Central Limit Theorem for the spiked Eigenvalues of Separable Sample Covariance Matrices

Zhang Bo

Division of Mathematical Sciences
School of Physical and Mathematical Sciences

2017

Central Limit Theorem for the spiked Eigenvalues of Separable Sample Covariance Matrices

Zhang Bo

Division of Mathematical Sciences
School of Physical and Mathematical Sciences

A thesis submitted to Nanyang Technological University
in fulfillment of the requirement for the degree of
Doctor of Philosophy in Mathematical Sciences

January 2017

Acknowledgements

I would like to express my deep appreciation to everyone who supported me during my Ph.D. studies at Nanyang Technology University.

- I would like to express my deepest gratitude to my supervisor, Pan Guangming, for his invaluable guidance, unfailing encouragement and patience over many years. He is not only an academic advisor, but also a mentor and a friend. He is always helpful whenever I need advice. I feel very fortunate for having the opportunity to work with him.

- I would like to thank the Department of Mathematics of Nanyang Technological University for providing scholarships to pursue research studies and opportunities to learn with outstanding academicians.

- I am grateful to Professor Gao Jiti who helps me a lot in doing research and providing me with the opportunity to learn with him.

- I am thankful to many teachers who have given me dedicated instructions during my long lasting student life. I am especially grateful to Prof. Lian Heng, Prof. Xiang Liming and Prof. Privault Nicolas for their tremendous help and guidance.

- I am so lucky to have so many amazing people as friends who make my four-year study at NTU a wonderful experience, Yang Yanrong, Chen Binbin, Bao Zhigang, Hu Jiang , Wang Shaochen, Yang Qing, Han Xiao, Liu Yiming and Zhang Zhixiang.

- Finally, I would like to thank my parents for their unwavering support during my time at Nanyang Technological University.

List of Works

Below is the list of the works during my PhD studies in NTU.

1. Han Xiao, Pan Guangming and Zhang Bo. The Tracy-Widom law for the largest eigenvalue of F type matrices. Annals of Statistics, 44(4), 1564-1592.

2. Zhang Bo, Pan Guangming and Gao Jiti. CLT for Largest Eigenvalues and Unit Root Tests for High-Dimensional Nonstationary Time Series. Submitted to Annals of Statistics. submitted to Annals of Statistics.(Under Review)

3. Pan Guangming and Zhang Bo. Change point detection in the preferential attachment graph, under revision.

4. Han Xiao, Pan Guangming and Zhang Bo. Central Limit Theorem for the spiked Eigenvalues of Separable Sample Covariance Matrices, under revision.

Contents

List of Tables

Abstract

This thesis is concerned about the central limit theorems for the spiked eigenvalues of separable sample covariance matrices and their applications.

The first problem is to test a p–dimensional time series model of the form: $\mathbf{x}_t = \mathbf{\Pi}\mathbf{x}_{t-1} + \Sigma^{1/2}\mathbf{y}_t,\ 1 \le t \le T$ where $\mathbf{y}_t = (Y_{t1}, \cdots, Y_{tp})'$ and $\Sigma^{1/2}$ is the square root of a symmetric positive definite matrix. Here $Y_{tj} = \sum_{k=0}^{\infty} b_k Z_{t-k,j}$ with $\sum_{i=0}^{\infty} |b_i| < \infty$ is a linear process where $\{Z_{ij}\}$ are independent and identically distributed (i.i.d.) random variables with $EZ_{ij} = 0$, $E|Z_{ij}|^2 = 1$ and $E|Z_{ij}|^4 < \infty$. Let $\mathbf{X} = (\mathbf{x_1}, \cdots, \mathbf{x_T})'$ and $\mathbf{X}^*$ be the conjugate transpose. We establish both the convergence in probability and the asymptotic joint distribution of the first k largest eigenvalues of $\mathbf{B} = (1/p)\mathbf{X}\mathbf{X}^*$ and $\bar{\mathbf{B}} = (1/p)(\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})^*$ when $\mathbf{\Pi} = \mathbf{I}$. Then we give two new unit root tests for high-dimensional time series as applications. We also provide some simulation results about the two tests.

Then we extend our theoretical results to the more general case. We study the separable sample covariance matrix $\Gamma\mathbf{X}\mathbf{\Omega}\mathbf{X}^T\Gamma^T$ with two different kinds of $\Gamma\Gamma^T$ and each of them has some extremely large eigenvalues. We prove the central limit theorems of the largest eigenvalues for the two cases and give two examples in time series data.

Chapter 1

Introduction

There have been an increasing interest and significant developments on the theory and methodologies for handling high-dimensional data in recent years. Understanding high-dimensional sample covariance matrices, including its eigenvalues and eigenvectors, has proved to be extremely useful for such developments. Indeed, random matrix theory has provided useful estimation and testing procedures for high-dimensional data analysis. Recent discussions on this topic can be found in Johnstone (2007), Paul and Aue (2014) and Yao et al. (2015).

Research towards understanding the eigenvalues of sample covariance matrices dates back to as early as the studies of Fisher (1937), Hsu (1939) and Roy (1939), and has become increasingly active since the publication of the celebrated work of Marcenko and Pastur (1967), in which the authors established a limiting spectral distribution (MP type distribution) for a sample covariance matrix for the case where p and T are comparable. More recent research has been devoted to establishing asymptotic properties for the eigenvalues and eigenvectors of high-dimensional sample covariance matrices.

There are currently two main lines of research about asymptotic distributions of the largest eigenvalues of high–dimensional random matrices. The first line of research is concerned with the Tracy-Widom law of the largest eigenvalues of random matrices. It is well known that limiting distributions of the largest eigenvalues of high-dimensional random matrices, such as Wigner matrices, follow the Tracy-Widom law, which was originally discovered by Tracy and Widom (1994) and Tracy and Widom (1996) for Gaussian Wigner ensembles. The largest eigenvalue of the Wishart matrix was investigated in Johnstone (2001). Several progresses for general sample covariance matrices have also been made, and we refer to Bao et al. (2015) and El Karoui (2007) among others.

Empirical data from wireless communication, finance and speech recognition often suggest that some extreme eigenvalues of sample covariance matrices are well separated from the rest. For example, the signals are often much bigger than noisy. This intrigues the second line of research about the spiked eigenvalues, which was first proposed in Johnstone (2001). Significant progresses have been made in recent years on the behaviour of these spiked eigenvalues. For instance, the CLTs of the largest eigenvalues of complex Gaussian sample covariance matrices with a spiked population were investigated in Baik et al. (2005), which also reported an interesting phase transition phenomenon. Baik and Silverstein (2006) further considered almost sure limits of the extreme sample eigenvalues of the general spiked population. Paul (2007) established a CLT for the spiked eigenvalues under the Gaussian population and the population spikes being simple. The fluctuation of the extreme sample eigenvalues of the general spiked population with arbitrary multiplicity numbers was further reported in Bai and Yao (2008).

Most of the above existing studies rely on the assumption that the observations of high dimensional data are independent, although dimensional correlation structure can be allowed. Observations of high–dimensional data in economics and finance are often highly dependent across time. We can't assume that the fluctuation of the stock market today is independent with the fluctuation yesterday. In view of this, Zhang (2006) investigated the empirical spectral distribution (ESD) of the sample covariance for the case where the data matrices are of the form $\mathbf{A}_1\mathbf{Z}\mathbf{A}_2$, where $\mathbf{A}_1$ and $\mathbf{A}_2$ are positive semidefinite matrices and $\mathbf{Z}$ has independent entries satisfying some moment assumptions. This model is referred to as the separable covariance model and allows for some dependence among observations recorded over different time points. Liu et al. (2015) studied the ESD of sample covariance matrices and symmetrized sample autocovariance matrices constructed from a linear process. Note that their setting also accommodates dependence among observations due to the fact that linear processes are built from the same innovation vectors. However, the above two papers considered the ESD only. To the best of our knowledge, there is no existing work available to deal with the largest eigenvalues of sample covariance matrices generated from high–dimensional nonstationary time series data. The main difficulty is that the properties of the population covariance matrices of the non-stationary data are unknown yet (even though we may make some assumptions about the error process). This thesis belongs to the second line of research about the spiked eigenvalues. We investigate the spiked eigenvalues of separable covariance matrices in some special cases. The main contribution of this paper is to establish several joint asymptotic distributions for the first several largest eigenvalues of separable covariance matrices in some special cases. An additional contribution of this paper is to propose two new unit root tests for testing nonstationarity of high–dimensional dependent time series.

The main content of the thesis is organized as follows.

- In Chapter 2, we establish several joint asymptotic distributions for the first several largest eigenvalues of sample covariance matrices of high–dimensional nonstationary time series data.
- In Chapter 3, we extend the CLT to the general case: the separable sample covariance matrix $\Gamma\mathbf{X}\mathbf{\Omega}\mathbf{X}^T\Gamma^T$ with two different kinds of $\Gamma\Gamma^T$ and each kind of $\Gamma\Gamma^T$ has some extremely large eigenvalues.
- In Chapter 4, we give some discussions about our results and future research.

Chapter 2

CLT for Largest Eigenvalues and Unit Root Testing for High–Dimensional Nonstationary Time Series

2.1 Introduction

The Chapter will be organized as follows. Section 2.2 establishes an asymptotic distributional theory for the first several largest eigenvalues of the covariance matrix of a high–dimensional dependent time series. Section 2.3 proposes two new unit root tests that are devoted to testing nonstationarity for high dimensional dependent data. Section 2.4 evaluates both the size and power properties of the proposed tests. Section 2.5 gives some concluding remarks. Appendix A establishes some useful results for truncated versions of sample covariance matrices by truncating linear processes. Appendix B gives the full proofs of the main theorems in Sections 2.2-2.3. The proofs of the results listed in Appendix A are given in Appendix C. Appendix D discusses some possible extensions of the main models to include a cointegrating structure and a deterministic trending component.

2.2 Asymptotic Theory

This section first introduces some necessary assumptions before we establish new asymptotic properties for the largest eigenvalues of the covariance matrix of a vector of high–dimensional time series.

2.2.1 Matrix models

The chapter is to consider high–dimensional covariance matrices for non-stationary time series. Specifically, define the following linear processes:

$$Y_{tj} = \sum_{k=0}^{\infty} b_k Z_{t-k,j} \tag{2.1}$$

with $\sum_{i=0}^{\infty} |b_i| < \infty$. Suppose that $\mathbf{y}_t = (Y_{t1}, \cdots, Y_{tp})'$ is a p–dimensional time series, where $\{Z_{ij}\}$ are independent and identically distributed (i.i.d.) random variables with $EZ_{ij} = 0$, $E|Z_{ij}|^2 = 1$ and $E|Z_{ij}|^4 < \infty$. Consider a p–dimensional time series model of the form:

$$\mathbf{x}_t = \mathbf{\Pi}\mathbf{x}_{t-1} + \mathbf{\Sigma}^{1/2}\mathbf{y}_t, \ 1 \le t \le T, \tag{2.2}$$

where the spectral norm of the coefficient matrix $\mathbf{\Pi}$ is bounded by one $(0 \le \|\mathbf{\Pi}\|_2 \le 1)$.

Define $\bar{\mathbf{X}} = \left(\frac{\sum_{\mathbf{t=1}}^{\mathbf{T}} \mathbf{x_t}}{\mathbf{T}}, \cdots, \frac{\sum_{\mathbf{t=1}}^{\mathbf{T}} \mathbf{x_t}}{\mathbf{T}}\right)'$ as a $T \times p$ matrix. Introduce the non-centered and centered sample covariance matrices

$$\mathbf{B} = \frac{1}{p}\mathbf{X}\mathbf{X}^* \tag{2.3}$$

and

$$\bar{\mathbf{B}} = \frac{1}{p}(\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})^* \tag{2.4}$$

with $\mathbf{X} = (\mathbf{x_1}, \cdots, \mathbf{x_T})'$. Here we point out that when $\mathbf{\Pi} = \mathbf{0}$, $\mathbf{\Sigma}$ satisfies some conditions and Y_{tj}'s are i.i.d random variables, the Tracy-Widom distribution has been established for the large eigenvalue of $\mathbf{B}$ in Bao et al. (2015). Also, when $\mathbf{\Pi} = \mathbf{0}$, $\mathbf{\Sigma}$ is a block matrix with spiked eigenvalues and Y_{tj}'s are i.i.d random variables, an asymptotic distribution (Gaussian distribution under some conditions) for the largest eigenvalues of $\mathbf{B}$ has been discussed in Paul (2007) and Bai and Yao (2008). It is not clear yet how the largest eigenvalues of $\mathbf{B}$ may behave when Y_{tj}'s have some dependence structure. One case is that $\mathbf{\Pi} = \mathbf{0}$, but $\mathbf{\Sigma}$ is involved in (2.1). When $\mathbf{\Pi} = \mathbf{I}$, (2.2) becomes nonstationary. The main motivation for considering such a model is the proposal of two unit root tests to be discussed in the next section.

This thesis is to investigate the largest eigenvalues of $\mathbf{B}$ and $\bar{\mathbf{B}}$ for the cases where $\mathbf{\Pi} = \mathbf{I}$ or $\|\mathbf{\Pi}\|_2 = \varphi < 1$. Throughout the thesis, we make the following assumptions about the coefficients b_i and $\mathbf{\Sigma}$:

(A1) $\sum_{i=0}^{\infty} i|b_i| < \infty$.

(A2) $\sum_{i=0}^{\infty} b_i = s \neq 0$.

(A3) There exist two positive constants M_0 and M_1 such that $\|\mathbf{\Sigma}\|_2 \leq M_0$ and $\frac{tr(\mathbf{\Sigma})}{p} \geq M_1$.

(A4) Let $T \to \infty$ and $p \to \infty$ such that $\lim_{T,\ p\to\infty} \frac{\sqrt{p}}{T} = 0$.

Here $\|\cdot\|_2$ stands for either the spectral norm of a matrix or the Euclidean norm of a vector. The linear process includes both MA(q) and AR(1) models. Assumption A2 is easily satisfied. Note that we do not require p and T to be of the same order, which is being commonly used in the random matrix theory literature. Assumption A3 covers some commonly used $\mathbf{\Sigma}$.

For example one may verify that the identity matrix $\mathbf{I}$ and the Toeplitz matrices satisfy it. However, we point out that Assumption A3 rules out the case where cross-sectional dependence has a factor model structure, which leads to very large eigenvalues of $\mathbf{\Sigma}$. We also need to make some assumptions about Z_{ij} and $\mathbf{x_0}$.

(A5) $\{Z_{i,j}\}$ are i.i.d random variables with mean zero, variance one and bounded fourth moment. Let $\mathbf{z}_t = (Z_{t1}, \cdots, Z_{tp})'$, where t can be either positive or negative integer (for the purpose of introducing A7 below).

(A6) $E\|\mathbf{x_0}\|_2^2 = O(p)$.

(A7) $\mathbf{x_0} = \sum_{k=0}^{\infty} \tilde{b}_k \mathbf{\Sigma_1}^{1/2}\mathbf{z_{-k}} + \tilde{b}_{-1}\mathbf{\Sigma_2}^{1/2}\tilde{\mathbf{z}} + \tilde{\mathbf{b}}_{-2}$, where $\|\mathbf{\Sigma_1}\|_2 \le M_0$, $\|\mathbf{\Sigma_2}\|_2 \le M_0$ and $\tilde{\mathbf{z}} = (\tilde{Z}_1, \cdots, \tilde{Z}_p)'$ is independent of $\mathbf{z_t}$ for any t, in which $\{\tilde{Z}_j\}$ are i.i.d random variables with mean zero, variance one and finite fourth moments. The coefficients satisfy $\sum_{k=0}^{\infty} |\tilde{b}_k| + |\tilde{b}_{-1}| < \infty$ and $\|\tilde{\mathbf{b}}_{-2}\|^2 = O(p)$.

2.2.2 Main results for non-centered sample covariance matrix B

To characterize the limits in probability of the eigenvalues of $\mathbf{B}$, define for $k = 1, \cdots, T$,

$$\lambda_k = \frac{1}{2(1+\cos\theta_k)} \quad \text{with} \quad \theta_k = \frac{2(T+1-k)\pi}{2T+1}, \tag{2.5}$$

and

$$\gamma_k = \lambda_k \left(a_0 + 2\sum_{j=1}^{\infty} a_j (-1)^j \cos(j\theta_k) \right), \tag{2.6}$$

where

$$a_i = \sum_{k=0}^{\infty} b_k b_{k+i}. \tag{2.7}$$

γ_k and λ_k are useful in studying the mean and variance of the largest eigenvalues. We first characterize the magnitude of λ_k and γ_k.

Proposition 1. *Let Assumptions A1 and A2 hold. For any fixed constant $k \geq 1$, there is a constant c_k such that*

$$\lim_{T\to\infty} \frac{\gamma_k}{T^2} = c_k > 0 \tag{2.8}$$

and

$$\lim_{T\to\infty} \frac{\gamma_k}{\gamma_1} = \lim_{T\to\infty} \frac{\lambda_k}{\lambda_1} = \frac{1}{(2k-1)^2}. \tag{2.9}$$

We are now at a position to state the main results; their proofs are given in Appendix B. The first theorem develops an upper bound in probability for the spectral norm of $\mathbf{B}$ for the stationary case. The second theorem gives a limit in probability and a joint distribution for the first k largest eigenvalues of $\mathbf{B}$ for nonstationary data.

Theorem 1. *Let Assumptions A1–A6 hold. When $0 \leq \|\mathbf{\Pi}\|_2 = \varphi < 1$, we obtain*

$$\|\mathbf{B}\|_2 = O_p\left(\frac{\left(1+\sqrt{\frac{T}{p}}\right)^2}{(1-\varphi)^2}\right). \tag{2.10}$$

Theorem 2. *Let Assumptions A1–A5 hold. Let ρ_k be the kth largest eigenvalue of $\mathbf{B}$. Let $\mathbf{\Pi} = \mathbf{I}$ and k is fixed.*

(1) If Assumption A6 holds, we have

$$\frac{\rho_k - \gamma_k \frac{tr(\mathbf{\Sigma})}{p}}{\gamma_1} \xrightarrow{i.p.} 0 \tag{2.11}$$

as p and T tend to infinity, where i.p. means convergence in probability.

(2) If Assumption A7 holds, the random vector

$$\frac{\sqrt{p}}{\gamma_1}\left(\rho_1-\gamma_1\frac{tr(\mathbf{\Sigma})}{p},\cdots,\rho_k-\gamma_k\frac{tr(\mathbf{\Sigma})}{p}\right)' \tag{2.12}$$

converges weakly to a zero-mean Gaussian vector $\mathbf{w}=(w_1,\cdots,w_k)'$ *with covariance function* $\mathrm{cov}(w_i,w_j)=0$ *for any* $i\neq j$ *and* $\mathrm{var}(w_i)=\frac{2\theta}{(2i-1)^4}$ *with* $\theta=\lim\limits_{p\to\infty}\frac{tr(\mathbf{\Sigma}^2)}{p}$*, as* p *and* T *tend to infinity.*

Remark 1. *The result holds for the complex case as well. In fact when* $\mathbf{Z}$ *is complex, set*

$$Re(Z_{jk})=Z_{ij}^R,\quad and\quad Im(Z_{jk})=Z_{ij}^I. \tag{2.13}$$

Let Z_{ij}^R *and* Z_{ij}^I *be independent. Then* $\frac{\sqrt{p}}{\gamma_1}\left(\rho_1-\gamma_1\frac{tr(\mathbf{\Sigma})}{p},\cdots,\rho_k-\gamma_k\frac{tr(\mathbf{\Sigma})}{p}\right)'$ *converges weakly to a zero-mean Gaussian vector* $\mathbf{w}=(w_1,\cdots,w_k)'$ *with* $\mathrm{var}(w_i)=\frac{2\theta}{(2i-1)^4}(1-2E(Z_{i1}^R)^2E(Z_{i1}^I)^2)$*, in which* $\theta=\lim\limits_{p\to\infty}\frac{tr(\mathbf{\Sigma}^2)}{p}$*. When* $i\neq j$*,* $\mathrm{cov}(w_i,w_j)=0$*.*

Remark 2. *If Assumption A7 does not hold but Assumption A6 is true, then Theorem 2 remains true under Assumptions A1-A3, A5 and* $\lim\limits_{T,\ p\to\infty}\frac{p}{T}=0$*.*

Remark 3. *We now compare our results with those in Bai and Yao (2008). Bai and Yao (2008) needs to assume that the observations are independent and that* $\mathbf{\Sigma}$ *has a spiked structure. In my thesis, the observations are highly dependent. Furthermore, we need not assume a spiked structure of* $\mathbf{\Sigma}$*, since the spiked eigenvalues come naturally from the random walk structure.*

2.2.3 Main results for centered sample covariance matrix $\bar{\mathbf{B}}$

We now consider the largest eigenvalues of $\bar{\mathbf{B}}$. To characterize the limits in probability of the eigenvalues of $\bar{\mathbf{B}}$, define for $k = 1, \cdots, T$,

$$\bar{\lambda}_k = \frac{1}{2(1+\cos\bar{\theta}_k)} \quad \text{with} \quad \bar{\theta}_k = \frac{(T-k)\pi}{T}, \tag{2.14}$$

and

$$\bar{\gamma}_k = \bar{\lambda}_k \left(a_0 + 2\sum_{j=1}^{\infty} a_j (-1)^j \cos(j\bar{\theta}_k) \right). \tag{2.15}$$

We below characterize the magnitude of $\bar{\lambda}_k$ and $\bar{\gamma}_k$. The result is similar to Proposition 1.

Proposition 2. *Let Assumptions A1 and A2 hold. For any fixed constant $k \geq 1$, there is a constant $\bar{c}_k$ such that*

$$\lim_{T\to\infty} \frac{\bar{\gamma}_k}{T^2} = \bar{c}_k > 0 \tag{2.16}$$

and

$$\lim_{T\to\infty} \frac{\bar{\gamma}_k}{\bar{\gamma}_1} = \lim_{T\to\infty} \frac{\bar{\lambda}_k}{\bar{\lambda}_1} = \frac{1}{k^2}. \tag{2.17}$$

We next list the results, which are similar to Theorems 1 and 2.

Theorem 3. *Let Assumptions A1–A6 hold. When $0 \leq \|\mathbf{\Pi}\|_2 = \varphi < 1$, we obtain*

$$\|\bar{\mathbf{B}}\|_2 = O_p\left(\frac{\left(1+\sqrt{\frac{T}{p}}\right)^2}{(1-\varphi)^2} \right). \tag{2.18}$$

Theorem 4. *Let Assumptions A1–A5 hold. Let $\bar{\rho}_k$ be the kth largest eigenvalue of $\bar{\mathbf{B}}$. Let $\mathbf{\Pi} = \mathbf{I}$ and k is fixed. We then have the following results:*

as p and T tend to infinity,

$$\frac{\bar{\rho}_k - \bar{\gamma}_k \frac{tr(\mathbf{\Sigma})}{p}}{\bar{\gamma}_1} \xrightarrow{i.p.} 0, \tag{2.19}$$

and the random vector

$$\frac{\sqrt{p}}{\bar{\gamma}_1}\left(\bar{\rho}_1 - \bar{\gamma}_1 \frac{tr(\mathbf{\Sigma})}{p}, \cdots, \bar{\rho}_k - \bar{\gamma}_k \frac{tr(\mathbf{\Sigma})}{p}\right)' \tag{2.20}$$

converges weakly to a zero-mean Gaussian vector $\bar{\mathbf{w}} = (\bar{w}_1, \cdots, \bar{w}_k)'$ with covariance function $\mathrm{cov}(\bar{w}_i, \bar{w}_j) = 0$ for any $i \neq j$ and $\mathrm{var}(\bar{w}_i) = \frac{2\theta}{i^4}$ with $\theta = \lim_{p\to\infty} \frac{tr(\mathbf{\Sigma}^2)}{p}$.

Remark 4. *It is noted that Theorem 4 doesn't need Assumptions A6 and A7 due to the structure of $\bar{\mathbf{B}}$.*

We are now ready to introduce two new unit root tests for the high–dimensional time series case before the proofs of the theorems are given in Appendix B below.

2.3 Unit Root Testing

This section is to explore an application of the main results to the proposal of a new unit root test for a high–dimensional time series setting.

Unit root testing is to check whether time series data are nonstationary or not. Existing studies on this topic can be found in Dickey and Fuller (1979), Chan and Wei (1988) and Phillips and Perron (1988). In the past two decades, unit root testing in panel data has received much attention. Many researchers (e.g. Choi (2001), Levin et al. (2002) and Im et al. (2003)) consider the time series case where the error process is independent across individuals. There are also some results (see, for example, Chang (2004), Pesaran (2003) and Pesaran et al. (2013)) that have considered the case where the error process is cross–sectional dependent. A book by Pesaran (2015) summarizes some recent developments about unit root testing for both time series and panel data settings. In the above literature, researchers often need to first estimate the covariance matrix of a panel of times associated with cross–sectional dependence. However, when the dimensionality of the time series becomes large, it is hard to consistently estimate it without imposing some structure on the covariance matrix. We therefore propose two new tests using the covariance matrices of high–dimensional time series under consideration.

To this end, a key observation is that Theorems 2 and 4 indicate that the largest eigenvalues of $\mathbf{B}$ and $\bar{\mathbf{B}}$ are of order T^2 in probability (the order of γ_1 and $\bar{\gamma}_1$, which are given in Propositions 1 and 2), while Theorems 1, 3 and Assumption (A4) imply that when $0 \leq \varphi < 1$, $\|\mathbf{B}\|_2 = o_p(T)$ and $\|\bar{\mathbf{B}}\|_2 = o_p(T)$. This motivates us to propose two new unit root tests based on the largest eigenvalues.

2.3.1 The model and test statistics

We consider the following model:

$$\mathbf{x_t} = (\mathbf{I} - \mathbf{\Pi})\phi + \mathbf{\Pi}\mathbf{x_{t-1}} + \mathbf{\Sigma}^{1/2}\mathbf{y_t},\ 1 \leq t \leq T, \tag{2.21}$$

where ϕ is a p-dimensional vector. The null hypothesis H_0 is $\mathbf{\Pi} = \mathbf{I}$ and the alternative hypothesis H_1 is $\|\mathbf{\Pi}\|_2 < 1$.

Theorem 2 states that under $H_0:\ \mathbf{\Pi} = \mathbf{I}$, the statistic $L_p = \frac{\sqrt{p}\left(\rho_1 - \gamma_1 \frac{tr(\mathbf{\Sigma})}{p}\right)}{\gamma_1\sqrt{2\theta}}$ converges weakly to a standard normal variable. Note that $\gamma_1 \frac{tr(\Sigma)}{p}$ and $\gamma_1\sqrt{2\theta}$ are both unknown in practice. We would like to emphasize that γ_1, $\frac{tr(\mathbf{\Sigma})}{p}$ and θ can not be estimated individually. However, it is possible to estimate their product as a whole. Specifically speaking, an estimator of $\frac{\gamma_1}{\lambda_1}\frac{tr(\mathbf{\Sigma})}{p}$ is proposed as follows.

Define $\breve{\mathbf{x}}_{f,g} = (\mathbf{x_f} - \mathbf{x_{f-1}})'(\mathbf{x_g} - \mathbf{x_{g-1}})$ for $1 \le f, g \le T$. A direct calculation yields $E\breve{\mathbf{x}}_{f,g} = a_{|f-g|}tr(\mathbf{\Sigma})$. Moreover, note that $\sum_{j=1}^{m_1} a_j(-1)^j \cos(j\theta_1)$ can be approximated by $\sum_{j=1}^{m_1} a_j$ for an appropriate m_1 to be specified below. In view of this, we propose an estimator of $\frac{\gamma_1}{\lambda_1}\frac{tr(\mathbf{\Sigma})}{p}$ as

$$\mu_{m_1} = \sum_{i=2}^{T} \frac{\breve{\mathbf{x}}_{i,i}}{p(T-1)} + 2\sum_{j=1}^{m_1}\sum_{i=2}^{T-j} \frac{\breve{\mathbf{x}}_{i,i+j}}{p(T-j-1)}. \tag{2.22}$$

We next find an estimator for $\gamma_1\sqrt{2\frac{tr(\mathbf{\Sigma^2})}{p}}$. The strategy is to find an estimator for the ratio of $\gamma_1\sqrt{2\frac{tr(\mathbf{\Sigma^2})}{p}}$ and $\gamma_1\frac{tr(\mathbf{\Sigma})}{p}$ first and then construct its estimator in conjunction with μ_{m_1}, the estimator of $\frac{\gamma_1}{\lambda_1}\frac{tr(\mathbf{\Sigma})}{p}$. To this end, we first find an estimator for $a_0^2\, tr(\Sigma^2)$. One may verify that $Var(\breve{\mathbf{x}}_{f,g}) = (a_{|f-g|}^2 + a_0^2)tr(\mathbf{\Sigma^2})$. It is also noted that $a_{|f-g|} = o(|f-g|)$ due to Assumption A1 so that the term $a_{|f-g|}$ in $Var(\breve{\mathbf{x}}_{f,g})$ can be negligible when choosing $|f-g|$ sufficiently large. We then propose an estimator for $a_0^2\, tr(\Sigma^2)$ as follows:

$$S_{\sigma^2,0} = \frac{\sum_{f=2}^{[T/2]} \sum_{g=f+[T/2]}^{T} \breve{\mathbf{x}}_{f,g}^2}{(T - \frac{3}{2}[T/2])([T/2]-1)}. \tag{2.23}$$

Furthermore, one may verify (by calculating the mean and variance) that

$$\frac{\sqrt{\frac{S_{\sigma^2,0}}{p}}}{\sum_{i=2}^{T}\frac{\breve{\mathbf{x}}_{i,i}}{p(T-1)}} - \frac{\sqrt{\frac{tr(\mathbf{\Sigma^2})}{p}}}{\frac{tr(\Sigma)}{p}} \xrightarrow{i.p.} 0$$

as p and T tend to infinity. We may then construct S_{σ^2,m_2}, the estimator of $\frac{\gamma_1}{\lambda_1}\sqrt{2\frac{tr(\mathbf{\Sigma^2})}{p}}$, as follows:

$$S_{\sigma^2,m_2} = \frac{|\mu_{m_2}|\sqrt{2\frac{S_{\sigma^2,0}}{p}}}{\sum_{i=2}^{T}\frac{\check{\mathbf{x}}_{i,i}}{p(T-1)}}$$

where m_2 is specified below.

Also, note that $\gamma_1/\lambda_1 = \bar{\gamma}_1/\bar{\lambda}_1$. Once the two estimators are available, we can construct the following test statistics, T_N and $\bar{T}_N$, of the form:

$$T_N = \sqrt{p}\frac{\rho_1 - \lambda_1\mu_{m_1}}{\lambda_1 S_{\sigma^2,m_2}} \tag{2.24}$$

and

$$\bar{T}_N = \sqrt{p}\frac{\bar{\rho}_1 - \bar{\lambda}_1\mu_{m_1}}{\bar{\lambda}_1 S_{\sigma^2,m_2}}, \tag{2.25}$$

where λ_1 and $\bar{\lambda}_1$ are given in (2.5) and (2.14), respectively.

Theorem 5. *Let Assumptions A1–A5 hold,* $m_1 = [\sqrt{p}]$ *and* m_2 *tends to infinity. Under* $H_0: \ \mathbf{\Pi} = \mathbf{I}$, *we have*

$$\bar{T}_N \xrightarrow{d} N(0,1) \tag{2.26}$$

as p *and* T *tend to infinity, where* $\xrightarrow{d}$ *stands for convergence in distribution.*

Furthermore, if Assumptions A7 also holds, under $H_0: \ \mathbf{\Pi} = \mathbf{I}$, *we have*

$$T_N \xrightarrow{d} N(0,1) \tag{2.27}$$

as p *and* T *tend to infinity.*

Remark 5. *The conditions imposed on* m_1 *and* m_2 *can be further relaxed. For example, if there exists a positive integer* s *such that* $b_i = 0$ *for any* $i > s$ *in (2.1), we find* $a_i = 0$ *for any* $i > s$ *in (2.7). So we don't need to estimate* a_i *for any* $i > s$ *and one can choose* $m_1 = m_2 = \min\{s, [\sqrt{p}]\}$ *in this case. This point helps us to simplify the design and the verifications of the assumptions for the simulation in Section 4 below.*

Now we investigate the power of T_N and $\bar{T}_N$ for the case where $\{Y_{tj}\}$ in (2.1) are i.i.d..

Theorem 6. *Let Assumptions A1–A5 hold with $b_i = 0$ for $i \geq 1$. Consider H_1 : $\mathbf{\Pi} = \varphi \mathbf{I}$ for $0 \leq \varphi < 1$. Then under the case of $m_1 = m_2 = 0$, we have*

$$\lim_{p,T\to\infty} P\left(\bar{T}_N > C_0 | H_1\right) = 1 \tag{2.28}$$

for some $C_0 > \ell_\alpha$, where ℓ_α is the α–level critical value of the standard normal distribution.

Furthermore, if $\|\phi\|_2^2 = O(p)$, then

$$\lim_{p,T\to\infty} P\left(T_N > C_0 | H_1\right) = 1. \tag{2.29}$$

Remark 6. *Although $\bar{T}_N$ and T_N may have the same asymptotic results when p and T are big enough, there may be differences under the small sample case. In fact under H_0, $\mathbf{x_0}$ affects the largest eigenvalues of $\mathbf{B}$ but doesn't affect the largest eigenvalues of $\bar{\mathbf{B}}$. So it may affect the size of T_N when the sample is small. Under H_1, ϕ affects the largest eigenvalues of $\mathbf{B}$ but doesn't affect the largest eigenvalues of $\bar{\mathbf{B}}$. It may affect the power of T_N when the sample is small. So $\bar{T}_N$ may be more useful than T_N when we don't have ϕ or $\mathbf{x_0}$. But when we have the condition that $\phi = 0$ and $\mathbf{x_0} = 0$, $\gamma_1 \approx 4\bar{\gamma}_1$ so that T_N can have a stronger power than $\bar{T}_N$ under small sample cases.*

Remark 7. *There are some well–known panel unit root tests (e.g. Choi (2001) and Levin et al. (2002)). They considered the case of $\mathbf{\Pi} = \mathrm{diag}(\varphi_1, \cdots, \varphi_N)$ and used the estimators of φ_i to test whether $\mathbf{\Pi} = \mathbf{I}$. Moreover, when the covariance matrix $\mathbf{\Sigma}$ is involved, it has to be estimated in order to test whether $\mathbf{\Pi} = \mathbf{I}$ (e.g. Chang (2004)). So such existing tests may only work* *for the finite-dimensional case. By contrast, our test makes the best use of the properties of the largest eigenvalues of B instead of estimating φ_i. In addition, we do not impose special structures, such as sparsity on the covariance matrix $\mathbf{\Sigma}$.*

Before the proofs of Theorems 5–6 are given in Appendix B, we evaluate the finite–sample performance of the proposed tests and also compare them with two natural competitors in Section 4 below.

2.4 Simulation

This section is to conduct some simulations to investigate the size and power of T_N and $\bar{T}_N$.

2.4.1 The selection of m_1 and m_2

Recalling Remark 5, we below propose a method to choose suitable m_1 and m_2. Note that

$$\zeta_j = \frac{\sum_{i=2}^{T-j} \frac{\breve{\mathbf{x}}_{i,i+j}}{p(T-j-1)}}{\sum_{i=2}^{T} \frac{\breve{\mathbf{x}}_{i,i}}{p(T-1)}} \xrightarrow{i.p.} \frac{a_j}{a_0}$$

with the rate $\frac{1}{\sqrt{pT}}$ as p and T tend to infinity. Particularly $\zeta_j = O(\frac{1}{\sqrt{pT}})$ if $a_j = 0$(one can verify it by calculating the mean and the variance of $\breve{\mathbf{x}}_{i,i+j}$). Moreover, if there exists a positive integer s such that $b_i = 0$ for any $i > s$ in (2.1), we find $a_i = 0$ for any $i > s$ in (2.7). So one can choose $m_1 = m_2 = \min\{s, [\sqrt{p}]\}$ in this case. In practice one can see whether $a_j = 0$ by comparing ζ_j with $p^{-1/2}T^{-1/4}$. Here $p^{-1/2}T^{-1/4}$ is used as a bound instead of $\frac{1}{\sqrt{pT}}$ since the convergence rate of μ_{m_1} to $\frac{\gamma_1}{\lambda_1}\frac{tr\Sigma}{p}$ should be

$o(p^{-1/2})$ to ensure that $\sqrt{p}(\mu_{m_1} - \frac{\gamma_1}{\lambda_1}\frac{tr\Sigma}{p}) = o(1)$. In view of this, we propose the following way of selecting m_1 and m_2:

$$\hat{m}_1 = \hat{m}_2 = \min\{\{0 \le i < [\sqrt{p}] : |\zeta_j| < p^{-1/2}T^{-1/4}, i < j < [\sqrt{p}]\} \cup \{[\sqrt{p}]\}\}. \tag{2.30}$$

The equation means that choosing an smaller integer between $[\sqrt{p}]$ and the smallest number i which satisfies $|\zeta_j| < p^{-1/2}T^{-1/4}$ for any $i < j < [\sqrt{p}]$.

Note that $\hat{m}_1$ and $\hat{m}_2$ work well when p and T are big enough. While when p and T are small, $\hat{m}_1$ and $\hat{m}_2$ may be affected by $\frac{a_j}{a_0}$. If $a_j \neq 0$ but $\frac{a_j}{a_0}$ is small, $\hat{m}_1$ and $\hat{m}_2$ may cause some problem when p and T are small.

2.4.2 The parametric bootstrap method

We also consider a parametric bootstrap method for our test statistics T_N and $\bar{T}_N$. Let $\dot{\mathbf{\Sigma}} = \frac{1}{T}\sum_{t=1}^{T}(\mathbf{x_t} - \mathbf{x_{t-1}})(\mathbf{x_t} - \mathbf{x_{t-1}})'$. If there is a constant $\dot{C} > 0$ such that $\frac{p}{T} \le \dot{C}$, we can find that $\| \dot{\mathbf{\Sigma}} \|_2 = O_p(1)$ and $\frac{tr(\dot{\mathbf{\Sigma}})}{p} = \dot{M}_1 + O_p(\frac{1}{\sqrt{p}})$, where $\dot{M}_1 > 0$. It is easily seen that Assumption A3 still holds for $\dot{\mathbf{\Sigma}}$. We then draw a new sample $\dot{\mathbf{x}}_\mathbf{t} = \dot{\mathbf{x}}_\mathbf{t-1} + \dot{\mathbf{\Sigma}}^{1/2}\dot{\mathbf{y}}_\mathbf{t}$ where $\dot{\mathbf{y}}_\mathbf{t}$ is a p-dimensional random vector from $N(0, \mathbf{I_p})$ and $\dot{\mathbf{y}}_\mathbf{t}$ is independent over t. Note that Assumptions A1-A7 still hold for $\dot{\mathbf{x}}_\mathbf{t}$. Let $\dot{\mathbf{X}} = (\dot{\mathbf{x}}_\mathbf{1}, \cdots, \dot{\mathbf{x}}_\mathbf{T})'$. We define $\dot{T}_N$ and $\dot{\bar{T}}_N$ from $\dot{\mathbf{X}}$, the analogues of T_N and $\bar{T}_N$, respectively. It follows from Theorem 5 that $\dot{T}_N \xrightarrow{d} N(0,1)$ and $\dot{\bar{T}}_N \xrightarrow{d} N(0,1)$. So for any p and T we can redraw $\dot{\mathbf{x}}_\mathbf{t}$ for many times (e.g. 200 times) to get an empirical distributions for each of $\dot{T}_N$ and $\dot{\bar{T}}_N$. Then we use the critical values from the empirical distributions to replace the critical values calculated from $N(0,1)$. When p and T are not big, the simulations show that $\dot{T}_N$ and $\dot{\bar{T}}_N$ based on the critical values from the empirical distributions perform better than these tests associated with the corresponding critical values calculated from $N(0,1)$.

2.4.3 Comparison with the existing tests

There are several existing unit root tests available for panel data. Some of them consider the case where there is no cross-sectional dependence (see, for example, the IPS test proposed in Im et al. (2003)). If there is cross-sectional dependence, the IPS test doesn't work. To test for nonstationarity in the panel data case with cross-sectional dependence, Chang (2004) showed that the Bootstrap method with estimation of $\mathbf{\Sigma}$ performs better for the case where p is fixed and T is large. Chang (2004) also stated that the Bootstrap-OLS performs better than Bootstrap-GLS when p is large. Furthermore, GLS doesn't work when $p \geq T$. We therefore compare T_N with the t–statistic corresponding to the Bootstrap-OLS t^*_{ols} and the F-statistic corresponding to Bootstrap-OLS F^*_{ols}.

We use the setting $\mathbf{y_t} = \mathbf{z_t}$ and $\mathbf{\Sigma} = \left(\Sigma_{i,j}\right) = \left(0.3^{|i-j|}\right)$. We compare the size performance of our test T_N with the two tests t^*_{ols} and F^*_{ols} under H_0 with $\mathbf{x_0} = 0$ and $\phi = 0$. Table 2.1 reports the results of the three tests based on 1000 replications, 500 bootstrap replications and different values of p and T. The nominal size throughout this section is set to be 0.05.

Then, we compare our test $\bar{T}_N$ with the two tests t^*_{ols} and F^*_{ols} under H_0 and $\mathbf{x_0} = 0, \mathbf{\Sigma} = \left(\Sigma_{i,j}\right) = \left(\frac{1}{(i-j)^2+1}\right)$. We sample each element of ϕ from the standard normal distribution. The results of the three test statistics based on 1000 replications, 500 bootstrap replications and different values of p and T are reported in Table 2.2.

One can observe that when p becomes large, both t^*_{ols} and F^*_{ols} have a poor size property even though $\mathbf{y_t}$ is independent over t. This indicates that their asymptotic distributions may not hold under the null hypothesis when

Table 2.1: *The empirical size of three tests*

the test	T \ p	5	10	20	40	60	80
T_N	40	0.057	0.057	0.041	0.048	0.050	0.040
t^*_{ols}	40	0.045	0.028	0.014	0.000	0.000	0.000
F^*_{ols}	40	0.054	0.044	0.027	0.000	0.003	0.001
T_N	60	0.053	0.050	0.048	0.055	0.048	0.044
t^*_{ols}	60	0.046	0.031	0.016	0.003	0.000	0.000
F^*_{ols}	60	0.044	0.047	0.024	0.007	0.000	0.002
T_N	80	0.053	0.048	0.041	0.052	0.048	0.041
t^*_{ols}	80	0.045	0.033	0.023	0.006	0.000	0.000
F^*_{ols}	80	0.064	0.035	0.027	0.011	0.003	0.000

Table 2.2: *The empirical size of three tests*

the test	T \ p	5	10	20	40	60	80
$\bar{T}_N$	40	0.070	0.056	0.062	0.052	0.038	0.043
t^*_{ols}	40	0.036	0.013	0.008	0.001	0.000	0.000
F^*_{ols}	40	0.056	0.029	0.014	0.001	0.000	0.000
$\bar{T}_N$	60	0.061	0.060	0.047	0.041	0.045	0.053
t^*_{ols}	60	0.041	0.037	0.011	0.001	0.000	0.000
F^*_{ols}	60	0.052	0.054	0.027	0.002	0.000	0.000
$\bar{T}_N$	80	0.055	0.058	0.053	0.048	0.041	0.048
t^*_{ols}	80	0.041	0.043	0.015	0.006	0.000	0.000
F^*_{ols}	80	0.041	0.045	0.035	0.008	0.000	0.000

p is large. One of the reasons is that when p is large and the population covariance matrix does not have any special structures, we cannot find any consistent estimates for the population covariance matrix and the unknown parameters involved. As a consequence, their asymptotic distributions may fail to hold under the null.

2.4.4 Simulation results for T_N under an MA(1) model

We now consider the setting where $\mathbf{y_t} = \psi \mathbf{z_{t-1}} + \mathbf{z_t}$, $\psi = 0.5$ and $\mathbf{\Sigma} = \left(\Sigma_{i,j}\right) = \left(0.3^{|i-j|}\right)$. To show the performance with the non–diagonal $\mathbf{\Pi}$, we design the following matrix as an alternative one:

$$(\mathbf{\Pi_2})_{ij} = \begin{cases} 0.5 & i = j, \\ 0.2 & |i-j| = 1, \\ 0 & |i-j| \geq 2. \end{cases}$$

We consider the performance of T_N and set $\phi = 0$. Under H_0 we set $\mathbf{x_0} = 0$. Under H_1 we generate the data by (2.21) with $t = -51, -50, \cdots, T$. Using an asymptotic critical value calculated from $N(0,1)$, the size and power results of T_N based on 1000 replications and different values of p, T and $\mathbf{\Pi}$ are reported in Table 2.3. We also use the parametric bootstrap method proposed in Section 4.2. The size and power results of T_N based on 1000 replications, 200 bootstrap replications and different values of p, T and $\mathbf{\Pi}$ are reported in Table 2.4.

Table 2.3: *The results for T_N and MA(1)*

p	T	**I** (size)	**0.95I** (power)	**0.9I** (power)	Π_2 (power)
20	20	0.019	0.102	0.216	0.510
20	30	0.037	0.109	0.672	0.830
20	40	0.036	0.346	0.951	0.935
20	60	0.043	0.896	1.000	0.997
20	80	0.039	0.997	1.000	1.000
40	20	0.019	0.102	0.580	0.710
40	30	0.028	0.301	0.964	0.938
40	40	0.031	0.752	0.999	0.974
40	60	0.034	0.997	1.000	0.998
40	80	0.033	1.000	1.000	1.000
60	20	0.021	0.100	0.766	0.876
60	30	0.029	0.421	0.998	0.981
60	40	0.033	0.905	1.000	0.989
60	60	0.045	1.000	1.000	0.998
60	80	0.046	1.000	1.000	1.000
80	20	0.020	0.116	0.870	0.932
80	30	0.029	0.561	1.000	0.996
80	40	0.032	0.966	1.000	0.997
80	60	0.036	1.000	1.000	1.000
80	80	0.034	1.000	1.000	1.000

Table 2.4: *The results for T_N and MA(1) with the parametric bootstrap method*

p	T	**I** (size)	**0.95I** (power)	**0.9I** (power)	Π_2 (power)
20	20	0.031	0.144	0.636	0.812
20	30	0.063	0.464	0.974	0.936
20	40	0.051	0.818	0.998	0.992
20	60	0.049	0.992	1.000	1.000
20	80	0.082	1.000	1.000	1.000
40	20	0.061	0.140	0.860	0.838
40	30	0.051	0.578	0.990	0.972
40	40	0.041	0.926	1.000	0.990
40	60	0.052	0.998	1.000	1.000
40	80	0.054	1.000	1.000	1.000
60	20	0.055	0.126	0.930	0.932
60	30	0.048	0.676	1.000	0.994
60	40	0.068	0.972	1.000	0.996
60	60	0.053	1.000	1.000	1.000
60	80	0.056	1.000	1.000	1.000
80	20	0.055	0.132	0.950	0.960
80	30	0.047	0.742	1.000	0.994
80	40	0.056	0.984	1.000	0.996
80	60	0.054	1.000	1.000	1.000
80	80	0.057	1.000	1.000	1.000

2.4.5 Simulation results for $\bar{T}_N$ under an MA(1) model

We still use the setting in Section 4.4 but sample each element of ϕ from the standard normal distribution. In each case, we use the critical value calculated from either $N(0,1)$ or by the parametric bootstrap method. The size and power results of $\bar{T}_N$ based on 1000 replications and different values of p, T and $\mathbf{\Pi}$ are reported in Tables 2.5-2.6.

When p is small, the size and power results of T_N and $\bar{T}_N$ based on the critical value either calculated from $N(0,1)$ or by the bootstrap method are reported in Tables 2.7 and 2.8. From Tables 2.7 and 2.8, one can observe that while $\bar{T}_N$ and T_N roughly have similar size values, the power of T_N is slightly better than that of $\bar{T}_N$. The power of the statistics of $\bar{T}_N$ and T_N improves when p and T increase. The parametric bootstrap (proposed in this thesis) based critical value in each case results in a stable size and better power than using an asymptotic critical value for the case where p is as small as $p = 5$ or $p = 10$.

In summary, for the case of $p = 5$ or $p = 10$, Tables 2.7 and 2.8 show that the size and power values of T_N and $\bar{T}_N$ based on the asymptotic critical value of $N(0,1)$ are much less stable and reasonable than those based on the parametric bootstrap critical value in each case. Tables 2.3–2.6 then show that when $p \geq 20$ and $T \geq 20$, there are stable sizes and reasonable power values for both T_N and $\bar{T}_N$ based on 1000 replications, 200 bootstrap replications and different values of p, T and $\mathbf{\Pi}$.

Table 2.5: *The results for $\bar{T}_N$ and MA(1)*

p	T	$\mathbf{I}$ (size)	$\mathbf{0.95I}$ (power)	$\mathbf{0.9I}$ (power)	Π_2 (power)
20	20	0.018	0.018	0.013	0.100
20	30	0.042	0.029	0.124	0.213
20	40	0.043	0.071	0.290	0.383
20	60	0.046	0.264	0.746	0.733
20	80	0.055	0.580	0.959	0.907
40	20	0.016	0.034	0.075	0.176
40	30	0.033	0.081	0.290	0.363
40	40	0.034	0.235	0.584	0.572
40	60	0.044	0.708	0.985	0.919
40	80	0.043	0.968	1.000	0.987
60	20	0.014	0.036	0.144	0.254
60	30	0.029	0.202	0.523	0.518
60	40	0.036	0.408	0.823	0.729
60	60	0.039	0.870	0.999	0.936
60	80	0.042	0.993	1.000	0.998
80	20	0.012	0.064	0.191	0.310
80	30	0.032	0.267	0.661	0.644
80	40	0.037	0.532	0.934	0.800
80	60	0.043	0.945	1.000	0.971
80	80	0.039	0.997	1.000	1.000

Table 2.6: *The results for $\bar{T}_N$ and MA(1) with the parametric bootstrap method*

p	T	**I** (size)	**0.95I** (power)	**0.9I** (power)	Π_2 (power)
20	20	0.034	0.144	0.239	0.326
20	30	0.051	0.310	0.606	0.596
20	40	0.061	0.502	0.837	0.782
20	60	0.067	0.824	0.986	0.971
20	80	0.078	0.946	1.000	0.996
40	20	0.040	0.188	0.352	0.412
40	30	0.049	0.412	0.695	0.660
40	40	0.045	0.604	0.873	0.782
40	60	0.054	0.950	0.999	0.972
40	80	0.053	0.994	1.000	0.998
60	20	0.034	0.232	0.452	0.504
60	30	0.049	0.506	0.807	0.704
60	40	0.047	0.728	0.961	0.850
60	60	0.048	0.980	1.000	0.980
60	80	0.062	1.000	1.000	0.998
80	20	0.033	0.276	0.512	0.534
80	30	0.040	0.548	0.903	0.826
80	40	0.046	0.816	0.986	0.910
80	60	0.052	0.990	1.000	0.986
80	80	0.064	1.000	1.000	0.998

Table 2.7: *The results for T_N and small p*

p	T	critical value	**I** (size)	**0.95I** (power)	**0.9I** (power)	Π_2 (power)
5	20	N(0,1)	0.040	0.056	0.008	0.070
		bootstrap	0.084	0.114	0.308	0.560
5	30	N(0,1)	0.039	0.014	0.010	0.038
		bootstrap	0.077	0.180	0.580	0.762
5	40	N(0,1)	0.051	0.002	0.012	0.030
		bootstrap	0.079	0.262	0.772	0.882
5	60	N(0,1)	0.055	0.000	0.006	0.002
		bootstrap	0.079	0.570	0.938	0.986
5	80	N(0,1)	0.049	0.000	0.002	0.002
		bootstrap	0.076	0.816	0.992	0.992
10	20	N(0,1)	0.023	0.078	0.048	0.202
		bootstrap	0.085	0.132	0.462	0.664
10	30	N(0,1)	0.031	0.016	0.142	0.330
		bootstrap	0.075	0.240	0.826	0.896
10	40	N(0,1)	0.039	0.042	0.322	0.416
		bootstrap	0.077	0.580	0.972	0.966
10	60	N(0,1)	0.037	0.126	0.558	0.502
		bootstrap	0.069	0.894	1.000	0.998
10	80	N(0,1)	0.049	0.246	0.678	0.598
		bootstrap	0.064	0.982	1.000	1.000

Table 2.8: *The results for $\bar{T}_N$ and small p*

p	T	critical value	$\mathbf{I}$(size)	$\mathbf{0.95I}$(power)	$\mathbf{0.9I}$(power)	Π_2(power)
5	20	N(0,1)	0.031	0.014	0.008	0.024
		bootstrap	0.046	0.086	0.155	0.238
5	30	N(0,1)	0.039	0.002	0.002	0.006
		bootstrap	0.066	0.132	0.284	0.422
5	40	N(0,1)	0.051	0.000	0.000	0.002
		bootstrap	0.071	0.170	0.417	0.518
5	60	N(0,1)	0.050	0.000	0.000	0.002
		bootstrap	0.063	0.328	0.712	0.754
5	80	N(0,1)	0.053	0.000	0.000	0.000
		bootstrap	0.069	0.466	0.896	0.926
10	20	N(0,1)	0.025	0.004	0.009	0.081
		bootstrap	0.049	0.098	0.218	0.308
10	30	N(0,1)	0.043	0.002	0.018	0.068
		bootstrap	0.056	0.214	0.450	0.471
10	40	N(0,1)	0.046	0.014	0.031	0.108
		bootstrap	0.073	0.300	0.653	0.684
10	60	N(0,1)	0.062	0.022	0.117	0.168
		bootstrap	0.069	0.560	0.904	0.880
10	80	N(0,1)	0.057	0.022	0.217	0.234
		bootstrap	0.075	0.748	0.991	0.960

2.5 Conclusions and Discussion

This thesis has developed an asymptotic theory for the largest eigenvalues of the covariance matrix of a high–dimensional time series vector. As an application, a new unit root test developed for testing nonstationarity in high–dimensional time series vectors has been proposed and then discussed both theoretically and numerically. The small sample properties discussed in Section 4 have offered the support to the theory established in Sections 2 and 3.

One possible extension involves the case where either a deterministic trending time series component or a factor model structure is included in model (2.21). As a consequence, it may be more appropriate to compare the corresponding versions of T_N and $\bar{T}_N$ with those proposed by Im et al. (2003), Pesaran (2003) and Pesaran et al. (2013). As suggested by the referees, another extension of model (2.21) is to take into account certain type of cointegrating structures. Appendix D of the supplementary document gives some brief discussion about possible extensions, which require developing new techniques and should be left for future research.

2.6 Acknowledgments

We would like to thank Professor Yoosoon Chang for kindly providing us with the computer code for the simulations in Chang (2004).

Appendix A: Results for truncated matrices

This section is to consider the truncated version of the sample covariance matrix. Let $\mathbf{Y} = (\mathbf{y}_1, \cdots, \mathbf{y}_T)'$ be a $T \times p$ random matrix. Define

$$Y_{ij,l} = \sum_{k=0}^{l} b_k Z_{i-k,j}$$

with $l = \max\{p, T\}$, a truncated version of Y_{tj} in (2.1). However, to simplify notation, we let $b_i = 0$ for all $i > l$ in this section, so that we can still use Y_{ij} instead of $Y_{ij,l}$. In this way a_i defined in (2.7) and Y_{tj} in (2.1) respectively become

$$a_i = \sum_{k=0}^{l-i} b_k b_{k+i}, \qquad Y_{tj} = \sum_{k=0}^{l} b_k Z_{t-k,j}.$$

Furthermore let $\mathbf{F} = (F_{ij})$ be a $T \times (T + l)$ matrix with

$$F_{ij} = \begin{cases} b_{l+i-j} & i \leq j \leq i + l, \\ 0 & \text{otherwise.} \end{cases} \tag{2.31}$$

It follows that $\mathbf{Y} = \mathbf{F}\mathbf{Z_p}$, where $\mathbf{Z_p}$ is a $(T + l) \times p$ random matrix with $(\mathbf{Z_p})_{i,j} = Z_{i-l,j}$. For the sake of notation simplicity, we below denote $\mathbf{Z_p}$ by $\mathbf{Z}$ and $(\mathbf{Z_p})_{i,j}$ by Z_{ij}. Let $\mathbf{A} = (A_{ij})_{T\times T} = (a_{|i-j|})_{T\times T}$. We then have $\mathbf{A} = \mathbf{F}\mathbf{F}'$. We would like to remind the readers that l depends on T, so that $a_{|i-j|}$ depends on T.

We also assume that $\mathbf{x_0} = \mathbf{0}$ in this section.

Appendix A.1: Upper bound of the spectral norm of B for stationary data This subsection is to investigate the upper bound of the spectral norm of $\mathbf{B}$ for stationary data.

Proposition 3. *Suppose that Assumptions A1-A5 hold. When* $0 \leq \|\mathbf{\Pi}\|_2 = \varphi < 1$,

$$\lim_{T\to\infty} P\left(\|\mathbf{B}\|_2 \leq \frac{8\sum_{i\geq 0}|a_i|}{(1-\varphi)^2} M_0 \left(1 + \sqrt{\frac{T}{p}}\right)^2\right) = 1.$$

The proof of the proposition is available from Appendix C.1.

Appendix A.2: Convergence in Probability and CLT of the first k largest eigenvalues when $\mathbf{\Pi} = \mathbf{I}$

Define $\mathbf{C} = (C_{ij})_{1\leq i,j\leq T}$ to be a $T \times T$ lower triangular matrix with

$$C_{ij} = 0 \ \text{ for } j > i \text{ and } \ C_{ij} = 1 \ \text{ for } 1 \leq j \leq i. \tag{2.32}$$

In this case one has

$$\mathbf{B} = (1/p)\mathbf{X}\mathbf{X}^* = (1/p)\mathbf{C}\mathbf{Y}\boldsymbol{\Sigma}\mathbf{Y}^*\mathbf{C}^* = (1/p)\mathbf{C}\mathbf{F}\mathbf{Z_p}\boldsymbol{\Sigma}\mathbf{Z_p^*}\mathbf{F}^*\mathbf{C}^*. \tag{2.33}$$

Proposition 4. *Suppose that Assumptions A1-A5 hold. Let* ρ_k *be the* k*th largest eigenvalue of* $\mathbf{B}$. *When* $\boldsymbol{\Pi} = \mathbf{I}$, $\frac{\rho_k - \gamma_k \frac{tr(\boldsymbol{\Sigma})}{p}}{\gamma_1} \to 0$ *in probability.*

Proposition 5. *Suppose that Assumptions A1-A5 hold. Let* ρ_k *be the* k*th largest eigenvalue of* $\mathbf{B}$. *When* $\boldsymbol{\Pi} = \mathbf{I}$, $(\sqrt{p}\frac{\rho_1-\gamma_1}{\gamma_1}, \cdots, \sqrt{p}\frac{\rho_k-\gamma_k}{\gamma_1})'$ *converges weakly to a zero-mean Gaussian vector* $\mathbf{w} = (w_1, \cdots, w_k)'$ *with covariance* $cov(w_i, w_j) = \delta_{ij}\frac{\theta}{(2i-1)^4}(2 - 4E(Z_{i1}^R)^2 E(Z_{i1}^I)^2)$ *and* $\theta = \lim_{p\to\infty} \frac{tr(\boldsymbol{\Sigma}^2)}{p}$.

The proofs of the propositions are available from Appendix C.2.

Appendix A.3: The results for $\bar{\mathbf{B}}$

The following results for $\bar{\mathbf{B}}$ are similar to those for $\mathbf{B}$. In view of (2.33), write

$$\bar{\mathbf{B}} = (1/p)\mathbf{H}\mathbf{C}\mathbf{F}\mathbf{Z_p}\boldsymbol{\Sigma}\mathbf{Z_p^*}\mathbf{F}^*\mathbf{C}^*\mathbf{H}^*, \tag{2.34}$$

where $\mathbf{H} = \mathbf{I} - \frac{\mathbf{1}\mathbf{1}'}{T}$ with the $p \times 1$ vector $\mathbf{1}$ consisting of all one.

Proposition 6. *Suppose that Assumptions A1-A5 hold. Let* $\bar{\rho}_k$ *be the* k*th largest eigenvalue of* $\bar{\mathbf{B}}$. *When* $\boldsymbol{\Pi} = \mathbf{I}$, $\frac{\bar{\rho}_k - \bar{\gamma}_k \frac{tr(\boldsymbol{\Sigma})}{p}}{\bar{\gamma}_1} \to 0$ *in probability.*

Proposition 7. *Suppose that Assumptions A1-A5 hold. Let* $\bar{\rho}_k$ *be the* k*th largest eigenvalue of* $\bar{\mathbf{B}}$. *When* $\boldsymbol{\Pi} = \mathbf{I}$, $(\sqrt{p}\frac{\bar{\rho}_1-\bar{\gamma}_1}{\bar{\gamma}_1}, \cdots, \sqrt{p}\frac{\bar{\rho}_k-\bar{\gamma}_k}{\bar{\gamma}_1})'$ *converges weakly to a zero-mean Gaussian vector* $\bar{\mathbf{w}} = (\bar{w}_1, \cdots, \bar{w}_k)'$ *with covariance* $cov(\bar{w}_i, \bar{w}_j) = \delta_{ij}\frac{\theta}{i^4}(2 - 4E(Z_{i1}^R)^2 E(Z_{i1}^I)^2)$ *and* $\theta = \lim_{p\to\infty} \frac{tr(\boldsymbol{\Sigma}^2)}{p}$.

The proofs of the propositions are available from the supplementary file.

Appendix B: Proofs of the Main Results

This section is to prove that the results obtained in Section 4 still hold for the general linear process (without the truncation step performed there) and the general initial vector $\mathbf{x}_0$. We define a $T \times p$ matrix $\mathbf{X_0} = (\mathbf{x_0}, \cdots, \mathbf{x_0})'$

consisting of the initial vector $\mathbf{x}_0$ of the time series. When $\mathbf{\Pi} = \mathbf{I}$, we may rewrite $\mathbf{X} = \mathbf{C}\mathbf{Y}\mathbf{\Sigma}^{1/2} + \mathbf{X_0}$ and $\bar{\mathbf{X}} = \frac{\mathbf{1}\mathbf{1}'}{T}\mathbf{C}\mathbf{Y}\mathbf{\Sigma}^{1/2} + \mathbf{X_0}$ so that the sample covariance matrices $\mathbf{B}$ and $\bar{\mathbf{B}}$ can be rewritten as follows:

$$\mathbf{B} = \frac{1}{p}\mathbf{X}\mathbf{X}^* = \frac{1}{p}\mathbf{C}\mathbf{Y}\mathbf{\Sigma}\mathbf{Y}^*\mathbf{C}^* + \frac{1}{p}\mathbf{C}\mathbf{Y}\mathbf{\Sigma}^{1/2}\mathbf{X_0}^* + \frac{1}{p}\mathbf{X_0}\mathbf{\Sigma}^{1/2}\mathbf{Y}^*\mathbf{C}^* * + \frac{1}{p}\mathbf{X_0}\mathbf{X_0}^* \tag{2.35}$$

and

$$\bar{\mathbf{B}} = \frac{1}{p}(\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})^* = \frac{1}{p}(\mathbf{I} - \frac{\mathbf{1}\mathbf{1}'}{T})\mathbf{C}\mathbf{Y}\mathbf{\Sigma}\mathbf{Y}^*\mathbf{C}^*(\mathbf{I} - \frac{\mathbf{1}\mathbf{1}'}{T})^*. \tag{2.36}$$

Lemma 1. *Recall the definitions of* $\mathbf{Y}$, λ_k *and* γ_k *in Section 2. Let* $l = \max\{p, T\}$ *and* $\mathbf{Y_l}$ *be the truncated matrix of* $\mathbf{Y}$ *in Section 4. Define*

$$\gamma_{k,l} = \lambda_k(a_{0,l} + 2\sum_{1\le j\le T-1} a_{j,l}(-1)^j\cos(j\theta_k))$$

where

$$a_{j,l} = \sum_{j\le k\le l} b_k b_{k-j}. \tag{2.37}$$

Then when $\mathbf{\Pi} = \mathbf{I}$*,*

$$\left\|\frac{(1/p)\mathbf{C}(\mathbf{Y}\mathbf{\Sigma}\mathbf{Y}^* - \mathbf{Y_l}\mathbf{\Sigma}\mathbf{Y_l}^*)\mathbf{C}^*}{\gamma_{1,l}}\right\|_2 = o_p(p^{-1/2}) \tag{2.38}$$

and

$$\frac{|\gamma_{k,l} - \gamma_k|}{\gamma_{1,l}} = o(1). \tag{2.39}$$

Proof of Lemma 1. We consider (2.39) first. To this end, observe that Assumption (A1) implies that

$$\sum_{i=0}^{\infty} i|a_i| < \infty, \tag{2.40}$$

because

$$\sum_{i=0}^{\infty} i|a_i| \le \sum_{i=0}^{\infty} i\sum_{k=0}^{\infty}|b_k||b_{k+i}| = \sum_{k=0}^{\infty}|b_k|(\sum_{i=0}^{\infty} i|b_{k+i}|) \le \sum_{k=0}^{\infty}|b_k|(\sum_{i=0}^{\infty} i|b_i|).$$

Write

$$\frac{|\gamma_{k,l}-\gamma_k|}{\gamma_{1,l}} \le \frac{\lambda_k}{\gamma_{1,l}}\left(\sum_{k>l} b_k^2 + 2\sum_{j=1}^{T-1}\sum_{k>l}|b_k||b_{k-j}| + 2\sum_{j\ge T}|a_j|\right)$$

$$\le \frac{\lambda_k}{\gamma_{1,l}}\left(\sum_{k>l} b_k^2 + 2\sum_{j=1}^{\infty}|b_j|\sum_{k>l}|b_k| + 2\sum_{j\ge T}|a_j|\right).$$

From (2.40) and Assumption (A1), we obtain that

$$\sum_{k>l} b_k^2 + 2\sum_{j=1}^{\infty}|b_j|\sum_{k>l}|b_k| + 2\sum_{j\ge T}|a_j| = o(1).$$

Moreover, Lemma 3 and Assumption (A1) (or (2.90)) imply that $\frac{\lambda_k}{\gamma_{1,l}}$ is bounded. So we conclude (2.39).

Now, we consider (2.38). Using Lemma 2 in the supplementary file, observe that

$$\left\|\frac{(1/p)\mathbf{C}\left(\mathbf{Y\Sigma Y}^* - \mathbf{Y_l\Sigma Y_l}^*\right)\mathbf{C}^*}{\gamma_{1,l}}\right\|_2 \le \frac{\|\mathbf{C}\|_2^2}{\gamma_{1,l}} \|(1/p)(\mathbf{Y\Sigma Y}^* - \mathbf{Y_l\Sigma Y_l}^*)\|_2$$

$$= \frac{\lambda_1}{\gamma_{1,l}} \|(1/p)(\mathbf{Y\Sigma Y}^* - \mathbf{Y_l\Sigma Y_l}^*)\|_2 .$$

As before $\frac{\lambda_1}{\gamma_{1,l}}$ is bounded. So we just need to consider $\|(1/p)(\mathbf{Y\Sigma Y}^* - \mathbf{Y_l\Sigma Y_l}^*)\|_2$. Let $\mathbf{K} = (K_{ij})_{1\le i\le T, 1\le j\le p} = \mathbf{Y} - \mathbf{Y_l}$. We can obtain that $K_{ij} = \sum_{k=l+1}^{\infty} b_k Z_{i-k,j}$ and

$$E|K_{ij}|^2 = \sum_{k=l+1}^{\infty} b_k^2.$$

By Assumption (A1), we can get

$$E|K_{ij}|^2 = \sum_{k=l+1}^{\infty} b_k^2 \le l^{-2}\sum_{k=l+1}^{\infty} k^2|b_k|^2 = o(l^{-2}),$$

which implies

$$E\left\|\frac{1}{\sqrt{p}}\mathbf{K}\right\|_{F^2} = o(Tl^{-2}).$$

This, together with (2.80), implies that

$$\begin{aligned} &\| (1/p)(\mathbf{Y\Sigma Y}^* - \mathbf{Y_l\Sigma Y_l}^*) \|_2 = \| (1/p)(\mathbf{K\Sigma Y_l}^* + \mathbf{Y_l\Sigma K}^* + \mathbf{K\Sigma K}^*) \|_2 \\ &\le 2 \| \frac{1}{\sqrt{p}}\mathbf{K} \|_F \|\mathbf{\Sigma}\|_2 \| \frac{1}{\sqrt{p}}\mathbf{Y_l} \|_2 + \| \frac{1}{\sqrt{p}}\mathbf{K} \|_F^2 \|\mathbf{\Sigma}\|_2 = o_p(p^{-1/2}). \end{aligned} \tag{2.41}$$

This concludes (2.38). □

Proof of Theorem 2. At first we prove (2.11). Recalling (2.35),

$$\mathbf{B} = \frac{1}{p}\mathbf{XX}^* = \frac{1}{p}\mathbf{CY\Sigma Y}^*\mathbf{C}^* + \frac{1}{p}\mathbf{CY\Sigma}^{1/2}\mathbf{X_0}^* + \frac{1}{p}\mathbf{X_0\Sigma}^{1/2}\mathbf{Y}^*\mathbf{C}^* + \frac{1}{p}\mathbf{X_0X_0}^*.$$

Assumption A6 implies that

$$\|\frac{1}{p}\mathbf{X_0X_0}^*\|_2 = O_p(T) \tag{2.42}$$

and that

$$\|\frac{1}{p}\mathbf{CY\Sigma}^{1/2}\mathbf{X_0}^*\|_2 = O_p\left(T^{1/2}\|\frac{1}{p}\mathbf{CY\Sigma Y}^*\mathbf{C}^*\|_2^{1/2}\right). \tag{2.43}$$

We can write $\frac{(1/p)\mathbf{CY\Sigma Y}^*\mathbf{C}^*}{\gamma_1}$ as

$$\begin{aligned} &\frac{(1/p)\mathbf{CY\Sigma Y}^*\mathbf{C}^*}{\gamma_1} = \frac{\gamma_{1,l}}{\gamma_1}\frac{(1/p)\mathbf{CY\Sigma Y}^*\mathbf{C}^*}{\gamma_{1,l}} \\ = &\frac{\gamma_{1,l}}{\gamma_1}\frac{(1/p)\mathbf{CY_l\Sigma Y_l^*C}^*}{\gamma_{1,l}} + \frac{\gamma_{1,l}}{\gamma_1}\frac{(1/p)\mathbf{C}(\mathbf{Y\Sigma Y}^* - \mathbf{Y_l\Sigma Y_l}^*)\mathbf{C}^*}{\gamma_{1,l}}. \end{aligned} \tag{2.44}$$

From (2.39) we have $\lim_{T\to\infty}\frac{\gamma_{1,l}}{\gamma_1} = 1$. This, together with (2.35), Proposition 4, (2.38), (2.42), (2.43) and Lemma 5 in the supplementary file, implies (2.11).

We next prove the CLT. In fact we just need to prove

$$\left\|\frac{1}{p}\mathbf{CY\Sigma}^{1/2}\mathbf{X_0}^*\right\|_2 = o_p(p^{-1/2}T^2). \tag{2.45}$$

Note that Equation (2.43) implies that $\|\frac{1}{p}\mathbf{CY\Sigma}^{1/2}\mathbf{X_0}^*\|_2 = O_p(T^{3/2})$. Remark 2 then follows.

The assumption A7 implies that

$$\|\frac{1}{p}\mathbf{X_0X_0}^*\|_2 = O_p(T). \tag{2.46}$$

Our aim is to prove (2.45). Note that $rank(\mathbf{C}\mathbf{Y}\boldsymbol{\Sigma}^{1/2}\mathbf{X_0}^*) = 1$. Recalling Assumption A7, we can then find

$$\left\|\frac{1}{p}\mathbf{C}\mathbf{Y}\boldsymbol{\Sigma}^{1/2}\mathbf{X_0}^*\right\|_2 = \frac{\sqrt{T}}{p}\sqrt{\sum_{t=1}^{T}\left(\sum_{i=1}^{t}\mathbf{y_i'}\boldsymbol{\Sigma}^{1/2}\mathbf{x_0}\right)^2}, \tag{2.47}$$

$$\sum_{i=1}^{t}\mathbf{y_i'}\boldsymbol{\Sigma}^{1/2}\mathbf{x_0} = \sum_{i=1}^{t}\mathbf{y_i'}\boldsymbol{\Sigma}^{1/2}\sum_{k=0}^{\infty}\tilde{b}_k\boldsymbol{\Sigma_1}^{1/2}\mathbf{z_{-k}} + \sum_{i=1}^{t}\mathbf{y_i'}\boldsymbol{\Sigma}^{1/2}\tilde{b}_{-1}\boldsymbol{\Sigma_2}^{1/2}\tilde{\mathbf{z}}$$
$$+\sum_{i=1}^{t}\mathbf{y_i'}\boldsymbol{\Sigma}^{1/2}\tilde{\mathbf{b}}_{-2}. \tag{2.48}$$

By (2.1) and a variable change we may write

$$\sum_{i=1}^{t}\mathbf{y_i'} = \sum_{j=1}^{t}\mathbf{z_j'}\left(\sum_{i=j}^{t}b_{i-j}\right) + \sum_{j=-\infty}^{0}\mathbf{z_j'}\left(\sum_{i=1}^{t}b_{i-j}\right). \tag{2.49}$$

Let $(\tilde{c}_{-2,1},\ldots,\tilde{c}_{-2,p})' = \tilde{\mathbf{c}}_{-2} = \boldsymbol{\Sigma}^{1/2}\tilde{\mathbf{b}}_{-2}$. Assumptions A3 and A7 imply $\|\tilde{\mathbf{c}}_{-2}\|^2 = O(p)$. Then

$$\sum_{i=1}^{t}\mathbf{y_i'}\boldsymbol{\Sigma}^{1/2}\tilde{\mathbf{b}}_{-2} = \sum_{i=1}^{t}\mathbf{y_i'}\tilde{\mathbf{c}}_{-2}.$$

It follows that

$$E\left(\sum_{i=1}^{t}\mathbf{y_i'}\boldsymbol{\Sigma}^{1/2}\tilde{\mathbf{b}}_{-2}\right) = 0 \tag{2.50}$$

and

$$Var\left(\sum_{i=1}^{t}\mathbf{y_i'}\boldsymbol{\Sigma}^{1/2}\tilde{\mathbf{b}}_{-2}\right) = \|\tilde{\mathbf{c}}_{-2}\|^2\left(\sum_{j=1}^{t}\left(\sum_{i=j}^{t}b_{i-j}\right)^2 + \sum_{j=-\infty}^{0}\left(\sum_{i=1}^{t}b_{i-j}\right)^2\right) = O(pt), \tag{2.51}$$

which imply

$$\sum_{i=1}^{t}\mathbf{y_i'}\boldsymbol{\Sigma}^{1/2}\tilde{\mathbf{b}}_{-2} = O_p(p^{1/2}t^{1/2}). \tag{2.52}$$

As in (2.49), write

$$\sum_{i=1}^{t}\mathbf{y_i'}\boldsymbol{\Sigma}^{1/2}\tilde{b}_{-1}\boldsymbol{\Sigma_2}^{1/2}\tilde{\mathbf{z}} = \tilde{b}_{-1}\left(\sum_{j=1}^{t}\mathbf{z_j'}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma_2}^{1/2}\tilde{\mathbf{z}}\left(\sum_{i=j}^{t}b_{i-j}\right) + \sum_{j=-\infty}^{0}\mathbf{z_j'}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma_2}^{1/2}\tilde{\mathbf{z}}\left(\sum_{i=1}^{t}b_{i-j}\right)\right).$$

Assumption A7 implies that $\tilde{\mathbf{z}}$ is independent of $\mathbf{z_t}$ and that $\tilde{b}_{-1}$ is bounded. It follows that

$$\sum_{i=1}^{t} \mathbf{y_i'}\mathbf{\Sigma}^{1/2}\tilde{b}_{-1}\mathbf{\Sigma_2}^{1/2}\tilde{\mathbf{z}} = O_p(p^{1/2}t^{1/2}). \tag{2.53}$$

Now we consider the first term of the right hand of (2.48). From (2.49), write

$$\begin{aligned}
&\sum_{i=1}^{t} \mathbf{y_i'}\mathbf{\Sigma}^{1/2}\sum_{k=0}^{\infty}\tilde{b}_k\mathbf{\Sigma_1}^{1/2}\mathbf{z_{-k}} \\
&= \sum_{j=1}^{t}\sum_{k=0}^{\infty}\mathbf{z_j'}\mathbf{\Sigma}^{1/2}\mathbf{\Sigma_1}^{1/2}\mathbf{z_{-k}}\tilde{b}_k\left(\sum_{i=j}^{t} b_{i-j}\right) + \sum_{j=-\infty}^{0}\sum_{k=0}^{\infty}\mathbf{z_j'}\mathbf{\Sigma}^{1/2}\mathbf{\Sigma_1}^{1/2}\mathbf{z_{-k}}\tilde{b}_k\left(\sum_{i=1}^{t} b_{i-j}\right).
\end{aligned}$$

Direct calculations imply

$$E\left(\sum_{i=1}^{t} \mathbf{y_i'}\mathbf{\Sigma}^{1/2}\sum_{k=0}^{\infty}\tilde{b}_k\mathbf{\Sigma_1}^{1/2}\mathbf{z_{-k}}\right) = \sum_{k=0}^{\infty}\mathrm{tr}\left(\mathbf{\Sigma}^{1/2}\mathbf{\Sigma_1}^{1/2}\right)\tilde{b}_k\left(\sum_{i=1}^{t} b_{i+k}\right) = O(p) \tag{2.54}$$

and

$$Var\left(\sum_{i=1}^{t} \mathbf{y_i'}\mathbf{\Sigma}^{1/2}\sum_{k=0}^{\infty}\tilde{b}_k\mathbf{\Sigma_1}^{1/2}\mathbf{z_{-k}}\right) = O(pt). \tag{2.55}$$

Equations (2.52)–(2.55) and Assumption A4 imply

$$\|\frac{1}{p}\mathbf{CY\Sigma}^{1/2}\mathbf{X_0}^*\|_2 = O_p(\max(p^{-1/2}T^{3/2}, T)) = o_p(p^{-1/2}T^2). \tag{2.56}$$

The proof of Theorem 2 is completed. □

Proof of Theorem 1. Define $\mathbf{X_{0\Pi}} = (\mathbf{\Pi x_0}, \cdots, \mathbf{\Pi^T x_0})'$ and $\mathbf{X_{1\Pi}} = \mathbf{X} - \mathbf{X_{0\Pi}}$. Write

$$\begin{aligned}
\mathbf{B} &= (1/p)\mathbf{XX}^* \qquad (2.57)\\
&= (1/p)\mathbf{X_{1\Pi}X_{1\Pi}}^* + (1/p)\mathbf{X_{1\Pi}X_{0\Pi}}^* + (1/p)\mathbf{X_{0\Pi}X_{1\Pi}}^* + (1/p)\mathbf{X_{0\Pi}X_{0\Pi}}^*.
\end{aligned}$$

Observe that

$$\|(1/p)\mathbf{X_{0\Pi}}^*\mathbf{X_{0\Pi}}\|_2 = \|(1/p)\sum_{t=1}^{T}\mathbf{\Pi}^t\mathbf{x_0}\mathbf{x_0'}\mathbf{\Pi'}^t\|_2 \le \frac{1}{p(1-\varphi^2)}\|\mathbf{x_0}\|^2. \tag{2.58}$$

This, together with Assumption A6, implies

$$\|(1/p)\mathbf{X_{0\Pi}}^*\mathbf{X_{0\Pi}}\|_2 = O_p(1). \tag{2.59}$$

Recalling (2.79), we have

$$\|(1/p)\mathbf{X_{1\Pi}}^*\mathbf{X_{1\Pi}}\|_2 \leq \frac{M_0}{(1-\varphi)^2}\|(1/p)\mathbf{Y}^*\mathbf{Y}\|_2.$$

We then conclude from (2.79), (2.80) and (2.41) that

$$\lim_{T\to\infty} P\left(\|(1/p)\mathbf{X_{1\Pi}}^*\mathbf{X_{1\Pi}}\|_2 \leq \frac{8\sum_{i\geq 0}|a_i|}{(1-\varphi)^2}M_0\left(1+\sqrt{\frac{T}{p}}\right)^2\right) = 1. \tag{2.60}$$

By Holder's inequality

$$\|(1/p)\mathbf{X_{0\Pi}}\mathbf{X_{1\Pi}}^*\|_2 \leq \sqrt{\|(1/p)\mathbf{X_{0\Pi}}^*\mathbf{X_{0\Pi}}\|_2\|(1/p)\mathbf{X_{1\Pi}}^*\mathbf{X_{1\Pi}}\|_2}. \tag{2.61}$$

Thus, equations (2.59)-(2.61) ensure Theorem 1.

□

The proof of Theorem 3 is simple since $\bar{\mathbf{B}} = \mathbf{HBH}$. So $\|\bar{\mathbf{B}}\|_2 \leq \|\mathbf{B}\|_2$ since $\|\mathbf{H}\|_2 = 1$.

Theorem 4 are similar to Theorem 2. We only need to replace the results of Appendix A.2 by those in Appendix A.3. Note that we don't need to prove (2.45) since (2.36) implies that $\mathbf{x_0}$ doesn't affect $\bar{\mathbf{B}}$.

Proof of Theorem 5. At first we prove that the error of the estimator μ_{m_1} is $o(p^{-1/2})$. Let $m_1 = [\sqrt{p}]$. From (2.5) and (2.40) we have

$$\begin{aligned}&\left|\left(a_0 + 2\sum_{1\leq j\leq m_1} a_j(-1)^j\cos(j\theta_1)\right) - \left(a_0 + 2\sum_{1\leq j\leq\infty} a_j(-1)^j\cos(j\theta_1)\right)\right|\\ &\leq 2\sum_{1+m_1\leq j\leq\infty}|a_j| = o(p^{-1/2})\end{aligned} \tag{2.62}$$

and

$$\begin{aligned}&\left|\left(a_0 + 2\sum_{1\leq j\leq m_1} a_j(-1)^j\cos(j\theta_1)) - (a_0 + 2\sum_{1\leq j\leq m_1} a_j\right)\right|\\ &\leq 2\sum_{1\leq j\leq m_1}|a_j|\left(1-\cos\frac{j\pi}{2T+1}\right) = O(p^{1/2}T^{-2}) = o(p^{-1/2}).\end{aligned} \tag{2.63}$$

In view of (2.6) it suffices to prove

$$|\mu_{m_1} - (a_0 + 2\sum_{1\le j\le m_1} a_j)\frac{tr(\mathbf{\Sigma})}{p}| = o_p(p^{-1/2}). \tag{2.64}$$

A direct calculation shows the following mean and variance

$$E\mu_{m_1} - \left(a_0 + 2\sum_{1\le j\le m_1} a_j\right)\frac{tr(\mathbf{\Sigma})}{p} = 0,$$

$$Var\left(\sum_{1\le j\le m_1}\frac{1}{T-j-1}\sum_{2\le i\le T-j}\frac{\mathbf{y_i'\Sigma y_{i+j}}}{p}\right)$$
$$= \sum_{1\le i,j\le m_1}\sum_{2\le f\le T-i}\sum_{2\le g\le T-j}\frac{Cov(\frac{\mathbf{y_f'\Sigma y_{f+i}}}{p},\frac{\mathbf{y_g'\Sigma y_{g+j}}}{p})}{(T-i-1)(T-j-1)}. \tag{2.65}$$

Moreover, we have

$$Cov\left(\frac{\mathbf{y_f'\Sigma y_{f+i}}}{p},\frac{\mathbf{y_g'\Sigma y_{g+j}}}{p}\right) = \frac{1}{p}\left[\frac{\sum_{i=1}^{p}\Sigma_{ii}^2}{p}E|Z_{ij}|^4\sum_{k=0}^{\infty}b_kb_{k+i}b_{k+g-f}b_{k+g-f+j}1_{(k+g-f\ge 0)}\right.$$
$$\left.+\frac{tr(\mathbf{\Sigma}^2)}{p}E|Z_{ij}|^2(a_{|f-g|}a_{|f+i-g-j|}+a_{|f+i-g|}a_{|f-g-j|})\right].$$

From the above, Assumption (A1) and (2.40), we conclude that

$$Var\left(\mu_{m_1}\right) = O(p^{-1}m_1T^{-1}) = o(p^{-1}). \tag{2.66}$$

Then (2.65) and (2.66) imply (2.64).

Now we prove

$$\frac{\sqrt{\frac{|S_{\sigma^2,0,0}|}{p}}}{\sum_{i=2}^{T}\frac{\check{\mathbf{x}}_{i,i}}{p(T-1)}} - \frac{\sqrt{\frac{tr(\mathbf{\Sigma^2})}{p}}}{\frac{tr(\Sigma)}{p}} \xrightarrow{i.p.} 0.$$

Let $\tilde{S}_{\sigma^2,0,0} = S_{\sigma^2,0,0} - a_0^2tr(\mathbf{\Sigma^2})$. It is then sufficient to show that

$$\frac{\tilde{S}_{\sigma^2,0,0}}{a_0^2tr(\mathbf{\Sigma^2})} = o_p(1).$$

From Assumptions A2 and A3 we have for large enough T,

$$a_0^2tr(\mathbf{\Sigma^2}) \ge a_0^2M_1^2p, \tag{2.67}$$

where we have used the fact that $\mathrm{tr}(\mathbf{\Sigma^2}) \geq \frac{(tr\mathbf{\Sigma})^2}{p}$. When T is large enough,

$$\left(T - \frac{3}{2}[T/2]\right)([T/2]-1) \geq \frac{T^2}{9}. \tag{2.68}$$

We next expand $\tilde{S}_{\sigma^2,0,0}$ in terms of Z_{ij} and write it a sum of the terms involving the high order of Z_{ij} and the terms involving the low order of Z_{ij}. Specifically, write $\tilde{S}_{\sigma^2,0,0} = \tilde{S}_{\sigma^2,0,0,h} + \tilde{S}_{\sigma^2,0,0,l}$, where

$$\begin{aligned} & \tilde{S}_{\sigma^2,0,0,h} \\ = \quad & \frac{1}{(T-\frac{3}{2}[T/2])([T/2]-1)} \sum_{f=2}^{[T/2]} \sum_{g=f+[T/2]}^{T} \\ & (\sum_{i_1,i_2=1}^{p} \Sigma_{i_1 i_1} \Sigma_{i_1 i_2} \sum_{s_1,s_2=-\infty}^{T} Z_{s_1 i_1}^3 Z_{s_2 i_2} (b_{f-s_1} b_{g-s_1} b_{f+i-s_1} b_{g+j-s_2} \\ & + b_{f-s_1} b_{g-s_1} b_{f+i-s_2} b_{g+j-s_1} + b_{f-s_1} b_{g-s_2} b_{f+i-s_2} b_{g+j-s_1} \\ & + b_{f-s_2} b_{g-s_1} b_{f+i-s_1} b_{g+j-s_1}) - 3 \sum_{i_1=1}^{p} \Sigma_{i_1 i_1}^2 \sum_{s_1=-\infty}^{T} Z_{s_1 i_1}^4 b_{f-s_1} b_{g-s_1} b_{f+i-s_1} b_{g+j-s_1}). \end{aligned} \tag{2.69}$$

Note that $b_k = 0$ when $k < 0$. We can then conclude from Assumption A1 that

$$E|\tilde{S}_{\sigma^2,0,0,h}| = o(p^2 T^{-2}). \tag{2.70}$$

(2.67) and (2.70) imply that

$$\frac{E|\tilde{S}_{\sigma^2,0,0,h}|}{a_0^2 tr(\mathbf{\Sigma^2})} = o(pT^{-2}) = o(1). \tag{2.71}$$

It can be derived that

$$\begin{aligned} & (T - \frac{3}{2}[T/2])([T/2]-1) E\tilde{S}_{\sigma^2,0,0,l} \\ & = \sum_{f=2}^{[T/2]} \sum_{g=f+[T/2]}^{T} (a_{g-f} a_{g-f} tr(\mathbf{\Sigma^2}) + a_{g-f} a_{g-f} (tr(\mathbf{\Sigma}))^2) \\ & = o(p^2 T^{-1}). \end{aligned} \tag{2.72}$$

This, together with (2.67) and (2.68), implies that

$$\frac{E\tilde{S}_{\sigma^2,0,0,l}}{a_0^2 tr(\mathbf{\Sigma^2})} = o(pT^{-3}) = o(1). \tag{2.73}$$

By (2.67), (2.68) and the assumption A1, one can also verify that

$$Var\left(\frac{\tilde{S}_{\sigma^2,0,0,l}}{a_0^2 tr(\mathbf{\Sigma^2})}\right) = o(pT^{-2} + p^{-1}) = o(1). \tag{2.74}$$

This, together with (2.71) and (2.73), shows that

$$\frac{\tilde{S}_{\sigma^2,0,0}}{a_0^2 tr(\mathbf{\Sigma^2})} = o_p(1).$$

This, together with (2.62)-(2.64), implies that $S_{\sigma^2,m_2} - \frac{a_0\sqrt{2tr(\mathbf{\Sigma^2})}}{\sqrt{p}} \xrightarrow{i.p.} 0$ when m_2 tends to infinity. Since the two estimators are available, it's easy to complete the proof with Theorems 2 and 4. □

Proof of Theorem 6. We claim that

$$\sum_{i=2}^{T} \frac{\breve{\mathbf{x}}_{i,i}}{p(T-1)} - \frac{2a_0 tr(\mathbf{\Sigma})}{p(1+\varphi)} \xrightarrow{i.p.} 0 \tag{2.75}$$

and

$$S_{\sigma^2,0} - \frac{2a_0\sqrt{2tr(\mathbf{\Sigma^2})}}{\sqrt{p}(1+\varphi)} \xrightarrow{i.p.} 0. \tag{2.76}$$

Indeed, the proofs of (2.75) and (2.76) are similar to that of Theorem 5 (replacing $m_1 = m_2$ there by 0). Moreover from Theorem 3 we have $\bar{\rho}_1 = o_p(T)$. This, together with (2.75) and (2.76), ensures that

$$\bar{T}_N + \sqrt{\frac{p}{2}}\frac{\frac{tr(\Sigma)}{p}}{\sqrt{\frac{tr(\mathbf{\Sigma^2})}{p}}} \xrightarrow{i.p.} 0, \tag{2.77}$$

which further yields (2.28).

When $\|\phi\|_2 = O(p)$, from Theorem 1 we have $\rho_1 = O_p(T)$. This, together with (2.75) and (2.76), ensures that

$$T_N + \sqrt{\frac{p}{2}}\frac{\frac{tr(\Sigma)}{p}}{\sqrt{\frac{tr(\mathbf{\Sigma^2})}{p}}} \xrightarrow{i.p.} 0, \tag{2.78}$$

which further implies (2.29). □

Appendix C: Proofs of the results in Appendix A

Appendix C.1: Proof of the result in Appendix A.1

Proof of Proposition 3. By (2.2) we may write

$$\mathbf{x_t} = \sum_{k_1=0}^{t-1} \mathbf{\Pi^{k_1} \Sigma^{1/2} y_{t-k_1}}.$$

This, together with (2.3), implies that

$$\begin{aligned}
\frac{1}{p}\mathbf{X}'\mathbf{X} &= \frac{1}{p}\sum_{t=1}^{T} \mathbf{x_t x_t'} = \frac{1}{p}\sum_{t=1}^{T}\sum_{k_1=0}^{t-1}\sum_{k_2=0}^{t-1} \mathbf{\Pi^{k_1}\Sigma^{1/2} y_{t-k_1} y_{t-k_2}' \Sigma^{1/2} \Pi'^{k_2}} \\
&= \frac{1}{p}\sum_{k_1=0}^{T-1}\sum_{k_2=0}^{T-1} \mathbf{\Pi^{k_1}\Sigma^{1/2}} \left(\sum_{t=\max(k_1,k_2)+1}^{T} \mathbf{y_{t-k_1} y_{t-k_2}'} \right) \mathbf{\Sigma^{1/2}\Pi'^{k_2}}.
\end{aligned}$$

Note that

$$\sum_{t=\max(k_1,k_2)+1}^{T} \mathbf{y_{t-k_1} y_{t-k_2}'} = \mathbf{Y}'\mathbf{\tilde{C}_{k_1}'\tilde{C}_{k_2}Y}$$

where $\tilde{C}_k$ is a $T \times T$ matrix with elements $\tilde{C}_{k,ij} = I(i-j=k)$. It's easy to know $\|\tilde{C}_k\|_2 \leq 1$. We then conclude that

$$\begin{aligned}
\|\frac{1}{p}\mathbf{X}^*\mathbf{X}\|_2 &\leq \sum_{k_1=0}^{T-1}\sum_{k_2=0}^{T-1} \varphi^{k_1+k_2} \|\frac{1}{p}\mathbf{Y}^*\mathbf{Y}\|_2 \|\mathbf{\Sigma}\|_2 \leq \frac{M_0}{(1-\varphi)^2}\|\frac{1}{p}\mathbf{Y}^*\mathbf{Y}\|_2 \\
&\leq \frac{M_0}{(1-\varphi)^2}\|\frac{1}{p}\mathbf{Z}^*\mathbf{Z}\|_2\|\mathbf{A}\|_2. \qquad (2.79)
\end{aligned}$$

As we know the matrix $\mathbf{A}$ is a Toeplitz matrix so that $\|\mathbf{A}\|_2 \leq 2\sum_{i\geq 0}|a_i|$ (see Pan et al. (2014)). From Assumption (A4) and the results in Chen and Pan (2012) and Bai and Silverstein (2006) we have

$$\lim_{T\to\infty} P\left(\|\frac{1}{p}ZZ^*\|_2 \leq 4\left(1+\sqrt{\frac{T}{p}}\right)^2 \right) = 1. \qquad (2.80)$$

Proposition then follows from (2.79).

□

Appendix C.2: Proofs of the results in Appendix A.2

We prove the results in Appendix A.2 in this section. We below investigate the eigenvalues and eigenvectors of $\mathbf{CFF}^*\mathbf{C}^* = \mathbf{CAC}^*$ at first.

C.2.1: Eigenvalues of $\mathbf{CAC}^*$

We consider the eigenvalues of $\mathbf{CAC}^*$ in this subsection. These are some crucial steps.

Since it is very hard to find the eigenvalues of $\mathbf{CAC}^*$ directly, we use the following strategy. At first, we note that the eigenvalues of $\mathbf{CAC}^*$ and $\mathbf{AC}^*\mathbf{C}$ are the same. We obtain the eigenvalues and eigenvectors of $\mathbf{C}^*\mathbf{C}$ by first studying $(\mathbf{C}^*\mathbf{C})^{-1}$. The next key step is to construct a matrix $\mathbf{A_m}$ defined in (2.102) in the supplementary file which has the same eigenvectors as $\mathbf{C}^*\mathbf{C}$. In the mean time, it is easy to find the eigenvalues of $\mathbf{A_m}\mathbf{C}^*\mathbf{C}$. We then use the eigenvalues of of $\mathbf{CA_m}\mathbf{C}^*$ to approximate those of $\mathbf{CAC}^*$. Our results are summarized in the following lemmas and theorems.

The first two lemmas describe the eigenvalues of $\mathbf{C}^*\mathbf{C}$ and determine their limits.

Lemma 2. $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_T \geq 0$ *are given in (2.5). One can verify that they are the eigenvalues of* $\mathbf{C}^*\mathbf{C}$.

Proof of Lemma 2.

Let $\mathbf{M_T} = (\mathbf{C}^*\mathbf{C})^{-1}$. Define the characteristic function of $\mathbf{M_T}$ by $g_T(\lambda) = \det(\lambda \mathbf{I_T} - \mathbf{M_T})$. We can verify that the entries of the inverse matrix $\mathbf{C}^{-1}$, a $T \times T$ lower triangular matrix, are as following

$$C_{ij}^{-1} = \begin{cases} 1 & i = j, \\ -1 & i = j+1, \\ 0 & \text{otherwise.} \end{cases} \tag{2.81}$$

It follows that $M_{i,j}$, the elements of $\mathbf{M_T} = (\mathbf{C}^*\mathbf{C})^{-1}$, satisfy

$$M_{ij} = \begin{cases} 1 & i = j = 1, \\ 2 & i = j > 1, \\ -1 & |i-j| = 1, \\ 0 & \text{otherwise}. \end{cases} \tag{2.82}$$

By the cofactor expansion we obtain a recurrence relation as following

$$g_T(\lambda) = (\lambda - 2)g_{T-1}(\lambda) - g_{T-2}(\lambda). \tag{2.83}$$

Consider $\lambda \in (0, 4)$ at first. Hence we may write $\lambda = \lambda(\theta) = 2 + 2\cos\theta$. We can further solve (2.83) to get

$$g_T(\lambda) = \frac{\sin T\theta + \sin(T+1)\theta}{\sin\theta}. \tag{2.84}$$

When $\sin\theta \neq 0$, $g_T(\lambda) = 0$ is equivalent to

$$\sin T\theta + \sin(T+1)\theta = 0. \tag{2.85}$$

Let $h_T(\theta) = \sin T\theta + \sin(T+1)\theta = 2\sin(T+1/2)\theta cos\frac{\theta}{2}$. Note that (2.5) gives T different solutions which satisfy $h_T(\theta) = 0$ and $\sin\theta \neq 0$. On the other hand, observe that there are at most T solutions for $g_T(\lambda) = 0$. The proof is complete. □

Lemma 3. *Using the notation in (2.5),*

$$\lim_{T\to\infty} \frac{\lambda_k}{T^2} = \frac{4}{\pi^2(2k-1)^2} \tag{2.86}$$

for any fixed k.

Lemma 4 below specifies the eigenvectors of $\mathbf{C}^*\mathbf{C}$.

Lemma 4. *Let $\tilde{\mathbf{x}}_\mathbf{k} = (x_{k,1}, \cdots, x_{k,T})'$ be a $T \times 1$ vector with*

$$x_{k,i} = (-1)^{T-i} \sin(T - i + 1)\theta_k, \ -l \le i \le T + l. \tag{2.87}$$

Then $\{\tilde{\mathbf{x}}_{\mathbf{k}}, 1 \le k \le T\}$ *are orthogonal and satisfy for any* k

$$\mathbf{C}^*\mathbf{C}\tilde{\mathbf{x}}_{\mathbf{k}} = \lambda_k \tilde{\mathbf{x}}_{\mathbf{k}}. \tag{2.88}$$

Lemmas 3 and 4 can be verified with some straightforward computations and a simple fact that

$$\sin(k+j)\theta + \sin(k-j)\theta = 2\sin k\theta \cos j\theta. \tag{2.89}$$

We omit the details here.

Lemma 5 below specifies the eigenvalues of $\mathbf{A_m}\mathbf{C}^*\mathbf{C}$ and gives their approximation to those of $\mathbf{A}\mathbf{C}^*\mathbf{C}$.

Lemma 5. *Define* γ_k *by*

$$\gamma_k = \lambda_k \left(a_0 + 2 \sum_{1 \le j \le T-1} a_j (-1)^j \cos(j\theta_k) \right). \tag{2.90}$$

For any fixed constant $k \ge 1$*, there is a constant* c_k *such that*

$$\lim_{T\to\infty} \frac{\gamma_k}{T^2} = c_k > 0 \tag{2.91}$$

and

$$\lim_{T\to\infty} \frac{\gamma_k}{\gamma_1} = \lim_{T\to\infty} \frac{\lambda_k}{\lambda_1} = \frac{1}{(2k-1)^2}. \tag{2.92}$$

Let $\beta_1 \ge \beta_2 \ge \cdots \ge \beta_T$ *be the eigenvalues of* $\mathbf{A}\mathbf{C}^*\mathbf{C}$*. If* $\mathbf{A}$ *satisfies Assumptions (A1) and (A2), then for any fixed integers* $i \ge 1$ *and* $j \ge 1$*, the following equation holds*

$$\left| \frac{\beta_i - \gamma_i}{\gamma_j} \right| = O(T^{-1}). \tag{2.93}$$

For any $\epsilon > 0$ *there exists* T_0 *and* k_0 *where* k_0 *is a fixed number independent of* T *such that when* $T \ge T_0$ *and* $k \ge k_0$*,*

$$\left| \frac{\beta_k}{\gamma_1} \right| \le \epsilon. \tag{2.94}$$

Proof of Lemma 5.

Let us prove (2.91) and (2.92) first. Note that

$$\left| \left(a_0 + 2 \sum_{1 \le j \le T-1} a_j (-1)^j \cos(j\theta_k) \right) - \left(a_0 + 2 \sum_{1 \le j \le \infty} a_j \right) \right|$$
$$\le 2 \sum_{1 \le j \le T-1} |a_j| \left| cos(\frac{j(2k-1)\pi}{2T+1}) - 1 \right| + 2 \sum_{T \le j} |a_j|.$$

For a fixed k, we can find a j_k to satisfy $\frac{\pi}{3} \le \frac{j_k(2k-1)\pi}{2T+1} \le \frac{\pi}{2}$. It follows that

$$2 \sum_{1 \le j \le j_k} |a_j| |cos(\frac{j(2k-1)\pi}{2T+1}) - 1| \le 2 \sum_{1 \le j \le j_k} |a_j| (\frac{j(2k-1)\pi}{2T+1})^2$$
$$\le \frac{2j_k(2k-1)^2\pi^2}{(2T+1)^2} \sum_{1 \le j \le j_k} j|a_j| \le \frac{(2k-1)\pi^2}{(2T+1)} \sum_{1 \le j \le \infty} j|a_j|$$

and that

$$2 \sum_{j_k < j \le T-1} |a_j| |cos(\frac{j(2k-1)\pi}{2T+1} - 1)| + 2 \sum_{T \le j} |a_j| \le 4 \sum_{j \ge j_k} |a_j|$$
$$\le \quad j_k^{-1} 4 \sum_{j \ge j_k} j|a_j| \le \frac{3(2k-1)}{2T+1} \sum_{1 \le j \le \infty} j|a_j|.$$

From the assumption (A2), (2.40) and truncation conditions we can find

$$\lim_{T \to \infty} \left(a_0 + 2 \sum_{1 \le j \le T-1} a_j (-1)^j \cos(j\theta_k) \right) = \lim_{T \to \infty} \left(a_0 + 2 \sum_{1 \le j \le \infty} a_j \right) = \left(\sum_{i=0}^{\infty} b_i \right)^2 = s^2 > 0. \tag{2.95}$$

In view of (2.86), (2.90) and (2.95), we can prove (2.91) and (2.92).

Now, we consider the eigenvalues of $\mathbf{AC^*C}$. From (2.5)

$$\sin(T-i)\theta_k = -\sin(T+i+1)\theta_k. \tag{2.96}$$

In view of (2.87) and (2.96), we obtain

$$x_{k,i} = x_{k,1-i}, \ -T \le i \le 0 \tag{2.97}$$

and

$$x_{k,i} = -x_{k,2T+2-i}, \ T+2 \le i \le 2T. \tag{2.98}$$

We construct a new matrix $\mathbf{A_m}$ whose the s–th row, $\mathbf{a_{m,s}}$, satisfies

$$\mathbf{a_{m,s}}\tilde{\mathbf{x}}_\mathbf{k} = a_0 x_{k,s} + \sum_{1\le j\le T-1} a_j(x_{k,s-j}+x_{k,s+j}) = \left(a_0 + 2\sum_{1\le j\le T-1} a_j(-1)^j cosj\theta_k\right) x_{k,s}. \tag{2.99}$$

Let $\mathbf{a_s}$ be the sth row of $\mathbf{A}$. We can find that

$$\mathbf{a_s}\tilde{\mathbf{x}}_\mathbf{k} = a_0 x_{k,s} + \sum_{1\le j\le s-1} a_j x_{k,s-j} + \sum_{1\le j\le T-s} a_j x_{k,s+j}. \tag{2.100}$$

We further define $T\times T$ matrix $\mathbf{A_l}$ by

$$(\mathbf{A_l})_{ij} = \begin{cases} a_{i+j-1} & i+j\le T, \\ -a_{2T-i-j+2} & i+j\ge T+3, \\ 0 & T+1\le i+j\le T+2. \end{cases} \tag{2.101}$$

Let

$$\mathbf{A_m} = \mathbf{A} + \mathbf{A_l}. \tag{2.102}$$

One can verify

$$\mathbf{A_m}\tilde{\mathbf{x}}_\mathbf{k} = \left(a_0 + 2\sum_{1\le j\le T-1} a_j(-1)^j cosj\theta_k\right)\tilde{\mathbf{x}}_\mathbf{k}. \tag{2.103}$$

It follows that

$$\mathbf{A_m}\mathbf{C}^*\mathbf{C}\tilde{\mathbf{x}}_\mathbf{k} = \lambda_k\mathbf{A_m}\tilde{\mathbf{x}}_\mathbf{k} = \gamma_k\tilde{\mathbf{x}}_\mathbf{k}, \tag{2.104}$$

which implies that γ_k is the eigenvalues of $\mathbf{A_m}\mathbf{C}^*\mathbf{C}$.

Now we consider $\mathbf{C}\mathbf{A_l}\mathbf{C}^*$. It is easily seen that,

$$\|\mathbf{C}\mathbf{A_l}\mathbf{C}^*\|_2 \le T\max_{i,j}\{|(\mathbf{C}\mathbf{A_l}\mathbf{C}^*)_{i,j}|\}. \tag{2.105}$$

Recalling (2.101) we can find that

$$\max_{i,j}\{|(\mathbf{C}\mathbf{A_l}\mathbf{C}^*)_{i,j}|\} \le 2\sum_{i=1}^{T-1} i|a_i|. \tag{2.106}$$

We conclude from (2.105) that

$$\|\mathbf{C}\mathbf{A_l}\mathbf{C}^*\|_2 \leq 2\sum_{i=0}^{T-1} i|a_i|T.$$

In view of (2.40), we have

$$\|\mathbf{C}\mathbf{A_l}\mathbf{C}^*\|_2 \leq 2\sum_{i=0}^{T-1} i|a_i|T = O(T). \tag{2.107}$$

Let $\tilde{\gamma}_1 \geq \tilde{\gamma}_2 \geq \cdots \geq \tilde{\gamma}_T$ be the eigenvalues of $\mathbf{C}\mathbf{A_m}\mathbf{C}^*$. For fixed integers i, β_i is the ith largest eigenvalue of $\mathbf{CAC}^*$. It follows that

$$\left|\frac{\beta_i - \tilde{\gamma}_i}{\gamma_j}\right| \leq \frac{\|\mathbf{C}(\mathbf{A_m} - \mathbf{A})\mathbf{C}^*\|_2}{\gamma_j} = \frac{\|\mathbf{C}\mathbf{A_l}\mathbf{C}^*\|_2}{\gamma_j}. \tag{2.108}$$

From (2.90) and (2.92) we can find T_i for any fixed i such that when $T > T_i$, $\tilde{\gamma}_i = \gamma_i$. By (2.91) and (2.107) we can prove (2.93).

(2.94) follows from Lemma 7 directly.

□

Lemma 6. *Suppose that* $\mathbf{A}$ *satisfies Assumptions (A1) and (A2). Then*

$$\mathrm{tr}\,(\mathbf{A}\mathbf{C}^*\mathbf{C}) = a_0\frac{(T+1)T}{2} + \sum_{1\leq j\leq T-1} a_j(T-j+1)(T-j) \tag{2.109}$$

and

$$\lim_{T\to\infty}\frac{\beta_k}{\mathrm{tr}\,(\mathbf{A}\mathbf{C}^*\mathbf{C})} = \lim_{T\to\infty}\frac{\gamma_k}{tr(\mathbf{A}\mathbf{C}^*\mathbf{C})} = \frac{8}{\pi^2(2k-1)^2}. \tag{2.110}$$

Proof of Lemma 6.

One can verify (2.109) with some computation. Observe that

$$\left|a_0 + \sum_{1\leq j\leq T-1} a_j\frac{(T-j+1)(T-j)}{\frac{(T+1)T}{2}} - (a_0 + 2\sum_{1\leq j\leq\infty} a_j)\right| = O(T^{-1}).$$

This, together with (2.86), (2.90), (2.95), (2.93) and the assumption (A2), implies (2.110). □

Lemma 7. *Suppose that* $\mathbf{A}$ *satisfies the assumptions (A1) and (A2). For any* $\epsilon > 0$*, we can find* T_0 *and* k_0*, where* k_0 *is a finite number independent of* T*, such that when* $T \geq T_0$*,*

$$\left| \frac{\sum_{k>k_0} \beta_k}{\gamma_1} \right| < \epsilon. \tag{2.111}$$

Proof of Lemma 7. Observe that

$$\sum_{k=1}^{\infty} \frac{1}{(2k-1)^2} = \frac{3}{4} \sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{8}. \tag{2.112}$$

For any $\epsilon > 0$, we can find k_0 such that

$$\left| \sum_{k=1}^{k_0} \frac{1}{(2k-1)^2} - \frac{\pi^2}{8} \right| < \frac{\epsilon}{3}. \tag{2.113}$$

From (2.92), (2.93) and (2.110), we can also find T_0 such that when $T \geq T_0$,

$$\left| \sum_{k=1}^{k_0} \frac{1}{(2k-1)^2} - \frac{\sum_{k=1}^{k_0} \beta_k}{\gamma_1} \right| < \frac{\epsilon}{3}. \tag{2.114}$$

and

$$\left| \frac{tr(\mathbf{A}\mathbf{C}^*\mathbf{C})}{\gamma_1} - \frac{\pi^2}{8} \right| < \frac{\epsilon}{3}. \tag{2.115}$$

It follows from (2.113)-(2.115) that

$$\left| \frac{\sum_{k>k_0} \beta_k}{\gamma_1} \right| = \left| \frac{tr(\mathbf{A}\mathbf{C}^*\mathbf{C})}{\gamma_1} - \frac{\sum_{k=1}^{k_0} \beta_k}{\gamma_1} \right| < \epsilon. \tag{2.116}$$

□

C.2.2:Eigenvectors of $\mathbf{CAC}^*$

This section is to investigate the eigenvectors of $\mathbf{CAC}^*$. At first we normalize $\{\tilde{\mathbf{x}}_\mathbf{k}\}_{1\leq k\leq T}$ to get $\{\tilde{\mathbf{y}}_\mathbf{k}\}_{1\leq k\leq T}$. Then we study the eigenvectors of of $\mathbf{AC}^*\mathbf{C}$ by representing them with $\{\tilde{\mathbf{y}}_\mathbf{k}\}_{1\leq k\leq T}$. At last we give some results about the eigenvectors of $\mathbf{CAC}^*$, which are necessary for the proof in Section 3.4. Our results are as follows.

Lemma 8. *Recall the eigenvectors* $\tilde{\mathbf{x}}_{\mathbf{k}}$ *defined in Lemma 4. Then*

$$\Sigma_{j=1}^{T}(x_{k,j})^2 = \frac{2T+1}{4}. \tag{2.117}$$

Let

$$\tilde{\mathbf{y}}_{\mathbf{k}} = \frac{\tilde{\mathbf{x}}_{\mathbf{k}}}{\|\tilde{\mathbf{x}}_{\mathbf{k}}\|}. \tag{2.118}$$

Then $\{\tilde{\mathbf{y}}_{\mathbf{k}}\}_{1\le k\le T}$ *are orthogonal and the* j*th element of* $\tilde{\mathbf{y}}_{\mathbf{k}}$, $y_{k,j}$, *satisfies*

$$|y_{k,j}| = \frac{|x_{k,j}|}{\sqrt{\frac{2T+1}{4}}} \le \frac{2}{\sqrt{2T+1}}. \tag{2.119}$$

Proof of Lemma 8. From (2.87) we obtain

$$|x_{k,j}| \le 1.$$

Lemma 8 can be then proved with some straightforward computations. We ignore details here. □

Lemma 9. *Let* $\{\mathbf{u}_{\mathbf{k}}\}_{1\le k\le T}$ *be orthogonal and real vectors such that* $\|\mathbf{u}_k\| = 1$ *and*

$$\mathbf{CAC}^*\mathbf{u}_{\mathbf{k}} = \beta_k \mathbf{u}_{\mathbf{k}}. \tag{2.120}$$

Define $\mathbf{f}_{\mathbf{k}} = \frac{\mathbf{C}^{-1}\mathbf{u}_{\mathbf{k}}}{\|\mathbf{C}^{-1}\mathbf{u}_{\mathbf{k}}\|}$ *such that*

$$\mathbf{f}_{\mathbf{k}} = \Sigma_{j=1}^{T}\alpha_{kj}\mathbf{y}_{\mathbf{j}} \tag{2.121}$$

with

$$\Sigma_{j=1}^{T}\alpha_{kj}^2 = 1. \tag{2.122}$$

Then when $k \ge 1$ *is fixed,*

$$\frac{\alpha_{kk}^2\lambda_k}{\Sigma_{j=1}^{T}\alpha_{kj}^2\lambda_j} = 1 + O(T^{-1}), \tag{2.123}$$

where $\{\lambda_j\}$ *are given in Lemma 2.*

Proof of Lemma 9. From $\mathbf{f_k} = \frac{\mathbf{C^{-1}u_k}}{\|\mathbf{C^{-1}u_k}\|}$ and (2.120), we have $\|\mathbf{f_k}\| = 1$ and

$$\mathbf{AC^*Cf_k} = \beta_k \mathbf{f_k}. \tag{2.124}$$

From (2.102) and (2.124), we have

$$\beta_k = \frac{\mathbf{f_k^*C^*CAC^*Cf_k}}{\|\mathbf{Cf_k}\|^2} = \frac{\mathbf{f_k^*C^*C(A_m - A_l)C^*Cf_k}}{\|\mathbf{Cf_k}\|^2}.$$

It follows that

$$\frac{|\mathbf{f_k^*C^*CA_mC^*Cf_k}| - |\mathbf{f_k^*C^*CA_lC^*Cf_k}|}{\|\mathbf{Cf_k}\|^2} \le \beta_k \le \frac{|\mathbf{f_k^*C^*CA_mC^*Cf_k}| + |\mathbf{f_k^*C^*CA_lC^*Cf_k}|}{\|\mathbf{Cf_k}\|^2}. \tag{2.125}$$

By (2.88), (2.118) and (2.121), we have

$$\|\mathbf{Cf_k}\| = \sqrt{\Sigma_{j=1}^T \alpha_{kj}^2 \lambda_j}. \tag{2.126}$$

Equations (2.88), (2.104), (2.118) and (2.121) imply that

$$\mathbf{C^*CA_mC^*Cf_k} = \mathbf{C^*C}\Sigma_{j=1}^T \alpha_{kj} \mathbf{A_m C^* C\tilde{y}_j} = \mathbf{C^*C}\Sigma_{j=1}^T \alpha_{kj}\gamma_j \mathbf{\tilde{y}_j} = \Sigma_{j=1}^T \alpha_{kj}\gamma_j\lambda_j \mathbf{\tilde{y}_j}.$$

This and (2.121) ensure

$$\mathbf{f_k^*C^*CA_mC^*Cf_k} = \Sigma_{j=1}^T \alpha_{kj}^2 \gamma_j \lambda_j. \tag{2.127}$$

From (2.107), we have

$$\frac{|\mathbf{f_k^*C^*CA_lC^*Cf_k}|}{\|\mathbf{Cf_k}\|^2} \le \|\mathbf{CA_lC^*}\|_2 = O(T).$$

This, together with (2.125)-(2.127), implies that

$$\frac{\Sigma_{j=1}^T \alpha_{kj}^2 \gamma_j \lambda_j}{\Sigma_{j=1}^T \alpha_{kj}^2 \lambda_j} - O(T) \le \beta_k \le \frac{\Sigma_{j=1}^T \alpha_{kj}^2 \gamma_j \lambda_j}{\Sigma_{j=1}^T \alpha_{kj}^2 \lambda_j} + O(T).$$

By Lemma 5, for any fixed k we have

$$\Sigma_{j=1}^T \frac{\alpha_{kj}^2 \lambda_j}{\Sigma_{j=1}^T \alpha_{kj}^2 \lambda_j} \frac{\gamma_j}{\beta_k} - O(T^{-1}) \le 1 \le \Sigma_{j=1}^T \frac{\alpha_{kj}^2 \lambda_j}{\Sigma_{j=1}^T \alpha_{kj}^2 \lambda_j} \frac{\gamma_j}{\beta_k} + O(T^{-1}). \tag{2.128}$$

Note that $\{\mathbf{u_k}\}_{1\le k\le T}$ are orthogonal and $\{\tilde{\mathbf{y}}_\mathbf{k}\}_{1\le k\le T}$ are orthogonal. When $k \neq m$, from (2.88), (2.118) and (2.121) we have

$$0 = U_k^* U_m = \frac{\mathbf{f_k^* C^* C f_m}}{\|\mathbf{Cf_k}\|\|\mathbf{Cf_m}\|} = \frac{\Sigma_{j=1}^T \alpha_{kj}\alpha_{mj}\lambda_j}{\|\mathbf{Cf_k}\|\|\mathbf{Cf_m}\|}.$$

This implies that

$$\Sigma_{j=1}^T \alpha_{kj}\alpha_{mj}\lambda_j = 0. \tag{2.129}$$

Moreover let $\upsilon_{kj} = \frac{\alpha_{kj}\sqrt{\lambda_j}}{\sqrt{\Sigma_{j=1}^T \alpha_{kj}^2 \lambda_j}}$. We have

$$\Sigma_{j=1}^T \upsilon_{kj}^2 = 1. \tag{2.130}$$

Note that (2.128) is equivalent to

$$\Sigma_{j=1}^T \upsilon_{kj}^2 \frac{\gamma_j}{\beta_k} - O(T^{-1}) \le 1 \le \Sigma_{j=1}^T \upsilon_{kj}^2 \frac{\gamma_j}{\beta_k} + O(T^{-1}). \tag{2.131}$$

Also (2.129) implies that

$$\Sigma_{j=1}^T \upsilon_{kj}\upsilon_{mj} = 0. \tag{2.132}$$

We consider υ_{kj} for fixed k below. When $k=1$ and T is big enough, Lemma 5, (2.130) and (2.131) imply

$$O(T^{-1}) = \left|1 - \Sigma_{j=1}^T \upsilon_{1j}^2 \frac{\gamma_j}{\beta_1}\right| \ge (1-\upsilon_{11}^2)\frac{\beta_1-\gamma_2}{\beta_1} - \upsilon_{11}^2 \frac{|\beta_1-\gamma_1|}{\beta_1}. \tag{2.133}$$

In view of (2.91)-(2.93), we have $\frac{\beta_1-\gamma_1}{\beta_1} = O(T^{-1})$ and $\frac{\beta_1-\gamma_2}{\beta_1} = \frac{8}{9} + o(1)$. It follows that (2.133) implies that $\upsilon_{11}^2 = 1 + O(T^{-1})$ and $\Sigma_{j=2}^T \upsilon_{1j}^2 = O(T^{-1})$. From (2.132), for any $k \neq 1$ we have

$$|\upsilon_{k1}\upsilon_{11}| = |\Sigma_{j=2}^T \upsilon_{kj}\upsilon_{1j}| \le \sqrt{\Sigma_{j=2}^T \upsilon_{kj}^2}\sqrt{\Sigma_{j=2}^T \upsilon_{1j}^2} = O(T^{-1/2}). \tag{2.134}$$

This implies $\upsilon_{k1}^2 = O(T^{-1})$. It's similar to obtain that $\upsilon_{22}^2 = 1 + O(T^{-1})$ and $\upsilon_{k2}^2 = O(T^{-1})$ for any $k \neq 2$.

By repeating these steps we conclude that $\upsilon_{kk}^2 = 1 + O(T^{-1})$ for any fixed k. This implies (2.123).

$\square$

Lemma 10. *Let* $(S_{k,1},\cdots,S_{k,T+l})' = \mathbf{s_k} = \frac{\mathbf{F^*C^*u_k}}{\sqrt{\gamma_1}}$. *Then* $\{\mathbf{s_k}\}_{1\le k\le T}$ *are orthogonal and*

$$\sum_{j=1}^{T+l} S_{k,j}^4 = O(T^{-1}). \tag{2.135}$$

Proof of Lemma 10. Note that $\{\mathbf{s_k}\}_{1\le k\le T}$ are orthogonal and real due to orthogonality of $\{\mathbf{u_k}\}_{1\le k\le T}$. We conclude from (2.88) and (2.121) that

$$\mathbf{s_k} = \frac{\mathbf{F^*C^*Cf_k}}{\sqrt{\gamma_1}\|\mathbf{Cf_k}\|} = \frac{1}{\sqrt{\gamma_1}\|\mathbf{Cf_k}\|}\Sigma_{j=1}^T\alpha_{kj}\lambda_j\mathbf{F}^*\tilde{\mathbf{y}}_\mathbf{j} = \mathbf{s_{k,M}} + \mathbf{s_{k,R}}, \tag{2.136}$$

where

$$\mathbf{s_{k,M}} = \frac{1}{\sqrt{\gamma_1}\|\mathbf{Cf_k}\|}\alpha_{kk}\lambda_k\mathbf{F}^*\tilde{\mathbf{y}}_\mathbf{k}, \quad \mathbf{s_{k,R}} = \frac{1}{\sqrt{\gamma_1}\|CL_k\|}\Sigma_{j\ne k}\alpha_{kj}\lambda_j\mathbf{F}^*\tilde{\mathbf{y}}_\mathbf{j}. \tag{2.137}$$

By Hölder's inequality, we have

$$\|\mathbf{s_{k,R}}\| = \|\frac{1}{\sqrt{\gamma_1}\|\mathbf{Cf_k}\|}\Sigma_{j\ne k}\alpha_{kj}\lambda_j\mathbf{F}^*\tilde{\mathbf{y}}_\mathbf{j}\| \le \frac{1}{\sqrt{\gamma_1}\|\mathbf{Cf_k}\|}\|\mathbf{F}\|_2\sqrt{\Sigma_{j\ne k}\alpha_{kj}^2\lambda_j^2}.$$

Recalling $\mathbf{A} = \mathbf{FF}^*$, we have

$$\|\mathbf{F}\|_2 = \sqrt{\|\mathbf{A}\|_2}.$$

Since A is a Hermitian Toeplitz matrix, from Pan et al. (2014),

$$\|\mathbf{A}\|_2 \le 2\sum_{0\le k\le l}|a_k|.$$

By (2.40) we can get

$$\|\mathbf{F}\|_2 = \sqrt{\|\mathbf{A}\|_2} < \infty.$$

From Lemma 3, (2.123), and (2.126) we can obtain that for any fixed k,

$$\frac{\sqrt{\Sigma_{j\ne k}\alpha_{kj}^2\lambda_j^2}}{\|\mathbf{Cf_k}\|} \le \sqrt{\frac{\Sigma_{j\ne k}\alpha_{kj}^2\lambda_j}{\Sigma_{j=1}^T\alpha_{kj}^2\lambda_j}}\sqrt{\lambda_1} = O(T^{1/2}).$$

This, together with (2.91), implies that for any fixed k,

$$\|\mathbf{s_{k,R}}\| = O(T^{-1/2}). \tag{2.138}$$

Similarly, we can also obtain that $\frac{1}{\sqrt{\gamma_1}\|\mathbf{Cf_k}\|}\alpha_{kk}\lambda_k$ is bounded for any fixed k.

Let $S_{k,M,j}$ be the jth element of $\mathbf{s_{k,M}}$ and $S_{k,R,j}$ be the jth element of $\mathbf{s_{k,R}}$. From (2.119), (2.123) and (2.137) and the assumption (A1) we can obtain that for any fixed k,

$$|S_{k,M,j}| \le \frac{1}{\sqrt{\gamma_1}\|\mathbf{Cf_k}\|}|\alpha_{kk}|\lambda_k \frac{2}{\sqrt{2T+1}}\sum_{h=0}^{l}|b_h| = O(T^{-1/2}). \quad (2.139)$$

It follows from (2.136), (2.137) and (2.139) that for any fixed k,

$$\begin{aligned}\sum_{j=1}^{T+l} S_{k,j}^4 &\le 8\sum_{j=1}^{T+l}(S_{k,R,j}^4 + S_{k,M,j}^4)\\ &\le 8\sum_{j=1}^{T+l} S_{k,M,j}^4 + 8\left(\sum_{j=1}^{T+l} S_{k,R,j}^2\right)^2 = O(T^{-1}). \quad (2.140)\end{aligned}$$

$\square$

C.2.3: Proof of Proposition 4

This subsection is to establish convergence in probability of the spiked eigenvalues of a kind of separable sample covariance matrices, which is enough for our purposes.

Lemma 11. *Let* $\mathbf{D} = \frac{1}{p}\mathbf{W}\mathbf{Z}\mathbf{\Sigma}\mathbf{Z}^*\mathbf{W}^*$, *where* $\mathbf{W}$ *is a* $T\times(T+l)$ *matrix,* $\mathbf{\Sigma}$ *is a* $p\times p$ *positive-definite matrix with* $\|\mathbf{\Sigma}\|_2 \le M_0$, *and* $\mathbf{Z}$ *is defined below (2.31). Order the eigenvalues of* $\mathbf{W}\mathbf{W}^*$ *as* $\tau_1 \ge \cdots \ge \tau_T$ *with* τ_1 *being bounded. Suppose that* $\{\tau_k\}_{1\le k\le T}$ *satisfy the following conditions.*

(C1) For any fixed k, *there is a constant* $c_k > 0$ *such that*

$$\lim_{T\to\infty}\tau_k = c_k. \quad (2.141)$$

(C2) For any $\epsilon > 0$ *there exist* T_0 *and* k_0, *where* k_0 *is a constant independent of* n *and* T, *such that when* $T \ge T_0$,

$$\left|\sum_{k>k_0}\tau_k\right| < \epsilon. \quad (2.142)$$

For any fixed k, denote the first k largest eigenvalues of $\mathbf{D}$ by $\rho_1 \geq \cdots \geq \rho_k$. Then $\rho_j - c_j \frac{tr(\mathbf{\Sigma})}{p} \to 0$ in probability.

Proof. We can find $\mathbf{V}$ such that

$$\mathbf{V}\mathbf{W}^*\mathbf{W}\mathbf{V}^* = diag(\tau_1, \cdots, \tau_{T+l}) = \mathbf{\Lambda_{T+l}},$$
$$\mathbf{V}\mathbf{V}^* = \mathbf{V}^*\mathbf{V} = \mathbf{I_{T+l}},$$

where $\tau_k = 0$ when $k > T$. So $\mathbf{D}^* = \frac{1}{p}\mathbf{V}\mathbf{Z}\mathbf{\Sigma}\mathbf{Z}^*\mathbf{V}^*\mathbf{\Lambda_{T+l}}$ has the same nonzero eigenvalues as $\frac{1}{p}\mathbf{Z}\mathbf{\Sigma}\mathbf{Z}^*\mathbf{W}^*\mathbf{W}$.

To prove this lemma, it suffices to prove that for any $\delta > 0$ and fix number k,

$$\lim_{T\to\infty} P\left(\left|\rho_k - c_k \frac{tr(\mathbf{\Sigma})}{p}\right| > \delta\right) = 0. \tag{2.143}$$

In view of (2.142), we can find $k_0 > 0$ such that

$$\left|\sum_{k>k_0} \tau_k\right| < \frac{\delta}{4M_0}. \tag{2.144}$$

Write $\mathbf{\Lambda_{T+l}} = \mathbf{\Lambda^M_{T+l}} + \mathbf{\Lambda^R_{T+l}}$, where

$$\mathbf{\Lambda^M_{T+l}} = diag\{\tau_1, \tau_2, \cdots, \tau_{k_0}, 0, \cdots, 0\}$$
$$\mathbf{\Lambda^R_{T+l}} = diag\{0, \cdots, 0, \tau_{k_0+1}, \tau_{k_0+2}, \cdots, \tau_{T+l}\}. \tag{2.145}$$

Let $h = tr(\frac{1}{p}\mathbf{V}\mathbf{Z}\mathbf{Z}^*\mathbf{V}^*\mathbf{\Lambda^R_{T+l}})$. Note that $E(\mathbf{V}\mathbf{Z}\mathbf{Z}^*\mathbf{V}^*)_{kk} = p$ and $\frac{Var(\mathbf{V}\mathbf{Z}\mathbf{Z}^*\mathbf{V}^*)_{kk}}{p} < \infty$. We can evaluate the mean and variance of h as follows

$$E(h) = \sum_{k=k_0+1}^{T+l} E\frac{(\mathbf{V}\mathbf{Z}\mathbf{Z}^*\mathbf{V}^*)_{kk}\tau_k}{p} = \sum_{k>k_0} \tau_k < \frac{\delta}{4M_0}$$

and

$$\mathrm{Var}(h) = \mathrm{Var}\left(\sum_{k=k_0+1}^{T}\left(\frac{(\mathbf{V}\mathbf{Z}\mathbf{Z}^*\mathbf{V}^*)_{kk}\tau_k}{p}\right)\right)$$
$$\leq \left(\sum_{k=k_0+1}^{T}\sqrt{\mathrm{Var}\left(\frac{(\mathbf{V}\mathbf{Z}\mathbf{Z}^*\mathbf{V}^*)_{kk}\tau_k}{p}\right)}\right)^2 = O\left(\frac{(\sum_{k>k_0}\tau_k)^2}{p}\right) = O\left(\frac{1}{p}\right).$$

As $\mathbf{\Sigma}$ is positive-definite, we have $\|\frac{1}{p}\mathbf{VZ\Sigma Z^*V^*\Lambda^R_{T+1}}\|_2 \leq tr(\frac{1}{p}\mathbf{VZ\Sigma Z^*V^*\Lambda^R_{T+1}}) \leq M_0 h$. It follows that

$$\lim_{T\to\infty} P\left(\left\|\frac{1}{p}\mathbf{VZ\Sigma Z^*V^*\Lambda^R_{T+1}}\right\|_2 > \frac{\delta}{2}\right) = 0. \tag{2.146}$$

Denote the kth largest eigenvalue of $\frac{1}{p}\mathbf{VZ\Sigma Z^*V^*\Lambda^M_{T+1}}$ by ρ_k^M. From the definition of $\mathbf{\Lambda^M_{T+1}}$, we conclude that $\frac{1}{n}\mathbf{VZ\Sigma Z^*V^*\Lambda^M_{T+1}}$ has the same nonzero eigenvalues as its upper left $k_0 \times k_0$ block. By Theorem 7.1 of Bai and Yao (2008), it's easy to prove that the limit of off-diagonal elements in the upper left $k_0 \times k_0$ block is 0 in probability. Recall that k_0 is a a constant which doesn't depend on T. We conclude that the nonzero eigenvalues of $\frac{1}{p}\mathbf{VZ\Sigma Z^*V^*\Lambda^M_{T+1}}$ are diagonal elements of the upper left $k_0 \times k_0$ block. From Theorem 7.1 of Bai and Yao (2008), the limit of diagonal elements in the upper left $k_0 \times k_0$ block can be obtained as follow

$$\lim_{T\to\infty} P\left(\left|\rho_k^M - \tau_k \frac{tr(\mathbf{\Sigma})}{p}\right| > \frac{\delta}{2}\right) = 0. \tag{2.147}$$

It follows from (2.139), (2.146) and (2.147)that

$$\begin{aligned}
& \lim_{T\to\infty} P(|\rho_k - \lambda_k \frac{tr(\mathbf{\Sigma})}{p}| > \delta) \\
\leq \quad & \lim_{T\to\infty} P(|\rho_k^M - \lambda_k \frac{tr(\mathbf{\Sigma})}{p}| + |\rho_k^M - \rho_k| > \delta) \\
\leq \quad & \lim_{T\to\infty} P(|\rho_k^M - \lambda_k \frac{tr(\mathbf{\Sigma})}{p}| > \frac{\delta}{2}) + \lim_{T\to\infty} P(|\rho_k^M - \rho_k| > \frac{\delta}{2}) \\
\leq \quad & \lim_{T\to\infty} P(|\rho_k^M - \lambda_k \frac{tr(\mathbf{\Sigma})}{p}| > \frac{\delta}{2}) + \lim_{T\to\infty} P(\|\frac{1}{p}\mathbf{VZ\Sigma Z^*V^*\Lambda^R_{p+1}}\|_2 > \frac{\delta}{2}) = 0.
\end{aligned}$$

□

We apply Lemma 11 with $\mathbf{D} = \frac{\mathbf{B}}{\gamma_1}$, where $\mathbf{B}$ is defined in (2.33). Lemmas 5 and 7 ensure that the conditions of Lemma 11 are satisfied, so that Proposition 4 holds.

C.2.4: Proofs of Proposition 5

Proof. Recalling the definitions of $\mathbf{u_k}$ and $\mathbf{s_k}$ in the above lemmas, we denote $(\mathbf{u_1}, \cdots, \mathbf{u_T})$ by $\mathbf{U}$ and $\frac{\mathbf{F^*C^*U}}{\sqrt{\gamma_1}}$ by $\mathbf{S}$. Note that $\{\mathbf{u_k}\}_{1\le k\le T}$ and $\{\mathbf{s_k}\}_{1\le k\le T}$ are both orthogonal and real. Since $\mathbf{s_j^*s_i} = 0$ for $i \neq j$ we have

$$\mathbf{SS}^* = \mathbf{\Lambda} = \mathrm{diag}\left\{\frac{\beta_1}{\gamma_1}, \cdots, \frac{\beta_T}{\gamma_1}\right\}. \tag{2.148}$$

In view of (2.33), let

$$\mathbf{D} = \frac{\mathbf{U^*BU}}{\gamma_1} = \frac{1}{p}\mathbf{S^*Z\Sigma Z^*S}. \tag{2.149}$$

The eigenvalues of $\mathbf{D}$ are ordered as $\frac{\rho_T}{\gamma_1} \le \cdots \le \frac{\rho_1}{\gamma_1}$.

To rewrite $\mathbf{D}$ as a block matrix we first introduce the following notation. For a fixed number $k > 0$, let $\mathbf{z_j} = (Z_{1j}, \cdots, Z_{(T+l)j})'$.

Set $\mathbf{V_1} = \frac{1}{\sqrt{p}}(\xi_\mathbf{1}, \cdots, \xi_\mathbf{p}) = \frac{1}{\sqrt{p}}\mathbf{Q_1Z} = \frac{1}{\sqrt{p}}(\mathbf{s_1}^*, \cdots, \mathbf{s_k}^*)'\mathbf{Z}$ and $\mathbf{V}_2 = \frac{1}{\sqrt{p}}(\eta_\mathbf{1}, \cdots, \eta_\mathbf{p}) = \frac{1}{\sqrt{p}}\mathbf{Q_2Z} = \frac{1}{\sqrt{p}}(\mathbf{s_{k+1}}^*, \cdots, \mathbf{s_T}^*)'\mathbf{Z}$ where $\mathbf{Q_1} = (\mathbf{s_1}^*, \cdots, \mathbf{s_k}^*)'$ and $\mathbf{Q_2} = (\mathbf{s_{k+1}}^*, \cdots, \mathbf{s_T}^*)'$. Then

$$\xi_\mathbf{j} = (\xi_j(1), \cdots, \xi_j(k))' = (\mathbf{s_1^*z_j}, \cdots, \mathbf{s_k^*z_j})' \tag{2.150}$$

and

$$\eta_\mathbf{j} = (\eta_j(k+1), \cdots, \eta_j(p))' = (\mathbf{s_{k+1}^*z_j}, \cdots, \mathbf{s_T^*z_j})'. \tag{2.151}$$

Let $\mathbf{\Lambda_1} = cov(\xi_\mathbf{j}) = \mathbf{Q_1Q_1}^*$ and $\mathbf{\Lambda_2} = cov(\eta_\mathbf{j}) = \mathbf{Q_2Q_2}^*$. In view of (2.148), we have

$$\mathbf{\Lambda_1} = \mathrm{diag}\left\{\frac{\beta_1}{\gamma_1}, \cdots, \frac{\beta_k}{\gamma_1}\right\}, \quad \mathbf{\Lambda_2} = \mathrm{diag}\left\{\frac{\beta_{k+1}}{\gamma_1}, \cdots, \frac{\beta_T}{\gamma_1}\right\}. \tag{2.152}$$

From Lemmas 5, 6 and (2.112) we can find a constant M_k such that

$$\lim_{T\to\infty} |\mathrm{tr}(\mathbf{\Lambda_2})| = \lim_{T\to\infty} |tr(\mathbf{\Lambda}) - tr(\mathbf{\Lambda_1})| = \left|\frac{\pi^2}{8} - \sum_{j=1}^{k} \frac{1}{(2i-1)^2}\right| < M_k. \tag{2.153}$$

In view of (2.149)-(2.151), we can rewrite $\mathbf{D}$ as

$$\mathbf{D} = \begin{pmatrix} \mathbf{V_1\Sigma V_1}^* & \mathbf{V_1\Sigma V_2}^* \\ \mathbf{V_2\Sigma V_1}^* & \mathbf{V_2\Sigma V_2}^* \end{pmatrix} \triangleq \begin{pmatrix} \mathbf{W_{11}} & \mathbf{W_{12}} \\ \mathbf{W_{21}} & \mathbf{W_{22}} \end{pmatrix}. \tag{2.154}$$

The characteristic polynomial of $\mathbf{D}$ is

$$0 = |\lambda \mathbf{I_T} - \mathbf{D}| = |\lambda \mathbf{I_{T-k}} - \mathbf{W_{22}}||\lambda \mathbf{I_k} - \mathbf{K_p}(\lambda)|\,, \tag{2.155}$$

where

$$\mathbf{K_p}(\lambda) = \mathbf{W_{11}} + \mathbf{W_{12}}(\lambda \mathbf{I_{T-k}} - \mathbf{W_{22}})^{-1}\mathbf{W_{21}}. \tag{2.156}$$

We conclude from Lemmas 5 and 7 that $\mathbf{W_{22}} = \mathbf{V_2 \Sigma V_2}^* = \frac{1}{p}\mathbf{Q_2 Z \Sigma Z^* Q_2^*}$ satisfies the conditions of Lemma 11. Lemma 11 immediately implies that the largest eigenvalue of $\mathbf{W_{22}}$, ρ, tends to $\frac{\gamma_{k+1}\frac{tr(\mathbf{\Sigma})}{p}}{\gamma_1}$ in probability. On the other hand, from Lemma 11, we also see that when $j \leq k$, $\frac{\rho_j - \gamma_j \frac{tr(\mathbf{\Sigma})}{p}}{\gamma_1} \to 0$ in probability. Since we want to study the first k largest eigenvalues, from Lemma 5 and (2.155), it's sufficient to consider the characteristic polynomial

$$0 = |\lambda \mathbf{I_k} - \mathbf{K_p}(\lambda)| = |\mathbf{G}(\lambda)|, \tag{2.157}$$

where

$$\mathbf{G}(\lambda) = \{G_{ij}(\lambda)\}_{1 \leq i,j \leq k} = \lambda \mathbf{I_k} - \mathbf{K_p}(\lambda). \tag{2.158}$$

From (2.156) we write

$$\begin{aligned}
\mathbf{K_p}(\lambda) &= \mathbf{W_{11}} + \mathbf{W_{12}}(\lambda \mathbf{I_{T-k}} - \mathbf{W_{22}})^{-1}\mathbf{W_{21}} \\
&= \mathbf{V_1 \Sigma V_1}^* + \mathbf{V_1 \Sigma V_2}^*(\lambda \mathbf{I_{T-k}} - \mathbf{W_{22}})^{-1}\mathbf{V_2 \Sigma V_1}^* \\
&= \mathbf{V_1}(\mathbf{\Sigma} + \mathbf{A_p}(\lambda))\mathbf{V_1}^*,
\end{aligned}$$

where

$$\mathbf{A_p}(\lambda) = \mathbf{\Sigma V_2}^*(\lambda \mathbf{I_{T-k}} - \mathbf{W_{22}})^{-1}\mathbf{V_2 \Sigma}. \tag{2.159}$$

It follows that

$$\mathbf{K_p}(\lambda) = \frac{1}{\sqrt{p}}\mathbf{R_p} + \mathbf{\Lambda_1}\frac{tr(\mathbf{\Sigma})}{p} + \mathbf{V_1 A_p}(\lambda)\mathbf{V_1}^*, \tag{2.160}$$

where

$$\mathbf{R_p} = \{R_{ij}\}_{1 \leq i,j \leq k} = \sqrt{p}\mathbf{V_1 \Sigma V_1}^* - \frac{tr(\mathbf{\Sigma})}{\sqrt{p}}\mathbf{\Lambda_1} = \sqrt{p}\mathbf{V_1 \Sigma V_1}^* - \sqrt{p}\mathbf{\Lambda_1}\frac{tr(\mathbf{\Sigma})}{p}. \tag{2.161}$$

Now we consider the Hermitian matrix $\mathbf{V_1 A_p}(\lambda)\mathbf{V_1}^*$ in (2.160). When λ is a solution of (2.157), we have $\lambda > \|\mathbf{W_{22}}\|_2$ in probability due to Lemma 11. Hence the eigenvalues of $(\lambda \mathbf{I_{T-k}} - \mathbf{W_{22}})^{-1}$ and $\sqrt{p}\mathbf{V_1 A_p}(\lambda)\mathbf{V_1}^*$ are non-negative in probability when λ is a solution of (2.157). Evidently we have

$$\|\sqrt{p}\mathbf{V_1 A_p}(\lambda)\mathbf{V_1}^*\|_2 \le \|\mathbf{\Sigma}\|_2^2 \|(\lambda \mathbf{I_{T-k}} - \mathbf{W_{22}})^{-1}\|_2 tr(\sqrt{p}\mathbf{V_1 V_2^* V_2 V_1}^*). \quad (2.162)$$

Note that eigenvalues of $\sqrt{p}\mathbf{V_1 V_2^* V_2 V_1}^*$ are also non-negative.

The next aim is to prove $E(\sqrt{p}\mathbf{V_1 V_2^* V_2 V_1}^*) = o_p(1)$.

Let $h_j = (\sqrt{p}\mathbf{V_1 V_2^* V_2 V_1}^*)_{jj} \ge 0$. we can claim that $h_j = o_p(1)$. In fact

$$h_j = \frac{1}{p\sqrt{p}} \mathbf{s_j^* Z Z^* Q_2^* Q_2 Z Z^* s_j}. \quad (2.163)$$

Let $EZ_{ij}^2 = y$, $E(Z_{ij}^*)^2 = z$ and $E|Z_{ij}|^4 = x + 1$. Write

$$\begin{aligned} E(h_j) &= \frac{1}{p\sqrt{p}} \mathbf{s_j}^* E(\mathbf{Z Z^* Q_2^* Q_2 Z Z^*})\mathbf{s_j} \quad (2.164)\\ &= \frac{1}{p\sqrt{p}} \sum_{i=1}^{p}\sum_{m=1}^{p} \mathbf{s_j}^* E(\mathbf{z_i z_i^* Q_2^* Q_2 z_m z_m^*})\mathbf{s_j} \\ &= \frac{1}{p\sqrt{p}} (\sum_{i=1}^{p}\sum_{m\neq i} \mathbf{s_j}^* E(\mathbf{z_i z_i^* Q_2^* Q_2 z_m z_m}^*)\mathbf{s_j} + \sum_{i=1}^{p} \mathbf{s_j}^* E(\mathbf{z_i z_i^* Q_2^* Q_2 z_i z_i}^*)\mathbf{s_j}). \end{aligned}$$

Consider the first term on the right hand of (2.164). When $i \neq m$,

$$\mathbf{s_j}^* E(\mathbf{z_i z_i^* Q_2^* Q_2 z_m z_m}^*)\mathbf{s_j} = \mathbf{s_j}^* E(\mathbf{z_i z_i}^*)\mathbf{Q_2^* Q_2} E(\mathbf{z_m z_m}^*)\mathbf{s_j} = \mathbf{s_j^* Q_2^* Q_2 s_j}.$$

Recall that $\{\mathbf{S}_k\}_{1\le k\le T}$ are orthogonal and real. Since $j \le k$, from the definition of $\mathbf{Q_2}$,

$$\mathbf{s_j}^* E(\mathbf{z_i z_i^* Q_2^* Q_2 z_m z_m}^*)\mathbf{s_j} = \mathbf{s_j^* Q_2^* Q_2 s_j} = 0, \quad (2.165)$$

which implies the first term of on the right hand of (2.164) equals 0.

Consider the second term on the right hand of (2.164) now. Let $\mathbf{P} = (P_{rt})_{1\le r,t\le T+l} = \mathbf{Q_2^* Q_2}$ and $\mathbf{H^v} = (H_{im}^v)_{1\le i,m\le T+l} = \mathbf{z_v z_v^* Q_2^* Q_2 z_v z_v}^* = \mathbf{z_v z_v^* P z_v z_v}^*$.

Then

$$
\begin{aligned}
E(H^v_{im}) &= E(Z_{iv}Z^*_{mv}\sum_{1\le r,t\le T+l}P_{rt}Z^*_{rv}Z_{tv})\\
&= \begin{cases} P_{ii}E|Z_{iv}|^4+\sum_{r\neq i}P_{rr}E|Z_{iv}|^2E|Z_{rv}|^2 & i=m,\\ P_{im}E|Z_{iv}|^2E|Z_{mv}|^2+P_{mi}EZ^2_{iv}E(Z^*_{mv})^2 & i\neq m.\end{cases}\\
&= \begin{cases} P_{ii}x+\sum_{r=1}^{T+l}P_{rr} & i=m,\\ P_{im}+P_{mi}yz & i\neq m.\end{cases}
\end{aligned}
$$

It follows that

$$
\begin{aligned}
&\sum_{v=1}^{p}\mathbf{s_j}^*E(\mathbf{z_v}\mathbf{z_v^*}\mathbf{P}\mathbf{z_v}\mathbf{z_v}^*)\mathbf{s_j}\\
=\ &\sum_{v=1}^{p}\mathbf{s_j}^*E(\mathbf{H^v})\mathbf{s_j}\\
=\ &\sum_{v=1}^{p}\sum_{i=1}^{T+l}\sum_{m=1}^{T+l}S^*_{ji}S_{jm}E(H^v_{im})\\
=\ &\sum_{v=1}^{p}\sum_{i=1}^{T+l}S^*_{ji}S_{ji}(P_{ii}x+\sum_{r=1}^{T+l}P_{rr})+\sum_{v=1}^{p}\sum_{i=1}^{T+l}\sum_{m=1,m\neq i}^{T+l}S^*_{ji}S_{jm}(P_{im}+P_{mi}yz)\\
=\ &\sum_{v=1}^{p}\sum_{i=1}^{T+l}S^*_{ji}S_{ji}(P_{ii}(x-1-yz)+\sum_{r=1}^{T+l}P_{rr})+\sum_{v=1}^{p}\sum_{i=1}^{T+l}\sum_{m=1}^{T+l}S^*_{ji}S_{jm}(P_{im}+P_{mi}yz)\\
\le\ &tr(\mathbf{P})(|x-1-yz|+1)\sum_{v=1}^{p}\sum_{i=1}^{T+l}S^*_{ji}S_{ji}+\sum_{v=1}^{n}\sum_{i=1}^{T+l}\sum_{m=1}^{T+l}S^*_{ji}S_{jm}(P_{im}+P_{mi}yz)\\
=\ &tr(\mathbf{P})(|x-1-yz|+1)p\|\mathbf{s_j}\|^2+p\sum_{i=1}^{T+l}\sum_{m=1}^{T+l}S^*_{ji}S_{jm}(P_{im}+P_{mi}yz)\\
=\ &tr(\mathbf{P})(|x-1-yz|+1)p\|\mathbf{s_j}\|^2+p\mathbf{s_j^*}\mathbf{P}\mathbf{s_j}+pyz\mathbf{s_j^*}\mathbf{P}\mathbf{s_j}.
\end{aligned}
$$

By (2.165) and $\mathbf{P}=\mathbf{Q_2^*Q_2}$, we have

$$
\sum_{v=1}^{p}\mathbf{s_j^*}E(\mathbf{z_v}\mathbf{z_v^*}\mathbf{P}\mathbf{z_v}\mathbf{z_v}^*)\mathbf{s_j}\le tr(\mathbf{P})(|x-1-yz|+1)p\|\mathbf{s_j}\|^2.
$$

Also, we have

$$
tr(\mathbf{P})=tr(\mathbf{Q_2^*Q_2})=tr(\mathbf{\Lambda_2}).
$$

From (2.153), (2.163) and (2.165) we can obtain

$$E(h_j) \leq \frac{1}{\sqrt{p}} tr(\mathbf{P})(|x-1-yz|+1)\|\mathbf{s_j}\|^2 \to 0, \tag{2.166}$$

as claimed.

Since k is a fixed number, we can obtain

$$E(tr(\sqrt{p}\mathbf{V_1}\mathbf{V_2^*}\mathbf{V_2}\mathbf{V_1}^*)) = \sum_{j=1}^{k} E(h_j) = o_p(1). \tag{2.167}$$

It follows from $\|(\lambda\mathbf{I_{T-k}} - \mathbf{W_{22}})^{-1}\|_2 = O_p(1)$, (2.162) and (2.167) that

$$\|\mathbf{V_1}\mathbf{A_p}(\lambda))\mathbf{V_1}^*\|_2 = o_p(p^{-1/2}). \tag{2.168}$$

Now we consider $\frac{1}{\sqrt{p}}\mathbf{R_p} = \frac{1}{\sqrt{p}}(R_{ij})$ in (2.160). From (2.161) and the definition of $\mathbf{V_1}$, $\mathbf{Q_1}$ and $\mathbf{\Lambda_1}$,

$$R_{ij} = \frac{1}{\sqrt{p}}(\mathbf{s_i^*}\mathbf{Z}\mathbf{\Sigma}\mathbf{Z}^*\mathbf{s_j} - \mathbf{s_i^*}\mathbf{s_j}tr(\mathbf{\Sigma})). \tag{2.169}$$

Note that $E(\mathbf{s_i^*}\mathbf{z_1}\mathbf{z_1^*}\mathbf{s_j}) = \mathbf{s_i^*}\mathbf{s_j}$. With Theorem 7.1 of Bai and Yao (2008), we can prove that R_{ij} converges weakly to a zero-mean Gaussian variable r_{ij} with bounded variance. It follows that

$$\frac{1}{\sqrt{p}} R_{ij} = O_p(p^{-1/2}). \tag{2.170}$$

Note that $\mathbf{\Lambda_1}\frac{tr(\mathbf{\Sigma})}{p}$ is a diagonal matrix and hence we can find that any off-diagonal element of $\lambda\mathbf{I_k} - \mathbf{K_n}(\lambda)$ is $O_p(p^{-1/2})$. This, together with (2.158), implies that for any $i \neq j$, $G_{ij}(\lambda) = O_p(p^{-1/2})$ with λ satisfying $|\mathbf{G}(\lambda)| = 0$. Similarly, $G_{ii}(\lambda) = \lambda - \frac{tr(\mathbf{\Sigma})\beta_i}{p\gamma_1} + O_p(p^{-1/2})$ is absolutely bounded with λ satisfying $|\mathbf{G}(\lambda)| = 0$.

Denote by Ω_k all the permutations σ of the set $\{1, 2, \cdots, k\}$. By the Laplace formula for a determinant we have

$$\begin{aligned} 0 \quad &= |\mathbf{G}(\lambda)| = \sum_{\sigma\in\Omega_k} \operatorname{sgn}(\sigma) \prod_{j=1}^{k} G_{j,\sigma_j}(\lambda) \\ &= \sum_{\sigma\in\Omega_k, \sigma\neq[1,2,\cdots,k]} \operatorname{sgn}(\sigma) \prod_{j=1}^{k} G_{j,\sigma_j}(\lambda) + \prod_{j=1}^{k} G_{jj}(\lambda). \end{aligned}$$

Recall that any off-diagonal element of $\mathbf{G}(\lambda)$ is $O_p(p^{-1/2})$. We conclude that when $\sigma \neq [1, 2, \cdots, k]$, $\prod_{j=1}^{k} G_{j,\sigma_j}(\lambda) = O_p(p^{-1})$ since there are at least two different j_1 and j_2 such that $\sigma_{j_1} \neq j_1$ and $\sigma_{j_2} \neq j_2$. Since k is fixed, we have

$$\prod_{1 \leq j \leq k} G_{jj}(\lambda) = -\sum_{\sigma \in \Omega_k, \sigma \neq [1,2,\cdots,k]} sgn(\sigma) \prod_{j=1}^{k} G_{j,\sigma_j}(\lambda) = O_p(p^{-1}). \qquad (2.171)$$

Then when λ satisfies $|\mathbf{G}(\lambda)| = 0$, there exists j (not bigger than k) such that $|G_{jj}(\lambda)| = o(1)$. When $i \neq j$, from (2.160), (2.168) and (2.170)

$$\begin{aligned} & |G_{jj}(\lambda) - G_{ii}(\lambda)| \\ \geq & \frac{tr(\mathbf{\Sigma})}{p}|(\Lambda_1)_{jj} - (\Lambda_1)_{ii}| - |\frac{1}{\sqrt{p}}R_{jj} - \frac{1}{\sqrt{p}}R_{ii}| - |(\mathbf{V_1 A_p}(\lambda)\mathbf{V_1}^*)_{jj} - (\mathbf{V_1 A_p}(\lambda)\mathbf{V_1}^*)_{ii}| \\ = & \frac{tr(\mathbf{\Sigma})|\beta_j - \beta_i|}{p\gamma_1} + O_p(p^{-1/2}). \end{aligned}$$

By Lemma 5 we can obtain that for any $i \neq j$, $|G_{ii}(\lambda)| \geq \frac{tr(\mathbf{\Sigma})}{p}(|\frac{1}{(2i-1)^2} - \frac{1}{(2j-1)^2}|) + o_p(1)$. This, together with (2.171), implies that $|G_{jj}(\lambda)| = O_p(p^{-1})$. Hence $|G_{jj}(\lambda_j)| = O_p(p^{-1})$ for any $\lambda_1 > \lambda_2 > \cdots > \lambda_k$ satisfying $|\mathbf{G}(\lambda_j)| = 0$ for $1 \leq j \leq k$. Write

$$(\lambda_1, \lambda_2, \cdots, \lambda_k) = (\lambda_1 - G_{11}(\lambda_1) + O_p(p^{-1}), \cdots, \lambda_k - G_{kk}(\lambda_k) + O_p(p^{-1})). \quad (2.172)$$

It follows that

$$\begin{aligned} &\left(\sqrt{p}\left(\lambda_1 - \frac{\gamma_1}{\gamma_1}\frac{tr(\mathbf{\Sigma})}{p}\right), \cdots, \sqrt{p}\left(\lambda_k - \frac{\gamma_k}{\gamma_1}\frac{tr(\mathbf{\Sigma})}{p}\right)\right) = \\ &\left(\sqrt{p}\left(\lambda_1 - \frac{\gamma_1}{\gamma_1}\frac{tr(\mathbf{\Sigma})}{p} - G_{11}(\lambda_1) + O_p(p^{-1})\right), \cdots, \sqrt{p}\left(\lambda_k - \frac{\gamma_k}{\gamma_1}\frac{tr(\mathbf{\Sigma})}{p} - G_{kk}(\lambda_k) + O_p(p^{-1})\right)\right) \end{aligned}$$

Using this, by (2.158), (2.160), (2.152), (2.161), (2.168) and Lemma 5 we further obtain

$$\begin{aligned} & \sqrt{p}[\lambda_j - \frac{\gamma_j}{\gamma_1}\frac{tr(\mathbf{\Sigma})}{p} - G_{jj}(\lambda_j) + O_p(p^{-1})] \\ = & \sqrt{p}[\lambda_j - \frac{\gamma_j}{\gamma_1}\frac{tr(\mathbf{\Sigma})}{p} - \lambda_j + \frac{\beta_j}{\gamma_1}\frac{tr(\mathbf{\Sigma})}{p} + \frac{1}{\sqrt{p}}R_{jj} + (\mathbf{V_1 A_n}(\lambda_j)\mathbf{V_1}^*)_{jj} + O_p(p^{-1})] \\ = & R_{jj} + O(\sqrt{p}T^{-1}) + o_p(1). \end{aligned} \qquad (2.173)$$

Recalling (2.169), we have

$$(R_{11}, \cdots, R_{kk})' = \left(\frac{1}{\sqrt{p}}(\mathbf{s_1^*Z\Sigma Z^*s_1} - \mathbf{s_1^*s_1}tr(\mathbf{\Sigma})), \cdots, \frac{1}{\sqrt{p}}(\mathbf{s_k^*Z\Sigma Z^*s_k} - \mathbf{s_k^*s_k}tr(\mathbf{\Sigma}))\right)'.$$

Note that $E(\mathbf{s_i^*z_1z_1^*s_j}) = \mathbf{s_i^*s_j}$. From Theorem 7.1 of Bai and Yao (2008), we can find that $(R_{11}, \cdots, R_{kk})'$ converges weakly to a zero-mean Gaussian vector $\mathbf{w} = (w_1, \cdots, w_k)'$.

We next determine the covariance between w_i and w_j for the complex case and the real case in a unified way. To this end, let $\omega = \lim_{p\to\infty} \frac{\sum_{1\le i\le p}\Sigma_{ii}^2}{p}$ and $\theta = \lim_{p\to\infty} \frac{tr(\Sigma^2)}{p} = \lim_{p\to\infty} \frac{tr(\Sigma\Sigma')}{p}$. When $i \neq j$, from Theorem 7.1 of Bai and Yao (2008), we have

$$\begin{aligned}
&cov(w_i, w_j)\\
&= \lim_{T\to\infty} \omega(E \mid \xi_1(i) \mid^2 \mid \xi_1(j) \mid^2 - (E \mid \xi_1(i) \mid^2)(E \mid \xi_1(j) \mid^2))\\
&+ \lim_{T\to\infty} (\theta - \omega)(E\bar{\xi}_1(i)\xi_1(j))(E\bar{\xi}_1(j)\xi_1(i)) + \lim_{T\to\infty} (\theta - \omega)(E\xi_1(i)\xi_1(j))(E\bar{\xi}_1(i)\bar{\xi}_1(j)).
\end{aligned}$$

In view of (2.150), we have

$$\begin{aligned}
& E(\mid \xi_1(i) \mid^2 \mid \xi_1(j) \mid^2) - E(\mid \xi_1(i) \mid^2)E(\mid \xi_1(j) \mid^2) \qquad (2.174)\\
= \quad & E(\mathbf{s_i^*Z_1Z_1^*s_is_j^*Z_1Z_1^*s_j}) - E(\mathbf{s_i^*Z_1Z_1^*s_i})E(\mathbf{s_j^*Z_1Z_1^*s_j}).
\end{aligned}$$

Recall that $\{\mathbf{s_k}\}_{1\le k\le T}$ are orthogonal and real. We obtain

$$
\begin{aligned}
& E(\mathbf{s_i^* z_1 z_1^* s_i s_j^* z_1 z_1^* s_j}) \\
= \quad & E\left(\sum_{f_1=1}^{T+l}\sum_{f_2=1}^{T+l} S_{if_1}S_{if_2}Z_{f_11}Z_{f_21}^*\right)\left(\sum_{f_1=1}^{T+l}\sum_{f_2=1}^{T+l} S_{jf_1}S_{jf_2}Z_{f_11}Z_{f_21}^*\right) \\
= \quad & \sum_{f_1=1}^{T+l}\sum_{f_2\neq f_1} S_{if_1}S_{if_2}S_{jf_1}S_{jf_2}\left(EZ_{f_11}^2\left(Z_{f_21}^*\right)^2 + E|Z_{f_11}|^2|Z_{f_21}|^2\right) \\
+ \quad & E\left(\sum_{f_1=1}^{T+l} S_{if_1}^2|Z_{f_11}|^2\right)\left(\sum_{f_2=1}^{T+l} S_{jf_2}^2|Z_{f_21}|^2\right) \\
= \quad & (yz+1)\sum_{f_1=1}^{T+l}\sum_{f_2\neq f_1} S_{if_1}S_{if_2}S_{jf_1}S_{jf_2} \\
+ \quad & E\left(\sum_{f_1=1}^{T+l} S_{if_1}^2|Z_{f_11}|^2\right)\left(\sum_{f_2=1}^{T+l} S_{jf_2}^2|Z_{f_21}|^2\right).
\end{aligned}
\tag{2.175}
$$

Since $\{\mathbf{s_k}\}_{1\le k\le T}$ are orthogonal, we conclude from (2.135) that

$$
\begin{aligned}
& (yz+1)\sum_{f_1=1}^{T+l}\sum_{f_2\neq f_1} S_{if_1}S_{if_2}S_{jf_1}S_{jf_2} \\
= \quad & (yz+1)\sum_{f_1=1}^{T+l} S_{if_1}S_{jf_1}\sum_{f_2=1}^{T+l} S_{if_2}S_{jf_2} - (yz+1)\sum_{f_1=1}^{T+l} S_{if_1}^2S_{jf_1}^2 \\
= \quad & -(yz+1)\sum_{f_1=1}^{T+l} S_{if_1}^2S_{jf_1}^2 = O(T^{-1})
\end{aligned}
\tag{2.176}
$$

and

$$
\begin{aligned}
& E\left(\sum_{f_1=1}^{T+l} S_{if_1}^2|Z_{f_11}|^2\right)\left(\sum_{f_2=1}^{T+l} S_{jf_2}^2|Z_{f_21}|^2\right) \\
= \quad & \sum_{f_1=1}^{T+l} S_{if_1}^2S_{jf_1}^2(E|Z_{f_11}|^4-1) + E(\mathbf{s_i^* z_1 z_1^* s_i})E(\mathbf{s_j^* z_1 z_1^* s_j}) \\
= \quad & E(\mathbf{s_i^* z_1 z_1^* s_i})E(\mathbf{s_j^* z_1 z_1^* s_j}) + O(T^{-1}).
\end{aligned}
\tag{2.177}
$$

Summarizing (2.174), (2.175), (2.176) and (2.177), we conclude that

$$\lim_{T\to\infty} \omega(E \mid \xi_1(i) \mid^2 \mid \xi_1(j) \mid^2 - (E \mid \xi_1(i) \mid^2)(E \mid \xi_1(j) \mid^2)) = 0.$$

Since $\{\mathbf{s_k}\}_{1\leq k\leq T}$ are orthogonal and real, we also have

$$E\bar{\xi}_1(i)\xi_1(j) = 0, E\bar{\xi}_1(j)\xi_1(i) = 0$$

and

$$E\xi_1(i)\xi_1(j) = 0, E\bar{\xi}_1(i)\bar{\xi}_1(j) = 0.$$

This implies

$$cov(w_i, w_j) = 0. \tag{2.178}$$

By (2.13) and (2.150) we can obtain

$$\begin{aligned}
& Var(w_i) \\
= \quad & \omega \lim_{T\to\infty}\{E \mid \xi_1(i) \mid^4 - 2(E \mid \xi_1(i) \mid^2)^2 - (E\xi_1(i)^2)(E\bar{\xi}_1(i)^2)\} \\
& +\theta \lim_{T\to\infty}(E \mid \xi_1(i) \mid^2)^2 + \theta \lim_{T\to\infty}(E\xi_1(i)^2)(E\bar{\xi}_1(i)^2) \\
= \quad & \omega \lim_{T\to\infty}\{E \mid \xi_1(i) \mid^4 - 2(E \mid \xi_1(i) \mid^2)^2 - (E\xi_1(i)^2)(E\bar{\xi}_1(i)^2)\} \\
& +\theta\frac{1}{(2i-1)^4} + \theta\frac{1}{(2i-1)^4}(1 - 4E(Z_{i1}^R)^2E(Z_{i1}^I)^2).
\end{aligned}$$

From Lemma 5

$$\begin{aligned}
& \lim_{T\to\infty}\{E \mid \xi_1(i) \mid^4 - 2(E \mid \xi_1(i) \mid^2)^2 - (E\xi_1(i)^2)(E\bar{\xi}_1(i)^2)\} \\
= \quad & \lim_{T\to\infty}\{E \mid \sum_{j=1}^{T+l} S_{ij}Z_{j1} \mid^4 - 2(\frac{\beta_i}{\gamma_1})^2 - (\frac{\beta_i}{\gamma_1})^2(E(Z_{i1}^R)^2 - E(Z_{i1}^I)^2)^2\} \\
= \quad & \lim_{T\to\infty}\{E(\sum_{j=1}^{T+l} S_{ij}Z_{j1}^R)^4 + E(\sum_{j=1}^{T+l} S_{ij}Z_{j1}^I)^4 + \\
& 2E(\sum_{j=1}^{T+l} S_{ij}Z_{j1}^R)^2(\sum_{j=1}^{T+l} S_{ij}Z_{j1}^I)^2\} - \frac{1}{(2i-1)^4}(2 + (E(Z_{i1}^R)^2 - E(Z_{i1}^I)^2)^2).
\end{aligned}$$

Recalling that Z_{j1}^R and Z_{j1}^I are independent, we have

$$2E(\sum_{j=1}^{T+l} S_{ij}Z_{j1}^R)^2(\sum_{j=1}^{T+l} S_{ij}Z_{j1}^I)^2 = 2E(\sum_{j=1}^{T+l} S_{ij}^2(Z_{j1}^R)^2)E(\sum_{j=1}^{T+l} S_{ij}^2(Z_{j1}^I)^2)$$

and

$$\begin{aligned}
& E(\sum_{j=1}^{T+l} S_{ij}Z_{j1}^R)^4 + E(\sum_{j=1}^{T+l} S_{ij}Z_{j1}^I)^4 + 2E(\sum_{j=1}^{T+l} S_{ij}Z_{j1}^R)^2(\sum_{j=1}^{T+l} S_{ij}Z_{j1}^I)^2 \\
= \quad & 3[(\sum_{j=1}^{T+l} S_{ij}^2 E(Z_{j1}^R)^2)^2 + (\sum_{j=1}^{T+l} S_{ij}^2 E(Z_{j1}^I)^2)^2] + \\
& \sum_{j=1}^{T+l} S_{ij}^4[E(Z_{j1}^R)^4 + E(Z_{j1}^I)^4 - 3(E((Z_{j1}^R)^2)^2 + (E(Z_{j1}^I)^2)^2)] \\
& +2E(\sum_{j=1}^{T+l} S_{ij}^2(Z_{j1}^R)^2)E(\sum_{j=1}^{T+l} S_{ij}^2(Z_{j1}^I)^2) \\
= \quad & 3[\sum_{j=1}^{T+l} S_{ij}^2 E(Z_{j1}^R)^2 + \sum_{j=1}^{T+l} S_{ij}^2 E(Z_{j1}^I)^2]^2 \\
& +\sum_{j=1}^{T+l} S_{ij}^4[E(Z_{j1}^R)^4 + E(Z_{j1}^I)^4 - 3((E(Z_{j1}^R)^2)^2 + (E(Z_{j1}^I)^2)^2)] - \\
& 4E(\sum_{j=1}^{T+l} S_{ij}^2(Z_{j1}^R)^2)E(\sum_{j=1}^{T+l} S_{ij}^2(Z_{j1}^I)^2) \\
= \quad & (2 + (E(Z_{i1}^R)^2 - E(Z_{i1}^I)^2)^2)(\sum_{j=1}^{T+l} S_{ij}^2)^2 \\
+ \quad & \sum_{j=1}^{T+l} S_{ij}^4[E(Z_{j1}^R)^4 + E(Z_{j1}^I)^4 - 3((E(Z_{j1}^R)^2)^2 + (E(Z_{j1}^I)^2)^2)].
\end{aligned}$$

In view of Lemma 5, (2.135) and (2.148), we have

$$(2 + (E(Z_{i1}^R)^2 - E(Z_{i1}^I)^2)^2)(\sum_{j=1}^{T+l} S_{ij}^2)^2 = (2 + (E(Z_{i1}^R)^2 - E(Z_{i1}^I)^2)^2)(\frac{1}{(2i-1)^4} + O(T^{-1}))$$

and

$$\sum_{j=1}^{T+l} S_{ij}^4[E(Z_{j1}^R)^4 + E(Z_{j1}^I)^4 - 3((E(Z_{j1}^R)^2)^2 + (E(Z_{j1}^I)^2)^2)] = O(T^{-1}).$$

So we can obtain

$$\lim_{T\to\infty}\{E \mid \xi_1(i) \mid^4 -2(E \mid \xi_1(i) \mid^2)^2 - (E\xi_1(i)^2)(E\bar{\xi}_1(i)^2)\} = 0.$$

It follows that

$$Var(w_i) = \theta \frac{1}{(2i-1)^4}(2 - 4E(Z_{i1}^R)^2 E(Z_{i1}^I)^2).$$

This, together with (2.173), (2.178) and the assumption (A4), implies Theorem 5.

□

Appendix C.3: Proofs of the results in Appendix A.3

We now list some results which are similar to those in the above section.

Lemma 12. *Recall $\bar{\lambda}_1 \geq \bar{\lambda}_2 \geq \cdots \geq \bar{\lambda}_{T-1} > 0$ in (2.14). One can verify that they are the positive eigenvalues of* $\mathbf{C}^*\mathbf{H}^*\mathbf{HC}$.

Proof of Lemma 12. We can write $\mathbf{C}^*\mathbf{H}^*\mathbf{HC} = diag\{0, \bar{\mathbf{M}}_{T-1}\}$, where $\bar{\mathbf{M}}_{T-1}$ is a $(T-1)\times(T-1)$ invertible matrix. Let $\ddot{\mathbf{M}}_{\mathbf{T-1}} = (\bar{\mathbf{M}}_{T-1})^{-1}$. Define the characteristic function of $\ddot{\mathbf{M}}_{\mathbf{T}}$ by $g_T(\lambda) = \det(\lambda\mathbf{I}_{\mathbf{T-1}} - \ddot{\mathbf{M}}_{\mathbf{T-1}})$. One can verify that $\{\ddot{M}_{i,j}\}$, the entries of $\ddot{\mathbf{M}}_{\mathbf{T-1}}$, satisfy

$$\ddot{M}_{ij} = \begin{cases} 2 & i = j, \\ -1 & |i-j| = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{2.179}$$

By the cofactor expansion we obtain a recurrence relation as following

$$g_T(\lambda) = (\lambda - 2)g_{T-1}(\lambda) - g_{T-2}(\lambda). \tag{2.180}$$

Consider $\lambda \in (0,4)$ at first. We then write $\lambda = \lambda(\theta) = 2 + 2\cos\theta$. We can further solve (2.180) to get

$$g_T(\lambda) = \frac{\sin T\theta}{\sin\theta}. \tag{2.181}$$

When $\sin\theta \neq 0$, $g_T(\lambda) = 0$ is equivalent to

$$\sin T\theta = 0. \tag{2.182}$$

Let $\bar{h}_T(\theta) = \sin T\theta$. Note that (2.14) gives $T-1$ different solutions which satisfy $h_T(\theta) = 0$ and $\sin\theta \neq 0$. On the other hand, observe that there are at most $T-1$ solutions for $g_T(\lambda) = 0$. The proof is complete.

□

Lemma 13. *Using the notation in (2.14),*

$$\lim_{T\to\infty} \frac{\bar{\lambda}_k}{T^2} = \frac{1}{\pi^2 k^2} \tag{2.183}$$

for any fixed k.

Lemma 14 below specifies the eigenvectors of $\mathbf{C^*H^*HC}$.

Lemma 14. *Let $\ddot{\mathbf{x}}_\mathbf{k} = (0, \ddot{x}_{k,1}, \cdots, \ddot{x}_{k,T-1})'$ be a $T \times 1$ vector with*

$$\ddot{x}_{k,i} = (-1)^{T-i} \sin(T-i)\bar{\theta}_k, \; -l \leq i \leq T+l. \tag{2.184}$$

Then $\{\ddot{\mathbf{x}}_\mathbf{k}, 1 \leq k \leq T-1\}$ are orthogonal and satisfy for any k

$$\mathbf{C^*H^*HC}\ddot{\mathbf{x}}_\mathbf{k} = \bar{\lambda}_k \ddot{\mathbf{x}}_\mathbf{k}. \tag{2.185}$$

Lemmas 13 and 14 can be verified with some straightforward computations and the fact (2.89).

Lemma 15 below specifies the eigenvalues of $\mathbf{A_m C^*H^*HC}$ and gives their approximation to those of $\mathbf{AC^*H^*HC}$.

Lemma 15. *Define $\bar{\gamma}_k$ by*

$$\bar{\gamma}_k = \bar{\lambda}_k \left(a_0 + 2 \sum_{1\leq j\leq T-2} a_j (-1)^j \cos(j\bar{\theta}_k) \right). \tag{2.186}$$

For any fixed constant $k \geq 1$, there is a constant c_k such that

$$\lim_{T\to\infty} \frac{\bar{\gamma}_k}{T^2} = c_k > 0 \tag{2.187}$$

and

$$\lim_{T\to\infty}\frac{\bar{\gamma}_k}{\bar{\gamma}_1}=\lim_{T\to\infty}\frac{\bar{\lambda}_k}{\bar{\lambda}_1}=\frac{1}{k^2}. \tag{2.188}$$

Let $\bar{\beta}_1 \geq \bar{\beta}_2 \geq \cdots \geq \bar{\beta}_T$ *be the eigenvalues of* $\mathbf{AC^*H^*HC}$. *If* $\mathbf{A}$ *satisfies Assumptions (A1) and (A2), then (2.93) and (2.94) still hold with* β_i *and* γ_i *replaced with* $\bar{\beta}_i$ *and* $\bar{\gamma}_i$ *respectively.*

Lemma 16. *Suppose that* $\mathbf{A}$ *satisfies Assumptions (A1) and (A2). Then*

$$\lim_{T\to\infty}\frac{\bar{\beta}_k}{tr(\mathbf{AC^*H^*HC})}=\lim_{T\to\infty}\frac{\bar{\gamma}_k}{tr(\mathbf{AC^*H^*HC})}=\frac{6}{\pi^2k^2}. \tag{2.189}$$

Lemma 17. *Suppose that* $\mathbf{A}$ *satisfies the assumptions (A1) and (A2). For any* $\epsilon > 0$*, we can find* T_0 *and* k_0*, where* k_0 *is a finite number independent of* T*, such that when* $T \geq T_0$*,*

$$\left|\frac{\sum_{k>k_0}\bar{\beta}_k}{\bar{\gamma}_1}\right|<\epsilon. \tag{2.190}$$

Lemma 18. *Recall the eigenvectors* $\ddot{\mathbf{x}}_\mathbf{k}$ *defined in Lemma 14. Then*

$$\Sigma_{j=1}^{T-1}(\ddot{x}_{k,j})^2=\frac{T}{2}. \tag{2.191}$$

Let

$$\ddot{\mathbf{y}}_\mathbf{k}=\frac{\ddot{\mathbf{x}}_\mathbf{k}}{\|\ddot{\mathbf{x}}_\mathbf{k}\|}. \tag{2.192}$$

Then $\{\ddot{\mathbf{y}}_\mathbf{k}\}_{1\leq k\leq T}$ *are orthogonal and the jth element of* $\ddot{\mathbf{y}}_\mathbf{k}$, $y_{k,j}$, *satisfies*

$$|\ddot{y}_{k,j}|=\frac{|\ddot{x}_{k,j}|}{\sqrt{\frac{T}{2}}}\leq\frac{\sqrt{2}}{\sqrt{T}}. \tag{2.193}$$

Lemma 19. *Let* $\{\ddot{\mathbf{u}}_\mathbf{k}\}_{1\leq k\leq T-1}$ *be orthogonal and real vectors such that* $\|\ddot{\mathbf{u}}_\mathbf{k}\|=1$ *and*

$$\mathbf{HCAC^*H^*}\ddot{\mathbf{u}}_\mathbf{k}=\bar{\beta}_k\ddot{\mathbf{u}}_\mathbf{k}. \tag{2.194}$$

Let $(\ddot{S}_{k,1},\cdots,\ddot{S}_{k,T+l})'=\ddot{\mathbf{s}}_\mathbf{k}=\frac{\mathbf{F^*H^*C^*}\ddot{\mathbf{u}}_\mathbf{k}}{\sqrt{\bar{\gamma}_1}}$. *Then* $\{\ddot{\mathbf{s}}_\mathbf{k}\}_{1\leq k\leq T-1}$ *are orthogonal and*

$$\sum_{j=1}^{T+l}\ddot{S}_{k,j}^4=O(T^{-1}). \tag{2.195}$$

The proofs of Lemma 15–Lemma 19 are similar to those of the counterparts in the above sections. We omit the details here.

Appendix D: Discussion about cointegrating structure and deterministic trending

Appendix D.1: Discussion how the rank of $\mathbf{\Pi}-\mathbf{I}$ affects performance

This section is to discuss how the rank of $\mathbf{\Pi}-\mathbf{I}$ affects performance of the statistics. To this end we first develop two theorems about the largest eigenvalues of centered and non-centered sample covariance matrices for the case when $rank(\mathbf{\Pi}-\mathbf{I})$ is not necessarily full. Some simulations are also conducted to see how the rank of $\mathbf{\Pi}-\mathbf{I}$ affects the performance of the statistics.

Theorem 7. *Suppose that Assumptions A1-A6 hold. The matrix $\mathbf{\Pi}$ is symmetric and $rank(\mathbf{\Pi}-\mathbf{I})=p-p_1$. The eigenvalues of $\mathbf{\Pi}$ satisfy*

$$\lambda_1(\mathbf{\Pi})=\cdots=\lambda_{p_1}(\mathbf{\Pi})=1>\varphi\geq\lambda_{p_1+1}(\mathbf{\Pi})\geq\lambda_p(\mathbf{\Pi})\geq-\varphi.$$

There exists a $p\times p$ matrix $\mathbf{U_\Pi}$ such that $\mathbf{U_\Pi}\mathbf{U_\Pi}'=I$ and $\mathbf{\Pi}=\mathbf{U_\Pi}diag\{\mathbf{I}_{p_1},\mathbf{\Pi_R}\}\mathbf{U_\Pi}'$, where $\mathbf{\Pi_R}=diag\{\lambda_{p_1+1}(\mathbf{\Pi}),\cdots,\lambda_p(\mathbf{\Pi})\}$. We can write $\mathbf{U_\Pi}=(\mathbf{U_{M1}}',\mathbf{U_{R1}}')'$, where U_{M1} is a $p_1\times p$ matrix and U_{R1} is a $(p-p_1)\times p$ matrix. Let ρ_k be the kth largest eigenvalue of $\mathbf{B}$ with k fixed.

(1) Let $\bar{\mathbf{\Sigma}}=\mathbf{U_{M1}}\mathbf{\Sigma}\mathbf{U}'_{\mathbf{M1}}$ be a $p_1\times p_1$ matrix. When $\frac{tr(\bar{\mathbf{\Sigma}})}{p}\geq c>0$,

$$\frac{\rho_k-\gamma_k\frac{tr(\bar{\mathbf{\Sigma}})}{p}}{\gamma_1}\xrightarrow{i.p.}0.$$

Furthermore, when $\lim\limits_{T\to\infty}\frac{p-p_1}{p}=0$,

$$\frac{\rho_k-\gamma_k\frac{tr(\mathbf{\Sigma})}{p}}{\gamma_1}\xrightarrow{i.p.}0.$$

(2) *Suppose that the Assumption A7 holds. Then when* $\frac{tr(\bar{\mathbf{\Sigma}})}{p} \geq c > 0$*, the random vector*

$$\left(\frac{p}{\sqrt{p_1}}\frac{\rho_1 - \gamma_1\frac{tr(\bar{\mathbf{\Sigma}})}{p}}{\gamma_1}, \cdots, \frac{p}{\sqrt{p_1}}\frac{\rho_k - \gamma_k\frac{tr(\bar{\mathbf{\Sigma}})}{p}}{\gamma_1}\right)'$$

converges weakly to a zero-mean Gaussian vector $\mathbf{w} = (w_1, \cdots, w_k)'$ *with covariance* $cov(w_i, w_j) = \delta_{ij}\frac{2\theta}{(2i-1)^4}$ *in which* $\theta = \lim_{p\to\infty}\frac{tr(\bar{\mathbf{\Sigma}}^2)}{p_1}$.

Furthermore, when $\lim\limits_{T\to\infty}\frac{p-p_1}{\sqrt{p}} = 0$*, the random vector*

$$\left(\sqrt{p}\frac{\rho_1 - \gamma_1\frac{tr(\mathbf{\Sigma})}{p}}{\gamma_1}, \cdots, \sqrt{p}\frac{\rho_k - \gamma_k\frac{tr(\mathbf{\Sigma})}{p}}{\gamma_1}\right)'$$

converges weakly to a zero-mean Gaussian vector $\mathbf{w} = (w_1, \cdots, w_k)'$ *with covariance* $cov(w_i, w_j) = \delta_{ij}\frac{2\theta}{(2i-1)^4}$*, in which* $\theta = \lim_{p\to\infty}\frac{tr(\mathbf{\Sigma}^2)}{p}$.

Proof of Theorem 7. In fact we only need to write $\mathbf{B} = \frac{1}{p}\mathbf{X}\mathbf{U}_{\mathbf{M1}}'\mathbf{U}_{\mathbf{M1}}\mathbf{X}^* + \frac{1}{p}\mathbf{X}\mathbf{U}_{\mathbf{R1}}'\mathbf{U}_{\mathbf{R1}}\mathbf{X}^*$. Then we can use Theorem 2.1 to study $\frac{1}{p}\mathbf{X}\mathbf{U}_{\mathbf{R1}}'\mathbf{U}_{\mathbf{R1}}\mathbf{X}^*$ and use Theorem 2.2 to study $\frac{1}{p}\mathbf{X}\mathbf{U}_{\mathbf{M1}}'\mathbf{U}_{\mathbf{M1}}\mathbf{X}^*$. From the differences between the orders of $\frac{1}{p}\mathbf{X}\mathbf{U}_{\mathbf{R1}}'\mathbf{U}_{\mathbf{R1}}\mathbf{X}^*$ and $\frac{1}{p}\mathbf{X}\mathbf{U}_{\mathbf{M1}}'\mathbf{U}_{\mathbf{M1}}\mathbf{X}^*$, we can complete the proof. □

Theorem 8. *Let Assumptions A1-A6 hold. The matrix* $\mathbf{\Pi}$ *is symmetric and* $rank(\mathbf{\Pi} - \mathbf{I}) = p - p_1$. *The eigenvalues of* $\mathbf{\Pi}$ *satisfy*

$$\lambda_1(\mathbf{\Pi}) = \cdots = \lambda_{p_1}(\mathbf{\Pi}) = 1 > \varphi \geq \lambda_{p_1+1}(\mathbf{\Pi}) \geq \lambda_p(\mathbf{\Pi}) \geq -\varphi.$$

There exists a $p\times p$ *matrix* $\mathbf{U_\Pi}$ *such that* $\mathbf{U_\Pi}\mathbf{U_\Pi}' = I$ *and* $\mathbf{\Pi} = \mathbf{U_\Pi} diag\{\mathbf{I}_{p_1}, \mathbf{\Pi_R}\}\mathbf{U_\Pi}'$*, where* $\mathbf{\Pi_R} = diag\{\lambda_{p_1+1}(\mathbf{\Pi}), \cdots, \lambda_p(\mathbf{\Pi})\}$*. We can write* $\mathbf{U_\Pi} = (\mathbf{U_{M1}}', \mathbf{U_{R1}}')'$*, where* $\mathbf{U_{M1}}$ *is a* $p_1 \times p$ *matrix and* U_{R1} *is a* $(p - p_1) \times p$ *matrix. Let* $\bar{\rho}_k$ *be the kth largest eigenvalue of* $\bar{\mathbf{B}}$ *with k fixed.*

(1) *Let* $\bar{\mathbf{\Sigma}} = \mathbf{U_{M1}}\mathbf{\Sigma}\mathbf{U}_{\mathbf{M1}}'$ *be a* $p_1 \times p_1$ *matrix. When* $\frac{tr(\bar{\mathbf{\Sigma}})}{p} \geq c > 0$,

$$\frac{\bar{\rho}_k - \bar{\gamma}_k\frac{tr(\bar{\mathbf{\Sigma}})}{p}}{\bar{\gamma}_1} \xrightarrow{i.p.} 0.$$

Furthermore, when $\lim\limits_{T\to\infty}\frac{p-p_1}{p}=0$,

$$\frac{\bar{\rho}_k-\bar{\gamma}_k\frac{tr(\boldsymbol{\Sigma})}{p}}{\bar{\gamma}_1}\xrightarrow{i.p.}0.$$

(2) When $\frac{tr(\bar{\boldsymbol{\Sigma}})}{p}\geq c>0$, *the random vector*

$$\left(\frac{p}{\sqrt{p_1}}\frac{\bar{\rho}_1-\bar{\gamma}_1\frac{tr(\bar{\boldsymbol{\Sigma}})}{p}}{\bar{\gamma}_1},\cdots,\frac{p}{\sqrt{p_1}}\frac{\bar{\rho}_k-\bar{\gamma}_k\frac{tr(\bar{\boldsymbol{\Sigma}})}{p}}{\bar{\gamma}_1}\right)'$$

converges weakly to a zero-mean Gaussian vector $\bar{\mathbf{w}}=(\bar{w}_1,\cdots,\bar{w}_k)'$ *with covariance* $cov(\bar{w}_i,\bar{w}_j)=\delta_{ij}\frac{2\theta}{i^4}$, *in which* $\theta=\lim_{p\to\infty}\frac{tr(\bar{\boldsymbol{\Sigma}}^2)}{p_1}$.

Furthermore, when $\lim\limits_{T\to\infty}\frac{p-p_1}{\sqrt{p}}=0$, *the random vector*

$$\left(\sqrt{p}\frac{\bar{\rho}_1-\bar{\gamma}_1\frac{tr(\boldsymbol{\Sigma})}{p}}{\bar{\gamma}_1},\cdots,\sqrt{p}\frac{\bar{\rho}_k-\bar{\gamma}_k\frac{tr(\boldsymbol{\Sigma})}{p}}{\bar{\gamma}_1}\right)'$$

converges weakly to a zero-mean Gaussian vector $\bar{\mathbf{w}}=(\bar{w}_1,\cdots,\bar{w}_k)'$ *with covariance* $cov(\bar{w}_i,\bar{w}_j)=\delta_{ij}\frac{2\theta}{i^4}$, *in which* $\theta=\lim_{p\to\infty}\frac{tr(\boldsymbol{\Sigma}^2)}{p}$.

The proof of Theorem 8 is similar to that of Theorem 7.

Theorem 9. *When* $\lim\limits_{T\to\infty}\frac{p-p_1}{\sqrt{p}}=0$, *Theorem 3.1 in the main paper still holds.*

Proof of Theorem 9. In fact we only need to study $\breve{\mathbf{x}}_{f,g}$ in this case. Write

$$\breve{x}_{f,g}=(\mathbf{x_f}-\mathbf{x_{f-1}})'\left(U'_{M1}U_{M1}+U'_{R1}U_{R1}\right)(\mathbf{x_g}-\mathbf{x_{g-1}}).$$

Since $p-p_1=o_p(p^{1/2})$, we can find that

$$E((\mathbf{x_f}-\mathbf{x_{f-1}})'U'_{M1}U_{M1}(\mathbf{x_g}-\mathbf{x_{g-1}}))=a_{|f-g|}tr(\bar{\boldsymbol{\Sigma}})=a_{|f-g|}tr(\boldsymbol{\Sigma})(1+o_p(p^{-1/2}))$$

and

$$Var((\mathbf{x_f}-\mathbf{x_{f-1}})'U'_{M1}U_{M1}(\mathbf{x_g}-\mathbf{x_{g-1}})=(a^2_{|f-g|}+a^2_0)tr(\bar{\boldsymbol{\Sigma}}^{\mathbf{2}}))$$
$$=(a^2_{|f-g|}+a^2_0)tr(\boldsymbol{\Sigma}^2)(1+o_p(p^{-1/2})).$$

Moreover, one can verify that

$$(\mathbf{x_f} - \mathbf{x_{f-1}})' U_{R1}' U_{R1} (\mathbf{x_g} - \mathbf{x_{g-1}}) = O_p(M_0(p - p_1)) = o_p(p^{1/2}).$$

So the two estimators still work in this case. The proof is complete.

$\square$

It is difficult to estimate $\gamma_1 \frac{tr\bar{\Sigma}}{p}$ in Theorem 7 or $\bar{\gamma}_1 \frac{tr\bar{\Sigma}}{p}$ in Theorem 8 and we do not know how to estimate them at this stage. As a consequence we have to pay a price and impose a condition that $(p - p_1)/\sqrt{p} \to 0$, which is a bit strict requirement.

From the above Theorems, we can find that the rank of $\mathbf{\Pi} - \mathbf{I}$ may play an important role. We below run some simulations to see this. We set the different matrices for $\mathbf{\Pi}$ and different ranks for $\mathbf{\Pi} - \mathbf{I}$ as follows:

$$\mathbf{\Pi_{c0}} = \mathbf{I} - \frac{\mathbf{11'}}{p}, \quad rank(\mathbf{\Pi_{c0}} - I) = 1, \tag{2.196}$$

$$\mathbf{\Pi_{c1}} = diag\{\mathbf{I_{p-[p/4]}}, 0.9\mathbf{I_{[p/4]}}\}, \quad rank(\mathbf{\Pi_{c1}} - I) = [p/4], \tag{2.197}$$

$$\mathbf{\Pi_{c2}} = diag\{\mathbf{I_{p-[p/2]}}, 0.9\mathbf{I_{[p/2]}}\}, \quad rank(\mathbf{\Pi_{c2}} - I) = [p/2], \tag{2.198}$$

and

$$\mathbf{\Pi_{c3}} = diag\{\mathbf{I_{p-[3p/4]}}, 0.9\mathbf{I_{[3p/4]}}\}, \quad rank(\mathbf{\Pi_{c3}} - I) = [3p/4], \tag{2.199}$$

We use the setting $\mathbf{y_t} = \psi \mathbf{z_{t-1}} + \mathbf{z_t}$ with $\psi = 0.5$ and $\mathbf{\Sigma}$ has the entries $\Sigma_{i,j} = 0.3^{|i-j|}$.

Observe that the finite–sample performance of $\mathbf{\Pi_{c0}}$ can be different when using an asymptotic critical value calculated from $N(0,1)$ and that based on the parametric bootstrap method when p is small. In fact although $rank(\mathbf{\Pi_{c0}} - \mathbf{I}) = 1$, $\frac{1}{\sqrt{p}}$ isn't small enough when p is 5 or 10. From the proofs of Theorems 7 and 9, we can find T_N (and $\bar{T}_N$) is the sum of two parts. When $\frac{p-p_1}{\sqrt{p}}$ isn't small enough, the second part(stationary part) can't be ignored and it leads to the power. Note that when p is small, if we use the critical value from $N(0,1)$, the power is very

Table 2.9: *The results of T_N*

p	T	$\mathbf{\Pi_{c0}}$(size)	$\mathbf{\Pi_{c1}}$(power)	$\mathbf{\Pi_{c2}}$(power)	$\mathbf{\Pi_{c3}}$(power)	$\mathbf{0.9I}$(power)
20	20	0.025	0.017	0.034	0.112	0.216
20	30	0.035	0.027	0.072	0.262	0.672
20	40	0.031	0.041	0.135	0.440	0.951
20	60	0.038	0.042	0.191	0.564	1.000
20	80	0.040	0.040	0.224	0.629	1.000
40	20	0.021	0.048	0.103	0.270	0.580
40	30	0.028	0.060	0.211	0.594	0.964
40	40	0.031	0.061	0.282	0.685	0.999
40	60	0.037	0.082	0.358	0.812	1.000
40	80	0.046	0.112	0.403	0.831	1.000
60	20	0.018	0.040	0.157	0.446	0.766
60	30	0.037	0.074	0.311	0.751	0.998
60	40	0.036	0.098	0.391	0.832	1.000
60	60	0.038	0.114	0.459	0.865	1.000
60	80	0.037	0.162	0.482	0.900	1.000
80	20	0.018	0.053	0.301	0.574	0.870
80	30	0.026	0.082	0.415	0.853	1.000
80	40	0.036	0.121	0.467	0.925	1.000
80	60	0.036	0.148	0.555	0.942	1.000
80	80	0.051	0.169	0.604	0.944	1.000

Table 2.10: *The results of T_N with the parametric bootstrap method*

p	T	$\mathbf{\Pi_{c0}}$(size)	$\mathbf{\Pi_{c1}}$(power)	$\mathbf{\Pi_{c2}}$(power)	$\mathbf{\Pi_{c3}}$(power)	$\mathbf{0.9I}$(power)
20	20	0.043	0.097	0.223	0.458	0.636
20	30	0.054	0.128	0.352	0.712	0.974
20	40	0.066	0.187	0.425	0.774	0.998
20	60	0.077	0.217	0.463	0.842	1.000
20	80	0.070	0.162	0.535	0.854	1.000
40	20	0.053	0.152	0.337	0.568	0.860
40	30	0.056	0.198	0.467	0.826	0.990
40	40	0.042	0.194	0.499	0.844	1.000
40	60	0.053	0.225	0.583	0.924	1.000
40	80	0.063	0.248	0.613	0.910	1.000
60	20	0.038	0.173	0.358	0.694	0.930
60	30	0.046	0.236	0.600	0.906	1.000
60	40	0.055	0.240	0.627	0.946	1.000
60	60	0.037	0.282	0.674	0.954	1.000
60	80	0.054	0.255	0.678	0.958	1.000
80	20	0.053	0.206	0.541	0.798	0.950
80	30	0.054	0.284	0.657	0.940	1.000
80	40	0.055	0.269	0.726	0.972	1.000
80	60	0.060	0.303	0.688	0.978	1.000
80	80	0.064	0.321	0.756	0.980	1.000

Table 2.11: *The results of T_N and small p*

p	T	$\mathbf{\Pi_{c0}}$(size)	$\mathbf{\Pi_{c1}}$(power)	$\mathbf{\Pi_{c2}}$(power)	$\mathbf{\Pi_{c3}}$(power)	$\mathbf{0.9I}$(power)
5	20	0.046	0.028	0.019	0.012	0.008
5	30	0.054	0.025	0.013	0.008	0.010
5	40	0.039	0.013	0.002	0.002	0.012
5	60	0.052	0.029	0.013	0.002	0.006
5	80	0.048	0.030	0.011	0.006	0.002
10	20	0.023	0.019	0.017	0.014	0.048
10	30	0.030	0.025	0.025	0.058	0.142
10	40	0.031	0.029	0.036	0.062	0.322
10	60	0.030	0.011	0.038	0.112	0.558
10	80	0.036	0.029	0.057	0.126	0.678

small. While if we use the critical value from the parametric bootstrap method, the power is much bigger (also see Tables 2.4 and 2.6 in main paper). So the performances are different when p is small.

The performances of $\mathbf{\Pi_{c1}}$, $\mathbf{\Pi_{c2}}$ and $\mathbf{\Pi_{c3}}$ show that the stationary part plays a more important role when $\frac{p-p_1}{\sqrt{p}}$ becomes larger.

Appendix D.2: Some extensive discussions about deterministic components

This section is to discuss how to deal with the deterministic components of the time series and the difficulties when establishing general theory for them.

There is another way of presenting models containing deterministic components. Specifically speaking, write

$$\mathbf{x_t} = \phi + \mathbf{\Pi x_{t-1}} + \mathbf{\Sigma}^{1/2}\mathbf{y_t},\ 1 \le t \le T. \tag{2.200}$$

Table 2.12: *The results of T_N with the parametric bootstrap method and small p*

p	T	$\mathbf{\Pi_{c0}}$(size)	$\mathbf{\Pi_{c1}}$(power)	$\mathbf{\Pi_{c2}}$(power)	$\mathbf{\Pi_{c3}}$(power)	$\mathbf{0.9I}$(power)
5	20	0.111	0.095	0.126	0.154	0.308
5	30	0.125	0.141	0.179	0.258	0.580
5	40	0.133	0.112	0.175	0.328	0.772
5	60	0.122	0.126	0.208	0.342	0.938
5	80	0.112	0.103	0.257	0.392	0.992
10	20	0.069	0.103	0.187	0.236	0.462
10	30	0.068	0.116	0.269	0.462	0.826
10	40	0.072	0.137	0.341	0.516	0.972
10	60	0.087	0.139	0.354	0.592	1.000
10	80	0.101	0.137	0.379	0.642	1.000

Table 2.13: *The results of $\bar{T}_N$*

p	T	$\mathbf{\Pi_{c0}}$(size)	$\mathbf{\Pi_{c1}}$(power)	$\mathbf{\Pi_{c2}}$(power)	$\mathbf{\Pi_{c3}}$(power)	$\mathbf{0.9I}$(power)
20	20	0.019	0.010	0.014	0.012	0.013
20	30	0.023	0.018	0.024	0.044	0.124
20	40	0.035	0.018	0.030	0.106	0.290
20	60	0.032	0.024	0.084	0.259	0.746
20	80	0.044	0.040	0.115	0.405	0.959
40	20	0.016	0.008	0.018	0.046	0.075
40	30	0.032	0.014	0.046	0.124	0.290
40	40	0.030	0.023	0.097	0.215	0.584
40	60	0.046	0.050	0.185	0.514	0.985
40	80	0.031	0.060	0.282	0.648	1.000
60	20	0.020	0.018	0.038	0.056	0.144
60	30	0.030	0.027	0.088	0.205	0.523
60	40	0.030	0.037	0.144	0.351	0.823
60	60	0.033	0.073	0.258	0.603	0.999
60	80	0.043	0.092	0.360	0.775	1.000
80	20	0.013	0.016	0.050	0.082	0.191
80	30	0.026	0.033	0.109	0.261	0.661
80	40	0.032	0.055	0.179	0.483	0.934
80	60	0.040	0.084	0.330	0.735	1.000
80	80	0.038	0.111	0.386	0.803	1.000

Table 2.14: *The results of $\bar{T}_N$ with the parametric bootstrap method*

p	T	$\mathbf{\Pi_{c0}}$(size)	$\mathbf{\Pi_{c1}}$(power)	$\mathbf{\Pi_{c2}}$(power)	$\mathbf{\Pi_{c3}}$(power)	$\mathbf{0.9I}$(power)
20	20	0.056	0.050	0.112	0.190	0.239
20	30	0.071	0.106	0.184	0.376	0.606
20	40	0.072	0.128	0.282	0.544	0.837
20	60	0.065	0.158	0.418	0.737	0.986
20	80	0.071	0.158	0.458	0.778	1.000
40	20	0.028	0.088	0.164	0.252	0.352
40	30	0.055	0.122	0.256	0.439	0.695
40	40	0.053	0.138	0.310	0.564	0.873
40	60	0.060	0.168	0.470	0.783	0.999
40	80	0.066	0.214	0.562	0.866	1.000
60	20	0.056	0.086	0.204	0.288	0.452
60	30	0.047	0.152	0.296	0.525	0.807
60	40	0.050	0.174	0.378	0.692	0.961
60	60	0.040	0.218	0.500	0.838	1.000
60	80	0.057	0.252	0.590	0.908	1.000
80	20	0.038	0.120	0.210	0.340	0.512
80	30	0.046	0.130	0.315	0.597	0.903
80	40	0.064	0.180	0.416	0.713	0.986
80	60	0.047	0.215	0.588	0.884	1.000
80	80	0.050	0.286	0.597	0.887	1.000

Table 2.15: *The results of $\bar{T}_N$ and small p*

p	T	$\mathbf{\Pi_{c0}}$(size)	$\mathbf{\Pi_{c1}}$(power)	$\mathbf{\Pi_{c2}}$(power)	$\mathbf{\Pi_{c3}}$(power)	$\mathbf{0.9I}$(power)
5	20	0.057	0.030	0.024	0.016	0.008
5	30	0.056	0.032	0.022	0.012	0.002
5	40	0.054	0.032	0.022	0.006	0.000
5	60	0.055	0.040	0.018	0.004	0.000
5	80	0.049	0.038	0.030	0.000	0.000
10	20	0.024	0.028	0.006	0.026	0.009
10	30	0.028	0.034	0.012	0.014	0.018
10	40	0.036	0.022	0.010	0.016	0.031
10	60	0.049	0.022	0.010	0.022	0.117
10	80	0.038	0.030	0.012	0.044	0.217

To deal with ϕ, we define

$$\tilde{\tilde{\mathbf{x}}}_{\mathbf{t}} = \mathbf{x_t} - \frac{t}{T}\mathbf{x_T}. \tag{2.201}$$

Under H_0 we can find that

$$\tilde{\tilde{\mathbf{x}}}_{\mathbf{t}} = \tilde{\tilde{\mathbf{x}}}_{\mathbf{t-1}} + \mathbf{\Sigma}^{1/2}(\mathbf{y_t} - \bar{\mathbf{y}}),\ 1 \leq t \leq T, \tag{2.202}$$

where $\bar{\mathbf{y}} = \frac{\sum_{t=1}^{T}\mathbf{y_t}}{T}$. Let $\tilde{\tilde{\mathbf{X}}} = (\tilde{\tilde{\mathbf{x}}}_{\mathbf{1}}, \cdots, \tilde{\tilde{\mathbf{x}}}_{\mathbf{T}})'$, we can find that

$$(1/p)\tilde{\tilde{\mathbf{X}}}\tilde{\tilde{\mathbf{X}}}^* = (1/p)\mathbf{CHY\Sigma Y^*HC^*}, \tag{2.203}$$

where $\mathbf{H} = \mathbf{I} - \mathbf{11}'/T$. It turns out that the largest eigenvalues of $(1/p)\mathbf{CHY\Sigma Y^*HC^*}$ have the same limit distributions as those of $(1/p)\mathbf{HCY\Sigma Y^*C^*H}$. So Theorems 2.3 and 2.4 hold for $(1/p)\tilde{\tilde{\mathbf{X}}}\tilde{\tilde{\mathbf{X}}}^*$. The sample covariance matrix $(1/p)\tilde{\tilde{\mathbf{X}}}\tilde{\tilde{\mathbf{X}}}^*$ can also remove the influence of ϕ under H_0. We can propose a test $\tilde{T}_N$ which is

Table 2.16: *The results of $\bar{T}_N$ with the parametric bootstrap method and small p*

p	T	$\mathbf{\Pi_{c0}}$(size)	$\mathbf{\Pi_{c1}}$(power)	$\mathbf{\Pi_{c2}}$(power)	$\mathbf{\Pi_{c3}}$(power)	$\mathbf{0.9I}$(power)
5	20	0.079	0.066	0.092	0.106	0.155
5	30	0.099	0.074	0.100	0.146	0.284
5	40	0.111	0.106	0.092	0.188	0.417
5	60	0.094	0.112	0.132	0.260	0.712
5	80	0.110	0.114	0.196	0.326	0.896
10	20	0.045	0.078	0.114	0.158	0.218
10	30	0.080	0.096	0.166	0.257	0.450
10	40	0.091	0.108	0.240	0.372	0.653
10	60	0.102	0.134	0.309	0.524	0.904
10	80	0.085	0.114	0.326	0.608	0.991

similar to T_N and $\bar{T}_N$. Moreover one can prove that Theorem 3.1 still holds for $\tilde{T}_N$. Indeed, its size is good as one can see from Table 13 below. Unfortunately, under H_1, $(1/p)\tilde{\mathbf{X}}\tilde{\mathbf{X}}^*$ can't remove the influence of ϕ. As a consequence, the power of the test may behave differently for different values of ϕ under H_1.

We run simulations for $\tilde{T}_N$ with the critical value from the limit distribution $N(0,1)$. The parameters are set to be the same as those in Tables 2.4, 2.6 and 2.8 in the main paper. The results of $\tilde{T}_N$ on 1000 replications and different values of p, T and $\mathbf{\Pi}$ are reported in Table 2.13. The nominal size is set to be 0.05. One can find that the size is good but the power is chaotic for different values of ϕ.

Alternatively, instead of using (2.201) we may also let $\hat{\mathbf{x}}_\mathbf{t} = \mathbf{x}_\mathbf{t} - \bar{\mathbf{x}} - (t - \frac{T+1}{2})\frac{2}{T-1}(\mathbf{x}_\mathbf{T} - \bar{\mathbf{x}})$, where $\bar{\mathbf{x}} = \frac{\sum_{t=1}^T \mathbf{x}_\mathbf{t}}{T}$. In this way, the influence of ϕ under both H_0 and H_1 can be removed. Under H_0,

$$\hat{\mathbf{x}}_\mathbf{t} = \hat{\mathbf{x}}_\mathbf{t-1} + \mathbf{\Sigma}^{1/2}\mathbf{y}_\mathbf{t} + \frac{2}{T-1}(T\mathbf{\Sigma}^{1/2}\bar{\mathbf{y}} - \sum_{i=1}^{T}\frac{T-i+1}{T}\mathbf{\Sigma}^{1/2}\mathbf{y}_\mathbf{i}),\ 1 \le t \le T. \quad (2.204)$$

Let $\hat{\mathbf{X}} = (\hat{\mathbf{x}}_\mathbf{1}, \cdots, \hat{\mathbf{x}}_\mathbf{T})'$. Note that $(1/p)\hat{\mathbf{X}}\hat{\mathbf{X}}^*$ also has the form

$$(1/p)\hat{\mathbf{X}}\hat{\mathbf{X}}^* = (1/p)\mathbf{C}\mathbf{H}_\mathbf{s}\mathbf{Y}\mathbf{\Sigma}\mathbf{Y}^*\mathbf{H}_\mathbf{s}^*\mathbf{C}^*, \quad (2.205)$$

where the definition of H_s comes from (2.204). Unfortunately, the property of $CH_sH_s^*C$ is more difficult to study and we have not obtained any result about it yet. Hence we fail to establish CLT for the largest eigenvalues of $(1/p)\hat{\mathbf{X}}\hat{\mathbf{X}}^*$.

A more complicated model can be considered similarly.

$$\mathbf{x}_\mathbf{t} = \alpha + \beta t + \mathbf{\Pi}\mathbf{x}_\mathbf{t-1} + \mathbf{\Sigma}^{1/2}\mathbf{y}_\mathbf{t},\ 1 \le t \le T. \quad (2.206)$$

Let $\mathbf{z}_\mathbf{t} = \mathbf{x}_\mathbf{t} - \mathbf{x}_\mathbf{t-1} - \frac{\mathbf{x}_\mathbf{T}}{T} - \frac{\mathbf{x}_\mathbf{T} - \mathbf{x}_\mathbf{T-1} - \mathbf{x}_\mathbf{1}}{T-1}(t - \frac{T+1}{2})$. It can remove the influence of α and β under both H_0 and H_1. Let $\tilde{\mathbf{z}}_\mathbf{t} = \tilde{\mathbf{z}}_\mathbf{t-1} + \mathbf{z}_\mathbf{t}$. Under H_0,

$$\tilde{\mathbf{z}}_\mathbf{t} = \tilde{\mathbf{z}}_\mathbf{t-1} + \mathbf{\Sigma}^{1/2}\mathbf{y}_\mathbf{t} - \mathbf{\Sigma}^{1/2}\bar{\mathbf{y}} - \frac{t - \frac{T+1}{2}}{T-1}\mathbf{\Sigma}^{1/2}(\mathbf{y}_\mathbf{T} - \mathbf{y}_\mathbf{1}).\ 1 \le t \le T, \quad (2.207)$$

Let $\tilde{\mathbf{Z}} = (\tilde{\mathbf{z}}_\mathbf{1}, \cdots, \tilde{\mathbf{z}}_\mathbf{T})'$. Define the sample covariance matrix:

$$(1/p)\tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^* = (1/p)\mathbf{C}\mathbf{H}_\mathbf{v}\mathbf{Y}\mathbf{\Sigma}\mathbf{Y}^*\mathbf{H}_\mathbf{v}^*\mathbf{C}^*, \quad (2.208)$$

Table 2.17: *The results for $\tilde{T}_N$ and MA(1)*

p	T	$\mathbf{I}$(size)	$\mathbf{0.95I}$(power)	$\mathbf{0.9I}$(power)	Π_2(power)
10	20	0.037	1.000	0.873	0.802
10	30	0.044	1.000	0.705	0.700
10	40	0.055	1.000	0.469	0.582
10	60	0.045	1.000	0.113	0.312
10	80	0.058	1.000	0.015	0.155
20	20	0.033	1.000	0.996	0.934
20	30	0.040	1.000	1.000	0.922
20	40	0.033	1.000	1.000	0.829
20	60	0.050	1.000	1.000	0.699
20	80	0.044	1.000	0.996	0.425
40	20	0.015	1.000	1.000	0.996
40	30	0.027	1.000	1.000	0.997
40	40	0.032	1.000	1.000	0.987
40	60	0.042	1.000	1.000	0.955
40	80	0.044	1.000	0.998	0.878
60	20	0.011	1.000	1.000	0.956
60	30	0.021	1.000	1.000	0.924
60	40	0.030	1.000	1.000	0.897
60	60	0.035	1.000	1.000	0.725
60	80	0.043	1.000	1.000	0.486
80	20	0.009	1.000	1.000	0.993
80	30	0.021	1.000	1.000	0.996
80	40	0.027	1.000	1.000	0.990
80	60	0.031	1.000	1.000	0.945
80	80	0.036	1.000	1.000	0.869

where the definition of H_v comes from (2.207). Unfortunately, studying the property of $CH_vH_v^*C$ is even more challenging. We haven't obtained any result for $(1/p)\tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^*$ yet. We are going to further pursue this in future research, however.

Chapter 3

Central Limit Theorem for the spiked Eigenvalues of Separable Sample Covariance Matrices

3.1 Introduction

Most of the existing studies in random matrix field rely on the assumption that the observations of high dimensional data are independent, although dimensional correlation structure can be allowed. However, high–dimensional data in economics and finance are often highly dependent across time. We can't assume that the fluctuation of the stock market today is independent with the fluctuation yesterday. In view of this, Zhang (2006) investigated the empirical spectral distribution (ESD) of the sample covariance for the case where the data matrices are of the form $\mathbf{A}_1\mathbf{Z}\mathbf{A}_2$, where $\mathbf{A}_1$ and $\mathbf{A}_2$ are positive semidefinite matrices and $\mathbf{Z}$ has independent entries satisfying some moment assumptions. This model is referred to as the separable covariance model and allows for some dependence among observations recorded over different time points.

In Chapter 2 we study the CLT of the largest eigenvalues for high-dimensional non-stationary time series. However it cannot handle the case when time series data contain a deterministic trend. In view of this we consider a kind of separable sample covariance matrices. Recalling the model in (2.33), we can find that the largest eigenvalues of $\mathbf{CFF}^*\mathbf{C}^*$ are very big and $\boldsymbol{\Sigma}$ is bounded. A direct question is whether the CLT can be extended to the separable sample covariance matrices with this form. In this Chapter, we study this question.

The Chapter will be organized as follows. Section 3.2 establishes the CLTs for two different cases. Section 3.3 provides two models in time series as two examples of the theoretical results. Section 3.4 and Section 3.5 prove the theoretical results in Section 3.2.

3.2 Theoretical Result

We consider the separable sample covariance matrix $\Gamma\mathbf{X}\boldsymbol{\Omega}\mathbf{X}^T\Gamma^T$, where $\Sigma_1 = \Gamma\Gamma^T$, $\mathbf{X} = (\mathbf{x}_1, ..., \mathbf{x}_n) = (\mathbf{x}_{ij})_{p\times n}$ and Γ are $p \times n$ and $(p-L) \times p$ matrices. Moreover, let the spectral decomposition of Γ be $\mathbf{V}\Lambda^{1/2}\mathbf{U}$, where $\mathbf{V}$ and $\mathbf{U}$ are $(p-L)\times(p-L)$ and $(p-L)\times p$ orthogonal matrices respectively($\mathbf{V}\mathbf{V}^T = \mathbf{U}\mathbf{U}^T = \mathbf{I}$), Λ is a diagonal matrix consisting of the descent ordered eigenvalues of Σ_1. Moreover, suppose that there are K eigenvalues of Σ_1 much bigger than others. Then we can write $\Lambda = \begin{pmatrix} \Lambda_S & 0 \\ 0 & \Lambda_P \end{pmatrix}$, where $\Lambda_S = diag(\mu_1, ..., \mu_K)$ and $\Lambda_P = diag(\mu_{K+1}, ..., \mu_{p-L})$ and $\mu_1, ..., \mu_K$ are the K spiked eigenvalues. In addition, we write $\mathbf{U} = \begin{pmatrix} \mathbf{U}_1 \\ \mathbf{U}_2 \end{pmatrix}$ and $\Sigma = \mathbf{U}_2^T\Lambda_P\mathbf{U}_2$. We focus on the asymptotic property of the largest K eigenvalues of $\Gamma\mathbf{X}\boldsymbol{\Omega}\mathbf{X}^T\Gamma^T$.

We need to make the following assumptions:

Assumption 1. $\mathbf{x}_j$, $j = 1, ..., n$ *are i.i.d. random vectors.* $\{\mathbf{x}_{ij}$: $i = 1, ..., p$, $j = 1, ..., n\}$ *are independent random variables with finite fourth moment such* *that* $\mathbb{E}\mathbf{x}_{ij} = 0$. $\mathbb{E}|\sqrt{n}\mathbf{x}_{ij}|^2 = 1$ *and* $\mathbb{E}|\sqrt{n}\mathbf{x}_{ij}|^4 = \gamma_{4i}$.

Assumption 2. *There exists* δ_n *satisfying*

$$\lim_{n\to\infty} \delta_n^{-4} \frac{1}{np} \sum_{i=1}^{p} \sum_{j=1}^{n} \mathbb{E}|\sqrt{n}\mathbf{x}_{ij}|^4 I(|\sqrt{n}\mathbf{x}_{ij}| > \delta_n \sqrt[4]{np}) = 0, \quad \delta_n \downarrow 0, \quad \delta_n \sqrt[4]{np} \uparrow \infty \tag{3.1}$$

Assumption 3. $\mathbf{\Omega}$ *is nonnegative. There exist two positive constants* M_0 *and* M_1 *such that* $\|\mathbf{\Omega}\|_2 \le M_0$ *and* $\frac{tr(\mathbf{\Omega})}{n} \ge M_1$.

Assumption 4. *The spiked eigenvalues of population covariance matrix are much bigger than others. Precisely speaking, for* $\forall \varepsilon > 0$, *there is* K_ε *which doesn't depend on* n *and* p *such that when* n *and* p *are big enough,*

$$\frac{\mu_{K_\varepsilon}(1+\sqrt{\frac{p}{n}})^2 M_0}{\mu_K \frac{tr(\mathbf{\Omega})}{n}} < \varepsilon. \tag{3.2}$$

Assumption 5. *The spiked eigenvalues of the population covariance matrix are much bigger than others. Precisely speaking, for* $\forall\varepsilon > 0$, *there is* K_ε *which doesn't depend on* n *and* p *such that when* n *and* p *are big enough,*

$$\frac{\sum_{i=K_\varepsilon}^{p-L} \mu_i M_0}{\mu_K \frac{tr(\mathbf{\Omega})}{n}} < \frac{\varepsilon}{2}. \tag{3.3}$$

Remark 8. *When* $\frac{n}{p} \to 0$, $\mu_i = i^{-1-\sigma}$ *and* $\sigma > 0$. *We can find that the Assumption 5 holds but the Assumption 4 doesn't hold. Conversely, When* $\frac{n}{p} \to c > 0$, $\mathbf{\Omega} = \mathbf{I}$, $\mu_K = \sqrt{p}$ *and* $\mu_i = 1$ *for any* $i > K$, *then the Assumption 4 holds but the Assumption 5 doesn't hold.*

Assumption 6. *There exists a small positive constant* c *such that* $\mu_K - \mu_{K+1} \ge c\mu_K$.

Assumption 7. *The spiked eigenvalues of the population covariance matrix are much bigger than* $\max\{1, \frac{p}{n}\}$. *Precisely speaking, the eigenvalues* $\mu_i = \frac{\max\{1,\frac{p}{n}\}}{d_i}$, *where* $d_i \to 0$, $i = 1, 2, ..., K$. *And for* $i = K+1, ..., p-L$, μ_i *are bounded by a constant* C.

Assumption 8. *$\alpha_{\mathcal{L}} = \mu_K = ... = \mu_{K-n_{\mathcal{L}}} < \alpha_{\mathcal{L}-1} = \mu_{K-n_{\mathcal{L}}+1} \cdots < \alpha_1 = \mu_{n_1} = ... = \mu_1$, Moreover, there exists a small constant c such that $\alpha_{i-1} - \alpha_i \geq c\alpha_i$, $i = 1, 2, ..., \mathcal{L}$. n_1,..., $n_{\mathcal{L}}$ are finite.*

Remark 9. *In fact, the Assumption 7 implies Assumptions 4 and 6.*

Denoting the i-th largest eigenvalue of $\Gamma\mathbf{X}\mathbf{\Omega}\mathbf{X}^T\Gamma^T$ by λ_i, we first prove the following proposition to ensure the convergence in probability:

Proposition 8. *For $\Gamma\mathbf{X}\mathbf{\Omega}\mathbf{X}^T\Gamma^T$ satisfying Assumptions 1, 3 and 4 or Assumptions 1, 3 and 5, the i-th largest eigenvalue λ_i converges to μ_i in probability after rescaling. Precisely speaking, we have uniformly for all $i = 1, ..., K$*

$$\frac{\lambda_i - \mu_i \frac{tr\mathbf{\Omega}}{n}}{\mu_i \frac{tr\mathbf{\Omega}}{n}} \to 0, \textit{in probability.} \tag{3.4}$$

Proof. Recall K_ε in the Assumption 4. For any $\varepsilon > 0$, we write $\Lambda = \begin{pmatrix} \Lambda_{S,\varepsilon} & 0 \\ 0 & \Lambda_{P,\varepsilon} \end{pmatrix}$, where $\Lambda_{S,\varepsilon} = diag(\mu_1, ..., \mu_{K_\varepsilon-1})$ and $\Lambda_{P,\varepsilon} = diag(\mu_{K_\varepsilon}, ..., \mu_{M-L})$. Note that the non-zero eigenvalues of $\Gamma\mathbf{X}\mathbf{\Omega}\mathbf{X}^T\Gamma^T$ equal those of $\mathbf{U}\mathbf{X}\mathbf{\Omega}\mathbf{X}^T\mathbf{U}^T\Lambda$. By weyl's inequality, we have

$$|\sigma_i(\Lambda^{1/2}\mathbf{U}\mathbf{X}\mathbf{\Omega}^{1/2}) - \sigma_i(\begin{pmatrix} \Lambda_{S,\varepsilon}^{1/2} & 0 \\ 0 & 0 \end{pmatrix}\mathbf{U}\mathbf{X}\mathbf{\Omega}^{1/2})| \leq \sigma_1(\begin{pmatrix} 0 & 0 \\ 0 & \Lambda_{P,\varepsilon}^{1/2} \end{pmatrix}\mathbf{U}\mathbf{X}\mathbf{\Omega}^{1/2}),$$

where $\sigma_i(\mathbf{A})$ is the i-th largest singular value of $\mathbf{A}$. Let $\mathbf{P}_\varepsilon = \begin{pmatrix} 0 & 0 \\ 0 & \Lambda_{P,\varepsilon}^{1/2} \end{pmatrix}\mathbf{U}\mathbf{X}\Omega^{1/2}$. By the results in Bai and Silverstein (2006) and Chen and Pan (2012), under Assumption 3, with probability tending to 1, we have $\|\mathbf{P}_\varepsilon{\mathbf{P}_\varepsilon}^T\| \leq \mu_{K_\varepsilon}(1+\sqrt{\frac{p}{n}})^2 M_0$.

We can also find that $\|\mathbf{P}_\varepsilon{\mathbf{P}_\varepsilon}^T\| \leq tr(\mathbf{P}_\varepsilon{\mathbf{P}_\varepsilon}^T)$ and $tr(\mathbf{P}_\varepsilon{\mathbf{P}_\varepsilon}^T) \leq 2\sum_{j=K_\varepsilon}^{M-L}\mu_j M_0$ with probability tending to 1. Then we find that when (3.2) or (3.3) holds, $\frac{\|\mathbf{P}_\varepsilon{\mathbf{P}_\varepsilon}^T\|}{\mu_K\frac{tr(\mathbf{\Omega})}{n}} \leq \varepsilon$ with probability tending to 1.

Moreover, we define $\mathbf{S}_\varepsilon = \begin{pmatrix} \Lambda_{S,\varepsilon}^{1/2} & 0 \\ 0 & 0 \end{pmatrix}\mathbf{U}\mathbf{X}\mathbf{\Omega}^{1/2}$. by Theorem 7.1 of Bai and Yao (2008), we can show that $\frac{\sigma_i(S_\varepsilon)^2}{\mu_i\frac{tr\mathbf{\Omega}}{n}} - 1 = o_p(1)$.

We then conclude that

$$\lim_{p,n\to\infty} P(|\frac{\lambda_i}{\mu_i \frac{tr\boldsymbol{\Omega}}{n}} - 1| \leq \varepsilon) = 1 \tag{3.5}$$

for any $\varepsilon > 0$. (3.4) is proved.

□

Theorem 10. *Suppose that Assumptions 1-3 and 7-8 hold. Define an event*

$$\mathfrak{E}_d = \{\|\boldsymbol{\Omega}^{1/2}\mathbf{X}^T\Sigma\mathbf{X}\boldsymbol{\Omega}^{1/2}\| \leq 4CM_0(1+\frac{p}{n})\}. \tag{3.6}$$

Let θ_i be the solution to

$$\mathbb{E}(\mathbf{u}_i^T\mathbf{X}\boldsymbol{\Omega}^{1/2}(\theta_i\mathbf{I} - \boldsymbol{\Omega}^{1/2}\mathbf{X}^T\Sigma\mathbf{X}\boldsymbol{\Omega}^{1/2})^{-1}\boldsymbol{\Omega}^{1/2}\mathbf{X}^T\mathbf{u}_i I(\mathfrak{E}_d)) = \alpha_i^{-1}.$$

Denote $m_i = \sum_{j=1}^{i-1} n_j$, for all $i = 1, 2, ..., \mathcal{L}$. We have

$$\frac{\sqrt{n}}{\alpha_i}(\lambda_{m_i+1} - \theta_i, \lambda_{m_i+2} - \theta_i, ..., \lambda_{m_i+n_i} - \theta_i) \xrightarrow{d} \mathcal{R}_i, \tag{3.7}$$

where $\mathcal{R}_i$ are the eigenvalues of $n_i \times n_i$ matrix $\mathfrak{R}_i$ with zero-mean Gaussian entries and the covariance of the $(\mathfrak{R}_i)_{k_1,l_1}$ and $(\mathfrak{R}_i)_{k_2,l_2}$ is

$$\lim_{n\to\infty} n^2 \times Cov(\mathbf{u}_{m_i+k_1}^T\mathbf{X}\boldsymbol{\Omega}\mathbf{X}^T\mathbf{u}_{m_i+l_1}, \mathbf{u}_{m_i+k_2}^T\mathbf{X}\boldsymbol{\Omega}\mathbf{X}^T\mathbf{u}_{m_i+l_2}). \tag{3.8}$$

Theorem 11. *Suppose that Assumptions 1, 3, 5, 6 and 8 hold. Denote $m_i = \sum_{j=1}^{i-1} n_j$, for all $i = 1, 2, ..., \mathcal{L}$. We have*

$$\frac{\sqrt{n}}{\alpha_i}(\lambda_{m_i+1} - \alpha_i\frac{tr\boldsymbol{\Omega}}{n}, \lambda_{m_i+2} - \alpha_i\frac{tr\boldsymbol{\Omega}}{n}, ..., \lambda_{m_i+n_i} - \alpha_i\frac{tr\boldsymbol{\Omega}}{n}) \xrightarrow{d} \mathcal{R}_i, \tag{3.9}$$

where $\mathcal{R}_i$ are the eigenvalues of $n_i \times n_i$ matrix $\mathfrak{R}_i$ with zero-mean Gaussian entries and the covariance of the $(\mathfrak{R}_i)_{k_1,l_1}$ and $(\mathfrak{R}_i)_{k_2,l_2}$ is

$$\lim_{n\to\infty} n^2 \times Cov(\mathbf{u}_{m_i+k_1}^T\mathbf{X}\boldsymbol{\Omega}\mathbf{X}^T\mathbf{u}_{m_i+l_1}, \mathbf{u}_{m_i+k_2}^T\mathbf{X}\boldsymbol{\Omega}\mathbf{X}^T\mathbf{u}_{m_i+l_2}). \tag{3.10}$$

Remark 10. *Actually, it follows from (3.62) that*

$$\frac{\theta_i}{\alpha_i} = \frac{tr\boldsymbol{\Omega}}{n} + \bar{c}c_i \cdot \frac{tr\boldsymbol{\Omega}^2}{n} + O(\frac{p}{n}\alpha_i^{-2}) + o(\frac{1}{\sqrt{n}}),$$

where

$$\bar{c} = \frac{1}{p-K-L}\sum_{j=K+1}^{p-L}\mu_j,\ c_i = \frac{p-K-L}{n\alpha_i}.$$

Remark 11. *By Theorem 7.2 of Bai and Yao (2008) it is not hard to calculate (3.8) and (3.10).*

When the population spiked eigenvalues are simple, we can conclude that the limiting distribution of extreme eigenvalues are Gaussian.

Proposition 9. *Suppose the conditions of Theorem 10 hold. Moveover,* $n_{k_1} = n_{k_2} = \cdots = n_{k_j} = 1$*. Define an event We have*

$$\sqrt{n}(\frac{\lambda_{m_{k_1}+1} - \theta_{k_1}}{\alpha_{k_1}}, \frac{\lambda_{m_{k_2}+1} - \theta_{k_2}}{\alpha_{k_2}}, ..., \frac{\lambda_{m_{k_j}+1} - \theta_{k_j}}{\alpha_{k_j}}) \xrightarrow{d} \mathcal{R}, \tag{3.11}$$

where $\mathcal{R}$ *are a vector with zero-mean Gaussian entries and the covariance of the* $\mathfrak{R}_f$ *and* $\mathfrak{R}_g$ *is*

$$\lim_{n\to\infty} n^2 \times Cov(\mathbf{u}_{m_{k_f}+1}^T\mathbf{X\Omega X}^T\mathbf{u}_{m_{k_f}+1}, \mathbf{u}_{m_{k_g}+1}^T\mathbf{X\Omega X}^T\mathbf{u}_{m_{k_g}+1}). \tag{3.12}$$

Proposition 10. *Suppose the conditions of Theorem 11 hold. Moveover,* $n_{k_1} = n_{k_2} = \cdots = n_{k_j} = 1$*. Define an event We have*

$$\sqrt{n}(\frac{\lambda_{m_{k_1}+1} - \alpha_{k_1}\frac{tr\mathbf{\Omega}}{n}}{\alpha_{k_1}}, \frac{\lambda_{m_{k_2}+1} - \alpha_{k_2}\frac{tr\mathbf{\Omega}}{n}}{\alpha_{k_2}}, ..., \frac{\lambda_{m_{k_j}+1} - \alpha_{k_j}\frac{tr\mathbf{\Omega}}{n}}{\alpha_{k_j}}) \xrightarrow{d} \mathcal{R}, \tag{3.13}$$

where $\mathcal{R}$ *are a vector with zero-mean Gaussian entries and the covariance of the* $\mathfrak{R}_f$ *and* $\mathfrak{R}_g$ *is*

$$\lim_{n\to\infty} n^2 \times Cov(\mathbf{u}_{m_{k_f}+1}^T\mathbf{X\Omega X}^T\mathbf{u}_{m_{k_f}+1}, \mathbf{u}_{m_{k_g}+1}^T\mathbf{X\Omega X}^T\mathbf{u}_{m_{k_g}+1}). \tag{3.14}$$

3.3 Applications in time series data

The data in time series often have the dependence over both the time and the cross-section. So the separable sample covariance may have potential applications in time series. In this section, we give two models in time series. One model is a special case for Theorem 10 and another is a special case for Theorem 11.

3.3.1 Stationary linear processes with factors

Consider a stationary linear process as follows.

$$y_{it} = \sum_{s=0}^{\infty} b_s z_{i,t-s} \tag{3.15}$$

and

$$z_{it} = \sum_{s=1}^{K} \phi_{is} f_{st} + \sum_{s=1}^{p-K} \Upsilon_{is} \zeta_{st}, \tag{3.16}$$

where $1 \le i \le p - K - L$ and $1 \le t \le n$.

We give some assumptions for the linear process.

Assumption 9. $\sum_{i=0}^{\infty} i|b_i| < \infty$.

From the Assumption 9 we can find that when l is big enough $\sum_{s=l+1}^{\infty} b_s z_{i,t-s}$ is very small. So we can only consider the case $y_{it} = \sum_{s=0}^{l} b_s z_{i,t-s}$. Let $\mathbf{Y}$ be a $(p - K - L) \times n$ matrix whose entries are y_{it}. We can write $\frac{1}{n}\mathbf{Y}\mathbf{Y}^T$ as the form of $\Gamma \mathbf{X} \mathbf{\Omega} \mathbf{X}^{\mathbf{T}} \Gamma^T$ and check the Assumptions in Theorem 10.

Assumption 10. *$(\mathbf{x}_{1t}, \cdots, \mathbf{x}_{pt})^T = \mathbf{x}_t = \frac{1}{\sqrt{n}}(f_{1t}, \cdots, f_{Kt}, \zeta_{1t}, \cdots, \zeta_{M-K,t})^T, t = -\infty, ..., n$ are i.i.d. random vectors. $\{\mathbf{x}_{ij}$: $i = 1, ..., p,\ j = -\infty, ..., n\}$ are independent random variables with finite fourth moment such that $\mathbb{E}\mathbf{x}_{ij} = 0$, $\mathbb{E}|\sqrt{n}\mathbf{x}_{ij}|^2 = 1$ and $\mathbb{E}|\sqrt{n}\mathbf{x}_{ij}|^4 = \gamma_{4i}$. There exists δ_n satisfying*

$$\lim_{n\to\infty} \delta_n^{-4} \frac{1}{np} \sum_{i=1}^{p} \sum_{j=1}^{n} \mathbb{E}|\sqrt{n}\mathbf{x}_{ij}|^4 I(|\sqrt{n}\mathbf{x}_{ij}| > \delta_n \sqrt[4]{np}) = 0, \quad \delta_n \downarrow 0, \quad \delta_n \sqrt[4]{np} \uparrow \infty \tag{3.17}$$

Write $\mathbf{X} = (\mathbf{x}_{1-l}, \cdots, \mathbf{x}_n)$.

The Assumption 10 ensures that Assumptions 1-2 hold.

Definition 1. *$\mathbf{W}$ is a $(n+l) \times n$ whose entries are $W_{ij} = b_{l-i+j} I\{0 \le i - j \le l\}$. Let $\mathbf{\Omega} = \mathbf{W}\mathbf{W}^T$.*

The Assumption 9 ensures that the Assumption 3 holds.

Assumption 11. *The* $(p-K-L)\times(p-K)$ *matrix* $\mathbf{\Upsilon}$ *consists of the entries* Υ_{is} *satisfing* $\|\mathbf{\Upsilon}\mathbf{\Upsilon}^T\|\le C$.

Assumption 12. *There exists a positive constant c such that* $\min_{1\le s\le K}\{\sum_{i=1}^{p-K-L}\phi_{is}^2\} > cp$. *Also* $\mathbf{\Phi}$ *is a* $(p-K-L)\times K$ *matrix whose entries are* ϕ_{is}.

Write $\Gamma=(\mathbf{\Phi},\mathbf{\Upsilon})$. Assumptions 11-12 ensure that the Assumption 7 holds. In fact $\Gamma\Gamma^T=\mathbf{\Phi}\mathbf{\Phi}^{\mathbf{T}}+\mathbf{\Upsilon}\mathbf{\Upsilon}^{\mathbf{T}}$. Assumption 12 implies that $rank(\mathbf{\Phi}\mathbf{\Phi}^{\mathbf{T}})=K$ and the K largest eigenvalue of $\mathbf{\Phi}\mathbf{\Phi}^{\mathbf{T}}$ is bigger than cp and Assumption 11 implies that $\|\mathbf{\Upsilon}\mathbf{\Upsilon}^T\|\le C$. Then we can find that the K largest eigenvalue of $\Gamma\Gamma^T$ is bigger that $cp-C$ and the $K+1$ largest eigenvalue of $\Gamma\Gamma^T$ is not bigger than C. Then we conclude that the Assumption 7 holds.

We then write

$$\frac{1}{n}\mathbf{Y}\mathbf{Y}^T=\Gamma\mathbf{X}\mathbf{\Omega}\mathbf{X}^{\mathbf{T}}\Gamma^T. \tag{3.18}$$

Moreover, Assumptions 1-3 and 7 hold. Then we can find that Theorem 10 works for the largest eigenvalues of $\frac{1}{n}\mathbf{Y}\mathbf{Y}^T$.

3.3.2 Nonstationary linear processes

Consider a nonstationary linear process as follows.

$$y_{it}=y_{i,t-1}+\sum_{s=0}^{l}b_s z_{i,t-s} \tag{3.19}$$

and

$$z_{it}=\sum_{s=1}^{n}\Upsilon_{is}\mathbf{x}_{st}, \tag{3.20}$$

where $1\le i\le n-L$ and $1-l\le t\le p-l$.

Let $\mathbf{Y}$ be a $(n-L)\times(p-l)$ matrix whose entries are y_{it}. We next give some assumptions and definitions to rewrite $\frac{1}{n}\mathbf{Y}^{\mathbf{T}}\mathbf{Y}$ in the form of $\Gamma\mathbf{X}\mathbf{\Omega}\mathbf{X}^{\mathbf{T}}\Gamma^T$. We also check the Assumptions in Theorem 11.

Assumption 13. $\mathbf{x}_i = (\mathbf{x}_{i,1-l}, \cdots, \mathbf{x}_{i,p-l})^T$,$i = 1, ..., n$ *are i.i.d. random vectors and* $\{\mathbf{x}_{ij}$*:* $i = 1, ..., n,\ j = 1 - l, ..., p - l\}$ *are independent random variables with finite fourth moment such that* $\mathbb{E}\mathbf{x}_{ij} = 0$*.* $\mathbb{E}|\sqrt{n}\mathbf{x}_{ij}|^2 = 1$ *and* $\mathbb{E}|\sqrt{n}\mathbf{x}_{ij}|^4 = \gamma_{4i}$*. Let* $\mathbf{X} = (\mathbf{x}_1, ..., \mathbf{x}_n)$*.*

Then we can find that the Assumption 1 holds.

Assumption 14. *The* $(n - L) \times n$ *matrix* $\boldsymbol{\Upsilon}$ *consists of the entries* Υ_{is}*.*

Assumption 15. $\sum_{i=0}^{\infty} i|b_i| < \infty$*.*

Definition 2. $\mathbf{W}$ *is a* $p \times (p-l)$ *whose entries are* $W_{ij} = b_{l-i+j} I\{0 \le i - j \le l\}$*. Let* $\mathbf{C} = (C_{ij})_{1\le i,j \le p-l}$ *be a* $(p - l) \times (p - l)$ *lower triangular matrix with*

$$C_{ij} = 0 \quad \textit{for } j > i \textit{ and } \quad C_{ij} = 1 \quad \textit{for } 1 \le j \le i. \tag{3.21}$$

Let $\Gamma = \mathbf{C}\mathbf{W}^T$. We can conclude that Assumptions 5-6 hold.

Write

$$\frac{1}{n}\mathbf{Y}^{\mathbf{T}}\mathbf{Y} = \Gamma\mathbf{X}\boldsymbol{\Omega}\mathbf{X}^{\mathbf{T}}\Gamma^T. \tag{3.22}$$

Moreover, Assumptions 1, 3 and 5-6 hold. Then we can find that Theorem 11 works for the largest eigenvalues of $\frac{1}{n}\mathbf{Y}^{\mathbf{T}}\mathbf{Y}$.

Remark 12. *One can see the details of the nonstationary linear processes in Chapter 2.*

3.4 Proof of Theorem 10

We give a new assumption

Assumption 16. *For any* $k \in \mathbb{N}^+$*, there exist constants* c_k *such that* $\mathbb{E}|\sqrt{n}\mathbf{x}_{ij}|^k \le c_k$*.*

We will prove a weaker version of Theorem 10 in Section 4.1. i.e. We assume Assumption 16 holds. Then we will show how to relax Assumption 16 in Section 3.4.3.

3.4.1 Proof of Theorem 10 under Assumption 16

Define $V_1 = \Lambda_S^{1/2}\mathbf{U}_1\mathbf{X}\mathbf{\Omega}^{1/2}$ and $V_2 = \Lambda_P^{1/2}\mathbf{U}_2\mathbf{X}\mathbf{\Omega}^{1/2}$. We prove CLT for a fixed λ_i, $i \in \{1, ..., K\}$. By the definition of λ_i, it solves the equation

$$\det(\lambda_i\mathbf{I} - \Lambda^{1/2}\mathbf{U}\mathbf{X}\mathbf{\Omega}\mathbf{X}^T\mathbf{U}^T\Lambda^{1/2}) = 0,$$

together with Proposition 8, with probability tending to 1. It is equivalent to

$$\det(\lambda_i\mathbf{I} - V_1\left[\mathbf{I} + V_2^T(\lambda_i\mathbf{I} - V_2V_2^T)^{-1}V_2\right]V_1^T = 0.$$

By schur's complement formula, it is also equivalent to

$$\det(\lambda_i\mathbf{I} - \lambda_i V_1(\lambda_i\mathbf{I} - V_2V_2^T)^{-1}V_1^T) = 0. \tag{3.23}$$

Noting that $\mathbf{U}_2^T\Lambda_P\mathbf{U}_2 = \Sigma$, the above equation becomes

$$\det(\lambda_i\mathbf{I} - \lambda_i\Lambda_S^{1/2}\mathbf{U}_1\mathbf{X}\mathbf{\Omega}^{1/2}(\lambda_i\mathbf{I} - \mathbf{\Omega}^{1/2}\mathbf{X}^T\Sigma\mathbf{X}\mathbf{\Omega}^{1/2})^{-1}\mathbf{\Omega}^{1/2}\mathbf{X}^T\mathbf{U}_1^T\Lambda_S^{1/2}) = 0.$$

i.e.

$$\det(\lambda_i\Lambda_S^{-1} - \lambda_i\mathbf{U}_1\mathbf{X}\mathbf{\Omega}^{1/2}(\lambda_i\mathbf{I} - \mathbf{\Omega}^{1/2}\mathbf{X}^T\Sigma\mathbf{X}\mathbf{\Omega}^{1/2})^{-1}\mathbf{\Omega}^{1/2}\mathbf{X}^T\mathbf{U}_1^T) = 0. \tag{3.24}$$

We abbreviate $\mathbf{Y} = \mathbf{X}\mathbf{\Omega}^{1/2}$, and then (3.24) becomes

$$\det(\alpha_i\Lambda_S^{-1} - \alpha_i\mathbf{U}_1\mathbf{Y}(\lambda_i\mathbf{I} - \mathbf{Y}^T\Sigma\mathbf{Y})^{-1}\mathbf{Y}^T\mathbf{U}_1^T) = 0. \tag{3.25}$$

Since the LHS of the equation above is a continue function of θ_i, we conclude that $\theta_i = \alpha_i\frac{tr\mathbf{\Omega}}{n}(1 + o(1))$. We denote $\lambda_i - \theta_i$ by ν_i. Then

$$\begin{aligned}\mathbf{U}_1\mathbf{Y}(\lambda_i\mathbf{I} - \mathbf{Y}^T\Sigma\mathbf{Y})^{-1}\mathbf{Y}^T\mathbf{U}_1^T &= \mathbf{U}_1\mathbf{Y}(\theta_i\mathbf{I} - \mathbf{Y}^T\Sigma\mathbf{Y})^{-1}\mathbf{Y}^T\mathbf{U}_1^T \\ &-\nu_i\mathbf{U}_1\mathbf{Y}(\lambda_i\mathbf{I} - \mathbf{Y}^T\Sigma\mathbf{Y})^{-1}(\theta_i\mathbf{I} - \mathbf{Y}^T\Sigma\mathbf{Y})^{-1}\mathbf{Y}^T\mathbf{U}_1^T.\end{aligned} \tag{3.26}$$

Similar to the proof of Theorem 3.1 in Bai and Yao (2008), the key step is to establish the central limit theorem for the entries of $\sqrt{n}\alpha_i\mathbf{U}_1\mathbf{X}(\theta_i\mathbf{I} - \mathbf{X}^T\Sigma\mathbf{X})^{-1}\mathbf{X}^T\mathbf{U}_1^T$ by the Leibniz formula for the determinant of a matrix. Let $\mathbf{u}_j^T$ be the jth row of

$\mathbf{U}_1$. In the sequel, we only prove the central limit theorem for $\sqrt{n}\alpha_i \mathbf{u}_j^T \mathbf{Y}(\theta_i \mathbf{I} - \mathbf{Y}^T \Sigma \mathbf{Y})^{-1}\mathbf{Y}^T \mathbf{u}_l$. Moreover, it is easy to see that

$$\|\theta_i^2 \mathbf{U}_1 \mathbf{X}(\lambda_i \mathbf{I} - \mathbf{X}^T \Sigma \mathbf{X})^{-1}(\theta_i \mathbf{I} - \mathbf{X}^T \Sigma \mathbf{X})^{-1}\mathbf{X}^T \mathbf{U}_1^T - \mathbf{I}\|_\infty \tag{3.27}$$
$$= o_p(1),$$

which implies that the randomness of $\theta_i^2 \mathbf{U}_1 \mathbf{Y}(\lambda_i \mathbf{I} - \mathbf{Y}^T \Sigma \mathbf{Y})^{-1}(\theta_i \mathbf{I} - \mathbf{Y}^T \Sigma \mathbf{Y})^{-1}\mathbf{Y}^T \mathbf{U}_1^T$ does not affect CLT. Before proceeding, we define another event

$$\mathfrak{C}_d^{(k)} = \{\|(\mathbf{Y}^{(k)})^T \Sigma \mathbf{Y}^{(k)}\| \le 4CM_0(1 + \frac{p}{n})\}, \tag{3.28}$$

where $\mathbf{Y}^{(k)} = \mathbf{Y} - \mathbf{x}_k \mathbf{e}_k^T \mathbf{\Omega}^{1/2}$, $\mathbf{x}_k$ is the k-th column of $\mathbf{X}$ and $\mathbf{e}_k = (0, .., 0, 1, 0, ..., 0)^T$ is p-dimensional vector with only k-th element being 1. Denote $\mathbf{y}_k = \mathbf{x}_k \mathbf{e}_k^T \mathbf{\Omega}^{1/2}$. We remind the readers that when we calculate the negligible terms, there are always several indicator functions such as $I(\mathfrak{C}_d)$ and $I(\mathfrak{C}_d^{(k)})$ involved in them. $\mathfrak{C}_d$ and $\mathfrak{C}_d^{(k)}$ hold with high probability by the results in Bai and Silverstein (2006) and Chen and Pan (2012), so they do not affect the CLT. One should notice that we assume that $\mathbb{E}|\mathbf{x}_{ij}|^2 = \frac{1}{n}$. Denote $\mathbb{E}_k = \mathbb{E}(.|\mathbf{x}_1, ..., \mathbf{x}_k)$ and $\vartheta_i = \frac{\alpha_i}{\theta_i}$. Then we have

$$\begin{aligned}
&\alpha_i \mathbf{u}_j^T \mathbf{Y}(\theta_i \mathbf{I} - \mathbf{Y}^T \Sigma \mathbf{Y})^{-1}\mathbf{Y}^T \mathbf{u}_l I(\mathfrak{C}_d) - \alpha_i \mathbb{E}\mathbf{u}_j^T \mathbf{Y}(\theta_i \mathbf{I} - \mathbf{Y}^T \Sigma \mathbf{Y})^{-1}\mathbf{Y}^T \mathbf{u}_l I(\mathfrak{C}_d)\\
&= \vartheta_i \sum_{k=1}^n (\mathbb{E}_k - \mathbb{E}_{k-1})\mathbf{u}_j^T \mathbf{Y}(\mathbf{I} - \mathbf{Y}^T \frac{\Sigma}{\theta_i}\mathbf{Y})^{-1}\mathbf{Y}^T \mathbf{u}_l I(\mathfrak{C}_d) \qquad (3.29)\\
&= \vartheta_i \sum_{k=1}^n (\mathbb{E}_k - \mathbb{E}_{k-1})\\
&\big(\mathbf{u}_j^T \mathbf{Y}(\mathbf{I} - \mathbf{Y}^T \frac{\Sigma}{\theta_i}\mathbf{Y})^{-1}\mathbf{Y}^T \mathbf{u}_l - \mathbf{u}_j^T \mathbf{Y}^{(k)}(\mathbf{I} - (\mathbf{Y}^{(k)})^T \frac{\Sigma}{\theta_i}\mathbf{Y}^{(k)})^{-1}(\mathbf{Y}^{(k)})^T \mathbf{u}_l\big) I(\mathfrak{C}_d).
\end{aligned}$$

Here we aim at showing the central limit theorem for $\alpha_i \mathbf{u}_j^T \mathbf{Y}(\theta_i \mathbf{I} - \mathbf{Y}^T \Sigma \mathbf{Y})^{-1}\mathbf{Y}^T \mathbf{u}_l I(\mathfrak{C}_d)$ instead of $\alpha_i \mathbf{u}_j^T \mathbf{Y}(\theta_i \mathbf{I} - \mathbf{Y}^T \Sigma \mathbf{Y})^{-1}\mathbf{Y}^T \mathbf{u}_l$. Actually, since $I(\mathfrak{C}_d) = 1$ with high probability, $\alpha_i \mathbf{u}_j^T \mathbf{Y}(\theta_i \mathbf{I} - \mathbf{Y}^T \Sigma \mathbf{Y})^{-1}\mathbf{Y}^T \mathbf{u}_l I(\mathfrak{C}_d)$ and $\alpha_i \mathbf{u}_j^T \mathbf{Y}(\theta_i \mathbf{I} - \mathbf{Y}^T \Sigma \mathbf{Y})^{-1}\mathbf{Y}^T \mathbf{u}_l$ share the same central limit theorem. We define $\frac{\Sigma}{\theta_i} = \tilde{\Sigma}$, $\mathbf{A} = \mathbf{I} - \mathbf{Y}^T \tilde{\Sigma}\mathbf{Y}$ and $\mathbf{A}_k = \mathbf{I} - (\mathbf{Y}^{(k)})^T \tilde{\Sigma}\mathbf{Y}^{(k)}$. Then $\mathbf{A} = \mathbf{A}_k - \mathbf{y}_k^T \tilde{\Sigma}\mathbf{Y}^{(k)} - (\mathbf{Y}^{(k)})^T \tilde{\Sigma}\mathbf{y}_k - \mathbf{y}_k^T \tilde{\Sigma}\mathbf{y}_k$.

Moreover, in case there are only $\mathbf{A}_k^{-1}$ involved in the equations we are calculating, we will replace $I(\mathfrak{C}_d)$ by $I(\mathfrak{C}_d^{(k)})$. This replacement aims at calculating the expectations about $\mathbf{x}_k$. For example, we replace $I(\mathfrak{C}_d)$ by $I(\mathfrak{C}_d^{(k)})$ at the first line of (3.41), which makes the expectation much easier.

We define $\mathbf{A}_{k,2} = \mathbf{y}_k^T\tilde{\Sigma}\mathbf{Y}^{(k)}$, $\mathbf{A}_{k,3} = (\mathbf{Y}^{(k)})^T\tilde{\Sigma}\mathbf{y}_k$ and $\mathbf{A}_{k,4} = \mathbf{y}_k^T\tilde{\Sigma}\mathbf{y}_k$. We also define $\zeta_{kk} = \mathbf{e}_k^T\mathbf{\Omega}^{1/2}\mathbf{A}_k^{-1}\mathbf{\Omega}^{1/2}\mathbf{e}_k$.

Noting that $\mathbf{u}_j^T\mathbf{Y} = \mathbf{u}_j^T\mathbf{Y}^{(k)} + \mathbf{u}_j^T\mathbf{y}_k$, we define

$$\mathbf{A}_{k,w} = \mathbf{A}_k - \mathbf{A}_{k,3}. \tag{3.30}$$

By the formula $\mathbf{A}^{-1} = \mathbf{B}^{-1} - \frac{\mathbf{B}^{-1}(\mathbf{A}-\mathbf{B})\mathbf{B}^{-1}}{1+tr(\mathbf{B}^{-1}(\mathbf{A}-\mathbf{B}))}$, we have

$$\begin{aligned}\mathbf{A}^{-1} \quad &= \mathbf{A}_{k,w}^{-1} + \frac{1}{a_k}\mathbf{A}_{k,w}^{-1}(\mathbf{A}_{k,2}+\mathbf{A}_{k,4})\mathbf{A}_{k,w}^{-1} \\ &= \mathbf{A}_k^{-1} + \frac{\mathbf{A}_k^{-1}\mathbf{A}_{k,3}\mathbf{A}_k^{-1}}{b_k} + \frac{\mathbf{A}_{k,w}^{-1}(\mathbf{A}_{k,2}+\mathbf{A}_{k,4})\mathbf{A}_{k,w}^{-1}}{a_k},\end{aligned} \tag{3.31}$$

where

$$a_k = 1 - \mathbf{x}_k^T\tilde{\Sigma}\mathbf{Y}^{(k)}\mathbf{A}_{k,w}^{-1}\mathbf{\Omega}^{1/2}\mathbf{e}_k - \mathbf{x}_k^T\tilde{\Sigma}\mathbf{x}_k\mathbf{e}_k^T\mathbf{\Omega}^{1/2}\mathbf{A}_{k,w}^{-1}\mathbf{\Omega}^{1/2}\mathbf{e}_k \tag{3.32}$$

and

$$b_k = 1 - \mathbf{e}_k^T\mathbf{\Omega}^{1/2}\mathbf{A}_k^{-1}(\mathbf{Y}^{(k)})^T\tilde{\Sigma}\mathbf{x}_k. \tag{3.33}$$

It follows that

$$(\mathbb{E}_k - \mathbb{E}_{k-1})\big(\mathbf{u}_j^T\mathbf{Y}\mathbf{A}^{-1}\mathbf{Y}^T\mathbf{u}_l\big) = (\mathbb{E}_k - \mathbb{E}_{k-1})(I_1+I_2+I_3+I_4), \tag{3.34}$$

where

$$\begin{aligned}I_1 \quad &= (\mathbf{u}_j^T\mathbf{x}_k)(\mathbf{u}_l^T\mathbf{x}_k)\mathbf{e}_k^T\mathbf{\Omega}^{1/2}\mathbf{A}^{-1}\mathbf{\Omega}^{1/2}\mathbf{e}_k \\ &= \frac{(\mathbf{u}_j^T\mathbf{x}_k)(\mathbf{u}_l^T\mathbf{x}_k)\mathbf{e}_k^T\mathbf{\Omega}^{1/2}\mathbf{A}_{k,w}^{-1}\mathbf{\Omega}^{1/2}\mathbf{e}_k}{a_k} \\ &= \frac{(\mathbf{u}_j^T\mathbf{x}_k)(\mathbf{u}_l^T\mathbf{x}_k)\mathbf{e}_k^T\mathbf{\Omega}^{1/2}\mathbf{A}_k^{-1}\mathbf{\Omega}^{1/2}\mathbf{e}_k}{a_k(1-\mathbf{e}_k^T\mathbf{\Omega}^{1/2}\mathbf{A}_k(\mathbf{Y}^{(k)})^T\tilde{\Sigma}\mathbf{x}_k)} \\ &= \frac{(\mathbf{u}_j^T\mathbf{x}_k)(\mathbf{u}_l^T\mathbf{x}_k)\zeta_{kk}}{a_kb_k},\end{aligned} \tag{3.35}$$

$$I_2 = \sum_{f\neq k} \mathbf{u}_l^T \mathbf{x}_k \mathbf{u}_j^T \mathbf{x}_f \mathbf{e}_f^T \mathbf{\Omega}^{1/2} \mathbf{A}^{-1} \mathbf{\Omega}^{1/2} \mathbf{e}_k \tag{3.36}$$
$$= \frac{\mathbf{u}_j^T \mathbf{Y}^{(k)} \mathbf{A}_k^{-1} \mathbf{y}_k^T \mathbf{u}_l}{a_k} + \frac{\mathbf{u}_j^T \mathbf{Y}^{(k)} \mathbf{A}_k^{-1} (\mathbf{Y}^{(k)})^T \tilde{\Sigma} \mathbf{y}_k \mathbf{A}_k^{-1} \mathbf{y}_k^T \mathbf{u}_l}{a_k b_k}$$
$$= \frac{\mathbf{u}_j^T \mathbf{Y}^{(k)} \mathbf{A}_k^{-1} \mathbf{y}_k^T \mathbf{u}_l}{a_k} + \frac{\mathbf{u}_j^T \mathbf{Y}^{(k)} \mathbf{A}_k^{-1} (\mathbf{Y}^{(k)})^T \tilde{\Sigma} \mathbf{x}_k \mathbf{x}_k^T \mathbf{u}_l \zeta_{kk}}{a_k b_k},$$

$$I_3 = \sum_{f\neq k} \mathbf{u}_j^T \mathbf{x}_k \mathbf{u}_l^T \mathbf{x}_f \mathbf{e}_f^T \mathbf{\Omega}^{1/2} \mathbf{A}^{-1} \mathbf{\Omega}^{1/2} \mathbf{e}_k \tag{3.37}$$
$$= \frac{\mathbf{u}_j^T \mathbf{y}_k \mathbf{A}_k^{-1} (\mathbf{Y}^{(k)})^T \mathbf{u}_l}{a_k} + \frac{\mathbf{u}_j^T \mathbf{y}_k \mathbf{A}_k^{-1} \mathbf{y}_k^T \tilde{\Sigma} \mathbf{Y}^{(k)} \mathbf{A}_k^{-1} (\mathbf{Y}^{(k)})^T \mathbf{u}_l}{a_k b_k}$$
$$= \frac{\mathbf{u}_j^T \mathbf{y}_k \mathbf{A}_k^{-1} (\mathbf{Y}^{(k)})^T \mathbf{u}_l}{a_k} + \frac{\mathbf{u}_j^T \mathbf{x}_k \mathbf{x}_k^T \tilde{\Sigma} \mathbf{Y}^{(k)} \mathbf{A}_k^{-1} (\mathbf{Y}^{(k)})^T \mathbf{u}_l \zeta_{kk}}{a_k b_k}$$

and

$$I_4 = \sum_{f,g\neq k} \mathbf{u}_j^T \mathbf{y}_f \mathbf{A}^{-1} \mathbf{y}_g^T \mathbf{u}_l \tag{3.38}$$
$$= \sum_{f,g\neq k} \mathbf{u}_j^T \mathbf{y}_f \mathbf{A}_k^{-1} \mathbf{y}_g^T \mathbf{u}_l + \sum_{f,g\neq k} \mathbf{u}_j^T \mathbf{y}_f \frac{\mathbf{A}_k^{-1} \mathbf{A}_{k,3} \mathbf{A}_k^{-1}}{b_k} \mathbf{y}_g^T \mathbf{u}_l$$
$$+ \sum_{f,g\neq k} \frac{1}{a_k} \mathbf{u}_j^T \mathbf{y}_f \mathbf{A}_{k,w}^{-1} (\mathbf{A}_{k,2} + \mathbf{A}_{k,4}) \mathbf{A}_{k,w}^{-1} \mathbf{y}_g^T \mathbf{u}_l$$
$$= \mathbf{u}_j^T \mathbf{Y}^{(k)} \mathbf{A}_k^{-1} (\mathbf{Y}^{(k)})^T \mathbf{u}_l + \frac{\mathbf{u}_j^T \mathbf{Y}^{(k)} \mathbf{A}_k^{-1} \mathbf{A}_{k,3} \mathbf{A}_k^{-1} (\mathbf{Y}^{(k)})^T \mathbf{u}_l}{b_k}$$
$$+ \frac{1}{a_k} \mathbf{u}_j^T \mathbf{Y}^{(k)} \mathbf{A}_{k,w}^{-1} (\mathbf{A}_{k,2} + \mathbf{A}_{k,4}) \mathbf{A}_{k,w}^{-1} (\mathbf{Y}^{(k)})^T \mathbf{u}_l.$$

Noting that $\mathbf{u}_j^T \mathbf{Y} = \sum_f \mathbf{u}_j^T \mathbf{y}_f$. One can verify that $I_2 + I_3$ is the $f \neq k$ term of $\mathbf{u}_j^T \mathbf{Y} \mathbf{A}^{-1} \mathbf{Y}^T \mathbf{u}_l$. Similarly, I_4 is the $f \neq g$ term of $\mathbf{u}_j^T \mathbf{Y} \mathbf{A}^{-1} \mathbf{Y}^T \mathbf{u}_l$.

By (3.35), we have $a_k b_k I(\mathfrak{E}_d) = \frac{\mathbf{e}_k^T \mathbf{\Omega}^{1/2} \mathbf{A}_k^{-1} \mathbf{\Omega}^{1/2} \mathbf{e}_k}{\mathbf{e}_k^T \mathbf{\Omega}^{1/2} \mathbf{A}^{-1} \mathbf{\Omega}^{1/2} \mathbf{e}_k} I(\mathfrak{E}_d) = 1 + o(1)$. Therefore, we conclude an important formula in this thesis, i.e.

$$a_k b_k I(\mathfrak{E}_d) = 1 + o(1), \tag{3.39}$$

with high probability. In fact, noticing that $\mathbf{x}_k^T\tilde{\Sigma}\mathbf{x}_k$ is the diagonal entry of $\mathbf{Y}^T\tilde{\Sigma}\mathbf{Y}$, we have with high probability

$$\begin{aligned} |1-b_k|I(\mathfrak{E}_d) &= |\mathbf{e}_k^T\mathbf{\Omega}^{1/2}\mathbf{A}_k^{-1}(\mathbf{Y}^{(k)})^T\tilde{\Sigma}\mathbf{x}_k|I(\mathfrak{E}_d) \\ &\leq \|\mathbf{\Omega}^{1/2}\mathbf{A}_k^{-1}(\mathbf{Y}^{(k)})^T\tilde{\Sigma}^{1/2}\|\|\tilde{\Sigma}^{1/2}\mathbf{x}_k\|I(\mathfrak{E}_d) \\ &= O(d_i). \end{aligned} \tag{3.40}$$

(3.39)-(3.40) imply that there is a constant $C_1 > 2$ such that $\max\{\frac{1}{a_k}, \frac{1}{b_k}\}I(\mathfrak{E}_d) \leq C_1 I(\mathfrak{E}_d)$.

Considering $(\mathbb{E}_k-\mathbb{E}_{k-1})(I_4-\mathbf{u}_j^T\mathbf{Y}^{(k)}\mathbf{A}_k^{-1}(\mathbf{Y}^{(k)})^T\mathbf{u}_l)$ first. The term $\mathbf{u}_j^T\mathbf{Y}^{(k)}\mathbf{A}_k^{-1}(\mathbf{Y}^{(k)})^T\mathbf{u}_l$ involved in I_4 disappears in $I_4-\mathbf{u}_j^T\mathbf{Y}^{(k)}\mathbf{A}_k^{-1}(\mathbf{Y}^{(k)})^T\mathbf{u}_l$. We claim the other two terms of I_4 at the RHS of (3.38) are negligible. In fact, consider the term

$$\sqrt{n}\sum_{k=1}^{n}(\mathbb{E}_k-\mathbb{E}_{k-1})\frac{\mathbf{u}_j^T\mathbf{Y}^{(k)}\mathbf{A}_k^{-1}\mathbf{A}_{k,3}\mathbf{A}_k^{-1}(\mathbf{Y}^{(k)})^T\mathbf{u}_l}{b_k}I(\mathfrak{E}_d),$$

we calculate the upper bound of the variance

$$\begin{aligned} &C_1^2 n\sum_{k=1}^{n}\mathbb{E}|\mathbf{u}_j^T\mathbf{Y}^{(k)}\mathbf{A}_k^{-1}\mathbf{A}_{k,3}\mathbf{A}_k^{-1}(\mathbf{Y}^{(k)})^T\mathbf{u}_l|^2 I(\mathfrak{E}_d) \\ &= C_1^2\sum_{k=1}^{n}\mathbb{E}\mathbf{u}_j^T\mathbf{Y}^{(k)}\mathbf{A}_k^{-1}(\mathbf{Y}^{(k)})^T\tilde{\Sigma}^2\mathbf{Y}^{(k)}\mathbf{A}_k^{-1}(\mathbf{Y}^{(k)})^T\mathbf{u}_l \\ &\mathbf{e}_k^T\mathbf{\Omega}^{1/2}\mathbf{A}_k^{-1}(\mathbf{Y}^{(k)})^T\mathbf{u}_l\mathbf{u}_j^T\mathbf{Y}^{(k)}\mathbf{A}_k^{-1}\mathbf{\Omega}^{1/2}\mathbf{e}_k I(\mathfrak{E}_d^{(k)}) + o(n^{-2}). \end{aligned} \tag{3.41}$$

By the definition of $I(\mathfrak{E}_d^{(k)})$, we have

$$\begin{aligned} &\mathbf{u}_j^T\mathbf{Y}^{(k)}\mathbf{A}_k^{-1}(\mathbf{Y}^{(k)})^T\tilde{\Sigma}^2\mathbf{Y}^{(k)}\mathbf{A}_k^{-1}(\mathbf{Y}^{(k)})^T\mathbf{u}_l I(\mathfrak{E}_d^{(k)}) \\ &\leq C_2\frac{d_i^2}{\max\{1,\frac{p}{n}\}}\mathbf{u}_j^T\mathbf{Y}^{(k)}(\mathbf{Y}^{(k)})^T\mathbf{u}_l I(\mathfrak{E}_d^{(k)}) = O(d_i^2) \to 0. \end{aligned} \tag{3.42}$$

Therefore, we only need to bound

$$\sum_{k=1}^{n}\mathbb{E}\mathbf{e}_k^T\mathbf{\Omega}^{1/2}\mathbf{A}_k^{-1}(\mathbf{Y}^{(k)})^T\mathbf{u}_l\mathbf{u}_j^T\mathbf{Y}^{(k)}\mathbf{A}_k^{-1}\mathbf{\Omega}^{1/2}\mathbf{e}_k I(\mathfrak{E}_d^{(k)}) \leq \sum_{k=1}^{n}\mathbb{E}|\mathbf{e}_k^T\mathbf{\Omega}^{1/2}\mathbf{A}_k^{-1}(\mathbf{Y}^{(k)})^T\mathbf{u}_1|^2 I(\mathfrak{E}_d^{(k)}). \tag{3.43}$$

Intuitively, if we can replace $\mathbf{A}_k$ and $\mathbf{Y}^{(k)}$ by $\mathbf{A}$ and $\mathbf{Y}$ respectively, then the expectation above is bounded by

$$\sum_{k=1}^{n} \mathbb{E}|\mathbf{e}_k^T \mathbf{\Omega}^{1/2}\mathbf{A}^{-1}\mathbf{Y}^T\mathbf{u}_1|^2 = \mathbb{E}\mathbf{u}_1^T\mathbf{Y}\mathbf{A}^{-1}\mathbf{\Omega}\mathbf{A}^{-1}\mathbf{Y}^T\mathbf{u}_1 \leq M_0\mathbb{E}\mathbf{u}_1^T\mathbf{Y}\mathbf{Y}^T\mathbf{u}_1 = O(1). \tag{3.44}$$

Therefore, we below replace $\mathbf{A}_k$ and $\mathbf{Y}^{(k)}$ by $\mathbf{A}$ and $\mathbf{Y}$ respectively. First, we replace $\mathbf{Y}^{(k)}$ by $\mathbf{Y}$. It follows that

$$\begin{aligned}\sum_{k=1}^{n} \mathbb{E}|\mathbf{e}_k^T \mathbf{\Omega}^{1/2}\mathbf{A}_k^{-1}(\mathbf{Y}^{(k)})^T\mathbf{u}_1|^2 I(\mathfrak{E}_d^{(k)}) \quad & \leq 2\sum_{k=1}^{n} \mathbb{E}|\mathbf{e}_k^T \mathbf{\Omega}^{1/2}\mathbf{A}_k^{-1}\mathbf{Y}^T\mathbf{u}_1|^2 I(\mathfrak{E}_d^{(k)}) \\ & +2\sum_{k=1}^{n} \mathbb{E}|\mathbf{e}_k^T \mathbf{\Omega}^{1/2}\mathbf{A}_k^{-1}\mathbf{\Omega}^{1/2}\mathbf{e}_k\mathbf{x}_k^T\mathbf{u}_1|^2 I(\mathfrak{E}_d^{(k)}).\end{aligned}$$

It is easy to see that

$$\sum_{k=1}^{n} \mathbb{E}|\mathbf{e}_k^T \mathbf{\Omega}^{1/2}\mathbf{A}_k^{-1}\mathbf{\Omega}^{1/2}\mathbf{e}_k\mathbf{x}_k^T\mathbf{u}_1|^2 I(\mathfrak{E}_d^{(k)}) \leq M_0 \sum_{k=1}^{n} |\mathbf{x}_k^T\mathbf{u}_1|^2 = M_0.$$

Therefore, it suffices to consider $\sum_{k=1}^{n} \mathbb{E}|\mathbf{e}_k^T\mathbf{A}_k^{-1}\mathbf{Y}^T\mathbf{u}_1|^2$. One should notice that in the sequel we will omit the indicator function $I(\mathfrak{E}_d)$ and $I(\mathfrak{E}_d^{(k)})$ involved in the equalities(inequalities). Recalling the definition below (3.29)

$$\mathbf{A} = \mathbf{A}_k - \mathbf{\Omega}^{1/2}\mathbf{e}_k\mathbf{x}_k^T\tilde{\Sigma}\mathbf{Y}^{(k)} - (\mathbf{Y}^{(k)})^T\tilde{\Sigma}\mathbf{x}_k\mathbf{e}_k^T\mathbf{\Omega}^{1/2} - \mathbf{\Omega}^{1/2}\mathbf{e}_k\mathbf{x}_k^T\tilde{\Sigma}\mathbf{x}_k\mathbf{e}_k^T\mathbf{\Omega}^{1/2}, \tag{3.45}$$

together with (3.31), we conclude that

$$\begin{aligned}\mathbf{e}_k^T\mathbf{\Omega}^{1/2}\mathbf{A}^{-1} \quad & = \frac{\mathbf{e}_k^T\mathbf{\Omega}^{1/2}(\mathbf{A}_k - \mathbf{\Omega}^{1/2}\mathbf{e}_k\mathbf{x}_k^T\tilde{\Sigma}\mathbf{Y}^{(k)})^{-1}}{a_k} \\ & = \frac{\mathbf{e}_k^T\mathbf{\Omega}^{1/2}\mathbf{A}_k^{-1}}{a_k} + \frac{\zeta_{kk}\mathbf{x}_k^T\tilde{\Sigma}\mathbf{Y}^{(k)}\mathbf{A}_k^{-1}}{a_k b_k}.\end{aligned}$$

Summarizing the above, we have

$$\begin{aligned}
&\sum_{k=1}^{n}\mathbb{E}|\mathbf{e}_k^T\mathbf{\Omega}^{1/2}\mathbf{A}_k^{-1}\mathbf{Y}^T\mathbf{u}_1|^2 I(\mathfrak{E}_d^{(k)}) \\
&\leq \sum_{k=1}^{n}\mathbb{E}|\mathbf{e}_k^T\mathbf{\Omega}^{1/2}\mathbf{A}_k^{-1}\mathbf{Y}^T\mathbf{u}_1|^2 I(\mathfrak{E}_d) + o(n^{-2}) \\
&\leq 2(\sum_{k=1}^{n}\mathbb{E}|\mathbf{e}_k^T\mathbf{\Omega}^{1/2}\mathbf{A}^{-1}\mathbf{Y}^T\mathbf{u}_1|^2 I(\mathfrak{E}_d) \\
&+\sum_{k=1}^{n}\mathbb{E}|\zeta_{kk}\mathbf{x}_k^T\tilde{\Sigma}\mathbf{Y}^{(k)}\mathbf{A}_k^{-1}\mathbf{Y}^T\mathbf{u}_1|^2) I(\mathfrak{E}_d^{(k)}) + o(n^{-2}).
\end{aligned} \tag{3.46}$$

The first term at the RHS of (3.46) is bounded by a constant referring to (3.44). To control the second term, recalling that $\mathbf{Y} = \mathbf{Y}^{(k)} + \mathbf{x}_k\mathbf{e}_k^T\mathbf{\Omega}^{1/2}$, similar to the previous conclusion, we have

$$\begin{aligned}
&\sum_{k=1}^{n}\mathbb{E}|\zeta_{kk}\mathbf{x}_k^T\tilde{\Sigma}\mathbf{Y}^{(k)}\mathbf{A}_k^{-1}\mathbf{Y}^T\mathbf{u}_1|^2 I(\mathfrak{E}_d^{(k)}) \\
&\leq \sum_{k=1}^{n}\mathbb{E}|\zeta_{kk}\mathbf{x}_k^T\tilde{\Sigma}\mathbf{Y}^{(k)}\mathbf{A}_k^{-1}(\mathbf{Y}^{(k)})^T\mathbf{u}_1|^2 I(\mathfrak{E}_d^{(k)}) \\
&+\sum_{k=1}^{n}\mathbb{E}|\zeta_{kk}\mathbf{x}_k^T\tilde{\Sigma}\mathbf{Y}^{(k)}\mathbf{A}_k^{-1}\mathbf{e}_k\mathbf{x}_k^T\mathbf{u}_1|^2 I(\mathfrak{E}_d^{(k)}) \\
&\to 0.
\end{aligned} \tag{3.47}$$

We turn to the more complex term

$$\frac{1}{a_k}\mathbf{u}_j^T\mathbf{Y}^{(k)}\mathbf{A}_{k,w}^{-1}(\mathbf{A}_{k,2}+\mathbf{A}_{k,4})\mathbf{A}_{k,w}^{-1}(\mathbf{Y}^{(k)})^T\mathbf{u}_l = \frac{z_k}{a_k}.$$

As before, we only need to control the case $j = l = 1$. Since a_k is close to 1 with high probability, it suffices to bound the variance of the following term

$$\sqrt{n}\sum_{k=1}^{n}(\mathbb{E}_k - \mathbb{E}_{k-1})z_k,$$

i.e.

$$n\sum_{k=1}^{n}\mathbb{E}|z_k|^2.$$

For convenience, we only consider one part of the above term, which is

$$n\sum_{k=1}^{n}\mathbb{E}|\mathbf{u}_1^T\mathbf{Y}^{(k)}\mathbf{A}_{k,w}^{-1}\mathbf{A}_{k,2}\mathbf{A}_{k,w}^{-1}(\mathbf{Y}^{(k)})^T\mathbf{u}_1|^2. \tag{3.48}$$

Since

$$\mathbf{A}_{k,w}^{-1}=\mathbf{A}_k^{-1}+\frac{\mathbf{A}_k^{-1}\mathbf{A}_{k,3}\mathbf{A}_k^{-1}}{b_k},$$

there are four terms involved in (3.48). We consider the first one

$$\begin{aligned}
&n\sum_{k=1}^{n}\mathbb{E}|\mathbf{u}_1^T\mathbf{Y}^{(k)}\mathbf{A}_k^{-1}\mathbf{\Omega}^{1/2}\mathbf{e}_k\mathbf{x}_k^T\tilde{\Sigma}\mathbf{Y}^{(k)}\mathbf{A}_k^{-1}(\mathbf{Y}^{(k)})^T\mathbf{u}_1|^2 \qquad (3.49)\\
&=\sum_{k=1}^{n}\mathbb{E}|\mathbf{u}_1^T\mathbf{Y}^{(k)}\mathbf{A}_k^{-1}\mathbf{\Omega}^{1/2}\mathbf{e}_k|^2|\mathbf{u}_1^T\mathbf{Y}^{(k)}\mathbf{A}_k^{-1}(\mathbf{Y}^{(k)})^T\tilde{\Sigma}^2\mathbf{Y}^{(k)}\mathbf{A}_k^{-1}(\mathbf{Y}^{(k)})^T\mathbf{u}_1|^2.
\end{aligned}$$

Similar to (3.41), we conclude that (3.49) tends 0. For the cross term, it is

$$\begin{aligned}
&n\sum_{k=1}^{n}\mathbb{E}|\mathbf{u}_1^T\mathbf{Y}^{(k)}\frac{\mathbf{A}_k^{-1}\mathbf{A}_{k,3}\mathbf{A}_k^{-1}}{b_k}\mathbf{\Omega}^{1/2}\mathbf{e}_k\mathbf{x}_k^T\tilde{\Sigma}\mathbf{Y}^{(k)}\mathbf{A}_k^{-1}(\mathbf{Y}^{(k)})^T\mathbf{u}_1|^2\\
&\le\frac{C_1^2}{n}\sum_{k=1}^{n}\mathbb{E}|\mathbf{u}_1^T\mathbf{Y}^{(k)}\mathbf{A}_k^{-1}(\mathbf{Y}^{(k)})^T\tilde{\Sigma}\mathbf{e}_k^T\mathbf{\Omega}^{1/2}\mathbf{A}_k^{-1}\mathbf{\Omega}^{1/2}\mathbf{e}_k\tilde{\Sigma}\mathbf{Y}^{(k)}\mathbf{A}_k^{-1}(\mathbf{Y}^{(k)})^T\mathbf{u}_1|^2\\
&\le\frac{C_1^2}{n}d_i^2\sum_{k=1}^{n}\mathbb{E}|\mathbf{u}_1^T\mathbf{Y}^{(k)}(\mathbf{Y}^{(k)})^T\mathbf{u}_1|^2=o(1). \qquad (3.50)
\end{aligned}$$

For the fourth term, we have

$$\begin{aligned}
&n\sum_{k=1}^{n}\mathbb{E}|\mathbf{u}_1^T\mathbf{Y}^{(k)}\frac{\mathbf{A}_k^{-1}\mathbf{A}_{k,3}\mathbf{A}_k^{-1}}{b_k}\mathbf{\Omega}^{1/2}\mathbf{e}_k\mathbf{x}_k^T\tilde{\Sigma}\mathbf{Y}^{(k)}\frac{\mathbf{A}_k^{-1}\mathbf{A}_{k,3}\mathbf{A}_k^{-1}}{b_k}(\mathbf{Y}^{(k)})^T\mathbf{u}_1|^2\\
&\le C_1^4n\sum_{k=1}^{n}\mathbb{E}|\mathbf{u}_1^T\mathbf{Y}^{(k)}\mathbf{A}_k^{-1}\mathbf{A}_{k,3}\mathbf{A}_k^{-1}\mathbf{\Omega}^{1/2}\mathbf{e}_k\mathbf{x}_k^T\tilde{\Sigma}\mathbf{Y}^{(k)}\mathbf{A}_k^{-1}\mathbf{A}_{k,3}\mathbf{A}_k^{-1}(\mathbf{Y}^{(k)})^T\mathbf{u}_1|^2\\
&=o(1), \qquad (3.51)
\end{aligned}$$

where in the last inequality we apply the same argument as (3.40) to show

$$|\mathbf{x}_k^T\tilde{\Sigma}\mathbf{Y}^{(k)}\mathbf{A}_k^{-1}(\mathbf{Y}^{(k)})^T\tilde{\Sigma}\mathbf{x}_k|=O(d_i^2)$$

with high probability. Next we consider I_1, I_2 and I_3, which follows from (3.35)-

(3.37) that

$$\sqrt{n}\sum_{k=1}^{n}(\mathbb{E}_k-\mathbb{E}_{k-1})\Big(\frac{(\mathbf{u}_j\mathbf{x}_k)(\mathbf{u}_l\mathbf{x}_k)\zeta_{kk}}{a_k} \\ +\frac{\mathbf{u}_j^T\mathbf{Y}^{(k)}\mathbf{A}_k^{-1}\mathbf{\Omega}^{1/2}\mathbf{e}_k\mathbf{x}_k^T\mathbf{u}_l}{a_k}+\frac{\mathbf{u}_j^T\mathbf{Y}^{(k)}\mathbf{A}_k^{-1}(\mathbf{Y}^{(k)})^T\tilde{\Sigma}\mathbf{x}_k\mathbf{x}_k^T\mathbf{u}_l\zeta_{kk}}{a_kb_k} \\ +\frac{\mathbf{u}_j^T\mathbf{x}_k\mathbf{e}_k^T\mathbf{\Omega}^{1/2}\mathbf{A}_k^{-1}(\mathbf{Y}^{(k)})^T\mathbf{u}_l}{a_k}+\frac{\mathbf{u}_j^T\mathbf{x}_k\mathbf{x}_k^T\tilde{\Sigma}\mathbf{Y}^{(k)}\mathbf{A}_k^{-1}(\mathbf{Y}^{(k)})^T\mathbf{u}_l\zeta_{kk}}{a_kb_k}\Big) \tag{3.52}$$

We claim that the third term of (3.52) is negligible. In fact, since $|a_k|$, $|b_k|$ and $\zeta_{kk}\sim 1$ with high probability, $\frac{1}{|a_k|},\frac{1}{|b_k|},|\zeta_{kk}|\le C_1$ with high probabiity. Therefore, we only need to calculate the following term

$$n\sum_{k=1}^{n}\mathbb{E}|\mathbf{u}_j^T\mathbf{Y}^{(k)}\mathbf{A}_k^{-1}(\mathbf{Y}^{(k)})^T\tilde{\Sigma}\mathbf{x}_k\mathbf{x}_k^T\mathbf{u}_l|^2. \tag{3.53}$$

Denote $\mathbf{c}_k=\tilde{\Sigma}\mathbf{Y}^{(k)}\mathbf{A}_k^{-1}(\mathbf{Y}^{(k)})^T\mathbf{u}_j$ and $\mathbf{h}=\mathbf{u}_l$. Then we have

$$\begin{aligned}(3.53) &\le \frac{\max_f\{\gamma_{4f}\}}{n}\sum_{k=1}^{n}\mathbb{E}\Big(\sum_{f=1}^{p}\mathbf{c}_{kf}^2\mathbf{h}_f^2+\sum_{f\neq g}^{p}\mathbf{c}_{kf}^2\mathbf{h}_g^2+\sum_{f\neq g}^{p}\mathbf{c}_{kf}\mathbf{h}_f\mathbf{c}_{kg}\mathbf{h}_g\Big)\\ &\le \frac{\max_f\{\gamma_{4f}\}}{n}\sum_{k=1}^{n}\mathbb{E}\Big(\sum_{f=1}^{p}\mathbf{c}_{kf}^2\mathbf{h}_f^2+\sum_{f,g=1}^{p}\mathbf{c}_{kf}^2\mathbf{h}_g^2+\sum_{f,g=1}^{p}\mathbf{c}_{kf}\mathbf{h}_f\mathbf{c}_{kg}\mathbf{h}_g\Big)\\ &\le \frac{\max_f\{\gamma_{4f}\}}{n}\sum_{k=1}^{n}\mathbb{E}\|\mathbf{c}_k\|^2. \end{aligned} \tag{3.54}$$

By the definition of $\mathbf{c}_k$, it is easy to see that

$$\begin{aligned}&\mathbb{E}\|\mathbf{c}_k\|^2=\mathbb{E}\mathbf{u}_j^T\mathbf{Y}^{(k)}\mathbf{A}_k^{-1}(\mathbf{Y}^{(k)})^T\tilde{\Sigma}^2\mathbf{Y}^{(k)}\mathbf{A}_k^{-1}(\mathbf{Y}^{(k)})^T\mathbf{u}_j\\ &\le d_i^2\mathbf{u}_j^T\mathbf{u}_j=d_i^2\to 0.\end{aligned}$$

Similarly, the fifth term of (3.52) is negligible. Therefore, the main term of (3.29) is

$$\sqrt{n}\sum_{k=1}^{n}(\mathbb{E}_k-\mathbb{E}_{k-1})\Big(\frac{(\mathbf{u}_j^T\mathbf{x}_k)(\mathbf{u}_l^T\mathbf{x}_k)\mathbf{e}_k^T\mathbf{\Omega}^{1/2}\mathbf{A}_k^{-1}\mathbf{\Omega}^{1/2}\mathbf{e}_k}{a_k}+ \\ \frac{\mathbf{u}_j^T\mathbf{Y}^{(k)}\mathbf{A}_k^{-1}\mathbf{\Omega}^{1/2}\mathbf{e}_k\mathbf{x}_k^T\mathbf{u}_l}{a_k}+\frac{\mathbf{u}_j^T\mathbf{x}_k\mathbf{e}_k^T\mathbf{\Omega}^{1/2}\mathbf{A}_k^{-1}(\mathbf{Y}^{(k)})^T\mathbf{u}_l}{a_k}\Big). \tag{3.55}$$

We below replace a_k by 1 in the second term. Note that $1-a_k = \mathbf{x}_k^T\tilde{\Sigma}\mathbf{Y}^{(k)}\mathbf{A}_{k,w}^{-1}\mathbf{\Omega}^{1/2}\mathbf{e}_k+$ $\mathbf{x}_k^T\tilde{\Sigma}\mathbf{x}_k\mathbf{e}_k^T\mathbf{\Omega}^{1/2}\mathbf{A}_{k,w}^{-1}\mathbf{\Omega}^{1/2}\mathbf{e}_k$. Write

$$\begin{aligned} &\frac{\mathbf{u}_j^T\mathbf{Y}^{(k)}\mathbf{A}_k^{-1}\mathbf{\Omega}^{1/2}\mathbf{e}_k\mathbf{x}_k^T\mathbf{u}_l}{a_k} \\ &= \mathbf{u}_j^T\mathbf{Y}^{(k)}\mathbf{A}_k^{-1}\mathbf{\Omega}^{1/2}\mathbf{e}_k\mathbf{x}_k^T\mathbf{u}_l + \frac{1-a_k}{a_k}\mathbf{u}_j^T\mathbf{Y}^{(k)}\mathbf{A}_k^{-1}\mathbf{\Omega}^{1/2}\mathbf{e}_k\mathbf{x}_k^T\mathbf{u}_l. \end{aligned} \tag{3.56}$$

We aim to prove that

$$\sqrt{n}\sum_{k=1}^n(\mathbb{E}_k-\mathbb{E}_{k-1})\frac{1-a_k}{a_k}\mathbf{u}_j^T\mathbf{Y}^{(k)}\mathbf{A}_k^{-1}\mathbf{\Omega}^{1/2}\mathbf{e}_k\mathbf{x}_k^T\mathbf{u}_l = o_p(1).$$

Since $\frac{1-a_k}{a_k} = o(1)$ with high probability, it suffices to prove that

$$n\sum_{k=1}^n\mathbb{E}|\mathbf{u}_j^T\mathbf{Y}^{(k)}\mathbf{A}_k^{-1}\mathbf{\Omega}^{1/2}\mathbf{e}_k\mathbf{x}_k^T\mathbf{u}_l|^2 = O(1). \tag{3.57}$$

By a simple calculation, we have

$$n\sum_{k=1}^n\mathbb{E}|\mathbf{u}_j^T\mathbf{Y}^{(k)}\mathbf{A}_k^{-1}\mathbf{\Omega}^{1/2}\mathbf{e}_k\mathbf{x}_k^T\mathbf{u}_l|^2 = \sum_{k=1}^n\mathbb{E}\mathbf{u}_j^T\mathbf{Y}^{(k)}\mathbf{A}_k^{-1}\mathbf{\Omega}^{1/2}\mathbf{e}_k\mathbf{e}_k^T\mathbf{\Omega}^{1/2}\mathbf{A}_k^{-1}(\mathbf{Y}^{(k)})^T\mathbf{u}_j,$$

which is exactly (3.43). Therefore, (3.57) has been proved. In view of (3.56), we only need to consider the term

$$\begin{aligned} &\sqrt{n}\sum_{k=1}^n(\mathbb{E}_k-\mathbb{E}_{k-1})\mathbf{u}_j^T\mathbf{Y}^{(k)}\mathbf{A}_k^{-1}\mathbf{\Omega}^{1/2}\mathbf{e}_k\mathbf{x}_k^T\mathbf{u}_l \\ &= \sqrt{n}\sum_{k=1}^n\mathbb{E}_k\mathbf{u}_j^T\mathbf{Y}^{(k)}\mathbf{A}_k^{-1}\mathbf{\Omega}^{1/2}\mathbf{e}_k\mathbf{x}_k^T\mathbf{u}_l. \end{aligned} \tag{3.58}$$

Similar to the proof of replacing a_k by 1, by $\|\mathbf{A}_k^{-1}-\mathbf{I}\| = o_p(1)$, we can show that

$$(3.58) = \sqrt{n}\sum_{k=1}^n\mathbb{E}_k\mathbf{u}_j^T\mathbf{Y}^{(k)}\mathbf{\Omega}^{1/2}\mathbf{e}_k\mathbf{x}_k^T\mathbf{u}_l + o_p(1).$$

We next consider the first term at the RHS of (3.52). Similar to (3.56), we can also replace a_k by 1. Therefore, by (3.55)-(3.58), we get

$$\begin{aligned} (3.55) \quad &= \sqrt{n}\sum_{k=1}^n\Big(((\mathbf{u}_j^T\mathbf{x}_k)(\mathbf{u}_l^T\mathbf{x}_k) - \frac{I(j=l)}{n})\mathbf{e}_k^T\mathbf{\Omega}\mathbf{e}_k + \mathbf{u}_j^T\mathbf{x}_k\mathbf{u}_l^T\mathbb{E}_k\mathbf{X}^{(k)}\mathbf{\Omega}\mathbf{e}_k \\ &+\mathbf{u}_l^T\mathbf{x}_k\mathbf{u}_j^T\mathbb{E}_k\mathbf{X}^{(k)}\mathbf{\Omega}\mathbf{e}_k\Big) \\ &= \sqrt{n}(\mathbf{u}_j^T\mathbf{X}\mathbf{\Omega}\mathbf{X}^T\mathbf{u}_l - \frac{I(j=l)tr\mathbf{\Omega}}{n}). \end{aligned} \tag{3.59}$$

Then the limit of the covariance is

$$\lim_{n\to\infty} n^2 \times Cov(\mathbf{u}_{m_i+k_1}^T \mathbf{X}\mathbf{\Omega}\mathbf{X}^T \mathbf{u}_{m_i+l_1}, \mathbf{u}_{m_i+k_2}^T \mathbf{X}\mathbf{\Omega}\mathbf{X}^T \mathbf{u}_{m_i+l_2}). \quad (3.60)$$

By Theorem 7.2 of Bai and Yao (2008) it is not hard to calculate them.

3.4.2 Calculation of Mean

In this section the aim is to calculate the expectation of $\mathbf{u}_j^T \mathbf{Y}\mathbf{A}^{-1}\mathbf{Y}^T \mathbf{u}_l I(\mathfrak{E}_d)$. In fact, we only need to consider the case $j = l$ since when $j \neq l$ the expectation is $o(\frac{\nu_i}{\theta_i^2})$. At first, we define new events $I(\tilde{\mathbf{E}}_d) = I(tr(\mathbf{Y}^T\Sigma\mathbf{Y}) \leq 4nCM_0(1+\frac{p}{n}))$ and $I(\tilde{\mathbf{E}}_d^{(k)}) = I(tr((\mathbf{Y}^{(k)})^T\Sigma\mathbf{Y}^{(k)}) \leq 4nCM_0(1+\frac{p}{n}))$. It is easy to see that

$$I(\tilde{\mathbf{E}}_d) = 1,\ I(\tilde{\mathbf{E}}_d^{(k)}) = 1$$

with high probability by the results in Bai and Silverstein (2006) and Chen and Pan (2012). Denote $\underline{\mathbf{A}} = \tilde{\Sigma}^{1/2}\mathbf{Y}\mathbf{Y}^T\tilde{\Sigma}^{1/2} - \mathbf{I}$, $\mathbf{r}_f = \tilde{\Sigma}^{1/2}\mathbf{x}_f$ and $\underline{\mathbf{A}}_g = \sum_{f\neq g}\mathbf{r}_f\mathbf{r}_f^T - \mathbf{I}$. Moreover, let $\mathbf{A}(\eta) = \mathbf{A} + \eta \mathrm{i}\mathbf{I}$, and $\underline{\mathbf{A}}(\eta) = \underline{\mathbf{A}} - \eta\mathrm{i}\mathbf{I}$, $\mathrm{i}^2 = -1$. Let $\eta = n^{-3/4}$. It is not hard to conclude that

$$\mathbb{E}\sqrt{n}\mathbf{u}_j^T\mathbf{Y}\mathbf{A}^{-1}\mathbf{Y}^T\mathbf{u}_j I(\mathfrak{E}_d) = \mathbb{E}\sqrt{n}\mathbf{u}_j^T\mathbf{Y}\mathbf{A}(\eta)^{-1}\mathbf{Y}^T\mathbf{u}_j I(\tilde{\mathbf{E}}_d) + o(1).$$

In order to calculate the derivatives, we introduce a smooth cutoff function of $I(\tilde{\mathbf{E}}_d)$:

$$\mathcal{X}(x) = \begin{cases} 1 & \text{if } |x| \leq 4nCM_0(1+\frac{p}{n}) \\ 0 & \text{if } |x| \geq 4nCM_0(1+\frac{p}{n}) + 1, \end{cases}.$$

whose derivatives satisfy $|\mathcal{X}^{(k)}| \leq 2$, $k \in \mathbb{N}^+$. Therefore we have

$$\mathbb{E}\sqrt{n}\mathbf{u}_j^T\mathbf{Y}\mathbf{A}(\eta)^{-1}\mathbf{Y}^T\mathbf{u}_j I(\tilde{\mathbf{E}}_d) = \mathbb{E}\sqrt{n}\mathbf{u}_j^T\mathbf{Y}\mathbf{A}(\eta)^{-1}\mathbf{Y}^T\mathbf{u}_j\mathcal{X}(tr\mathbf{Y}^T\Sigma\mathbf{Y}) + o(1).$$

Up to now, we are ready to prove the universality result, which means that

$$\mathbb{E}\sqrt{n}\mathbf{u}_j^T\mathbf{Y}\mathbf{A}(\eta)^{-1}\mathbf{Y}^T\mathbf{u}_j\mathcal{X}(tr\mathbf{Y}^T\Sigma\mathbf{Y})$$

does not depend on the distribution but the first two moments. In the sequel, we define $\tilde{\mathbf{A}} = \mathbf{A}(\eta)$ and $F(\mathbf{Y}) = \mathbf{u}_j^T \mathbf{Y} \tilde{\mathbf{A}}^{-1} \mathbf{Y}^T \mathbf{u}_j$. Similar to the previous section, we omit the term $\mathcal{X}(tr\mathbf{Y}^T \Sigma \mathbf{Y})$ and the relative arguments since $\frac{\partial^k \mathcal{X}(tr\mathbf{Y}^T \Sigma \mathbf{Y})}{\partial \mathbf{x}_{f\mu}^k} = 0$ with high probability. Moreover, we also need to define the following interpolation matrix:

Definition 3. *For $t \in \{0, 1\}$, $1 \le f \le p$ and $1 \le \mu \le n$, denote the density(or probability mass function) of $X_{f\mu}^t$ by $\rho_{f\mu}^t$. For $t \in [0, 1]$, we define the distribution function by*

$$\rho_{f\mu}^t = t\rho_{f\mu}^1 + (1-t)\rho_{f\mu}^0.$$

The interpolation matrix $\mathbf{Y}^t$ is $(X_{f\mu}^t)$ with $F_{f\mu}^t$ being the distribution of $X_{f\mu}^t$ and the entries $\{X_{f\mu}^t\}$ are mutually independent for all f, μ. Moreover, we introduce the matrix

$$\mathbf{Y}_{(f\mu)}^{t,\lambda} = \mathbf{Y}^t + (\lambda - X_{f\mu}^\theta)\mathbf{e}_f \mathbf{e}_\mu^T, \tag{3.61}$$

which differs from $\mathbf{Y}^t$ at the (f, μ) position only.

By Lemma 7.9 of Knowles and Yin (2014), we have

$$\sqrt{n}(\mathbb{E}F(\mathbf{Y}^1) - \mathbb{E}F(\mathbf{Y}^0)) = \sqrt{n} \int dt \sum_{f=1}^{p} \sum_{\mu=1}^{n} \left[\mathbb{E}F(\mathbf{Y}_{(f\mu)}^{t,x_{f\mu}^1}) - \mathbb{E}F(\mathbf{Y}_{(f\mu)}^{t,x_{f\mu}^0})\right],$$

Since we assume that the first two moments of $x_{i\mu}$ are 0 and $\frac{1}{N}$ respectively, it suffices to prove

$$n^{-1} \sum_{f=1}^{p} \sum_{\mu=1}^{n} \mathbb{E}F^{(3)}(\mathbf{Y}_{(f\mu)}^{t,0}) \to 0.$$

We need to calculate the derivative carefully. Since $F(\mathbf{Y}) = \mathbf{u}_j^T \mathbf{Y} \tilde{\mathbf{A}}^{-1} \mathbf{Y}^T \mathbf{u}_j$, we split the derivative into 3 parts:

$$\mathbb{E}F^{(3)}(\mathbf{Y}_{(f\mu)}^{t,0}) = B_1 + B_2 + B_3.$$

B_1: We differentiate $\mathbf{u}_j^T \mathbf{Y}$ twice. In this case, the derivative is equal to

$$(\mathbf{u}_j^T \mathbf{\Omega}^{1/2} \mathbf{e}_f)^2 \mathbf{e}_\mu^T \frac{\partial \tilde{\mathbf{A}}^{-1}}{\partial x_{f\mu}} \mathbf{e}_\mu.$$

It is easy to see that $\mathbf{e}_\mu^T \frac{\partial \tilde{\mathbf{A}}^{-1}}{\partial x_{f\mu}} \mathbf{e}_\mu = o(1)$. Together with $\sum_{f=1}^p (\mathbf{u}_j^T \mathbf{\Omega}^{1/2} \mathbf{e}_f)^2 = \mathbf{u}_j^T \mathbf{\Omega} \mathbf{u}_j$, we have proved in this case, $n^{-1} \sum_{f=1}^p \sum_{\mu=1}^n B_1 = o(1)$.

B_2: We differentiate $\mathbf{u}_j^T \mathbf{Y}$ once. Then the derivative equals

$$\begin{aligned} 2\mathbf{u}_j^T \mathbf{\Omega}^{1/2} \mathbf{e}_f \mathbf{e}_\mu^T \frac{\partial \tilde{\mathbf{A}}^{-2}}{\partial x_{f\mu}^2} \mathbf{Y}^T \mathbf{u}_j \qquad &= 2\mathbf{u}_j^T \mathbf{\Omega}^{1/2} \mathbf{e}_f \mathbf{e}_\mu^T \tilde{\mathbf{A}}^{-1} \frac{\partial \tilde{\mathbf{A}}}{\partial x_{f\mu}} \tilde{\mathbf{A}}^{-1} \frac{\partial \tilde{\mathbf{A}}}{\partial x_{f\mu}} \tilde{\mathbf{A}}^{-1} \mathbf{Y}^T \mathbf{u}_j \\ &+ 2\mathbf{u}_j^T \mathbf{\Omega}^{1/2} \mathbf{e}_f \mathbf{e}_\mu^T \tilde{\mathbf{A}}^{-1} \mathbf{e}_\mu \breve{\Sigma}_{ff} \mathbf{e}_\mu^T \tilde{\mathbf{A}}^{-1} \mathbf{Y}^T \mathbf{u}_j, \end{aligned}$$

where $\breve{\Sigma} = \mathbf{\Omega}^{1/2} \tilde{\Sigma} \mathbf{\Omega}^{1/2}$. By cauchy's inequality, we have

$$\begin{aligned} & n^{-1} \sum_{f=1}^p \sum_{\mu=1}^n \mathbb{E} \mathbf{u}_j^T \mathbf{\Omega}^{1/2} \mathbf{e}_f \mathbf{e}_\mu^T \tilde{\mathbf{A}}^{-1} \mathbf{e}_\mu \breve{\Sigma}_{ff} \mathbf{e}_\mu^T \tilde{\mathbf{A}}^{-1} \mathbf{Y}^T \mathbf{u}_j \\ & \leq n^{-1} \sqrt{np} \sqrt{\sum_{f=1}^p \mathbf{u}_{jf}^2 \breve{\Sigma}_{ff}^2} \sqrt{\mathbb{E} \mathbf{u}_j^T \mathbf{Y} \tilde{\mathbf{A}}^{-2} \mathbf{Y}^T \mathbf{u}_j} \\ & = O(d_i) = o(1). \end{aligned}$$

We consider one part involved in the other term, i.e. $\mathbf{u}_{jf} \mathbf{e}_\mu^T \tilde{\mathbf{A}}^{-1} \mathbf{e}_\mu \mathbf{e}_f^T \breve{\Sigma} \mathbf{Y} \tilde{\mathbf{A}}^{-1} \mathbf{Y}^T \breve{\Sigma} \mathbf{e}_f \mathbf{e}_\mu^T \tilde{\mathbf{A}}^{-1} \mathbf{Y}^T \mathbf{u}_j$. Similar to the previous inequality, we have

$$\begin{aligned} & n^{-1} \sum_{f=1}^p \sum_{\mu=1}^n \mathbb{E} |\mathbf{u}_j^T \mathbf{\Omega}^{1/2} \mathbf{e}_f \mathbf{e}_\mu^T \tilde{\mathbf{A}}^{-1} \mathbf{e}_\mu \mathbf{e}_f^T \breve{\Sigma} \mathbf{Y} \tilde{\mathbf{A}}^{-1} \mathbf{Y}^T \breve{\Sigma} \mathbf{e}_f \mathbf{e}_\mu^T \tilde{\mathbf{A}}^{-1} \mathbf{Y}^T \mathbf{u}_j| \\ & \leq O(d_i^2) \sqrt{\sum_f (\mathbf{u}_j^T \mathbf{\Omega}^{1/2} \mathbf{e}_f)^2} \sqrt{\mathbb{E} \mathbf{u}_j^T \mathbf{Y} \tilde{\mathbf{A}}^{-2} \mathbf{Y}^T \mathbf{u}_j} = o(1). \end{aligned}$$

Therefore, $n^{-1} \sum_{f=1}^p \sum_{\mu=1}^n B_2 = o(1)$. B_3: We can handle B_3 similar to B_1 and B_2, therefore, we have shown the universality result. By the arguments of this section, $\mathbb{E} \sqrt{n} \mathbf{u}_j^T \mathbf{Y} \mathbf{A}^{-1} \mathbf{Y}^T \mathbf{u}_j I(\mathfrak{E}_d)$ is not affected by the distribution of $\mathbf{X}$ asymptotically. Therefore, we can find an asymptotic explicit expression of

$\mathbb{E}\mathbf{u}_j^T\mathbf{Y}\mathbf{A}^{-1}\mathbf{Y}^T\mathbf{u}_jI(\mathfrak{E}_d)$ by calculating the special case $\mathbf{X}=\mathbf{X}^0$. i.e.

$$\begin{aligned}
& \mathbb{E}\mathbf{u}_j^T\mathbf{Y}\mathbf{A}^{-1}\mathbf{Y}^T\mathbf{u}_jI(\mathfrak{E}_d) \qquad (3.62)\\
= \; & \mathbb{E}\mathbf{u}_j^T\mathbf{X}^0\Omega^{1/2}\mathbf{A}^{-1}\mathbf{\Omega}^{1/2}(\mathbf{X}^0)^T\mathbf{u}_jI(\mathfrak{E}_d)+o(\frac{1}{\sqrt{n}})\\
= \; & \frac{1}{n}\mathbb{E}tr(\mathbf{A}^{-1}\mathbf{\Omega})I(\mathfrak{E}_d)+o(\frac{1}{\sqrt{n}})\\
= \; & \frac{1}{n}\sum_{g=0}^{\infty}tr\mathbb{E}(\mathbf{\Omega}^{1/2}(\mathbf{X}^0)^T\tilde{\Sigma}\mathbf{X}^0\mathbf{\Omega}^{1/2})^g\mathbf{\Omega}I(\mathfrak{E}_d)+o(\frac{1}{\sqrt{n}})\\
= \; & \frac{tr\mathbf{\Omega}}{n}+\bar{c}c_1\cdot\frac{tr\mathbf{\Omega}^2}{n}+O(\frac{p}{n}\alpha_i^{-2})+o(\frac{1}{\sqrt{n}}),
\end{aligned}$$

where

$$\bar{c}=\frac{1}{p-K-L}\sum_{j=K+1}^{p-L}\mu_j,\; c_i=\frac{p-K-L}{n\alpha_i}.$$

3.4.3 Relax Assumption 16: Truncation and Centralization

In this section, we relax Assumption 16. We need to truncate and centralize $\mathbf{x}_{fg}$. Under Assumption 1, we first truncate $\mathbf{x}_{fg}$ to $\hat{\mathbf{x}}_{fg}=\mathbf{x}_{fg}I(\sqrt{n}|\mathbf{x}_{fg}|<\delta_n\sqrt[4]{np})$ and then centralize it to $\tilde{\mathbf{x}}_{fg}=\frac{\hat{\mathbf{x}}_{fg}-\mathbb{E}\hat{\mathbf{x}}_{fg}}{\sigma_i}$, where σ_f is the standard deviation of $\sqrt{n}\hat{\mathbf{x}}_{fg}$. It is easy to see that

$$\begin{aligned}
\mathbb{P}(\mathbf{Y}\neq\hat{\mathbf{Y}}) &\leq \sum_{f=1}^{p}\sum_{g=1}^{n}\mathbb{P}(|\sqrt{n}\mathbf{x}_{fg}|\geq\delta_n\sqrt[4]{np})\\
&\leq \frac{1}{np}\delta_n^{-4}\sum_{f=1}^{p}\sum_{g=1}^{n}\mathbb{E}|\sqrt{n}\mathbf{x}_{fg}|^4I(|\sqrt{n}\mathbf{x}_{fg}|>\delta_n\sqrt[4]{np})\to 0. \qquad (3.63)
\end{aligned}$$

Then we have

$$\mathbb{P}(\alpha_i\mathbf{U}_1\mathbf{Y}(\lambda_i\mathbf{I}-\mathbf{Y}^T\Sigma\mathbf{Y})^{-1}\mathbf{Y}^T\mathbf{U}_1^T=\alpha_i\mathbf{U}_1\hat{\mathbf{Y}}(\lambda_i\mathbf{I}-\hat{\mathbf{Y}}^T\Sigma\hat{\mathbf{Y}})^{-1}\hat{\mathbf{Y}}^T\mathbf{U}_1^T)\to 1.$$

Therefore, together with (3.24), we only need to consider

$$\det(\alpha_i\Lambda_S^{-1}-\alpha_i\mathbf{U}_1\hat{\mathbf{Y}}(\lambda_i\mathbf{I}-\hat{\mathbf{Y}}^T\Sigma\hat{\mathbf{Y}})^{-1}\hat{\mathbf{Y}}^T\mathbf{U}_1^T)=0. \qquad (3.64)$$

Considering $\tilde{\mathbf{x}}_{fg}$, we have

$$\begin{aligned} |1-\sigma_f^2| &\leq 2|\mathbb{E}(n\mathbf{x}_{fg}^2)I(|\sqrt{n}\mathbf{x}_{fg}|>\delta_n\sqrt[4]{np})| \\ &\leq 2(np)^{-1/2}\delta_n^{-2}\mathbb{E}|\sqrt{n}\mathbf{x}_{fg}|^4 I(|\sqrt{n}\mathbf{x}_{fg}|>\delta_n\sqrt[4]{np}), \end{aligned} \tag{3.65}$$

$$|\mathbb{E}\sqrt{n}\hat{\mathbf{x}}_{fg}| \leq \delta_n^{-3}(np)^{-3/4}\mathbb{E}|\sqrt{n}\mathbf{x}_{fg}|^4 I(|\sqrt{n}\mathbf{x}_{fg}|>\delta_n\sqrt[4]{np}), \tag{3.66}$$

and

$$\begin{aligned} &\mathbb{E}tr(\hat{\mathbf{Y}}-\tilde{\mathbf{Y}})(\hat{\mathbf{Y}}-\tilde{\mathbf{Y}})^T \leq \sum_{f=1}^{p}\sum_{g=1}^{n}\mathbb{E}|\hat{\mathbf{x}}_{fg}-\tilde{\mathbf{x}}_{fg}|^2 \\ &\leq 2\sum_{f=1}^{p}\sum_{g=1}^{n}(\frac{(1-\sigma_f)^2}{\sigma_f^2}Var(\hat{\mathbf{x}}_{fg})+\frac{1+(1-\sigma_f)^2}{\sigma_f^2}|\mathbb{E}\hat{\mathbf{x}}_{fg}|^2). \end{aligned} \tag{3.67}$$

From (3.65) and the Assumption 2, we can find that

$$\begin{aligned} &\sum_{f=1}^{p}\sum_{g=1}^{n}\frac{(1-\sigma_f)^2}{\sigma_f^2}Var(\hat{\mathbf{x}}_{fg}) \\ &\leq \frac{C_4}{n}\sum_{f=1}^{p}\sum_{g=1}^{n}(np)^{-1}\delta_n^{-4}(\mathbb{E}|\sqrt{n}\mathbf{x}_{fg}|^4 I(|\sqrt{n}\mathbf{x}_{fg}|>\delta_n\sqrt[4]{np}))^2 \\ &= o(\frac{1}{n}). \end{aligned} \tag{3.68}$$

From (3.66) and the Assumption 2, we can see that

$$\begin{aligned} &\sum_{f=1}^{p}\sum_{g=1}^{n}|\mathbb{E}\hat{\mathbf{x}}_{fg}|^2 \\ &\leq \frac{1}{n}\sum_{f=1}^{p}\sum_{g=1}^{n}(np)^{-3/2}\delta_n^{-6}(\mathbb{E}|\sqrt{n}\mathbf{x}_{fg}|^4 I(|\sqrt{n}\mathbf{x}_{fg}|>\delta_n\sqrt[4]{np}))^2 \\ &= o(\frac{1}{n}). \end{aligned} \tag{3.69}$$

These, together with (3.67), implies that

$$\mathbb{E}tr(\hat{\mathbf{Y}}-\tilde{\mathbf{Y}})(\hat{\mathbf{Y}}-\tilde{\mathbf{Y}})^T \leq \sum_{f=1}^{p}\sum_{g=1}^{n}\mathbb{E}|\hat{\mathbf{x}}_{fg}-\tilde{\mathbf{x}}_{fg}|^2 = o(\frac{1}{n}). \tag{3.70}$$

By (3.65)-(3.70), replacing $\hat{\mathbf{Y}}$ by $\tilde{\mathbf{Y}}$, we have

$$\|\alpha_i \mathbf{U}_1 \hat{\mathbf{Y}}(\lambda_i \mathbf{I} - \hat{\mathbf{Y}}^T \Sigma \hat{\mathbf{Y}})^{-1} \hat{\mathbf{Y}}^T \mathbf{U}_1^T - \alpha_i \mathbf{U}_1 \tilde{\mathbf{Y}}(\lambda_i \mathbf{I} - \tilde{\mathbf{Y}}^T \Sigma \tilde{\mathbf{Y}})^{-1} \tilde{\mathbf{Y}}^T \mathbf{U}_1^T\| = o_p(n^{-1/2}),$$

and

$$\|\alpha_i \mathbf{u}_j^T \hat{\mathbf{Y}}(\lambda_i \mathbf{I} - \hat{\mathbf{Y}}^T \Sigma \hat{\mathbf{Y}})^{-1} \hat{\mathbf{Y}}^T \mathbf{u}_l - \alpha_i \mathbf{u}_j^T \tilde{\mathbf{Y}}(\lambda_i \mathbf{I} - \tilde{\mathbf{Y}}^T \Sigma \tilde{\mathbf{Y}})^{-1} \tilde{\mathbf{Y}}^T \mathbf{u}_l\| = o_p(n^{-1/2}).$$

Therefore, (3.24) can be rewritten as

$$\det(\alpha_i \Lambda_1^{-1} - \alpha_i \mathbf{U}_1 \tilde{\mathbf{Y}}(\lambda_i \mathbf{I} - \tilde{\mathbf{Y}}^T \Sigma \tilde{\mathbf{Y}})^{-1} \tilde{\mathbf{Y}}^T \mathbf{U}_1^T + o_p(n^{-1/2})(\mathbf{1}\mathbf{1}^T - \mathbf{e}_i \mathbf{e}_i^T) + o_p(n^{-1/2}) \mathbf{e}_i \mathbf{e}_i^T) = 0. \tag{3.71}$$

From the proof of Theorem 10, it is easy to see that the small term $o_p(n^{-1/2})$ does not affect CLT. Therefore, we can prove Theorem 10 based on

$$\det(\alpha_i \Lambda_1^{-1} - \alpha_i \mathbf{U}_1 \tilde{\mathbf{Y}}(\lambda_i \mathbf{I} - \tilde{\mathbf{Y}}^T \Sigma \tilde{\mathbf{Y}})^{-1} \tilde{\mathbf{Y}}^T \mathbf{U}_1^T) = 0. \tag{3.72}$$

Recalling the proof of Theorem 10, all arguments hold for $\tilde{\mathbf{Y}}$ and hence Assumption 16 can be removed.

3.5 Proof of Theorem 11

We prove CLT for a fixed λ_i, $i \in \{1, ..., K\}$. Define $V_1 = \Lambda_S^{1/2} \mathbf{U}_1 \mathbf{X}$ and $V_2 = \Lambda_P^{1/2} \mathbf{U}_2 \mathbf{X}$. By the definition of λ_i, it solves the equation

$$\det(\lambda_i \mathbf{I} - \Lambda^{1/2} \mathbf{U} \mathbf{X} \mathbf{\Omega} \mathbf{X}^T \mathbf{U}^T \Lambda^{1/2}) = 0.$$

Together with Proposition 8, with probability tending to 1, it is equivalent to

$$\det(\lambda_i \mathbf{I} - V_1 \mathbf{\Omega} V_1^T - V_1 \mathbf{\Omega} V_2^T (\lambda_i \mathbf{I} - V_2 \mathbf{\Omega} V_2^T)^{-1} V_2 \mathbf{\Omega} V_1^T) = 0.$$

It's easy to see $\|(\lambda_i\mathbf{I} - V_2\mathbf{\Omega}V_2^T)^{-1}\| = O(\frac{M_0}{\mu_K})$ and $\|\mathbf{\Omega}\| \le M_0$. Then we consider the part $\Lambda_S^{-1/2}V_1V_2^TV_2V_1^T\Lambda_S^{-1/2}$. Note that

$$\begin{aligned}
&tr(\Lambda_S^{-1/2}V_1V_2^TV_2V_1^T\Lambda_S^{-1/2})\\
&= tr(\mathbf{U}_1\mathbf{X}\mathbf{X}^TU_2^T\Lambda_PU_2\mathbf{X}\mathbf{X}^TU_1^T)\\
&= \sum_{j=1}^{K}(\mathbf{U}_1\mathbf{X}\mathbf{X}^TU_2^T\Lambda_PU_2\mathbf{X}\mathbf{X}^TU_1^T)_{jj}\\
&= \sum_{j=1}^{K} h_j,
\end{aligned} \tag{3.73}$$

where $h_j = (\mathbf{U}_1\mathbf{X}\mathbf{X}^TU_2^T\Lambda_PU_2\mathbf{X}\mathbf{X}^TU_1^T)_{jj}$. Note that $\mathbf{u}_j^T\Sigma\mathbf{u}_j = 0$. Write

$$\begin{aligned}
Eh_j &= E(\mathbf{u}_j^T\mathbf{X}\mathbf{X}^TU_2^T\Lambda_PU_2\mathbf{X}\mathbf{X}^T\mathbf{u}_j)\\
&= E(\mathbf{u}_j^T\mathbf{X}\mathbf{X}^T\Sigma\mathbf{X}\mathbf{X}^T\mathbf{u}_j)\\
&= \frac{1}{n^2}\sum_{i_1=1}^{n}\sum_{i_2\neq i_1}\mathbf{u}_j^T\Sigma\mathbf{u}_j + \sum_{i_1=1}^{n}\mathbf{u}_j^TE(\mathbf{x}_{i_1}\mathbf{x}_{i_1}^T\Sigma\mathbf{x}_{i_1}\mathbf{x}_{i_1}^T)\mathbf{u}_j\\
&\le \frac{1}{n}tr(\Sigma)\max_i\{\gamma_{4i}\}.
\end{aligned} \tag{3.74}$$

Combining Assumption 5, (3.73) and (3.74) we can conclude that

$$\begin{aligned}
Etr(\Lambda_S^{-1/2}V_1V_2^TV_2V_1^T\Lambda_S^{-1/2}) &\le \sum_{j=1}^{K}\frac{1}{n}tr(\Sigma)\max_i\{\gamma_{4i}\}\\
&= o(\frac{\mu_K}{\sqrt{n}}).
\end{aligned} \tag{3.75}$$

It follows that

$$\|\Lambda_S^{-1/2}V_1\mathbf{\Omega}V_2^T(\lambda_i\mathbf{I} - V_2\mathbf{\Omega}V_2^T)^{-1}V_2\mathbf{\Omega}V_1^T\Lambda_S^{-1/2}\| = o(\frac{1}{\sqrt{n}}). \tag{3.76}$$

Define

$$H = \sqrt{n}\Lambda_S^{-1/2}(V_1\mathbf{\Omega}V_1^T - diag\{\mu_1\frac{tr\mathbf{\Omega}}{n}, \cdots, \mu_K\frac{tr\mathbf{\Omega}}{n}\})\Lambda_S^{-1/2}. \tag{3.77}$$

and the $n_i \times n_i$ matrix R which has the element

$$R_{fg} = H_{m_i+f, m_i+g}. \tag{3.78}$$

By Theorem 7.2 of Bai and Yao (2008) it is not hard to conclude that

$$R \xrightarrow{d} \mathfrak{R}_i. \tag{3.79}$$

Chapter 4

Discussions and Future Research

This research work develops CLTs of the largest eigenvalues of separable sample covariance matrices $\Gamma\mathbf{X}\mathbf{\Omega}\mathbf{X}^T\Gamma^T$ with two different kinds of $\Gamma\Gamma^T$. These theoretical results have potential applications in time series data.

As a special case of the second kind of $\Gamma\Gamma^T$ (see Section 3.3.2), we study the high–dimensional nonstationary time series and propose two new unit root tests in Chapter 2. The two tests can work when the high-dimensional time series data don't have the structure of factors. As we know, the existing tests work poor on this case.

The example in Section 3.3.1 seems to have the same potential applications in unit root tests. The trouble is how to estimate the parameters. The estimations are one of our main future research work.

Another future work may be to consider the data with more complicated deterministic components. It may extend the applications of the results.

Bibliography

Bai, Z. and Silverstein, J. "Spectral Analysis of Large Dimensional Random Matrices." *2nd Edition*, Springer:New York. (2006).

Bai, Z. and Yao, J. "Central limit theorems for eigenvalues in a spiked population model." *Ann. Inst. H. Poincaré*, 44:447–474. (2008).

Baik, J., Ben Arous, G., and Péché, S. "Phase transition of the largest eigenvalue for non–null complex sample covariance matrices." *Ann. Probab.*, 33:1643–1697. (2005).

Baik, J. and Silverstein, J. W. "Eigenvalues of large sample covariance matrices of spiked population models." *J. Multivar. Anal.*, 97:1382–1408. (2006).

Bao, Z. G., PAN, G. M., and ZHOU, W. "Universality for the largest eigenvalue of sample covariance matrices with general population." *Ann. Statist.*, 43:382–421. (2015).

Chan, N. and Wei, C. "Limiting distributions of least squares estimates of unstable autoregressive processes." *Ann. Statist.*, 16:367–401. (1988).

Chang, Y. "Bootstrap unit root tests in panels with cross sectional dependency." *J. Econom.*, 120:263–293. (2004).

Chen, B. B. and Pan, G. M. "Convergence of the largest eigenvalue of normalized sample covariance matrices when p and n both tend to infinity with their ratio converging to zero." *Bernoulli*, 18:1405–1420. (2012).

Choi, I. "Unit root tests for panel data." *J. Internat. Money & Finan.*, 20:249–272. (2001).

Dickey, D. and Fuller, W. "Distribution of the estimators for autoregressive time series with a unit root." *J. Amer. Statist. Assoc.*, 74:423–431. (1979).

El Karoui, N. "Tracy–Widom limit for the largest eigenvalue of a large class of complex sample covariance matrices." *Ann. Probab.*, 35:663–714. (2007).

Fisher, R. A. "The sampling distribution of some statistics obtained from non-linear equations." *Ann. Eugenics.*, 9:238–249. (1937).

Hsu, P. L. "On the distribution of roots of certain determinant equations." *Ann. Eugenics.*, 9:250–258. (1939).

Im, K., Pesaran, M., and SHIN, Y. "Testing for unit roots in heterogeneous panels." *J. Econom.*, 115:53–74. (2003).

Johnstone, I. "On the distribution of the largest eigenvalue in principal component analysis." *Ann. Statist.*, 29:295–327. (2001).

—. "High dimensional statistical inference and random matrices." *In International Congress of Mathematicians I*, 307–333 (2007).

Knowles, A. and Yin, J. "Anisotropic local laws for random matrices." *arXiv:1410.3516v3* (2014).

Levin, A., Lin, C., and Chu, C. "Unit root tests in panel data: asymptotic and finite-sample properties." *J. Econom.*, 108:1–24. (2002).

Liu, H. Y., A., A., and D., P. "On the Marcenko–Pastur law for linear time series." *Ann. Statist.*, 43:675–712. (2015).

Marcenko, V. A. and Pastur, L. A. "Distribution for some sets of random matrices." *Math. USSR-Sb.*, 1:457–483. (1967).

Pan, G., Gao, J., and Yang, Y. "Testing independence among a large number of high dimensional random vectors." *J. Amer. Statist. Assoc.*, 109:600–612. (2014).

Paul, D. "Asymptotics of sample eigen–structure for a large dimensional spiked covariance model." *Statist. Sinica.*, 17:1617–1642. (2007).

Paul, D. and Aue, A. "Random matrix theory in statistics: a review." *J. Statist. Plan. & Infer.*, 150:1–29. (2014).

Pesaran, M. "A simple panel unit root test in the presence of cross–sectional dependence." *J. Appl. Econom.*, 22:265–312. (2003).

—. "Time Series and Panel Data Econometrics." *Oxford: Oxford University Press.* (2015).

Pesaran, M., Smith, L. V., and Yamagata, T. "Panel unit root tests in the presence of a multifactor error structure." *J. Econom.*, 175:94–115. (2013).

Phillips, P. and Perron, P. "Testing for a unit root in time series regression." *Biometrik*, 75:335–346. (1988).

Roy, S. N. "*P*–statistics and some generalizations in analysis of variance appropriate to multivariate problems." *Sankhyā*, 4:381–396. (1939).

Tracy, C. A. and Widom, H. "Level-spacing distributions and the Airy kernel." *Comm. Math. Phys.*, 159:151–174. (1994).

—. "On orthogonal and symplectic matrix ensembles." *Comm. Math. Phys.*, 177:727–754. (1996).

Yao, J., Zheng, S. R., and Bai, Z. "Large Sample Covariance Matrices and High-Dimensional Data Analysis." *Cambridge: Cambridge University Press* (2015).

Zhang, L. "Spectral analysis of large dimensional random matrices." *Ph.D. thesis, National University of Singapore, Singapore.* (2006).