การประมาณค่าแวลูแอทริสก์ด้วยวิธีการอนุภาคเอ็มซีเอ็มซีปรับได้และการถดถอยซับพอร์ทเวกเตอร์

นางซินเฉีย หยาง

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรดุษฎีบัณฑิต
สาขาวิชาคณิตศาสตร์ประยุกต์
มหาวิทยาลัยเทคโนโลยีสุรนารี
ปีการศึกษา 2559

ESTIMATION OF VALUE AT RISK USING AN ADAPTIVE MCMC METHOD AND SUPPORT VECTOR REGRESSION

Xinxia Yang

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

Suranaree University of Technology

Academic Year 2016

ESTIMATION OF VALUE AT RISK USING AN ADAPTIVE PARTICLE MCMC METHOD AND SUPPORT VECTOR REGRESSION

Suranaree University of Technology has approved this thesis submitted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy.

Thesis Examining Committee

(Asst. Prof. Dr. Eckart Schulz)
Chairperson

(Prof. Dr. Pairote Sattayatham)
Member (Thesis Adivisor)

(Prof. Dr. Suthep Suantai)
Member

(Asst. Prof. Dr. Arjuna Chaiyasena)
Member

(Asst. Prof. Dr. Benjawan Rodjanadid)
Member

(Prof. Dr. Sukit Limpijumnong)
Vice Rector for Academic Affairs and Innovation

(Prof. Dr. Santi Maensiri)
Dean of Institute of Science

ซินเฉีย หยาง : การประมาณค่าแวลูแอทริสก์ด้วยวิธีการอนุภาคเอ็มซีเอ็มซีปรับได้และการถดถอยซับพอร์ทเวกเตอร์ (ESTIMATION OF VALUE AT RISK USING AN ADAPTIVE MCMC METHOD AND SUPPORT VECTOR REGRESSION)
อาจารย์ที่ปรึกษา : ศาสตราจารย์ ดร.ไพโรจน์ สัตยธรรม, 87 หน้า.

จุดประสงค์ของการศึกษา คือ การประมาณค่าแวลูแอทริสก์ (วาร์) จากข้อมูลอนุกรมเวลาทางการเงินด้วยตัวแบบความผันผวนและวิธีการเรียนรู้ทางสถิติ วิทยานิพนธ์ฉบับนี้นำเสนอวิธีการคำนวณ วาร์ สองวิธี วิธีแรก คือ การคำนวณด้วยเทคนิคที่มีชื่อว่า "โซ่มาร์คอฟ มอนติคาร์โล ที่ปรับตัวได้" (A-PMCMC) และวิธีที่สองใช้เทคนิคที่เรียกว่า "การวิเคราะห์หลายสเกลด้วยนอร์มัลอินเวอร์ส เกาซียน" (NIG-MSA)

ในวิธีแรกนั้นผู้วิจัยได้สร้างตัวแบบความผันผวนรูปแบบใหม่ขึ้น โดยตัวแบบความผันผวนใหม่นี้ได้รวมถึงผลกระทบที่สำคัญในตลาดการเงินจริงสามประเภท คือ ผลกระทบจากตัวเร่ง จากค่าเฉลี่ยแบบมีเงื่อนไขไม่คงที่ และ จากการเปลี่ยนแปลงอย่างกะทันหัน จากนั้นจึงใช้เทคนิคโซ่มาร์คอฟ มอนติคาร์โล ที่ปรับตัวได้เพื่อประมาณค่าพารามิเตอร์ในตัวแบบดังกล่าว

สำหรับวิธีที่สองนั้นผู้วิจัยได้นำเสนอตัวแบบลูกผสม ซึ่งมีการใช้การแจกแจงนอร์มัลอิเวอร์สเกาซียนไปฟิตการแจกแจงของเศษ ของตัวแบบการ์ช และได้ผนวก เอมไพริคัล โหมด ดีคอมโพซิชัน กับซัพพอร์ต เวกเตอร์ แมชีน เพื่อทำการวิเคราะห์หลายสเกลโดยการรวมกันของวิธีเหล่านี้จะสามารถอธิบายสเกลของความผันผวนได้มากกว่าหนึ่งสเกลได้ทำการทดลองด้วยการจำลองและด้วยข้อมูลจริงเพื่อทดสอบตัวแบบที่ได้นำเสนอข้างต้น

สาขาวิชาคณิตศาสตร์ ลายมือชื่อนักศึกษา Xinxia Yang
ปีการศึกษา 2559 ลายมือชื่ออาจารย์ที่ปรึกษา

XINXIA YANG : VALUE AT RISK ESTIMATION BASED ON VOLATILITY MODEL AND STATISTICAL LEARNING APPROACH. THESIS ADVISOR : PROF.PAIROTE SATTAYATHAM, Ph.D. 87 PP.

VALUE AT RISK / STOCHASTIC VOLATILITY MODEL / GARCH MODEL / PARTICLE FILTER / MCMC / SUPPORT VECTOR REGRESSION

The objective of this study is to estimate Value-at-Risk of financial time series based on two volatility models and statistical learning approach. In this thesis, we propose two VaR estimation methods, referred to as A-PMCMC (Adaptive Particle Markov Chain Monte Carlo) approach and NIG-MSA (Normal Inverse Gaussian-Multi Scale Analysis) approach respectively. In the A-PMCMC approach, a new stochastic volatility model with leverage effect, non-constant conditional mean and jump is presented, and then the particle filter and adaptive MCMC algorithms are integrated to estimate the volatility model. The NIG-MSA is a hybrid approach. It exploits the normal inverse Gaussian distribution to fit the residual distribution of the GARCH model, and combines the empirical mode decomposition with support vector regression to structure a multi-scale analysis approach. The simulation experiment and empirical application shows the applicability of our proposed methods.

School of Mathematics

Academic Year 2016

Student's Signature XinXia Yang

Advisor's Signature P. Sattayatham

ACKNOWLEDGEMENTS

This thesis has involved the help and invaluable advice of many people. I am deeply indebted to Prof. Dr. Pairote Sattayatham, Dr. Bhusana Premanode and Dr. Ratthachat Chatpatanasiri, my thesis advisors for their support, patient help and offering many useful suggestions. I wish to express my sincere gratitude to them for patiently listening to my talking and never failing to give me encouragement.

I am grateful to Asst. Prof. Dr. Eckart Schulz for teaching and helping me during the course of studies in Mathematics. I would like to express my appreciation to Asst. Prof. Dr. Eckart Schulz, Prof. Dr. Pairote Sattayatham, Dr. Bhusana Premanode, Dr. Ratthachat Chatpatanasiri, Prof. Dr. Suthep Suantai, Asst. Prof. Dr. Arjuna Chaiyasena, and Asst. Prof. Dr. Benjawan Rodjanadid, for their valuable discussions and comments.

In addition, I would like to acknowledge the professional support I received from the faculty of the School of Mathematics.

Lastly, I am forever grateful to my family and my friends for constantly giving me support, understanding, encouragement and love.

Xinxia Yang

CONTENTS

Page

CONTENTS (Continued)

Page

CONTENTS (Continued)

Page

LIST OF TABLES

LIST OF FIGURES

Figure **Page**

LIST OF FIGURE (Continued)

Figure **page**

CHAPTER I

INTRODUCTION

1.1 Introduction

Over recent years, financial markets have become much more volatile compared to previous decades. The most difficult task for investors and regulators is to accurately measure financial market risks. Financial market risk measurement has been addressed by an increasing number of researches (Szegö, 2002; Yamai and Yoshiba, 2005; Tsukahara, 2014). Value-at-Risk (VaR), as a statistical approach to measure financial market risk, has become a market standard for measuring, managing, and reporting market risk (Basak and Shapiro, 2001). VaR provides users with a summary measure of market risk and is truly a forward-looking risk measure and has turned into the most frequently used risk measure (Alexander, 2009). VaR is widely used by banks, securities firms, commodity and energy merchants, and other trading organizations.

VaR is generally defined as the worst expected loss for a given position or portfolio within a known confidence interval over a specific time horizon under normal market conditions. We use an intuitive example here to understand the VaR. Figure 1.1 shows the distribution of the daily returns of NASDAQ100 (Harper, 2008). Notice the red bars that compose the “left tail” of the histogram. The red bars run from daily losses of 4% to 8%. If given a loss level of 5%, i.e 95% confidence level, we expect that our worst daily loss will not exceed 4%, then 4% is VaR. For example,

if we invest \$100, we are 95% confident that our worst daily loss will not exceed \$4, then \$4 is VaR. If given a loss level of 1%, i.e 99% confidence level, we expect that the worst daily loss will not exceed 7%, then 7% is VaR. For example, if we invest \$100, we are 99% confident that our worst daily loss will not exceed \$7, then \$7 is VaR.

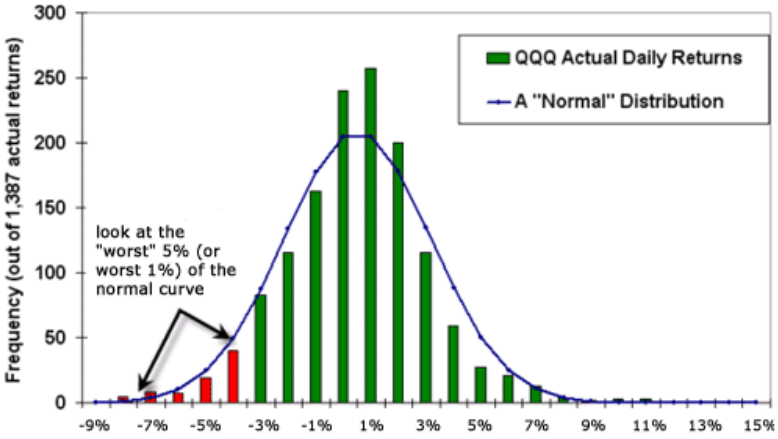


Figure 1.1 Distribution of Daily Return, NASDAQ 100.

Mathematically VaR at τ loss probability level is commonly defined as:

$$VaR_{\tau,t} = F_t^{-1}(\tau) = \sigma_t q_\tau \tag{1.1}$$

where F_t^{-1} is the inverse function of the conditional cumulative Gaussian distribution function at time t (Franke et al.,2004). q_τ denotes the τ -th quantile of the distribution of innovation term ε_t, i.e. $P(\varepsilon_t < q_\tau) = \tau$, and σ_t denotes the volatility. Since the VaR can be expressed as $\sigma_t \times q_\tau$, so it is crucial for the calculation of VaR to model the distribution of the innovation term and estimate the volatility accurately.

It is an obvious truth that the volatility of returns plays an important role when estimating VaR of returns. Volatility forecasts are important inputs into risk

management models. However, a number of empirical studies have concluded that the assumption of constant volatility is inadequate to describe stock returns, based on three findings: (1) volatilities of stock returns vary over time, but persist at certain levels (mean-reversion property); (2) volatilities are correlated with stock returns (Li et al., 2005); (3) many financial time series exhibit volatility clustering whereby volatility is likely to be high when it has recently been high and volatility is likely to be low when it has recently been low. Therefore, modeling time-varying conditional volatility σ_t is an important and interesting topic to study.

The Stochastic Volatility (SV) model (Taylor, 1986) and the Generalized Autoregressive Conditional Heteroscedastic (GARCH) model (Engle, 1982; Bollerslev, 1986) are the two classes of models that are often used to estimate and forecast unobserved time-varying conditional volatility σ_t. On the one hand, the SV model fits the time varying variance as an unobserved component that follows a particular stochastic process which is not restricted to follow a deterministic process. On the other hand, the GARCH model fits the time varying variance as a deterministic function of lagged squared residuals and lagged conditional variance and is particularly useful for modeling time-varying conditional volatility.

However, a voluminous literature has demonstrated that the occurrence of a financial crisis or a change of the economic environment will cause the stock price to exhibit sharp volatility, leading to the structural instability of financial markets (Cogley and Sargent, 2001; Koop and Potter, 2007). Therefore, the basic SV model and GARCH model have difficulties in accurately estimating the time-varying volatility. In this study, to address this issue, we improve the basic SV model and GARCH model using different tactics to model the time-varying volatility of financial

time series respectively.

First, we propose a new SV model with leverage effect, non-constant conditional mean and jump, which considers the nature of structural instability of stochastic volatility models in order to avoid the misspecification in the volatility process. This SV model is structured based on the following facts:

(i) The leverage effect describes the relationship between returns and conditional variances. It means that bad news in the markets lead to a boost in the variance. The underlying reasoning is that bad news tend to decrease prices, thus leading to an increase in debt-to-equity ratio (Figlewski et al., 2000). The firms are hence riskier and this translates into an increase in expected future volatility as captured by a negative relationship between volatility and return.

(ii) The stochastic volatility model can describe the serial correlation on the return process explained by unexpected stochastic dividend effects (Eberlein et al., 2003). In addition, it is important to consider the conditional mean in applications related to optimal portfolio choices. The economic theory shows that an investor gains from market predictability and volatility timing (Merton, 1971). In order to accurately describe the stock returns process, we model the conditional expected value of the returns together with the dynamic of the volatilities. Moreover, we model the conditional mean of returns via an autoregressive process, which can be explained by the non-synchronous trading and unexpected stochastic dividend effects (Chernov et al., 2003).

(iii) There is substantial evidence in favor of jumps on returns and volatilities (Merton, 1976; Eraker et al., 2003; Raggi, 2005). Jumps are an important feature of financial markets, which can basically be described as rare events and infrequent but

extreme movements in returns. Furthermore, there is some evidence that an extreme and rare event influences the conditional mean and the volatility (Liu et al. 2001), so we take into account a jumps dynamic in the stochastic volatility model.

Second, we propose a GARCH model with Normal Inverse Gaussian (NIG) distribution. The distribution assumption of the innovative term influences the performance of the VaR. It is clear that the accuracy of VaR depends heavily on the assumption of the underlying distribution, which often assumes that the involved risk factors are normally distributed for reasons of stochastic and numerical simplicity. However, many empirical studies have shown that the financial returns have leptokurtic distribution with high peak and fat tails (Peiro, 1999; Verhoeven, 2004). The NIG distribution is a heavy-tailed distribution that can well replicate the empirical distribution of the financial risk factors (Nielsen, 1977; Chen et al., 2008). Therefore, in this thesis, we discuss the application of the NIG distribution in financial risk measurement.

Apart from improving the volatility models, we employ the statistical learning approach to estimate the proposed volatility models. Statistical learning is a branch of statistics aimed at modeling and understanding complex models and datasets. Generally speaking, statistical learning aims at estimating a target variable, based on a set of inputs, or predictors. In this thesis, we use Bayes statistical inference and support vector regression to enhance the VaR estimation models. Based on the Bayes statistical inference theory, we construct an A-PMCMC method used to estimate the unknown parameters and latent variables of the stochastic volatility model. Based on the support vector regression theory, we proposed a hybrid NIG-MSA model to estimate the time-varying volatility of financial return series.

The proposed stochastic volatility model with leverage effect, non-constant conditional mean and jump belong to the wide class of nonlinear state-space models; classical parameter estimation is difficult due to lacking the tractable form of the likelihood. Statistical estimation of stochastic volatility models is, however, greatly complicated by the stochastic evolution of volatility which implies that the likelihood here can not be obtained in closed form. In order to estimate the proposed SV model, we present an A-PMCMC (Adaptive Particle Markov China Monte Carlo) algorithm to calculate the model parameters and latent variables.

The motivation why the A-PMCMC algorithm is proposed is that although the Bayesian MCMC methods are widely employed to estimate the stochastic volatility models with jumps (Jacquier et al., 1994; Eraker et al., 2003; Li et al., 2008; Golightly, 2009), in real time applications we have to restart the inferential procedure from scratch causing a considerable amount of wasted time. The A-PMCMC algorithm adaptively samples using information obtained from precious draws to turn the proposal distribution automatically, and contains the advantages of the sequential Monte Carlo (Particle Filter) method and MCMC algorithm. The particle filter algorithm can update the unobserved process given a specific value for the parameters, and MCMC moves will be used to update the parameter values.

In addition, considering that financial data is usually not constant or absolute scale and is usually found with multiple time-scale characteristics (Skjeltorp, 2000), we use multi-scale analysis for financial data (Guhathakurta et al., 2008; Huang et al., 2003). Multi-scale analysis is a comprehensive analysis approach and specially developed for non-stationary processes. In general, the multi-scale analysis consists of two steps: (1) Decompose the original signal according to the time scale and (2)

integrate the analysis results of subsystems. Particularly, we adopt empirical mode decomposition (EMD) method to decomposition the process, and support vector regression (SVR) method for the integration process.

The EMD method proposed by Huang (1998) can adaptively decompose the original signal into a series of intrinsic mode function components with different time-scale. The method is applicable to nonlinear and non-stationary processes since it is based on the local characteristic time scale of the data. Compared to wavelet decomposition and Fourier decomposition, EMD decomposition has been reported to have worked better in describing the local time scale. The EMD method has been applied to analyze the non-stationary financial time series (Huang et al., 2003; Premanode et al., 2013; Hong, 2011). Support Vector Regression (SVR), proposed by Vapnik (1996), is a non-linear kernel-based regression method which tries to find the best regression hyper-plane based on the structural risk minimization principle (Yeh et al., 2011). The solution of SVR is unique and globally optimal since SVR is formulated to linearly constrain quadratic programming problem. The SVR method has been empirically shown to produce a better generalization performance than artificial neural networks which is a commonly used machine learning method (Rosillo, 2014). The SVR approach has been widely used in volatility forecasting (Wang et al., 2013; Tang et al., 2009). For non-stationary financial time series, the multi-scale method has been addressed by an increasing number of researches. Most of the literature has concentrated on the prediction of crude oil price and stock index (Yu et al., 2008; Kazem et al., 2013). To the best of our knowledge, using the multi-scale method to forecast VaR of financial market has not been studied so far.

Integrating the GARCH model with NIG distribution and multi scale analysis,

we propose the NIG-GARCH model, which following the steps: Firstly, we decompose the financial time series into several intrinsic mode functions by the empirical mode decomposition. Secondly, the GARCH model with NIG distribution is used to forecast the volatility of the each intrinsic mode function components respectively. Finally, the volatility that has been predicted before will be integrated by the support vector regression (SVR) method. We call the VaR estimation process as NIG-MSA approach.

1.2 Outline of the thesis

This thesis is organized as the follows.

In Chapter II, we introduce the model description & mathematical background.

In Chapter III, we propose an A-PMCMC VaR estimation approach based a new stochastic volatility model using adaptive PMCMC method.

In Chapter IV, we propose a new hybrid VaR estimation approach, named the NIG-MSA approach, that integrates the GARCH model with NIG distribution and multi scale analysis.

The conclusion of the thesis is presented in the last chapter.

CHAPTER II

MODEL DESCRIPTION & MATHEMATICAL BACKGROUND

In this chapter, we present the model description and mathematical background related with our study contents. First, we introduce the elementary knowledge of Value-at-Risk. Second, we briefly present some volatility models including stochastic volatility model and GARCH model. Third, we review the background knowledge of Bayesian inference and support vector regression. They are the two primary statistical learning approaches.

2.1 Value at Risk

In finance, the Value-at-Risk (VaR) is a risk measure used to estimate how the value of an asset or of a portfolio of assets will decrease over a certain time period (usually over one day or 10 days) under usual conditions. VaR has two parameters: (1) the significance level α (or confidence level $1-\alpha$) at which we plan to make the estimate; (2) the risk horizon, denoted h, the length of time over which the assets in the portfolio will be held, also called holding period or forecast horizon.

2.1.1 VaR definition

Value-at-Risk (VaR) is a statistical approach to measure financial risk. In short, VaR is the maximum loss over a target horizon for a given confidence level.

In financial mathematics and financial risk management, VaR is defined as: given a loss level $\alpha \in (0,1)$, which corresponds to a $1-\alpha$ level of confidence, the VaR of the portfolio at the loss level α is given by the largest number x_α such that the probability that the return X is less than x_α, is at most α. Mathematically, if X is the return of a portfolio, then $VaR_\alpha(X)$ is the level $\alpha-$quantile, i.e.

$$\mathrm{VaR}_\alpha(X) = \sup\{x_\alpha \in R : \Pr(X < x_\alpha) \le \alpha\} \qquad (2.1)$$

2.1.2 VaR models

There are three basic types of VaR models: the normal linear VaR model, the historical simulation model and the Monte Carlo VaR model. The only differences between the three VaR models are due to the manner in which this distribution is constructed. All three approaches may be developed and generalized.

(1) Normal Linear VaR model.

Suppose we only seek to measure the VaR of a portfolio without attributing the VaR to different risk factors. We also make the simplifying assumption that the portfolio's discounted $h-$day returns are $i.i.d.$ and normally distributed. For simplicity of notation we shall, in this section, write the return as X, dropping the dependence on both time and risk horizon. Thus we have

$$X \sim i.i.d.(\mu, \sigma^2).$$

We will derive a formula for x_α, the α quantile return, *i.e.* the return such that $\Pr(X < x_\alpha) = \alpha$. Then the α VaR, expressed as a percentage of the portfolio value, is minus this α quantile. Using the standard normal transformation, we have

$$\Pr\left(X < x_\alpha\right) = \Pr\left(\frac{X-\mu}{\sigma} < \frac{x_\alpha - \mu}{\sigma}\right) = \Pr\left(Z < \frac{x_\alpha - \mu}{\sigma}\right),$$

where $Z \sim N(0,1)$, So if $\Pr\left(X < x_\alpha\right) = \alpha$, then

$$\Pr\left(Z < \frac{x_\alpha - \mu}{\sigma}\right) = \alpha .$$

By definition, $\Pr\left(Z < \phi^{-1}(\alpha)\right) = \alpha$, so

$$\frac{x_\alpha - \mu}{\sigma} = \phi^{-1}(\alpha),$$

where ϕ is the standard normal distribution function.

But $x_\alpha = -\mathrm{VaR}_\alpha$ by definition, and $\phi^{-1}(\alpha) = -\phi^{-1}(1-\alpha)$ by the symmetry of the standard normal distribution. Then we can yield an analytic formula for the VaR for a portfolio with an *i.i.d*. normal return, *i.e.*

$$VaR_\alpha = \phi^{-1}(1-\alpha)\sigma - \mu . \tag{2.2}$$

If we want to be more precise about the risk horizon of our VaR estimate, we may write

$$VaR_{h,\alpha} = \phi^{-1}(1-\alpha)\sigma_h - \mu_h \tag{2.3}$$

This is a simple formula for the $100\alpha\%$ $h-$day VaR, as a percentage of the portfolio value, when the portfolio's discounted returns are *i.i.d*. normally distributed with expectation μ_h and standard deviation σ_h.

To obtain the VaR in value terms, we simply multiply the percentage VaR by the current value of the portfolio:

$$VaR_{h,t,\alpha} = \left(\phi^{-1}(1-\alpha)\sigma_h - \mu_h\right)P_t , \tag{2.4}$$

where P_t is the value of the portfolio at the time t when the VaR is measured.

(2) Historical VaR model.

The historical VaR model is a nonparametric method that uses the empirical distribution of past returns to generate a VaR. It assumes that all possible future variations have been experienced in the past, and that the historically simulated distribution is identical to the returns distribution over the forward looking risk horizon. Historical Simulation is the procedure for predicting value at risk by "simulating" or constructing the cumulative distribution function of asset returns over time. It does not require any statistical assumption beyond stationary of the distribution of returns or, in particular, their volatility. The limitation of the historical simulation lies in its i.i.d. assumption of returns. From empirical evidence, it is known that asset returns are clearly not independent as they exhibit certain patterns such as volatility clustering. Unfortunately Historical Simulation does not take into account such patterns.

Historical Simulation is a good re-sampling method because of its simplicity and lack of distributional assumption about underlying process of returns. It is based on the assumption that history is repeating itself. Unlike other parametric methods, the historical simulation makes no specific distribution assumption about return distributions. However, the historical simulation implicitly assumes that the distribution of past returns is a good and complete representation of expected future returns. This method also relies on the specified short historical moving window.

(3) Monte Carlo VaR model

Monte Carlo VaR model uses random samples from known populations of simulated data to track a statistic's behavior. With Monte Carlo VaR measures, an inference procedure typically characterizes the distribution of returns by assuming

some standard joint distribution-such as the joint-normal distribution-and specifying a covariance matrix and mean vector.

The Monte Carlo simulation method is similar to the historical simulation method, except that the movements in risk factors are generated from estimated distribution. When we need to specify the dependency structure of the risk factors, we can determine the marginal distributions as well as their copula.

The Monte Carlo VaR model requires users to make assumptions about the stochastic process. It is subject to model risk. It also creates inherent sampling variability because of the randomization. Different random numbers will lead to different results. It may take a large number of iterations to converge to a stable VaR measure. Unlike other methods, Monte Carlo simulation makes explicit the sampling variability in the risk numbers. Although the Monte Carlo Simulation can be time-consuming according to the properties of problem, the main benefit of running it is that it can model instruments with non-linear and path dependent payoff functions, especially complex derivatives.

2.2 Volatility model

In this section, we briefly introduce the two most common volatility models, i.e. stochastic volatility model and GARCH model.

2.2.1 Stochastic volatility model

We only present the three basic stochastic volatility models related with our study.

(i) The basic stochastic volatility model

$$\begin{aligned} y_t &= \varepsilon_t \exp\left(h_t/2\right), \\ h_{t+1} &= \mu + \left(h_t - \mu\right)\phi + \sigma_\eta \eta_t, \end{aligned} \tag{2.5}$$

where

$$\begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix} \sim N\left(0, \Sigma\right) \text{ and } \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Here y_t is the observed stock return, h_t are unobserved log-volatilities, μ is the drift in the state equation, σ_η is the volatility of volatility factor η_t and ϕ is the persistence clustering and heavy tailed distribution of returns (Micciche et al., 2002; Taylor, 1986; Liesenfeld and Jung, 2000).

(ii) The stochastic volatility model with leverage effects

The stochastic volatility model with leverage effect can be written as

$$\begin{aligned} y_t &= \varepsilon_t \exp\left(h_t/2\right), \\ h_{t+1} &= \mu + \left(h_t - \mu\right)\phi + \sigma_\eta \eta_t, \end{aligned} \tag{2.6}$$

where

$$\begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix} \sim N\left(0, \Sigma\right) \text{ and } \Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}.$$

Note that the disturbances are conditionally Gaussian, we can write $\eta_t = \rho\varepsilon_t + \sqrt{\left(1-\rho^2\right)}\xi_t$, where $\xi_t \sim N\left(0,1\right)$. The state equation can be reformulated as $h_{t+1} = \mu + \left(h_t - \mu\right)\phi + \sigma_\eta \rho \varepsilon_t + \sigma_\eta \sqrt{\left(1-\rho^2\right)}\xi_t$.

(iii) The stochastic volatility model with jumps

The stochastic volatility model with jumps can be written as

$$\begin{aligned} y_t &= \varepsilon_t \exp\left(h_t/2\right) + J_t \omega_t, \\ h_{t+1} &= \mu + \left(h_t - \mu\right)\phi + \sigma_\eta \eta_t, \end{aligned} \tag{2.7}$$

where

$$\begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix} \sim N(0, \Sigma) \text{ and } \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Here $J_t = j$ is the time$-t$ jump arrival where $j = 0,1$ is a Bernoulli counter with intensity p, $\omega_t \sim N(0, \sigma_J^2)$ dictates the jump size. The leverage effect is incorporated as noted earlier.

2.2.2 GARCH model

It is well known that financial time series exhibit time-varying volatility. Bollerslev (1986) and Taylor (1986) independently generalised Engle's model to make it more realistic; the generalisation was called "GARCH". GARCH is probably the most commonly used financial time series model and has inspired dozens of more sophisticated models.

The elementary GARCH(p, q) model is defined by,

$$\begin{aligned} y_t &= \sigma_t \varepsilon_t, \\ \sigma_t^2 &= \omega + \sum_{i=1}^{p} \alpha_i y_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2, \end{aligned} \tag{2.8}$$

where $\omega > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$, and the innovation sequence $\{\varepsilon\}_{i=-\infty}^{\infty}$ is independent and identically distributed with $E(\varepsilon_0) = 0$ and $E(\varepsilon_0^2) = 1$.

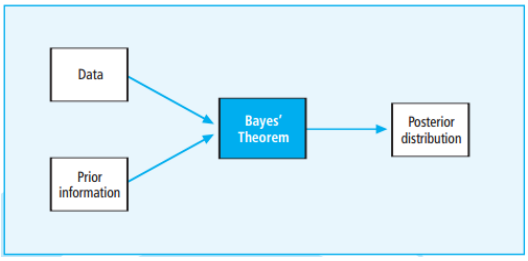
The main idea is that σ_t^2, the conditional variance of y_t given information available up to time $t-1$, has an autoregressive structure and is positively correlated to its own recent past and to recent values of the squared returns y^2. This captures the idea of volatility being "persistent": large (small) values of y_t^2 are likely to be followed by large (small) values.

In general, GARCH models are the most appropriate models to use when one has to evaluate the volatility of the returns of groups of stocks with large amounts (thousands) of observations. GARCH models are becoming widely used in econometrics and finance because they have randomly varying volatility. The GARCH method is one of the techniques based on the assumption that the random component of the model shows changes in variability. The GARCH models were applied successfully in modeling the volatility of the variable in time series, with the applications being taken in large measure from the area of financial investments.

2.3 Bayesian inference

2.3.1 Fundamental of Bayesian Inference

Bayesian inference is an important technique in statistics, in which all forms of uncertainty are expressed in terms of probability. The Bayesian approach regards probability as a measure of the degree of personal belief about the value of an unknown parameter. Therefore, it is possible to ascribe probability to any event or proposition about which we are uncertain. In frequentist inference, parameters are not repeatable random things but are fixed (albeit unknown) quantities, which means that they can’t be considered as random variables. In contrast, in Bayesian inference anything about which we are uncertain, including the true value of a parameter, can be thought of as being a random variable which we can assign a probability distribution, known specifically as prior information. A fundamental feature of the Bayesian inference is the use of prior information in addition to the sample data. The core principle of Bayesian inference is shown in the following Figure:



Bayesian inference synthesises two sources of information about the unknown parameters of interest. The first of these is the sample data, expressed formally by the likelihood function. The second is the prior distribution, which represents additional information that is available to the investigator. Note that the likelihood function is also fundamental to frequentist inference, the prior distribution is used only in the Bayesian inference.

We represent the data by the symbol X and denote the set of unknown parameters by θ; then the likelihood function is $p(X|\theta)$, the probability of observing the data X being conditional on the values of the parameter θ. In addition, we further represent the prior distribution for θ as $\pi(\theta)$, giving the probability that θ takes any particular value based on whatever additional information might be available to the investigator. Then, with the application of Bayes' theorem, we can obtain a posterior distribution for θ, $p(X|\theta)$, which expresses what is now known about θ based on both the sample data and prior information.

Now we introduce the elementary Bayes' theorem. Suppose x and y are random variables, and $p(x,y)$ denotes the joint probability distribution function (PDF) of x and y, then

$$p(x,y) = p(x|y)\,p(y) = p(y|x)\,p(x), \tag{2.9}$$

is called the product rule. Moreover, if assuming y is a discrete random variable, then

$$p(x) = \sum_b p(x, y = b) = \sum_b p(x|y = b)\, p(y = b), \quad (2.10)$$

is called the sum rule. Combining the definition of conditional PDF with the product and sum rules, we can obtain the Bayes' rule:

$$p(x|y) = \frac{p(x, y)}{p(y)} = \frac{p(y|x)\, p(x)}{\sum_b p(x' = d)\, p(y' = b)}, \quad (2.11)$$

which is also called Bayes' theorem. Note that Bayes' theorem is also true for probability distribution functions, not just for probabilities (Bishop, 2006). Moreover, following standard Bayesian textbook (Gelman et al., 2003), we use the two words distribution and density interchangebly.

Based on the Bayes theorem, the posterior distribution over parameters having a set of observed values of realized random variables X is given by:

$$p(\theta|X) = \frac{p(X|\theta)\pi(\theta)}{P(X)} \propto p(X|\theta)\pi(\theta), \quad (2.12)$$

The proportionality symbol $\propto$ expresses the fact that the product of the likelihood function and the prior distribution must be scaled to integrate to one over the range of plausible θ values for it to be a proper probability distribution.

The posterior distribution for θ is a weighted compromise between the prior information and the sample data. In particular, if for some value of θ the likelihood, $p(X|\theta)$, is small, so that the data suggests that this value of θ is implausible, then the posterior distribution will also give small probability to this θ value. Similarly, if for some value of θ the prior distribution, $\pi(\theta)$ is small, so that

the prior information suggests that this value of θ is implausible, then, again, the posterior distribution will also give small probability to this θ value. Therefore, the posterior probability will be high for some θ only when both information sources support that value. The simple and intuitive nature of Bayes' theorem as a mechanism for synthesising information and updating personal beliefs about unknown parameters is an attractive feature of the Bayesian inference.

2.3.2 MCMC algorithm

To calculate $p(\theta|X)$, in practice, some approximation approach such as Markov Chain Monte Carlo is usually needed. Markov Chain Monte Carlo (MCMC) methods are a class of algorithms for sampling from a probability distribution based on constructing a Markov chain that has the desired distribution as its equilibrium distribution. The state of the chain after a number of steps is then used as a sample of the desired distribution (Chernozhukov and Hong, 2003).

Suppose that the specified distribution (the desired stationary distribution of the MCMC sampler we are constructing) has an un-normalized distribution $p(\theta)$. MCMC is an iterative procedure, such that given the current state of the chain, θ^i, the algorithm makes a probabilistic update to $\theta^{(i+1)} = q\left(\theta^{(i)}\right)$. Note that the update, $q(\cdot)$, is made in such a way that the distribution $p(\theta^i) \to p(\theta)$, the target distribution, as $i \to \infty$, for any starting value $\theta^{(0)}$. We shall review two of the most general procedures for MCMC simulation from a target distribution, namely, the Metropolis-Hastings algorithm and the Gibbs sampler.

2.3.3 Adaptive MCMC Algorithm

MCMC algorithms are extremely widely used in statistical inference, to

sample from complicated high-dimensional distributions. Tuning of associated parameters of the proposal density is crucial to achieve efficient mixing, but can also be very difficult to guarantee convergence to the desire distribution. Adaptive MCMC algorithms attempt to deal with this problem by automatically "learning" better parameter values of MCMC algorithms while they run. We will review the Adaptive Metropolis algorithm (Roberts and Rosenthal, 2009) in this section, which is the basic adaptive MCMC algorithm. A common adaptive MCMC algorithm is the adaptive random walk Metropolis. This method performs a Metropolis algorithm with proposal distribution at iteration j denoted by $q(\theta \mid \theta_j)$, which can be written as

$$q\left(\tilde{\theta}\middle|\theta_j\right) = \omega_{1j}\phi_d\left(\theta\middle|\theta_j, k_1\sum_1\right) + \omega_{2j}\phi_d\left(\theta\middle|\theta_j, k_2\sum_2\right), \qquad (2.13)$$

where d is the dimension of θ and $\phi_d\left(\theta\middle|\mu,\sum\right)$ is a multivariate d dimensional normal density in θ with mean μ and covariance matrix $\sum$. ω_{1j} is a preset parameter which belong the interval $[0,1]$, $\omega_{2j} = 1-\omega_{1j}$. $\sum_1$ is a constant covariance matrix, which is taken as the identity matrix. The matrix $\sum_{2j}$ is the sample covariance matrix of the first j iterates. The scalar k_1 is meant to achieve a high acceptance rate by moving the sampler locally, while the scalar k_2 is considered to be optimal for a random walk proposal when the target is a multivariate normal. The acceptance probability for the adaptive random walk Metropolis simplifies to

$$\alpha(\theta \mid \theta_j) = \min\left\{\frac{p(\theta)}{p(\theta_j)}\right\}.$$

2.3.4 Particle filter algorithm

The particle filters which are also known as sequential Monte Carlo

methods, are simulation based algorithms used to compute the high-dimensional and/or complex integrals that arise regularly in applied work. SMC is useful for inference on Latent Variable Model (LVM), where LVM will be introduced below. These methods are becoming increasingly popular in economics and finance; from dynamic stochastic general equilibrium models in macro-economics to option pricing. In this section, we will review the basic ideas and algorithm of particle filter.

We consider the general LVM which consists of two equations: the observation or measurement equation and the transition equation. They are respectively given by

$$\begin{aligned} y_n &= m_n\left(x_n, \varepsilon_n\right), \\ x_n &= h_n\left(x_{n-1}, \eta_n\right). \end{aligned} \tag{2.14}$$

The functions m_n and h_n are possibly nonlinear but of known form. These densities $p\left(y_n \middle| x_n; \theta\right)$ and $p\left(x_n \middle| x_{n-1}; \theta\right)$ corresponding to the above equations respectively are called the observation and transition densities. The sequence of state variables are generally unobserved and it is the aim of the researcher to estimate them using the observed data. Uncertainty about the state variable is formulated as a joint conditional probability distribution $p\left(x_{0:n} \middle| y_{1:n}; \theta\right)$ known as the joint smoothing distribution. It is defined as

$$p\left(x_{0:n} \middle| y_{1:n}; \theta\right) = \frac{p\left(x_{0:n}, y_{1:n}; \theta\right)}{p\left(y_{1:n}; \theta\right)},$$

where the constant of integration $p\left(y_{1:n}; \theta\right)$ is the likelihood of the state space model. Consider approximating the entire joint distribution $p\left(x_{0:n} \middle| y_{1:n}; \theta\right)$. Given a function f of the state variable, a standard Monte Carlo estimator of the integral

$$E\left[f\left(x_{0:n}\middle|y_{1:n}\right)\right] = \int f\left(x_{0:n}\right) p\left(x_{0:n}\middle|y_{1:n};\theta\right) dx_{0:n}, \qquad (2.15)$$

consists of drawing sequences $x_{0:n}$ directly from the target distribution $p\left(x_{0:n}\middle|y_{1:n};\theta\right)$. However, this strategy is generally impossible for complex models because the target distribution is non-standard and it is unknown how to draw directly from it. Therefore, we use the particle filter to sequentially sample from the posterior probability.

2.4 Support vector regression

Support Vector Regression (SVR), as a common statistical learning method, has been applied in various fields – time series and financial (noisy and risky) prediction, approximation of complex engineering analyses, convex quadratic programming and choices of loss functions, etc. (Basak, Pal, Patranabis, 2007). The SVR based on the computation of a linear regression function in a high dimensional feature space where the input data are mapped via a nonlinear function.

2.4.1 The basic idea

Statistical Learning Theory has provided a very effective framework for classification and regression tasks involving features. Support Vector Regression (SVR) is directly derived from this framework and works by solving a constrained quadratic problem where the convex objective function for minimization is given by the combination of a loss function with a regularization term (the norm of the weights).

Traditional statistical regression procedures are often stated as the processes deriving a function $f(x)$ that has the least deviation between predicted and experimentally observed responses for all training examples. One of the main

characteristics of Support Vector Regression (SVR) is that instead of minimizing the observed training error, SVR attempts to minimize the generalized error bound so as to achieve generalized performance. This generalization error bound is the combination of the training error and a regularization term that controls the complexity of the hypothesis space. Support Vector Regression (SVR) is the most common application form of SVMs. An overview of the basic ideas underlying support vector machines (SVM) for regression and function estimation has been given in Smola, and Schölkopf, 1998. Support vector regression (SVR) is a powerful technique for predictive data analysis.

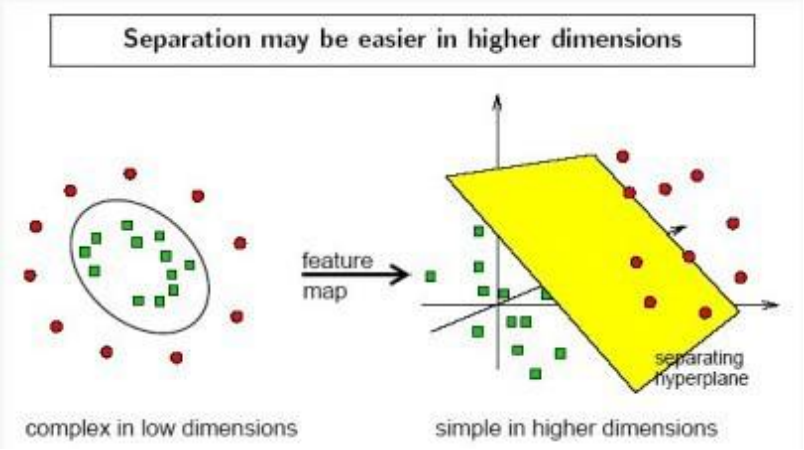


Figure 2.1 The nonlinear mapping from low dimensional feature space to high dimensional feature space.

The main idea of SVR is to construct a hyper plane as the decision surface. It performs by nonlinearly mapping the input space into a high dimensional feature space and then runs the linear regression in the output space. The nonlinear mapping process is intuitively shown in Figure 2.1. The Support Vector Regression (SVR) uses

the same principles as the SVM for classification, with only a few minor differences. First of all, because the output is a real number it becomes very difficult to predict the information at hand, which has infinite possibilities. In the case of regression, the SVR (Support Vector Regression) algorithm is more complicated therefore to be taken in consideration. However, the main idea of SVM and SVR is always the same: to minimize error, individualizing the hyperplane which maximizes the margin.

2.4.2 The estimation method

Let $L=\{(x_1,y_1),...,(x_L,y_L)\}$ be a set of training data, where $x_i \in R^n$ denotes the value of input variables, $y_i \in R$ the corresponding output value for $i=1,2,...,l$, where l corresponds to the size of the training dataset, By using non-linear transformation φ from R^n to a high dimensional space F to structure optimal decision function, that is $f(x)=(w\cdot\varphi(x))+b,\ \varphi: R^n \to F,\ w\in F$, where b denote threshold value.

Using structural risk minimization criterion and introducing slack variables ξ_i, ξ_i^*, w and b can be estimated by minimizing the following optimization problem:

$$\min_{w\in F} \quad \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}(\xi_i+\xi_i^*)$$
$$s.t. \quad \begin{cases} y_i-(w\cdot\varphi(x_i))-b\le\varepsilon+\xi_i, \\ (w\cdot\varphi(x))+b-y_i\le\varepsilon+\xi_i^* \\ \xi_i,\xi_i^*\ge 0 \end{cases} \tag{2.16}$$

where the constants $C>0$ and $\varepsilon>0$ are specified parameters in advance. C is referred to as a regularized constant and it determines the trade-off between model complexity and training error, ε is known as channel width which controls the

number of support vectors. By introducing Lagrange multiplier a_i and a_i^* to structure Lagrange function, we can obtain the dual representation:

$$\min_{a_i, a_i^* \in R} \quad \frac{1}{2}\sum_{i,j=1}^{l}(a_i - a_i^*)(a_j - a_j^*)\varphi(x_i)\cdot\varphi(x_j) + \sum_{i=1}^{l} a_i(\varepsilon - y_i) + \sum_{i=1}^{l} a_i^*(\varepsilon - y_i)$$
$$s.t. \quad \begin{cases} \sum_{i=1}^{l}(a_i - a_i^*) = 0 \\ a_i, a_i^* \in [0, C] \end{cases}, \qquad (2.17)$$

where y_i is the corresponding output value of training data set. Thus, the SVR problem can be converted into a quadratic programming problem, and the parameter vector w can be written in terms of data as follows:

$$w = \sum_{i=1}^{l}(a_i - a_i^*)\varphi(x_i), \qquad (2.18)$$

the regression function of SVR can be represented in the following equation:

$$\begin{aligned} f(x) &= \sum_{i=1}^{l}(a_i - a_i*)(\varphi(x_i)\cdot\varphi(x)) + b \\ &= \sum_{i=1}^{l}(a_i - a_i*)k(x_i, x) + b \end{aligned}, \qquad (2.19)$$

where $k(x_i, x) = \varphi(x_i)\cdot\varphi(x)$ known as the kernel function. Common kernel functions include the following several kinds:

Linear kernel function: $k(x, x_i) = x \cdot x_i$;

Polynomial kernel function: $k(x, x_i) = (x \cdot x_i)^d$;

RBF kernel function: $k(x, x_i) = \exp(-\gamma \|x - x_i\|^2),\ \gamma > 0$;

Gaussian RBF kernel function: $k(x, x_i) = \exp(-\|x - x_i\|^2 / 2\sigma^2)$.

The kernel functions transform the data into a higher dimensional feature space to make it possible to perform the linear separation. Figure 2.2 demonstrates

how the kernel function $k(x, x_i)$ transform the data into a linear separable higher dimensional feature space. Although requiring flatness only in feature space, one can observe that the functions also are very flat in input space. This is due to the fact, that kernels can be associated with flatness properties via regularization operators.

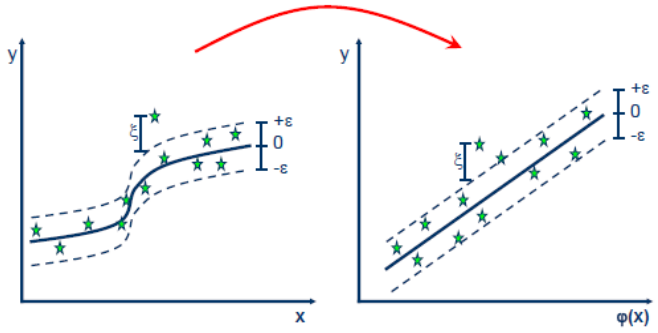


Figure 2.2 The kernel function transforms process.

In this thesis, the SVR model is used to integrate the volatility process of IMF components. However, the true volatility process is usually unobservable, so it is difficult to train the SVR. We propose to use the volatility estimation, $\sigma_t^2 = (R - \bar{R})^2$, as a proxy of the true volatility, where $\bar{R}$ denote the mean of returns. The proxy volatility is able to describe the volatility process of the returns series, but it can not predict the volatility.

CHAPTER III

VALUE AT RISK ESTIMATION UNDER STOCHASTIC VOLATILITY MODELS USING ADAPTIVE PMCMC METHODS

In this chapter, we propose a value-at-risk (VaR) estimation technique based on a new stochastic volatility model with leverage effect, non-constant conditional mean and jumps. In order to estimate the model parameters and latent state variables, we integrate the particle filter and adaptive MCMC algorithms to develop a novel adaptive particle MCMC (A-PMCMC) algorithm. Comprehensive simulation experiments based on three stock indices and two foreign exchange time series show the effectiveness of the A-PMCMC algorithm and the proposed VaR estimation technique.

3.1 Stochastic volatility model with leverage effects and jump

3.1.1 Model description

Risk in the sense of the possibility of losses is an inherent ingredient of financial markets. To measure and monitor risk as accurately as possible has become a competitive factor for financial institutions. Value-at-risk (VaR) (Choudhry, 2013), a quantile measure, has become the preferred tool in the financial industry. The importance of VaR was further highlighted after it was used by the central banks to

govern and supervise the capital adequacy of financial institutions in the Group of Ten countries in 1995. The selection of the internal VaR model as well as the volatility estimation is essential to the VaR based risk management.

Volatility which is the most essential parameter in VaR estimation, can be considered as the temperature of the market and as such can change rapidly. Risk management approaches which do not take these fluctuations into account tend to under- or overestimate risk depending on the current market situation. The key idea of using stochastic volatility in risk management is to devolatilize the observed return series and to forecast future returns (Eberlein et al., 2003). Stochastic volatility models have gradually emerged as a useful way of modeling time-varying volatility with significant potential for applications, especially in finance (Taylor, 1994; Shephard, 1996; Ghysels et al., 1996). In this thesis, we propose a new stochastic volatility model with leverage effect, jumps and non-constant conditional mean to describe the real world stochastic volatility process.

The leverage effect has been investigated in a stochastic volatility framework (Yu, 2005). This leverage effect describes the relationship between returns and conditional variances. It means that bad news in the markets lead to a boost on the variance. The underlying reason is that bad news tend to decrease price thus leading to an increase in debt-to-equity ratio. Furthermore, the leverage effect is important to explain some characteristics of the data on financial derivatives. Conditional mean is regarded as the predictable component of the returns and is important in applications related to optimal portfolio choices. Economic theory shows that an investor gains from market predictability and volatility timing (Merton, 1971). In order to accurately describe the stock returns process, we model the conditional expected value of the

returns together with the dynamic of the volatilities.

The conditional mean of returns is modeled via an autoregressive process, which can be explained by the non-synchronous trading and unexpected stochastic dividend effects (Chernov et al., 2003). In addition, there is substantial evidence in favor of jumps on returns and volatilities (Merton, 1976; Eraker et al., 2003; Raggi, 2005). Jumps can basically be described as rare events; large, infrequent movements in returns which are an important feature of financial markets. These have been widely documented to be important in characterizing the non-Gaussian tail behaviour of conditional distribution of returns.

The stochastic volatility model with leverage effect, jumps and non-constant conditional mean is a Gaussian non-linear state-space model, which contains unknown parameters and unobserved latent state variables. In this complicated model, it is highly non-straightforward to efficiently and simultaneously estimate the model parameters and latent state variables. In existing literatures, Bayesian MCMC methods are commonly applied for inferences on stochastic volatility models. A general stochastic volatility model is as follows:

$$y_t = G(X_t; \Theta; \varepsilon_t), \tag{3.1}$$

$$X_t = H(X_{t-1}; \Theta; \eta_t), \tag{3.2}$$

where X_t is the latent state vector, Θ is the unknown parameter vector, ε_t and η_t denote the innovation terms. Given the prior information about X_t and Θ, the goal of the Bayesian analysis is to learn about the unknown parameters and latent variables from the augmented posterior distribution

$$\pi(\mathrm{X}, \Theta \mid y) \propto \pi(\Theta) \prod_{t=1}^{T} f(y_t \mid X_t, \Theta) f(X_t \mid X_{t-1}, \Theta). \tag{3.3}$$

The Bayesian MCMC inference on the parameters and latent variables are conducted by producing a sample $\left\{X^i, \Theta^i\right\}$ from this augmented posterior density. In particular, the MCMC algorithm can be used for the stochastic volatility models with jumps process (Chib et al., 2002; Eraker et al., 2003; Li et al., 2008; Golightly, 2011). The main limitation of MCMC in applications is that every time we observe new data, we have to restart the inference procedure from scratch, which is causing a considerable amount of wasted time.

Particle filters, also known as sequential Monte Carlo methods, can reduce the computational burden required for estimating latent state variables. The great advantage of particle filtering is that in many implementations we simply simulate forward in time from our data generating process for the state variables. The typical particle filter methods include the Sampling Importance Re-sampling (SIR) particle filter (Gordon et al., 1993), the auxiliary particle filter (Pitt and Shephard, 1999), Liu and West filter (Liu and West, 2001), Storvik filter (Storvik, 2002) and Particle learning (Carvalho et al., 2010). However as mentioned above, particle filters are efficient for inference on the unobserved latent processes given known parameter values, but struggle when dealing with unknown parameters.

Recently, an advanced method called “Particle MCMC” (PMCMC) (Andrieu et al., 2010) was proposed. PMCMC involves using a particle filter within an MCMC algorithm, which takes advantage of the strength of its two components. It is most naturally applied to inference on state space models, where there is an unobserved stochastic process. The idea of PMCMC is that the particle filter will update the unobserved process given specific values for the parameters, and MCMC moves will be used to update the parameter values. The PMCMC algorithm has

already been applied in some areas (Peters et al., 2000; Rasmussen et al., 2011; Golightly and Wilkinson, 2011; Pitt et al., 2012; Wood et al., 2014). There is still room to improve the MCMC, however, due to a large Monte Carlo error in the particle filter, and due to slow mixing of the MCMC moves. Therefore, we further integrate an adaptive MCMC algorithm (Andrieu and Thoms, 2008) to PMCMC in order to improve the performance of PMCMC. The adaptive MCMC samplers use information obtained from previous draws to tune the proposal distribution automatically and repeatedly. The new algorithm is called A-PMCMC.

To summarize, our contributions are three fold. Firstly, we propose a new volatility model including three important features which appear in the real-world data. Secondly, we propose an efficient A-PMCMC algorithm which is most naturally applied to inference on state space models. It is capable of simultaneously estimating parameters and inferencing the latent state variables. Thirdly, we apply the A-PMCMC algorithm to a VaR estimation technique based on a new stochastic volatility model. To prove the efficiency of our method, we apply the proposed VaR technique to predict the single-period and multi-period VaR for stock index and exchange rate returns. The back-testing results illustrate that the proposed VaR estimation technique consistently outperforms the compared methods.

The remainder of this chapter is organized as follows. The stochastic volatility model is described in Section 3.1.2. Section 3.2-3.3 proposes Bayesian inference for A-PMCMC and the proposed VaR technique. Section 3.4 conducts a Monte Carlo simulation experiment to validate the availability of A-PMCMC method for volatility estimation. Section 3.5 illustrates the proposed VaR estimate technique using stock indices and exchange rate daily data and demonstrates its effectiveness via

Back-testing.

3.1.2 Model deformation

In order to capture most characteristics of the financial time series data, we extend the standard stochastic volatility model as

$$y_t = \mu_t + \varepsilon_t \exp(h_t / 2) + J_t \varpi_t , \tag{3.4}$$

$$h_{t+1} = \theta + (h_t - \theta)\phi + \sigma_\eta \eta_t , \tag{3.5}$$

$$\mu_{t+1} = \alpha + (\mu_t - \alpha)\beta + \sigma_\zeta \zeta_t , \tag{3.6}$$

where

$$\begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \sim N(0, \Sigma), \ \Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \tag{3.7}$$

and

$$J_t \sim \text{Bernoulli}(\lambda), \ \varpi_t \sim N(0, \sigma_J), \ \zeta_t \sim N(0,1). \tag{3.8}$$

The novelty of the proposed model is that it expresses three essential features of return process, i.e., stochastic volatility, conditional mean and jump. To our knowledge, we are the first to propose a model which includes all three features. One drawback of our model is that it is quite complicated, so that we have to invent an efficient procedure to make a statistical inference, as explained in Section 3.2. In this specification, y_t is the observed asset log-return at time t, i.e. $y_t = \log p_t - \log p_{t-1}$, where p_t is an asset price. The conditional mean μ_t and the log-volatility h_t are described by two non observable latent processes. In order to properly describe extreme events such as crashes in the markets, we allow jumps in the return process. J_t is the time-t jump arrival where its value $j \in \{0,1\}$ is a Bernoulli counter with

intensity λ. ϖ_t dictates the jump size. μ_t is the drift component of the volatility process and σ_η can be interpreted as the volatility of the volatility factor η_t. The parameter ϕ is the persistence of the volatility process that allows the volatility clustering (Yu, 2005). Typically we would impose that $|\phi|<1$, so that we have a stationary process with the initial condition that

$$h_0 \sim N(\theta, \sigma_\eta^2/(1-\phi^2)). \tag{3.9}$$

We assume that $Cov(\varepsilon_t, \eta_t)$ are correlated with correlation ρ, which can describe the leverage effect. Since the disturbances are conditionally Gaussian, we can write $\eta_t = \rho\varepsilon_t + \sqrt{(1-\rho^2)}\,\xi_t$, where $\xi_t \sim N(0,1)$. Therefore, the volatility equation (3.5) can then be reformulated as

$$h_{t+1} = \theta + (h_t - \theta)\phi + \sigma_\eta \rho \varepsilon_t + \sigma_\eta \sqrt{(1-\rho^2)}\xi_t. \tag{3.10}$$

In Eq. (3.6), we also directly model the conditional mean of the returns via an autoregressive process μ_{t+1}. α is the drift component of the mean process and σ_ζ can be interpreted as the volatility of the innovation term ζ_t. The parameter β is the persistence of the regression process and it is assumed that $|\beta|<1$. As for the conditional variance, the conditional mean at time 0 is distributed as,

$$\mu_0 \sim N(\alpha, \sigma_\zeta^2/(1-\beta^2)). \tag{3.11}$$

We emphasize that the serial correlation on the returns induced by μ_{t+1} can be explained by the non-synchronous trading and unexpected stochastic dividend effects (Chernov et al., 2003). We assume that the noise process ζ_t is uncorrelated with ε_t and η_t. From the above discussion, the stochastic volatility model with jumps

and leverage effects can be reformulated the Gaussian nonlinear state space form

$$\begin{cases} y_t = \mu_t + \varepsilon_t \exp(h_t/2) + J_t \varpi_t \\ h_{t+1} = \theta + (h_t - \theta)\phi + \sigma_\eta \rho \varepsilon_t + \sigma_\eta \sqrt{(1-\rho^2)}\xi_t \\ \mu_{t+1} = \alpha + (\mu_t - \alpha)\beta + \sigma_\zeta \zeta_t \end{cases}, \tag{3.12}$$

where

$$\varepsilon_t, \xi_t, \zeta_t \sim iid\ N(0,1),\ J_t \sim \text{Bernoulli}(\lambda),\ \varpi_t \sim N(0, \sigma_J^2).$$

In our state space model, the complete specification of the latent state includes not just only the unknown h_t and μ_t but also any random variable generated in the process of obtaining the h_t and μ_t that may have an influence on the distribution of y_t. We thus have the latent state $\tilde{X}_t = (h_t, \mu_t, J_t, \varepsilon_t)$. In order to simplify the inference process of unknown parameters and latent variables, we make the following independent assumptions: (i) J_t is unconditionally independent with h_t and μ_t; (ii) ε_t is unconditionally independent with J_t, h_t and μ_t.

3.2 Adaptive particle MCMC algorithm

In order to simplify the description of the Adaptive Particle MCMC (A-PMCMC) algorithm, we assign the latent state vector $\tilde{X}_t = (X_t, J_t, \varepsilon_t)$, where $X_t = (h_t, \mu_t)$, and the parameter vector $\Theta = (\lambda, \sigma_J, \theta, \phi, \sigma_\eta, \rho, \alpha, \beta, \sigma_\zeta)$.

3.2.1 Inference state variables with known parameters: particle filtering

In the first step, assuming that Θ is known, we use a particle filter to approximate the latent state variables. From the state space model (3.12), we obtain the measurement density $f(y_t \mid X_t; \Theta)$ and the Markov transition density

$f(X_{t+1} \mid X_t; \Theta)$. As the parameter vector Θ is considered to be known, the measurement density and the transition density are rewritten as $f_\Theta(y_t \mid X_t)$ and $f_\Theta(X_{t+1} \mid X_t)$ respectively. The particle filter is an algorithm to propagate and update the Monte Carlo samples, usually called "particles", in order to obtain a new sample which is approximately distributed as the posterior density $f_\Theta(X_t \mid y_{1:t})$, in which $y_{1:t} = (y_1, y_2, ..., y_t)$ denotes the past history of the observable process. The posterior, or filtering, density can be formulated as

$$\begin{aligned} f_\Theta(X_t \mid y_{1:t}) &\propto f_\Theta(y_t \mid X_t) f_\Theta(X_t \mid y_{1:t-1}) \\ &= f_\Theta(y_t \mid X_t) \int f_\Theta(X_t \mid X_{t-1}) f_\Theta(X_{t-1} \mid y_{1:t-1}) dX_{t-1}. \end{aligned}$$

In order to sample from this filtering density we employ the SIR (Sampling Importance Re-sampling) particle filter algorithm (Gordon et al., 1993). The details are illustrated in Algorithm 1 below.

Algorithm 1: SIR particle filter algorithm for latent states sampling.

Draw $h_1^i \sim f_\Theta(h_1)$ and $\mu_1^i \sim f_\Theta(\mu_1)$ for $i = 1, ..., M$.

For $t = 1:T-1$

Step 1: sample $\{\varepsilon_t^i\}_{i=1}^M \sim f_\Theta(\varepsilon_t \mid h_t^i, \mu_t^i, y_t)$;

Step 2: sample $\{\tilde{h}_{t+1}^i\}_{i=1}^M \sim f_\Theta(h_{t+1} \mid h_t^i, \varepsilon_t^i)$;

Step 3: sample $\{\tilde{\mu}_{t+1}^i\}_{i=1}^M \sim f_\Theta(\mu_{t+1} \mid \mu_t^i)$;

Step 4: calculate normalized weights,

$\{\omega_{t+1}^i\}_{i=1}^M = \dfrac{\tilde{\omega}_{t+1}^i}{\sum_{k=1}^K \tilde{\omega}_{t+1}^i}$, *where* $\tilde{\omega}_{t+1}^i = f_\Theta(y_{t+1} \mid \tilde{h}_{t+1}^i, \tilde{\mu}_{t+1}^i)$.

Step 5: sample $\{h_{t+1}^i\}_{i=1}^M \sim \sum_{k=1}^M \omega_{t+1}^k \delta(h_{t+1} - \tilde{h}_{t+1}^k)$ and $\{\mu_{t+1}^i\}_{i=1}^M \sim \sum_{k=1}^M \omega_{t+1}^k \delta(\mu_{t+1} - \tilde{\mu}_{t+1}^k)$.

Here $\delta(\cdot)$ is the Dirac-delta function. This will yield an approximation of the desired posterior density, $f_\Theta(X_t \mid y_{1:t})$ as t varies. In Algorithm 1, the density function $f_\Theta(\varepsilon_t \mid h_t, \mu_t, y_t)$ in Step 1 is a mixture of the form

$$f_\Theta(\varepsilon_t \mid h_t, \mu_t, y_t) = \sum_{j=0}^{1} f_\Theta(\varepsilon_t \mid J_t = j; h_t, \mu_t, y_t) \times P(J_t = j \mid h_t, \mu_t, y_t). \tag{3.13}$$

Using the independence assumption stated at the end of Section 2, the conditional probability of a jump is given by

$$P(J_t = 1 \mid h_t, \mu_t, y_t) = \frac{N(y_t \mid \mu_t, \exp(h_t) + \sigma_J^2) \times \lambda}{N(y_t \mid \mu_t, \exp(h_t) + \sigma_J^2) \times \lambda + N(y_t \mid \mu_t, \exp(h_t)) \times (1-\lambda)}. \tag{3.14}$$

Here we denote $p_t = P(J_t = 1 \mid h_t, \mu_t, y_t)$. Hence

$$P(J_t = 0 \mid h_t, \mu_t, y_t) = 1 - P(J_t = 1 \mid h_t, \mu_t, y_t) = 1 - p_t. \tag{3.15}$$

Now we have

$$\begin{aligned} f_\Theta(\varepsilon_t \mid J_t = 1, h_t, \mu_t, y_t) &\propto N(y_t \mid \mu_t + \varepsilon_t \exp(h_t/2), \sigma_J^2) \times N(\varepsilon_t \mid 0,1) \\ &\propto \exp\left(\frac{(y_t - \mu_t - \varepsilon_t \exp(h_t/2))^2}{\sigma_J^2} + \varepsilon_t^2 \right) \\ &\propto \exp\left(\frac{(\varepsilon_t - (y_t - \mu_t)\exp(h_t/2)/(\exp(h_t) + \sigma_J^2))^2}{\sigma_J^2/(\exp(h_t) + \sigma_J^2)} \right). \end{aligned} \tag{3.16}$$

We hence establish that

$$f_\Theta(\varepsilon_t \mid J_t = 1, h_t, \mu_t, y_t) = N(\upsilon_{\varepsilon_t}, \sigma_{\varepsilon_t}^2), \tag{3.17}$$

where

$$\upsilon_{\varepsilon_t} = \frac{(y_t - \mu_t)\exp(h_t/2)}{\exp(h_t) + \sigma_J^2}, \ \sigma_{\varepsilon_t}^2 = \frac{\sigma_J^2}{\exp(h_t) + \sigma_J^2}.$$

If the process does not jump, there is a Dirac-delta mass at the point $\varepsilon_t = (y_t - \mu_t)\exp(-h_t/2)$, hence we can establish that

$$f_\Theta(\varepsilon_t \mid J_t = 0, h_t, \mu_t, y_t) = \delta((y_t - \mu_t)\exp(-h_t/2) - \varepsilon_t). \tag{3.18}$$

According to (3.13)-(3.18), this mixture is

$$f_{\Theta}(\varepsilon_t \mid h_t, \mu_t, y_t) = \delta((y_t - \mu_t)\exp(-h_t/2) - \varepsilon_t) \times (1-p_t) + N(\varepsilon_t \mid \upsilon_{\varepsilon_t}, \sigma^2_{\varepsilon_t}) \times p_t \,. \quad (3.19)$$

Moreover, the non-normalized weights for $\{\tilde{\omega}^i_{t+1}\}^M_{i=1}$ in step 4 are of the form

$$f_{\Theta}(y_t \mid h_t, \mu_t) = N(y_t \mid \mu_t, \exp(h_t)) \times (1-\lambda) + N(y_t \mid \mu_t, \exp(h_t) + \sigma^2_J) \times \lambda \,. \quad (3.20)$$

3.2.2 Maximum likelihood estimation

The previous section shows an inference process given that Θ is known. In the next section, we will show how to make an inference with unknown parameters Θ. To do that, one important ingredient is the calculation of likelihood, which shall be explained here. The log-likelihood of Θ is

$$\log L(\Theta) = \log f(y_1, ..., y_T \mid \Theta) = \sum\nolimits_{t=1}^{T} \log f(y_{t+1} \mid \Theta; y_{1:t}) \,. \quad (3.21)$$

In order to estimate this function, via the prediction decomposition (Harvey, 1993), we employ the predictive density

$$f(y_{t+1} \mid \Theta; y_{1:t}) = \int f(y_{t+1} \mid X_{t+1}; \Theta) f(X_{t+1} \mid y_{1:t}; \Theta) dX_{t+1} \,. \quad (3.22)$$

In order to estimate the integral, we sample from the transition density $f(X_{t+1} \mid X_t; \Theta)$. Then we estimate the predictive density as

$$\tilde{f}(y_{t+1} \mid \Theta; y_{1:t}) = \frac{1}{M} \sum\nolimits_{k=1}^{M} f(y_{t+1} \mid \tilde{X}^k_{t+1}; \Theta) = \frac{1}{M} \sum\nolimits_{k=1}^{M} \tilde{\omega}^k_{t+1} \,. \quad (3.23)$$

The estimation of the log-likelihood $\log L(\Theta)$ is therefore a by-product of a single run of the particle filter. Then the estimator for the log-likelihood would therefore be

$$\log \tilde{L}(\Theta) = \sum\nolimits_{t=1}^{T} \log \tilde{f}(y_t \mid \Theta; y_{1:t-1}) = \sum\nolimits_{t=1}^{T} \log(\frac{1}{M} \sum\nolimits_{k=1}^{M} \tilde{\omega}^k_t) \,. \quad (3.24)$$

3.2.3 Inference with unknown parameters: A-PMCMC algorithm

As mentioned, the particle filter is efficient for inference on the unobserved process given known parameter values of Θ, but struggle when dealing with unknown parameters. Therefore, in this work, we invent the improved adaptive PMCMC(A-PMCMC) methods to update the parameter values. To simultaneously sample both latent states and parameters, we consider the following density

$$f(\Theta, X_{1:T} \mid y_{1:T}) = f(\Theta \mid y_{1:T}) f_{\Theta}(X_{1:T} \mid y_{1:T}).$$

We suggest the following form of proposal density for a Metropolis-Hastings update

$$q((\Theta^*, X_{1:T}^*) \mid (\Theta, X_{1:T})) = q(\Theta^* \mid \Theta) f_{\Theta^*}(X_{1:T}^* \mid y_{1:T}),$$

for which the proposed $X_{1:T}^*$ is perfectly adapted to the proposed Θ^*. The resulting Metropolis-Hastings acceptance ratio is given by

$$\begin{aligned} \alpha(\Theta, X_{1:T}; \Theta^*, X_{1:T}^*) &= \frac{f(\Theta^*, X_{1:T}^* \mid y_{1:T})}{f(\Theta, X_{1:T} \mid y_{1:T})} \times \frac{q((\Theta, X_{1:T}) \mid (\Theta^*, X_{1:T}^*))}{q((\Theta^*, X_{1:T}^*) \mid (\Theta, X_{1:T}))} \\ &= \frac{f_{\Theta^*}(y_{1:T}) p(\Theta^*)}{f_{\Theta}(y_{1:T}) p(\Theta)} \times \frac{q(\Theta \mid \Theta^*)}{q(\Theta^* \mid \Theta)}. \end{aligned} \tag{3.25}$$

The expression for this ratio suggests that the algorithm effectively targets the marginal density $f(\Theta \mid y_{1:T}) \propto f_{\Theta}(y_{1:T}) f(\Theta)$ (Andrieu and Roberts, 2009), where we can use the likelihood formula explained in the previous section. In order to improve the performance of MCMC algorithm, we further employ the adaptive Metropolis-Hastings sampling for $q(\Theta^* \mid \Theta)$ (Roberts and Rosenthal, 2009). It is defined as

$$q(\Theta^* \mid \Theta_{i-1}) = \tau_1 N_d(\Theta^* \mid \Theta_{i-1}, \kappa_1 \Sigma_1) + \tau_2 N_d(\Theta^* \mid \Theta_{i-1}, \kappa_2 \Sigma_2), \tag{3.26}$$

where i is the iterations, d is the dimension of Θ and $N(\Theta^* \mid \Theta_{i-1}, \Sigma)$ is a multivariate d dimensional normal density in Θ^* with mean Θ_{i-1} and covariance

matrix Σ ; τ_1, τ_2, κ_1 and κ_2 are the system setting parameters. By the symmetric property of Eq. (3.26), we note that the acceptance probability for the adaptive Metropolis-Hastings simplifies to

$$\alpha(\Theta, X_{1:T}; \Theta^*, X_{1:T}^*) = \min\left\{1\ ,\ \frac{f_{\Theta^*}(y_{1:T})\,p(\Theta^*)}{f_{\Theta}(y_{1:T})\,p(\Theta)}\right\}. \tag{3.27}$$

To summarize, the A-PMCMC method is described in Algorithm 2:

Algorithm 2: Adaptive PMCMC methods for parameter values estimation

---

Step 1: set the prior distribution for the parameter vector Θ and to initial value Θ_0;

For $n = 1:N$

Step 2: sample $\Theta^* \sim q(\Theta^* \mid \Theta_{i-1})$;

Step 3: run an SIR particle filter algorithm targeting $\tilde{f}_{\Theta^*}(X_{1:T} \mid y_{1:T})$, sample

$X_{1:T}^* \sim \tilde{f}_{\Theta_*}(X_{1:T} \mid y_{1:T})$ and $\tilde{f}_{\Theta^*}(y_{1:T})$.

Step 4: with probability

$$\alpha(\Theta, X_{1:T}; \Theta^*, X_{1:T}^*) = \min\left\{1\ ,\ \frac{\tilde{f}_{\Theta^*}(y_{1:T})\,p(\Theta^*)}{\tilde{f}_{\Theta}(y_{1:T})\,p(\Theta)}\right\} \tag{3.28}$$

set $\Theta_i = \Theta^*$, $X_{1:T}^i = X_{1:T}^*$ and $\tilde{f}_{\Theta_i}(y_{1:T}) = \tilde{f}_{\Theta^*}(y_{1:T})$; otherwise

set $\Theta_i = \Theta_{i-1}$, $X_{1:T}^i = X_{1:T}^{i-1}$ and $\tilde{f}_{\Theta_i}(y_{1:T}) = \tilde{f}_{\Theta_{i-1}}(y_{1:T})$.

Note that under mild assumptions the acceptance probability (3.28) converges to equation (3.27) as $N \to \infty$(Andrieu et al., 2010). Note also that, in Step 3, running an SIR particle filter algorithm, we can only obtain $f_{\Theta^*}(X_t \mid y_{1:t})$. By means of $f_{\Theta^*}(X_t \mid y_{1:t}) \propto f_{\Theta^*}(y_t \mid X_t) f_{\Theta^*}(X_t \mid X_{t-1}) f_{\Theta^*}(X_{1:t-1} \mid y_{1:t-1})$, we can get $\tilde{f}_{\Theta^*}(X_{1:T} \mid y_{1:T})$

iteratively.

3.3 Dynamic Value at Risk

3.3.1 Dynamic Value at Risk definition

Value-at-Risk estimates often move through time. To make the Value-at-Risk estimation more precise, we need to use dynamic strategies to measure the Value-at-Risk. The dynamic strategy can be most appropriately illustrated by applying it to a real situation in portfolio management. The market conditions are changing every day. That is why we must adapt the Value-at-Risk estimation to the daily, weekly or monthly fluctuations. The problem of dynamic Value-at-Risk deals with the questions, how one can define the general trading rules and build a single adaptation scheme for risk estimations. Generally speaking, the problems of a dynamic Value-at-Risk measurement are difficult to solve, but very interesting and still actual in modern risk research. (Rogachev Andrey, 2002.)

3.3.2 Dynamic Value at Risk estimation

Value at risk (VaR) is one of the most often used risk measurement in finance. It measures the possible loss level over a given horizon at a given confidence level $1-\tau$. Mathematically VaR at τ probability level is defined as:

$$VaR_{\tau,t} = F_t^{-1}(\tau), \tag{3.28}$$

where F_t^{-1} is the inverse function of the conditional cumulative distribution function $F_{\Theta}(y_t \mid X_t)$, and as shown in previous sections, the corresponding probability density function $f_{\Theta}(y_t \mid X_t)$ is

$$f_{\Theta}(y_t \mid X_t) = p_t \left\{ 2\pi(\exp(h_t) + \sigma_J^2) \right\}^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2}(y_t - u_t)^2 (\exp(h_t) + \sigma_J^2)^{-1} \right\} + (1 - p_t) \left\{ 2\pi \exp(h_t) \right\}^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2}(y_t - u_t)^2 \exp(-h_t) \right\}. \quad (3.29)$$

Based on the A-PMCMC algorithm, we can describe the dynamic VaR estimation technique as follows:

Algorithm 3: A-PMCMC Value at Risk estimation

Step 1: conduct the algorithm 1 and algorithm 2 to on line estimate the latent state X_t and parameters Θ;

Step 2: calculate the conditional probability of a jump for predict $\tilde{J}_{t+1}$ by Eq. (3.14) and predict the $\tilde{h}_{t+1}$ and $\tilde{\mu}_{t+1}$ by Eq. (3.5) and Eq. (3.6) respectively;

Step 3: compute $F_{\Theta}(y_{t+1} \mid X_{t+1})$ by Eq. (3.29);

Step 4: estimate $VaR_{\tau,t+1} = F_{t+1}^{-1}(\tau)$.

3.4 Simulation experiment

To understand the capabilities of both PMCMC and the improved A-PMCMC methods, we illustrate their performance on learning parameters from synthetic data. The synthetic data were simulated from the model with the following the parameters

Volatility process: $\theta = 0.08,\ \phi = 0.975,\ \sigma_\eta = 0.2,\ \rho = -0.3$;

Conditional mean: $\alpha = 0.001,\ \beta = 0.95,\ \sigma_\zeta = 0.1$;

Jump process: $\lambda = 0.01,\ \sigma_J = 2$.

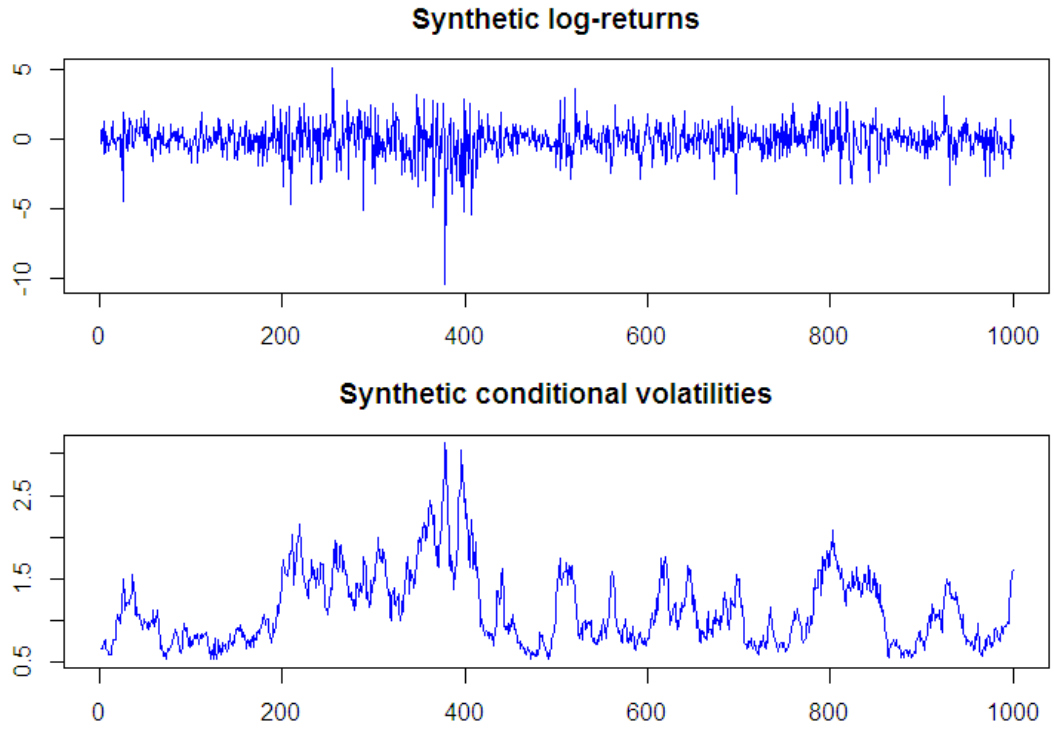


Figure 3.1 Synthetic log-returns and synthetic conditional volatilities.

These parameter values are consistent with empirical findings on similar jump diffusion models (Chib et al., 2002).

In the simulation procedure, we set observations numbers $T = 1000$, particles numbers $\mathrm{M} = 5000$, MCMC iterations $N = 3000$, initial iterations $i_0 = 1000$. Following Roberts and Rosenthal (2001), we set $\tau_1 = 1$ for $i \le i_0$, with i_0 representing the initial iterations, $\tau_1 = 0.05$ for $i > i_0$ with $\tau_2 = 1 - \tau_1$; $\kappa_1 = 0.01/d$, $\kappa_2 = 5.6644/d$; Σ_1 is an identity covariance matrix and Σ_2 is the sample covariance matrix of the first $i-1$ iterates. Moreover, in accordance with Eraker et al. (2003), we hypothesize the prior distribution for the parameter vector Θ as follows:

Volatility process: $\theta \sim N(0,10)$, $\phi \sim Beta(25,2)$, $\sigma_\eta^2 \sim IG(2.5,0.1)$, $\rho \sim U(-1,1)$;

Conditional mean: $\alpha \sim N(0,4)$, $\beta \sim Beta(25,2)$, $\sigma_\zeta^2 \sim IG(2.5,0.1)$;

Jump process: $\lambda \sim Beta(2,40)$, $\sigma_J^2 \sim IG(2.5,0.05)$.

Figure 3.2 shows the parameter density estimations of the A-PMCMC and PMCMC algorithms. The corresponding plot in Figure 3.2 suggests that estimated parameters converge towards the given parameters. Figure 3.3 describes the synthetic conditional volatilities, estimated volatility by A-PMCMC method and PMCMC method respectively, which shows that the A-PMCMC method outperforms the PMCMC method by better tracking the synthetic volatilities.

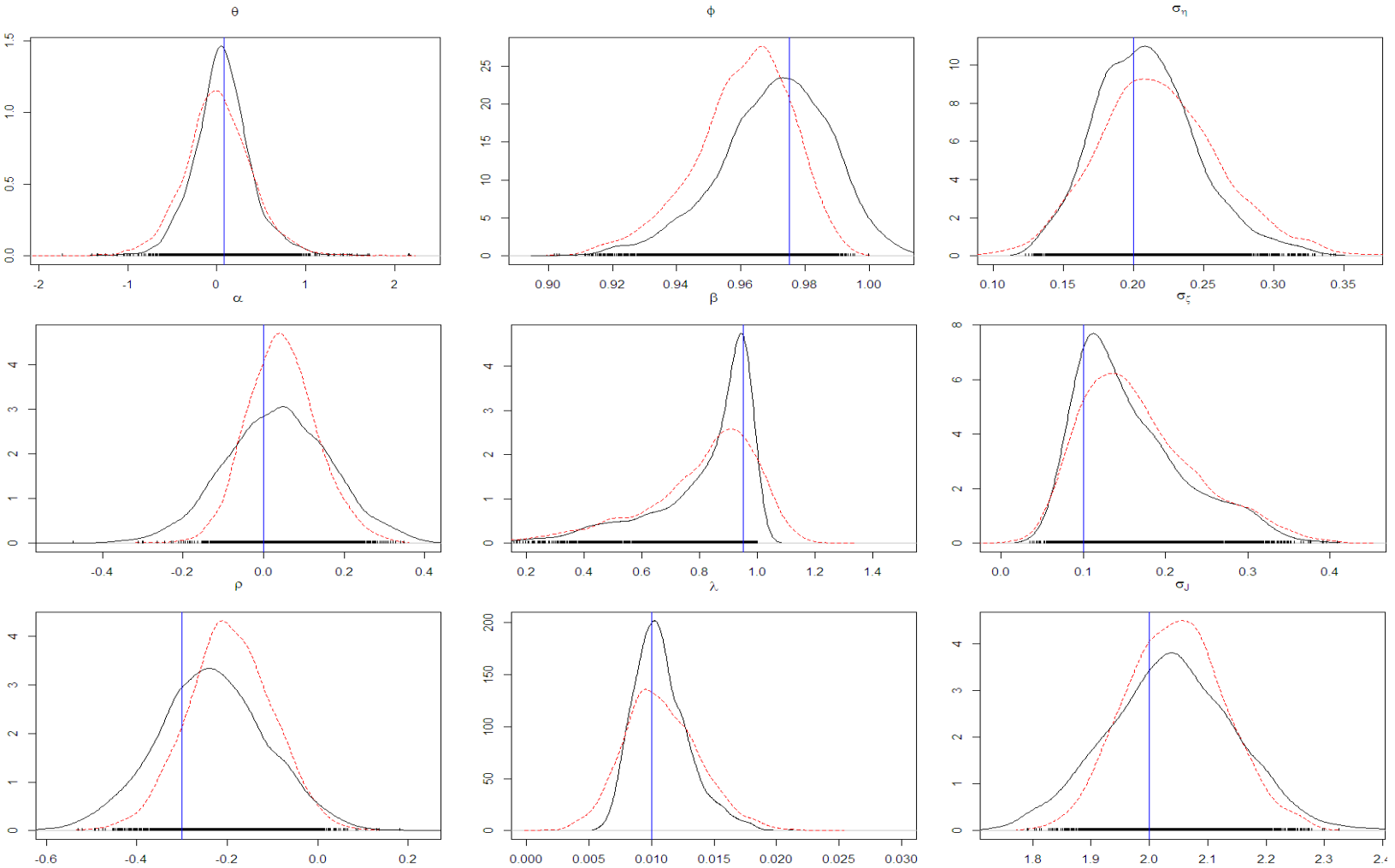


Figure 3.2 Estimated density of unknown parameters of stochastic volatility model. The blue line is the true value; the black line and the red line denote the posterior densities estimated by A-PMCMC and PMCMC, respectively.

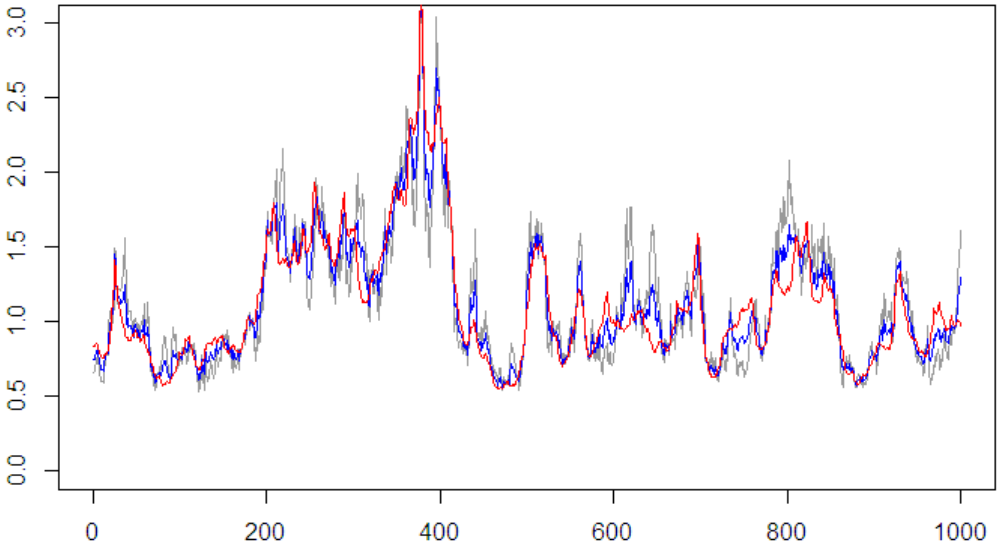


Figure 3.3 Estimated volatilities of A-PMCMC method (blue line) and PMCMC method (red line) and the synthetic conditional volatilities (gray line).

The parameter estimation results of A-PMCMC method and PMCMC method are described in Table 3.1, which contains the means, standard deviations and estimated quantiles at 5%, 50%, 95% respectively. In order to increase the robustness of the parameter estimation, the A-PMCMC method and PMCMC method are run 50 times using different random seeds but keeping the prior parameter values fixed. The estimated parameters are the average of 50 times individual running results. In testing for bias we find very encouraging results. We point out that all the parameters estimated by both A-PMCMC method and PMCMC method are within the boundary of their 90% confidence limits. Moreover, we find that the parameter values estimated by A-PMCMC method are closer to the true value compared to the estimated parameters of the PMCMC method.

Table 3.1 Posterior means, standard deviations and quantiles for parameters.

		A-PMCMC Method					PMCMC Method				
	True	Posterior					Posterior				
	Value	Mean	SD	5%	50%	95%	Mean	SD	5%	50%	95%
θ	0.08	0.069	0.242	-0.460	0.060	0.639	0.019	0.386	-0.595	0.0102	0.649
ϕ	0.975	0.970	0.017	0.932	0.962	0.982	0.961	0.015	0.932	0.964	0.983
σ_η	0.2	0.210	0.036	0.154	0.207	0.275	0.218	0.044	0.149	0.216	0.295
α	0.001	0.035	0.132	-0.179	0.037	0.253	0.046	0.088	-0.087	0.044	0.200
β	0.95	0.899	0.079	0.731	0.925	0.975	0.887	0.120	0.671	0.897	1.069
σ_ζ	0.1	0.158	0.066	0.076	0.142	0.293	0.167	0.070	0.0733	0.155	0.305
ρ	-0.3	-0.234	0.121	-0.437	-0.235	-0.035	-0.195	0.094	-0.348	-0.197	-0.043
λ	0.01	0.011	0.002	0.007	0.010	0.150	0.011	0.003	0.006	0.011	0.016
σ_J	2	2.037	0.109	1.856	2.037	2.215	2.050	0.084	1.83	2.045	2.188

3.5 Application to financial risk

In this section, the above described methodology is applied to real world applications in order to estimate the VaRs of three major stock indices and two currency exchange rates. To test the performance of our method, we estimate the 1-period VaR and 10-period VaR using daily data of stock indices and currency exchange rates, respectively. The two common significance levels $\tau = 0.01$ and $\tau = 0.05$ are chosen. In order to validate the VaR calculation, the back testing procedures are employed. Let N denote the number of exceptions at time $t, t = 1, 2, ..., T$. It should be the case that the proportion of exceptions N/T approximately agrees with the fixed probability

level τ. Hence, the hypothesis test is given by:

$$H_0 : E[N] = T\tau, \qquad H_1 : E[N] \neq T\tau.$$

Alexander (2009) proposed to employ the following likelihood ratio statistic,

$$LR = -2\log[(1-\tau)^{T-N}\tau^N] + 2\log[(1-N/T)^{T-N}(N/T)^N],$$

to test this hypothesis. Under the null hypothesis H_0, The statistic LR follows an asymptotically $\chi^2(1)$ distribution. The comparison is based on the asymptotically valid $p-$value for the likelihood ratio statistic, and methods with higher $p-$value are preferred.

3.5.1 Application to estimate 1-period VaR of stock market

The S&P 500, Nikkei 255 and Shanghai composite are three major stock market indices. Our methodology is applied to estimate the 1-day VaR on these stock indices from July 21, 2005, to December 31, 2014，containing 2115 daily observations. In the sample period, two major events occurred in the global financial market: the US subprime crisis and the Eurozone sovereign debt crisis, which caused sharp fluctuations in the global stock markets. Figure 3.4 shows the stock price indices and their log-returns. It is evident that the US stock market, Japanese stock market and Chinese stock market underwent the largest market volatility in September 2008, when Lehman Brothers collapsed. During 2007-2010, the volatility of stock markets fluctuated widely. Additionally, it shows that the three stock markets experienced a large volatility in September 2011. On September 30, 2011, investors' panic mood for the Euro-zone sovereign debt crisis had reached the peak, which resulted in a sharp fluctuation in the stock market.

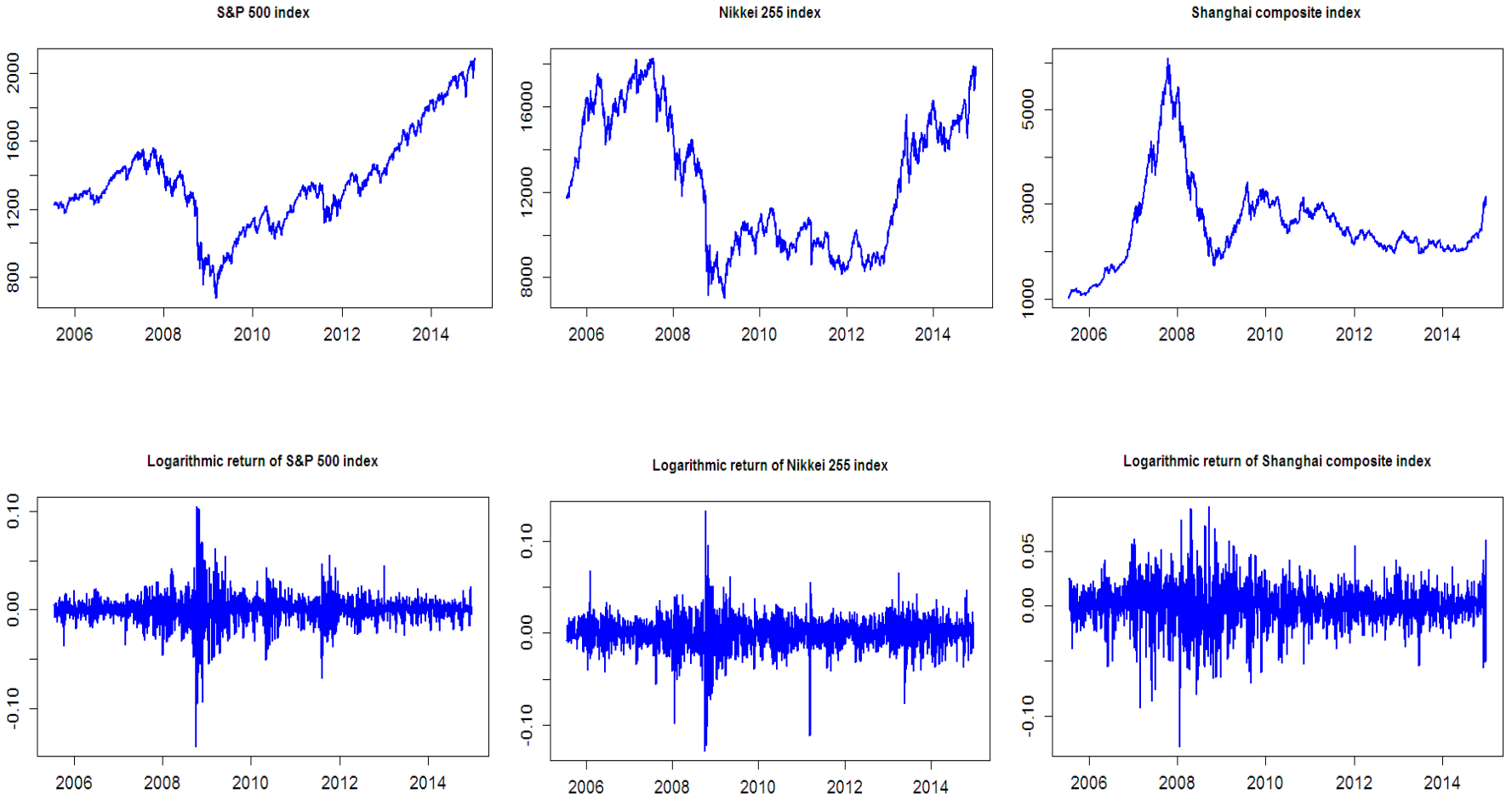
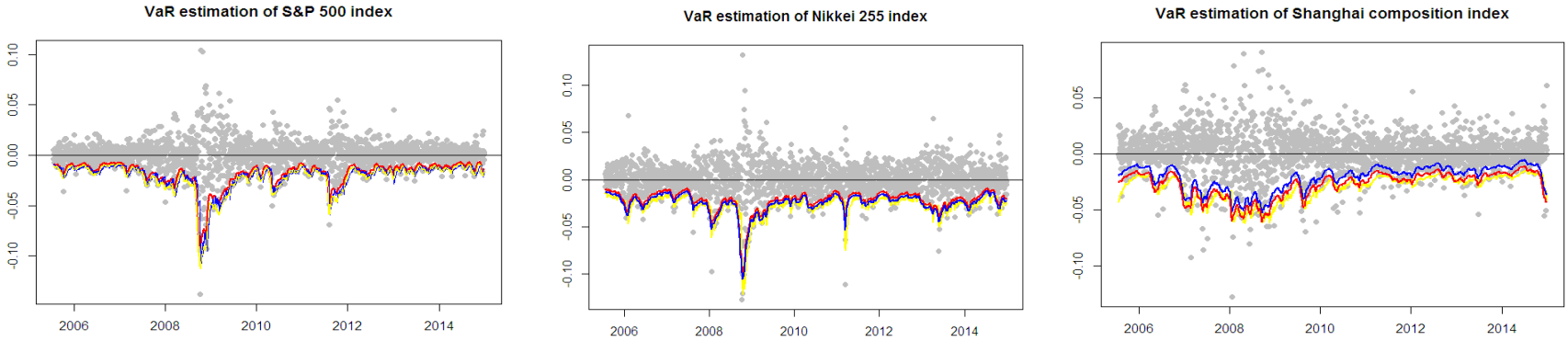


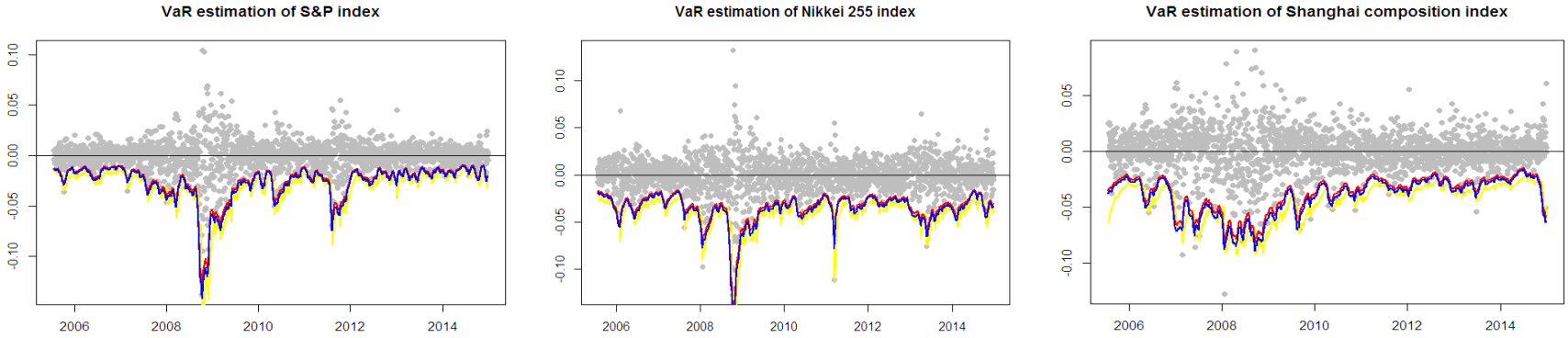
Figure 3.4 S&P 500, Nikkei 255 and Shanghai composite indices and their log-returns from 2005/07/21 to 2014/12/30.

In this section, we compare our A-PMCMC method with the PMCMC method. We also compare them with the GARCH (1, 1) method, which is current state-of-the-art used by most practitioners. Figure 3.5 shows VaR estimation results at probability levels $\tau = 0.05$ and $\tau = 0.01$ in which all three methods seem to do very good jobs on tracking the extreme VaR to the real loss. However, note that GARCH (1, 1) tends to overestimate the extreme fluctuations.

In the VaR back-testing, we observe the fact that GARCH (1, 1) and PMCMC fail to provide acceptable results for some stock markets. Table 3.2 presents the back-testing results, which indicates that at significance levels $\tau = 0.05$ and $\tau = 0.01$, the A-PMCMC method gives accurate estimations at each probability level. Nevertheless, the PMCMC method fails to provide acceptable results at 95% confidence level for the Japanese and Chinese stock markets at $\tau = 0.05$ risk level and for the Chinese stock market at $\tau = 0.01$ risk level. Particularly, at $\tau = 0.05$ risk level, the GARCH (1, 1) method fails to provide acceptable results at 95% confidence level for the Japanese and Chinese stock markets, and at $\tau = 0.01$ risk level, it fails to provide acceptable results for all three stock markets. Therefore, the proposed A-PMCMC method is the only one which is able to provide reliable VaR estimation in real-world stock markets.



Panel A: $\tau = 0.05$



Panel B: $\tau = 0.01$

Figure 3.5 The 1-day VaR estimations of S&P 500, Nikkei 255 and Shanghai composition indices at $\tau = 0.05$ and $\tau = 0.01$. The dots are the actual log-returns; the blue line denotes the A-PMCMC method, the red line denotes the PMCMC method and the yellow line denotes the GARCH (1, 1) method.

Table 3.2 Back-testing results for the 1-day VaR estimations.

	A-PMCMC Method				PMCMC Method				GARCH(1, 1) Method			
	N	N/T	LR	p-value	N	N/T	LR	p-value	N	N/T	LR	p-value
$\tau = 0.05$												
S&P 500	105	0.0497	0.0049	0.9443	96	0.0454	0.9655	0.3258	89	0.04210	2.9280	0.0871
Nikkei 255	99	0.0468	0.4562	0.4993	82	0.0388	6.0413	0.0139	80	0.0378	7.1549	0.007
Shanghai composition	107	0.0506	0.0167	0.8969	85	0.0402	4.5605	0.0327	83	0.0393	5.5227	0.0187
$\tau = 0.01$												
S&P 500	24	0.0113	0.3744	0.5405	18	0.0085	0.4960	0.4812	12	0.0057	4.7296	0.0296
Nikkei 255	19	0.0089	0.2265	0.6341	15	0.0071	2.0044	0.1568	11	0.0052	5.9570	0.0147
Shanghai composition	27	0.0127	1.5085	0.2193	31	0.0146	4.0613	0.0438	12	0.0057	4.7296	0.00296

3.5.2 Application to estimate 10-period VaR of exchange rate market

In this section we perform a predictive performance for 10-day VaR estimation. For the multi-period VaR estimation, the common method is the "static" estimation method. We first introduce the static method of 10-day VaR. Under the assumption that the daily returns are independent identically distributed (i.i.d) and normal, the $h-$day VaR can be calculated by the square-root scaling rule, that is

$$VaR_{h,\tau} \approx \sqrt{h} \times VaR_{1,\tau}. \tag{3.30}$$

Although the assumption that the return series are i.i.d. and normal is usually not justified in practice, the static method is widely applied by practitioners so it is used as a benchmark. It provides a sort of "plain vanilla" estimation for $h-$day VaR. Comparing with the "static" estimation method, the proposed A-PMCMC method and PMCMC methods can be regarded as "dynamic" estimation method that can predict all volatilities in the next h days directly. We will use the Japanese-Yen / US Dollar and Chinese-RMB / US Dollar time series to compare the three estimation methods.

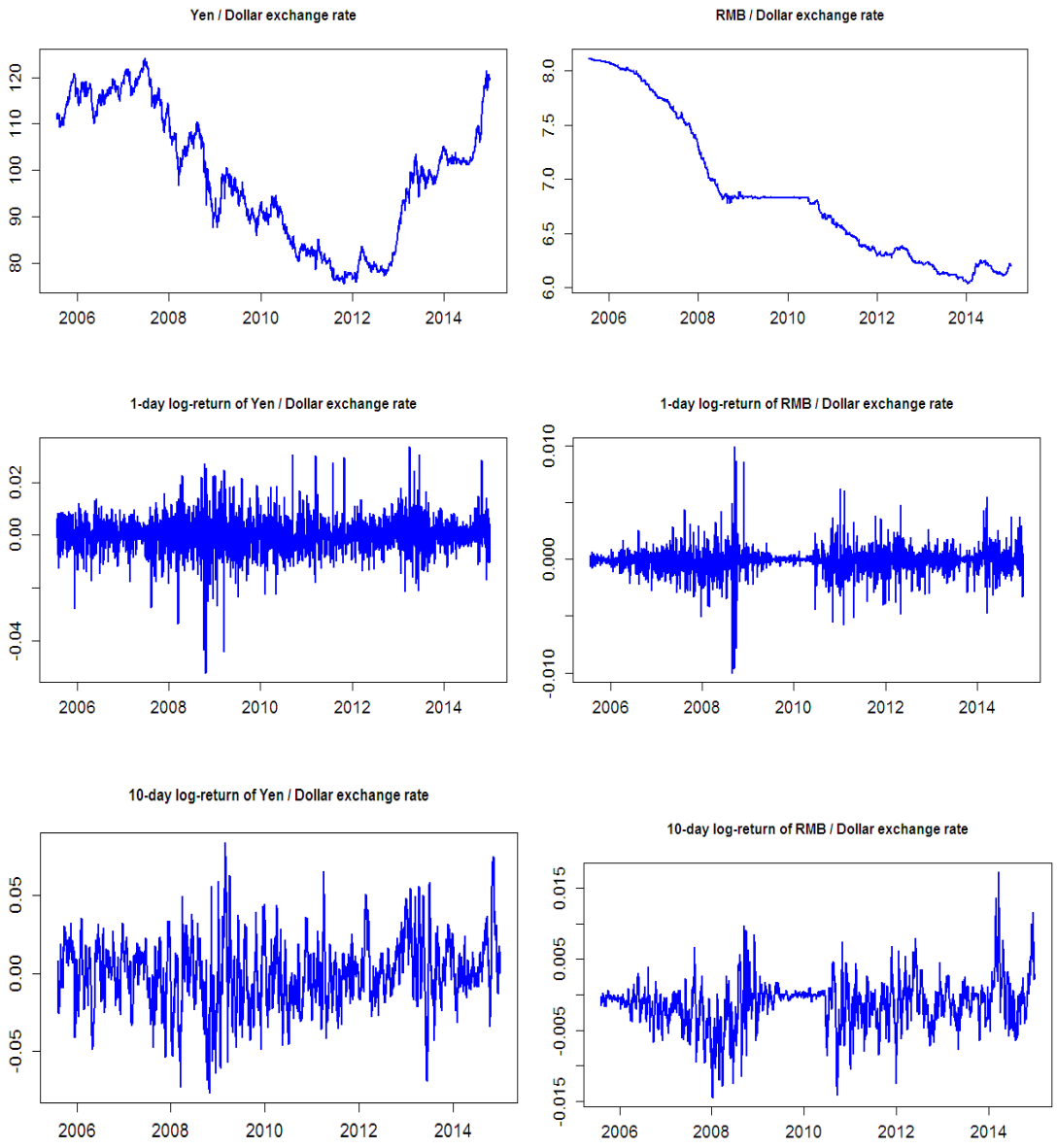
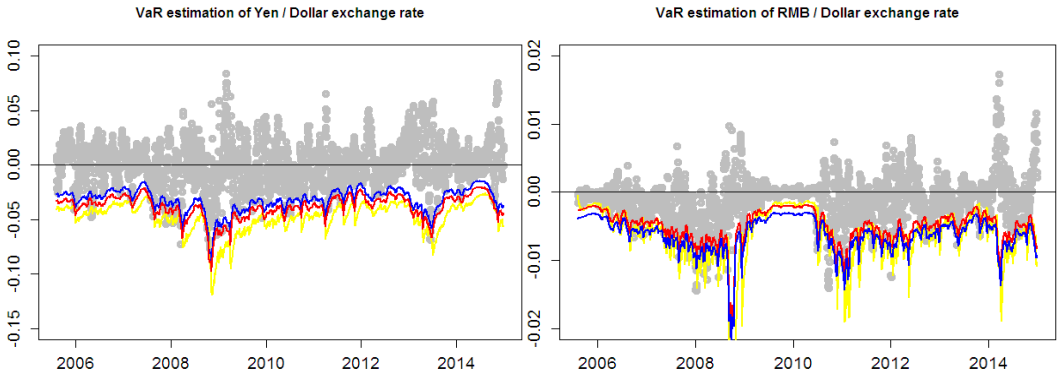


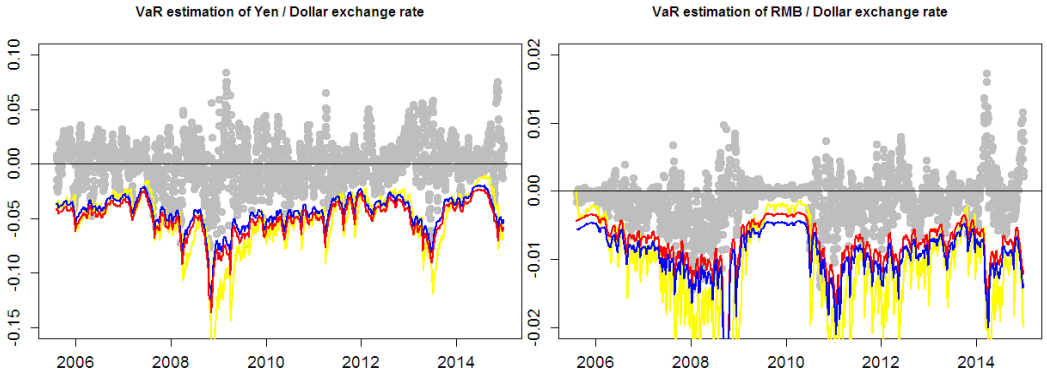
Figure 3.6 Japanese-Yen/US-Dollar and Chinese-RMB/US-Dollar exchange rates, 1-day log-returns from 2005/07/21 to 2014/12/30 and 10-day log-return from 2005/07/31 to 2014/12/30.

The Japanese-Yen / US Dollar and Chinese-RMB / US Dollar are two major currency exchange rates. The sample set contains 2372 daily observations period from July 21, 2005, to December 31, 2014. Figure 3.6 shows the Yen / Dollar and RMB /

Dollar exchange rates and their 1-day and 10-day log-returns, respectively. It is obvious that the two exchange rates underwent the largest market volatility during the US subprime crisis and the Eurozone sovereign debt crisis. Moreover, the fluctuations of 10-day log-returns are greater than 1-day log-returns. Therefore, from a methodological perspective, the 10-day VaR estimation is much harder than the 1-day VaR estimation.



Panel A: $\tau = 0.05$



Panel B: $\tau = 0.01$

Figure 3.7 The 10-day VaR estimations of Yen/Dollar and RMB/Dollar exchange rates at $\tau = 0.05$ and $\tau = 0.01$. The dots are the 10-day log-returns, the blue line denotes the A-PMCMC VaR method, the red line denotes the PMCMC VaR method, and the yellow line denotes the Static VaR Method.

Figure 3.7 presents the 10-day VaR estimation results for Yen / Dollar and RBM / Dollar at $\tau = 0.05$ and $\tau = 0.01$ risk levels. The VaR estimations are different among the "dynamic" A-PMCMC and PMCMC method and the "static" method. It is obvious that as the probability level decreases from $\tau = 0.05$ to some extreme values such as $\tau = 0.01$, the gaps of these three methods get larger and larger. Table 3.3 summarizes the results of the back-testing for the Yen / Dollar and RMB / Dollar exchange rates. On average, the A-PMCMC method gives the most accurate forecasts at each probability level compared to the PMCMC method and the "static" method. The test of VaR is not rejected for the A-PMCMC method at all risk levels. In contrast, the PMCMC method fails to provide acceptable results for the Chinese stock market at $\tau = 0.05$ and $\tau = 0.01$ risk levels. Similarly, the "static" estimation method also fails to provide acceptable results for RMB / Dollar exchange rate at the $\tau = 0.05$ level and for the two exchange rates at extreme the $\tau = 0.01$ level. The comparative result indicates that in multi-period VaR estimation, the "dynamic" A-PMCMC and PMCMC methods have a better performance than "static" estimation method. Further, in the "dynamic" estimation methods, our A-PMCMC method is superior to the PMCMC method.

Table 3.3 Back-testing results for the 10-day VaR estimations.

	A-PMCMC Method				PMCMC Method				Static Method			
	N	N/T	LR	p-value	N	N/T	LR	p-value	N	N/T	LR	p-value
$\tau = 0.05$												
Yen	122	0.0516	0.1341	0.7141	113	0.0478	0.2350	0.6278	94	0.0397	5.5493	0.0184
RMB	145	0.0614	6.0323	0.0140	154	0.0652	10.5271	0.0011	168	0.07112	19.7352	0.0000
$\tau = 0.01$												
Yen	32	0.0135	2.7032	0.1001	20	0.0084	0.5911	0.4419	12	0.0051	7.0452	0.0079
RMB	22	0.0093	0.1148	0.7346	36	0.0152	5.6482	0.0171	48	0.0203	19.5693	0.0000

CHAPTER IV

DYNAMIC RISK MEASUREMENT OF FINANCIAL TIME SERIES WITH HEAVY-TAILED: A NEW HYBRID APPROACH

In this chapter, a new hybrid approach to measure dynamic risk of financial time series with heavy-tailed distribution is presented. The proposed method, hereafter referred to as NIG-MSA, exploits the normal inverse Gaussian (NIG) distribution to fit the heavy-tailed distribution, and combines the empirical mode decomposition with support vector regression to structure a multi-scale analysis (MSA) methodology. The validity of the NIG-MSA method for volatility prediction is confirmed through Monte Carlo simulation. This method is illustrated with an application to the risk measurement of returns on the S&P 500 index, and our results show that the proposed NIG-MSA approach provides more precise Value at Risk calculation than the traditional single-scale model.

4.1 The dynamic risk measurement model

4.1.1 The GARCH (1, 1) model with NIG distribution

The GARCH (1, 1) model is a parsimonious model in volatility forecasting models (Eberlein, 2003). The model provides a simple representation of the main statistical characteristics of a return process, such as autocorrelation and volatility

clustering. The GARCH (1, 1) model is the most popular structure for volatility forecasting and, consequently, it is extensively used to model real financial time series.

Let $R_t = \log P_t - \log P_{t-1}$ denote the logarithm of return, where P_t is the asset price at time t. The return process is modeled in the GARCH (1, 1):

$$\begin{aligned} R_t &= \sigma_t \varepsilon_t \\ \sigma_t^2 &= \omega + \phi R_{t-1}^2 + \varphi \sigma_{t-1}^2 \end{aligned}, \tag{4.1}$$

where the innovation term ε_t is assumed to be an independently and identically distributed random variable. The volatility σ_t^2 is time varying and unobservable in the market. To ensure that the conditional variance is positive, we assume that the parameters ω, ϕ and φ all satisfy $\omega > 0, \phi, \varphi \geq 0$.

The NIG distribution is a heavy-tailed distribution which is rich enough to model financial time series and has the benefit of numerical tractability (Eberlein et al., 1995; Nielsen, 1997). The density function of the NIG distribution for x is

$$f_{NIG}(x;\alpha,\beta,\delta,\mu) = \frac{\alpha\delta}{\pi} \cdot \frac{K\{\alpha\sqrt{\delta^2 + (x-\mu)^2}\}}{\sqrt{\delta^2 + (x-\mu)^2}} \exp\{\delta\sqrt{\alpha^2 - \beta^2} + \beta(x-\mu)\}, \tag{4.2}$$

where, $\delta > 0$ and $|\beta| \leq \alpha$, $K(x) = \frac{1}{2}\int_0^\infty \exp\{-\frac{x}{2}(y + y^{-1})\}dy$.

The location and scale of the density are mainly controlled by paramters μ and δ respectively, whereas α and β play roles in the skewness and kurtosis of the distribution. Thus all moments of $NIG(\alpha,\beta,\delta,\mu)$ have simple explicit expressions, in particular, the mean and variance are $E(x) = \mu + \beta\delta / \sqrt{\alpha^2 - \beta^2}$ and $Var(x) = \alpha^2\delta / \sqrt{(\alpha^2 - \beta^2)^3}$. Furthermore, if $\mu = 0$, the NIG distribution has the tail-

behavior

$$f_{NIG}(x,\alpha,\beta,\delta,\mu=0)\sim x^{-\frac{3}{2}}e^{-(\alpha-\beta)x} \qquad as \quad x\to\infty, \tag{4.3}$$

which shows that the NIG distribution has an exponential decaying speed. As compared to the normal distribution, the NIG distribution decays more slowly and the NIG distribution often appears in modeling the return process. In this paper, we propose that the stochastic term ε_t is assumed to possess the NIG distribution. The parameters in GARCH (1, 1) model are estimated using quasi-maximum likelihood method.

4.1.2 Empirical Mode Decomposition (EMD)

The decomposition is based on the local characteristic time scale of the data. So any non-stationary dataset can be adaptively decomposed into a finite and often small number of Intrinsic Mode Functions (IMF) with individual intrinsic time scale properties. The IMFs satisfy the following two prerequisites: (1) In the whole data series, the number of extreme points and the number of zero crossings must be equal or differ at most by one. (2) The mean value of the envelopes defined by local maxima and minima must be zero at all points. Each IMF component has a clear physical meaning and contains a certain characteristic range of time scale (Huang et al., 1998). As compared with the original data, the IMF components are more stationary, which is advantageous to forecast volatility of return process. The generic EMD algorithm is described by the following steps:

(i) Identify all the maximum points and all the minimum points of original signal $x(t)$.

(ii) Fit the maxima envelop $x_u(t)$ and minima envelope $x_l(t)$ with cubic

spline function.

(iii) Calculate the mean value $m_1(t) = (x_l(t) + x_u(t))/2$.

(iv) Calculate the quasi-IMF $h_1(t) = x(t) - m_1(t)$ and test whether $h_1(t)$ satisfies the two prerequisites of an IMF property. If they are satisfied, we obtain the first IMF. If not, we regard $h_1(t)$ as $x(t)$ and repeat steps (i)-(iii) until $h_1(t)$ becomes an IMF.

(v) Calculate the first residual term $res(t) = x(t) - h_1(t)$. The $res(t)$ is treated as new input $x(t)$ in the next loop to derive the next IMF. We stop the decomposition procedure until the residual term $res(t)$ becomes a monotonic function from which no further IMF can be extracted.

From the above decomposition process, it is obvious that the original time series $x(t)$ can be reconstructed by summing up all the IMF components together with the last residue component, that is $x(t) = \sum h_i(t) + res(t)$. In this thesis, the residual term is seen as the last IMF.

The EMD method adaptively obtains the local IMF components with the shortest cycle by screening the local characteristics from the original signal and each component also includes a corresponding section of different frequency component.

4.1.3 The hybrid dynamic risk model

In this thesis, the overall process of formulating the dynamic risk measurement model is presented. Here we name this new VaR method as the Normal Inverse Gaussian- Multi-Scale Analysis (NIG-MSA) method. The NIG-MSA model ensemble paradigm can be formulated as illustrated in Figure 4.1.

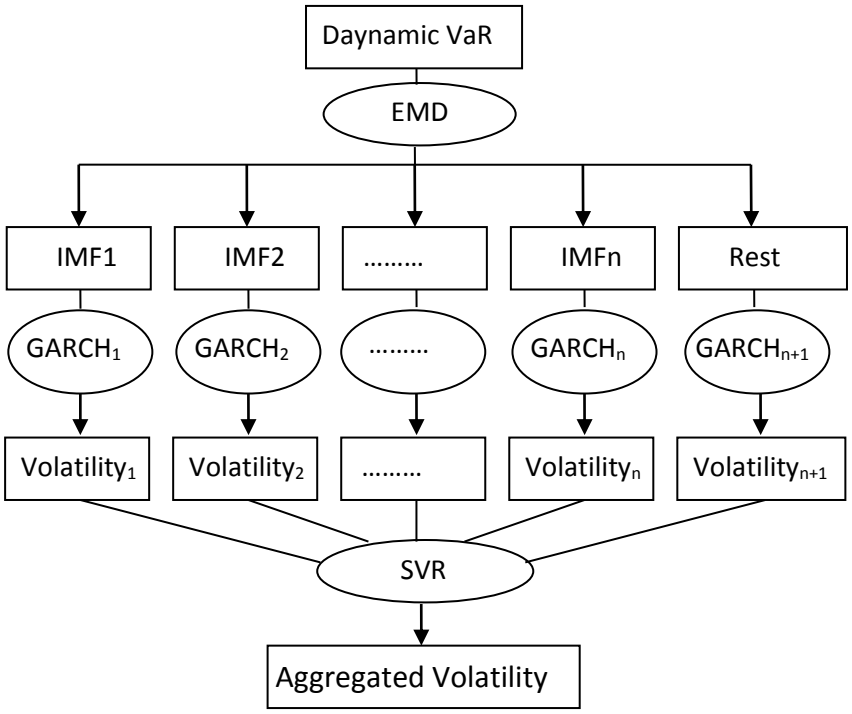


Figure 4.1 The overall process of the NIG-MSA model.

As can be seen from Figure 4.1, the NIG-MSA model generally consists of the following four main steps:

(1) The returns series $R(t),\ t=1,2,\ldots,T$ is adaptively decomposed into a finite number of IMF (Intrinsic Mode Function) components employing the EMD method.

(2) The GARCH (1, 1) model is used as a prediction tool to model the volatility process of each extracted IMF component and to predict the corresponding volatility, in which we assume the innovation term is NIG distribution.

(3) The volatility forecasting results of all extracted IMF components in step (2) are integrated to generate an aggregated volatility estimation using a SVR model.

(4) Using the aggregated volatility to calculate the devolatilized return, then the NIG distribution parameters can be estimated.

4.2 Simulation experiment

4.2.1 Experimental design

The NIG-MSA technique consists of two main parts: predict the volatility using the multi-scale methodology and dynamically estimate the quantile of innovation. The calculation procedure can be described as:

(i) Set up the data generating model and select the kernel functions in SVR model.

(ii) Estimate the aggregated volatility $\hat{\sigma}_t$ using the multi-scale methodology.

(iii) Calculate the innovation terms $\varepsilon_t = R_t / \hat{\sigma}_t$ and fit the NIG distribution parameters and estimate the quantile $\hat{q}_\tau$.

(v) Calculate the $VaR_t = \hat{\sigma}_t \cdot \hat{q}_\tau$.

From the above calculation steps, we can see that the key pillar for the NIG-MSA technique is the accurate estimation of the volatility. In the simulation, we only focus on the volatility forecasting. The Monte Carlo simulation is applied to evaluate the performance of the NIG-MSA method. The simulated data set is generated by the following model

$$R_t = \sigma_t \varepsilon_t$$
$$\sigma_t^2 = \begin{cases} 0.1 + 0.4R_{t-1}^2 + 0.5\sigma_{t-1}^2 & 1 \le t \le 400 \\ 0.5 + 0.1R_{t-1}^2 + 0.8\sigma_{t-1}^2 & 400 < t \le 750 \\ 0.1 + 0.7R_{t-1}^2 + 0.2\sigma_{t-1}^2 & 750 < t \le 1000 \end{cases}, \quad (4.4)$$

where R_t is return and ε_t is innovation distributed as normal inverse Gaussian with zero mean and unit variance. Notice that if the data generating process is $R_t = \sigma_t \varepsilon_t$ then $VaR_t = \sigma_t q_\tau$ (Franke et al., 2004).

4.2.2 Experimental result analysis

The purpose of this experiment is to evaluate the four volatility forecast methods: (i) GARCH (1, 1) with normal distribution (Nor-GAR), (ii) GARCH (1,1) with normal inverse Gaussian distribution (NIG-GAR), (iii) Multi-scale analysis with normal distribution (Nor-MSA) and (iv) Multi-scale analysis with normal inverse Gaussian distribution (NIG-MSA). We use the four models to respectively forecast the real volatility generated in (3) and the simulation results are shown in Figure 4.2.

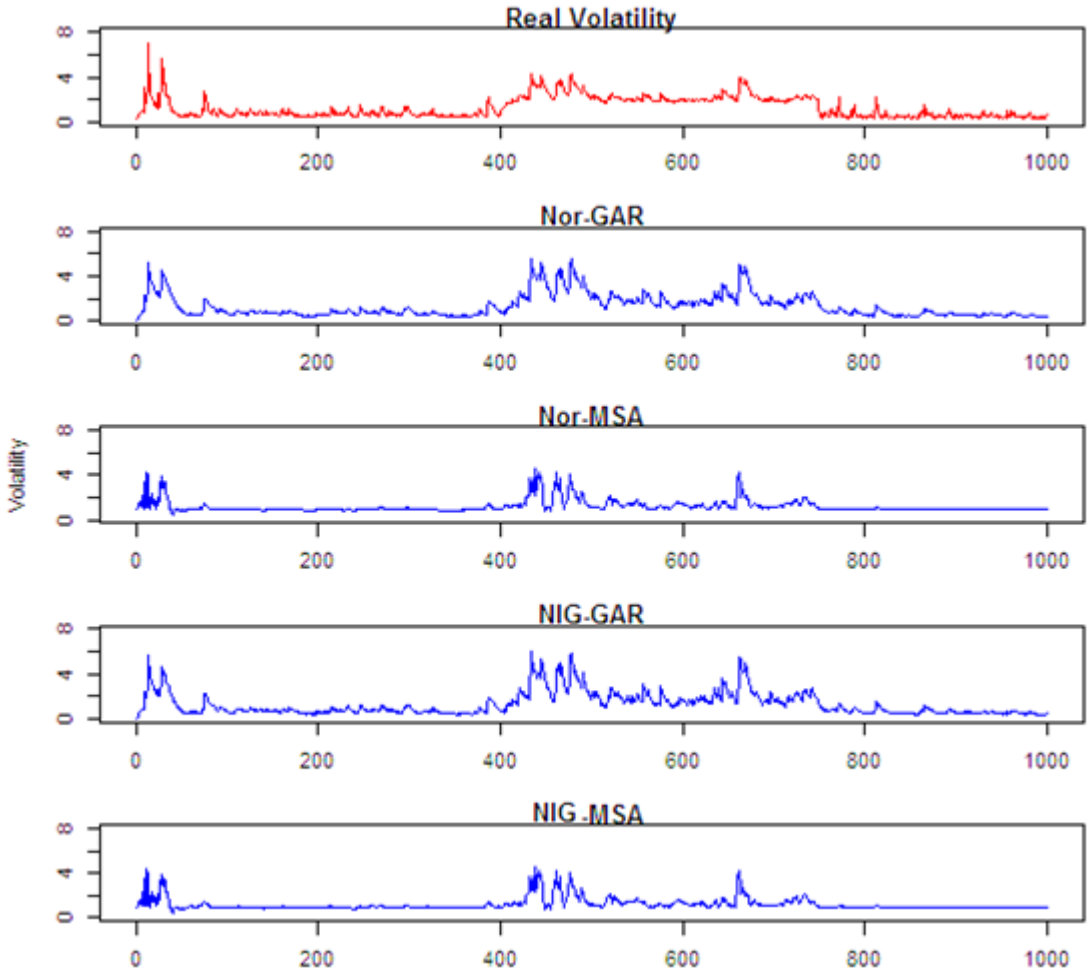


Figure 4.2 The comparison of volatility forecasting.

The volatility forecasting performance is evaluated using the following statistical metrics:

Normalized mean squared error (NMSE):

$$NMSE = \sqrt{\sum_{t=1}^{N}(\hat{\sigma}_t^2 - R_t^2)^2 \Big/ \sum_{t=1}^{N}(R_{t-1}^2 - R_t^2)^2} \tag{4.5}$$

Normalized mean absolute error (NMAE):

$$NMAE = \sum_{t}^{N}\left|\hat{\sigma}_t^2 - R_t^2\right| \Big/ \sum\left|R_{t-1}^2 - R_t^2\right| \tag{4.6}$$

Hit rate (HR)

$$HR = \frac{1}{N}\sum_{t}^{N}\Delta_t, \qquad \Delta_t = \begin{cases} 1 \ , \ (\hat{\sigma}_t^2 - R_{t-1}^2)(R_t^2 - R_{t-1}^2) \geq 0 \\ 0 \ , \ else. \end{cases} \tag{4.7}$$

The three statistical metrics relate the predicted volatility $\hat{\sigma}_t^2$ to the proxy volatility estimation R_{t-1}^2. The NMSE and NMAE are the measures of the deviation between the proxy and predicted values. The smaller their values, the closer the predicted volatility is to the actual values. The HR is a measure of how often the model gives the correct direction of change of volatility. The larger the value of HR, the better is the performance of prediction.

Additionally, the volatility of the return process can not be observed, so we evaluate the performance of the volatility prediction in the model following the criterion: the better the forecasting performance of volatility model, the better the standardized observation ($\hat{\varepsilon}_t = R_t / \hat{\sigma}_t$) is fitting the normal inverse Gaussian distribution. The Kolmogorov– Smirnov distance (KS) is usually used to test whether a given $F(x)$ is the underlying probability distribution of $F_n(x)$, so we use the Kolmogorov-Smirnov distance as the criterion for the goodness of fit testing. It is defined as

$$KS = \sup_{x \in R} \left| F(x) - F_n(x) \right|, \tag{4.8}$$

where $F(x)$ is the empirical sample distribution and $F_n(x)$ is the cumulative distribution function. The smaller the values of KS distances, the closer are the predicted volatility to the actual values.

Table 4.1 gives the descriptive statistics of the simulation results and shows the superiority of the NIG-MSA model over the other models. The table reports the computed KS distance, NMSE, NASE and HR statistics metrics for the four

models: Nor-GAR, Nor-MSA, NIG-GARCH, NIG-MSA. It shows that the values of KS distance, NMSE, NMAE for the NIG-MSA model are below the others, and that the value of HR for NIG-MSA model is the highest. Further, it indicates that multi-scale analysis has more influence on the volatility forecasting performance compared with the NIG distribution assumption. For an example of KS distance, the value of Nor-MSA reduced to 0.0174 and NIG-GAR only down to 0.0402 relative to the value of Nor-GAR 0.0581. As for the other three statistical metrics, we can draw the same conclusion. The NIG-MSA model can give better predictions because of its good time-frequency property which can describe non-stationary financial time series.

Table 4.1 Results of the simulation.

Methods	Nor-GAR	Nor-MSA	NIG-GAR	NIG-MSA
KS distance	0.0581	0.0174	0.0402	0.0165
NMSE	0.8015	0.7954	0.7028	0.6905
NASE	0.8943	0.8873	0.6674	0.6367
HR	0.65	0.694	0.671	0.76

4.3 Empirical analysis

The data set S&P500 index was used in our empirical analysis. The index is daily registered from 2000/01/03 to 2014/10/28. There are 3729 observations. The first 2768 observations (from 2000/01/03 to 2010/12/31) are used as a basis to train the multi-scale analysis system and estimate the NIG distribution parameters. The residual 961 observations are used as a test set to evaluate the prediction of the

dynamic VaR calculated by the NIG-MSA dynamic risk measurement model. The graphics of the return processes of train set are displayed in Figure 4.3.

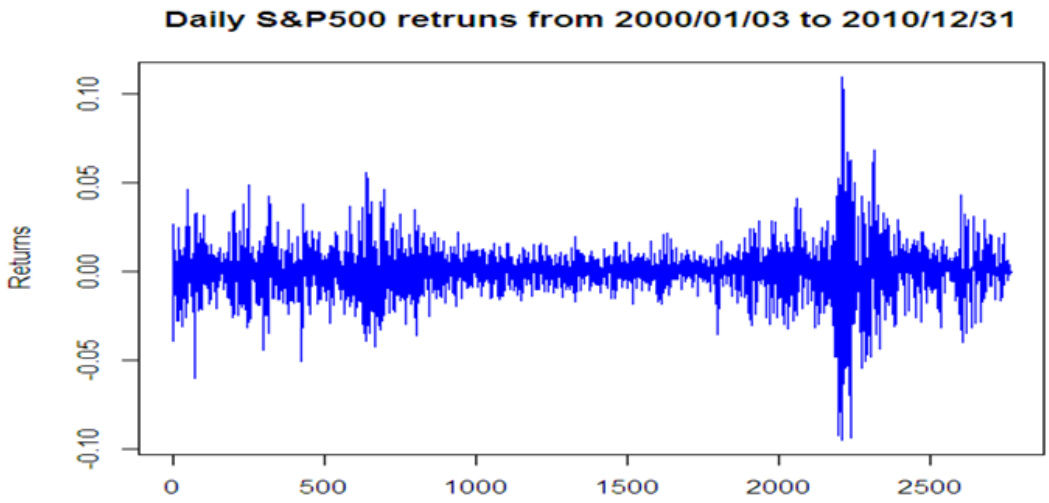


Figure 4.3 The logarithmic return process of S&P 500 index.

4.3.1 NIG-MSA model training and volatility prediction

The NIG-MSA model can be trained according to the multi-scale methodology shown in 4.1. Firstly, the training set (2768 observations) is decomposed into ten IMF components (the last IMF is residual term) using the EMD technique, as illustrated in Figure 4.4. Then, the GARCH (1, 1) model was used to model the every IMF component and to estimate corresponding volatility. The estimation results of the parameters is shown in Table 4.2. Finally, we select the Gaussian RBF kernel function to train the SVR model and the aggregated volatility as given in Figure 4.5.

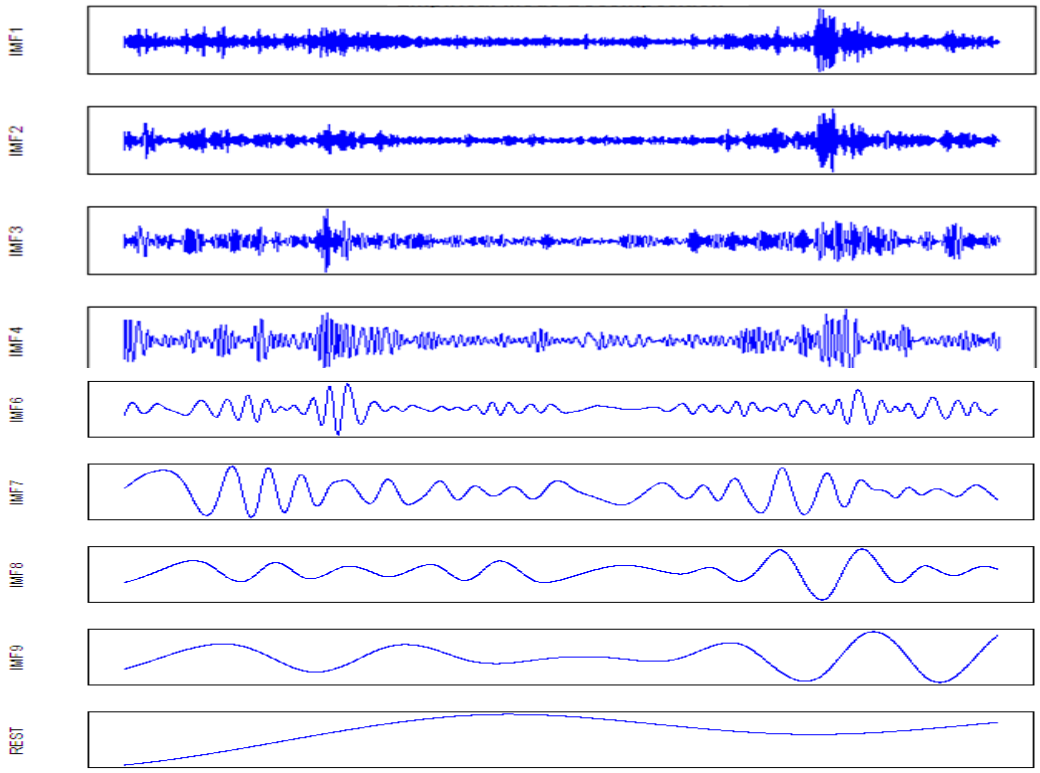


Figure 4.4 The decomposition of the training set.

Table 4.2 Parameter estimation results of the GARCH (1, 1) for IMF components.

		Estimate	Std. Error	t-statistic	Prob.(>|t|)
IMF1					
	ω	0.000003	4.9912e-07	6.01063	0.00000
	ϕ	0.499008	0.028298	17.63383	0.00000
	φ	0.499992	0.022552	22.17095	0.00000
IMF2					
	ω	0.000001	1.3844e-07	7.223301	0.00000
	ϕ	0.485806	0.085176	19.296566	0.00000
	φ	0.513194	0.018489	27.756293	0.00000
IMF3					
	ω	0.000001	2.8603e-07	3.496137	0.00000
	ϕ	0.665307	0.030181	22.04393	0.00000
	φ	0.333693	0.010341	32.26893	0.00000
IMF4					
	ω	0.000003	4.7507e-07	6.314859	0.00000
	ϕ	0.915177	0.153852	5.948424	0.00000
	φ	0.083811	0.004509	18.58749	0.00000
IMF5					
	ω	0.000001	1.4110e-07	7.087172	0.00000
	ϕ	0.86854	0.176395	4.923836	0.00000
	φ	0.12394	0.020874	5.937530	0.00000
IMF6					
	ω	0.000001	1.3554e-07	7.377896	0.00000
	ϕ	0.76570	0.114766	6.671837	0.00000
	φ	0.23093	0.018947	12.18821	0.00000
IMF7					
	ω	0.000001	1.0645e-07	9.394082	0.00000
	ϕ	0.09652	0.008564	11.27043	0.00000
	φ	0.88455	0.043197	20.47712	0.00000
IMF8					
	ω	0.000002	1.8609e-07	10.74749	0.00000
	ϕ	0.10291	0.012609	8.161631	0.00000
	φ	0.89145	0.033281	26.78555	0.00000
IMF9					
	ω	0.0000001	1.5428e-07	6.481722	0.00000
	ϕ	0.094704	0.009503	9.965695	0.00000
	φ	0.899026	0.047287	19.01212	0.00000
REST					
	ω	0.000001	1.3697e-07	7.301785	0.00000
	ϕ	0.095354	0.010075	9.464417	0.00000
	φ	0.901476	0.063218	14.2598	0.00000

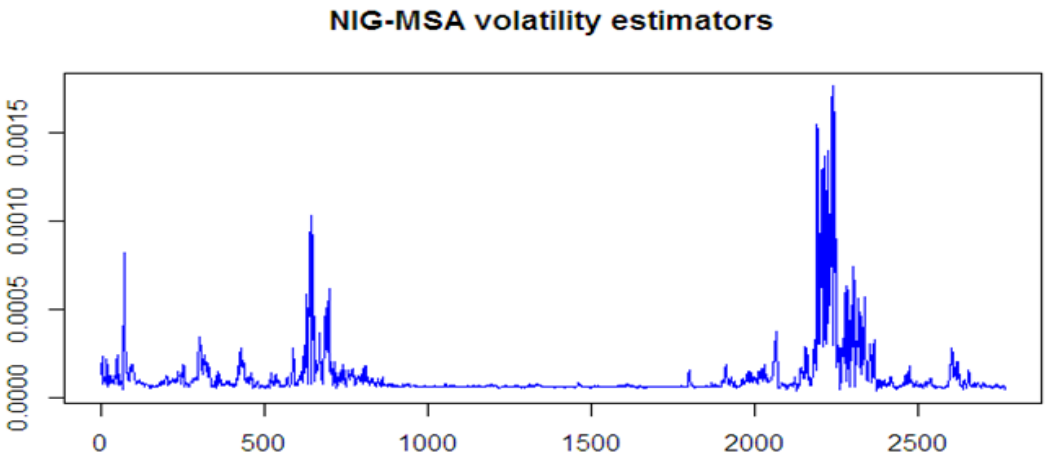


Figure 4.5 The volatility estimation of the training set.

The NIG-MSA model that has been trained can be used to predict the volatility of the test set series. The basic idea of the volatility estimation comes from the assumption that although the returns series is non-stationary in a long time period, its volatility structure is relatively stationary. So, we suppose that the test set consists of 10 IMF components, and use the GARCH (1, 1) model which has been modeled to forecast the volatility of the each component. Then the 10 volatility prediction series are integrated to generate the final volatility prediction. In order to evaluate the performance of the NIG-MSA model, the Nor-GAR model is selected as the reference method, their prediction results of volatility as shown in Figure 4.6.

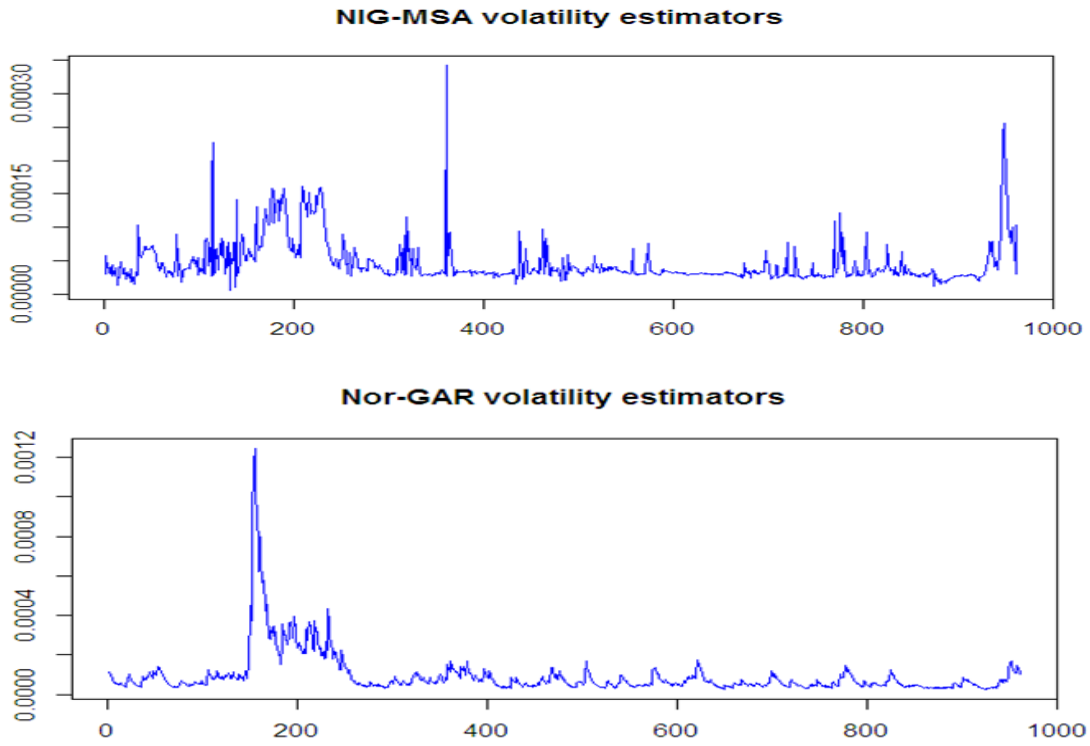


Figure 4.6 The volatility estimation of the test set.

4.3.2 Time-varying quantile estimation

In the NIG-MSA model, the distribution parameters could be time-variant as well. Figure 4.7 show how the quantile varies as time passes, which means that we could not keep the how assumption that the devolatilized returns are identically distributed. Instead, we estimated the dynamic quantiles based on the test set data. In Figure 4.7, we show the dynamic quantile estimations of the three probability levels, from the top the evolving NIG quantiles for $\tau = 0.5$, $\tau = 0.05$ and $\tau = 0.005$. A more detailed description is shown in Table 4.3 which contains four statistical metrics: Minimum, Maximum, Mean and Standard deviation. It gives the descriptive statistics of dynamic quantiles estimated by NIG-MSA technique. This provides evidence that the more extreme the probability levels, the greater the quantile varies as time passes.

For the extreme probability $\tau = 0.005$, the variety range value is 0.3538 and the standard deviation is 0.1049. However, for the probability $\tau = 0.5$, the variety range value is 0.0323 and the standard deviation is 0.0053. This inspires us to consider that we should use dynamic quantiles to calculate the VaR, especially for the extreme events.

Table 4.3 Descriptive statistics of the dynamic quantiles under different confidence levels.

	τ =0.005	τ =0.01	τ =0.025	τ =0.05	τ =0.25	τ =0.5
Min	-3.6719	-3.2147	-2.5134	-1.9833	-0.6840	0.0329
Max	-3.3181	-2.8895	-2.3133	-1.8621	-0.6605	0.0651
Mean	-3.5937	-3.1385	-2.4645	-1.9515	-0.6751	0.0489
S.d	0.1049	0.1033	0.0627	0.0370	0.0082	0.0053

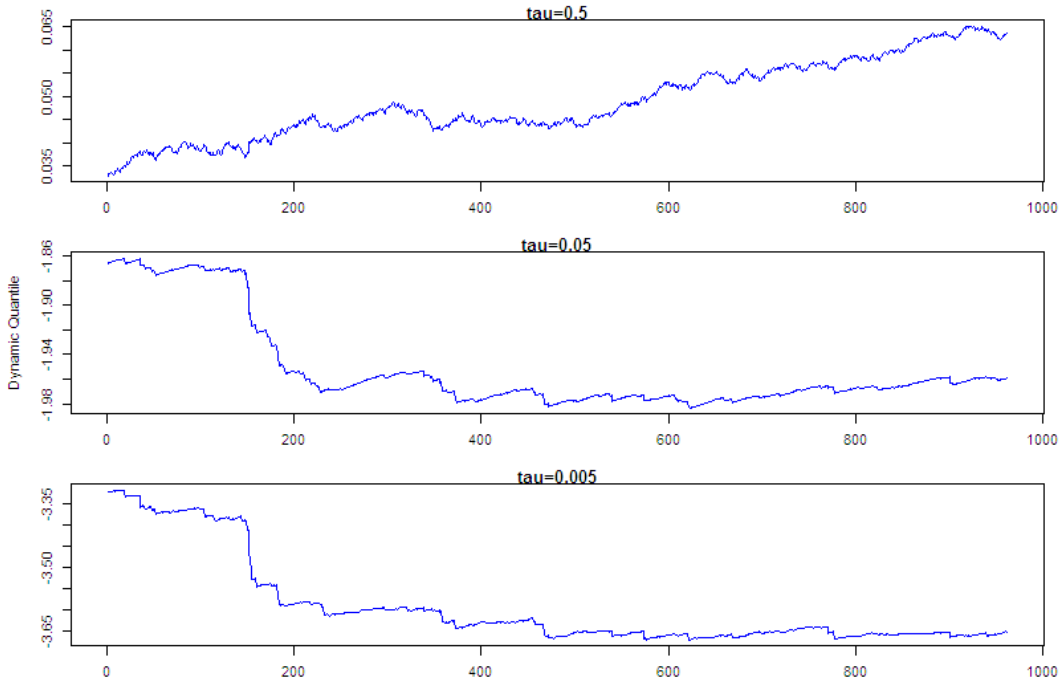


Figure 4.7 Dynamic quantiles estimation of test set.

4.3.3 Value at risk and backtesting

Value at risk (VaR) can answers the question: How much can one lose with τ probability over the pre-set horizon. The volatility estimation as well as the distribution assumption of the devolatilized returns is essential to the VaR based risk management. We can calculate VaR using the formula $VaR_{\tau,t} = \sigma_t q_\tau$. But in practice, one is interested in the prediction of VaR. In the NIG-MSA approach, we robustly estimated the volatility $\hat{\sigma}_t$. Because the volatility process is a supermartingale, so we use the estimate today as the volatility forecast $\tilde{\sigma}_{t+1}$ for tomorrow, i.e. $\tilde{\sigma}_{t+1} = \hat{\sigma}_t$. Further, we calculate the NIG distribution parameters of the devolatilized returns and estimated the dynamic quantile $\hat{q}_\tau$. Then, the VaR at the probability level τ was predicted as

$$VaR_{\tau,t+1} = \hat{\sigma}_{t+1}\hat{q}_{\tau}\,. \tag{4.9}$$

The daily VaR predictions of S&P 500 returns test set are displayed in Figure 4.8. The VaR forecasts are different between the NIG-MSA model and the Nor-GAR model. At the 5% probability level, there are more than 53 exceptions observed in Nor-GAR model and more than 50 exceptions observed in NIG-MSA model. Their exception rate respectively is 5.44% and 5.19%, both are very close to the probability level 5%. However, at the 0.5% probability level, the exceptions rate of the two models is 0.83% and 0.42% respectively. This means that as the probability level decreases to some extreme level, the gaps of these two models get larger. Figure 4.8 gives the quantitative statistics of the testing set.

We employ the back testing procedures to evaluate the validation of the VaR calculation. The standard is that a VaR calculation should not underestimate the market risk. Let N denote the number of exceptions at time t, $t = 1, 2, ..., T$. We hope that the proportion of exceptions N/T equal with the fixed probability level τ. The hypothesis test is given as:

$$H_0 : E[N] = T\tau, \qquad H_1 : E[N] \neq T\tau.$$

Jorion (2001) proposed using the likelihood ratio statistic,

$$LR = -2\log[(1-\tau)^{T-N}\tau^{N}] + 2\log[(1-N/T)^{T-N}(N/T)^{N}], \tag{4.10}$$

to test this hypothesis. Under H_0, the statistic LR is asymptotically $\chi^2(1)$ distributed.

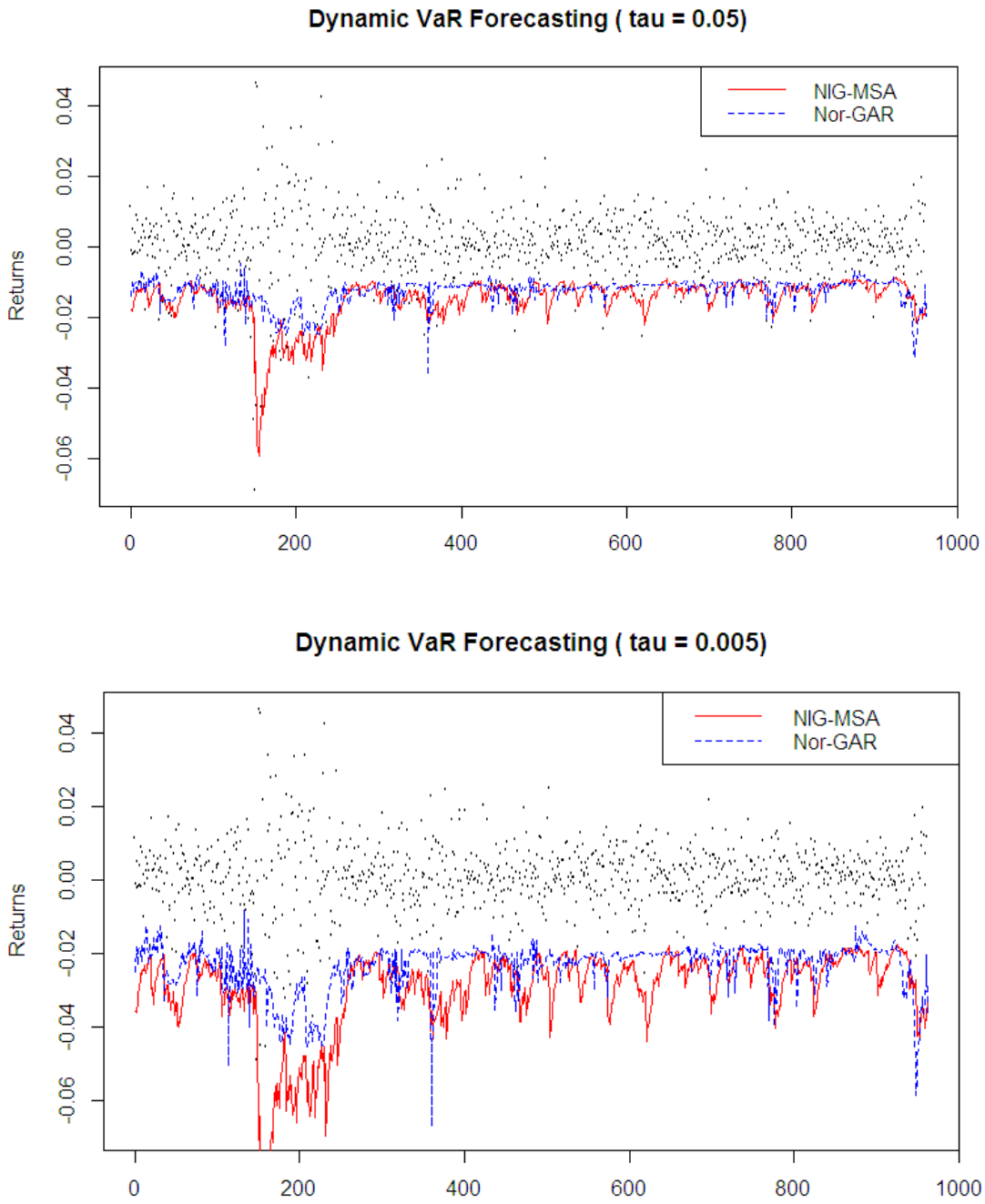


Figure 4.8 Dynamic Value at Risk forecasting of test set.

Table 4.4 summarizes the results of the backtesting for the test data. We compare the NIG-MSA model with the Nor-GAR model under four probability levels: 0.5%, 1%, 2.5% and 5%. It shows that the NIG-MSA model gives more

accurate predictions at each probability level than the Nor-GAR model. Especially under extreme probability level 1% and 0.5%, the Nor-GAR model fails to provide acceptable results under 95% confidence level.

Table 4.4 Backtesting results for the S&P 500 returns.

Model	τ	N/T	LR	p-value
Nor-GAR	0.005	0.0083	6.1576	0.0131*
	0.01	0.0135	3.8902	0.0486*
	0.025	0.0284	1.0807	0.2986
	0.05	0.0544	1.7706	0.1833
NIG-MSA	0.005	0.0042	0.1455	0.7029
	0.01	0.0104	0.0149	0.9029
	0.025	0.0260	0.1581	0.6909
	0.05	0.0519	0.0780	0.7801

* indicates the rejection of the model under confidence level 95%.

CHAPTER V

CONCLUSION AND FURTHER RESEARCH

This thesis proposes two Value-at-Risk estimation models, based on two volatility models and statistical learning method for financial time series data. The empirical analysis illustrates that our Value-at-Risk estimation models are available and effective.

The first is the A-PMCMC (Adaptive Particle Markov China Monte Carlo) approach. We primarily propose a new stochastic volatility model with leverage effect, non-constant conditional mean and jumps. Then, we propose an A-PMCMC algorithm to simultaneously estimate the model parameters and the latent variables. The A-PMCMC approach is used to estimate the VaR of stock and exchange rate returns, and the estimated results show that the proposed A-PMCMC model is excellent.

The second is a hybrid model that integrates the GARCH model with NIG distribution and multi scale analysis method. It is called NIG-MSA model. In the multi scale analysis method, the EMD is used to decompose the financial time series data and the SVR is used to integrate predicted volatility. The NIG-GARCH model fully considers the non-stationary feature of financial time series and the non-normal distribution of the returns. Comprehensive simulation experiments illustrate that the proposed model has superior performance.

For the future research, we will investigate the risk spillover between financial

time series based on the multivariable volatility model and statistical learning method. Specially, we expect to apply the idea of multiple time-scale analysis to forecast volatility of multivariable financial time series.

REFERENCES

REFERENCES

Alexander, C. (2009). **Market Risk Analysis, Value at Risk Models (Vol. 4)**. John Wiley & Sons.

Alexander, C. (2009). **Market Risk Analysis, Value at Risk Models (Vol. 4).** John Wiley & Sons.

Andrieu, C., and Roberts, G. O. (2009). The pseudo-marginal approach for efficient Monte Carlo computations. **The Annals of Statistics**. 697-725.

Andrieu, C., and Thoms J. (2008). A tutorial on adaptive MCMC. **Statistics and Computing**. 18(4): 343-373.

Andrieu, C., Doucet, A., and Holenstein, R. (2010). Particle Markov chain Monte Carlo methods. **Journal of the Royal Statistical Society: Series B** (Statistical Methodology). 72(3): 269-342.

Basak, D., Pal, S., and Patranabis, D. C. (2007). Support vector regression [J]. **Neural Information Processing-Letters and Reviews**. 11(10): 203-224.

Basak, S., and Shapiro, A. (2001). Value-at-risk-based risk management: Optimal policies and asset prices. **Review of Financial Studies**. 14: 371-405.

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

Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. **Journal of Econometrics**. 31: 307-327.

Carvalho, C. M., Johannes, M. S., Lopes, H. F., and Polson, N. G. (2010). Particle learning and smoothing. **Statistical Science**. 25(1): 88-106.

Chen, Y., Härdle, W., and Jeong, S. O. (2008). Nonparametric risk management with generalized hyperbolic distributions. **Journal of the American Statistical Association**. 103(483): 910-923.

Chernov, M., Gallant, A., Ghysels, E., and Tauchen, G. (2003). Alternative models for stock price dynamics. **Journal of Econometrics**. 116: 225-257.

Chernozhukov, V., and Hong, H. (2003). An MCMC approach to classical estimation. **Journal of Econometrics**. 115: 293-346.

Chib, S., Nardari, F., and Shephard, N. (2002). Markov chain Monte Carlo methods for stochastic volatility models. **Journal of Econometrics**. 108(2): 281-316.

Choudhry, M. (2013). **An Introduction to Value-at-Risk.** Wiley & Sons, xv–xvii.

Cogley, T., and Sargent, T. J. (2001). Evolving post-world war Π US inflation dynamics. **NBER Macroeconomics Annual**. 16: 331-388.

Eberlein, E., and Keller U. (1995). **Hyperbolic Distributions in Finance.** Bernoulli. 1(3): 281- 299.

Eberlein, E., Kallsen, J., and Kristen, J. (2003). Risk management based on stochastic volatility. **Journal of Risk**. 5: 19-44.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. **Econometrica**, 50: 987-1007.

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

Figlewski, S., and Wang, X. (2000). Is the 'Leverage Effect' a Leverage Effect? **Available at SSRN 256109.**

Franke, J., Härdle, W. K., and Hafner, C. M. (2004). **Statistics of Financial Markets (Vol. 2).** Heidelberg, Germany: Springer.

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

Ghysels, E., Harvey, A. C., and Renault, E. (1996). 5 Stochastic volatility. **Handbook of Statistics**. 14: 119-191.

Golightly, A. (2009). Bayesian filtering for jump-diffusions with application to stochastic volatility. **Journal of Computational and Graphical Statistics**. 18(2): 384-400.

Golightly, A., and Wilkinson, D. J. (2011). Bayesian parameter inference for stochastic biochemical network models using particle Markov chain Monte Carlo. **Interface Focus**. rsfs20110047.

Gordon, N. J., Salmond, D. J., and Smith, A. F. (1993, April). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. **In IEE Proceedings F-Radar and Signal Processing** (Vol. 140, No. 2, pp. 107-113). IET.

Guhathakurta, K., Mukherjee, I., and Chowdhury, A. R. (2008). Empirical mode decomposition analysis of two different financial time series and their comparison. **Chaos, Solitons & Fractals**. 37(4): 1214-1227.

Harper, D. (2008). **Introduction to Value at Risk (VaR).** (Online). Available: http://www. investopedia.com/articles/04/092904. Asp.

Hong, L. (2011). Decomposition and forecast for financial time series with high-frequency based on empirical mode decomposition**. Energy Procedia**. 5: 1333-1340.

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., and Liu, H. H. (1998, March). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. **In Proceedings of the**

Royal Society of London A: Mathematical, Physical and Engineering Sciences. **The Royal Society.** 454: 903-995.

Huang, N. E., Wu, M. L., Qu, W., Long, S. R., and Shen, S. S. (2003). Applications of Hilbert–Huang transform to non-stationary financial time series analysis. **Applied Stochastic Models in Business and Industry**. 19(3): 245-268.

Jacquier, E., Polson, N. G., and Rossi, P. E. (1994). Bayesian analysis of stochastic volatility models. **Journal of Business and Economic Statistics**. 12: 371-389.

Jorion, P. (2001). **Financial Risk Manager Handbook 2001-2002** (Vol. 96). Wiley.

Kazem, A., Sharifi, E., Hussain, F. K., Saberi, M., and Hussain, O. K. (2013). . Support vector regression with chaos-based firefly algorithm for stock market price forecasting. **Applied Soft Computing**. 13(2): 947-958.

Koop, G., and Potter, S. M. (2007). Estimation and forecasting in models with multiple breaks. **Review of Economic Studies**. 74: 763-789.

Li, H., Wells, M. T., and Cindy, L. Y. A. (2008). Bayesian analysis of return dynamics with levy jumps. **Review of Financial Studies**. 21(5): 2345-2378.

Li, Q., Yang, J., Cheng, H., and Chang, Y. J. (2005). The relationship between stock returns and volatility in international stock markets. **Journal of Empirical Finance**, 12: 650-665.

Liesenfeld, R., and Jung, R. C. (2000). Stochastic volatility models: conditional normality versus heavy-tailed distributions. **Journal of Applied Econometrics.** 15: 137-160.

Liu, J., and West, M. (2001). Combined parameter and state estimation in simulation-based filtering. **In Sequential Monte Carlo Methods in Practice**. Springer New York. 197-223.

Merton, R. C. (1971). Optimum consumption and portfolio rules in a continuous-time model. **Journal of Economic Theory**. 3(4): 373-413.

Merton, R. C. (1976). Option pricing when underlying stock returns are discontinuous. **Journal of Financial Economics**. 3(1): 125-144.

Micciche, S., Bonanno, G., Lillo, F., and Mantegna, R. N. (2002). Volatility in financial markets: stochastic models and empirical results. **Physica A: Statistical Mechanics and its Applications**. 314: 756-761.

Nielsen, B. O. (1977). Beech seeds as an ecosystem component. **Oikos**. 268-274.

Peiro, A. (1999). Skewness in financial returns. **Journal of Banking and Finance.** 23(6): 847-862.

Peters, G. W., Balkrishnan, K., and Lasscock, B. (2000). Model selection and adaptive Markov chain Monte Carlo for Bayesian cointegrated VAR models. **Bayesian Analysis**. 5(3): 465-492.

Pitt, M. K., and Shephard, N. (1999). Filtering via simulation: Auxiliary particle filters. **Journal of the American Statistical Association**. 94(446): 590-599.

Pitt, M. K., dos Santos Silva, R., Giordani, P., and Kohn, R. (2012). On some properties of Markov chain Monte Carlo simulation methods based on the particle filter. **Journal of Econometrics**. 171(2): 134-151.

Premanode, B., and Toumazou C. (2013). Improving prediction of exchange rates using Differential EMD. **Expert Systems with Applications**. 40(1): 377-384

Raggi, D. (2005). Adaptive MCMC methods for inference on affine stochastic volatility models with jumps. **The Econometrics Journal**. 8(2): 235-250.

Rasmussen, D. A., Ratmann, O., and Koelle, K. (2011). Inference for nonlinear epidemiological models using genealogies and time series. **PLoS Comput Biol**. 7(8): e1002136.

Roberts, G. O., and Rosenthal, J. S. (2009). Examples of adaptive MCMC. **Journal of Computational and Graphical Statistics**. 18: 349-367.

Roberts, G. O., and Rosenthal J. S. (2001). Optimal scaling for various Metropolis-Hastings algorithms. **Statistical Science**. 16(4): 351-367.

Rogachev, A. (2002). **Dynamic Value-at-risk.** University of St. Gallen.

Rosillo, R., Giner, J., and De la Fuente, D. (2014). Stock market simulation using support vector machines. **Journal of Forecasting**. 33(6): 488-500.

Shephard, N. (1996). Statistical aspects of ARCH and stochastic volatility. **Monographs on Statistics and Applied Probability**. 65: 1-68.

Skjeltorp, J. A. (2000). Scaling in the Norwegian stock market. **Physica A: Statistical Mechanics and its Applications**. 283(3): 486-528.

Smola, A. J., and Schölkopf, B. (2004). A tutorial on support vector regression[J]. **Statistics and Computing**. 14(3): 199-222.

Storvik, G. (2002). Particle filters for state-space models with the presence of unknown static parameters. **IEEE Transactions on Signal Processing**. 50(2): 281-289.

Szegö, G. (2002). Measures of risk. **Journal of Banking and Finance**. 26(7): 1253-1272.

Tang, L. B., Tang, L. X., and Sheng, H. Y. (2009). Forecasting volatility based on wavelet support vector machine. **Expert Systems with Applications**. 36(2): 2901-2909.

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

Taylor, S. J. (1994). Modeling stochastic volatility: A review and comparative study. **Mathematical Finance**. 4(2): 183-204.

Tsukahara, H. (2014). Estimation of distortion risk measures. **Journal of Financial Econometrics**. 12(1): 213-235.

Vapnik, V., Golowich, S. E., and Smola, A. (1996). Support vector method for function approximation, regression estimation, and signal processing. **In Advances in Neural Information Processing Systems 9**.

Verhoeven, P., and McAleer, M. (2004). Fat tails and asymmetry in financial volatility models. **Mathematics and Computers in Simulation**. 64(3): 351-361.

Wang, B., Huang, H., and Wang, X. (2013). A support vector machine based MSM model for financial short-term volatility forecasting. **Neural Computing and Applications**. 22(1): 21-28

Wood, F., van de Meent, J. W., and Mansinghka, V. (2014). A New Approach to Probabilistic Programming Inference. In *AISTATS*, 1024-1032.

Yamai, Y., and Yoshiba, T. (2005). Value-at-risk versus expected shortfall: A practical perspective. **Journal of Banking and Finance**. 29: 997-1015.

Yeh, C. Y., Huang, C. W., and Lee, S. J. A. (2011). Multiple-kernel support vector regression approach for stock market price forecasting. **Expert Systems with Applications**. 38(3): 2177-2186.

Yu, L., Wang, S., and Lai, K. K. (2008). Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm. **Energy Economics**. 30(5): 2623-2635.

Yu, J. (2005). On leverage in a stochastic volatility model. **Journal of Econometrics**. 127: 165-178.

CURRICULUM VITAE

NAME: Xinxia Yang

GENDER: Female

NATIONALITY: Chinese

DATE OF BIRTH: November 30, 1979

EDUCATION BACKGROUND :

B. Sc. School of Mathematics, Guizhou University, Guizhou, China, 2002.

M. Sc. in Economics, Wuhan University, Wuhan, China, 2008.

AFFILIATION AND ADDRESS:

School of Mathematics and Statistics, Guizhou University of Finance and Economics, Guiyang, 550025, China

PUBLICATIONS :

X. Yang, R. Chatpatanasiria, P. Sattayathama. Value at risk estimation under stochastic volatility models using adaptive PMCMC methods. **Communications in Statistics - Simulation and Computation** (Submitted)**.**

X. Yang, R. Chatpatanasiria, P. Sattayathama. Dynamic Risk Measurement of Financial Time Series with Heavy-tailed: A New Hybrid Approach. **Thai Journal of Mathematics** (Submitted).