Real Time Estimation of Multivariate Stochastic Volatility Models

Jian Wang

Supervisor: Kostas Triantafyllopoulos
Advisor: Tim Heaton

Submitted to the Department of Probability and Statistics
in partial fulfillment of the requirements for
the degree of Doctor of Philosophy
at the University of Sheffield

Registration: October 2012

Abstract

Abstract of Real time estimation of multivariate stochastic volatility models

Author: Jian Wang

Date: September 2016

Keywords: multivariate volatility, sequential Monte Carlo, Bayesian inference, Wishart autoregressive process, Newton-Rapshon method

This thesis firstly considers a modelling framework for multivariate volatility in financial time series. As most financial returns exhibit heavy tails and skewness, we are considering a model for the returns based on the skew-t distribution, while the volatility is assumed to follow a Wishart autoregressive process. We define a new type of Wishart autoregressive process and highlight some of its properties and some of its advantages. Particle filter based inference for this model is discussed and a novel approach of estimating static parameters is provided. Furthermore, an alternative methodology for estimating higher dimension data is developed.

Secondly, inspired from the idea of Ulig's Wishart process, a new Wishart-Newton model is developed. The approach combines conjugate Bayesian inference while the hyperparameters are estimated by a Newton-Raphson method and here an online volatility estimate algorithm is proposed.

The two proposed models are compared with the benchmarking GO-GARCH model in both function execution time and cumulative returns of

different dimensional datasets.

Acknowledgements

I would like to take this opportunity to thank my supervisor, Dr Kostas Triantafyllopoulos, for devoting both his time and effort into guiding me through this project. The thesis has benefitted by discussion with Dr Farhat Iqbal. I would also like to thank all my lecturers, the secretarial staff and my fellow students for what has been a thoroughly enjoyable journey, both academically and socially. Finally, I would like to thank my parents, relatives and Yijia Zuo for their continual support and love.

Contents

List of Tables

List of Figures

List of Algorithms

Chapter 1

Introduction

1.1 Financial time series and volatility

Financial time series analysis aims to evaluate assets over time using statistical theory and practice. It has drawn huge attention recently. In the fields of, for example, stochastic volatility and high frequency finance, financial time series analysis has grown fast. Besides the mathematical analysis, it is also believed to be a very empirical subject, meaning that there are stylised facts which are assumed based on experience and long-term observation. For example, the heteroskedasticity and non-linearity of financial time series data.

As an important factor of modern financial time series analysis, volatility has been used to describe the variation or fluctuation of financial instrument. Volatility generally has a crucial contribution in portfolio construc-

tion and asset management. For example high volatility is usually associated to increased uncertainty. Sometimes this can be exploited to construct market-neutral arbitrage opportunities, such as short selling and prices trading. Nowadays it is also possible for investors to trade volatility directly by using derivative securities, for example options and variance swaps.

In the famous Black-Scholes option price formula Black and Scholes (1973), volatility is one of the parameters in the function, however it is treated as a constant due to the fact that it is not directly observable and hard to evaluate. With further research, people now begin to believe that volatility is not constant, but a parameter changing over time Tsay (2002). As a time-varying parameter, time series theories and methods can play an important role in both describing and forecasting volatility behaviour.

There are well-established time series models for univariate volatility, for example, ARCH Engle (1982) and GARCH Bollerslev (1986) models. Both models are performed under the assumption of a Gaussian distribution of returns and they are often estimated by offline methods, such as the Maximum Likelihood method and Markov Chain Monte Carlo method.

1.2 Aims of the thesis

In real-world scenarios, challenges are emerging for volatility modelling. Firstly, univariate modelling of volatility is plain and simple but does not investigate the correlation between different components. The correlation coefficient

contains important information of correlation and dependence; it can reveal positive or negative relationships between different assets. Lack of such information could cause serious damage for investors and decision makers. Therefore multivariate volatility and return estimation and forecasting is attractive.

Secondly, online estimation methods has been developed rapidly in the past decade. In comparision to the offline estimation which requires all input data, online estimation methods can estimate the model whenever new observation is available using only current measurements and previous parameter estimates. Since online methods always involve recursive algorithms, they are efficient in terms of memory usage and also require smaller computational effort. This efficiency makes them suited to online and embedded applications. Offline estimation methods are often not efficient enough for real traffic flow of time series data. For example, modern high frequency financial data are usually collected over 10 minutes or even shorter time interval. This makes offline estimation algorithms practically unsuitable for such kind of data, particular in modern automated financial systems which implement algorithmic trading. For these systems online estimation procedures, such as these proposed in this thesis, are attractive alternatives to their offline counterparts.

Therefore, there is a desire to study and explore multivariate volatility models with online estimation methods. In this thesis we will focus on developing multivariate volatility models. We will then estimate both the

volatility matrix and unknown parameters using sequential Monte Carlo and other online methods. We hope our work can push multivariate volatility modelling forward to medium and high data dimensions. We will also compare our models with the established benchmarking model in both function execution time and application performance. We would expect our models to have relatively shorter function execution time and similar or even better performance in application, especially for medium or high data dimensions. The rest of the thesis has the following structure:

Chapter 2 is the literature review; it introduces basic concepts of time series, financial time series and modern portfolio theory, together with the introduction of volatility and well-known volatility models. Estimation methods including MCMC, MLE and SMC are also mentioned.

In Chapter 3, we develop a series of multivariate volatility models, including the Gaussian, student-t and skew-t Wishart autoregressive models. We integrate the estimation of unknown parameters into the process of SMC so that both the autoregressive parameter and the volatility matrix can be estimated within the algorithm. Furthermore, we simplify the autoregressive parameter to be a diagonal matrix and carry out the estimation. The application and results are shown with both real and simulated data.

In Chapter 4, we develop another multivariate volatility model, the Wishart-Newton model. We were able to estimate both the volatility matrix and the unknown parameter (as a diagonal matrix) sequentially using the Newton-Raphson method. The application and results are shown with real data.

Chapter 5 compares each model through their application. Firstly we compare the efficiency of the Wishart autoregressive model, Wishart-Newton model and GO-GARCH model by their function execution time. Secondly, the performance of the three models are compared by their cumulative returns on 3, 5 and 10-dimensional real data. Finally, the performance of both the Wishart autoregressive model and Wishart-Newton model are compared by their cumulative returns on high dimensional Dow Jones 30 data.

Chapter 2

Literature Review

2.1 Time series and financial time series

2.1.1 Time series

The definition of time series can be found in many time series books, for example, Hamilton (1994) and Brockwell and Davis (2009). They define the time series as 'a sequence of numerical data points in successive order, usually occurring in uniform intervals'. Put simply, it is a sequence of data collected at regular intervals over time.

Since time series data are collected in time space, time series are often demonstrated via line charts. Figure 2.1 shows an example of a common time series plot: it shows the daily stock prices of IBM and Microsoft collected from 2 January 2008 to 31 December 2009. According to Fuller (2009), time series can be used in many areas, such as statistics, econometrics, mathemat-

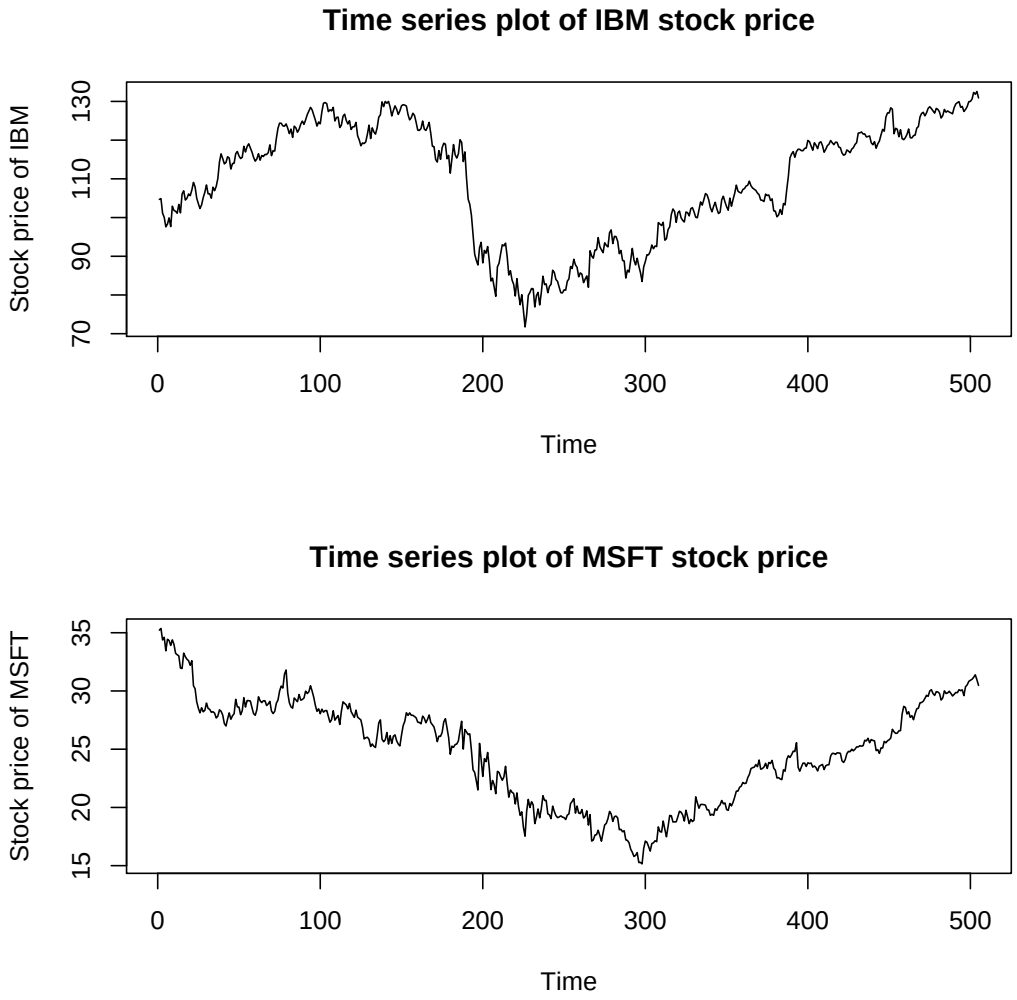

Figure 2.1: Time series plot of IBM and Microsoft stock prices

ical finance, weather forecasting, control engineering, signal processing and earthquake prediction.

Time series data

Time series data track information through time, showing trends of what has happened from one time to another. Time series data are different from other types of data simply because of their natural ordering property, see Lutkepohl (2006). For example, stock prices can only be input and studied in time sequence, rather than doing so in a random time line. Figure 2.2, details see Brillinger and David (2000), shows an instance of different types of time series data. They can be seasonal or have increasing or decreasing trending, random noise, etc.

Time series analysis

Time series analysis consists of models and estimation methods for analysing time series data in order to extract meaningful statistics and other properties of the data, see Wei and William (1994). The usage of time series analysis is twofold: 'one is to obtain an understanding for the underlying forces and structure that produced the observed data, the other one is to fit a model and proceed to forecasting, monitoring or even feedback and feed-forward control.', see Pham and Hoang (2006). In the first category are, for example, exploratory analysis (Shumway and Stoffer, 2000), autocorrelation analysis (Hamilton, 1994), statistical classification (Michie et al., 1994) and

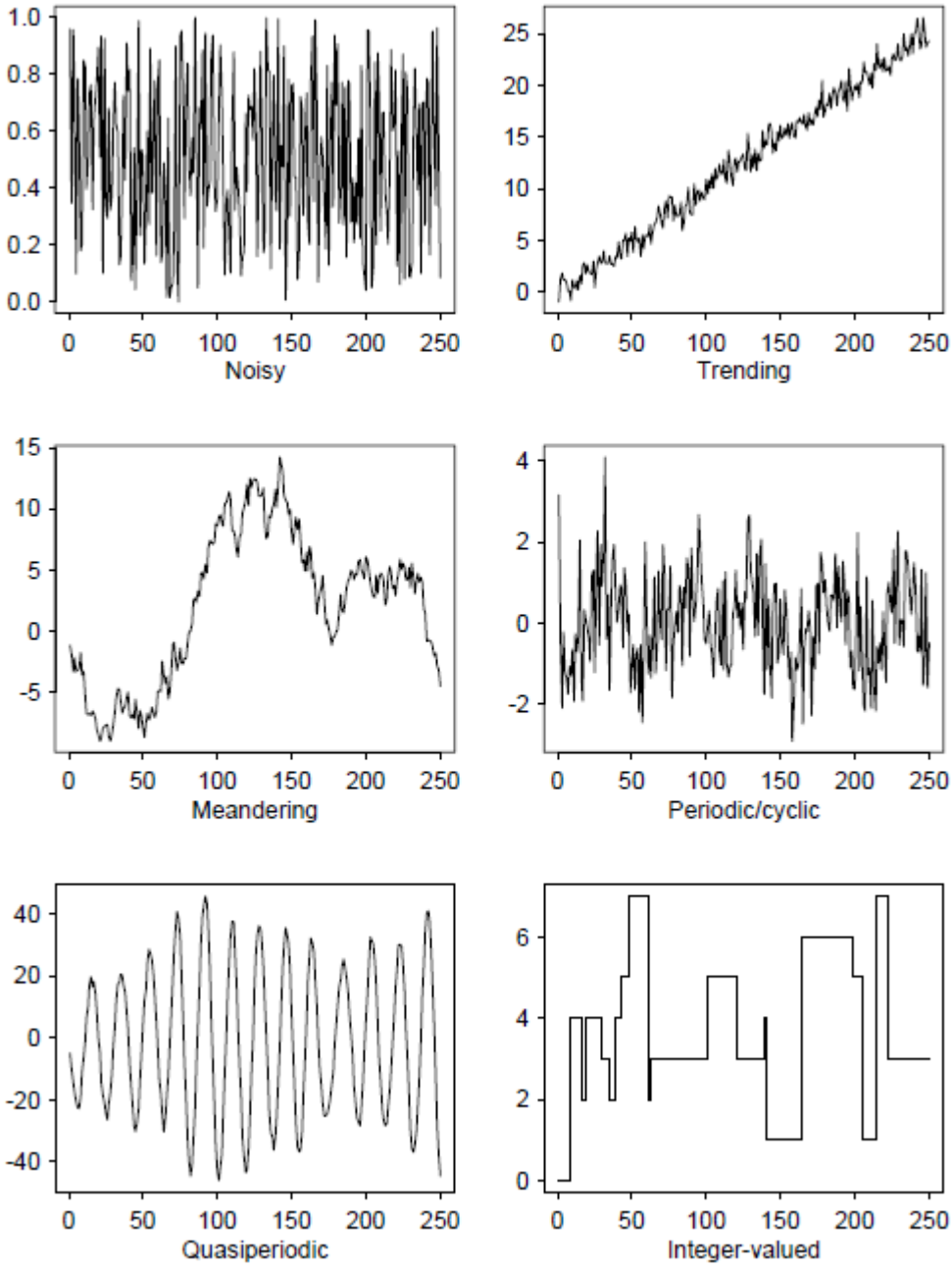

Figure 2.2: Different types of time series data

spectrum analysis (Kay et al., 1981). Examples of the second category include regression analysis, which is mainly based on the pioneered Box-Jenkins methodology (Box and Jenkens, 1970), prediction and forecasting (Weigend and Andreas, 1994) and signal estimation (Poor and Vincent, 1988).

In another perspective, time series analysis can also be divided into two sides, a theoretical side and an applied side. According to Brillinger and David (2000), 'the former side is part of the theory of stochastic process (representations, prediction, information and limit theory for instance), while the application side often involves extensions of techniques of "ordinary" statistics (for example regression, analysis of variance and sampling).'

2.1.2 Financial time series

Econometrics is one of the most important application areas of time series analysis. One of the reasons is that there are lots of time-varying financial data, for example, the stock price, option price, interest rate and gross domestic product (GDP), see Taylor and Stephen (2008). Furthermore, Tsay (2002) advocates that both financial theory and empirical time series contain an element of uncertainty and in order to monitor the uncertainty, statistical theory and methods play an important role in financial time series analysis.

It should be noted that financial time series have some characteristics that are different from other time series data. Most of the characteristics are prescribed as stylised facts. According to Sewell (2011), the term stylised facts is often used in economics to refer to empirical findings that are so

consistent that they are accepted as truth.

Distribution assumption

It is known that in economics, most of the pioneering studies, including the famous Box-Jenkins methodology, see Box and Jenkens (1970), which applies autoregressive moving average (ARMA) models to time series analysis, the models rely heavily on the assumption of normality. However, this assumption is often violated for financial time series. See for instance Ball et al. (1972), Jondeau et al. (2007). They believe that as a stylised fact, the distributions of asset data have tails that are much heavier than a normal distribution and the distribution is also negatively skewed. The negative skewness indicates there are more extreme negative values than positive values, which has a very serious implication for risk management and portfolio selection, for details see Peiro and Amado (1999). The skew-t distribution, see Azzalini and Capitanio (2003) and Appendix for details, are becoming more and more popular in financial time series analysis. Figure 2.3 shows an example of a negatively skewed t-distribution with simulated data.

Heteroskedasticity

According to Lux et al. (2000) and Krawiecki et al. (2002), asset returns are time dependent. Furthermore, squared returns, absolute returns and all measures and proxies of volatility exhibit strong serial correlation. Engle (1982) named this property volatility clustering or conditional heteroskedasticity.

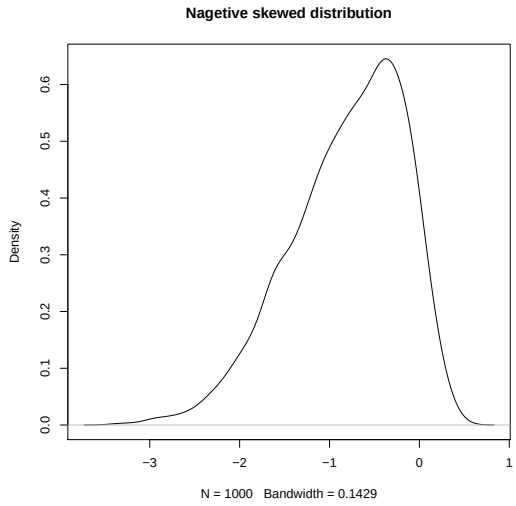

Figure 2.3: Density plot of negative skewed t-distribution

The details of volatility are discussed in later sections.

2.1.3 Modern portfolio theory

According to Muller and Heinz (1988), modern portfolio theory (MPT) indicates a thoery of finance that either, for a given portfolio risk, tries to maximise the portfolio return, or for a given level of portfolio return, tries to minimise the risk by selecting the proportions of various assets. Lubatkin et al. (1994) believe that MPT is a form a diversification. Under certain assumptions and for specific definitions of risk and return, MPT shows how to get the best diversification strategy in asset allocation.

Specifically, MPT uses the normal distribution to model assets' returns and it uses the standard deviation of assets' returns as the definition of risk.

It models the return of a portfolio as a weighted combination of the assets' returns. Then by combining different assets, MPT attempts to find the lowest total variance of total portfolio return.

MPT was first proposed by Markowitz in his article Markowitz (1952) and following book Markowitz (1959). MPT is a widely used theory in the finance industry and a Nobel memorial prize was even awarded to Markowitz for his contribution on MPT.

Some important assumptions about both markets and investors are made as follows for MPT:

- Assets' returns are assumed to be normally distributed.
- Correlations between assets are considered to be constant.
- Market information is shared by all investors at the same time.
- No tax or transaction fee for each trade.
- Unlimited amounts of money can be borrowed by all investors at the risk free rate of interest.
- Volatility is known in advance and it is assumed to be a constant.
- All investors are believed to be rational and risk-averse.
- All investors are aiming to make as much money as possible, which is also known as economic utility maximisation.
- No actions from investors will affect market prices.

However, both theoretical and practical criticisms against these assumptions have been raised since MPT has been proposed. In fact, almost all the assumptions above are now believed to be inaccurate or untrue. For example, Jondeau et al. (2007) suggest there is evidence supporting the fact that financial returns do not follow the normal distribution. Furthermore, they argue that the distribution of financial returns is not even symmetric. Also, Elton et al. (1997) propose that correlations between assets are not constant over time. For instance, in a financial crisis, all assets appear to be positively correlated since they all go down together with the whole market.

Although MPT has been shown to have some restrictions, it is still a very useful tool in the financial area and it was the first theory that used variance to quantify risk. It is a pioneering framework for financial mathematics and there are many extensions of MPT, for details see post-modern portfolio theory (Rom and Ferguson, 1994) and the Black-Litterman model (Black and Litterman, 1992). Applications to asset allocation will be demonstrated in later chapters.

2.2 Volatility and popular volatility models

2.2.1 Notations

The notations that are used throughout the thesis are introduced as follows:

- The p-dimensional vector of stock prices at time t are defined as $P_{1t}, \ldots, P_{pt}$.

- r_t is the return of the data and y_t is the log-return of the data, i.e. $r_t = P_t/P_{t-1}$ and $y_t = \log r_t$.
- σ is the volatility in univariate model and Σ is the volatility in multivariate model.
- a's are the parameters in univariate volatility model and A's are the parameters in multivariate volatility model.
- p is the dimension of the data.
- k is the degree of freedom of Wishart distribution; v is the degree of freedom of Inverse Wishart distribution; v_{st} is the degree of freedom of skew-t distribution.
- μ is the mean of the distributions.
- α is the skewness coefficient of the skew-t distribution.

2.2.2 Volatility

Tsay (2002) states that volatility is the conditional variance of a financial instrument over time. This is an important study direction in econometrics because it can affect decision making in finance, for example options trading. Also, in risk management, such as Value at Risk (VaR): one of the simple approaches to calculating the VaR of a financial position is by using volatility modelling. For more information about VaR, see Jorion (1997). Finally, in

time series inference and forecasting, volatility modelling can improve both forecast efficiency and forecast accuracy.

It should be noted that in the well-known Black-Scholes option price formula Black and Scholes (1973), volatility was assumed to be a constant, which has proved to be untrue in most cases. Volatility is now believed to be a time-varying parameter.

Properties of volatility

- Volatility is believed to be unobservable. Jacquier et al. (2002) suggest this property makes it difficult to evaluate model forecasting performance.

- There exist volatility clusters. This property is noted by Mandelbrot (1963): 'large changes tend to be followed by large changes, and small changes tend to be followed by small changes.' which means in some time periods, volatility can be low, while in other time periods, it can be high.

- Generally, time-varying volatility is believed to continuously evolve. It is quite rare to see volatility jumps.

- As a stylised fact, if we see volatility as a stochastic process, then it is always assumed to be a stationary process with fixed unconditional mean and variance. Because volatility quite often varies within some fixed range, it is rarely to see volatility breaks the limit and reaches

extremely small or large values. However, there are some studies discussing non-stationarity of the volatility, such as Cavaliere et al. (2007).

- According to Tsay (2002), volatility seems to be asymmetric, that is, volatility is higher in falling markets than in rising markets. This might be because investors are more afraid of losing money than gaining it, so that they are more likely to sell when the market goes down and this selling behaviour makes the market goes down even more.

2.2.3 Volatility models

ARCH model

Many models have been set up for volatility and it has become a popular research direction in recent times. However, it was the autoregressive conditional heteroscedastic (ARCH) model that first provided a systematic framework for the study of volatility modelling. It was proposed by Engle (1982).

The ARCH model assumes that the mean-corrected asset return r_t is serially uncorrelated, but it can be dependent with its past values. Then Engle (1982) uses a quadratic function of its lagged values to imply the dependence. Assume the return r_t follows a normal distribution with zero mean and σ_t^2 volatility. The mathematical description of the ARCH model

is demonstrated as follows:

$$\begin{aligned} r_t &= \sigma_t \epsilon_t \\ \sigma_t^2 &= a_0 + a_1 r_{t-1}^2 + \cdots + a_m r_{t-m}^2 \end{aligned} \tag{2.1}$$

where ϵ_t is set up to be a sequence of independent and identically distributed (iid) random variables with mean zero and variance one, the parameters $a_0 > 0$ and $a_i \geq 0$ for all $i > 0$. It should be noted that in most cases ϵ_t is assumed to a normal or a student-t distribution.

Tsay (2002) suggests the ARCH model implies that a large shock tends to be followed by another large shock because variance σ_t^2 is related to squared shocks $\{r_{t-i}^2\}$, and the current return r_t is related to variance σ_t^2. This property indicates the volatility clustering feature in assets returns. Furthermore, the tail distribution of an ARCH process is heavier than that of a normal distribution, which reflects the fact of the heavy tail of volatility.

There are also some drawbacks to the ARCH model, such as it is likely to over-estimate volatility, and it responds equally to both positive and negative shocks.

GARCH model

The GARCH model is an extension to the ARCH model. It was proposed by Bollerslev (1986). The model is given by:

$$\begin{aligned} r_t &= \sigma_t \epsilon_t \\ \sigma_t^2 &= a_0 + \sum_{i=1}^{m} a_i r_{t-i}^2 + \sum_{j=i}^{s} b_j \sigma_{t-j}^2 \end{aligned} \tag{2.2}$$

where, similarly to ARCH model, $\{\epsilon_t\}$ is a sequence of iid mean zero and variance one random variables and it is always assumed to be a normal or student-t distribution, the parameters are set to be $a_0 > 0$, $a_i \geq 0$, $b_j \geq 0$ and $\sum_{i-1}^{max(m,s)}(b_j + a_i) < 1$.

The main difference from the ARCH model is that GARCH takes the past information of variance, $\{\sigma_{t-j}^2\}$, into consideration when it models current variance σ_t^2.

From equation (2.2), we can see that either a large shock r_{t-i} or past variance σ_{t-j}^2 might lead to a large variance σ_t^2, and it is similar to the clustering property of volatility.

However, GARCH has almost the same disadvantages as ARCH. For instance, both positive and negative shocks have same influence on volatility. Ghose et al. (1995) believe the tail distribution of the GARCH process is still too short even with the assumption of ϵ_t having a student-t distribution.

Stochastic volatility model

Generally speaking, the difference between GARCH and Stochastic Volatility model is that, for GARCH model, it is a determinist function, if given the parameters, volatility is fixed. However for the stochastic volatility, volatility is assumed to be a random variable, mostly be described by probability distribution function.

The stochastic volatility (SV) model is an alternative method of modelling volatility by adding another innovation to the conditional variance equation. This model was developed by Melino and Turnbull (1990), Harvey et al. (1994) and Jacquier et al. (1994). The mathematical model is given as follows:

$$\begin{aligned} r_t &= \sigma_t \epsilon_t \\ log(\sigma_t^2) - a_0 &= (1 - a_1 B - \cdots - a_m B^m)^{-1} h_t \end{aligned}$$

where ϵ_t are i.i.d $N(0,1)$, h_t are i.i.d $N(0, \sigma_v^2)$ and normally it is assumed that the two innovations ϵ_t and h_t are independent. B is called back-shift operator such that $B \log(\sigma_t^2) = \log(\sigma_{t-1}^2)$. Also, it should be noted that a_0 is a constant and all zeros of the polynomial $1 - \sum_{i-1}^{m}(a_i B^i)$, are greater than 1 in modulus.

Tsay (2002) thought that adding an additional innovation h_t to the model gave more flexibility for the model to illustrate the evolution of volatility σ_t^2. However, introducing another term, especially a random variable into the

model increases the difficulty of model parameter estimation dramatically.

Jacquier et al. (1994) believed that a SV model needed a quasi-likelihood method via Kalman filering or a MCMC method to estimate the parameters. They used both methods and gave a comparison in their research.

Furthermore, relatively recently, sequential Monte Carlo (SMC) methods, also known as particle filtering, have been developed. Thes methods have been proved to be fairly efficient and accurate for solving the optimal estimation problems in non-linear and non-Gaussian scenarios. For more details see Liu (1991), Doucet et al. (2001) and Cappe et al. (2005).

2.2.4 Multivariate GARCH model

The constant conditional correlation (CCC)-GARCH model (Bollerslev (1990)) is the simplest multivariate volatility model which is nested in other conditional correlation models. As a natural extension of univariate GARCH, the return and volatility matrix of a CCC-GARCH(m,s) in a vector form can be described as follows:

$$\begin{aligned} r_t &= \Sigma_t \epsilon_t \\ \Sigma_t &= A_0 + \sum_{i=1}^{m} A_i r_{t-i}^2 + \sum_{j=i}^{s} B_j \Sigma_{t-j}^2 \end{aligned}$$

where ϵ_t follows multivariate normal distribution, A_0 is a vector, A_i and B_j are diagonal matrix with positively definite diagonal elements.

With further extension to treat parameter matrix A_i and B_j as full ma-

trix instead of diagonal matrix, extended CCC (ECCC)-GARCH was introduced by Jeantheau (1998). As another extension of CCC-GARCH, Engle (2002) also developed dynamic conditional correlation (DCC)-GARCH model, which has a dynamic volatility structure.

2.2.5 Rotated Baba-Engle-Kraft-Kroner(RBEKK) model

The RBEKK model is an extension from famous Baba-Engle-Kraft-Kroner (BEKK) model(see Engle and Kroner (1995)). RBEKK model added an extra term, rotated log-return $\tilde{y}_t$, as:

$$\begin{aligned} y_t &= \Sigma_t^{1/2}\epsilon_t \\ \tilde{y}_t &= H^{-1/2}y_t = P\Lambda^{-1/2}P'r_t \end{aligned}$$

where H is the unconditional covariance of y_t and $P\Lambda^{-1/2}P'$ is the spectral decomposition of $H^{-1/2}$. ϵ_t is a i.i.d process with zero mean and identity covariance matrix.

With the rotated returns defined, the conditional covariance matrix of y_t can be defined as:

$$\Sigma_t = (I_p - AA' - BB') + A\tilde{y}_t\tilde{y}_t'A' + B\Sigma_{t-1}B'$$

where $(I_p - AA' - BB') > 0$. A and B are two parameters.
In the later chapter, the data simulation is carried out by setting $A = B$ and

they are diagonal matrix with diagonal elements equal to 0.6 and $\Sigma_0 = I_p$.

2.2.6 GO-GARCH model

VanDerWeide (2002) developed generalized orthogonal GARCH (GO-GARCH) as an extension of orthogonal GARCH model, with having the log-return y_t as:

$$y_t = H_t^{1/2}\epsilon_t$$

where H_t is a diagonal matrix with $H_t = diag(h_{1i}, \ldots, h_{pt})$, p is the dimension of the data. ϵ_t are a sequence of i.i.d. random variable with zero mean. Then have the covariance matrix and the diagonal elements of H_t as:

$$\begin{aligned} \Sigma_t &= ZH_tZ' \\ h_{it} &= (1 - \alpha_i - \beta_i) + \alpha_i y_{i(t-1)} + \beta_i h_{i(t-1)} \end{aligned}$$

where Z is a non singular matrix.

Silvennoinen and Terasvirta (2009) think GO-GARCH is one of the most popular and advanced model in GARCH family and Noureldin et al. (2014) suggests that the GO-GARCH model has better performance than RBEKK model on medium dimensional data.

2.2.7 Wishart autoregressive process

The Wishart autoregressive process is a newly developed method to estimate the multivariate stochastic volatility by Gourieroux et al. (2009) and Triantafyllopoulos (2012). The model is given as follows:

$$\begin{aligned} r_t &= \sum_{k=1}^{K} x_{kt} x_{kt}' \\ x_{k,t+1} &= M x_{kt} + \epsilon_{k,t+1}, \epsilon_{k,t+1} \sim N(0, \Sigma) \end{aligned}$$

where $N(0, \Sigma)$ denotes a multivariate Gaussian distribution with mean zero and covariance matrix Σ. M is a autoregressive parameter which links between x_t and x_{t+1}.

From the model, we can see that $\{x_{kt}\}$ follow AR(1) processes hence they have multivariate Gaussian distributions. The main idea is that, in the univariate case, we have the property that the sum of the square Gaussian distributions follows a gamma distribution, then its multivariate extension can be used here and the log-return y_t follows a Wishart distribution, $W_n(K, M\Sigma M')$.

The model provides an alternative to the standard GARCH and stochastic volatility models. It shows flexibility in the way of changing the autoregressive lags and providing factor representation. Furthermore, the idea of using Wishart distribution to describe the covariance matrix is inspiring, and further work will be introduced in later chapters of the thesis.

2.3 Estimation and inference

2.3.1 Markov chain Monte Carlo

The Markov chain Monte Carlo space (MCMC) method is a widely used simulation based algorithm for calculating numerical approximations, for example in Bayesian inference, computational biology and physics. It samples from a constructed Markov chain which has the target distribution as its equilibrium distribution. After steps of simulations, the chain converges and the current state of the chain is used as a sample of the target distribution. So that the more steps or simulations the Markov chain runs, the closer it gets to the desire distribution, therefore the better result will get. For details see Gilks (2005).

The famous Metropolis algorithm of MCMC was first invented by Metropolis et al. (1953), when they tried to simulate a liquid in equilibrium with its gas phase. They realised that the simulation does not need to have the exact dynamics: it is sufficient for the Markov chain to share the same equilibrium distribution. Then decades later, Hastings (1970) generalised the algorithm and formed the so-called Metropolis-Hastings algorithm. Geman and Geman (1984) introduced a special case of the Metropolis-Hastings algorithm without knowledge of the previous work; the simulation is called the Gibbs sampler. Further work was also done by Gourierouxa et al. (1987) and Gelfand and Smith (1990).

It was not until the 1990s that people realised most Bayesian inference can

be done by the Metropolis-Hastings algorithm and the Gibbs sampler. Since then, both research and applications have grown rapidly. Further work has been done by Geyer (1994) and Green (1995), adaptive rejection Metropolis sampling by Gilks et al. (1995) and the acceptance-rejection Metropolis-Hastings algorithm was developed by Tierney (1994) for sampling in MCMC. Asai (2006) believes that the acceptance-rejection Metropolis-Hastings algorithm is one of the most efficient algorithm at the time.

2.3.2 Maximum likelihood estimation

Maximum likelihood estimation (MLE) is probably the most well-known estimation method in statistics. Although MLE has been used in other articles, Fisher (1925) vastly popularised this method.

Assume that we have $x_1, \ldots, x_n$ as n independent and identically distributed data. The method requires the joint density function of all observations conditional on the parameter θ:

$$f(x_1, x_2 \ldots, x_n|\theta) = f(x_1|\theta) \times f(x_2|\theta) \times \cdots \times f(x_n|\theta),$$

then the observations are considered to be fixed 'parameters' and the likelihood function is shown as:

$$l(\theta; x_1, \ldots, x_n) = f(x_1, x_2 \ldots, x_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$$

The MLE works by finding the θ that maximises the likelihood. For example, the log-likelihood function for the ARCH model is shown as follows:

$$l(\sigma; r_{m+1}, \ldots, r_T) = -\sum_{t=m+1}^{T} \left[\frac{1}{2} ln(\sigma_t^2) + \frac{1}{2} \frac{r_t^2}{\sigma_t^2} \right]$$

where l is log-likelihood, σ_t is the volatility and r_t are the return of data. Then by maximising the function, we can obtain the estimators for σ.

MLE has proved to be useful in many cases because it can deliver an explicit function of the estimators. However, in some situations, a close-formed solution may not be possible for the likelihood function, there might be multiple solutions available to maximise the likelihood, or no maximum likelihood estimates may exist (Hamilton, 1994).

2.3.3 Sequential Monte Carlo

The SMC method, which is also known as particle filtering, was first named by Moral and Pierre (1996) in the application to fluid mechanics. The particle filter technique was actually developed as early as 1955, in molecular chemistry, see Rosenbluth et al. (1955). The method was not widespread at that time due to its computationally expensive mechanics. However, with the rapid development of computational power these days, SMC has become a popular numerical estimation method, especially for solving non-linear and non-Gaussian problems. It is widely used in such fields as chemical engineering, computer vision, financial econometrics and robotics, see Doucet and

Johansen (2009).

Arulampalam et al. (2002) suggests that the main advantage of SMC is its sequential property. As the data are received at each time point they can be used to estimate the model sequentially instead of being used as a batch. In this way, the whole data set does not need to be stored and reprocessed every time. The current state can be updated based only on knowledge of the new data by using Bayes theorem.

There are many different available algorithms in SMC, for example the bootstrap filter method, sequential importance sampling, sequential importance resampling, the hybrid method and auxiliary particle filter (see Appendix for details).

Chapter 3

Multivariate Stochastic Volatility Estimation Using Particle Filter

3.1 Introduction

This chapter mainly focuses on developing Wishart autoregressive models for multivariate volatility estimation. The model is inherited from Philipov et al. (2006)'s factor stochastic volatility model. However we neglect the mean term and scalar parameter in the model and push the model for high dimension data. Furthermore, we also added student-t and skew-t into our model to describe both symmetric and heavy-tailed asymmetric observations. The factor stochastic volatility model is estimated by an offline MCMC,

while we estimate the volatility matrix and autoregressive parameter with an online SMC algorithm. Then our model is further simplified by treating the autoregressive parameter as a diagonal matrix so that higher dimensional data can be applied. The applications on both real-world data and simulated data are shown in the end.

3.2 Naive model

3.2.1 Gaussian Wishart autoregressive model

Let us consider a p-dimensional time series vector $\{y_t\}$, which usually is the log-return of the asset prices or stock prices or any other financial instrument. The log-return of the prices can be interpreted as, for example, if we have the p-dimensional vector of stock prices $(P_{1t}, P_{2t}, \cdots, P_{pt})$, then:

$$y_{it} = \log(P_{it}) - \log(P_{i(t-1)}) = \log\left(\frac{P_{it}}{P_{i(t-1)}}\right)$$

where $i = 1, ..., p$.

There are several reasons that we use log-return instead of raw data in this research. The first reason is that log-return can normalization the data, which means measuring all variables in a comparable metric. The second reason is that log-return data can often prevent serial correlation both for the series and error terms. The third reason is that log-return data can often stationary which is vital for autoregressive models. The fourth reason is that

when returns are very small, the log-returns ensures they are close to the raw returns.

For the basic model, we inherit the classical idea from econometrics that conditional on the volatility matrix Σ_t, $\{y_t\}$ is believed to follow a multivariate normal distribution, i.e.,

$$y_t|\Sigma_t = \mu + \Sigma_t^{1/2}\epsilon_t, \quad \epsilon_t \sim N_p(0, I_p), \tag{3.1}$$

where μ is the expectation vector of the observations and $\{\epsilon_t\}$ is a sequence which follows a normal distribution with zero means and diagonal variances matrix, also known as white noise.

Given the properties of the normal distribution, equation (3.1) can also be written as:

$$y_t|\Sigma_t \sim N_p(\mu, \Sigma_t) \tag{3.2}$$

It should be noted that the volatility matrix Σ_t, is usually assumed to be a strictly positive-definite and $p \times p$ symmetric matrix.

Then we assume the evolution of the volatility precision matrix Σ_t^{-1} as:

$$\Sigma_t^{-1}|\Sigma_{t-1}, A \sim W_p\left(k, \frac{1}{k}\Sigma_{t-1}^{-\frac{1}{2}} A \Sigma_{t-1}^{-\frac{1}{2}}\right) \tag{3.3}$$

where W denotes a Wishart distribution, which is the multivariate extension of the gamma distribution (details in Appendix). k is the degree of freedom parameter and is defined to be greater than $p - 1$.

The reasons equation (3.3) is proposed are:

- By the definition of volatility, Σ_t is a positive-definite symmetric matrix. As an extension of the gamma distribution, the Wishart distribution also inherits the positive-definite symmetric property.

- The Wishart distribution is well-known as a conjugate distribution. This property proves very convenient in the Bayesian inference later on.

- The scale matrix of the Wishart distribution $\frac{1}{k}\Sigma_{t-1}^{-\frac{1}{2}}A\Sigma_{t-1}^{-\frac{1}{2}}$ is motived by the autoregressive model, i.e. the conditional expectation of the volatility $E(\Sigma_t^{-1}|A,\Sigma_{t-1}) = \Sigma_{t-1}^{-\frac{1}{2}}A\Sigma_{t-1}^{-\frac{1}{2}}$ which matches the mean of the AR(1) model. A is the autoregressive coefficient in the model, it interprets the relation between current volatility state with its past. So that the closer eigenvalue of A to 1, the more dependence in the volatility and we don't expect the eigenvalue close to 0. Therefore it is our interest to develop a way of estimating A.

Given equation (3.2) and equation (3.3), the basic mean-volatility model we propose is:

$$\begin{aligned} \Sigma_t^{-1}|\Sigma_{t-1},A &\sim W_p(k,\frac{1}{k}\Sigma_{t-1}^{-\frac{1}{2}}A\Sigma_{t-1}^{-\frac{1}{2}}) \\ A &\sim IW(v_0,\beta_0) \\ y_t|\Sigma_t &\sim N(\mu,\Sigma_t), \end{aligned} \tag{3.4}$$

where $p(A)$ is the prior distribution of A and follows an inverse Wishart distribution (for details see Appendix).

The advantage of this model set-up is that, first, in this model the description of the volatility matrix is focussed on the single autoregressive parameter matrix A, while other stochastic volatility models, for example, the ARCH and GARCH models, have two or more parameter matrices. The reduction of parameters makes a huge difference in the fight against the curse of dimensionality with possible sacrifice of accuracy; furthermore, A is a symmetric matrix by definition, which further reduces the number of unknown parameters in high dimensions by nearly a half. Second, the model structure depends on the Wishart distribution; the conjugate property makes it much easier to adapt and facilitate Bayesian inference in the estimation process. In later sections (3.2.3 and 3.2.4), we will derive the posterior distribution for both the optimal importance function and the parameter matrix A of the model.

3.2.2 Extended models

However, in real-world scenarios, financial time series are believed not to be normally distributed (see Literature Review): financial forecasters report returns exhibiting heavy tails. Therefore, the basic model defined above should be extended to describe data with heavy-tailed distributions. A naive approach to achieving this would be to use a student-t distribution to reflect

the heavy-tailedness:

$$\begin{aligned} \Sigma_t^{-1}|\Sigma_{t-1}, A &\sim W(k, \frac{1}{k}\Sigma_{t-1}^{-\frac{1}{2}} A \Sigma_{t-1}^{-\frac{1}{2}}) \\ A &\sim IW(v_0, \beta_0) \\ y_t|\Sigma_t &\sim t(v, \Sigma_t), \end{aligned} \tag{3.5}$$

where $t(v)$ is a student-t distribution with v degrees of freedom.

In equation (3.5), the volatility function and the prior function remain the same. A student-t distribution is used instead of a normal distribution to describe the heavy-tailed returns.

However, recent finance literature suggests that skewness can be found in most financial time series data. So in order to reflect not only the heavy-tailed returns but also the skewness, we propose the skew-t distribution(details in Appendix). The student-t model (equation (3.5)) can be further extended as follows:

$$\begin{aligned} \Sigma_t^{-1}|\Sigma_{t-1}, A &\sim W(k, \frac{1}{k}\Sigma_{t-1}^{-\frac{1}{2}} A \Sigma_{t-1}^{-\frac{1}{2}}) \\ A &\sim IW(v_0, \beta_0) \\ y_t|\Sigma_t &\sim St(\mu, \Phi(\Sigma_t), \alpha, v_{st}), \end{aligned} \tag{3.6}$$

where $St(\mu, \Phi, \alpha, v_{st})$ stands for a skew-t distribution with v_{st} degree of freedom. It should be noted that μ is defined as a 'location' parameter instead of the expectation and Φ is defined as a 'shape' parameter, which is a function

of the variance Σ_t of the skew-t distribution. α is the skewness factor. The mean (E(X)) and variance (Σ_t) of the skew-t distribution can be related to μ and Σ as follows:

$$\begin{aligned} \mu &= E(X) - \alpha(\frac{v_{st}}{v_{st}-2}) \\ \Phi &= \big(\Sigma_t - \alpha^2 \frac{2v_{st}^2}{(v_{st}-2)^2(v_{st}-4)}\big)\frac{v_{st}}{(v_{st}-2)} \end{aligned} \tag{3.7}$$

In equation (3.6), the volatility function and prior function again remain unchanged, and the skew-t distribution is used to reflect the properties of heavy-tailedness and skewness of the data.

3.2.3 Bayesian inference for the autoregressive parameter, A

We have defined the naive volatility model and its extended models with heave tails and skewness in the previous subsection. Since our model follows a wishart autoregressive process and we neglect the mean term to further simplified the model, the autoregressive parameter matrix A is the most important parameter matrix in all our three models, used to describe the relationship between present and past volatility. In this case, based on the proposed models and the fact that the Wishart distribution is the conjugate prior for the Wishart likelihood function, Bayesian inference can be applied to derive the posterior probability function of A for the Gaussian model, the

student-t model and the skew-t model:

$$\begin{aligned}
p(A|\Sigma_t^{-1}, \Sigma_{t-1}) &= \frac{p(\Sigma_t^{-1}|\Sigma_{t-1}, A)p(A|\Sigma_{t-1})}{p(\Sigma_t^{-1}|\Sigma_{t-1})} \\
&\propto p(\Sigma_t^{-1}|A, \Sigma_{t-1})p(A|\Sigma_{t-1}) \\
&= W(k, \frac{1}{k}\Sigma_{t-1}^{-\frac{1}{2}}A\Sigma_{t-1}^{-\frac{1}{2}})IW(v_{t-1}, \beta_{t-1}) \\
&= |\frac{1}{k}\Sigma_{t-1}^{-\frac{1}{2}}A\Sigma_{t-1}^{-\frac{1}{2}}|^{-\frac{k}{2}}|\Sigma_t^{-1}|^{\frac{k-p-1}{2}} \\
&\quad e^{-\frac{1}{2}tr((\frac{1}{k}\Sigma_{t-1}^{-\frac{1}{2}}A\Sigma_{t-1}^{-\frac{1}{2}})^{-1}\Sigma_t^{-1})}|\beta_{t-1}|^{\frac{v}{2}}A^{-\frac{v+p+1}{2}}e^{-\frac{1}{2}tr(\beta_{t-1}A^{-1})} \\
&\propto |A|^{-\frac{k+v+P+1}{2}}e^{-\frac{1}{2}tr((k\Sigma_{t-1}^{\frac{1}{2}}\Sigma_t^{-1}\Sigma_{t-1}^{\frac{1}{2}}+\beta_{t-1})A^{-1})} \\
&= IW(k + v_{t-1}, k\Sigma_{t-1}^{\frac{1}{2}}\Sigma_t^{-1}\Sigma_{t-1}^{\frac{1}{2}} + \beta_{t-1}) \\
&= IW(v_t, \beta_t) \\
v_t &= v_0 + tk \\
\beta_t &= \sum_{i=1}^{t}(k\Sigma_{i-1}^{\frac{1}{2}}\Sigma_t^{-1}\Sigma_{i-1}^{\frac{1}{2}}) + \beta_0 \qquad (3.8)
\end{aligned}$$

where v_t and β_t are the degree of freedom and scale matrix parameters for A.

Since we have derived the posterior probability function of unknown parameter A to be an inverse Wishart distribution. We treat A as a fixed term in the model but we would like to estimate A sequentially, because in this way we can integrate the estimation process of A into the estimation process of SMC, the details will be shown in later algorithms. We will have to use the idea of Storvik (2002). which is treating the posterior probability function of A at previous time $t-1$ as the prior probability function at current time t

so that at time point t, the Bayesian inference can be performed and we can obtain the posterior probability function of A sequentially.

3.2.4 Bayesian inference for importance function

Recall that the SMC has been introduced in the literature review and the details of SMC algorithms can be found in the Appendix. We know that the selection of the importance function is a crucial part of SMC estimation. A good choice of importance function can lead to minimal variance of the importance weights, which results in better particle generation for the next time point and a better estimation at the current time point. The optimal importance function of SMC can be defined as $p(\Sigma_t^{-1}|\Sigma_{t-1}, y_t)$.

It should be noted that the optimal importance function is often difficult to obtain. This is because firstly, the calculation may involve complex high dimensional integrations of Σ_t, which most of time it is rather tedious or impossible to perform. Secondly, it may also be impossible to directly sample from the derived complex importance function.

But for the Gaussian model (equation(3.4)), based on the conjugation property of the Wishart distribution, Bayesian inference can also be applied

directly to derive the optimal importance function as follows:

$$
\begin{aligned}
p(\Sigma_t^{-1}|\Sigma_{t-1}, y_t) &= \frac{p(y_t|\Sigma_t^{-1}, \Sigma_{t-1})p(\Sigma_t^{-1}|\Sigma_{t-1})}{p(y_t|\Sigma_{t-1})} \\
&\propto p(y_t|\Sigma_t^{-1})p(\Sigma_t^{-1}|\Sigma_{t-1}) \\
&\propto N(\mu, \Sigma_t)W(k, \frac{1}{k}\Sigma_{t-1}^{-\frac{1}{2}}A\Sigma_{t-1}^{-\frac{1}{2}}) \\
&\propto |\Sigma_t^{-\frac{1}{2}}|e^{-\frac{1}{2}tr(y_t y_t'\Sigma_t^{-1})}|\Sigma_t|^{-\frac{k+p+1}{2}}e^{-\frac{1}{2}tr((\frac{1}{k}\Sigma_{t-1}^{-\frac{1}{2}}A\Sigma_{t-1}^{-\frac{1}{2}})^{-1}\Sigma_t^{-1})} \\
&\propto |\Sigma_t|^{-\frac{k+1+p+1}{2}}e^{-\frac{1}{2}tr((y_t y_t'+(\frac{1}{k}\Sigma_{t-1}^{-\frac{1}{2}}A\Sigma_{t-1}^{-\frac{1}{2}})^{-1})\Sigma_t^{-1})} \\
&\propto W\left(k+1, \left((y_t y_t') + (\frac{1}{k}\Sigma_{t-1}^{-\frac{1}{2}}A\Sigma_{t-1}^{-\frac{1}{2}})^{-1}\right)^{-1}\right), \qquad (3.9)
\end{aligned}
$$

where p is the dimension of the data. Equation (3.9) gives us a closed-formed optimal importance function of SMC. It is also rather easy to directly sample from such a Wishart distribution.

It should be noted that, for the student-t and skew-t model, although equation (3.9) is not the optimal importance function for them, it is still believed to be a promising choice of importance function for their SMC algorithms.

3.2.5 Exponentially weighted moving average for skewness coefficient α

Recall equation (3.6): we have four parameters, μ, Φ, α, v, in the skew-t distribution. The plan is to estimate these parameters within the framework of the SMC method. The most nature idea is that we can perform full

Bayesian Inference on all the parameters in the skew-t distribution, however that would be further complicated the algorithm and adding more computation. Our interest is in pushing the model onto high-dimensional data so that we choose to simplify the algorithm and reduce the computation.

- The location parameter, μ, can be interpreted as a function of the expectation of the data. Since typically the log-return of a financial instrument is used as data, either the historical mean of the data or even simply value zero can be used as the estimation of the expectation.

- The shape parameter, Φ, is a function of the volatility matrix, which will be estimated by the SMC method at each time.

- The degrees of freedom parameter, v, can be easily determined from the shape of the density of the data. If the data distribution has heavy tails, then a small value (e.g. 3 or 5) can be used. However if the data show no sign of heavy-tailedness, then a large value (e.g. 20 or 30) can be used.

- The skewness coefficient, α, is the parameter used to describe the skewness of the data, which is one of the most important parameters in the skew-t distribution. The definition of the skewness α in univariate statistics is:

$$\alpha = \frac{E(y-m)^3}{Var(y)^3} = \frac{E(y^3) - m^3 + 3m^2E(y) - 3mE(y^2)}{\Sigma^3}$$

Here m is the expectation of the log-return of the data, normally it is close to zero; Σ is the variance of the data. Therefore the terms with m can be neglected and the skewness can be rewritten as:

$$\alpha = \frac{E(y^3)}{\Sigma^3}$$

In order to integrate the estimation of the skewness coefficient into the SMC method, the estimation process has to be sequential. Therefore we propose the exponential moving average (EMA) (for details see Lawrance and Lewis (1977)) to estimate the skewness coefficient at each time point:

$$\hat{\alpha}_t = \lambda \hat{\alpha}_{t-1} + (1-\lambda)\frac{E(y_{1:t}^3)}{\Sigma_t^3}, \tag{3.10}$$

where $0 < \lambda < 1$ is the coefficient that represents the degree of weighting decrease and it affects the results of the moving average. $\lambda < 0.5$ means skewness is more related to the $\frac{E(y_{1:t}^3)}{\Sigma_t^3}$ at each time, while $\lambda > 0.5$ makes the previous $\hat{\alpha}_{t-1}$ more important in the estimation.

However, in the multivariate scenario, in order to deal with the curse of dimensionality, further simplify the calculation and reduce the number of unknown parameters as much as possible, the equation (3.10) is replaced by:

$$\hat{\alpha}_t = \lambda \hat{\alpha}_{t-1} + (1-\lambda) y_t \tag{3.11}$$

Compared to equation (3.10), equation (3.11) is reduced from a cubic func-

tion to a linear function. It simplifies the calculation dramatically, especially for the high dimensional matrix calculation. We link the skewness coefficient with the observation is because we think observations have great influence on the parameter. For example, when the observations are mostly positive and large, it is very likely to lead the distribution to a positive skewness coefficient. Inevitably, there are losses in estimation accuracy, but that is the tradeoff we have to make to deal with the curse of dimensionality.

3.2.6 Algorithms

Algorithm 1 shows the application of the SMC method for the initial model (equation (3.4)). At each time point, we update the posterior probability function for the autoregressive coefficient matrix A and the optimal importance function. Then the updated information can be used to calculate the particle weights. After that the weights can be normalised which leads to an estimation of the volatility matrix.

Algorithm 2 is the application of the SMC method for the skew-t model (Equation (3.6)) with the two unknown parameters: autoregressive coefficient A and skewness coefficient α. In the same way as algorithm 1, the posterior function of A is updated at each time, then the updated information can be used to estimate the skewness coefficient. With the estimated skewness coefficient, we can calculate the particle weights and finally move on to the estimation of the volatility matrix.

Algorithm 1 Algorithm of SMC for normal model with A

When $t = 0$, Σ_0 and A can be simulated from their prior distributions:

$$p(\hat{A}_0) = IW(v_0, \beta_0)$$

and

$$p(\Sigma_0^{-1}) = W(k_0, B_0)$$

For t=1, Σ_1^{-1} can be simulated from its own distribution using the simulated parameters from time 0 and A_1 can also be derived:

$$p(\Sigma_1^{-1}|\Sigma_0, \hat{A}_0) = W(k, \frac{1}{k}\Sigma_0^{-\frac{1}{2}}\hat{A}_0\Sigma_0^{-\frac{1}{2}})$$

$$p(\hat{A}_1|\Sigma_1^{-1}, \Sigma_0) \propto IW(k, k\Sigma_0^{\frac{1}{2}}\Sigma_1^{-1}\Sigma_0^{\frac{1}{2}} + \beta_0)$$

for $t = 2, \ldots, T$ **do**

The unknown parameter A_t is updated at every time point t as:

$$\begin{aligned} p(\hat{A}_t|\Sigma_t^{-1}, \Sigma_{t-1}) &\propto IW(k + (t-1)v, \phi_{t-1}) \\ \phi_1 &= k\Sigma_0^{\frac{1}{2}}\Sigma_1^{-1}\Sigma_0^{\frac{1}{2}} + \beta_0 \\ \phi_t &= k\Sigma_{t-1}^{\frac{1}{2}}\Sigma_t^{-1}\Sigma_{t-1}^{\frac{1}{2}} + \phi_{t-1} \end{aligned}$$

After obtaining the unknown parameter A at time t, the SMC method can be applied as follows:

for $i = 1, \ldots, N$ **do**

Draw Σ_t^i from $p_{t|t-1}(\Sigma_t|\Sigma_{t-1}, y_{t-1}, A)$.

Set

$$\hat{w}_t^i = \frac{p(y_t|\Sigma_t)p(\Sigma_t^i|\Sigma_{t-1}, A)}{g(\Sigma_t^i|\Sigma_{t-1}, y_t, A)}\hat{w}_{t-1}^i$$

Normalise the weights:

$$w_t^i = \frac{\hat{w}_t^i}{\sum_{j=1}^N \hat{w}_t^j}$$

end for

Compute:

$$\hat{\Sigma}_t = \sum_{i=1}^N w_t^i \delta_{\Sigma_t^i}$$

where δ_Σ denotes the Dirac delta mass located at Σ_t.

end for

Algorithm 2 Algorithm of SMC for skew-t model with A and α

1: At $t = 0$, X_0 and A can be simulated from their prior distributions:

$$p(\hat{A}_0) = IW(v, \beta) \qquad p(\Sigma_0^{-1}) = W(k, B)$$

2: At t=1, Σ_1^{-1} can be simulated from its own distribution using the simulated parameters from time 0 and A can also be derived:

$$p(\Sigma_1^{-1}|\Sigma_0, \hat{A}_0) = W(k, \frac{1}{k}\Sigma_0^{-\frac{1}{2}}\hat{A}_0\Sigma_0^{-\frac{1}{2}}) \quad and \quad p(\hat{A}_1|\Sigma_1^{-1}, \Sigma_0) \propto IW(k, kX_0^{\frac{1}{2}}\Sigma_1^{-1}\Sigma_0^{\frac{1}{2}} + \beta)$$

Also the initial value for $\hat{\alpha}_t$ can be set up to be $\hat{\alpha}_1$.

3: **for** $t = 2, \ldots, T$ **do**

4: The unknown parameter A is updated at every time point t as:

$$\begin{aligned} p(\hat{A}_t|\Sigma_t^{-1}, \Sigma_{t-1}) &\propto IW(k + (t-1)v, \phi_{t-1}) \\ \phi_1 &= k\Sigma_0^{\frac{1}{2}}\Sigma_1^{-1}\Sigma_0^{\frac{1}{2}} + \beta \\ \phi_t &= k\Sigma_{t-1}^{\frac{1}{2}}\Sigma_t^{-1}\Sigma_{t-1}^{\frac{1}{2}} + \phi_{t-1} \end{aligned}$$

After obtained the unknown parameter A at time t, then the SMC method can be applied as follows:

5: **for** $i = 1, \ldots, N$ **do**

6: Draw Σ_t^i from $p_{t|t-1}(\Sigma_t^i|\Sigma_{t-1}, y_{t-1}, A)$.

7: The unknown parameter α is updated as:

$$\hat{\alpha}_t^i = \lambda\hat{\alpha}_{t-1}^i + (1-\lambda)y_t$$

8: Set

$$\hat{w}_t^i = \frac{p(y_t|\Sigma_t, \hat{\alpha}_t^i)p(\Sigma_t^i|\Sigma_{t-1}, A)}{g(\Sigma_t^i|\Sigma_{t-1}, y_t, A)}\hat{w}_{t-1}^i$$

9: Normalise the weights:

$$w_t^i = \frac{\hat{w}_t^i}{\sum_{j=1}^N \hat{w}_t^j}$$

10: **end for**

11: Compute:

$$\hat{\Sigma}_t = \sum_{i=1}^N w_t^i \delta_{\Sigma_t^i}$$

where δ_Σ denotes the Dirac delta mass located at Σ_t.

12: **end for**

3.3 Estimation and applications

3.3.1 Data analysis

In this section, the proposed Gaussian Wishart autoregressive model and the extended skew-t-Wishart autoregressive model were applied to real data by using the SMC method. It should be noted that for the purposes of demonstration, two-dimensional data were used. The data set was selected from the stock prices of Cairn Energy and Anglo American over the period from 2006 to 2008 at daily frequency. The set of stock prices consisted of 500 observations in total. The dataset can be found in the appendix document 'CairnEnergy and AngloAmerica data.txt'. A time series plot of the data is shown in Figure 3.1. From the plot, though each stock seems to follow a random walk, the overall trend of the two stock prices are quite opposite. Cairn Energy starts off high and goes down in the middle and ends up going up, while Anglo American starts off low, goes up in the middle and ends up in a relatively high position. It is natural for us to assume that the two stock prices share more negative correlations than positive.

We haven't use simulated data to fit our model is because the main structure of our model is set to be an autoregressive structure. However in order to push the model to higher dimensions, reduce the number of parameters and overall calculation, we neglect the volatility mean term in our model, which causes the problem that when we simulate the data, the volatility converges to zero very quickly after certain numbers of simulation. It should be noted

that even with the mean term, it just shift the trend but the degeneration dose not change. Therefore we are not able to simulate a long term dataset.

The density plot of log-returns of the data is shown in Figure 3.2. Firstly, both panels of the plot indicate heavy-tailed distributions. Second, we suspect that Cairn Energy shows a slight negative skewness and Anglo American seems more symmetric. We expect such information can be obtained from the estimation of the skewness coefficient vector of the skew-t distribution.

We are mainly interested in revealing the volatility changes over time given the time series data and showing the estimation of the unknown parameters autoregressive coefficient, A and skewness coefficient, α. The starting values for the parameters were set to be: the number of iterations $N = 500$, the number of time periods $T = 500$, the degrees of freedom of the inverse Wishart of the volatility $k = 5$, the degrees of freedom of the inverse Wishart of the autoregressive coefficient $v = 5$, the degrees of freedom of the skew-t distribution $df = 5$. We think 5 is a moderate choice for degree of freedom because it assumes the distribution neither has extreme heavy tails nor shows no evidence of heavy tails at all. The mean $\mu = (0.001, 0001)'$ is because the log-return of the data are very close to 0, the starting value of the scale matrix $\beta = \mathbf{I}$ and the starting value of the skewness coefficient vector$\alpha_{\mathbf{0}} = (0, 0)'$ is because this start value has little impact on the estimation. The R code can be found in the appendix document 'Wishart-naive.R'.

Figure 3.1: Stock prices of Anglo American and Cairn Energy

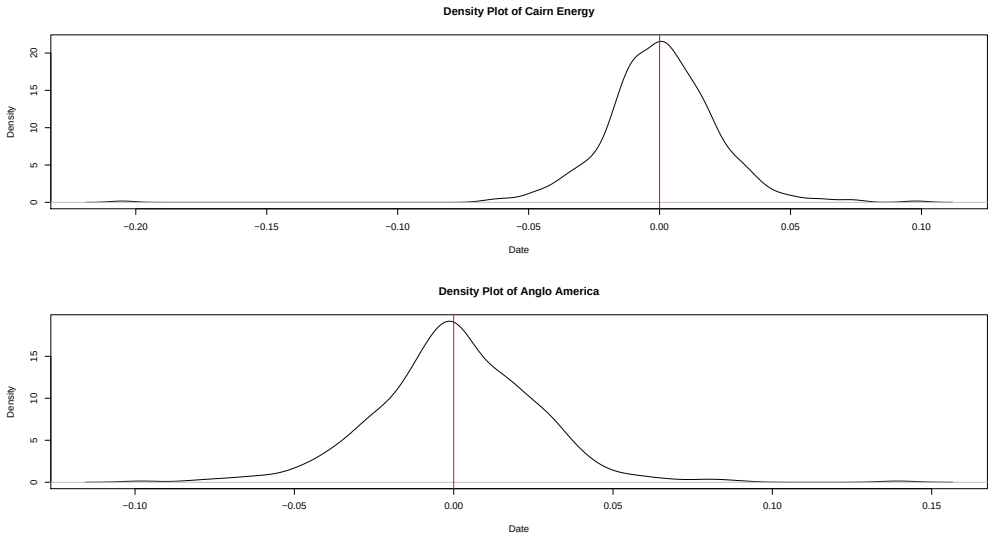

Figure 3.2: Density plot of log-returns of the data

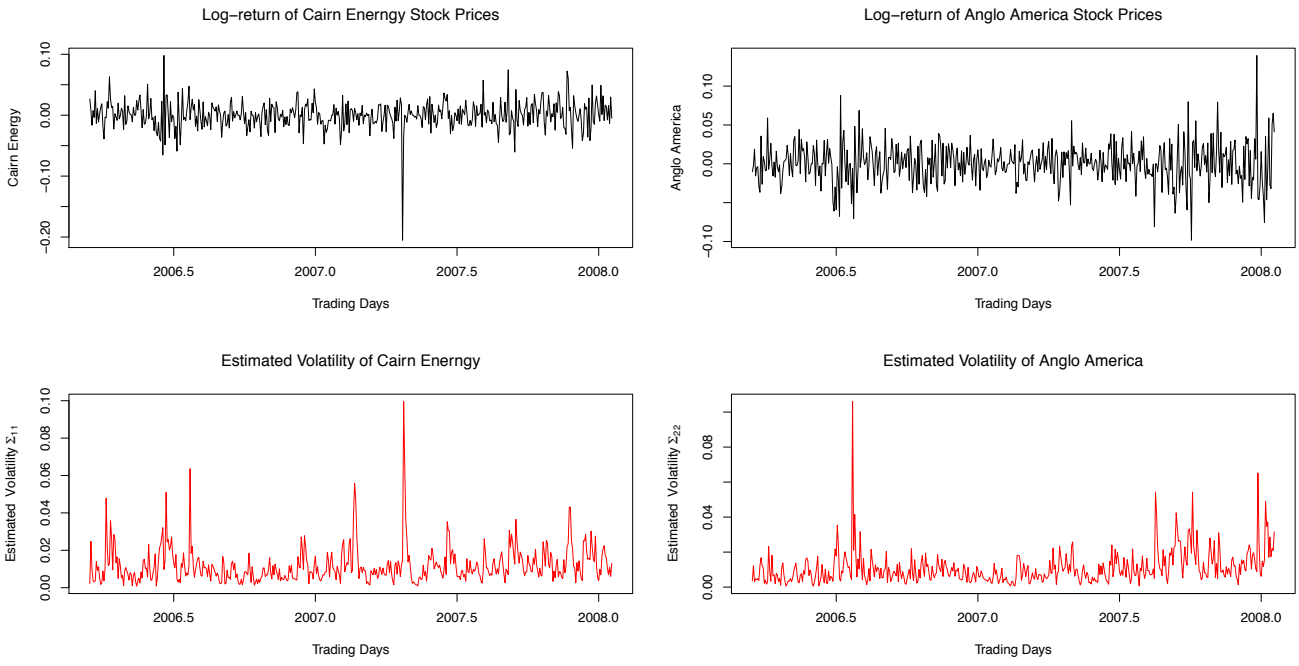

Figure 3.3: Log-return of the data vs. estimated volatility (Gaussian)

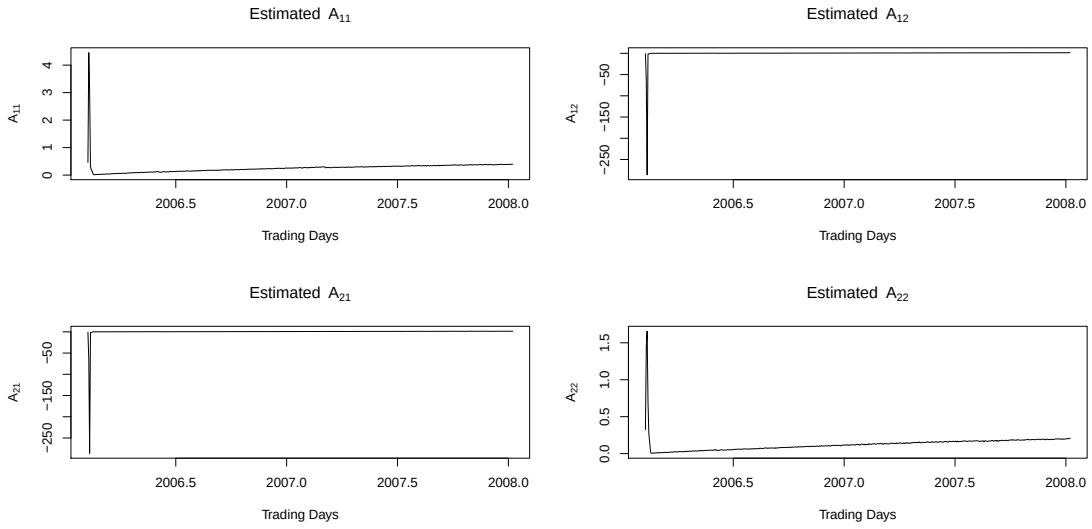

Figure 3.4: Estimation of autoregressive coefficient A

3.3.2 SMC on Gaussian model

We start from the naive set-up and the SMC method shown in algorithm 1 is used on the Gaussian Wishart autoregressive model. Figure 3.3 shows the log-returns of the data and the estimation of the volatility for each stock price. From the plot we see that the estimated volatility can capture the fluctuation pattern of the data. For example, the low value of the returns of Cairn Energy at 2007.3 is reflected by a high volatility estimation and the high fluctuation of Anglo American at 2008.0 is demonstrated well by the high volatility estimations in the end. However, we do notice that there might be an overreaction of the volatility estimation at around 2006.5. The peak of the estimation appears to be too high to interpret the data fluctuation at that time.

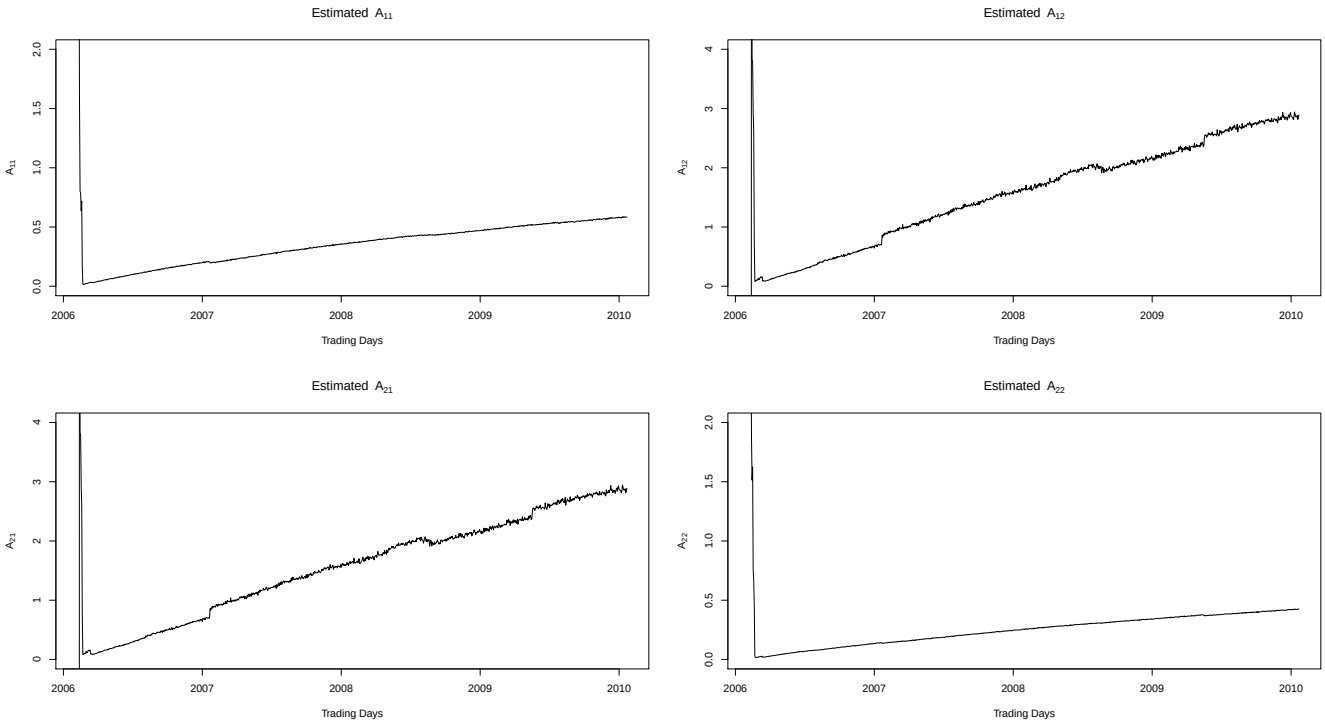

Figure 3.5: Estimation of A on 1000 Cairn Energy and Anglo America datapoints

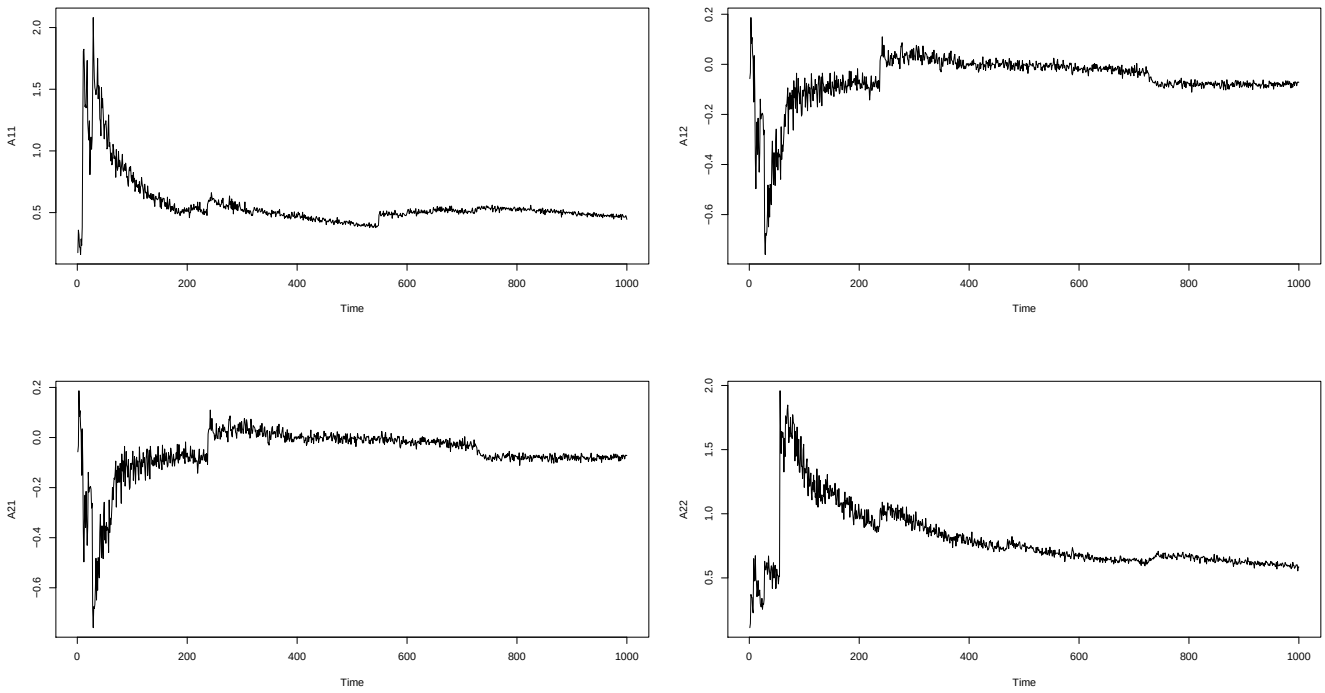

Figure 3.6: Estimation of A on 1000 simulated datapoints

The estimation for the unknown autoregressive coefficient A is shown in Figure 3.4. The four plots represent the four elements of matrix A respectively. We can see from the plots that after a short time at the beginning, the estimations start to stay much more stable. However there are some evidences that they don't converge, especially the A_{22} estimation. Also, if we look at the off-diagonal elements, they have influence on volatility matrix as well. A relatively large value on the off-diagonal elements can be a concern of the estimate since that could cause large eigenvalues of A and it could suggest that the volatility is not stationary. Then we fitted the model for much longer data(increased from 500 datapoint to 1000 datapoint, still Carin Energy and Anglo America data). The estimation of A shows in Figure 3.5 with a closer scale. We can see that same as before, after a short time at the beginning, the estimation start to stay stable but doesn't seem to converge. Then we are fitting the model to artificial data, which are simulated from RBEKK model(details see section 3.5.3). We have 1000 datapoint and the estimation of A shows in Figure 3.6. It is clear that after around 500 data, the estimation starts to converge with a noisy trend.

It is believed that the reasons that estimation of A does not seem to converge on Cairn Energy and Anglo America data are firstly, the log-return transformation makes the data much more smaller(close to 0) so that the algorithm takes much longer time and effort to adapt the information and make the estimate. Within 1000 points of time, the estimate of A is still unstable exhibiting a monotonic increase. However the artificial data has a

much larger scale, so that the estimation process adapts much quicker(takes around 500 data) and starts to converge after that. Secondly, in some cases, the behaviour of A might be as a time varying variable so that we should treat and estimate it in a time vary fashion. It is an interesting object that we could focus in the future and we will discuss in the future work section.

3.3.3 SMC on skew-t model

In order to further interpret the heavy-tailedness and the potential skewness of the given data, the SMC method shown in algorithm 2 is used on the skew-t Wishart autoregressive model. The R code can be found in appendix document 'Wishart-skewt.R'.

Figure 3.7 shows the log-returns of the data and the estimated volatility for the stock prices. The volatility can capture the fluctuation pattern of the data just fine. For example, it reflects the very low value of Cairn Energy at 2007.3 to have a high volatility estimation. Especially compared to Figure 3.3, the estimation of the high value of Anglo American at 2006.5 is an improvement. It reflects on the fluctuation of the returns but there is no clear evidence of overreaction this time.

The estimation for the unknown autoregressive coefficient $\mathbf{A}$ is shown in Figure 3.8. The four plots represent the four elements of matrix $\bar{\mathbf{A}}$ respectively. We can see from the plots that after a short time at the beginning, the estimations start to converge to certain values and stay stable afterwards.

The top plot of Figure 3.9 is the estimated skewness coefficient for Cairn

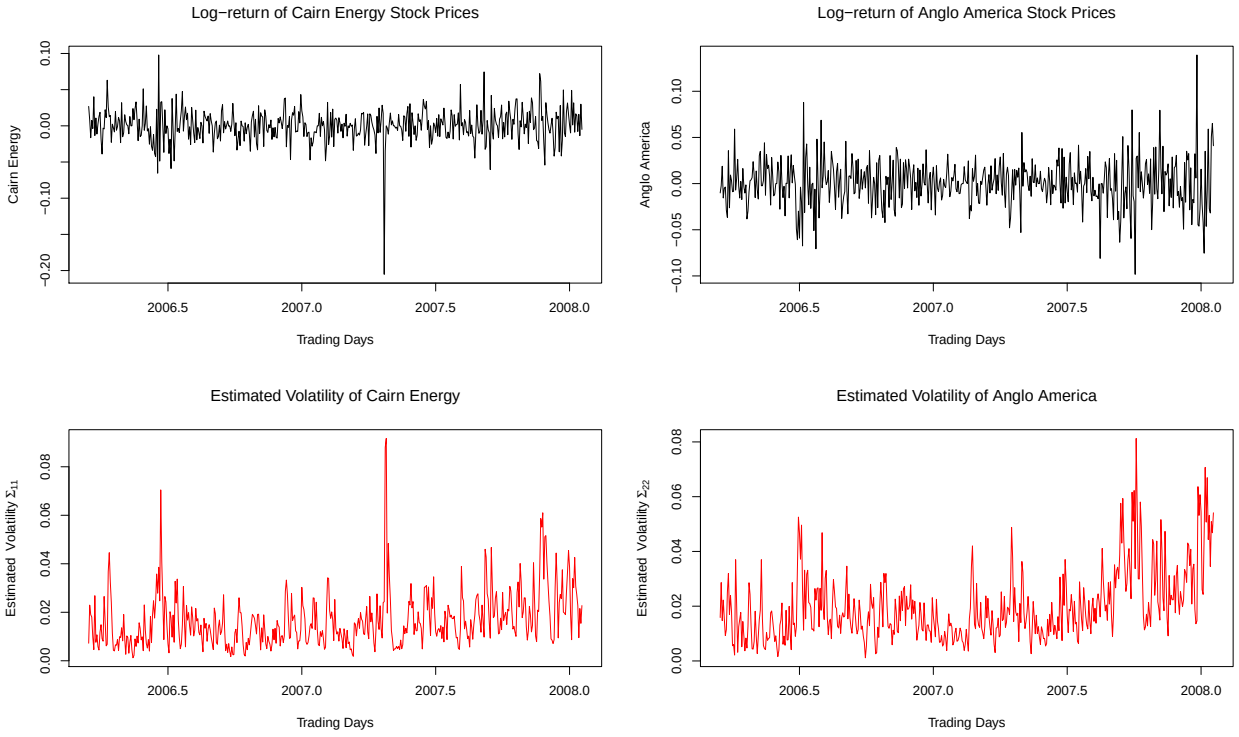

Figure 3.7: Log-return of the data vs. estimated volatility (skew-t)

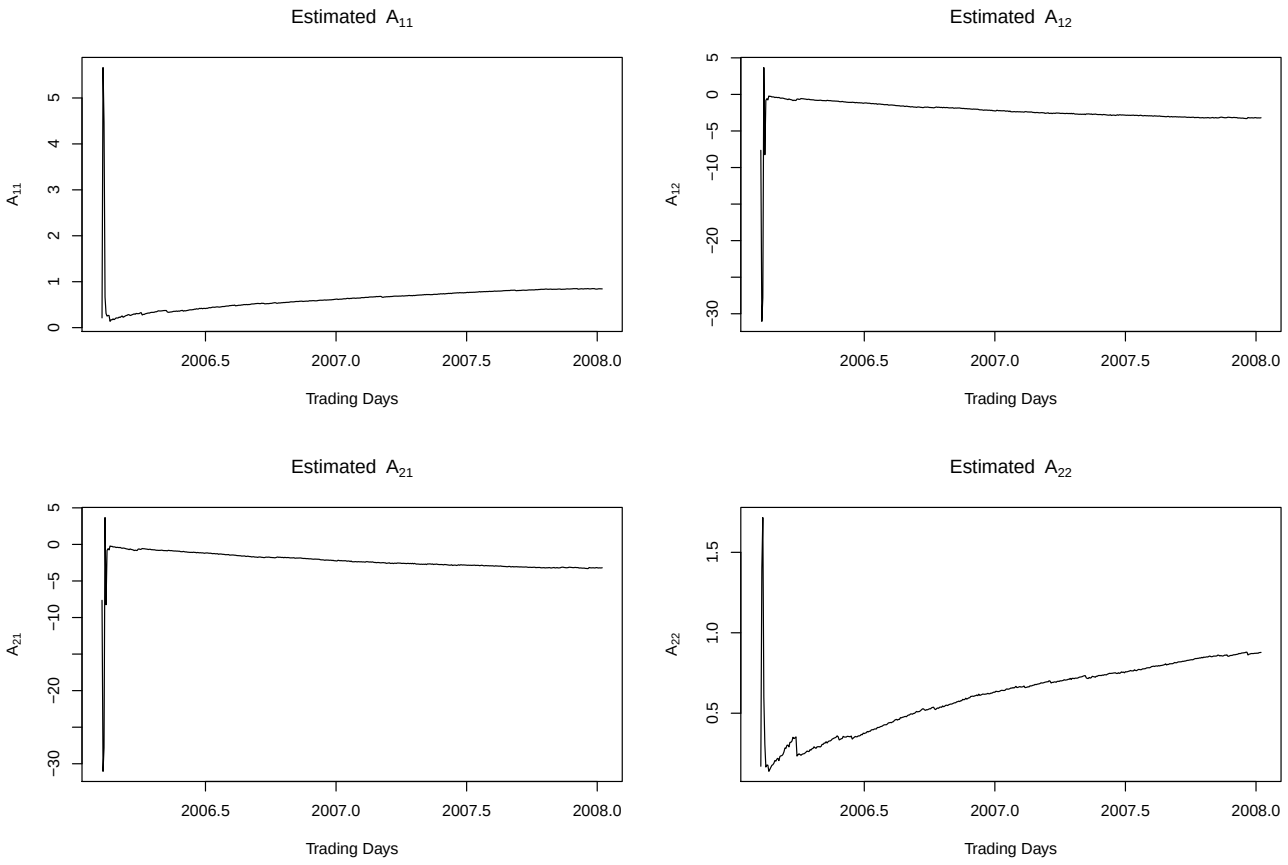

Figure 3.8: Estimations for autoregressive coefficient A

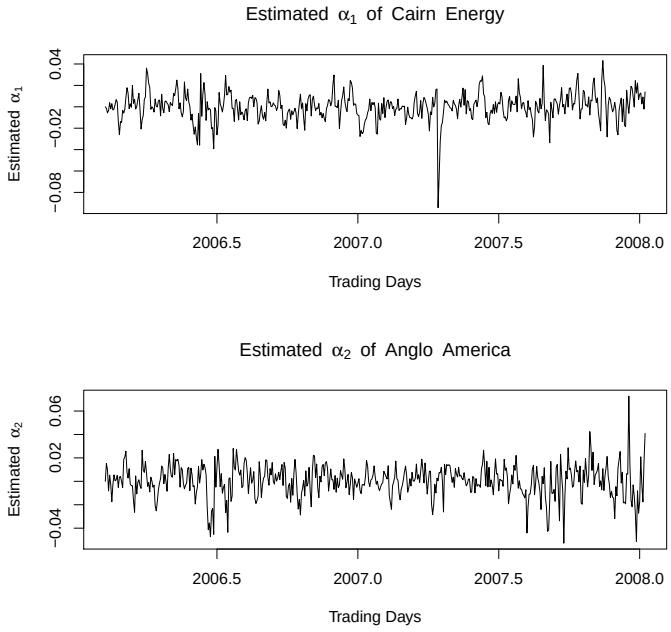

Figure 3.9: Estimations for skewness coefficient α

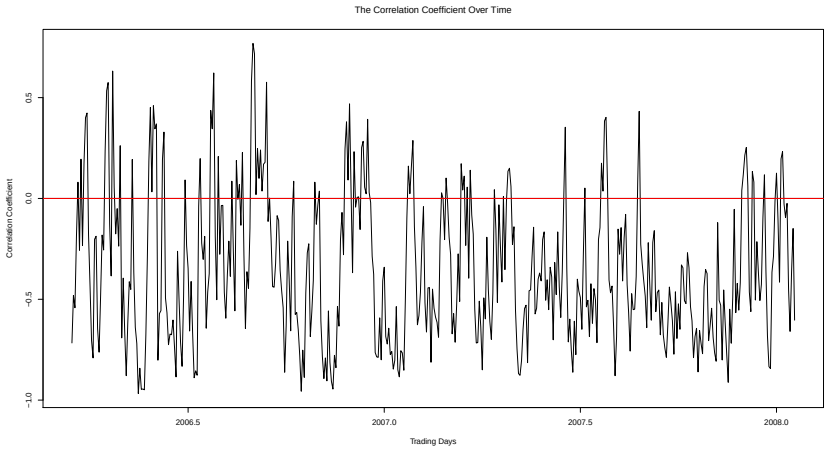

Figure 3.10: Estimated correlation coefficient

Energy and the bottom plot shows the estimated skewness coefficient for Anglo American. As we expected, the estimations of α change dramatically over time and should not be treated as a constant.

Figure 3.10 shows the estimated cross-correlation coefficient of the two returns. Notice that all the estimated values are ranged within the $[-1, 1]$ interval, which suits the properties of correlation coefficients. Also from the figure we can tell that there are more negative values than positive ones, which indicates the fact that the two stock prices are more often negatively related than staying in a positive relation. From the given two dimensional data, we can calculate the overall correlation coefficient is -0.448 and the mean of our estimated correlation coefficient is -0.349 which is close to the real value.

Figure 3.11 gives density plots for both the estimated volatility and the correlation coefficient at different random dates. The dates were chosen as 23 February 2007 for the top panel, 28 July 2006 for the middle panel and 19 May 2006 for the bottom panel. The figure shows that, first of all, none of the densities are normally distributed. Secondly, most only have one peak. Thirdly, most seem to have heavy tails. Then finally, most of the densities are believed to be asymmetric.

Overall, this subsection shows that the performance of the proposed models are good at reflecting the fluctuation of the data. Both the volatility and correlations can be estimated by the extended particle filter. The autoregressive coefficient matrix can be valued sequentially too. Due to the

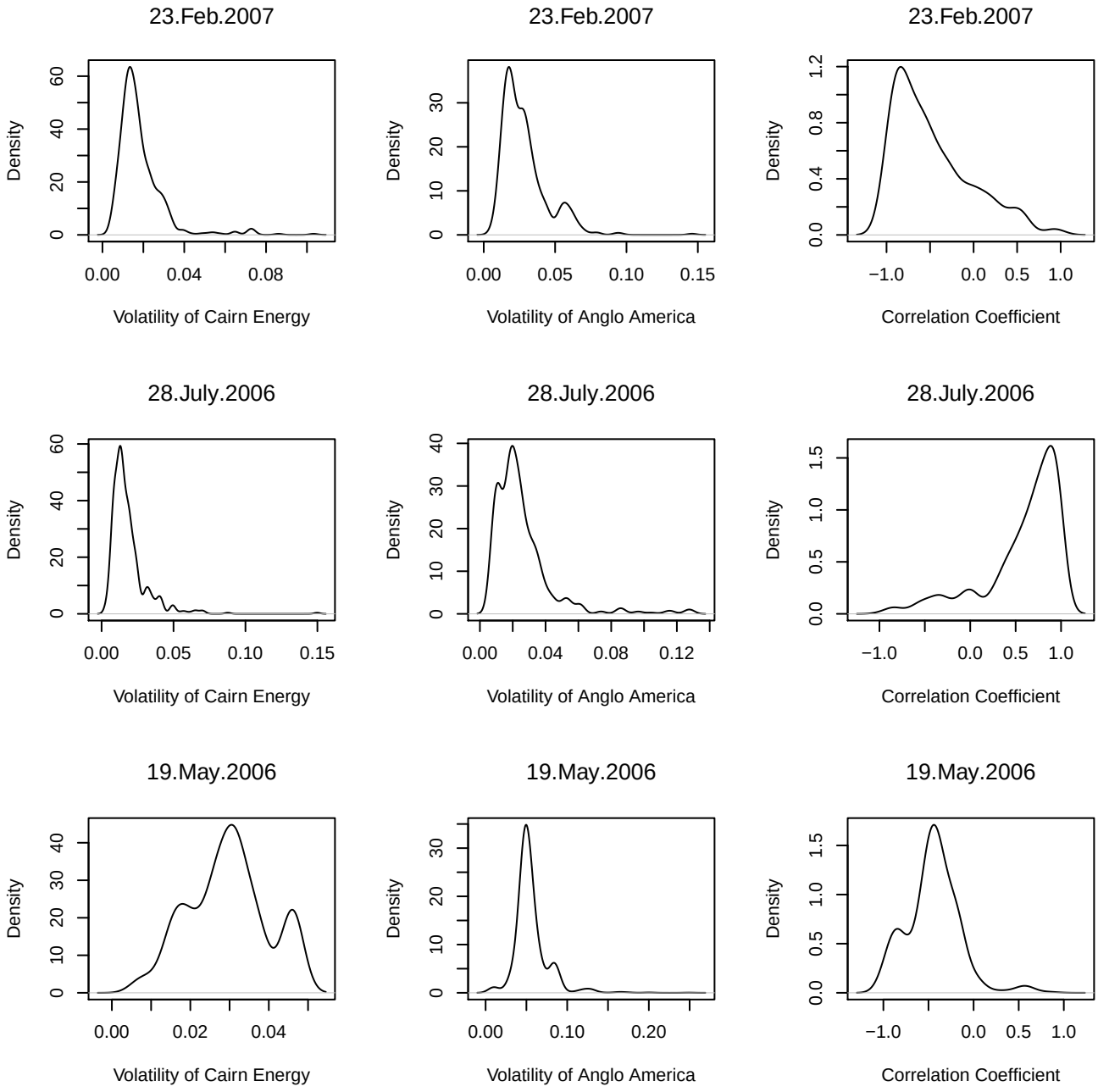

Figure 3.11: Density plots for covariance matrix at different dates

heavy-tailedness and skewness of the data, the skew-t autoregressive model shows a better performance over the Gaussian autoregressive model.

3.4 Gaussian Wishart autoregressive model with diagonal autoregressive coefficient

The Gaussian and skew-t Wishart autoregressive models in the last subsection have plausible performance over two-dimensional financial time series data. However, as the dimensionality of the data increases, the matrix computation involved in the estimation gets complicated, especially for the skew-t Wishart model. In order to deal with the curse of dimensionality, the proposed models need further parameter reduction, from which however a sacrifice of estimation accuracy is also expected.

We consider the volatility function in the Gaussian Wishart autoregressive model: $p(\Sigma_t^{-1}|\Sigma_{t-1}, A) = W(k, \frac{1}{k}\Sigma_{t-1}^{-\frac{1}{2}} A \Sigma_{t-1}^{-\frac{1}{2}})$. In the two dimensional case,

the conditional mean of the precision is as follows:

$$
\begin{aligned}
E(\Sigma_t^{-1}|\Sigma_{t-1}, A) &= \Sigma_{t-1}^{-\frac{1}{2}} A \Sigma_{t-1}^{-\frac{1}{2}} \\
&= \begin{bmatrix} x_1 & x_2 \\ x_2 & x_3 \end{bmatrix} \begin{bmatrix} a_1 & a_2 \\ a_2 & a_3 \end{bmatrix} \begin{bmatrix} x_1 & x_2 \\ x_2 & x_3 \end{bmatrix} \\
&= \begin{bmatrix} a_1x_1 + a_2x_2 & a_2x_1 + a_3x_2 \\ a_1x_2 + a_2x_3 & a_2x_2 + a_3x_3 \end{bmatrix} \begin{bmatrix} x_1 & x_2 \\ x_2 & x_3 \end{bmatrix} \\
&= \begin{bmatrix} a_1x_1^2 + a_2x_1x_2 + a_2x_1x_2 + a_3x_2^2 & a_1x_1x_2 + a_2x_2^2 + a_2x_1x_3 + a_3x_2x_3 \\ a_1x_1x_2 + a_2x_1x_3 + a_2x_2^2 + a_3x_2x_3 & a_1x_2^2 + a_2x_2x_3 + a_2x_2x_3 + a_3x_3^2 \end{bmatrix} \\
&= \begin{bmatrix} a_1x_1^2 + 2a_2x_1x_2 + a_3x_2^2 & a_1x_1x_2 + a_2(x_2^2x_1x_3) + a_3x_2x_3 \\ a_1x_1x_2 + a_2(x_1x_3 + x_2^2) + a_3x_2x_3 & a_1x_2^2 + 2a_2x_2x_3 + a_3x_3^2 \end{bmatrix}
\end{aligned} \tag{3.12}
$$

In order to simplify equation (3.12), we assume A to be a diagonal matrix, i.e. $a_2 = 0$. Then the conditional mean of the precision is:

$$
\begin{aligned}
E(\Sigma_t^{-1}|\Sigma_{t-1}, A) &= \Sigma_{t-1}^{-\frac{1}{2}} A \Sigma_{t-1}^{-\frac{1}{2}} \\
&= \begin{bmatrix} a_1x_1^2 + a_3x_2^2 & a_1x_1x_2 + a_3x_2x_3 \\ a_1x_1x_2 + a_3x_2x_3 & a_1x_2^2 + a_3x_3^2 \end{bmatrix}
\end{aligned} \tag{3.13}
$$

Compare equation (3.13) to equation (3.12): the parameter calculation is reduced by nearly a half. This will hugely decrease the complexity of the parameter estimation and the time consumption of the calculation. However, with less parameters describing the volatility, the accuracy of the estimation

will be slightly compromised.

Based on the Gaussian and skew-t autoregressive Wishart model, we propose a simplified model, which assumes A to be a diagonal matrix:

$$\begin{aligned} \Sigma_t|\Sigma_{t-1}, A &\sim IW(k, (k-n-1)A\Sigma_{t-1}A') \\ a_{ii} &\sim Gamma(c_i, d_i), \quad i = 1, \cdots, n \\ y_t|\Sigma_t &\sim N(\mu, \Sigma_t) \ \ or \ \ St(\mu, \Sigma, \alpha, v), \end{aligned} \tag{3.14}$$

where a_{ii} is the i-th diagonal element of A, n is the dimension of the data and $Gamma(c_i, d_i)$ is the prior density for the diagonal elements. IW is an inverse Wishart distribution, for details see Appendix.

It should be noted that, firstly, we focus on the volatility matrix instead of its precision here, therefore it follows an inverse Wishart distribution. Secondly, compared to the previous Gaussian Wishart model (equation (3.6)), the scale matrix has changed from $\Sigma_{t-1}^{-\frac{1}{2}} A \Sigma_{t-1}^{-\frac{1}{2}}$ to $A\Sigma_{t-1}A'$. The change of positions between Σ and A is because $A\Sigma_{t-1}A'$ is a natural autoregressive structure. $\Sigma_{t-1}^{-\frac{1}{2}} A \Sigma_{t-1}^{-\frac{1}{2}}$ was used before for the purpose of obtaining a close-formed optimal importance function.

The proposed model inherits the autoregressive Wishart set up from the previous model (equation (3.6)) and the data can be demonstrated by either a Gaussian distribution or a skew-t distribution. We also assume the prior distribution for the diagonal elements of A to follow a gamma distribution, which is a conjugate distribution so that Bayesian inference can be applied

to derive the posterior distribution of a_{ii}. As before, SMC will be used to estimate the volatility and the autoregressive coefficient A.

3.4.1 Estimation of A

Firstly, we introduce a lemma (for details see Johnson et al. (2002)):

$$\begin{aligned} If \;\; X &\sim IW(k, S), \\ then \;\; x_{ii} &\sim IG(\frac{k-2n}{2}, \frac{s_{ii}}{2}), \end{aligned}$$

where x_{ii} is the i-th diagonal elements of matrix X. IW is the inverse Wishart distribution and IG is the inverse gamma distribution.

Recall the volatility equation (3.14), where the conditional volatility follows an inverse Wishart distribution, then we apply the lemma and have:

$$x_{ii(t)}|a_{ii} \sim IG\left(\frac{k-2n}{2}, \frac{(k-n-1)a_{ii}^2 x_{ii(t-1)}}{2}\right), \tag{3.15}$$

where $x_{ii(t)}$ is the i-th diagonal element of the volatility matrix Σ_t at time t.

It should be noted that both diagonal elements and off-diagonal elements of covariance matrix give information on a_{ii}, however in order to reduce the calculation and simplify the estimation process, we think it is better to neglect the off-diagonal element and focus on the diagonal element. With the density function of $x_{ii(t)}|a_{ii}, x_{ii(t-1)}$, and a prior for a_{ii}, then using the conjugate property of the inverse gamma distribution, we can apply the Bayesian

rule and derive the posterior distribution of $a_{ii}|x_{ii(t)}, x_{ii(t-1)}$:

$$
\begin{aligned}
P(a_{ii}|x_{ii(t)}, x_{ii(t-1)}) &= \frac{P(x_{ii(t)}|x_{ii(t-1)}, a_{ii})P(a_{ii}|x_{ii(t-1)})}{P(x_{ii(t)}|x_{ii(t-1)})} \\
&\propto P(x_{ii(t)}|a_{ii}, x_{ii(t-1)})P(a_{ii}|x_{ii(t-1)}) \\
&\propto IG(\frac{k-2n}{2}, \frac{(k-n-1)a_{ii}^2 x_{ii(t-1)}}{2}) \cdot \\
&\quad Gamma(c_{t-1}, d_{t-1}) \\
&\propto \frac{(\frac{(k-n-1)a_{ii}^2 x_{ii(t-1)}}{2})^{\frac{k-2n}{2}}}{\Gamma(\frac{k-2n}{2})} x_{ii(t)}^{-\frac{k-2n}{2}-1} e^{-\frac{(k-n-1)a_{ii}^2 x_{ii(t-1)}}{2} / x_{ii(t)}} \\
&\quad \frac{1}{\Gamma(c_{t-1}) d_{t-1}^{c_{t-1}}} a_{ii}^{c_{t-1}-1} e^{-\frac{a_{ii}}{d_{t-1}}} \\
&\propto a_{ii}^{\frac{k-2n}{2}+c_{t-1}-1} e^{a_{ii}(\frac{1}{d_{t-1}} + \frac{(k-n-1)x_{ii(t-1)}}{2x_{ii(t)}})} \\
&\propto Gamma(\frac{k-2n}{2} + c_{t-1}, \frac{1}{\frac{1}{d_{t-1}} + \frac{(k-n-1)x_{ii(t-1)}}{2x_{ii(t)}}}) \\
&\propto Gamma(c_t, d_t) \qquad (3.16)
\end{aligned}
$$

Algorithm 3 shows the estimation process of SMC on the Wishart autoregressive model with diagonal A. At each time point, the posterior distribution of the diagonal elements a_{ii} and the skewness coefficient α can be estimated using past information and observations. Then we calculate the particle weights with the help of the estimated unknown parameters, and finally estimate the volatility matrix.

Algorithm 3 Algorithm of SMC for model with diagonal A

1: At $t = 0$, Σ_0 and A can be simulated from their prior distributions:

$$p(\hat{a}_{ii(0)}) = Gamma(c_0, d_0) \qquad p(\Sigma_0) = IW(k, B)$$

2: At t=1, Σ_1 can be simulated from the inverse Wishart distribution using the simulated parameters from time 0 and A_1 can also be estimated:

$$p(\Sigma_1|\Sigma_0, A) = IW(k, (k-n-1)A_0\Sigma_0 A) \quad and \quad p(\hat{a}_{11}|\Sigma_1, \Sigma_0) \propto Gamma(c_1, d_1)$$

3: **for** $t = 2, \ldots, T$ **do**

4: $\hat{a}_{ii(t)}$ is updated at every time point t as:

$$p(\hat{a}_{ii(t)}|\Sigma_t, \Sigma_{t-1}) \quad \propto \quad Gamma(c_t, d_t)$$

where $c_t = \frac{k-2n}{2} + c_{t-1}$ and $d_t = \frac{1}{\frac{1}{d_{t-1}} + \frac{(k-n-1)x_{ii(t-1)}}{2x_{ii(t)}}}$

After obtaining the unknown parameter A at time t, the SMC method can be applied as follows:

5: **for** $i = 1, \ldots, N$ **do**

6: Draw Σ_t^i from $p_{t|t-1}(\Sigma_t^i|\Sigma_{t-1}, y_{t-1}, A)$.

7: The skewness coefficient α is updated as:

$$\hat{\alpha}_t^i = \lambda\hat{\alpha}_{t-1}^i + (1-\lambda)y_t$$

8: Set

$$\hat{w}_t^i = \frac{p(y_t|\Sigma_t, \hat{\alpha}_t^i)p(\Sigma_t^i|\Sigma_{t-1}, A)}{g(\Sigma_t^i|\Sigma_{t-1}, y_t, A)}\hat{w}_{t-1}^i$$

9: Normalise the weights:

$$w_t^i = \frac{\hat{w}_t^i}{\sum_{j=1}^N \hat{w}_t^j}$$

10: **end for**

11: Compute:

$$\hat{\Sigma}_t = \sum_{i=1}^N w_t^i \delta_{\Sigma_t^i}$$

where δ_Σ denotes the Dirac delta mass located at Σ_t.

12: **end for**

3.5 Application

3.5.1 Data analysis

In this section, the proposed diagonal autoregressive coefficient model will be applied to the data using the SMC method. Compared to the former sections, the data used in this section are higher 6-dimensional currency exchange rate data. The data consist of 700 data points of the currency exchange rate between Canadian dollar (CAD) and US dollar (USD), Japanese yen (JPY) and USD, Euro (EUR) and USD, Australian dollar (AUD) and USD, pound sterling (GBP) and USD and New Zealand dollar (NZD) and USD, from 2008 to 2012, which is collected at daily frequency. The dataset can be found in appendix document 'Currency exchange rate data.txt'.

The time series plots of these data are shown in Figure 3.12. The plot shows that most of the currency exchange rate are correlated. For example the CAD and EUR share a similar pattern, especially after 400 time points. We will explore more of their correlations after the estimation. Secondly, the time series are not stationary, therefore the log-return of the data was used. The time series plots of the log-return are shown in Figure 3.13. Finally, we investigate the normality and skewness of the data. The density plot log-return and the quantile-quantile (Q-Q) plot of the log-return of the data are shown in Figures 3.14 and 3.15, respectively. The density plots suggest the data has skewness, for example JPY and AUD. The Q-Q plot indicates that most of the data are not normally distributed, especially JPY, AUD and

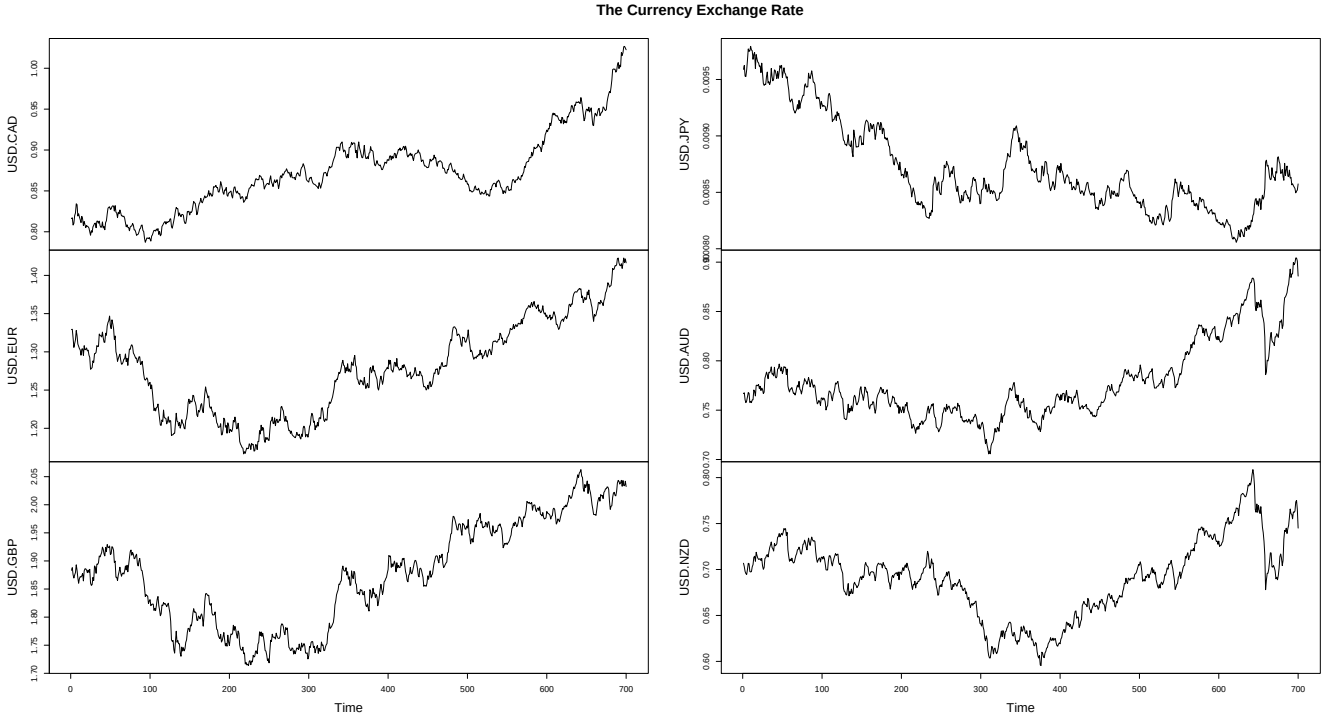

Figure 3.12: Currency exchange rate data

NZD. Therefore, the skew-t distribution will be used to describe the data.

The starting values for the parameters were set to be: the number of iterations $N = 500$, the number of time periods $T = 700$, the degrees of freedom of the inverse Wishart of the volatility $k = 5$, the degrees of freedom of the inverse Wishart of the autoregressive coefficient $v = 5$, the mean $\mu = (0.001, 0001)$, the degrees of freedom of the skew-t distribution $df = 5$, the starting value of the scale matrix $\beta = \mathbf{I}$, the starting value of the shape parameter of the gamma distribution $c_0 = 10$, the starting value of the scale parameter of the gamma distribution $d_0 = 0.1$ and the starting value of the skewness coefficient $\alpha_{\mathbf{0}} = (0, 0)$. The R code can be found in appendix document 'Wishart-diagonal.R'.

3.5.2 Estimation

We apply algorithm 3 of the SMC method on the diagonal autoregressive coefficient Wishart autoregressive model given the currency exchange rate data.

Figure 3.16 shows the log-return of the data on the top panel and the estimated volatility on the bottom. The estimated volatility follows a similar pattern as the log-return data. For example the log-return of CAD indicates a high volatility at around 80 time points at the beginning and at around 380 time points at the end. The estimated volatility captures the high fluctuation of the data and shows high value in the corresponding positions.

The estimation of the diagonal autoregressive coefficient is illustrated

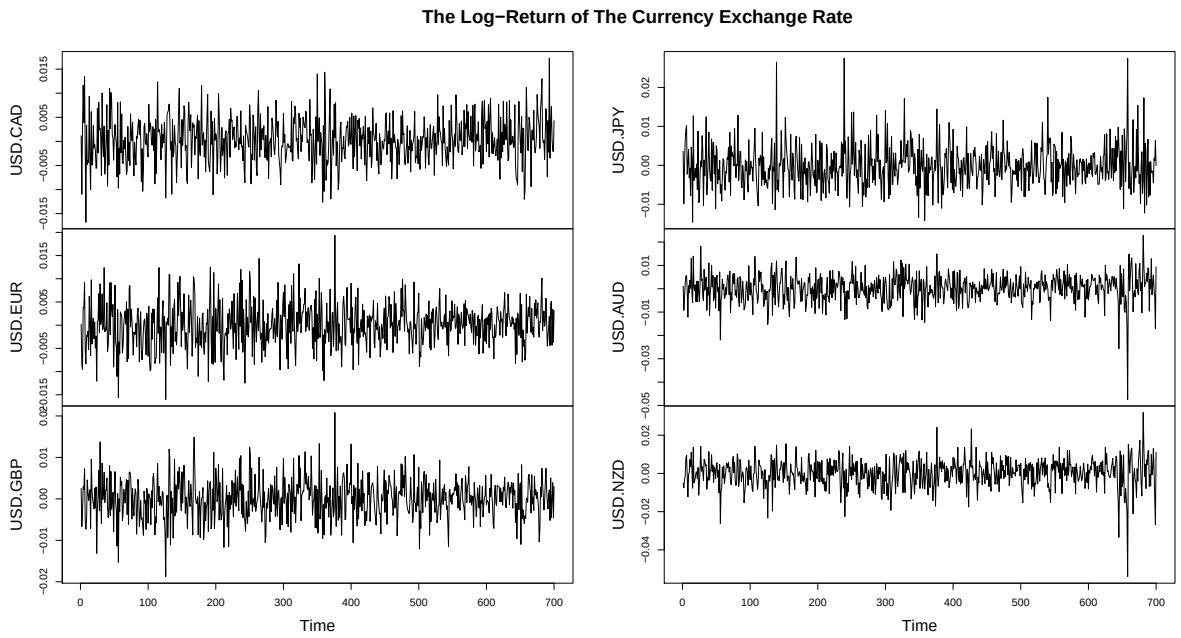

Figure 3.13: Log-return of exchange rate data

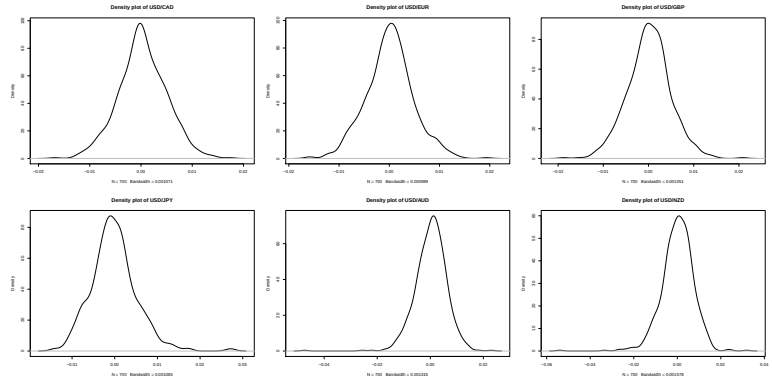

Figure 3.14: Density plots of returns for each currency

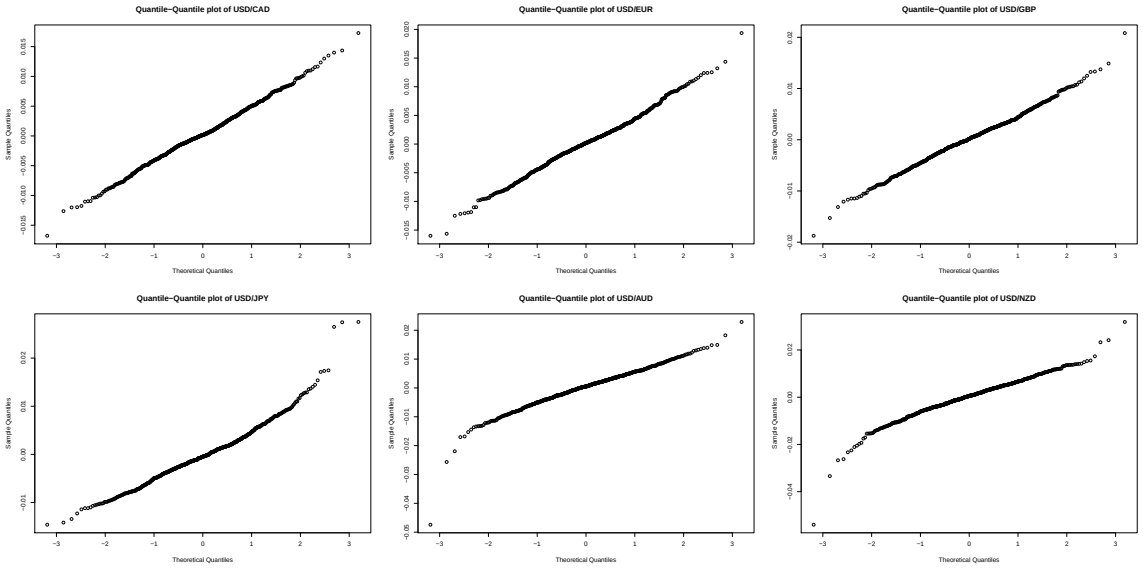

Figure 3.15: QQ-plot of returns for each currency

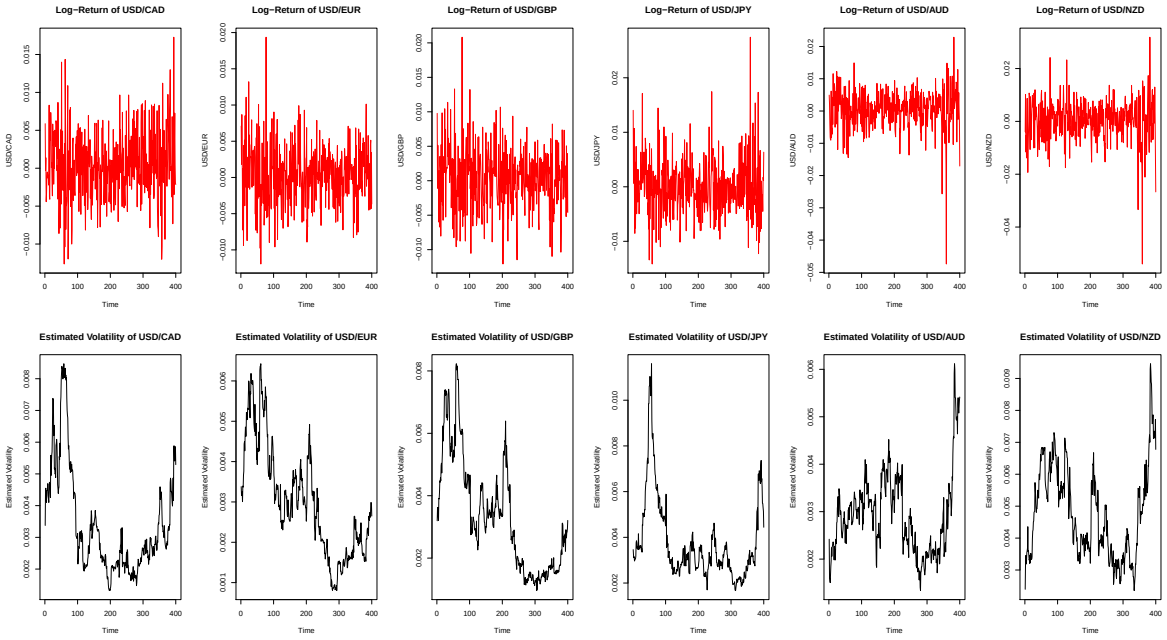

Figure 3.16: Log-return vs. estimated volatility

in Figure 3.17. The coefficient is a 6-dimensional diagonal matrix in this case, therefore the 6 diagonal elements are shown. The estimations have a decreasing trend, but it should be noted that the drop occurs over such a small scale (from $0.9986 - 0.9974$). It could mean that the process is still adapting in a slow pace. We have discussion about the monotonic behaviour of A in earlier section (section(3.3.2)).

The estimation of the volatility matrix allows us to look into the correlation between each currency exchange rate. For example, the correlation coefficient between CAD and EUR is shown in Figure 3.18. We can see that all the estimated values are ranged within the $[-1, 1]$ interval, which meets the requirements for a correlation coefficient. Also there are more positive values than negative values, which indicates the fact that the two currency exchange rates are more often positively related than negatively related.

Overall, this subsection demonstrates the performance of the estimation of the SMC on the diagonal autoregressive coefficient model. The volatility estimation is good at reflecting the fluctuation of the data, the autoregressive coefficient estimation is solid and the correlations between each currency exchange rate reveals extra information.

3.5.3 Comparison with multivariate GARCH

This subsection is joint work with Farhat Iqbal (Department of Statistics, University of Balochistan).

The comparison is between the Wishart autoregressive model with diago-

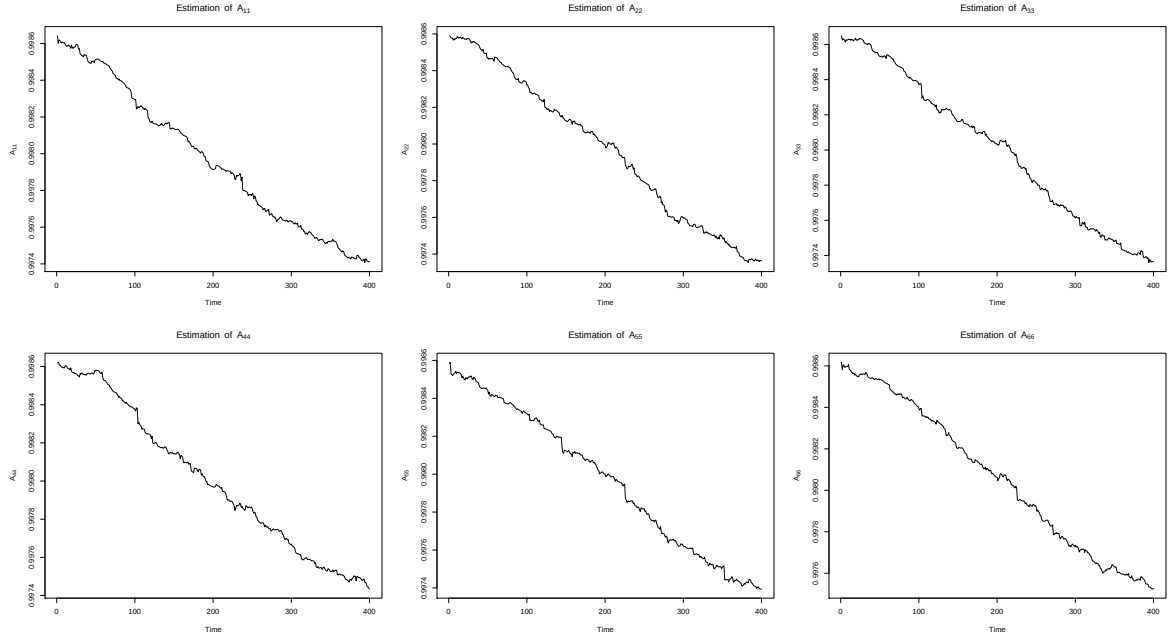

Figure 3.17: Estimation for a_{ii}

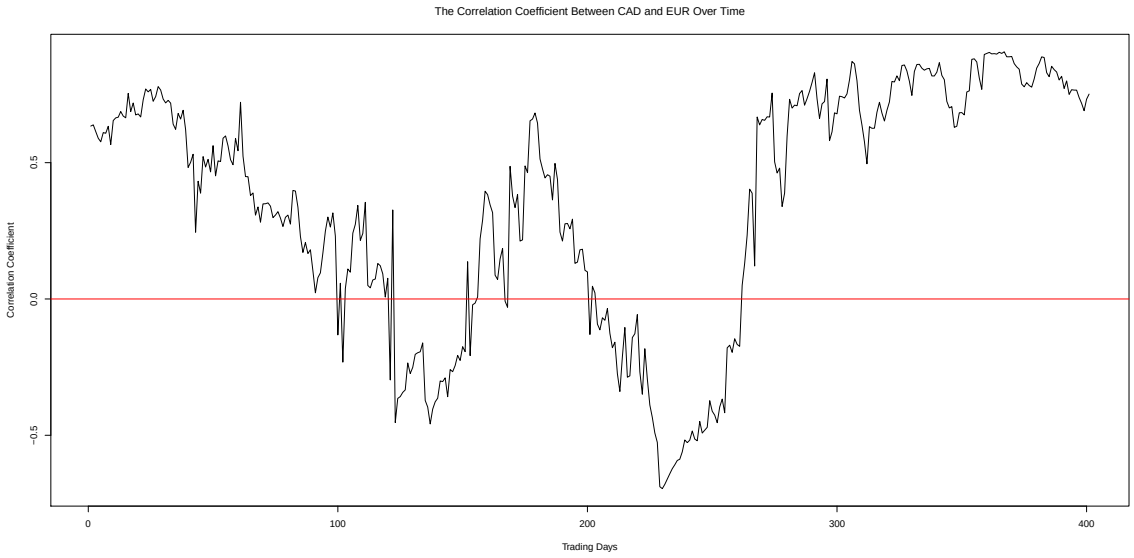

Figure 3.18: Estimated correlation coefficient between CAD and EUR

	Wishart-RMSE	RBEKK-RMSE	Wishart-MAE	RBEKK-MAE
Mean	0.36	0.14	0.32	0.13
Variance	0.2	0.023	0.16	0.020

Table 3.1: Mean and variance of RMSE and MAE between simulated and estimated covariance matrix of Wishart autoregressive model and RBEKK model

nal autoregressive coefficient (section 3.5.1) and the multivariate generalised autoregressive conditional heteroscedastic model (MGARCH), more specifically, the rotated Baba-Engle-Kraft-Kroner model (RBEKK) – see Iqbal and Triantafyllopoulos (2016), Engle and Kroner (1995) and Noureldin et al. (2014), more details see Introduction. The RBEKK model is an extension of the MGARCH model for medium dimensional data by exploiting returns rotation and covariance targeting. The estimation method of RBEKK is carried out by MCMC for reasons of extra accuracy and stability.

The data are simulated from RBEKK model, the simulated return and volatility can be found in appendix document 'Simulated return.txt' and 'Simulated Volatility.txt'. We cannot simulate data from Wishart model is because the main structure of our model is set to be an autoregressive structure. However in order to push the model to higher dimensions, reduce the number of parameters and overall calculation, we neglect the volatility mean term in our model, which causes the problem that when we simulate the data, the volatility converges to zero very quickly after certain numbers of simulation. It should be noted that even with the mean term, it just shift the trend but the degeneration dose not change. Therefore we are not able

to simulate a long term dataset.

The simulated data are used for both models and the covariance matrix are estimated by SMC and MCMC for the Wishart autoregressive model and RBEKK model, respectively. In order to obtain the distance between the simulated variance matrix and estimated variance matrix, mean absolute error (MAE) and root mean squared error (RMSE) are used to calculate the matrix norm. The definitions of MAE and RMSE are as follows:

$$\begin{aligned} MAE &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}|e_{ij}| \\ RMSE &= \sqrt{\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}|e_{ij}|^2} \\ e_{ij} &= (a_{ij}-b_{ij}), \end{aligned} \tag{3.17}$$

where a_{ij} and b_{ij} are the ijth element of matrix A and B.

In order to make the comparison more stable and accurate, we use a Monte Carlo approach, i.e. we simulated 100 sets of bivariate data with 1000 data points each using a standard RBEKK model, details see Appendix. Because of the convergence of MCMC and SMC, we neglect the first 300 data points and only count the last 700. Then both models were applied to the simulated data and both MAE and RMSE between simulated and estimated covariance matrix calculated. The comparison R code can be found in appendix document 'Comparison-Wishart and REBBEK.R'.

The mean and variance of MAEs and RMSEs of 100 data sets for both

models are shown in Table 5.2. From the table, we can see that the mean and variance of both MAEs and RMSEs of the RBEKK model are smaller than for the Wishart model. However, the differences are not significant. The RBEKK model outperformed the Wishart autoregressive model by a small margin. Figure 3.19 shows the first set of simulated data vs. simulated volatility and estimated volatility of both models. The RBEKK has close estimations to the simulations; the Wishart autoregressive presents less close but still acceptable estimations.

Overall, taken into account that firstly the data are simulated from RBEKK model, secondly the GARCH family models are much more well developed, thirdly the well-known accurate and stable MCMC was performed on RBEKK model, the Wishart autoregressive model with online SMC estimation method achieves an inferior but close enough performance on the simulated data.

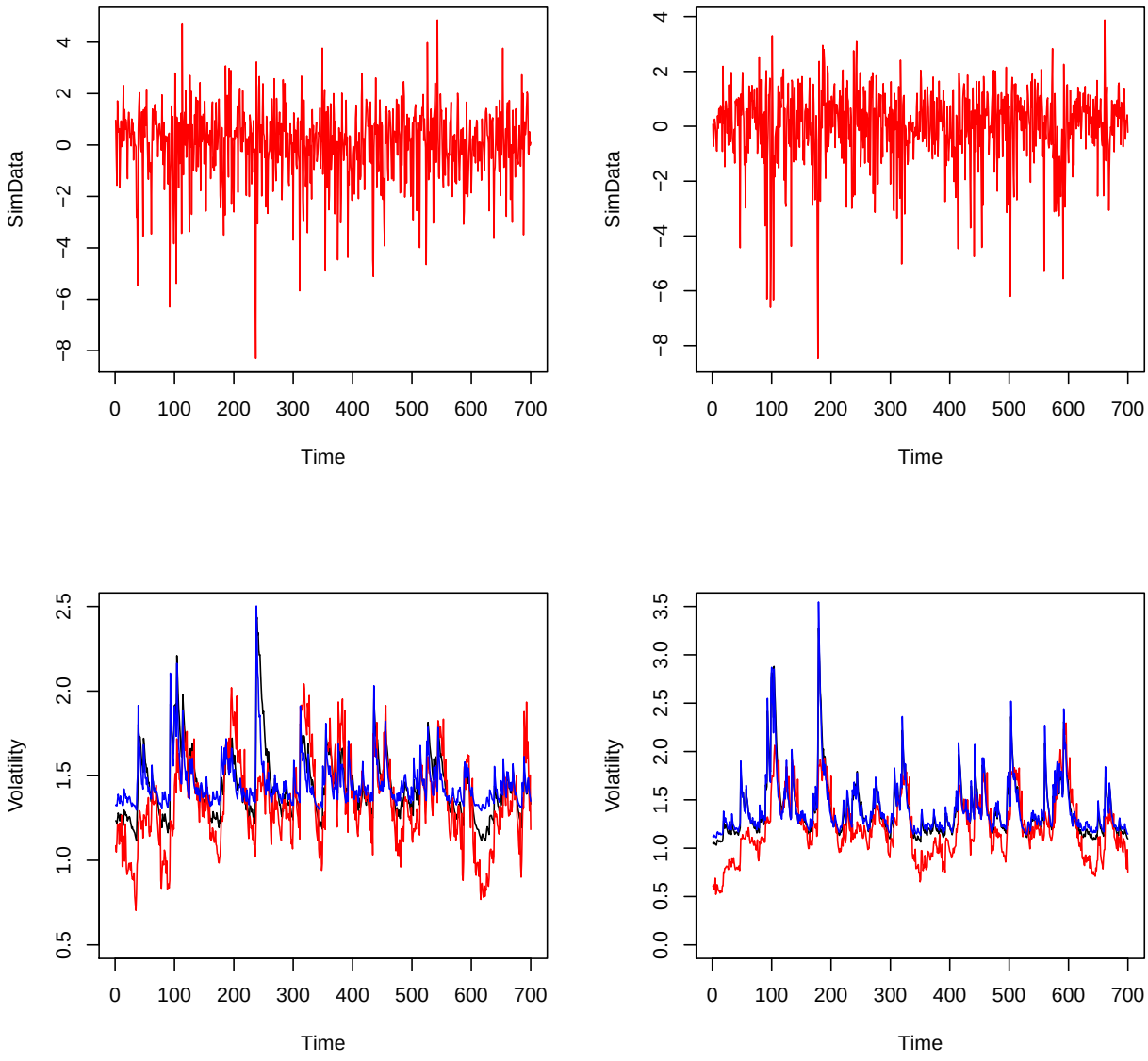

Figure 3.19: Simulated data and simulated volatility (black) vs. Wishart (red) vs. RBEKK (blue)

Chapter 4

Wishart-Newton Model

4.1 Introduction

Inspired by Uhlig's (Uhlig (1997)) Wishart autoregressive process, a Wishart-Newton model is developed. The hyperparameter is treated as a diagonal matrix and is estimated by the Newton-Raphson method. We manage to integrate parameter estimation into an online volatility estimation algorithm and the application is shown at the end of the chapter.

4.2 Model definition

In the same way as for the Gaussian Wishart autoregressive model (see Section 3.2.1), in this chapter we consider the p-dimensional time series data $\{y_t\}$, which is the log-return of the prices of any financial instrument. The

data can be described by a multivariate normal distribution with a mean vector μ and a p-dimensional covariance matrix Σ_t:

$$y_t|\Sigma_t \sim N(\mu, \Sigma_t)$$

Our main interest is to estimate the covariance matrix Σ_t under the assumption that its $p \times p$ precision matrix $\Phi_t = \Sigma_t^{-1}$ exists, i.e. Σ_t is a strictly positive defined matrix. First, Uhlig's (Uhlig, 1997) proposed Wishart process as follows:

$$\Phi_t = kU(\Phi_{t-1})'BU(\Phi_{t-1})$$

However the process has a random walk structure which results the estimation go wild sometimes.Therefore we decided to add another parameter matrix A in order to restrict the behaviour of the model and makes the model follows a more stable autoregressive structure, the precision matrix can be described as follows:

$$\begin{aligned} \Phi_t &= kAU(\Phi_{t-1})'BU(\Phi_{t-1})A' \\ B &\sim Beta_p(a/2, b/2), \end{aligned} \tag{4.1}$$

where k is a constant to keep the autoregressive property of the model – it needs to be determined later on. A is the autoregressive parameter matrix which we will estimate. It should be noted that in our model, we treat autoregressive coeffcient A as the main parameter. It represents the relationship

between current volatility state and its past. Since we treat A as a diagonal matrix, the diagonal elements have major impact on the model. We would expect the diagonal elements have range from 0 to 1. The closer the diagonal elements are close to 1, the more dependence between current volatility with its past state. $U(\Phi_{t-1})$ is the upper triangular matrix of the Choleski decompostion of the precision matrix Φ_{t-1}. B, assumed to be independent from Φ_{t-1}, follows a singular multivariate beta distribution with two positive shape parameters $a/2$ and $b/2$, both of which will be decided later.

If we look at the expectation of the precision matrix:

$$\begin{aligned} E(\Phi_t|\Phi_{t-1}) &= kAE(U(\Phi_{t-1})'BU(\Phi_{t-1}))A' \\ &= k\frac{a}{a+b}A(U(\Phi_{t-1})'U(\Phi_{t-1}))A' \\ &= k\frac{a}{a+b}A\Phi_{t-1}A' \quad , \end{aligned}$$

the expectation has the same structure as a autoregressive model of order one, which ensures the stationarity and the accuracy of the volatility estimation.

4.2.1 Inference conditional on A

Although the expectation of the volatility suits our purpose, the proposed model in equation (4.1) is still too complicated to estimate, especially with the upper triangular matrix structure and the singular beta distribution. (It should be noted that in this subsection, we consider the autoregressive coefficient A to be known; the estimation of A will be discussed in the next

subsection.)

We shall exploit the conjugacy of the multivariate beta and Wishart distributions to facilitate computationally efficient inference; here we introduce a lemma: If we have $\Phi \sim W_p(a+b, F)$, where $a > p+1$ and b is a positive integer, and $B \sim B_p(a/2, b/2)$, where B is independent from Φ, then we have:

$$\Phi^* = U(\Phi)' B U(\Phi) \sim W_p(a, F) \tag{4.2}$$

for details of the lemma, see Uhlig (1997). The lemma combines the multiplication of the upper triangular matrix of the Wishart distribution and the singular beta distribution into a Wishart distribution which helps us greatly with model simplification and calculation.

With the help of the lemma, we can work on the details of the model parameters. Assume the posterior distribution for the precision volatility matrix at $t-1$ is $\Phi_{t-1}|A, D_{t-1} \sim W_p(n+p-1, F_{t-1})$, where $D_t = (y_1, ..., y_t)$ denotes the available information and the data at time t and $n > 0$. Here we set $n = \frac{1}{1-\delta}$, where $0.67 < \delta < 1$ is a discount factor to make sure that $n > 3$, i.e. the degrees of freedom of the Wishart distribution as:

$$n+p-1 = \frac{1}{1-\delta} + p - 1 = (\frac{1}{1-\delta} - 1 + p - 1) + 1 = (\frac{\delta}{1-\delta} + p - 1) + 1$$

Then we specify $a = \frac{\delta}{1-\delta} + p - 1$ and $b = 1$ and apply the lemma, the prior

for Φ_t given A and D_{t-1} can be derived as:

$$\begin{aligned} k^{-1}A^{-1}\Phi_t A'^{-1}|A, D_{t-1} &\sim W_p(a, F_{t-1}) \\ \Phi_t &\sim W_p(a, kAF_{t-1}A') \\ &= W_p(\frac{\delta}{1-\delta} + p - 1, kAF_{t-1}A') \\ &= W_p(\delta n + p - 1, kAF_{t-1}A') \end{aligned} \tag{4.3}$$

Since the model is set up to be an autoregressive model, therefore the autoregressive property need to be satisfied by: $E(\Phi_t|A, D_{t-1}) = AE(\Phi_{t-1}|A, D_{t-1})A'$. From the posterior distribution of Φ_{t-1} and equation (4.3) we have:

$$\begin{aligned} E(\Phi_t|A, D_{t-1}) &= (\delta n + p - 1)kAF_{t-1}A' \\ AE(\Phi_{t-1}|A, D_{t-1})A' &= (n + p - 1)AF_{t-1}A' \\ (\delta n + p - 1)kAF_{t-1}A' &= (n + p - 1)AF_{t-1}A' \\ (\delta n + p - 1)k &= (n + p - 1) \\ k &= \frac{n + p - 1}{\delta n + p - 1} \end{aligned} \tag{4.4}$$

Thus k in equation (4.4) can ensure the autoregressive property of the model.

Now we have the prior distribution of $\Phi_t|A, D_{t-1}$ following $W_p(\delta n + p - 1, \frac{n+p-1}{\delta n+p-1}AF_{t-1}A')$, we can derive the posterior distribution of $\Phi_t|A, D_t$ based on Bayes' theorem and the conjugate property of the Wishart distribution

as follows:

$$
\begin{aligned}
p(\Phi_t|A, D_t) &\propto p(y_t|\Phi_t, A, D_{t-1})p(\Phi_t|A, D_{t-1}) \\
&= N(\mu, \Phi_t^{-1})W_p(\delta n + p - 1, \frac{n+p-1}{\delta n + p - 1}AF_{t-1}A') \\
&= \frac{1}{\sqrt{2\pi}|\Phi_t|^{-1/2}} e^{-\frac{1}{2}tr((y_t-\mu)(y_t-\mu)'\Phi_t)} \times \\
&\quad |\Phi_t|^{\frac{\delta n+p-1}{2}} e^{-\frac{1}{2}tr((kAF_{t-1}A')^{-1}\Phi_t)} \\
&\propto |\Phi_t|^{\frac{\delta n+p-1+1}{2}} e^{-\frac{1}{2}tr(((y_t-\mu)(y_t-\mu)'+(kAF_{t-1}A')^{-1})\Phi_t)} \\
&= W_p(\delta n + 1 + p - 1, ((y_t-\mu)(y_t-\mu)' + (kAF_{t-1}A')^{-1})^{-1}) \\
&= W_p((1-\frac{1}{n})n + 1 + p - 1, F_t) \\
&= W_p(n + p - 1, F_t) \qquad (4.5) \\
F_t &= (e_t e_t' + (kAF_{t-1}A')^{-1}))^{-1} \\
e_t &= y_t - \mu
\end{aligned}
$$

With the posterior distribution of $p(\Phi_t|D_t, A)$ derived, we can redefine our model as follows:

$$
\begin{aligned}
\Sigma_t|\Sigma_{t-1}, A, y_t &\sim IW(n + p - 1, F_t) \\
A &\sim MVN(M, V) \\
y_t|\Sigma_t &= N(\mu, \Sigma_t), \qquad (4.6)
\end{aligned}
$$

where IW is an inverse Wishart distribution and the prior distribution for the autoregressive parameter A is a diagonal matrix and its diagonal elements

follow a multivariate normal distribution with mean M and covariance matrix V.

4.3 Estimation of hyperparameter matrix A

From the proposed model (equation (4.6)), it is clear that the autoregressive coefficient matrix A is the most important unknown parameter. Since our purpose for the model is to estimate high dimensional data, therefore the autoregressive coefficient is treated as a fix diagonal matrix, as before, which can simplify the calculation and reduce the estimation time.

4.3.1 Estimation of A

Let us consider the joint prior distribution of Φ_t and A given D_{t-1}; applying the joint distribution definition we have:

$$f(\Phi_t, A|D_{t-1}) = f(\Phi_t|D_{t-1}, A)f(A|D_{t-1}) \tag{4.7}$$

Then we have the posterior joint distribution of Φ_t and A given D_t and apply Bayes theorem as follows:

$$f(\Phi_t, A|D_t) = f(\Phi_t, A|D_{t-1})f(y_t|\Phi_t, A) = f(\Phi_t, A|D_{t-1})f(y_t|\Phi_t) \tag{4.8}$$

Taking equation (4.7) into equation (4.8), we have:

$$f(\Phi_t, A|D_t) = f(y_t|\Phi_t)f(\Phi_t|D_{t-1}, A)f(A|D_{t-1})$$

In order to obtain the density function of $f(A|D_t)$, we can integrate Φ_t out of the joint distribution of Φ_t and A given D_t as follows:

$$f(A|D_t) \propto f(A|D_{t-1}) \int f(y_t|\Phi_t)f(\Phi_t|D_{t-1}, A)d\Phi_t \tag{4.9}$$

From last section we have $y_t|\Phi_t \sim N(\mu, \Phi_t^{-1})$ and $\Phi_t|D_{t-1}, A \sim W_p(\delta n + p - 1, kAF_{t-1}A')$, therefore we can integrate the factor by conjugacy and the integral of equation (4.9) is:

$$\begin{aligned} f(A|D_t) &\propto f(A|D_{t-1}) \int f(y_t|\Phi_t)f(\Phi_t|D_{t-1}, A)d\Phi_t \\ &\propto f(A|D_{t-1})|e_t e_t' + (kAF_{t-1}A')^{-1})|^{-(\delta n+p)/2} \end{aligned}$$

Then for each time, we can repeat the calculation of the integration:

$$\begin{aligned} f(A|D_t) &\propto f(A|D_{t-2}) \prod_{j=t-1}^{t} |e_j e_j' + (kAF_{j-1}A')^{-1})|^{-(\delta n+p)/2} \\ &\propto \cdots\cdots \\ &\propto \cdots\cdots \\ &\propto f(A) \prod_{j=1}^{t} |e_j e_j' + (kAF_{j-1}A')^{-1})|^{-(\delta n+p)/2} \quad , \end{aligned} \tag{4.10}$$

where $f(A)$ is the prior density of A which is defined in the model and follows a multivariate normal distribution.

We are interested in estimating the mode of A. However, it is clearly difficult or even impossible to obtain a closed-form solution of the first derivative for $f(A|y_t)$; therefore we apply the Newton-Raphson method to approximate the mode, $\hat{A}$. According to the method, at each time point t, for iteration $i = 1, 2, \ldots,$ the estimation of the mode $\hat{A}$ can be computed as follows:

$$vec(\hat{A}^i) = vec(\hat{A}^{i-1}) + \frac{\frac{\partial \log f(A|y_t)}{\partial vec(A)}}{\frac{\partial^2 \log f(A|y_t)}{\partial vec(A) \partial vec(A)'}}|_{A=\hat{A}^{i-1}} \quad , \tag{4.11}$$

where the initial value of $\hat{A}^0$ will be given and $vec(')$ is the column stacking operator of an unrestricted matrix. It has been proved that, under some restrictive conditions, the estimation $\hat{A}^i$ will finally converge to the true value of the mode A.

From equation (4.10) we have:

$$\begin{aligned} \log f(A|D_t) = {} & \log(c) + \log f(A) - \frac{\delta n + p}{2} \sum_{j=1}^{t} \log |e_j e_j' + \\ & (kAF_{j-1}A')^{-1}| \end{aligned} \tag{4.12}$$

Equation (4.11) shows that, in order for the Newton-Raphson method to work, we need to calculate the first and second derivatives of $\log f(A|y_t)$.

From equation (4.12) we have:

$$\frac{\partial \log f(A|y_t)}{\partial A} = \frac{\partial \log(f(A))}{\partial A} - \frac{\delta n + p}{2} \sum_{j=1}^{t} \frac{\partial \log |e_j e_j' + (kAF_{j-1}A')^{-1}|}{\partial A} \quad (4.13)$$

and

$$\frac{\partial^2 \log(A|y_t)}{\partial A \partial A'} = \frac{\partial^2 \log(f(A))}{\partial A \partial A'} - \frac{\delta n + p}{2} \sum_{j=1}^{t} \frac{\partial^2 \log |e_j e_j' + (kAF_{j-1}A')^{-1}|}{\partial A \partial A'} \quad (4.14)$$

The equation (4.13) is the first derivative of $\log(A|y_t)$ and the equation (4.14) is the second derivative of $\log(A|y_t)$. In the following subsections, these two derivatives will be calculated.

4.3.2 First derivative of $\log f(A|y_t)$

Here we focus on obtaining the first derivative for $\log |e_j e_j' + (kAF_{j-1}A')^{-1}|$

$$\frac{\partial \log |e_j e_j' + (kAF_{j-1}A')^{-1}|}{\partial A} = \frac{\partial \log |e_j e_j' + (k^{-1}(A^{-1})' F_{j-1}^{-1} A^{-1})|}{\partial A} \quad (4.15)$$

Since A is a diagonal matrix, we have $A = A'$.

We assume that $e_j e_j' = K$, $k^{-1} = c$, $A = X$ and $F_{j-1}^{-1} = B$, so that:

$$\begin{aligned} \frac{\partial \log |K + cX^{-1}BX^{-1}|}{\partial x_i} &= \frac{\partial \log |K + cX^{-1}BX^{-1}|}{\partial x_i^{-1}} \frac{\partial x_i^{-1}}{\partial x_i} \\ &= -\frac{\partial \log |K + cX^{-1}BX^{-1}|}{\partial x_i^{-1}} \frac{1}{x_i^2} \end{aligned} \quad (4.16)$$

We have a lemma as follows (for details see Johnson et al. (2002)):

$$\frac{\partial \log(detF(x))}{\partial x_i} = tr(F(x)^{-1}\frac{\partial F(x)}{\partial x_i}) \tag{4.17}$$

So that:

$$\begin{aligned}\frac{\partial \log|K + cX^{-1}BX^{-1}|}{\partial x_i^{-1}} &= tr((K + cX^{-1}BX^{-1})^{-1}\frac{\partial (K + cX^{-1}BX^{-1})}{\partial x_i^{-1}}) \\ &= tr((K + cX^{-1}BX^{-1})^{-1}c\frac{\partial (X^{-1}BX^{-1})}{\partial x_i^{-1}}) \end{aligned} \tag{4.18}$$

We have a lemma as follows (for details see Johnson et al. (2002)):

$$\frac{\partial F(x)G(x)}{\partial x_i} = \frac{\partial F(x)}{\partial x_i}G(x) + F(x)\frac{\partial G(x)}{\partial x_i} \tag{4.19}$$

Therefore equation (4.18) becomes:

$$\begin{aligned} & tr((K + cX^{-1}BX^{-1})^{-1}c(\frac{\partial X^{-1}}{\partial x_i^{-1}}BX^{-1} + X^{-1}B\frac{\partial X^{-1}}{\partial x_i^{-1}})) \\ = \;& tr(cBX^{-1}(K + cX^{-1}BX^{-1})^{-1}\frac{\partial X^{-1}}{\partial x_i^{-1}}) \\ + \;& tr(c(K + cX^{-1}BX^{-1})^{-1}X^{-1}B\frac{\partial X^{-1}}{\partial x_i^{-1}}) \\ = \;& tr(\mu_i' cBX^{-1}(K + cX^{-1}BX^{-1})^{-1}\mu_i)\frac{\partial X^{-1}}{\partial x_i^{-1}} + tr(\mu_i' c(K + cX^{-1}BX^{-1})^{-1}X^{-1}B\mu_i)\frac{\partial X^{-1}}{\partial x_i^{-1}} \\ = \;& tr\Big(\mu_i' cBX^{-1}(K + cX^{-1}BX^{-1})^{-1}\mu_i + \mu_i' c(K + cX^{-1}BX^{-1})^{-1}X^{-1}B\mu_i\Big)\frac{\partial X^{-1}}{\partial x_i^{-1}} \end{aligned} \tag{4.20}$$

where μ_i is defined as:

$$\mu_i = \frac{\partial X}{\partial x_i}$$

It represents the ith partial derivatives of vector X. In a matrix notation, ijth partial derivatives of matrix X is $\frac{\partial X}{\partial x_{ij}} = \mu_i \mu_j'$. One of the important properties of μ_i is the free position switch i.e. $tr(A\mu_i\mu_j') = tr(\mu_j' A \mu_i) = a_{ij}$, details see Johnson et al. (2002)

Because of the property and all the metric are symmetric, we have: $tr(\mu_i' cBX^{-1}(K + cX^{-1}BX^{-1})^{-1}\mu_i) = tr(\mu_i' c(K + cX^{-1}BX^{-1})^{-1}X^{-1}B\mu_i) = tr(\mu_i Z \mu_i') = z_{ii}$, so that equation (4.16) becomes:

$$\frac{\partial \log|K + cX^{-1}BX^{-1}|}{\partial x_i} = -2z_{ii}\frac{1}{x_i^2} \quad , \tag{4.21}$$

where $Z = cBX^{-1}(K + cX^{-1}BX^{-1})^{-1}$ and z_{ii} is the ith diagonal element of Z.

Note that $f(A)$ follows a multivariate normal distribution, therefore the first derivatives of $log(f(A))$ can be derived as follows:

$$\begin{aligned}
\frac{\partial f(\log(A))}{\partial A} &= \frac{\partial(-\frac{p}{2}\log(2\pi) - \frac{1}{2}\log|V| - \frac{1}{2}(A-M)'V^{-1}(A-M))}{\partial A} \\
&= \frac{\partial(-\frac{1}{2}(A-M)'V^{-1}(A-M))}{\partial A} \\
&= -\frac{1}{2}((A-M)'V^{-1} + V^{-1}(A-M))
\end{aligned} \tag{4.22}$$

Due to the fact that matrices A, M and V are symmetric matrices, the

equation (4.22) becomes:

$$\frac{\partial f(log(A))}{\partial A} = -\frac{1}{2}(2V^{-1}(A-M)) = -V^{-1}(A-M) \tag{4.23}$$

4.3.3 Second derivative of $\log f(A|y_t)$

Here we focus on obtaining the second derivative for $\log|e_j e_j' + (kAF_{j-1}A')^{-1}|$. First, introduce the lemma (for details see Johnson et al. (2002)):

$$\frac{\partial^2 \log det(F)}{\partial x_i \partial x_j} = tr(F^{-1}\frac{\partial^2 F}{\partial x_i \partial x_j}) - tr(F^{-1}\frac{\partial F}{\partial x_i}F^{-1}\frac{\partial F}{\partial x_j}) \quad , \tag{4.24}$$

where $F = K + cX^{-1}BX^{-1}$.

From equation (4.24), we need to calculate $\frac{\partial^2 F}{\partial x_i \partial x_j}$, $\frac{\partial F}{\partial x_i}$ and $\frac{\partial F}{\partial x_j}$ respectively.

$$\begin{aligned}\frac{\partial^2 F}{\partial x_i \partial x_j} &= \frac{\partial^2 (K + cX^{-1}BX^{-1})}{\partial x_i \partial x_j} = \frac{\partial^2 (cX^{-1}BX^{-1})}{\partial x_i \partial x_j} = \frac{\partial^2 (c^{-1}XB^{-1}X)^{-1}}{\partial x_i \partial x_j} \\ &= \frac{\partial^2 G^{-1}}{\partial x_i \partial x_j} \quad ,\end{aligned} \tag{4.25}$$

where $G = (c^{-1}XB^{-1}X)^{-1}$.

Then we have the lemma as follows (for details see Johnson et al. (2002)):

$$\begin{aligned}
\frac{\partial^2 G^{-1}}{\partial x_i \partial x_j} &= -G^{-1}\frac{\partial^2 G}{\partial x_i \partial x_j}G^{-1} + G^{-1}\frac{\partial G}{\partial x_i}G^{-1}\frac{\partial G}{\partial x_j}G^{-1} \\
&+ G^{-1}\frac{\partial G}{\partial x_j}G^{-1}\frac{\partial G}{\partial x_i}G^{-1} \\
\frac{\partial^2 G}{\partial x_i \partial x_j} &= \frac{\partial^2 (c^{-1}XB^{-1}X)}{\partial x_i \partial x_j} = c^{-1}\frac{\partial^2 (XB^{-1}X)}{\partial x_i \partial x_j} \\
&= c^{-1}2b_{ij} \quad , \qquad (4.26)
\end{aligned}$$

where b_{ij} is the i^{th} row, j^{th} column element of B.

$$\begin{aligned}
\frac{\partial G}{\partial x_i} &= \frac{\partial (c^{-1}XB^{-1}X)}{\partial x_i} = c^{-1}\frac{\partial (XB^{-1}X)}{\partial x_i} \\
&= c^{-1}(X\frac{\partial B^{-1}X}{\partial x_i} + \frac{\partial X}{\partial x_i}B^{-1}X) \\
&= c^{-1}(XB^{-1}u_i u_i' + u_i u_i' B^{-1}X) \\
\frac{\partial G}{\partial x_j} &= c^{-1}(XB^{-1}u_j u_j' + u_j u_j' B^{-1}X) \\
\frac{\partial F}{\partial x_i} &= \frac{\partial G^{-1}}{\partial x_i} = -G^{-1}\frac{\partial G}{\partial x_i}G^{-1} \\
&= -cG^{-1}(XB^{-1}u_i u_i' + u_i u_i' B^{-1}X)G^{-1} \\
\frac{\partial F}{\partial x_j} &= \frac{\partial G^{-1}}{\partial x_j} = -G^{-1}\frac{\partial G}{\partial x_j}G^{-1} \\
&= -cG^{-1}(XB^{-1}u_j u_j' + u_j u_j' B^{-1}X)G^{-1} \qquad (4.27)
\end{aligned}$$

Therefore we have:

$$\begin{aligned}
\frac{\partial^2 F}{\partial x_i \partial x_j} &= -cX^{-1}BX^{-1}c^{-1}2b_{ij}cX^{-1}BX^{-1} + cX^{-1}BX^{-1}c^{-1}(XB^{-1}u_iu_i' + u_iu_i'B^{-1}X) \\
&\times \; cX^{-1}BX^{-1}c^{-1}(XB^{-1}u_ju_j' + u_ju_j'B^{-1}X)cX^{-1}BX^{-1} \\
&+ \; cX^{-1}BX^{-1}c^{-1}(XB^{-1}u_ju_j' + u_ju_j'B^{-1}X) \\
&\times \; cX^{-1}BX^{-1}c^{-1}(XB^{-1}u_iu_i' + u_iu_i'B^{-1}X)cX^{-1}BX^{-1} \\
&= -2cb_{ij}X^{-1}BX^{-1}X^{-1}BX^{-1} + cX^{-1}BX^{-1}(XB^{-1}u_iu_i' + u_iu_i'B^{-1}X) \\
&\times \; X^{-1}BX^{-1}(XB^{-1}u_ju_j' + u_ju_j'B^{-1}X)X^{-1}BX^{-1} \\
&+ \; cX^{-1}BX^{-1}(XB^{-1}u_ju_j' + u_ju_j'B^{-1}X) \\
&\times \; X^{-1}BX^{-1}(XB^{-1}u_iu_i' + u_iu_i'B^{-1}X)X^{-1}BX^{-1} \qquad (4.28)
\end{aligned}$$

Finally from equation (4.27) and equation (4.28), we have:

$$\begin{aligned}
\frac{\partial^2 logdet(F)}{\partial x_i \partial x_j} &= tr(c^{-1}XB^{-1}X\frac{\partial^2 F}{\partial x_i \partial x_j}) - tr(c^{-1}XB^{-1}X\frac{\partial F}{\partial x_i}c^{-1}XB^{-1}X\frac{\partial F}{\partial x_j}) \\
&= tr\Big(- 2b_{ij}X^{-1}BX^{-1} + (XB^{-1}u_iu_i' + u_iu_i'B^{-1}X) \\
&\times \; X^{-1}BX^{-1}(XB^{-1}u_ju_j' + u_ju_j'B^{-1}X)X^{-1}BX^{-1} \\
&+ \; (XB^{-1}u_ju_j' + u_ju_j'B^{-1}X)X^{-1}BX^{-1}(XB^{-1}u_iu_i' + u_iu_i'B^{-1}X)X^{-1}BX^{-1}\Big) \\
&- \; tr\Big(c^2(XB^{-1}u_iu_i' + u_iu_i'B^{-1}X)cX^{-1}BX^{-1}(XB^{-1}u_ju_j' + u_ju_j'B^{-1}X)cX^{-1}BX^{-1}\Big) \\
& \qquad (4.29)
\end{aligned}$$

Also the second derivative of $\log(f(A))$ can be calculated based on equation (4.23):

$$\frac{\partial^2 \log(f(A))}{\partial A \partial A'} = \frac{\partial(-V^{-1}(A-M))}{\partial A'} = -V^{-1} \tag{4.30}$$

4.3.4 Algorithm

The algorithm of the Newton-Raphson method is shown in algorithm 4. The idea of Storvik (2002) has been used: at each time, we can estimate F_t and A based the information at previous time and then use the estimated F_t and A to estimate the volatility Σ_t and F_{t+1}. As the time moves forward, we repeat the process above.

Algorithm 4 Algorithm of Newton-Raphson method

1: At t=1, Σ_1 and F_2 can be calculated using the start value F_1 and $\hat{A}_1$:

$$p(\Sigma_1|F_1) = IW(n+p-1, F_1) \quad and \quad p(F_2) = (e_2 e_2' + (k\hat{A}_1 F_1 \hat{A}_1')^{-1})^{-1}$$

2: **for** $t = 2, \ldots, T$ **do**, $\hat{A}_t$ is updated at every time point t by employing the Newton-Raphson method.

3: **Repeat**:

$$vec(\hat{A}_t^j) \quad = \quad vec(\hat{A}_t^{j-1}) + \frac{\frac{\partial log f(A|y_t)}{\partial vec(A)}}{\frac{\partial^2 log f(A|y_t)}{\partial vec(A) \partial vec(A)'}}|_{A=\hat{A}_t^{j-1}}$$

where $\frac{\partial log f(A|y_t)}{\partial vec(A)}$ and $\frac{\partial^2 log f(A|y_t)}{\partial vec(A) \partial vec(A)'}$ can be calculated from equation (4.21), (4.23), (4.29) and (4.30) respectively. $\hat{A}_t^j$ is the estimated j-th iteration of mode of $f(A|y_t)$

4: **Until**:

$$|\hat{A}_t^j - \hat{A}_t^{j-1}| < \phi$$

where ϕ is a certain threshold that is close to 0.

5: After obtaining the estimation for the unknown parameter A at time t, we can calculate the volatility from its posterior distribution:

$$p(\Sigma_t|\Sigma_{t-1}, A, y_t) = IW(n+p-1, F_t)$$

6: Update F_{t+1} as:

$$F_{t+1} = (e_{t+1} e_{t+1}' + (k\hat{A}_t F_t \hat{A}_t')^{-1})^{-1}$$

7: **end for**

4.4 Application

4.4.1 Currency exchange rate data

In this subsection, the proposed Newton method will be applied to currency exchange rate data. The dataset is the currency exchange rates between Canadian dollar and US dollar, and between Euro and US dollar. The data were collected at daily frequency from 2010 to 2012 with 400 data points in total.

The starting values for the parameters were set to be: the number of time periods $T_1 = 400$, the coefficient $\delta = 0.9$, the starting value of A is an identity matrix, the starting value of F is also an identity matrix, the mean of the observation is $\mu = (0.001, 0.001)$, the degrees of freedom of the inverse Wishart distribution $n = 10$, the threshold of the norm between $\{\hat{A}_t^j\}$, is set to be $\phi = 0.01$.

4.4.2 Estimation

We apply algorithm 4 of the Newton-Raphson method to the proposed Wishart Autoregressive model (equation (4.6)). The R code can be found in appendix document 'Newton.R'.

The estimated volatility and the log-return of the data is shown in Figure 4.1. It indicates that the estimated volatility reflects the fluctuation of the log-return of the data well. For example, both the CAD and the EUR have big swings at the beginning, around 60 to 80 data points; the estimated

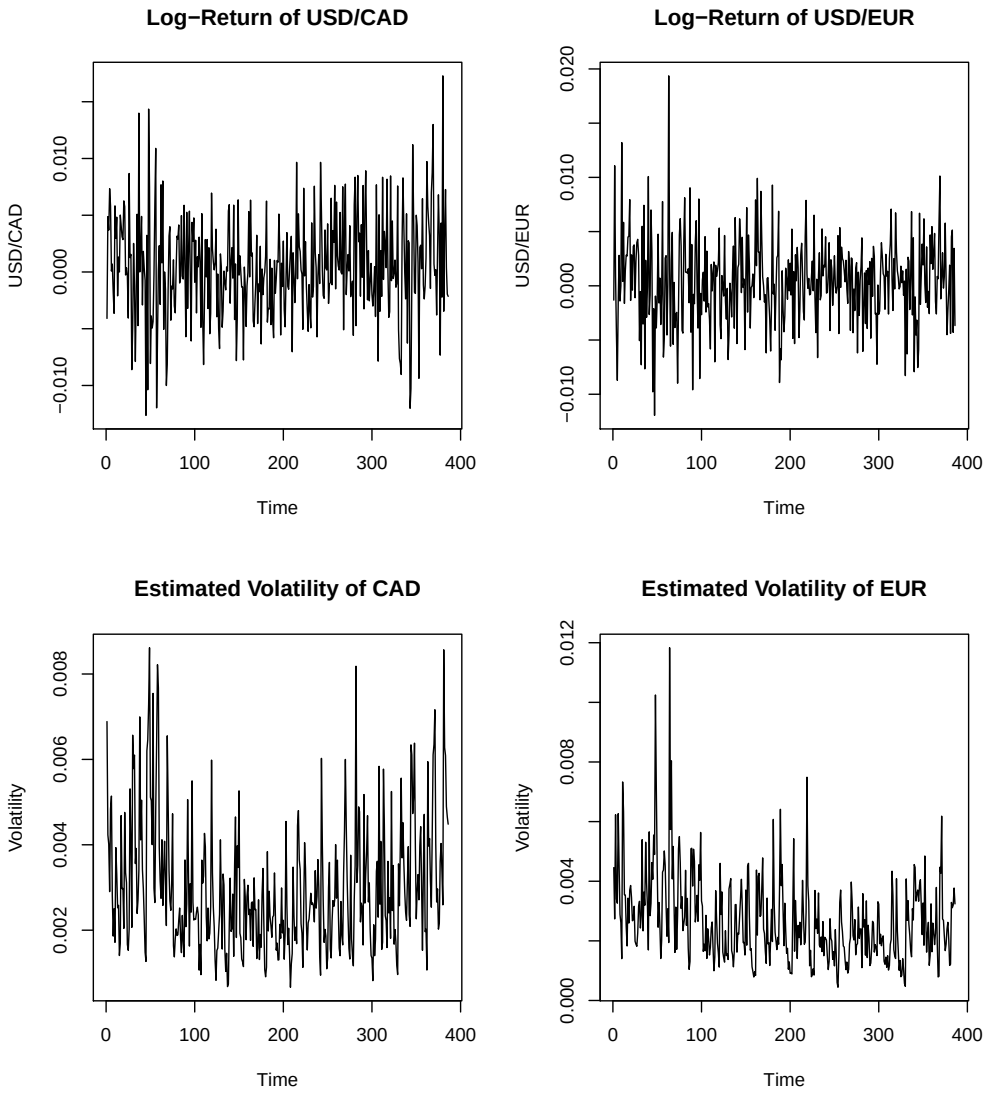

Figure 4.1: Log-return vs. estimated volatility

volatility captures the above movements and responds with high estimated volatility values at the corresponding times.

The estimations of the diagonal elements of the Wishart autoregressive coefficient A is shown in Figure 4.2. We can see that both a_{11} and a_{22} are converge after time 50 and the converging values are 0.9999 and 0.9999 respectively.

One of the advantages of estimating the covariance matrix as a whole is that we can learn the correlations between each component. The correlation coefficient between CAD and EUR is shown in Figure 4.3. First of all, as expected, all the estimated correlation coefficients are within the range $[-1, 1]$. Then there are more positive correlation coefficients than negative correlation coefficients in the plot, which means that most of time the exchange rate of CAD/USD and EUR/USD are positively correlated.

Figure 4.4 shows the change of norm between the matrices $\{\hat{A}^j\}s$ under each iteration at time point 2. We can see that the norm of the matrices starts at around 0.1; as the number of iterations increases, the norm decreases and so does the rate of decrease. Eventually, after around 9000 iterations, the norm reaches the threshold 0.01 which means that the $\{\hat{A}^j\}s$ have converged so that the estimation has been made.

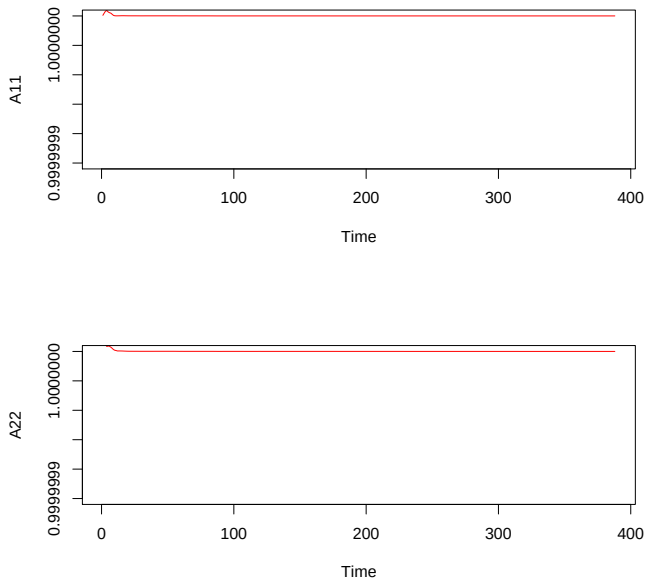

Figure 4.2: Estimation of a_{ii}

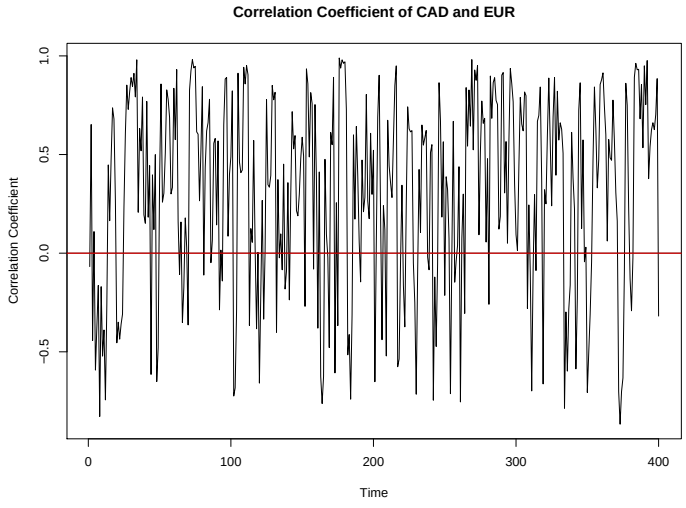

Figure 4.3: Estimated correlation coefficient between CAD and EUR

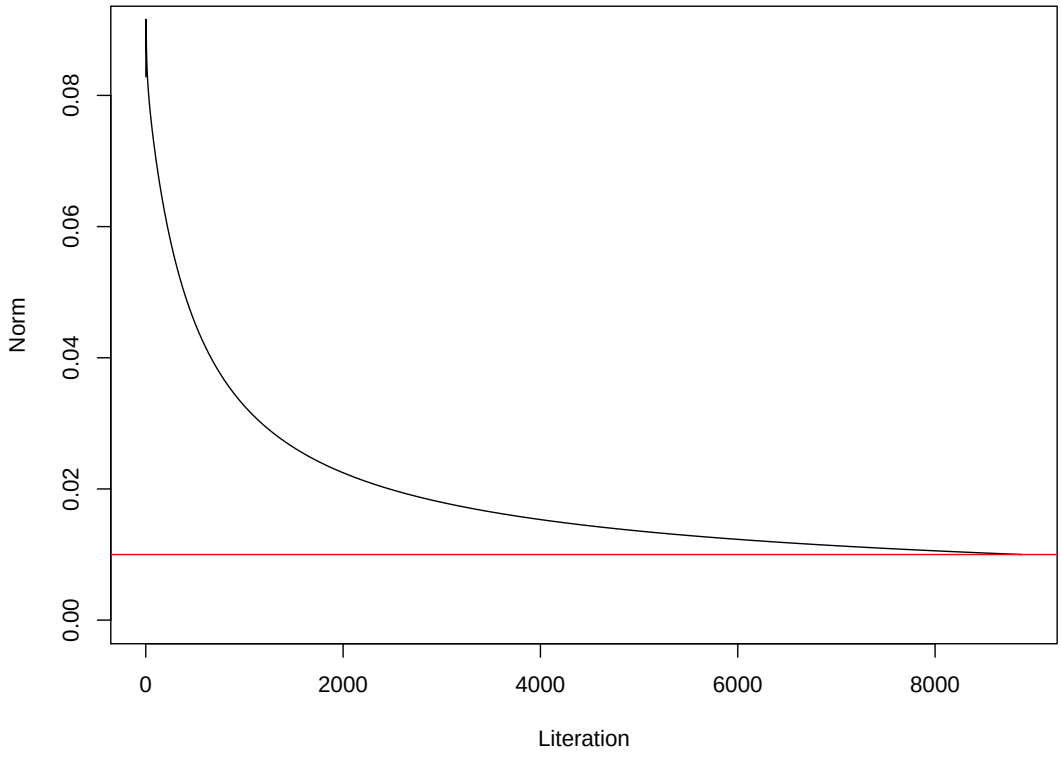

Figure 4.4: The norm of $\hat{A}^j$ and $\hat{A}^{j-1}$ at time 2

Chapter 5

Portfolio Returns and Applications

5.1 Introduction

In this chapter, we will use the previously established Wishart autoregressive model (Section 3.2) and Wishart-Newton model (Section 4.2) and then compare them with the generalised orthogonal (GO) GARCH model (for details see VanDerWeide (2002)). The GO-GARCH model is a well-known multivairate GARCH family model and it has proved to be a standard multivariate volatility model with accuracy and efficiency in fairly low and medium dimensions. We choose GO-GARCH over RBEKK model is because according to Noureldin et al (2014), GO-GARCH model has silghtly better performance than RBEKK model on medium dimensional data. The comparison

includes: firstly, the function execution time in R of each model; secondly, the cumulative returns of each model; thirdly, the Dow Jones Industrial 30 portfolio selection and returns of the Wishart autoregressive model and Wishat-Newton models. Throughout the comparisons, we will be able to see the efficiency and performance of each of the models in different dimensions.

5.2 Function execution time

The function execution time can be affected by different factors:

- The complexity of the algorithm. This includes the amount of calculation, the number of iterations and the different types of operation involved in the algorithm.
- The programming language, for example the $C++$ or Java languages are believed to be more efficient in coding and calculation compared to R; however R is more statistics-friendly.
- Code optimisation. Optimisation of the code can make it more efficient so that functions use less resources and time during the calculation.
- Computer hardware. A good computer with high performance CPU, large memory capacity and fast disk can boost the speed of the function calculation.
- Operating system. There are certain OS which are designed for programmers, such as Ubuntu and Linux, that are more flexible and faster

for running code and complex calculations than Windows systems.

So, by fixing the elements above, the time consumption of each program then reflects the complexity of each method. It is very important that the method is simple and takes little time to operate, especially in real-world situations.

We run the optimised code for the Wishart autoregressive model, the GO-GARCH model and the Newton model in the programming language R on the same computer and the same operating system. The execution time of the three models for 400 data points in dimensions 3, 5 and 10 is shown in Table 5.1. The 3 and 5 dimensional data are the currency exchange rate data and the 10 dimensional data are FTSE stock prices. It should be noted that the Wishart autoregressive mode is estimated by the hybrid SMC method and the GO-GARCH model is by default estimated by the offline ML method. In most practical situations, new data arrive in fixed intervals and at each interval we estimate the volatility from the given data.

From Table 5.1, we can see that the Wishart autoregressive model has the shortest execution time, and that increase in the dimensionality of the data has the least impact on its time. The Wishart-Newton model, conversely, has the longest execution time of the three by a large margin, and increase in the dimensionality of the data also has the largest impact on its execution time. The GO-GARCH model lies in the middle on both comparisons.

Method	3-dimensional data	5-dimensional data	ratio	10-dimensional data	ratio
WIS	16.54 sec	24.19 sec	1.46	36.42 sec	2.20
GO-GARCH	37.32 sec	46.48 sec	1.24	82.67 sec	2.21
Newton	746.39 sec	1072.43 sec	1.43	1891.91 sec	2.54

Table 5.1: Execution time of the methods for 400 data points in different dimensions

5.3 Portfolio allocation and cumulative return

The volatility model comparison is normally made by dynamic portfolio allocation, which involves the one-step forecast of both the volatility of the returns $Q_t = Var(y_t|D_{t-1})$ and the mean of the returns $f_t = E(y_t|D_{t-1})$. Dynamic portfolio allocation means the the allocation decisions can be made sequentially. At time $t-1$, the investor can calculate the optimal allocation vector a_t for time t, with the help of forecasting Q_t and f_t, and then reallocate the investment accordingly. For reasons of simplicity, we assume no transaction costs in the trade, and the reallocation of the portfolio happens instantaneously both on long and short positions.

The task of dynamic portfolio allocation is to find the allocation vector or the optimal weight, w_t, at each time, which can balance the two ends of both high return and low risk. Here we implement the Markowitz mean-variance optimisation allocation to determine w_t.

The basic idea of Markowitz mean-variance optimisation is that either (1) by fixing the expected return $m = w_t' f_t$, we can find a_t that minimises the variance of the realised portfolio return $r_t = w_t' y_t$, that is minimising the portfolio volatility $w_t' Q_t w_t$; or (2) by fixing the portfolio volatility $w_t' Q_t w_t =$

ϕ^2, we can find the same a_t that maximises the expected portfolio return m.

The w_t that suits the dual property above can be calculated as follows:

$$w_t = \frac{mQ_t^{-1}f_t}{f_t'Q_t^{-1}f_t} \tag{5.1}$$

It should be noted that here we use an unconstrained portfolio strategy which indicates that the allocation can be chosen freely without regard to resources, permitting arbitrary long or short positions across the suitable financial instruments.

With the optimisation portfolio allocation, we can perform volatility analysis by applying the Wishart autoregressive, Wishart-Newton and GO-GARCH models, and compare the cumulative returns, ($R_T = \sum_{t=1}^{T} r_t$), of each model. Obviously, we are expecting the best performing model should have the largest cumulative returns. The selection of the returns portfolio is done online, i.e. we estimate Σ_t based on past state of volatility Σ_{t-1} and observation y_{t-1} only then, at each time we choose a portfolio to hold until new observation y_t come in and then we repeat for each time $t = 1, \ldots, T$ and work out the overall return. For the WIS and Newton model, since we have developed online estimate algorithm for them, so that it is straightforward to estimate the volatility online and make one-step forecast. However for the GO-GARCH model, for each time $t = 1, \ldots, T$, we use all the past observations $y_1, \ldots, y_{t-1}$ for the model to estimate and make one-step forecast.

Since the data that we used are financial data, so that I've taken the log-

return of the data and therefore the mean of the log-returns were set to be zero for all methods.The cumulative returns of WIS (green), Newton (blue) and GO-GARCH (red) models estimating the 3-dimensional data is shown in Figure 5.1, 5-dimensional data is shown in Figure 5.3 and 10-dimensional data is shown in Figure 5.5. The portfolio allocation of the 3-dimensional data is shown in Figure 5.2, 5-dimensional data is shown in Figure 5.4, 10-dimensional data is shown in Figure 5.6. The 3-dimensional data and the 5-dimensional data are currency exchange rates and the 10-dimensional data are FT-500 stock prices.

From Figure 5.1, we can see firstly that all three models give positive cumulative returns in the end. Secondly, GO-GARCH has the highest cumulative returns most of the time, WIS has lower cumulative returns but it is relatively close to GO-GARCH, while Newton has the lowest cumulative returns, especially over the time period 300 to 400. Figure 5.3 shows that GO-GARCH still has the highest cumulative returns, WIS is in the middle and the performance of Newton has improved. Figure 5.5 illustrates that WIS has the best performance of the three, Newton catches up to second place and GO-GARCH is the one with the lowest cumulative returns. We can see that as the dimensionality of the data increases, the performance of the WIS and Newton models is getting better, and both eventually outperform GO-GARCH on 10-dimensional data. The dataset can be found in appendix document 'FTSE data.csv'.

It should be noted in the portfolio weight figures, the weights are mostly

stable but occasionally erratic. For example in Figure 5.6, the erratic occurred from time 260 to 280. After checking the estimated volatility, we found the large estimated volatility on the same period of time. I think the large volatility could cause the unstable behaviour of the portfolio weights.

Also given the optimal weight, see equation (5.1), we calculated the minimized variance of return for each portfolio. It should be noted that the lower the minimized variance means the better estimated volatility it is to the portfolio and we also want the minimized variance close to the average variance of the data as much as possible. They are shown in table 5.2 along with the simple average variance of the data. We can see that for 3-dimensional data, GO-GARCH has the lowest minimized variance while Newton method has the highest. The minimized variance for each method are fairly close to the average variance of the data which suggests the estimate are accurate. For the 5-dimensional data, Wishart method has the lowest minimized variance and historical variance has the highest. However all the minimized variance are quite far from the average variance of the data which could lead to the conclusion of poor estimate. For 10-dimensional data, Wishart method again has the lowest minimized variance and historical variance has the highest. It is also good to see that the minimized variance for each method are close to the average variance which suggests the estimate are accurate for 10-dimensional data.

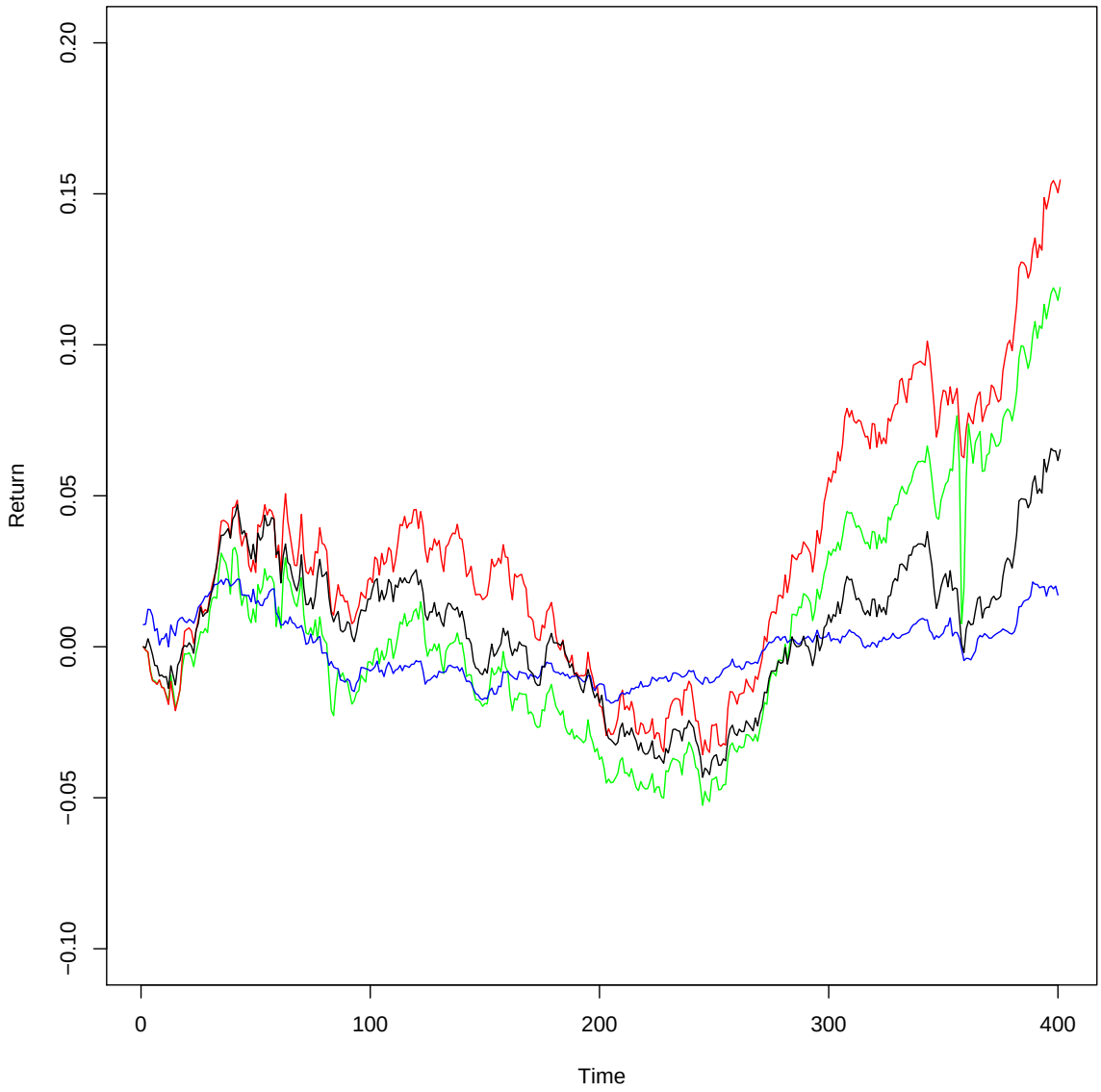

Figure 5.1: The cumulative return of Historical volatility(black), GO-GARCH(red), Wishart(green) and Wishart-Newton(blue) of 3-dimensional data

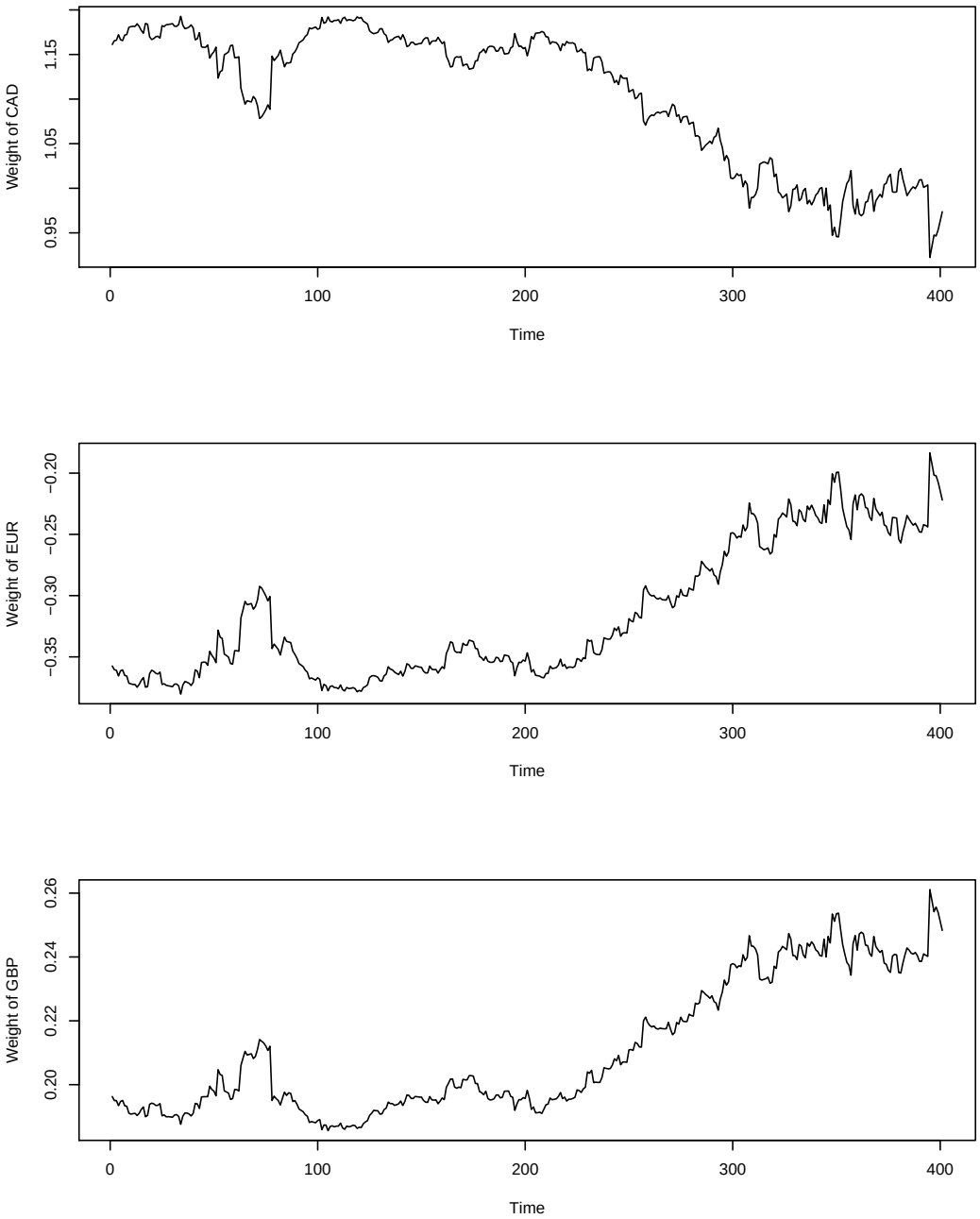

Figure 5.2: The allocated weight for each column of 3-dimensional data

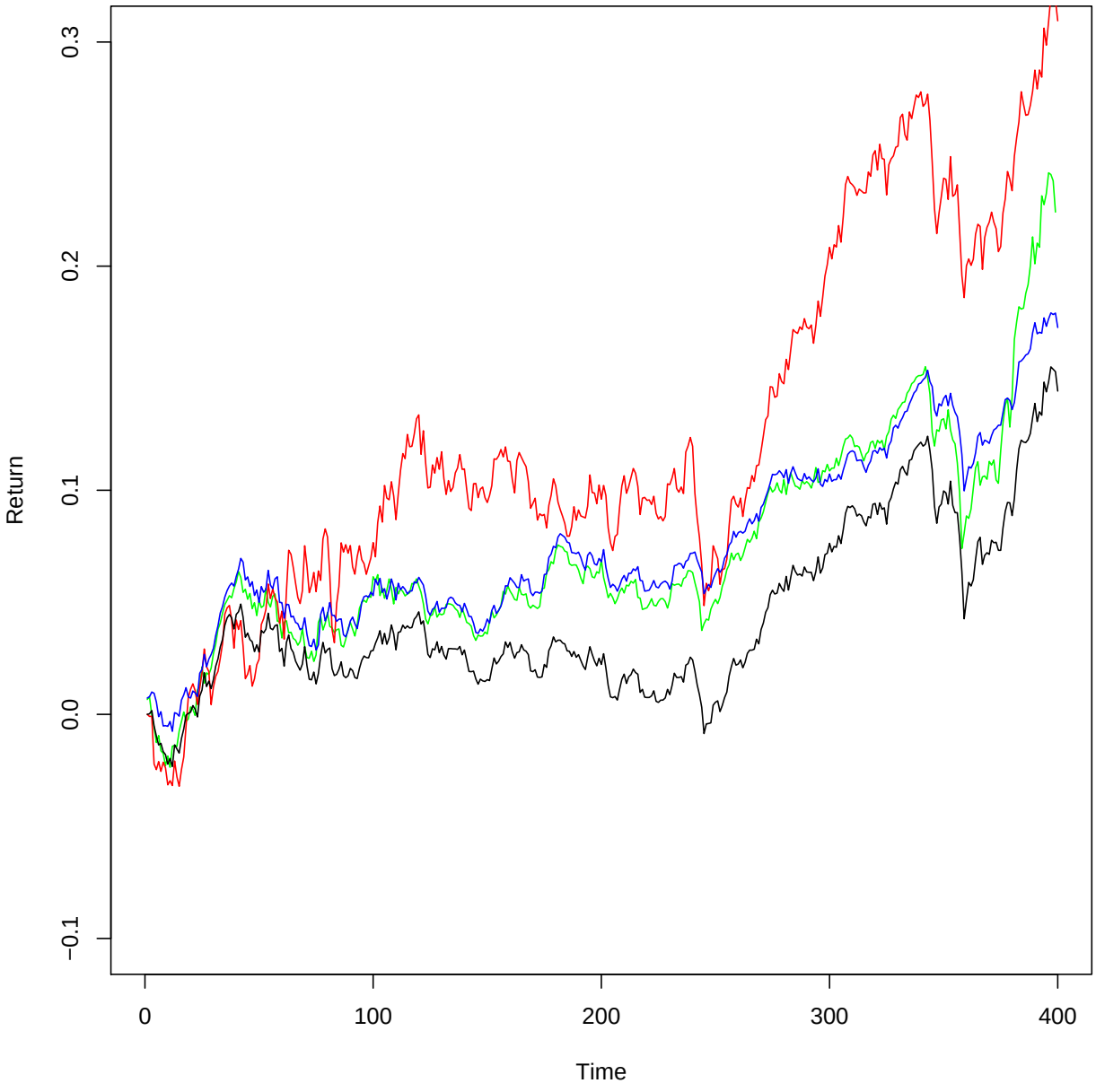

Figure 5.3: The cumulative return of Historical volatility(black), GO-GARCH(red), Wishart(green) and Wishart-Newton(blue) of 5-dimensional data

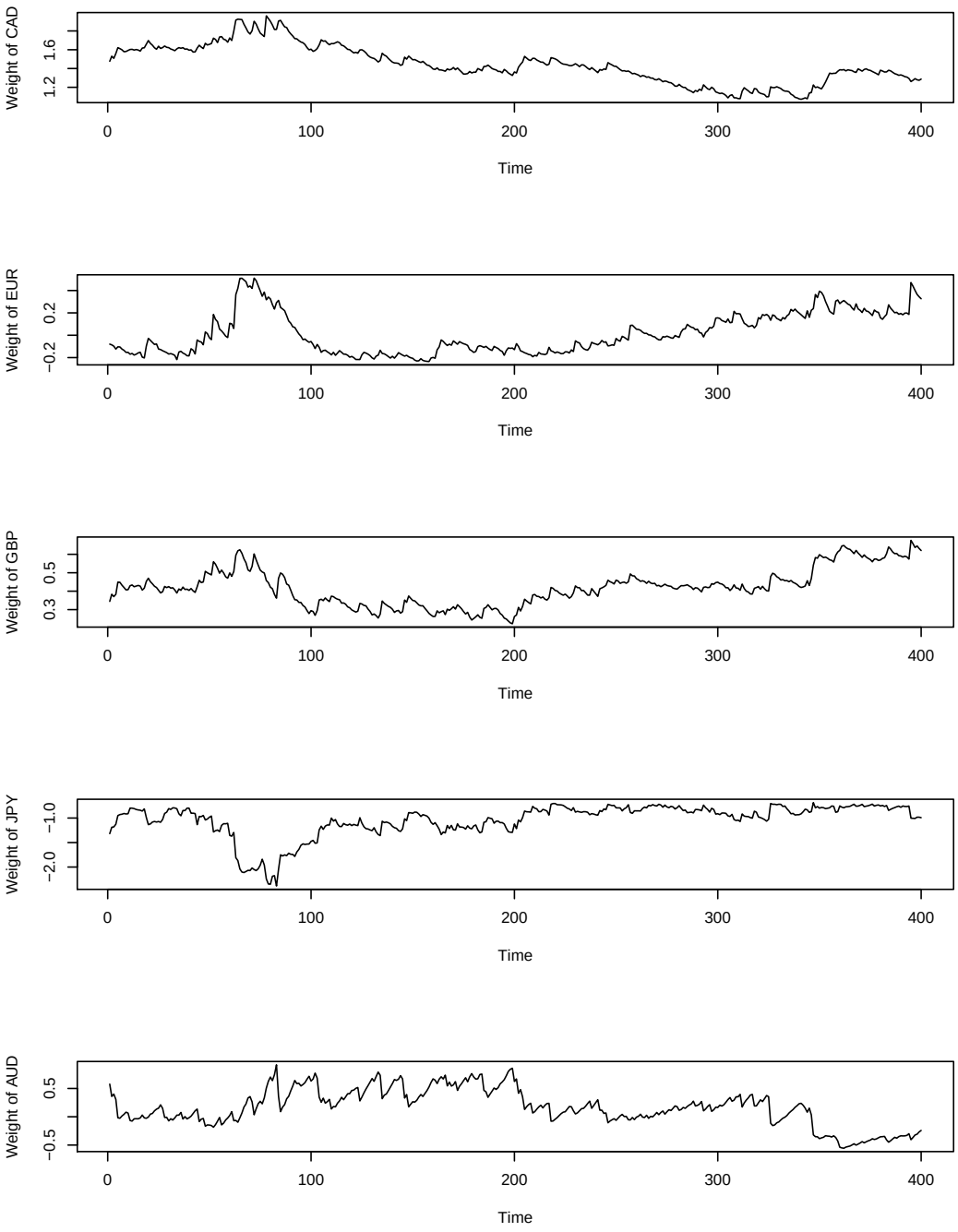

Figure 5.4: The allocated weight for each column of 5-dimensional data

Method/ Min var	3-dimensional data	5-dimensional data	10-dimensional data
WIS	0.0061	0.0099	0.0401
GO-GARCH	0.0050	0.0092	0.0577
Newton	0.0121	0.0128	0.0493
Historical	0.0074	0.0143	0.0742
Average variance	0.0046	0.0047	0.0319

Table 5.2: Minimized variance of the portfolio on different dimensions

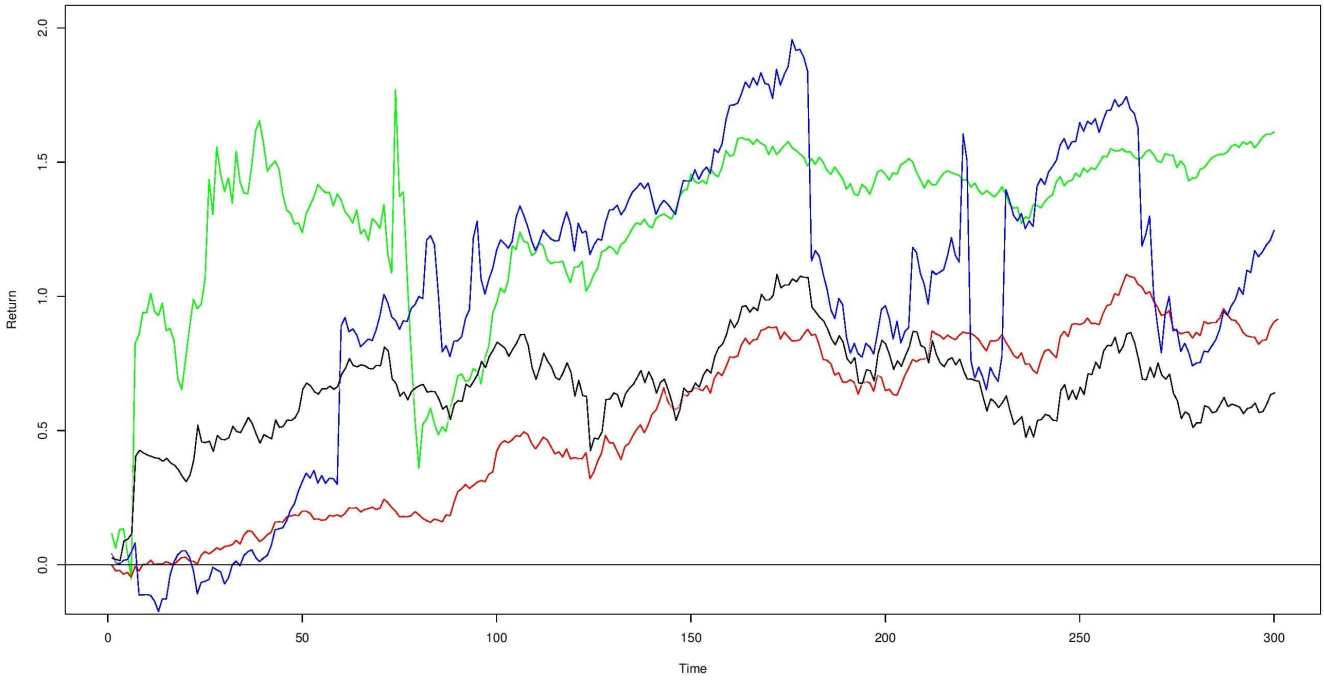

Figure 5.5: The cumulative return of Historical volatility(black), GO-GARCH (red), Wishart (green) and Wishart-Newton (blue) of 10-dimensional data

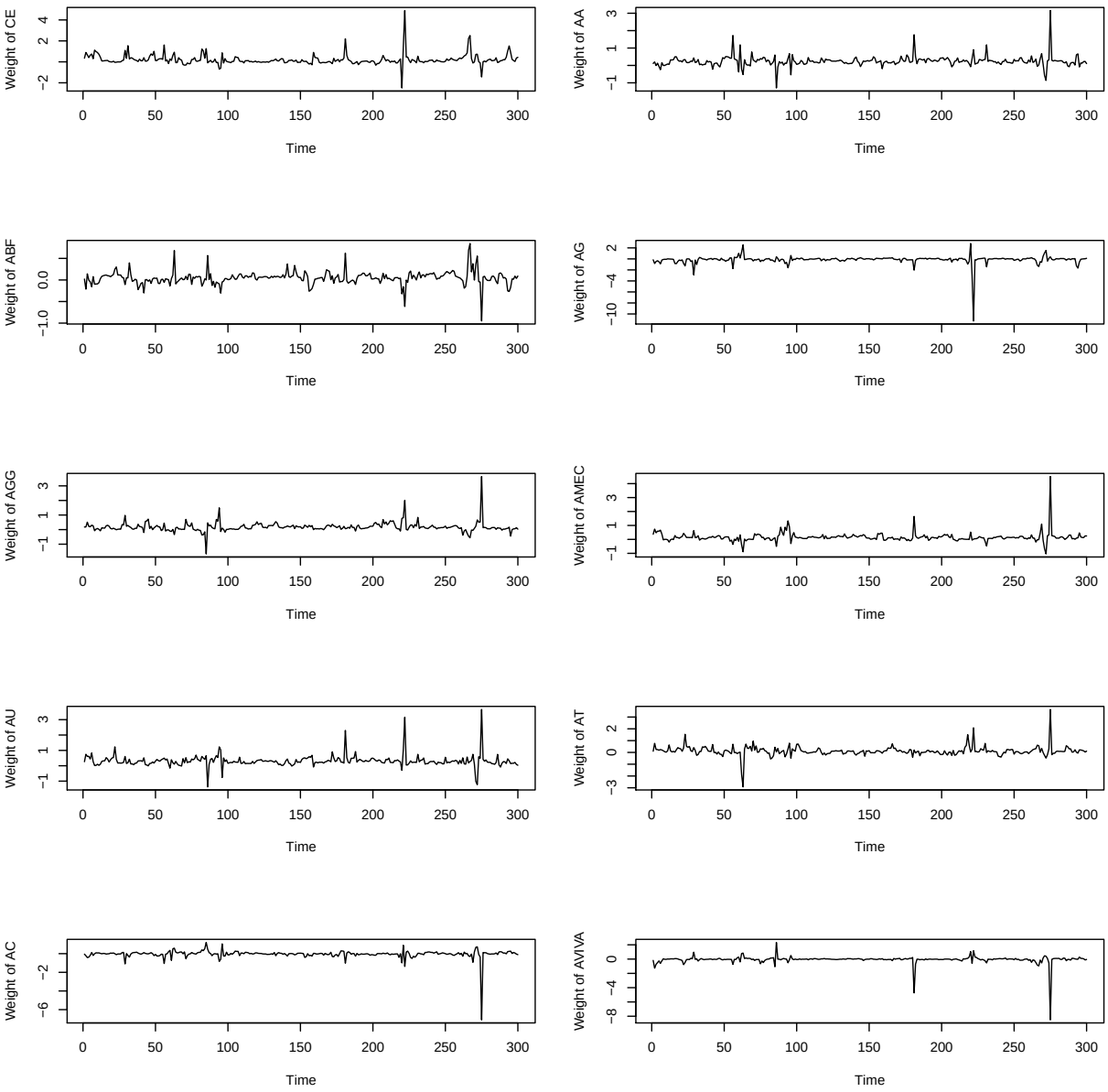

Figure 5.6: The allocated weight for each column of 10-dimensional data

5.4 Dow Jones 30 portfolio dimension reduction and allocation

The size of portfolios is growing larger and larger nowadays. One of the reasons that large size portfolio strategy is popular in investment is that it can reduce risk by diversifying the investment target. Diversification includes portfolio spread into different investment vehicles, choice of securities with varying risk, and spread of securities in different industries to minimise unsystematic risk. Most of the time, the size of the portfolio is much larger than 10, sometimes even larger than 100. Such large dimensions cause various problems: for example, it is difficult to track down all the performance and potential of the stocks; it takes a long time to collect and analyse the data; the transaction cost of the trade is relatively high. So the key here would be portfolio dimension reduction. It would save a lot of time and money if we can describe a high dimension portfolio with a lower-dimensional one without much loss of information.

Here we take the Dow Jones 30 industrial average (DJIA) for example; the data are collected from 2013 to 2014 – 256 trading days in one year. Since the data has 30 dimensions, we will reduce the dimensions to 20, then apply the WIS and Newton model to estimation the data and forecast the cumulative returns. The dataset can be found in appendix document 'DowJones-30 data.xlsx'.

The mathematic definition of DJIA is:

$$DJIA_t = \frac{\sum_{i=1}^{30} p_{it}}{d} \quad ,$$

where p_{it} are the prices of component stocks and d is called the Dow Divisor, which is a dynamic number changing with the prices. d was equal to 0.146 at the time the data was collected.

According to this definition, the DJIA is a price-weighted index, which is criticised a lot because it gives stocks with high prices more influence than low prices stocks. However, the DJIA is still the most cited and used stock market index. Given its price-weighted property, we can choose its top 20 priced stocks and reduce the dimensions with the least cost. Then the Wishart autoregressive and Wishart-Newton models can be applied to the data and the cumulative returns are shown in Figure 5.8. Both methods delivered positive cumulative returns and the results are almost the same in the end. The Wishart-Newton model has better returns by a very small margin. The Dow-Jones 30 Index at the same period of time has bee shown in Figure (5.7). We can see that the Index has a similar trend as the cumulative returns, however the returns of the Index over the year is much smaller than the cumulative returns of Wishart and Newton method.

It should be noted that for the 20-dimensional data, we did not include GO-GARCH as a comparison simply because for high dimensional data, the GARCH family models are known to have relatively poor and unstable per-

formance (for details of arguments see Bauwens et al. (2006) and Silvennoinen and Terasvirta (2009)), but recently there is composite likelihood estimation of large dimension GARCH model, details see Engle et al. (2016) which has been developed to deal with large dimensional data.

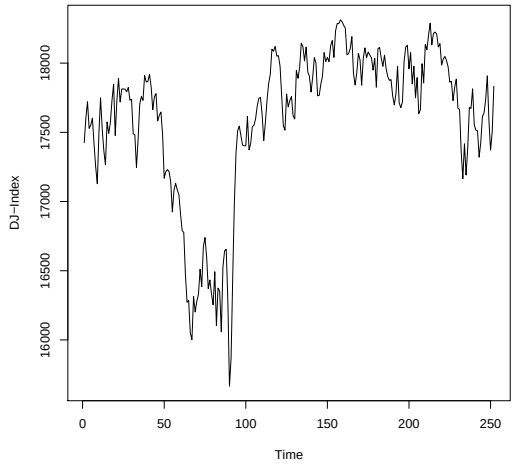

Figure 5.7: The Dow-Jones 30 Index from 2015 to 2016

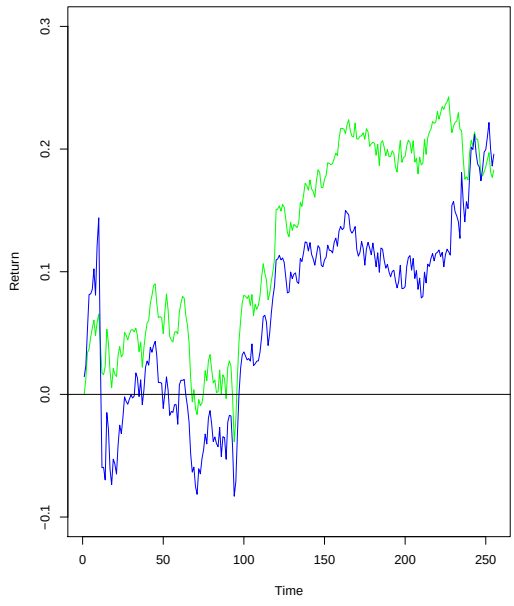

Figure 5.8: The cumulative return of Wishart (green) and Wishart-Newton (blue) method of DJ30 portfolio data from 2015 to 2016

Chapter 6

Conclusions and Future Work

6.1 Conclusions

In this thesis we developed two different sets of models. The first is Wishart autoregressive models which can describe both Gussian returns and returns with heavy tails and skewness. Then we simplified the model by treating the autoregressive parameter A as a diagonal matrix. This makes a huge reduction in the number of parameters and estimate computation. Both the parameter and the volatility matrix can be estimated by the SMC algorithm, which is an online estimation method. The applications show that both the original model and the simplified model have plausible volatility and parameter estimates.

The second model used is the Wishart-Newton model with a diagonal hyperparameter. This model is inspired by Uhlig's Wishart autoregressive

process but improved with an autoregressive structure. The estimation of the hyperparamter is carried out by the Newton-Raphson method and integrated into an online volatility estimation algorithm. The application shows reasonable volatility and parameter estimates.

We compare the two developed models with a well-established multivariate volatility model, GO-GARCH. The comparison of function execution time shows that the Wishart autoregressive model use least time and is also least impacted by the dimensionality of the data. Then given 3, 5 and 10 dimensional data, we compare the cumulative returns of the three models. The GO-GARCH model has the best cumulative returns in 3 and 5 dimensional data but it is outperformed by the Wishart autoregressive and Wishart-Newton models in 10 dimensional data. Furthermore, we reduce 30-dimensional Dow Jones 30 industrial index data to 20 dimensions and apply both Wishart autoregressive model and Wishart-Newton model to them. The cumulative returns show that both models have similar returns, and the Wishart-Newton model has the better performance by a small margin.

Overall, both models show reliable estimates for volatility and unknown parameters in a multivariate set-up using online algorithms. The dimensionality of the data is improved to medium or high level (20 to 30 and even more). The comparison reveals that the performance of both models improves as the dimensionality of the data increases, and the benchmark model is outperformed by both models in medium and high dimensional data.

6.2 Future work

The methodology proposed in this thesis can be extended and expanded in various directions. Below we outline a few of those.

In the thesis so far we assumed that A the matrix of coefficients of the inverse Wishart process used to generate the volatility process, is time-invariant. This assumption aids the support of the autoregressive Wishart process as a means of generating the volatilities which are mean-stationary. Indeed, this assumption on similar Wishart processes is adopted in the literature, see e.g. Philipov and Glickman (2006). However, as this thesis demonstrates there might be financial returns for which this assumption is weak or even invalid. There might be a case for an evolution of A, in particular considering an application over a long time-span. A future direction in research can be devoted to extending the models by considering a process for A. One possibility is to assume a relatively smooth evolution of A, considering that it is generated by some random walk process with low variance. Another possibility is to consider a jump process of A which can allow of structural changes on the volatility. The latter approach can be seen as a threshold AR model for covariance matrices and hence it will generalize the well known threshold autoregressive models for the mean.

A second direction of research involves inference around the mean of the returns. In this thesis (Chapters 3 and 5) a time-invariant mean is used and this is estimated just as the mean of historical returns. The rationale

of this set-up is that this mean should be close to zero and it is rather common to set it equal to the historical mean, see e.g. Aguilar and West (2000) and references therein. However, as new data come in, it is expected that structural changes of the volatility process may well affect the return process. Hence it is appropriate to consider a more structural model for the returns. The methodology proposed in this thesis can be extended by considering a state-space evolution of the returns, typically consisting of a vector autoregressive component. This will allow simultaneous inference of the returns (inference around the mean) and the volatility (inference around the covariance and correlation structure). Computationally, this approach will be more expensive and the challenge will be to retain as much as possible of the complexity without compromising too much on the efficiency, e.g. by introducing some sparsity in the autoregressive structure.

A third line of future research involves the extension of the proposed methodology in order to cater for higher dimensional data. In recent years big data have dominated the finance industry, in particular provided the automation of data collection in hedge funds and investment boutiques. Nowadays portfolio and asset management require the consideration of a large number of assets, sometimes from different markets. The models proposed in this thesis, provide an advantage to traditional MCMC inference for stochastic volatility, which is slower and not adaptive. However, the particle-based filtering is still slow when considering a large number of assets. The proposed methodology can be extended by considering sparsity on the parameters in

order to cater for higher dimensional financial data. For example a factor model or a model with sparse components may be developed.

Chapter 7

Appendix

7.1 Sequential Monte Carlo methods

SMC method is developed to solve space-state-structure or hidden Markov chain and nonlinear filtering models. These models often show the properties of nonlinear, non-Gaussian and with hidden variables (latent variable). Firstly, assume we have the following models:

- We have the observation data y_t and we assume X_t as the latent variable.
- Observation model: $f = p(y_t|X_t)$, where $p(y_t|X_t)$ could be a non-Gaussian function.
- Latent variable model: $\pi = p(X_t|X_{t-1})$, where $p(X_t|X_{t-1})$ could be both nonlinear and non-Gaussian.

- Initial state: $p(X_0)$

In this setup, the observable variables are the observations and they are related to the latent variable by $p(y_t|X_t)$, the latent variables are also described by the dynamical system $p(X_t|X_{t-1})$. The object is to approximate the posterior distribution for latent variable given the observations. The SMC algorithm provides the sequential estimate of X_t given the data y_t at any time t.

7.1.1 Importance sampling

Basically, we can regard SMC as an extension of importance sampling (IS). So we will recall the IS method first as follows:

- Assume we are interested in approximating an integral:

$$I = \int f(x)\pi(x)dx = E[f(x)]$$

- If we can sample $x_1, x_2, \ldots, x_n$ from $\pi(x)$, then we can approximate I by a Monte Carlo estimate:

$$\hat{I} = \frac{1}{n}\sum_{i=1}^{n} f(x_i)$$

- However, usually it is difficult to directly sample from $\pi(x)$.

- If we can sample from another function $g(x)$, then we can rewrite the

integral as:

$$I = \int f(x)\pi(x)dx = \int f(x)g(x)\frac{\pi(x)}{g(x)} = E_g[f(x)w(x)],$$

where $w(x) = \frac{\pi(x)}{g(x)}$, $g(x)$ is known as an importance density function and $w(x)$ is the weight.

- then we will have:

$$\hat{I} = \frac{1}{n}\sum_{i=1}^{n} f(x_i)w(x_i)$$

- However, in Bayesian statistics, the target distribution $\pi(x)$ is usually evaluated up to a normalising constant. Thus we can rewrite the expectation integral as

$$\begin{aligned} I &= \frac{\int f(x)\pi(x)dx}{\int \pi(x)dx} \\ &= \frac{\int f(x)\frac{\pi(x)}{g(x)}g(x)dx}{\int \frac{\pi(x)}{g(x)}g(x)dx} \end{aligned}$$

- Then we can sample N times $x^{(i)}$ from $g(x)$ and define the weights as:

$$w_i = \frac{\frac{\pi(x_i)}{g(x_i)}}{\sum_i^N \frac{\pi(x_i)}{g(x_i)}}$$

and we can get:

$$\sum_{i=1}^{N} w_i = 1$$

- The importance sampling provides pairs X_{it} and w_{it} so that integrals $\hat{I}$ are allowed. Finally we have:

$$\hat{I} = \sum_{i=1}^{N} w_i f(x_i).$$

7.1.2 SIS and hybrid method

Now applying importance sampling to the model, we have the weight to be calculated as follows:

- $$w_t = \frac{p(\Sigma_{0:t}|y_{1:t})}{g_t(\Sigma_{0:t}|y_{1:t})} \tag{7.1}$$

- Because of the definition of the conditional probability distribution, equation (7.1) can be written as:

$$w_t \propto \frac{p(\Sigma_{0:t}, y_t)|y_{1:t-1})}{g_t(\Sigma_{0:t}|y_{1:t})} \tag{7.2}$$

- Recall the Bayesian rule:

$$p(x|y_1, y_2, \ldots, y_t) = \frac{p(y_t|x_t)p(x_t|y_1, y_2, \ldots, y_{t-1})}{p(y_t|y_1, y_2, \ldots, y_{t-1})}$$

- Applying the Bayesian rule above to equation (7.2), we have:

$$\begin{aligned} w_t &\propto \frac{p(\Sigma_t, y_t|\Sigma_{0:t-1}, y_{1:t-1})p(\Sigma_{0:t-1}|y_{1:t-1})/p(y_t|y_1,\ldots,y_{t-1})}{g_t(\Sigma_t, y_t|\Sigma_{0:t-1}, y_{1:t-1})g_t(\Sigma_{0:t-1}|y_{1:t-1})/g_t(y_t|y_1,\ldots,y_{t-1})} \\ &\propto \frac{p(\Sigma_t, y_t|\Sigma_{0:t-1}, y_{1:t-1})p(\Sigma_{0:t-1}|y_{1:t-1})}{g_{t|t-1}(\Sigma_t|\Sigma_{0:t-1}, y_{1:t-1})g_{t|t-1}(\Sigma_{0:t-1}|y_{1:t-1})} \\ &\propto \frac{p(\Sigma_t, y_t|\Sigma_{0:t-1}, y_{1:t-1})}{g_{t|t-1}(\Sigma_t|\Sigma_{0:t-1}, y_{1:t-1})}w_{t-1} \end{aligned} \tag{7.3}$$

- Again we can apply the definition of the conditional probability distribution to equation (7.3):

$$\begin{aligned} w_t &\propto \frac{p(y_t|\Sigma_t)p(\Sigma_t|\Sigma_{t-1})}{g_{t|t-1}(\Sigma_t|\Sigma_{0:t-1}, y_{1:t-1})}w_{t-1} \\ &\propto \frac{p(y_t|\Sigma_t)p(\Sigma_t|\Sigma_{t-1})}{g(\Sigma_t|\Sigma_{t-1}, y_t)}w_{t-1} \end{aligned} \tag{7.4}$$

Equation (7.4) shows that the current weight, w_t, can be derived from the past weight w_{t-1}, using Σ_t, which is drawn from importance density $g(\Sigma_t|\Sigma_{t-1}, y_t)$ and the data y_t.

With this weight updating process, we can have the sequential importance sampling (SIS) algorithm: However, SIS is proved to be not an efficient method in practice. This is because, after performing a certain numbers of updating processes, some points will have relatively large weights and other points will have almost no weight at all. This situation will lead to a deterioration in the follow-up Monte Carlo estimation.

In order to solve this problem, effective sample size (ESS) has been created

Algorithm 5 Seqential Importance Sampling algorithm

1: Initialise: Draw $\Sigma_0^1, \ldots, \Sigma_0^N$ independently from $\pi(\Sigma_0)$ and set:

$$w_0^i = \frac{1}{N}$$

2: **for** $t = 1, \ldots, T$ **do**

3: **for** $i = 1, \ldots, N$ **do**

4: Draw Σ_t^i from $g_{t|t-1}(\Sigma_t|\Sigma_{0:t-1}, y_{1:t-1})$.

5: Set

$$\hat{w}_t^i = \frac{p(y_t|\Sigma_t)p(\Sigma_t|\Sigma_{t-1})}{g(\Sigma_t|\Sigma_{t-1}, y_t)}\hat{w}_{t-1}^i$$

6: Normalise the weights:

$$w_t^i = \frac{\hat{w}_t^i}{\sum_{j=1}^N w_t^j}$$

7: **end for**

8: Compute:

$$\hat{\pi}_t = \Sigma_{i=1}^N w_t^i \delta_{\Sigma_{0:t}^i}$$

where δ_x denotes the Dirac delta mass located at x.

9: **end for**

as a criterion. ESS is defined as follows:

$$N_{ESS} = \frac{1}{\sum_{i=1}^{N}(w_t^i)^2}$$

When the value of ESS is below a certain threshold, say N_0, then a re-sampling step will be performed among $\{\Sigma_t^i\}$ and the weights will also be reset to $1/N$, which forces all the re-sampled points to have same weight again. It should be noted that in practice, the threshold N_0 is often set up to be $N/2$. This method is called hybrid sequential Monte Carlo method and the algorithm is demonstrated in algorithm 6.

7.1.3 Bootstrapping filter

The bootstrapping filter, which is also known as condensation, is probably the simplest particle filter.

The main idea of this filter is that we set the importance density function to be:

$$g(\Sigma_t|\Sigma_{t-1}, y_t) = p(\Sigma_t|\Sigma_{t-1})$$

so that the weight function becomes:

$$\begin{aligned} \hat{w}_t^i &= \frac{p(y_t|\Sigma_t)p(\Sigma_t|\Sigma_{t-1})}{g(\Sigma_t|\Sigma_{t-1}, y_t)}\hat{w}_{t-1}^i \\ &= p(y_t|\Sigma_t)\hat{w}_{t-1}^i \end{aligned}$$

In this way it is obvious that the particles are drawn from the observed model

directly and the information provided from the latent variable model does not affect the particles. Therefore, Petris et al. (2009) suggest that 'most of the time the generated particles will fall in regions of low posterior density.' This leads to an inaccurate posterior density and high Monte Carlo variance for the simulated particles.

7.2 Probability distribution

7.2.1 Skew-t distribution

The skew-t distribution was invented to describe data with both heavy tails and skewness. It should be noted that there is no 'standard' version of the skew-t distribution. There are some popular definitions of the skew-t distribution which are proposed by Azzalini et al. (2003), Aas and Ingrid (2006) and Carmen and Steael (1998). In my research I chose Azzalini et al. (2003)'s distribution for two reasons. Firstly, they used the same idea as a skew-normal distribution to define a skew-t distribution, which makes it easy to understand and operate during my research. Secondly, there is a whole R package, package sn, supporting this definition so that it is convenient to code and calculate later on.

$st(\mu, \Phi, \alpha, v)$ stands for the skew-t distribution with four parameters, where μ is defined as a 'location' parameter related to the expectation, Φ is defined as a 'shape' parameter related to the standard variance, α is the skewness factor and v is the degrees of freedom. The probability density

Algorithm 6 Hybrid algorithm

1: Initialise: Draw $\Sigma_0^1, \ldots, \Sigma_0^N$ independently from $\pi(\Sigma_0)$ and set:

$$w_0^i = \frac{1}{N}$$

2: Set $N_0 = N/2$

3: **for** $t = 1, \ldots, T$ **do**

4: **for** $i = 1, \ldots, N$ **do**

5: Draw Σ_t^i from $g_{t|t-1}(\Sigma_t|\Sigma_{0:t-1}, y_{1:t-1})$.

6: Set

$$\hat{w}_t^i = \frac{p(y_t|\Sigma_t)p(\Sigma_t|\Sigma_{t-1})}{g(\Sigma_t|\Sigma_{t-1}, y_t)}\hat{w}_{t-1}^i$$

7: Normalise the weights:

$$w_t^i = \frac{\hat{w}_t^i}{\sum_{j=1}^N w_t^j}$$

8: **end for**

9: Compute:

$$N_{ESS} = \frac{1}{\sum_{i=1}^N (w_t^i)^2}$$

10: If $N_{ESS} < N_0$, then re-sample:

- Sample from previous $\{\Sigma_t^i\}$ with probabilities equal to normalised weights $\{w_t^i\}$
- Reset the weight to be $w_t^i = 1/N$

11: Compute:

$$\hat{\pi}_t = \Sigma_{i=1}^N w_t^i \delta_{\Sigma_{0:t}^i}$$

where δ_x denotes the Dirac delta mass located at x.

12: **end for**

function of the skew-t distribution is as follows:

$$f(x|\mu,\Phi,\alpha,v) = \frac{\Gamma(\frac{v+p}{2})}{|\Phi|^{\frac{1}{2}}(\pi v)^{\frac{p}{2}}\Gamma(v/2)}(1+\frac{Q}{v})^{-(\frac{v+p}{2})},$$

where p is the dimensionality of the data and $Q = (x-\mu)^{-1}\Phi^{-1}(x-\mu)$. The mean and variance of a skew-t distribution can be calculated as follows:

$$\begin{aligned} E(X) &= \mu + \alpha(\frac{v}{v-2}) \\ var(X) &= \frac{v}{v-2}\Phi + \alpha^2\frac{2v^2}{(v-2)^2(v-4)} \end{aligned}$$

7.2.2 Gamma distribution

The gamma distribution is a commonly used two parameter probability distribution; the well-known exponential distribution and chi-squared distribution are special cases of it. The probability density function of the gamma distribution is as follows:

$$f(x|k,\theta) = \frac{1}{\Gamma(k)\theta^k}x^{k-1}e^{-\frac{x}{\theta}},$$

where k is the shape parameter and θ is the scale parameter.

The mean of the distribution can be calculated as $E(X) = k\theta$. The variance of the distribution can be calculated as $Var(X) = k\theta^2$. One of the most important properties of the gamma distribution is that it can be a conjugate prior distribution in Bayesian inference.

7.2.3 Wishart distribution

The Wishart distribution is the multivariate extension of the gamma distribution. It has two parameters. Anderson (1962) believes it is a very important distribution for estimation of covariance matrices for multivariate statistics study. Assume we have a matrix $\mathbf{X}$ follows a Wishart distribution, then the probability density function of the Wishart distribution is as follows:

$$f(\mathbf{X}|k,\Sigma) = \frac{|\mathbf{X}|^{\frac{n-p-1}{2}} e^{-\frac{tr(\Sigma^{-1}\mathbf{X})}{2}}}{2^{\frac{np}{2}}|\Sigma|^{\frac{n}{2}}\Gamma(\frac{n}{2})},$$

where n is the degrees of freedom, p is the dimensionality of the data and Σ is the scale matrix.

The mean of the distribution can be calculated as $E(\mathbf{X}) = n\Sigma$. The covariance matrix is complex so it will not be discussed here. As an inheritance from the gamma distribution, the Wishart distribution also has the conjugation property and it is a suitable conjugate prior for the precision matrix in Bayesian statistics.

7.2.4 Inverse Wishart distribution

The inverse Wishart distribution is the multivariate extension of the inverse gamma distribution. The inverse Wishart distribution is related to the Wishart distribution: if we have $\mathbf{X} \sim W(k,\Sigma)$, then $\mathbf{X}^{-1} \sim IW(k,\Sigma^{-1})$. It is a conjugate prior for the precision matrix of the multivariate Gaussian distribution in Bayesian statistics. The probability density function of the

Inverse Wishart distrbution is as follows:

$$f(\mathbf{X}|k,\Sigma) = \frac{|\Sigma|^{\frac{v}{2}}}{2^{\frac{vp}{2}}\Gamma_p(\frac{v}{2})}|\mathbf{X}|^{\frac{v+p+1}{2}}e^{-\frac{1}{2}tr(\Sigma\mathbf{X}^{-1})} \quad ,$$

where p is the dimensionality of the data. The mean of the inverse Wishart distribution is $E(\mathbf{X}) = \frac{\Sigma}{v-p-1}$ and the mode is $M(\mathbf{X}) = \frac{\Sigma}{v+p+1}$.

newpage

Bibliography

Aas, K. and Ingrid, H. (2006). The generalized hyperbolic skew students t-distribution.

Anderson, T. (1962). *An introduction to multivariate statistical analysis*, Wiley.

Arulampalam, M. S., Maskell, S. and Gordon, N. (2002). A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking, *IEEE Transactions on signal processing*, **50**: 174–188.

Asai, M. (2006). Comparison of mcmc methods for estimating garch models.

Azzalini, Adelchi and Antonella, C. (2003). Distributions gererated by perturbation of symmetry with emphasis on a multivariate skew t-distribution, *Statistical methodology*, **65**: 367–389.

Azzalini, A. and Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution.

Ball, Ray and Ross, W. (1972). Some time series properties of accounting income, *The journal of Finance*, **27**: 663–681.

Bauwens, Laurent, L. S. and Rombouts, J. V. (2006). Multivariate garch models: a survey., *Journal of Applied Econometrics*, **21.1**: 79–109.

Black, F. and Litterman, R. (1992). Global portfolio optimization, *Financial Analysts Journal*, **5**: 11–17.

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities, *Journal of Political Economy*, **81**: 637–654.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticitys, *The Journal of Econometrics*, **31**: 307–327.

Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: A multivariate generalized arch model, *Review of Economics and Statistics*, **72**: 498–505.

Box and Jenkens (1970). *Time series analysis: Forcasting and control*, San Francisco:Holden-Day.

Brillinger and David, R. (2000). *Time series : General*, Berkeley.

Brockwell, P. and Davis, R. (2009). *Time series: Theory and Methods*, springer.

Cappe, O., Moulines and Ryden, T. (2005). *Inference in Hidden Markov Models*, Springer.

Carmen, F. and Steael, M. (1998). On bayesian modeling of fat tails and skewness.

Cavaliere, Giuseppe and Taylor, A. M. (2007). Testing for unit roots in time series models with non-stationary volatility, *Journal of Econometrics*, **140**: 919–947.

Doucet, Freitas, D. and Gardon, N. (2001). *Sequential Monte Carlo in Practice*, Springer.

Doucet, A. and Johansen, A. M. (2009). A tutorial on particle filtering and smoothing: Fifteen years later., *Handbook of nonlinear filtering*, **12**: 656–704.

Elton, Edwin and Gruber, M. (1997). Modern portfolio theory, 1950 to date, *Journal of Banking and Finance*, **21**: 1743–1759.

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

Engle, R. and Kroner, K. (1995). Multivariate simultaneous generalized arch, *Econometric Theory*, **11**: 122–150.

Engle, R. F. (2002). Dynamic conditional correlationa simple class of multivariate garch models. journal=Journal of Business and Economic Statistics, volume=20, pages=339-350.

Engle, R. F., Ledoit, O. and Wolf, M. (2016). Large dynamic covariance matrices.

Fisher, R. (1925). *Statistical methods for research workers*, Genesis Publishing Pvt Ltd.

Fuller, W. (2009). *Introduction to statistical time series*, Wiley.

Gelfand, A. and Smith, A. (1990). Sampling-based approaches to calculating marginal densities, *Journal of the American Statistical Association*, **85**: 398–409.

Geman, S. and Geman, D. (1984). Stochastic relaxation, gibbs distribution, and the bayesian restoration of images, *JIEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**: 721–741.

Geyer, C. (1994). On the convergence of monte carlo maximum likelihood calculations, *Journal of the Royal Statistical Society*, **56**: 261–274.

Ghose, Devajyoti and Kroner, K. (1995). The relationship between garch and symmetric stable processes: Finding the source of fat tails in financial data, *Journal of Empirical Finance*, **2**: 225–251.

Gilks, Best and Tan (1995). Adaptive rejection metropolis sampling within gibbs sampling.

Gilks, W. (2005). *Markov chain monte carlo*, John Wiley and Sons.

Gourieroux, C., Jasiakb, J. and Sufanab, R. (2009). The wishart autoregressive process of multivariate stochastic volatility, *Journal of Econometrics*, **150**: 167–181.

Gourierouxa, C., Jasiakb, J. and Sufanab, R. (1987). The calculation of posterior distributions by data augmentation(with discussion), *Journal of the American Statistical Association*, **82**: 528–550.

Green, P. (1995). Reversible jump markov chain monte carlo computation and bayesian model determination, *Biometrika*, **82**: 711–732.

Hamilton, J. (1994). *Time Series Analysis*, Princeton Univ.

Harvey, A., Ruiz and Shephard, N. (1994). Multivariate stochastic variance model.

Hastings, W. (1970). Monte carlo sampling methods using markov chains and their applications, *Biometrika*, **57**: 97–109.

Jacquier, E., Nicholas, P. G. and Peter, R. E. (2002). Bayesian analysis of stochastic volatility models, *Journal of Business andd Economic statistics*, **20**: 69–87.

Jacquier, E., Polson, N. and Rossi, P. (1994). Bayesian analysis of stochastic volatility modelsl.

Jeantheau, T. (1998). Strong consistency of estimators for multivariate arch models, *Econometric Theory*, **14**: 70–86.

Johnson, Norman, L., Kotz, S. and Balakrishnan, N. (2002). *Continuous multivariate distributions*, John Wiley and sons.

Jondeau, Eric, Poon, S.-H. and Rockinger, M. (2007). *Financial modeling under non-Gaussian distributions*, Springer.

Jorion, P. (1997). *Value at Risk: The new benchmark for conrtolling market risk*, The McGraw-Hill Company.

Kay, Steven, M. and Stanley, L. (1981). Spectrum analysis- a modern perspective, *Proceedings of the IEEE*, **69**: 1380–1419.

Krawiecki, Andrzej, Holyst and Helbing (2002). Volatility clustering and scaling for financial time series due to attractor bubbling, *Physical review letters*, **89**: 158–701.

Lawrance, A. J. and Lewis, P. A. W. (1977). An exponential moving-average sequence and point process, *Journal of Applied Probability*, **20**: 98–113.

Liu, J. (1991). *Monte Carlo strategies in scientific computing*, Springer.

Lubatkin, Michael and Chatterjee, S. (1994). Extending modern portfolio theory into the domain of corporate diversification: does it apply?, *Academy of Management Journal*, **37**: 109–136.

Lutkepohl, H. (2006). *New Introduction to Multiple Time Series Analysis*, Springer.

Lux, Thomas and Marchesi, M. (2000). Volatility clustering in financial markets:a miscrosimulation of interacting agents, *International Journal of Theoretical and Applied Finance*, **3**: 675–702.

Mandelbrot, B. (1963). The variation of certain speculative prices, *The Journal of Business*, **36**: 394–419.

Markowitz, M. H. (1952). Portfolio selection, *The journal of finance*, **7**: 77–91.

Markowitz, M. H. (1959). *Portfolio selection: Efficient Diversificaton of investments*, Wiley.

Melino, A. and Turnbull, S. (1990). Pricing foreign currency options with stochastic volatility.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953). Equation of state calculations by fast computing machines, *Journal of Chemical Physics*, **21**: 1087–1092.

Michie, Donald, David, J. and Charles, C. (1994). *Machine learning, neural and statistical classification*, Springer.

Moral, D. and Pierre (1996). Non linear filtering: Interacting particle solution, *Markov Processes and Related Fields*, **2**: 555–580.

Muller and Heinz, H. (1988). Modern portfolio theory, *Astin Bulletin*, **18**: 127–145.

Noureldin, D., Shephard, N. and Sheppard, K. (2014). Multivariate rotated arch models, *Journal of Econometrics*, **179**: 16–30.

Peiro and Amado (1999). Skewness in financial returns, *Jounal of Banking and Finance*, **23**: 847–862.

Petris, G., Petrone, S. and Campagnoli, P. (2009). *Dynamic linear models with R*, Springer.

Pham and Hoang (2006). *Springer handbook of engneering statistics*, Springer.

Philipov, Alexander and Glickman, M. E. (2006). Multivariate stochastic volatility via wishart processes, *Journal of Business and Economic Statistics*, **24**: 311–334.

Poor and Vincent, H. (1988). *An introduction to signal detection and estimation*, Springer.

Rom and Ferguson, K. (1994). Post-modern portfolio theory comes of age, *Journal of Investing*, **3.3**: 11–17.

Rosenbluth, Marshall, N. and Arianna, W. (1955). Monte-carlo calculations of the average extension of macromolecular chains, *J. Chem. Phys*, **23**: 356–359.

Sewell, M. (2011). *Characterization of financial time series*, UCL.

Shumway, R. and Stoffer, D. (2000). *Time Series Analysis and Its Applications*, Springer.

Silvennoinen, A. and Terasvirta, T. (2009). *Handbook of financial time series*, Springer Berlin Heidelberg.

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

Taylor and Stephen, J. (2008). *Modelling financial time series*, World scientific publishing.

Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion).

Triantafyllopoulos, K. (2012). Multivariate stochastic volatility modeling using wishart autoregressive processes, *Journal of Time Series Analysis*, **33**: 48–60.

Tsay, R. (2002). *Analysis of Financial Time Series Analysis*, Wiley.

Uhlig, H. (1997). Bayesian vector autoregressions with stochastic volatility, *Econometrica*, **65**: 59–73.

VanDerWeide, R. (2002). Go-garch: a multivariate generalized orthogonal garch model., *Journal of Applied Econometrics*, **17.5**: 549–564.

Wei, W. and William, W. (1994). *Time series Analysis*, Wesley.

Weigend and Andreas, S. (1994). *Time series prediction: forecasting the future and understanding the past*, Santa Fe Institute.