

2014

FORECASTING STOCK MARKET RETURNS VOLATILITY

Yanan Li
University of Rhode Island, yanan_li@my.uri.edu

Follow this and additional works at: <https://digitalcommons.uri.edu/theses>

Terms of Use

All rights reserved under copyright.

Recommended Citation

Li, Yanan, "FORECASTING STOCK MARKET RETURNS VOLATILITY" (2014). *Open Access Master's Theses*. Paper 346.
<https://digitalcommons.uri.edu/theses/346>

This Thesis is brought to you by the University of Rhode Island. It has been accepted for inclusion in Open Access Master's Theses by an authorized administrator of DigitalCommons@URI. For more information, please contact digitalcommons-group@uri.edu. For permission to reuse copyrighted content, contact the author directly.

FORECASTING STOCK MARKET RETURNS
VOLATILITY

BY

YANAN LI

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
IN
STATISTICS

UNIVERSITY OF RHODE ISLAND

2014

MASTER OF SCIENCE THESIS

OF

YANAN LI

APPROVED:

Thesis Committee:

Major Professor Gavino Puggioni

Liliana Gonzalez

Tong Yu

Nasser H. Zawia

DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND

2014

ABSTRACT

It has been widely known that the stock market is always volatile and full of risk. How to better capture the volatility and decrease risk accordingly has become a main concern for both investors and researchers. In this thesis, the stochastic volatility model with offset mixture of normal distribution is fitted for financial dataset NASDAQ:LLTC daily stock market returns volatility and one-step-ahead prediction is made based on the AR(1) SV model. Bayesian analysis is fully applied for model fitting and parameter estimation. The Markov Chain Monte Carlo algorithm, using the Metropolis Hasting method, the Forward Filtering Backward Sampling and the Gibbs Sampler is well developed to fit the real data. A small improvement incorporated is the resampling of weights in the discrete normal mixture distribution which is used to approximate a non-normal distribution. Estimated parameters when having weights sampled are compared with the results when weights are fixed. The predictive distribution for one-step-ahead log volatility z_{T+1} and log transformed stock return y_{T+1} is given in the graphs. Mean and 95% posterior interval are also provided for both z_{T+1} and y_{T+1} . FFBS algorithm is first applied to a simulated dataset with normal mixture structure in Dynamic Linear Models. Visual plots with posterior mean and 95% posterior interval are given. Autoregressive model with application of Monte Carlo approximation is also included to model LLTC stock returns.

ACKNOWLEDGMENTS

Foremost, I would like to express my sincere gratitude to my advisor Professor Gavino Puggioni for the continuous support of my master's study and research, for his guidance on my job interviews, for his patience, motivation, enthusiasm and immense knowledge. His guidance helped me in all the time of research and writing of this thesis but not limited to that. He is the best advisor I could've ever had for my master's study. I also feel proud by the fact that I am the first student that he has supervised.

Besides my advisor, I would like to thank the rest of my thesis committee: Prof. Gonzalez, Prof. Yu, Prof. Finizio for their encouragement, patience and great suggestions. My special thanks go to Prof. Gonzalez, for admitting me into the Statistics Program, and for offering me funding for my second year of study at URI.

My sincere thanks goes to Prof. Yu, for his continuous support on my thesis, and for the great opportunity he provides me to get exposed to the finance area.

My sincere thanks also goes to my committee chair Prof. Finizio, for teaching me two math classes which I enjoy a lot, also for agreeing to be part of my committee.

I thank my friends Zonghao Zhu , Si Yang, Mike Piserchia, Aihong Xi, Yun Chu for being my friends and for whatever they have done for me to make my life warmer and brighter.

Last but not the least, I would like to thank my family: my parents Xianming Li and Yan Wang, my brother Xiaolin Li, for their selfless love for me and their support both financially and spiritually throughout my life.

TABLE OF CONTENTS

ABSTRACT.....	ii
ACKNOWLEDGMENTS	iii
TABLE OF CONTENTS.....	iv
LIST OF TABLES	v
LIST OF FIGURES	vi
CHAPTER 1	1
INTRODUCTION	1
CHAPTER 2	13
REVIEW OF LITERATURE	13
CHAPTER 3	20
METHODOLOGY.....	20
CHAPTER 4	35
FINDINGS	35
CHAPTER 5	50
CONCLUSION.....	50
BIBLIOGRAPHY	52

LIST OF TABLES

TABLE	PAGE
Table 1. Discrete normal mixture approximation to log chi-square distribution.....	27
Table 2. Estimated posterior mean and 95% posterior interval	29
Table 3. Mean and 95% posterior interval for one-step-ahead volatility prediction ..	47
Table 4. Comparison of fitted results for SV models when having weights fixed and sampled	49

LIST OF FIGURES

FIGURE	PAGE
<p>Figure 1. Simulated series from AR(1) with mixture observational errors(circles and solid circles). The red solid line corresponds to the posterior mean of the smoothing distribution over time $(\mu_t D_{200})$ obtained from a model that ignores the mixture structure in the observational errors. The dotted lines are 95% posterior bands for $(\mu_t D_{200})$</p>	25
<p>Figure 2. Residual analysis obtained from fitting the model $\{1,0.9,1,1\}$ to the simulated data The top and bottom plots display, respectively, the standardized residuals and the sample ACF of the residuals.</p>	25
<p>Figure 3. Simulated series. Posterior mean and 95% posterior bands for $(\mu_t D_{200})$ obtained from the model described by Equations (1) and (2).....</p>	26
<p>Figure 4. Posterior mean of the latent process λ_t</p>	26
<p>Figure 5. A snapshot of the raw data.....</p>	34
<p>Figure 6. Top: Time series and histogram for LLTC stock price; Bottom: Density and autocorrelation function for stock price</p>	36
<p>Figure 7. Top: Time series plot for LLTC stock market returns. Bottom:A closer look at the return series from the fist 200 observations.</p>	37
<p>Figure 8. Top: Histogram and QQ plot for LLTC Daily Returns; Bottom: Density and Correlogram for LLTC Returns</p>	38
<p>Figure 9.Top: Simulated posterior variance for innovation terms; Bottom:Histogram for simulated posterior variances</p>	40

Figure 10. The absolute LLTC stock returns	42
Figure 11. Estimated volatility process σ_t in the standard univariate SV model	42
Figure 12. Top: Complete traceplot and traceplot after burn in period for ϕ ; Bottom: Histogram and correlogram for sampled ϕ after burn in period	43
Figure 13. Top: Complete traceplot and traceplot after burn in period for μ ; Bottom: Histogram and correlogram for sampled μ after burn in period	44
Figure 14. Top: Complete traceplot and traceplot after burn in period for ν ; Bottom: Histogram and correlogram for sampled ν after burn in period	44
Figure 15. One step ahead volatility prediction histograms.....	46
Figure 16. One step ahead prediction for y_{T+1} , mean and 95% posterior interval.....	47
Figure 17. Red curve is the posterior mean of the predictive distribution for log transformed stock returns; Yellow curves represent the simulated predictive distribution after burn in.	48

CHAPTER 1

INTRODUCTION

Stock is a type of security that signifies an ownership position in a corporation. A company can be divided into a number of shares and each share of stock is entitled to a proportional share of profit or loss made by the company. It is a representative of the claim as part of the corporation's assets and earnings. There are many options for people who want to make investments. The buying and selling of stock is always the most popular option for public trading. When investors buy stocks, they become shareholders, which means that they own a part of the company. If the company's profits go up, they will share those increased profits with the company. Similarly, if the company's profits go down, the stock price goes down accordingly and the loss in profits will be shared with investors too. The logic to make money is that investors buy the stock, hold it for some time, and then sell it at a higher price than the purchasing price. Suppose they sold their stock at a price lower than the price they have paid for it, they would lose money.

It is widely known that stock price is very changeable, even on a daily basis. The reason for that is because of supply and demand. In stock markets, a large volume of stocks are traded every day. If there are more people who buy a stock than the people who sell it, out of the expectation that the price will go up in the future, then the price will rise. Conversely if more people want to sell it than to buy it, the stock price will fall dramatically. However, investors' expectation for the market is in a

permanent state of fluctuation due to all kinds of information obtained over time that strongly affect their decision-making. That's also why the stock trading has been processed so often over a short period of time.

Most stocks are traded on stock exchanges, which are the places where buyers and sellers meet and make a deal on a price. One of the most famous stock exchanges is the New York Stock Exchange (NYSE), in which much of the trading is done face to face on the trading floor and prices are determined using the auction method. The NASDAQ stock exchange, commonly known as NASDAQ, is the second type of stock exchange, where the trading is done through a computer and telecommunication network of dealers. This is also called the over-the-counter (OTC) market. NASDAQ used to be the largest company listed only on the NYSE. Due to the late 90s' technology boom, NASDAQ is now becoming home to several big technology companies such as Microsoft, Intel, Dell etc. LLTC is also one of them, which I am going to talk about in detail in my thesis.

Linear Technology Corporation (LLTC), a member of the S&P 500, has been designing, manufacturing and marketing a broad line of high performance analog integrated circuits for major companies worldwide for three decades. The Company's products provide an essential bridge between our analog world and the digital electronics in communications, networking, industrial, automotive, computer, medical, instrumentation, consumer, and military and aerospace systems. Linear Technology produces power management, data conversion, signal conditioning, RF and interface ICs, μ Module subsystems, and wireless sensor network products. (see LLTC website)

As a master's candidate, I justify my research on "Forecasting Stock Market

Returns Volatility” based on the following introduction and overview.

1.1 Stock Market Returns

In a stock market, four stock prices will usually be provided in a day which include open price, close price, high price and low price. Open price is the price at which a specified stock first trades upon the opening of an exchange on a given trading day. Close price is the final price at which a stock is traded. Close price is important because it represents the most up-to-date valuation of a security before the next trading day. The closing price of one day can be compared to the previous close price in order to measure market sentiment.

In the thesis, stock daily return is calculated by taking the natural logarithm of the ratio of stock closing price at time $t+1$ and time t , denoted by $\ln(p_t/p_{t-1})$, which is a commonly used approximation of the percentage of return $r_t = (p_t - p_{t-1})/p_{t-1}$. There are several reasons for doing so. One reason is that when return r_t is very small, we have $\ln(1+r_t) \approx r_t$, $r_t \ll 1$ based on the Taylor expansion in mathematics. Another reason is that the multiplication of small numbers $\prod_{i=1}^n r_i$ will cause arithmetic underflow and taking the logarithm is a modification to this potential problem by the summation of log values $\sum_{i=1}^n \log(1+r_i)$. Hence, loosely speaking, log return means the ratio of money gained or lost on a stock. Positive returns reflect a rising market (bull market) where people make money, while negative returns are usually referred as a “bear market” where people lose money.

1.2 Stock Market Volatility

Volatility is a statistical measure of the dispersion or variability of returns for a given security or market index. It refers to how uncertain one is about the size of change in a security's value over time. Volatility, which is always positive, can either be measured by variance or standard deviation (the square root of variance), between returns over a specified period of time. A high volatility means that the returns can potentially take values in a large range of values, so the uncertainty about the stock returns is also high which represents high risk. Hence high volatility implies that the returns can change dramatically in values over a short period of time, since large amounts of stocks are traded within any minute. Conversely, a low volatility means the returns don't fluctuate dramatically over a short time period. In general, the higher the volatility, the riskier the stock market is. And the more risk investors take, the greater the potential for higher gain or loss.

1.3 Why Stock Market Returns?

Stock price is the price of a single share of many saleable stocks for a company or other financial institution. It essentially, is a function of the amount of dividends that can be expected in the future. Therefore, the current price at a given time point t reflects the whole investment community's expectation and confidence towards a stock in the future. If the underlying expectation for the market stays the same, then stock price will not make any change. But nobody is really sure about the change of stock price. Different expectations for future stock price lead to frequent buying and selling of stocks, which in reverse, causes the constant change of price over time. As we can observe directly from the stock markets, even in a single day, four different prices are given to reflect the volatility of stock prices. Basically stock price depends

on demands and supply driven by buyers and sellers. The interest of the buyers and sellers for stocks depends on market sentiment. Market sentiment relies on the domestic and international economy and other factors, which are very complex and unpredictable. So it is extremely hard to predict stock prices directly, if not impossible. Obtaining stock returns, however, makes it more convenient to analyze changes in stock prices by making a transformation on them. Changes in stock returns can easily be used to make inference on stock prices. After a transformation, we don't have to deal with the high correlations among stock prices since low correlations exist among stock returns and volatility in returns is also lower than in stock prices.

For LLTC stock returns (see Figure 7), the means of the series seem constant over the long run and only the variances keep changing. Some big spikes show up every now and then. What I am aiming to do in this thesis is to fit a model that can best capture the volatility and make a short-term prediction on future volatility with the assumption that the possible values of the future stock returns is within the range of what we have observed in the data.

1.4 Time Series

A univariate time series is a chronological sequence of observations about a particular variable, which is usually denoted as y_t ($t = \dots -2, -1, 0, 1, 2, \dots$), e.g., exchange rate, inflation rate, product sales, unemployment etc. Usually time t is taken at equally spaced intervals, and the unit of time may be anything from seconds to years. Time series analysis can be useful to detect the change over time for a security or other time series variables. Five common features in economic and business time series include trends, seasonality, aberrant observation, conditional heteroskedasticity and non-

linearity, see Franses' book "Time series models for business and economic forecasting" (2000) at Chapter Two for details. Besides that, a comparison can be made on multiple time series over the same period of time. As for financial time series, for example, co-integration was raised to specifically investigate the co-movement/common trend between two or more financial time series. Financial time series analysis (FTSA), is concerned with the theory and practice of asset valuation over time. What makes financial time series analysis different is that it is highly volatile and empirical such as stock market indices, market shares. Furthermore, it relies more on statistical theory and methods for the development of robust models since there is no universal model that will fit every financial time series.

Stock returns series is a typical financial time series with high volatility. Due to the sensitivity of stock prices to economic events or interest rates, it is important to build a volatility model with high accuracy to better capture the change in stock returns over time. Bayesian analysis can be applied to accomplish the goal here.

1.5 Bayesian Analysis

In a precise mathematical sense, it has been shown that probabilities can numerically represent a set of rational beliefs, that is we claim that probabilities are a way to numerically express rational beliefs. Bayesian statistics is thus founded on the fundamental premise that all uncertainties about quantities should be represented and measured by probabilities. We use statistical induction to learn about the general characteristics of a population from a subset of the population.

1.5.1 Notations and Definitions

Stochastic Process: A stochastic process is a family of random variables

defined on a given probability space, indexed by the time variable t , which is used to represent the evolution of some random values or system over time. It is also known as a random process. A time series process is a stochastic process.

Stationarity: A strict stationary process is a stochastic process whose joint probability distribution doesn't change when shifted in time or space. Thus, stationarity explores the time invariant behavior of a time series. Determining the stationarity condition of the time series allows for proper identification and development of forecasting models. There are two types of stationarities: strict stationarity whose distribution is time invariant, and weak stationarity for which only the first two moments are time invariant, that is the data values fluctuate with constant variation around a constant level. Most financial time series exhibit a weak form stationarity.

Autocorrelation: Autocorrelation describes the serial correlation between values of a stochastic process at different times, as a function of the time lag. Let $\{y_t, t \in T\}$ be a time series process. Then the sample autocorrelation function is given by

$$\rho_l = \frac{\sum_{t=1}^{T-l} (y_t - \bar{y})(y_{t+l} - \bar{y})}{\sum_{t=1}^T (y_t - \bar{y})^2},$$

where \bar{y} is the sample mean and T is the sample size. Existence of autocorrelation implies the return is predictable, indicating market inefficiency.

White Noise: White noise is a simple type of stochastic process whose terms are identically independently distributed (iid) with zero mean. A Gaussian white noise is a stochastic process with zero mean, finite variance and zero autocorrelation.

1.5.2 Bayes' Rule

Numerical description of population characteristics are typically expressed in terms of a set of parameters, θ s, and a numerical description of the sample make up a data set y . Before a data set is obtained, the numerical values of θ, y are uncertain. After a dataset is obtained, the information it contains can be used to decrease our uncertainty about the population characteristics, which is the goal of Bayesian inference. Bayes' rule (1.1) provides us with a rational method to update the uncertainty as new information about y is collected.

For each numerical parameter $\theta \in \Theta$ (Θ is the set of possible parameter values for θ), our prior distribution $p(\theta)$ describes our belief that θ represents the true population characteristics. For each $\theta \in \Theta$ and $y \in Y$, our sampling model $p(y|\theta)$ describes our belief that y would be the outcome of our study if we know that θ is true. Once we obtain the data set y , the last step is to update our belief about θ . For each $\theta \in \Theta$, our posterior distribution $p(\theta|y)$ describes our belief that θ is the true value, after observing data set y . The posterior distribution is obtained from the prior distribution and the sampling model via Bayes' rule:

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} = \frac{p(y|\theta)p(\theta)}{\int_{\Theta} p(y|\theta)p(\tilde{\theta})} \quad (1.1)$$

It is important to note that Bayes' rule does not tell us what our beliefs should be, it tells us how they should change after observing new information. Even if a particular prior distribution does not exactly reflect our prior information, the corresponding posterior distribution can still be a useful means of providing stable

inference and estimation for both large and small sample sizes.

1.5.3 Metropolis-Hasting Algorithm

In a generic situation where we have a sampling model $Y \sim p(y|\theta)$ and the prior distribution $p(\theta)$, our target distribution is

$$p(\theta|y) = \frac{p(\theta)p(y|\theta)}{\int p(\theta')p(y|\theta')d\theta'} \quad (1.2)$$

What Bayesian simulation usually does is to sample from the posterior distribution and obtain a Monte Carlo approximation to posterior quantities, if $p(\theta|y)$ is tractable with an analytical form. But if we can't sample directly from $p(\theta|y)$, the Metropolis algorithm needs to be used to approximate the posterior distribution. In terms of estimating the posterior distribution, the key is to be able to construct a large collection of θ values, whose empirical distribution approximates $p(\theta|y)$. Suppose we already have an acceptable collection $\{\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(s)}\}$, and we are considering to add another value $\theta^{(s+1)}$. With a proposed value θ^* , whether we add it to the set or not depends on the ratio $r = p(\theta^*|y)/p(\theta^{(s)}|y)$. How the Metropolis algorithm generates a value $\theta^{(s+1)}$ is as follows:

1. Sample $\theta^* \sim J(\theta|\theta^{(s)})$. $J(\theta|\theta^{(s)})$ is a symmetric proposal distribution.
2. Compute the acceptance ratio

$$r = \frac{p(\theta^*|y)}{p(\theta^{(s)}|y)} = \frac{p(y|\theta^*)p(\theta^*)}{p(y|\theta^{(s)})p(\theta^{(s)})}$$

3. Let

$$\theta^{(s+1)} = \begin{cases} \theta^* & \text{with the probability } \min(r, 1) \\ \theta^{(s)} & \text{with the probability } 1 - \min(r, 1) \end{cases}$$

Step 3 can be accomplished by sampling $u \sim \text{uniform}(0, 1)$ and setting $\theta^{(s+1)} = \theta^*$ if $u < r$ and setting $\theta^{(s+1)} = \theta^{(s)}$ otherwise.

1.5.4 Gibbs Sampler

Suppose we have a vector of parameters $\Phi = \{\phi_1, \phi_2, \dots, \phi_p\}$, and the prior information is given by $p(\Phi) = p(\phi_1, \phi_2, \dots, \phi_p)$. Given a starting point $\Phi^{(0)} = \{\phi_1^{(0)}, \phi_2^{(0)}, \dots, \phi_p^{(0)}\}$, the Gibbs sampler generates $\Phi^{(s)}$ from $\Phi^{(s-1)}$ by sampling from the full conditional distribution as follows:

1. Sample $\phi_1^{(s)} \sim p(\phi_1 | \phi_2^{(s-1)}, \dots, \phi_p^{(s-1)})$
2. Sample $\phi_2^{(s)} \sim p(\phi_2 | \phi_1^{(s)}, \phi_3^{(s-1)}, \dots, \phi_p^{(s-1)})$
- ⋮
- p. Sample $\phi_p^{(s)} \sim p(\phi_p | \phi_1^{(s)}, \phi_2^{(s)}, \dots, \phi_{p-1}^{(s)})$

This algorithm generates a dependent sequence of vectors

$$\begin{aligned} \Phi^{(1)} &= \{\phi_1^{(1)}, \phi_2^{(1)}, \dots, \phi_p^{(1)}\} \\ \Phi^{(2)} &= \{\phi_1^{(2)}, \phi_2^{(2)}, \dots, \phi_p^{(2)}\} \\ &\vdots \\ \Phi^{(s)} &= \{\phi_1^{(s)}, \phi_2^{(s)}, \dots, \phi_p^{(s)}\} \end{aligned}$$

In this sequence, if $\Phi^{(s)}$ depends on $\Phi^{(0)}, \Phi^{(1)}, \dots, \Phi^{(s-1)}$ only through $\Phi^{(s-1)}$ the sequence is called a Markov Chain. With the Gibbs sampler, we can approximate posterior means, quantile of interest using the empirical distribution of $\{\Phi^{(1)}, \Phi^{(2)}, \dots, \Phi^{(s)}\}$.

1.5.5 Standard MCMC approximation

The standard practice in the MCMC approximation, using either the Metropolis algorithm or the Gibbs sampler, is as follows:

1. Run the algorithm until some iteration B for which it looks like the Markov Chain has achieved stationarity;
2. Run the algorithm S more times, generating $\{\theta^{(B+1)}, \dots, \theta^{(B+S)}\}$;
3. Discard $\{\theta^{(1)}, \dots, \theta^{(B)}\}$ and use the empirical distribution of $\{\theta^{(B+1)}, \dots, \theta^{(B+S)}\}$ to approximate $p(\theta|y)$.

The iterations up to and including B are called the “burn-in” period, in which the Markov Chain created after burn in has higher posterior probability. Another reason for that is to weaken the influence of the initial values, especially when we don’t have a good idea about the prior belief.

1.6 Application of Stochastic Volatility Models

Many models exist in literature and in practice for the uncertainty of unrealized volatility. Model calibration is made to find a set of parameters that minimizes the difference between the model predictions and realized market data. The principle of model selecting is to provide the most ease with respect to market calibration, instead of capturing the particular dynamic features within the related structure. Stochastic volatility models have been widely used in derivative pricing and hedging in the past decade since a non-constant implied volatility was brought to attention and become more acceptable, especially after the 1987 crash. A natural extension of SV models is to modify the specification of volatility in the stochastic dynamics of the underlying asset price model. A variety of effects are considered in different modeling domains,

such as leverage effects, supply and demand, declined stock prices leading to massive portfolio volatility than increased stock price. Stochastic volatility correction has been made to Black-Scholes (The Journal of Political Economy, 1973) by a new approach which requires volatility to be mean-reverting and transforms the slope and intercept of the implied volatility skew into information about the correlation between volatility and stock price shocks and the market's volatility risk premium. When volatility persists, the derivative price can be approximated in the SV environment by pricing a more complicated security in the Black-Scholes constant volatility environment. The payoff structure of the new security depends on the Black-Scholes pricing formula for the original one and accounts appropriately for volatility risk. (Fouque et al., 2000)

When it comes to volatility forecasting, there are many practical applications. Since volatility is the essential risk aspect of the market, a large part of financial risk management is to capture volatility in tractable statistical models and to measure and manage the potential future losses. Asset allocation is the way you allocate your investment in bills, stocks, bonds etc. To balance the risk and reward for individuals, volatility forecasting is also a problem that can't be ignored. Besides that, the most challenging application of volatility forecasting is to use it for developing a volatility trading strategy.

CHAPTER 2

REVIEW OF LITERATURE

2.1 Review of Models for Volatility

The rapid growth in the financial market and the continual development of more complex financial instruments requires advanced statistical methods to gain the theoretical and empirical knowledge of financial time series. It is widely known that the daily returns of financial assets, especially stock returns, are difficult to predict, even though the volatility of stock returns seems easier to forecast. The time-varying volatility models have been used in various contexts of a time series analysis.

The simplest model is the ARCH model, which stands for Autoregressive Conditional Heteroscedasticity, first developed by Engle (1982). The AR comes from the fact that these models are autoregressive models in squared returns. The conditional comes from the fact that next period's volatility is conditional on information from this period. Heteroscedasticity means non constant volatility. Let us assume that the return on an asset is

$$r_t = \mu + \sigma_t \varepsilon_t \quad (2.1)$$

where $\varepsilon_t \stackrel{i.i.d}{\sim} N(0,1)$. In an ARCH(1) model, the residual return at time t is defined as $a_t = r_t - \mu$,

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_t^2 \quad (2.2)$$

where $\alpha_0 > 0$ and $\alpha_1 > 0$ to ensure positive variance and $\alpha_1 < 1$ to ensure stationarity.

The conditional variance is expressed as a function of the squares of past observations and past variances.

An extension of the ARCH model leads to the GARCH model or Generalized ARCH model. The GARCH model was proposed by Bollerslev (1986) and Taylor (1986) primarily to overcome the large number of ARCH parameters that were needed to model the volatility process. The fundamental idea of the GARCH model is to describe the evolution of the time-varying variance σ_t^2 with a moving average structure. Let $r_t = \mu_t + a_t$, then a_t follows a GARCH (p, q) model if

$$\begin{aligned} a_t &= \sigma_t \varepsilon_t \\ \sigma_t^2 &= \alpha_0 + \alpha_1 a_{t-1}^2 + \dots + \alpha_p a_{t-p}^2 + \beta_1 \sigma_{t-1}^2 + \dots + \beta_q \sigma_{t-q}^2 \\ \varepsilon_t &\sim N(0,1) \end{aligned} \quad (2.3)$$

When p=q=1, the GARCH (1,1) model is expressed as $\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2$, in this model the next period forecast of variance is a combination of the last period forecast and last period's return. A special case of the GARCH (1,1) model arises when $\alpha_1 + \beta_1 = 1$ and $\alpha_0 = 0$, which is known as IGARCH model or Integrated GARCH model. In this case it is common to use the symbol λ for β_1 and the equation takes a simpler form

$$\sigma_t^2 = (1-\lambda)\varepsilon_{t-1}^2 + \lambda\sigma_{t-1}^2 = (1-\lambda)\sum_{i=0}^{\infty} \lambda^i \varepsilon_{t-i}^2 \quad (2.4)$$

The variance in this case can be interpreted as a weighted average of all previous squared returns with the weights decreasing exponentially over time.

The stochastic volatility models (proposed by Taylor, 1986) assume volatility follows some latent stochastic process and introduce the innovation in the conditional

variance. It was developed out of a need to modify the Black Scholes model for option pricing in the theoretical finance literature, which failed to effectively take the volatility in the price of the underlying security into account. The Black Scholes model assumed that the volatility of the underlying security was constant, while stochastic volatility models categorized the price of the underlying security as a random variable. Allowing the price to vary in the stochastic volatility models improved the accuracy of calculations and forecasts. The canonical AR(1) SV model for regularly spaced data is as follows

$$\begin{aligned}
 y_t &= \beta e^{h_t/2} \varepsilon_t \\
 h_{t+1} &= \mu + \phi(h_t - \mu) + \sigma_\eta \eta_t, \\
 h_1 &\sim N\left(\mu, \frac{\sigma^2}{1-\phi^2}\right)
 \end{aligned} \tag{2.5}$$

where h_t is the log-volatility following a stationary autoregressive process with order 1, the parameter β or $\exp(\mu/2)$ is a constant factor and can be thought of as modal instantaneous volatility. This is not a linear model, and transformation is needed to proceed with the analysis.

2.2 ARCH/GARCH Models and SV Models Comparison

Modeling volatility plays a crucial role in risk management in banks or other financial institutions since volatility is considered to be a measure of risk. It is now widely agreed that financial asset returns volatilities are time-varying, with persistent dynamics. (Andersen et al 2007). A comparison of the popular GARCH models and less known SV(or ARSV) models, both of which is capable of modeling time varying volatility and capturing the volatility clustering, will be discussed in this section.

SV models is usually considered a successful alternative of ARCH models in

modeling financial return series. The distinctive advantage of SV models is that they incorporate leverage effect (volatility tends to increase when prices go down) and also capture the main empirical properties often observed in daily return series in a more appropriate way. The reason that SV models are less popular in practice is mainly because of the complexity and difficulty of parameter estimation. SV models are non-linear and non-Gaussian and the computations are more demanding than for GARCH models. For instance, there are two error/noise terms: observation error and state error in SV models due to the assumption of latent process, but there is only one error process in GARCH models. The problem of SV models is obvious from the likelihood function where we have to integrate over the latent factor a T dimensional integral. This can not be solved analytically, so the numerical simulation is required.

However, GARCH-type models have a poor forecasting ability. Besides that, there are other problems with GARCH, such as inconsistent parameter estimate results for different time scales, inefficiency to capturing outliers and large moves and failure to distinguish the association between large moves and earnings announcements and other news. The theoretical examination of GARCH and SV comparison provided by Carnero et al. (2001) shows that SV models can better explain the excess kurtosis, low first order autocorrelation and high persistence of volatility. They also show that SV model is less dependent on the choice of returns distribution. In the paper by Mapa et al (2010), they conclude that SV models capture more aspect of volatility than GARCH model due to its sources of variability and produce lower forecast errors by comparing basic GARCH volatility forecasts with SV models which are computed through Kalman Filter or MCMC method. Hence, if we ignore the calculation

difficulties for SV models, SV models are adequate substitute to GARCH models.

2.3 SV Models Literature Summary

In the process of writing this thesis, I referred to many books and papers. The book I would like to summarize here is “Time Series: Modeling, Computation, and Inference” by Raquel Prado and Mike West. In this book, mainstream modeling approaches in time series with a range of significant recent developments in methodology and application of time series analysis are integrated. The overview of several models such as traditional time domain models, state space TVAR models, dynamic linear models, mixture models etc., and related methodology for inference, statistical computation for model fitting and forecasting is present. The estimation and forecasting are based on Bayesian analysis which involves likelihood and Bayesian methodologies, with a strong emphasis on using simulation-based approaches for model estimation, inference and forecasting.

Posterior inference and forecasting can be easily achieved in the normal DLM framework. When more general models are considered, such as non-linear and non-normal dynamic models, Markov Chain Monte Carlo algorithms can be implemented for posterior estimation. In a general framework, a nonlinear/non-Gaussian dynamic model is defined by the densities $p(y_t | \theta_t, D_{t-1})$, $p(\theta_t | \theta_{t-1}, D_{t-1})$ and the prior density $p(\theta_0 | D_0)$. Our interest is to obtain samples from the filtering distribution $p(\theta_t | D_t)$ and the joint posterior distribution $p(\theta_{0:T} | D_T)$. In a Gibbs sampling framework, we would sample from the conditional posterior $p(\theta_t | \theta_{(-t)}, D_T)$, where $\theta_{(-t)}$ consists of $\theta_{0:T}$ except the t -th element θ_t , we sequentially sample through $t = 0:T$ as follows:

1. Set initial values $\theta_{0:T}^{(0)}$.
2. For each iteration m , sample $\theta_{0:T}^{(m)}$ component by component, i.e., for each t ,
sample $\theta_t^{(m)}$ from $p(\theta_t | \theta_{0:(t-1)}^{(m)}, \theta_{(t+1):T}^{(m-1)}, D_T)$
3. Repeat the previous step until MCMC convergence.

However, the posterior $p(\theta_t | \theta_{0:(t-1)}^{(m)}, \theta_{(t+1):T}^{(m-1)}, D_T)$ is rarely calculable in practice. In this case, Metropolis-Hastings steps within the Gibbs iteration could be used. In the normal dynamic linear model, a useful MCMC method Forward Filtering Backward Sampling method was developed and proved to be powerful. For univariate stochastic volatility model, discrete mixture of normal distribution was introduced to approximate the non-linear log chi-square distribution and the number of mixtures to be chosen is seven, even though more mixtures can refine the approximation. Zero-mean AR(1) SV model and the MCMC analysis applied to real data were discussed in detail.

In the paper “Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models” by Kim, Shephard et al, Markov Chain Monte Carlo sampling methods are exploited to provide a practical likelihood-based framework for the analysis of stochastic volatility models. To sample all unobserved volatility all at once using offset mixture model followed by an importance reweighting procedure is proved to be an effective method. This approach is compared with several alternative methods with application on real data. Simulation-based methods for filtering, likelihood evaluation and model diagnostic are also developed. The fit of stochastic volatility and GARCH models are compared.

The first complete Markov Chain Monte Carlo simulation-based analysis of the SV model which covers efficient methods for Bayesian inference, likelihood evaluation, computation of filtered volatility estimates, diagnostics for model failure, and computation of statistics for comparing non-nested volatility models are provided. A very simple Bayesian method for estimating SV model, which is based on one-at-a-time updating of the volatilities, is shown to be quite inefficient from a simulation perspective. An improved method that relies on an offset mixture of normal approximation to a log-chi-square distribution coupled with an importance reweighting procedure is shown to be strikingly more effective. Additional refinements of the latter method are developed to reduce the number of blocks in the Markov Chain sampling. The paper also develops formal tools for comparing the basic SV and Gaussian and t-GARCH models and finds that the simple SV model typically fits the data as well as more heavily parameterized GARCH models. In the end, a number of extensions of the SV model are considered that can be fitted using the developed methodology.

CHAPTER 3

METHODOLOGY

3.1 Dynamic Linear Models

3.1.1 Model Introduction

When parameters of time series model are indexed in time $\theta_t = (\theta_{t,1}, \dots, \theta_{t,p})'$, the dynamic linear models for univariate time series with equally spaced observations are given by

$$y_t = F_t' \theta_t + v_t \quad (3.1)$$

$$\theta_t = G_t \theta_{t-1} + w_t \quad (3.2)$$

y_t is the observation at time t ;

θ_t is the state vector $(\theta_{t,1}, \theta_{t,2}, \dots, \theta_{t,p})$ at time t ;

F_t is a p -dimensional column vector of known constants at time t ;

v_t is the observation noise with $N(v_t | 0, v_t)$;

G_t is a $p \times p$ matrix known as evolution or transition matrix;

w_t is the state noise, or innovation, with $N(w_t | 0, W_t)$.

v_t and w_t are independent and mutually independent.

Then y_t has a DLM representation given by $\{F_t, G_t, v_t, W_t\}$. (3.1) is often referred to as the “Observation Equation”, and (3.2) as the “Evolution Equation”.

In general, Dynamic Linear Models are given by probability density functions

$p(y_t|\theta_t)$ and $p(\theta_t|\theta_{t-1})$, which define a conditional dependence structure between observations y_t and parameters θ_t . DLMs has a sequential nature and one of the main targets from a Bayesian approach is $p(\theta_t|D_t)$: the posterior distribution of θ_t given all the information available at time t , i.e., $D_t = \{y_1, y_2, \dots, y_t\}$.

3.1.2 DLMs with Normal Mixture Structure for Simulated Data

Mixture of normal distribution was proposed to accommodate the non-normality and asymmetric characteristic of financial time series. A mixture of two normal distribution is given by

$$f(x_t) = \pi f_1(\mu_1, \sigma_1^2) + (1 - \pi) f_2(\mu_2, \sigma_2^2) \quad (3.3)$$

where $f_1(\mu_1, \sigma_1^2)$ is the PDF of a normal distribution with mean μ_1 and variance σ_1^2 , $f_2(\mu_2, \sigma_2^2)$ is the PDF of a normal distribution with mean μ_2 and variance σ_2^2 . The weight is the probability of π when the first regime occurs while the second regime occurs with the probability $1 - \pi$.

Consider the model

$$y_t = \mu_t + \nu_t, \quad (3.4)$$

$$\mu_t = \phi \mu_{t-1} + \omega_t, \quad (3.5)$$

with $\nu_t \sim \pi N(0, \nu) + (1 - \pi) N(0, k^2 \nu)$ and $\omega_t \sim N(0, w)$, and they can be written as a conditionally Gaussian DLM given by $\{F_t, G_t, \nu \lambda_t, w\}$. In here a latent variable λ_t is introduced. It takes the value of 1 with the probability of π and the value of k^2 with the probability of $1 - \pi$, which also explains the meaning of normal mixture.

I simulated 200 points from the given model with $\phi = \pi = 0.9, k^2 = 4, \nu = 1,$

$\omega=1$. Then two different models were fitted to the simulated data. First I fit a model that ignores the mixture structure in the innovations at the observational level with $v_t \sim N(0, v)$ and assume that all parameters $\{\phi, v, w\}$ are given. Then I fit a model taking the normal mixtures on the observational error into consideration with $\{\phi, v, w\}$ unknown (π, k^2 are given). To make inference on the model, a simulation-based method such as Markov Chain Monte Carlo algorithm with emphasis on Forward Sampling Backward Filtering and Gibbs sampling will be applied for model fitting. The mean posterior distribution of the state parameter μ_t and its 95% posterior bands will be displayed in Figure 1 and Figure 3.

Forward Filtering Backward Sampling method

Assuming F_t, G_t, v_t, w_t are known, we have the following distribution at each time t :

Forward Filtering

- The prior for the state vector at time t $p(\theta_t | D_{t-1})$ is $N(\theta_t | a_t, R_t)$ with $a_t = G_t m_{t-1}$ and $R_t = G_t C_{t-1} G_t' + W_t$. D_{t-1} represents all the information available at time $t-1$, i.e., $D_{t-1} = \{y_{t-1}, y_{t-2}, \dots, y_0\}$
- One-step-ahead predictive distribution at time $t-1$ is given by $(y_t | D_{t-1}) \sim N(\theta_t | f_t, q_t)$ with $f_t = F_t' a_t$ and $q_t = F_t' R_t F_t + v_t$.
- Posterior distribution for θ_t given current information set D_t $p(\theta_t | D_t)$ is $N(\theta_t | m_t, C_t)$ with $m_t = a_t + A_t e_t$ and $C_t = R_t - A_t A_t' q_t$ where $e_t = y_t - F_t$ and $A_t = R_t F / q_t$.

Backward Sampling

- The distribution of the past state vector conditional on all future state vectors and information up to time T is $p(\theta_t | \theta_{t+1}, \dots, \theta_T, D_T) = p(\theta_t | \theta_{t+1}, D_t) \sim N(m_t^*, C_t^*)$ with $m_t^* = m_t + B_t(\theta_{t+1} - a_{t+1})$, $C_t^* = C_t - B_t R_{t+1} B_t'$ where $B_t = C_t G_{t+1}' R_{t+1}^{-1}$.

Posterior simulation with MCMC algorithm

Assume the model defined by $\{F_t, G_t, v_t, W_t\}$ depends on some latent variables λ_t . After calculating all the quantities about parameter θ_t ($t=1:T$) with the above equations, we can start to simulate from its posterior distribution. That is, to sample θ_t from the posterior distribution over and over again until the samplers converge. For each MCMC iteration i , we will get a sequence of $\theta_{0:T}$. The way to obtain $\theta_{0:T}^{(i)}$ conditional on $\lambda_{1:T}^{(i)}$ is given as follows:

- 1) Use the DLM filtering equations to compute m_t, a_t, C_t and R_t for $t=1:T$.
- 2) At time $t=T$ sample $\theta_T^{(i)}$ from $N(\theta_T | m_T, C_T)$.
- 3) For $t=(T-1):0$ sample $\theta_t^{(i)}$ from $N(\theta_t | m_t^*, C_t^*)$

MCMC Algorithm assuming unknown parameters

Assume ϕ, v, w are unknown, and the normal mixture structure relies on latent variables λ_t where $\Pr(\lambda_t = 1) = \pi$, $P(v_t | \lambda_t = 1) \sim N(0, v)$; $\Pr(\lambda_t = k^2) = 1 - \pi$, $P(v_t | \lambda_t = k^2) \sim N(0, k^2 v)$. For each MCMC iteration i , the simulation steps are as follows:

- Sample $(v | \omega, \phi, \mu_{0:200}, \lambda_{1:200}, y_{1:200})$. This reduced to sample v from an inverse-

gamma distribution ,i.e., $IG(\nu|\alpha_\nu, \beta_\nu)$ with $\alpha_\nu = \alpha_{0,\nu} + T/2$, $\beta_\nu = \beta_{0,\nu} + s_\nu^2/2$

$\alpha_{0,\nu}$ and $\beta_{0,\nu}$ are prior fixed values, $T = 200$, and s_ν^2 is given by

$$s_\nu^2 = \sum_{\{\lambda_t=1\}} (y_t - \mu_t)^2 + \sum_{\{\lambda_t=k^2\}} (y_t - \mu_t)^2 / k^2 .$$

- Sample $(\omega|\nu, \phi, \mu_{0:200}, \lambda_{1:200}, y_{1:200})$. ω is sampled from an inverse-gamma

distribution $IG(\omega|\alpha_\omega, \beta_\omega)$ with $\alpha_\omega = \alpha_{0,\omega} + T/2$, $\beta_\omega = \beta_{\omega,0} + \sum_{t=1}^T (\mu_t - \mu_{t-1})^2 / 2$

where $\alpha_{0,\omega}$ and $\beta_{\omega,0}$ are prior fixed values.

- Sample $(\phi|\nu, \omega, \mu_{0:200}, y_{1:200})$. ϕ is sampled from $N(\phi|m_\phi, C_\phi)$, with

$$m_\phi = \left(\sum_{t=1}^{200} \mu_t \mu_{t-1} \right) / \sum_{t=1}^{200} \mu_{t-1}^2 \text{ and } C_\phi = \omega / \sum_{t=1}^{200} \mu_{t-1}^2 .$$

- Sample $(\mu_{0:200}|\nu, \omega, \phi, \lambda_{1:200}, y_{1:200})$. A forward filtering backward sampling algorithm was used to obtain a sample of $\mu_{0:200}$.

- Sample $(\lambda_{1:200}|\nu, \omega, \phi, \mu_{0:200}, y_{1:200})$. At each time t , λ_t is sampled from a discrete distribution, i.e., λ_t is set to 1 or k^2 with probabilities defined in terms of ratio

$$\frac{\Pr(\lambda_t = 1|\nu, \omega, \mu_{0:200})}{\Pr(\lambda_t = k^2|\nu, \omega, \mu_{0:200})} = \frac{\pi}{1-\pi} k \times \exp\left\{-\frac{(y_t - \mu_t)^2 (1-k^{-2})}{2\nu}\right\}$$

3.1.3 Fitted AR(1) with Normal Mixture DLMs for Simulated Data

In Figure 1 and Figure 3, the points simulated from $N(0, \nu)$ are shown with circles

and the points simulated from $N(0, k^2\nu)$ with solid circles to distinguish.

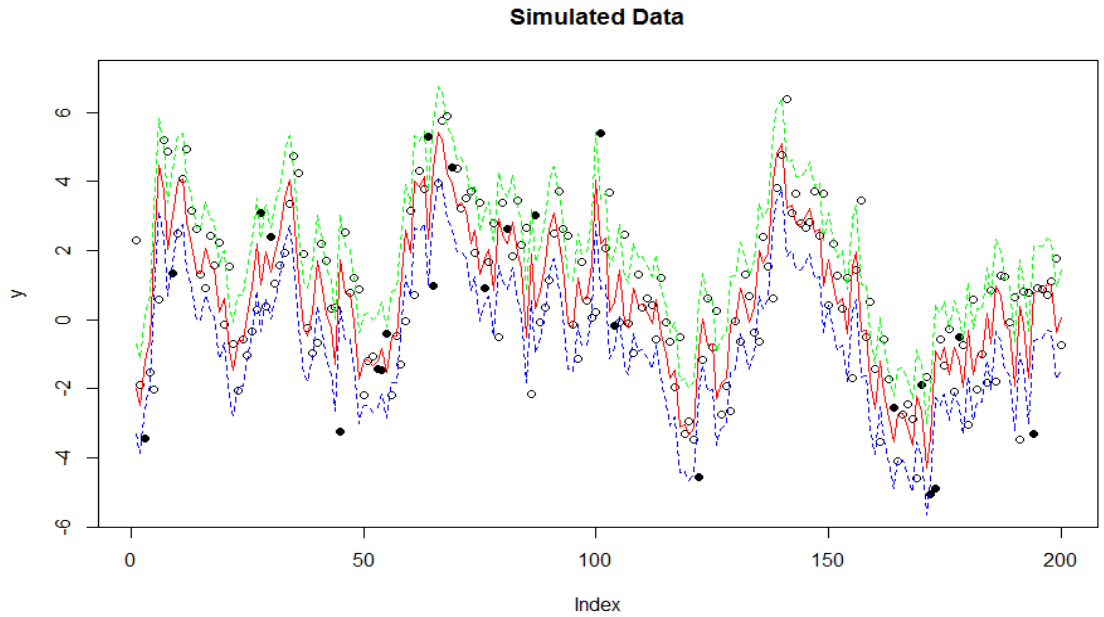


Figure1 : Simulated series from AR(1) with mixture observational errors(circles and solid circles). The red solid line corresponds to the posterior mean of the smoothing distribution over time $(\mu_t | D_{200})$ obtained from a model that ignores the mixture structure in the observational errors. The dotted lines are 95% posterior bands for $(\mu_t | D_{200})$.

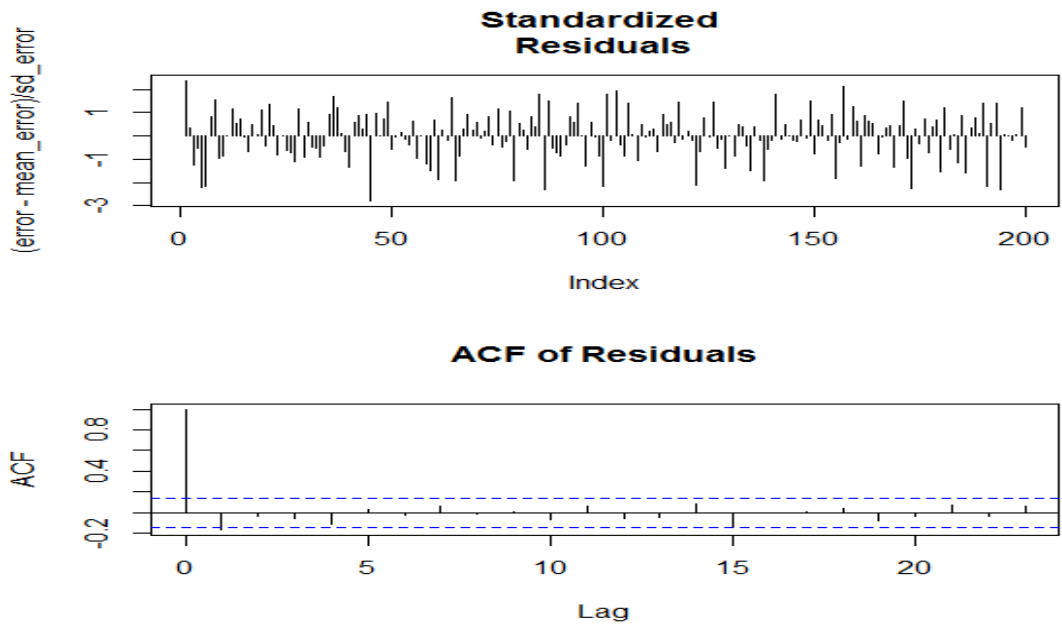


Figure2: Residual analysis obtained from fitting the model $\{1,0.9,1,1\}$ to the simulated data. The top and bottom plots display, respectively, the standardized residuals and the sample ACF of the residuals.

Simulated Data inferred with unknown parameters

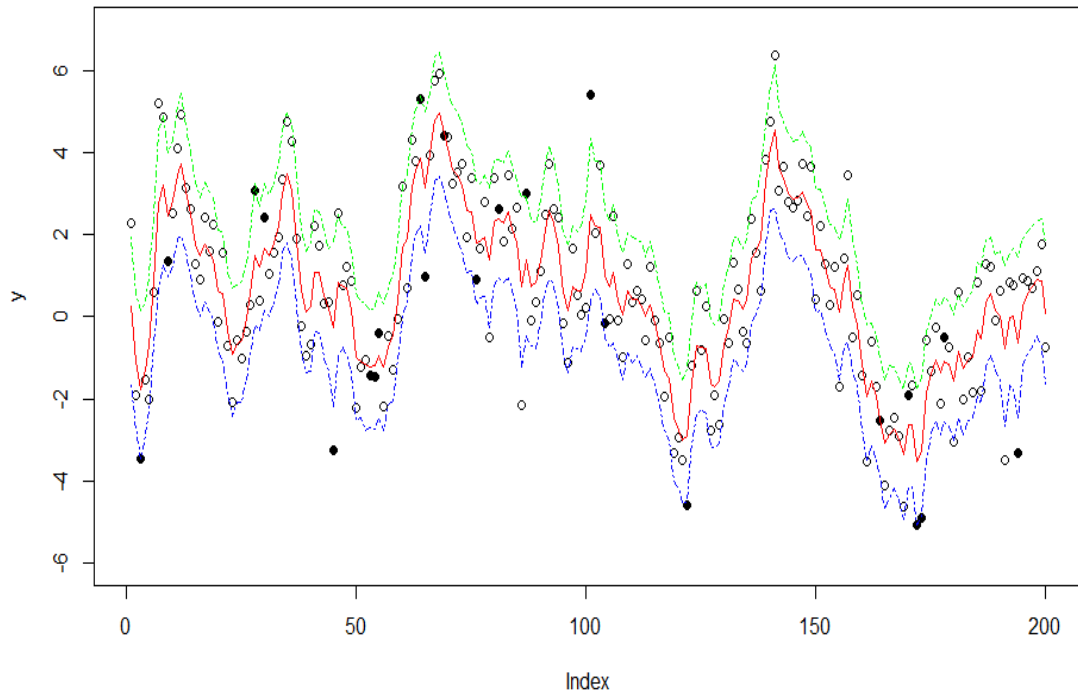


Figure3: Simulated series. Posterior mean and 95% posterior bands for $(\mu_t | D_{200})$ obtained from the model described by Equations (1) and (2).

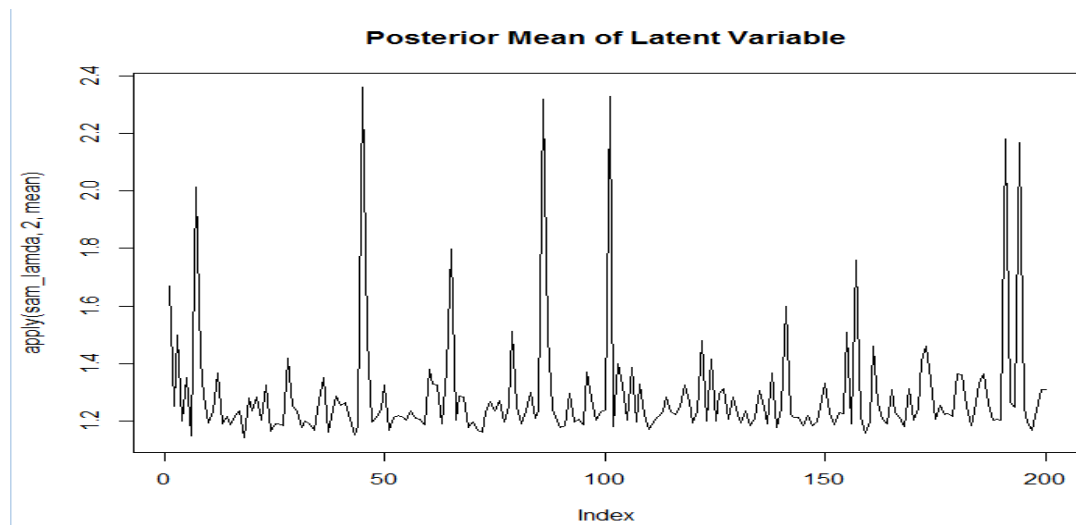


Figure4 :posterior mean of the latent process λ_t

By comparison of the plots displayed above, when the parameters are not given and normal mixtures are taken into consideration, the estimated mean and posterior interval using MCMC simulation is as well as the estimation when parameters are all

given ignoring normal mixture, which shows the efficiency of MCMC method in model fitting.

In addition, the posterior mean and 95% posterior interval for the parameters are given in the table below:

	ϕ	ν	ω
Posterior mean	0.882	1.302	1.025
True value	0.9	1.0	1.0
95% Posterior Interval	(0.789,0.958)	(0.875,1.865)	(0.582,1.640)

Table1: Estimated posterior mean and 95% posterior interval

We can see in the table that posterior estimation for the unknown parameters are very close to the true values with small errors, and each of the 95% posterior interval contains the true value.

3.2 Stochastic Volatility Model

3.2.1 AR(1) zero-mean Stochastic Volatility Model

We are modeling the stock market returns as a zero mean with time-varying variance process

$$\begin{aligned}
 r_t &\sim N(r|0, \sigma_t^2) \\
 \sigma_t &= \exp(\mu + x_t) \\
 x_t &= \phi x_{t-1} + \varepsilon_t
 \end{aligned}
 \tag{3.6}$$

where $\varepsilon_t \sim N(\varepsilon_t|0, \nu)$, r_t is the stock return at time t , x_t is the log volatility at time t which is assumed to follow a stationary process ($|\phi| < 1$). ε_t is uncorrelated white noise and $N(\cdot, \cdot)$ is the normal distribution.

The parameter μ defines the baseline log-volatility; the AR(1) parameter ϕ defines persistence in deviations in volatility from the baseline, and the innovation variance ν “drives” the levels of activity in the volatility process. Typically ϕ is close to one, and in any case the AR(1) process is assumed stationary, so that the marginal distribution for the missing initial value is $N(x_0 | 0, \nu / (1 - \phi^2))$.

Since we have non-normality inherent in the observation equation, a transformation is needed to linearize it. Let $y_t = \frac{1}{2} \log(\xi_t^2)$, so that

$$\begin{aligned} y_t &= \mu + x_t + v_t \\ v_t &= \log(\xi_t^2) / 2 = \log(\kappa_t) / 2, \\ \kappa_t &\sim \chi_1^2 \end{aligned} \tag{3.7}$$

Here x_t is assumed independent of the v_t . We approximate the density of v_t by a discrete mixture of normal components of the form $p(v_t) \approx \sum_{j=1}^J q_j N(v_t | b_j, w_j)$. Very good approximation can be obtained by choosing J as low as five (we choose seven in this thesis) and with approximate choices of the component weights, means and variances. This strategy converts the model into a conditionally Gaussian DLM. Introduce a latent indicator variable $r_t \in \{1: J\}$, at each time t the mixture of normal distribution can be constructed from the conditionals

$$\Pr(r_t = j) = q_j \quad \text{and} \quad (v_t | r_t = j) \sim N(v_t | b_j, w_j), \quad j = 1: J.$$

v_t are independent over time t .

When $J=7$, we have the means and variances given as follows:

q_j	0.00730	0.00002	0.10556	0.25750	0.34001	0.24566	0.04395
b_j	-5.7002	-4.9186	-2.6216	-1.1793	-0.3255	0.2624	0.7537
ω_j	0.7537	1.2949	0.6534	0.3157	0.1600	0.0851	0.0418

Table2: Discrete normal mixture approximation to log chi- square distribution.

Extending the analysis to include inference on the sequence of γ_t ($t = 1:T$) opens up the model fitting strategy of conditional Gaussian models, so that posterior inference can be performed via standard MCMC methods.

However, Gibbs sampling based MCMC applied to the original parameterization suffers from low convergence due to the inherent negative correlations always evident in conditional posteriors $(\mu | x_{0:T}, -)$ $(x_{0:T} | \mu, -)$ between μ and each of the x_t . This is resolved by developing the posterior simulation on the posterior for $z_{0:T}$ and μ where $z_t = x_t + \mu$. We re-express the model with the equivalent model representation

$$\begin{aligned}
 y_t &= z_t + v_t \\
 z_t &= \mu + \phi(z_{t-1} - \mu) + \varepsilon_t
 \end{aligned}
 \tag{3.8}$$

where z_t is the AR(1) volatility process centered around the baseline level μ .

3.2.2 MCMC Analysis

Inference for SV model is based on Markov Chain Monte Carlo methods, which is also known as the Metropolis-Hastings and Gibbs Sampling algorithm. These methods are widely used in the theory and practice of Bayesian inference. The idea behind MCMC methods is to produce variates from a given multivariate density (the

posterior density distribution) by repeatedly sampling a Markov chain whose invariant distribution is the target density of interest. Sample variates from a MCMC algorithm are a high-dimensional (correlated) sample from the target density of interest. Posterior moments and marginal densities can be estimated by averaging the relevant function of interest over the sampled variates. For example, the posterior mean of θ is simply estimated by the sample mean of the simulation θ values. And these estimates can be made arbitrarily accurate by increasing the simulation size.

In this thesis, I first applied the MCMC analysis to the model in which the weights, means and variances, of the normal mixture distribution are all fixed as shown in the table above. Then I included weight samplers in each MCMC iteration, that is to sample weights $q_{t:j}$ (denoted by $\eta_{t:j}$ instead) before sampling the latent variables $\gamma_{1:T}$ in each iteration. The priors on the parameters are defined by

$$\mu \sim N(g, G),$$

$$\phi \sim N(c, C)I(0 < \phi < 1)$$

$$v^{-1} \sim G(a/2, av_0/2).$$

Iterating through these steps provides the iterate of the overall MCMC without sampling weights.

1. Resample $\gamma_{1:T}$ from conditionally independent posteriors for each γ_t with

$$\Pr(\gamma_t = j | y_t, z_t) = q_{t,j}^* \quad , \text{where } q_{t,j}^* \propto q_j \exp\left\{-\frac{(y_t - b_j - z_t)^2}{2w_j}\right\} / w_j^{1/2} \quad \text{for}$$

$$j = 1: J \quad \text{and} \quad \sum_{j=1}^J q_{t,j}^* = 1.$$

2. Resample ϕ under the implied conditional posterior distribution

$$p(\phi|z_{0:T}, \mu, \nu) \propto a(\phi)p^*(\phi)I \quad (0 < \phi < 1)$$

where $p^*(\phi) = N(\phi|c^*, C^*)$ is the normal density given by

$$p^*(\phi) \propto N(\phi|c^*, C^*) \prod_{t=1}^T N(z_t | \mu + \phi(z_{t-1} - \mu), \nu)$$

$$a(\phi) \propto p(z_0 | \mu, \phi, \nu) = (1 - \phi^2)^{1/2} \exp(\phi^2(z_0 - \mu)^2 / (2\nu))$$

3. Resample μ from the implied conditional normal posterior proportional to

$$p(\mu)N(z_0 | \mu, \nu / (1 - \phi^2)) \prod_{t=1}^T N(z_t | \mu + \phi(z_{t-1} - \mu), \nu)$$

4. Resample ν from the implied conditional inverse-gamma posterior proportional to

$$p(\nu)N(z_0 | \mu, \nu / (1 - \phi^2)) \prod_{t=1}^T N(z_t | \mu + \phi(z_{t-1} - \mu), \nu).$$

5. Resample from the conditional posterior for the full volatility sequence $z_{0:T}$.

The modified observation equation now is

$$y_t = z_t + b_{\gamma_t} + v_t^*, v_t^* \sim N(v_t^* | 0, w_{\gamma_t})$$

$$z_t = \mu + \phi(z_{t-1} - \mu) + \varepsilon_t, \varepsilon_t \sim N(\varepsilon_t | 0, \nu)$$

Using forward filtering backward sampling (FFBS) algorithm:

- *Forward Filtering*: Sequentially compute and update the on-line posteriors

$$(z_t | y_{1:t}, \gamma_{1:t}, \mu, \phi, \nu) \sim N(z_t | m_t, M_t), \quad t = 1:T$$

where $m_0 = \mu$ and $M_0 = \nu / (1 - \phi^2)$.

- *Backward Sampling*: Sample from $N(z_T | m_T, M_T)$. Then for each $t = (T-1):0$ sample from the implied sequence of normal distribution

$$p(z_t | z_{t+1}, y_{1:t}, \gamma_{1:t}, \mu, \phi, \nu) \propto N(z_t | m_t, M_t) N(z_{t+1} | \mu + \phi(z_t - \mu), \nu).$$

Assume the (b_j, w_j) are appropriately fixed, we now treat the mixing normal probabilities as uncertain. A more general model is generated by

$$(y_t | z_t, \eta_{1:J}) \sim \sum_{j=1}^J \eta_j N(b_j + z_t, w_j),$$

where $\eta_{1:J}$ are also parameters to be estimated. The original $q_{1:J}$ given in Table 2 are referred to as a good first guess and used as prior means for a conjugate Dirichlet prior

$$\text{with } p(\eta_{1:J}) \propto \prod_{j=1}^J \eta_j^{\alpha q_j - 1} \quad (0 < \eta_j < 1 \text{ and } \sum_{j=1}^J \eta_j = 1).$$

We now need to resample η_j in each MCMC iteration for SV Model in order to incorporate this uncertainty about the normal mixture structure. In the following iteration steps, I add one more function to sample the weights, and in Gibbs sampling process weights will be sampled in each iteration like the other samplers.

1. Resample $\eta_{1:J}$ from the conditional posterior

$$p(\eta_{1:J} | \gamma_t) \sim \text{Dir}(\mathbb{C} + \alpha q),$$

where $\mathbb{C} = \{c_j | j=1, 2, \dots, J\}$ with c_j defined as the number of j 's in $\gamma_t (t=1:T)$, α is a known value, $q = \{q_j | j=1:J\}$. Here, the latent variable γ_t has a multinomial distribution with $1, 2, \dots, J$ as possible outcomes for each trial.

2. Resample $\gamma_{1:T}$ from conditionally independent posteriors for each γ_t with

$$\Pr(\gamma_t = j | y_t, z_t, \eta_{1:J}) = \eta_{t,j}^*$$

where $\eta_{t,j}^* \propto \eta_j \exp\{- (y_t - b_j - z_t)^2 / (2w_j)\} / w_j^{1/2}$ for $j = 1: J$ and $\sum_{j=1}^J \eta_{t,j}^* = 1$.

3. Resample ϕ under the implied conditional posterior

$$p(\phi | z_{0:T}, \mu, \nu) \propto a(\phi) p^*(\phi) I(0 < \phi < 1)$$

where $p^*(\phi) = N(\phi | c^*, C^*)$ is the normal density given by

$$p^*(\phi) \propto N(\phi | c, C) \prod_{t=1}^T N(z_t | \mu + \phi(z_{t-1} - \mu), \nu)$$

$$a(\phi) \propto p(z_0 | \mu, \phi, \nu) = (1 - \phi^2)^{1/2} \exp(\phi^2(z_0 - \mu)^2 / (2\nu))$$

4. Resample μ from the implied conditional normal posterior proportional to

$$p(\mu) N(z_0 | \mu, \nu / (1 - \phi^2)) \prod_{t=1}^T N(z_t | \mu + \phi(z_{t-1} - \mu), \nu)$$

5. Resample ν from the implied conditional inverse-gamma posterior proportional to

$$p(\nu) N(z_0 | \mu, \nu / (1 - \phi^2)) \prod_{t=1}^T N(z_t | \mu + \phi(z_{t-1} - \mu), \nu).$$

6. Resample from the conditional posterior for the full volatility sequence $z_{0:T}$.

The modified observation equation now is

$$y_t = z_t + b_{\gamma_t} + \nu_t^*, \nu_t^* \sim N(\nu_t^* | 0, w_{\gamma_t})$$

$$z_t = \mu + \phi(z_{t-1} - \mu) + \varepsilon_t, \varepsilon_t \sim N(\varepsilon_t | 0, \nu)$$

Using forward filtering backward sampling (FFBS) algorithm:

- *Forward Filtering*: Sequentially compute and update the on-line posteriors

$$(z_t | y_{1:t}, \gamma_{1:t}, \mu, \phi, \nu) \sim N(z_t | m_t, M_t), \quad t = 1:T$$

where $m_0 = \mu$ and $M_0 = \nu / (1 - \phi^2)$.

- *Backward Sampling:* Sample from $N(z_T | m_T, M_T)$. Then for each $t = (T - 1) : 0$ sample from the implied sequence of normal distribution

$$p(z_t | z_{t+1}, y_{1:t}, \gamma_{1:t}, \mu, \phi, \nu) \propto N(z_t | m_t, M_t) N(z_{t+1} | \mu + \phi(z_t - \mu), \nu) .$$

CHAPTER 4

FINDINGS

4.1 Data Description

Linear Technology Corporation (LLTC), a member of the S&P 500, designs manufactures and markets a broad line of high performance analog integrated circuits for major companies worldwide. The daily stock price data -open, high, low and close price, for NASDAQ-LLTC was provided by Google Finance.

1	Date	Open	High	Low	Close	Volume
2	3/25/2013	37.54	37.54	37	37.31	1327106
3	3/22/2013	37.15	37.5	36.79	37.39	1518421
4	3/21/2013	37.21	37.37	36.85	36.98	1608393
5	3/20/2013	37.58	37.61	37.26	37.51	1889351
6	3/19/2013	37.4	37.6	36.83	37.2	2811081
7	3/18/2013	37.21	37.59	37.16	37.24	2658738
8	3/15/2013	38.23	38.38	37.64	37.72	3201032
9	3/14/2013	38.46	38.72	38.38	38.42	1483144
10	3/13/2013	38.38	38.38	37.96	38.25	3281134
11	3/12/2013	38.32	38.36	38.02	38.23	2415751
12	3/11/2013	38.42	38.46	38.12	38.29	2975601
13	3/8/2013	38.46	38.75	38.21	38.42	1726564
14	3/7/2013	38.9	38.9	38.32	38.36	1886759
15	3/6/2013	38.64	38.71	38.37	38.48	2002674

Figure5 : A snapshot of the raw data

Descriptive plots for the daily stock price are given below. We can see that stock price is changing over time with an obvious pattern. It has time-varying means and time-varying variances. A large spike shows up. Very high autocorrelation exists within the series over a long period of time, and it decays extremely slowly.

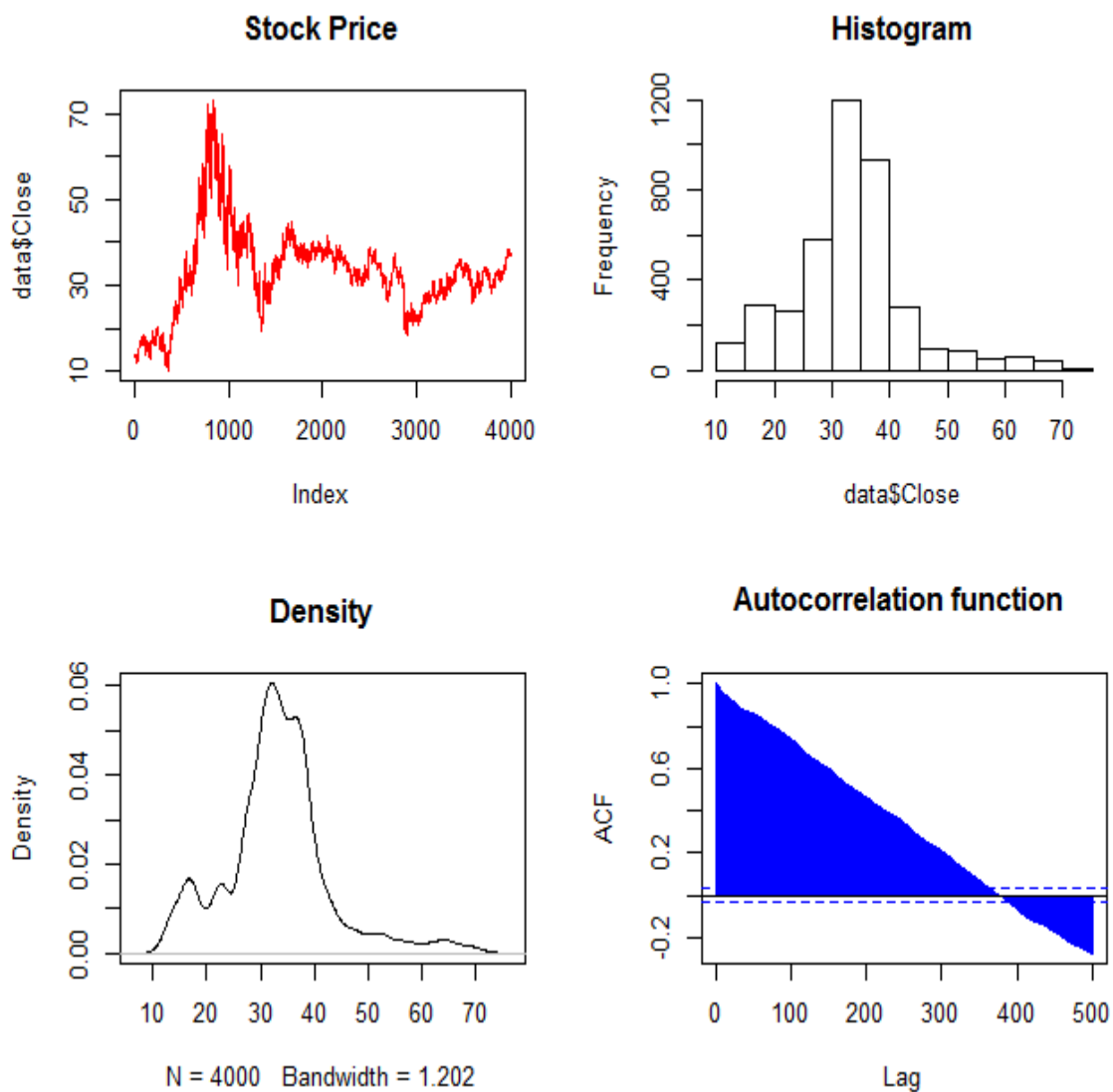


Figure6: Left Top and Right Top: time series and histogram for LLTC Daily Stock Price; Left Bottom and Right Bottom: Density and Autocorrelation Plot for Stock Price.

The only variable used in this thesis is close price. There are no missing values in the raw data. Stock market returns need to be calculated from stock close price before performing any analysis. In the raw data, we have 4000 observations, each of which represents the stock price on a business day. After obtaining stock returns data, there are 3999 observations, each of which represents the stock return on a specific trading day. The visualization of daily log returns is as follows.

I first plot the returns series with 3999 data points connected by lines, then a closer look at the data is taken by plotting the first 200 observations. As we can see from the plots below, stock returns are quite volatile and seem to have a constant mean around 0 over the long run. The variances of the series are changing over time, with some large spikes.

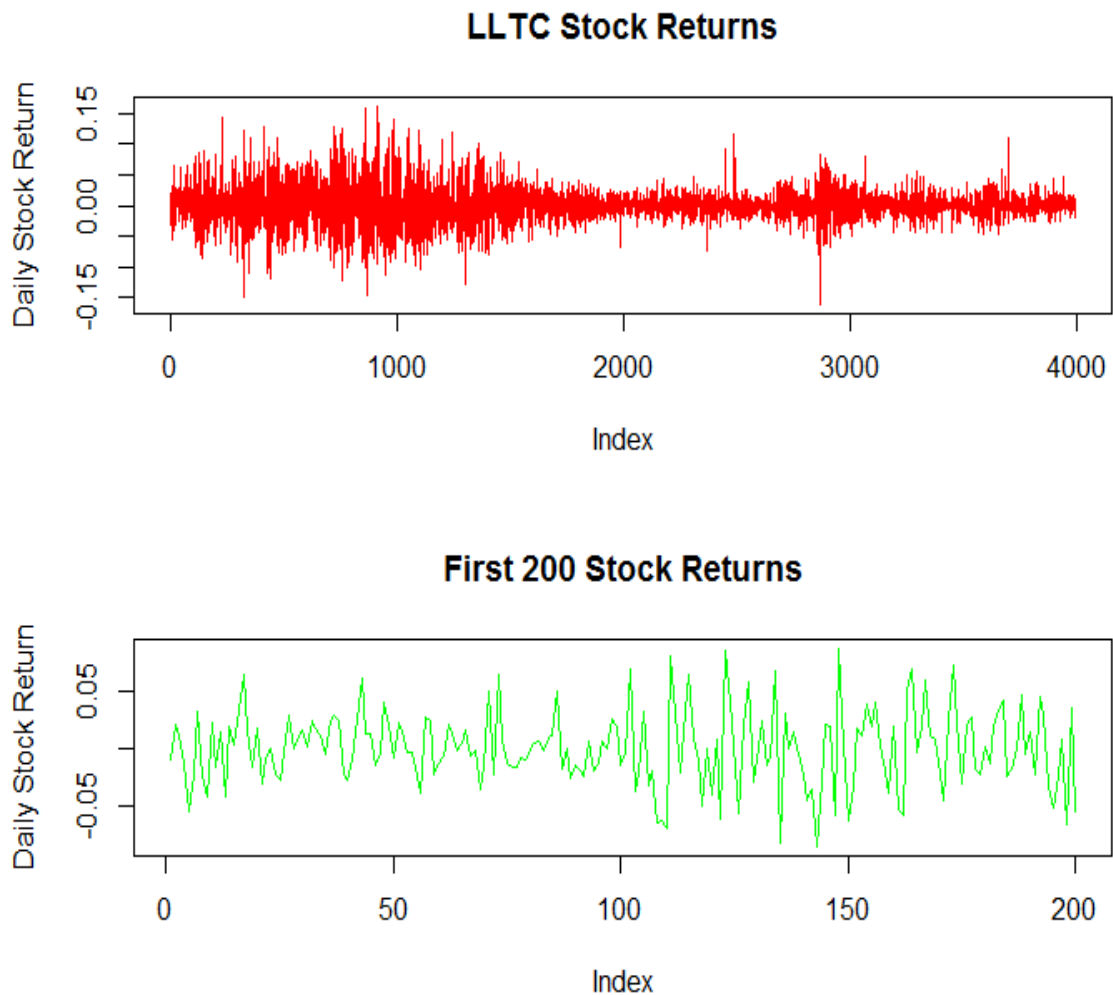


Figure7: Top:Time series plot for LLC stock market returns from 5/22/97 to 3/25/13. Bottom:A closer look at the return series from the first 200 observations.

Basic visualizations are given below. The data seems to be have a normal distribution but not exactly, as we can see there is a little peak on one tail from the density plot. In the QQ plot, the data displays a systematic departure from the straight

line at both ends. From the ACF plot, the returns have a much lower autocorrelation compared with stock price. By calculating log returns, we detrended the stock price series .

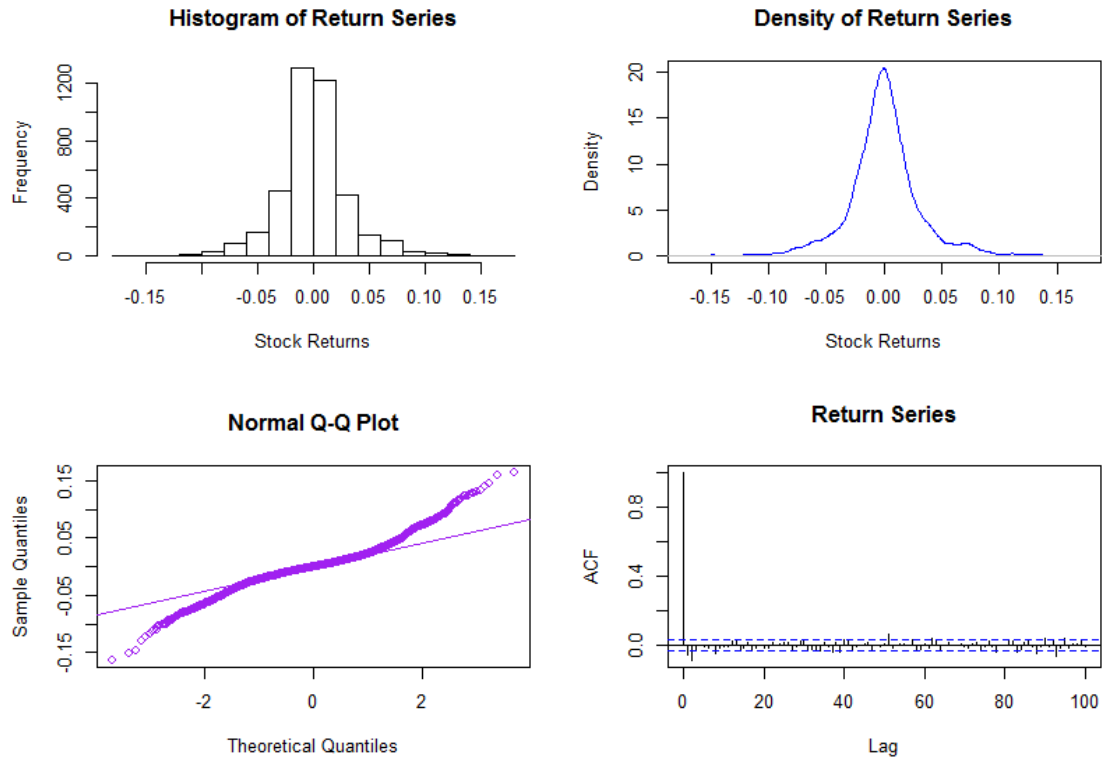


Figure8: Left Top and Left Bottom: Histogram and QQ plot for LLTC Daily Returns; Right Top and Right Bottom: Density and Correlogram for LLTC Returns

4.2 AR(8) Model for NASDAQ:LLTC Stock Market Returns

4.2.1 Structure of Autoregressive Model AR(P)

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t \quad (4.1)$$

$$\varepsilon_t \sim N(0, \sigma^2)$$

ϕ_j s are constant parameters.

4.2.2 Order Selection

From the autocorrelation plot given above on the right bottom, spikes showed up at Lag 1,2 and 8. So the order chosen for the autoregressive model is 8.

The model to be fitted has the form as follows :

$$\begin{aligned} y_t &= \beta_0 - \beta_1 y_{t-1} - \beta_2 y_{t-2} - \beta_8 y_{t-8} + \varepsilon_t \\ \varepsilon_t &\sim N(0, \sigma^2) \end{aligned} \quad (4.2)$$

4.2.3 Monte Carlo Approximation

Since AR(8) is a multiple linear regression model with the error normally distributed, we can apply the Bayesian method to estimate unknown parameters. The parameters of interest are $\{\beta, \sigma^2\}$. Our goal is to simulate from the joint posterior distribution $p(\beta, \sigma^2 | y, X)$ by the Monte Carlo method. We assume a prior distribution for β is multivariate normal distribution with mean β_0 and variance-covariance Σ_0 . In here a popular specification “g-prior” was used with $\beta_0 = 0$ and $\Sigma_0 = g\sigma^2(X^T X)^{-1}$. the posterior distributions for the two unknown quantities are calculated by

$$\begin{aligned} p(\sigma^2 | y, X) &\sim IG((\nu_0 + n) / 2, (\nu_0 \sigma_0^2 + SSR_g) / 2), \\ p(\beta | \sigma^2, y, X) &\sim MVN\left(-\frac{g}{g+1} \hat{\beta}_{ols}, \frac{g}{g+1} \sigma^2 (X^T X)^{-1}\right), \end{aligned}$$

where $SSR_g = y^T (I - \frac{g}{g+1} X(X^T X)^{-1} X^T) y$ and $\beta = (X^T X)^{-1} X^T y$.

A sample value of (β, σ^2) from $p(\beta, \sigma^2 | y, X)$ can be made as follows:

1. Sample $\sigma^2 \sim IG((\nu_0 + n) / 2, (\nu_0 \sigma_0^2 + SSR_g) / 2)$,
2. Sample $\beta \sim MVN\left(-\frac{g}{g+1} \hat{\beta}_{ols}, \frac{g}{g+1} \sigma^2 (X^T X)^{-1}\right)$

In this way, we can generate multiple independent Monte Carlo samples from the posterior distribution.

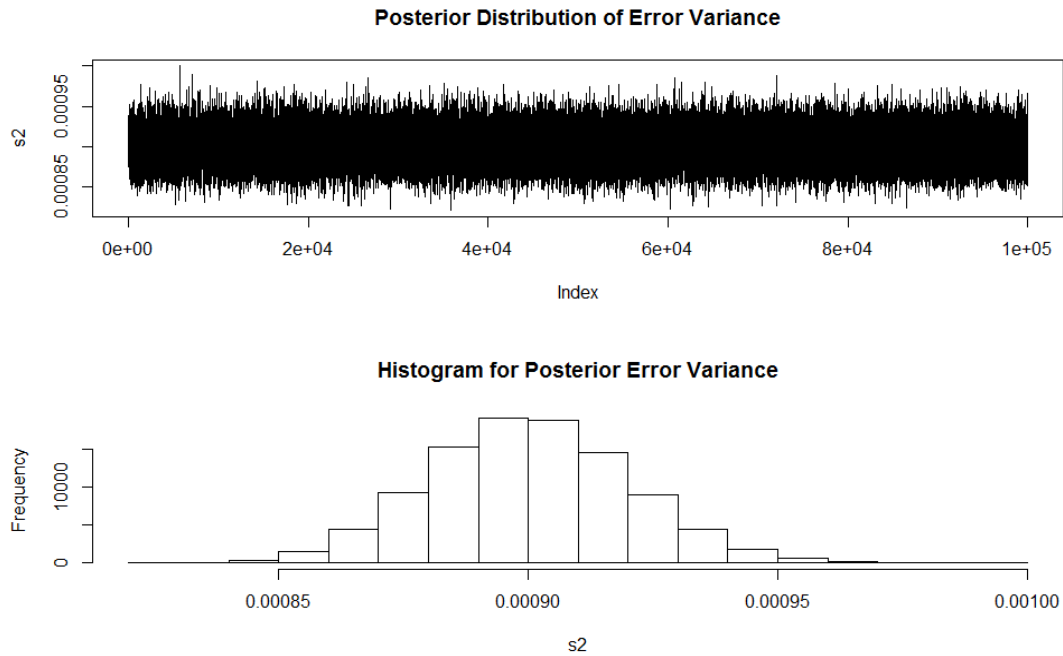
4.2.4 Estimated AR(8) model for NASDAQ:LLTC Stock Market Returns

$$y_t = 0.000324 - 0.057734y_{t-1} - 0.089153y_{t-2} - 0.046243y_{t-8} + \varepsilon_t$$

$$\varepsilon_t \sim N(0, 0.0009^2) \quad (4.3)$$

4.2.5 Error Variance Plots

The estimated error variance σ^2 specified above is the mean value of all error variances drawn from its posterior distribution. We can see below that the variance of the error seems to be normally distributed with constant mean.



*Figure9 :Top Graph : Simulated posterior variance for innovation terms;
Bottom Graph: Histogram for simulated posterior error variances*

4.3 AR(1) SV Models for NASDAQ:LLTC Stock Returns

4.3.1 Fitted AR(1) SV Models with fixed weights for Normal Mixtures

I ran the MCMC simulation for 20000 times and discarded the first 10000 iterations. The estimated parameters for ϕ, μ, ν are

$$\hat{\phi} = 0.9931$$

$$\hat{\mu} = -3.2050$$

$$\hat{\nu} = 0.0301 \quad (4.4)$$

The fitted model can be written as

$$\begin{aligned}
 y_t &= z_t + \nu_t \\
 z_t &= -3.2050 + 0.9931(z_{t-1} + 3.2050) + \xi_t, \\
 \xi_t &\sim N(0, \sqrt{0.0301})
 \end{aligned} \tag{4.5}$$

where $y_t = 0.5 \log(r_t^2 + 0.001)$, r_t represents NASDAQ-LLTC daily stock market returns, 0.01 is the offset part, $\nu_t = 0.5 \log(\varepsilon_t^2)$, $\varepsilon_t \sim N(0,1)$. We approximate ν_t by the mixture of seven normal distributions with fixed weights given in Table 1.

4.3.2 Fitted AR(1) SV Models with Sampled Weights for Normal Mixtures

I ran the Gibbs Sampling 50000 times and discarded the first 20000 iterations. The burn in period was intended to ensure that the effect of the initial values becomes insignificant, which means that the different starting values that could've been set don't actually provide any additional information or play a role in the parameter estimation when running multiple Markov Chains. The estimated results for the parameters are

$$\begin{aligned}
 \hat{\phi} &= 0.9862 \\
 \hat{\mu} &= -2.9368 \\
 \hat{\nu} &= 0.0262
 \end{aligned} \tag{4.6}$$

Hence, the fitted AR(1) SV models can be expressed as

$$\begin{aligned}
 y_t &= z_t + \nu_t \\
 z_t &= -2.9368 + 0.9862(z_{t-1} + 2.9368) + \xi_t, \\
 \xi_t &\sim N(0, \sqrt{0.0262})
 \end{aligned} \tag{4.7}$$

Applying the exponential function to the sampled z_t ($t = 0:T$), we get the stock market returns volatility, namely $\sigma_t = \exp(z_t)$.

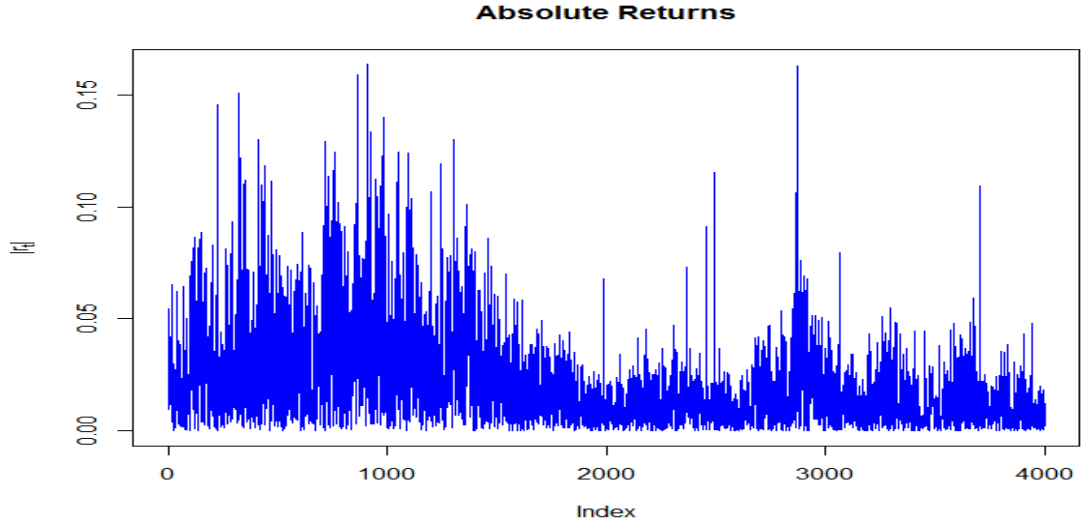


Figure 10: The absolute LLTC Stock Market Returns $|r_t|$.

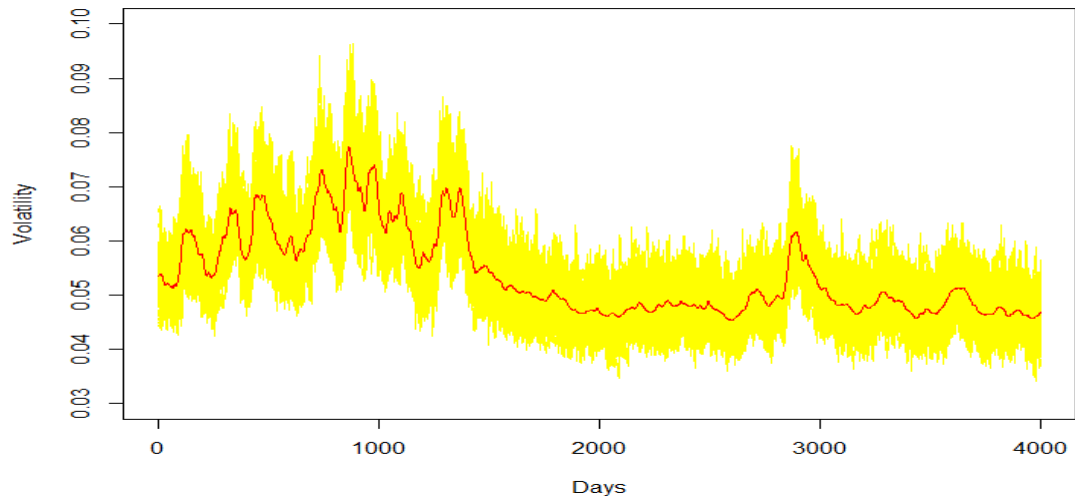
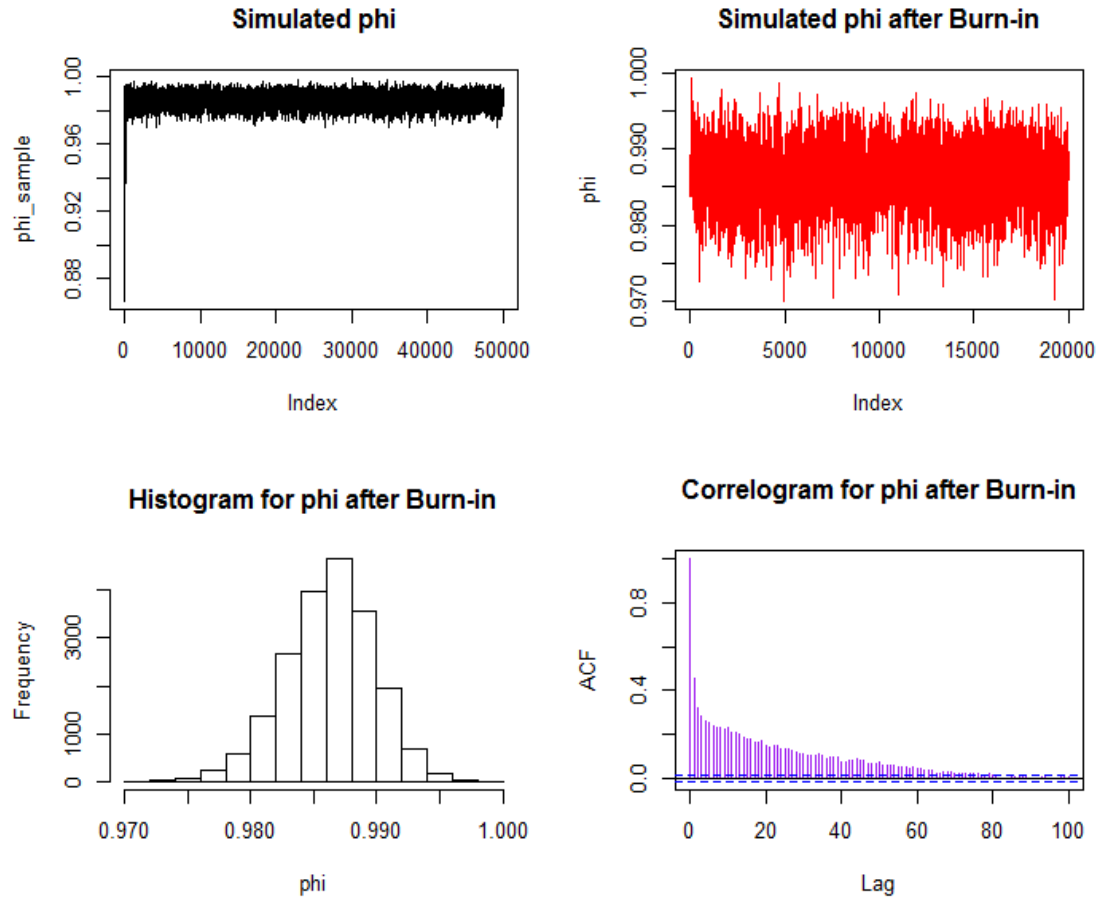


Figure 11: Estimated volatility process σ_t in the standard univariate SV model. The full line indicates the posterior mean of $\exp(z_t)$, plotted over days=1:T, from the MCMC analysis; the yellow shading is 50 similar time plots representing 50 randomly selected trajectories from the posterior.

In each of the following three figures, four plots--complete iterations, iterations after burn in period, histogram, correlogram, are displayed to keep track of the single move Gibbs samplers for parameters ϕ, μ and ν . The correlogram (autocorrelation function) indicates important autocorrelations for ϕ, μ and ν at large lag lengths.



*Figure12: Top : Complete traceplot and traceplot after burn in period for ϕ
 Bottom: Histogram and correlogram for sampled ϕ after burn in period.*

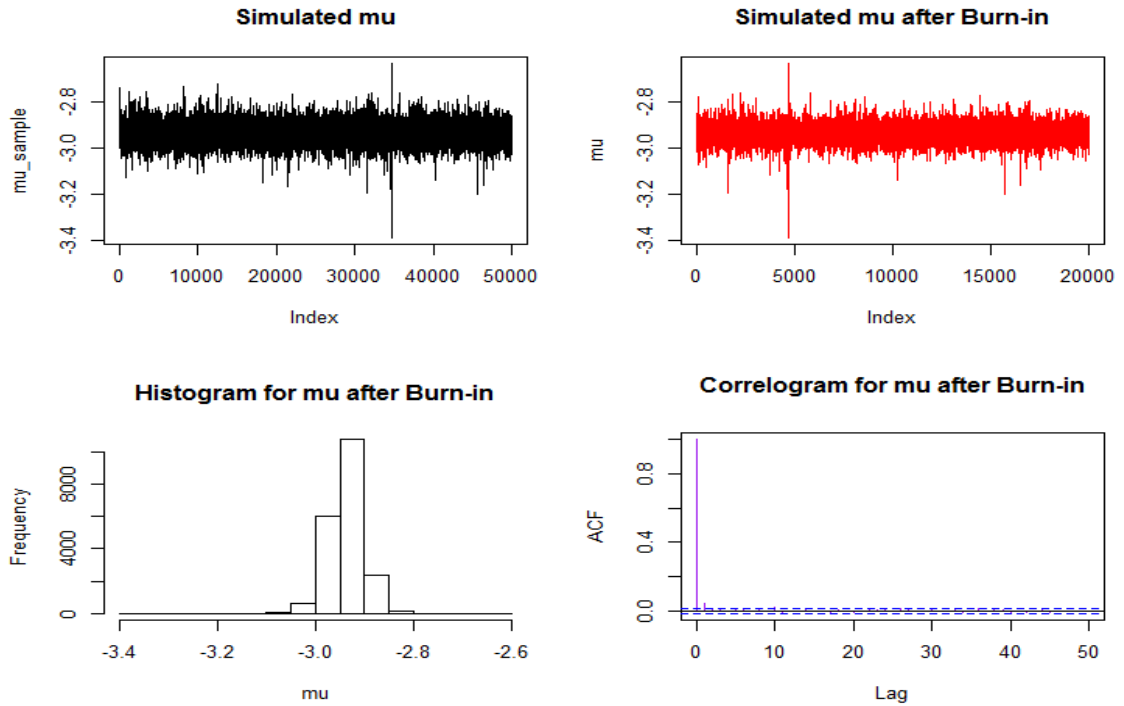


Figure13: Top : Complete traceplot and traceplot after burn in period for μ
 Bottom: Histogram and correlogram for sampled μ after burn in period.

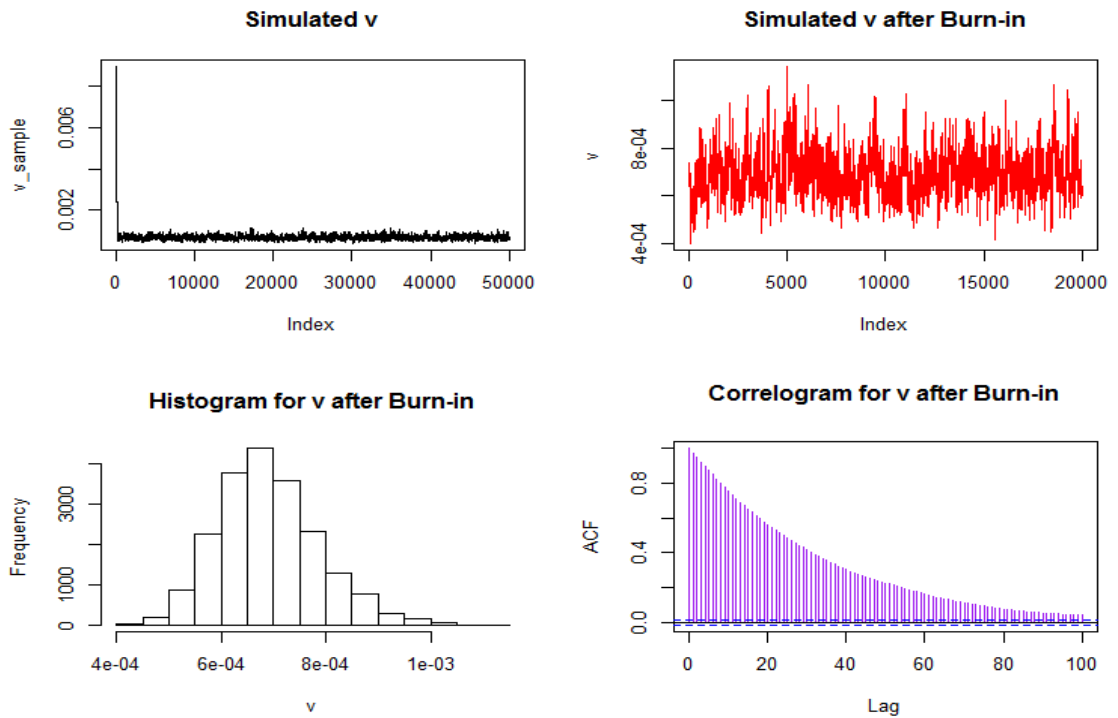


Figure14: Top : Complete traceplot and traceplot after burn in period for v
 Bottom: Histogram and correlogram for sampled v after burn in period.

4.3.3 NASDAQ:LLTC One-step-ahead Stock Market Return Forecast

4.3.3.1 Definitions

Predictive Distribution: When it comes to forecasting, the event of interest is to predict the value of a future observation y_{t+h} , ($h=1,2,3\dots$), given the data y , $y=(y_1, y_2, \dots, y_t)$. This can be solved by computing conditional density $y_{t+h}|y$. Assume y_1, y_2, \dots, y_t are conditionally independent given parameter θ . (θ can be a random vector, finite or infinite dimensional). When $h=1$,

$$\begin{aligned} p(y_{t+1}|y_1, y_2, \dots, y_t) &= \int f(y_{t+1}, \theta|y_1, y_2, \dots, y_t) d\nu(\theta) \\ &= \int f(y_{t+1}|\theta, y_1, y_2, \dots, y_t) \pi(\theta|y_1, y_2, \dots, y_t) d\nu(\theta) \quad . \\ &= \int f(y_{t+1}|\theta) \pi(\theta|y_1, y_2, \dots, y_t) d\nu(\theta) \end{aligned}$$

where $\pi(\theta|y_1, y_2, \dots, y_t)$ is the posterior distribution of θ given (y_1, y_2, \dots, y_t) . Note that

$$\pi(\theta|y_1, y_2, \dots, y_t) \propto \pi(\theta|y_1, y_2, \dots, y_{t-1}) f(y_t|\theta) \propto \prod_{i=1}^{t-1} f(y_i|\theta) f(y_t|\theta) = \prod_{i=1}^t f(y_i|\theta)$$

based on Bayes' Rule, so $\pi(\theta|y_1, y_2, \dots, y_t)$ can be computed recursively through

$\pi(\theta|y_1, y_2, \dots, y_{t-1}) \cdot p(y_{t+1}|y_1, y_2, \dots, y_t)$ is also called One-step-ahead prediction.

95% Posterior Interval: In Bayesian inference, the 95% posterior interval for a parameter, is estimated by the samples drawn from its conditional posterior distribution, which means that the probability that the posterior interval will contain the true value is 95%. This is a Bayesian analogue of confidence interval in frequentist statistics.

4.3.3.2 One-step-ahead Prediction Results

The one-step-ahead log transformed volatility for stock returns is defined

as $z_{T+1} = \log(\sigma_{T+1})$. After transformation, original volatility $\sigma_{T+1} = \exp(z_{T+1})$. Using the parameterized model (3.7), when $t = T + 1$, we can get z_{T+1} by calculating $\mu + \phi(z_T - \mu)$ in the first equation. Below is the estimated distribution for log volatility and volatility at $t = T + 1$. They have the same shape because the exponential function is monotone increasing.

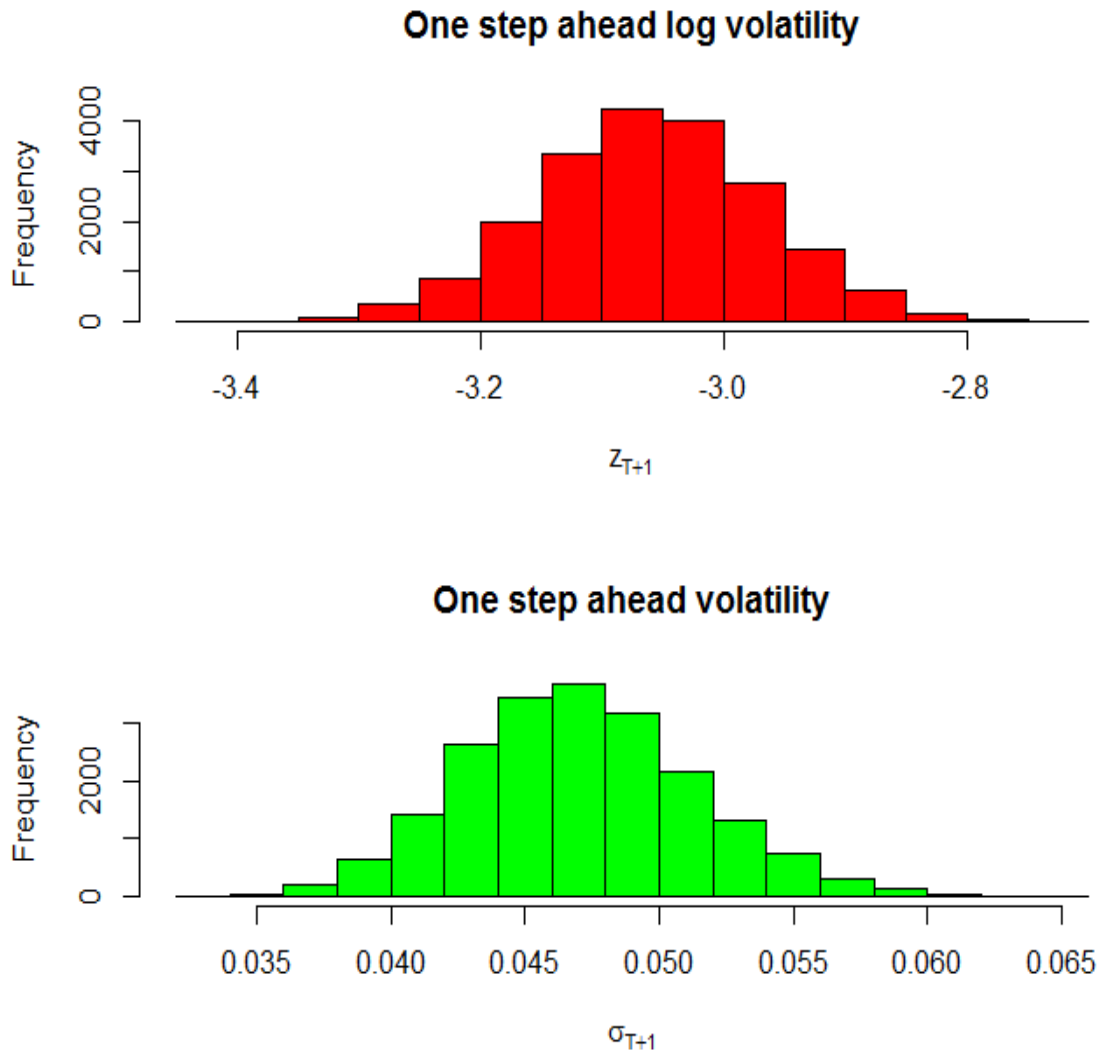


Figure15: One step ahead volatility prediction histograms

The mean and 95% posterior interval for z_{T+1} and σ_{T+1} :

Posterior Estimation	z_{T+1}	σ_{T+1}
Mean	-3.0611	0.0470
95% PI	(-3.2471,-2.8807)	(0.0388,0.0560)

Table3 : Mean and 95% posterior interval for One-step-ahead volatility prediction

Given predicted z_{T+1} , we can predict the log transformed stock returns by

$$y_{T+1} = z_{T+1} + v_{T+1}.$$

Predicted one-step-ahead log transformed stock return

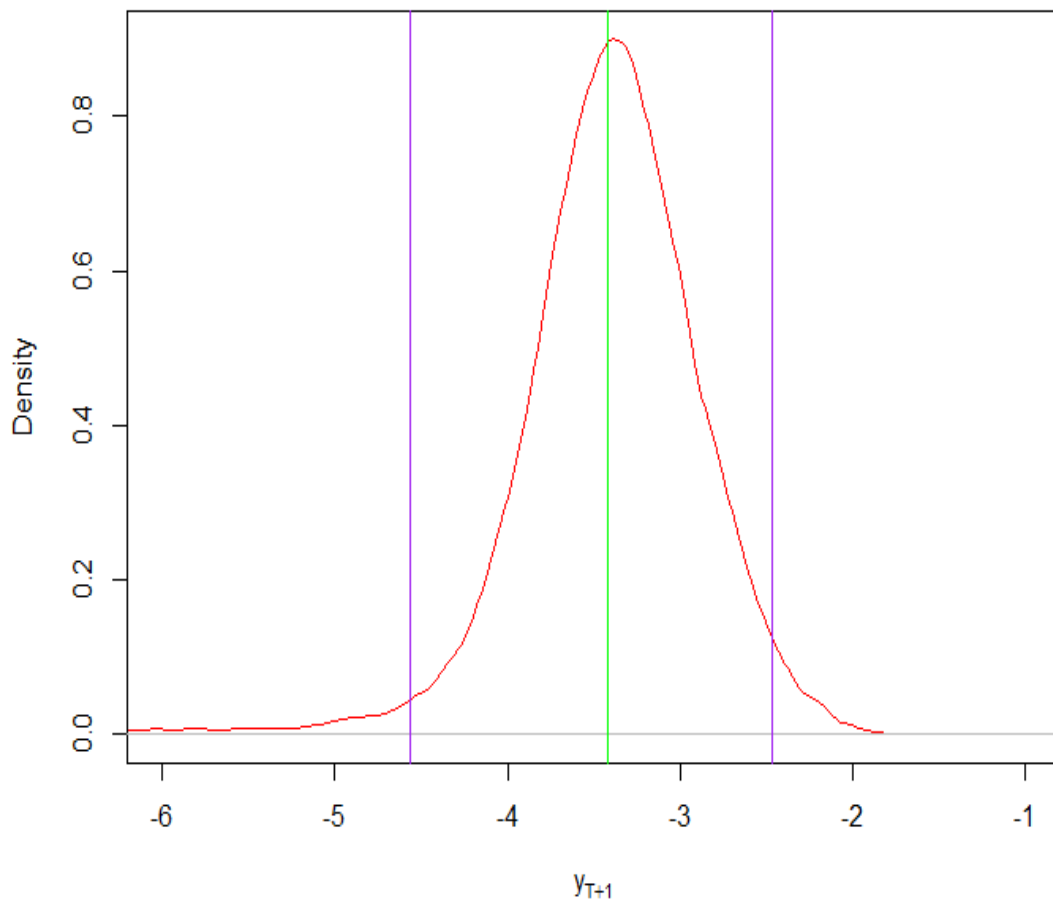


Figure16 : One step ahead prediction distribution for y_{T+1} , its mean (green line) and 95% posterior interval (purple lines).

The posterior mean of one-step-ahead prediction for log transformed stock market return y_{T+1} is -3.4151 (compared with the previous observation $y_T = -3.451589$) and the 95% posterior interval is (-4.5613, -2.4645), as shown above in the plot.

Predictive distribution for log transformed stock returns

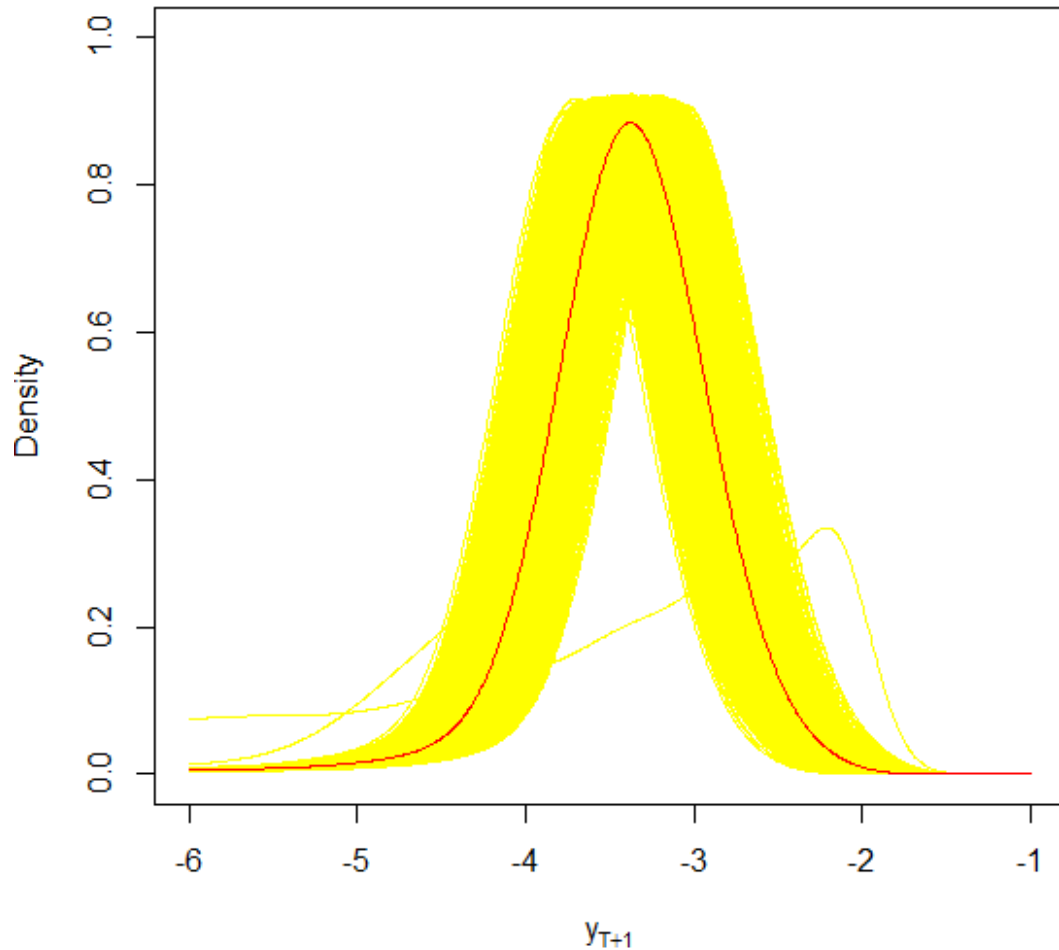


Figure 17: Red curve is the posterior mean of the predictive distribution for log transformed stock market returns; Yellow curves represent the simulated predictive distribution after burn in.

4.3.4 Parameter Estimation Summary

MCMC analysis is fully applied to our data when building the SV model. Two specified models are given above with slight differences in parameter estimation and

the estimation tends to be more accurate when having weights sampled for the normal mixture approximation structure. The table below lists the estimated values for each parameter under both conditions.

Normal Mixture Approximation	Estimated ϕ	Estimated μ	Estimated ν
Fixed Weights	0.9931	-3.2050	0.0301
Sampled Weights	0.9862	-2.9368	0.0262

Table 4 : Comparison of fitted results for SV models when having weights fixed and sampled.

CHAPTER 5

CONCLUSION

To better understand how to fit stochastic volatility models, which are non-linear and non-Gaussian, for NASDAQ-LLTC daily stock market returns, we need to get familiar with Dynamic Linear Models. Normal Dynamic Linear Model with emphasis on Forward Filtering Backward Sampling is applied to simulated data for parameter estimation and is shown to be an efficient simulation-based method with small errors. For the real data, we first fit a simple autoregressive model with order eight. Using this AR model, we can make a prediction on the future stock returns given the returns for the past eight business days. In stochastic volatility models, stock returns are assumed to be normally distributed with mean zero and time-varying variance or volatility. What we are aiming to do is to model this volatility and volatility is assumed to have a non-linear form. In this thesis, the log transformation with offset term is used to linearize the model and avoid meaningless definitions. Besides, normal mixture approximation to log chi-square distribution is fully applied with our data.

Model fitting and forecasting are realized with the application of a well developed Markov Chain Monte Carlo algorithm. In the MCMC context, Gibbs Sampling is applied to SV models to sample multiple unknown parameters from their posterior distribution and it also allows us to sample the log volatilities $z_{0,T}$ all at once for each iteration i . The goal of the MCMC method is to sample quantities from their joint posterior distribution. To simplify this process, Bayes' theorem is fully applied

which makes the sequential conditional posterior sampling plausible and effective. One step ahead prediction of log volatility and log transformed stock market returns are both talked about and displayed in the graphs. The mean and 95% posterior interval are also given as part of the prediction result. Two model fitting results are given in the context and the differences arise from whether the weights for the normal mixture structure are sampled or not. We make a prediction using the model estimated from MCMC procedure with sampled weights.

BIBLIOGRAPHY

- Barnett, G., R. Kohn, and S. Sheather. "Bayesian Estimation of an autoregressive model using Markov chain Monte Carlo". *Journal of Economics* 74, 237-254. (1996)
- Berzuini, C., N. Best, W. R. Gilks, and C. Larizza. "Dynamic conditional independence models and Markov chain Monte Carlo methods". *Journal of the American Statistical Association* 92, 1403-1412. (1997)
- Bollerslev T.. "Generalized Autoregressive Conditional Heteroskedasticity", *Journal of Econometrics* 31, 307-327. (1986)
- Box, G. E. P., G. M. Jenkins, and G. C. Reinsel. *Time series Analysis Forecasting and Control* (Fourth ed.). New Jersey: John Wiley & Sons. (2008)
- Box, G. E. P. and G. C. Tiao. *Bayesian Inference in statistical Analysis*. Massachusetts: Addison-Wesley. (1973)
- Black F, Scholes M, "The Pricing of Options and Corporate Liability", *The Journal of Political Economy*, Volume 81, Issue 3, 637-654. (May-Jun., 1973)
- Brockwell, P. J. and R.A. Davis, "Introduction to Time Series and Forecasting (Second ed.)," New York: Springer-Verlag. (2002)
- Carlin, B. P., N. G. Polson, and D.S. Stoffer. "A Monte Carlo approach to nonnormal nonlinear state-space modeling". *Journal of the American Statistical Association* 87, 493-500. (1992)
- Carter, C.K. and R. Kohn. Gibbs sampling for state space models. *Biometrika* 81, 541-553. (1994)
- Chatfield, C.. *The Analysis of Time Series: An Introduction* (Fifth ed.). London:

- Chapman and Hall. (1996)
- Chib, S., F. Nadari, and N. Shephard, “Markov chain Monte carlo methods for stochastic volatility models”, *Journal of Econometrics* 108, 281-316.
- Del Moral, P., A. Jasra, and A. Doucet. Sequential Monte Carlo for Bayesian Computation (with discussion). In *Bayesian Statistics 8*. Oxford:Oxford University Press. (2007)
- Durbin, J.. “Estimation of parameters in time series regression models”, *Journal of the Royal Statistical Society, Series B* 22, 139-153.(1960)
- Escobar, M. and West M.. “Bayesian density estimation and inference using mixtures”, *Journal of the American Statistical Association* 90,577-588. (1995)
- Franses H. P., “Time series models for business and economic forecasting” , Cambridge University Press. (1998)
- Hoff D. P., “A First Course in Bayesian Statistical Methods”, Springer. (2009)
- Kim S., N. Shephard, S. Chib, “Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models”, *The Review of Economic Studies*, Vol.65, No.3, Jul. 1998, pp. 361-393.(1998)
- Pesaran M.H., Timmermann A.. Predictability of Stock Returns: Robustness and Economic Significance. *The Journal of Finance*, Vol. 50, No. 4. pp. 1201-1228 (Sep., 1995)
- Petris G., Petrone S., Campagnoli P., “Dynamic Linear Models with R”, Springer, Aug. 2007.
- Prado R., West M., “Time Series Modeling, Computation, and Inference”, CRC Press. (2010)

- Rachev T. S., Hsu S.J. J., Bagasheva S. B., Fabozzi J. F., "Bayesian Methods in Finance", John Wiley & Sons, Inc.(2008)
- Raftery, A.E. and S. M. Lewis. "How many iterations in the Gibbs sampler?" In J. M. Reinsel, G. C. "Elements of Multivariate Time Series Analysis (2nd ed)," New York: Springer-Verlag. (1992)
- Robert, C. P. and G. Casella, "Monte Carlo Statistical Methods (Second ed.)," New York: Springer-Verlag.(2005)
- Shephard, N. (Ed.), "Stochastic Volatility :Selected Readings," Oxford University Press.(2005)
- West M., Harrison J., "Bayesian Forecasting and Dynamic Models," Springer. (1997)
- West, M. and P. J. Harrison, "Bayesian Forecasting and Dynamic Models (2nd ed.)," New York: Springer- Verlag.(1997)
- West, M., P. J. Harrison, and H. S. Migon, "Dynamic generalized linear models and Bayesian forecasting (with discussion)," *Journal of the American Statistical Association* 80, 73-87.(1985)
- West, M., "Approximating posterior distribution by mixtures," *Journal of the Royal Statistical Society, Series B* 55, 409-422.(1993a)