이학석사학위논문

Multiple change point detection in high dimensional multivariate series data using trace of the beta matrix

AUGUST 2021

서울대학교 대학원

통계학과

조 영 현

Multiple change point detection in high dimensional multivariate series data using trace of the beta matrix

지도교수 임 요 한

이 논문을 이학석사 학위논문으로 제출함

2021년 7월

서울대학교 대학원

통계학과

조 영 현

조 영 현의 이학석사 학위 논문을 인준함

2021년 7월

위 원 장 **이 재 용**

부위원장 **임 요 한**

위 원 **원 중 호**

Abstract

This research proposes a method to test and estimate change points in the covariance structure of high-dimensional multivariate series data. The method uses the trace of the beta matrix, known as the Pillai's statistics, at each time point. This paper extends the asymptotic normality of the Pillai's statistics for testing the equality of two covariance matrices when both n (sample size) and p (dimension) increase at the same rate, introduced in Koo et al. (2019). The method computes the Pillai's statistics and its p-value for each time point. Then the method tests the existence of single change point by combining individual p-values by using Cauchy combination test by Liu and Xie (2020) and estimates the change point as the point whose statistic is the greatest. To test and estimate multiple change points, the idea of the wild binary segmentation by Fryzlewicz (2014) is applied. The above procedure is applied to each segmented series until no significant change point exists. This paper numerically provides the size and power of the proposed method. Finally, the proposed method is applied to find abnormal behavior in the investment of a private equity.

Keywords: beta matrix, high-dimensional covariance matrix, multiple change points, Pillai's statistic, private equity, random matrix theory

Student Number: 2019-20319

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Detecting change points in covariance structures in multivariate time series data is of a great importance. However, it has received much less attention than mean change point detection problem. Furthermore, proper method to deal with high-dimensional data does not exist even though such data is frequent in recent days.

Therefore, the main aim of this paper is to propose a method to detect change points in the covariance structure of high-dimensional multivariate series data. To do so, the trace of the beta matrix, known as the Pillai's statistics and its asymptotic normality under increasing both n (sample size) and p (dimension) are used.

The rest of the paper is organized as follows. Chapter 2 presents review on existing methods. In chapter 3, main method is introduced. Chapter 4 provides numerical study on simulations to evaluate the size and power of the method with comparison to those of existing method. In chapter 5, real data analysis is conducted to detect potential abnormal behavior on managing private equity fund by using the method.

Chapter 2

Review

In this section, the existing methods of detecting structural change points via detecting the changes in covariance matrix are reviewed.

Suppose $\mathbf{y}_1, \ldots, \mathbf{y}_T$ are a sequence of p-dimensional vector sample with mean $\mathbf{0}$ and convariance matrix Σ_i. Consider the null hypothesis

$$\mathbf{H}_0 : \Sigma_1 = \Sigma_2 = \cdots = \Sigma_T$$

To find the change point(CP) k with $[\epsilon \cdot T] + 1 \leq k \leq [(1-\epsilon)T]$ for some $\epsilon > 0$, consider the following alternative hypothesis

$$\mathbf{H}_1 : \Sigma_1 = \cdots = \Sigma_{k-1} = \Sigma_k \neq \Sigma_{k+1} = \cdots = \Sigma_T$$

for some k

To revisit a likelihood ratio test for detecting change points in the covariace matrix, assume for a moment that each $\mathbf{y}_j$ is from multivariate normal with expectation μ_j and covariace matrix Σ_j. A likelihood ratio test is introduced in Zamba and Hawkins (2009) and Anderson (2009). The likelihood ratio test statistic for detecting a change point at k is

$$\Lambda_k = \frac{|S_{1,k}|^{\frac{k-1}{2}} |S_{k+1,T}|^{\frac{T-k-1}{2}}}{|S_{1,T}|^{\frac{T-1}{2}}},$$

where $S_{i,j}$ denotes sample covariance matrix of data from y_i to y_j and $|\cdot|$ is the matrix determinant operator.

It is known that $-2\log(\Lambda_k)$ follows a chi-square distribution for large T and for large $T-k$. It should be noted that this method assumes that the location of the change point is given to be at k. To remedy this restriction, the method can be alleviated to allow an unknown change point location by considering $\max_{[\epsilon \cdot n]+1 \leq k \leq [(1-\epsilon)n]} \Lambda_k$.

On top of the method with parametric assumptions, there are non-parametric methods. The method proposed by Aue et al. (2009) neglects assumption on multivariate normal, and instead assume finite second moment condition with homogeneous mean, that is, $\mathbf{E}(\mathbf{y}_i) = \mu$ and $\mathbf{E}(|\mathbf{y}_i|^2) < \infty$ for all $i = 1, \cdots, T$. Aue et al. (2009) constructed CUSUM statistic using the estimator

$$S_k = \frac{1}{\sqrt{T}}\left(\sum_{j=1}^{k} \text{vech}(\tilde{\mathbf{y}}_j \tilde{\mathbf{y}}_j^T) - \frac{k}{n}\sum_{j=1}^{n} \text{vech}(\tilde{\mathbf{y}}_j \tilde{\mathbf{y}}_j^T)\right), \qquad k = 1, \ldots, n,$$

where $\text{vech}(\cdot)$ is the half vectorization operator that stacks the columns below the diagonal of a symmetric $p \times p$ matrix as a vector with dimension $\mathcal{D} = p(p+1)/2$, and $\tilde{\mathbf{y}}_j = \mathbf{y}_j - \bar{\mathbf{y}}_T$ with $\bar{\mathbf{y}}_T = \frac{1}{n}\sum_{j=1}^{T} \mathbf{y}_j$.

The test statistic is defined by

$$\Omega_n = \frac{1}{n}\sum_{k=1}^{T} S_k \hat{\Sigma}_n^{-1} S_k,$$

where $\hat{\Sigma}_n$ is a Bartlett estimator of

$$\Sigma = \sum_{j} \text{Cov}(\text{vech}(\mathbf{y}_0 \mathbf{y}_0^T), \text{vech}(\mathbf{y}_j \mathbf{y}_j^T))$$

Then Ω_n satisfies asymptotic distribution that

$$\Omega_n \xrightarrow{D} \Omega(\mathcal{D}) = \sum_{l}^{\mathcal{D}} \int_0^1 B_l^2(t)dt,$$

where $\{B_l(t) : t \in [0,1], 1 \leq l \leq \mathcal{D}\}$ are independent Brownian bridges and $\xrightarrow{D}$ indicates convergence in distribution.

If the computed test statistic exceeds the predetermined critical value, $\mathbf{H}_0$ is rejected, and one can estimate the change point by $k = [\hat{\theta}_n T]$, where

$$\hat{\theta}_n = \frac{1}{T} \underset{1 \leq k \leq T}{\operatorname{argmax}} S_k^T \hat{\Sigma}_T^{-1} S_k.$$

On the other hand, Barnett and Onnela (2016) proposed simulation based change point detection without distributional assumption. Let $\mathbf{Y}^{(b)}$ be one of the bootstrap resamples from $\mathbf{Y} = [\mathbf{y}_1, \cdots, \mathbf{y}_T]$, where the resampling is repeated for $b = 1, \ldots, B$. For each k such that $[\epsilon \cdot T] + 1 \leq k \leq [(1-\epsilon)T]$, $d^{(b)}(k) = |S_{1,k}^{(b)} - S_{k+1,T}\|_F$ is calculated, and in turn, calculate so called z-score by

$$z^{(b)}(k) = \frac{d^{(b)}(k) - \hat{\mu}_0(k)}{\sqrt{\hat{\sigma}_0^2(k)}},$$

where $\hat{\mu}_0(k) = \frac{1}{B} \sum_{b=1}^{B} d^{(b)}(k)$ and $\hat{\sigma}_0^2(k) = \frac{1}{B-1} \sum_{b=1}^{B} (d^{(b)}(k) - \hat{\mu}_0(k))^2$.

Let $Z^{(b)} = \max_k \{z^{(b)}(k)\}$ for each bootstrap sample b, and derive $z(k)$ and the test statistic $Z = \max_k \{z(k)\}$ from the observed data. The corresponding p-value from bootstrapping is

$$p - \text{value} = \frac{1}{B} |\{b : Z^{(b)} \geq Z\}|,$$

where $|\cdot|$ is the cardinality of the set. If the p-value is significant, then $\mathbf{H}_0$ is rejected and the change point is estimated by $\underset{k}{\operatorname{argmax}}\, z(k)$.

Note that if $\mathbf{y}_j$ are all independent and come from the same distribution under $\mathbf{H}_0$, bootstrapping method is appropriate, while bootstrapping gives rise to bias in approximation of the null distribution if $\mathbf{y}_j$ are autocorrelated, which is frequent in many time series applications. To handle the autocorrelation in the resampling, one can use the sieve bootstrap instead.

However, each of these method has own innated drawback. First, the likelihood ratio test requires multivariate normal assumption so this method is not applicable to data that do not follow normal distribution. These days, there are a lot of high-dimensional data which deviate from normal distribution and so such restriction is not desirable. The method proposed by Aue et al. (2009) requires half-vectorization and this can cause a huge obstacle in analysis. In particular, under high-dimensional data with large p, the dimension of $\hat{\Sigma}_n$ is $\mathcal{D} = p(p+1)/2$, thus considerably large amount of data is required to calculate $\hat{\Sigma}_n^{-1}$. On the other hand, the simulation based method proposed by Barnett and Onnela (2016) can survive in such high-dimensional setting. However, this method has innated drawback in that bootstrapping method requires huge amount of computation resources.

Since the goal of this paper is to detect structural change points in high-dimensional setting, this paper compares the simulation method by Barnett and Onnela (2016) with method introduced in the next chapter.

Chapter 3

Proposed method

3.1 Single point detection

This section proposes the test $\mathcal{T}_{max}$ to detect the structural change point. This method is an extension of invariant test for equality of two large scale covariance matrix proposed by Koo et al. (2019)

For a sequence of p-dimensional vector valued data $\mathbf{y}_1, \ldots, \mathbf{y}_T$ with mean $\mathbf{0}$ and convariance matrix Σ_i, let $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_T]^\top$. For each $k = [\epsilon \cdot T] + 1, \ldots, [(1-\epsilon) \cdot T]$, with predetermined window length wl, let

$$Y_{k,1} = [\mathbf{y}_{a_k}, \ldots, \mathbf{y}_k]^\top \text{ and } Y_{k,2} = [\mathbf{y}_{k+1}, \ldots, \mathbf{y}_{b_k}]^\top,$$

where $a_k = \max\{k - wl + 1, 1\}$, $b_k = \min\{k + wl, T\}$

Then compute the Beta-matrix $B_k = n_{k,1}S_{k,1}(n_{k,1}S_{k,1} + n_{k,2}S_{k,2})^{-1}$, where $n_{k,1}$, $n_{k,2}$ are the numbers of columns of $Y_{k,1}$, $Y_{k,2}$, respectively and $S_{k,1}$, $S_{k,2}$ are the covariance matrices of $Y_{k,1}$, $Y_{k,2}$, respectively.

Statistic $\mathcal{K}_k$ is defined by

$$\mathcal{K}_k = \frac{\sum \lambda_i^{\mathbf{B}_k} - pl_k - \mu_k}{\sigma_k},$$

where $\lambda_i^{\mathbf{B}_k}$ denotes the i-th smallest eigenvalue of $\mathbf{B}_k$, $l_k = \frac{h^2\delta_{y_2>1}+y_2^2\delta_{y_2<1}}{y_2(y_1+y_2)}$, $\mu_k = -\frac{\Delta_1 h^2 y_1^2 y_2^2}{(y_1+y_2)^4} + \frac{\Delta_2 h^2 y_1^2 y_2^2}{(y_1+y_2)^4}$, $\sigma_k = \frac{2h^2 y_1^2 y_2^2}{(y_1+y_2)^4} + (\Delta_1 y_1 + \Delta_2 y_2)\frac{h^4 y_1^2 y_2^2}{(y_1+y_2)^6}$, and $y_1 = \frac{p}{n_{k,1}}$, $y_2 = \frac{p}{n_{k,2}}$ and Δ_i stands for skewness.

Intuitively, $\mathcal{K}_k$ is the sum of eigenvalues of the beta matrix. Koo et al. (2019) proved asymptotic normality under suitable conditions on moments and dimensionality,

$$\mathcal{K}_k \xrightarrow{D} N(0,1)$$

The proposed statistic for detecting CP is based on the series of Beta-matrices as

$$k^* = \underset{[\epsilon\cdot n]+1,\ldots,[(1-\epsilon)\cdot n]}{\operatorname{argmax}} \mathcal{K}_k, \text{ and } T_{max} = \mathcal{K}_{k^*}$$

For the observed value t_{max} of $\mathcal{T}_{max}$, the p-value for testing the existence of the CP is bounded by

$$P(\mathcal{T}_{max} > t_{max}) \leq \sum_{k=[\epsilon\cdot T]+1}^{[(1-\epsilon)\cdot T]} P(\mathcal{K}_k > t_{max}).$$

However, Bonferroni type bound tends to be too conservative so Cauchy combination test by Liu and Xie (2020) is adopted as:

$$T = \sum_k \omega_k \tan\{(0.5 - P(\mathcal{K}_k > t_{max})\pi\}$$

The p-value is approximated by

$$p - \text{value} = \frac{1}{2} - (\arctan t_0)/\pi,$$

where t_0 is the observed value of T.

3.2 Multiple points detection

The proposed method can be easily extended to multiple detection problem followed the idea in Fryzlewicz (2014). If an estimated change point from the entire data is statistically significant, split the data into two sub-data with the estimated point. Detection is applied on both sub-data, which possibly result in further splits. The recursion on a given segment continues until there is no significant change point.

Chapter 4

Numerical study

4.1 Single point detection

This section numerically provides powers and empirical sizes of the proposed test, while comparing to the existing method: Barnett and Onnela (2016)

In this study, the sample sizes under consideration are $T \in \{200, 500, 800, 1000\}$ and a change point is prefixed at $k^* = \frac{T}{2} + 1$ for each T. 1000 data sets are generated as:

- Case1: $\mathbf{y}_j \sim \mathbf{MVN}(\mathbf{0}, \Sigma_{0.5})$ for all $1 \leq j \leq T$

- Case2: $\begin{cases} \mathbf{y}_j \sim \mathbf{MVN}(\mathbf{0}, \Sigma_{0.4}) & \text{for all } 1 \leq j \leq k^* \\ \mathbf{y}_j \sim \mathbf{MVN}(\mathbf{0}, \Sigma_{0.6}) & \text{for all } k^* + 1 \leq j \leq T \end{cases}$

Powers and sizes are evaluated by counting the number of rejection under significant level 0.05.

Note that the fourth moments of y_i in Case1, 2 are known by $\Delta_1 = 0$, $\Delta_2 = 0$. Note also the choice of ϵ and window length must be deliberate. Too small value of ϵ leads to imbalance between $n_{k,1}$ and $n_{k,2}$. It is empirically

observed that this imbalance can lead to malfunction of the method. Moreover, small value of wl gives small $n_{k,1}$ and $n_{k,2}$. Computation of the inverse of beta matrix cannot work under too small $n_{k,1}$ and $n_{k,2}$. This paper recommends to choose ϵ which is enough to discard data as many as p and enough wl to guarantee the computation of the inverse of beta matrix. Therefore, ϵ is chosen that $[\epsilon \cdot T] \simeq wl$ and $[(1-\epsilon) \cdot T] \simeq wl$, while wl is at least larger than p. In this study, wl is varied by $wl \in \{30, 50\}$. On the other hand, 10,000 resampling is done for the simulation method.

Table 4.1 Empirical size and power of each test

	Size			Power		
Data	Proposed		Simulation	Proposed		Simulation
size	wl=30	wl=50		wl=30	wl=50	
200	0.129	0.067	0.035	0.870	0.998	0.274
500	0.062	0.051	0.050	0.705	0.991	0.646
800	0.050	0.055	0.051	0.699	0.983	0.840
1000	0.063	0.045	0.048	0.630	0.983	0.911

Table 4.1 summarizes the results. For the empirical sizes, the proposed method shows appropriate size. It is notable that the proposed method with $wl = 50$ outperforms the simulation test in every setting in terms of power. In addition, longer window length gives enhanced power. To evaluate the method, the locations of significant estimated change points should also be considered. The estimated locations are shown on Figure 4.1.

Figure 4.1 provides histograms of the locations of significant estimated change points and it implies that the proposed method gives much accurate location than the simulation method. On top of that it is impressive that the results of estimated locations are robust under the choices of window length. Note additionally that the proposed method runs faster than the simulation method in that bootstrapping requires substantial amount of computing.

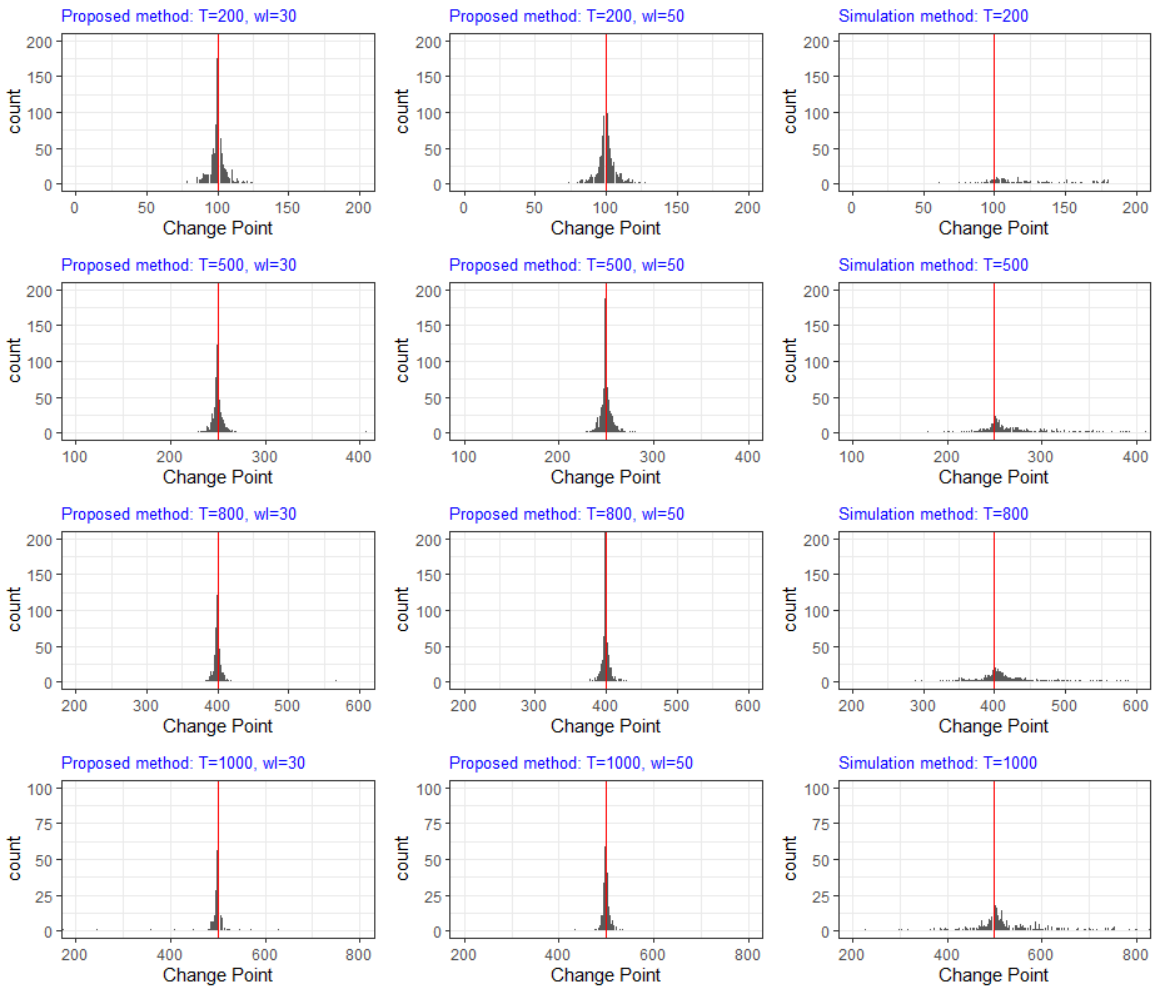

Figure 4.1 The results of single point detection

4.2 Multiple points detection

In this section, simulation results to observe the performance on multiple detection are provided. 1000 set of data are generated as:

$$\mathbf{y}_j \sim \begin{cases} \mathbf{MVN}(\mathbf{0}, \Sigma_{0.3}) & \text{for all } 1 \leq j \leq k_1^* \\ \mathbf{MVN}(\mathbf{0}, \Sigma_{0.7}) & \text{for all } k_1^* + 1 \leq j \leq k_2^* \\ \mathbf{MVN}(\mathbf{0}, \Sigma_{0.5}) & \text{for all } k_2^* + 1 \leq j \leq T, \end{cases}$$

where $T \in \{500, 800, 1000\}$, $k_1^* = \lfloor \frac{T}{3} \rfloor$ and $k_2^* = \lfloor \frac{2T}{3} \rfloor$

Table 4.2 summarizes specifications on ϵ's and wl's that are used in this study. In particular, ϵ_1 and wl_1 are used when detecting the initial change point. ϵ_2 and wl_2 are used when detecting change point in subdata 1, while ϵ_3 and wl_3 are used when detecting change point in subdata 2. On the other hand, 10,000 resampling is used for the simulation method.

Table 4.2 Specifications on ϵ's and wl's

T	ϵ_1	wl_1	ϵ_2	wl_2	ϵ_3	wl_3
500	0.06	60	0.2	40	0.1	40
800	0.04	70	0.15	40	0.06	40
1000	0.03	80	0.1	40	0.05	40

Figure 4.2 shows histograms of the locations of significant estimated change points and it shows the proposed method outperforms the simulation method in that the estimated points from the proposed method are more accurately located in the real change points.

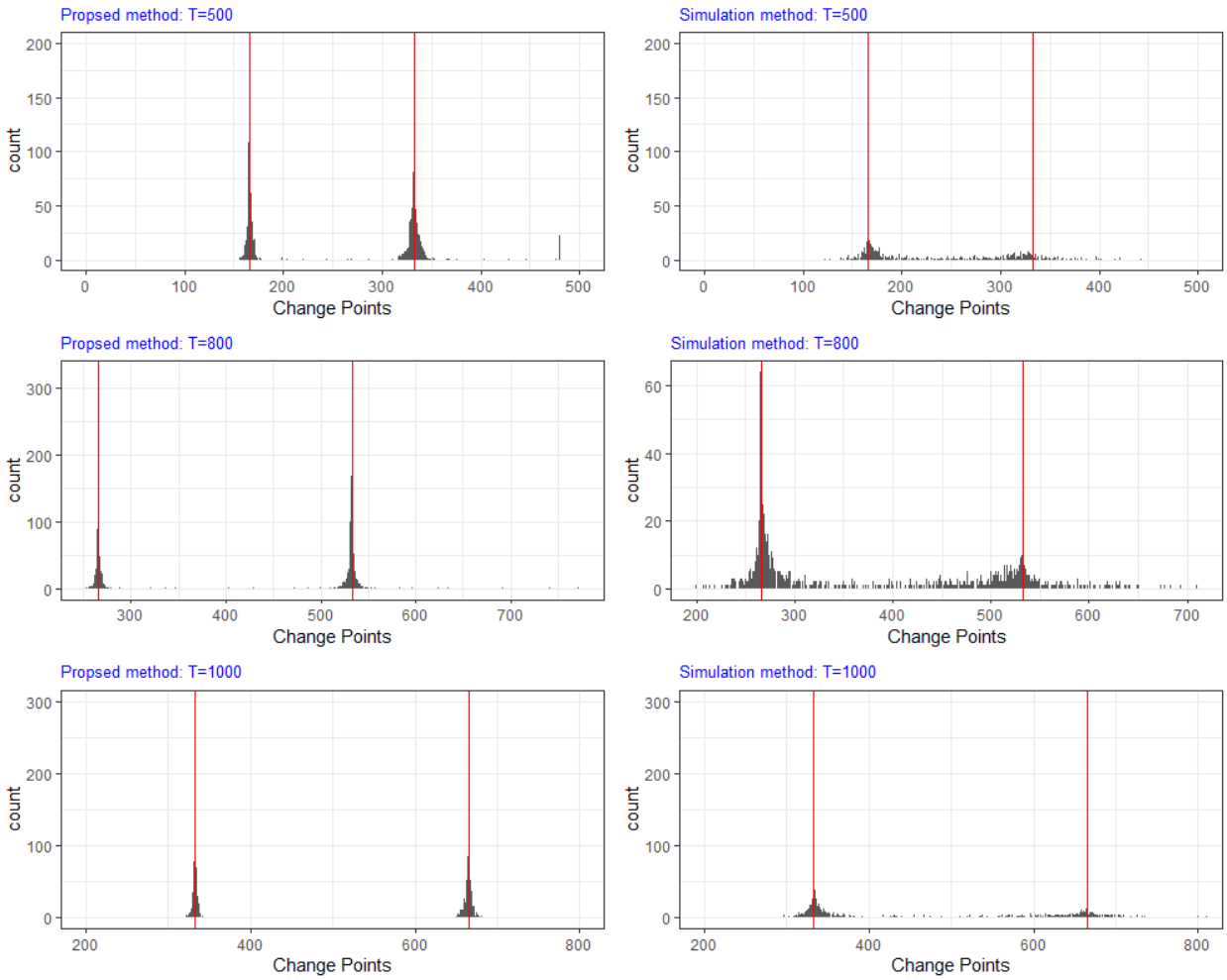

Figure 4.2 The results of multiple points detection

Chapter 5

Real data analysis

In this chapter, the method is applied to the portfolio data of a private equity fund in Korea. The data are daily records of the holdings of stocks in Korean stock market of the fund from 2013-01-02 to 2018-01-25, $n = 1252$ working days. The number of stocks constitute the fund is $p = 258$. Note however that stocks are grouped by 11 sectors, followed by the Global Industry Classification Standard. Remark that the fund is evaluated every six months by the firm and, if it is graded as D in sequel, then the fund is terminated at six months later after the evaluation. The fund is terminated at 2018-01-25 and this implies that it is evaluated as D about one year earlier and again graded as D about the evaluate six months earlier than the termination. The researchers want to know whether the managers of the fund may take a risk when the fund is first graded as D. In other words, they change the investment strategy and do not to follow their initial public offerings (IPOs), the portfolios initially designed for rasing the fund.

Table 5.1 summarizes the sector classification used during the analysis. Note that the Global Industry Classification Standard also has Sector 60 for

real estate. However, the portfolio has no stock in real estate and contains considerable amount of bank deposit. On top of that some stocks didn't belong to any of the Sector classification. This mismatch occured because the latest 2019 version of Global Industry Classification Standard is used while the portfolio consists of stocks from 2013 to 2018. Therefore bank deposit and mismatched stocks are grouped by etc.

Table 5.1 Sector classifications

Sector Number	Industry Group
10	Energy
15	Materials
20	Industrials
25	Consumer Discretionary
30	Consumer Staples
35	Health Care
40	Financials
45	Information Technology
50	Communication Service
55	Utility
etc	Deposit and stocks not included in the above

During the analysis, $wl = 25$ is chosen because it means the method compares the structures of the portfolio with 25 business days, that is, approximately a month. After detecting the most significant change point, the data is divided into two, by before and after the change point. Then the detection is repeated on each subdata.

Table 5.2 summarizes the estimated change points in the data. The most significant change point(CP1) is detected to exist in the 904^{th} business day, which corresponds to 2016-08-25. Indeed, Figure 5.1 shows deviated correlation

structures among sectors. In particular, stocks in Sector 50, 55 and etc show considerable changes.

Table 5.2 The estimated change points in the structure of the portfolio

Initial Change Point	Change Point in Subdata 1	Change Point in Subdata 2
904	233	1120

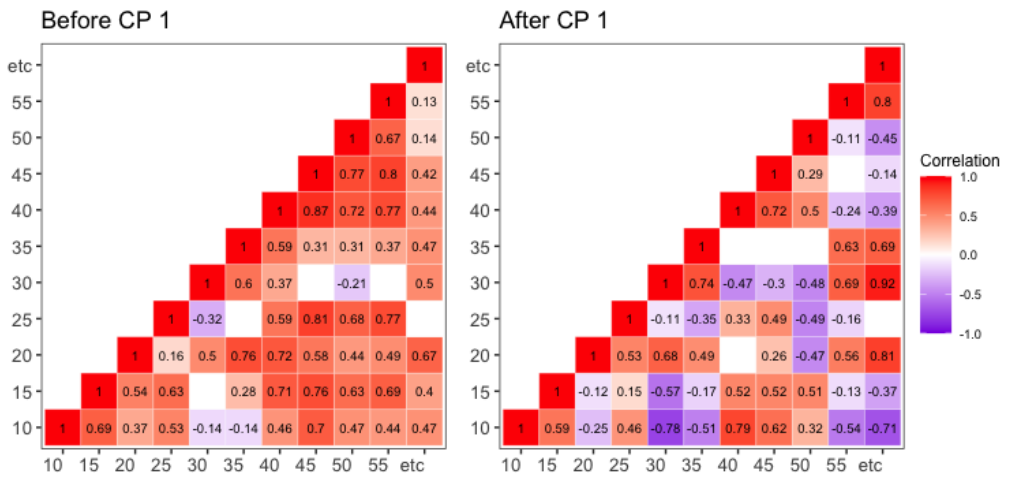

Figure 5.1 The correlation heatmaps for before and after the most significant change point

In subdata 1, change point is detected on the 233^{th} business day, which correspond to 2013-12-06. Figure 5.2 implies that the overall structure among stocks changed notably. This deviation can be explained by Figure 5.3. Figure 5.3 depicts trend of Vollatility Index of KOSPI 200(VKOSPI) and that of the variance of the portfolio, denoted by risk. According to Figure 5.3, before the late 2013, the risk of the portfolio was lower than VKOSPI and showed decoupled trends between them. However, as VKOSPI got lower and eventually met the risk in late 2013, the risk and VKOSPI began to show similar trend. One hypothetical explanation is that when the market had high risk, fund man-

agers reacted by their optimal managing strategy. However, as the market's risk got alleviated, they began to put little effort on managing strategy.

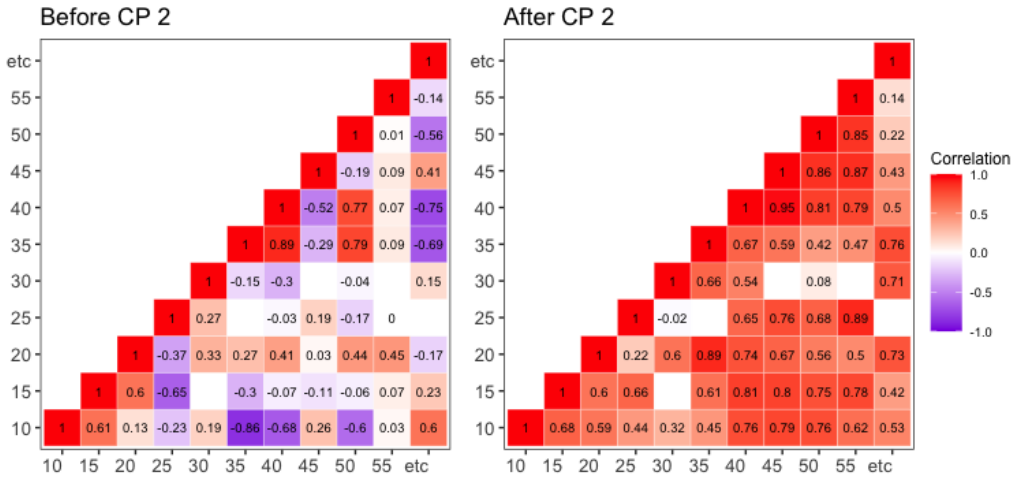

Figure 5.2 The correlation heatmaps for before and after the change point in the subdata 1

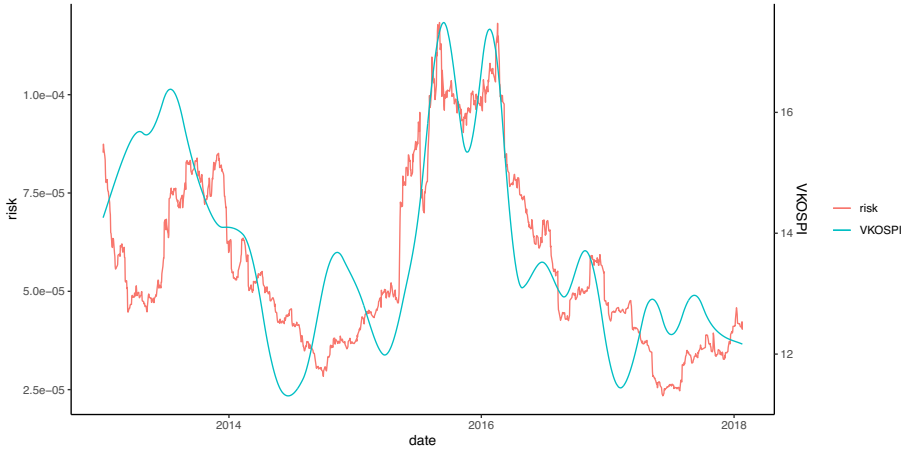

Figure 5.3 Trends of VKOSPI and the variance of the portfolio

In subdata 2, change point is estimated on the 1120^{th} business day, which correspond to 2017-07-11. Figure 5.4 shows there were overall changes in correlations among sectors. Since 2017-07-11 is to the second evaluation day, it seems that the fund managers tried to change their portfolio for improved return on investment to avoid D grade.

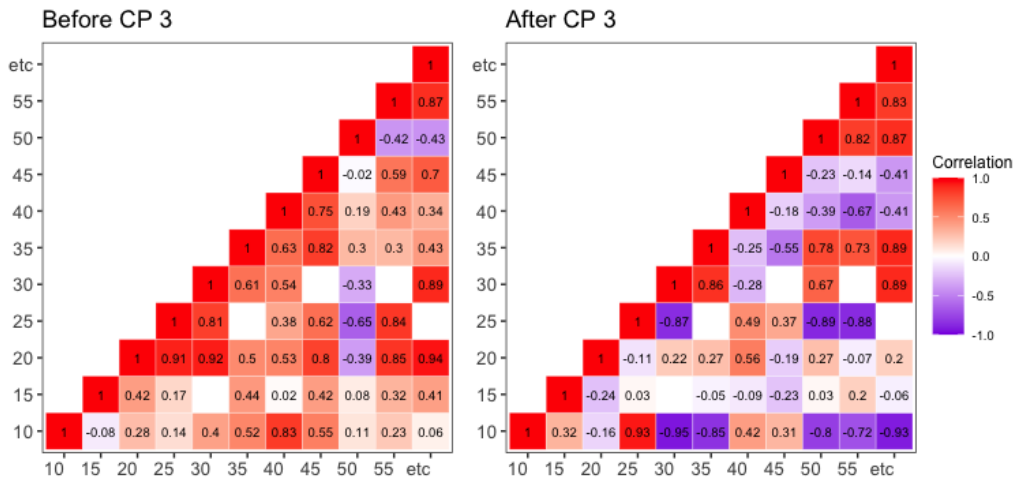

Figure 5.4 The correlation heatmaps for before and after the change point in the subdata 2

To sum up, this analysis was able to detect change points in the structure in the portfolio, and therefore the fund managers are suspected to have deviated their initial investment strategy. Especially, CP3, or the change point detected in the subdata 2 implies the managers have had taken additional risk to improve their return on investment, in order not to be terminated.

Chapter 6

Conculsion

Existing methods on detecting change points in covariance structure have drawbacks so that they are not suitable for arbitrary high-dimensional multivariate data. However, the proposed method is robust under high-dimension and the simulation study shows the proposed method outperforms the existing method.

Applying the proposed method to the private equity data, change points were detected. This provides that fund managers tried to take additional risk to improve their return which is a deviation from their initial managing strategy.

However, more discussions on the choice of ϵ and wl are needed as there is no golden rule. However, as a rule of thumb, wl is recommended to be at least p or greater. Moreover, once wl is determined, recommended level of ϵ is such that guarantees to drop the data as many as the wl. On the other hand, trying a various choices on wl does no harm as it is shown in the simulation study, thus one may try various level of wl until the optimal one is chosen.

Bibliography

Anderson, T. (2009). *An Introduction to Multivariate Statistical Analysis*.

Aue, A., Hörmann, S., Horváth, L., and Reimherr, M. (2009). Break detection in the covariance structure of multivariate time series models. *The Annals of Statistics*, 37(6B):4046–4087.

Barnett, I. and Onnela, J.-P. (2016). Change point detection in correlation networks. *Scientific reports*, 6(1):1–11.

Fryzlewicz, P. (2014). Wild binary segmentation for multiple change-point detection. *The Annals of Statistics*, 42(6):2243–2281.

Koo, T., Cho, S., and Lim, J. (2019). An invariant test for equality of two large scale covariance matrices.

Liu, Y. and Xie, J. (2020). Cauchy combination test: a powerful test with analytic p-value calculation under arbitrary dependency structures. *Journal of the American Statistical Association*, 115(529):393–402.

Zamba, K. and Hawkins, D. M. (2009). A multivariate change-point model for change in mean vector and/or covariance structure. *Journal of Quality Technology*, 41(3):285–303.

국문초록

본 연구는 고차원 다변량 시계열 데이터에서의 공분산 구조의 변화점을 탐지하는 방법을 제안한다. 해당 방법은 필라이 통계량이라고도 알려져 있는 베타 행렬의 트레이스를 활용한다. 이를 위해 Koo et al. (2019) 논문에서 제시하고 있는 필라이 통계량의 점근적 정규성을 활용한다. 본 방법은 각 시점에서 필라이 통계량을 계산하고 그에 따른 유의확률을 계산한다. 그 후 Liu and Xie (2020) 논문에서 제안하는 Cauchy combination test를 활용하여 각 시점에서 계산된 유의확률을 합해주며, 가장 큰 통계량을 보인 시점을 변화점이 발생한 시점이라 추정한다. 나아가, 다중 변화점 감지를 위해 Fryzlewicz (2014) 논문에서 제안하는 아이디어를 착안한다. 즉, 추정된 변화점을 기준으로 데이터를 쪼개고 유의한 변화점이 존재하지 않는다고 나올때까지 본 방법을 각 쪼개진 데이터에 적용한다. 본 논문은 시뮬레이션을 통해 본 방법의 사이즈와 검정력을 수치적으로 제시한다. 마지막으로 본 방법을 사모펀드 투자내역에 적용하여 펀드 운용 간 일탈행위의 존재여부를 확인해본다.

주요어: 베타 행렬, 고차원 공분산 행렬, 다중 변화점 감지, 필라이 통계량, 사모펀드, 확률 행렬 이론

학번: 2019-20319