Estimation and Detection of Network Variation in Intraday Stock Market

A Dissertation Presented

by

Shanshan Li

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Applied Mathematics and Statistics

Stony Brook University

November 2016

Stony Brook University

The Graduate School

Shanshan Li

We, the dissertation committee for the above candidate for the

Doctor of Philosophy degree, hereby recommend

acceptance of this dissertation.

Haipeng Xing - Dissertation Advisor
Associate Professor, Department of Applied Mathematics and
Statistics

Raphael Douady - Chairperson of Defense
Professor, Department of Applied Mathematics and Statistics

Jiaqiao Hu - Member
Associate Professor, Department of Applied Mathematics and
Statistics

Keli Xiao - Outside Member
Assistant Professor of Finance, College of Business, SUNYSB

This dissertation is accepted by the Graduate School.

Charles Taber
Dean of the Graduate School

ii

Abstract of the Dissertation

Estimation and Detection of Network Variation in Intraday Stock Market

by

Shanshan Li

Doctor of Philosophy

in

Applied Mathematics and Statistics

Stony Brook University

2016

The last few years have witnessed an exponential growth in the collection and analysis of financial market data. Investigating the interactions between the dynamics of the financial system and extracting useful information from these multivariate data streams can help us in improving our understanding of the underlying backbone in the financial market. These massive noisy data sets require the application of suitable and efficient dependency measurements for their analysis in a real-time environment. And that is why network analysis has emerged recently, which is a plausible representation helps interpret the hidden interconnection between the elements in large datasets. However, most frequently used methods in this area have certain limitations, such as the com-

putational complexity or the assumption of a temporally invariant network. This thesis has two major purposes; firstly, to construct time-varying networks by presenting two new approaches to dynamically measure symmetric and asymmetric interactions; and secondly, to detect the structural breaks in the high dimensional time series of the financial market.

Building on previous work, we propose two computationally efficient approaches based on partial correlation network and vector autoregressive adjacency network. Since both of these estimators are under the high-dimension-low-sample-size setting, we develop a penalized kernel smoothing method for the problem of selecting non-zero elements of the time-varying matrix. The network structure of multivariate financial time series are established for the first time for such estimators and displayed in a graphical representation. Furthermore, we consider the problem of efficient financial surveillance aimed at prompt detection of structural breaks in the market. Assuming the model evolves in a piece-wise constant fashion, we study four types of detection rules, including statistical process control chart, generalized likelihood ratio detection rule, a detection method based on an extension of Shiryaev's Bayesian single change point model and a sequential detection rule for multiple change points. The efficiency of the proposed methods is demonstrated on both simulation studies and the empirical analysis focusing primarily on the intraday stock market. Our findings shed a new light on uncovering the hidden interactions between the financial dynamics and present new insight into market structure and market stability.

I dedicate this thesis to my parents, my grandma and

Tongfei.
I love you all dearly.

Table of Contents

List of Figures

x

List of Tables

Acknowledgements

Firstly, I would like to express my sincere gratitude to my advisor, Professor Haipeng Xing, for the continuous support of my Ph.D study and related research, for his patience, motivation, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my Ph.D study.

Besides my advisor, I would like to thank the rest of my thesis committee: Professor Raphael Douady, Professor Jiaqiao Hu, and Professor Keli Xiao, for their insightful comments and encouragement, but also for the hard questions which incented me to widen my research from various perspectives.

My sincere thanks also goes to Professor James Hicks and Professor Michael Wigler, who provided me an opportunity to join their team as intern, and who gave access to the laboratory and research facilities. Without their precious support it would not be possible for me to improve my quantitative skills and achieve large-data analysis experience.

I especially thank Dr. Tongfei Guo, who has been a constant source of strength and inspiration. I can honestly say that it was only his determination and constant encouragement that ultimately made it possible for me to see this project through to the end. Also I owe a lot to my parents Dr. Jisheng Li and

Part I

Time-Varying Network Analysis of Price and

Volume Change in Intraday Stock Market

Chapter 1

Introduction

1.1 Market Macrostructure

Market macrostructure studies the price formation process, and how this process is affected by the organization of the market. Normally, we analyze trades and price dynamics when there are significant turbulences observed in the market. The certain significant events induce a transient drop or raise in the willingness and ability of financial institutions to hold assets such as stocks or bonds. Alternatively, an aggregate turbulence in market can reflect events affecting the overall financial situation of a category of institutions, e.g., funds experiencing large outflows or losses, banks incurring large losses in Berndt (2005), or specialists building extreme positions in Comerton (2010).

Technical analysts of market data believe that trading price and volume data provide indicators of future price movement. In particular, Brown and Jennings(1989) consider rational expectation models in which a single price does not reveal the underlying information but a sequence of security price does. They demonstrate that technical analysis of price patterns may be valu-

able because stock markets behave as complex dynamic systems, and as such, it is critical to investigate the dependencies between the dynamics of the system variables. It is common to associate such interactions with the notion of correlation. Indeed, much effort is dedicated to study and understand such stock cross-correlations in an attempt to extract maximum market latent information that is embedded in the interactions between the market variables, see Podobnik (2009), Morck (2000), Krishna (2009), Fizeau (2001) and Huang (2013).

One prominent feature in financial markets is the presence of an observed correlation, no matter positive or negative, between the price movements of different financial dynamics. The presence of a high degree of cross-correlation between the synchronous time evolution of a set of equity returns is a well known empirical fact in Markowitz (1952) and Eltonet (2009). The Pearson(1895) correlation coefficient provides information about the similarity in the price change behavior of a given pair of stocks. Much effort has been devoted to extracting meaningful information from the observed correlations in order to gain insights into the underlying structure and dynamics of financial markets, see Embrechtset (2002), Morcket (2000), Campbellet (2008), Krishanet (2009), Asteet (2010), Campbellet (2008), Cizeauet (2001) and Huanget (2013), Forbes and Rigobon (2002).

1.2 Network Analysis

Nowadays, giant datasets are collected with lots of empirical information about the functioning of almost every field of study, at a cost much lower than

a few decades ago, for instance biotechnology in McBride (2012), medical science in Groves (2013), and in particular business and economics study in Einav and Levin (2013). Especially after financial crisis of 2007, there are a number of researchers make contributions on network estimation including Pellizzon (2012), Diebold and Yilmaz (2013) and Hautsch and Schienle (2012, 2013).

One can be interested in the existing linkages between any different elements, namely stock market returns of S&P 500, inflation rates in top economies and so on, which included in a collection of the dataset. And that is the reason why network analysis has emerged in recent years. Network analysis is used to help interpret the hidden interconnections between different elements in large dataset. With the application of proper statistical tools, analysts can not only get the statistical results about the data, but also plot the interconnections in a graphical representation that eases the explanation of the real market observations. That is to say, network analysis allows us to construct graphs representing the reality behind those complex empirical datasets.

In mathematics, the traditional way of representing networks is using graphs, which can be generally defined as a collection of nodes connected by lines. Here we will introduce a bit of graph theory. A graph is an ordered pair as following:

$$G = (V; \epsilon) \tag{1.1}$$

The first one represents nodes while the second one is the edge connecting nodes. In our study, we consider two main characteristics of a network: direc-

4

tionality and weight. First, If an edge from node i to node j is different from an edge from j to i, then the graph is directed. On the other hand, if all the links between nodes don't have a particular direction, then the graph is undirected. Next, The difference between weighted and unweighted networks has to do with the relative weight of each edge. In weighted networks, the thickness of the edge depends on the intensity of the correlation between two nodes. As described later on, due to the specific resources and purpose of our study, we use undirected and weighted graphs for partial correlation networks while directed and weighted graphs for vector autoregressive adjacency networks .

The graphical representation of a network has its origin in what is known as the adjacency matrix A. it is the compact representation of a network with the form of a matrix. It vertices i and j are connected by an edge, then the A_{ij} element has a value different form zero. In the case of weighted graphs, A_{ij} can take any value as the network represent does not identify just the existence of a linkage, but also specifies the weight of it.

As briefly introduced in the above section, the final aim of network analysis is to represent large data collection as a network with which would be easy to interpret the linkages between different elements. In this article, our data collection is a multivariate time series, which consists of sequences of values of several contemporaneous variables changing with time. That is to say, we will apply network analysis using values of a determined number of variables taken in successive periods of time. In our study, the stock prices and trading volumes at different time points for each stock can be regarded as a time series. Without loss of generality, we suppose that we have N variables for T periods

of time. So the input data will be

$$y_t = \begin{bmatrix} y_{t1} \\ y_{t2} \\ ... \\ y_{tN} \end{bmatrix} \; for \; t = 1, ...T \tag{1.2}$$

And the final output of the network analysis of the particular multivariate time series could be displayed in a graph, in which all the target stocks are represented by nodes and the interconnections between them are plotted as the edges linking the nodes.

In our research, we consider two kinds of network model: partial correlation network and VAR adjacency network. Building on previous work, we propose two computationally efficient approaches under the high-dimension-low-sample-size setting. we develop a penalized kernel smoothing method for the problem of selecting non-zero elements of the time-varying matrix. As the size of samples T and the number of variables N increases, the computation cost turns out to be a serious problem. So we make formulate a novel algorithm named Multi-Active-Shooting(MAS) algorithm. As far as we know, it is the first time that this algorithm is used to solve the regression problem with dataset of hundreds of dimensions. The network structure of multivariate financial time series are established for the first time for such estimators and displayed in a graphical representation. Hence, we confirm the efficiency and validity of our proposed methods, especially in the case of presenting cross-sectional interconnections between stocks in financial market. Our findings

6

shed a new light on uncovering the hidden interactions between the financial dynamics and present new insight into market structure and market stability.

1.3   Outline

The rest of the thesis is organized as follows. In Chapter 2, we explain the characteristics of the studied models, including partial correlation network in Section 2.1 and VAR adjacency network in Section 2.2, from the basic idea up to the features of the model behind it. Afterwards, in Chapter 3, we give full details about the proposed methods with kernel smoothing approach in Section 3.1 and computational efficient algorithms to estimate interconnections between elements in large data sets, including partial correlation network in Section 3.2 and VAR adjacency network in Section 3.3. In Chapter 4, we performed an illustrated simulation in order to show the power and efficiency of the model presented. Moreover, the proposed models are put into practice by analyzing a relatively large dat set of real world data, with the objective of assessing whether the proposed statistical instrument is valid and useful when applied to a real multivariate time series. Also, a summary of the main results and proposal about the future work are given. In short, the findings all along this part of thesis suggest the partial correlation network and VAR adjacency network both perform well under the assumption of sparsity of data. They are shown to be very valid tools to represent cross-sectional interconnections in between elements in large data sets and allow us to observe and analyze the existing linkages that could have been omitted otherwise.

Chapter 2

Two Models to Estimate Network Variation

2.1 Partial Correlation Network

To study the relationship between two stock returns, say y_i and y_j, the common method is to calculate the Pearson correlation coefficient:

$$\rho_{Pearson}(i,j) = \frac{(y_i - \mu_i) \times (y_j - \mu_j)}{\sigma_i \times \sigma_j} \tag{2.1}$$

where μ_* represents average value of y_* and σ_* denotes the standard deviation.

Despite the meaningful information provided by investigating the correlation coefficient, it lacks the capacity to provide information about whether a different variable eventually influence the observed relationship between these two variables. That is to say, in some cases, a strong correlation coefficient does not necessarily mean strong direct relation between two variables. As we know, two financial dynamics in the same market can be affected by common macroeconomic force and investor psychological factors, such as bandwagon effect. It is well known that there exists significant cross-correlation between

the synchronous time evolution of a pair of variables, for instance stock re-turns. To overcome this issue, we introduce partial correlation coefficient, which quantifies the correlation between variables when conditioned on one or several other variables, to get rid of the common driving factors.

Suppose that $(y_1, ...y_p)^T$ has a joint distribution with mean μ and covari-ance Σ, in which Σ is a p by p positive definite matrix. Denote the partial correlation between y_i and y_j by $\rho^{ij}(1 \leq i < j \leq p)$. ε_i and ε_j are re-spectively the prediction errors of the linear predictors of y_i and y_j based on $y_{-(i,j)} = \{y_k : 1 \leq k \neq i, j \leq p\}$. Thus, y_i is expressed as

$$y_i = \sum_{j \neq i} \beta_{ij} y_j + \varepsilon_i \tag{2.2}$$

where

$$var(\varepsilon_i) = \frac{1}{\sigma^{ii}} cov(\varepsilon_i, \varepsilon_j) = \frac{\sigma^{ij}}{\sigma^{ii}\sigma^{jj}} \beta_{ij} = -\frac{\sigma_{ij}}{\sigma_{ii}} = \rho_{ij} \times \sqrt{\frac{\sigma_{jj}}{\sigma_{ii}}} \tag{2.3}$$

Here we define concentration matrix $K \equiv \Sigma^{-1}$ by $(\sigma^{ij})_{p \times p}$. And the partial correlation ρ^{ij} is denotes as

$$\rho^{ij} = Corr(y_i, y_j | y_{-(i,j)}) = -\frac{\sigma_{ij}}{\sqrt{\sigma_{ii} \times \sigma_{ij}}})$$

In our article, we primarily deal with the particular case of high dimension. In other words, the number of variables, namely p, is always larger than the sample size, namely n. This high-dimension-low-sample-size assumption of the partial correlation network model is called the sparsity of data set. Sparsity

9

refers to the fact that the given network is not complete. That is to say, not every node is connected with any other node in the dataset. Recall that a zero element in the correlation matrix implies the absence of an edge between two variables, thus the correlation matrix of a sparse network contains a great number of zeros. Actually, this kind of sparse networks have been studied in many areas, e.g. genetics networks, social networks and so on. It is widely believed that most equity pairs are not directly interacting with each other, so we assume that the data from stock market can also be considered as sparse.

Based on the relationship between ρ^{ij} and β_{ij}, we find that the problem of searching for nonzero partial correlations can be regarded as a model selection problem under the regression setting in equation 2.2. Even if the number of variables is much larger than the sample size, we can still solve the the problem with effective dimension in a suitable range. In our research, we take advantage of Lasso for detecting pairs of stocks having nonzero partial correlations among a large dataset, by imposing the l_1 penalty item on the associated loss function to solve the problem.

Lasso, stands for Least Absolute Shrinkage and Selection Operator, presented by Robert (1996), has been a very effective tool to obtain the estimations of correlations. It allows to shrink the number of nonzero estimated coefficients low enough to end up with a sparse network. We consider the usual setup for linear regression. We have a response variable y_i and a predictor vector $y_{-(i)}$, in which $y_{-(i)} = \{y_k : 1 \leq k \neq i \leq p\}$. Lasso estimators are

calculated as following:

$$\beta_{lasso} = \arg\min_{\theta} \sum_{i=1}^{p} (y_i - y_{-(i)}^T \beta)^2 + \lambda \sum_{i=1}^{p-1} \mid \beta_i \mid \quad \lambda \geq 0 \qquad (2.4)$$

Thus, the Lasso estimators β_{lasso} are those coefficients that minimize the l_1 form of errors in the regression of each element on all the others. The estimators are calculated in a similar way to ordinary least squares(OLS) estimators, but take absolute values in the penalty item instead of squares. By shrinking some of the estimated coefficients to exact zeros, it perfectly fit for our purpose of ending up with a sparse network.

There is another great thing about Lasso estimation. We take into account the parameter denoted as λ, which controls for the amount of shrinkage in the estimation procedure. In other words, with proper λ, Lasso estimation selects the variables which better explain the linkage between them. It shrinks the estimator β_i corresponding to the certain variable which is not so explanatory to exact zero, while keeping the parameter of the worthy variables different from zero. In short, Lasso estimator turns out to be an optimal instrument for estimating the existence of partial correlations for the reason that it allows to shrink non-useful coefficients of the regression model to zeros and select worthy ones that can be used to represent sparse networks of multivariate time series, even in a scenario where there are much more variables than observations. Therefore, we introduce Lasso in the set up of our model.

In our research, we have n observations of p stocks over a time period. Each observation at a certain time point can be regarded as a sample in the

above problem settings. Thus, we suppose $Y^k = (y_1^k, y_2^k, ..., y_p^k)^T$ for $k = 1, ..., n$. Denote the sample of the ith variable as $Y_i = (y_i^1, y_i^2, ..., y_i^p)^T$ for $i = 1, ..., p$. Based on equation 2.2, we propose to estimate the partial correlation ρ^{ij} by the following loss function:

$$
\begin{aligned}
L_n(\mathrm{P}, \sigma, Y) &= \frac{1}{2} \left(\sum_{i=1}^{p} \omega_i \| Y_i - \sum_{j \neq i} \beta_{ij} Y_j \|^2 \right) \\
&= \frac{1}{2} \left(\sum_{i=1}^{p} \omega_i \| Y_i - \sum_{j \neq i} \rho^{ij} \times \sqrt{\frac{\sigma_{jj}}{\sigma_{ii}}} Y_j \|^2 \right) \quad (2.5)
\end{aligned}
$$

where

$$
\sigma = \{\sigma^{ii}\}_{i=1}^{p} \tag{2.6}
$$

$$
Y = \{Y^k\}_{k=1}^{n} \tag{2.7}
$$

$$
\omega = \{\omega_i\}_{i=1}^{p} \tag{2.8}
$$

$$
\mathrm{P} = (\rho^{12}, ..., \rho^{(p-1)p})^T \tag{2.9}
$$

As stated above, we impse an l_1 penalty item in the penalized loss function for estimating the partial correlation P:

$$
\begin{aligned}
\mathrm{L}_n(\mathrm{P}, \sigma, Y) &= L_n(\mathrm{P}, \sigma, Y) + \lambda \| \mathrm{P} \|_{l_1} \\
&= L_n(\mathrm{P}, \sigma, Y) + \lambda \sum_{1 \leq i < j \leq p} | \rho_{ij} | \quad (2.10)
\end{aligned}
$$

In this way, we formulate an Lasso-like regression which is performed separately for each variable on the rest of the variables. The outcome of the application of this model is the matrix containing the estimations of the partial

correlations between variables. Also, the network graphs related to the certain partial correlation matrix have a simple and intuitive way to transforming potentially complex and large dataset into a plain and clear representation of vertices and edges showing the existence or not of the interconnections in between the variables from a multivariate time series.

Although the above partial correlation network analysis has the advantage of easing the interpretation to representing the dependency between a large number of stocks over a short time period, it has an important drawback. The model is based on the interconnections between stocks only at the limited period of time, which may be regarded as contemporaneous dependence in some cases. Therefore, partial correlation network has limitations in the case of time series processes exhibiting serial dependence or spillover effect. Also, partial correlation network model comes out with the undirected partial correlation matrices. However, we may want to look into the relationship between the stocks with direction. In the next chapter, we will introduce vector autoregressive adjacency network to capture the lead or lag effects in the multivariate time series.

2.2 Multivariate Autoregressive Adjacency Network

As stated by Andrews and Monahan (1992), multivariate time series process can be represented as a Vector Autoregression (VAR) model with suitable time lags and for such processes the long run covariance matrix is in a closed form whose inverse is available. In this section, we propose a novel network definition for multivariate time series which attempts to overcome the limita-

tions of partial correlations.

Suppose that we have $X^t = (x_1^t, ..., x_p^t)'$ as the $t-$th observed vector over the time period $(1, ..., n)$, which is a vector autoregressive process of d, denoted as $VAR(d)$. It satisfies the recursion

$$X^t = \sum_{k=1}^{d} A_k X^{t-k} + V_t, \quad t = 1, ..., n \tag{2.11}$$

where $A_1, ..., A_d$ are $p \times p$ matrices of autoregression coefficients, k denotes the number of time lags. V_t are $p-$variate normal distribution with mean μ_V and covariance matrix Σ_V. Moreover, we assume that V_t is unrelated to X^s for $s < t$. We can see that given large number of variables p and time lags d, we have to make the estimation of parameters for a total number of $p^2 d$, which is greater than the number of observations np. Additionally, as there may be a number of structural change-point occurred in financial dynamics time series, the effective number of observations used for parameter estimation would be much smaller than the whole period $1, ..., n$. Thus, we still have to work out the method in high-dimension-low-sample-size settings. From the partial correlation network model, we take advantage of Lasso estimation for shrinking low enough the number of variables. We select the dominant variables which are more valuable in interpretation of the network and get rid of the ones with less influence on the parameters. Therefore, we also formulate a Lasso-type estimation for VAR adjacency network.

Let $\mathcal{X}$ be a $n \times p$ matrix of observations and $\mathcal{S} = \frac{1}{n}\mathcal{X}^T\mathcal{X}$ be the associated covariance matrix. And $\mathcal{X}^t$ is corresponding to the $t-$th time point in $\mathcal{X}$

14

while $\mathcal{X}_i^t$ be the $i-$th column. Under the general weighted lasso penalty, the estimation of adjacency matrix is formulated by the following l_1 regularized problems, for $i = 1, ..., p$:

$$\hat{A}_i = \arg\min_{\theta^t} \frac{1}{n} \parallel \mathcal{X}_i^n - \sum_{k=1}^{d} \mathcal{X}^{n-k} \theta^k \parallel^2 + \lambda \sum_{j=1}^{p} \mid \theta_j^k \mid \omega_j^k \qquad (2.12)$$

In this way, we have constructed an directed coefficient network which has similar representation as in partial correlation network. In the next section, we implement an iterative algorithm to complete the estimation of VAR adjacency matrix.

Chapter 3

Computational Efficient Methods in Time-Varying Network Analysis

3.1 Kernel Smoothing Approach

In this section, we introduce a class of regression techniques that achieve flexibility in estimating the regression function, say $f(X)$, by using only those observations close to the target point x_0. In such a way, the resulting estimated function $\widehat{f}(X)$ is smoothed. This localization is achieved via a weighting function of kernel $K_\lambda(x_0, x_i)$, which assigns a weight to x_i based on its distance from x_0. Recall Nadaraya-Watson kernel-weighted average

$$\widehat{f}(x_0) = \frac{\sum_{i=1}^{N} K_\lambda(x_0, x_i) y_i}{\sum_{i=1}^{N} K_\lambda(x_0, x_i)} \tag{3.1}$$

with the Epanechnikov quadratic kernel

$$K_\lambda(x_0, x) = D\left(\frac{\mid x - x_0 \mid}{\lambda}\right) \tag{3.2}$$

with

$$D_t = \begin{cases} \frac{3}{4}(1 - t^2) & if \mid t \mid \leq 1 \\ \\ 0 & otherwise \end{cases} \tag{3.3}$$

Thus, we have progressed from the raw moving average to a smoothly varying locally weighted average by using kernel weighting. Furthermore, locally weighted regression solves a separate weighted least square problem at each target point x_0:

$$\min_{\alpha(x_0),\beta(x_0)} \sum_{i=1}^{N} K_\lambda(x_0, x_i)[y_i - \alpha(x_0) - \beta(x_0)x_i]^2 \tag{3.4}$$

The estimate is then $\widehat{f}(x_0) = \widehat{\alpha}(x_0) + \widehat{\beta}(x_0)x_0$. Without stopping local linear fits, we can fit local polynomial fits of any degree d,

$$\min_{\alpha(x_0),\beta_j(x_0),\ j=1,...,d} \sum_{i=1}^{N} K_\lambda(x_0, x_i)[y_i - \alpha(x_0) - \sum_{j=1}^{d} \beta_j(x_0)x_i]^2 \tag{3.5}$$

with solution $\widehat{f}(x_0) = \widehat{\alpha}(x_0) + \sum_{j=1}^{d} \widehat{\beta}_j(x_0)x_0^j$. Since local linear fits can help bias dramatically at the boundaries at a modest cost in variance, we will apply this kernel smoother to the regression function of both the partial correlation network model and VAR adjacency model as noted in the previous section.

3.2 Proposed Method 1: Estimation of Partial Correlation Network

3.2.1 Implementation with MAS algorithm

In this section, we present a two-step iterative algorithm to estimate the parameters P and σ noted in the loss function in equation 2.10. The main idea is to specify an initial value for $\sigma^{(initial)}$ at first, and then obtain P by minimizing the loss function of equation 2.10. We implement the iteration of the stated two-step procedure until the updated parameters is converged to the previous ones. The implementation details is given in the following.

Recall that we have multivariate normal distributed observations $Y^k = (y_1^k, y_2^k, ..., y_p^k)^T$ for $k = 1, ..., n$. Denote the sample of the ith variable as $Y_i = (y_i^1, y_i^2, ..., y_i^p)^T$ for $i = 1, ..., p$. Then we denote $\tilde{Y}_i = \sqrt{\omega_i} Y_i$, $i = 1, ..., p$ with predefined ω. It is also another advantage in our model that we can make use of prior knowledge of the network structure by assigning different weights ω_i to different variables. Normally we take $\omega_i = \frac{1}{var(\varepsilon_i)}$ to consideration in the residual variances of the variables. Moreover, we take $\hat{\sigma}_{ii} = \frac{1}{(n-1)\sum_{k=1}^{n}(y_i^k - \bar{y}_i)^2}$, which is the sample variance of y_i, as the initial value of σ^{ii}. Since we have setup the initial value of σ and variable weight ω, we now denote

$$\mathcal{Y} = (\tilde{Y}_1', \tilde{Y}_2', ..., \tilde{Y}_n')^T, \tag{3.6}$$

$$\hat{\mathcal{K}}_{i,j} = (0, ..., 0, \sqrt{\frac{\tilde{\sigma}^{jj}}{\tilde{\sigma}^{ii}}})^T, 0, ..., 0, \sqrt{\frac{\tilde{\sigma}^{jj}}{\tilde{\sigma}^{ii}}}, 0, ..., 0)^T \tag{3.7}$$

where $\tilde{\sigma}^{ii} = \frac{\sigma^{ii}}{\omega_i}$ for $i = 1, ..., p$. Then we rewrite the l_1 penalized loss function

18

in equation 2.10 as

$$\min_{P} = \parallel \mathcal{Y} - \mathcal{K}P \parallel^2 + \lambda \parallel P \parallel_{l_1} \tag{3.8}$$

where $\mathcal{K} = (\mathcal{K}_{1,2}, ..., \mathcal{K}_{p-1,p})$. As the size of samples n and the number of variables p increases, the computation cost turns out to be a serious problem. So we make use of active-shooting approach, which is proved to be a very computational efficient algorithm. In the following empirical study, we make a modification of active-shooting to deal with stock variables of more than 200 tickers. Here we name it Multi-Active-Shooting(MAS) algorithm. As far as we know, it is the first time that this algorithm is used to solve the regression problem with dataset of hundreds of dimensions. Here we present details of the MAS algorithm as below.

Recall the loss function with respect to P as in equation 3.8. Firstly, with given weights w and initialized σ, we have known $\mathcal{Y}$ as well as $\mathcal{K}$. Thus, for $j = 1, ..., p$:

$$\begin{aligned}
P_j^{(initial)} &= \arg\min_{P_j} \left\{ \parallel \mathcal{Y} - P_j \mathcal{K}_j \parallel^2 + \lambda \sum_j \mid P_j \mid \right\} \\
&= sign(\mathcal{Y}^T \mathcal{K}_j) \frac{(\mid \mathcal{Y}^T \mathcal{K}_j \mid - \lambda)_+}{\mathcal{K}_j^T} \mathcal{K}_j
\end{aligned} \tag{3.9}$$

where $P_j = (0, ..., 0, \rho^{1j}, 0, ...0, \rho^{hj}, 0, ..., 0, \rho^{jj}, 0, ..., 0)^T$ for $1 \leq h \leq j$, and $\mathcal{K}_j = (0, ..., 0, \mathcal{K}_{1,j}, 0, ...0, \mathcal{K}_{h,j}, 0, ..., 0, \mathcal{K}_{j,j}, 0, ..., 0)^T$ for $1 \leq h \leq j$. $(.)_+$ is used to take only the positive elements in the matrix.

Then, we update P_j recursively until the distance between the current one and the previous one lies in a certain limit, which is decided by the tuning pa-

19

rameter λ. In the updated formulation, we define active set as $\Lambda := k : P_k \neq 0$. Thus, for each $k \in \Lambda$, we have

$$
\begin{aligned}
P_i^{(current)} &= P_i^{(previous)}, \ i \neq k \qquad\qquad\qquad\qquad\qquad (3.10) \\
P_k^{(current)} &= \arg\min_{P_k} \left\{ \parallel \mathcal{Y} - \sum_{i \neq k} P_i^{(previous)} \mathcal{K}_i - P_k \mathcal{K}_k \parallel^2 + \lambda \sum_k \mid P_k \mid \right\} \\
&= sign\left(\frac{\varepsilon^{(previous)^T} \mathcal{K}_k}{\mathcal{K}_k^T \mathcal{K}_k} + P_k^{(previous)} \right) \times \qquad\qquad (3.11) \\
&\quad \left(\mid \frac{\varepsilon^{(previous)^T} \mathcal{K}_k}{\mathcal{K}_k^T \mathcal{K}_k} + P_k^{(previous)} \mid - \frac{\lambda}{\mathcal{K}_k^T \mathcal{K}_k} \right)_+
\end{aligned}
$$

where $\varepsilon^{(previous)} = \mathcal{Y} - \mathcal{K} P^{(previous)}$. And we repeat the updating process until $P^{(current)}$ is converged to $P^{(previous)}$ in acceptable arrange.

With the help of MAS algorithm, we obtain the updated $\hat{P}^{(current)}$. Then we update σ as following:

$$
\frac{1}{(\hat{\sigma}^{ii})^{(current)}} = \frac{1}{n} \parallel Y_i - \sum_{j \neq i} \hat{P}_{ij}^{current} Y_j \parallel^2 \qquad\qquad (3.12)
$$

Therefore, we calculate $\hat{P}^{(current)}$ and $(\hat{\sigma}^{ii})^{(current)}$ in each iteration until they converge. However, as stated before, the performance of this model depends much on the choice of the tuning parameter λ. We will next discuss about estimation of the tuning parameter.

3.2.2 Tuning Parameter

The tuning parameter λ controls for the amount of shrinkage in the estimation procedure. So λ can be seen as the penalty to pay when constructing

a sparse network form empirical data which is not sparse enough. First, in the case λ takes a value equal to zero, no shrinkage is produced and the estimators are exactly the same as in the ordinary least square case. Second, in the case λ takes a value big enough, all the lasso estimators might be shrunk to zeros so that there is no estimator different from zero. Thus, only when we pick a suitable λ, we could have a proper number of parameters been shrunk to zero.

In practice, different values of λ are estimated for the partial correlation networks, and afterwards, information criteria like AIC(Akaike Information Criterion) or BIC(Bayesian Information Criterion) are applied to determine the optimal value of λ. In general, BIC is preferred to AIC due to its simplicity and computational easiness. According to Zou (2007), we make use of BIC criterion for selecting the tuning parameter for the sake of simplicity and computational easiness. For a predefined λ, we have estimators calculated using the above procedures and denoted as $\hat{P}_\lambda = \{\hat{\rho}_\lambda^{ij} : 1 \leq i < j \leq p\}$ and $\hat{\sigma}_\lambda = \{\hat{\sigma}_\lambda^{ii} : 1 \leq i \leq p\}$. Recall equation 2.2, we denote the residual sum of squares as:

$$RSS_i(\lambda) = \sum_{k=1}^{n} \left(y_i^k - \sum_{j \neq i} \hat{\rho}_\lambda^{ij} \sqrt{\frac{\hat{\sigma}_\lambda^{jj}}{\hat{\sigma}_\lambda^{ii}}} y_j^k \right)^2 \tag{3.13}$$

For the $i-$th regression, we have

$$BIC_i(\lambda) = n \times \log(RSS_i(\lambda)) + \log n \times \sum_{j \neq i} e_{ij} \tag{3.14}$$

where $e_{ij} = 0$ if $\hat{\rho}_\lambda^{ij} = 0$, and $e_{ij} \neq 0$ if $\hat{\rho}_\lambda^{ij} \neq 0$. Therefore, we minimize $BIC(\lambda) = \sum_{i=1}^{p} BIC_i(\lambda)$ to decide which λ is our choice in the model.

3.3 Proposed Method 2: Estimation of VAR Adjacency Network

3.3.1 Implementation with Recursive Pathwise Approach

In this section, we first present an iterative algorithm for estimation of the VAR coefficient matrix. In the recursive procedure, we apply kernel smoothers and deduct a pathwise approach to the sparse regression model. Moreover, we discuss the selection of initial parameters and related parameter tuning process.

We assume that the initial value of VAR coefficient is represented as $A^{k\,(initial)}$, suppose that $k_{\backslash d} = 1, ..., d-1$, and for $k = 1, ..., d$:

$$R^{k\,(initial)} = \mathcal{X}^n - \sum_{j \in k_{\backslash d}} A^{k\,(initial)} \mathcal{X}^{n-j} \tag{3.15}$$

Then, we calculate the updated estimator recursively until convergence, for $i = 1, ..., p$:

$$\hat{A}_i^{(cur)} = \arg\min_\theta \frac{1}{n} \parallel R_i^{k\,(pre)} - \mathcal{X}^{n-k}\theta^{k\,(pre)} \parallel^2 + \lambda \sum_{j=1}^p \mid \theta_j^{k\,(pre)} \mid \omega_j^k \tag{3.16}$$

During the recursive calculation, the penalized regression function at $k-$ stage can be regarded as a general Lasso optimization problem. Thus, we utilize kernel smoother and rewrite the updated function simply as:

$$\arg\min \sum_{i=1}^p D\left(\frac{\mid t - t_0 \mid}{h_\lambda(X_0)}\right) \parallel R_i^k - \mathcal{X}_i^{n-k}\theta^k \parallel^2 + \lambda \sum_{j=1}^p \mid \theta_j^k \mid \tag{3.17}$$

22

Consider a coordinate descent step for solving the above problem. That is, suppose we have estimates $\tilde{\theta}_0$ and $\tilde{\theta}_l$ for $l \neq j$, and we wish to partially optimize with respect to θ_j. Denote by $R(\beta_0, \beta)$, the objective function in equation 3.4. We would like to compute the gradient at $\theta_j = \tilde{\theta}_j$, which only exists if $\tilde{\theta}_j \neq 0$. If $\tilde{\theta}_j > 0$, then

$$\frac{\partial R}{\partial \theta_j} \Big|_{\theta = \tilde{\theta}} = -2 \times \sum_{i=1}^{p} D\left(\frac{\mid t - t_0 \mid}{h_\lambda(X_0)}\right) X_{ij}(R_i^k - \mathcal{X}_i^{n-k}\theta^k) + \lambda \qquad (3.18)$$

A similar expression exists if $\tilde{\theta}_j < 0$, and $\tilde{\theta}_j = 0$ is treated separately. Simple calculation shows that the coordinate-wise update has the form

$$S\left(-2 \times \sum_{i=1}^{p} D(\frac{\mid t - t_0 \mid}{h_\lambda(X_0)}) X_{ij}(R_i^k - \tilde{R}_i^{k^{(j)}}), \lambda\right) \rightarrow \tilde{\theta}_j \qquad (3.19)$$

where

$$\tilde{R}_i^{k^{(j)}} = \tilde{\theta}_0 + \sum_{l \neq j} X_{il}\tilde{\theta}_l \qquad (3.20)$$

is the fitted value excluding the contribution from X_{ij}, and hence $R_i^k - \tilde{R}_i^{k^{(j)}}$ the partial residual for fitting θ_j. $S(z, \gamma)$ is the soft thresholding operator with value

$$\begin{cases} z - \gamma & if \ z > 0 \ and \ \gamma <\mid z \mid \\ z + \gamma & if \ z < 0 \ and \ \gamma <\mid z \mid \\ 0 & if \ \gamma \geq \mid z \mid \end{cases} \qquad (3.21)$$

Thus, with smoothing on time period, we compute the simple autoregressive coefficient on the partial residual, apply soft-thresholding to take care of

23

the lasso contribution to penalty, and then apply a proportional shrinkage for the ridge penalty. Furthermore, we move on to the selection of initial value and tuning process of parameters.

3.3.2   Selection of Initial Value and Tuning Parameter

According to Wang (2006), to obtain the consistent estimator for the iterative process, we introduce the ordinary least squares(OLS) estimator as an initial value for the adjacency matrix A_k, in which k denotes the number of time logs:

$$\hat{A}_k^{(initial)} = \frac{\mathcal{X}_{n-k:n}{}^T \mathcal{X}_n}{\mathcal{X}_{n-k:n}{}^T \mathcal{X}_{n-k:n}} \tag{3.22}$$

With the assumption of V_t in equation 2.11 is independent of X_i, it is proved that $\hat{A}_k^{(initial)}$ is a consistent estimator of initialized parameter.

To complete the whole iteration process, we need to select the tuning parameters as well. Recall that BIC perform better than other information criterion, we formulate BIC-type parameter in the case of k time log as:

$$BIC_i = \arg\min \sum_{i=1}^{p} D\left(\frac{|\, t - t_0\,|}{h_\lambda(X_0)}\right) \parallel R_i^k - \mathcal{X}_i^{n-k}\theta^k \parallel^2 + \lambda \sum_{j=1}^{p} |\, \theta_j^k\,| \tag{3.23}$$

where $\Lambda(.)$ stands for the number of nonzero elements in the matrix. Now we have completed both of the implementation of methods for estimating the partial correlation network and VAR adjacency network. We will illustrate the performance of our proposed method in both simulation study and real data analysis in the next section.

Chapter 4

Numerical Study and Result

4.1 Illustrated Simulation

In order to check the performance of the proposed method to estimate partial correlation and VAR coefficient networks, we process with an illustrated simulation. Based on the assumption of sparsity, we randomly create a true network with 10 elements and a total of 9 edges, see Figure 4.1. Therefore, the concentration matrix is a 10×10 matrix. Though it doesn't represent a very big dataset, it is enough to perform a practical simulation to analyze and get useful results.

First, we make n random observations from a multivariate normal distribution. The number of variables is 10 and their covariance matrix used to generate those n observations is the inverse of the concentration matrix created in the first step. After that, our proposed method is used to generate a concentration matrix from the random observations and the previously defined λ. Finally, we get the adjacency matrix after several iterations. We then
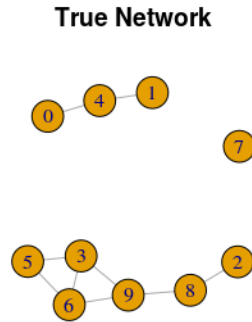
Figure 4.1: Ture Network Plot with 10 nodes and 9 edges

evaluate each method at a series of different values of the tuning parameter λ. Our target is to present how the change of λ will effect the shrinkage result of the variables, which is showed in the number of edges left in the plot.

In our simulation, we study for four different sample sizes $n = \{50; 100; 150; 200\}$. The tuning parameter λ, always proportional to the sample size, takes values from $\lambda = 0.00 \times n$ to $\lambda = 1.00 \times n$ with interval of 0.05. Since the simulation is a random process, it must be done many times in order to get accurate and unbiased results. To reach that end, for each pair of λ and sample size values, we calculate the network estimator for 1000 times. In order to analyze the results obtained from the simulation, one graph will be showed to study the features of the regression model. Figure 4.2 will show how changes in the penalty value affect the sparsity of the estimated network.

From Figure 4.2, we can see that there is a positive relationship between the penalty value λ and shrinkage result, in other words is the sparsity. When λ

26
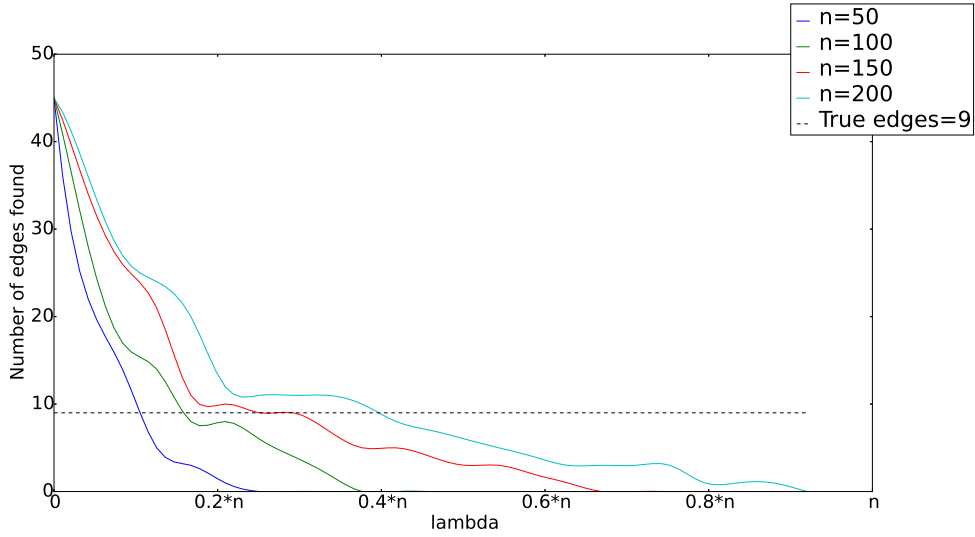
Figure 4.2: How does λ affect the shrinkage result

increased, the total number of edges found decreases, so the estimated network becomes sparser, moving from far more above the number of true edges to below it when the penalization is too big. When the penalty value is zero, the number of edges found is $45 = \frac{10 \times 9}{2}$ for every sample size, which is the maximum possible number of existing edges in a network of 10 variables. On the other hand, when the penalty is $1.00 \times n$, sparsity converges again for all sample sizes at a level below the true number of the edges in the network. For different sample sizes, it is observed that for large values of n, the estimated networks become very sparse and the number of edges estimated fall to the number of true edges via a mild path. That is to say, in the case n is lower, the estimated networks become sparser following a steep path until the true number of edges is reached, and once in there, the number of edges estimated fall below the optimal value faster than with large sample sizes as λ increases.

4.2 Empirical Analysis

4.2.1 Data Description

The S&P 500 index is probably the most commonly referenced U.S equity benchmark for determining the state of the overall economy. S&P Dow Jones Indices updates the components of the S&P 500 periodically, typically in response to acquisitions, or to keep the index up to date as various companies grow or shrink in value. Between 1/1/2005 and 1/1/2015, 188 index components were replaced by other components. In our paper, we propose to track the performance of the largest and most dominant American companies included in the index. So, we pick 223 stocks as our target stocks, which have been included in the index for 12 years, from the year 2002 till the year 2013.

Through Daily TAQ(Trade and Quote), which provides us with FTP access to all trades and quotes for all issues traded on NYSE in all the trading days, we download the 1-minute stock data using our own target stock lists. So our dataset includes associated key financial dynamics such as time of trading, stock price, market capitalization in the certain time period.

Before we apply the model to our data, we shall ensure that the supplied data is clean, correct and useful. In our raw data, there are included 1-minute price, time of trading and trading volume of all the 223 stocks. However, there are some issues worthy of our attention. Firstly, we have, inevitably, lost some data on the certain trading time points during one month. Secondly, the 1-minute trading volume should be within a certain range, to make sure that the future calculation will not spill over. So before the introduction of our

28

theoretical method, we have to process with data loss and data cleaning.

For every one of our target stocks, there are no more than 2 percent of data has been left blank. To deal with the data loss, we use linear interpolation to fill the gap in our raw data. For instance, if there is no record at the certain time during a trading day, then we will use two closest stock prices and trading volume to get the estimated value. To calculate the weighted average as unknown prices and trading volume, the weights are inversely related to the distance from the known points to the unknown point. In this way, we successfully get all the data at all the time points filled.

In both of two proposed methodologies, we use an iterative algorithm to calculate the estimated parameters of partial correlation matrix and VAR adjacency matrix. So we have to get rid of those trading volume records which are beyond the limit of acceptability or fairness. First, all the trading volume data should be positive. Secondly, the value of trading volume should not exceed a certain range, since unreasonable large value will limit the calculation of partial correlation matrix and VAR adjacency matrix. Without loss of common sense, we eliminate those 1-minute trading volumes whose values are negative or more than 1 million.

After processing with the raw data, we have more concerns before application of real data to our model. In our article, we will use 30 minute as interval of two neighbor records. That is to say, we calculate the average values of stock prices and trading volume in half an hour. During one trading day, whose trading period last from 9:30 in the morning till 4:00 in the afternoon, we have 14 records of prices and trading volumes for every target stock.

29

Furthermore, to have a better understanding of the macrostructure of the stock market, we choose two measurements to do the quantitative research: stock returns and volume-price trend indicators. Stock return is a traditional measure of a company's performance over time. Here we use single period return. Next, to measure the balance between a stock' demand and supply, we introduce volume-price trend (VPT), which is an technical analysis indicator to relate price and volume in the stock market. VPT is based on a running cumulative volume that adds or subtracts a multiple of the percentage change in share price trend and current volume, depending upon the investment's upward or downward movements. We have the formula as following:

$$VPT \ = \ VPT_{prev} + volume \times \frac{price_{current} - price_{prev}}{price_{prev}} \qquad (4.1)$$

$$VPT \ = \ VPT_{prev} + volume \times return \qquad (4.2)$$

Thus, we use stock returns and VPT indicators as our measurement to the performance of target stocks and stock market balance.

In this way, we have both of two kinds of financial dynamics related to intraday stock market, which are setup in good quality and reasonable range. In the following section, we will deduct two empirical studies with our proposed methods. In Chapter 4.2.2, we choose the data in September of year 2013 as a simple test sample. For September 2013, we have 20 trading days. Thus, we have 280 records of stock prices and trading volumes for 223 target stocks in the whole. And in Chapter 4.2.3, we challenge our methods with the data over time period of April 2013 to September 2013, which has 1786 time points in

total. That is to say, we apply our network analysis with 223 variables of 1786 observations. As far as we know, this is the first time that one deal with such high dimensional time-varying network in such long period of time to interpret the hidden dependency in the view of the whole stock market.

4.2.2 Empirical Study 1

In the previous section, we have achieved the associated trading data of 223 target stocks. For September 2013, we have 279 30-minute stock returns for every stock. Thus, our real dataset is composed of 223 variables containing 279 time points for each one, standing as a reasonable size for our estimations to end up with significant results. The procedure to reach the estimated network is based on calculating the partial correlation matrices and VAR adjacency matrices of the dataset and plotting its network.

Firstly, we import the 223×279 matrix, where each column is named with the time points and each row is named with the ticker stock symbol. Go through each time point, we apply kernel smoother on time to get the estimated partial correlation matrix and VAR adjacency matrix which are generated by our proposed method respectively. Here, we choose 30 as the moving window size. Afterwards, we use Qgraph package in R program to plot each network on each time point. Thus, we have 279 plots of partial correlation networks and VAR adjacency networks of stock returns. Here in this article, we picked 4 of them at different Tuesday in September 2013.The interval of these 4 figures are the same, which is $14 \times 5 = 70$. One thing to mention, we choose VAR coefficient network with one time lag in empirical study. See Figure 4.3 for

31

the output of partial correlation network analysis. In the pictures, each node represents a stock and each edge represents a partial correlation between two stocks. As well as in Figure 4.4, each node represents a stock and each edge represents a VAR(1) coefficient between two stocks. Green edges indicate positive correlations, red edges indicate negative correlations, and the width and color of the edges correspond to the absolute value of the correlations: the higher the correlation, the thicker and more saturated is the edge.

As mentioned in previous section, there are quantitative methods as BIC and AIC which enable to find out the optimal λ for the model. In practice, the penalty value of the real data application has been chosen on the results obtained from several of tryings. That is to say, the tuning process has been based on observing the networks from a wide range of λ values and choosing the one that looks more efficient for our study. For instance, we take the adjacency matrix at the first time point of September 2013, if we choose $\lambda = 7e^{-6}$, the VAR(1) adjacency network at time 9:30 am on September 3rd, 2013 is shown in Figure 4.5(a); while if $\lambda = 7e^{-5}$ is chosen, then the corresponding VAR(1) adjacency network at exact the same time, is shown in Figure 4.5(b).

From Figure 4.5(a), we can see that the choice of $\lambda = 7e^{-6}$ must be too small, since the shrinkage of Lasso doesn't have much effect on coefficient matrix. We can almost read no useful information from the redundant VAR(1) adjacency network in Figure 4.5(a). For Figure 4.5(b), we can read clear information of the significant adjacency matrix between two stocks. So the choice of $\lambda = 7e^{-5}$ is a good one. In practice, remind that the analysis of the causes behind the nature of the dependency matrices are not a goal of the

201309_1 (a)

201309_71 (b)

201309_141 (c)

201309_211 (d)
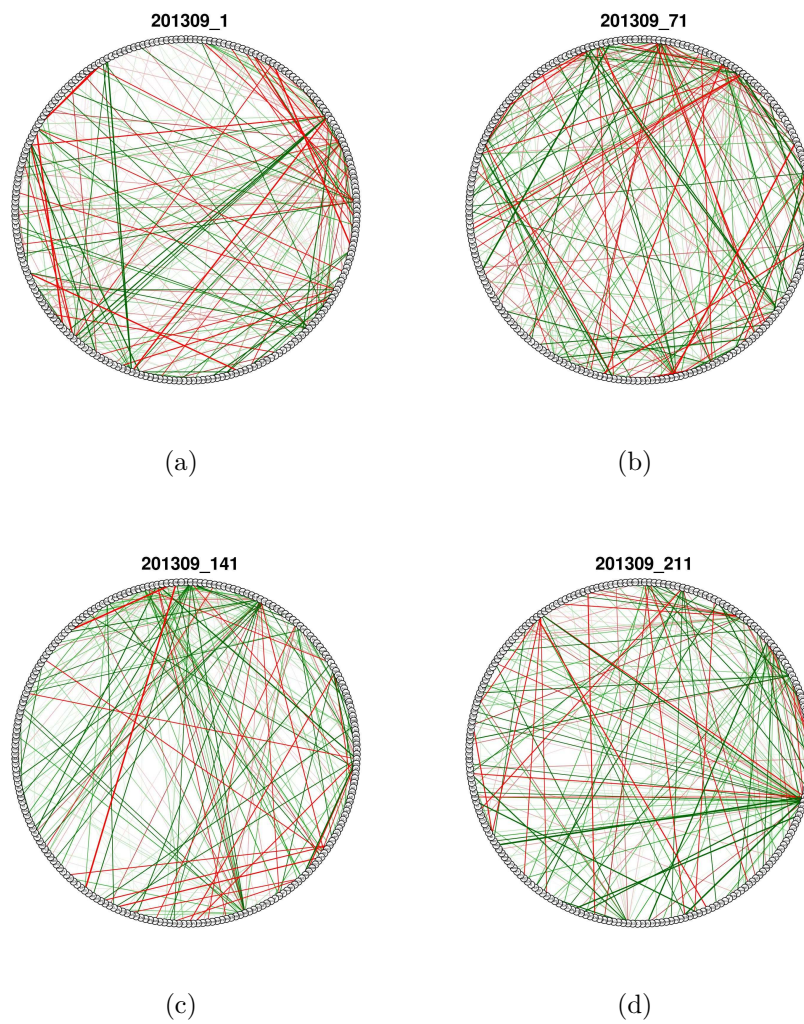
Figure 4.3: Partial Correlation Network of Stock Return at 9:30am on (a)September 3rd, 2013; (b)September 10th; (c)September 17th; (d)September 24th. (From left to right, top to down)

201309_1 (a)

201309_71 (b)

201309_141 (c)

201309_211 (d)
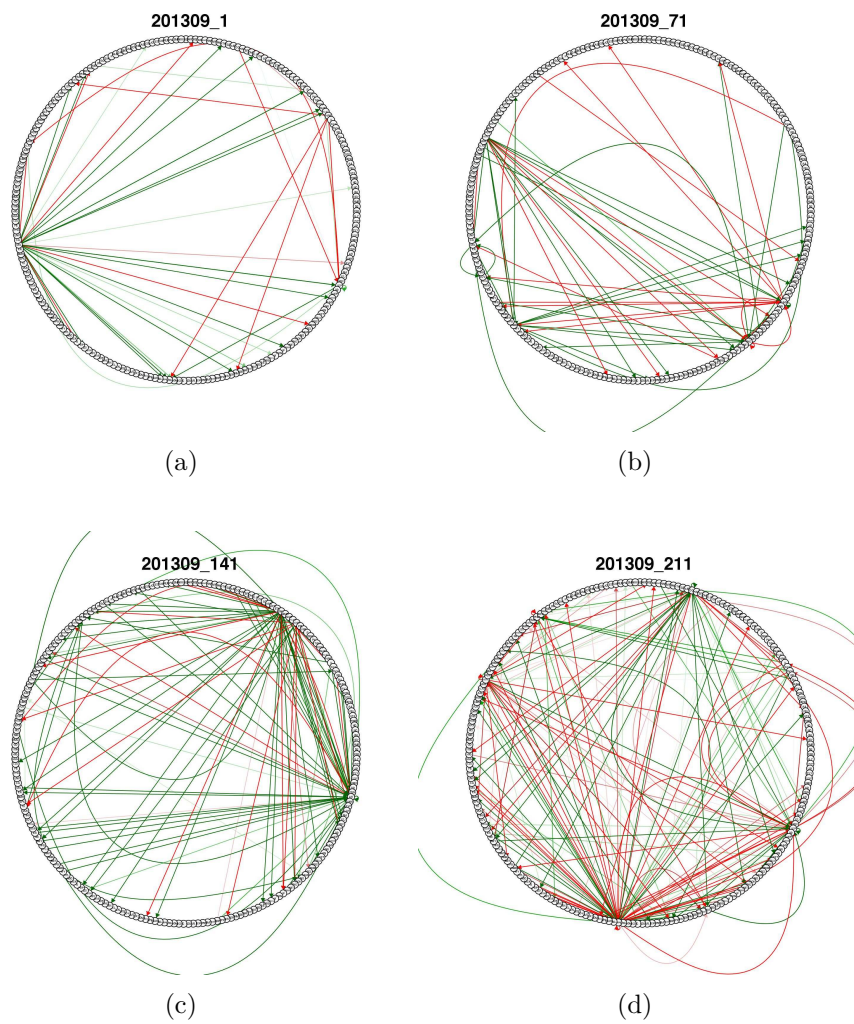
Figure 4.4: VAR(1) Adjacency Network of Stock Return at 9:30am on (a)September 3rd, 2013; (b)September 10th; (c)September 17th; (d)September 24th. (From left to right, top to down)
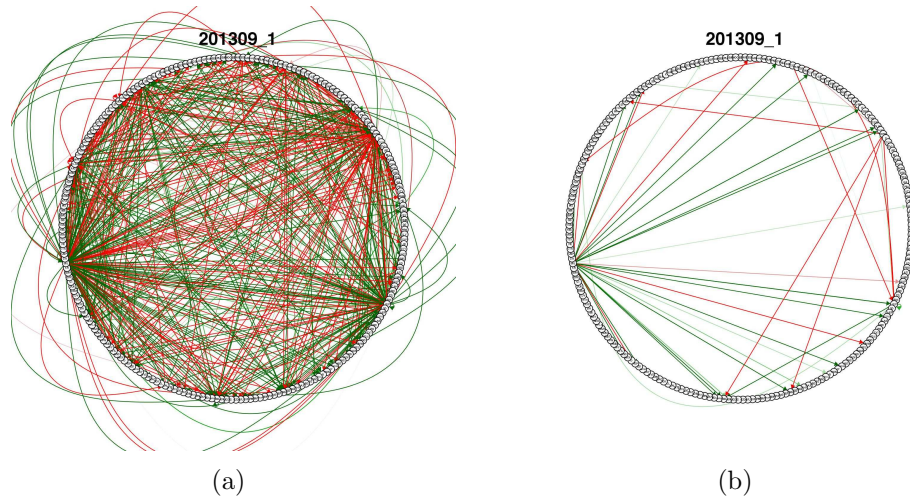
Figure 4.5: The VAR(1) Adjacency Network at time 9:30 am on September 3rd, 2013 when (a)$\lambda = 7 \times e^{-6}$; (b) $\lambda = 7 \times e^{-5}$

project, therefore the selection of the optimal λ is not a crucial point in our study, as long as we are working in a close interval from it.

Furthermore, we are interested in maximum eigenvalue of the dependency matrix at the certain time point. In this part, we take the partial correlation matrix in September 2013 for instance. For the time being, the change of maximum eigenvalues is shown in Figure 4.6(a), while the percentage of the maximum eigenvalue over the sum of all the eigenvalues in the partial correlation matrix is shown in Figure 4.6(b).

As we know, the maximum eigenvalue changes through time. So it could be regarded as a time series data. Here we fit this time series data with AR(1) model. The Autocorrelation plot is shown in Figure 4.7(a) while the partial Autocorrelation plot is shown in Figure 4.7(b).

Moreover, we want to have a general look at the performance of the maximum eigenvalue time series with respect to different λ. We chose 20 values of λ,
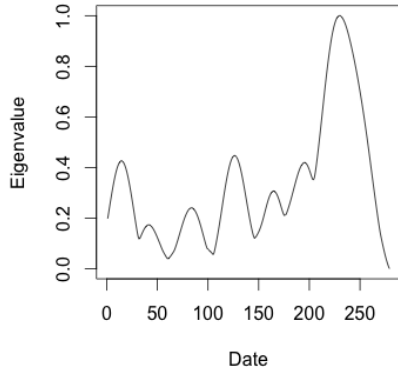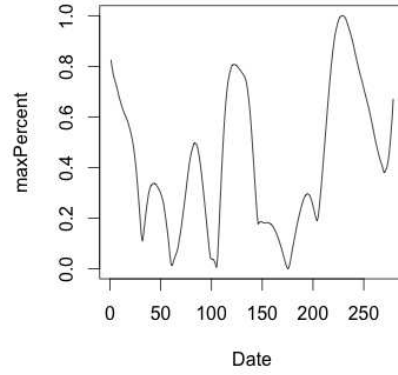
(a)   (b)

Figure 4.6: Partial Correlation Matrix in September 2013: (a)Maximum Eigenvalues; (b)The Percentage of Maximum Eigenvalues
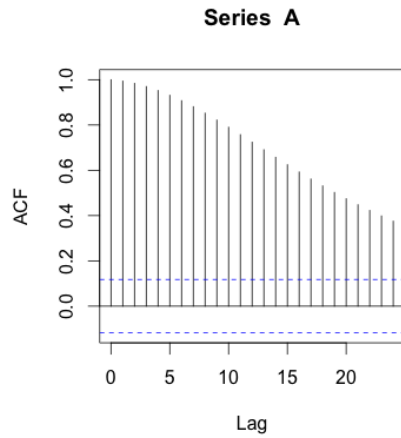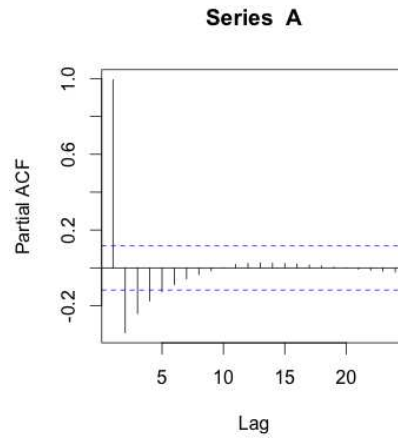
(a)   (b)

Figure 4.7: Partial Correlation Matrix in September 2013: (a)ACF plots of maximum eigenvalues; (b)PACF plots of maximum eigenvalues

which is respectively $1e^{-3}$, $3e^{-3}$,$5e^{-3}$,$7e^{-3}$,$9e^{-3}$, $1e^{-4}$, $3e^{-4}$,$5e^{-4}$,$7e^{-4}$,$9e^{-4}$,$1e^{-5}$, $3e^{-5}$,$5e^{-5}$,$7e^{-5}$,$9e^{-5}$,$1e^{-6}$, $3e^{-6}$,$5e^{-6}$,$7e^{-6}$,$9e^{-6}$. Next, we draw the 3 dimension surface of the results, from different angle. See Figure 4.8. With the help of 3-dimension version plots, we have a full view of how the change of penalty parameter λ influence the maximum eigenvalue over the time period.
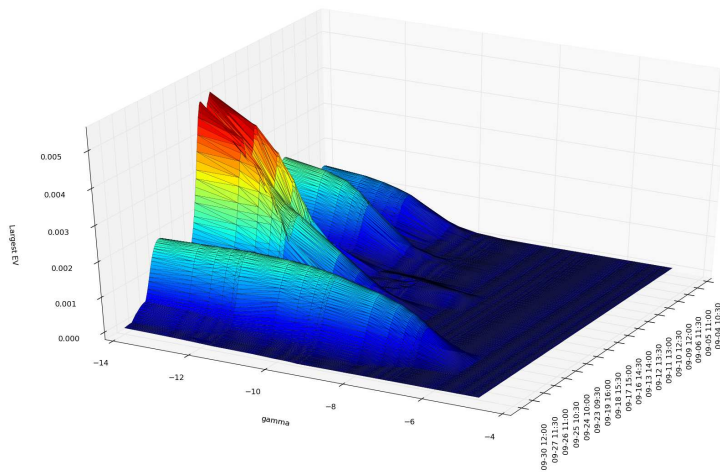
Figure 4.8: Three dimension plot of maximum eigenvalues with different λ

From the above 3D version plots of maximum eigenvalues with different λ, we can see that there are some peak points. Specially at the 121st time point and 165th time point. We are interested in such points so that we track back to these two exact times. For the 121st time point, it refers to 1:30pm on September 13rd, 2013; while the 165th time points refers to 2:30pm on September 18th, 2013. Tracking back to that date, a meeting of the Federal Open Market Committee was held in the offices of the Board of Governors of the Federal Reserve System in Washington, D.C., on Tuesday, September 17,

2013, at 1:00 p.m. and continued on Wednesday, September 18, 2013, at 8:30
a.m.

4.2.3   Empirical Study 2

In the previous section, we have tested our proposed methods to estimate
both of time-varying partial correlation network and VAR adjacency network
with 223 target stock returns in September 2013. The results are presented
in different kinds of network plots and several related statistics plot. In this
section, we will deduct a more complicated and comprehensive empirical study
as well as some insightful concluding remarks.

Except for stock returns, we introduce VPT, which is believed to be
another widely used indicator of macro market structure. Moreover, we have
1785 30-minute stock returns and 1785 30-minute VPTs for every stock. Thus,
our real dataset is composed of 223 variables containing 1785 time points for
each one, standing as a reasonable size for our estimations to end up with
significant results. The procedure to reach the estimated network is based on
calculating the partial correlation matrices and VAR adjacency matrices for
both stock returns and VPT indicators and plotting its network.

Network Plots

Firstly, we import the 223×1785 matrix, where each column is named
with the time points and each row is named with the ticker stock symbol. Go
through each time point, we apply kernel smoother weighted on time to get
the estimated partial correlation matrix and VAR adjacency matrix which are

generated by our proposed method respectively. Here, we choose 50 as the moving window size. Afterwards, we use Q-graph package in R program to plot each network on each time point. Thus, we have 1785 plots of partial correlation networks and VAR adjacency networks of stock returns and VPTs. The procedure of dependency matrix estimation and the associated network plots is the same as we did in Chapter 4.2.2. However, we apply two matrix estimation methods we proposed on two kinds of financial dynamics, stock returns and VPTs. Therefore, for each time point, we have four network plots. Furthermore, we display more plots indicating how maximum eigenvalues and the related maximum eigenvectors change through time.

Firstly, we picked 6 time points for each month in the year 2013. The interval of these 6 figures are the same, which is 300, approximately 21 and a half trading days. As stated above, we have four different network plots including two proposed methods applied on two kinds of stock indicators. See Figure 4.9, Figure 4.10, Figure 4.11 and Figure 4.12. In the pictures, each node represents a stock and each edge represents a partial correlation between two stocks. Green edges indicate positive correlations, red edges indicate negative correlations, and the width and color of the edges correspond to the absolute value of the correlations: the higher the correlation, the thicker and more saturated is the edge.
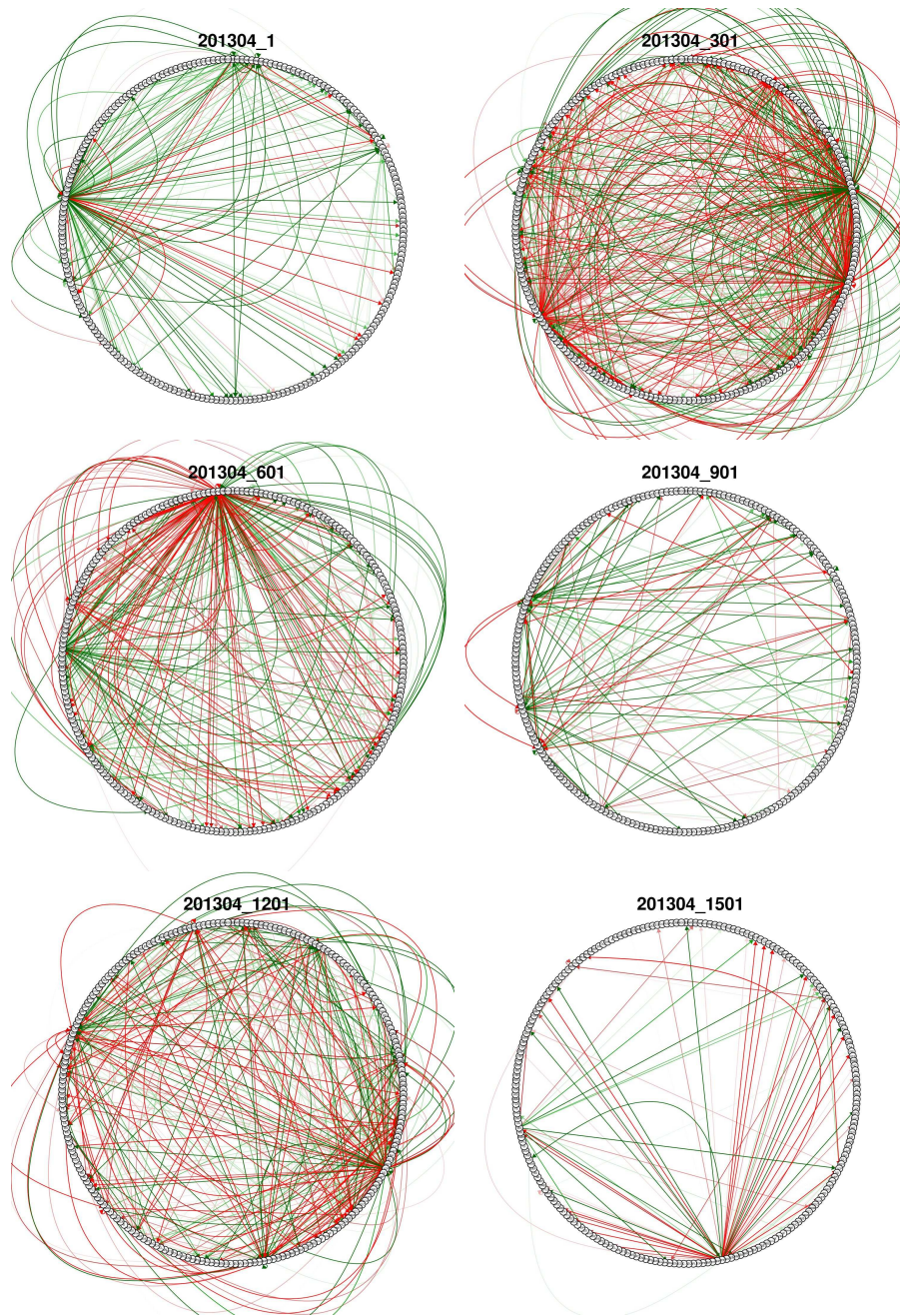
Figure 4.9: VAR Adjacency Network of Stock Return Plot at time: (a)1, (b)301, (c)601, (d)901, (e)1201, (f)1501. (From left to right, top to bottom)
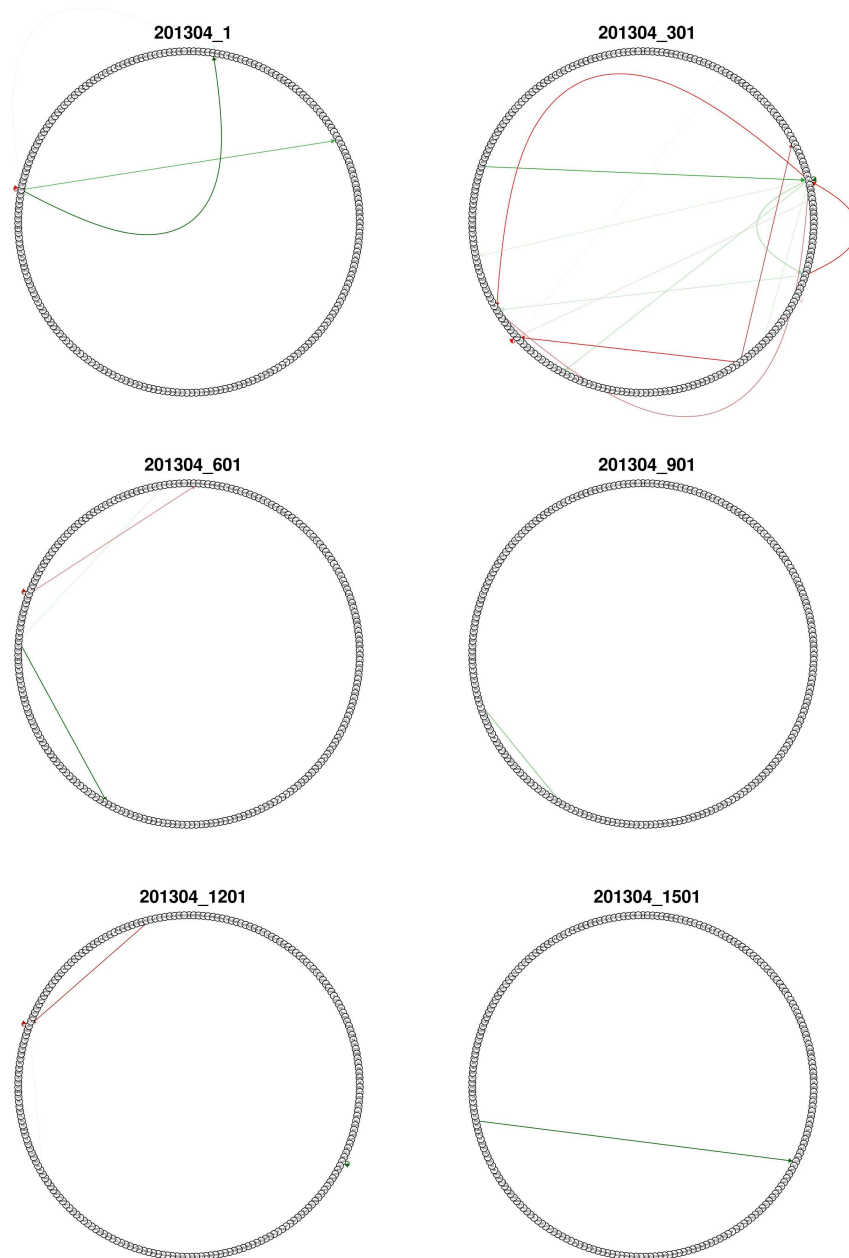
Figure 4.10: VAR Adjacency Network of Stock VPTs plot at time: (a)1, (b)301, (c)601, (d)901, (e)1201, (f)1501. (From left to right, top to down)

Figure 4.11: Partial Correlation Network Stock Return plot at time: (a)1, (b)301, (c)601, (d)901, (e)1201, (f)1501. (From left to right, top to down)

42

Figure 4.12: Partial Correlation Network Stock VPT plot at time: (a)1, (b)301, (c)601, (d)901, (e)1201, (f)1501. (From left to right, top to down)

The VAR(1) adjacency network for stock returns and VPTs of six time points evenly distributed from April 2013 to September 2013 are respectively plotted in Figure 4.9 and Figure 4.10. While The partial correlation network for stock returns and VPTs of six time points evenly distributed from April 2013 to September 2013 are respectively plotted in Figure 4.11 and Figure 4.12. To make a fair comparison with these networks, we make use of same penalty parameter λ in all the four dimension shrinkage process. Also, the order of the tickers are fixed in all the network plots.

Firstly, we take a look in the number of edges left after dimension shrinkage process in each method. As we know, the number of edges left is an indicator of how much the method works on selecting dominant variables. If the number of edges is too great, the network plotting will be presented in a massive way and difficult to discover the most important dependencies in it. If the number of edges is too small, it means that we have done too much on variable shrinkage and may lose the useful information. From the comparison of Figure 4.9 and Figure 4.10 with Figure 4.11 and Figure 4.12, we see that under the assumption of the same penalty parameter, proposed method 1 for partial correlation network has better performance in dimension shrinkage than proposed method 2 for VAR adjacency network. Also, the stock returns network has less response on the dimension shrinkage than VPTs in both proposed estimation method. Moreover, the number of edges changes through time. Great number of edges means that there are strong dependencies between the stocks. As shown in Figure 4.9, it is clear to see that strong VAR(1) coefficients between stock returns took place in May, June and August. And

44

in Figure 4.12, it is clear to see that strong partial correlation between stock VPTs took place in April, May and September.

Secondly, we consider the color and thickness of the edges in the network plots. As we know, the red lines refer to the negative dependency between two nodes, while the green lines refer to the positive relationship. In a general look, the green lines are much more than red lines, which shows that the positive dependencies between stocks are much more than the negative ones. This result indicates that there exists common driven force in the macro market, so that the stocks return and VPT increase or decrease in the same direction simultaneously. Also, at the same time point, all the four networks plots indicate almost the same distribution of red and green colors. See Figure 4.9 and Figure 4.12 for instance, red lines are more significant than green lines in the beginning of May 2013. In fact, after comparison of four sets of networks, we can see that the ratio of red lines to green lines changes in the same trend in all network plots.

Furthermore, we focus on some special tickers of great interest on the certain time point. In Figure 4.9, we summarize interesting tickers: HUM in health care sector in April; PBI in industrial sector, CAH in health care sector and BBY consumer discretionary sector in May; PLD and SLG which are both in real estate sector in June; JCP in service sector in August; APA in energy sector in September. In Figure 4.12, we summarize interesting tickers: PNC in financial sector in April; CAH in health care sector in May; HRB in financial sector in September.

45

Eigenvalue and Eigenvector Plots

As stated in the previous section, we are interested in the eigenvalues and the associated eigenvectors at the certain time point. For each time point, we have four output plots: VAR(1) adjacency network estimation with stock returns as in Figure 4.13, Figure 4.17 and Figure 4.21; VAR(1) adjacency network estimation with VPTs as in Figure 4.14, Figure 4.18 and Figure 4.22; partial correlation network estimation with stock returns as in Figure 4.15, Figure 4.19 and Figure 4.23; and partial correlation network estimation with VPTs as in Figure 4.16, Figure 4.20 and Figure 4.24. In each set of plots, we choose six time points evenly distributed from April 2013 to September 2013 respectively.

Among of these pictures, absolute value of eigenvalues for 223 tickers are plotted in Figure 4.13 for VAR(1) adjacency network with stock returns, Figure 4.14 for VAR(1) adjacency network with VPTs, Figure 4.15 for partial correlation network with stock returns, and Figure 4.16 for partial correlation network with VPTs. With the comparison of Figure 4.13 and Figure 4.9, Figure 4.14 and Figure 4.10, Figure 4.15 and Figure 4.11, and Figure 4.16 and Figure 4.12, we find that when there are more edges in network plots, which indicates that the tickers are more influenced by each other in the market, the number of single points in eigenvalues plots increase almost in each model.

For each set of plots, there are less than ten tickers represented by single point, which means that eigenvalues are almost zeros for most of the tickers. This indicates that our dimension shrinkage methods work well in four networks through time by leaving only a few of tickers, which are the ones of

46

interest. In this way, one can ignore most of the tickers and only focus on the ones with great eigenvalues. However, since the absolute value of eigenvalues for each ticker at the certain time is too small, we want to take a look at what is the proportion of the tickers with great eigenvalues. Thus, we introduce the plots of eigenvalue percentages in the following.

Figure 4.13: VAR Adjacency Network Stock Return eigenvalue plot at time: (a)1, (b)301, (c)601, (d)901, (e)1201, (f)1501. (From left to right, top to down)
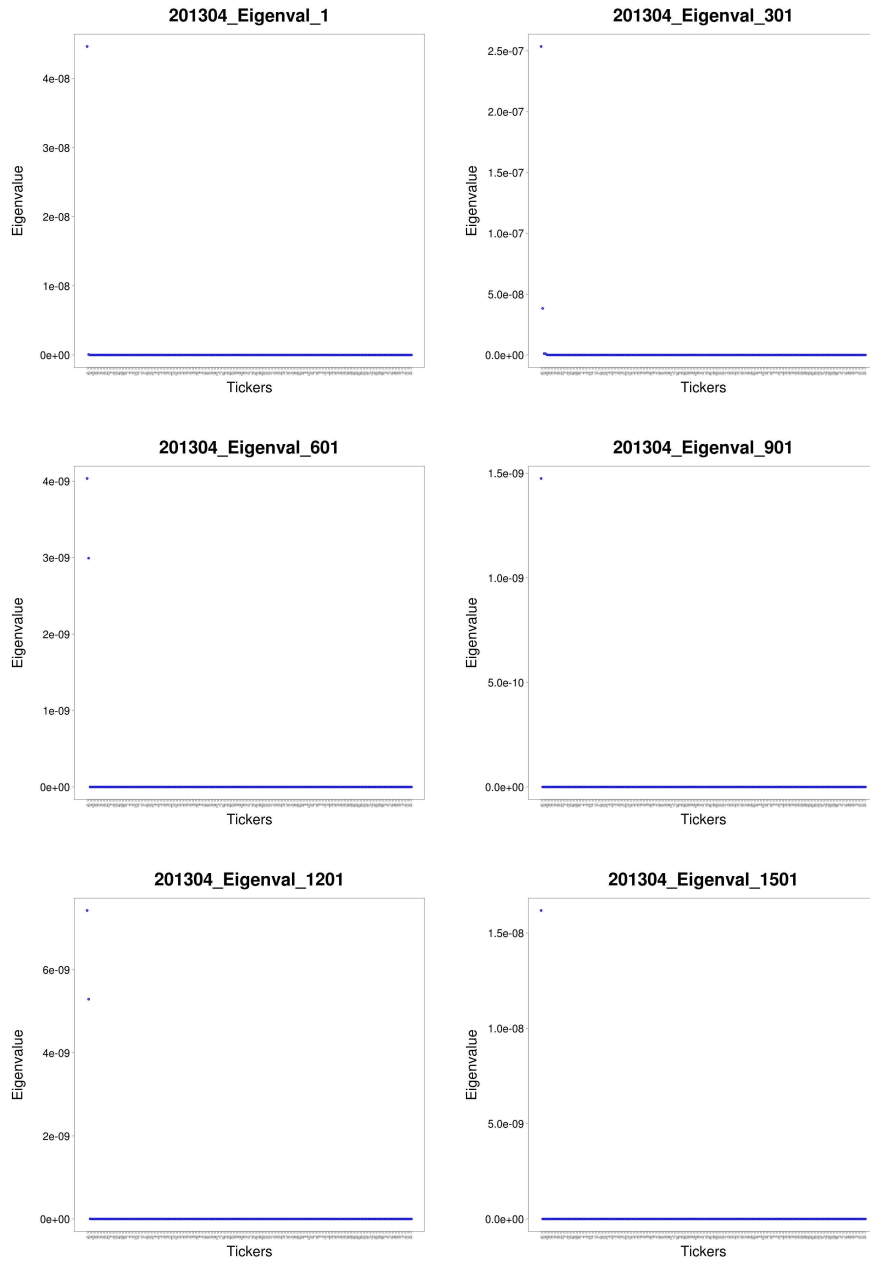
Figure 4.14: VAR Adjacency Network Stock VPT eigenvalue plot at time: (a)1, (b)301, (c)601, (d)901, (e)1201, (f)1501. (From left to right, top to down)
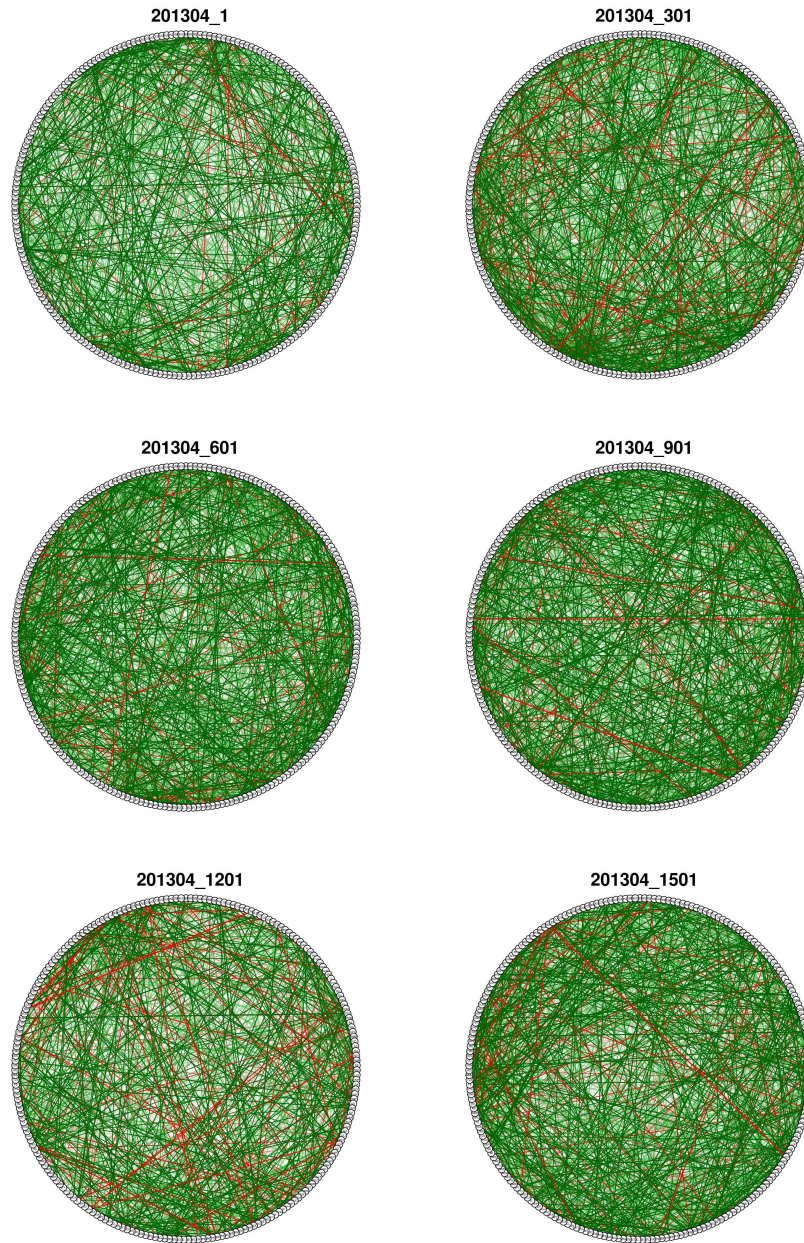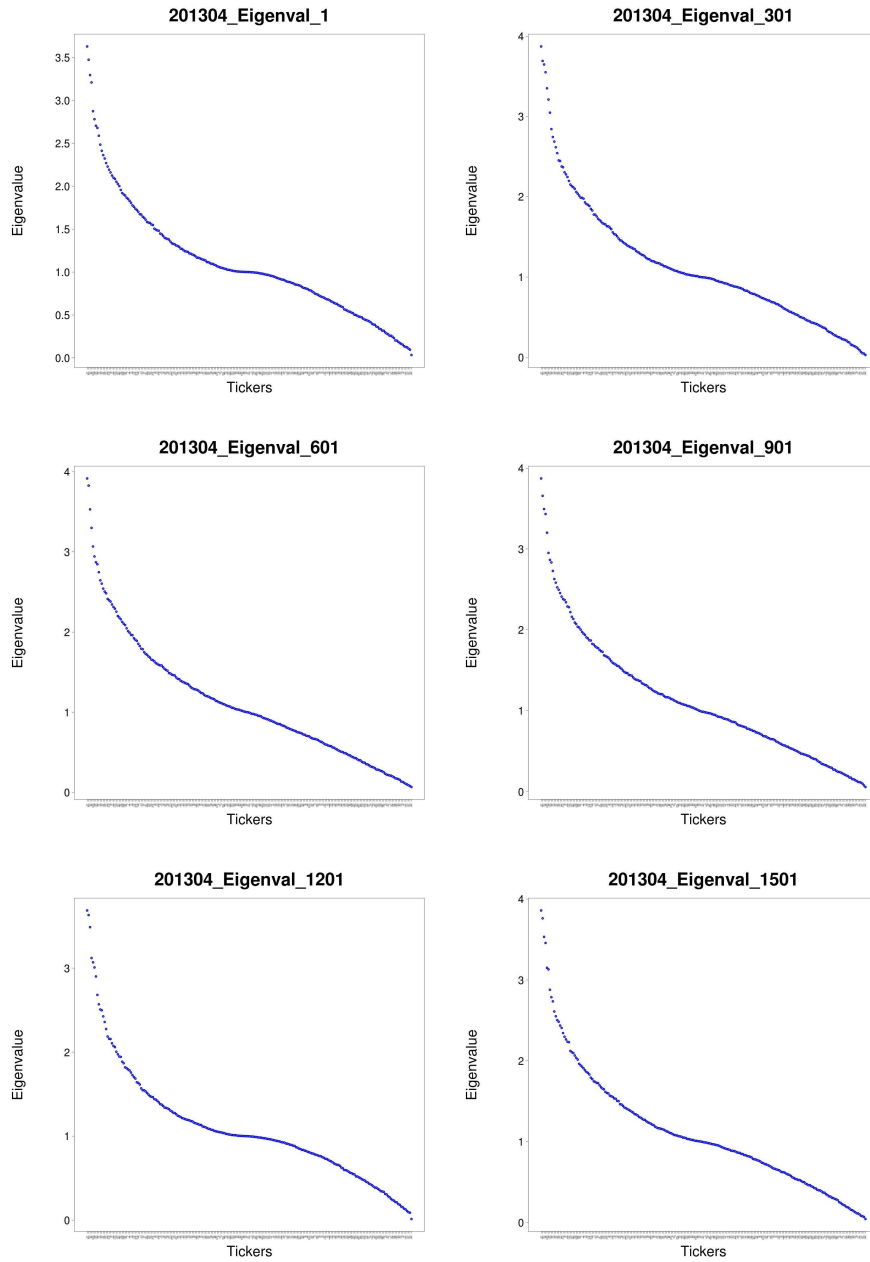
Figure 4.15: Partial Correlation Network Stock Return eigenvalue plot at time: (a)1, (b)301, (c)601, (d)901, (e)1201, (f)1501. (From left to right, top to down)
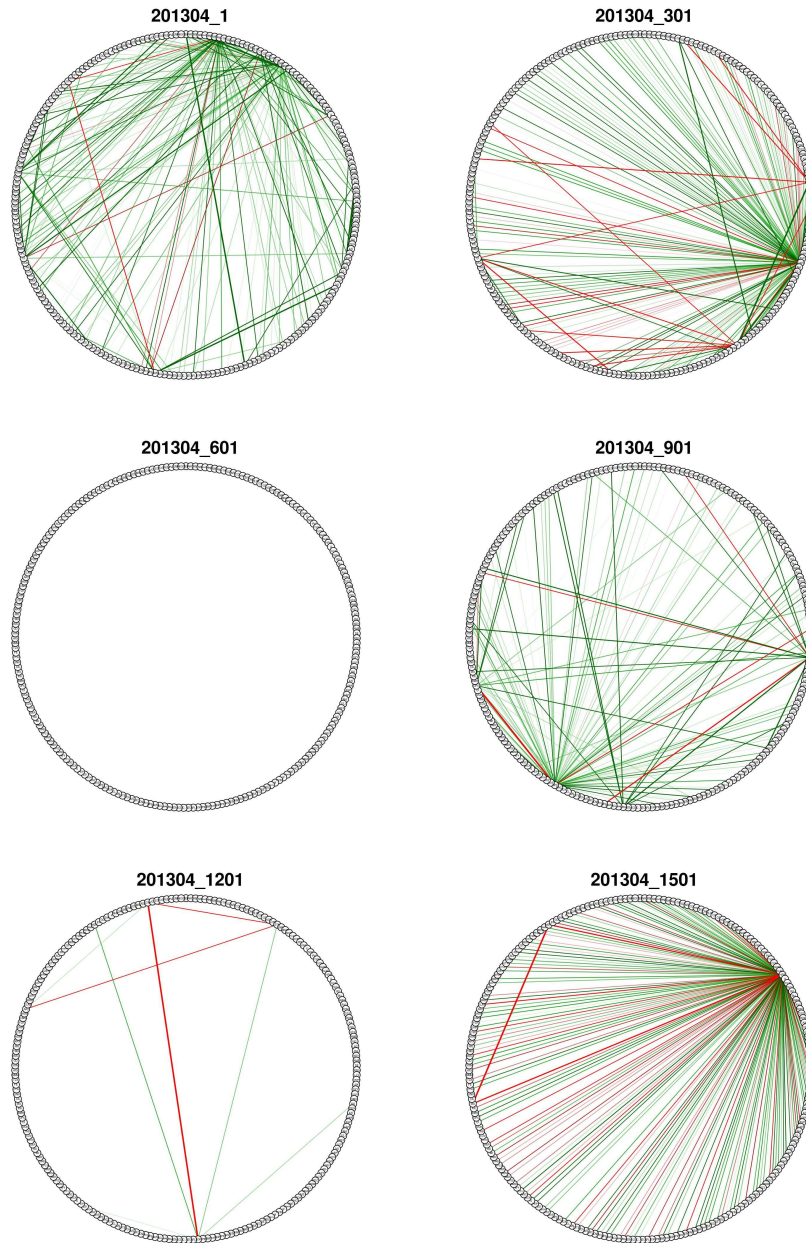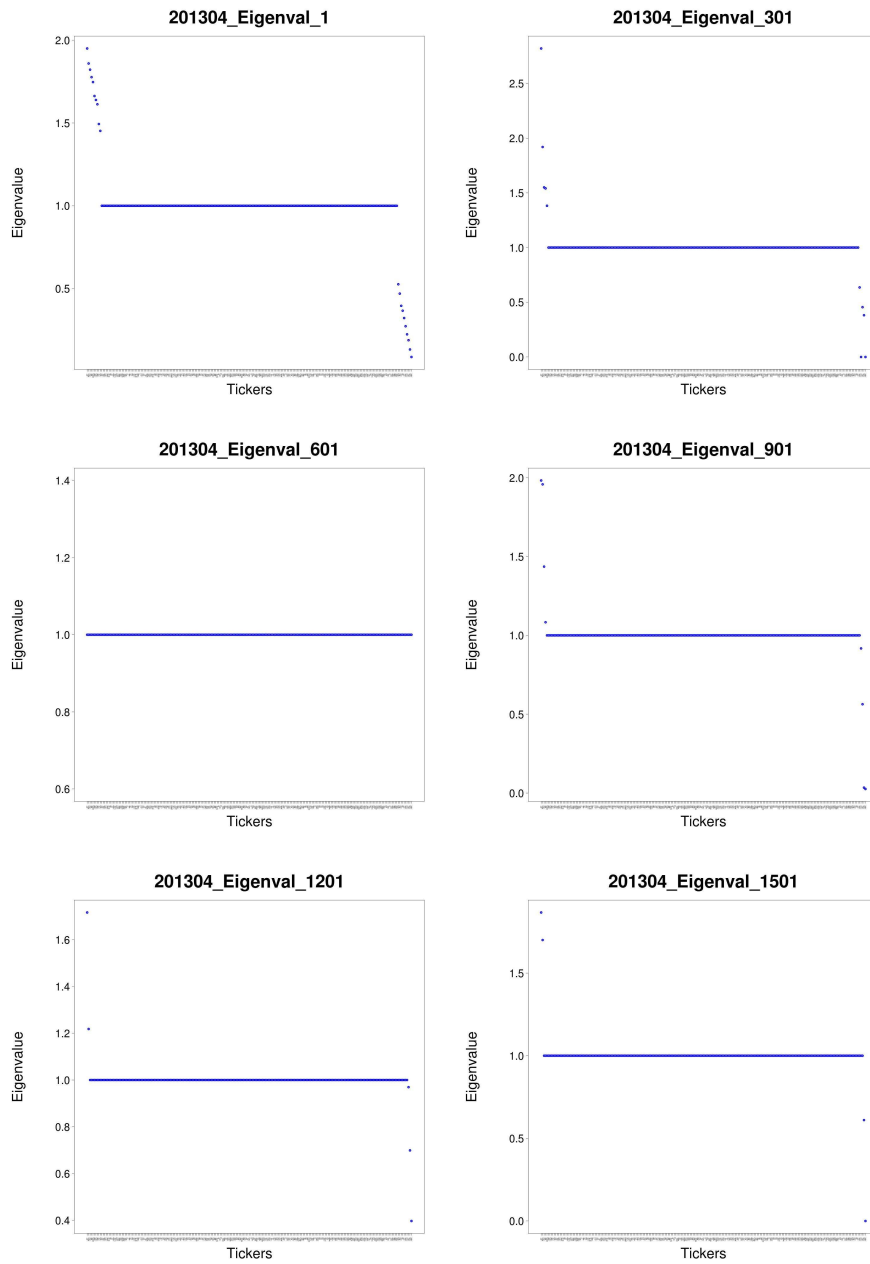
Figure 4.16: Partial Correlation Network Stock VPT eigenvalue plot at time: (a)1, (b)301, (c)601, (d)901, (e)1201, (f)1501. (From left to right, top to down)

Next, we move on to calculate the percentage of each eigenvalue over the sum of all eigenvalues. The percentage of maximum eigenvalue also indicates how large the proportion of the stock with maximum eigenvalue is compared to the other tickers. In the following pictures, the percentage of eigenvalues for 223 tickers are plotted in Figure 4.17 for VAR(1) adjacency network with stock returns, Figure 4.18 for VAR(1) adjacency network with VPTs, Figure 4.19 for partial correlation network with stock returns, and Figure 4.20 for partial correlation network with VPTs. It is with no doubt that the distribution of eigenvalue percentages is the same as absolute eigenvalues. One thing to mention, once the percentage of the maximum eigenvalue is greater than 0.5, then the related ticker is of great interest. See Figure 4.17 for instance, the percentage of eigenvalue of ticker HUM is more than 80%, which is also a node with a lot of edges which can be found out in Figure 4.9. We find that almost all tickers of interest have a large proportion over the sum of eigenvalue percentages, which means that the ticker with maximum eigenvalue might be a hidden driven force of the whole market.
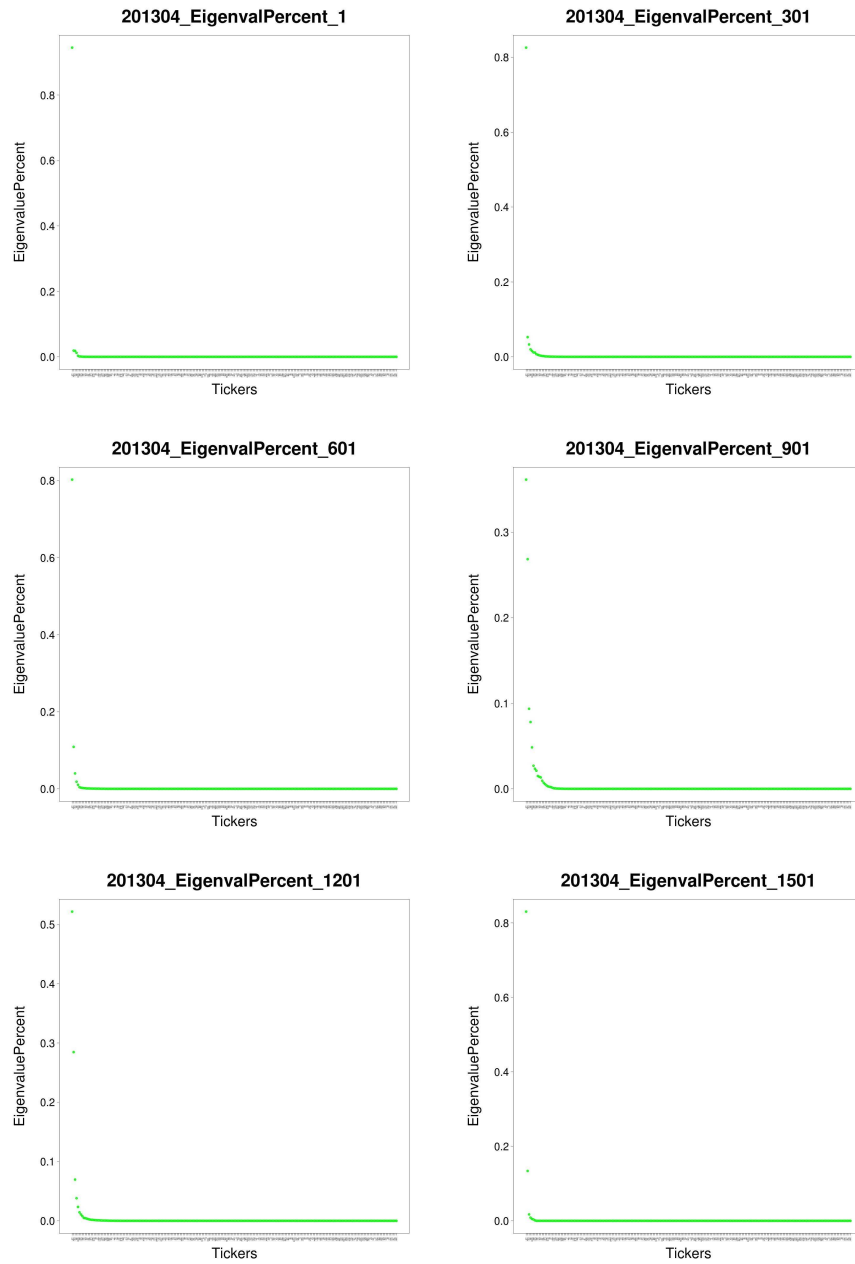
Figure 4.17: VAR Adjacency Network Stock Return eigenvalue percent plot at time: (a)1, (b)301, (c)601, (d)901, (e)1201, (f)1501. (From left to right, top to down)
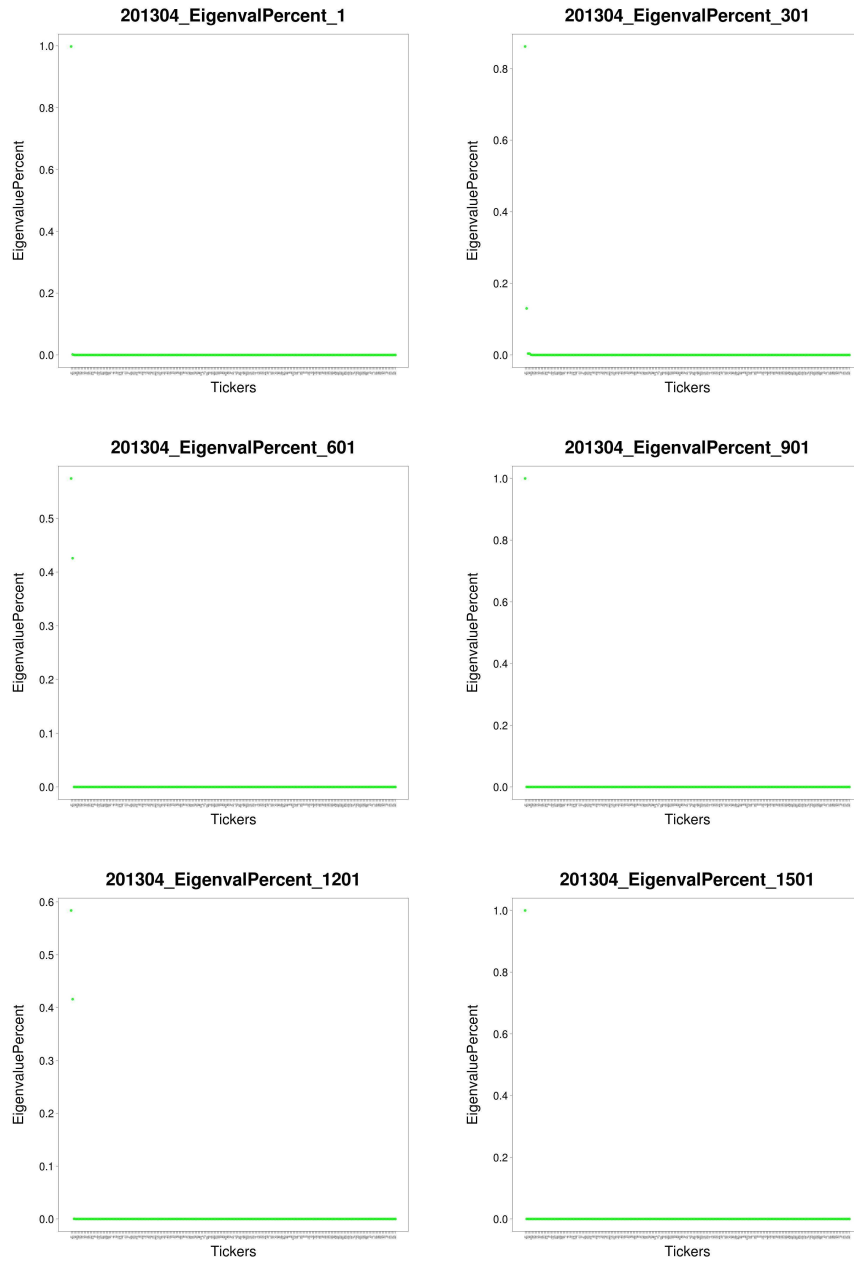
Figure 4.18: VAR Adjacency Network Stock VPT eigenvalue percent plot at time: (a)1, (b)301, (c)601, (d)901, (e)1201, (f)1501. (From left to right, top to down)
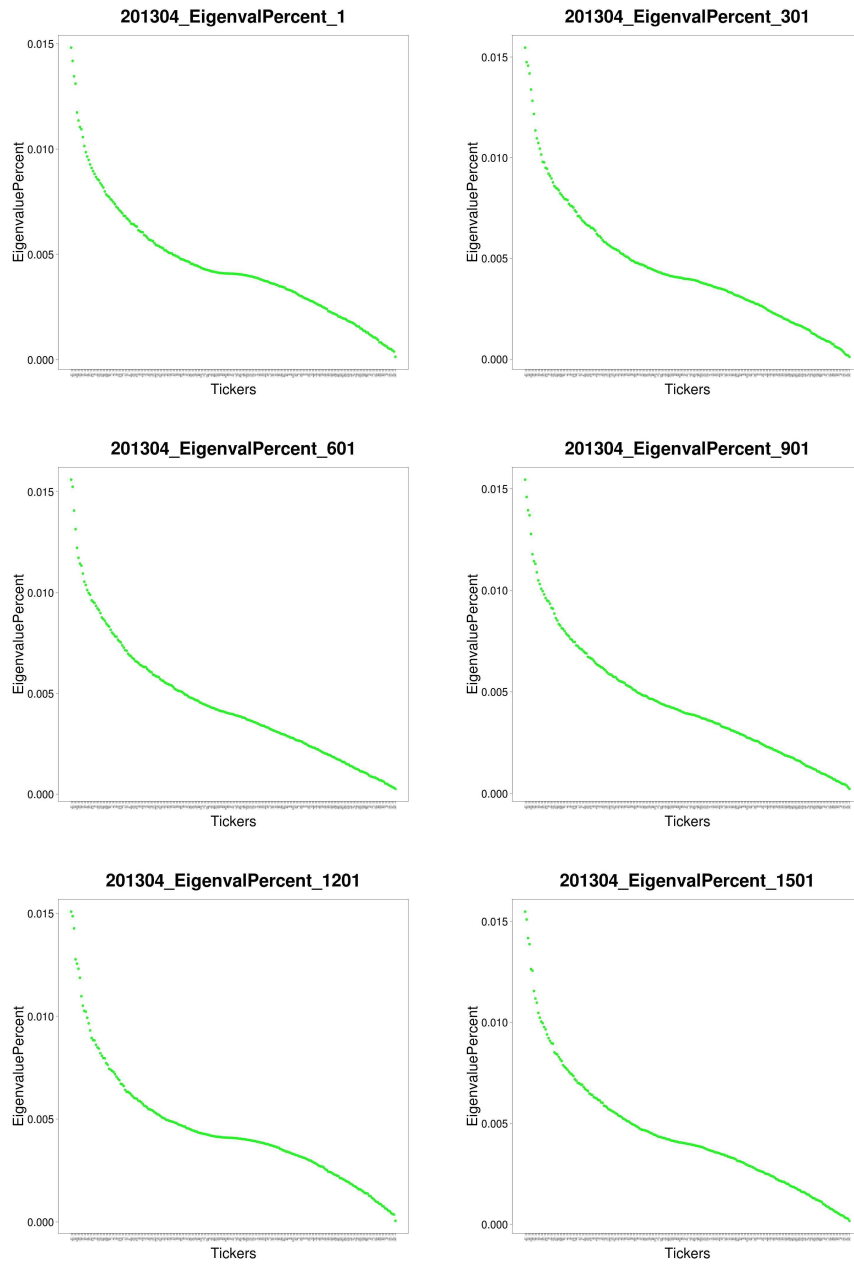
Figure 4.19: Partial Correlation Network Stock Return eigenvalue percent plot at time: (a)1, (b)301, (c)601, (d)901, (e)1201, (f)1501. (From left to right, top to down)
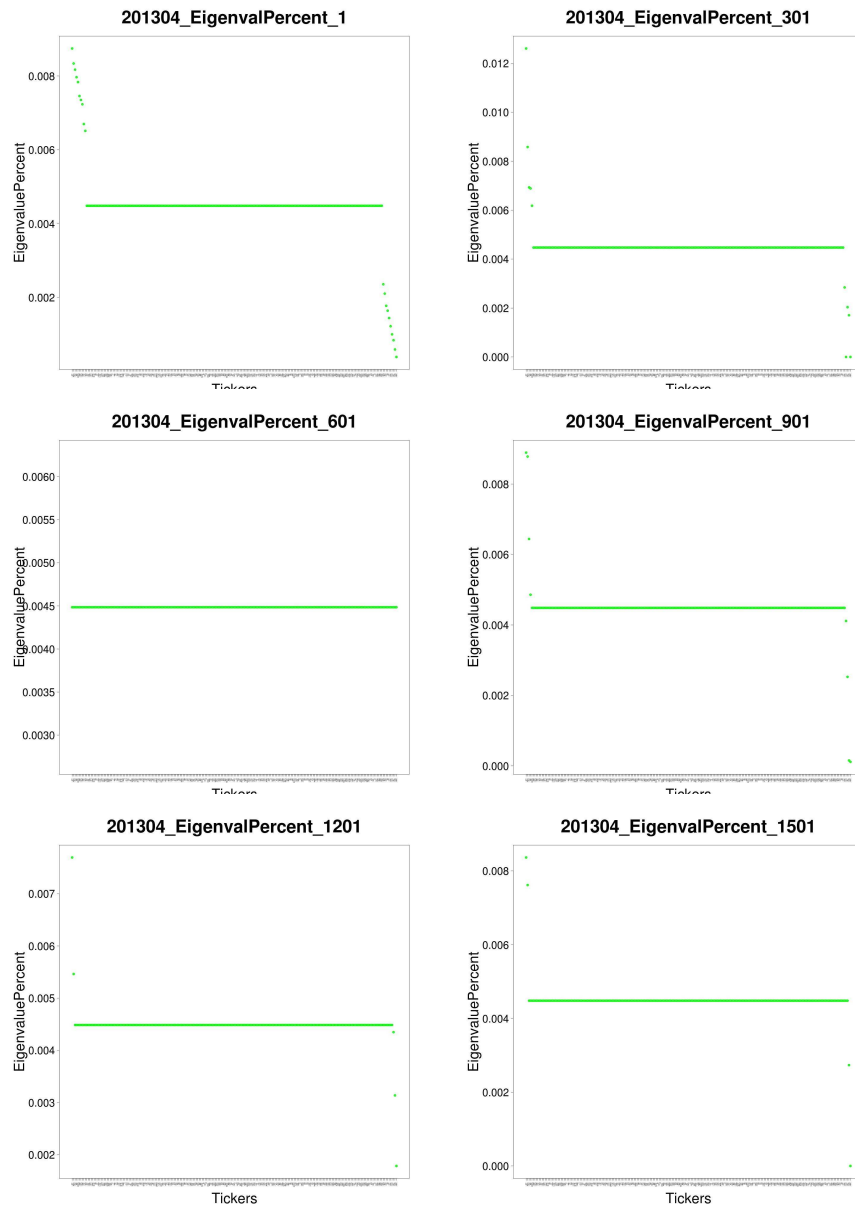
Figure 4.20: Partial Correlation Network Stock VPT eigenvalue percent plot at time: (a)1, (b)301, (c)601, (d)901, (e)1201, (f)1501. (From left to right, top to down)

56

Furthermore, we look into the eigenvectors related to maximum eigenvalue. Among of the following pictures, the eigenvectors related to the maximum eigenvalue for 223 tickers are plotted in Figure 4.21 for VAR(1) adjacency network with stock returns, Figure 4.22 for VAR(1) adjacency network with VPTs, Figure 4.23 for partial correlation network with stock returns, and Figure 4.24 for partial correlation network with VPTs. From this set of plots, we find that as the value of maximum eigenvalue increases, the distribution of the related eigenvectors get more condense, most of whose value are zeros, see Figure 4.21 and Figure 4.22. On the other hand, if the value of the maximum eigenvalue is not significantly larger than the other eigenvalues, the distribution of the related eigenvectors get very scattered, see Figure 4.23 and Figure 4.24.
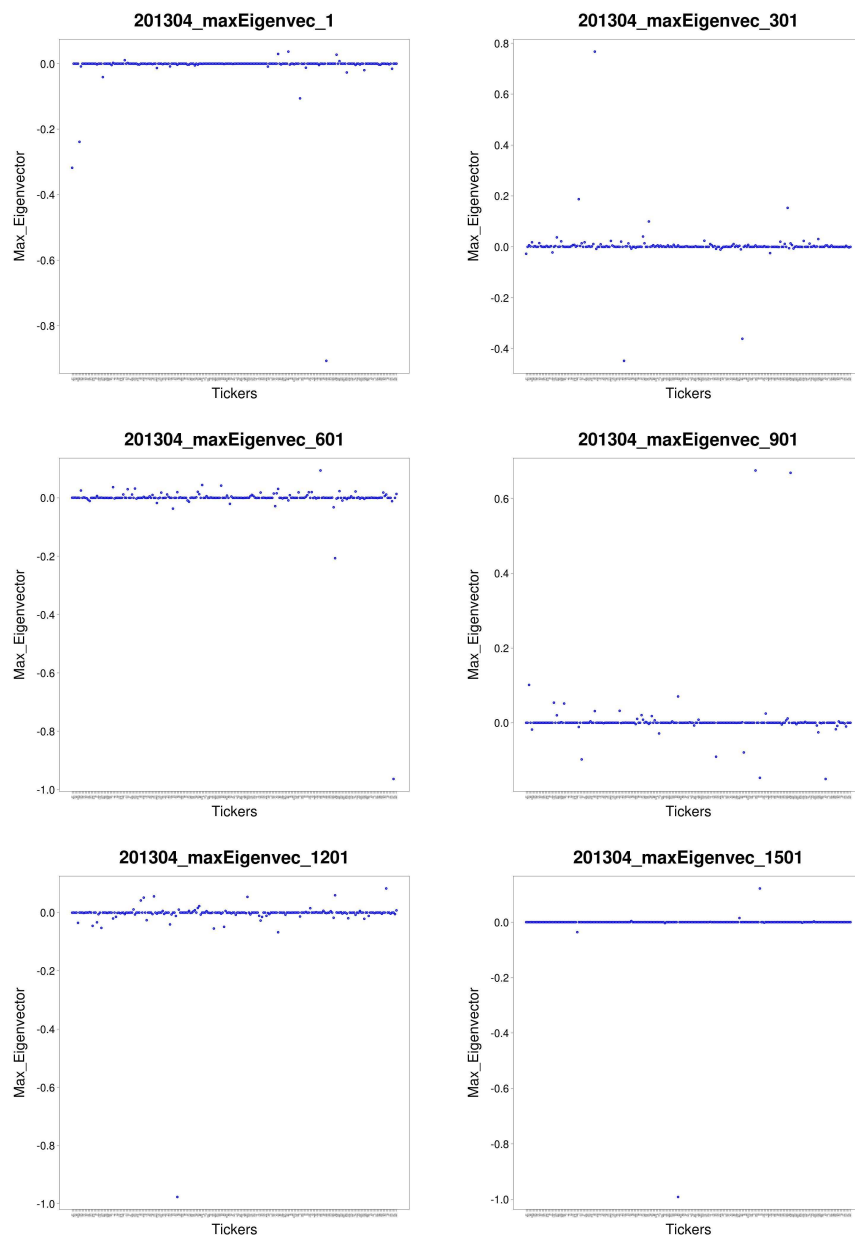
Figure 4.21: VAR Adjacency Network Stock Return eigenvector plot at time: (a)1, (b)301, (c)601, (d)901, (e)1201, (f)1501. (From left to right, top to down)
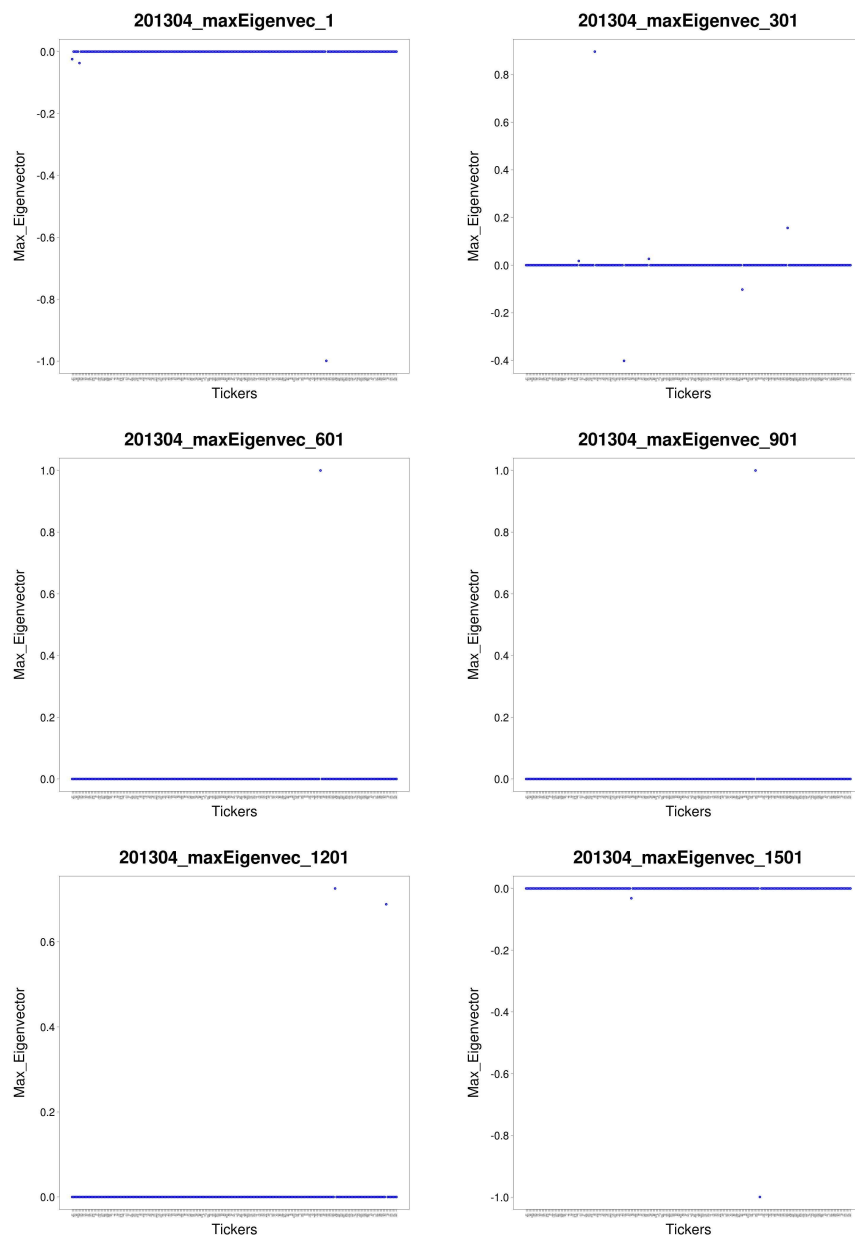
Figure 4.22: VAR Adjacency Network Stock VPT eigenvector plot at time: (a)1, (b)301, (c)601, (d)901, (e)1201, (f)1501. (From left to right, top to down)
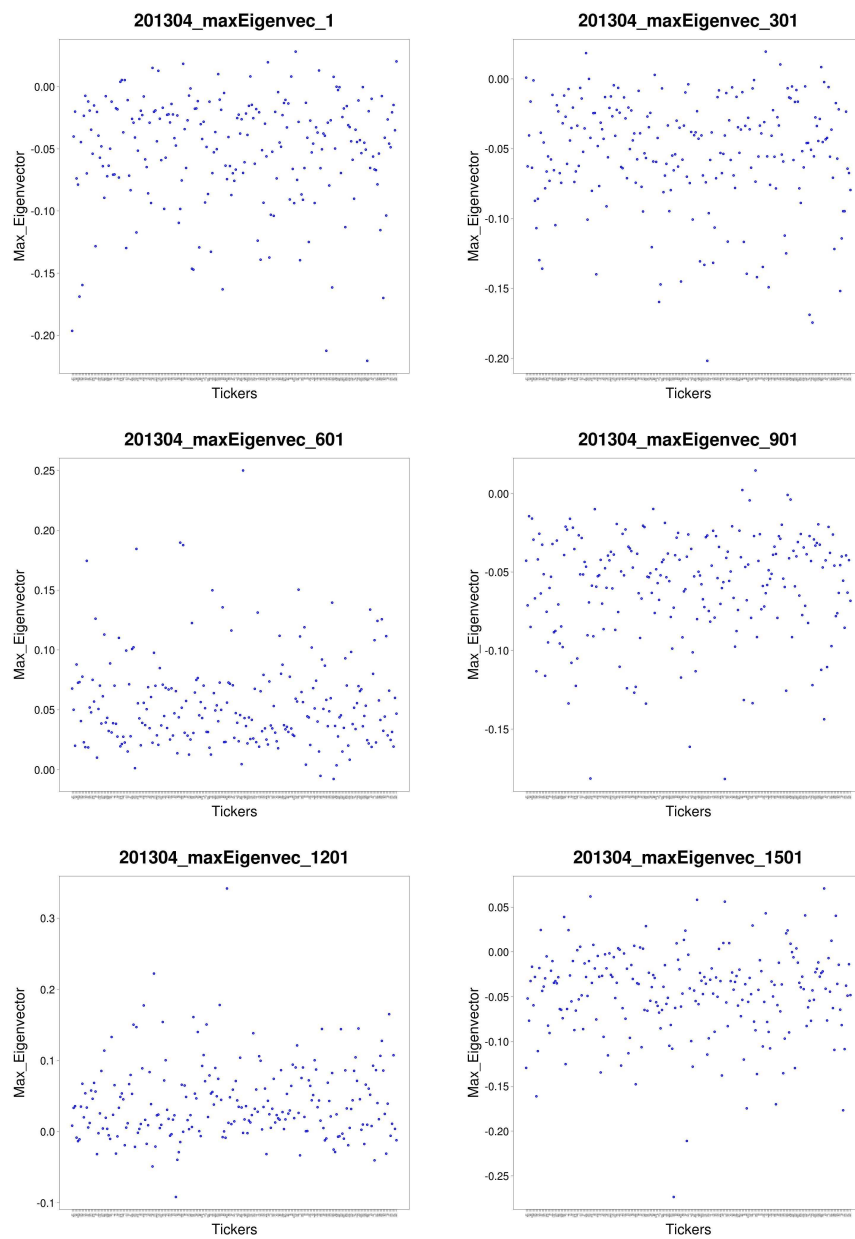
Figure 4.23: Partial Correlation Network Stock Return eigenvector plot at time: (a)1, (b)301, (c)601, (d)901, (e)1201, (f)1501. (From left to right, top to down)

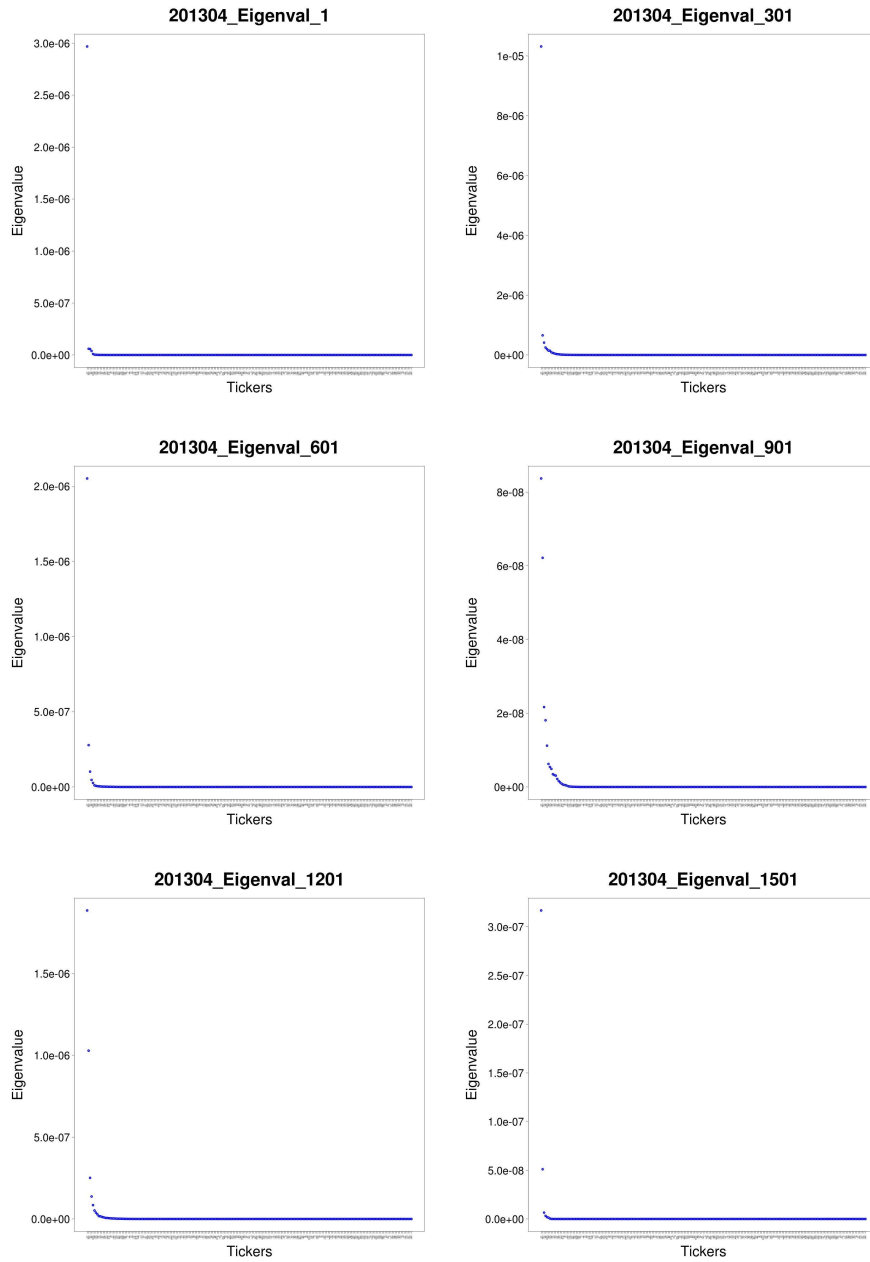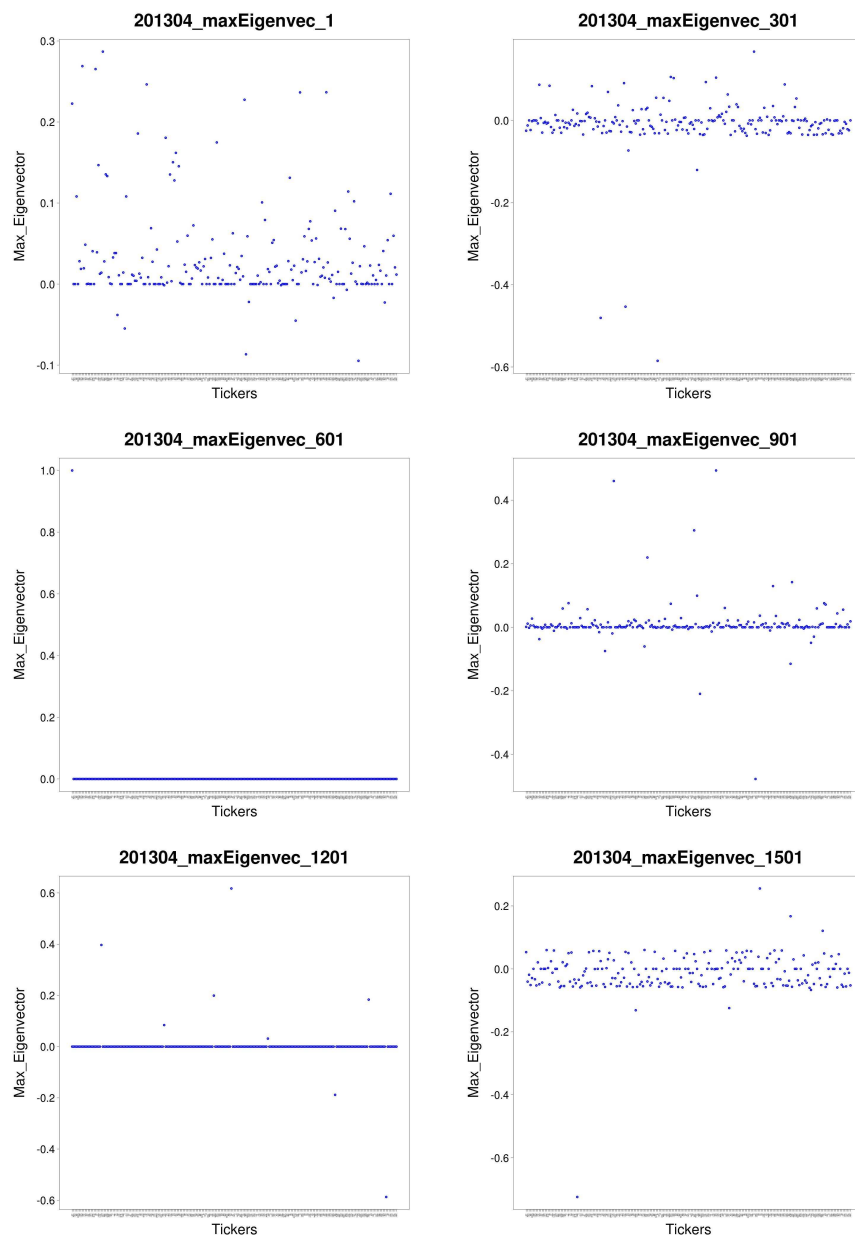Figure 4.24: Partial Correlation Network Stock VPT eigenvalvector plot at time: (a)1, (b)301, (c)601, (d)901, (e)1201, (f)1501. (From left to right, top to down)

Maximum Eigenvalues Evolve through Time

In the previous part, we compare four different models at a certain time point. Now we consider the change of maximum eigenvalues over the whole time period from April 2013 to September 2013. See Figure 4.25 for the comparison of maximum eigenvalues, Figure 4.26 for the comparison of maximum eigenvalues quantile, and Figure 4.27 for the comparison of maximum eigenvectors quantile.

Among these plots, Figure 4.25(a) shows how the maximum eigenvalues of VAR(1) adjacency stock return matrix with change as time goes by, Figure 4.25(b) shows the evolution of maximum eigenvalues of VAR(1) adjacency VPT matrix, Figure 4.25(c) shows the evolution of maximum eigenvalues of partial correlation stock return matrix, and Figure 4.25(d) shows the evolution of maximum eigenvalues of partial correlation VPT matrix. From the four plots, we summarize some time points of interest, which are (80) in April, (485) in May, (725) and (805) on June, (1055) and (1185) on July, (1235) and (1405) in August, (1625) in September. Actually, in the next part of thesis, we will carry out change point detection method over the same dataset and consider if these points of interest are the structural breaks over the whole time period.

Moreover, we have another two sets of figures plotted named with quantile, which means that we take the 95%, 50% and 5% quantile respectively from the distribution of the eigenvalues and eigenvectors related to the maximum eigenvalues. Figure 4.26(a) shows how the 95%, 50% and 5% quantile of eigenvalues of VAR(1) adjacency stock return matrix with change as time goes by, Figure 4.26(b) shows the evolution of the 95%, 50% and 5% quantile of eigen-

values of VAR(1) adjacency VPT matrix, Figure 4.26(c) shows the evolution of the 95%, 50% and 5% quantile of eigenvalues of partial correlation stock return matrix, and Figure 4.26(d) shows the evolution of the 95%, 50% and 5% quantile of eigenvalues of partial correlation VPT matrix. Also, see Figure 4.27(a) shows how the 95%, 50% and 5% quantile of eigenvectors related to the maximum eigenvalue of VAR(1) adjacency stock return matrix with change as time goes by, Figure 4.27(b) shows the evolution of the 95%, 50% and 5% quantile of eigenvectors related to the maximum eigenvalue of VAR(1) adjacency VPT matrix, Figure 4.27(c) shows the evolution of the 95%, 50% and 5% quantile of eigenvectors related to the maximum eigenvalue of partial correlation stock return matrix, and Figure 4.27(d) shows the evolution of the 95%, 50% and 5% quantile of eigenvectors related to the maximum eigenvalue of partial correlation VPT matrix.
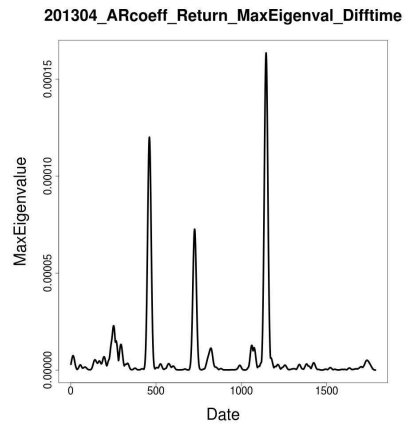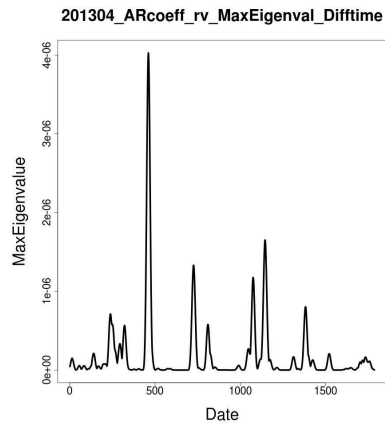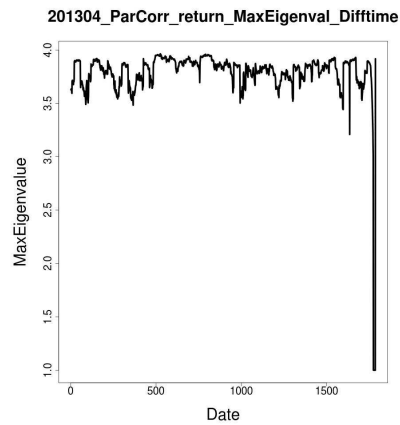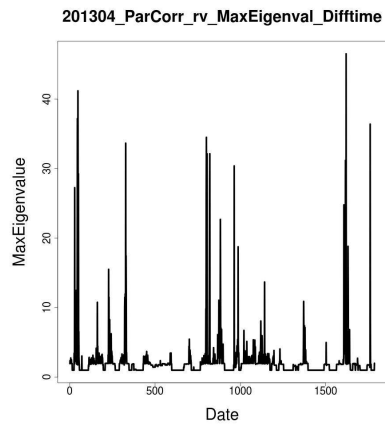
Figure 4.25: (a)VAR Adjacency Network Stock Return maxEigenval; (b)VAR Adjacency Network Stock VPT maxEigenval; (c)Partial Correlation Network Stock Return maxEigenval; (d)Partial Correlation Network Stock VPT maxEigenval. (From left to right, top to down)
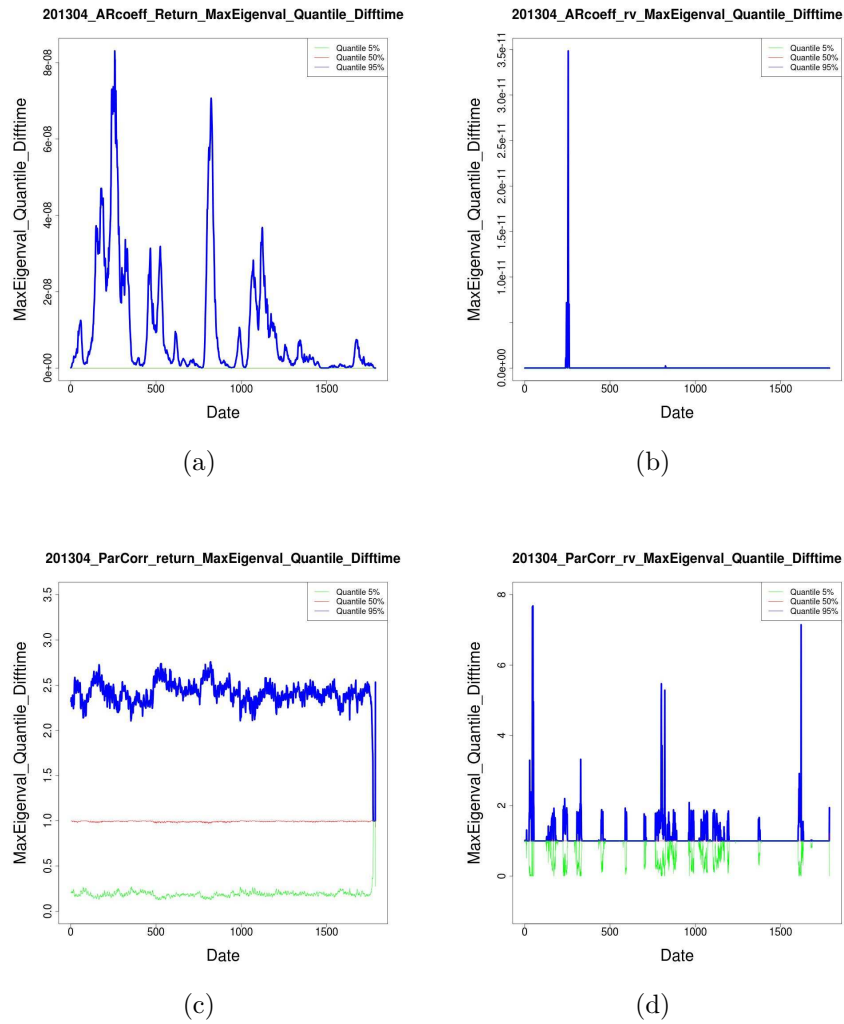
Figure 4.26: (a)VAR Adjacency Network Stock Return maxEigenval quantile; (b)VAR Adjacency Network Stock VPT maxEigenval quantile; (c)Partial Correlation Network Stock Return maxEigenval quantile; (d)Partial Correlation Network Stock VPT maxEigenval quantile. (From left to right, top to down)
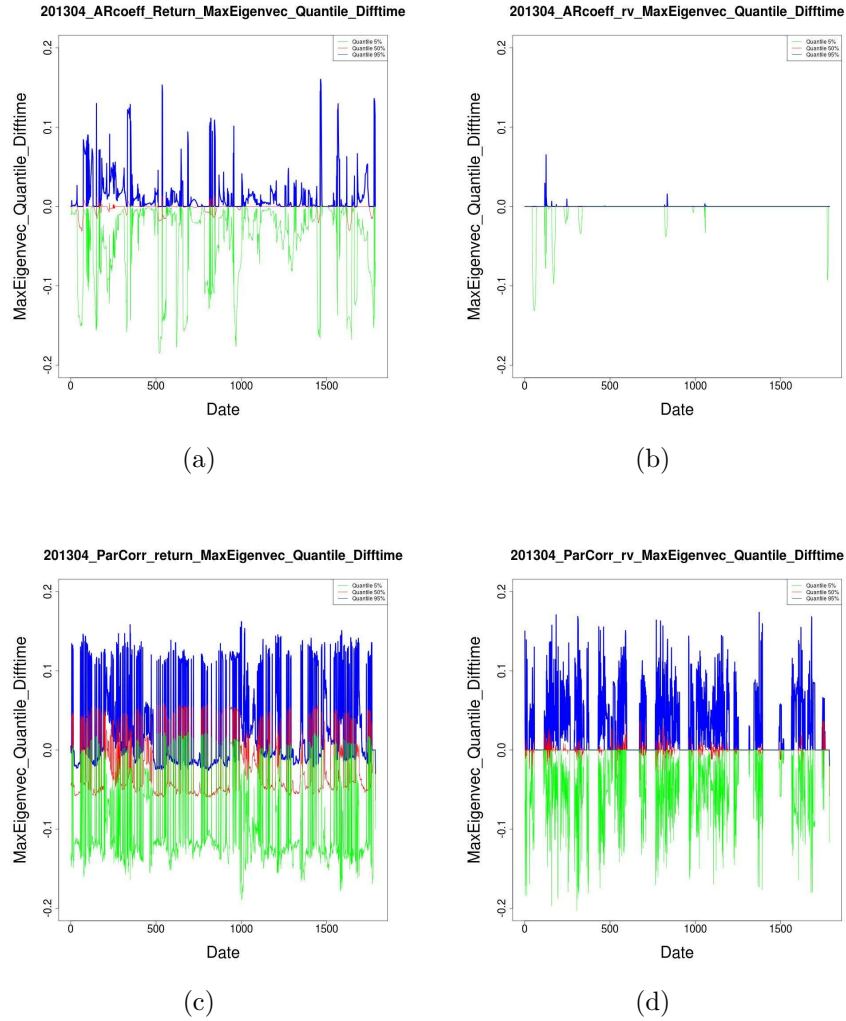
Figure 4.27: (a)VAR Adjacency Network Stock Return maxEigenvec quantile; (b)VAR Adjacency Network Stock VPT maxEigenvec quantile; (c)Partial Correlation Network Stock Return maxEigenvec quantile; (d)Partial Correlation Network Stock VPT maxEigenvec quantile. (From left to right, top to down)

66

4.3 Concluding Remark and Future Work

Throughout this part of thesis, two computational efficient methods are proposed to estimate both time-varying partial correlation networks and VAR adjacency network under the assumption of sparsity. We offer two thorough sets of methodologies including recursive algorithm, parameter tuning, and smoothing method for the estimation of networks. Moreover, after the setting of the theoretical deduction and data processing, we carry out simulation study and two empirical studies. Hence, we confirm the efficiency and validity of our proposed methods, especially in the case of presenting cross-sectional interconnections between stocks in financial market.

We have applied our proposed methods to 30-minute stock returns and VPTs over the time period from April 2013 to September 2013 and presented different kinds of output plots to discover the dependency network in the whole market, including partial correlation network and VAR adjacency network. It is proved that our methods works well in estimation both the directed and undirected time-varying network and can help us in interpretation of market stability and liquidity. However, it is just a start in this topic. In the future work, we plan to include more time periods in our model and investigate more details about the tickers of interest.

Part II

Detection Rules of Network Variations in

Intraday Stock Market

Chapter 5

Introduction

5.1 Motivation for Financial Surveillance

As stated by the former researchers: "Finance is, as it were, the stomach of the country, from which all the other organs take their tone." It expressed the importance of finance for the economy. Nowadays, financial markets play a more and more important role in economic growth. However, financial markets not always play a positive role in our society. Therefore, it is important for every one of financial market participants to ensure the quality of the markets and the fair access to it. Generally speaking, market efficiency and market integrity principles mean a lot for the whole society. The market efficiency states the condition where all market participants have the ability to transact easily at low costs, while the market integrity describes the ability to transact in a fair and informed market where prices reflect all publicly available information (Rydge, 2006).

Indeed, continuous monitoring and surveillance is in strong need to ensure the normal operation of financial market systems. In fact, the structure

of the financial markets has changed significantly and unfortunately, market efficiency and market integrity principles have been suffered a lot over the past 40 years. Firstly, market structural breaks are caused by several reasons, such as changes in the number and type of trading orders, the number and type of financial products and so on. However, the explosive growth of financial products and overwhelming developed financial market have not been matched by appropriate surveillance structures to allow regulators and market participants to collect and uncover information on market activities. For most people, the structural breaks in the market are difficult to forecast and figure out. Nowadays, transactions are more easy to made and the related costs decrease dramatically. Thus, improved connectivity and convenience to the markets also imply increased complexity and contagious effects especially when things go wrong. In fact, the economic crises, including the latest and still ongoing global financial recession started in 2008, make it important and urgent to formulate efficient methods for continuous financial surveillance, see Frisen(2008).

According to (Cumming, 2008), monitoring and surveillance systems in the financial market are far less understood and much less efficient than one could expect. Both academic researchers and market participants have taken an active interest and trials to the development of this field. Polansky (2004) stated in his paper that "market surveillance is defined to encompass the processes and technologies that support the detection and investigation of potential trading rule violations, whether defined in statute or marketplace rules" and "the components of market monitoring and surveillance can be

broken down into detection, investigation and enforcement tasks, which are supported by different processes and technologies".

In our research, we primarily focus on the "detection" task in financial market surveillance. As stated in the first part of thesis, the amount of data collected in financial markets has been growing exponentially during the last few years, compared with other fields. The information overload makes it very difficult to handle even intraday databases since one single stock may have huge amount of trades in a day. Therefore, finding an appropriate way to statistically analyze and extract useful information from this huge amount of data is very important for financial surveillance, according to Ngai (2011). In previous part of thesis, we present two new approaches to dynamically estimate symmetric and asymmetric dependency network between stocks. In this part of thesis, we consider the problem of efficient financial surveillance aimed at prompt detection of structural breaks in multivariate normal distributed time series in macromarket. Assuming the model evolves in a piece-wise constant fashion, we carry out a detailed literature review and analytical comparison about all the detection rules in the next section.

5.2 Literature Review of Detection Methods

In this section, we provide an overview of the detection methods of sequential change points assuming unknown pre- and post-change distributions. We cover almost all major formulations of the stochastic structural break model and pay particular attention to the cutting-edge advances in each.

The problem of change detection is comprised of three elements: a stochas-

tic process under observation, a change-point at which the statistical properties of the process suffer a change, and a decision mechanism whose target is to detect this structural change by making use of the statistical properties of the ongoing process. If the change is observed by the decision mechanism before the change actually takes place, we call it a false alarm event. The mean idea of change-point detection is to implement algorithm which can be applied to detect the change as soon as possible, and continue to move on for the next change-point. There are a lot of important applications in change-point detect problem, including quality control engineering, financial markets, intrusion detection in computer networks and security systems and so on.

The subject of change-point detection began to emerge in 1920s for the sake of quality control. Earliest results date back to the work of Shewhart (1931), Wald(1947) and Page (1954). Shewhart proposed to use a control chart to detect a change-point, in which the statistics taken in real time are plotted on a chart and an alarm is raised when the first time the statistics fall outside some pre-defined control limits. The statistics in control chart are calculated with a function of only the measurements at that time. However, it may result in a loss of information in the past. Page(1954) stated that a weighted sum (moving average chart) or a cumulative sum of the past statistics can be applied in the control chart to detect the change more efficiently. Studied by Shiryaev(1963), under the further assumption that the change-point time is a random variable with a known geometric distribution, one obtains an optimal algorithm on minimizing the expected detection delay over all stopping times. This is the first time when change-point detection method

is formulated in the Bayesian framework. As suggested by Roberts(1966), an algorithm called the Shiryaev-Roberts algorithm is obtained by taking a limit on Shiryaevs Bayesian solution as the geometric parameter of the change point goes to zero. However, it is originally impossible to obtain an comprehensive algorithm which can work efficiently over all possible values of the change-point. Therefore, a minimax approach is required. According to Lorden(1971), the first minimax theory is proposed in which detection delay is obtained by taking the supremum over all possible change-point of a worst-case delay over all possible observations. Pollak(1985) modified Lordens minimax theory by replacing the double maximization by a single maximization over all possible change-point of the detection delay conditioned on the change point. As pointe out by Lai (1998), Lorden's highly conservative worse-case expected deal is related to the lower bound imposed on the average run length as a false alarm criterion. Motivated by fault detection in control systems, Lai(1995, 2000) proposed to a new evaluation criteria which not only involve expected delay but also the stead-state false alarm probability.

We consider in the case where the pre- and post-change of multivariate observations are not pre-defined. In other words, the theoretical setup of the change-point detection problem is for a series of observations in which, conditioned on the change point, the observations are independent and identically distributed with some unknown distribution before the change point, and independent and identically distributed with some other unknown distribution after the change point. Mei(2008)'s discussions of this topic are considered too simple for the surveillance application, according to Frisen(2008). The

bayesian approach developed by Lai and Xing (2009), involves a class of hidden Markov models that have tractable filters, which can be approximated by parallel recursive algorithms. With the application of Elliott (2005)'s general jump Markov model, one can generalize to multiple change-point. And for realizing the assumption of unknown pre- and post-change distribution, one can replace the density function by a parametric family and a sequence of positive i.i.d random variables. In Xing (2012), he provided a general framework for multiple change-point in complex system and the related recursive algorithms.

Recently, testing on structural breaks has been an interested research topic. Levine(2001) uses the test in Perron(1989) 's paper on structural breaks to evaluate changes of stock market liquidity after the policy change date. According to Noel (2008), people have realized that the structural breaks has contributed to weakening risk management. Finding a better way to assess systemic risk has been an important area to reduce the risk of crises. The importance and advantages of detection on structural breaks has been more obvious. Valentinyi Endresz(2004)'s study has given the result of whether detecting structural breaks in the volatility model can improve the Value at Risk forecast. Also, we should note that the model proposed by Xing (2012) essentially identify historical market structural breaks or sharp changes base only on credit rating records. In Xing (2012)'s model, rating transition process as piecewise homogeneous Markov chains with unobserved structural breaks. Their proposed model provides explicit formulas for the posterior distribution of the time-varying rating transition generator matrices, the probability of structural break at each period and prediction of transition matrices in the

74

presence of possible structural breaks. Estimating the model by credit rating history, Xing (2012) show that the structural breaks in rain transitions can be captured by the proposed model, and compare the prediction performance of their proposed and time homogeneous Markov chain models. Motivated by the surveillance model of credit rating system in Xing(2012), we conduct research on detecting the structural breaks in intraday stock market using high dimensional multivariate normal distributed data in the market. In more details, the contribution of our research can be summarized in several different aspects. First, in contrast to the most frequently used measurements for change-point detection, we offer statistical techniques for sequentially detecting a structural change in a multidimensional data process. Instead of trying to identify an unknown structural break point in the past, we aim at providing a decision rule whether we observe a change at the current point. Second, the thesis extends the existing literature of four types of detection rules on multivariate data, including statistical process control chart, generalized likelihood ratio detection rule, a detection method based on an extension of Shiryaev's Bayesian single change point model and a sequential detection rule for multiple change points initiated by Xing(2012) and apply the modified methods to the high dimension of observations in multivariate normal distribution. Third, we make use of the conjugacy property of Inverse Wishart family and study the efficiency of the proposed methods on both simulation studies and the empirical analysis focusing primarily on the intraday stock market.

5.3 Outline

The thesis is organized as follows. Chapter 6 presents stochastic structural break model and the associated detection rules on single change point and multiple change points: Chapter 6.1 present the statistical process control chart detection method; Chapter 6.2 describes generalized likelihood ratio (GLR) detection rule; Chapter 6.3 describes the extension of Shiryaev's Bayesian change-point and detection rule; Chapter 6.4 shows a sequential detection full for multiple structural breaks. In Chapter 7, we carry out the simulation studies of four detection methods proposed: all the change point generating process is listed in Chapter 7.1; eight different scenarios are described in details in Chapter 7.2; a discussion about how the theoretical thresholds are chosen is given in Chapter 7.3; and Chapter 7.4 present the simulation results of four detection rules to illustrate the performance of our proposed detection rule and carry out a detailed theoretical and analytical comparison of our proposed approaches to detect the change-point. In Chapter 8, we present the performance of different detection rules with real dataset: we state the hyper-parameter selection procedure in Chapter 8.1.1; and threshold determined method is given in Chapter 8.1.2; Chapter 8.2 summarize our results in comparison with the concluding remarks of Part 1.

Chapter 6

Stochastic Structural Break Model and Associated Detection Rules

In this section, to detect the structural breaks in multivariate normal distributed data stream, we present four different kinds of methods including statistical process control chart detection method, generalized likelihood ratio detection rule, detection method based on an extension of Shiryaev's Bayesian single change-point model and a sequential detection rule for multiple structural breaks. We take advantage of the above methods to formulate novel surveillance models in detecting structural breaks of time-varying multivariate normal distributed dataset.

6.1   Statistical Process Control Chart Detection Method

The methods of statistical process control (SPC) are frequently used in order to detect changes in model parameters of multivariate time series. The applications of SPC can be found in different kinds of fields, such as quality

control in production, risky indicators of macroeconomics and environmental evaluations. The main idea of statistical process control is to detect variations during an on-line process from a predefined target process, as soon as a change-point occurred. The most important tools of SPC are control charts, which were introduced for the first time by Shewhart(1982), for the original purpose of surveillance in production process. A Shewhart-type control chart works for controlling the mean of an industrial process based on the so-called Mahalanobis distance statistic.

A typical control chart is consisted of the control statistic and the control limits.The observed incoming data stream is examined in the form of control statistic sequentially. If at a certain time point, the control statistic keeps within the control limits, we conclude that the process is still in control and the procedure will move on to the next time point. If the control statistic exceeds the control limit, the procedure stops for the reason that the process is thought to be out of control.

In our research, we consider a special control chart based on the exponentially weighted moving average(EWMA) statistic, which is proposed for the first time by Roberts(1959) using the exponential smoothing method. Lowry(1992) extended the application of multivariate EMWA charts to monitoring the multivariate time series, which has the form as following:

$$Z_t = (I - R)Z_{t-1} + R\tau_t, \quad t \geq 1 \tag{6.1}$$

where Z_t is the n dimensional veto of EWMA statistic at time t, I denotes the

$n \times n$ dimensional identity matrix, τ_t is the vector of the observed parameter at time t and $R = diag[\alpha_i]_{1 \leq i \leq n}$ is a diagonal weighting matrix with elements indicating a smoothing parameter for the ith component of the vector τ_t.

Schmid and Tzothchev(2004) formulated a one-factor Cos Ingersoll Rox (CIR) model, which is a celebrated model of the general approach to modeling the term-structure. They transformed the multivariate statistics into a univariate quantity with the application of multivariate EWMA recursion. In other words, they presented a univariate statistic and then applied the EWMA recursion to it. The EWMA statistic is also based on the Mahalanobis distance denoted as T_t:

$$T_t = (\tau_t - \mu_t)^T \Sigma_t^{-1} (\tau_t - \mu_t) \tag{6.2}$$

where μ_t refers to the mean value of τ_t and Σ_t denotes the covariance of τ_t. Thus the EWMA statistic is given by

$$Z_t = (1 - \alpha)Z_{t-1} + \alpha T_t, \quad t \geq 1 \tag{6.3}$$

In our research, we make use of the above control statistic to sequentially detect a single change-point. We choose one-factor version of the CIR model and build up suitable statistics for uncovering the structural change in the parameters of the model. Suppose that we have multivariate normal distributed observations at time t, written as $X_1^t, ..., X_n^t \sim N(\mu, \Sigma_t)$. Then based on equation 6.3, we modify the decision statistic Z_t based on an EWMA recursion on singular

value decomposition (SVD) of $M(\Sigma)$ as following:

$$Z_t = (1 - \alpha)Z_{t-1} + \alpha M(\Sigma)_t, \ t \geq 1 \tag{6.4}$$

with $0 < \alpha \leq 1$ being the smoothing parameter, where the SVD metric for a covariance matrix P is defined as

$$M(\Sigma)_t = \frac{1}{n} \sum_{i=1}^{n} \sqrt{e_i(\Sigma_t - I)'(\Sigma_t - I))} \tag{6.5}$$

in which I is the $n \times n$ identity matrix and $e_i(.)$ is the ith eigenvalue of the matrix.

Combining the above model, we derive a simple change-point detection rule, with a change-point at t, but not before t_p as

$$T = \inf \{t > t_p : Z_t \geqslant \theta\} \tag{6.6}$$

A signal is provided when $Z_t > \theta$, where $\theta > 0$ stands for the control limit. If the Z_t statistic lies within the rejection area, the observed process is considered to be out-of-control at time t. Alternatively, if the Z_t statistic lies within the the acceptance area, the observed process is considered to be the in-control state at time t. The acceptance and the rejection areas are one-sided interval determined by a critical value θ. And average run length (ARL), which denotes the average number of observations until a "false" signal is obtained, is the most important criterion to determine the control limit. θ is chosen such that the in-control average run length of the chart is equal to a fixed value.

From the above, we have formulated a control chart to detect the change-point in multivariate time series dataset. However, there are some differences in the purpose of control chart and that of change point estimation. Control chart has typically been focused more on detecting changes as quickly as possible, while in change-point estimation, the focus is usually on estimating the time when a most significant change takes place in the process.

6.2 Generalized Likelihood Ratio (GLR) Detection Rule

Let the observations, $X_1, X_2, ..., X_n$, be independent random vectors with a common density function f_0 for $t < v$, and another common density function f_1 when $t \geq v$. To solve the problem of optimal sequential detection of the change-time v, Shiryaev (1978) proposed Bayesian approach by putting a geometric prior distribution on v and assume that there is a loss of c for each observation taken at or after v and a loss of 1 for a false alarm before v. Using optimal stopping theory, he proved that when a change has occurred, the Bayes rule would trigger an alarm as soon as the posterior probability exceeds some fixed threshold. Since

$$P\{v \leq n | X_1, ..., X_n\} = R_{p,n}/(R_{p,n} + p^{-1}) \tag{6.7}$$

where p is the parameter of the geometric distribution $P\{v = n\} = p(1-p)^{n-1}$ and

$$R_{p,n} = \sum_{k=1}^{n} \prod_{i=k}^{n} \{\frac{f_1(X_i)}{(1-p)f_0(X_i)}\}, \tag{6.8}$$

the Bayes rule declares that at time

$$N(\gamma) = \inf\{n \geq 1 : R_{p,n} \geq \gamma\} \tag{6.9}$$

that a change has occurred. Roberts(1966) introduced the case of $p = 0$ in equation 6.9 and formulated the Shiryaev-Roberts rule, which can be expressed as

$$\tilde{N}(\gamma) = \inf\left\{n \geq 1 : \sum_{k=1}^{n}\prod_{i=k}^{n}\frac{f_1(X_i)}{f_0(X_i)} \geq \tilde{\gamma}\right\} \tag{6.10}$$

Pollak(1985) also proved that when $p \to 0$, it is asymptotically Bayes risk efficient. Instead of the Bayesian approach, Lorden(1971) used the minimax approach of minimizing the worse-case expected delay

$$\bar{E}_1(T) = \sup_{v \geq 1} ess\ supE[(T - v + 1)^+|X_1, ..., X_{v-1}] \tag{6.11}$$

over the class $\mathcal{F}_\gamma$ of all rules T satisfying the constraint $E_0(T) \geq \gamma$ on the expected duration to false alarm. He showed that as $\gamma \to \infty$, Page's (1954) CUSUM rule

$$\tau = \inf\left\{n \geq 1 : \max_{1 \leq k \leq n}\sum_{i=k}^{n}\log(\frac{f_1(X_i)}{f_0(X_i)}) \geq c\right\} \tag{6.12}$$

with c so chosen that $E_0(\tau) = \gamma$, is asymptotically minimax in the sense that

$$\bar{E}_1(\tau) \sim \inf_{T \in \mathcal{F}_\gamma} \bar{E}_1(T) \sim \frac{\log \gamma}{E_1\{\log \frac{f_1(X_t)}{f_0(X_t)}\}} \tag{6.13}$$

82

Note that equation 6.12 essentially replaces $\sum_{k=1}^{n}$ in equation 6.10 by $\max_{1 \leq k \leq n}$, which can be regarded as using maximum likelihood to detect the unknown change-point.

According to Lai and Xing (2009), without assuming both of the pre- and post-change density functions to be specified in advance, suppose the observations belong to a multivariate exponential family with density function

$$f_\theta(x) = exp\left\{\theta' x - \psi(\theta)\right\} \tag{6.14}$$

They proposed to replace the likelihood ratio statistic in the CUSUM rule by the generalized likelihood ratio (GLR) statistic, which leads to GLR rule for testing the null hypothesis of no change-point, versus the alternative hypothesis of a single change-point prior to n but not before n_0. Thus, the GLR statistic based on $X_1, ..., X_n$ is

$$
\begin{aligned}
&\max_{n_0 \leq k \leq n} \left\{ \sup_{\theta} \sum_{i=1}^{k} \log f_\theta(X_i) + \sup_{\tilde{\theta}} \sum_{i=k+1}^{n} \log f_{\tilde{\theta}}(X_i) - \sup_{\lambda} \sum_{i=1}^{n} \log f_\lambda(X_i) \right\} \\
&= \max_{n_0 \leq k \leq n} \left\{ kL(\overline{X}_{1,k}) + (n-k)L(\overline{X}_{k+1,n}) - nL(\overline{X}_{1,n}) \right\}
\end{aligned} \tag{6.15}
$$

where

$$
\begin{aligned}
\overline{Y}_{m,n} &= \sum_{i=m}^{n} Y_i/(n-m+1), \tag{6.16} \\
L(\mu) &= \sup_{\theta} \left\{\theta^T \mu - \psi(\theta)\right\} = \left\{\theta_\mu^T \mu - \psi(\theta_\mu)\right\} \tag{6.17}
\end{aligned}
$$

and $\theta_\mu = (\triangledown \psi)^{-1}(\mu)$, $\triangledown$ denotes the gradient vector of partial derivatives. In equation 6.15, sup_λ is related to maximizing likelihood under the null hypoth-

esis, and sup_θ and $sup_{\tilde{\theta}}$ are obtained by maximizing the likelihood under the hypothesis of a single change-point occurring at $k+1$. For a simple representation, let

$$g(\alpha, x, y) = \alpha L(x) + (1-\alpha)L(y) - L(\alpha x + (1-\alpha)y) \qquad (6.18)$$

Therefore, we can rewrite the CUSUM rule in equation 6.12 by taking the GLR statistics as in equation 6.15 instead of the likelihood ratio statistics, which results in the GLR rule for detecting a change in θ of the multivariate exponential family when the pre- and post-parameters are unknown:

$$\hat{N} = \inf\left\{n > n_0 : \max_{n_0 \le k < n} ng(k/n, \overline{X}_{1,k}, \overline{X}_{k+1,n}) \ge c\right\} \qquad (6.19)$$

Let us now suppose that we have an independently, identically distributed data set $\mathcal{D} = X_1, ..., X_n$ sampled from a multivariate normal distribution $N(\mu, \Sigma)$. We want to make a quick detection of the change-points in this data stream. According to the above GLR detection rule, we form the log likelihood function by taking the logarithm of the product of n Gaussians, which is:

$$l(D|\mu, \Sigma) = -\frac{n}{2}\log|\Sigma| - \frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^T\Sigma^{-1}(x_i - \mu) + const \qquad (6.20)$$

Taking the derivative with respect to μ and setting it to zero, we have the maximum likelihood estimate of the mean:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n}X_i \qquad (6.21)$$

84

Similarly, we can also obtain the maximum likelihood estimate of the covariance matrix:

$$\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} (X_i - \hat{\mu})^T (X_i - \hat{\mu}) \tag{6.22}$$

Thus, we have the maximum log likelihood function based on $\mathcal{D} = X_1, ..., X_n$ as

$$
\begin{aligned}
l_{1,n} &= -\frac{n}{2} \log |\hat{\Sigma}| - \frac{1}{2} Tr(\hat{\Sigma}^{-1} \frac{1}{n} \sum_{i=1}^{n} (X_i - \hat{\mu})^T (X_i - \hat{\mu})) \\
&= -\frac{n}{2} \log |\frac{1}{n} \sum_{i=1}^{n} (X_i - \frac{1}{n} \sum_{i=1}^{n} X_i)^T (X_i - \frac{1}{n} \sum_{i=1}^{n} X_i)| - \frac{1}{2} Tr(nI_n) \\
&= -\frac{n}{2} \log |\frac{1}{n} \sum_{i=1}^{n} (X_i - \frac{1}{n} \sum_{i=1}^{n} X_i)^T (X_i - \frac{1}{n} \sum_{i=1}^{n} X_i)| - \frac{1}{2} n^2 \tag{6.23}
\end{aligned}
$$

Here $|.|$ denotes the determinant of the matrix and $Tr(.)$ refers to the trace of the matrix. In general case, the GLR statistic function $L(\bar{X}_{i,j})$ can be written as

$$L(\bar{X}_{i,j}) = -\frac{j-i+1}{2} \log |\frac{1}{j-i+1} \sum_{r=i}^{j} (X_r - \bar{X}_{i,j})^T (X_r - \bar{X}_{i,j})| - \frac{1}{2}(j-i+1)^2 \tag{6.24}$$

Rewrite the GLR statistics in equation 6.19, we can get the GLR detection method especially in the case of multivariate normal distribution.

We have presented the GLR approach in the case of multivariate normal distributed time series. However, as pointed out by Lai (2008), it is much more natural to adopt a full Bayesian change-point model to detect the unknown change-point time with the unknown pre- and post- change parameters. In the next section, we will provide a full Bayesian framework to estimate the

85

change-point time without specified parameters in advance.

6.3 Detection Method Based on an Extension of Shiryaev's Bayesian Single Change-Point Model

Lai and Xing (2010) used a mixing distribution which can be regarded as a Bayesian approach to handle the unspecified pre- and post-change parameters. The generalization form of the extension of Shiryaev's Bayesian change-point model and detection rule is proposed as following: for the multiparameter exponential family as defined in equation 6.14, let π be a prior density function on $\Theta := \left\{ \theta : \int e^{\theta' X} d\omega(X) < \infty \right\}$ given by

$$\pi(\theta; a_0, \mu_{\mathbf{0}}) = c(a_0, \mu_{\mathbf{0}}) exp \left\{ a_0 \mu_{\mathbf{0}}' \theta - a_0 \psi(\theta) \right\}, \ \theta \in \Theta \qquad (6.25)$$

where

$$\begin{aligned}
1/c(a_0, \mu_{\mathbf{0}}) &= \int_{\Theta} exp \left\{ a_0 \mu_{\mathbf{0}}' \theta - a_0 \psi(\theta) \right\} d\theta \\
\mu_0 &\in (\nabla \psi)(\Theta)
\end{aligned} \qquad (6.26)$$

According to Diaconis and Ylvisaker (1979), the posterior density of θ given the observations $X_1, ..., X_n$ drawn from f_θ is

$$\pi(\theta; a_0 + n, (a_0 \mu_{\mathbf{0}} + \sum_{i=1}^{n} X_i)/(a_0 + n)) \qquad (6.27)$$

Moreover,

$$\int_{\Theta} f_\theta(\mathbf{X})\pi(\theta; a, \mu)d\theta = \frac{c(a, \mu)}{c(a + 1, (a\mu + \mathbf{X})/(a + 1))} \qquad (6.28)$$

Suppose the change-time v and the pre- and post-change values of parameters are unknown. And for $t < v$, the parameter θ is the value of θ_0 and θ_1 for $t \geq v$. Following Shiryaev(1978), Lai and Xing(2010) used the Bayesian approach that assumes v to be geometric with parameter p but constrained to be larger than n_0, and θ_0, θ_1 are independent and identically distributed of equation 6.25, and are also independent of v. Since $\pi_n = P\{v \leq n | X_1, ... X_n\}$ is no longer Markovian in the setting of unknown pre- and post-change parameters, Zacks(1991) suggested using dynamic programming to find the stopping rule. Thus, Lai and Xing(2010) introduced a modification of Shiryaev's rule. Let $\mathcal{F}_t$ denote the σ-field generated by $X_1, ..., X_t$. Let

$$\pi_{0,0} = c(a_0, \mu_0) \qquad (6.29)$$

$$\pi_{i,j} = c(a_0 + j - i + 1, a_0\mu_0 + \sum_{t=i}^{j} \mathbf{X}_t)/(a_0 + j - i + 1) \qquad (6.30)$$

For $n_0 < i < n$,

$$P\{v = i | \mathcal{F}_n\} \propto p(1 - p)^{i-1}\frac{\pi_{0,0}^2}{\pi_{1,i-1}\pi_{i,n}}, \qquad (6.31)$$

$$P\{v > n | \mathcal{F}_n\} \propto p(1 - p)^n\frac{\pi_{0,0}}{\pi_{1,n}} \qquad (6.32)$$

It is known that the sum of all the probabilities in equation 6.31 is 1. So we have

$$P(v \leq n | v \geq n^* - k_p, \mathcal{F}_n) = \sum_{i=n-k_p}^{n} \frac{P(v = i | \mathcal{F}_n)}{\sum_{i=n-k_p}^{n} P(v = i | \mathcal{F}_n) + P(v > n | \mathcal{F}_n)}$$

(6.33)

In this way, we can express $P\{n_0 < v \leq n | \mathcal{F}_n\} = \sum_{i=n_0+1}^{n} P\{v = i | \mathcal{F}_n\}$ in terms of $\pi_{i,j}$. Therefore, the Shiryaev's stopping rule for unknown pre- and post-change parameters can be modified in the form of equation 6.9 with replacement of $R_{p,n} = \sum_{i=n_0+1}^{n} \frac{\pi_{0,0}\pi_{1,n}}{(1-p)^{n-i}\pi_{1,i}\pi_{i,n}}$. With suitable choice of k_p, η_p and n_p, the modified stopping rule is proved to be asymptotically optimal as $p \to 0$, which is

$$N = \inf\{n > n_p : P(v \leq n | v \geq n - k_p, \mathcal{F}_n) \geq \eta_p\}$$

(6.34)

Combining with equation 6.33 and equation 6.34, we obtain the modified Shiryaev's rule for detecting a change in θ of the multivariate exponential family when the pre- and post-parameters are unknown:

$$N = \inf\left\{n > n_p : \sum_{i=n-k_p}^{n} \frac{\pi_{0,0}\pi_{1,n}}{(1-p)^{n-i+1}\pi_{1,i}\pi_{i,n}} \geq \gamma_p\right\}$$

(6.35)

Specifically in our research, we introduce Inverse Wishart distribution, which is frequently used as the prior on the covariance matrix parameter Σ of a multivariate normal distribution. Note that the inverse gamma distribution is the conjugate prior for the variance parameter σ^2 of a univariate normal

distribution, while the Inverse Wishart distribution extends conjugacy to the multivariate normal distribution.

Consider $X_1, ... X_n$ be i.i.d observations from an $k-$variate normal distribution $N(\mu, \Sigma)$. If we put an Inverse Wishart prior distribution on the parameter Σ such that

$$\Sigma \sim IW_k(\Psi, n_0) \tag{6.36}$$

After the new observations have been collected, which is represented by $X_i, ..., X_j$. And $n_{i,j}$ denotes the number of observations, which is $j - i + 1$. The posterior distribution of Σ will also be Inverse Wishart distributed as

$$\Sigma | \mathcal{D}_{i,j} \sim IW_k(\Psi + S_{i,j}, n_0 + n_{i,j}) \tag{6.37}$$

where $S_{i,j}$ is the sample sums of squares matrix $S_{i,j} := \sum_{r=i}^{j}(X_r - \bar{X})((X_r - \bar{X})^T$. And $\mathcal{D}_{i,j}$ is a simpler representation of the incoming observations $X_i, ..., X_j$. Given $\mathcal{D}_{i,j}$, the posterior distribution of Σ is written as $\Sigma_{i,j}$. Therefore, n_0 essentially acts as the number of observations we had observed prior to collecting the data, or, alternatively, the number of observations on which our prior sums of squares matrix Ψ is based.

Now it can be characterized as following: let $R_l = \max \{t_{m-1} | I_m = 1, m \leq l\}$, which represents the most recent structural break up to time t_{l-1}. From the above, given $R_l = t_{m-1}$ and $\mathcal{D}_{t_{m-1}, t_l}$, we can get the conditional distribution of Σ_{t_{m-1}, t_l}, which is $IW(\Psi + S_{t_{m-1}, t_l}, n_0 + n_{t_{m-1}, t_l})$. Let $p_{m,l} = P(R_l = t_{m-1} | \mathcal{D}_{t_{m-1}, t_l})$. Then the posterior distribution of Σ_{t_{l-1}, t_l} given $\mathcal{D}_{0, t_l}$ can be expressed as a mixture of Inverse Wishart distributions,

$$\Sigma_{t_{l-1},t_l}|\mathcal{D}_{0,t_l} \sim \sum_{i=1}^{l} p_{i,l} IW(\Psi_0 + S_{t_{i-1},t_l}, n_0 + n_{t_{i-1},t_l}) \tag{6.38}$$

where the mixture weight is recursively calculated by $p_{m,l} = p_{m,l}^* / \sum_{i=1}^{l} p_{i,l}^*$, in which

$$p_{m,l}^* = \begin{cases} p\dfrac{f_{l,l}}{f_{0,0}} & if \ m = l \\[3mm] (1-p)p_{m,l-1}\dfrac{f_{m,l}}{f_{m,l-1}} & if \ m \le l-1 \end{cases} \tag{6.39}$$

the terms $f_{m,l}$ and $f_{0,0}$ above are expressed as following:

$$
\begin{aligned}
f_{m,l} &= \frac{2^{(n_0+n_{t_{m-1},t_l})p/2}\Gamma_k\big((n_0 + n_{t_{m-1},t_l})/2\big)}{|\Psi + S_{t_{m-1},t_l}|^{(n_0+n_{t_{m-1},t_l})/2}}, \\[3mm]
f_{0,0} &= \frac{2^{n_0 p/2}\Gamma_k(n_0/2)}{|\Psi|^{n_0/2}}
\end{aligned}
\tag{6.40}
$$

The proof of the above posterior distributions can be briefly expressed as following: Let $f(.|\mathcal{D}_{(0,t_l]})$ denote the density function of Σ_{t_{l-1},t_l} given $\mathcal{D}_{(0,t_l]}$. We have

$$
\begin{aligned}
f(\Sigma_{t_{l-1},t_l}|\mathcal{D}_{(0,t_l]}) &\sim f(\Sigma_{t_{l-1},t_l}, \mathcal{D}_{(t_{l-1},t_l]}|\mathcal{D}_{(0,t_{l-1}]}) \\[2mm]
&= pf(\Sigma_{t_{l-1},t_l}, \mathcal{D}_{(t_{l-1},t_l]}|\mathcal{D}_{(0,t_{l-1}]}, I_l = 1) + \\[2mm]
&\quad (1-p)f(\Sigma_{t_{l-1},t_l}, \mathcal{D}_{(t_{l-1},t_l]}|\mathcal{D}_{(0,t_{l-1}]}, I_l = 0) \tag{6.41}
\end{aligned}
$$

in which

$$p f(\Sigma_{t_{l-1},t_l}, \mathcal{D}_{(t_{l-1},t_l]}|\mathcal{D}_{(0,t_{l-1}]}, I_l = 1)$$

$$= p_{l,l}^* f(\Sigma_{t_{l-1},t_l}|\mathcal{D}_{(0,t_l]}, I_l = 1)$$

$$= p_{l,l}^* IW(\Psi_0 + S_{t_{l-1},t_l}, n_0 + n_{t_{l-1},t_l}) \tag{6.42}$$

where

$$p_{l,l}^* = p f(\mathcal{D}_{(t_{l-1},t_l]}|\mathcal{D}_{(0,t_{l-1}]}, I_l = 1)$$

$$= p \int f(\mathcal{D}_{(t_{l-1},t_l]}|\Sigma_{t_{l-1},t_l}) g(\Sigma_{t_{l-1},t_l}) d\Sigma_{t_{l-1},t_l}$$

$$= p f_{l,l}/f_{0,0} \tag{6.43}$$

and

$$(1-p) f(\Sigma_{t_{l-1},t_l}, \mathcal{D}_{(t_{l-1},t_l]}|\mathcal{D}_{(0,t_{l-1}]}, I_l = 0)$$

$$= (1-p) \sum_{m=1}^{l-1} P(R_{l-1} = t_{m-1}|\mathcal{D}_{(0,t_{l-1}]}, I_l = 0) \times$$

$$f(\Sigma_{t_{l-1},t_l}, \mathcal{D}_{(t_{l-1},t_l]}|R_{l-1} = t_{m-1}, \mathcal{D}_{(0,t_{l-1}]}, I_l = 0)$$

$$= \sum_{m=1}^{l-1} p_{m,l}^* f(\Sigma_{t_{l-1},t_l}|R_{l-1} = t_{m-1}, \mathcal{D}_{(0,t_l]}, I_l = 0)$$

$$= \sum_{m=1}^{l-1} p_{m,l}^* IW(\Psi_0 + S_{t_{m-1},t_l}, n_0 + n_{t_{m-1},t_l}) \tag{6.44}$$

where

$$\begin{aligned} p_{m,l}^* &= (1-p)p_{m,l-1}f(\mathcal{D}_{(t_{l-1},t_l]}|R_{l-1}=t_{m-1},\mathcal{D}_{(0,t_{l-1}]}, I_l=0) \\ &= (1-p)p_{m,l-1}\frac{f(\mathcal{D}_{(t_{m-1},t_l]}), R_l=t_{m-1}}{f(\mathcal{D}_{(t_{m-1},t_{l-1}]}), R_{l-1}=t_{m-1}} \\ &= (1-p)p_{m,l-1}f_{m,l}/f_{m,l-1} \end{aligned} \qquad (6.45)$$

Combining the above model with the rule in equation 6.35, let $\mathcal{F}_t$ dente the σ-field generated by $I_1, I_2, ..., I_t$. Let m be the time of change point, the associated detection rule for testing the null hypothesis of no change-point, versus the althernative hypothesis of a single change-point to n but not before n_0 is

$$N = \inf\left\{n > n_p : P(v \le n|v \ge n - k_p, \mathcal{F}_n) \ge \gamma_p\right\} \qquad (6.46)$$

and

$$P\left\{v = r|\mathcal{F}_n\right\} \propto p(1-p)^{r-1}f_{1,r-1}f_{r,l}/f_{0,0}^2, \qquad (6.47)$$

$$P\left\{v > n|\mathcal{F}_n\right\} \propto p(1-p)^n f_{1,l}/f_{0,0} \qquad (6.48)$$

we can use equation 6.47 to rewrite equation 6.46 in the form

$$N = \inf\left\{n > n_p : \sum_{i=n-k_p}^{n} \frac{f_{1,i-1}f_{i,n}}{(1-p)^{n-i+1}f_{0,0}f_{1,n}} \ge \gamma_p\right\} \qquad (6.49)$$

6.4 A Sequential Detection Rule for Multiple Structural Breaks

In the above section, we tried to use a mixing Inverse Wishart distribution to represent the posterior distribution of the covariance matrix parameter of a multivariate normal distributed time series when the pre- and post- change of it are both unknown. However, as pointed out by Zacks(2006), it might not be convenient if one has to keep all the observations when doing sequential surveillance. So in this section, we introduce a sequential detection rule for multiple structural breaks which does not have to keep all past observations and the posterior probability can be calculated by explicit recursive formulas. Consider a multivariate normal distribution model $(X_1, ..., X_n) \sim N(\mathbf{0}, \Sigma_t)$, in which $\Sigma \sim IW(\Psi, n_0)$. Suppose the parameter vector Σ_t may come across with changes for $t > 1$, the indicator variables

$$I_t := \mathbf{1}_{\theta_t \neq \theta_{t-1}} \tag{6.50}$$

are independent Bernoulli random variables with $P(I_t = 1) = p$. According to Yao (1984), when a parameter changed at time t, in other words, $I_t = 1$, the changed parameter θ_t is assumed to be derived from π as in equation 6.29. Let $\mathcal{D}_{i,j}$ denote the observations $X_i, ..., X_j$, and K_t refer to the most recent change-time up to t, which is, $K_t = \max\{s \leq t : I_s = 1\}$, and $f(.|.)$ as the

93

conditional densities, such that

$$f(\mathbf{\Sigma_t}|\mathcal{D}_{1,t}) = \sum_{i=1}^{t} p_{i,t} f(\mathbf{\Sigma_t}|\mathcal{D}_{i,t}, K_t = i) \tag{6.51}$$

where $p_{i,t} = P(K_t = i|\mathcal{D}_t)$. In the last section, we have showed that the Inverse Wishart distribution is a popular conjugate family for estimation of covariance matrices, so

$$f(\mathbf{\Sigma_t}|\mathcal{D}_{i,t}, K_t = i) = IW(\Psi + S_{i,t}, n_0 + n_{i,t}) \tag{6.52}$$

where $\overline{\mathbf{S}}_{i,j}$ is the sample sum of square matrix for $j \geq i$. Combining equation 6.51 with equation 6.52, we obtain that

$$f(\mathbf{\Sigma_t}|\mathcal{D}_t) = \sum_{i=1}^{t} p_{i,t} IW(\Psi + S_{i,t}, n_0 + n_{i,t}) \tag{6.53}$$

As we know, $\sum_{i=1}^{t} p_{i,t} = 1$, so the recursive formula can be characterized as

$$p_{i,t} \propto p_{i,t}^* = \begin{cases} pf(\mathbf{X}_t|I_t = 1) & if\ i = t \\ (1-p)p_{i,t-1}f(\mathbf{X}_t|\mathcal{D}_{i,t-1}, K_t) & if\ i \leq t-1 \end{cases} \tag{6.54}$$

when combining $f(\mathbf{X}_t|\mathcal{D}_{i,t-1}, K_t = i) = \int f_{\theta_t}(\mathbf{X})f(\theta_t|\mathcal{D}_{i,t-1}, K_t = i)d\theta_t$ with equation 6.53 and equation 6.53 yields

$$p_{i,t}^* = \begin{cases} p\frac{IW_{0,0}}{IW_{i,t}} & if\ i = t \\ (1-p)p_{i,t-1}\frac{IW_{i,t-1}}{IW_{i,t}} & if\ i \leq t-1 \end{cases} \tag{6.55}$$

94

From the above section, we have

$$IW_{i,j} = \frac{|\Psi + S_{t_{i-1},t_j}|^{(n_0 + n_{t_{i-1},t_j})/2}}{2^{(n_0 + n_{t_{i-1},t_j})p/2}\Gamma_k((n_0 + n_{t_{i-1},t_j})/2)}, \qquad (6.56)$$

$$IW_{0,0} = \frac{|\Psi|^{n_0/2}}{2^{n_0 p/2}\Gamma_k(n_0/2)} \qquad (6.57)$$

Note that $\sum_{t-m}^{t} p_{i,t}$ is the posterior mean number of change-points in the time interval between $t - m$ and t given $\mathcal{D}_{1,t}$.

According to Lai and Xing (2008), for small p, we can use bounded complexity mixture(BCMIX) approximation for the recursive formulas to get $\Sigma_t | \mathcal{D}_{1,t}$. We suppose that there are $M(p)$ components. Most of the recent $m(p)$ $(m(p) < M(p))$ weights $p_{s,l}$(with $l - m(p) < s \leq l$) are kept. Then the posterior density can be obtained as following.

Let $\kappa_{t-1}(p)$ be the set of indices i for which $p_{i,t-1}$ is kept at stage $l - 1$, then we have $\{t-1, ..., t-m(p)\} \subset \kappa_{t-1}(p)$. At stage t, define $p_{i,t}^*$ as in equation 6.55 for $i \in \{t\} \cup \kappa_{t-1}(p)$ and $i_t \notin \{t, ..., t - m(p) + 1\}$ such that

$$p_{i_t,t}^* = min\left\{p_{j,l}^* : j \in \kappa_{t-1}(p), j \leq t - m(p)\right\} \qquad (6.58)$$

here choosing i_t to be the minimizer farthest form t if there are two or more minimizers. Define $\kappa_t(p) = \{t\} \cup (\kappa_{t-1}(p) - \{i_t\})$ and let

$$p_{i,t} = (p_{i,t}^* / \sum_{j \in \kappa_t(p)} p_{j,l}^*), l \in \kappa_t(p) \qquad (6.59)$$

Recall that Lai, Liu and Xing (2009) has proposed a modified Shiryaev's

95

rule to detect a change-point occurred in the time interval between $n - k(p)$ and n, declaring at time n if

$$\sum_{i=n-k(p)}^{n} p_{in} \geq \eta_p \tag{6.60}$$

With the combination of equation 6.60 and equation 6.59, and suitable choice of $k(p)$, η_p and n_p, we obtain the modified Shiryaev's rule for sequentially detecting multiple change points in the time interval between $n - k(p)$ and n when the pre- and post-parameters are unknown:

$$N = \inf \left\{ n > n_p : \sum_{i=n-k(p)}^{n} p_{i,n} \geq \eta_p \right\} \tag{6.61}$$

for suitably chosen η_p.

Chapter 7

Simulation Study and Results

In this section, we present the simulation studies of four detection methods proposed in the last section. All the change point generating process is listed in Chapter 7.1. Then eight different scenarios are considered in Chapter 7.2. Furthermore, we discuss about how the theoretical thresholds are chosen in Chapter 7.3. In Chapter 7.4, we present the simulation results of four detection rules with a multivariate normal distributed time series both of returns of 10 stocks, and a more computational complicated model with 20 stocks, whose covariance matrix is Inverse Wishart distributed.

7.1 Data Generation

For the $p-$stock ($p = 20, 30$) multivariate normal distribution model, we assume that there are p multivariate distributed time series of stock returns over the sample time period $(0, 300)$. We consider an evenly spaced partition in the time period of $(0, 300)$, $0 = t_0 < t_1 < ... < t_n = 300$ and let each time period be 1. So there are totally 300 time periods in our model. In each

scenario, 1000 simulations of the returns of these p stocks are generated to detect the change-point time.

The parameters of structural breaks in this model is generated as following:

Step 1. Pre-specify the change-point time. In our study, to avoid the impact from the previous change-point in the sequential detection methods, we assume that there are more than 80 time periods between adjacent structural breaks. In our data, the number of structural breaks are no more than 3 for each simulation. The details of the change-point time will be given in the scenarios statement.

Step 2. Generate samples of Inverse Wishart distributed matrix to obtain different covariance matrix parameters. Suppose that we have an initialized Inverse Wishart distribution as $IW_p(\Psi, n_0)$. According to Sawyer(2007), we can create samples from Inverse Wishart distribution, which note as $\Sigma^{(1)}, \Sigma^{(2)}, ..., \Sigma^{(n)}$. Afterwards, we take $\Sigma^{(k)}(1 \leq k \leq n)$ as the covariance matrix of multivariate normal distribution, which is

$$X_i, ..., X_j \sim N(0, \Sigma^{(k)}) \tag{7.1}$$

Then use the conjugacy of the Inverse Wishart distribution, note the posterior distribution of the covariance matrix parameter $\Sigma | X_i, ..., X_j$ as $\Sigma_{i,j}$, we have

$$\Sigma_{i,j} \sim IW_p(\Psi + S_{i,j}, n_0 + n_{i,j}) \tag{7.2}$$

Where $S_{i,j}$ is the sample sums of squares matrix $S_{i,j} := \sum_{r=i}^{j}(X_r - \bar{X})((X_r -$

98

$\bar{X})^T$. And $n_{i,j}$ is the number of incoming observations $X_i, ..., X_j$.

7.2 Scenarios Statement

In this section, we consider four scenarios which are: no change-point model, single change-point model, two change points model, and multiple change points model. The detailed description of the scenarios generating process will be given. For each scenario, we run 1000 simulations to evaluate the performance of all the four detection rules.

Scenario 1. No change-point model. The data are generated from a constant parameter model. In other words, we use a constant $\Sigma^{(1)}$ as the covariance matrix of multivariate normal distribution to generate all the observation vectors $(X_0, ..., X_{300}) \sim N(0, \Sigma^{(1)})$. This model will also be regarded as null hypothesis in critical value determined section.

Scenario 2. Single change-point model. The data are generated from a single change-point model with a change-point at $t = 100$. The covariance matrix parameter to generate the multivariate normal distributed vector are: $\Sigma^{(1)}$ for $0 \leq t < 100$, and $\Sigma^{(2)}$ for $100 \leq t \leq 300$. Thus, we have $X_0, ..., X_{99} \sim N(0, \Sigma^{(1)})$ and $X_{100}, ..., X_{300} \sim N(0, \Sigma^{(2)})$.

Scenario 3. Two change points model. The data are generated from a two change-point model with the first change-point at $t = 100$, and the second change-point at $t = 200$. The covariance matrix parameter to generate the multivariate normal distributed vector are: $\Sigma^{(1)}$ for $0 \leq t < 100$, $\Sigma^{(2)}$ for $100 \leq t < 200$ and $\Sigma^{(3)}$ for $200 \leq t \leq 300$. Thus, we have $X_0, ..., X_{99} \sim N(0, \Sigma^{(1)})$, $X_{100}, ..., X_{199} \sim N(0, \Sigma^{(2)})$ and $X_{200}, ..., X_{300} \sim N(0, \Sigma^{(3)})$.

Scenario 4. Multiple change points model. Different the above models, we don't specify the change-point times in advance in this model. Firstly, we have to simulate a Poisson process with a constant rate η during the time period $(0, 300)$, such that $\{\tau_m - \tau_{m-1}\} \sim exp(\eta)$, in which τ_i is the time when a structural break takes place. Then, we follow the similar data generating process as the above. Since we have stated that we take at most 3 change-point in our model, so the covariance matrix parameter to generate the multivariate normal distributed vector are: $\Sigma^{(1)}$ for $0 \leq t < \tau_1$, $\Sigma^{(2)}$ for $\tau_1 \leq t < \tau_2$, $\Sigma^{(3)}$ for $\tau_2 \leq t \leq \tau_3$ and $\Sigma^{(4)}$ for $\tau_3 \leq t \leq \tau_4$. Thus, we have $X_0, ..., X_{\tau_1} \sim N(0, \Sigma^{(1)})$, $X_{\tau_1}, ..., \tau_2 \sim N(0, \Sigma^{(2)})$, $X_{\tau_2}, ..., X_{\tau_3} \sim N(0, \Sigma^{(3)})$ and $X_{\tau_3}, ..., X_{\tau_4} \sim N(0, \Sigma^{(4)})$.

Scenario 5 $\sim$ Scenario 8. To testify the performance of our detection rules in higher dimension, we consider a set of more computational complicated data stream with 20 stocks. All the data generating process and simulation setup will be the same as *Scenario 1 $\sim$ Scenario 4*.

The prior Inverse Wishart distributed matrix parameter Ψ is created from the covariance matrix of 10 stock returns in April 2013, noted as Ψ_{20}, which is listed in Table 7.1.

The prior Inverse Wishart distributed matrix parameter Ψ is created from the covariance matrix of 20 stock returns in April 2013, noted as Ψ_{20}, which is listed in Table 7.2.

Table 7.1: Covariance Matrix of 10 Stock Returns in April 2013

1.29	-3.e-2	1.23e-1	1.85e-2	-2.69e-2	7.e-1	3.76e-1	4.24e-2	1.33e-1	2.52e-2
-3.e-2	3.14e-2	-6.44e-3	2.07e-2	4.26e-2	4.98e-2	3.31e-3	3.55e-2	-1.57e-2	3.18e-2
1.23e-1	-6.44e-3	1.45e-1	2.23e-3	1.04e-2	1.07e-1	-5.13e-2	5.43e-3	-1.22e-3	4.05e-2
1.85e-2	2.07e-2	2.23e-3	5.83e-2	9.49e-2	1.14e-1	3.94e-2	5.69e-2	-4.71e-3	8.13e-3
-2.69e-2	4.26e-2	1.04e-2	9.49e-2	3.01e-1	2.17e-1	3.66e-2	9.65e-2	-4.19e-2	7.54e-3
7.e-1	4.98e-2	1.07e-1	1.14e-1	2.17e-1	1.21	1.61e-1	1.28e-1	3.32e-2	7.25e-2
3.76e-1	3.31e-3	-5.13e-2	3.94e-2	3.66e-2	1.61e-1	3.02e-1	3.98e-2	5.e-2	-2.76e-2
4.24e-2	3.55e-2	5.43e-3	5.69e-2	9.65e-2	1.28e-1	3.98e-2	9.57e-2	-9.68e-3	4.05e-2
1.33e-1	-1.57e-2	-1.22e-3	-4.71e-3	-4.19e-2	3.32e-2	5.e-2	-9.68e-3	5.93e-2	-5.91e-3
2.52e-2	3.18e-2	4.05e-2	8.13e-3	7.54e-3	7.25e-2	-2.76e-2	4.05e-2	-5.91e-3	1.16e-1

Table 7.2: Covariance Matrix of 20 Stock Returns in April 2013

1.3	-3.5e-2	1.2e-1	1.9e-2	-2.6e-2	7.9e-1	3.7e-1	4.2e-2	1.3e-1	2.5e-2	-2e-2	1.3e-1	3.6e-2	2e-2	8e-2	5.9e-2	1.1e-1	2.5e-1	7.2e-2	3.6e-3
-3.5e-2	3.1e-2	-6.4e-3	2.7e-2	4.2e-2	4.9e-2	3.3e-3	3.5e-2	-1.5e-2	3.1e-2	4.8e-3	-1.4e-2	1.1e-2	1.5e-2	5.4e-3	1.8e-3	3e-3	1.2e-2	9.1e-3	2.4e-2
1.2e-1	-6.4e-3	1.5e-1	2.2e-3	1e-2	1e-1	-5.1e-2	5.4e-2	-1.2e-3	4e-2	9.1e-3	6.4e-2	-2.4e-2	4.1e-2	3.1e-2	2.8e-3	2.7e-2	1e-2	2.7e-2	2e-2
1.9e-2	2.1e-2	2.2e-3	5.8e-2	9.4e-2	1.1e-1	3.9e-2	5.6e-2	-4.7e-3	8.1e-3	1.7e-2	3.4e-3	1.6e-2	4.4e-3	2.7e-2	2e-2	2.2e-2	2.9e-2	1.6e-2	3e-2
-2.7e-2	4.3e-2	1e-2	9.4e-2	3e-1	2.1e-1	3.6e-2	9.6e-2	-4.1e-2	7.5e-3	3.9e-2	-3.7e-2	3e-2	7.9e-3	4.5e-2	1.5e-2	3e-2	-1.6e-3	-3e-3	5.5e-2
7.9e-1	5e-2	1.1e-1	1.1e-1	2.1e-1	1.2	1.6e-1	1.2e-1	3.3e-2	7.2e-2	4.9e-2	1.5e-2	-3e-2	2e-2	9.8e-2	4.9e-2	1.3e-1	7.9e-2	9.5e-2	1.1e-1
3.7e-1	3.3e-3	-5.1e-2	3.9e-2	3.6e-2	1.6e-1	3e-1	3.9e-2	5e-2	-2.7e-2	-1.2e-2	2.5e-2	6e-2	-3.6e-2	1.9e-2	3.6e-2	2.1e-2	1.1e-1	1.2e-2	-4e-2
4.2e-2	3.6e-2	5.4e-2	5.6e-2	9.6e-2	1.2e-1	3.9e-2	9.5e-2	-9.6e-3	4e-2	1.7e-2	1e-2	3.7e-2	4.5e-2	2.7e-2	2.1e-2	2.4e-2	6.3e-2	3.3e-2	4.5e-2
1.3e-1	-1.6e-2	-1.2e-3	-4.7e-3	-4.19-2	3.3e-2	5e-2	-9.6e-3	5.9e-2	-5.9e-3	4.8e-2	3.2e-2	8.5e-4	-1.1e-2	2.1e-2	2.7e-2	1.6e-2	4.7e-2	2.3e-2	-1.1e-2
2.5e-2	3.2e-2	4.1e-2	8.1e-3	7.5e-3	7.2e-2	-2.7e-2	4e-2	-5.9e-3	1.1e-1	3.9e-3	-1.2e-2	1.1e-3	7.1e-2	8.3e-3	-1.9e-3	1.2e-2	3.3e-2	3e-2	3.9e-2
-2e-2	4.8e-3	9.1e-3	1.7e-2	3.9e-2	4.9e-2	-1.2e-2	1.7e-2	4.8e-3	3.9e-3	2.5e-2	1.5e-2	-2.7e-3	5.5e-3	2.1e-2	1.1e-2	1.1e-2	-2.9e-3	1.2e-2	1e-2
1.4e-1	-1.4e-2	6.4e-2	3.4e-3	-3.7e-2	1.5e-2	2.5e-2	1e-2	3.2e-2	-1.2e-2	1.5e-2	1.2e-1	-3.6e-3	2e-2	2.5e-2	2.2e-2	1.6e-2	6.7e-2	4.1e-2	-5.7e-3
3.7e-2	1.1e-2	-2.4e-2	1.6e-2	3e-2	-3e-2	6e-2	3.7e-2	8.5e-4	1.1e-3	-2.7e-3	-3.6e-3	6.3e-2	1.6e-2	1.4e-2	1e-2	5.9e-3	2.2e-2	-2.7e-2	3.7e-3
2.1e-2	1.6e-2	4.2e-2	4.4e-3	7.9e-3	2e-2	-3.6e-2	4.5e-2	-1.1e-2	7.1e-2	5.5e-3	2e-2	1.6e-2	8.1e-2	5.2e-3	8.3e-3	1.6e-2	5.1e-2	2.5e-2	2.8e-2
8.1e-2	5.5e-3	3.1e-2	2.7e-2	4.5e-2	9.8e-2	1.9e-2	2.7e-2	2.1e-2	8.3e-3	2.1e-2	2.5e-2	1.4e-2	5.2e-3	6.9e-2	2.4e-2	2.9e-2	4.8e-2	1.8e-2	1.6e-2
6e-2	1.8e-2	2.9e-2	2e-2	1.5e-2	4.9e-2	3.6e-2	2.1e-2	2.7e-2	-1.9e-3	1.1e-2	2.2e-2	1e-2	8.3e-3	2.4e-2	5e-2	2.8e-2	6.3e-2	2.2e-2	8.1e-3
1.1e-1	3.1e-3	2.8e-2	2.2e-2	3e-2	1.3e-1	2.1e-2	2.4e-2	1.6e-2	1.2e-2	1.1e-2	1.6e-2	5.9e-3	1.6e-2	2.9e-2	2.8e-2	3.6e-2	4e-2	2e-2	2e-2
2.6e-1	1.3e-2	1.1e-2	2.9e-2	-1.6e-3	7.9e-2	1.1e-1	6.3e-2	4.7e-2	3.3e-2	-2.9e-3	6.7e-2	2.2e-2	5.1e-2	4.8e-2	6.3e-2	4e-2	3.1e-1	5.1e-2	-8.4e-3
7.2e-2	9.2e-3	2.8e-2	1.6e-2	-3e-3	9.5e-2	1.2e-2	3.3e-2	2.3e-2	3e-2	1.2e-2	4.1e-2	-2.7e-3	2.5e-2	1.8e-2	2.2e-2	2e-2	5.1e-2	4.8e-2	2.1e-2
3.7e-3	2.4e-2	2.1e-2	3e-2	5.5e-2	1.1e-1	-4e-2	4.5e-2	-1.1e-2	3.9e-2	1e-2	-5.7e-3	3.7e-3	2.8e-2	1.6e-2	8.1e-3	2e-2	-8.4e-3	2.1e-2	7.5e-2

7.3 Critical Value Determined Procedure

In our study, the critical value, which is akin to the threshold θ in equation 6.6, c in equation 6.19, γ_p in equation 6.49 and η_p in equation 6.61 are determined by using Monte Carlo simulation. For each detection rule, we calculate the associated statistics under null hypothesis of no change-point based on the data of *Scenario 1* and *Scenario 5* for 1000 times. Then we take the 95%, 97% and 98% quantile of the 1000 statistics as the thresholds in four detection rules. The details of the critical value determined procedure is listed seperately as following.

7.3.1 Statistical Process Control Chart Detection Method

As stated in the above section, *Scenario 1* and *Scenario 5* are generated under the hypothesis of no change-point during the time period $(0, 300)$. Since we do the simulation for 1000 times, the constant covariance matrices are separately $\Sigma_1, \Sigma_2, ..., \Sigma_{1000}$, using which we can calculate the statistics under the null hypothesis denoted as $Z_{300}^{(1)}, Z_{300}^{(2)}, ..., Z_{300}^{(1000)}$. From equation 6.6, we have the recursive formula as following:

$$Z_t^{(i)} = (1 - \alpha)Z_{t-1}^{(i)} + \alpha \frac{1}{p} \sum_{i=1}^{p} \sqrt{e_i(\Sigma_i - I)'(\Sigma_i - I))}, \quad 1 \leq t \leq 300 \qquad (7.3)$$

in which I is the $p \times p$ identity matrix and $e_i(.)$ is the ith eigenvalue of the matrix. And the covariance matrix stay constant over the time period. The smoothing parameter α is related to the quantile we choose in advance. Thus, we have 1000 statistics under no change-point hypothesis. In the following

102

study, we take 95%, 97% and 98% quantile of the distribution of $Z_{300}^{(i)}$ as the thresholds for statistical process control chart detection method, which simply noted as Rule 1.

7.3.2   Generalized Likelihood Ratio (GLR) Detection Rule

Similarly, for 1000 simulations, the constant covariance matrices are separately $\Sigma_1, \Sigma_2, ..., \Sigma_{1000}$. So the observations vector generated with constant Σ_i are written as

$$X_1^{(i)}, ..., X_{300}^{(i)} \sim N(0, \Sigma^{(i)}) \tag{7.4}$$

We denote the statistics under the null hypothesis as $G^{(1)}, G^{(2)}, ..., G^{(1000)}$. From equation 6.15, we have

$$G^{(i)} = \max_{n_0 \leq k \leq T} \left\{ kL(\overline{X}_{1,k}) + (T-k)L(\overline{X}_{k+1,T}) - nL(\overline{X}_{1,T}) \right\}, \quad T = 300 \tag{7.5}$$

where $L(\overline{X}_{i,j})$ is the maximum log likelihood as stated in equation 6.24. Thus, we have 1000 statistics under no change-point hypothesis. In the following study, we take 95%, 97% and 98% quantile of the distribution of $G^{(i)}$ as the thresholds for generalized likelihood ratio (GLR) detection rule, which simply noted as Rule 2.

103

7.3.3 Detection Method Based on an Extension of Shiryaev's Bayesian Single Change-Point Model

Similarly, using the constant covariance matrices $\Sigma_1, \Sigma_2, ..., \Sigma_{1000}$, we generate $X_1^{(i)}, ..., X_{300}^{(i)}$ for 1000 simulations. We denote the statistics under the null hypothesis as $V^{(1)}, V^{(2)}, ..., V^{(1000)}$. From equation 6.49, we have

$$V^{(i)} = \sum_{r=T-k_p}^{T} \frac{f_{1,r-1} f_{r,T}}{(1-p)^{T-r+1} f_{0,0} f_{1,T}}, \quad T = 300 \tag{7.6}$$

where $f_{i,j}$ is referred to equation 6.40. Thus, we have 1000 statistics under no change-point hypothesis. In the following study, we take 95%, 97% and 98% quantile of the distribution of $V^{(i)}$ as the thresholds for detection method based on an extension of Shiryaev's Bayesian single change-point model, which simply noted as Rule 3.

7.3.4 A Sequential Detection Rule for Multiple Structural Break

With the constant covariance matrices $\Sigma_1, \Sigma_2, ..., \Sigma_{1000}$, we generate $X_1^{(i)}, ..., X_{300}^{(i)}$ for 1000 simulations. We denote the statistics under the null hypothesis as $W^{(1)}, W^{(2)}, ..., W^{(1000)}$. From equation 6.61, we have

$$W^{(i)} = \sum_{i=T-k(p)}^{T} p_{i,T}, \quad T = 300 \tag{7.7}$$

where $p_{i,j}$ is given in equation 6.60 and equation 6.59. Thus, we have 1000 statistics under no change-point hypothesis. In the following study, we take 95%, 97% and 98% quantile of the distribution of $W^{(i)}$ as the thresholds for detection method based on a sequential detection rule for multiple structural breaks, which simply noted as Rule 4.

7.4   Simulation Results

In the previous sections, we generate 4 scenarios for $10-$ and $20-$variate normal distributed stock returns for 1000 simulations over time period of $(0, 300)$. In this part, we present the simulation results to compare the performance of the four detection rules in different dimensions.

7.4.1   Statistics Plot of *Scenario 2*

Firstly, we take a general look at the path of the statistics in four detection rules changing over time period with *Scenario 2*, which is the single change-point model.

From Figure 7.1, we can see that all the detection methods work well on detecting the change-point at $t = 100$. The statistics in Rule 1 has a delay around 40, which is much bigger than the ones in Rule 3 and Rule 4, both of which are less than 20. The statistics in Rule 2 has the best performance since the delay of detection point is the least one. However, these results are just from one simulation. We will show the results of 1000 simulations to compare the stable and average performance in next part.
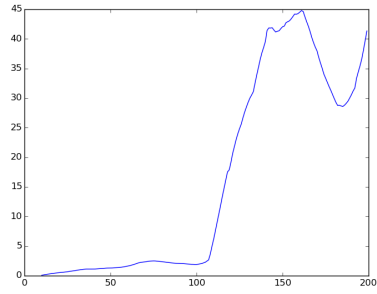
(a)
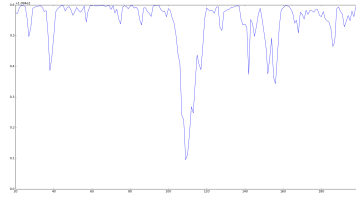
(b)
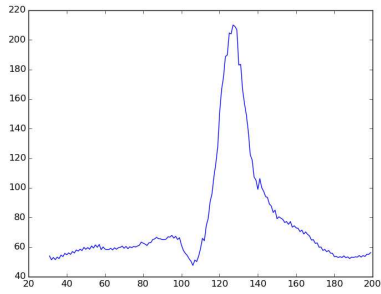
(c)
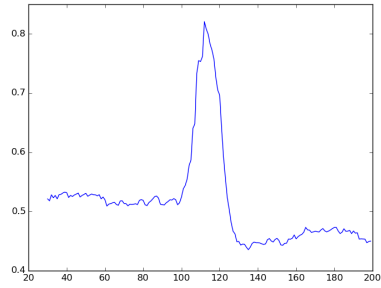
(d)

Figure 7.1: When taking the 95% quantile, the value of statistics of different methods change over time in (a)Rule 1; (b)Rule 2; (c)Rule 3; (d)Rule 4. (From left to right, top to down)

7.4.2 Detailed Comparison of Detection Rules

Table 7.3 compares the performance of the four methods for the single change-point model with change-point time at $t = 100$. We choose $n_p = 10$ and $k_p = 20$ in the detection method of the extension of Shiryaev's Bayesian single change-point model and $k(p) = 20$ in the sequential detection rule for multiple structural breaks. To get rid of the influence of the previous change-point, once a change-point time is detected, the data before this time point is not included for next change-point detection. For each detection method, we list the expected detection delay based on 1000 simulations, as well as the related standard errors are shown in the parentheses.

Table 7.3: Expected Detection Delays for *Scenario 2* (Single Change-point Model)

Quantile	Detection Rule	Change-Point time t=100	Detected Number
95%	Rule 1	35.25(24.75)	687
	Rule 2	48.03(20.79)	916
	Rule 3	5.497(6.597)	976
	Rule 4	12.92(19.53)	913
97%	Rule 1	32.70(17.57)	647
	Rule 2	44.61(20.21)	842
	Rule 3	5.505(6.834)	964
	Rule 4	3.365(8.269)	945
98%	Rule 1	31.61(18.64)	666
	Rule 2	43.62(19.15)	843
	Rule 3	5.638(7.073)	959
	Rule 4	3.087(7.937)	946

In Table 7.3, the third column is the expected detection delay of $t = 100$, and the associated standard error is given in the following parentheses. The expected delay is in the form of $E[\tau - t | \tau \leq 300]$, where τ is the delayed

detected time, and t denotes the true change-point time. Based on 1000 simulations, the total number of change-point that we can detect is listed in the third column, which is noted as detected number. Except for the expected detection delay and the associated standard errors, detected number is also an important measurement indicator to the performance of the detection rule. First, we denote two kinds of detection error. One is the error which is not able to show change-point when it does have, and the other one is falsely showing the change-point when it does not have. For most of time, the second kind of error takes place before the time of the first change-point happens. So the detected number in the right list of the tables equals to the theoretical change-point numbers minus the number of the first kind of error.

From this comparison, we can see that Rule 1 and Rule 2 shows a great value of expected detection delay with very large variance. Rule 3 and Rule 4 have both acceptable expected detection deal and low standard errors.

For the comparison of Rule 2, Rule 3 and Rule 4, we have a more detailed analysis based on *Scenario 3* (two change-point model) and *Scenario 4* (two change-point model). See Table 7.4 and Table 7.5.

In Table 7.4, the change-point times are $t_1 = 100$ and $t_2 = 200$, the theoretical change-point number for each change-point is 1000. It is clearly to see that detection Rule 2 has a worse expected detection delay and the standard error comparing with the other two detection rules. And the expected detection delay and the standard errors of Rule 3 and Rule 4 are almost the same. But Rule 4 has a little bit more detected number than Rule 3. For a comprehensive analysis of this table, the result of Rule 4 is the best.

108

Table 7.4: Expected Detection Delays for *Scenario 3* (Two Change-point Model)

	Rule 2	Rule 3	Rule 4
Quantile = 95%			
Expected Delay at t_1	33.68(18.04)	3.478(7.350)	4.572(11.51)
Detected Number at t_1	344	913	896
Expected Delay at t_2	44.54(20.29)	25.44(10.97)	8.799(15.20)
Detected Number at t_2	865	996	932
Quantile = 97%			
Expected Delay at t_1	32.31(16.46)	3.412(6.685)	2.782(6.597)
Detected Number at t_1	334	890	947
Expected Delay at t_2	43.49(21.00)	24.09(11.56)	4.319(8.415)
Detected Number at t_2	877	998	978
Quantile = 98%			
Expected Delay at t_1	38.85(18.10)	2.792(5.851)	2.560(7.179)
Detected Number at t_1	280	889	957
Expected Delay at t_2	46.17(20.09)	26.59(10.06)	4.277(9.524)
Detected Number at t_2	860	1000	988

Table 7.5 shows the result for the multiple change-point based on *Scenario 4*. We can get almost the same result as in Table 7.4. For this multiple change-point model, the correct detected number of change-point is acceptable for all the three change-point, which shows that our detection methods have stable performance even in the case of detecting multi change-point.

To more simply and intuitively compare for Rule 2, Rule 3 and Rule 4, we also present some figures to show the performance of different detection methods with different scenarios.

Figure 7.2 plots the detected time of Rule 2, Rule 3 and Rule 4 for *Scenario 2*. We can see that the green dots which represent the result of Rule 4 are closest to the line of change-point time at $t = 100$, and the red dots are farthest among the three rules. It is the same as the results shown in Table

Table 7.5: Expected Detection Delays for *Scenario 4* (Three Change-point Model)

	Rule 2	Rule 3	Rule 4
Quantile = 95%			
Expected Delay at τ_1	29.38(13.12)	5.485(7.610)	4.392(9.576)
Detected Number at τ_1	424	989	935
Expected Delay at τ_2	31.45(14.58)	20.68(10.40)	5.253(10.64)
Detected Number at τ_2	921	997	960
Expected Delayat τ_3	34.49(14.06)	25.54(8.417)	7.899(10.99)
Detected Number at τ_3	925	1000	964
Quantile = 97%			
Expected Delay at τ_1	27.28(13.88)	3.324(4.690)	4.266(10.10)
Detected Number at τ_1	466	889	938
Expected Delay at τ_2	31.57(14.67)	19.60(10.87)	5.348(9.627)
Detected Number at τ_2	931	999	976
Expected Delay at τ_3	34.40(14.77)	25.22(9.014)	6.173(10.74)
Detected Number at τ_3	912	1000	986
Quantile = 98%			
Expected Delay at τ_1	29.22(13.60)	2.403(4.316)	4.372(8.891)
Detected Number at τ_1	435	892	949
Expected Delay at τ_2	33.01(13.72)	19.91(9.464)	6.250(11.00)
Detected Number at τ_2	926	999	963
Expected Delay at τ_3	33.50(14.58)	24.66(8.301)	6.694(11.37)
Detected Number at τ_3	922	999	984

Figure 7.2: Plots of Detected Change Points for *Scenario 2*: single change–point model with change-point at $t = 100$ in Rule2, Rule3 and Rule4

7.3 that Rule 3 and Rule 4 over performance Rule 2 in this scenario.

Figure 7.3 Plots the Detection Results for *Scenario 3*. And we can get the same result with Table 7.4. In this figure, the green dots which represent Rule 4 and the blue dots which represent Rule 3 show very acceptable results for both change time at $t_1 = 100$ and $t_2 = 200$. The red dots for Rule 2 is still the farthest line among three detection rules.

Figure 7.4 plots the detection results for *Scenario 4*. And we can get the same result with Table 7.5. In this figure, the green dots which represent Rule 4 and the blue dots which represent Rule 3 show very acceptable results for both change time at $t_1 = \tau_1$, $t_2 = \tau_2$ and $t_3 = \tau_3$. The red dots for Rule 2 is still the farthest line among three detection rules.
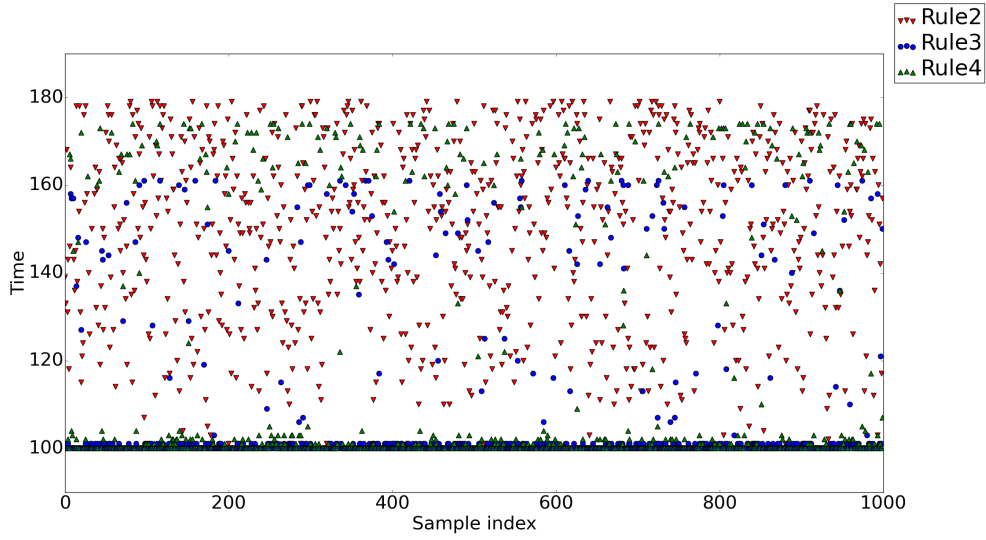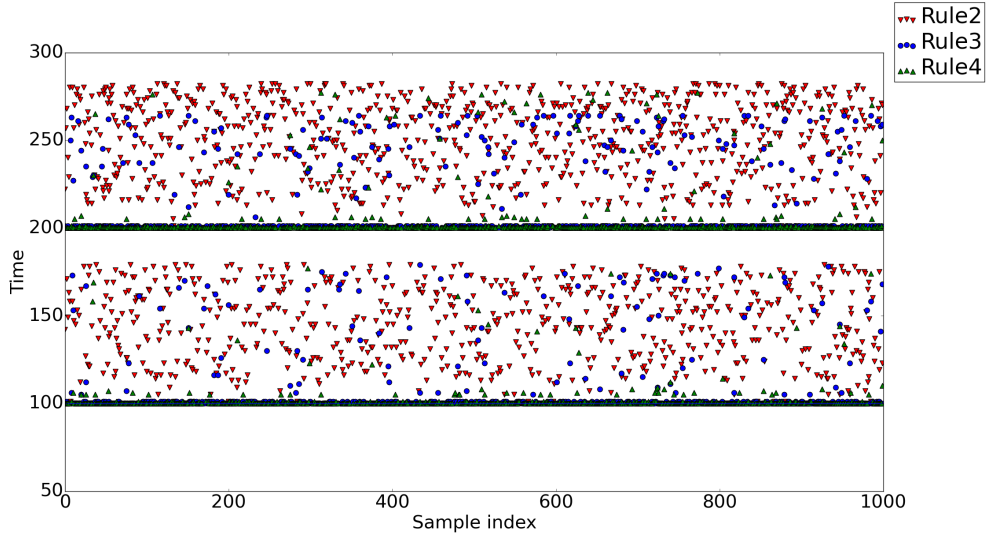
Figure 7.3: Plots of Detected Change Points for *Scenario 3*: two change-point model with change-point at $t_1 = 100$ and $t_2 = 200$ in Rule2, Rule3 and Rule4
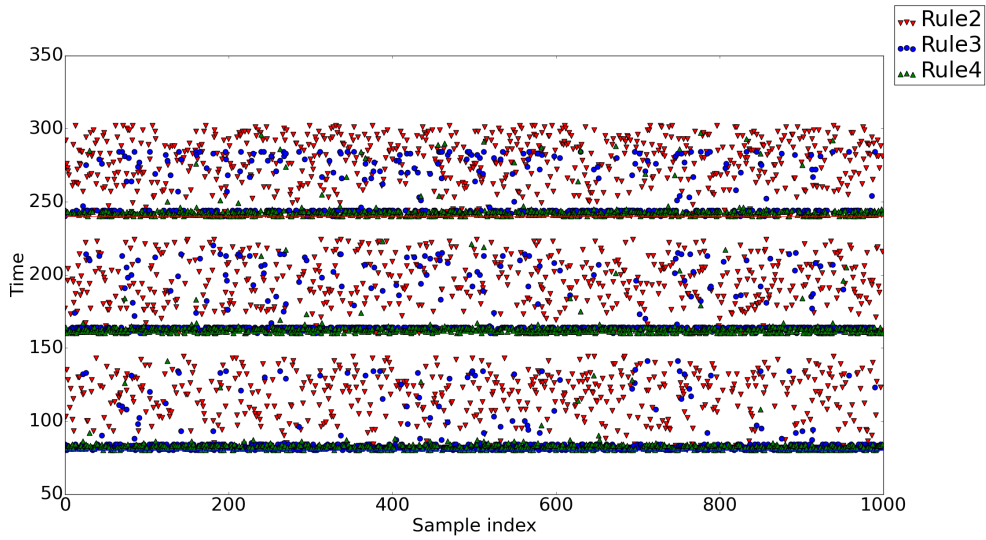

Figure 7.4: Plots of Detected Change Points for *Scenario 4*: three change-point model with change-points at $t_1 = \tau_1$, $t_2 = \tau_2$ and $t_3 = \tau_3$ in Rule2, Rule3 and Rule4

112

7.4.3   More Complicated Simulation Study with 20 Stocks

Furthermore, we move on to a more computational complicated model with a higher dimension. Similar with the case of 10 stock returns, we generate 20-variate normal distributed vectors over time period (0,300) for 1000 simulations. We choose $p = 0.01$, $n_p = 10$ and $k_p = 20$ in Rule 3 and $k(p) = 30$ in Rule 4. To ensure sequentially detecting the change-point, we get rid of the data before the detection of next change-point once the previous change-point is found. All the simulations are considered based on the threshold for 95%, 97% and 98% quantile in the distribution of statistics calculated with null hypothesis.

We also use some tables to compare the performance of the four detection methods. Here we consider two extra indexes: the detected rate and the accurate rate. Firstly, the detected rate is the one that represents the number of simulations that we can detect the change-point over 1000 simulations. In our study, we pre-specify all the time points we detect before $t = 150$ as the first change-point detected number, while all the change-point detect after $t = 150$ belong to the second change-point detected number. From this assumption, if there is no change-point detected before $t = 150$, then there is zero detected number for the first change-point, so is detected rate. Secondly, we define the accurate rate as the number of correct detection result which arrives exactly at the change-point time over the total simulation number 1000.

Table 7.6: False Alarm Rate and Expected Detection Delays for *Scenario 6* (Single Change-point Model)

Quantile	Detection	Expected Delay	Detected Number	False Alarm Rate	Accurate Rate
95%	Rule 1	30.11(16.53)	622	37.8%	25.6%
	Rule 2	42.80(18.02)	958	4.2%	15.6%
	Rule 3	0.0(0.0)	965	3.5%	96.5%
	Rule 4	0.809(0.259)	851	14.9%	63.9%
97%	Rule 1	25.01(16.09)	682	31.8%	35.0%
	Rule 2	47.70(19.52)	852	14.8%	4.0%
	Rule 3	0.0(0.0)	951	4.9%	95.1%
	Rule 4	1.210(0.522)	873	12.7%	63.0%
98%	Rule 1	28.16(17.20)	696	30.4%	31.9%
	Rule 2	48.71(19.18)	847	15.3%	3.4%
	Rule 3	0.0(0.0)	959	4.1%	95.9%
	Rule 4	1.154(0.536)	896	10.4%	66.2%

From Table 7.6, Table 7.7 and Table 7.8, which are the result for *Scenario 6*, *Scenario 7* and *Scenario 8*, we can see that both Rule 3 and Rule 4 give acceptable pre-detection rate and low expected delay. The results of Rule 2 either have very high pre-detection rate or high expected detection delay. There is no missed change-point detected in each simulation of all the detection rules. In fact, in the case of both $10-$variate data and $20-$variate data, Rule 3 and Rule 4 have better performance than that in Rule 2.

To more simply and intuitively compare for Rule 2, Rule 3 and Rule 4, we also present some figures to show the performance of different detection methods with different scenarios.

Figure 7.5, Figure 7.6 and Figure 7.7 plots the detection results for *Scenario 6*, *Scenario 7* and *Scenario 8*. And we can get the same result with the low dimension figures. In these figures, the green dots which represent

114

Table 7.7: False Alarm Rate and Expected Detection Delays for *Scenario 7* (Two Change-point Model)

	Rule 2	Rule 3	Rule 4
Quantile = 95%			
Expected Delay at t_1	38.37(18.03)	0.0(0.0)	0.791(0.187)
Detected Number at t_1	299	997	969
False alarm Rate att_1	70.1%	0.3%	3.1%
Accurate Rate at t_1	3.9%	96.5%	82.7%
Expected Delay at t_2	46.32(19.63)	2.404(1.733)	0.754(0.182)
Detected Number at t_2	865	998	991
False Alarm Rate at t_2	13.5%	0.2%	0.9%
Accurate Rate at t_2	4.9%	87.8%	85.3%
Quantile = 97%			
Expected Delay at t_1	32.20(17.42)	0.0(0.0)	0.769(0.185)
Detected Number at t_1	342	999	975
False Alarm Rate at t_1	65.8%	0.1%	2.5%
Accurate Rate at t_1	8.7%	95.5%	83.6%
Expected Delay at t_2	43.46(20.46)	0.904(0.0)	0.821(0.230)
Detected Number at t_2	849	1000	997
False alarm rate at t_2	15.1%	0%	0.3%
Accurate Rate at t_2	8.5%	88.7%	84.5%
Quantile = 98%			
Expected Delay at t_1	37.93(17.83)	0.0(0.0)	0.765(0.186)
Detected Number at t_1	272	999	977
False alarm rate at t_2	72.8%	0.1%	2.3%
Accurate Rate at t_1	3.80%	97.20%	84.10%
Expected Delay at t_2	45.99(28.94)	0.4(0.0)	0.785(0.220)
Detected Number at t_2	851	1000	989
False alarm rate at t_2	14.9%	0%	1.1%
Accurate Rate at t_2	3.90%	90.00%	85.50%

Table 7.8: False Alarm rate and Expected Detection Delays for *Scenario 8* (three change-point model)

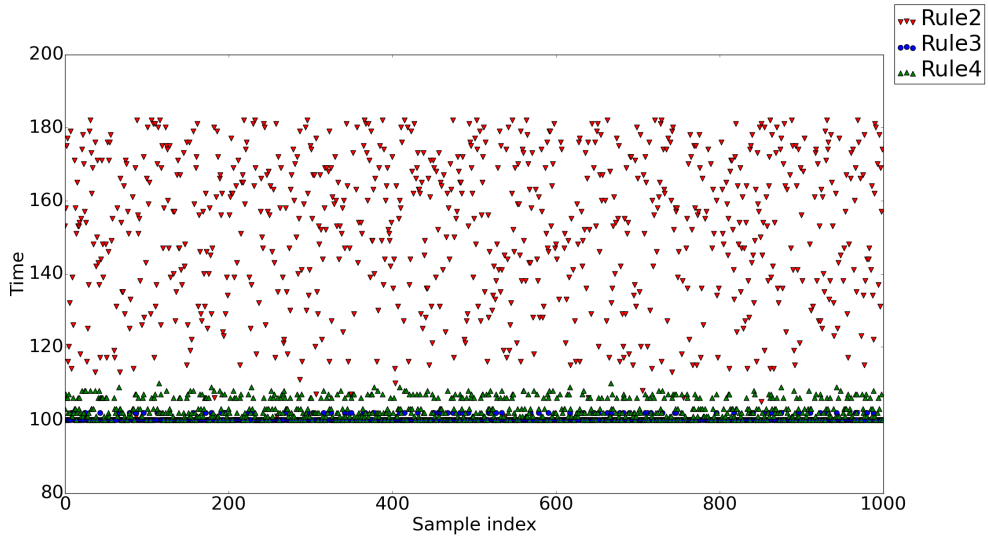|	Rule 2	Rule 3	Rule 4
Quantile = 95%			
Expected Delay at τ_1	32.82(13.60)	0.0(0.0)	1.534(0.749)
Detected Number at τ_1	371	999	963
False alarm rate at τ_2	62.9%	0.1%	3.7%
Accurate Rate at τ_1	4.2%	94.7%	64.8%
Expected Delay at τ_2	35.25(14.54)	1.512(0.0)	1.928(0.821)
Detected Number at τ_2	913	1000	983
False alarm rate at τ_2	8.7%	0%	1.7%
Accurate Rate at τ_2	4.3%	89.2%	62.4%
Expected Delay at τ_3	35.60(14.16)	4.246(0.0)	1.739(0.823)
Detected Number at τ_3	913	1000	978
False alarm rate at τ_2	8.7%	0%	2.2%
Accurate Rate at τ_3	4.5%	80.7%	63.5%
Quantile = 97%			
Expected Delay at τ_1	31.80(13.95)	0.0(0.0)	1.629(0.805)
Detected Number at τ_1	375	1000	973
False alarm rate at τ_2	62.5%	0%	2.7%
Accurate Rate at τ_1	5%	95.4%	63.7%
Expected Delay at τ_2	34.41(14.33)	0.0(0.0)	1.397(0.671)
Detected Number at τ_2	915	1000	976
False alarm rate at τ_2	8.5%	0%	2.4%
Accurate Rate at τ_2	4.1%	90.2%	67.5%
Expected Delay at τ_3	37.38(14.71)	0.0(0.0)	1.493(0.681)
Detected Number at τ_3	907	1000	990
False alarm rate at τ_2	9.3%	0%	1%
Accurate Rate at τ_3	3.10%	83.7%	66.5%
Quantile = 98%			
Expected Delay at τ_1	29.22(13.60)	0.0(0.0)	1.548(0.581)
Detected Number at τ_1	435	999	964
False alarm rate at τ_2	56.5%	0.1%	3.6%
Accurate Rate at τ_1	9.1%	95.8%	63.8%
Expected Delay at τ_2	33.21(13.76)	0.0(0.0)	1.603(0.756)
Detected Number at τ_2	926	999	993
False alarm rate at τ_2	7.4%	0.1%	0.7%
Accurate Rate at τ_2	9.1%	91.4%	65.7%
Expected Delay at τ_3	33.50(14.58)	0.0(0.0)	1.525(0.730)
Detected Number at τ_3	922	1000	983
False alarm rate at τ_2	7.8%	0%	1.7%
Accurate Rate at τ_3	10.3%	83.5%	66.8%

Figure 7.5: Plots of Detected Change Points for *Scenario 6*: single change–point model with change-point at $t = 100$ in Rule2, Rule3 and Rule4
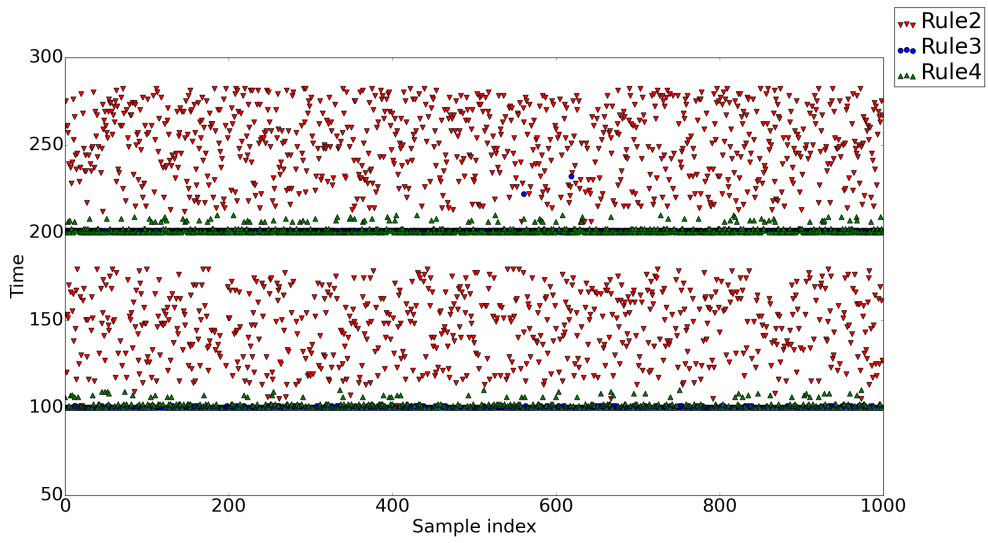

Figure 7.6: Plots of Detected Change Points for *Scenario 3*: two change-point model with change-point at $t_1 = 100$ and $t_2 = 200$ in Rule2, Rule3 and Rule4
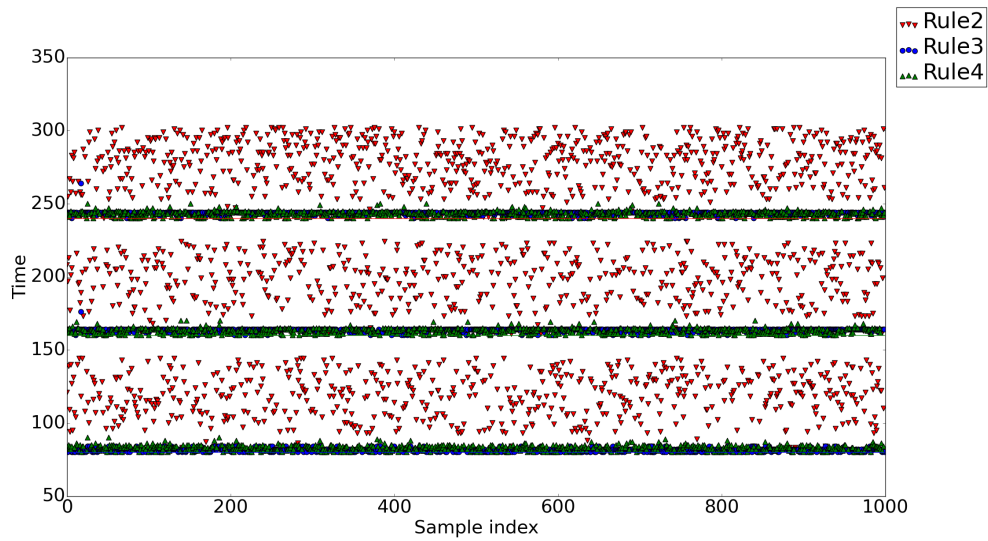
117

Figure 7.7: Plots of Detected Change Points for *Scenario 8*: three change-point model with change-points at $t_1 = \tau_1$, $t_2 = \tau_2$ and $t_3 = \tau_3$ in Rule2, Rule3 and Rule4

Rule 4 and the blue dots which represent Rule 3 show very acceptable results at change-point times. The red dots for Rule 2 is still the farthest line among three detection rules.

Chapter 8

Real Data Analysis

In this section, we use exactly the same empirical data set as in Chapter 4.2.3. Thus, we assume $223-$variate normal distributed observations $X_1, ..., X_T$, $T = 1786$ over time period from April 2013 to September 2013. Specifically, X_i take the value of both returns and VPTs. We present the hyper-parameter selection in Chapter 8.1.1 and threshold determined in Chapter 8.1.2. In Chapter 8.2, we present the performance of different detection rules and summarize our results in comparison with the concluding remarks from Part 1.

8.1　Parameter Initialization

8.1.1　Prior Distribution Determined using EM Algorithm

In our study, we use Expectation Maximization (EM) algorithm to estimate the prior parameters in the model. Let $X_i = (x_{i,1}, x_{i,2}, ..., x_{i,223})$ be the

vector of 223 stock returns at time $t = i$. As stated above, the covariance matrix of these multivariate Gaussian vector follows Inverse Wishart distribution: $\Sigma \sim IW_{223}(\Psi, n_0)$. By the independence assumptions, the log likelihood function for (Ψ, n_0) is given by

$$
\begin{aligned}
l(\Psi, n_0 | X_s, ..., X_t) &= \log f(X_s, ..., X_t | \Psi, n_0) \\
&= \log \prod_{i=s}^{t} \int \frac{|\Psi|^{n_0/2} \Gamma_{223}(\frac{n_0+n_i}{2})}{|\Psi + S_{s,t}|^{\frac{n_0+n_i}{2}} \Gamma_{223}(\frac{n_0}{2})} \\
&= \sum_{i=s}^{t} [\frac{n_0}{2} \log |\Psi| - \frac{n_0 + n_i}{2} \log |\Psi + S_i| + \log \frac{\Gamma_{223}(\frac{n_0+n_i}{2})}{\Gamma_{223}(\frac{n_0}{2})}]
\end{aligned}
\tag{8.1}
$$

Where $S_i = X_i^T X_i$ and the studies sample size is n_0. We now compute the expectation step of the EM algorithm. Let $\Delta_i = \Sigma_i^{-1}$ and $\Theta = \Psi^{-1}$, then we equivalently have that

$$
\Delta_i \sim W_p(\Theta, n_0) S_i | \Delta_i \sim W_p(\Delta_i^{-1}, n_i)
\tag{8.2}
$$

Since the conjugacy holds for both Inverse Wishart and Wishart distribution, then we have

$$
\Delta_i | S_i \sim W_p((\Theta^{-1} + S_i)^{-1}, n_0 + n_i)
\tag{8.3}
$$

With knowing that the expectation of the Wishart distribution,

$$
E[\Delta_i | S_i] = (n_0 + n_i)(\Theta^{-1} + S_i)^{-1}
\tag{8.4}
$$

The maximization step, in which the log likelihood $l(\Theta | \Delta_s, ..., \Delta t)$ is maximized and the estimate $\hat{\Theta} = \frac{1}{(t-s+1)n_0} \sum_{i=s}^{t} \Delta_i$. Then, let $\hat{\Theta}_{(m)}$ be the current

120

estimate of Θ. This yields the updating estimate as

$$\hat{\Theta}_{(m+1)} = \frac{1}{(t-s+1)n_0} \sum_{i=s}^{t} (n_0 + n_i)(\hat{\Theta}_{(m)}^{-1} + S_i)^{-1} \tag{8.5}$$

In this way, we can get the $\hat{\Psi}$, which is the estimate calculated by repeated iteration of equation 8.5.

8.1.2   Threshold Determined from Simulation Data

From the description of the real data set in Chapter 4.2.3, we have totally $T = 1786$ time points of 30 minutes returns and VPTs of 223 stocks from April 2013 to September 2013. To determine the thresholds for different detection rules in the real data analysis, we first partition the data set into two parts. For the first part, we take use of the first 500 observations $X_1, ..., X_{500}$ and follow EM algorithm in the above section to estimate the initial values of the prior parameters. In the second part, we decide the thresholds with the application of simulated data which are created from the prior distribution. Afterwards, we testify the performance of detection rules using the second part of the data.

Follow the data generating procedures in Seciton 7.1, we first obtain an estimation prior of Ψ_{500} based on the observations during time period $t = 1, ..., 500$. Then, we generate 1000 groups of simulated time series whose covariance matrix follow Inverse Wishart distribution with prior parameter Ψ_{500}. With 1000 simulation data, we have a 95%, 97% and 98% quantile critical value for all four detection rules, namely Z_i in equation 7.3, G_i in equation 7.5, V_i in equation 7.6 and W_i in equation 7.7. Then we detect the real data

using Rule 1, Rule 2, Rule 3 and Rule 4 for the time period $t = 1, ..., 501$, and check if there is a change-point at $t = 501$.

Then we repeat the process for $t = 502, ..., 1786$ following the same steps as in $t = 501$. If there is a change-point occurred at time t, the data before t is supposed to be deleted from the data set for the following detection, to get rid of the influence from the previous structural break.

8.2 Concluding Remarks and Future Work

8.2.1 Empirical Study Results

In this section, we present plots of how change-point detection statistics in each rule under null hypothesis change through time. Then, we summarize the structural breaks found in the market by using our detection rules and compare the detection results with the ones in Chapter 4.2.3. Moreover, we consider about the shortcomings in our methods and discuss about the future plan.

Firstly, in order to have a general knowledge about how the statistics in each rule change over the whole time period, we plot a set of figures under the assumption of no change-point. As stated in Chapter 4.2.3, we introduce two kinds of financial dynamics as indicators of the macro market, which are stock returns and VPTs. From the simulation study in Chapter 7, we conclude that Rule 1 has the worst performance compared to the other three detection rules. Thus, we only consider Rule 2, Rule 3 and Rule 4 in empirical study. Among the following figures, Figure 8.1 is the application of Rule 2 with stock returns,

Figure 8.2 is the application of Rule 3 with stock returns, Figure 8.3 is the application of Rule 4 with stock returns, Figure 8.4 is the application of Rule 2 with stock VPTs, Figure 8.5 is the application of Rule 3 with stock VPTs, and Figure 8.6 is the application of Rule 4 with stock VPTs.

From Figure 8.1, Figure 8.2 and Figure 8.2, we can see that there are approximately three change-point in each detection. For Rule 2, there are peaks at 476, 1169 and 1752. For Rule 3, the peaks are at 782, 1222 and 1691. For Rule 4, the peaks are at 824 and 1176. Considering that there are expected delays or false alarm occurred in the detection, the second change point detected in Rule 2, Rule 3 and in the same time period. Also, the third change-point found in Rule 2 and Rule 3 are in the same time zone. In the case of VPTs dataset, we make a comparison of Figure 8.1 with Figure 8.4, Figure 8.2 with Figure 8.5, Figure 8.3 with Figure 8.6. It is quite obvious that the change-point detection results in stock returns are almost the same as the ones with VPTs.

Furthermore, recall the plots of maximum eigenvalue evolution in Chapter 4.2.3 of Part 1, we make a comparison of the change of maximum eigenvalue through time, see the red lines in Figure 8.7, with the change of detection statistics in Rule 2, see the blue lines in Figure 8.7. Both of them are based on stock returns dataset. It shows that in both plots, there are significant changes around 480 and 1190. This result may show that there are connections in the results calculated in Part 1 with the one in this part of thesis. Therefore, we move on to compare these two sets of result in details.

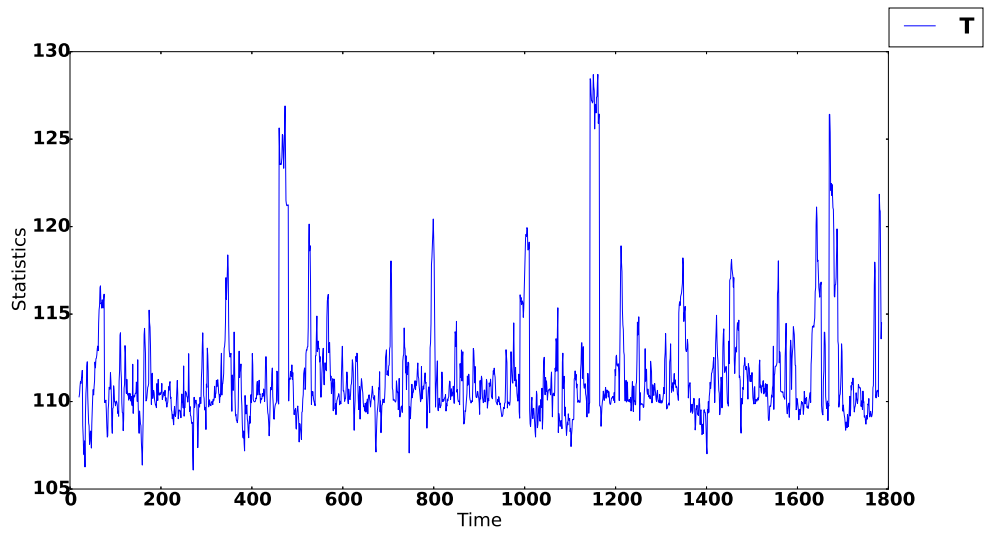Then, we list all the structural breaks detected in Table 8.1 by applying

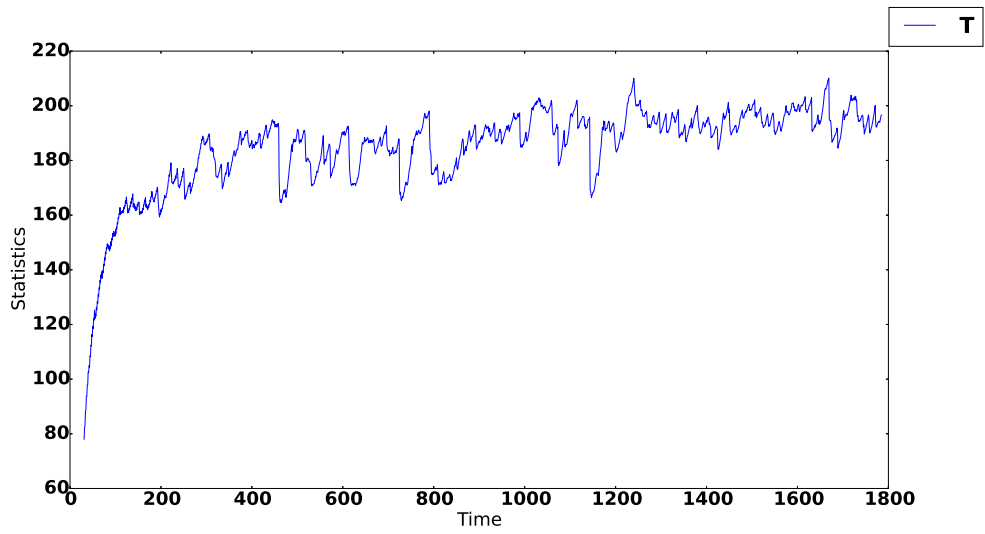Figure 8.1: Plots of Statistics in Rule 2 for S&P 500 Intraday Stock Returns

Figure 8.2: Plots of Statistics in Rule 3 for S&P 500 Intraday Stock Returns
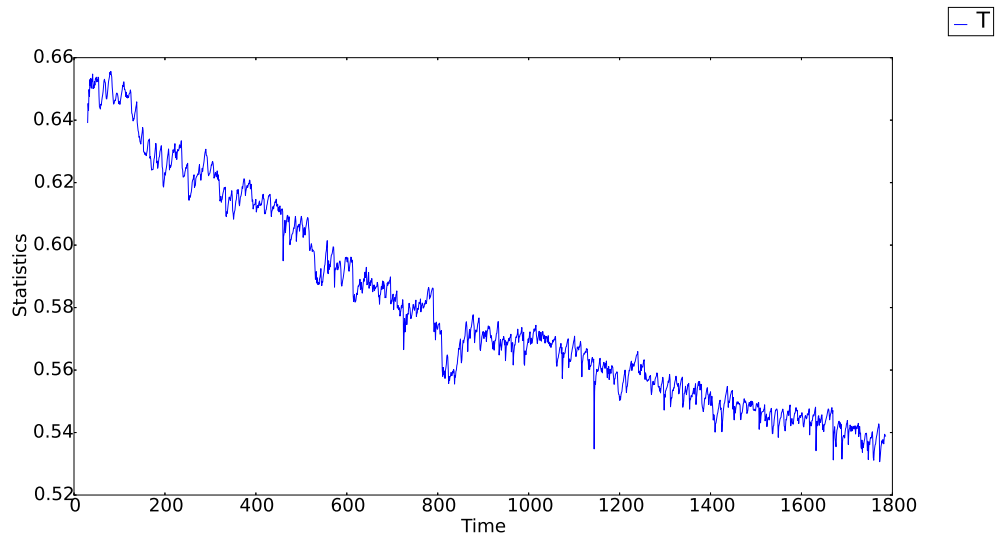
124

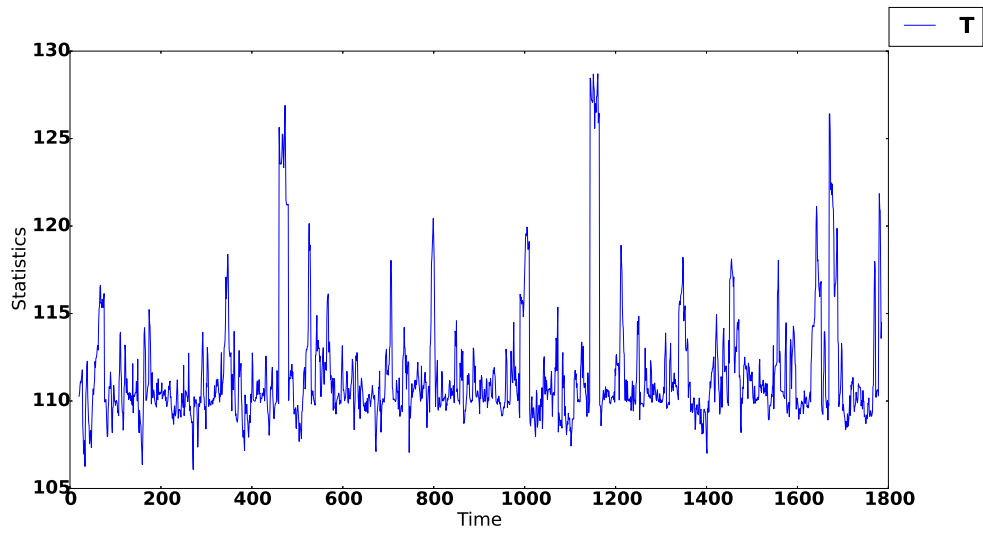Figure 8.3: Plots of Statistics in Rule 4 for S&P 500 Intraday Stock Returns

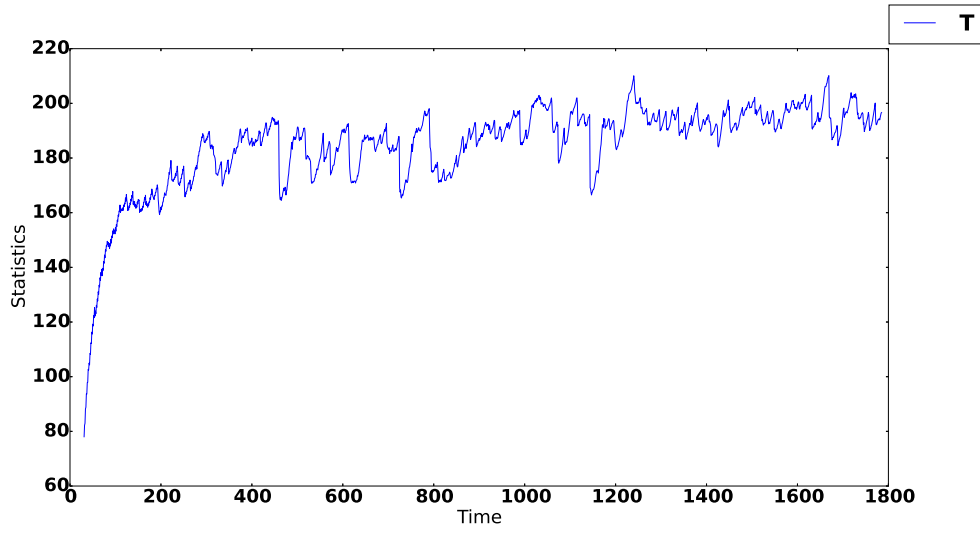Figure 8.4: Plots of Statistics in Rule 2 for S&P 500 Intraday Stock VPTs

125

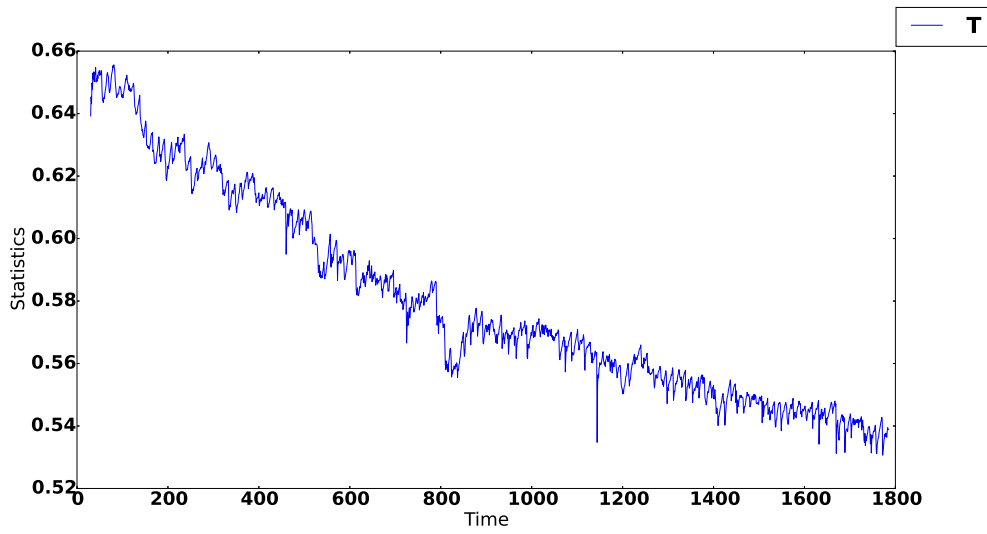Figure 8.5: Plots of Statistics in Rule 3 for S&P 500 Intraday Stock VPTs

Figure 8.6: Plots of Statistics in Rule 4 for S&P 500 Intraday Stock VPTs

126
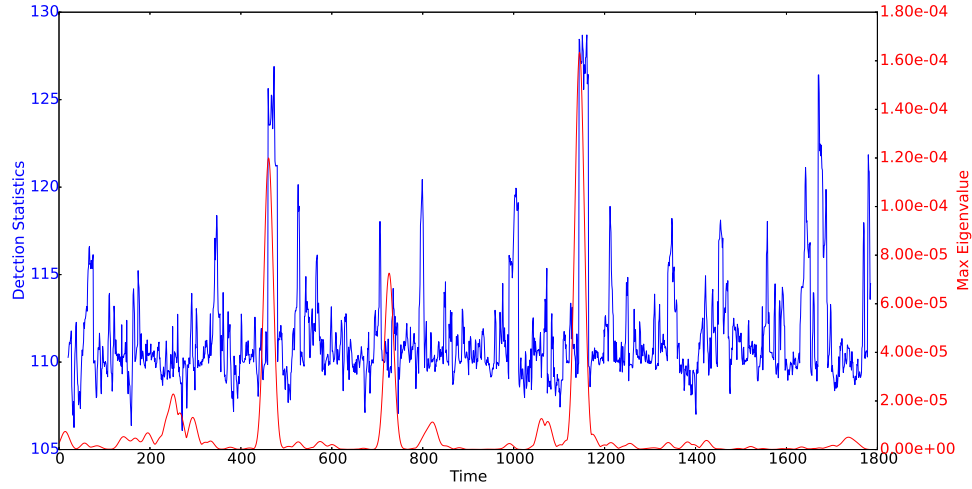
Figure 8.7: Comparison of Detection Statistics and Maximum Eigenvalues Calculated with Stock Returns from April 2013 to September 2013

determined thresholds with 98% quantile in Chapter 8.1.2. Also, we include the time points of interest we obtained in the first part of thesis. In fact, the time points of interest are those maximum eigenvalues which are the local optimums through time. From the Table 8.1, we can see that some of the time points of interest found with our proposed estimation methods are almost simultaneously with the change-point detected with our proposed change-point detection rules.

127

Table 8.1: Structural Breaks Detected in Part 2 and Time Points of Interest Found in Part 1

		Structural Breaks
Rule 2	Stock Returns	489, 1155, 1689
	VPTs	481, 811, 1150, 1708
Rule 3	Stock Returns	430, 782, 1064, 1281, 1655
	VPTs	446, 781, 1139, 1285, 1677
Rule 4	Stock Returns	461, 720, 809, 1148
	VPTs	479, 723, 838, 1160
Maximum Eigenvalue	Stock Returns	455, 729, 1179
	VPTs	469, 750, 1093, 1121

8.2.2 Summary and Future Plan

In this part of thesis, we make use of four kinds of change-point detection rules to find out the structural breaks in multivariate normal distributed data stream over a time period. Then, we carry out a detailed theoretical and empirical comparison of our proposed approaches to detect the change-point with high dimensional dataset. Moreover, we compare the structural breaks we find using these detection rules with the ones in Part one.

However, there are still limitations in this part of empirical study. Considering the complexity of computation, we only take a half year of intraday stock returns and VPTs in real data analysis. And it is difficult to make a conclusion about the structural breaks for a whole market in such a relatively short period of time. In future study, one may want to take a longer time for detection the structural break for the whole financial market and has to overcome the computation complexity in the study. What's more, since we have found that some of the time points of interest found in Part 1 are almost

simultaneously with the change-point detected in Part 2, one may get interested in discovering the relationship between these time points with significant events in the real world.

Bibliography

[1] Admati, Anat R., and Paul Pfleiderer. "A theory of intraday patterns: Volume and price variability." Review of Financial studies 1.1 (1988): 3-40.

[2] Alarcon Heras, Oriol, and Albert Gordi Margalef. "Partial correlation: Network analysis." (2014).

[3] Anufriev, Mikhail, and Valentyn Panchenko. "Connecting the dots: Econometric methods for uncovering networks with an application to the Australian financial institutions." Journal of Banking & Finance 61 (2015): S241-S255.

[4] Barigozzi, Matteo, and Christian T. Brownlees. "Nets: Network estimation for time series." Available at SSRN 2249909 (2014).

[5] Basu, Sumanta, and George Michailidis. "Regularized estimation in sparse high-dimensional time series models." The Annals of Statistics 43.4 (2015): 1535-1567.

[6] Bhadra, Anindya, and Bani K. Mallick. "Joint High?Dimensional Bayesian Variable and Covariance Selection with an Application to eQTL Analysis." Biometrics 69.2 (2013): 447-457.

[7] Biais, Bruno, Larry Glosten, and Chester Spatt. "Market microstructure: A survey of microfoundations, empirical results, and policy implications." Journal of Financial Markets 8.2 (2005): 217-264.

[8] Bilgrau, Anders Ellern, et al. "Estimating a common covariance matrix for network meta-analysis of gene expression datasets in diffuse large B-cell lymphoma." arXiv preprint arXiv:1503.07990 (2015).

[9] Billio, Monica, et al. Network connectivity and systematic risk. Working Paper, 2015.

[10] Blume, Lawrence, David Easley, and Maureen O'hara. "Market statistics and technical analysis: The role of volume." The Journal of Finance 49.1 (1994): 153-181.

[11] Bouriga, Mathilde, and Olivier Fron. "Estimation of covariance matrices based on hierarchical inverse-Wishart priors." Journal of Statistical Planning and Inference 143.4 (2013): 795-808.

[12] Brida, Juan Gabriel, David Matesanz, and Maria Nela Seijas. "Network analysis of returns and volume trading in stock markets: The Euro Stoxx case." Physica A: Statistical Mechanics and its Applications 444 (2016): 751-764.

[13] Brownlees, Christian, Christina Hans, and Eulalia Nualart. Bank Credit Risk Networks: Evidence from the Eurozone Crisis. Technical Report, 2014.

[14] Cheifetz, Nicolas, et al. "A CUSUM approach for online change-point detection on curve sequences." ESANN. 2012.

[15] Cheng, Jie, et al. "High-dimensional mixed graphical models." arXiv preprint arXiv:1304.2810 (2013).

[16] Comerton-Forde, Carole, and James Rydge. "Market integrity and surveillance effort." Journal of Financial Services Research 29.2 (2006): 149-172.

[17] Cumming, Douglas, and Sofia Johan. "Global market surveillance." American Law and Economics Review 10.2 (2008): 454-506.

[18] Davis, Richard A., Pengfei Zang, and Tian Zheng. "Sparse vector autoregressive modeling." Journal of Computational and Graphical Statistics just-accepted (2015): 1-53.

[19] Donoho, David L., and Iain M. Johnstone. "Adapting to unknown smoothness via wavelet shrinkage." Journal of the american statistical association 90.432 (1995): 1200-1224.

[20] Friedman, Jerome, et al. "Pathwise coordinate optimization." The Annals of Applied Statistics 1.2 (2007): 302-332.

[21] Epskamp, Sacha, et al. "qgraph: Network visualizations of relationships in psychometric data." Journal of Statistical Software 48.4 (2012): 1-18.

[22] Fan, Jianqing, Chunming Zhang, and Jian Zhang. "Generalized likelihood ratio statistics and Wilks phenomenon." Annals of Statistics (2001): 153-193.

[23] Friedman, Jerome, Trevor Hastie, and Rob Tibshirani. "Regularization paths for generalized linear models via coordinate descent." Journal of statistical software 33.1 (2010): 1.

[24] Frisen, Marianne, ed. Financial surveillance. Vol. 71. John Wiley & Sons, 2008.

[25] Garthoff, Robert, Iryna Okhrin, and Wolfgang Schmid. "Control charts for multivariate nonlinear time series." REVSTATStatistical Journal 13.2 (2015): 131-144.

[26] Giudici, Paolo, and Alessandro Spelta. "Graphical network models for international financial flows." Journal of Business & Economic Statistics 34.1 (2016): 128-138.

[27] Golosnoy, Vasyl, and Wolfgang Schmid. "Statistical process control in asset management." Applied Quantitative Finance. Springer Berlin Heidelberg, 2009. 399-416.

[28] Guo, Jian, et al. "Joint structure estimation for categorical Markov networks." Unpublished manuscript 3.5.2 (2010): 6.

[29] Han, Fang, and Han Liu. "Transition Matrix Estimation in High Dimensional Time Series." ICML (2). 2013.

[30] Han, Fang, Huanran Lu, and Han Liu. "A direct estimation of high dimensional stationary vector autoregressions." Journal of Machine Learning Research 16 (2015): 3115-3150.

[31] Kazemilari, Mansooreh, and Maman Abdurachman Djauhari. "Correlation network analysis for multi-dimensional data in stocks market." Physica A: Statistical Mechanics and its Applications 429 (2015): 62-75.

[32] Kenett DY, Tumminello M, Madi A, Gur-Gershgoren G, Mantegna RN, Ben-Jacob E (2010) Dominating Clasp of the Financial Sector Revealed by Partial Correlation Analysis of the Stock Market. PLoS ONE 5(12): e15032.

[33] Kenett, Dror Y., et al. "Dominating clasp of the financial sector revealed by partial correlation analysis of the stock market." PloS one 5.12 (2010): e15032.

[34] Kenett, Dror Y., et al. "Quantifying meta-correlations in financial markets." EPL (Europhysics Letters) 99.3 (2012): 38001.

[35] Kolar, Mladen, et al. "Estimating time-varying networks." The Annals of Applied Statistics (2010): 94-123.

[36] Kooij, Anita J. Prediction accuracy and stability of regression with optimal scaling transformations. Child & Family Studies and Data Theory (AGP-D), Department of Education and Child Studies, Faculty of Social and Behavioural Sciences, Leiden University, 2007.

[37] Lai, T. L., Liu, H., and Xing, H. (2005) Autoregressive models with piecewise constant volatility and regression parameters. Statistica Sinica. 15, 279-301.

[38] Lai, T. L., and Xing, H. (2006) Structural change as an alternative to long memory in financial time series. Advances in Econometrics (H. Carter and T. Fomby, eds.), Vol 20 (Econometric Analysis of Economic and Financial Time Series), 209-228

[39] Lai, T.L., Liu, T. and Xing, H. (2009). A Bayesian approach to sequential surveillance in exponential families. Communications in Statistics, Theory and Methods. 38, 2958-2968.

[40] Lai, T.L. and Xing H. (2010). Sequential change-point detection when the pre- and post-change parameters are unknown. Sequential Analysis. 29, 162-175.

[41] Lai, T.L., Xing, H. and Chen, Z. (2011). Mean-variance portfolio optimization when means and covariances of asset returns are unknown. The Annals of Applied Statistics. 5, 798-823.

[42] Lai, T.L. and Xing, H. (2011). A simple Bayesian approach to multiple change-points. Statistica Sinica. 21, 539-569.

[43] Lavielle, Marc, and Gilles Teyssiere. "Detection of multiple change-points in multivariate time series." Lithuanian Mathematical Journal 46.3 (2006): 287-306.

133

Leng, Chenlei, and Cheng Yong Tang. "Sparse matrix graphical models." Journal of the American Statistical Association 107.499 (2012): 1187-1200.

[44] MarketCapitalizations.com "Changes in S&P 500 components during the past ten years", 25 March 2015.

[45] Medeiros, Marcelo C., and Eduardo F. Mendes. "?1-regularization of high-dimensional time-series models with non-Gaussian and heteroskedastic errors." Journal of Econometrics 191.1 (2016): 255-271.

[46] Mei, Yajun. "Sequential change-point detection when unknown parameters are present in the pre-change distribution." The Annals of Statistics (2006): 92-122.

[47] Ngai, E. W. T., et al. "The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature." Decision Support Systems 50.3 (2011): 559-569.

[48] O'Hara, Maureen. "Liquidity and financial market stability." National Bank of Belgium Working Paper 55 (2004).

[49] Patel, Alpaben K., and Jyoti Divecha. "Modified exponentially weighted moving average (EWMA) control chart for an analytical process data." Journal of Chemical Engineering and Materials Science 2.1 (2011): 12-20.

[50] Peng, Jie, et al. "Partial correlation estimation by joint sparse regression models." Journal of the American Statistical Association 104.486 (2009).

[51] Pepelyshev, Andrey, and Aleksey S. Polunchenko. "Real-time financial surveillance via quickest change-point detection methods." arXiv preprint arXiv:1509.01570 (2015).

[52] Peralta, Gustavo, and Abalfazl Zareei. "A network approach to portfolio selection." Journal of Empirical Finance (2016).

[53] Polansky, S., M. Kulczak, and L. Fitzpatrick. "NASD Market Surveillance Assessment and Recommendations. Final Report." Achievement of Market Friendly Initiatives and Results Program (AMIR 2.0 Program) (2004).

[54] Sawyer, S. "Wishart Distributions and Inverse-Wishart Sampling." Washington University, St. Louis (2007).

[55] Schmid, Wolfgang, and Dobromir Tzotchev. "Statistical Surveillance of the Parameters of a One-Factor Cox?Ingersoll?Ross Model." Sequential Analysis 23.3 (2004): 379-412.

[56] Shojaie, Ali, and George Michailidis. "Discovering graphical Granger causality using the truncating lasso penalty." Bioinformatics 26.18 (2010): i517-i523.

[57] Siegmund, David, and E. S. Venkatraman. "Using the generalized likelihood ratio statistic for sequential detection of a change-point." The Annals of Statistics (1995): 255-271.

[58] Snoussi, Hichem, and Ali Mohammad-Djafari. "Estimation of structured Gaussian mixtures: the inverse EM algorithm." IEEE transactions on signal processing 55.7 (2007): 3185-3191.

[59] Verdier, Ghislain, Nadine Hilgert, and Jean-Pierre Vila. "Optimality of cusum rule approximations in change-point detection problems: application to nonlinear statespace systems." IEEE Transactions on Information Theory 54.11 (2008): 5102-5112.

[60] Woodall, William H., and Douglas C. Montgomery. "Research issues and ideas in statistical process control." Journal of Quality Technology 31.4 (1999): 376.

[61] Worthington, Andrew, and Helen Higgs. "Transmission of equity returns and volatility in Asian developed and emerging markets: a multivariate GARCH analysis." International Journal of Finance & Economics 9.1 (2004): 71-80.

[62] Wu, Sen, Mengjiao Tuo, and Deying Xiong. "Network structure detection and analysis of Shanghai stock market." Journal of Industrial Engineering and Management 8.2 (2015): 383.

[63] Xie, Yao, and David Siegmund. "Sequential multi-sensor change-point detection." Information Theory and Applications Workshop (ITA), 2013. IEEE, 2013.