-1)} \log f_{ijm} . \quad \text{(C.6)}$$

When $t' = T$ , using the CI rules, that is $v_T \perp s_T, h_{T-1}, s_{T-1}, v_1^{T-1}, y_1^{T-1}, y_T \,|\, h_T$ and $y_T \perp s_T, h_{T-1}, s_{T-1}, v_1^{T-1}, y_1^{T-1} \,|\, h_T$ and $h_T \perp s_{T-1}, v_1^{T-1}, y_1^{T-1} \,|\, h_{T-1}, s_T$ and $s_T \perp h_{T-1}, v_1^{T-1}, y_1^{T-1} \,|\, s_{T-1}$, we can get as

$q^{(w)}(h_{t'}, s_{t'}, h_{t'-1}, s_{t'-1})$

$= q^{(w)}(h_T, s_T, h_{T-1}, s_{T-1})$

$= P(h_T, s_T, h_{T-1}, s_{T-1} \,|\, v_1^T, y_1^T, \theta^{(w-1)})$

$= P(h_T, s_T, h_{T-1}, s_{T-1} \,|\, v_1^{T-1}, y_1^{T-1}, v_T, y_T, \theta^{(w-1)})$

$$= \frac{p(v_T, y_T \,|\, h_T, h_{T-1}, s_T, s_{T-1}, v_1^{T-1}, y_1^{T-1}, \theta^{(w-1)})}{p(v_T, y_T \,|\, v_1^{T-1}, y_1^{T-1}, \theta^{(w-1)})} p(h_{T-1}, s_{T-1}, h_T, s_T \,|\, v_1^{T-1}, y_1^{T-1}, \theta^{(w-1)})$$

$$= \frac{p(v_T \,|\, h_T, \theta^{(w-1)}) p(y_T \,|\, h_T, \theta^{(w-1)})}{p(v_T, y_T \,|\, v_1^{T-1}, y_1^{T-1}, \theta^{(w-1)})} p(h_T \,|\, h_{T-1}, s_T, s_{T-1}, v_1^{T-1}, y_1^{T-1}, \theta^{(w-1)}) \times$$

$$p(s_T \,|\, h_{T-1}, s_{T-1}, v_1^{T-1}, y_1^{T-1}, \theta^{(w-1)}) p(h_{T-1}, s_{T-1} \,|\, v_1^{T-1}, y_1^{T-1}, \theta^{(w-1)})$$

$$= \frac{p(v_T \,|\, h_T, \theta^{(w-1)}) p(y_T \,|\, h_T, \theta^{(w-1)})}{p(v_T, y_T \,|\, v_1^{T-1}, y_1^{T-1}, \theta^{(w-1)})} p(h_T \,|\, h_{T-1}, s_T, \theta^{(w-1)}) p(s_T \,|\, s_{T-1}, \theta^{(w-1)}) \times$$

$$p(h_{T-1}, s_{T-1} \,|\, v_1^{T-1}\ y_1^{T-1}, \theta^{(w-1)}. \quad \text{(C.7)}$$

By equations (4.1), (4.2), (4.3) and (4.4), then (C.7) can be rewritten as

$$q^{(w)}(h_T = a_i, s_T = b_m, h_{T-1} = a_j, s_{T-1} = b_n) = \frac{\gamma_{li}^{(w-1)} g_i^{(w-1)}(y_T)}{k_{Tl}^{(w-1)}} f_{ijm}^{(w-1)} p_{mn}^{(w-1)} \alpha_{(T-1)jn}^{(w-1)}$$

$$= \frac{\gamma_{li}^{(w-1)} g_i^{(w-1)}(y_T)}{\sum_{i'}\sum_{m'} \alpha_{Ti'm'}^{\prime(w-1)}} f_{ijm}^{(w-1)} p_{mn}^{(w-1)} \alpha_{(T-1)jn}^{(w-1)} \quad \text{(C.8)}$$

Hence when $t' = T$, (C.8) can be obtained as

$$\sum_H \sum_S q^{(w)}(H,S) \log p(h_T = a_i \,|\, h_{T-1} = a_j, s_T = b_m, \theta)$$

$$= \sum_i \sum_j \sum_m \sum_n q^{(w)}(h_T = a_i, s_T = b_m, h_{T-1} = a_j, s_{T-1} = b_n) \log p(h_T \,|\, h_{T-1}, s_T, \theta)$$

$$= \frac{\gamma_{li}^{(w-1)} g_i^{(w-1)}(y_T)}{\sum_{i'} \sum_{m'} \alpha_{Ti'm'}'^{(w-1)}} f_{ijm}^{(w-1)} p_{mn}^{(w-1)} \alpha_{(T-1)jn}^{(w-1)} \log f_{ijm}.$$

So according to (C.6) and (C.8), (C.1) can be obtained as

When $2 \le t' \le T$,

$$\sum_H \sum_S q^{(w)}(H,S) \sum_{t'=2}^{T} \log p(h_{t'} \,|\, s_{t'}, h_{t'-1}, \theta)$$

$$= \sum_{t'=2}^{T-1} \sum_i \sum_j \sum_m \sum_n \beta_{t'im}^{(w-1)} \frac{\gamma_{li}^{(w-1)} g_i^{(w-1)}(y_{t'})}{\sum_{i'} \sum_{m'} \alpha_{t'i'm'}'^{(w-1)}} f_{ijm}^{(w-1)} p_{mn}^{(w-1)} \alpha_{(t'-1)jn}^{(w-1)} \log f_{ijm} +$$

$$\frac{\gamma_{li}^{(w-1)} g_i^{(w-1)}(y_T)}{\sum_{i'} \sum_{m'} \alpha_{Ti'm'}'^{(w-1)}} f_{ijm}^{(w-1)} p_{mn}^{(w-1)} \alpha_{(T-1)jn}^{(w-1)} \log f_{ijm}.$$

Hence we have obtained equation (4.26).

□

# CURRICULUM VITAE

**Name** Lingling Luo

**Date of Birth** 28/Jan./1984

**Place of Birth** Hunan Province, China

**Education**

M. Sc. School of Science, Guizhou Minzu University, 2010. Major: Probability and Statistics.

B. Sc. School of Science, Guizhou Minzu University, 2007. Major: Mathematics and Applied Mathematics.

**Publications**

[1] Lingling Luo, Sattayatham Pairote and Ratthachat Chatpatanasiri. GARCH-type Forecasting Models for Volatility of Stock Market and MCS Test. Communications in Statistics - Simulation and Computation, DOI: 10.1080/03610918.2016.1152366.

**Position and Place of Work**

School of Mathematics and Statistics, Guizhou University of Finance and Economics, Guiyang 550025, China.