High dimensional independence test based on random matrix theory

Yang, Yanrong

2013

https://hdl.handle.net/10356/54765

https://doi.org/10.32657/10356/54765

High Dimensional Independence Test Based On Random Matrix Theory

Yanrong Yang

Division of Mathematical Sciences
School of Physical and Mathematical Sciences

2013

High Dimensional Independence Test Based On Random Matrix Theory

Yanrong Yang

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

March 2013

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, Guangming Pan, 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 Jiti Gao 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. Wang Zhou, Prof. Heng Lian, Prof. Liming Xiang and Prof. Zhen Pang 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, Binbin Chen, Zhigang Bao, Zhouping Li, Jiang Hu, Peng Lai, Qing Yang, Xiao Han and Bo Zhang.

- 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 work done, in chronological order, during my PhD studies in NTU.

1. Yanrong Yang and Guangming Pan, The convergence of the empirical distribution of canonical correlation coefficients, Electronic Journal of Probability, 17(64), 2012, 1-13.

2. Haifeng Fu, Xing Jin, Guangming Pan and Yanrong Yang, Estimating multiple option Greeks simultaneously using parameter regression, Journal of Computational Finance, forthcoming 16(2), 2013, 1-33.

3. Yanrong Yang and Guangming Pan, Independence test for high dimensional data based on regularized canonical correlation coefficients, first round of revision.(Under Review)

4. Guangming Pan, Jiti Gao, Yanrong and Meihui Guo, Independence test in a class of large panel data models, submitted.(Under Review)

5. Yanrong Yang and Guangming Pan, Determine the number of factors in factor models with an application to economic data, submitted.(Under Review)

6. Jiti Gao, Guangming Pan and Yanrong Yang, Capturing dependence between a large number of covariance stationary time series, submitted.(Under Review)

7. Jiti Gao, Guangming Pan and Yanrong Yang, Exponent of cross-sectional dependence with dynamic factor model structures, under revision.

Contents

List of Figures

List of Tables

Abstract

This thesis is concerned about statistical inference for high dimensional data based on large dimensional random matrix theory, especially, independence tests for high dimensional data.

The first problem we discussed is an independence test between two high dimensional random vectors $\mathbf{x} : p_1 \times 1$ and $\mathbf{y} : p_2 \times 1$, each of which has n random samples, i.e. $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ and $\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n\}$ respectively. A statistic is proposed based on the sum of squares of sample canonical correlation coefficients. Fortunately, the squares of the sample canonical correlation coefficients $r_1^2, r_2^2, \ldots, r_{p_1}^2$ are eigenvalues of the matrix $\mathbf{S}_{xy} = (\frac{1}{n}\mathbf{X}\mathbf{X}^T)^{-1}\frac{1}{n}\mathbf{X}\mathbf{Y}^T(\frac{1}{n}\mathbf{Y}\mathbf{Y}^T)^{-1}\frac{1}{n}\mathbf{Y}\mathbf{X}^T$, where $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$ and $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n)$. From this point, the proposed statistic is a linear spectral statistic for the matrix $\mathbf{S}_{xy}$(This matrix is called canonical correlation matrix). For this matrix, we investigate its limiting spectral distribution(LSD) and the central limit theorem(CLT) for its linear spectral statistics.

Under the case of $\mathbf{X}$ and $\mathbf{Y}$ being Gaussian distributed and independent, the LSD of $\mathbf{S}_{xy}$ has been provided in Wachter (1980). By using the Stieltjes transform method and Lindeberg's method, under the finite second moment condition, we derive that the LSD of $\mathbf{S}_{xy}$ under the general case is the same as that under the Gaussian case. The CLT for linear spectral statistics of $\mathbf{S}_{xy}$ is also provided by a similar approach. Under the Gaussian case, the empirical spectral distribution(ESD) of the matrix $\mathbf{S}_{xy}$ can be related to the ESD of an F-matrix. Under the general case, by the interpolation method and the general Stein's equation provided in Lytova and Pastur (2009), we conclude that the CLT is the same as that under the Gaussian case with the assumption $\mathbb{E}X_{11}^4 = 3$.

Apparently, the matrix $\mathbf{S}_{xy}$ only can be used under the restricted condition of $p_1, p_2 < n$ in order to make the matrices $\frac{1}{n}\mathbf{X}\mathbf{X}^T$ and $\frac{1}{n}\mathbf{Y}\mathbf{Y}^T$ invertible. To overcome this drawback, we propose regularized canonical correlation coefficients whose squares are eigenvalues of the regularized matrix $\mathbf{T}_{xy} = (\frac{1}{n}\mathbf{X}\mathbf{X}^T + t\mathbf{I}_{p_1})^{-1}\frac{1}{n}\mathbf{X}\mathbf{Y}^T(\frac{1}{n}\mathbf{Y}\mathbf{Y}^T)^{-}\frac{1}{n}\mathbf{Y}\mathbf{X}^T$, where $t > 0$, $\mathbf{I}_{p_1}$ is a $p_1 \times 1$ identity matrix and $(\frac{1}{n}\mathbf{Y}\mathbf{Y}^T)^{-}$ denotes the Moore-Penrose pseudoinverse matrix of $\frac{1}{n}\mathbf{Y}\mathbf{Y}^T$. The LSD and CLT for the matrix $\mathbf{T}_{xy}$ have been provided in a similar way to those of $\mathbf{S}_{xy}$.

Moreover, in order to derive the asymptotic theorem for the matrix $\mathbf{T}_{xy}$, we develop a CLT for linear spectral statistics of a kind of random matrix which is a sample covariance matrix plus a nonnegative definite matrix.

The second independence test is about testing independence among a large number of high dimensional random vectors, i.e. detecting independence among $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$, where $\mathbf{x}_i$ is a p-dimensional random vector and p is comparable to n. For this test, we propose a linear spectral statistic of the sample covariance matrix $\mathbf{S} = \frac{1}{n}\mathbf{X}\mathbf{X}^T$ with $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$.

When each random vector $\mathbf{x}_i, i = 1, 2, \ldots, n$ consists of independent and identical distributed(i.i.d.) components, based on the idea of the LSD of $\mathbf{S}$ being Marcenko-Pastur law under the null hypothesis, we use the characteristic function of the ESD of $\mathbf{S}$ to capture dependence between $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$. Moreover, since the characteristic function contains a parameter t, we provide the asymptotic theorem for the linear spectral statistic process with respect to the parameter t in any closed interval.

We also deal with this problem for random vectors which have relatively more complicated structures. Each random vector $\mathbf{x}_i, i = 1, 2, \ldots, n$ has the linear dependent structure $\mathbf{x}_i = \mathbf{T}\mathbf{w}_i$ or $\mathbf{x}_i$ is a covariance stationary linear process, where $\mathbf{T}$ is a nonnegative definite Hermitian matrix and $\mathbf{w}_i$ is a p-

dimensional random vector with i.i.d. components. We utilize the first two moments of the ESD of the matrix $\mathbf{S}$ to construct test statistics. The CLT for linear spectral statistics of $\mathbf{S}$ when $\mathbf{x}_i = \mathbf{T}\mathbf{w}_i$ has been provided in Bai and Silverstein (2004) while the CLT for those when each $\mathbf{x}_i$ is covariance stationary is provided in this thesis by the interpolation trick proposed in Lytova and Pastur (2009).

In summary, this thesis proposes some linear spectral statistics for independence tests for high dimensional data and develop the asymptotic theoretical results. Some simulation results are also provided to demonstrate the effectiveness of the proposed test statistics.

Chapter 1

Introduction

Recent technological innovations have brought explosion of data into many scientific disciplines, including genomics, image processing, microarray, proteomics and finance, to name but a few. In these areas the dimensionality of the data p can be much larger than or at least comparable to the sample size n. We focus on the scenario of p/n tending to a constant. This type of data poses great challenges because traditional multivariate approaches do not necessarily work, which were established for the case of the sample size n tending to infinity and the dimension p remaining fixed (See Anderson (1984)). There have been a substantial body of research work dealing with high dimensional data, e.g. Bai and Saranadasa (1996), Fan et al. (2012), Huang et al. (2008), Fan and Fan (2008), Bai and Ng (2002), Birke and Dette (2005), etc.

The importance of the independence assumption for inference arises in many aspects of multivariate analysis. For example, it is often the case in multivariate analysis that a number of variables can be rationally classified into several mutually exclusive categories. When variables can be grouped in such a way, a natural question is whether there is any significant relationship between the groups of variables. In other words, can we claim that the groups are mutually independent so that further statistics analysis such as classification and testing hypothesis of equality of mean vectors and covariance matrices could be conducted ? When the dimension p is fixed, Wilks (1935) used the likelihood ratio statistic to test independence for k sets of normal distributed random variables and one may also refer to Chapter 12 of Anderson (1984) regarding to this point.

In this thesis, we will investigate the independence test for high dimensional data, including testing independence for any two large dimensional random vectors and independence test between a large number of high dimensional random vectors. Large dimensional random matrix theory provides us with good tools to construct novel test statistics and develop their asymptotic theory for high dimensional independence test. Thus we come to the main purpose of the thesis in detail.

1.1 Canonical Correlation Analysis

The aim is to test the hypothesis

$$\mathbb{H}_0 : \mathbf{x} \textit{ and } \mathbf{y} \textit{ are independent}; \ \textit{against } \mathbb{H}_1 : \mathbf{x} \textit{ and } \mathbf{y} \textit{ are dependent}, \quad (1.1)$$

where $\mathbf{x} = (x_1, \ldots, x_{p_1})^T$ and $\mathbf{y} = (y_1, \ldots, y_{p_2})^T$. Without loss of generality, suppose that $p_1 \leq p_2$.

It is well known that canonical correlation analysis (CCA) deals with the correlation structure between two random vectors (see Chapter 12 of Anderson (1984)). Draw n independent and identically distributed (i.i.d.) observations from these two random vectors $\mathbf{x}$ and $\mathbf{y}$, respectively and group them into $p_1 \times n$ random matrix $\mathbf{X} = (\mathbf{x}_1, \cdots, \mathbf{x}_n) = (X_{ij})_{p_1 \times n}$ and $p_2 \times n$

random matrix $\mathbf{Y} = (\mathbf{y}_1, \cdots, \mathbf{y}_n) = (Y_{ij})_{p_2 \times n}$ respectively. CCA seeks the linear combinations $\mathbf{a}^T\mathbf{x}$ and $\mathbf{b}^T\mathbf{y}$ that are most highly correlated, that is to maximize

$$\gamma = Corr(\mathbf{a}^T\mathbf{x}, \mathbf{b}^T\mathbf{y}) = \frac{\mathbf{a}^T\mathbf{\Sigma}_{\mathbf{xy}}\mathbf{b}}{\sqrt{\mathbf{a}^T\mathbf{\Sigma}_{\mathbf{xx}}\mathbf{a}}\sqrt{\mathbf{b}^T\mathbf{\Sigma}_{\mathbf{yy}}\mathbf{b}}}, \tag{1.2}$$

where $\mathbf{\Sigma}_{\mathbf{xx}}$ and $\mathbf{\Sigma}_{\mathbf{yy}}$ are the population covariance matrices for $\mathbf{x}$ and $\mathbf{y}$ respectively and $\mathbf{\Sigma}_{\mathbf{xy}}$ is the population covariance matrix between $\mathbf{x}$ and $\mathbf{y}$. After finding the maximal correlation r_1 and associated vectors $\mathbf{a}_1$ and $\mathbf{b}_1$, CCA continues to seek a second linear combination $\mathbf{a}_2^T\mathbf{x}$ and $\mathbf{b}_2^T\mathbf{y}$ that has the maximal correlation among all linear combinations uncorrelated with $\mathbf{a}_1^T\mathbf{x}$ and $\mathbf{b}_1^T\mathbf{y}$. This procedure can be iterated and successive canonical correlation coefficients $\gamma_1, \ldots, \gamma_{p_1}$ can be found.

It turns out that the population canonical correlation coefficients $\gamma_1, \ldots, \gamma_{p_1}$ can be recast as the roots of the determinant equation

$$\det(\mathbf{\Sigma}_{\mathbf{xy}}\mathbf{\Sigma}_{\mathbf{yy}}^{-1}\mathbf{\Sigma}_{\mathbf{xy}}^T - \gamma^2\mathbf{\Sigma}_{\mathbf{xx}}) = 0. \tag{1.3}$$

About this point, one may refer to page 284 of Mardia et al. (1979). The roots of the determinant equation above go under many names, because they figure equally in discriminant analysis, canonical correlation analysis, and invariant tests of linear hypotheses in the multivariate analysis of variance.

Traditionally, sample covariance matrices $\hat{\mathbf{\Sigma}}_{xx}$, $\hat{\mathbf{\Sigma}}_{xy}$ and $\hat{\mathbf{\Sigma}}_{yy}$ are used to replace the corresponding population covariance matrices to solve the nonnegative roots $\rho_1, \rho_2, \ldots, \rho_{p_1}$ to the determinant equation

$$\det(\hat{\mathbf{\Sigma}}_{\mathbf{xy}}\hat{\mathbf{\Sigma}}_{\mathbf{yy}}^{-1}\hat{\mathbf{\Sigma}}_{\mathbf{xy}}^T - \rho^2\hat{\mathbf{\Sigma}}_{\mathbf{xx}}) = 0$$

where

$$\hat{\mathbf{\Sigma}}_{\mathbf{xx}} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T, \ \hat{\mathbf{\Sigma}}_{\mathbf{xy}} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{y}_i - \bar{\mathbf{y}})^T,$$

$$\hat{\mathbf{\Sigma}}_{\mathbf{yy}} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})^T,\ \bar{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i,\ \bar{\mathbf{y}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{y}_i.$$

However, it is inappropriate to use these types sample covariance matrices to replace population covariance matrices to test (1.1) in some cases. We demonstrate such an example in Section 3.4.4.

Therefore, in this part we instead consider the nonnegative roots $r_1, r_2, \ldots, r_{p_1}$ of an alternative determinant equation as follows

$$\det(\mathbf{A}_{\mathbf{xy}}\mathbf{A}_{\mathbf{yy}}^{-1}\mathbf{A}_{\mathbf{xy}}^T - r^2\mathbf{A}_{\mathbf{xx}}) = 0, \tag{1.4}$$

where

$$\mathbf{A}_{\mathbf{xx}} = \frac{1}{n}\mathbf{X}\mathbf{X}^T,\ \mathbf{A}_{\mathbf{yy}} = \frac{1}{n}\mathbf{Y}\mathbf{Y}^T,\ \mathbf{A}_{\mathbf{xy}} = \frac{1}{n}\mathbf{X}\mathbf{Y}^T.$$

We also call $\mathbf{A}_{\mathbf{xx}}$, $\mathbf{A}_{\mathbf{yy}}$ and $\mathbf{A}_{\mathbf{xy}}$ sample covariance matrices, as in the random matrix community. However, whichever sample covariance matrices are used they are not consistent estimators of population covariance matrices, which is called 'curses of dimensionality', when the dimensions p_1 and p_2 are both comparable to the sample size n. As a consequence it is conceivable that the classical likelihood ratio statistic (see Wilks (1935) and Anderson (1984)) does not work well in the high dimensional case (in fact, it is not well defined and we will discuss this point in the later section).

Moreover, from (1.4), one can see that $r_1^2, r_2^2, \ldots, r_{p_1}^2$ are the eigenvalues of the matrix

$$\mathbf{S}_{xy} = \mathbf{A}_{\mathbf{xx}}^{-1}\mathbf{A}_{\mathbf{xy}}\mathbf{A}_{\mathbf{yy}}^{-1}\mathbf{A}_{\mathbf{xy}}^T. \tag{1.5}$$

Evidently $\mathbf{A}_{\mathbf{xx}}^{-1}$ and $\mathbf{A}_{\mathbf{yy}}^{-1}$ do not exist when $p_1 > n$ and $p_2 > n$. For this reason, we also consider the eigenvalues of the regularized matrix

$$\mathbf{T}_{\mathbf{xy}} = \mathbf{A}_{t\mathbf{x}}^{-1}\mathbf{A}_{\mathbf{xy}}\mathbf{A}_{\mathbf{yy}}^{-}\mathbf{A}_{\mathbf{xy}}^T, \tag{1.6}$$

where $\mathbf{A}_{t\mathbf{x}}^{-1} = (\frac{1}{n}\mathbf{X}\mathbf{X}^T + t\mathbf{I}_{p_1})^{-1}$, t is a positive constant number and $\mathbf{I}_{p_1}$ is a $p_1 \times p_1$ identity matrix, and $\mathbf{A}_{\mathbf{yy}}^{-}$ denotes the Moore-Penrose pseudoinverse matrix of $\mathbf{A}_{\mathbf{yy}}$. Since $\mathbf{T}_{\mathbf{xy}}$ is not a symmetric matrix we consider its symmetric version

$$\mathbf{B}_n = \widetilde{\mathbf{P}}_{\mathbf{y}}\mathbf{P}_{t\mathbf{x}}\widetilde{\mathbf{P}}_{\mathbf{y}}, \tag{1.7}$$

where $\widetilde{\mathbf{P}}_{\mathbf{y}} = \frac{1}{n}\mathbf{Y}^T(\frac{1}{n}\mathbf{Y}\mathbf{Y}^T)^{-}\mathbf{Y}$ and $\mathbf{P}_{t\mathbf{x}} = \frac{1}{n}\mathbf{X}^T(\frac{1}{n}\mathbf{X}\mathbf{X}^T + t\mathbf{I}_{p_1})^{-1}\mathbf{X}$. The projection matrix $\widetilde{\mathbf{P}}_{\mathbf{y}}$ is unique when $p_2 > n$. Hence $\mathbf{T}_{\mathbf{xy}}$ is well defined even in the case of $p_1, p_2 \geq n$. Moreover $\mathbf{T}_{\mathbf{xy}}$ reduces to $\mathbf{S}_{xy}$ when p_1, p_2 are both smaller than n and $t = 0$. Here a natural question may be asked: why we use a regularized version in $\mathbf{P}_{t\mathbf{x}}$ and a generalized inverse in $\widetilde{\mathbf{P}}_{\mathbf{y}}$? This choice totally comes from technical convenience. In the proofs of asymptotic theorems, we will need treat $\widetilde{\mathbf{P}}_{\mathbf{y}}$ as a project matrix and wonder an identity matrix from multiplying $\frac{1}{n}\mathbf{X}\mathbf{X}^T$ by its inverse.

We now look at CCA from another perspective. The original random vectors $\mathbf{x}$ and $\mathbf{y}$ can be transformed into new random vectors $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ as

$$\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \rightarrow \begin{pmatrix} \boldsymbol{\xi} \\ \boldsymbol{\eta} \end{pmatrix} = \begin{pmatrix} \mathcal{A}' & \mathbf{0} \\ \mathbf{0} & \mathcal{B}' \end{pmatrix}\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \tag{1.8}$$

such that

$$\begin{pmatrix} \mathcal{A}' & \mathbf{0} \\ \mathbf{0} & \mathcal{B}' \end{pmatrix}\begin{pmatrix} \boldsymbol{\Sigma}_{\mathbf{xx}} & \boldsymbol{\Sigma}_{\mathbf{xy}} \\ \boldsymbol{\Sigma}_{\mathbf{yx}} & \boldsymbol{\Sigma}_{\mathbf{yy}} \end{pmatrix}\begin{pmatrix} \mathcal{A} & \mathbf{0} \\ \mathbf{0} & \mathcal{B} \end{pmatrix} = \begin{pmatrix} \mathbf{I}_{p_1} & \mathcal{P} \\ \mathcal{P}' & \mathbf{I}_{p_2} \end{pmatrix}, \tag{1.9}$$

where $\mathcal{P} = (\mathcal{P}_1, \mathbf{0})$, $\mathcal{P}_1 = diag(\gamma_1, \ldots, \gamma_{p_1})$ and $\mathcal{A} = \boldsymbol{\Sigma}_{\mathbf{xx}}^{-1/2}\mathbf{Q}_1$, $\mathcal{B} = \boldsymbol{\Sigma}_{\mathbf{yy}}^{-1/2}\mathbf{Q}_2$, with $\mathbf{Q}_1 : p_1 \times p_1$ and $\mathbf{Q}_2 : p_2 \times p_2$ being orthogonal matrices satisfying

$$\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1/2}\boldsymbol{\Sigma}_{\mathbf{xy}}\boldsymbol{\Sigma}_{\mathbf{yy}}^{-1/2} = \mathbf{Q}_1\mathcal{P}\mathbf{Q}_2.$$

Hence testing independence between $\mathbf{x}$ and $\mathbf{y}$ is equivalent to testing independence between $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$. The covariance between $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ has the following simple expression

$$Var\begin{pmatrix} \boldsymbol{\xi} \\ \boldsymbol{\eta} \end{pmatrix} = \begin{pmatrix} \mathbf{I}_{p_1} & \mathcal{P} \\ \mathcal{P}' & \mathbf{I}_{p_2} \end{pmatrix}. \tag{1.10}$$

In view of this, independence between $\mathbf{x}$ and $\mathbf{y}$ is equivalent to asserting that the population canonical correlations all vanish: $\gamma_1 = \cdots = \gamma_{p_1} = 0$ if the joint distribution of $\mathbf{x}$ and $\mathbf{y}$ is Gaussian. Details can be referred to Chapter 11 of Fujikoshi et al. (2010). A natural criteria for this test should be $\sum_{i=1}^{p_1} \gamma_i^2$.

As pointed out, r_i is not a consistent estimator of the corresponding population version γ_i in the high dimensional case. However, fortunately, the classical sample canonical correlation coefficients $r_1, r_2, \ldots, r_{p_1}$ or its regularized analogous still contain important information so that hypothesis testing for (1.1) is possible although the classical likelihood ratio statistic does not work well in the high dimensional case. This is due to the fact that the limits of the empirical spectral distribution (ESD) of $r_1, \cdots, r_{p_1}$ under the null and the alternative are different so that we may use it to distinguish dependence from independence (one may see the next section). Our approach essentially makes use of the integral of functions with respect to the ESD of canonical correlation coefficients. The proposed statistic turns out a trace of the corresponding matrices, i.e. $\sum_{i=1}^{p_1} r_i^2$.

In addition to proposing a statistic for testing (1.1), another contribution of this part is to establish the limit of the ESD of regularized sample canonical correlation coefficients and central limit theorems (CLT) of linear functionals of the classical and regularized sample canonical correlation coefficients $r_1, r_2, \ldots, r_{p_1}$ respectively. This is of an independent interest in its own right in addition to providing asymptotic distributions for the proposed statistics.

To derive the CLT for linear spectral statistics of classical and regularized sample canonical correlation coefficients, the strategy is to first establish the CLT under the Gaussian case, i.e. the entries of $\mathbf{X}$ are Gaussian distributed. In the Gaussian case, the CLT for linear spectral statistics of the matrix $\mathbf{S}_{xy}$ can be linked to that of an F-matrix, which has been investigated in Zheng (2012). We then extend CLT to the general distributions by bounding the difference between the characteristic functions of the respective linear spectral statistics of $\mathbf{S}_{xy}$ under the Gaussian case and nonGaussian case. To bound such a difference and handle the inverse of a random matrix we use an interpolation approach and a smooth cutoff function. The approach of developing the CLT for linear spectral statistics of the matrix $\mathbf{T}_{xy}$ is similar except we first have to develop CLT of perturbed sample covariance matrices in another chapter for establishing CLT of the matrix $\mathbf{S}_{xy}$ when the entries of $\mathbf{X}$ are Gaussian.

One point to be stressed is that, in order to derive the asymptotic theorem for the matrix $\mathbf{T}_{xy}$, we need the CLT for linear spectral statistics of a kind of random matrix which is the sum of one sample covariance matrix and a nonnegative definite matrix. This kind of matrix plays an important role in random matrix theory and we have also provided it in this thesis.

Here we would point out some works on canonical correlation coefficients under high dimensional scenario. In the high dimensional case Wachter (1980) investigated the limit of the empirical spectral distribution function of the classical sample canonical correlation coefficients $r_1, r_2, \ldots, r_{p_1}$ and Johnstone (2008) established the Tracy-Widom law of the maximum of sample correlation coefficients when $\mathbf{A_{xx}}$ and $\mathbf{A_{yy}}$ are Wishart matrices and $\mathbf{x}$, $\mathbf{y}$ are independent.

1.2 Independence test for random vectors with i.i.d components

Suppose that $\{X_{ji}, j = 1, \ldots, n; i = 1, \ldots, p\}$ are real–valued random variables. For $1 \leq i \leq p$, let $\mathbf{x}_i = (X_{1i}, \cdots, X_{ni})^T$ denote the i–th vector of random variables and $(\mathbf{x}_1, \cdots, \mathbf{x}_p)$ be a matrix of p vectors of random variables, where n usually denotes the sample size in each of the time series data. In both theory and practice, it may be unrealistic to assume that $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_p$ are independent or even uncorrelated. This is because there is no natural ordering for cross–sectional indices. There are such cases in various disciplines. In economics and finance, for example, it is not unreasonable to expect that there is significant evidence of cross–sectional dependence in output innovations across p countries and regions in the world. In the field of climatology, there is also some evidence to show that climatic variables at different stations may be cross–sectionally dependent and the level of cross–sectional dependence may be determined by some kind of physical distance. Moreover, one would expect that climatic variables, such as temperature and rainfall variables, in a station in Australia have higher–level dependence with the same type of climatic variables in a station in New Zealand than those in the United States.

In such situations, it may be necessary to test whether $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_p$ are independent before a statistical model is used to fit such data. In the econometrics and statistics literature, several papers have basically considered testing for cross–sectional uncorrelatedness for the residuals involved in some specific regression models. Such studies include Pesaran (2004) for the parametric linear model case, Hsiao, Pesaran and Pick (2007) for the parametric nonlinear case, and Chen, Gao and Li (2012) for the nonparametric nonlinear case. Other related papers include Su and Ullah (2009) for testing conditional uncorrelatedness through examining a covariance matrix in the case where p is fixed. The main purpose of this part is to propose using an empirical spectral distribution function based test statistic for independence of $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_p$.

The aim is to test

$$\mathbf{H}_0 : \mathbf{x}_1, \cdots, \mathbf{x}_p \textit{ are independent}; \text{ against } \mathbf{H}_1 : \mathbf{x}_1, \cdots, \mathbf{x}_p \textit{ are not independent}, \tag{1.11}$$

where $\mathbf{x}_i = (X_{1i}, \ldots, X_{ni})^T$ for $i = 1, \ldots, p$.

In time series analysis, mutual independence test for multiple time series has long been of interest. Moreover, time series always display various kinds of dependence. For example, an autoregressive conditional heteroscedastic (ARCH(1)) model involves a martingale difference sequence (MDS); a nonlinear moving average (MA) model is not a MDS, but its autocorrelations are zero; a linear moving average (MA) model and an autoregressive (AR) model are both models with correlated structures. In this part, we also employ the proposed statistic to test dependence for multiple time series. Section 8.5 of Anderson (1984) also considers a similar problem but with the dimensionality being fixed. His problem and approach are as follows. Let the pm-component vector $\mathbf{x}$ be distributed according to $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Partition $\mathbf{x}$ into p subvectors with m components respectively, that is, $\left(\mathbf{h}_1^T, \cdots, \mathbf{h}_p^T\right)^T$, and partition $\boldsymbol{\Sigma}$ as p^2 submatrices, that is, $\boldsymbol{\Sigma} = (\boldsymbol{\Sigma}_{ij})$ with each $\boldsymbol{\Sigma}_{ij}$ being $m \times m$.

To tackle the problem above, draw n observations from the population $\mathbf{x}$ to form a sample covariance matrix. The likelihood ratio criterion proposed is

$$\lambda = \frac{|\mathbf{Q}|^{\frac{1}{2}n}}{\prod_{i=1}^{p} |\mathbf{Q}_{ii}|^{\frac{1}{2}n}},$$

where $\mathbf{Q} = (\mathbf{Q}_{ik})$ with $\mathbf{Q}_{ik}$ being the sample covariance matrix of the random vectors $\mathbf{h}_i$ and $\mathbf{h}_k$.

Our approach essentially uses the characteristic function of the empirical spectral distribution of sample covariance matrices in large dimensional random matrix theory. Unlike the Anderson's test, we need not re–draw observations from the set of vectors of $\mathbf{x}_1, \cdots, \mathbf{x}_p$ due to the high dimensionality.

1.3 Independence test for linear dependent structures and covariance stationary processes

Testing cross-sectional dependence between a large number of high–dimensional random vectors attracts great interest in high dimensional statistical analysis, especially in longitudinal data and panel data analysis (Frees (1995); Mundlak (1978); Hsiao et al. (2009); Sarafidis et al. (2009); Chen et al. (2012)). In longitudinal data or panel data analysis, one of the key reasons of pooling the data together is to overcome the aggregation problems that arise with dependent data in modelling the behaviour of heterogenous agents on the basis of the representative assumption. In multivariate time series analysis, elucidation of various causalities between time series is vital to forecasting and prediction. Compared with the literature focusing on detecting serial dependence within a univariate time series, relatively few studies have been done to capture dependence between time series (Haugh (1976); Geweke (1993); Hong (1996)). Moreover, the goal of these papers is restricted to investigating dependence between two covariance stationary time series.

Mutual independence is difficult to test while nonlinear dependence is also not easy to detect. Mutual independence is more demanded than pairwise independence. One conventional measure of linear dependence is the correlation function, which may overlook nonlinear dependent structures that have zero correlations, e.g. Hong (1996). Another useful tool is to utilize the equivalence of the joint distribution and the product of the corresponding marginal distributions under independent case (see Hong (2000); Hong (2005)). Of course, this method can capture all kinds of dependence types since it is a sufficient and necessary condition of independence. However, it is just applicable to pairwise independence test rather than mutual independence test for a large number of high–dimensional random vectors. Hong (1999) developed a generalized spectral density approach via the empirical characteristic function for serial independence test of one time series. This method is also applicable to some types of linear and nonlinear dependencies but only works for detecting pairwise dependence.

In this part, we propose a novel test statistic to test mutual independence for n random vectors of length p when n and p are comparable. Since there are $n \times p$ observed data available, we pool them together to form a data matrix so that some features of the data matrix to investigate independence among the initial n random vectors can be utilized. Large dimensional random matrix theory then serves as a powerful tool to investigate such a matrix. Specifically speaking, we group the n random vectors into a matrix $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$ and then consider the empirical spectral distribution (ESD) of the eigenvalues of the corresponding sample covariance matrix $\mathbf{S} = \frac{1}{n}\mathbf{X}\mathbf{X}^T$, where $\mathbf{x}_i, i = 1, 2 \ldots, n$ are the observed n time series, each being of length p, i.e. $\mathbf{x}_i = (X_{1i}, X_{2i}, \ldots, X_{pi})'$. Here we would like to point out that there have been a substantial set of research works dealing with high dimensional data by random matrix theory (see, for example, Ledoit and Wolf (2002), Johnstone (2001), Birke and Dette (2005) and Yao (2012)). Our approach essentially uses the ESD of the sample covariance matrix $\mathbf{S}$ for n random vectors to distinguish dependence from independence. Our discussion covers both the case where the random vectors are n covariance stationary time series and the case where the random vectors are vectors of linear combinations of independent random variables.

To study the size of the proposed test we first establish the limiting spectral distribution(LSD), i.e. the limit of the ESD of the sample covariance matrix $\mathbf{S}$ under the finite second moment condition on the components. This generalizes the result of Yao (2012), which obtained the LSD under the finite fourth moment condition. Moreover, for the first time we establish a central limit theorem (CLT) for linear spectral statistics of the sample covariance matrices whose columns are covariance stationary time series under the finite fourth moment condition on the time series components. This CLT complements the classical result of linear spectral statistics of the sample covariance matrices of the independent random vectors with i.i.d. components or linear independent structure (see Bai and Silverstein (2009) and Lytova and Pastur (2009)).

As stated above, correlation functions are useful enough for describing linear dependence but can not detect all sorts of nonlinear dependencies. To some extent, our proposed test statistic is also based on a correlation structure, i.e. the sample covariance matrix. A natural question is how our test performs under all sorts of dependent structures. For the Gaussian case, the sample covariance matrix of a linear covariance stationary time series can be written in the form of $\mathbf{S}_1 = \frac{1}{n}\mathbf{T}_1^{1/2}\mathbf{Y}\mathbf{Y}'\mathbf{T}_1^{1/2}$, where $\mathbf{T}_1$ is a $p \times p$ nonnegative positive Hermitian deterministic matrix and $\mathbf{Y}$ is a $p \times n$ random matrix with i.i.d. components. If the cross–sectional dependence can be described as $\frac{1}{n}\mathbf{T}_1^{1/2}\mathbf{Y}\mathbf{T}_2\mathbf{Y}'\mathbf{T}_1^{1/2}$ with $\mathbf{T}_2$ being an $n \times n$ Hermitian deterministic matrix, the limit of its ESD is then given in Theorem 1.2.1 of Zhang (2006), which is different from the limit of the ESD of $\mathbf{S}_1$ corresponding to the independent case. In view of this, our test is able to capture this type of dependent structure. In panel data analysis, the issue of how to characterize cross–sectional dependence attracts great attention among researchers. Spatial models and factor models are two commonly used dependent structures. The simulation given in Section 4 below shows that the proposed test can be applied to these two types of dependence. Finite sample simulations illustrate that the proposed test can also detect some kinds of nonlinear dependence with zero correlations except the "ARCH" dependence. To detect the ARCH dependence we use high power of entries X_{ji} instead of X_{ji} so that the test statistic still works.

1.4 Thesis Outline

The main content of the thesis is organized as follows.

- In Chapter 2, we provide the LSD of the canonical correlation matrix as the dimensions p_1, p_2 are comparable to the sample size n.

- In Chapter 3, the CLT for the linear spectral statistics of the classical canonical correlation matrix is provided. Moreover, the regularized canonical correlation matrix is proposed. The LSD and CLT of the regularized matrix are investigated.

- In Chapter 4, in order to satisfy the necessity of Chapter 3, we investigate a perturbation matrix which is a sample covariance matrix plus a nonnegative definite matrix and provide the CLT for linear spectral statistics of this kind of matrices.

- In Chapter 5, independence test for a large number of high dimensional random vectors with i.i.d. components is investigated and the proposed statistic is based on the characteristic function of the ESD of the corresponding sample covariance matrix.

- In Chapter 6, we provide the LSD and CLT for linear spectral statistics of sample covariance matrices whose columns are independent covariance stationary processes. Based on the first two moments of the ESD of the sample covariance matrices, we propose a novel statistic for independence test for a large number of covariance stationary time series. Moreover, the proposed test can be applied to a large number of random vectors with linear dependent structures.

Chapter 2

Limiting Spectral Distribution Of Canonical Correlation Coefficients

2.1 Introduction

Canonical correlation analysis(CCA) deals with the relationship between two random variable sets. Suppose that there are two random variable sets: $\mathbf{x} = \{x_1, \ldots, x_{p_1}\}$, $\mathbf{y} = \{y_1, \ldots, y_{p_2}\}$, where $p_1 \le p_2$. Assume that there are n observations for each of the $p_1 + p_2$ variables and they are grouped into $p_1 \times n$ random matrix $\mathbf{X} = (X_{ij})_{p_1 \times n}$ and $p_2 \times n$ random matrix $\mathbf{Y} = (Y_{ij})_{p_2 \times n}$ respectively. CCA seeks the linear combinations $\mathbf{a}^T\mathbf{x}$ and $\mathbf{c}^T\mathbf{y}$ that are most highly correlated, that is to maximize

$$r = Corr(\mathbf{a}^T\mathbf{x}, \mathbf{c}^T\mathbf{y}) = \frac{\mathbf{a}^T\Sigma_{\mathbf{xy}}\mathbf{c}}{\sqrt{\mathbf{a}^T\Sigma_{\mathbf{xx}}\mathbf{a}}\sqrt{\mathbf{c}^T\Sigma_{\mathbf{yy}}\mathbf{c}}}, \tag{2.1}$$

where $\Sigma_{\mathbf{xx}}$, $\Sigma_{\mathbf{yy}}$ are population covariance matrices for $\mathbf{x}$, $\mathbf{y}$ respectively; $\Sigma_{\mathbf{xy}}$ is the population covariance matrix between $\mathbf{x}$ and $\mathbf{y}$.

After finding the maximal correlation r_1 and associated combination vectors $\mathbf{a}_1$, $\mathbf{c}_1$, CCA considers seeking a second linear combination $\mathbf{a}_2^T\mathbf{x}$,

$\mathbf{c}_2^T\mathbf{y}$ that has the maximal correlation among all linear combinations uncorrelated with $\mathbf{a}_1^T\mathbf{x}$, $\mathbf{c}_1^T\mathbf{y}$. This procedure can be iterated and successive canonical correlation coefficients $r_1, \ldots, r_{p_1}$ can be found. Substituting population covariance matrices with sample covariance matrices, $r_1, \ldots, r_{p_1}$ can be recast as the roots of the determinant equation

$$det(\mathbf{A}_{xy}\mathbf{A}_y^{-1}\mathbf{A}_{xy}^T - r^2\mathbf{A}_x) = 0, \tag{2.2}$$

where

$$\mathbf{A}_x = \frac{1}{n}\mathbf{X}\mathbf{X}^T,\ \mathbf{A}_y = \frac{1}{n}\mathbf{Y}\mathbf{Y}^T,\ \mathbf{A}_{xy} = \frac{1}{n}\mathbf{X}\mathbf{Y}^T.$$

About this point, one may refer to page 284 of Mardia et al. (1979). The roots of the determinant equation above go under many names, because they figure equally in discriminant analysis, canonical correlation analysis, and invariant tests of linear hypotheses in the multivariate analysis of variance. These are standard techniques in multivariate statistical analysis. Section 4 of Wachter (1980) described how to transform these statistical settings to the determinant equation form. Johnstone (2008) also gave its applications in these aspects in multivariate statistical analysis.

The empirical distribution of the canonical correlation coefficients $r_1, r_2, \cdots, r_{p_1}$ is defined as

$$F(x) = \frac{1}{p_1}\#\{i : r_i \le x\}, \tag{2.3}$$

where $\#\{\cdots\}$ denotes the cardinality of the set $\{\cdots\}$. When the two variable sets $\mathbf{x}$ and $\mathbf{y}$ are independent and each set consists of i.i.d Gaussian random variables, Wachter (1980) proved that the empirical distribution of $r_1, r_2, \cdots, r_{p_1}$ converges in probability and obtained an explicit expression for the limit of the empirical distribution when p_1, p_2 and n are all approaching infinity. From the determinant equation (2.2), it can be seen that $\lambda_1 = r_1^2, \lambda_2 = r_2^2, \ldots, \lambda_{p_1} = r_{p_1}^2$ are eigenvalues of the matrix $\mathbf{S}_{xy} = \mathbf{A}_x^{-1}\mathbf{A}_{xy}\mathbf{A}_y^{-1}\mathbf{A}_{xy}^T$. Hence the analysis of the empirical distribution of $r_1, r_2, \cdots, r_{p_1}$ is equivalent to analyzing the ESD of the matrix $\mathbf{S}_{xy}$. Here for any $p \times p$ matrix $\mathbf{A}$ with real eigenvalues $x_1 \leq x_2 \leq \ldots \leq x_p$, its ESD is defined as

$$F^{\mathbf{A}}(x) = \frac{1}{p}\#\{i : x_i \leq x\}. \tag{2.4}$$

The aim of this chapter is to prove that the result in Wachter (1980) remains true when the entries of $\mathbf{X}$ and $\mathbf{Y}$ have finite second moments but not necessarily Gaussian distribution.

Theorem 1. *Assume that*

(a) $\mathbf{X} = (X_{ij})_{1\leq i\leq p_1, 1\leq j\leq n}$ *where* $X_{ij}, 1 \leq i \leq p_1, 1 \leq j \leq n$, *are i.i.d real random variables with* $EX_{11} = 0$ *and* $E|X_{11}|^2 = 1$.

(b) $\mathbf{Y} = (Y_{ij})_{1\leq i\leq p_2, 1\leq j\leq n}$ *where* $Y_{ij}, 1 \leq i \leq p_2, 1 \leq j \leq n$ *are i.i.d real random variables with* $EY_{11} = 0$ *and* $E|Y_{11}|^2 = 1$.

(c) $p_1 = p_1(n)$ *and* $p_2 = p_2(n)$ *with* $\frac{p_1}{n} \to c_1$ *and* $\frac{p_2}{n} \to c_2$, $c_1, c_2 \in (0, 1)$, *as* $n \to \infty$.

(d) $\mathbf{S}_{xy} = \mathbf{A}_x^{-1}\mathbf{A}_{xy}\mathbf{A}_y^{-1}\mathbf{A}_{xy}^T$ *where* $\mathbf{A}_x = \frac{1}{n}\mathbf{X}\mathbf{X}^T$, $\mathbf{A}_y = \frac{1}{n}\mathbf{Y}\mathbf{Y}^T$ *and* $\mathbf{A}_{xy} = \frac{1}{n}\mathbf{X}\mathbf{Y}^T$.

(e) $\mathbf{X}$ *and* $\mathbf{Y}$ *are independent.*

Then as $n \to \infty$ *the empirical distribution of the matrix* $r_1, r_2, \cdots, r_{p_1}$ *converges almost surely to a fixed distribution function whose density is*

$$\rho(r) = ((r-L)(r+L)(H-r)(H+r))^{\frac{1}{2}}/[\pi c_1 r(1-r)(1+r)],\ r \in [L, H], \tag{2.5}$$

where $L = |(c_2 - c_2c_1)^{\frac{1}{2}} - (c_1 - c_1c_2)^{\frac{1}{2}}|$ *and* $H = (c_2 - c_2c_1)^{\frac{1}{2}} + (c_1 - c_1c_2)^{\frac{1}{2}}$; *and atoms of size* $max(0, 1 - c_2/c_1)$ *at zero and size* $max(0, 1 - (1 - c_2)/c_1)$ *at unity.*

Remark 1. *The inverse of a matrix, such as $\mathbf{A}_x^{-1}$ and $\mathbf{A}_y^{-1}$, is the moore-penrose pseudoinverse, i.e. in the spectral decomposition of the initial matrix, replace each nonzero eigenvalue by its reciprocal and leave the zero eigenvalues alone. This is because under the finite second moment condition, the matrices $\mathbf{A}_x$ and $\mathbf{A}_y$ may be not invertible under the classical inverse matrix definition. However, with the additional assumption that $EX_{11}^4 < \infty$ and $EY_{11}^4 < \infty$, we have the conclusion that the smallest eigenvalues of the sample matrices $\mathbf{A}_x$ and $\mathbf{A}_y$ converge to $(1-\sqrt{c_1})^2$ and $(1-\sqrt{c_2})^2$ respectively[Theorem 5.11 of Bai and Silverstein (2009)], which are not zero since $c_1, c_2 \in (0,1)$. So $\mathbf{A}_x$ and $\mathbf{A}_y$ are invertible with probability one under the finite fourth moment condition.*

As stated previously, it is sufficient to analyze the limiting spectral distribution(LSD) of the matrix $\mathbf{S}_{xy}$, where LSD denotes the limit of the empirical spectral distribution as $n \to \infty$.

The strategy of the proof of Theorem 1 is as follows. Since the matrix $\mathbf{S}_{xy}$ is not symmetric, it is difficult to work on it directly. Instead we consider the $n \times n$ symmetric matrix

$$\mathbf{P}_y\mathbf{P}_x\mathbf{P}_y \tag{2.6}$$

where

$$\mathbf{P}_x = \mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-1}\mathbf{X},\ \mathbf{P}_y = \mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T)^{-1}\mathbf{Y}.$$

Note that $\mathbf{P}_x$ and $\mathbf{P}_y$ are projection matrices. Let $\mathbf{O}_{xy} = (\mathbf{X}\mathbf{X}^T)^{-1}\mathbf{X}\mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T)^{-1}\mathbf{Y}$. Then $\mathbf{S}_{xy} = \mathbf{O}_{xy}\mathbf{X}^T$ and $\mathbf{P}_x\mathbf{P}_y = \mathbf{X}^T\mathbf{O}_{xy}$. By the property that for any two matrices $\mathbf{A}$ and $\mathbf{B}$, the nonzero eigenvalues of $\mathbf{AB}$ are the same as the nonzero eigenvalues of $\mathbf{BA}$, we can derive that, the eigenvalues of the matrix $\mathbf{P}_x\mathbf{P}_y$ are the same as those of the matrix $\mathbf{S}_{xy}$ other than $n-p_1$ zero eigenvalues. Moreover by the fact that $\mathbf{P}_y^2 = \mathbf{P}_y$, $\mathbf{P}_y\mathbf{P}_x\mathbf{P}_y$ and $\mathbf{P}_x\mathbf{P}_y$

also have the same nonzero eigenvalues. So the eigenvalues of the matrix $\mathbf{P}_y\mathbf{P}_x\mathbf{P}_y$ are the same as those of the matrix $\mathbf{S}_{xy}$ other than $n-p_1$ zero eigenvalues, i.e.

$$F^{\mathbf{P}_y\mathbf{P}_x\mathbf{P}_y}(x) = \frac{p_1}{n}F^{\mathbf{S}_{xy}}(x) + \frac{n-p_1}{n}I_{[0,+\infty)}(x). \tag{2.7}$$

By (2.7) and the result in Wachter (1980), one can easily obtain the limit of $F^{\mathbf{P}_y\mathbf{P}_x\mathbf{P}_y}(x)$ when the entries of $\mathbf{X}$ and $\mathbf{Y}$ are Gaussian distributed. To move from the Gaussian case to non-Gaussian case, we mainly use Lindeberg's method (see Lindeberg (1922) and Chatterjee (2006)) and the Stieltjes transform. The Stieltjes transform for any probability distribution function $G(x)$ is defined as

$$m_G(z) = \int \frac{1}{x-z}dG(x), \quad z \in \mathbb{C}^+ \equiv \{z \in \mathbb{C},\ v = \mathrm{Im} z > 0\}. \tag{2.8}$$

An additional key technique is to introduce a perturbation matrix in order to deal with the random matrix $(\mathbf{X}\mathbf{X}^T)^{-1}$ under the finite second moment condition.

2.2 Proof of Theorem 1

We divide the proof of Theorem 1 into 4 parts:

2.2.1 Step 1: Introducing a perturbation matrix

Let

$$\mathbf{A} = \mathbf{P}_y\mathbf{P}_x\mathbf{P}_y.$$

In view of (2.7) it is enough to investigate $F^{\mathbf{A}}$ to prove Theorem 1. In order to deal with the matrix $(\mathbf{X}\mathbf{X}^T)^{-1}$, we make a perturbation of the matrix $\mathbf{A}$

and obtain a new matrix

$$\mathbf{B} = \mathbf{P}_y \mathbf{P}_{tx} \mathbf{P}_y,$$

where $\mathbf{P}_{tx} = \frac{1}{n}\mathbf{X}^T(\frac{1}{n}\mathbf{X}\mathbf{X}^T + t\mathbf{I}_{p_1})^{-1}\mathbf{X}$, $t > 0$ is a small constant number and $\mathbf{I}_{p_1}$ is the identity matrix of the size p_1.

We claim that, with probability one,

$$\lim_{t\to 0}\lim_{n\to\infty} L\Big(F^{\mathbf{A}}, F^{\mathbf{B}}\Big) = 0. \tag{2.9}$$

where $L(F^{\mathbf{A}}, F^{\mathbf{B}})$ is Levy distance between two distribution functions $F^{\mathbf{A}}(\lambda)$ and $F^{\mathbf{B}}(\lambda)$. By Lemma 6 in the Appendix,

$$\begin{aligned} L^3(F^{\mathbf{A}}, F^{\mathbf{B}}) &\le \frac{1}{n}tr(\mathbf{A}-\mathbf{B})^2 \le \frac{1}{n}tr(\mathbf{P}_x - \mathbf{P}_{tx})^2 \\ &= \frac{1}{n}tr\big(\frac{1}{n}\mathbf{X}\mathbf{X}^T[(\frac{1}{n}\mathbf{X}\mathbf{X}^T)^{-1} - (\frac{1}{n}\mathbf{X}\mathbf{X}^T + t\mathbf{I}_{p_1})^{-1}]\big)^2 \\ &\le \frac{t^2}{n}tr(\frac{1}{n}\mathbf{X}\mathbf{X}^T + t\mathbf{I}_{p_1})^{-2}, \end{aligned} \tag{2.10}$$

where the second inequality uses the fact that $||\mathbf{P}_y|| = 1$ with the norm being the spectral norm and the last inequality uses the spectral decomposition of the matrix $\frac{1}{n}\mathbf{X}\mathbf{X}^T$, i.e.

$$\begin{aligned} &\frac{1}{n}\mathbf{X}\mathbf{X}^T[(\frac{1}{n}\mathbf{X}\mathbf{X}^T)^{-1} - (\frac{1}{n}\mathbf{X}\mathbf{X}^T + t\mathbf{I}_{p_1})^{-1}] \\ = \ &\mathbf{U}^T \begin{pmatrix} \mu_1 & & & & & \\ & \ddots & & & & \\ & & \mu_m & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix} \mathbf{U}\mathbf{U}^T \begin{pmatrix} \frac{t}{\mu_1(\mu_1+t)} & & & & & \\ & \ddots & & & & \\ & & \frac{t}{\mu_m(\mu_m+t)} & & & \\ & & & -\frac{1}{t} & & \\ & & & & \ddots & \\ & & & & & -\frac{1}{t} \end{pmatrix} \mathbf{U} \end{aligned}$$

$$= \mathbf{U}^T \begin{pmatrix} \frac{t}{\mu_1+t} & & & & & \\ & \ddots & & & & \\ & & \frac{t}{\mu_m+t} & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix} \mathbf{U} \tag{2.11}$$

with that $\mu_1, \ldots, \mu_m$ are the nonzero eigenvalues of the matrix $\frac{1}{n}\mathbf{X}\mathbf{X}^T$ and the columns of $\mathbf{U}^T$ are the eigenvectors of the matrix $\frac{1}{n}\mathbf{X}\mathbf{X}^T$.

Given $t > 0$, by Theorem 3.6 in Bai and Silverstein (2009) (or see Jonsson (1982) and Marcenko and Pastur (1967)) and the Helly-Bray theorem, we have with probability one

$$\frac{1}{n}tr(\frac{1}{n}\mathbf{X}\mathbf{X}^T + t\mathbf{I}_{p_1})^{-2} = \frac{p_1}{n}\int \frac{1}{(\lambda+t)^2}dF_{p_1}(\lambda) \to c_1\int_a^b \frac{1}{(\lambda+t)^2}dF_{c_1}(\lambda)$$
$$= \int_a^b \frac{\sqrt{(b-\lambda)(\lambda-a)}}{(\lambda+t)^2 2\pi\lambda}d\lambda \le \int_a^b \frac{\sqrt{(b-\lambda)(\lambda-a)}}{\lambda^3 2\pi}d\lambda \le M,$$

where F_{p_1} is the ESD of the sample matrix $\frac{1}{n}\mathbf{X}\mathbf{X}^T$, F_{c_1} is the Marcenko-Pastur Law, $b = (1+\sqrt{c_1})^2$ and $a = (1-\sqrt{c_1})^2$. Here and in what follows M stands for a positive constant number and it may be different from line to line. This, together with (2.10), implies (2.9), as claimed.

Let $\bar{\mathbf{B}}$ and $\bar{\mathbf{A}}$, respectively, denote analogues of the matrices $\mathbf{B}$ and $\mathbf{A}$ with the elements of $\mathbf{X}$ replaced by i.i.d. Gaussian distributed random variables, independent of the entries of $\mathbf{Y}$. By (2.9) and the fact that, for any $\lambda \in \mathbb{R}$,

$$|F^{\mathbf{A}}(\lambda) - F^{\bar{\mathbf{A}}}(\lambda)| \le |F^{\mathbf{A}}(\lambda) - F^{\mathbf{B}}(\lambda)| + |F^{\mathbf{B}}(\lambda) - F^{\bar{\mathbf{B}}}(\lambda)| + |F^{\bar{\mathbf{B}}}(\lambda) - F^{\bar{\mathbf{A}}}(\lambda)|,$$

in order to prove that, for any fixed $t > 0$, with probability one,

$$\lim_{n\to\infty} |F^{\mathbf{A}}(\lambda) - F^{\bar{\mathbf{A}}}(\lambda)| = 0, \tag{2.12}$$

it suffices to prove with probability one,

$$\lim_{n\to\infty} |F^{\mathbf{B}}(\lambda) - F^{\bar{\mathbf{B}}}(\lambda)| = 0. \tag{2.13}$$

If we have (2.12), then for any $\lambda \in \mathbb{R}$, with probability one,

$$\lim_{n\to\infty} |F^{\mathbf{P}_x\mathbf{P}_y}(\lambda) - F^{\mathbf{P}_x^g\mathbf{P}_y}(\lambda)| = 0. \tag{2.14}$$

Since $\mathbf{P}_y$ and $\mathbf{P}_x$ stand symmetric positions in the matrix $\mathbf{P}_x\mathbf{P}_y$, as in (2.12) and (2.14), one can similarly prove that for any $\lambda \in \mathbb{R}$, with probability one,

$$\lim_{n\to\infty} |F^{\mathbf{P}_x^g\mathbf{P}_y}(\lambda) - F^{\mathbf{P}_x^g\mathbf{P}_y^g}(\lambda)| = 0, \tag{2.15}$$

where $\mathbf{P}_y^g$ is obtained from the matrix $\mathbf{P}_y$ with all the entries of $\mathbf{Y}$ replaced by i.i.d Gaussian distributed random variables, independent of $\mathbf{P}_x^g$. Then (2.14) and (2.15) imply that for any $\lambda \in \mathbb{R}$, with probability one,

$$\lim_{n\to\infty} |F^{\mathbf{P}_x\mathbf{P}_y}(\lambda) - F^{\mathbf{P}_x^g\mathbf{P}_y^g}(\lambda)| = 0. \tag{2.16}$$

With the theorem obtained in Wachter (1980) and (2.16), our theorem is easily derived.

Hence the subsequent parts are devoted to proving (2.13).

2.2.2 Step 2: Truncation, Centralization, Rescaling and Tightness of $F^{\mathbf{B}}$

With (1.8) of Bai and Silverstein (2004) and the arguments above and below, we can choose $\varepsilon_n > 0$ such that $\varepsilon_n \to 0$, $n^{1/2}\varepsilon_n \to \infty$ as $n \to \infty$, and $P(|X_{ij}| \geq n^{1/2}\varepsilon_n) \leq \frac{\varepsilon_n}{n}$. Define

$$\tilde{X}_{ij} = X_{ij}I(|X_{ij}| < n^{1/2}\varepsilon_n), \;\; \hat{X}_{ij} = \tilde{X}_{ij} - E\tilde{X}_{11},$$

$$\mathbf{P}_{tx} = \frac{1}{n}\mathbf{X}^T(\frac{1}{n}\mathbf{X}\mathbf{X}^T + t\mathbf{I}_{p_1})^{-1}\mathbf{X},\ \tilde{\mathbf{P}}_{tx} = \frac{1}{n}\tilde{\mathbf{X}}^T(\frac{1}{n}\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T + t\mathbf{I}_{p_1})^{-1}\tilde{\mathbf{X}},$$

$$\hat{\mathbf{P}}_{tx} = \frac{1}{n}\hat{\mathbf{X}}^T(\frac{1}{n}\hat{\mathbf{X}}\hat{\mathbf{X}}^T + t\mathbf{I}_{p_1})^{-1}\hat{\mathbf{X}},\ \tilde{\mathbf{B}} = \mathbf{P}_y\tilde{\mathbf{P}}_{tx}\mathbf{P}_y,\ \hat{\mathbf{B}} = \mathbf{P}_y\hat{\mathbf{P}}_{tx}\mathbf{P}_y,$$

where $\tilde{\mathbf{X}} = (\tilde{X}_{ij})_{1\le i\le p_1;1\le j\le n}$ and $\hat{\mathbf{X}} = (\hat{X}_{ij})_{1\le i\le p_1;1\le j\le n}$.

Let $\eta_{ij} = 1-I(|X_{ij}| < n^{1/2}\varepsilon_n)$. We then get by Lemma 4 in the appendix

$$\sup_\lambda |F^{\mathbf{B}}(\lambda) - F^{\tilde{\mathbf{B}}}(\lambda)| \le \frac{1}{n}rank(\mathbf{P}_y\mathbf{P}_{tx}\mathbf{P}_y - \mathbf{P}_y\tilde{\mathbf{P}}_{tx}\mathbf{P}_y) \le \frac{1}{n}rank(\mathbf{P}_{tx} - \tilde{\mathbf{P}}_{tx})$$

$$\le \frac{1}{n}[rank(\mathbf{X}^T - \tilde{\mathbf{X}}) + rank(\mathbf{X}\mathbf{X}^T - \tilde{\mathbf{X}}\tilde{\mathbf{X}}^T) + rank(\mathbf{X} - \tilde{\mathbf{X}}^T)] \le \frac{4}{n}\sum_{i=1}^{p_1}\sum_{j=1}^{n}\eta_{ij}.$$

Denote $q = P(\eta_{ij} = 1) = P(|X_{ij}| \ge n^{1/2}\varepsilon_n)$. We conclude from Lemma 5 that for any $\delta > 0$,

$$P(\sup_\lambda |F^{\mathbf{B}}(\lambda) - F^{\tilde{\mathbf{B}}}(\lambda)| \ge \delta) \le P(\frac{1}{n}\sum_{i=1}^{p_1}\sum_{j=1}^{n}\eta_{ij} \ge \delta)$$

$$= P(\sum_{i=1}^{p_1}\sum_{j=1}^{n}\eta_{ij} - np_1q \ge np_1(\frac{\delta}{p_1} - q))$$

$$\le 2exp\Big(- \frac{n^2p_1^2(\frac{\delta}{p_1} - q)^2}{2np_1q + np_1(\frac{\delta}{p_1} - q)}\Big) \le 2exp(-nh),$$

for some positive h. It follows from Borel-Cantelli's lemma that

$$\sup_\lambda |F^{\mathbf{B}}(\lambda) - F^{\tilde{\mathbf{B}}}(\lambda)| \to 0, \quad a.s. \quad \text{as } n \to \infty.$$

Next, we prove that

$$\sup_\lambda |F^{\hat{\mathbf{B}}}(\lambda) - F^{\tilde{\mathbf{B}}}(\lambda)| \to 0, \quad a.s. \quad \text{as } n \to \infty. \tag{2.17}$$

Again by Lemma 4 we have

$$\sup_\lambda |F^{\hat{\mathbf{B}}}(\lambda) - F^{\tilde{\mathbf{B}}}(\lambda)| \le \frac{1}{n}rank(\hat{\mathbf{B}} - \tilde{\mathbf{B}}) \le \frac{1}{n}rank\Big[\hat{\mathbf{P}}_{tx} - \tilde{\mathbf{P}}_{tx}\Big]$$

$$\leq \frac{1}{n}rank\Big[\frac{1}{n}\tilde{\mathbf{X}}^T\Big((\frac{1}{n}\hat{\mathbf{X}}\hat{\mathbf{X}}^T + t\mathbf{I}_{p_1})^{-1} - (\frac{1}{n}\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T + t\mathbf{I}_{p_1})^{-1}\Big)\tilde{\mathbf{X}}\Big]$$

$$+\frac{1}{n}rank\Big[\frac{1}{n}\tilde{\mathbf{X}}^T(\frac{1}{n}\hat{\mathbf{X}}\hat{\mathbf{X}}^T+t\mathbf{I}_{p_1})^{-1}E\tilde{\mathbf{X}}\Big]+\frac{1}{n}rank\Big[\frac{1}{n}(E\tilde{\mathbf{X}}^T)(\frac{1}{n}\hat{\mathbf{X}}\hat{\mathbf{X}}^T+t\mathbf{I}_{p_1})^{-1}\tilde{\mathbf{X}}\Big]$$

$$+\frac{1}{n}rank\Big[\frac{1}{n}(E\tilde{\mathbf{X}}^T)(\frac{1}{n}\hat{\mathbf{X}}\hat{\mathbf{X}}^T + t\mathbf{I}_{p_1})^{-1}E\tilde{\mathbf{X}}\Big].$$

Since all elements of $E\tilde{\mathbf{X}}$ are identical, $rank\Big(E\tilde{\mathbf{X}}\Big) = 1$. Moreover, from (2.19)

$$\begin{aligned}
& (\frac{1}{n}\hat{\mathbf{X}}\hat{\mathbf{X}}^T + t\mathbf{I}_{p_1})^{-1} - (\frac{1}{n}\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T + t\mathbf{I}_{p_1})^{-1} \\
= \; & (\frac{1}{n}\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T + t\mathbf{I}_{p_1})^{-1}(\frac{1}{n}\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T - \frac{1}{n}\hat{\mathbf{X}}\hat{\mathbf{X}}^T)(\frac{1}{n}\hat{\mathbf{X}}\hat{\mathbf{X}}^T + t\mathbf{I}_{p_1})^{-1} \\
= \; & \frac{1}{n}(\frac{1}{n}\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T + t\mathbf{I}_{p_1})^{-1}(-E\tilde{\mathbf{X}}E\tilde{\mathbf{X}}^T + \tilde{\mathbf{X}}E\tilde{\mathbf{X}}^T + (E\tilde{\mathbf{X}})\tilde{\mathbf{X}}^T)(\frac{1}{n}\hat{\mathbf{X}}\hat{\mathbf{X}}^T + t\mathbf{I}_{p_1})^{-1}.
\end{aligned}$$

Hence

$$\sup_\lambda |F^{\hat{\mathbf{B}}}(\lambda) - F^{\tilde{\mathbf{B}}}(\lambda)| \leq \frac{M}{n} \to 0.$$

Let $\hat{\sigma}^2 = E(|\hat{X}_{ij}|^2)$ and $\hat{\mathbf{B}} = \frac{1}{n\hat{\sigma}^2}\hat{\mathbf{X}}^T(\frac{1}{n\hat{\sigma}^2}\hat{\mathbf{X}}\hat{\mathbf{X}}^T + t\mathbf{I}_{p_1})^{-1}\hat{\mathbf{X}}$. Then by Lemma 6, we have

$$\begin{aligned}
& L^3(F^{\hat{\mathbf{B}}}, F^{\hat{\mathbf{B}}}) \leq \frac{1}{n}tr(\hat{\mathbf{B}} - \hat{\mathbf{B}})^2 \\
= \; & \frac{(\hat{\sigma}^2-1)^2t^2}{n}tr\Big(\frac{1}{n}\hat{\mathbf{X}}\hat{\mathbf{X}}^T(\frac{1}{n}\hat{\mathbf{X}}\hat{\mathbf{X}}^T + \hat{\sigma}^2t\mathbf{I}_{p_1})^{-1}(\frac{1}{n}\hat{\mathbf{X}}\hat{\mathbf{X}}^T + t\mathbf{I}_{p_1})^{-1}\Big)^2 \\
= \; & \frac{(\hat{\sigma}^2-1)^2t^2}{n}tr\Big((\frac{1}{n}\hat{\mathbf{X}}\hat{\mathbf{X}}^T + \hat{\sigma}^2t\mathbf{I}_{p_1} - \hat{\sigma}^2t\mathbf{I}_{p_1})(\frac{1}{n}\hat{\mathbf{X}}\hat{\mathbf{X}}^T + \hat{\sigma}^2t\mathbf{I}_{p_1})^{-1}(\frac{1}{n}\hat{\mathbf{X}}\hat{\mathbf{X}}^T + t\mathbf{I}_{p_1})^{-1}\Big)^2 \\
= \; & \frac{(\hat{\sigma}^2-1)^2t^2}{n}tr\Big((\frac{1}{n}\hat{\mathbf{X}}\hat{\mathbf{X}}^T + t\mathbf{I}_{p_1})^{-1} - \hat{\sigma}^2t(\frac{1}{n}\hat{\mathbf{X}}\hat{\mathbf{X}}^T + \hat{\sigma}^2t\mathbf{I}_{p_1})^{-1}(\frac{1}{n}\hat{\mathbf{X}}\hat{\mathbf{X}}^T + t\mathbf{I}_{p_1})^{-1}\Big)^2 \\
\leq \; & \frac{(\hat{\sigma}^2-1)^2t^2}{n}p_1\Big(||(\frac{1}{n}\hat{\mathbf{X}}\hat{\mathbf{X}}^T + t\mathbf{I}_{p_1})^{-1}|| \\
& +\hat{\sigma}^2t||(\frac{1}{n}\hat{\mathbf{X}}\hat{\mathbf{X}}^T + \hat{\sigma}^2t\mathbf{I}_{p_1})^{-1}|| \cdot ||(\frac{1}{n}\hat{\mathbf{X}}\hat{\mathbf{X}}^T + t\mathbf{I}_{p_1})^{-1}||\Big)^2 \\
\leq \; & \frac{(\hat{\sigma}^2-1)^2t^2}{n}p_1\frac{4}{t^2} \to 0,
\end{aligned}$$

because $\hat{\sigma}^2 \to 1$ and $p_1/n \to c_1$ as $n \to \infty$; where the first equality uses the formula (2.19); the second inequality uses the matrix inequality that

$$tr(\mathbf{C}) \leq p_1 ||\mathbf{C}||,$$

holding for any $p_1 \times p_1$ normal matrix $\mathbf{C}$; and the last inequality uses the fact that

$$||(\frac{1}{n}\hat{\mathbf{X}}\hat{\mathbf{X}}^T + \hat{\sigma}^2 t\mathbf{I}_{p_1})^{-1}|| \leq \frac{1}{\hat{\sigma}^2 t}, \ ||(\frac{1}{n}\hat{\mathbf{X}}\hat{\mathbf{X}}^T + t\mathbf{I}_{p_1})^{-1}|| \leq \frac{1}{t}.$$

In view of the truncation, centralization and rescaling steps above, in the sequel, we shall assume that the underlying variables satisfy

$$|X_{ij}| \leq n^{1/2}\varepsilon_n, \quad EX_{ij} = 0, \quad EX_{ij}^2 = 1, \tag{2.18}$$

and for simplicity we shall still use notation X_{ij} instead of $\hat{X}_{ij}$.

We now turn to investigating the tightness of $F^{\mathbf{B}}$. For any constant number $K > 0$,

$$\int_{\lambda > K} dF^{\mathbf{B}} \leq \frac{1}{K}\int \lambda dF^{\mathbf{B}} = \frac{1}{K}\frac{1}{n} tr[\mathbf{P}_y \mathbf{P}_{tx} \mathbf{P}_y]$$

Since the largest eigenvalue of $\mathbf{P}_y$ is 1 and $\mathbf{P}_{tx}$ is a nonnegative matrix we obtain

$$\begin{aligned} tr[\mathbf{P}_y \mathbf{P}_{tx} \mathbf{P}_y] &= tr[\mathbf{P}_y \mathbf{P}_{tx}] \\ &\leq tr[\mathbf{P}_{tx}] = tr[\frac{1}{n}\mathbf{X}\mathbf{X}^T(\frac{1}{n}\mathbf{X}\mathbf{X}^T + t\mathbf{I}_{p_1})^{-1}] \leq n. \end{aligned}$$

The last inequality has used the facts that $t > 0$ and that all the eigenvalues of $\frac{1}{n}\mathbf{X}\mathbf{X}^T(\frac{1}{n}\mathbf{X}\mathbf{X}^T + t\mathbf{I}_{p_1})^{-1}$ are less than 1.

It follows that $F^{\mathbf{B}}$ is tight.

2.2.3 Step 3: Convergence of the random part

Let

$$\mathbf{B}^{-1}(z) = (\mathbf{P}_y \mathbf{P}_{tx} \mathbf{P}_y - z\mathbf{I})^{-1}.$$

The aim in this section is to prove that

$$\frac{1}{n} tr \mathbf{B}^{-1}(z) - E\frac{1}{n} tr \mathbf{B}^{-1}(z) \to 0 \quad \text{a.s. as } n \to \infty.$$

To this end we introduce some notation. Let $\mathbf{x}_k$ denote the kth column of $\mathbf{X}$ and $\mathbf{e}_k$ the column vector of the size of p_1 with the kth element being 1 and otherwise 0. Moreover, define $\mathbf{X}_k$ to be the matrix obtained from $\mathbf{X}$ by replacing the elements of the kth column of $\mathbf{X}$ with 0.

Fix $v = \Im z > 0$. Define $\mathcal{F}_k$ to be the σ-field generated by $\mathbf{x}_1, \cdots, \mathbf{x}_k$. Let $E_k(\cdot)$ denote the conditional expectation with respect to $\mathcal{F}_k$ and E_0 denote expectation. That is, $E_k(\cdot) = E(\cdot|\mathcal{F}_k)$ and $E_0(\cdot) = E(\cdot)$. Let

$$\mathbf{B}_k = \mathbf{P}_y \mathbf{P}_k^{tx} \mathbf{P}_y, \ \mathbf{B}_k^{-1}(z) = (\mathbf{P}_y \mathbf{P}_k^{tx} \mathbf{P}_y - z\mathbf{I})^{-1},$$

where $\mathbf{P}_{tx} = \frac{1}{n}\mathbf{X}^T(\frac{1}{n}\mathbf{X}\mathbf{X}^T + t\mathbf{I}_{p_1})^{-1}\mathbf{X}$, $\mathbf{P}_k^{tx} = \frac{1}{n}\mathbf{X}_k^T(\frac{1}{n}\mathbf{X}_k\mathbf{X}_k^T + t\mathbf{I}_{p_1})^{-1}\mathbf{X}_k$. Define $\mathbf{H}_k^{-1} = (\frac{1}{n}\mathbf{X}_k\mathbf{X}_k^T + t\mathbf{I}_{p_1})^{-1}$ and $\mathbf{H}^{-1} = (\frac{1}{n}\mathbf{X}\mathbf{X}^T + t\mathbf{I}_{p_1})^{-1}$.

Note that $\mathbf{X} = \mathbf{X}_k + \mathbf{x}_k\mathbf{e}_k^T$, that the elements of $\mathbf{X}_k\mathbf{e}_k$ are all zero and hence that

$$\mathbf{X}\mathbf{X}^T - \mathbf{X}_k\mathbf{X}_k^T = \mathbf{x}_k\mathbf{x}_k^T.$$

This implies that

$$\mathbf{H}_k^{-1} - \mathbf{H}^{-1} = \frac{1}{n}\mathbf{H}^{-1}\mathbf{x}_k\mathbf{x}_k^T\mathbf{H}_k^{-1} = \frac{1}{1+\frac{1}{n}\mathbf{x}_k^T\mathbf{H}_k^{-1}\mathbf{x}_k}\frac{1}{n}\mathbf{H}_k^{-1}\mathbf{x}_k\mathbf{x}_k^T\mathbf{H}_k^{-1},$$

where we make use of the formula

$$\mathbf{A}_1^{-1} - \mathbf{A}_2^{-1} = \mathbf{A}_2^{-1}(\mathbf{A}_2 - \mathbf{A}_1)\mathbf{A}_1^{-1}, \tag{2.19}$$

holding for any two invertible matrices $\mathbf{A}_1$ and $\mathbf{A}_2$;

and

$$(\mathbf{U} + \mathbf{u}\mathbf{v}^T)^{-1}\mathbf{u} = \frac{\mathbf{U}^{-1}\mathbf{u}}{1 + \mathbf{v}^T\mathbf{U}^{-1}\mathbf{u}}, \tag{2.20}$$

holding for any invertible matrices $\mathbf{U}$ and $(\mathbf{U} + \mathbf{u}\mathbf{v}^T)$, vectors $\mathbf{u}$ and $\mathbf{v}$. We then write

$$\mathbf{B}_k - \mathbf{B} = \mathbf{P}_y(\mathbf{P}_k^{tx} - \mathbf{P}_{tx})\mathbf{P}_y = \mathbf{P}_y(C_1 + C_2 + C_3 + C_4)\mathbf{P}_y, \tag{2.21}$$

where

$$C_1 = \frac{1}{n}\frac{\mathbf{X}_k^T\mathbf{H}_k^{-1}\frac{1}{n}\mathbf{x}_k\mathbf{x}_k^T\mathbf{H}_k^{-1}\mathbf{X}_k}{1 + \frac{1}{n}\mathbf{x}_k^T\mathbf{H}_k^{-1}\mathbf{x}_k}, \quad C_2 = -\frac{1}{n}\frac{\mathbf{X}_k^T\mathbf{H}_k^{-1}\mathbf{x}_k\mathbf{e}_k^T}{1 + \frac{1}{n}\mathbf{x}_k^T\mathbf{H}_k^{-1}\mathbf{x}_k},$$

$$C_3 = -\frac{1}{n}\frac{\mathbf{e}_k\mathbf{x}_k^T\mathbf{H}_k^{-1}\mathbf{X}_k}{1 + \frac{1}{n}\mathbf{x}_k^T\mathbf{H}_k^{-1}\mathbf{x}_k}, \quad C_4 = -\frac{1}{n}\frac{\mathbf{e}_k\mathbf{x}_k^T\mathbf{H}_k^{-1}\mathbf{x}_k\mathbf{e}_k^T}{1 + \frac{1}{n}\mathbf{x}_k^T\mathbf{H}_k^{-1}\mathbf{x}_k}. \tag{2.22}$$

Now write

$$\frac{1}{n}tr\mathbf{B}^{-1}(z) - E\frac{1}{n}tr\mathbf{B}^{-1}(z) = \frac{1}{n}\sum_{k=1}^{n}[E_k tr\mathbf{B}^{-1}(z) - E_{k-1}tr\mathbf{B}^{-1}(z)]$$

$$= \frac{1}{n}\sum_{k=1}^{n}(E_k - E_{k-1})(tr\mathbf{B}^{-1}(z) - tr\mathbf{B}_k^{-1}(z))$$

$$= \frac{1}{n}\sum_{k=1}^{n}(E_k - E_{k-1})\Big[\sum_{i=1}^{4} tr\Big(\mathbf{B}_k^{-1}(z)\mathbf{P}_y C_i \mathbf{P}_y \mathbf{B}^{-1}(z)\Big)\Big],$$

where the last step uses (2.19) and (2.21). Let $||\cdot||$ denote the spectral norm of matrices or the Euclidean norm of vectors. It is observed that

$$||\mathbf{B}^{-1}(z)|| \le \frac{1}{v}, \quad ||\mathbf{B}_k^{-1}(z)|| \le \frac{1}{v}, \quad ||\mathbf{P}_y|| \le 1, \quad \frac{1}{p_1}tr\mathbf{H}_k^{-1} \le \frac{1}{t}. \tag{2.23}$$

and since $\mathbf{x}_k^T\mathbf{H}_k^{-1}\mathbf{x}_k \ge 0$ we have

$$\frac{1}{1 + \frac{1}{n}\mathbf{x}_k^T\mathbf{H}_k^{-1}\mathbf{x}_k} \le 1. \tag{2.24}$$

It follows that

$$|tr\mathbf{B}_k^{-1}(z)\mathbf{P}_y\mathbf{C}_1\mathbf{P}_y\mathbf{B}^{-1}(z)| = \frac{1}{n^2}\Big|\frac{\mathbf{x}_k^T\mathbf{H}_k^{-1}\mathbf{X}_k\mathbf{P}_y\mathbf{B}^{-1}(z)\mathbf{B}_k^{-1}(z)\mathbf{P}_y\mathbf{X}_k^T\mathbf{H}_k^{-1}\mathbf{x}_k}{1+\frac{1}{n}\mathbf{x}_k^T\mathbf{H}_k^{-1}\mathbf{x}_k}\Big|$$
$$\leq \frac{1}{v^2n^2}||\mathbf{x}_k^T\mathbf{H}_k^{-1}\mathbf{X}_k||^2 \leq \frac{1}{v^2n}|\mathbf{x}_k^T\mathbf{H}_k^{-1}\mathbf{x}_k| + \frac{t}{v^2n}|\mathbf{x}_k^T\mathbf{H}_k^{-2}\mathbf{x}_k|, \tag{2.25}$$

where the last inequality uses the facts that $||\mathbf{x}_k^T\mathbf{H}_k^{-1}\mathbf{X}_k||^2 = \mathbf{x}_k^T\mathbf{H}_k^{-1}\mathbf{X}_k\mathbf{X}_k^T\mathbf{H}_k^{-1}\mathbf{x}_k$ and $\mathbf{H}_k^{-1}\mathbf{X}_k\mathbf{X}_k^T\mathbf{H}_k^{-1} = n\mathbf{H}_k^{-1}(\frac{1}{n}\mathbf{X}_k\mathbf{X}_k^T + t\mathbf{I}_{p_1} - t\mathbf{I}_{p_1})\mathbf{H}_k^{-1} = n\mathbf{H}_k^{-1} - nt\mathbf{H}_k^{-2}$.

We then conclude from Lemma 2, (2.23)-(2.25) that

$$\begin{aligned} & E\left|\frac{1}{n}\sum_{k=1}^n(E_k - E_{k-1})tr\mathbf{B}_k^{-1}(z)\mathbf{P}_y\mathbf{C}_1\mathbf{P}_y\mathbf{B}^{-1}(z)\right|^4 \\ \leq\ & \frac{M}{n^3}\sum_{k=1}^n E\Big|tr\mathbf{B}_k^{-1}(z)\mathbf{P}_y\mathbf{C}_1\mathbf{P}_y\mathbf{B}^{-1}(z)\Big|^4 \\ \leq\ & \frac{M}{n^7}\sum_{k=1}^n E\Big|\mathbf{x}_k^T\mathbf{H}_k^{-1}\mathbf{x}_k\Big|^4 + \frac{M}{n^7}\sum_{k=1}^n E\Big|\mathbf{x}_k^T\mathbf{H}_k^{-2}\mathbf{x}_k\Big|^4 \\ =\ & O(\frac{1}{n^2}), \end{aligned}$$

where the last step uses the facts that via Lemma 3 and (2.18)

$$\frac{1}{n^4}E\Big|\mathbf{x}_k^T\mathbf{H}_k^{-1}\mathbf{x}_k\Big|^4 \leq \frac{1}{n^4}ME\Big|\mathbf{x}_k^T\mathbf{H}_k^{-1}\mathbf{x}_k - tr\mathbf{H}_k^{-1}\Big|^4 + \frac{1}{n^4}ME|tr\mathbf{H}_k^{-1}|^4 \leq M \tag{2.26}$$

and that

$$\frac{1}{n^4}E\Big|\mathbf{x}_k^T\mathbf{H}_k^{-2}\mathbf{x}_k\Big|^4 \leq M. \tag{2.27}$$

Similarly, we can also obtain for $i = 2, 3, 4$,

$$E|\frac{1}{n}\sum_{k=1}^n(E_k - E_{k-1})tr\mathbf{B}^{-1}(z)\mathbf{P}_y\mathbf{C}_i\mathbf{P}_y\mathbf{B}_k^{-1}(z)|^4 \leq \frac{M}{n^2}. \tag{2.28}$$

It follows from Borel-Cantelli's lemma that

$$\frac{1}{n}tr\mathbf{B}^{-1}(z) - E\frac{1}{n}tr\mathbf{B}^{-1}(z) \quad \text{a.s. } n \to \infty. \tag{2.29}$$

2.2.4 Step 4: From Gaussian distribution to general distributions

This section is to prove that

$$E[\frac{1}{n}tr\mathbf{B}^{-1}(z)] - E[\frac{1}{n}tr\mathbf{D}^{-1}(z)] \to 0 \quad as \quad n \to \infty, \tag{2.30}$$

where $\mathbf{D}^{-1}(z) = (\mathbf{P}_y\mathbf{P}_{tx}^g\mathbf{P}_y - z\mathbf{I})^{-1}$, $\mathbf{P}_{tx}^g = \frac{1}{n}\mathbf{G}^T(\frac{1}{n}\mathbf{G}\mathbf{G}^T + t\mathbf{I}_{p_1})^{-1}\mathbf{G}$ and $\mathbf{G} = (G_{ij})_{p_1\times n}$ consists of i.i.d. Gaussian random variables. We would point out that (2.13) follows immediately from (2.29), (2.30), tightness of $F^{\mathbf{B}}$ and the well-known inversion formula for Stieltjes transform[Theorem B.8 of Bai and Silverstein (2009)]. We use Lindeberg's method in Chatterjee (2006) to prove this result.

To facilitate statements, denote

$$X_{11}, \cdots, X_{1n}, X_{21}, \cdots, X_{p_1n} \; respectively \; by \; \hat{X}_1, \cdots, \hat{X}_n, \hat{X}_{n+1}, \cdots, \hat{X}_{p_1n}$$

and

$$G_{11}, \cdots, G_{1n}, G_{21}, \cdots, G_{p_1n} \; respectively \; by \; \hat{G}_1, \cdots, \hat{G}_n, \hat{G}_{n+1}, \cdots, \hat{G}_{p_1n}.$$

For each j, $0 \le j \le p_1n$, set

$$\mathbf{Z}_j = (\hat{X}_1, \cdots, \hat{X}_j, \hat{G}_{j+1}, \cdots, \hat{G}_{p_1n}) \; and \; \mathbf{Z}_j^0 = (\hat{X}_1, \cdots, \hat{X}_{j-1}, 0, \hat{G}_{j+1}, \cdots, \hat{G}_{p_1n}). \tag{2.31}$$

Note that $\mathbf{X}$ in $\mathbf{B}^{-1}(z)$ consists of the entries of $\mathbf{Z}_{p_1n}$. Hence we denote $\frac{1}{n}tr\mathbf{B}^{-1}(z)$ by $\frac{1}{n}tr(\mathbf{B}(\mathbf{Z}_{p_1n}) - zI)^{-1}$. Define the mapping f from R^{np_1} to C as

$$f(\mathbf{Z}_{p_1n}) = \frac{1}{n}tr(\mathbf{B}(\mathbf{Z}_{p_1n}) - z\mathbf{I})^{-1}. \tag{2.32}$$

Furthermore we use the entries of $\mathbf{Z}_j$, $j = 0, 1, \cdots, p_1n - 1$, respectively, to replace $\hat{X}_1, \cdots, \hat{X}_{p_1n}$, the entries of $\mathbf{X}$ in $\mathbf{B}$, to constitute a series of new matrices. For these new matrices, we define $f(\mathbf{Z}_j)$, $j = 0, 1, \cdots, p_1n-1$ as $f(\mathbf{Z}_{p_1n})$ is defined for the matrix $\mathbf{B}$. For example, $f(\mathbf{Z}_0) = \frac{1}{n}tr\mathbf{D}^{-1}(z)$. We then write

$$E[\frac{1}{n}tr\mathbf{B}^{-1}(z)] - E[\frac{1}{n}tr\mathbf{D}^{-1}(z)] = \sum_{j=1}^{p_1n} E\Big(f(\mathbf{Z}_j) - f(\mathbf{Z}_{j-1})\Big).$$

A third Taylor expansion yields

$$f(\mathbf{Z}_j) = f(\mathbf{Z}_j^0) + \hat{X}_j\partial_j f(\mathbf{Z}_j^0) + \frac{1}{2}\hat{X}_j^2\partial_j^2 f(\mathbf{Z}_j^0) + \frac{1}{2}\hat{X}_j^3\int_0^1 (1-\tau)^2\partial_j^3 f(\mathbf{Z}_j^{(1)}(\tau))d\tau,$$

$$f(\mathbf{Z}_{j-1}) = f(\mathbf{Z}_j^0) + \hat{G}_j\partial_j f(\mathbf{Z}_j^0) + \frac{1}{2}\hat{G}_j^2\partial_j^2 f(\mathbf{Z}_j^0) + \frac{1}{2}\hat{G}_j^3\int_0^1 (1-\tau)^2\partial_j^3 f(\mathbf{Z}_{j-1}^{(2)}(\tau))d\tau,$$

where $\partial_j^r f(\cdot)$, $r = 1, 2, 3$, stand for the r-fold derivative of the function f in the j-th coordinate, and

$$\mathbf{Z}_j^{(1)}(\tilde{t}) = (\hat{X}_1, \cdots, \hat{X}_{j-1}, \tau\hat{X}_j, \hat{G}_{j+1}, \cdots, \hat{G}_{pn}),$$

$$\mathbf{Z}_{j-1}^{(2)}(\tilde{t}) = (\hat{X}_1, \cdots, \hat{X}_{j-1}, \tau\hat{G}_j, \hat{G}_{j+1}, \cdots, \hat{G}_{pn}).$$

Since $\hat{X}_j$ and $\hat{G}_j$ are both independent of $\mathbf{Z}_j^0$, $E[\hat{X}_j] = E[\hat{G}_j] = 0$ and $E[\hat{X}_j^2] = E[\hat{G}_j^2] = 1$, we obtain

$$\begin{aligned} & E[\frac{1}{n}tr\mathbf{B}^{-1}(z)] - E[\frac{1}{n}tr\mathbf{D}^{-1}(z)] \\ = \; & \frac{1}{2}\sum_{j=1}^{p_1n} E\Big[\hat{X}_j^3\int_0^1 (1-\tau)^2\partial_j^3 f(\mathbf{Z}_j^{(1)}(\tau))d\tau - \hat{G}_j^3\int_0^1 (1-\tau)^2\partial_j^3 f(\mathbf{Z}_{j-1}^{(2)}(\tau))d\tau\Big]. \end{aligned}$$

Next we evaluate $\partial_j^3 f(\mathbf{Z}_{p_1n}^{(1)}(\tau))$. Note that

$$\frac{\partial\mathbf{H}^{-1}}{\partial X_{ij}} = -\mathbf{H}^{-1}\frac{\partial\mathbf{H}}{\partial X_{ij}}\mathbf{H}^{-1}. \tag{2.33}$$

A simple but tedious calculation indicates that

$$\frac{\partial\mathbf{B}}{\partial X_{ij}} \;=\; \frac{1}{n}\mathbf{P}_y\mathbf{e}_j\mathbf{e}_i^T\mathbf{H}^{-1}\mathbf{X}\mathbf{P}_y + \frac{1}{n}\mathbf{P}_y\mathbf{X}^T\mathbf{H}^{-1}\mathbf{e}_i\mathbf{e}_j^T\mathbf{P}_y$$

$$-\frac{1}{n^2}\mathbf{P}_y\mathbf{X}^T\mathbf{H}^{-1}(\mathbf{e}_i\mathbf{e}_j^T\mathbf{X}^T+\mathbf{X}\mathbf{e}_j\mathbf{e}_i^T)\mathbf{H}^{-1}\mathbf{X}\mathbf{P}_y,$$

$$\begin{aligned}\frac{\partial^2\mathbf{B}}{\partial X_{ij}^2} &= \frac{2}{n}\mathbf{P}_y\mathbf{e}_j\mathbf{e}_i^T\mathbf{H}^{-1}\mathbf{e}_i\mathbf{e}_j^T\mathbf{P}_y-\frac{2}{n^2}\mathbf{P}_y\mathbf{e}_j\mathbf{e}_i^T\mathbf{H}^{-1}(\mathbf{e}_i\mathbf{e}_j^T\mathbf{X}^T+\mathbf{X}\mathbf{e}_j\mathbf{e}_i^T)\mathbf{H}^{-1}\mathbf{X}\mathbf{P}_y\\ &-\frac{2}{n^2}\mathbf{P}_y\mathbf{X}^T\mathbf{H}^{-1}(\mathbf{e}_i\mathbf{e}_j^T\mathbf{X}^T+\mathbf{X}\mathbf{e}_j\mathbf{e}_i^T)\mathbf{H}^{-1}\mathbf{e}_i\mathbf{e}_j^T\mathbf{P}_y-\frac{2}{n^2}\mathbf{P}_y\mathbf{X}^T\mathbf{H}^{-1}\mathbf{e}_i\mathbf{e}_i^T\mathbf{H}^{-1}\mathbf{X}\mathbf{P}_y\\ &+\frac{2}{n^3}\mathbf{P}_y\mathbf{X}^T[\mathbf{H}^{-1}(\mathbf{e}_i\mathbf{e}_j^T\mathbf{X}^T+\mathbf{X}\mathbf{e}_j\mathbf{e}_i^T)]^2\mathbf{H}^{-1}\mathbf{X}\mathbf{P}_y,\end{aligned}$$

$$\begin{aligned}\frac{\partial^3\mathbf{B}}{\partial X_{ij}^3} &= -\frac{6}{n^2}\mathbf{P}_y\mathbf{e}_j\mathbf{e}_i^T\mathbf{H}^{-1}(\mathbf{e}_i\mathbf{e}_j^T\mathbf{X}^T+\mathbf{X}\mathbf{e}_j\mathbf{e}_i^T)\mathbf{H}^{-1}\mathbf{e}_i\mathbf{e}_j^T\mathbf{P}_y\\ &-\frac{6}{n^2}\mathbf{P}_y\mathbf{e}_j\mathbf{e}_i^T\mathbf{H}^{-1}\mathbf{e}_i\mathbf{e}_i^T\mathbf{H}^{-1}\mathbf{X}\mathbf{P}_y\\ &+\frac{6}{n^3}\mathbf{P}_y\mathbf{e}_j\mathbf{e}_i^T[\mathbf{H}^{-1}(\mathbf{e}_i\mathbf{e}_j^T\mathbf{X}^T+\mathbf{X}\mathbf{e}_j\mathbf{e}_i^T)]^2\mathbf{H}^{-1}\mathbf{X}\mathbf{P}_y\\ &-\frac{6}{n^2}\mathbf{P}_y\mathbf{X}^T\mathbf{H}^{-1}\mathbf{e}_i\mathbf{e}_i^T\mathbf{H}^{-1}\mathbf{e}_i\mathbf{e}_j^T\mathbf{P}_y\\ &+\frac{6}{n^3}\mathbf{P}_y\mathbf{X}^T[\mathbf{H}^{-1}(\mathbf{e}_i\mathbf{e}_j^T\mathbf{X}^T+\mathbf{X}\mathbf{e}_j\mathbf{e}_i^T)]^2\mathbf{H}^{-1}\mathbf{e}_i\mathbf{e}_j^T\mathbf{P}_y\\ &-\frac{6}{n^4}\mathbf{P}_y\mathbf{X}^T[\mathbf{H}^{-1}(\mathbf{e}_i\mathbf{e}_j^T\mathbf{X}^T+\mathbf{X}\mathbf{e}_j\mathbf{e}_i^T)]^3\mathbf{H}^{-1}\mathbf{X}\mathbf{P}_y\\ &+\frac{6}{n^3}\mathbf{P}_y\mathbf{X}^T\mathbf{H}^{-1}(\mathbf{e}_i\mathbf{e}_j^T\mathbf{X}^T+\mathbf{X}\mathbf{e}_j\mathbf{e}_i^T)\mathbf{H}^{-1}\mathbf{e}_i\mathbf{e}_i^T\mathbf{H}^{-1}\mathbf{X}\mathbf{P}_y\\ &+\frac{6}{n^3}\mathbf{P}_y\mathbf{X}^T\mathbf{H}^{-1}\mathbf{e}_i\mathbf{e}_i^T\mathbf{H}^{-1}(\mathbf{e}_i\mathbf{e}_j^T\mathbf{X}^T+\mathbf{X}\mathbf{e}_j\mathbf{e}_i^T)\mathbf{H}^{-1}\mathbf{X}\mathbf{P}_y.\end{aligned}$$

Also, by the formula

$$\frac{1}{n}\frac{\partial tr\mathbf{B}^{-1}(z)}{\partial X_{ij}}=-\frac{1}{n}tr(\frac{\partial\mathbf{B}}{\partial X_{ij}}\mathbf{B}^{-2}(z)),$$

it is easily seen that

$$\begin{aligned}\frac{1}{n}\frac{\partial^3 tr\mathbf{B}^{-1}(z)}{\partial X_{ij}^3} &= -\frac{6}{n}tr(\frac{\partial\mathbf{B}}{\partial X_{ij}}\mathbf{B}^{-1}(z)\frac{\partial\mathbf{B}}{\partial X_{ij}}\mathbf{B}^{-1}(z)\frac{\partial\mathbf{B}}{\partial X_{ij}}\mathbf{B}^{-2}(z))\\ &-\frac{1}{n}tr(\frac{\partial^3\mathbf{B}}{\partial X_{ij}^3}\mathbf{B}^{-2}(z))+\frac{3}{n}tr(\frac{\partial^2\mathbf{B}}{\partial X_{ij}^2}\mathbf{B}^{-2}(z)\frac{\partial\mathbf{B}}{\partial X_{ij}}\mathbf{B}^{-1}(z))\\ &+\frac{3}{n}tr(\frac{\partial^2\mathbf{B}}{\partial X_{ij}^2}\mathbf{B}^{-1}(z)\frac{\partial\mathbf{B}}{\partial X_{ij}}\mathbf{B}^{-2}(z)).\end{aligned}$$

There are lots of terms in the expansion of $\frac{1}{n}\frac{\partial^3 tr\mathbf{B}^{-1}(z)}{\partial X_{ij}^3}$ and therefore we do not enumerate all the terms here. By using the formula that, for any matrices $\mathbf{A}$, $\mathbf{B}$ and column vectors $\mathbf{e}_j$ and $\mathbf{e}_k$,

$$tr(\mathbf{A}\mathbf{e}_j\mathbf{e}_k^T\mathbf{B}) = \mathbf{e}_k^T\mathbf{B}\mathbf{A}\mathbf{e}_j, \tag{2.34}$$

all the terms of $\frac{1}{n}\frac{\partial^3 tr\mathbf{B}^{-1}(z)}{\partial X_{ij}^3}$ can be dominated by a common expression. That is

$$\begin{aligned}
||\frac{1}{n}\frac{\partial^3 tr\mathbf{B}^{-1}(z)}{\partial X_{ij}^3}|| \;\leq\; & \frac{M}{n^3}||\mathbf{H}^{-1}||\cdot||\mathbf{X}^T\mathbf{H}^{-1}|| + \frac{M}{n^4}||\mathbf{X}^T\mathbf{H}^{-1}||^3 \\
& + \frac{M}{n^4}||\mathbf{H}^{-1}||\cdot||\mathbf{X}^T\mathbf{H}^{-1}||^2 \\
& + \frac{M}{n^4}||\mathbf{H}^{-1}||\cdot||\mathbf{X}^T\mathbf{H}^{-1}||\cdot||\mathbf{X}^T\mathbf{H}^{-1}\mathbf{X}|| \\
& + \frac{M}{n^5}||\mathbf{H}^{-1}||\cdot||\mathbf{X}^T\mathbf{H}^{-1}||\cdot||\mathbf{X}^T\mathbf{H}^{-1}\mathbf{X}||^2 \\
& + \frac{M}{n^5}||\mathbf{X}^T\mathbf{H}^{-1}||^3\cdot||\mathbf{X}^T\mathbf{H}^{-1}\mathbf{X}|| \\
& + \frac{M}{n^6}||\mathbf{X}^T\mathbf{H}^{-1}||^3\cdot||\mathbf{X}^T\mathbf{H}^{-1}\mathbf{X}||^2 \\
& + \frac{M}{n^7}||\mathbf{X}^T\mathbf{H}^{-1}||^3\cdot||\mathbf{X}^T\mathbf{H}\mathbf{X}||^3.
\end{aligned} \tag{2.35}$$

Obviously

$$||\mathbf{H}^{-1}|| \leq \frac{1}{t}. \tag{2.36}$$

It is observed that

$$\begin{aligned}
||\mathbf{X}^T\mathbf{H}^{-1}\mathbf{X}||^2 &= \lambda_{\max}(\mathbf{X}^T\mathbf{H}^{-1}\mathbf{X}\mathbf{X}^T\mathbf{H}^{-1}\mathbf{X}) = \lambda_{\max}(\mathbf{H}^{-1}\mathbf{X}\mathbf{X}^T\mathbf{H}^{-1}\mathbf{X}\mathbf{X}^T) \\
&\leq n^2[1 + 2t||\mathbf{H}^{-1}|| + t^2||\mathbf{H}^{-2}||] \leq Mn^2,
\end{aligned} \tag{2.37}$$

where $\lambda_{\max}(\cdot)$ denotes the maximum eigenvalue of the corresponding matrix; and the first inequality above utilizes the fact that $\mathbf{H}^{-1}\mathbf{X}\mathbf{X}^T = n\mathbf{H}^{-1}(\frac{1}{n}\mathbf{X}\mathbf{X}^T + t\mathbf{I}_{p_1} - t\mathbf{I}_{p_1}) = n\mathbf{I}_{p_1} - nt\mathbf{H}^{-1}$.

Similarly we can obtain

$$||\mathbf{X}^T\mathbf{H}^{-1}|| \leq M\sqrt{n}. \tag{2.38}$$

We conclude from (2.35)-(2.38) that

$$||\frac{1}{n}\frac{\partial^3 tr\mathbf{B}^{-1}(z)}{\partial X_{ij}^3}|| \leq \frac{M}{n^{5/2}}. \tag{2.39}$$

This implies that

$$E|X_{ij}^3 \cdot \frac{1}{n}\frac{\partial^3 tr\mathbf{B}^{-1}(z)}{\partial X_{ij}^3}| \leq \frac{M}{n^{5/2}}E[X_{ij}^3] \leq \frac{M\varepsilon_n}{n^2}. \tag{2.40}$$

Since all X_{ij} and W_{ij} play a similar role in their corresponding matrices, the above argument works for all matrices. Hence we obtain

$$\begin{aligned}
& |E[\frac{1}{n}tr\mathbf{B}^{-1}(z)] - E[\frac{1}{n}tr\mathbf{D}^{-1}(z)]| \\
\leq\ & M\sum_{j=1}^{p_1 n}[\int_0^1 (1-\tau)^2 E|\hat{X}_j^3\partial_j^3 f(\mathbf{Z}_j^{(1)}(\tau))|d\tau + \int_0^1 (1-\tau)^2 E|\hat{G}_j^3\partial_j^3 f(\mathbf{Z}_{j-1}^{(2)}(\tau))|d\tau] \\
\leq\ & M\varepsilon_n.
\end{aligned}$$

This ensures that

$$E[\frac{1}{n}tr\mathbf{B}^{-1}(z)] - E[\frac{1}{n}tr\mathbf{D}^{-1}(z)] \to 0 \quad as \quad n \to \infty.$$

Therefore the proof of Theorem 1 is completed.

2.3 Conclusion

Canonical correlation coefficients play an important role in the analysis of correlations between random vectors[Anderson (1984)]. Nowadays, investigations of large dimensional random vectors attract a substantial research works, e.g. Fan and Lv (2010). As future works, we plan to develop central limit theorems for the empirical distribution of canonical correlation coefficients and make statistical applications of the developed asymptotic theorems for large dimensional random vectors.

2.4 Appendix

Lemma 1 (Burkholder (1973)). *Let $\{X_k, 1 \le k \le n\}$ be a complex martingale difference sequence with respect to the increasing σ-field $\{\mathcal{F}_k\}$. Then, for $p \ge 2$,*

$$E|\sum_{k=1}^{n} X_k|^p \le K_p(E(\sum_{k=1}^{n} E(|X_k|^2|\mathcal{F}_{k-1}))^{p/2} + E\sum_{k=1}^{n} |X_k|^p).$$

Lemma 2 (Burkholder (1973)). *With $\{X_k, 1 \le k \le n\}$ as above, we have, for $p > 1$,*

$$E|\sum_{k=1}^{n} X_k|^p \le K_p E(\sum_{k=1}^{n} |X_k|^2)^{p/2}.$$

Lemma 3 (Lemma B.26 of Bai and Silverstein (2009)). *For $\mathbf{X} = (X_1, \cdots, X_n)^T$ i.i.d standardized entries, $\mathbf{C}$ $n \times n$ matrix, we have, for any $p \ge 2$,*

$$E|\mathbf{X}^*\mathbf{C}\mathbf{X} - tr\mathbf{C}|^p \le K_p((E|X_1|^4 tr\mathbf{C}\mathbf{C}^*)^{p/2} + E|X_1|^{2p} tr(\mathbf{C}\mathbf{C}^*)^{p/2}).$$

Lemma 4 (Theorem A.43 of Bai and Silverstein (2009)). *Let $\mathbf{A}$ and $\mathbf{B}$ be two $n \times n$ symmetric matrices. Then*

$$||F^{\mathbf{A}} - F^{\mathbf{B}}|| \le \frac{1}{n} rank(\mathbf{A} - \mathbf{B}),$$

where $||f|| = sup_x |f(x)|$.

Lemma 5 (Hoeffding (1963)). *Let $Y_1, Y_2, \ldots$ be i.i.d random variables, $P(Y_1 = 1) = q = 1 - P(Y_1 = 0)$. Then*

$$P(|Y_1 + \cdots + Y_n - nq| \ge n\varepsilon) \le 2e^{-\frac{n^2\varepsilon^2}{2nq+n\varepsilon}}$$

for all $\varepsilon > 0$, $n = 1, 2, \ldots$.

Lemma 6 (Corollary A.41 of Bai and Silverstein (2009)). *Let $\mathbf{A}$ and $\mathbf{B}$ be two $n \times n$ symmetric matrices with their respective ESDs of $F^{\boldsymbol{A}}$ and $F^{\boldsymbol{B}}$. Then,*

$$L^3(F^{\boldsymbol{A}}, F^{\boldsymbol{B}}) \le \frac{1}{n} tr(\boldsymbol{A} - \boldsymbol{B})^2.$$

Chapter 3

Regularized Canonical Correlation Coefficients

3.1 Methodology and Theory

Throughout this chapter we make the following assumptions.

Assumption 1. $p_1 = p_1(n)$ and $p_2 = p_2(n)$ with $\frac{p_1}{n} \to c_1$ and $\frac{p_2}{n} \to c_2$, $c_1, c_2 \in (0, 1)$, as $n \to \infty$.

Assumption 2. $p_1 = p_1(n)$ and $p_2 = p_2(n)$ with $\frac{p_1}{n} \to c_1^{'}$ and $\frac{p_2}{n} \to c_2^{'}$, $c_1^{'} \in (0, +\infty)$ and $c_2^{'} \in (0, +\infty)$, as $n \to \infty$.

Assumption 3. $\mathbf{X} = (X_{ij})_{i,j=1}^{p_1,n}$ and $\mathbf{Y} = (Y_{ij})_{i,j=1}^{p_2,n})$ satisfy $\mathbf{X} = \mathbf{\Sigma}_{\mathbf{xx}}^{1/2}\mathbf{W}$ and $\mathbf{Y} = \mathbf{\Sigma}_{\mathbf{yy}}^{1/2}\mathbf{V}$, where $\mathbf{W} = (\mathbf{w}_1, \cdots, \mathbf{w}_n) = (W_{ij})_{i,j=1}^{p_1,n}$ consists of i.i.d real random variables $\{W_{ij}\}$ with $EW_{11} = 0$ and $E|W_{11}|^2 = 1$; $\mathbf{V} = (\mathbf{v}_1, \cdots, \mathbf{v}_n) = (V_{ij})_{i,j=1}^{p_1,n}$ consists of i.i.d real random variables with $EV_{11} = 0$ and $E|V_{11}|^2 = 1$; $\mathbf{\Sigma}_{\mathbf{xx}}^{1/2}$ and $\mathbf{\Sigma}_{\mathbf{yy}}^{1/2}$ are Hermitian square roots of positive definite matrices $\mathbf{\Sigma}_{\mathbf{xx}}$ and $\mathbf{\Sigma}_{\mathbf{yy}}$ respectively so that $(\mathbf{\Sigma}_{\mathbf{xx}}^{1/2})^2 = \mathbf{\Sigma}_{\mathbf{xx}}$ and $(\mathbf{\Sigma}_{\mathbf{yy}}^{1/2})^2 = \mathbf{\Sigma}_{\mathbf{yy}}$.

Assumption 4. $F^{\mathbf{\Sigma_{xx}}} \xrightarrow{D} H$, a proper cumulative distribution function.

Remark 2. *By the definition of the matrix* $\mathbf{S}_{xy}$*, the classical canonical correlation coefficients between* $\mathbf{x}$ *and* $\mathbf{y}$ *are the same as those between* $\mathbf{w}$ *and* $\mathbf{v}$ *when* $\mathbf{w}$ *and* $\{\mathbf{w}_i\}$ *are i.i.d, and* $\mathbf{v}$ *and* $\{\mathbf{v}_i\}$ *are i.i.d.*

We now introduce some results from random matrix theory. Denote the empirical spectral distribution function (ESD) of any $n \times n$ matrix $\mathbf{A}$ with real eigenvalues $\mu_1 \leq \mu_2 \leq \cdots \leq \mu_n$ by

$$F^{\mathbf{A}}(x) = \frac{1}{n}\#\{i : \mu_i \leq x\}, \tag{3.1}$$

where $\#\{\cdots\}$ denotes the cardinality of the set $\{\cdots\}$.

When the two random vectors $\mathbf{x}$ and $\mathbf{y}$ are independent and each of them consists of i.i.d Gaussian random variables, under Assumptions 1 and 3, Wachter (1980) proved that the empirical measure of the classical sample canonical correlation coefficients $r_1, r_2, \cdots, r_{p_1}$ converges in probability to a fixed distribution whose density is given by

$$\rho(x) = \frac{\sqrt{(x-L)(x+L)(H-x)(H+x)}}{\pi c_1 x(1-x)(1+x)}, \ x \in [L, H], \tag{3.2}$$

and atoms size of $\max(0, (1-c_2)/c_1)$ at zero and size $\max(0, 1-(1-c_2)/c_1)$ at unity where $L = |\sqrt{c_2 - c_2c_1} - \sqrt{c_1 - c_1c_2}|$ and $H = |\sqrt{c_2 - c_2c_1} + \sqrt{c_1 - c_1c_2}|$. Here the empirical measure of $r_1, r_2, \cdots, r_{p_1}$ is defined as in (3.1) with μ_i replaced by r_i.

In Chapter 2, we have proved that (3.2) also holds for classical sample canonical correlation coefficients when the entries of $\mathbf{x}$ and $\mathbf{y}$ are not necessarily Gaussian distributed. For easy reference, we state the result in the following proposition.

Proposition 1. *In addition to Assumptions 1 and 3, suppose that* $\{X_{ij}, 1 \le i \le p_1, 1 \le j \le n\}$ *and* $\{Y_{ij}, 1 \le i \le p_2, 1 \le j \le n\}$ *are independent. Then the empirical measure of* $r_1, r_2, \ldots, r_{p_1}$ *converges almost surely to a fixed distribution function whose density is given by (3.2).*

Under Assumptions 2-4, instead of $F^{\mathbf{S}_{xy}}$, we analyze the ESD, $F^{\mathbf{T_{xy}}}$, of the regularized random matrix $\mathbf{T_{xy}}$ given in (1.6). To this end, define the Stieltjes transform of any distribution function $G(x)$ by

$$m_G = \int \frac{1}{x-z} dG(x),\ z \in \mathbb{C}^+ = \{z \in \mathbb{C}, \mathrm{I}mz > 0\},$$

where $\mathrm{I}mz$ denotes the imaginary part of the complex number z.

It turns out that the limit of the empirical spectral distribution (LSD) of $\mathbf{T_{xy}}$ is connected to that of the LSD of $\mathbf{S}_1\mathbf{S}_{2t}^{-1}$ defined below. Let

$$\mathbf{S}_1 = \frac{1}{p_2}\sum_{k=1}^{p_2} \mathbf{w}_k\mathbf{w}_k^T, \quad \mathbf{S}_{2t} = \frac{1}{n-p_2}\sum_{k=p_2+1}^{n} \mathbf{w}_k\mathbf{w}_k^T + t\frac{n}{n-p_2}\mathbf{\Sigma}_{\mathbf{xx}}^{-1},$$

$$y_1 = \frac{c_1^{'}}{c_2^{'}}, \quad y_2 = \frac{c_1^{'}}{1-c_2^{'}}.$$

The LSD of $\mathbf{S}_{2t}$ and its Stieltjes transform are denoted by F_{y_2t} and $m_{y_2t}(z)$ respectively. Under Assumptions 2- 4 from Silverstein and Bai (1995) and Pan (2010) $m_{y_2t}(z)$ is the unique solution in $\mathbb{C}^+$ to

$$m_{y_2t}(z) = m_{H_t}\Big(z - \frac{1}{1+y_2 m_{y_2t}(z)}\Big), \tag{3.3}$$

where $m_{H_t}(z)$ denotes the Stieltjes transform of the LSD of the matrix $t\frac{n}{n-p_2}\mathbf{\Sigma}_{\mathbf{xx}}^{-1}$ (one may also see (4.4) in Chapter 4). Let $\mathbf{n} = (n_1, n_2)$ and $\mathbf{y} = (y_1, y_2)$ with $n_1 = p_1$ and $n_2 = n - p_2$. The Stieltjes transforms of the ESD and LSD of the matrix $\mathbf{S}_1\mathbf{S}_{2t}^{-1}$ are denoted by $m_{\mathbf{n}}(z)$ and $m_{\mathbf{y}}(z)$ respectively while those of the ESD and LSD of the matrix $\frac{1}{p_2}\sum_{k=1}^{p_2}\mathbf{w}_k^T\mathbf{S}_{2t}^{-1}\mathbf{w}_k$ are denoted by $\underline{m}_{\mathbf{n}}(z)$ and $\underline{m}_{\mathbf{y}}(z)$ respectively. Observe that the spectra of

$\mathbf{S}_1\mathbf{S}_{2t}^{-1}$ and $\frac{1}{p_2}\sum_{k=1}^{p_2}\mathbf{w}_k^T\mathbf{S}_{2t}^{-1}\mathbf{w}_k$ are the same except (p_1-p_2) zero eigenvalues and this leads to

$$\underline{m}_{\mathbf{y}}(z) = -\frac{1-y_1}{z} + y_1 m_{\mathbf{y}}(z). \tag{3.4}$$

We are now in a position to state the LSD of $\mathbf{T}_{\mathbf{xy}}$.

Theorem 2. *In addition to Assumptions 2-4, suppose that* $\{X_{ij}, 1 \le i \le p_1, 1 \le j \le n\}$ *and* $\{Y_{ij}, 1 \le i \le p_2, 1 \le j \le n\}$ *are independent.*

a) If $c_2^{'} \in (0,1)$*, then the ESD,* $F^{\mathbf{T}_{\mathbf{xy}}}(\lambda)$*, converges almost surely to a fixed distribution* $\tilde{F}(\frac{\lambda}{q(1-\lambda)})$*, where* $q = \frac{c_2}{1-c_2}$ *and* $\tilde{F}(\lambda)$ *is a nonrandom distribution and its Stieltjes transform* $m_{\mathbf{y}}(z)$ *is the unique solution in* $\mathbb{C}^+$ *to*

$$m_{\mathbf{y}}(z) = -\int \frac{dF_{y_2 t}(1/\lambda)}{\lambda(1-y_1'-y_1' z m_{\mathbf{y}}(z)) - z}. \tag{3.5}$$

b) If $c_2^{'} \in [1,\infty)$*, then* $F^{\mathbf{T}_{\mathbf{xy}}}(\lambda)$*, converges almost surely to a fixed distribution* $\tilde{G}(\frac{t}{1-x} - t)$ *where* $\tilde{G}(\lambda)$ *is a nonrandom distribution and its Stieltjes transform satisfies the equation*

$$m_{\tilde{G}}(z) = \int \frac{dH(\lambda)}{\lambda(1-c_1'-c_1' z m_{\tilde{G}}(z)) - z}. \tag{3.6}$$

Remark 3. *Indeed, taking* $t = 0$ *in (3.5) recovers Wachter (1980)'s result (one may refer to the result of F matrix in Bai and Silverstein (2009)).*

Let us now introduce the test statistic. Under Assumption 1 and Assumption 3, behind our test statistic is the observation that the limit of $F^{\mathbf{S}_{xy}}(x)$ can be obtained from (3.2) when $\mathbf{x}$ and $\mathbf{y}$ are independent, while the limit of $F^{\mathbf{S}_{xy}}(x)$ could be different from (3.2) when $\mathbf{x}$ and $\mathbf{y}$ have correlation. For example, if $\mathbf{y} = \Sigma_1\mathbf{w}$ and $\mathbf{x} = \Sigma_2\mathbf{w}$ with $p_1 = p_2$ and both Σ_1 and Σ_2 being invertible, then

$$\mathbf{S}_{xy} = \mathbf{I},$$

which implies that the limit of $F^{\mathbf{S}_{xy}}(x)$ is a degenerate distribution. This suggests that we may make use of $F^{\mathbf{S}_{xy}}(x)$ to construct a test statistic. Thus we consider the following statistic

$$\int \phi(x) dF^{\mathbf{S}_{xy}}(x) = \frac{1}{p_1} \sum_{i=1}^{p_1} \phi(r_i^2). \tag{3.7}$$

A perplexing problem is how to choose an appropriate function $\phi(x)$. For simplicity we choose $\phi(x) = x$ in this work. That is, our statistic is

$$S_n = \int x dF^{\mathbf{S}_{xy}}(x) = \frac{1}{p_1} \sum_{i=1}^{p_1} r_i^2. \tag{3.8}$$

Indeed, extensive simulations based on Theorems 3 and 4 below have been conducted to help select an appropriate function $\phi(x)$. We find that other functions such as $\phi(x) = x^2$ does not have an advantage over $\phi(x) = x$.

In the classical CCA, the maximum likelihood ratio test statistic for (1.1) with fixed dimensions is

$$MLR_n = \sum_{i=1}^{p_1} \log(1 - r_i^2) \tag{3.9}$$

(see Wilks (1935) and Aderson (1984)). That is, $\phi(x)$ in (3.7) takes $\log(1 - x)$. Note that the density $\rho(x)$ has atom size of $\max(0, 1 - (1 - c_2)/c_1)$ at unity by (3.2). Thus the normalized statistic MLR_n is not well defined when $c_1 + c_2 > 1$ (because $\int \log(1 - x^2)\rho(x)dx$ is not meaningful). In addition, even when $c_1 + c_2 \leq 1$, the right end point of $\rho(x)$, H, can be equal to one so that some sample correlation coefficients r_i are close to one. For example $H = 1$ when $c_1 = c_2 = 1$. This in turns causes a big value of the corresponding $\log(1 - r_i^2)$. Therefore, MLR_n is not stable and this phenomenon is also confirmed by our simulations.

Under Assumptions 2 and 3, we substitute $\mathbf{T_{xy}}$ for $\mathbf{S}_{xy}$ and use the statistic

$$T_n = \int x dF^{\mathbf{T_{xy}}}(x). \tag{3.10}$$

We next establish the CLTs of the statistics (3.7) and (3.10). To this end, write

$$G^{(1)}_{p_1,p_2}(\lambda) = p_1\big(F^{\mathbf{S}_{xy}}(\lambda) - F^{c_{1n},c_{2n}}(\lambda)\big), \tag{3.11}$$

and

$$G^{(2)}_{p_1,p_2}(\lambda) = p_1\big(F^{\mathbf{T}_{xy}}(\lambda) - F^{c'_{1n},c'_{2n}}(\lambda)\big), \tag{3.12}$$

where $F^{c_{1n},c_{2n}}(\lambda)$ and $F^{c'_{1n},c'_{2n}}(\lambda)$ are obtained from $F^{c_1,c_2}(\lambda)$ and $F^{c'_1,c'_2}$ with c_1, c_2 and c'_1, c'_2 replaced by $c_{1n} = \frac{p_1}{n}, c_{2n} = \frac{p_2}{n}$ and $c'_{1n} = \frac{p_1}{n}, c'_{2n} = \frac{p_2}{n}$ respectively; $F^{c_1,c_2}(\lambda)$ and $F^{c'_1,c'_2}(\lambda)$ are the limiting spectral distributions of the matrices $\mathbf{S}_{xy}$ and $\mathbf{T_{xy}}$ respectively, whose densities can be obtained from $\rho(x)$ in (3.2) and (3.5). We re-normalize (3.7) and (3.10) as

$$\int \phi(\lambda) dG^{(1)}_{p_1,p_2}(\lambda) := p_1\Big(\int \phi(\lambda) dF^{\mathbf{S}_{xy}}(\lambda) - \int \phi(\lambda) dF^{c_{1n},c_{2n}}(\lambda)\Big), \tag{3.13}$$

and

$$\int \phi(\lambda) dG^{(2)}_{p_1,p_2}(\lambda) := p_1\Big(\int \phi(\lambda) dF^{\mathbf{T}_{xy}}(\lambda) - \int \phi(\lambda) dF^{c'_{1n},c'_{2n}}(\lambda)\Big). \tag{3.14}$$

Also, let

$$\bar{y}_1 := \frac{c_1}{1-c_2} \in (0, +\infty),\ \bar{y}_2 := \frac{c_1}{c_2} \in (0,1),\ h = \sqrt{\bar{y}_1 + \bar{y}_2 - \bar{y}_1\bar{y}_2},\ a_1 = \frac{(1-h)^2}{(1-\bar{y}_2)^2},$$

$$a_2 = \frac{(1+h)^2}{(1-\bar{y}_2)^2},\ g_{\bar{y}_1,\bar{y}_2}(\lambda) = \frac{1-\bar{y}_2}{2\pi\lambda(\bar{y}_1 + \bar{y}_2\lambda)}\sqrt{(a_2-\lambda)(\lambda - a_1)},\ a_1 < \lambda < a_2. \tag{3.15}$$

Theorem 3. *Let $\phi_1, \cdots, \phi_s$ be functions analytic in an open region in the complex plane containing the interval $[a_1, a_2]$. In addition to Assumptions 1 and 3, suppose that*

$$EX^4_{11} = EY^4_{11} = 3. \tag{3.16}$$

Then, as $n \to \infty$, the random vector

$$\Big(\int \phi_1(\lambda) dG^{(1)}_{p_1,p_2}(\lambda), \ldots, \int \phi_s(\lambda) dG^{(1)}_{p_1,p_2}(\lambda) \Big) \tag{3.17}$$

converges weakly to a Gaussian vector $(X_{\phi_1}, \ldots, X_{\phi_s})$ with mean

$$EX_{\phi_i} = \lim_{r \downarrow 1} \frac{1}{4\pi i} \oint_{|\xi|=1} f_i\Big(\frac{1 + h^2 + 2h\Re(\xi)}{(1-\bar{y}_2)^2} \Big) \Big[\frac{1}{\xi - r^{-1}} + \frac{1}{\xi + r^{-1}} - \frac{2}{\xi + \frac{\bar{y}_2}{h}} \Big] d\xi, \tag{3.18}$$

and covariance function

$$cov(X_{\phi_i}, X_{\phi_j}) = -\lim_{r \downarrow 1} \frac{1}{4\pi^2} \oint_{|\xi_1|=1} \oint_{|\xi_2|=1} \frac{f_i\big(\frac{1+h^2+2h\Re(\xi_1)}{(1-\bar{y}_2)^2}\big) f_j\big(\frac{1+h^2+2h\Re(\xi_2)}{(1-\bar{y}_2)^2}\big)}{(\xi_1 - r\xi_2)^2} d\xi_1 d\xi_2, \tag{3.19}$$

where $f_i(\lambda) = \phi_i(\frac{1}{1+(\frac{1-c_2}{c_2})\lambda})$; $\Re$ denotes the real part of a complex number; and $r \downarrow 1$ means that r approaches to 1 from above.

Remark 4. *When $\phi(x) = x$, the mean of the limit distribution in Theorem 3 is 0 and the variance is $\frac{2h^2 y_1^2 y_2^2}{(y_1+y_2)^4}$. These are calculated in Example 4.2 of Zheng (2012).*

Before stating the CLT of the linear spectral statistics for the matrix $\mathbf{T}_{\mathbf{xy}}$, we make some notation. Let r be a positive integer and introduce

$$m_r(z) = \int \frac{dH(x)}{(x - z + \varpi(z))^r}, \quad \varpi(z) = \frac{1}{1 + y_2 m_{y_2 t}(z)},$$
$$g(z) = \frac{y_2 (m_{y_2 t}(-\underline{m}_{\mathbf{y}}(z)))'}{(1 + y_2 m_{y_2 t}(-\underline{m}_{\mathbf{y}}(z)))^2}, \ h(z) = \frac{-\underline{m}^2_{\mathbf{y}}(z)}{1 - y_1 \underline{m}^2_{\mathbf{y}}(z) \int \frac{dF_{y_2 t}(x)}{(x + \underline{m}_{\mathbf{y}}(z))^2}},$$
$$s(z_1, z_2) = \frac{1}{1 + y_2 m_{y_2 t}(z_1)} - \frac{1}{1 + y_2 m_{y_2 t}(z_2)},$$

where $(m_{y_2 t}(z))'$ stands for the derivative with respect to z.

Theorem 4. *Let $\phi_1, \cdots, \phi_s$ be functions analytic in an open region in the complex plane containing the support of the LSD $\tilde{F}(\lambda)$ whose stieltjes transform is (3.5). In addition to Assumptions* 2-4*, suppose that*

$$EX_{11}^4 = EY_{11}^4 = 3. \tag{3.20}$$

a) If $c_2^{'} \in (0,1)$, then the random vector

$$\Big(\int \phi_1(\lambda) dG^{(2)}_{p_1,p_2}(\lambda), \ldots, \int \phi_s(\lambda) dG^{(2)}_{p_1,p_2}(\lambda) \Big) \tag{3.21}$$

converges weakly to a Gaussian vector $(X_{\phi_1}, \ldots, X_{\phi_s})$ with mean

$$\begin{aligned} EX_{\phi_i} &= -\frac{1}{2\pi i} \oint_{\mathcal{C}} \phi_i\big(\frac{z}{1+z}\big)\Big(\frac{y_1 \int \underline{m}_{\mathbf{y}}(z)^3 x[x+\underline{m}_{\mathbf{y}}(z)]^{-3} dF_{y_2 t}(x)}{[1 - y_1 \int \underline{m}_{\mathbf{y}}(z)^2 (x+\underline{m}_{\mathbf{y}}(z))^{-2} dF_{y_2 t}(x)]^2} \\ &+h(z) \frac{y_2 \varpi^2(-\underline{m}_{\mathbf{y}}(z)) m_3(-\underline{m}_{\mathbf{y}}(z)) + y_2^2 \varpi^4(-\underline{m}_{\mathbf{y}}(z)) m^{'}_{y_2 t}(-\underline{m}(z)) m_3(-\underline{m}_{\mathbf{y}}(z))}{1 - y_2 \varpi^2(-\underline{m}_{\mathbf{y}}(z)) m_2(-\underline{m}_{\mathbf{y}}(z))} \\ &-h(z) \frac{y_2^2 \varpi^3(-\underline{m}_{\mathbf{y}}(z)) m^{'}_{y_2 t}(-\underline{m}(z)) m_2(-\underline{m}_{\mathbf{y}}(z))}{1 - c\varpi^2(-\underline{m}_{\mathbf{y}}(z)) m_2(-\underline{m}_{\mathbf{y}}(z))} \Big) dz \end{aligned} \tag{3.22}$$

and covariance

$$\begin{aligned} Cov(X_{\phi_i}, X_{\phi_j}) &= -\frac{1}{2\pi^2} \oint_{\mathcal{C}} \oint_{\mathcal{C}} \phi_i\big(\frac{z_1}{1+z_1}\big) \phi_j\big(\frac{z_2}{1+z_2}\big) \Big(\frac{\underline{m}^{'}_{\mathbf{y}}(z_1) \underline{m}^{'}_{\mathbf{y}}(z_2)}{(\underline{m}_{\mathbf{y}}(z_1) - \underline{m}_{\mathbf{y}}(z_2))^2} \\ &- \frac{1}{(z_1 - z_2)^2} - \frac{h(z_1)h(z_2)}{(-\underline{m}_{\mathbf{y}}(z_2) + \underline{m}_{\mathbf{y}}(z_1))^2} \\ &+ \frac{h(z_1)h(z_2)[1 + g(z_1) + g(z_2) + g(z_1)g(z_2)]}{[-\underline{m}_{\mathbf{y}}(z_2) + \underline{m}_{\mathbf{y}}(z_1) + s(-\underline{m}_{\mathbf{y}}(z_1), -\underline{m}_{\mathbf{y}}(z_2))]^2} \Big) dz_1 dz_2. \end{aligned} \tag{3.23}$$

Here all the contour integrals can be evaluated on any contour enclosing the support of the LSD $\tilde{F}(\lambda)$ whose stieltjes transform is (3.5).

b) If $c_2^{'} \in [1, +\infty)$, (3.21) converges weakly to a Gaussian vector $(X_{\phi_1}, \ldots, X_{\phi_s})$ with mean and covariance illustrated in (9.7.5) and (9.7.6) of Bai and Silverstein (2009), where the parameter y equals $c_1^{'}$, σ^2 equals t^{-1} and $\mathbf{T}_n = \mathbf{\Sigma}_{\mathbf{xx}}$.

Here we would like to point out that the idea of testing independence between two random vectors $\mathbf{x}$ and $\mathbf{y}$ by CCA is based on the fact that the uncorrelatedness between $\mathbf{x}$ and $\mathbf{y}$ is equivalent to independence between them when the random vector of size (p_1+p_2) consisting of the components of $\mathbf{x}$ and $\mathbf{y}$ is a Gaussian random vector. See Wilks (1935) and Anderson (1984). For nonGaussian random vectors $\mathbf{x}$ and $\mathbf{y}$, uncorrelatedness is not equivalent to independence. CCA may fail in this case. Yet, since Theorems 3 and 4 hold for nonGaussian random vectors $\mathbf{x}$ and $\mathbf{y}$ CCA can be still utilized to capture dependent but uncorrelated $\mathbf{x}$ and $\mathbf{y}$ such as ARCH type of dependence by considering higher power of their entries. See Section 3.4.6 for the further discussion.

Following Lytova and Pastur (2009) condition (3.20) can be removed. However it will significantly increase the length of this work and we will not pursue it here.

3.2 The power under local alternatives

This section is to evaluate the power of S_n or T_n under a kind of local alternative. Recall the definitions of $G^{(i)}_{p_1,p_2}, i = 1, 2$ in (3.11) and (3.12) and let $R^{(i)}_n = \int \lambda dG^{(i)}_{p_1,p_2}$.

Theorem 5. *In addition to assumptions in Theorem 3 or Theorem 4, for any $i = 1, 2$, suppose that in probability*

$$\left| tr\left(\mathbf{S}^{\mathbb{H}_1}_{xy} - \mathbf{S}^{\mathbb{H}_0}_{xy}\right)\right| \to \infty, \quad \left| tr\left(\mathbf{T}^{\mathbb{H}_1}_{xy} - \mathbf{T}^{\mathbb{H}_0}_{xy}\right)\right| \to \infty, \tag{3.24}$$

where $\mathbf{S}^{\mathbb{H}_j}_{xy}$ is $\mathbf{S}_{xy}$ under $\mathbb{H}_j$, with $j = 0, 1$. $\mathbf{T}^{\mathbb{H}_j}_{xy}, j = 0, 1$ are defined similarly. Then

$$\lim_{n\to\infty} P(R^{(i)}_n > z^{(i)}_{1-\alpha} \text{ or } R^{(i)}_n < z^{(i)}_{\alpha}|\mathbb{H}_1) = 1, \tag{3.25}$$

where $z_{1-\alpha}^{(i)}$ *and* $z_{\alpha}^{(i)}$ *are* $(1-\alpha)$ *and* α *quantiles of the asymptotic distribution of the statistic* $R_n^{(i)}$ *under the null hypothesis.*

Remark 5. *If* $\mathbf{S}_{xy}^{\mathbb{H}_1} = (\mathbf{X}\mathbf{X}^T)^{-1}\mathbf{X}\mathbf{T}\mathbf{P}_y\mathbf{X}^T$ *and* $\mathbf{S}_{xy}^{\mathbb{H}_0} = (\mathbf{X}\mathbf{X}^T)^{-1}\mathbf{X}\mathbf{P}_y\mathbf{X}^T$ *with*$\mathbf{T}$ *being a nonnegative definite matrix, then this implies the covariance matrix between* $\mathbf{x}$ *and* $\mathbf{y}$ *is* $\mathbf{T}$. *Particularly, if* $\mathbf{T} = \mathbf{I} + \mathbf{e}\mathbf{e}^T$ *with* $\mathbf{e} = (1, 1, \cdots, 1)$ *then under assumptions in Theorem 3 or Theorem 4 it can be proved that*

$$tr\left(\mathbf{S}_{xy}^{\mathbb{H}_1} - \mathbf{S}_{xy}^{\mathbb{H}_0}\right) = \mathbf{e}^T\mathbf{P}_y\mathbf{P}_x\mathbf{e} = n\frac{\mathbf{e}^T\mathbf{P}_y\mathbf{P}_x\mathbf{e}}{\|\mathbf{e}\|} = O_p(n)$$

satisfying (3.24).

3.3 Applications of CCA

This section explores some applications of the proposed test. We consider two examples from multivariate analysis and time series analysis respectively.

3.3.1 Multivariate Regression test with CCA

Consider the multivariate regression(MR) model as follows:

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{E}, \tag{3.26}$$

where

$$\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{p_1}]_{n\times p_1},\ \mathbf{X} = [\mathbf{1}_n, \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{p_2}]_{n\times p_2},$$

$$\mathbf{B} = [\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_{p_1}]_{p_2\times p_1},\ \mathbf{E} = [\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_{p_1}]_{n\times p_1},$$

and each of the vectors $\mathbf{y}_j$, $\mathbf{x}_j$, $\mathbf{e}_j$, for $j = 1, 2, \ldots, p_1$ is $n \times 1$ vectors and $\{\boldsymbol{\beta}_i,\ i = 1, 2, \ldots, p_1\}$ are $p_2 \times 1$ vectors.

Let $\mathbf{A_{xy}} = \frac{1}{n}\mathbf{X}'\mathbf{Y}$ and $\mathbf{A_{xx}} = \frac{1}{n}\mathbf{X}'\mathbf{X}$. We have the least square estimate of $\mathbf{B}$

$$\hat{\mathbf{B}} = \mathbf{A}_{\mathbf{xx}}^{-1}\mathbf{A}_{\mathbf{xy}}. \tag{3.27}$$

The most common hypothesis testing is to test whether there exists linear relationship between the two sets of variables (response variables and predictor variables) or the overall regression test

$$\mathbb{H}_0 : \mathbf{B} = \mathbf{0}. \tag{3.28}$$

To test $\mathbb{H}_0 : \mathbf{B} = \mathbf{0}$, Wilks' Λ criterion is

$$\Lambda = \frac{|\mathbf{E}|}{|\mathbf{E} + \mathbf{H}|} = \prod_{i=1}^{s}(1 + \lambda_i)^{-1}, \tag{3.29}$$

where

$$\mathbf{E} = \mathbf{Y}'\big(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\big)\mathbf{Y} \tag{3.30}$$

and

$$\mathbf{H} = \hat{\mathbf{B}}'(\mathbf{X}'\mathbf{X})\hat{\mathbf{B}}; \tag{3.31}$$

and $\{\lambda_i : i = 1, \ldots, s\}$ are the roots of $|\mathbf{H} - \lambda\mathbf{E}| = 0$, $s = min(k, p)$.

Remark 6. *Wilk's Λ criterion indeed assumes Gaussian observations and derive the exact lambda distribution. In this thesis, we derive the asymptotic distribution of Wilk's criterion under high dimensional case without Gaussian assumption.*

An alternative form for Λ is to employ sample covariance matrices. That is, $\mathbf{H} = \mathbf{A_{yx}}\mathbf{A_{xx}^{-1}}\mathbf{A_{xy}}$ and $\mathbf{E} = \mathbf{A_{yy}} - \mathbf{A_{yx}}\mathbf{A_{xx}^{-1}}\mathbf{A_{xy}}$, so that $|\mathbf{H} - \lambda\mathbf{E}| = 0$ becomes $|\mathbf{A_{yx}}\mathbf{A_{xx}^{-1}}\mathbf{A_{xy}} - \lambda(\mathbf{A_{yy}} - \mathbf{A_{yx}}\mathbf{A_{xx}^{-1}}\mathbf{A_{xy}})| = 0$. From Theorem 2.6.8 of Timm (2001) we have $|\mathbf{H} - \theta(\mathbf{H} + \mathbf{E})| = |\mathbf{A_{yx}}\mathbf{A_{xx}^{-1}}\mathbf{A_{xy}} - \theta\mathbf{A_{yy}}| = 0$ so that

$$\Lambda = \prod_{i=1}^{s}(1+\lambda_i)^{-1} = \prod_{i=1}^{s}(1-\theta_i) = \frac{|\mathbf{A_{yy}} - \mathbf{A_{yx}}\mathbf{A_{xx}^{-1}}\mathbf{A_{xy}}|}{|\mathbf{A_{yy}}|}. \tag{3.32}$$

Evidently, the quantities $r_i^2 = \theta_i, i = 1, \ldots, s$ are sample canonical correlation coefficients. Therefore the test statistic (3.29) can be rewritten as

$$log\Lambda = \sum_{i=1}^{s} log(1 - r_i^2). \tag{3.33}$$

From this point of view, the multiple regression test is equivalent to the independence test based on canonical correlation coefficients. As stated in the last section, the statistic $log\Lambda$ is not stable in the high dimensional cases. Hence our test statistic S_n or T_n can be applied in the MR test.

3.3.2 Testing for Cointegration with CCA

Consider an n-dimensional vector process $\{\mathbf{y}_t\}$ that has a first-order error correction representation

$$\Delta\mathbf{y}_t = -\boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t, \ t = 1, \ldots, T, \tag{3.34}$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are full rank $n \times r$ matrices $(r < n)$ and the n-dimensional innovation $\{\varepsilon_t\}$ is i.i.d. with zero mean and positive covariance matrix $\boldsymbol{\Omega}$. Select $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ so that the fact that $|\mathbf{I}_n - (\mathbf{I}_n - \boldsymbol{\alpha}\boldsymbol{\beta}')z| = 0$ implies that either $|z| > 1$ or $z = 1$ and that $\boldsymbol{\alpha}_\perp'\boldsymbol{\beta}_\perp$ is of full rank, where $\boldsymbol{\alpha}_\perp$ and $\boldsymbol{\beta}_\perp$ are full rank $n \times (n - r)$ matrices orthogonal to $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. Under these assumptions, $\{\mathbf{y}_t\}$ is $I(1)$ with r cointegration relations among its elements; that is $\{\boldsymbol{\beta}'\mathbf{y}_t\}$ is $I(0)$.

The goal is to test

$$\mathbb{H}_0 : r = 0 \ (\boldsymbol{\alpha} = \boldsymbol{\beta} = \mathbf{0}); \ \textit{against} \ \mathbf{H}_1 : r > 0; \tag{3.35}$$

i.e. whether there exists cointegration relationships among the elements of the time series $\{\mathbf{y}_t\}$.

This cointegration test is equivalent to testing

$$\begin{aligned} &\mathbb{H}_0 : \Delta\mathbf{y}_t \ \textit{is independent with} \ \Delta\mathbf{y}_{t-1}; \\ &\textit{against} \\ &\mathbb{H}_1 : \Delta\mathbf{y}_t \ \textit{is dependent with} \ \Delta\mathbf{y}_{t-1}. \end{aligned} \tag{3.36}$$

In order to apply canonical correlation coefficients to cointegration test (3.35), we construct random matrices

$$\mathbf{X} = \big(\Delta\mathbf{y}_2, \Delta\mathbf{y}_4, \ldots, \Delta\mathbf{y}_{2t-2}, \Delta\mathbf{y}_{2t}, \ldots, \Delta\mathbf{y}_T\big), \tag{3.37}$$

$$\mathbf{Y} = \big(\Delta\mathbf{y}_1, \Delta\mathbf{y}_3, \ldots, \Delta\mathbf{y}_{2t-1}, \Delta\mathbf{y}_{2t+1}, \ldots, \Delta\mathbf{y}_{T-1}\big). \tag{3.38}$$

3.4 Simulation results

This section reports some simulated examples to show the finite sample performance of the proposed test.

3.4.1 Bootstrap Test Statistic

We should address one question about how to use the propose statistics T_n. There is an unknown parameter H in the LSD $F^{c_1', c_2'}$ and the asymptotic distribution of

$$\int \lambda dG^{(2)}_{p_1,p_2}(\lambda). \tag{3.39}$$

Here H is the LSD of the matrix $\mathbf{\Sigma}^{-1}_{\mathbf{xx}}$.

To overcome this difficulty, we consider a bootstrap method as follows. We redraw samples $\mathbf{X}^* = (\mathbf{x}^*_1, \mathbf{x}^*_2, \ldots, \mathbf{x}^*_n)$ and $\mathbf{Y}^* = (\mathbf{y}^*_1, \mathbf{y}^*_2, \ldots, \mathbf{y}^*_n)$ from the original samples $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$ and $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n)$ respectively. Then consider the bootstrap linear spectral statistic

$$\int \lambda dG^{(2)*}_{p_1,p_2}(\lambda), \tag{3.40}$$

where $G^{(2)*}_{p_1,p_2}(\lambda) = p_1\Big(F^{\mathbf{S}^*_{xy}}(\lambda) - F^{c'_{1n},c'_{2n}}_*(\lambda)\Big)$, $\mathbf{S}^*_{xy}$ is $\mathbf{S}_{xy}$ with $\mathbf{X}$ and $\mathbf{Y}$ replaced by $\mathbf{X}^*$ and $\mathbf{Y}^*$ respectively, and $F^{c'_{1n},c'_{2n}}_*$ is $F^{c'_{1n},c'_{2n}}$ with the bootstrap version $(\mathbf{X}^*, \mathbf{Y}^*)$.

From Theorem 4, the asymptotic distribution of the bootstrap statistic (3.40) is normal distribution with mean (3.22) and (3.23) in which the LSD H should be replaced by the LSD of $\mathbf{A}_{xx}$.

3.4.2 Empirical sizes and empirical powers

First we introduce the method of calculating empirical sizes and empirical powers. Let $z_{1-\alpha}$ be the $100(1-\alpha)\%$ quantile of the asymptotic null distribution of the test statistic S_n. With K replications of the data set simulated under the null hypothesis, we calculate the empirical size as

$$\hat{\alpha} = \frac{\{\sharp \ of \ S^H_n \geq z_{1-\alpha}\}}{K}, \tag{3.41}$$

where S^H_n represents the values of the test statistic S_n based on the data simulated under the null hypothesis.

The empirical power is calculated as

$$\hat{\beta} = \frac{\{\sharp\ of\ S_n^A \geq \hat{z}_{1-\alpha}\}}{K}, \tag{3.42}$$

where S_n^A represents the values of the test statistic S_n based on the data simulated under the alternative hypothesis.

In our simulation, we choose $K = 1000$ as the number of repeated simulations. The significance level is $\alpha = 0.05$.

3.4.3 Testing independence

Consider the data generating process

$$\mathbf{x} = \mathbf{\Sigma}_{\mathbf{xx}}^{1/2}\mathbf{w}, \quad \mathbf{y} = \mathbf{\Sigma}_{\mathbf{yy}}^{1/2}\mathbf{v}, \tag{3.43}$$

and two cases are investigated as

$$\begin{aligned} &(a) \quad \mathbf{\Sigma}_{\mathbf{xx}} = \mathbf{I}_{p_1}, \ \mathbf{\Sigma}_{\mathbf{yy}}^{1/2} = \mathbf{I}_{p_2}; \\ &(b) \quad \mathbf{\Sigma}_{\mathbf{xx}} = (\sigma_{kh}^{MA})_{k,h=1}^{p_1}, \ \mathbf{\Sigma}_{\mathbf{yy}}^{1/2} = \mathbf{I}_{p_2}, \end{aligned}$$

with

$$\sigma_{kh}^{MA} = \begin{cases} (1+\theta^2), & k = h; \\ \theta, & |k-h| = 1, \\ 0, & |k-h| > 1. \end{cases}$$

and $\theta = 0.6$.

The empirical sizes of the proposed statistics S_n for cases (a) and (b) are listed in Table 3.1. Moreover, the empirical sizes for re-normalized statistic MLR_n are included as comparison with S_n. Note that the re-normalized statistic MLR_n means that we use the statistic

$$p_1 \int \log(1-\lambda) d\big(F^{\mathbf{S}_{xy}}(\lambda) - F^{c_{1n},c_{2n}}(\lambda)\big).$$

The empirical sizes of T_n for cases (a) and (b) are listed in Table 3.2. From the results in Table 3.1 and 3.2, the proposed statistics S_n and T_n work well under Assumption 1 and 2 respectively.

3.4.4 Factor model dependence

We consider the factor model as follows:

$$\mathbf{x}_t = \Lambda_1 \mathbf{f}_t + \mathbf{u}_t, \quad \mathbf{y}_t = \Lambda_2 \mathbf{f}_t + \mathbf{v}_t, t = 1, 2, \ldots, n, \tag{3.44}$$

where Λ_1 and Λ_2 are $p_1 \times r$ and $p_2 \times r$ deterministic matrices respectively; $\mathbf{f}_t, t = 1, 2, \ldots, n$ are $r \times 1$ random vectors with i.i.d Gaussian distributed elements and $\mathbf{u}_t$ and $\mathbf{v}_t, t = 1, 2, \ldots, n$ are independent random vectors whose elements are all Gaussian distributed.

For this model, $\mathbf{x}_t$ and $\mathbf{y}_t$ are not independent if $r \neq 0$. The proposed test statistic S_n and T_n can be used to detect this dependent structure. Table 3.3 and 3.4 illustrate the powers of the proposed statistic S_n and T_n respectively, as r increases from 1 to 4. Results in these tables indicate that for one pair (p_1, p_2, n), the power increases as the number of factors r increases. This phenomenon makes sense since the dependence between $\mathbf{x}_t$ and $\mathbf{y}_t$ is described by the r common factors contained in the factor vector $\mathbf{f}_t$. Stronger dependence between $\mathbf{x}_t$ and $\mathbf{y}_t$ exists while more common factors are included in the model.

Here would like to point out that using CCA based on the sample covariance matrices with sample mean will incorrectly conclude that $\mathbf{x}_t$ and $\mathbf{y}_t$ are independent even if $r \neq 0$ but $\mathbf{f}_t = \mathbf{f}$ independent of t because CCA of $\mathbf{x}_t$ and $\mathbf{y}_t$ is the same as that of $\mathbf{u}_t$ and $\mathbf{v}_t$. This is why (1.4) and (1.6) are used.

3.4.5 Uncorrelated case

The construction of (3.8) is based on the idea that the limit of $F^{\mathbf{S}_{xy}}(x)$ could not be determined from (3.2) when $\mathbf{x}$ and $\mathbf{y}$ have correlation. Thus, a natural question is whether our statistic works in the uncorrelated but dependent case. Below is such an example to demonstrate the power of the test statistic in detecting uncorrelatedness.

Let $\mathbf{x}_t = (X_{1t}, X_{2t}, \dots, X_{p_1 t}), t = 1, 2, \dots, n$ be i.i.d normally distributed random vectors with zero means and unit variances. Define $\mathbf{y}_t = (Y_{1t}, Y_{2t}, \dots, Y_{p_2 t})$, $t = 1, 2, \dots, n$ by $Y_{it} = (X_{it}^{2k} - EX_{it}^{2k}), i = 1, 2, \dots, p_1$ and $Y_{jt} = \varepsilon_{jt}, j = p_1 + 1, \dots, p_2; t = 1, \dots, n$, where $\varepsilon_{jt}, j = p_1 + 1, \dots, p_2; t = 1, \dots, n$ are i.i.d normal distributed random variables and independent with $\mathbf{x}_t$ and k is an positive integer.

Remark 7. *For standard normal random variable X_{it}, the $2k$-th moment is $EX_{it}^{2k} = 2^{-k}\frac{(2k)!}{k!}$.*

For this model, $\mathbf{x}_t$ and $\mathbf{y}_t$ are uncorrelated since $Cov(X_{it}, Y_{it}) = EX_{it}^{2k+1} - EX_{it}EX_{it}^{2k} = 0$. Simulation results in Table 3.7 and Table 3.8 provide the empirical powers of S_n and T_n by taking $k = 2$ and $k = 5$ respectively. They show that S_n and T_n can distinguish this kind of dependent relationship well when $k = 5$.

3.4.6 ARCH type dependence

The statistic works in the above example because the limit of $F^{\mathbf{S}_{xy}}$ can not be determined from (3.2) if $\mathbf{x}$ and $\mathbf{y}$ are uncorrelated. However the limit of $F^{\mathbf{S}_{xy}}(x)$ might be the same as (3.2) when $\mathbf{x}$ and $\mathbf{y}$ are uncorrelated. We consider such an example as follows.

Consider two random vectors $\mathbf{x}_t = (X_{1t}, X_{2t}, \ldots, X_{p_1t})$ and $\mathbf{y}_t = (Y_{1t}, Y_{2t}, \ldots, Y_{p_2t})$ as follows:

$$Y_{it} = Z_{it}\sqrt{\alpha_0 + \alpha_1 X_{it}^2}, i = 1, 2, \ldots, p_1; Y_{jt} = Z_{jt}, j = p_1 + 1, \ldots, p_2, \quad (3.45)$$

where $\mathbf{z}_t = (Z_{1t}, Z_{2t}, \ldots, Z_{p_2t})$ is a random vector consisting of i.i.d elements generated from Normal (0,1) and $\{\mathbf{z}_t\}$ are independent across t; $\mathbf{x}_t = (X_{1t}, X_{2t}, \ldots, X_{p_1t})$ is also a random vector with i.i.d elements generated from Normal(0,1); Moreover, $\{\mathbf{z}_t : t = 1, \ldots, T$ are independent of $\{\mathbf{x}_t : t = 1, \ldots, T\}$.

For this model, $\mathbf{x}_t$ and $\mathbf{y}_t$ are dependent but uncorrelated. Simulation results indicate that the proposed test statistic S_n can not detect the dependence between them. Nevertheless, if we substitute the elements X_{it}^2 and Y_{it}^2 for X_{it} and Y_{jt}, respectively, in the matrix $\mathbf{S}_{xy}$, then the new resulting statistic S_n can capture the dependence of this type. This efficiency is due to the correlation between the high powers of X_{it} and Y_{it}.

Table 3.5 and 3.6 list the powers of the proposed statistics S_n and T_n for testing model (3.45) in several cases, i.e. α_0 and α_1 take different values. From the table, we can find the phenomenon that as α_1 increases, the powers also increase. This is consistent with our intuition because larger α_1 brings about larger correlation between Y_{it} and X_{it}.

3.5 Empirical applications

As an application of the proposed independence test, we test the cross-sectional dependence of daily closed stock prices sets between two different sections from New York Stock Exchange(NYSE) during the period 2000.1.1 − 2002.1.1, including consumer service section, consumer duration section, consumer nonduration section, energy section, finance section, transport section, healthcare section, capital goods section, basic industry section and public utility section. The data set is obtained from Wharton Research Data Services (WRDS) database.

We randomly choose p_1 and p_2 companies from two different sections respectively, such as the transport and finance section. At each time t, denote the closed stock prices of these companies from the two different sections as $\mathbf{x}_t = (x_{1t}, x_{2t}, \ldots, x_{p_1 t})$ and $\mathbf{y}_t = (y_{1t}, y_{2t}, \ldots, y_{p_2 t})$ respectively. For each company, there exist 1000 daily closed stock prices, i.e. $t = 1, 2, \ldots, 1000$. The goal is to test dependence between $\mathbf{x}_t$ and $\mathbf{y}_t$. From the time series $\{\mathbf{x}_t : t = 1, 2, \ldots, 1000\}$, we construct i.i.d samples and group them into a matrix $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_{1+20}, \ldots, \mathbf{x}_{1+20n})$, where $n \leq 50$. Similarly, we can derive the sample matrix $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_{1+20}, \ldots, \mathbf{y}_{1+20n})$. The construction of $\mathbf{X}$ and $\mathbf{Y}$ is based on the idea that the daily closed stock prices tend to being independent as the length of the time between them is long, i.e. current price is independent of the price after 20 days.

The proposed test S_n is applied to testing dependence of $\mathbf{x}_t$ and $\mathbf{y}_t$. For each (p_1, p_2, n), we randomly choose p_1 and p_2 companies from two different sections, construct the corresponding sample matrices $\mathbf{X}$ and $\mathbf{Y}$, and then calculate the P-value by applying the proposed test. Repeat this procedure 100 times and derive 100 P-values to see whether the cross-sectional 'dependence' feature is popular between the tested two sections.

We test independence of closed stock prices of companies from three pairs of sections, i.e. basic industry section and capital goods section, public utility section and capital goods section, finance section and healthcare section. From Table 3.9, Table 3.10, and Table 3.11, we can see that, as the pair of numbers of companies (p_1, p_2) increases, more experiments are rejected in terms of the P-values below 0.05. It shows that cross-sectional dependence exists and is popular for different sections in NYSE. This suggests that the assumption that cross-sectional independence in such empirical studies may not be appropriate.

3.6 Appendix

Throughout this chapter, M, M_1, M_2, K and K_1 denote positive constants which may change from line to line, $o(1)$ means the term converging to zero and $O(n^{-k})$ means the term divided by n^{-k} bounded in absolute value.

3.6.1 Some Useful Lemmas

Lemma 7 (Duhamel formula)**.** *Let* $\mathbf{M}_1, \mathbf{M}_2$ *be* $n \times n$ *matrices and* $t \in \mathbb{R}$*. Then we have*

$$e^{(\mathbf{M}_1+\mathbf{M}_2)t} = e^{\mathbf{M}_1 t} + \int_0^t e^{\mathbf{M}_1(t-s)}\mathbf{M}_2 e^{(\mathbf{M}_1+\mathbf{M}_2)s} ds. \tag{3.46}$$

Moveover, if $(A_{ij}(t))_{1\le i,j\le n}$ *is a matrix-valued function of* $t \in \mathbb{R}$ *that is* C^∞ *in the sense that each matrix element* $A_{ij}(t)$ *is* C^∞*. Then*

$$\frac{d}{dt}e^{\mathbf{A}(t)} = \int_0^1 e^{s\mathbf{A}(t)}\mathbf{A}'(t)e^{(1-s)\mathbf{A}(t)} ds. \tag{3.47}$$

Lemma 8. *Assume that* $F(\mathbf{X})$ *is a differentiable function of each of the elements of the matrix* $\mathbf{X}$*, it then holds that*

$$\frac{\partial Tr(F(\mathbf{X}))}{\partial \mathbf{X}} = f(\mathbf{X})^T,$$

where $f(\cdot)$ *is the scalar derivative of* $F(\cdot)$*.*

Lemma 9. *Let* $\mathbf{U} = f(\mathbf{X})$ *be a matrix, then the derivative of the function* $g(\mathbf{U}) : \mathbb{R}^{m\times n} \to \mathbb{R}^1$ *with respect to the element* X_{ij} *of* $\mathbf{X}$ *is*

$$\frac{\partial g(\mathbf{U})}{\partial X_{ij}} = Tr[(\frac{\partial g(\mathbf{U})}{\partial \mathbf{U}})^T \frac{\partial \mathbf{U}}{\partial X_{ij}}]. \tag{3.48}$$

Lemma 10 (Stein's equation)**.** *Let* $\xi = \{\xi_\ell\}_{\ell=1}^p$ *be independent Gaussian random variables of zero mean, and* $\Phi : \mathbb{R}^p \to \mathbb{C}$ *be a differentiable function with polynomially bounded partial derivatives* $\Phi'_\ell, \ell = 1, \ldots, p$. *Then we have*

$$E\{\xi_\ell \Phi(\xi)\} = E\{\xi_\ell^2\} E\{\Phi'_\ell(\xi)\}, \ \ell = 1, \ldots, p, \tag{3.49}$$

and

$$Var\{\Phi(\xi)\} \leq \sum_{\ell=1}^{p} E\{\xi_\ell^2\} E\{|\Phi'_\ell(\xi)|^2\}. \tag{3.50}$$

Lemma 11 (Generalized Stein's equation of Lytova and Pastur (2009))**.** *Let* ξ *be a random variable such that* $E|\xi|^{p+2} < \infty$ *for a certain nonnegative integer* p. *Then for any function* $\Phi : \mathbb{R} \to \mathbb{C}$ *of the class* C^{p+1} *with bounded derivative* $\Phi^{(\ell)}, \ell = 1, \ldots, p+1$, *we have*

$$E\{\xi \Phi(\xi)\} = \sum_{\ell=0}^{p} \frac{\kappa_{\ell+1}}{\ell!} E\{\Phi^{(\ell)}(\xi)\} + \varepsilon_p, \tag{3.51}$$

where the remainder term ε_p *admits the bound*

$$|\varepsilon_p| \leq C_p \int_0^1 E\Big|\xi^{p+2}\Phi^{(p+1)}(\xi v)\Big|(1-v)^p dv, \ C_p \leq \frac{1+(3+2p)^{p+2}}{(p+1)!}, \tag{3.52}$$

and $\kappa_{\ell+1}$ *is the* $\ell+1$*-th cumulant.*

Lemma 12 (Theorem A.37 of Bai and Silverstein (2009))**.** *If* $\mathbf{A}$ *and* $\mathbf{B}$ *are two* $n \times p$ *matrices and* $\lambda_k, \delta_k, \ k = 1, 2, \ldots, n$ *denote their singular values. If the singular values are arranged in descending order, then we have*

$$\sum_{k=1}^{\nu} |\lambda_k - \delta_k|^2 \leq tr[(\mathbf{A}-\mathbf{B})(\mathbf{A}-\mathbf{B})^*], \tag{3.53}$$

where $\nu = \min\{p, n\}$.

3.6.2 Proof of Theorem 2

Since the matrix $\mathbf{T}_{xy}$ is not symmetric, it is difficult to work on it directly. Instead we consider the $n \times n$ symmetric matrix

$$\mathbf{B}_n = \tilde{\mathbf{P}}_y \mathbf{P}_{tx} \tilde{\mathbf{P}}_y. \tag{3.54}$$

It is easily seen that the eigenvalues of the matrix $\mathbf{B}_n$ are the same as those of the matrix $\mathbf{T}_{xy}$ other than $(n - p_1)$ zero eigenvalues. It follows that the ESDs of $\mathbf{B}_n$ and $\mathbf{T}_{\mathbf{xy}}$ satisfy the equality

$$F^{\mathbf{B}_n}(x) = \frac{p_1}{n} F^{\mathbf{T}_{\mathbf{xy}}}(x) + \frac{n - p_1}{n} I_{[0,+\infty)}(x). \tag{3.55}$$

Below we first consider the case when the entries of $\mathbf{X}$ and $\mathbf{Y}$ ($\mathbf{W}$ and $\mathbf{V}$) are normal random variables. Write

$$\mathbf{X}^T = \mathbf{X}_1^T + \mathbf{X}_2^T, \tag{3.56}$$

where $\mathbf{X}_1^T = \tilde{\mathbf{P}}_y \mathbf{X}^T$ and $\mathbf{X}_2^T = (\mathbf{I} - \tilde{\mathbf{P}}_y)\mathbf{X}^T$ is the corresponding residual matrix. Let

$$\mathbf{W}_1^T = \tilde{\mathbf{P}}_y \mathbf{W}^T, \quad \mathbf{W}_2^T = (\mathbf{I}_n - \tilde{\mathbf{P}}_y)\mathbf{W}^T.$$

Then

$$\mathbf{X}_1 = \mathbf{\Sigma}_{\mathbf{xx}}^{1/2} \mathbf{W}_1, \quad \mathbf{X}_2 = \mathbf{\Sigma}_{\mathbf{xx}}^{1/2} \mathbf{W}_2.$$

Since $\tilde{\mathbf{P}}_y$ is a projection matrix, the entries of $\mathbf{W}_1$ are independent of those of $\mathbf{W}_2$ and $\mathbf{X}_1$ is independent of $\mathbf{X}_2$. Note that by the definition of Moore-Penrose pseudoinverse

$$\tilde{\mathbf{P}}_y = \tilde{\mathbf{P}}_v = \mathbf{V}^T(\mathbf{V}\mathbf{V}^T)^{-}\mathbf{V}. \tag{3.57}$$

The ESD of $\mathbf{B}_n$ can be then written as

$$F^{\mathbf{B}_n}(x) = F^{\frac{1}{n}\mathbf{X}_1^T(\frac{1}{n}\mathbf{X}\mathbf{X}^T + t\mathbf{I})^{-1}\mathbf{X}_1}(x)$$

$$
\begin{aligned}
&= F^{\frac{1}{n}\mathbf{W}_1^T(\frac{1}{n}\mathbf{W}\mathbf{W}^T+t\mathbf{\Sigma}_{\mathbf{xx}}^{-1})^{-1}\mathbf{W}_1}(x) \\
&= \frac{p_1}{n}F^{\left(\frac{1}{n}\mathbf{W}_1\mathbf{W}_1^T+\frac{1}{n}\mathbf{W}_2\mathbf{W}_2^T+t\mathbf{\Sigma}_{\mathbf{xx}}^{-1}-(\frac{1}{n}\mathbf{W}_2\mathbf{W}_2^T+t\mathbf{\Sigma}_{\mathbf{xx}}^{-1})\right)(\frac{1}{n}\mathbf{W}_1\mathbf{W}_1^T+\frac{1}{n}\mathbf{W}_2\mathbf{W}_2^T+t\mathbf{\Sigma}_{\mathbf{xx}}^{-1})^{-1}}(x) \\
&\quad +\frac{n-p_1}{n}I_{[0,+\infty)}(x) \\
&= \frac{p_1}{n}F^{\mathbf{I}-\left(\frac{1}{n}\mathbf{W}_1\mathbf{W}_1^T(\frac{1}{n}\mathbf{W}_2\mathbf{W}_2^T+t\mathbf{\Sigma}_{\mathbf{xx}}^{-1})^{-1}+\mathbf{I}\right)^{-1}}(x)+\frac{n-p_1}{n}I_{[0,+\infty)}(x). \qquad (3.58)
\end{aligned}
$$

This, together with (3.55), yields

$$F^{\mathbf{T_{xy}}}(x) = F^{\mathbf{I}-\left(\frac{1}{n}\mathbf{W}_1\mathbf{W}_1^T(\frac{1}{n}\mathbf{W}_2\mathbf{W}_2^T+t\mathbf{\Sigma}_{\mathbf{xx}}^{-1})^{-1}+\mathbf{I}\right)^{-1}}(x). \qquad (3.59)$$

If $p_2 \geq n$, then $Rank(\tilde{\mathbf{P}}_y) = tr\tilde{\mathbf{P}}_y = tr\tilde{\mathbf{P}}_v = n$ with probability one by the definition of Moore-Penrose pseudoinverse because $\mathbf{V}\mathbf{V}'$ has $(p_2 - n)$ zero eigenvalues and from Theorem 1.1 of Rudelson and Vershynin (2011) with probability one

$$\frac{\lambda_{\min}(\mathbf{V}'\mathbf{V})}{n} \geq \left(\frac{\sqrt{p_2}-\sqrt{n-1}}{\sqrt{n}}\right)^2\frac{1}{n^2}. \qquad (3.60)$$

It follows that there exists a unitary matrix $\mathbf{U}$ such that with probability one

$$\mathbf{U}^*\tilde{\mathbf{P}}_v\mathbf{U} = diag(1,\dots,1), \qquad (3.61)$$

where $diag(\cdot)$ denotes a diagonal matrix. Since $\mathbf{U}$ is a unitary matrix and all the elements of $\mathbf{W}$ are independent Gaussian random variables we obtain

$$\mathbf{W}^T \stackrel{d}{=} \mathbf{U}\mathbf{W}^T, \quad \mathbf{W}_1\mathbf{W}_1^T \stackrel{d}{=} \mathbf{W}\mathbf{W}^T, \qquad (3.62)$$

where $\stackrel{d}{=}$ denotes equality in distribution of two random variables. Hence $\frac{1}{n}\mathbf{W}_1\mathbf{W}_1^T(\frac{1}{n}\mathbf{W}_2\mathbf{W}_2^T + t\mathbf{\Sigma}_{\mathbf{xx}}^{-1})^{-1} \stackrel{d}{=} t^{-1}\frac{1}{n}\mathbf{W}\mathbf{W}^T\mathbf{\Sigma}_{\mathbf{xx}}$. This is a sample covariance matrix and its LSD and CLT have been provided in (6.1.2) and Theorem 9.10 of Bai and Silverstein (2009) respectively.

If $p_2 < n$ then $Rank(\tilde{\mathbf{P}}_y) = tr\tilde{\mathbf{P}}_y = tr\tilde{\mathbf{P}}_v = p_2$ with probability one by an inequality similar to (3.60). Therefore there exists a unitary matrix $\mathbf{U}$

such that with probability one

$$\mathbf{U}^*\tilde{\mathbf{P}}_y\mathbf{U} = diag(1,\ldots,1,0,\ldots,0), \tag{3.63}$$

where $diag(\cdot)$ denotes a diagonal matrix and the number of the entries 1 on the diagonal is p_2. This implies that

$$\mathbf{W}_1\mathbf{W}_1^T \stackrel{d}{=} \sum_{k=1}^{p_2}\mathbf{w}_k\mathbf{w}_k^T, \quad \mathbf{W}_2\mathbf{W}_2^T \stackrel{d}{=} \sum_{k=p_2+1}^{n}\mathbf{w}_k\mathbf{w}_k^T,$$

where $\mathbf{w}_k$ is the k-th column of $\mathbf{W}$. Therefore, with $q_n := \frac{p_2}{n-p_2}$ we then have

$$\frac{1}{n}\mathbf{W}_1\mathbf{W}_1^T(\frac{1}{n}\mathbf{W}_2\mathbf{W}_2^T + t\mathbf{\Sigma}_{\mathbf{xx}}^{-1})^{-1} \stackrel{d}{=} q_n\mathbf{S}_1 S_{2t}^{-1}, \tag{3.64}$$

where

$$\mathbf{S}_1 = \frac{1}{p_2}\sum_{k=1}^{p_2}\mathbf{w}_k\mathbf{w}_k^T, \quad \mathbf{S}_{2t} = \frac{1}{n-p_2}\sum_{k=p_2+1}^{n}\mathbf{w}_k\mathbf{w}_k^T + t\frac{n}{n-p_2}\mathbf{\Sigma}_{\mathbf{xx}}^{-1}.$$

Denote by $\mu_1,\mu_2,\ldots,\mu_{p_1}$ the eigenvalues of $\mathbf{S}_1 S_{2t}^{-1}$. In view of (3.59) the eigenvalues of $\mathbf{T}_{\mathbf{xy}}$ can be written as $\frac{q_n\mu_i}{1+q_n\mu_i}$, $i = 1,2,\ldots,p_1$. Note that (6.1.2) of Bai and Silverstein (2009) has provided the equation satisfied by the Stieltjes transform of the LSD of the matrix $\mathbf{ST}$, where $\mathbf{S}$ is a sample covariance matrix and $\mathbf{T}$ is a matrix which is independent of $\mathbf{S}$. Moreover the Stieltjes transform of the LSD of S_{2t} is provided in Silverstein and Bai (1995). By taking $\mathbf{S} = S_1$ and $\mathbf{T} = q_n\mathbf{S}_{2t}^{-1}$, we see that (3.5) follows from (6.1.2) of Bai and Silverstein (2009).

As for the nonGaussian case, write

$$\mathbf{P}_{tx} = \frac{1}{n}\mathbf{X}^T(\frac{1}{n}\mathbf{X}\mathbf{X}^T + t\mathbf{I}_{p_1})^{-1}\mathbf{X} = \frac{1}{n}\mathbf{W}^T(\frac{1}{n}\mathbf{W}\mathbf{W}^T + t\mathbf{\Sigma}_{\mathbf{xx}}^{-1})^{-1}\mathbf{W}. \tag{3.65}$$

Then the proof of Theorem 1 in Chapter 2 indeed shows that replacing Gaussian entries in $\mathbf{W}$ (or $\mathbf{X}$) by nonGaussian entries does not affect the

LSD of $\mathbf{B}_n$ and one may refer to (2.13). In view of (3.57), to replace Gaussian entries in $\mathbf{V}$ by nonGaussian entries, as in (2.9), one can first prove that the Levy distance, as $n \to \infty$, *then* $u \to 0$,

$$L^3\Big(F^{\mathbf{P}_{tx}^{1/2}\tilde{\mathbf{P}}_y\mathbf{P}_{tx}^{1/2}}, F^{\mathbf{P}_{tx}^{1/2}\mathbf{P}_{uy}\mathbf{P}_{tx}^{1/2}}\Big) \le \frac{Mu^2}{n} tr(\frac{1}{n}\mathbf{V}\mathbf{V}^T + u\mathbf{I}_{p_2})^{-2} \le Mu^2 \xrightarrow{a.s.} 0,$$

where $(\mathbf{P}_{tx}^{1/2})^2 = \mathbf{P}_{tx}$ and $\mathbf{P}_{uy} = \frac{1}{n}\mathbf{V}^T(\frac{1}{n}\mathbf{V}\mathbf{V}^T + u\mathbf{I}_{p_2})^{-1}\mathbf{V}$, $u > 0$. Moreover, we see that conclusion (2.13) still holds if we replace $\mathbf{P}_y$ and $\mathbf{P}_{tx}$ there by $\mathbf{P}_{tx}^{1/2}$ and $\mathbf{P}_{uy}$ respectively and check on its argument carefully. Therefore (2.13) ensures that replacing Gaussian entries in $\mathbf{Y}$ by nonGaussian entries does not affect the LSD of $\mathbf{P}_{tx}^{1/2}\mathbf{P}_{uy}\mathbf{P}_{tx}^{1/2}$ when the entries of $\mathbf{X}$ are nonGaussian. The proof is now complete.

3.6.3 Proof of Theorem 3

The strategy of the proof is to first associate sample correlation coefficients with the F matrix when the entries of $\mathbf{x}$ are Gaussian distributed, whose CLT was provide by Zheng (2012). Then by an interpolation trick first adopted in Lytova and Pastur (2009), we extend the result to the non-Gaussian distributions. When applying such an interpolation method, an additional key technique is to introduce a smooth cut function so that we can handle the expectation of the trace of the inverse of the sample covariance matrix.

3.6.3.1 The Gaussian case

Since the classical sample canonical correlation coefficients between $\mathbf{x}$ and $\mathbf{y}$ are the same with those between $\mathbf{w}$ and $\mathbf{v}$, we assume that $\mathbf{\Sigma_{xx}} = \mathbf{\Sigma_{yy}} = \mathbf{I}$ in this theorem.

Assume that the entries of $\mathbf{X}$ are Gaussian distributed. We below demonstrate how the eigenvalues of the matrix $\mathbf{S}_{xy}$ are connected to those of an F-matrix.

We would remind the readers that the matrix $\mathbf{S}_{xy}$ consists of the project matrix $\mathbf{P}_x$ rather than it perturbation matrix $\mathbf{P}_{tx}$ and $\mathbf{P}_y$ rather than $\tilde{\mathbf{P}}_y$ where

$$\mathbf{P}_x = \mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-1}\mathbf{X},\ \mathbf{P}_y = \mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T)^{-1}\mathbf{Y}.$$

As before, since the matrix $\mathbf{S}_{xy}$ is not symmetric we instead consider the $n \times n$ symmetric matrix

$$\mathbf{A}_n = \mathbf{P}_y\mathbf{P}_x\mathbf{P}_y. \tag{3.66}$$

Then we have

$$F^{\mathbf{A}_n}(x) = \frac{p_1}{n}F^{\mathbf{S}_{xy}}(x) + \frac{n-p_1}{n}I_{[0,+\infty)}(x). \tag{3.67}$$

Note that under Assumption 1 $Rank(\tilde{\mathbf{P}}_y) = tr\tilde{\mathbf{P}}_y = p_2$ with probability one because $\lambda_{\min}(\mathbf{Y}'\mathbf{Y})/n \xrightarrow{a.s.} (1-\sqrt{c_2})^2$. Therefore, with a little abuse of notation, as in (3.64), we obtain

$$\mathbf{X}_1\mathbf{X}_1^T \stackrel{d}{=} \sum_{k=1}^{p_2}\mathbf{x}_k\mathbf{x}_k^T,\ \mathbf{X}_2\mathbf{X}_2^T \stackrel{d}{=} \sum_{k=p_2+1}^{n}\mathbf{x}_k\mathbf{x}_k^T, \tag{3.68}$$

where $\mathbf{X}_1$ and $\mathbf{X}_2$ are similarly defined as in (3.56) with $\tilde{\mathbf{P}}_y$ replaced by $\mathbf{P}_y$. As in (3.59) we conclude that

$$\begin{aligned} F^{\mathbf{A}_n}(x) &= F^{\mathbf{X}_1^T(\mathbf{X}\mathbf{X}^T)^{-1}\mathbf{X}_1}(x) = \frac{p_1}{n}F^{(\mathbf{X}_1\mathbf{X}_1^T+\mathbf{X}_2\mathbf{X}_2^T)^{-1}\mathbf{X}_1\mathbf{X}_1^T}(x) + \frac{n-p_1}{n}I_{[0,+\infty)}(x) \\ &= \frac{p_1}{n}F^{\left(\mathbf{I}+\mathbf{X}_2\mathbf{X}_2^T(\mathbf{X}_1\mathbf{X}_1^T)^{-1}\right)^{-1}}(x) + \frac{n-p_1}{n}I_{[0,+\infty)}(x). \end{aligned} \tag{3.69}$$

This, together with (3.67), yields

$$F^{\mathbf{S}_{xy}}(x) = F^{\left(\mathbf{I}+\mathbf{X}_2\mathbf{X}_2^T(\mathbf{X}_1\mathbf{X}_1^T)^{-1}\right)^{-1}}(x). \tag{3.70}$$

Since $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent the matrix $\frac{1}{n-p_2}\mathbf{X}_2\mathbf{X}_2^T(\frac{1}{p_2}\mathbf{X}_1\mathbf{X}_1^T)^{-1}$ is an $F-$ matrix. The limiting spectral distribution of the F-matrix is

$$F_{\bar{y}_1,\bar{y}_2}(dx) = g_{\bar{y}_1,\bar{y}_2}(x)I_{[a_1,a_2]}(x)dx + (1-\frac{1}{\bar{y}_1})I_{\{\bar{y}_1>1\}}\delta_0(dx), \tag{3.71}$$

where $g_{\bar{y}_1,\bar{y}_2}$ is given in (3.15) (one may see Section 4 of Bai and Silverstein (2009)).

Denoting the eigenvalues of $\frac{1}{n-p_2}\mathbf{X}_2\mathbf{X}_2^T(\frac{1}{p_2}\mathbf{X}_1\mathbf{X}_1^T)^{-1}$ by $\lambda_1,\ldots,\lambda_{p_1}$, then the eigenvalues of the matrix $\mathbf{S}_{xy}$ can be expressed as $\frac{1}{1+\frac{n-p_2}{p_2}\lambda_1},\ldots,\frac{1}{1+\frac{n-p_2}{p_2}\lambda_{p_1}}$. Therefore the statistic (3.13) can be expressed as

$$\int \phi(\lambda)dG_{p_1,p_2}(\lambda) = \int \phi(\frac{1}{1+\frac{n-p_2}{p_2}\lambda})dp_1[F^{\frac{1}{n-p_2}\mathbf{X}_2\mathbf{X}_2^T(\frac{1}{p_2}\mathbf{X}_1\mathbf{X}_1^T)^{-1}}(\lambda)-F_{\bar{y}_{1n},\bar{y}_{2n}}(\lambda)], \tag{3.72}$$

where $F_{\bar{y}_{1n},\bar{y}_{2n}}$ is obtained from $F_{\bar{y}_1,\bar{y}_2}$ with the substitution of $(\bar{y}_{n1},\bar{y}_{n2})$ for $(\bar{y}_1,\bar{y}_2)$ and the associated constants (h_n,a_{n1},a_{n2}) for (h,a_1,a_2), i.e.

$$\begin{aligned} \bar{y}_{n1} &= \frac{p_1}{n-p_2},\ \bar{y}_{n2} = \frac{p_1}{p_2},\ h_n = \sqrt{\bar{y}_{n1}+\bar{y}_{n2}-\bar{y}_{n1}\bar{y}_{n2}}, \\ a_{n1} &= \frac{(1-h_n)^2}{(1-\bar{y}_{n2})^2},\ a_{n2} = \frac{(1+h_n)^2}{(1-\bar{y}_{n2})^2}. \end{aligned}$$

In view of (3.72), it suffices to provide the CLT for the F-matrix $\mathbf{C}_n = \frac{1}{n-p_2}\mathbf{X}_2\mathbf{X}_2^T(\frac{1}{p_2}\mathbf{X}_1\mathbf{X}_1^T)^{-1}$. Zheng (2012) has established the CLTs for linear spectral statistics of F-matrices, which yields Theorem 3 for the Gaussian distribution ((3.20) holds in the Gaussian case).

3.6.3.2 The general case

We next consider the CLT for the general distribution by the interpolation trick. By (3.67), we have

$$\int \phi(\lambda)dG_{p_1,p_2}(\lambda) = n[\int \phi(\lambda)d\big(F_n^{\mathbf{P}_y\mathbf{P}_x\mathbf{P}_y}(\lambda) - F_n^{yxy}(\lambda)\big)], \tag{3.73}$$

where $F_n^{yxy}(\lambda)$ is obtained from the limit, F^{yxy}, of $F^{\mathbf{P}_y\mathbf{P}_x\mathbf{P}_y}$ with c_1 and c_2 replaced by p_1/n and p_2/n respectively.

We start with the truncation of the underlying random variables. Define

$$\widetilde{\mathbf{X}}_n = (\tilde{X}_{ij})_{p_1\times n},\ \breve{\mathbf{X}} = (\breve{X}_{ij})_{p_1\times n} \tag{3.74}$$

where $\tilde{X}_{ij} = (\breve{X}_{ij} - E\breve{X}_{ij})/\sigma_{ij}$, $\breve{X}_{ij} = X_{ij}I_{|X_{ij}|<\sqrt{n}\varepsilon}$ and $\sigma_{ij}^2 = E|\breve{X}_{ij} - E\breve{X}_{ij}|^2$. Choose $\varepsilon_n > 0$ such that $\varepsilon_n \to 0$, $n^{1/2}\varepsilon_n \to \infty$ and $\frac{K}{\varepsilon_n}EX_{11}^4 I_{(|X_{11}|>\sqrt{n}\varepsilon)} \to 0$ as $n \to \infty$. Denote $\varepsilon = \varepsilon_n$ and we have

$$P\Big(\mathbf{P}_x \neq \breve{\mathbf{P}}_x\Big) \leq \sum_{i,j=1}^{p_1,n} P\Big(X_{ij} \neq \breve{X}_{ij}\Big) \leq \frac{K}{\varepsilon^4}EX_{11}^4 I_{(|X_{11}|>\sqrt{n}\varepsilon)} \to 0, \tag{3.75}$$

where $\breve{\mathbf{P}}_x$ is obtained from $\mathbf{P}_x$ with $\mathbf{X}$ replaced by $\breve{\mathbf{X}}$.

Let $\lambda_k^{\mathbf{A}}$ denote the i-th smallest eigenvalue of an Hermitian matrix $\mathbf{A}$. We use $\breve{G}_{p_1,p_2}(x)$ and $\widetilde{G}_{p_1,p_2}(x)$ to denote the analogues of $G_{p_1,p_2}(x)$ with the matrix $\mathbf{C}_n = \mathbf{P}_y\mathbf{P}_x\mathbf{P}_y$ replaced by $\breve{\mathbf{C}}_n = \mathbf{P}_y\breve{\mathbf{P}}_x\mathbf{P}_y$ and $\widetilde{\mathbf{C}}_n = \mathbf{P}_y\widetilde{\mathbf{P}}_x\mathbf{P}_y$ with $\widetilde{\mathbf{P}}_x = \widetilde{\mathbf{X}}_n^T(\widetilde{\mathbf{X}}_n\widetilde{\mathbf{X}}_n^T)^{-1}\widetilde{\mathbf{X}}_n$, respectively. By Lemma 12, we have, for each j=1,2,. . . ,s,

$$\begin{aligned}
&\Big|\int \phi_j(x)d\breve{G}_n(x) - \int \phi_j(x)d\widetilde{G}_n(x)\Big| \leq K\sum_{k=1}^n |\lambda_k^{\breve{\mathbf{C}}_n} - \lambda_k^{\widetilde{\mathbf{C}}_n}| \\
\leq\ & \sqrt{n}\Big(\sum_{k=1}^n |\lambda_k^{\breve{\mathbf{C}}_n} - \lambda_k^{\widetilde{\mathbf{C}}_n}|^2\Big)^{1/2} \leq \sqrt{n}\Big(tr(\breve{\mathbf{C}}_n - \widetilde{\mathbf{C}}_n)(\breve{\mathbf{C}}_n - \widetilde{\mathbf{C}}_n)^T\Big)^{1/2} \\
\leq\ & \sqrt{n}\Big(tr(\breve{\mathbf{P}}_x - \widetilde{\mathbf{P}}_x)(\breve{\mathbf{P}}_x - \widetilde{\mathbf{P}}_x)^T\Big)^{1/2},
\end{aligned} \tag{3.76}$$

where K is a bound on $|f_j'(z)|$. Moreover, one can check that

$$(\sigma_{11}^{-1} - 1)^2 = o(n^{-2}), \qquad |E\breve{X}_{11}| = o(n^{-\frac{3}{2}}). \tag{3.77}$$

By the formula

$$\mathbf{A}^{-1} - \mathbf{B}^{-1} = \mathbf{A}^{-1}(\mathbf{B} - \mathbf{A})\mathbf{B}^{-1}, \tag{3.78}$$

we obtain

$$\breve{\mathbf{P}}_x - \widetilde{\mathbf{P}}_x = \mathbf{Q}_1 + \mathbf{Q}_2 + \mathbf{Q}_3 + \mathbf{Q}_4,$$

where

$$\mathbf{Q}_1 = \frac{1}{n}\mathbf{X}_d\breve{\mathbf{H}}^{-1}\breve{\mathbf{X}}_n^T,\ \mathbf{Q}_2 = -\frac{1}{n}\widetilde{\mathbf{X}}_n\widetilde{\mathbf{H}}^{-1}\mathbf{X}_d,\quad \mathbf{Q}_3 = -\frac{1}{n}\widetilde{\mathbf{X}}_n\breve{\mathbf{H}}^{-1}\frac{1}{n}\mathbf{X}_d\widetilde{\mathbf{X}}_n\widetilde{\mathbf{H}}^{-1}\breve{\mathbf{X}}_n^T,$$
$$\mathbf{Q}_4 = -\frac{1}{n}\widetilde{\mathbf{X}}_n\breve{\mathbf{H}}^{-1}\frac{1}{n}\breve{\mathbf{X}}_n\mathbf{X}_d\widetilde{\mathbf{H}}^{-1}\breve{\mathbf{X}}_n^T,$$

with $\mathbf{X}_d = \breve{\mathbf{X}}_n - \widetilde{\mathbf{X}}_n$, $\breve{\mathbf{H}}^{-1} = (\frac{1}{n}\breve{\mathbf{X}}_n^T\breve{\mathbf{X}}_n)^{-1}$ and $\widetilde{\mathbf{H}}^{-1} = (\frac{1}{n}\widetilde{\mathbf{X}}_n^T\widetilde{\mathbf{X}}_n)^{-1}$. Note that

$$tr(\breve{\mathbf{P}}_x - \widetilde{\mathbf{P}}_x)(\breve{\mathbf{P}}_x - \widetilde{\mathbf{P}}_x)^T \leq K\sum_{i,j=1}^{4} tr\mathbf{Q}_i\mathbf{Q}_j^T.$$

We obtain from (3.77)

$$\begin{aligned} tr\mathbf{Q}_1\mathbf{Q}_1^T &\leq \tfrac{||\breve{\mathbf{H}}^{-1}||}{n}\Big(tr\mathbf{X}_d^T\mathbf{X}_d\Big) \leq K||\breve{\mathbf{H}}^{-1}||\Big[(1-1/\sigma_{11})^2 tr\breve{\mathbf{H}}\breve{\mathbf{X}}_n + \sigma_{11}^{-2}n|E\breve{X}_{11}|^2\Big] \\ &\leq K||\breve{\mathbf{H}}^{-1}||\Big[(1-1/\sigma_{11})^2 n\lambda_{\max}\big(\breve{\mathbf{H}}^{-1}\big) + \sigma_{11}^{-2}n|E\breve{X}_{11}|^2\Big] = o(n^{-1}). \end{aligned}$$

Similarly, one may verify that $tr\mathbf{Q}_j\mathbf{Q}_j^T = o(n^{-1})$, j=2,3,4. It follows that

$$\Big|\int \phi_j(x)d\breve{G}_n(x) - \int \phi_j(x)d\widetilde{G}_n(x)\Big| \xrightarrow{i.p.} 0.$$

In what follows, for simplicity we still use notation X_{ij} rather than $\widetilde{X}_{ij}$ and can assume that

$$|X_{ij}| \leq \sqrt{n}\varepsilon,\ EX_{ij} = 0,\ EX_{ij}^2 = 1. \tag{3.79}$$

To employ the interpolation trick we first introduce some notation. Let

$$\mathcal{N}_n[\phi] = n\int \phi(\lambda)dF^{\mathbf{A}_n}(\lambda),\ \mathcal{N}_n^{\circ}[\phi] = n\int \phi(\lambda)d[F^{\mathbf{A}_n}(\lambda) - F^{yxy}(\lambda)].$$

Moreover we introduce the following interpolating matrices

$$\mathbf{A}_n(s) = \mathbf{P}_y\mathbf{P}_x(s)\mathbf{P}_y,\ \mathbf{X}(s) = s^{1/2}\mathbf{X} + (1-s)^{1/2}\hat{\mathbf{X}},$$

$$\mathbf{P}_x(s) = \frac{1}{n}\mathbf{X}^T(s)\mathbf{H}^{-1}(s)\mathbf{X}(s), \quad \mathbf{H}^{-1}(s) = (\mathbf{H}(s))^{-1} = (\frac{1}{n}\mathbf{X}(s)\mathbf{X}^T(s))^{-1},$$

where $\hat{\mathbf{X}} = (\hat{X}_{kj})$ is obtained from $\mathbf{X} = (X_{kj})$ but consisting of standardized normal random variables. Define

$$\begin{aligned} e_n(s,x) &= exp\Big(ixTr\phi(\mathbf{A}_n(s))\Big), \quad \mathbf{U}(t,s) = e^{it\mathbf{A}_n(s)}, \\ e_n^{\circ}(s,x) &= exp\Big(ix[Tr\phi(\mathbf{A}_n(s)) - n\int \phi(\lambda)dF_n^{yxy}(\lambda)]\Big). \end{aligned} \tag{3.80}$$

By the continuous theorem of characteristic functions and Subsection 3.6.3.1 it suffices to prove that

$$\hat{R}_n(x) = E\Big(e^{ix\mathcal{N}_n^{\circ}[\phi]}\Big) - E\Big(e^{ix\hat{\mathcal{N}}_n^{\circ}[\phi]}\Big) \to 0, \quad as\ n \to \infty, \tag{3.81}$$

where $\hat{\mathcal{N}}_n^{\circ}[\phi]$ is the analogue of $\mathcal{N}_n^{\circ}[\phi]$ with all entries of $\mathbf{X}$ replaced by i.i.d standardized normal random variables.

For technical requirements, we introduce a smooth cut off function $\chi(x) : \mathbb{R} \to \mathbb{R}$:

$$\chi(x) = \begin{cases} 1, & |x| \le K_1 n^{-2} \\ 0, & |x| \ge 2K_1 n^{-2}, \end{cases} \tag{3.82}$$

whose first four derivatives satisfy $|\chi^{(j)}(x)| \le Mn^{2j}, j = 1,2,3,4.$

To prove (3.81) we first claim that

$$\tilde{R}_n(x) = E\Big(e^{ix\mathcal{N}_n^{\circ}[\phi]}\Big) - E\Big(e^{ix\mathcal{N}_n^{\circ}[\phi]}\chi(\mathrm{I}m(m_n(in^{-2})))\Big) \to 0,\ as\ n \to \infty, \tag{3.83}$$

where $m_n(z)$ is the Stieltjes transform of $\mathbf{H} = \frac{1}{n}\mathbf{X}\mathbf{X}^T$. Indeed, let $\tilde{\lambda}_1, \ldots, \tilde{\lambda}_{p_1}$ be the eigenvalues of $\mathbf{H}$. Since

$$\mathrm{I}m(m_n(in^{-2})) = n^{-3}\sum_{i=1}^{n} \frac{1}{\tilde{\lambda}_i^2 + n^{-4}}, \tag{3.84}$$

we conclude that

$$\tilde{\lambda}_{p_1} > \frac{M_2}{n} \quad \text{if} \quad |\mathrm{I}m(m_n(in^{-2}))| \le M_1 n^{-2}, \tag{3.85}$$

where M_1 may be the same as or different from K_1 given in (3.82). From Theorem 9.13 of Bai and Silverstein (2009), under our truncation, we have, for any $x > 0$ and any integers $k \geq 2$,

$$P(\tilde{\lambda}_{p_1} \leq (1-\sqrt{c_1})^2 - x) = O(n^{-k}). \tag{3.86}$$

By (3.84) and taking an appropriate x we have

$$P\big(|\mathrm{Im}(m_n(in^{-2}))| \leq K_1 n^{-2}\big) \geq P\big(\tilde{\lambda}_{p_1} > (1-\sqrt{c_1})^2 - x\big) = 1 - O(n^{-k}). \tag{3.87}$$

This is equivalent to

$$P\big(\chi(\mathrm{Im}(m_n(in^{-2}))) = 1\big) = 1 - O(n^{-k}). \tag{3.88}$$

It follows that

$$\begin{aligned} |\tilde{R}_n(x)| &= |E\Big(e^{ix\mathcal{N}_n^\circ[\phi]}\big(1-\chi(\mathrm{Im}(m_n(in^{-2})))\big)\Big)| \\ &\leq P\Big(\chi(\mathrm{Im}(m_n(in^{-2}))) \neq 1\Big) = O(n^{-k}) \to 0, \ as \ n \to \infty. \end{aligned} \tag{3.89}$$

Thus (3.83) is true.

Evidently, (3.83) holds as well if $\mathbf{X}$ is replaced by its normal analogue, $\hat{\mathbf{X}}$. In view of (3.83), to prove (3.81), it suffices to prove that, as $n \to \infty$,

$$R_n(x) = E\Big(e^{ix\mathcal{N}_n^\circ[\phi]}\chi\big(\mathrm{Im}(m_n(in^{-2}))\big)\Big) - E\Big(e^{ix\hat{\mathcal{N}}_n^\circ[\phi]}\chi\big(\mathrm{Im}(\hat{m}_n(in^{-2}))\big)\Big) \to 0, \tag{3.90}$$

where $\hat{m}_n(z)$ is the Stieltjes transform of $\hat{\mathbf{H}} = \frac{1}{n}\hat{\mathbf{X}}\hat{\mathbf{X}}^T$.

We show here for future use the moment of $(\lambda_{\min}^{-r}\chi(\mathrm{Im}(m_n^s(in^{-2})))$ where $m_n^s(z)$ denotes the Stieltjes transform of $\mathbf{H}(s)$ and $\lambda_{\min}$ denotes the minimum eigenvalue of $\mathbf{H}(s)$. Note that (3.85)-(3.88) still hold for $\mathbf{H}(s)$ (replace eigenvalues of $\mathbf{H}$ correspondingly by $\mathbf{H}(s)$) because the truncation steps for

$\{X_{kj}\}$ are applicable to $\{\hat{X}_{kj}\}$. In what follows we shall directly quote them for $\mathbf{H}(s)$. By (3.85) and (3.86) we have, for any integer $r>0$,

$$\begin{aligned} E\Big[\frac{\chi(\mathrm{Im}(m_n^s(in^{-2})))}{\lambda_{\min}^r}\Big] \quad &\leq \quad Mn^r\cdot P\big(\frac{M_2}{n}<\lambda_{\min}<(1-\sqrt{c_1})^2-x\big) \\ &\qquad +M\big((1-\sqrt{c_1})^2-x\big)^{-r}=O(1). \qquad (3.91)\end{aligned}$$

We now consider (3.90). In what follows, to simplify notation denote $\chi(\mathrm{Im}(m_n^s(in^{-2})))$ by χ_{ns}. By the inverse Fourier transform

$$\phi(\lambda)=\int e^{it\lambda}\hat{\phi}(t)dt, \qquad (3.92)$$

where $\hat{\phi}(t)$ is the Fourier transform of $\phi(\lambda)$, i.e. $\hat{\phi}(t)=\frac{1}{2\pi}\int e^{-it\lambda}\phi(\lambda)d\lambda$, we obtain

$$\begin{aligned} &R_n(x)=\int_0^1\frac{\partial}{\partial s}E\Big(e_n^\circ(s,x)\chi_{ns}\Big)ds \\ = \quad & ixe^{-ixn\int\phi(\lambda)dF_n^{yxy}(\lambda)}\times\int_0^1 ds\int\hat{\phi}(\theta)\theta d\theta\cdot E\Big(Tr\mathbf{U}(\theta,s)\mathbf{P}_y\frac{\partial\mathbf{P}_x(s)}{\partial s}\mathbf{P}_y e_n(s,x)\chi_{ns}\Big) \\ &+\int_0^1 E\Big[e_n^\circ(s,x)\frac{\partial}{\partial s}\Big(\chi_{ns}\Big)\Big]ds. \qquad (3.93)\end{aligned}$$

We next prove that the last term in (3.93) converges to zero.

To this end, we first list formulas for matrix derivatives. By the matrix derivative formula

$$\frac{\partial\mathbf{H}^{-1}(s)}{\partial s}=-\mathbf{H}^{-1}(s)\frac{\partial\mathbf{H}(s)}{\partial s}\mathbf{H}^{-1}(s), \qquad (3.94)$$

and the chain rule of matrix derivatives, we have

$$\begin{aligned}\frac{\partial\mathbf{P}_x(s)}{\partial s} \quad &= \quad \frac{1}{2n}\mathbf{X}_{ds}^T\mathbf{H}^{-1}(s)\mathbf{X}(s)+\frac{1}{2n}\mathbf{X}^T(s)\mathbf{H}^{-1}(s)\mathbf{X}_{ds} \\ &\qquad -\frac{1}{2n^2}\mathbf{X}^T(s)\mathbf{H}^{-1}(s)[\mathbf{X}_{ds}\mathbf{X}^T(s)+\mathbf{X}(s)\mathbf{X}_{ds}^T]\mathbf{H}^{-1}(s)\mathbf{X}(s), \\ & \qquad (3.95)\end{aligned}$$

where $\mathbf{X}_{ds} = (\frac{1}{\sqrt{s}}\mathbf{X} - \frac{1}{\sqrt{1-s}}\hat{\mathbf{X}})$. Denote the first derivative with respect to $\frac{1}{\sqrt{n}}X_{kj}(s)$ by

$$D_{kj} = \partial/\partial(\frac{1}{\sqrt{n}}X_{kj}(s)).$$

Similar to (3.94) we obtain

$$D_{kj}(\mathbf{H}^{-1}(s)) = -\mathbf{H}^{-1}(s)\mathbf{W}_n(s,k,j)\mathbf{H}^{-1}(s), \qquad D_{kj}(\frac{1}{\sqrt{n}}\mathbf{X}(s)) = \mathbf{e}_k\mathbf{e}_j^T, \tag{3.96}$$

where

$$\mathbf{W}_n(s,k,j) = \mathbf{e}_k\mathbf{e}_j^T\frac{1}{\sqrt{n}}\mathbf{X}^T(s) + \frac{1}{\sqrt{n}}\mathbf{X}(s)\mathbf{e}_j\mathbf{e}_k^T.$$

From Lemma 9, Lemma 8, (3.87), (3.79), (3.96) and (3.88), we have

$$\begin{aligned} &\Big|E\Big(D_{kj}\chi_{ns}\Big)\Big| = \Big|\frac{1}{n}E\Big(\chi_{ns}^{'}D_{kj}\mathrm{Im}\Big(tr(\mathbf{H}(s) - in^{-2}\mathbf{I})^{-1}\Big)\Big)\Big| \\ = &\ \Big|\frac{1}{n}E\Big[\chi_{ns}^{'}\mathrm{Im}\Big(Tr\big[(\mathbf{H}(s) - in^{-2}\mathbf{I})^{-2}\mathbf{W}_n(s,k,j)\big]\Big)\Big]\Big| \\ \le &\ Mn^7P(\chi_{ns} \ne 1) = O(n^{-k}), \quad \text{for any} \ \ k, \end{aligned} \tag{3.97}$$

where the last inequality uses the fact that $\chi_{ns}^{'} \ne 0$ occurs only when $K_1 n^{-2} \le \mathrm{Im}(m_n^s(n^{-2})) \le 2K_1 n^{-2}$. This ensures the last term in (3.93) converges to zero.

In view of (3.93), (3.95) and (3.97) we may write $R_n(x)$ as

$$R_n(x) = \frac{ixe^{-ixn\int\phi(\lambda)dF_n^{yxy}(\lambda)}}{2}\int_0^1 ds\int\hat{\phi}(\theta)\theta d\theta\sum_{i=1}^{2}[Q_n^{(i)} - V_n^{(i)}] + o(1), \tag{3.98}$$

where

$$Q_n^{(1)} = \frac{1}{\sqrt{ns}}\sum_{j,k=1}^{n,p_1}E(X_{kj}\Phi_{kj}^{(1)}), \quad V_n^{(1)} = \frac{1}{\sqrt{n(1-s)}}\sum_{j,k=1}^{n,p_1}E(\hat{X}_{kj}\Phi_{kj}^{(1)}),$$

with

$$\Phi_{kj}^{(1)} = \Phi_{kj}^{(1)}(X_{kjs}) = \big(\mathbf{H}^{-1}(s)\frac{1}{\sqrt{n}}\mathbf{X}(s)\mathbf{P}_y\mathbf{U}(\theta,s)\mathbf{P}_y\big)_{kj}e_n(s,x)\chi_{ns},$$

$$X_{kjs} = s^{1/2} X_{kj} + (1-s)^{1/2} \hat{X}_{kj}; \tag{3.99}$$

and

$$Q_n^{(2)} = \frac{1}{\sqrt{ns}} \sum_{k,j=1}^{p_1,n} E(X_{kj} \Phi_{kj}^{(2)}), \quad V_n^{(2)} = \frac{1}{\sqrt{n(1-s)}} \sum_{k,j=1}^{p_1,n} E(\hat{X}_{kj} \Phi_{kj}^{(2)}),$$

with

$$\Phi_{kj}^{(2)} = \Phi_{kj}^{(2)}(X_{kjs}) = \big(\mathbf{P}_x(s) \mathbf{P}_y \mathbf{U}(\theta, s) \mathbf{P}_y \frac{1}{\sqrt{n}} \mathbf{X}^T(s) \mathbf{H}^{-1}(s)\big)_{jk} e_n(s, x) \chi_{ns}.$$

Now, the aim is to prove that (3.98) $\to 0$ as $n \to \infty$. To this end, we first further simplify $Q_n^{(i)}$ and $V_n^{(i)}, i = 1, 2$. Applying stein's equation in Lemma 10 to the terms $V_n^{(1)}$ and $V_n^{(2)}$ respectively, we can obtain

$$V_n^{(1)} = \frac{1}{n} \sum_{j,k=1}^{n,p_1} E(D_{kj} \Phi_{kj}^{(1)}), \; V_n^{(2)} = \frac{1}{n} \sum_{j,k=1}^{n,p_1} E(D_{kj} \Phi_{kj}^{(2)}). \tag{3.100}$$

Similarly, by generalized stein's equation in Lemma 11 with $p = 3$, we have

$$Q_n^{(i)} = \sum_{\ell=0}^{3} T_{\ell\varepsilon}^{(i)} + \xi_3^{(i)}, \quad i = 1, 2; \tag{3.101}$$

where

$$T_{\ell\varepsilon}^{(i)} = \frac{s^{\frac{\ell-1}{2}}}{\ell! n^{\frac{\ell+1}{2}}} \sum_{j,k=1}^{n,p_1} \kappa_{\ell+1,kj}^{\varepsilon} E(D_{kj}^{\ell} \Phi_{kj}^{(i)}), \; \ell = 0, 1, 2, 3;$$

with $\kappa_{\ell,kj}^{\varepsilon}$ being the ℓ-th cumulant of the truncated random variable X_{kj} and

$$|\xi_3^{(i)}| \le \frac{K}{n^{5/2}} \sum_{k,j=1}^{n,p_1} \int_0^1 E\Big| X_{kjs}^5 D_{kj}^4 \Phi_{kj}^{(i)}(v X_{kjs}) \Big| dv,$$

where $\Phi_{kj}^{(i)}(vX_{kjs})$ is obtained from $\Phi_{kj}^{(i)}(X_{kjs})$ given in (3.99) with X_{kjs} replaced by vX_{kjs}.

We next prove that $E\Big| D_{kj}^{\ell} \Phi_{kj}^{(i)} \Big|^2$ is bounded for $\ell = 1, 2, 3, 4, \; i = 1, 2$. To this end, we below develop the expansion of $D_{kj}(s)\Phi_{kj}^{(1)}(s)$ first. Let $\mathbf{e}_k$ be the unit vector with the k th entry being 1 and zero otherwise. Recalling the definition of the matrix $\mathbf{U}(\theta, s)$ in (3.80) and applying the Duhamel formula (3.47) and (3.96) we have

$$\begin{aligned} D_{kj}(\mathbf{U}(\theta, s)) &= \int_0^1 e^{it\theta\mathbf{A}_n(s)} D_{kj}\big(i\theta\mathbf{A}_n(s)\big) e^{i(1-t)\theta\mathbf{A}_n(s)} dt \\ &= i\int_0^\theta \mathbf{U}(\tau, s) D_{kj}\Big(\mathbf{A}_n(s)\Big)\mathbf{U}(\theta-\tau, s) d\tau \\ &= i\int_0^\theta \mathbf{U}(\tau, s)\mathbf{P}_y\mathbf{B}_{ns}\mathbf{P}_y\mathbf{U}(\theta-\tau, s) d\tau, \end{aligned} \tag{3.102}$$

where

$$\begin{aligned} \mathbf{B}_{ns} &= \mathbf{e}_j\mathbf{e}_k^T\mathbf{H}^{-1}(s)\frac{1}{\sqrt{n}}\mathbf{X}(s) - \frac{1}{\sqrt{n}}\mathbf{X}^T(s)\mathbf{H}^{-1}(s)\mathbf{W}_n(s,k,j)\mathbf{H}^{-1}(s)\frac{1}{\sqrt{n}}\mathbf{X}(s) \\ &\quad + \frac{1}{\sqrt{n}}\mathbf{X}^T(s)\mathbf{H}^{-1}(s)\mathbf{e}_k\mathbf{e}_j^T. \end{aligned} \tag{3.103}$$

It follows from (3.92), (3.102) and the chain rule of calculating matrix derivatives that

$$D_{kj}(e_n(s,x)) = -xe_n(s,x)\int \hat{\phi}(\theta)\theta Tr\Big[\mathbf{U}(\theta, s)\mathbf{P}_y\mathbf{B}_{ns}\mathbf{P}_y\Big]d\theta, \tag{3.104}$$

where we also use the fact that

$$\int_0^\theta \mathbf{U}(\theta-\tau, s)\mathbf{U}(\tau, s) d\tau = \theta\mathbf{U}(\theta, s).$$

From (3.99) and (3.96) we have

$$\begin{aligned} D_{kj}(\Phi_{kj}^{(1)}) &= -\mathbf{e}_k^T\mathbf{H}^{-1}(s)\mathbf{W}_n(s,k,j)\mathbf{Q}_{ns}\mathbf{U}(\theta, s)\mathbf{P}_y\mathbf{e}_j e_n(s,x)\chi_{ns} \qquad (3.105) \\ &\quad + \mathbf{e}_k^T\mathbf{H}^{-1}(s)\mathbf{e}_k\mathbf{e}_j^T\mathbf{P}_y\mathbf{U}(\theta, s)\mathbf{P}_y\mathbf{e}_j e_n(s,x)\chi_{ns} \\ &\quad + \mathbf{e}_k^T\mathbf{Q}_{ns}\Big(D_{kj}(\mathbf{U}(\theta, s))\Big)\mathbf{P}_y\mathbf{e}_j e_n(s,x)\chi_{ns} \\ &\quad + \mathbf{e}_k^T\mathbf{Q}_{ns}\mathbf{U}(\theta, s)\mathbf{P}_y\mathbf{e}_j\Big(D_{kj}(e_n(s,x))\Big)\chi_{ns} \\ &\quad + \mathbf{e}_k^T\mathbf{Q}_{ns}\mathbf{U}(\theta, s)\mathbf{P}_y\mathbf{e}_j e_n(s,x) D_{kj}(\chi_{ns}), \end{aligned}$$

where $\mathbf{Q}_{ns} = \mathbf{H}^{-1}(s)\frac{1}{\sqrt{n}}\mathbf{X}(s)\mathbf{P}_y$.

Although there are many terms in the expansion of $D_{kj}(\Phi_{kj}^{(1)})$, from (3.105), (3.102) and (3.104) we see that each term must be products of some of the factors and their transposes below

$$\begin{aligned}&\mathbf{e}_k^T\mathbf{H}^{-1}(s)\mathbf{e}_k,\ \mathbf{e}_k^T\mathbf{H}^{-1}(s)\frac{1}{\sqrt{n}}\mathbf{X}(s),\ \mathbf{e}_j^T\mathbf{P}_x(s),\\ &\mathbf{P}_y,\ \mathbf{U}(\theta,s),\ \chi_{ns},\ \mathbf{e}_k,\ e_n(s,x),\ \mathbf{e}_j,\ D_{kj}(\chi_{ns}).\end{aligned}\tag{3.106}$$

By the facts that $|e_n(s,x)| \le 1$, $|\chi_{ns}| \le M$, $||\mathbf{P}_x(s)|| = ||\mathbf{P}_y|| = ||\mathbf{U}(\theta,s)|| = ||\mathbf{e}_k|| = ||\mathbf{e}_j|| = 1$ and (3.106), we conclude from (3.105) that

$$\begin{aligned}\left|D_{kj}\Phi_{kj}^{(1)}\right| &\le K||\lambda_{\min}||^{-r}||\mathbf{e}_k^T\mathbf{H}^{-1}(s)\frac{1}{\sqrt{n}}\mathbf{X}(s)||^d|\chi_{ns}|\\ &\quad +K||\mathbf{e}_k^T\mathbf{H}^{-1}(s)\frac{1}{\sqrt{n}}\mathbf{X}(s)|||D_{kj}(\chi_{ns})|\\ &\le \frac{K}{\lambda_{\min}^{r+d/2}}|\chi_{ns}| + \frac{K}{\lambda_{\min}}|D_{kj}(\chi_{ns})|,\end{aligned}\tag{3.107}$$

where r, d are some nonnegative integers independent of n, and $||\cdot||$ stands for the spectral norm of a matrix or the Euclidean norm of a vector. From the argument of (3.97), (3.91) and (3.85) we see

$$E\Big(\frac{1}{\lambda_{\min}^2}|D_{kj}(\chi_{ns})|^2\Big) \le K,\tag{3.108}$$

In view of (3.107), (3.110) and (3.91) we conclude that $E|D_{kj}\Phi_{kj}^{(1)}|^2$ is bounded.

We now claim that $E\big(D_{kj}^{\ell}\Phi_{kj}^{(1)}\big)^2$, $\ell = 2, 3, 4$ are bounded as well. Indeed, from (3.102) to (3.105) we see that each higher derivative of $E\big(D_{kj}^1\Phi_{kj}^{(1)}\big)$ must be a sum of the products of some of the derivatives $D_{kj}(\mathbf{U}(\theta,s))$, $D_{kj}(e_n(s,x)), D_{kj}(\mathbf{H}^{-1}(s))$, $D_{kj}(\frac{1}{\sqrt{n}}\mathbf{X}(s))$ and $D_{kj}^{\ell}(\chi_{ns})$. From (3.96)-(3.104) we see such derivatives must be formed by some of the factors listed in (3.106) as well as $D_{kj}^{\ell}(\chi_{ns})$. Here we would point out that the trace involved in (3.104) is handled in the way that $tr\mathbf{C}\mathbf{e}_k\mathbf{e}_j^T\mathbf{D} = \mathbf{e}_j^T\mathbf{D}\mathbf{C}\mathbf{e}_k$. Therefore, as in (3.107), we have for $\ell = 2, 3, 4$

$$\begin{aligned}\Big|D_{kj}^{\ell}\Phi_{kj}^{(\ell)}\Big| &\le K||\lambda_{\min}||^{-r_1}||\mathbf{e}_k^T\mathbf{H}^{-1}(s)\frac{1}{\sqrt{n}}\mathbf{X}(s)||^{d_1}\sum_{m=0}^{\ell}|D_{kj}^m(\chi_{ns})| \\ &\le \frac{K}{\lambda_{\min}^{r_1+d_1/2}}\sum_{m=0}^{\ell}|D_{kj}^m(\chi_{ns})|, \qquad (3.109)\end{aligned}$$

where r_1, d_1 are some nonnegative integers, independent of n. Again, from the argument of (3.97), (3.91) and (3.85) one can verify that

$$E\Big(\frac{1}{\lambda_{\min}^{r_1+d_1/2}}|D_{kj}^{\ell}(\chi_{ns})|^2\Big) \le K. \qquad (3.110)$$

Hence $E\Big|D_{kj}^{\ell}\Phi_{kj}^{(1)}\Big|^2 \le K$. Likewise one may verify that $E\Big|D_{kj}^{\ell}\Phi_{kj}^{(2)}\Big|^2$ is bounded. Summarizing the above we have proved that

$$E\Big|D_{kj}^{\ell}\Phi_{kj}^{(i)}\Big|^2 \le K, \quad \ell = 1, 2, 3, 4, \quad i = 1, 2. \qquad (3.111)$$

Consider $\xi_3^{(i)}$ in (3.101) now. Define the event

$$B = \Big(\lambda_{\min} \ge (1-\sqrt{c_1})^2/2\Big). \qquad (3.112)$$

Write

$$E\Big|X_{kjs}^5 D_{kj}^4\Phi_{kj}^{(i)}(vX_{kjs})\Big| = E\Big|X_{kjs}^5 D_{kj}^4\Phi_{kj}^{(i)}(vX_{kjs})\Big|I(B) + E\Big|X_{kjs}^5\Phi^{(4)}(vX_{kjs})\Big|I(B^c).$$

From (3.109) and (3.79) we see that on the event B

$$|X_{kjs}^5 D_{kj}^4\Phi_{kj}^{(i)}(vX_{kjs})| \le K\sqrt{n}\varepsilon|X_{kjs}^4| + K(\sqrt{n}\varepsilon)^5\sum_{m=1}^{4}|D_{kj}^m(\chi_{ns})|.$$

Moreover, as in (3.97) one may verify that

$$E\Big|(\sqrt{n}\varepsilon)^5\sum_{m=1}^{4}|D_{kj}^m(\chi_{ns})\Big| = O(n^{-k}).$$

While (3.111) and (3.86) imply

$$\begin{aligned} & E\Big|X_{kjs}^5 D_{kj}^4 \Phi_{kj}^{(i)}(vX_{kjs})\Big| I(B^c) \\ \leq\ & (E|D_{kj}^4 \Phi_{kj}^{(i)}(vX_{kjs})|^2)^{1/2}(E|X_{kjs}|^{4(p+2)}P(B^c))^{1/4} = O(n^{-k}). \end{aligned}$$

It follows that

$$|\xi_3^{(i)}| \leq \frac{K}{n^{5/2}} \sum_{k,j=1}^{n,p_1} \int_0^1 E\Big|X_{kjs}^5 D_{kj}^4 \Phi_{kj}^{(i)}(vX_{kjs})\Big| dv \leq K\varepsilon \to 0. \tag{3.113}$$

For $\ell = 1, 2, 3.4$, let $\mu_{\ell,kj}^{\varepsilon}(\mu_{\ell,kj})$ and $\kappa_{\ell,kj}^{\varepsilon}(\kappa_{\ell,kj})$ be the ℓ-th moment and ℓ-th cumulant of the truncated variables X_{kj}^{ε} (the original variables X_{kj}) respectively. Then

$$|\mu_{\ell,kj}^{\varepsilon} - \mu_{\ell,kj}| \leq KE|X_{kj}|^{\ell} I(|X_{kj}| > \varepsilon\sqrt{n}) \leq \frac{K}{(\sqrt{n}\varepsilon)^{4-\ell}} E|X_{kj}|^4 I(|X_{kj}| > \varepsilon\sqrt{n}). \tag{3.114}$$

It is well-known that the ℓ-th cumulant κ_ℓ can be written in terms of the moments μ_λ as

$$\kappa_\ell = \sum_\lambda c_\lambda \mu_\lambda, \tag{3.115}$$

where the sum is over all additive partitions λ of the set $\{1, \dots, \ell\}$, $\{c_\lambda\}$ are known coefficients and $\mu_\lambda = \prod_{\ell \in \lambda} \mu_\ell$. We then obtain from (3.114) and (3.115),

$$|\kappa_{\ell,kj}^{\varepsilon} - \kappa_{\ell,kj}| \leq \frac{K}{(\sqrt{n}\varepsilon)^{4-\ell}} E|X_{kj}|^4 I(|X_{kj}| > \varepsilon\sqrt{n}). \tag{3.116}$$

Recalling the definition of $T_{\ell\varepsilon}^{(1)}$ in (3.6.3.2), from (3.116) and (3.111) we may write

$$T_{\ell\varepsilon}^{(1)} = T_\ell^{(1)} + r_\ell^{(1)}, \tag{3.117}$$

where the error term $r_\ell^{(1)}$ satisfies

$$|r_\ell^{(1)}| \leq \frac{s^{(\ell-1)/2}}{\ell! n^{(\ell+1)/2}} \sum_{k,j=1}^{p_1,n} |\kappa_{\ell+1,kj}^{\varepsilon} - \kappa_{\ell+1,kj}||E(D_{kj}^{\ell} \Phi_{kj}^{\varepsilon(i)})| \leq \frac{K\varepsilon^{\ell-4}}{\sqrt{n}} \tag{3.118}$$

and $T_{\ell}^{(1)}$ is the analogue of $T_{\ell\varepsilon}^{(1)}$ with $\kappa_{\ell,kj}^{\varepsilon}$ replaced by $\kappa_{\ell,kj}$. Note that $T_0^{(1)} = T_3^{(1)} = 0$, $T_1^{(1)} = V_n^{(1)}$ because $\kappa_1 = \kappa_4 = 0$. In view of Lemma 13 below, $T_2^{(1)} = o(1)$, and hence

$$Q_n^{(1)} = V_n^{(1)} + \xi_3^{(1)} + o(1). \tag{3.119}$$

With the same proof as above, we can obtain

$$Q_n^{(2)} = V_n^{(2)} + \xi_3^{(2)} + o(1). \tag{3.120}$$

This, together with (3.119), (3.113) and (3.98), completes the proof of this theorem.

Lemma 13. *Under the assumptions of Theorem 3,*

$$T_2^{(i)} = \frac{s^{\frac{1}{2}}}{2n^{\frac{3}{2}}} \sum_{j,k=1}^{n,p_1} \kappa_{3,kj}^{\varepsilon} E\Big(D_{kj}^2(\Phi_{kj}^i)\Big) = o(1),\ i = 1, 2, \tag{3.121}$$

as $n \to \infty$.

By taking a further derivative of (3.105) we may obtain the expansion of $D_{kj}^2(\Phi_{kj}^i)$. However since such an expansion is rather complicated we do not list all the terms here. Note that each term of its expansion must be a product or a convolution of some of the following factors

$$C_1 = (\mathbf{V}_n(s))_{kj},\ C_2 = (\mathbf{V}_n(s)\frac{1}{\sqrt{n}}\mathbf{X}(s)\mathbf{H}^{-1}(s))_{kk},\ C_3 = (\mathbf{P}_x(s)\mathbf{P}_y\mathbf{U}\mathbf{P}_y)_{jj}, \tag{3.122}$$

$$C_4 = (\mathbf{P}_y\mathbf{U}\mathbf{P}_y)_{jj},\ C_5 = e_n(s,x),\ C_6 = (\mathbf{V}_n(s)\mathbf{P}_x(s))_{kj},\ C_7 = (\mathbf{H}^{-1}(s))_{kk},$$

$$C_{10} = (\mathbf{P}_x(s))_{jj},\ C_8 = (\mathbf{P}_x(s)\mathbf{P}_y\mathbf{U}\mathbf{P}_y\mathbf{P}_x(s))_{jj},$$

$$C_9 = (\frac{1}{\sqrt{n}}\mathbf{X}^T(s)\mathbf{H}^{-1}(s))_{jk},\ C_{11} = \chi_{ns},\ C_{12} = D_{kj}^{\ell}(\chi_{ns}), \ell = 1, 2,$$

where $\mathbf{V}_n(s) = \mathbf{H}^{-1}(s)\frac{1}{\sqrt{n}}\mathbf{X}(s)\mathbf{P}_y\mathbf{U}\mathbf{P}_y$ and $\mathbf{U}$ stands for $\mathbf{U}(\theta, s)$ or $\mathbf{U}(\theta - \tau, s)$. Moreover, each term of the expression of $D_{kj}^2(s)\Phi_{kj}^{(1)}(s)$ must contain

$C_5 = e_n(s,x)$ and at least one of C_{11} and C_{12}; and moreover, it contains at least one of C_1, C_6 and C_9. For example we see that $D_{kj}(e_n(s,x))$ contains C_1 or C_6 from (3.104) and $D_{kj}((\mathbf{H}^{-1}(s))_{kk})$ includes C_9 from (3.96).

Thus, to prove (3.121), it suffices to estimate the following term

$$\frac{1}{n^{3/2}}\sum_{k,j=1}^{p_1,n} E\Big(C_i^{r_i+1}\prod_{h\in D, h\neq i} C_h^{r_h}\Big) = o(1),\ i = 1,6,9, \tag{3.123}$$

where all $r_h, h \in D = \{1,\cdots,12\}$ are nonnegative integers, independent of n. As in (3.97) one may verify that (3.123) converges to zero if C_{12} is contained in (3.123). Below we consider only the case when C_{12} is not contained in (3.123) and as a result it must contain C_{11}.

We first prove (3.123) holds for the case when there are at least two of $C_i, i = 1,6,9$ contained in the expectation sign of (3.123). Moreover for concreteness we consider the case when C_1 and C_6 are both contained in (3.123) and all the remaining cases can be proved similarly. With $D_1 = \{2,\cdots,5,7,\cdots,10\}$ by the Schwartz inequality and arguments similar to (3.91) and (3.107) we obtain

$$\begin{aligned}
&\Big|\frac{1}{n^{3/2}}\sum_{k,j=1}^{p_1,n} E\Big(C_1^{r_1+1}C_6^{r_6+1}C_{11}\prod_{h\in D_1} C_h^{r_h}\Big)\Big| \\
\leq\ & \frac{K}{n^{3/2}}E\Big(\sum_{k,j=1}^{p_1,n}|(\mathbf{V}_n(s))_{kj}|^{2(r_1+1)}\sum_{k,j=1}^{p_1,n}|(\mathbf{V}_n(s)\mathbf{P}_x(s))_{kj}|^{2(r_6+1)}I(B)\Big)^{1/2} \\
&+\frac{Kn^{4+r_1+r_6+r_2+r_7}}{n^{3/2}}P\Big(\frac{M_2}{n}\leq\lambda_{\min}\leq\frac{(1-\sqrt{c_1})^2}{2}\Big) = O(\frac{1}{\sqrt{n}}),
\end{aligned} \tag{3.124}$$

where we also use the fact that recalling the definition of the event B in (3.112),

$$\sum_{k,j=1}^{p_1,n}|(\mathbf{V}_n(s))_{kj}|^{2(r_1+1)}I(B)\leq K\sum_{k,j=1}^{p_1,n}|(\mathbf{V}_n(s))_{kj}|^2 I(B)\leq Ktr(\mathbf{V}_n(s))^2 I(B)\leq nK.$$

If there is only one of $C_i, i = 1,6,9$ contained in (3.123) but its corresponding r_i being greater than zero, then repeating the argument of

(3.124) ensures that (3.123) holds. We now consider the case when one of $C_i, i = 1, 6, 9$ is contained in (3.123) but its corresponding r_i equals zero. For concreteness we consider C_1 contained in (3.123) and the remaining cases can be proved similarly. Let $D_2 = \{2, \cdots, 5, 7, 8, 10\}$. By the Schwartz inequality

$$
\begin{aligned}
&\Big|\frac{1}{n^{3/2}}\sum_{k,j=1}^{p_1,n} E\Big(C_1 C_{11}\prod_{h\in D_2} C_h^{r_h}\Big)\Big|^2 \\
\leq\ & \frac{K}{n^3}E\Big[\sum_{j=1}|C_3^{r_3}C_4^{r_4}C_8^{r_8}C_{10}^{r_{10}}|^2\sum_{j=1}\Big|\sum_k(\mathbf{V}_n(s))_{kj}C_2^{r_2}C_7^{r_7}\Big|^2 I(B)\Big] \\
&+\frac{Kn^{8+2r_1+2r_2+2r_7}}{n^3}P\Big(\frac{M_2}{n}\leq\lambda_{\min}\leq\frac{(1-\sqrt{c_1})^2}{2}\Big) \\
\leq\ & \frac{K}{n^2}E\Big[\sum_{j=1}\sum_{k_1,k_2}(\mathbf{V}_n(s))_{k_1 j}(\bar{\mathbf{V}}_n(s))_{k_2 j}C_{2k_1k_1}^{r_2}C_{7k_1k_1}^{r_7}\bar{C}_{2k_2k_2}^{r_2}\bar{C}_{7k_2k_2}^{r_7}I(B)\Big] \qquad (3.125)\\
&+\frac{Kn^{8+2r_1+2r_2+2r_7}}{n^3}P\Big(\frac{M_2}{n}\leq\lambda_{\min}\leq\frac{(1-\sqrt{c_1})^2}{2}\Big) = O(\frac{1}{\sqrt{n}}), \qquad (3.126)
\end{aligned}
$$

where we use C_{2kk} and $C_{7kk}, k = k_1, k_2$, respectively, to denote C_2 and C_7 to emphasize their dependence on k and the notation $(\bar{\cdot})$ denotes its corresponding complex conjugate. As for (3.125) we use the following fact that

$$
(3.125) = \frac{K}{n^2}E\Big[\sum_{k_1,k_2}(\bar{\mathbf{V}}_n(s)\mathbf{V}_n^T(s))_{k_2k_1}C_{2k_1k_1}^{r_2}C_{7k_1k_1}^{r_7}\bar{C}_{2k_2k_2}^{r_2}\bar{C}_{7k_2k_2}^{r_7}I(B)\Big]
$$

$$
\begin{aligned}
&\leq \frac{K}{n^2}E\Big[\sum_{k_1}|C_{2k_1k_1}^{r_2}C_{7k_1k_1}^{r_7}|^2\sum_{k_1}\Big|\sum_{k_2}(\bar{\mathbf{V}}_n(s)\mathbf{V}_n^T(s))_{k_2k_1}\bar{C}_{2k_2k_2}^{r_2}\bar{C}_{7k_2k_2}^{r_7}\Big|^2 I(B)\Big]^{1/2} \\
&\leq \frac{K}{n^{3/2}}E\Big[\sum_{k_1}\sum_{k_2,k_3}(\bar{\mathbf{V}}_n(s)\mathbf{V}_n^T(s))_{k_2k_1}(\mathbf{V}_n^*(s)\mathbf{V}_n(s))_{k_3k_1}\bar{C}_{2k_2k_2}^{r_2}\bar{C}_{7k_2k_2}^{r_7}C_{2k_3k_3}^{r_2}C_{7k_3k_3}^{r_7}I(B)\Big]^{1/2} \\
&= \frac{K}{n^{3/2}}E\Big[\sum_{k_2,k_3}(\bar{\mathbf{V}}_n(s)(\mathbf{V}_n^T(s))^2\bar{\mathbf{V}}_n(s))_{k_2k_3}\bar{C}_{2k_2k_2}^{r_2}\bar{C}_{7k_2k_2}^{r_7}C_{2k_3k_3}^{r_2}C_{7k_3k_3}^{r_7}I(B)\Big]^{1/2} \leq \frac{K}{\sqrt{n}},
\end{aligned}
$$

where $\mathbf{V}_n^*(s)$ stands for the complex conjugate transpose of $\mathbf{V}_n(s)$. Therefore (3.123) holds for all cases and the proof of Lemma 13 is complete.

3.6.4 Proof of Theorem 4

3.6.4.1 The Gaussian case

The CLT under the case of $p_2 \geq n$ has been discussed in the proof of Theorem 2. Consider $c_2' \in (0,1)$ next.

We remind readers that we below use the same notations as those in Theorem 2. Recall $q_n = \frac{p_2}{n-p_2}$. From (3.59) we can see that the statistic (3.10) can be expressed as

$$\int \phi(\lambda) dG^{(2)}_{p_1,p_2}(\lambda) = \int \phi(\frac{q_n\mu}{1+q_n\mu}) dp_1[F^{\mathbf{S}_1\mathbf{S}_{2t}^{-1}}(\mu) - \tilde{F}_{y_{1n},y_{2n}}(\mu)], \quad (3.127)$$

where $\tilde{F}_{y_{1n},y_{2n}}(\mu)$ is obtained from $\tilde{F}_{y_1,y_2}(\mu)$, whose stieltjes transform is defined in (3.5), with the substitution of (y_{n1}, y_{n2}) for (y_1, y_2). Here $y_{n1} = \frac{p_1}{p_2}$ and $y_{n_2} = \frac{p_1}{n-p_2}$.

From (3.127), it suffices to provide the CLT for generalized F-matrix $\mathbf{K}_n = \mathbf{S}_1\mathbf{S}_{2t}^{-1}$. When $t = 0$, the CLT of the linear spectral statistics of $\mathbf{K}_n$ is provided in Zheng (2012). Following a line similar to the proof of Theorem 3.1 of Zheng (2012), we next provide the CLT for the linear spectral statistics of the matrix $\mathbf{K}_n$ in the case of $t > 0$.

Let $\mathbf{n} = (n_1, n_2)$ and $\mathbf{y} = (y_1, y_2)$ with $n_1 = p_1$ and $n_2 = n - p_2$. The Stieltjes transforms of the ESD and LSD of the matrix $\mathbf{S}_1\mathbf{S}_{2t}^{-1}$ are denoted by $m_{\mathbf{n}}(z)$ and $m_{\mathbf{y}}(z)$ respectively while those of the ESD and LSD of the matrix $\frac{1}{p_2}\mathbf{W}_1^T\mathbf{S}_{2t}^{-1}\mathbf{W}_1$ are denoted by $\underline{m}_{\mathbf{n}}(z)$ and $\underline{m}_{\mathbf{y}}(z)$ respectively. The ESD and LSD of $\mathbf{S}_{2t}$ are written as F_{n_2t} and F_{y_2t} respectively while those of $\mathbf{S}_{2t}^{-1}$ are written as $H_{n_2t}(x)$ and $H_{y_2t}(x)$ respectively. The Stieltjes transforms of F_{n_2t} and F_{y_2t} are denoted by $m_{n_2t}(z)$ and $m_{y_2t}(z)$ respectively. The Stieltjes transforms of ESD and LSD of the matrix $\mathbf{S}_2 = \frac{1}{n-p_2}\mathbf{W}_2\mathbf{W}_2^T$ are written as $m_{n_2}(z)$ and $m_{y_2}(z)$ respectively while those of the ESD and LSD of the matrix $\frac{1}{n-p_2}\mathbf{W}_2^T\mathbf{W}_2$ are denoted by $\underline{m}_{n_2}(z)$ and $\underline{m}_{y_2}(z)$ respectively.

Moreover, $m_{\mathbf{y_n}}, \underline{m}_{\mathbf{y_n}}$ are obtained from $m_{\mathbf{y}}, \underline{m}_{\mathbf{y}}$ respectively with $\mathbf{y} = (y_1, y_2)$ replaced by $\mathbf{y_n} = (y_{1n}, y_{2n})$. Also $F_{y_{n_2}t}, m_{y_{n_2}t}, \underline{m}_{y_{n_2}t}, F_{y_{n_2}}, m_{y_{n_2}}$ and $\underline{m}_{y_{n_2}}$ are obtained from $F_{y_2t}, m_{y_2t}, \underline{m}_{y_2t}, F_{y_2}, m_{y_2}$ and $\underline{m}_{y_2}$ with y_2 replaced by y_{2n}.

Some of the Stieltjes transforms and ESDs above have the following relations:

$$\underline{m}_{\mathbf{n}}(z) = -\frac{1-y_{n1}}{z} + y_{n1}m_{\mathbf{n}}(z), \quad \underline{m}_{\mathbf{y}}(z) = -\frac{1-y_1}{z} + y_1 m_{\mathbf{y}}(z); \tag{3.128}$$

and for all $x > 0$,

$$H_{n_2t}(x) = 1 - F_{n_2t}(\frac{1}{x}), \quad H_{y_2t}(x) = 1 - F_{y_2t}(\frac{1}{x}).$$

This, together with Theorem 4.3 of Bai and Silverstein (2009), indicates that $\underline{m}_{\mathbf{y}}(z)$ satisfies the following equation

$$z = -\frac{1}{\underline{m}_{\mathbf{y}}(z)} + \int \frac{y_1 dF_{y_2t}(x)}{x + \underline{m}_{\mathbf{y}}(z)}. \tag{3.129}$$

Replacing $F_{y_2t}(x)$ by $F_{y_{n_2t}}(x)$ we have a similar expression (see (6.2.15) of Bai and Silverstein (2009) as well)

$$z = -\frac{1}{\underline{m}_{\mathbf{y}_n}} + \int \frac{y_{n_1} dF_{y_{n_2t}}(x)}{x + \underline{m}_{\mathbf{y_n}}}. \tag{3.130}$$

Write

$$n_1[\underline{m}_{\mathbf{n}}(z) - \underline{m}_{\mathbf{y_n}}(z)] = n_1[\underline{m}_{\mathbf{n}}(z) - \underline{m}^{y_{n_1},H_{n_2t}}(z)] + n_1[\underline{m}^{y_{n_1},H_{n_2t}}(z) - \underline{m}_{\mathbf{y_n}}(z)], \tag{3.131}$$

where $\underline{m}^{\{y_{n_1},H_{n_2t}\}}(z)$ is the unique root to the following equation

$$z = -\frac{1}{\underline{m}^{\{y_{n_1},H_{n_2t}\}}} + \int \frac{y_{n_1} dF_{n_2t}(x)}{x + \underline{m}^{\{y_{n_1},H_{n_2t}\}}}. \tag{3.132}$$

Roughly speaking, $\underline{m}^{y_{n_1},H_{n_2t}}(z)$ is the Stieltjes transform of the LSD of $\frac{1}{n}\mathbf{W}_1^T\mathbf{S}_{2t}^{-1}\mathbf{W}_1$ when $\mathbf{W}_2$ is given.

Step 1: Given $\mathbf{W}_2$, consider the conditional distribution of

$$n_1[\underline{m}_{\mathbf{n}} - \underline{m}^{\{y_{n_1}, H_{n_2 t}\}}(z)]. \tag{3.133}$$

For simplicity, write $\underline{m}_{\mathbf{y}}(z)$ as $\underline{m}(z)$. By Lemma 9.11 of Bai and Silverstein (2009), we can obtain the conditional distribution of (3.133) given $\mathbf{W}_2$ converges to a Gaussian process $M_1(z)$ on some contour $\mathcal{C}$ (see Lemma 9.11 of Bai and Silverstein (2009)) with mean function

$$E(M_1(z)|\mathbf{W}_2) = \frac{y_1 \int \underline{m}(z)^3 x[x + \underline{m}(z)]^{-3} dF_{y_2 t}(x)}{[1 - y_1 \int \underline{m}(z)^2 (x + \underline{m}(z))^{-2} dF_{y_2 t}(x)]^2} \tag{3.134}$$

for $z \in \mathcal{C}$ and covariance function

$$Cov(M_1(z_1), M_1(z_2)|\mathbf{W}_2) = 2\Big(\frac{\underline{m}'(z_1)\underline{m}'(z_2)}{(\underline{m}(z_1) - \underline{m}(z_2))^2} - \frac{1}{(z_1 - z_2)^2}\Big) \tag{3.135}$$

for $z_1, z_2 \in \mathcal{C}$.

Step 2: Consider the limit distribution of

$$n_1[\underline{m}^{\{y_{n_1}, H_{n_2 t}\}}(z) - \underline{m}_{\mathbf{y}_n}(z)]. \tag{3.136}$$

By the definition of the Stieltjes transform, rewrite the equations of (3.130) and (3.132) as

$$z = -\frac{1}{\underline{m}_{\mathbf{y}_n}} + y_{n_1} m_{y_{n_2 t}}(-\underline{m}_{\mathbf{y}_n}),\ z = -\frac{1}{\underline{m}^{\{y_{n_1}, H_{n_2 t}\}}} + y_{n_1} m_{n_2 t}\big(-\underline{m}^{\{y_{n_1}, H_{n_2 t}\}}\big). \tag{3.137}$$

Taking a difference of the above two identities we obtain

$$\begin{aligned} 0 &= \frac{\underline{m}^{\{y_{n_1}, H_{n_2 t}\}} - \underline{m}_{\mathbf{y}_n}}{\underline{m}_{\mathbf{y}_n} \underline{m}^{\{y_{n_1}, H_{n_2 t}\}}} + y_{n1}[m_{n_2 t}(-\underline{m}^{\{y_{n_1}, H_{n_2 t}\}}) - m_{n_2 t}(-\underline{m}_{\mathbf{y}_n}) \\ &\quad + m_{n_2 t}(-\underline{m}_{\mathbf{y}_n}) - m_{y_{n_2 t}}(-\underline{m}_{\mathbf{y}_n})] \\ &= \frac{\underline{m}^{\{y_{n_1}, H_{n_2 t}\}} - \underline{m}_{\mathbf{y}_n}}{\underline{m}_{\mathbf{y}_n} \underline{m}^{\{y_{n_1}, H_{n_2 t}\}}} - y_{n_1} \int \frac{(\underline{m}^{\{y_{n_1}, H_{n_2 t}\}} - \underline{m}_{\mathbf{y}_n}) dF_{n_2 t}(x)}{(x + \underline{m}^{\{y_{n_1}, H_{n_2 t}\}})(x + \underline{m}_{\mathbf{y}_n})} \\ &\quad + y_{n_1}[m_{n_2 t}(-\underline{m}_{\mathbf{y}_n}) - m_{y_{n_2 t}}(-\underline{m}_{\mathbf{y}_n})]. \end{aligned}$$

From the above equality, we can obtain

$$\begin{aligned} & n_1[\underline{m}^{\{y_{n_1},H_{n_2t}\}}(z)-\underline{m}_{\mathbf{y}_n}(z)] \\ = & -y_{n_1}\underline{m}_{\mathbf{y}_n}\underline{m}^{\{y_{n_1},H_{n_2t}\}}\frac{n_1[m_{n_2t}(-\underline{m}_{\mathbf{y}_n})-m_{y_{n_2t}}(-\underline{m}_{\mathbf{y}_n})]}{1-y_{n_1}\int\frac{\underline{m}_{\mathbf{y}_n}\underline{m}^{\{y_{n_1},H_{n_2t}\}}dF_{n_2t}(x)}{(x+\underline{m}_{\mathbf{y}_n})(x+\underline{m}^{\{y_{n_1},H_{n_2t}\}})}}. \end{aligned} \quad (3.138)$$

From the fact that $\underline{m}_{\mathbf{y}_n}(z)\to\underline{m}(z)$ and Theorem 3.9 of Billingsley (1999), the limiting distribution of

$$p_1[m_{n_2t}(-\underline{m}_{\mathbf{y}_n})-m_{y_{n_2t}}(-\underline{m}_{\mathbf{y}_n})]$$

is the same as that of

$$p_1[m_{n_2t}(-\underline{m})-m_{y_{n_2t}}(-\underline{m})].$$

Recall the definition of $g(z)$ before Theorem 4. By Theorem 6 in Chapter 4, we see that $n_1[m_{n_2t}(-\underline{m}(z))-m_{y_{n_2t}}(-\underline{m}(z))]$ converges to a Gaussian process $M_2(\cdot)$ on $z\in\mathcal{C}$ with mean function

$$\begin{aligned} \mathbb{E}M_2(z) = & \frac{y_2\varpi^2(-\underline{m}(z))m_3(-\underline{m}(z))+y_2^2\varpi^4(-\underline{m}(z))m'_{y_2t}(-\underline{m}(z))m_3(-\underline{m}(z))}{1-y_2\varpi^2(-\underline{m}(z))m_2(-\underline{m}(z))} \\ & -\frac{y_2^2\varpi^3(-\underline{m}(z))m'_{y_2t}(-\underline{m}(z))m_2(-\underline{m}(z))}{1-y_2\varpi^2(-\underline{m}(z))m_2(-\underline{m}(z))} \end{aligned}$$

and covariance

$$\begin{aligned} Cov(M_2(z_1),M_2(z_2)) = & -\frac{2}{(-\underline{m}(z_2)+\underline{m}(z_1))^2} \\ & +\frac{2[1+g(z_1)+g(z_2)+g(z_1)g(z_2)]}{[-\underline{m}(z_2)+\underline{m}(z_1)+s(-\underline{m}(z_1),-\underline{m}(z_2))]^2}, \end{aligned}$$

Since $\frac{-\underline{m}_{\mathbf{y}_n}(z)\underline{m}^{\{y_{n_1},H_{n_2t}\}}(z)}{1-y_{n_1}\int\frac{\underline{m}_{\mathbf{y}_n}(z)\underline{m}^{\{y_{n_1},H_{n_2t}\}}(z)dF_{n_2t}(x)}{(x+\underline{m}_{\mathbf{y}_n}(z))(x+\underline{m}^{\{y_{n_1},H_{n_2t}\}})}}$ converges to $h(z)=\frac{-\underline{m}^2(z)}{1-y_1\underline{m}^2(z)\int\frac{dF_{y_2t}(x)}{(x+\underline{m}(z))^2}}$, we have (3.138) converges weakly to a Gaussian process $M_3(\cdot)=h(z)M_2(z)$ with mean $E(M_3(z))=h(z)EM_2(z)$ and covariance

$$Cov(M_3(z_1),M_3(z_2))=h(z_1)h(z_2)Cov(M_2(z_1),M_2(z_2)).$$

Since the limit of

$$n_1[\underline{m}_{\mathbf{n}}(z) - \underline{m}^{\{y_{n_1},H_{n_2t}\}}(z)]$$

conditioning on $\mathbf{W}_2$ is independent of the ESD of S_{n_2}, the limits of

$$n_1[\underline{m}_{\mathbf{n}}(z) - \underline{m}^{\{y_{n_1},H_{n_2t}\}}(z)] \ and \ n_1[\underline{m}^{\{y_{n_1},H_{n_2t}\}}(z) - \underline{m}_{\mathbf{y}_n}(z)]$$

are asymptotically independent. Therefore $n_1[\underline{m}_{\mathbf{n}}(z) - \underline{m}_{\mathbf{y}_n}(z)]$ converges weakly to $M_1(z) + M_3(z)$ with mean function

$$\begin{aligned}
&E(M_1(z)+M_3(z)) = \frac{y_1 \int \underline{m}(z)^3 x[x+\underline{m}(z)]^{-3} dF_{y_2t}(x)}{[1 - y_1 \int \underline{m}(z)^2 (x+\underline{m}(z))^{-2} dF_{y_2t}(x)]^2}\\
&+h(z)\frac{y_2\varpi^2(-\underline{m}(z))m_3(-\underline{m}(z)) + y_2^2\varpi^4(-\underline{m}(z))m_{y_2t}^{'}(-\underline{m}(z))m_3(-\underline{m}(z))}{1 - c\varpi^2(-\underline{m}(z))m_2(-\underline{m}(z))}\\
&-h(z)\frac{y_2^2\varpi^3(-\underline{m}(z))m_{y_2t}^{'}(-\underline{m}(z))m_2(-\underline{m}(z))}{1 - y_2\varpi^2(-\underline{m}(z))m_2(-\underline{m}(z))}
\end{aligned} \tag{3.139}$$

and covariance function

$$\begin{aligned}
&Cov(M_1(z_1)+M_3(z_1), M_1(z_2)+M_3(z_2)) = 2\Big(\frac{\underline{m}^{'}(z_1)\underline{m}^{'}(z_2)}{(\underline{m}(z_1)-\underline{m}(z_2))^2} - \frac{1}{(z_1-z_2)^2}\Big)\\
&-\frac{2h(z_1)h(z_2)}{(-\underline{m}(z_2)+\underline{m}(z_1))^2} + \frac{h(z_1)h(z_2)2[1+g(z_1)+g(z_2)+g(z_1)g(z_2)]}{[-\underline{m}(z_2)+\underline{m}(z_1)+s(-\underline{m}(z_1),-\underline{m}(z_2))]^2}.
\end{aligned} \tag{3.140}$$

By the Cauchy integral formula, we have with probability one for all n large

$$\int f(x) dG_{p_1,p_2}^{(2)}(x) = -\frac{1}{2\pi i}\int f(z) m_G(z) dz. \tag{3.141}$$

Then

$$\Big(\int f_1(x) dG_{p_1,p_2}^{(2)}(x), \ldots, \int f_k(x) dG_{p_1,p_2}^{(2)}(x)\Big)$$

converges to a Gaussian vector $(X_{f_1}, \ldots, X_{f_k})$ where

$$EX_{f_i} = -\frac{1}{2\pi i}\oint f_i(z) E(M_1(z)+M_3(z)) dz \tag{3.142}$$

and

$$Cov(X_{f_i}, X_{f_j}) = -\frac{1}{4\pi^2} \oint \oint f_i(z) f_j(z) Cov(M_1(z_1) + M_3(z_1), M_1(z_2) + M_3(z_2)) dz_1 dz_2. \tag{3.143}$$

As for the non-Gaussian case, under the assumption that $EX_{11}^4 = 3$, one can verify that the CLT is the same as that in the Gaussian case by repeating the method in Section 3.6.3.2 (replacing $\mathbf{P}_x$ there by $\mathbf{P}_{tx}$ in (3.65)). We omit the details here.

3.6.5 The proof of Theorem 5

Set

$$D^{(i)} = p_1 \int \lambda d\Big(F_{\mathbb{H}_1}^{\mathbf{R}_{xy}^{(i)}}(\lambda) - F_{\mathbb{H}_0}^{\mathbf{R}_{xy}^{(i)}}(\lambda)\Big),$$

where $\mathbf{R}_{xy}^{(1)}$ represents the matrix $\mathbf{S}_{xy}$ while $\mathbf{R}_{xy}^{(2)}$ represents the matrix $\mathbf{T}_{xy}$; and $F_{\mathbb{H}_0}^{\mathbf{R}_{xy}^{(i)}}, F_{\mathbb{H}_1}^{\mathbf{R}_{xy}^{(i)}}$ stand for the ESDs of $\mathbf{R}_{xy}^{(i)}$ under $\mathbb{H}_0$ and $\mathbb{H}_1$, respectively. The power can be then calculated as

$$\begin{aligned} \beta_n &= P\Big(R_n^{(i)} > z_{1-\alpha} \ or\ R_n^{(i)} < z_\alpha \Big| \mathbb{H}_1\Big) \\ &= P\Big(D^{(i)} + R_n^{(i)0} > z_{1-\alpha} \ or\ D^{(i)} + R_n^{(i)0} < z_\alpha \Big| \mathbb{H}_1\Big) \\ &= P\Big(R_n^{(i)0} > z_{1-\alpha} - D^{(i)} \ or\ R_n^{(i)0} < z_\alpha - D^{(i)} \Big| \mathbb{H}_1\Big), \end{aligned} \tag{3.144}$$

where $R_n^{(i)0}$ stands for $R_n^{(i)}$ under $\mathbb{H}_0$. Under the condition (3.24), we have

$$\beta_n \to 1, \quad as \quad n \to \infty.$$

Table 3.1: *Empirical sizes of the proposed test S_n and the re-normalized likelihood ratio test MLR_n at 0.05 significance level for DGP(a) and DGP(b).*

(p_1, p_2, n)	S_n DGP(a)	S_n DGP(b)	MLR_n DGP(a)	MLR_n DGP(b)
(10,20,40)	0.0458	0.0461	0.0481	0.0490
(20,30,60)	0.0480	0.0488	0.0440	0.0448
(30,60,120)	0.0475	0.0480	0.0530	0.0520
(40,80,160)	0.0464	0.0466	0.0420	0.0420
(50,100,200)	0.0503	0.0504	0.0487	0.0500
(60,120,240)	0.0490	0.0490	0.0574	0.0572
(70,140,280)	0.0524	0.0520	0.0570	0.0582
(80,160,320)	0.0500	0.0500	0.0632	0.0583
(90,180,360)	0.0521	0.0511	0.0559	0.0580
(100,200,400)	0.0501	0.0503	0.0482	0.0589
(110,220,440)	0.0504	0.0500	0.0440	0.0590
(120,240,480)	0.0513	0.0511	0.0400	0.0432
(130,260,520)	0.0511	0.0511	0.0520	0.0560
(140,280,560)	0.0469	0.0474	0.0582	0.0580
(150,300,600)	0.0495	0.0500	0.0590	0.0593
(160,320,640)	0.0514	0.0517	0.0437	0.0559
(170,340,680)	0.0498	0.0500	0.0428	0.0430
(180,360,720)	0.0509	0.0510	0.0580	0.0577
(190,380,760)	0.0488	0.0485	0.0388	0.0499
(200,400,800)	0.0491	0.0491	0.0462	0.0499
(210,420,840)	0.0491	0.0500	0.0450	0.0555
(220,440,880)	0.0515	0.0510	0.0572	0.0588
(230,460,920)	0.0493	0.0498	0.0470	0.0488
(240,480,960)	0.0482	0.0479	0.0521	0.0561
(250,500,1000)	0.0452	0.0450	0.0527	0.0545

Table 3.2: *Empirical sizes of the proposed test T_n at 0.05 significance level for DGP(a) and DGP(b).*

(p_1, p_2, n)	T_n DGP(a)	T_n DGP(b)
(100,50,60)	0.0582	0.0579
(140,70,80)	0.0591	0.0571
(180,90,100)	0.0549	0.0568
(200,100,120)	0.0561	0.0558
(240,120,130)	0.0571	0.0572
(280,140,150)	0.0540	0.0569
(320,160,170)	0.0551	0.0559
(360,180,190)	0.0542	0.0572
(400,190,200)	0.0571	0.0553
(440,220,230)	0.0532	0.0561
(480,240,250)	0.0540	0.0557

*The parameter t in the statistic T_n takes value of 0.5.

Table 3.3: *Empirical powers of the proposed test S_n at 0.05 significance level for factor models.*

(p_1, p_2, n)	r=1	r=2	r=3	r=4
(10,20,40)	0.2690	0.6460	0.9420	0.9980
(30,60,120)	0.2930	0.8010	0.9760	0.9990
(50,100,200)	0.3110	0.7650	0.9770	1.0000
(70,140,280)	0.3240	0.7710	0.9830	0.9980
(90,180,360)	0.3450	0.7940	0.9870	1.0000
(110,220,440)	0.3330	0.7980	0.9800	0.9990
(130,260,520)	0.3460	0.7820	0.9780	0.9990
(150,300,600)	0.3510	0.7980	0.9720	0.9990
(170,340,680)	0.3250	0.7780	0.9750	1.0000
(190,380,760)	0.3480	0.7810	0.9810	1.0000
(210,420,840)	0.3210	0.7900	0.9700	1.0000
(230,460,920)	0.3300	0.7810	0.9790	1.0000
(250,500,1000)	0.3370	0.7890	0.9790	1.0000

*The powers are under the alternative hypothesis that $\mathbf{x}$ and $\mathbf{y}$ satisfy the factor model (3.44). r is the number of factors.

Table 3.4: *Empirical powers of the proposed test T_n at 0.05 significance level for factor models.*

(p_1, p_2, n)	r=1	r=2	r=3	r=4
(10,20,40)	0.3150	0.7440	0.9500	0.9830
(30,60,120)	0.4230	0.8550	0.9740	1.0000
(50,100,200)	0.3990	0.8760	0.9890	1.0000
(70,140,280)	0.4120	0.8690	0.9990	0.9980
(90,180,360)	0.4050	0.8820	0.9840	1.0000
(110,220,440)	0.4000	0.8740	0.9850	0.9990
(130,260,520)	0.3990	0.8910	0.9880	0.9990
(150,300,600)	0.4110	0.8570	0.9870	0.9990
(170,340,680)	0.4100	0.8800	0.9920	1.0000
(190,380,760)	0.3980	0.8690	0.9860	1.0000
(210,420,840)	0.3490	0.8790	0.9770	1.0000
(230,460,920)	0.3990	0.8730	0.9730	1.0000
(250,500,1000)	0.4100	0.8770	0.9800	1.0000

*The powers are under the alternative hypothesis that $\mathbf{x}$ and $\mathbf{y}$ satisfy the factor model (3.44). r is the number of factors.

Table 3.5: *Empirical powers of the proposed test S_n at 0.05 significance level for $\mathbf{x}$ and $\mathbf{y}$ with ARCH(1) dependent type.*

(p_1, p_2, n)	$(0.9, 0.1)$	$(0.8, 0.2)$	$(0.7, 0.3)$	$(0.6, 0.4)$	$(0.5, 0.5)$
(10,20,40)	0.3480	0.4670	0.6380	0.7650	0.8500
(30,60,120)	0.4840	0.8090	0.9820	0.9990	1.0000
(50,100,200)	0.6190	0.9730	1.0000	1.0000	1.0000
(70,140,280)	0.7020	0.9980	1.0000	1.0000	1.0000
(90,180,360)	0.7900	1.0000	1.0000	1.0000	1.0000
(110,220,440)	0.8620	1.0000	1.0000	1.0000	1.0000
(130,260,520)	0.8970	1.0000	1.0000	1.0000	1.0000
(150,300,600)	0.9440	1.0000	1.0000	1.0000	1.0000
(170,340,680)	0.9520	1.0000	1.0000	1.0000	1.0000
(190,380,760)	0.9810	1.0000	1.0000	1.0000	1.0000
(210,420,840)	0.9880	1.0000	1.0000	1.0000	1.0000
(230,460,920)	0.9950	1.0000	1.0000	1.0000	1.0000
(250,500,1000)	0.9980	1.0000	1.0000	1.0000	1.0000

*The powers are under the alternative hypothesis that $Y_{it} = Z_{it}\sqrt{\alpha_0 + \alpha_1 X_{it}^2}, i = 1, 2, \ldots, p_1; Y_{jt} = Z_{jt}, j = p_1 + 1, \ldots, p_2$. The pair of two numbers in this table is the value of (α_0, α_1).

Table 3.6: *Empirical powers of the proposed test T_n at 0.05 significance level for* $\mathbf{x}$ *and* $\mathbf{y}$ *with ARCH(1) dependent type.*

(p_1, p_2, n)	$(0.9, 0.1)$	$(0.8, 0.2)$	$(0.7, 0.3)$	$(0.6, 0.4)$	$(0.5, 0.5)$
(100,50,60)	0.5620	05720	0.6880	0.8030	0.9430
(140,70,80)	0.6330	0.7590	0.8200	0.9170	0.9590
(180,90,100)	0.7190	0.8240	0.9560	0.9920	1.0000
(200,100,110)	0.7990	0.8370	0.9890	1.0000	1.0000
(240,120,130)	0.8810	0.9450	1.0000	1.0000	1.0000
(280,140,150)	0.9470	0.9780	1.0000	1.0000	1.0000
(320,160,170)	0.9790	0.9920	1.0000	1.0000	1.0000
(360,180,190)	0.9890	0.9990	1.0000	1.0000	1.0000
(400,200,210)	0.9860	0.9990	1.0000	1.0000	1.0000
(440,220,230)	0.9930	1.0000	1.0000	1.0000	1.0000
(480,240,250)	0.9970	0.9990	1.0000	1.0000	1.0000

*The powers are under the alternative hypothesis that $Y_{it} = Z_{it}\sqrt{\alpha_0 + \alpha_1 X_{it}^2}, i = 1, 2, \ldots, p_1; Y_{jt} = Z_{jt}, j = p_1 + 1, \ldots, p_2$. The pair of two numbers in this table is the value of (α_0, α_1). The parameter t in the statistic T_n takes value of 0.5.

Table 3.7: *Empirical powers of the proposed test S_n at 0.05 significance level for uncorrelated but dependent case.*

(p_1, p_2, n)	$\omega = 4$	$\omega = 10$
(10,20,40)	0.8140	0.9690
(30,60,120)	0.8200	0.9510
(50,100,200)	0.8220	0.9600
(70,140,280)	0.8100	0.9610
(90,180,360)	0.8210	0.9640
(110,220,440)	0.8110	0.9670
(130,260,520)	0.8320	0.9740
(150,300,600)	0.8420	0.9740
(170,340,680)	0.8450	0.9760
(190,380,760)	0.8580	0.9680
(210,420,840)	0.8420	0.9670
(230,460,920)	0.8440	0.9810
(250,500,1000)	0.8620	0.9810

*The powers are under the alternative hypothesis that $Y_{it} = X_{it}^{\omega}, i = 1, 2, \ldots, p_1$ and $Y_{jt} = \varepsilon_{jt}, j = p_1 + 1, \ldots, p_2; t = 1, \ldots, n$, where $\varepsilon_{jt}, j = p_1 + 1, \ldots, p_2; t = 1, \ldots, n$ are standard normal distributed and independent with X_{it} and $\omega = 4, 10$.

Table 3.8: *Empirical powers of the proposed test T_n at 0.05 significance level for uncorrelated but dependent case.*

(p_1, p_2, n)	$\omega = 4$	$\omega = 10$
(100,50,60)	0.6270	0.8130
(140,70,80)	0.6920	0.8430
(180,90,100)	0.7010	0.8380
(200,100,110)	0.7920	0.8460
(240,120,130)	0.8240	0.9330
(280,140,150)	0.8660	0.9650
(320,160,170)	0.9040	0.9780
(360,180,190)	0.9060	0.9830
(400,200,210)	0.9310	0.9920
(440,220,230)	0.9690	1.0000
(480,240,250)	0.9920	1.0000

*The powers are under the alternative hypothesis that $Y_{it} = X_{it}^{\omega}, i = 1, 2, \ldots, p_1$ and $Y_{jt} = \varepsilon_{jt}, j = p_1 + 1, \ldots, p_2; t = 1, \ldots, n$, where $\varepsilon_{jt}, j = p_1 + 1, \ldots, p_2; t = 1, \ldots, n$ are standard normal distributed and independent with X_{it} and $\omega = 4, 10$. The parameter t in the statistic T_n takes value of 0.5.

Table 3.9: *P-values for* (p_1, p_2) *companies from basic industry section and capital goods section of NYSE.*

P-values	(p_1, p_2, n)	(p_1, p_2, n)
	$(10, 15, 20)$	$(15, 20, 25)$
P-value Interval	No. of Exp.	No. of Exp.
$[0, 0.05]$	41	53
$[0.05, 0.1]$	30	29
$[0.1, 0.2]$	12	10
$[0.2, 0.3]$	6	5
$[0.3, 0.4]$	8	2
$[0.4, 0.5]$	0	1
$[0.6, 0.7]$	2	0
$[0.8, 0.9]$	1	0
$[0.9, 1]$	0	0

*These are P-values for (p_1, p_2) companies from different two sections of NYSE: basic industry section and capital goods section, each of which has n closed stock prices during the period $1990.1.1 - 2002.1.1$. The number of repeated experiments are 100. All the closed stock prices are from WRDS database. No. of Exp. is the number of experiments whose P-values are in the corresponding interval.

Table 3.10: *P-values for (p_1, p_2) companies from public utility section and capital goods section of NYSE.*

P-values	(p_1, p_2, n)	(p_1, p_2, n)
	$(10, 15, 20)$	$(15, 20, 25)$
P-value Interval	No. of Exp.	No. of Exp.
$[0, 0.05]$	82	80
$[0.05, 0.1]$	9	16
$[0.1, 0.2]$	4	3
$[0.2, 0.3]$	2	0
$[0.3, 0.4]$	0	1
$[0.4, 0.5]$	0	0
$[0.6, 0.7]$	2	0
$[0.8, 0.9]$	1	0
$[0.9, 1]$	0	0

*These are P-values for (p_1, p_2) companies from different two sections of NYSE: basic industry section and capital goods section, each of which has n closed stock prices during the period 1990.1.1 − 2002.1.1. The number of repeated experiments are 100. All the closed stock prices are from WRDS database. No. of Exp. is the number of experiments whose P-values are in the corresponding interval.

Table 3.11: *P-values for (p_1, p_2) companies from finance section and healthcare section of NYSE.*

P-values	(p_1, p_2, n)	(p_1, p_2, n)
	$(10, 15, 20)$	$(15, 20, 25)$
P-value Interval	No. of Exp.	No. of Exp.
$[0, 0.05]$	84	88
$[0.05, 0.1]$	7	10
$[0.1, 0.2]$	2	2
$[0.2, 0.3]$	4	1
$[0.3, 0.4]$	1	0
$[0.4, 0.5]$	0	0
$[0.6, 0.7]$	1	0
$[0.8, 0.9]$	1	0
$[0.9, 1]$	0	0

*These are P-values for (p_1, p_2) companies from different two sections of NYSE: basic industry section and capital goods section, each of which has n closed stock prices during the period 1990.1.1 − 2002.1.1. The number of repeated experiments are 100. All the closed stock prices are from WRDS database. No. of Exp. is the number of experiments whose P-values are in the corresponding interval.

Chapter 4

CLT for a sample covariance matrix plus a perturbation

As stated in last Chapter, we need the central limit theorem for linear spectral statistics of a perturbation matrix. This chapter is devoted to providing the CLT for linear spectral statistics, quantities of the form

$$\frac{1}{n}\sum_{j=1}^{n} f(\lambda_j) = \int f(x) dF^{\mathbf{B}_n}(x), \tag{4.1}$$

where f is a function on $[0, \infty)$, $\lambda_1, \ldots, \lambda_n$ denote the eigenvalues of random matrices $\mathbf{B}_n$ and

$$\mathbf{B}_n = \frac{1}{N}\mathbf{X}\mathbf{X}^* + \mathbf{T}_n. \tag{4.2}$$

Here $\mathbf{X} = (X_{ij})$ is $n \times N$ with independent and identically distributed (i.i.d) complex (real) standardized entries, $\mathbf{T}_n$ is a nonnegative Hermitian matrix, and the empirical spectral distribution (ESD) of any square matrix $\mathbf{A}$ with real eigenvalues $\mu_1 \le \mu_2 \le \cdots \le \mu_n$ is denoted by

$$F^{\mathbf{A}}(x) = \frac{1}{n}\#\{i : \mu_i \le x\}, \tag{4.3}$$

where $\#\{\cdots\}$ denotes the cardinality of the set $\{\cdots\}$.

Silverstein (1995) discovers the limiting spectral distribution(LSD) $F_{c,H}$, the limit of $F^{\mathbf{B}_n}$, which is given in Lemma 14 below for easy reference. The Stieltjes transform of any distribution function $G(x)$ is defined by

$$m_G(z) \equiv \int \frac{1}{\lambda - z} dG(\lambda), \quad \mathrm{Im}(z) \neq 0. \tag{4.4}$$

Lemma 14. *Assume that*

1. *For each n, $\mathbf{X}_n = (X_{ij}^n)$, $\{X_{ij}^n : i = 1, \ldots, n; j = 1, \ldots, N\}$ are i.d.; for all n, i, j, $\{X_{ij}^n : n = 1, 2, \ldots; i = 1, \ldots, n; j = 1, \ldots, N\}$ are independent. Moreover, $\mathbb{E}X_{11} = 0$ and $\mathbb{E}|X_{11}|^2 = 1$.*

2. *$n = n(N)$ with $n/N \to c > 0$ as $N \to \infty$.*

3. *$\mathbf{T}_n$ is an $n \times n$ Hermitian nonrandom matrix for which $F^{\mathbf{T}_n}(x)$ converges vaguely to a nonrandom distribution $H(x)$,*

then almost surely, $F^{\mathbf{B}_n}$, the ESD of $\mathbf{B}_n$, converges vaguely, as $N \to \infty$, to a nonrandom distribution $F_{c,H}$, whose Stieltjes transform $m^0(z), z \in \mathbb{C}^+$ satisfies

$$m^0(z) = m_H\Big(z - \frac{1}{1 + cm^0(z)}\Big), \tag{4.5}$$

where $m_H(z)$ denotes the Stieltjes transform of $H(x)$.

Remark 8. *Indeed, Silverstein (1995) derives a more general equation than (4.5) for the matrix $\frac{1}{n}\mathbf{X}\mathbf{A}_n\mathbf{X}^* + \mathbf{T}_n$, where $\mathbf{A}_n$ is a diagonal matrix. If we take $\mathbf{A}_n = diag\Big(\frac{n}{N}, \frac{n}{N}, \ldots, \frac{n}{N}\Big)$ then the equation (4.5) for the matrix $\mathbf{B}_n = \frac{1}{n}\mathbf{X}\mathbf{X}^* + \mathbf{T}_n$ follows. A similar result covering more general matrices $\mathbf{A}_n$ can be found in Pan (2010).*

Before stating Theorem 6, we introduce some notation. Set

$$G_n(x) = n[F^{\mathbf{B}_n}(x) - F_{c_n,H_n}(x)], \tag{4.6}$$

where $H_n \equiv F^{\mathbf{T}_n}$, $c_n = n/N$ and $F_{c_n,H_n}(x)$ can be obtained from $F_{c,H}(x)$ with c and $H(x)$ replaced by c_n and $H_n(x)$, respectively.

Let

$$\begin{aligned} m_r(z) &= \int \frac{dH(x)}{(x - z + \varpi(z))^r}, \qquad \varpi(z) = \frac{1}{1 + cm^0(z)}, \\ s(z_1, z_2) &= \frac{1}{1 + cm^0(z_1)} - \frac{1}{1 + cm^0(z_2)}. \end{aligned} \tag{4.7}$$

where r is a positive integer.

The main result of this chapter is Theorem 6.

Theorem 6. *Assume that*

(a) $\{X_{ij}, i \leq n, j \leq N\}$ are i.i.d. with $EX_{11} = 0$, $E|X_{11}|^2 = 1$ and $E|X_{11}|^4 < \infty$.

(b) $\mathbf{T}_n$ is $n \times n$ nonrandom Hermitian nonnegative definite with spectral norm bounded in n, and with $F^{\mathbf{T}_n} \xrightarrow{D} H$, a proper c.d.f.

(c) $n = n(N)$ with $n/N \to c > 0$ as $N \to \infty$.

Let $f_1, \ldots, f_k$ be functions on $\mathbb{R}$ analytic on an open interval containing

$$\Big[I_{(0,1)}(c)(1 - \sqrt{c})^2 + \liminf_n \lambda_{min}^{\mathbf{T}_n}, (1 + \sqrt{c})^2 + \limsup_n \lambda_{max}^{\mathbf{T}_n}\Big], \tag{4.8}$$

where $\lambda_{min}^{\mathbf{T}_n}$ and $\lambda_{max}^{\mathbf{T}_n}$ denote the maximum and minimum eigenvalues of $\mathbf{T}_n$ respectively. Then

(i) the random vector

$$\Big(\int f_1(x) dG_n(x), \ldots, \int f_k(x) dG_n(x)\Big) \tag{4.9}$$

forms a tight sequence in n.

(ii) If X_{11} and $\mathbf{T}_n$ are real and $EX_{11}^4 = 3$, then (4.9) converges weakly to a

Gaussian vector $(X_{f_1}, \ldots, X_{f_k})$ *with mean*

$$EX_f = \frac{1}{-2\pi i} \oint_{\mathcal{C}} f(z) \frac{c\varpi^2(z)m_3(z) + c^2\varpi^4(z)\big(m^0(z)\big)' m_3(z) - c^2\varpi^3(z)\big(m^0(z)\big)' m_2(z)}{1 - c\varpi^2(z)m_2(z)} dz \tag{4.10}$$

and covariance function

$$\begin{aligned} Cov(X_{f_i}, X_{f_j}) &= -\frac{1}{2\pi^2} \oint_{\mathcal{C}} \oint_{\mathcal{C}} f_i(z_1) f_j(z_2) [1 + \frac{c(m^0(z_1))'}{(1 + cm^0(z_1))^2} + \frac{c(m^0(z_2))'}{(1 + cm^0(z_2))^2} \\ &\quad + \frac{c(m^0(z_1))'}{(1 + cm^0(z_1))^2} \frac{c(m^0(z_2))'}{(1 + cm^0(z_2))^2}] \frac{1}{(z_2 - z_1 + s(z_1, z_2))^2} dz_1 dz_2. \end{aligned} \tag{4.11}$$

The contours in (4.10) and (4.11) are closed and are taken in the positive direction in the complex plane, each enclosing the support of $F_{c,H}$*.*
(iii) If X_{11} *is complex with* $\mathbb{E}(X_{11}^2) = 0$ *and* $\mathbb{E}(|X_{11}|^4) = 2$*, then the result above also holds, except the mean is zero and the covariance function is* $1/2$ *the function given in (4.11).*

Remark 9. *We investigate the matrix* $\mathbf{B}_n = \frac{1}{N}\mathbf{X}\mathbf{X}^* + \mathbf{T}_n$ *while Bai and Silverstein (2004) studies the matrix of the form* $\mathbf{S}_n = \frac{1}{N}\mathbf{R}_n^{1/2}\mathbf{X}\mathbf{X}^*\mathbf{R}_n^{1/2}$*, where* $\mathbf{R}_n^{1/2}$ *is a Hermitian square root of the nonnegative definite Hermitian matrix* $\mathbf{R}_n$*. The two matrices* $\mathbf{B}_n$ *and* $\mathbf{S}_n$ *are the same when the matrix* $\mathbf{T}_n$ *becomes a zero matrix and* $\mathbf{R}_n$ *becomes an identity matrix. In this case, the asymptotic means and covariances in Bai and Silverstein (2004) and in Theorem 6 are the same, which is verified in the last part of this chapter.*

4.1 Proof of Theorem 6

The proof of Theorem 6 follows a line similar to that in Bai and Silverstein (2004). Throughout the proof K denotes a constant which may change from line to line.

4.1.1 Truncation, centralization and renormalization

We begin the proof by replacing the entries of $\mathbf{X}_n$ with truncated and centralized variables. Since the argument for (1.8) in Bai and Silverstein (2004) can be carried directly over to the present case, we can then select positive sequences δ_n such that

$$\delta_n \to 0, \quad \delta_n^{-4} \int_{\{|X_{11}| \geq \delta_n \sqrt{n}\}} |X_{11}|^4 \to 0. \tag{4.12}$$

Set $\hat{\mathbf{B}}_n = \frac{1}{N}\hat{\mathbf{X}}_n\hat{\mathbf{X}}_n^* + \mathbf{T}_n$ with $\hat{\mathbf{X}}_n$ (of size $n \times N$) having the (i,j)th entry $X_{ij}I_{|X_{ij}|<\delta_n\sqrt{n}}$. Then we have

$$P(\mathbf{B}_n \neq \hat{\mathbf{B}}_n) \leq nNP(|X_{11}| \geq \delta_n\sqrt{n}) \leq K\delta_n^{-4} \int_{\{|X_{11}| \geq \delta_n \sqrt{n}\}} |X_{11}|^4 = o(1).$$

Define $\widetilde{\mathbf{B}}_n = \frac{1}{N}\widetilde{\mathbf{X}}_n\widetilde{\mathbf{X}}_n^* + \mathbf{T}_n$ with $\widetilde{\mathbf{X}}_n$ having (i,j)th entry $(\hat{X}_{ij} - \mathbb{E}\hat{X}_{ij})/\sigma_n$, where $\sigma_n = \mathbb{E}|\hat{X}_{ij} - \mathbb{E}\hat{X}_{ij}|^2$. From Bai and Silverstein (2004) we know that both $\limsup_n \lambda_{max}^{\hat{\mathbf{C}}_n}$ and $\limsup_n \lambda_{max}^{\widetilde{\mathbf{C}}_n}$ are almost surely bounded by $(1+\sqrt{c})$, where $\hat{\mathbf{C}}_n = \frac{1}{N}\hat{\mathbf{X}}_n\hat{\mathbf{X}}_n^*$ and $\widetilde{\mathbf{C}}_n = \frac{1}{N}\widetilde{\mathbf{X}}_n\widetilde{\mathbf{X}}_n^*$. By Weyl's inequality and the assumption $||\mathbf{T}_n|| \leq M$, we have that $\limsup_n \lambda_{max}^{\hat{\mathbf{B}}_n}$ and $\limsup_n \lambda_{max}^{\widetilde{\mathbf{B}}_n}$ are almost surely bounded by $[(1+\sqrt{c}) + M]$. We use $\hat{G}_n(x)$ and $\widetilde{G}_n(x)$ to denote the analogues of $G_n(x)$ with the matrix $\mathbf{B}_n$ replaced by $\hat{\mathbf{B}}_n$ and $\widetilde{\mathbf{B}}_n$ respectively.

Since $\mathbf{T}_n$ is a nonnegative definite matrix, we can write $\mathbf{T}_n = \mathbf{T}_n^{1/2}\mathbf{T}_n^{1/2} = \sum_{i=1}^n \mathbf{t}_i\mathbf{t}_i^*$, where $\mathbf{t}_i$ is the ith column of $\mathbf{T}_n^{1/2}$. We may then write

$$\mathbf{B}_n = \mathbf{F}_n\mathbf{F}_n^*, \tag{4.13}$$

where

$$\mathbf{F}_n = (\mathbf{r}_1, \ldots, \mathbf{r}_N, \mathbf{t}_1, \ldots, \mathbf{t}_n) \tag{4.14}$$

with $\mathbf{r}_i = \frac{1}{N}\mathbf{X}_{.i}$, $i = 1, \ldots, N$ and $\mathbf{X}_{.i}$ standing for the ith column of $\mathbf{X}_n$. Define $\hat{\mathbf{F}}_n$ and $\widetilde{\mathbf{F}}_n$ to be the analogues of $\mathbf{F}_n$ with the matrix $\mathbf{X}_n$ replaced by $\hat{\mathbf{X}}_n$ and $\widetilde{X}_n$ respectively. For each $j = 1, 2, \ldots, k$,

$$\begin{aligned}\Big|\int f_j(x)d\hat{G}_n(x) - \int f_j(x)d\widetilde{G}_n(x)\Big| &\le K_j \sum_{k=1}^{n} \Big|\lambda_k^{\hat{\mathbf{B}}_n} - \lambda_k^{\widetilde{\mathbf{B}}_n}\Big| \\ &\le 2K_j\Big(tr(\hat{\mathbf{F}}_n - \widetilde{\mathbf{F}}_n)(\hat{\mathbf{F}}_n - \widetilde{\mathbf{F}}_n)^*\Big)^{1/2}\Big(n(\lambda_{max}^{\hat{\mathbf{B}}_n} + \lambda_{max}^{\widetilde{\mathbf{B}}_n})\Big)^{1/2},\end{aligned}$$

where K_j is a bound on $|f_j'(z)|$ and $\lambda_k^{\mathbf{A}}$ denotes the ith smallest eigenvalue of the matrix $\mathbf{A}$.

By the fact that

$$tr(\hat{\mathbf{F}}_n - \widetilde{\mathbf{F}}_n)(\hat{\mathbf{F}}_n - \widetilde{\mathbf{F}}_n)^* = N^{-1}tr(\hat{\mathbf{X}}_n - \widetilde{\mathbf{X}}_n)(\hat{\mathbf{X}}_n - \widetilde{\mathbf{X}}_n)^*,$$

and the result on page 560 of Bai and Silverstein (2004), i.e.

$$\Big(N^{-1}tr(\hat{\mathbf{X}}_n - \widetilde{\mathbf{X}}_n)(\hat{\mathbf{X}}_n - \widetilde{\mathbf{X}}_n)^*\Big)^{1/2} = o(\delta_n n^{-1/2})(\lambda_{max}^{\hat{\mathbf{B}}_n})^{1/2} + o(\delta_n n^{-1}),$$

we obtain

$$\int f_j(x)dG_n(x) = \int f_j(x)d\widetilde{G}_n(x) + o_P(1).$$

Therefore, in the sequel, we shall assume

$$|X_{ij}| < \delta_n\sqrt{n}, \quad \mathbb{E}X_{ij} = 0, \quad \mathbb{E}|X_{ij}|^2 = 1, \quad \mathbb{E}|X_{ij}|^4 < \infty,$$

and for the real case, $\mathbb{E}|X_{11}|^4 = 3 + o(1)$ while for the complex case, $\mathbb{E}X_{11}^2 = o(1/n)$ and $\mathbb{E}|X_{11}|^4 = 2 + o(1)$. For simplicity we suppress all the subscripts and superscripts on variables.

4.1.2 From linear spectral statistics to Stieltjes transforms

With notation $\mathbf{C}_n = \frac{1}{N}\mathbf{X}\mathbf{X}^*$, by Weyl's inequality we have

$$\lambda_{max}^{\mathbf{B}_n} \le \lambda_{max}^{\mathbf{C}_n} + \lambda_{max}^{\mathbf{T}_n}, \quad \lambda_{min}^{\mathbf{B}_n} \ge \lambda_{min}^{\mathbf{C}_n} + \lambda_{min}^{\mathbf{T}_n}. \tag{4.15}$$

From (1.9a) and (1.9b) of Bai and Silverstein (2004), we have

$$P(\lambda_{max}^{\mathbf{B}_n} \geq \eta) = o(n^{-\ell}), \quad P(\lambda_{min}^{\mathbf{B}_n} \leq \theta) = o(n^{-\ell}), \tag{4.16}$$

for any $\eta > \Big((1+\sqrt{c})^2 + \limsup_n \lambda_{max}^{\mathbf{T}_n}\Big)$, any $0 < \theta < \Big(I_{(0,1)}(c)(1-\sqrt{c})^2 + \liminf_n \lambda_{min}^{\mathbf{T}_n}\Big)$ and any positive ℓ.

Write

$$M_n(z) = n\big(m_n(z) - m_n^0(z)\big)$$

where $m_n(z)$ denotes the Stieltjes transform of $F^{\mathbf{B}_n}$ and $m_n^0(z)$ is $m^0(z)$ with c, H replaced by c_n, H_n respectively. By Cauchy's integral formula

$$f_\ell(x) = \frac{1}{2\pi i} \oint \frac{f_\ell(z)}{z-x} dz, \tag{4.17}$$

we have for $k \geq 1$, any complex constants $a_1, \cdots, a_k$, and for all n large with probability one,

$$\sum_{\ell=1}^{k} a_\ell \int f_\ell(x) dG_n(x) = -\sum_{\ell=1}^{k} \frac{a_\ell}{2\pi i} \oint_{\mathcal{C}} f_\ell(z) M_n(z) dz, \tag{4.18}$$

where the contour $\mathcal{C}$ is specified below. Let $v_0 > 0$ be arbitrary. Let x_r be any number greater than the right end point of interval (4.8). Let x_ℓ be any negative number if the left end point of (4.8) is zero. Otherwise, choose $x_\ell \in \big(0, (1-\sqrt{c})^2 + \liminf_n \lambda_{\min}^{\mathbf{T}_n}\big)$. Let

$$\mathcal{C}_u = \{x + iv_0 : x \in [x_\ell, x_r]\}.$$

Set

$$\mathcal{C}^+ \equiv \{x_\ell + iv : v \in [0, v_0]\} \cup \mathcal{C}_u \cup \{x_r + iv : v \in [0, v_0]\}.$$

Let $\mathcal{C}^-$ be the symmetric part of $\mathcal{C}^+$ about the real axis. Then set $\mathcal{C} = \mathcal{C}^+ \cup \mathcal{C}^-$.

We define now the subsets $\mathcal{C}_n^+$ and its symmetric part $\mathcal{C}_n^-$ of $\mathcal{C}$ when $M_n(\cdot)$ agrees with $\hat{M}_n(\cdot)$, a truncated version of $M_n(\cdot)$ to be defined below. Select a sequence $\{\varepsilon_n\}$ such that for some $\rho \in (0,1)$

$$\varepsilon_n \downarrow 0, \quad \varepsilon_n \geq n^{-\rho}.$$

Let

$$\mathcal{C}_\ell = \begin{cases} \{x_\ell + iv : v \in [n^{-1}\varepsilon_n, v_0]\}, & x_\ell > 0; \\ \{x_\ell + iv : v \in [0, v_0]\}, & x_\ell < 0, \end{cases}$$

and

$$\mathcal{C}_r = \{x_r + iv : v \in [n^{-1}\varepsilon_n, v_0]\}.$$

Set $\mathcal{C}_n^+ = \mathcal{C}_\ell \cup \mathcal{C}_u \cup \mathcal{C}_r$. The process $\hat{M}_n(\cdot)$ can now be defined. For $z = x+iv$, we have

$$\hat{M}_n(z) = \begin{cases} M_n(z), \quad if\ z \in \mathcal{C}_n^+ \cup \mathcal{C}_n^-; \\ \frac{nv+\varepsilon_n}{2\varepsilon_n} M_n(x_r + in^{-1}\varepsilon_n) + \frac{\varepsilon_n - nv}{2\varepsilon_n} M_n(x_r - in^{-1}\varepsilon_n), \\ \qquad if\ x = x_r, v \in [-n^{-1}\varepsilon_n, n^{-1}\varepsilon_n]; \\ \frac{nv+\varepsilon_n}{2\varepsilon_n} M_n(x_l + in^{-1}\varepsilon_n) + \frac{\varepsilon_n - nv}{2\varepsilon_n} M_n(x_l - in^{-1}\varepsilon_n), \\ \qquad if\ x = x_\ell > 0, v \in [-n^{-1}\varepsilon_n, n^{-1}\varepsilon_n]. \end{cases} \tag{4.19}$$

With probability one, for all n large,

$$\begin{aligned} &\Big| \oint_{\mathcal{C}} f(z)\big(M_n(z) - \hat{M}_n(z)\big) dz \Big| \\ \leq \quad & K\varepsilon_n \Big(\Big| \max\big((1+\sqrt{c_n})^2 + \lambda_{max}^{\mathbf{T}_n}, \lambda_{max}^{\mathbf{B}_n}\big) - x_r \Big|^{-1} \Big) \\ &\quad + \Big| \min\Big(I_{(0,1)}(c)(1-\sqrt{c})^2 + \lambda_{min}^{\mathbf{T}_n}, \lambda_{min}^{\mathbf{B}_n} \Big) - x_\ell \Big|^{-1} \to 0. \end{aligned} \tag{4.20}$$

In view of this and (4.18), as discussed in Bai and Silverstein (2004), it is enough to consider the limiting distribution of $\sum_{\ell=1}^k a_\ell \hat{M}_n(z_\ell)$.

4.1.3 CLT of the Stieltjes transform $m_n(z)$ of $F^{\mathbf{B}_n}$

Recall the definitions of $m(z,r), \varpi(z)$ and $s(z_1,z_2)$ in the introduction.

Lemma 15. *Under conditions (a)-(c) of Theorem 6, $\{\hat{M}_n(z)\}$ forms a tight sequence on $\mathcal{C}$. Moreover, if assumptions in (ii) or (iii) of Theorem 6 on X_{11} hold, then $\hat{M}_n(z)$ converges weakly to a Gaussian process $M(z)$ for $z \in \mathcal{C}$ under the assumptions in (ii),*

$$\mathbb{E}M(z) = \frac{c\varpi^2(z)m(z,3) + c^2\varpi^4(z)\big(m^0(z)\big)' m(z,3) - c^2\varpi^3(z)\big(m^0(z)\big)' m(z,2)}{1 - c\varpi^2(z)m(z,2)} \tag{4.21}$$

and for $z_1, z_2 \in \mathcal{C}$

$$\begin{aligned} Cov(M(z_1), M(z_2)) \quad = \quad & -\frac{2}{(z_2-z_1)^2} + 2[1 + \frac{c(m^0(z_1))'}{(1+cm^0(z_1))^2} + \frac{c(m^0(z_2))'}{(1+cm^0(z_2))^2} \\ & + \frac{c(m^0(z_1))'}{(1+cm^0(z_1))^2}\frac{c(m^0(z_2))'}{(1+cm^0(z_2))^2}]\frac{1}{(z_2-z_1+s(z_1,z_2))^2}, \end{aligned} \tag{4.22}$$

while under the assumptions in (iii) $\mathbb{E}M(z) = 0$, and the covariance function similar to (4.22) is half of the right hand side of (4.22).

We first list (2.3) of Bai and Silverstein (2004) below as Proposition 2, which holds as well in our setting.

Proposition 2. *For any nonrandom $n \times n$ matrices $\mathbf{A}_k, k = 1, \ldots, p$ and $\mathbf{B}_\ell, \ell = 1, \ldots, q$, there exists*

$$\begin{aligned} & \Big|\mathbb{E}\Big(\prod_{k=1}^{p} \mathbf{r}_1^*\mathbf{A}_k\mathbf{r}_1 \prod_{\ell=1}^{q}(\mathbf{r}_1^*\mathbf{B}_\ell\mathbf{r}_1 - N^{-1}tr\mathbf{B}_\ell)\Big)\Big| \\ \leq \quad & KN^{-(1\wedge q)}\delta_n^{(2q-4)\vee 0}\prod_{k=1}^{p}||\mathbf{A}_k||\prod_{\ell=1}^{q}||\mathbf{B}_\ell||, \quad p \geq 0, \quad q \geq 0. \end{aligned} \tag{4.23}$$

Proof. We now start the proof of Lemma 15. Write $M_n(z) = M_n^{(1)}(z) + M_n^{(2)}(z)$, where

$$M_n^{(1)}(z) = n\Big(m_n(z) - \mathbb{E}m_n(z)\Big), \quad M_n^{(2)}(z) = n\Big(\mathbb{E}m_n(z) - m_n^0(z)\Big).$$

By the discussion in Bai and Silverstein (2004), it suffices to prove the following four statements.

1. Finite dimension convergence of $M_n^{(1)}(z)$ on $\mathcal{C}_n$.

2. $M_n^{(1)}(z)$ is tight on $\mathcal{C}_n$ where $\mathcal{C}_n = \mathcal{C}_n^+ \cup \mathcal{C}_n^-$.

3. $M_n^{(2)}(z) \to \mathbb{E}M(z)$, for $z \in \mathbb{C}_n$, where $M(z)$ is the limit of $M_n(z)$ as $n \to \infty$.

4. $\{M_n^{(2)}(z)\}$ for $z \in \mathcal{C}_n$ is bounded and equicontinuous.

4.1.3.1 Step 1: Convergence of $M_n^{(1)}(z)$

Let $v_0 = \mathrm{I}m(z)$. To facilitate analysis we consider the case of $v_0 > 0$ only. We first introduce some notation as follows.

$$\begin{aligned}
&\mathbf{r}_j = \frac{1}{\sqrt{N}}\mathbf{X}_{\cdot j},\ \mathbf{D}(z) = \mathbf{B}_n - z\mathbf{I},\ \mathbf{D}_j(z) = \mathbf{D}(z) - \mathbf{r}_j\mathbf{r}_j^*,\\
&\gamma_j(z) = \mathbf{r}_j^*\mathbf{D}_j^{-1}(z)\mathbf{r}_j - \frac{1}{N}Etr\mathbf{D}_j^{-1}(z),\ \varepsilon_j(z) = \mathbf{r}_j^*\mathbf{D}_j^{-1}(z)\mathbf{r}_j - \frac{1}{N}tr\mathbf{D}_j^{-1}(z),\\
&\delta_j(z) = \mathbf{r}_j^*\mathbf{D}_j^{-2}(z)\mathbf{r}_j - \frac{1}{N}tr\mathbf{D}_j^{-2}(z) = \frac{d}{dz}\varepsilon_j(z), \quad \beta_j(z) = \frac{1}{1+\mathbf{r}_j^*\mathbf{D}_j^{-1}(z)\mathbf{r}_j},\\
&\beta_j^{tr}(z) = \frac{1}{1+N^{-1}tr\mathbf{D}_j^{-1}(z)}, \quad b_n(z) = \frac{1}{1+N^{-1}\mathbb{E}tr\mathbf{D}_1^{-1}(z)}.
\end{aligned}$$

As pointed out by Bai and Silverstein (2004), the later three variables are all bounded by $|z|/v_0$. Let $\mathbb{E}_0(\cdot)$ denote expectation and $\mathbb{E}_j(\cdot)$ denote conditional expectation with respect to the σ-field generated by $\mathbf{r}_1, \ldots, \mathbf{r}_j$.

Write

$$\begin{aligned} n\Big(m_n(z)-\mathbb{E}m_n(z)\Big) &= tr\Big(\mathbf{D}^{-1}(z)-\mathbb{E}\mathbf{D}^{-1}(z)\Big)=\sum_{j=1}^{N} tr\mathbb{E}_j\mathbf{D}^{-1}(z)-tr\mathbb{E}_{j-1}\mathbf{D}^{-1}(z) \\ &= \sum_{j=1}^{N} tr\mathbb{E}_j\Big(\mathbf{D}^{-1}(z)-\mathbf{D}_j^{-1}(z)\Big)-tr\mathbb{E}_{j-1}\Big(\mathbf{D}^{-1}(z)-\mathbf{D}_j^{-1}(z)\Big) \\ &= -\sum_{j=1}^{N}(\mathbb{E}_j-\mathbb{E}_{j-1})\beta_j(z)\mathbf{r}_j^*\mathbf{D}_j^{-2}(z)\mathbf{r}_j, \end{aligned} \tag{4.24}$$

where the last equality uses

$$\mathbf{D}^{-1}(z)-\mathbf{D}_j^{-1}(z)=-\mathbf{D}_j^{-1}(z)\mathbf{r}_j\mathbf{r}_j^*\mathbf{D}_j^{-1}(z)\beta_j(z). \tag{4.25}$$

By the identity

$$\beta_j(z)=\beta_j^{tr}(z)-\beta_j(z)\beta_j^{tr}(z)\varepsilon_j(z)=\beta_j^{tr}(z)-(\beta_j^{tr}(z))^2\varepsilon_j(z)+(\beta_j^{tr}(z))^2\beta_j(z)\varepsilon_j^2(z), \tag{4.26}$$

we have

$$\begin{aligned} &(\mathbb{E}_j-\mathbb{E}_{j-1})\beta_j(z)\mathbf{r}_j^*\mathbf{D}_j^{-2}(z)\mathbf{r}_j=\mathbb{E}_j\Big(\beta_j^{tr}(z)\delta_j(z)-(\beta_j^{tr}(z))^2\varepsilon_j(z)\frac{1}{N}tr\mathbf{D}_j^{-2}(z)\Big) \\ &-(\mathbb{E}_j-\mathbb{E}_{j-1})(\beta_j^{tr}(z))^2\Big(\varepsilon_j(z)\delta_j(z)-\beta_j(z)\mathbf{r}_j^*\mathbf{D}_j^{-2}(z)\mathbf{r}_j\varepsilon_j^2(z)\Big). \end{aligned}$$

By Proposition 2 one can prove that $(\mathbb{E}_j-\mathbb{E}_{j-1})(\beta_j^{tr}(z))^2\Big(\varepsilon_j(z)\delta_j(z)-\beta_j(z)\mathbf{r}_j^*\mathbf{D}_j^{-2}(z)\mathbf{r}_j\varepsilon_j^2(z)\Big)$ converges to zero in probability (One can refer to page 569 of Bai and Silverstein (2004) for similar arguments).

Therefore it is sufficient to consider the sum $\sum_{\ell=1}^{k}a_\ell\sum_{j=1}^{N}Y_j(z_\ell)$, where

$$Y_j(z)=\mathbb{E}_j\Big(\beta_j^{tr}(z)\delta_j(z)-(\beta_j^{tr}(z))^2\varepsilon_j(z)\frac{1}{N}tr\mathbf{D}_j^{-2}(z)\Big)=-\mathbb{E}_j\frac{d}{dz}\beta_j^{tr}(z)\varepsilon_j(z). \tag{4.27}$$

We next utilize Lemma 2.4 of Bai and Silverstein (2004), CLT for martingale differences. By Proposition 2 and using the same arguments as those above (2.4) on page 570 of Bai and Silverstein (2004), we see that condition

2 of Lemma 2.4 of Bai and Silverstein (2004) is satisfied and it is therefore enough to find the limit in probability of

$$\sum_{j=1}^{N} \mathbb{E}_{j-1}\Big(Y_j(z_1)Y_j(z_2)\Big). \tag{4.28}$$

Consider the sum

$$\sum_{j=1}^{N} \mathbb{E}_{j-1}\Big(\mathbb{E}_j\big(\beta_j^{tr}(z_1)\varepsilon_j(z_1)\big)\mathbb{E}_j\big(\beta_j^{tr}(z_2)\varepsilon_j(z_2)\big)\Big). \tag{4.29}$$

Since

$$\frac{\partial^2}{\partial z_2 \partial z_1}(4.29) = (4.28), \tag{4.30}$$

by the same arguments as those on page 571 of Bai and Silverstein (2004) we only need to show (4.29) converges in probability and to determine its limit.

Note that the derivation above (4.3) of Bai and Silverstein (1998) is true in the present case and hence

$$E|\frac{1}{N}tr\mathbf{D}_j^{-1}(z) - \frac{1}{N}Etr\mathbf{D}_j^{-1}(z)|^p \leq KN^{-p/2}. \tag{4.31}$$

By the discussions above (2.7) of Bai and Silverstein (2004), we then have

$$\sum_{j=1}^{N} \mathbb{E}_{j-1}\Big(\mathbb{E}_j\big(\beta_j^{tr}(z_1)\varepsilon_j(z_1)\big)\mathbb{E}_j\big(\beta_j^{tr}(z_2)\varepsilon_j(z_2)\big)\Big)$$
$$-b_n(z_1)b_n(z_2)\sum_{j=1}^{N} \mathbb{E}_{j-1}\Big(\mathbb{E}_j\big(\varepsilon_j(z_1)\big)\mathbb{E}_j\big(\varepsilon_j(z_2)\big)\Big) \xrightarrow{i.p.} 0.$$

Thus it remains to prove that

$$b_n(z_1)b_n(z_2)\sum_{j=1}^{N} \mathbb{E}_{j-1}\Big(\mathbb{E}_j\big(\varepsilon_j(z_1)\big)\mathbb{E}_j\big(\varepsilon_j(z_2)\big)\Big) \tag{4.32}$$

converges in probability and to determine its limit.

In the complex case, namely $EX_{11}^2 = o(1/n)$ and $E|X_{11}|^4 = 2 + o(1)$, by the identity

$$\begin{aligned}
&\mathbb{E}(\mathbf{X}_{.1}^*\mathbf{A}\mathbf{X}_{.1} - tr\mathbf{A})(\mathbf{X}_{.1}^*\mathbf{B}\mathbf{X}_{.1} - tr\mathbf{B}) \\
= &\ (\mathbb{E}|X_{11}|^4 - |EX_{11}^2|^2 - 2)\sum_{i=1}^{n} a_{ii}b_{ii} + |EX_{11}^2|^2 tr\mathbf{A}\mathbf{B}^T + tr\mathbf{A}\mathbf{B}
\end{aligned} \tag{4.33}$$

valid for $n \times n$ nonrandom matrices $\mathbf{A} = (a_{ij})$ and $\mathbf{B} = (b_{ij})$, (4.32) becomes

$$b_n(z_1)b_n(z_2)\frac{1}{N^2}\sum_{j=1}^{N}\Big(tr\mathbb{E}_j\big(\mathbf{D}_j^{-1}(z_1)\big)\mathbb{E}_j\big(\mathbf{D}_j^{-1}(z_2)\big) + o(1)A_n\Big), \tag{4.34}$$

where

$$|A_n| \le K\Big(tr\mathbb{E}_j\big(\mathbf{D}_j^{-1}(z_1)\big)\mathbb{E}_j\big(\bar{\mathbf{D}}_j^{-1}(z_1)\big) \times tr\mathbb{E}_j\big(\mathbf{D}_j^{-1}(z_2)\big)\mathbb{E}_j\big(\bar{\mathbf{D}}_j^{-1}(z_2)\big)\Big)^{1/2} = O(N).$$

Thus it is sufficient to study

$$b_n(z_1)b_n(z_2)\frac{1}{N^2}\sum_{j=1}^{N} tr\mathbb{E}_j\big(\mathbf{D}_j^{-1}(z_1)\big)\mathbb{E}_j\big(\mathbf{D}_j^{-1}(z_2)\big). \tag{4.35}$$

In the real case, namely $E|X_{11}|^4 = 3 + o(1)$, (4.32) should be double the limit of (4.35).

The next aim is to investigate (4.35). To this end, set $\mathbf{D}_{ij}(z) = \mathbf{D}(z) - \mathbf{r}_i\mathbf{r}_i^* - \mathbf{r}_j\mathbf{r}_j^*$,

$$\begin{aligned}
\beta_{ij}(z) &= \frac{1}{1 + \mathbf{r}_i^*\mathbf{D}_{ij}^{-1}(z)\mathbf{r}_i},\ \ b_1(z) = \frac{1}{1 + N^{-1}\mathbb{E}tr\mathbf{D}_{12}^{-1}(z)}, \\
\mathbf{H}^{-1}(z) &= \big(z\mathbf{I} - \frac{N-1}{N}b_1(z)\mathbf{I} - \mathbf{T}_n\big)^{-1}.
\end{aligned}$$

Write

$$\mathbf{D}_j(z_1) + z_1\mathbf{I} - \frac{N-1}{N}b_1(z_1)\mathbf{I} - \mathbf{T}_n = \sum_{i\neq j}^{N}\mathbf{r}_i\mathbf{r}_i^* - \frac{N-1}{N}b_1(z_1)\mathbf{I}.$$

Multiplying by $\mathbf{H}^{-1}(z_1)$ on the left hand side, $\mathbf{D}_j^{-1}(z_1)$ on the right hand side and using

$$\mathbf{r}_i^*\mathbf{D}_j^{-1}(z_1) = \beta_{ij}(z_1)\mathbf{r}_i^*\mathbf{D}_{ij}^{-1}(z_1), \tag{4.36}$$

we conclude that

$$\begin{aligned}
\mathbf{D}_j^{-1}(z_1) &= -\mathbf{H}^{-1}(z_1) + \sum_{i \neq j}^{N} \beta_{ij}(z_1)\mathbf{H}^{-1}(z_1)\mathbf{r}_i\mathbf{r}_i^*\mathbf{D}_{ij}^{-1}(z_1) \\
&\quad - \frac{N-1}{N} b_1(z_1)\mathbf{H}^{-1}(z_1)\mathbf{D}_j^{-1}(z_1) \\
&= -\mathbf{H}^{-1}(z_1) + b_1(z_1)A(z_1) + B(z_1) + C(z_1), \qquad (4.37)
\end{aligned}$$

where

$$\begin{aligned}
A(z_1) &= \sum_{i \neq j} \mathbf{H}^{-1}(z_1)(\mathbf{r}_i\mathbf{r}_i^* - N^{-1}\mathbf{I})\mathbf{D}_{ij}^{-1}(z_1), \\
B(z_1) &= \sum_{i \neq j} \big(\beta_{ij}(z_1) - b_1(z_1)\big)\mathbf{H}^{-1}(z_1)\mathbf{r}_i\mathbf{r}_i^*\mathbf{D}_{ij}^{-1}(z_1), \\
C(z_1) &= N^{-1} b_1(z_1)\mathbf{H}^{-1}(z_1) \sum_{i \neq j} \big(\mathbf{D}_{ij}^{-1}(z_1) - \mathbf{D}_j^{-1}(z_1)\big).
\end{aligned}$$

It is easy to verify for any real t,

$$\begin{aligned}
&\Big|1 - \frac{1}{z\big(1 + N^{-1}\mathbb{E}tr\mathbf{D}_{12}^{-1}(z)\big)} - \frac{t}{z}\Big|^{-1} = \Big|\frac{z\big(1 + N^{-1}\mathbb{E}tr\mathbf{D}_{12}^{-1}(z)\big)}{(z-t)\big(1 + N^{-1}\mathbb{E}tr\mathbf{D}_{12}^{-1}(z)\big) - 1}\Big| \\
&\le \frac{\big|z\big(1 + N^{-1}\mathbb{E}tr\mathbf{D}_{12}^{-1}(z)\big)\big|}{\mathrm{Im}\Big[(z-t)\big(1 + N^{-1}\mathbb{E}tr\mathbf{D}_{12}^{-1}(z)\big)\Big]} \le \frac{|z|(1 + n/(Nv_0))}{v_0},
\end{aligned}$$

where the last inequality uses

$$\begin{aligned}
&\mathrm{Im}\Big[(z-t)\big(1 + N^{-1}\mathbb{E}tr\mathbf{D}_{12}^{-1}(z)\big)\Big] = v_0 + \mathrm{Im}\Big[(z-t)N^{-1}\mathbb{E}tr\mathbf{D}_{12}^{-1}(z)\Big] \\
= \;& v_0 + \mathrm{Im}\Big[(z-t)N^{-1}\sum_{i=1}^{n}\mathbb{E}\frac{1}{\lambda_i - z}\Big] = v_0 + \mathrm{Im}\Big[N^{-1}\sum_{i=1}^{n}\mathbb{E}\frac{(z-t)(\lambda_i - t - (\bar{z} - t))}{|\lambda_i - z|^2}\Big] \\
= \;& v_0 + N^{-1}\sum_{i=1}^{n}\mathbb{E}\frac{(\lambda_i - t)v_0}{|\lambda_i - z|^2} \ge v_0,
\end{aligned}$$

with the fact that

$$\lambda_i \ge t, \forall i = 1, 2, \ldots, n,$$

where $\lambda_i, i = 1, 2, \ldots, n$ are eigenvalues of $\mathbf{D}_{12} = \sum_{i\neq 1,2}^{n} \mathbf{r}_i\mathbf{r}_i^T + \mathbf{T}_n$. It follows that

$$\left|\left|\mathbf{H}^{-1}(z)\right|\right| = \left|\left|\left(z\mathbf{I} - \frac{N-1}{N}b_1(z)\mathbf{I} - \mathbf{T}_n\right)^{-1}\right|\right| \leq \frac{1 + n/(Nv_0)}{v_0}. \quad (4.38)$$

Moreover from (4.31) and (4.23) we have

$$E|\gamma_j(z)|^p \leq KN^{-1}\delta_n^{2p-4}, \quad p \geq 2. \quad (4.39)$$

Therefore the discussions for (2.11)-(2.13) of Bai and Silverstein (2004) still work in our case. That is,

$$\mathbb{E}|tr\mathbf{B}(z_1)\mathbf{M}| \leq K_{\mathbf{M}}KN^{1/2}, \quad \mathbb{E}|tr\mathbf{C}(z_1)\mathbf{M}| \leq K_{\mathbf{M}}K, \quad (4.40)$$

when $K_{\mathbf{M}}$ denotes the nonrandom bound of the spectral norm of $\mathbf{M}$, an $n \times n$ matrix; When $\mathbf{M}$ is non-random, we also have for any j,

$$\mathbb{E}|tr\mathbf{A}(z_1)\mathbf{M}| \leq K\|\mathbf{M}\|N^{1/2}, \quad (4.41)$$

where $\|\mathbf{M}\|$ denotes the spectral norm of a matrix.

Using an identity similar to (4.25) yields

$$tr\mathbb{E}_j\big(\mathbf{A}(z_1)\big)\mathbf{D}_j^{-1}(z_2) = A_1(z_1, z_2) + A_2(z_1, z_2) + A_3(z_1, z_2), \quad (4.42)$$

where

$$A_1(z_1, z_2) = -tr\sum_{i<j}\mathbf{H}^{-1}(z_1)\mathbf{r}_i\mathbf{r}_i^*\mathbb{E}_j\big(\mathbf{D}_{ij}^{-1}(z_1)\big)\beta_{ij}(z_2)\mathbf{D}_{ij}^{-1}(z_2)\mathbf{r}_i\mathbf{r}_i^*\mathbf{D}_{ij}^{-1}(z_2)$$

$$= \quad -\sum_{i<j}\beta_{ij}(z_2)\mathbf{r}_i^*\mathbb{E}_j\big(\mathbf{D}_{ij}^{-1}(z_1)\big)\mathbf{D}_{ij}^{-1}(z_2)\mathbf{r}_i\mathbf{r}_i^*\mathbf{D}_{ij}^{-1}(z_2)\mathbf{H}^{-1}(z_1)\mathbf{r}_i,$$

$$A_2(z_1, z_2) = -tr\sum_{i<j}\mathbf{H}^{-1}(z_1)N^{-1}\mathbb{E}_j\big(\mathbf{D}_{ij}^{-1}(z_1)\big)\big(\mathbf{D}_j^{-1}(z_2) - \mathbf{D}_{ij}^{-1}(z_2)\big),$$

$$A_3(z_1, z_2) = tr\sum_{i<j}\mathbf{H}^{-1}(z_1)(\mathbf{r}_i\mathbf{r}_i^* - N^{-1}\mathbf{I})\mathbb{E}_j\big(\mathbf{D}_{ij}^{-1}(z_1)\big)\mathbf{D}_{ij}^{-1}(z_2).$$

By arguments similar to (2.15) in Bai and Silverstein (2004), (4.38) and (4.23) we have

$$|A_2(z_1,z_2)| \le K, \quad \mathbb{E}|A_3(z_1,z_2)| \le KN^{1/2}.$$

The arguments above (2.16) of Bai and Silverstein (2004) can be carried over to the present setting and therefore we obtain

$$\mathbb{E}\Big|A_1(z_1,z_2) + \frac{j-1}{N^2}b_1(z_2)tr\Big(\mathbb{E}_j\big(\mathbf{D}_j^{-1}(z_1)\big)\mathbf{D}_j^{-1}(z_2)\Big)tr\mathbf{D}_j^{-1}(z_2)\mathbf{H}^{-1}(z_1)\Big| \le KN^{1/2}. \tag{4.43}$$

We conclude from (4.37)-(4.43) that

$$\begin{aligned} & tr\Big(\mathbb{E}_j\big(\mathbf{D}_j^{-1}(z_1)\big)\mathbf{D}_j^{-1}(z_2)\Big)\Big(1+\frac{j-1}{N^2}b_1(z_1)b_1(z_2)tr\Big(\mathbf{D}_j^{-1}(z_2)\mathbf{H}^{-1}(z_1)\Big) \\ = & -tr\Big(\mathbf{H}^{-1}(z_1)\mathbf{D}_j^{-1}(z_2)\Big) + A_4(z_1,z_2), \end{aligned} \tag{4.44}$$

where

$$\mathbb{E}|A_4(z_1,z_2)| \le KN^{1/2}.$$

Applying the expression for $\mathbf{D}_j^{-1}(z_2)$ in (4.37), (4.40) and (4.41), we obtain

$$\begin{aligned} & tr\Big(\mathbb{E}_j\big(\mathbf{D}_j^{-1}(z_1)\big)\mathbf{D}_j^{-1}(z_2)\Big) \times \Big(1-\frac{j-1}{N^2}b_1(z_1)b_1(z_2)tr\mathbf{H}^{-1}(z_1)\mathbf{H}^{-1}(z_2)\Big) \\ = & \; tr\Big(\mathbf{H}^{-1}(z_1)\mathbf{H}^{-1}(z_2)\Big) + A_5(z_1,z_2), \end{aligned} \tag{4.45}$$

where

$$E|A_5(z_1,z_2)| \le KN^{1/2}.$$

Since $(b_1(z) - \frac{1}{1+c_n m_n^0(z)}) \to 0$ (indeed, the next subsection proves $Em_n(z) - m_n^0(z) = O(N^{-1})$), we have

$$\frac{1}{N}tr\Big(\mathbb{E}_j\big(\mathbf{D}_j^{-1}(z_1)\big)\mathbf{D}_j^{-1}(z_2)\Big)$$

$$\times\Big(1-\frac{j-1}{N}c_n\frac{1}{\big(1+cm_n^0(z_1)\big)\big(1+cm_n^0(z_2)\big)}$$

$$\int\frac{dH_n(t)}{\big(z_2-\frac{1}{1+cm_n^0(z_2)}-t\big)\big(z_1-\frac{1}{1+cm_n^0(z_1)}-t\big)}\Big)$$

$$= \quad c_n\int\frac{dH_n(t)}{\big(z_2-\frac{1}{1+cm_n^0(z_2)}-t\big)\big(z_1-\frac{1}{1+cm_n^0(z_1)}-t\big)}+A_6(z_1,z_2), \tag{4.46}$$

where $E|A_6(z_1,z_2)|=o(1)$. Let

$$a_n(z_1,z_2)=c_n\frac{1}{\big(1+cm_n^0(z_1)\big)\big(1+cm_n^0(z_2)\big)}\int\frac{dH_n(t)}{\big(z_2-\frac{1}{1+cm_n^0(z_2)}-t\big)\big(z_1-\frac{1}{1+cm_n^0(z_1)}-t\big)}.$$

We claim that

$$|a_n(z_1,z_2)|<1. \tag{4.47}$$

Indeed, by the Cauchy-Schwartz inequality, we have

$$\Big|\frac{c_n}{\big(1+c_nm_n^0(z_1)\big)\big(1+c_nm_n^0(z_2)\big)}\int\frac{dH_n(t)}{\big(z_2-\frac{1}{1+c_nm_n^0(z_2)}-t\big)\big(z_1-\frac{1}{1+c_nm_n^0(z_1)}-t\big)}\Big|$$

$$\le \quad \Big(\int\frac{c_ndH_n(t)}{|1+c_nm_n^0(z_1)|^2|z_1-\frac{1}{1+c_nm_n^0(z_1)}-t|^2}\Big)^{1/2}$$

$$\Big(\int\frac{c_ndH_n(t)}{|1+c_nm_n^0(z_2)|^2|z_2-\frac{1}{1+c_nm_n^0(z_2)}-t|^2}\Big)^{1/2}. \tag{4.48}$$

Note that $m_n^0(z)$ satisfies an equality similar to (4.5)

$$m_n^0(z)=\int\frac{dH_n(t)}{t-z+\frac{1}{1+c_nm_n^0(z)}}. \tag{4.49}$$

Taking the imaginary part of the both sides of (4.49) leads to

$$\mathrm{Im}\big(m_n^0(z)\big)=\int\frac{\mathrm{Im}\Big(\overline{t-z+\frac{1}{1+c_nm_n^0(z)}}\Big)dH_n(t)}{|t-z+\frac{1}{1+c_nm_n^0(z)}|^2}$$

$$= \quad v_0\int\frac{dH_n(t)}{|t-z+\frac{1}{1+c_nm_n^0(z)}|^2}+\frac{c_n\mathrm{Im}(m_n^0(z))}{|1+c_nm_n^0(z)|^2}\int\frac{dH_n(t)}{|t-z+\frac{1}{1+c_nm_n^0(z)}|^2}.$$

Dividing by $\mathrm{Im}\big(m_n^0(z)\big)$ on both sides, we have

$$\frac{c_n}{|1+c_n m_n^0(z)|^2}\int \frac{dH_n(t)}{|t-z+\frac{1}{1+c_n m_n^0(z)}|^2} = 1-\frac{v_0}{\mathrm{Im}\big(m_n^0(z)\big)}\int \frac{dH_n(t)}{|t-z+\frac{1}{1+c_n m_n^0(z)}|^2} < 1.$$

This, together with (4.48), yields (4.47).

It follows from (4.46) and (4.47) that (4.35) can be written as

$$a_n(z_1,z_2)\frac{1}{N}\sum_{j=1}^{N}\frac{1}{1-\big((j-1)/N\big)a_n(z_1,z_2)} + A_7(z_1,z_2),$$

where $E|A_7(z_1,z_2)| = o(1)$. We then conclude that

$$(4.35) \stackrel{i.p.}{\to} a(z_1,z_2)\int_0^1 \frac{1}{1-ta(z_1,z_2)}dt = \int_0^{a(z_1,z_2)} \frac{1}{1-z}dz,$$

where

$$\begin{aligned}
a(z_1,z_2) &= \frac{c}{\big(1+cm^0(z_1)\big)\big(1+cm^0(z_2)\big)}\int \frac{dH(t)}{\big(z_2-\frac{1}{1+cm^0(z_2)}-t\big)\big(z_1-\frac{1}{1+cm^0(z_1)}-t\big)} \\
&= \frac{c\big(m^0(z_2)-m^0(z_1)\big)}{\big(1+cm^0(z_1)\big)\big(1+cm^0(z_2)\big)}\frac{1}{z_2-z_1+\frac{1}{1+cm^0(z_1)}-\frac{1}{1+cm^0(z_2)}} \\
&= \frac{s(z_1,z_2)}{z_2-z_1+s(z_1,z_2)} = 1-\frac{z_2-z_1}{z_2-z_1+s(z_1,z_2)},
\end{aligned}$$

where the second equality uses (4.5) and $s(z_1,z_2) = \frac{1}{1+cm^0(z_1)} - \frac{1}{1+cm^0(z_2)}$. Therefore the limit of (4.28) under the complex case is

$$\frac{\partial^2}{\partial z_2 \partial z_1}\int_0^{a(z_1,z_2)} \frac{1}{1-z}dz = \frac{\partial}{\partial z_2}\Big(\frac{\partial a(z_1,z_2)/\partial z_1}{1-a(z_1,z_2)}\Big).$$

$$= \frac{\partial}{\partial z_2}\Big[\frac{s(z_1,z_2)+(z_1-z_2)\frac{c(m^0(z_1))'}{(1+cm^0(z_1))^2}}{(z_2-z_1+s(z_1,z_2))(z_2-z_1)}\Big]$$

$$= \frac{\partial}{\partial z_2}\Big[\frac{1}{z_2-z_1} - (1+\frac{c(m^0(z_1))'}{(1+cm^0(z_1))^2})\frac{1}{z_2-z_1+s(z_1,z_2)}\Big]$$

$$= -\frac{1}{(z_2-z_1)^2} + [1+\frac{c(m^0(z_1))'}{(1+cm^0(z_1))^2} + \frac{c(m^0(z_2))'}{(1+cm^0(z_2))^2} + \frac{c(m^0(z_1))'}{(1+cm^0(z_1))^2}\frac{c(m^0(z_2))'}{(1+cm^0(z_2))^2}] \times \frac{1}{(z_2-z_1+s(z_1,z_2))^2}.$$

4.1.3.2 Step 2: Tightness of $\hat{M}_n^{(1)}(z)$

The tightness of $\{\sum_{\ell=1}^k a_\ell \hat{M}_n^{(1)}(z)\}$ on $z \in \mathcal{C}$ can be proved in the same way as that in Bai and Silverstein (2004).

4.1.3.3 Step 3: Convergence of $M_n^{(2)}(z)$

We first list some results from Sections 3 and 4 in Bai and Silverstein (2004), which hold in the present setting as well. Consider $z \in \mathcal{C}_n^+$. As in (3.5), (3.6) and the argument below (3.6) of Bai and Silverstein (2004) we have

$$\mathbb{E}|\gamma_j|^p \le KN^{-1}\delta_n^{2p-4}, \quad p \ge 2 \tag{4.50}$$

and

$$\mathbb{E}|\beta_1(z)|^p \le K, \quad p \ge 1, \quad |b_n(z)| \le K. \tag{4.51}$$

Similar to (3.1) and (3.2) in Bai and Silverstein (2004), by (4.16), we have for any positive p

$$\max\Big(\mathbb{E}||\mathbf{D}^{-1}(z)||^p, \mathbb{E}||\mathbf{D}_j^{-1}(z)||^p, \mathbb{E}||\mathbf{D}_{ij}^{-1}(z)||^p\Big) \le K \tag{4.52}$$

and via (4.23) and (4.12)

$$\left|\mathbb{E}\Big(a(v)\prod_{m=1}^{q}\big(\mathbf{r}_1^*\mathbf{B}_m(v)\mathbf{r}_1 - N^{-1}tr\mathbf{B}_m(v)\big)\Big)\right| \le KN^{-(1\wedge q)}\delta_n^{(2q-4)\vee 0}, \quad q \ge 0, \tag{4.53}$$

where the matrices $\mathbf{B}_m(v)$ are independent of $\mathbf{r}_1$ and

$$\max(|a(v)|, ||\mathbf{B}_m(v)||) \le K\big(1 + n^s I(||\mathbf{B}_n|| \ge \eta_r \ or \ \lambda_{\min}^{\widetilde{\mathbf{B}}} \le \eta_\ell)\big)$$

for some positive s, with $\widetilde{\mathbf{B}}$ being $\mathbf{B}_n$ or $\mathbf{B}_n$ with one or two of the $\mathbf{r}_j$'s removed. Here $\eta_r \in \big((1+\sqrt{c})^2 + \limsup_n ||\mathbf{T}_n||, x_r\big)$. If $x_\ell > 0$, then

$\eta_\ell \in \big(x_\ell, (1-\sqrt{c})^2 + \liminf_n \lambda_{\min}^{\mathbf{T}_n}\big)$; if $x_\ell < 0$, then $\eta_\ell < 0$. Similar to (4.1) in Bai and Silverstein (2004), one may prove as $n \to \infty$,

$$\sup_{z\in\mathcal{C}_n^+} |\mathbb{E}m_n(z) - m^0(z)| \to 0. \tag{4.54}$$

Let $\mathbf{M}$ be an $n\times n$ non-random matrix. With the same arguments as (4.7) in Bai and Silverstein (2004) we obtain

$$\mathbb{E}|tr\mathbf{D}_1^{-1}(z)\mathbf{M} - \mathbb{E}tr\mathbf{D}_1^{-1}(z)\mathbf{M}|^2 \le K||\mathbf{M}||^2. \tag{4.55}$$

We next show

$$\sup_{z\in\mathcal{C}_n^+} ||\Big((\mathbb{E}\beta_1)\mathbf{I} - z\mathbf{I} + \mathbf{T}_n\Big)^{-1}|| < \infty. \tag{4.56}$$

Denote the supports of the distributions H and $F_{c,H}$ by S_H and $S_{F_{c,H}}$ respectively. We see that $||\Big((\mathbb{E}\beta_1)\mathbf{I} - z\mathbf{I} + \mathbf{T}_n\Big)^{-1}||$ is bounded by $2\frac{1+n/(Nv_0)}{v_0}$ on $\mathcal{C}_u$ by (4.38) and (4.39).

Consider $x = x_\ell$ or x_r now. So $x \in S^c_{F_{c,H}}$, where $S^c_{F_{c,H}}$ denotes the complement of $S_{F_{c,H}}$. We next prove that $t - x + \frac{1}{1+cm^0(x)} \neq 0$ for any $t \in S_H$ and $x \in I \subset S^c_{F_{c,H}}$ where I is an open interval by following a line similar to Theorem 4.1 of Silverstein and Choi (1995). For any $x_0 \in I$, let $m_0 = m^0(x_0)$ and $D = \{z \in \mathbb{C} : \mathrm{Im}z > 0\}$. Let $m = m(z) = z - \frac{1}{1+cm^0(z)} \in D$ (for $z \in D$). From (4.5) we have

$$z(m) = m + \frac{1}{1+cm_H(m)}. \tag{4.57}$$

Since $m'(x_0) = 1 + \frac{(m^0(x_0))'}{(1+cm^0(x_0))^2} > 0$, $m(z)$ has an inverse $\tilde{z}(m)$ in a neighborhood V of x_0 by the inverse function theory. By the open mapping theorem $m(V)$ is open and includes $(x_0 - \frac{1}{1+cm_0})$. It follows that $\tilde{z}(m) \to x_0$ as $m \in m(V) \to (x_0 - \frac{1}{1+cm_0})$. However we must have $\tilde{z}(m) = z(m)$ on

$m(V \cap D) = m(V) \cap D$ due to (4.57) and (4.5). Therefore we have $z(m) \to x_0$ as $m \in D \to (x_0 - \frac{1}{1+cm_0})$.

(4.57) can be further rewritten as

$$m_H(m) = \frac{1}{c(z(m) - m)} - \frac{1}{c}.$$

Hence $m_H(m)$ converges to a real number when $m \in D \to (x_0 - \frac{1}{1+cm_0})$. By Theorem 2.1 of Silverstein and Choi (1995) $H'(x_0 - \frac{1}{1+cm_0}) = 0$. This implies $H^{'} = 0$ on the set $J \equiv \{x - \frac{1}{1+cm^0(x)} : x \in I \subset S^c_{F_{c,H}}\}$ which is open due to the monotonicity of $(x - \frac{1}{1+cm^0(x)})$ on I. Hence H is constant on J which implies that $J \subset S^c_H$. Therefore if t is in the support of H, we then have $t \neq x - \frac{1}{1+cm^0(x)}$, i.e. $t - x + \frac{1}{1+cm^0(x)} \neq 0$.

Since $m^0(z)$ is continuous on $\mathcal{C}^0 \equiv \{x + iv : v \in [0, v_0]\}$, there exist positive constants η and κ such that for t_0 in the support of $H(x)$

$$\inf_{z \in \mathcal{C}^0} |t_0 - z + \frac{1}{1 + cm^0(z)}| > \eta \quad \textit{and} \quad \sup_{z \in \mathcal{C}^0} |m^0(z)| < \kappa. \tag{4.58}$$

Also from (4.50), (4.51) and (4.54) we have

$$\sup_{z \in \mathcal{C}_n^+} |E\beta_1 - \frac{1}{1 + cm^0(z)}| \to 0. \tag{4.59}$$

Moreover, since $F^{\mathbf{T}_n} \xrightarrow{D} H(x)$, for all large n, there exists an eigenvalue μ of $\mathbf{T}_n$ such that

$$|\mu - t_0| < \eta/4. \tag{4.60}$$

We conclude from (4.60), (4.59) and (4.58) that

$$\inf_{z \in \mathcal{C}_\ell \cup \mathcal{C}_r} |\mu - z + E\beta_1| > \eta/2, \tag{4.61}$$

which ensures (4.56).

With $\mathbf{H}_1 = \mathbb{E}\beta_1(z)\mathbf{I} - z\mathbf{I} + \mathbf{T}_n$, write

$$\mathbf{D}(z) - \mathbf{H}_1 = \sum_{j=1}^{N} \mathbf{r}_j\mathbf{r}_j^* - (\mathbb{E}\beta_1(z))\mathbf{I}. \tag{4.62}$$

Postmultiplying $\mathbf{D}^{-1}(z)$ and premuliplying $\mathbf{H}_1^{-1}$ on the both sides, taking expectation and using an equality similar to (4.36) we get

$$\begin{aligned}
&\mathbf{H}_1^{-1} - \mathbb{E}\mathbf{D}^{-1}(z) = \mathbf{H}_1^{-1}\mathbb{E}\Big[\Big(\sum_{j=1}^{N} \mathbf{r}_j\mathbf{r}_j^* - (\mathbb{E}\beta_1(z))\mathbf{I}\Big)\mathbf{D}^{-1}(z)\Big] \\
= \;& \mathbf{H}_1^{-1}\sum_{j=1}^{N}\mathbb{E}\Big(\mathbf{r}_j\beta_j(z)\mathbf{r}_j^*\mathbf{D}_j^{-1}(z)\Big) - \mathbf{H}_1^{-1}(\mathbb{E}\beta_1(z))\mathbb{E}\mathbf{D}^{-1}(z) \\
= \;& N\mathbb{E}\Big[\beta_1(z)\Big(\mathbf{H}_1^{-1}\mathbf{r}_1\mathbf{r}_1^*\mathbf{D}_1^{-1}(z) - \frac{1}{N}\mathbf{H}_1^{-1}\mathbb{E}\mathbf{D}^{-1}(z)\Big)\Big]. \qquad (4.63)
\end{aligned}$$

Taking trace on both sides, we have

$$\begin{aligned}
&n\Big(\int \frac{dH_n(x)}{x - (z - \mathbb{E}\beta_1(z))} - \mathbb{E}m_n(z)\Big) \\
= \;& N\mathbb{E}\Big[\beta_1(z)\Big(\mathbf{r}_1^*\mathbf{D}_1^{-1}(z)\mathbf{H}_1^{-1}\mathbf{r}_1 - \frac{1}{N}tr\mathbf{H}_1^{-1}\mathbb{E}\mathbf{D}^{-1}(z)\Big)\Big]. \qquad (4.64)
\end{aligned}$$

When there is no confusion, we below drop z from $\beta_1(z), \gamma_1(z), b_n(z)$, etc. By (4.25), we have

$$\begin{aligned}
&\mathbb{E}tr\mathbf{H}_1^{-1}\mathbf{D}_1^{-1}(z) - \mathbb{E}tr\mathbf{H}_1^{-1}\mathbf{D}^{-1}(z) = \mathbb{E}\Big[\beta_1(z)tr\mathbf{H}_1^{-1}\mathbf{D}_1^{-1}(z)\mathbf{r}_1\mathbf{r}_1^*\mathbf{D}_1^{-1}(z)\Big] \\
= \;& b_n(z)\mathbb{E}\Big[(1 - \beta_1\gamma_1)\mathbf{r}_1^*\mathbf{D}_1^{-1}(z)\mathbf{H}_1^{-1}\mathbf{D}_1^{-1}(z)\mathbf{r}_1\Big], \qquad (4.65)
\end{aligned}$$

where the last equality uses $\beta_1 = b_n - \beta_1 b_n\gamma_1$. In view of (4.53), (4.50) and (4.56), we obtain

$$\Big|\mathbb{E}\beta_1(z)\gamma_1\mathbf{r}_1^*\mathbf{D}_1^{-1}(z)\mathbf{H}_1^{-1}\mathbf{D}_1^{-1}(z)\mathbf{r}_1\Big| \le KN^{-1}, \tag{4.66}$$

which implies that

$$\Big|(4.65) - N^{-1}b_n\mathbb{E}\mathbf{D}_1^{-1}(z)\mathbf{H}_1^{-1}\mathbf{D}_1^{-1}(z)\Big| \le KN^{-1}.$$

Since $\beta_1 = b_n - b_n^2\gamma_1 + \beta_1 b_n^2\gamma_1^2$ we may write

$$\begin{aligned}
& N\mathbb{E}(\beta_1\mathbf{r}_1^*\mathbf{D}_1^{-1}(z)\mathbf{H}_1^{-1}\mathbf{r}_1) - \mathbb{E}\beta_1\mathbb{E}tr\mathbf{H}_1^{-1}\mathbf{D}_1^{-1}(z) \\
= \; & -b_n^2 N\mathbb{E}(\gamma_1\mathbf{r}_1^*\mathbf{D}_1^{-1}(z)\mathbf{H}_1^{-1}\mathbf{r}_1) \\
& + b_n^2\Big(N\mathbb{E}(\beta_1\gamma_1^2\mathbf{r}_1^*\mathbf{D}_1^{-1}\mathbf{H}_1^{-1}\mathbf{r}_1)\big(\mathbb{E}\beta_1\gamma_1^2\big)\mathbb{E}tr\mathbf{H}_1^{-1}\mathbf{D}_1^{-1}(z)\Big) \\
= \; & -b_n^2 N\mathbb{E}(\gamma_1\mathbf{r}_1^*\mathbf{D}_1^{-1}(z)\mathbf{H}_1^{-1}\mathbf{r}_1) + b_n^2(z)Cov\Big(\beta_1\gamma_1^2, tr\mathbf{D}_1^{-1}(z)\mathbf{H}_1^{-1}\Big) \\
& + b_n^2\Big(\mathbb{E}[N\beta_1\gamma_1^2\mathbf{r}_1^*\mathbf{D}_1^{-1}(z)\mathbf{H}_1^{-1}\mathbf{r}_1 - \beta_1\gamma_1^2 tr\mathbf{D}_1^{-1}\mathbf{H}_1^{-1}]\Big).
\end{aligned}$$

One may refer to a similar expansion on page 587 of Bai and Silverstein (2004). It follows from (4.50), (4.56) and (4.53) that

$$\Big|\mathbb{E}[N\beta_1(z)\gamma_1^2(z)\mathbf{r}_1^*\mathbf{D}_1^{-1}(z)\mathbf{H}_1^{-1}\mathbf{r}_1 - \beta_1\gamma_1^2 tr\mathbf{D}_1^{-1}\mathbf{H}_1^{-1}]\Big| \le K\delta_n^2.$$

By (4.51), (4.50), (4.56) and (4.55) we have

$$\begin{aligned}
& \Big|Cov\Big(\beta_1\gamma_1^2, tr\mathbf{D}_1^{-1}(z)\mathbf{H}_1^{-1}\Big)\Big| \\
\le \; & (\mathbb{E}|\beta_1|^4)^{1/4}(\mathbb{E}|\gamma_1|^8)^{1/4}\Big(\mathbb{E}\Big|tr\mathbf{D}_1^{-1}(z)\mathbf{H}_1^{-1} - \mathbb{E}tr\mathbf{D}_1^{-1}(z)\mathbf{H}_1^{-1}\Big|^2\Big)^{1/2} \\
\le \; & K\delta_n^3 N^{-1/4}.
\end{aligned}$$

We conclude from (4.50), (4.51) and $\beta_1 = b_n - \beta_1 b_n\gamma_1$ that

$$\mathbb{E}\beta_1 = b_n + O(N^{-1/2}).$$

By the definition of γ_1 we have

$$\begin{aligned}
& \mathbb{E}N\gamma_1\mathbf{r}_1^*\mathbf{D}_1^{-1}\mathbf{H}_1^{-1}\mathbf{r}_1 \\
= \; & N\mathbb{E}\Big[\Big(\mathbf{r}_1^*\mathbf{D}_1^{-1}\mathbf{r}_1 - N^{-1}tr\mathbf{D}_1^{-1}\Big)\Big(\mathbf{r}_1^*\mathbf{D}_1^{-1}\mathbf{H}_1^{-1}\mathbf{r}_1 - N^{-1}tr\mathbf{D}_1^{-1}\mathbf{H}_1^{-1}\Big)\Big] \\
& + N^{-1}Cov\Big(tr\mathbf{D}_1^{-1}, tr\mathbf{D}_1^{-1}(z)\mathbf{H}_1^{-1}\Big). \qquad (4.67)
\end{aligned}$$

In view of (4.55), we see the second term above is $O(N^{-1})$. We conclude from (4.64)-(4.67) that

$$n\Big(\int \frac{dH_n(x)}{x - (z - \mathbb{E}\beta_1(z))} - \mathbb{E}m_n(z)\Big)$$

$$
\begin{aligned}
&= o(1) && (4.68)\\
&\quad +b_n^2(z)N^{-1}\mathbb{E}tr\mathbf{D}_1^{-1}\mathbf{H}_1^{-1}\mathbf{D}_1^{-1} && (4.69)\\
&\quad -b_n^2 N\mathbb{E}\Big[\Big(\mathbf{r}_1^*\mathbf{D}_1^{-1}\mathbf{r}_1 - N^{-1}tr\mathbf{D}_1^{-1}\Big)\Big(\mathbf{r}_1^*\mathbf{D}_1^{-1}\mathbf{H}_1^{-1}\mathbf{r}_1 - N^{-1}tr\mathbf{D}_1^{-1}\mathbf{H}_1^{-1}\Big)\Big] && (4.70)
\end{aligned}
$$

Using (4.33) on (4.68) and by the assumptions under the complex case, we have

$$
n\Big(\int\frac{dH_n(x)}{x-(z-\mathbb{E}\beta_1(z))} - \mathbb{E}m_n(z)\Big) \to 0, \quad as \quad N\to\infty,
$$

while under the real case

$$
n\Big(\int\frac{dH_n(x)}{x-(z-\mathbb{E}\beta_1(z))} - \mathbb{E}m_n(z)\Big) = -b_n^2N^{-1}\mathbb{E}tr\mathbf{D}_1^{-1}\mathbf{H}_1^{-1}\mathbf{D}_1^{-1} + o(1). \tag{4.71}
$$

It is sufficient to find the limit of $N^{-1}\mathbb{E}tr\mathbf{D}_1^{-1}\mathbf{H}_1^{-1}\mathbf{D}_1^{-1}$. Applications of (4.25),(4.51),(4.53) and (4.56) ensure that

$$
\mathbb{E}tr\mathbf{D}_1^{-1}\mathbf{H}_1^{-1}\mathbf{D}_1^{-1} - \mathbb{E}tr\mathbf{D}^{-1}\mathbf{H}_1^{-1}\mathbf{D}_1^{-1}
$$

and

$$
\mathbb{E}tr\mathbf{D}^{-1}\mathbf{H}_1^{-1}\mathbf{D}_1^{-1} - \mathbb{E}tr\mathbf{D}^{-1}\mathbf{H}_1^{-1}\mathbf{D}^{-1}
$$

are bounded. Hence it then reduces to considering the limit of

$$
N^{-1}\mathbb{E}tr\mathbf{D}^{-1}\mathbf{H}_1^{-1}\mathbf{D}^{-1}. \tag{4.72}
$$

From (4.62), similar to (4.63) we have

$$
\begin{aligned}
\mathbf{D}^{-1}(z) &= \mathbf{H}_1^{-1} - \sum_{j=1}^{N}\beta_j\mathbf{H}_1^{-1}\mathbf{r}_j\mathbf{r}_j^*\mathbf{D}_j^{-1}(z) + \big(\mathbb{E}\beta_1\big)\mathbf{H}_1^{-1}\mathbf{D}^{-1}(z)\\
&= \mathbf{H}_1^{-1} + \big(\mathbb{E}\beta_1\big)A(z) + B(z) + C(z), && (4.73)
\end{aligned}
$$

where

$$A(z) = -\sum_{j=1}^{N} \mathbf{H}_1^{-1}\Big(\mathbf{r}_j\mathbf{r}_j^* - N^{-1}\mathbf{I}\Big)\mathbf{D}_j^{-1}(z),$$

$$B(z) = -\sum_{j=1}^{N} \Big(\beta_j - \mathbb{E}\beta_1\Big)\mathbf{H}_1^{-1}\mathbf{r}_j\mathbf{r}_j^*\mathbf{D}_j^{-1}(z),$$

$$C(z) = -N^{-1}\big(\mathbb{E}\beta_1\big)\mathbf{H}_1^{-1}\sum_{j=1}^{N}\Big(\mathbf{D}_j^{-1}(z) - \mathbf{D}^{-1}(z)\Big)$$

$$= \quad -N^{-1}\big(\mathbb{E}\beta_1\big)\mathbf{H}_1^{-1}\sum_{j=1}^{N}\beta_j\mathbf{D}_j^{-1}(z)\mathbf{r}_j\mathbf{r}_j^*\mathbf{D}_j^{-1}(z).$$

It follows from (4.50) and (4.51) that

$$\mathbb{E}|\beta_1 - \mathbb{E}\beta_1|^2 \leq KN^{-1}. \tag{4.74}$$

For any $n \times n$ matrix $\mathbf{M}$, by (4.52), (4.53), (4.51) and (4.74) we obtain

$$\begin{aligned} |N^{-1}\mathbb{E}tr\mathbf{B}(z)\mathbf{M}| &\leq K\Big(\mathbb{E}|\beta_1(z) - \mathbb{E}\beta_1(z)|^2\Big)^{1/2}\Big(\mathbb{E}\big|\mathbf{r}_1^*\mathbf{r}_1||\mathbf{D}_1^{-1}\mathbf{M}||\big|^2\Big)^{1/2} \\ &\leq KN^{-1/2}(\mathbb{E}||\mathbf{M}||^4)^{1/4} \end{aligned} \tag{4.75}$$

and

$$\begin{aligned} |N^{-1}\mathbb{E}tr\mathbf{C}(z)\mathbf{M}| &\leq KN^{-1}\mathbb{E}|\beta_1(z)|\mathbf{r}_1^*\mathbf{r}_1||\mathbf{D}_1^{-1}(z)||^2||\mathbf{M}|| \\ &\leq KN^{-1}(\mathbb{E}||\mathbf{M}||^2)^{1/2}. \end{aligned} \tag{4.76}$$

For any $n \times n$ nonrandom matrix $\mathbf{M}$ with a bounded spectral norm, we write

$$tr\mathbf{A}(z)\mathbf{D}^{-1}(z)\mathbf{M} = A_1(z) + A_2(z) + A_3(z), \tag{4.77}$$

where

$$A_1(z) = -tr\sum_{j=1}^{N}\mathbf{H}_1^{-1}\mathbf{r}_j\mathbf{r}_j^*\mathbf{D}_j^{-1}(z)\Big(\mathbf{D}^{-1}(z) - \mathbf{D}_j^{-1}(z)\Big)\mathbf{M},$$

$$A_2(z) = -tr\sum_{j=1}^{N}\mathbf{H}_1^{-1}\Big(\mathbf{r}_j\mathbf{r}_j^*\mathbf{D}_j^{-2}(z) - N^{-1}\mathbf{D}_j^{-2}(z)\Big)\mathbf{M},$$
$$A_3(z) = -tr\sum_{j=1}^{N}\mathbf{H}_1^{-1}N^{-1}\mathbf{D}_j^{-1}(z)\Big(\mathbf{D}_j^{-1}(z) - \mathbf{D}^{-1}(z)\Big)\mathbf{M}.$$

Obviously $\mathbb{E}A_2(z) = 0$ and similar to (4.76), we obtain

$$|\mathbb{E}N^{-1}A_3(z)| \le KN^{-1}. \tag{4.78}$$

From (4.50) and (4.51)

$$\mathbb{E}|\beta_1 - b_n|^2 \le KN^{-1}. \tag{4.79}$$

Using (4.53), (4.79) and (4.25) yields

$$\begin{aligned}
\mathbb{E}N^{-1}A_1(z) &= \mathbb{E}\Big[\beta_1\mathbf{r}_1^*\mathbf{D}_1^{-2}(z)\mathbf{r}_1\mathbf{r}_1^*\mathbf{D}_1^{-1}(z)\mathbf{M}\mathbf{H}_1^{-1}\mathbf{r}_1\Big]\\
&= b_n\mathbb{E}\Big[\Big(N^{-1}tr\mathbf{D}_1^{-2}(z)\Big)\Big(N^{-1}tr\mathbf{D}_1^{-1}(z)\mathbf{M}\mathbf{H}_1^{-1}\Big)\Big] + o(1)\\
&= b_n\mathbb{E}\Big[\Big(N^{-1}tr\mathbf{D}^{-2}(z)\Big)\Big(N^{-1}tr\mathbf{D}^{-1}(z)\mathbf{M}\mathbf{H}_1^{-1}\Big)\Big] + o(1).
\end{aligned}$$

By (4.55) and (4.52), we have

$$\begin{aligned}
&\Big|Cov\Big(N^{-1}tr\mathbf{D}^{-2}(z), N^{-1}tr\mathbf{D}^{-1}(z)\mathbf{M}\mathbf{H}_1^{-1}\Big)\Big|\\
\le\ &\Big(\mathbb{E}|N^{-1}tr\mathbf{D}^{-2}(z)|^2\Big)^{1/2}N^{-1}\Big(\mathbb{E}\Big|tr\mathbf{D}^{-1}(z)\mathbf{M}\mathbf{H}_1^{-1} - \mathbb{E}tr\mathbf{D}^{-1}\mathbf{M}\mathbf{H}_1^{-1}\Big|^2\Big)^{1/2}\\
\le\ &KN^{-1}.
\end{aligned}$$

We thus have

$$\mathbb{E}N^{-1}A_1(z) = b_n\Big(\mathbb{E}N^{-1}tr\mathbf{D}^{-2}(z)\Big)\Big(\mathbb{E}N^{-1}tr\mathbf{D}^{-1}(z)\mathbf{M}\mathbf{H}_1^{-1}\Big) + o(1). \tag{4.80}$$

Moreover, by (4.73), (4.75) and (4.76), we have

$$\mathbb{E}N^{-1}tr\mathbf{D}^{-1}(z)\mathbf{H}_1^{-2} = N^{-1}tr\Big(\mathbf{H}_1^{-1} + \mathbb{E}B(z) + \mathbb{E}C(z)\Big)\mathbf{H}_1^{-2}$$

$$= c_n \int \frac{dH_n(x)}{\big(x - z + \mathbb{E}\beta_1\big)^3} + o(1). \tag{4.81}$$

From (4.73)-(4.81) we conclude that

$$\begin{aligned}
& N^{-1}\mathbb{E}tr\mathbf{D}^{-1}(z)\mathbf{H}_1^{-1}\mathbf{D}^{-1}(z) \\
= \; & \mathbb{E}N^{-1}tr\mathbf{D}^{-1}(z)\mathbf{H}_1^{-2} + b_n^2\Big(\mathbb{E}N^{-1}tr\mathbf{D}^{-2}(z)\Big)\Big(\mathbb{E}N^{-1}tr\mathbf{D}^{-1}(z)\mathbf{H}_1^{-2}\Big) + o(1) \\
= \; & c_n \int \frac{dH_n(x)}{\big(x - z + \mathbb{E}\beta_1\big)^3} + b_n^2 c_n^2 \mathbb{E}\int \frac{dF_n(x)}{(x-z)^2} \int \frac{dH_n(x)}{\big(x - z + \mathbb{E}\beta_1\big)^3} + o(1).
\end{aligned} \tag{4.82}$$

This, together with (4.71), (4.56), (4.54) and (4.59), leads to

$$\begin{aligned}
& n\Big(\int \frac{dH_n(x)}{x - (z - \mathbb{E}\beta_1)} - \mathbb{E}m_n(z)\Big) \qquad (4.83) \\
= \; & -c_n b_n^2 \int \frac{dH_n(x)}{\big(x - z + \mathbb{E}\beta_1\big)^3} - b_n^4 c_n^2 \mathbb{E}\int \frac{dF_n(x)}{(x-z)^2} \int \frac{dH_n(x)}{\big(x - z + \mathbb{E}\beta_1\big)^3} + o(1) \\
= \; & -c\varpi^2(z) \int \frac{dH(x)}{\big(x - z + \varpi(z)\big)^3} - \varpi^4(z)c^2 \int \frac{dF_{c,H}(x)}{(x-z)^2} \int \frac{dH(x)}{\big(x - z + \varpi(z)\big)^3} + o(1),
\end{aligned}$$

where the last step uses

$$\sup_{z\in\mathcal{C}_n} \Big|\mathbb{E}\int \frac{dF_n(x)}{(x-z)^2} - \int \frac{F_{c,H}(x)}{(x-z)^2}\Big| \to 0, \quad as \quad n \to \infty, \tag{4.84}$$

which can be proved similarly to (4.1) in Bai and Silverstein (2004).

Let $\varpi_n(z) = 1/(1 + c_n m_n^0(z))$. By (4.49) we then write

$$\begin{aligned}
& n\Big(\mathbb{E}m_n(z) - m_n^0(z)\Big) = n\Big(\mathbb{E}m_n(z) - \int \frac{dH_n(x)}{x - (z - \varpi_n(z))}\Big) \qquad (4.85) \\
= \; & n\Big(\mathbb{E}m_n(z) - \int \frac{dH_n(x)}{x - (z - \mathbb{E}\beta_1)}\Big) \\
& + n\Big(\int \frac{dH_n(x)}{x - (z - \mathbb{E}\beta_1)} - \int \frac{dH_n(x)}{x - (z - \varpi_n(z))}\Big) \\
= \; & n\Big(\mathbb{E}m_n(z) - \int \frac{dH_n(x)}{x - (z - \mathbb{E}\beta_1)}\Big) \\
& + n\Big(\varpi_n(z) - \mathbb{E}\beta_1\Big) \int \frac{dH_n(x)}{\big(x - (z - \varpi_n(z))\big)\big(x - (z - \mathbb{E}\beta_1)\big)}.
\end{aligned}$$

We next find the limit of $n\Big(\varpi_n(z)-\mathbb{E}\beta_1\Big)$. Recall that $\beta_j^{tr}(z)=1/(1+\frac{1}{N}tr\mathbf{D}_j^{-1}(z))$ and let $\beta^{tr}(z)=1/(1+\frac{1}{N}tr\mathbf{D}^{-1}(z))$ and $b(z)=1/(1+\frac{1}{N}Etr\mathbf{D}^{-1}(z))$. Write

$$\begin{aligned} n\Big(\varpi_n(z)-\mathbb{E}\beta_1\Big) &= n\Big(\varpi_n(z)-\mathbb{E}\beta^{tr}(z)\Big)+n\Big(\mathbb{E}\beta^{tr}(z)-\mathbb{E}\beta_1^{tr}(z)\Big)\\ &\quad +n(E\beta_1^{tr}(z)-E\beta_1(z)). \end{aligned} \tag{4.86}$$

First, by the fact that

$$\beta^{tr}(z)=b(z)+\beta^{tr}(z)b(z)\big(c_nEm_n(z)-c_nm_n(z)\big) \tag{4.87}$$

we have

$$\begin{aligned} & n\Big(\varpi_n(z)-\mathbb{E}\beta^{tr}(z)\Big)=n\mathbb{E}\Big[\varpi_n(z)\beta^{tr}(z)(c_nm_n(z)-c_nm_n^0(z))\Big]\\ = {} & n\mathbb{E}\Big[\varpi_n(z)b(z)(c_nm_n(z)-c_nm_n^0(z))\Big]\\ & +n\mathbb{E}\Big[\varpi_n(z)b(z)\beta^{tr}(z)\Big(c_nm_n(z)-c_nm_n^0(z)\Big)\Big(\mathbb{E}c_nm_n(z)-c_nm_n(z)\Big)\Big]\\ = {} & n\varpi_n(z)b(z)\mathbb{E}\big(c_nm_n(z)-c_nm_n^0(z)\big)+o(1), \end{aligned} \tag{4.88}$$

where via (4.54), (4.55), (4.51) and (4.87)

$$\begin{aligned} & nc_n^2b(z)\varpi_n(z)\mathbb{E}\Big[\beta^{tr}(z)\Big(m_n(z)-m_n^0(z)\Big)\Big(\mathbb{E}m_n(z)-m_n(z)\Big)\Big] \qquad (4.89)\\ = {} & nc_n^2b(z)\varpi_n(z)\Big[\mathbb{E}\Big(\beta^{tr}(z)\big(Em_n(z)-m_n^0(z)\big)\big(\mathbb{E}m_n(z)-m_n(z)\big)\Big)\\ & \quad -\mathbb{E}\Big(\beta^{tr}(z)\big(\mathbb{E}m_n(z)-m_n(z)\big)^2\Big)\Big]\\ = {} & o(1)+nc_n^2b^2(z)\varpi_n(z)\Big[\mathbb{E}\Big(\mathbb{E}m_n(z)-m_n(z)\Big)^2-\mathbb{E}\Big(\beta^{tr}(z)\big(\mathbb{E}m_n(z)-m_n(z)\big)^3\Big)\Big]\\ = {} & o(1), \end{aligned}$$

the last step using $\mathbb{E}|m_n(z)-\mathbb{E}m_n(z)|^6=O(n^{-3})$ (see the argument above (3.5) of Bai and Silverstein (2004)).

As for the second term on the right side of (4.86), by (4.25), we obtain

$$n\Big(\mathbb{E}\beta^{tr}(z)-\mathbb{E}\beta_1^{tr}(z)\Big)=\frac{n}{N}\mathbb{E}\Big[\beta^{tr}(z)\beta_1^{tr}(z)tr\Big(\mathbf{D}_1^{-1}(z)-\mathbf{D}^{-1}(z)\Big)\Big]$$
$$=\frac{n}{N}\mathbb{E}\Big[\beta^{tr}(z)\beta_1^{tr}(z)\mathbf{r}_1^*\mathbf{D}_1^{-2}(z)\mathbf{r}_1\beta_1(z)\Big]=c^2\varpi^3(z)\int\frac{dF_{c,H}(x)}{\big(x-z\big)^2}+o(1),\tag{4.90}$$

where the last step uses (4.54), (4.55), (4.51), (4.53), (4.84) and (4.79).

As for the third term on the right side of (4.86) we conclude from (4.26) and (4.53) that

$$n\Big(\mathbb{E}\beta_1^{tr}(z)-\mathbb{E}\beta_1(z)\Big)=-n\mathbb{E}\Big((\beta_j^{tr}(z))^2\beta_j(z)\varepsilon_j^2(z)\Big)\tag{4.91}$$
$$=-n\mathbb{E}\Big(\varepsilon_1^2(z)\big(\beta_1^{tr}(z)\big)^3\Big)+n\mathbb{E}\Big(\varepsilon_1^3(z)\beta_j(z)\big(\beta_1^{tr}(z)\big)^3\Big)=-n\mathbb{E}\Big(\varepsilon_1^2(z)\big(\beta_1^{tr}(z)\big)^3\Big)+o(1).$$

Moreover by (4.53), (4.55), (4.33) and (4.54) we have for the real case

$$n\mathbb{E}\Big(\varepsilon_1^2(z)\big(\beta_1^{tr}(z)\big)^3\Big)=-\frac{\mathbb{E}\varepsilon_1^2(z)}{(1+c_n\mathbb{E}m_n(z))^3}+o(1)=-2c^2\varpi^3(z)\int\frac{dF_{c,H}(x)}{(x-z)^2}+o(1),$$

while the limit is half of the above in the complex case. This implies that in the real case

$$n\Big(\mathbb{E}\beta_1^{tr}(z)-\mathbb{E}\beta_1(z)\Big)\to-2c^2\varpi^3(z)\int\frac{dF_{c,H}(x)}{(x-z)^2},\quad as\ N\to\infty,\tag{4.92}$$

while the limit is half of the above in the complex case.

Summarizing the above we conclude that

$$n\Big(\varpi_n(z)-\mathbb{E}\beta_1(z)\Big)\tag{4.93}$$
$$=\begin{cases}c_n\varpi_n(z)b(z)n\mathbb{E}\big(m_n(z)-m_n^0(z)\big)-c^2\varpi^3(z)\int\frac{dF_{c,H}(x)}{(x-z)^2}+o(1) & \text{in the real case}\\ c_n\varpi_n(z)b(z)n\mathbb{E}\big(m_n(z)-m_n^0(z)\big)+o(1) & \text{in the complex case.}\end{cases}$$

The proof for (4.47) also shows that

$$|c_n\varpi_n^2(z)\int\frac{dH_n(x)}{\big(x-(z-\varpi_n(z))\big)^2}|<1.$$

This, together with (4.85), (4.93), (4.56) and (4.54), yields

$$
\begin{aligned}
&n\Big(\mathbb{E}m_n(z)-m_n^0(z)\Big)\\
=&\begin{cases}
\dfrac{n\Big(\mathbb{E}m_n(z)-\int\frac{dH_n(x)}{x-(z-\mathbb{E}\beta_1)}\Big)-c^2\varpi^3(z)\int\frac{dF_{c,H}(x)}{(x-z)^2}\int\frac{dH_n(x)}{\big(x-(z-\varpi_n(z))\big)^2}}{1-c_n\varpi_n^2(z)\int\frac{dH_n(x)}{\big(x-(z-\varpi_n(z))\big)^2}}+o(1), & \text{in the real case}\\
o(1), & \text{in the complex case}
\end{cases}\\
\to&\begin{cases}
\dfrac{c\varpi^2(z)m_3(z)+c^2\varpi^4(z)\Big(m^0(z)\Big)'m_3(z)-c^2\varpi^3(z)\Big(m^0(z)\Big)'m_2(z)}{1-c\varpi^2(z)m_2(z)}, & \text{in the real case}\\
0, & \text{in the complex case}
\end{cases}
\end{aligned}
$$

where we use

$$
m_r(z)=\int\frac{dH(x)}{(x-z+\varpi(z))^r},\quad \Big(m^0(z)\Big)'=\int\frac{dF_{c,H}(x)}{(x-z)^2}.
$$

4.1.3.4 Step 4: Boundness and equicontinuous of $M_n^{(2)}(z)$

Boundness and equicontinuous of $M_n^{(2)}(z)$ can be similarly proved as in the last paragraph of Section 4 in Bai and Silverstein (2004).

□

4.2 Verification of Remark 2

This section is to verify the asymptotic means and covariances in Theorem 1.1 of Bai and Silverstein (2004) and in Theorem 6 are the same when $\mathbf{T}_n$ and $\mathbf{R}_n$ become zero matrix and identity matrix respectively, as pointed out in Remark 2.

Consider (4.11) first. When $\mathbf{T}_n$ is a zero matrix, by (4.5) the Stieltjes transform $m^0(z)$ satisfies the following equation

$$
m^0(z)=\frac{1}{1-z-c-cm^0(z)}. \tag{4.94}
$$

Define $\underline{B}_n = \frac{1}{N}\mathbf{X}^*\mathbf{X}$ and denote its limiting Stieltjes transform by $\underline{m}^0(z)$. Then $\underline{m}^0(z)$ and $m^0(z)$ have the relation

$$\underline{m}^0(z) = -\frac{1-c}{z} + cm^0(z). \tag{4.95}$$

By (4.94) and (4.95), we have

$$\underline{m}^0(z) = -\frac{1}{zm^0(z)} - 1. \tag{4.96}$$

Moreover, from (4.5)

$$\frac{1}{m^0(z)} = \frac{1}{1+cm^0(z)} - z. \tag{4.97}$$

Combining (4.96) with (4.97), we get

$$z\underline{m}^0(z) = -\frac{1}{1+cm^0(z)}. \tag{4.98}$$

We then conclude from (4.98) that

$$\frac{c\big(m^0(z)\big)'}{\big(1+cm^0(z)\big)^2} = -\Big(\frac{1}{1+cm^0(z)}\Big)' = \big(z\underline{m}^0(z)\big)'. \tag{4.99}$$

It follows that

$$\begin{aligned} & 1 + \frac{c(m^0(z_1))'}{(1+cm^0(z_1))^2} + \frac{c(m^0(z_2))'}{(1+cm^0(z_2))^2} + \frac{c(m^0(z_1))'}{(1+cm^0(z_1))^2}\frac{c(m^0(z_2))'}{(1+cm^0(z_2))^2} \\ = \; & 1 + \big(z_1\underline{m}^0(z_1)\big)' + \big(z_2\underline{m}^0(z_2)\big)' + \big(z_1\underline{m}^0(z_1)\big)'\big(z_1\underline{m}^0(z_1)\big)'. \end{aligned} \tag{4.100}$$

On the other hand, since (4.95) has an inverse (one may also see (1.2) in Bai and Silverstein (2004))

$$z = -\frac{1}{\underline{m}^0(z)} + \frac{c}{1+\underline{m}^0(z)}, \tag{4.101}$$

we have

$$z\big(1+\underline{m}^0(z)\big) = -1 + c - \frac{1}{\underline{m}^0(z)}. \tag{4.102}$$

From this, we have

$$\left(z\underline{m}^0(z)\right)' = \frac{\left(\underline{m}^0(z)\right)'}{\left(\underline{m}^0(z)\right)^2} - 1. \tag{4.103}$$

Thus by (4.98) and (4.102), we have

$$\begin{aligned} & z_2 - z_1 + s(z_1, z_2) = z_2\left(1 + \underline{m}^0(z_2)\right) - z_1\left(1 + \underline{m}^0(z_1)\right) \\ = \quad & -\frac{1}{\underline{m}^0(z_2)} + \frac{1}{\underline{m}^0(z_1)} = \frac{\underline{m}^0(z_2) - \underline{m}^0(z_1)}{\underline{m}^0(z_1)\underline{m}^0(z_2)}. \end{aligned} \tag{4.104}$$

We then conclude from (4.100), (4.103) and (4.104) that

$$\begin{aligned} & [1 + \frac{c(m^0(z_1))'}{(1+cm^0(z_1))^2} + \frac{c(m^0(z_2))'}{(1+cm^0(z_2))^2} + \frac{c(m^0(z_1))'}{(1+cm^0(z_1))^2}\frac{c(m^0(z_2))'}{(1+cm^0(z_2))^2}] \\ & \times \frac{1}{(z_2 - z_1 + s(z_1, z_2))^2} \\ = \quad & \Big[1 + \frac{\left(\underline{m}^0(z_1)\right)'}{(\underline{m}^0(z_1))^2} - 1 + \frac{\left(\underline{m}^0(z_2)\right)'}{(\underline{m}^0(z_2))^2} - 1 + \Big(\frac{\left(\underline{m}^0(z_1)\right)'}{(\underline{m}^0(z_1))^2} - 1\Big)\Big(\frac{\left(\underline{m}^0(z_2)\right)'}{(\underline{m}^0(z_2))^2} - 1\Big)\Big] \\ & \times \Big(\frac{\underline{m}^0(z_2) - \underline{m}^0(z_1)}{\underline{m}^0(z_1)\underline{m}^0(z_2)}\Big)^2 \\ = \quad & \frac{\left(\underline{m}^0(z_1)\right)'\left(\underline{m}^0(z_2)\right)'}{(\underline{m}^0(z_1) - \underline{m}^0(z_2))^2}. \end{aligned} \tag{4.105}$$

In view of (4.105) we see that (1.7) in Bai and Silverstein (2004) and (4.11) are the same when $\mathbf{T}_n$ is a zero matrix and $\mathbf{R}_n$ is an identity matrix.

We next consider the asymptotic mean (4.10). When $\mathbf{T}_n = \mathbf{0}$, by (4.5), we get

$$m_r(z) = (m^0(z))^r. \tag{4.106}$$

Moreover we obtain from (4.96) and (4.98)

$$m^0(z) = -\frac{1}{z(\underline{m}^0(z) + 1)}, \quad \varpi(z) = -z\underline{m}^0(z). \tag{4.107}$$

From (4.106) and (4.107), it follows that

$$\varpi^r(z)m_r(z) = \left(\underline{m}^0(z)\right)^r\left(1 + \underline{m}^0(z)\right)^{-r}. \tag{4.108}$$

This ensures that $\mathbb{E}M(z)$ in (4.21) can be written as

$$\mathbb{E}M(z) = \frac{c(\underline{m}^0(z))^3(1+\underline{m}^0(z))^{-3}\Big(\frac{1}{\varpi(z)} + c\varpi(z)(m^0(z))' - c(m^0(z))'\frac{1}{m^0(z)}\Big)}{1-c(\underline{m}^0(z))^2(1+\underline{m}^0(z))^{-2}}. \tag{4.109}$$

Comparing (4.109) with (1.6) in Bai and Silverstein (2004), it is sufficient to prove that

$$\frac{1}{\varpi(z)} + c\varpi(z)(m^0(z))' - c(m^0(z))'\frac{1}{m^0(z)} = \frac{1}{1-c(\underline{m}^0(z))^2(1+\underline{m}^0(z))^{-2}}. \tag{4.110}$$

In view of (4.107) we have

$$\frac{1}{\varpi(z)} + c\varpi(z)(m^0(z))' - c(m^0(z))'\frac{1}{m^0(z)} = -\frac{1}{z\underline{m}^0(z)} + c\big(m^0(z)\big)'z. \tag{4.111}$$

Taking derivative with respect to z on the both sides of (4.95) we have

$$c\big(m^0(z)\big)'z = \big(\underline{m}^0(z)\big)'z - \frac{1-c}{z} = \frac{c\big(\underline{m}^0(z)\big)^2\big(1+\underline{m}^0(z)\big)^{-1} - \underline{m}^0(z)}{1-c\big(\underline{m}^0(z)\big)^2\big(1+\underline{m}^0(z)\big)^{-2}} - \frac{1-c}{z}, \tag{4.112}$$

the last step using the expression (4.101) for z.

In view of (4.110), (4.111) and (4.112) it is enough to show

$$-\frac{1}{z}\Big(\frac{1}{\underline{m}^0(z)} + 1 - c\Big) = \frac{1+\underline{m}^0(z) - c\big(\underline{m}^0(z)\big)^2\big(1+\underline{m}^0(z)\big)^{-1}}{1-c\big(\underline{m}^0(z)\big)^2\big(1+\underline{m}^0(z)\big)^{-2}}. \tag{4.113}$$

From (4.102) the left hand side of (4.113) becomes $1+\underline{m}^0(z)$. Because it is easy to check that

$$\Big(1+\underline{m}^0(z)\Big)\Big(1-c\big(\underline{m}^0(z)\big)^2\big(1+\underline{m}^0(z)\big)^{-2}\Big) = 1+\underline{m}^0(z) - c\big(\underline{m}^0(z)\big)^2\big(1+\underline{m}^0(z)\big)^{-1},$$

we get (4.113). The proof is completed.

Chapter 5

Independence Test For A Large Panel Data

5.1 Theory and Methodology

Before we establish the main theory and methodology, we first introduce the following assumptions:

Assumption 1. *For each* $i = 1, \ldots, p$, $X_{1i}, \cdots, X_{ni}$ *are independent and identically distributed (i.i.d) random variables with mean zero and variance one. When* X_{ji}*'s are complex random variables, we require* $EX_{ji}^2 = 0$.

Assumption 2. $p = p(n)$ *with* $\frac{p}{n} \to c \in (0, \infty)$ *as* $n \to \infty$.

We stack p time series one by one to form a data matrix $\mathbf{X} = (\mathbf{x}_1, \cdots, \mathbf{x}_p)$. Moreover, denote the sample covariance matrix by

$$\mathbf{A}_n = \frac{1}{n}\mathbf{X}^*\mathbf{X},$$

where $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_p) = (\mathbf{y}_1, \ldots, \mathbf{y}_n)^T$, $\mathbf{y}_j^T$ denotes the j-th row of the matrix $\mathbf{X}$ and $\mathbf{X}^*$ is the Hermitian transform of the matrix $\mathbf{X}$. The empirical spectral distribution (ESD) of the sample covariance matrix $\mathbf{A}_n$ is defined as

$$F^{\mathbf{A}_n}(x) = \frac{1}{p}\sum_{i=1}^{p} I(\lambda_i \le x), \tag{5.1}$$

where $\lambda_1 \le \lambda_2 \le \ldots \le \lambda_p$ are eigenvalues of $\mathbf{A}_n$.

It is well–known that under Assumptions 1 and 2, if $\mathbf{x}_1, \cdots, \mathbf{x}_p$ are independent then $F^{\mathbf{A}_n}(x)$ converges with probability one to the Marcenko-Pastur Law $F^c(x)$ (see Marcenko-Pastur (1967)) whose density has an explicit expression of the form

$$f_c(x) = \begin{cases} \frac{1}{2\pi x c}\sqrt{(b-x)(x-a)}, & a \le x \le b; \\ 0, & otherwise; \end{cases} \tag{5.2}$$

and a point mass $1 - 1/c$ at the origin if $c > 1$, where $a = (1-\sqrt{c})^2$ and $b = (1+\sqrt{c})^2$.

When there is some correlation structure among $\mathbf{x}_1, \cdots, \mathbf{x}_p$, denote by $\mathbf{T}_p$ the covariance matrix of each row, $\mathbf{y}_j^T$, of $\mathbf{X}$ (more specifically, we need to assume $\mathbf{y}_j = \mathbf{T}_p^{1/2}\mathbf{w}_j$ whose definitions are given in Assumption 3 below). Then, under Assumptions 1 and 2, if $F^{\mathbf{T}_p}(x) \xrightarrow{D} H(x)$, then $F^{\mathbf{A}_n}$ converges with probability one to a non random distribution function $F^{c,H}$ whose Stieltjes transform satisfies (see Silverstein (1995))

$$m(z) = \int \frac{1}{x(1-c-czm(z)) - z} dH(x), \tag{5.3}$$

where the Stieltjes transform m_G for any c.d.f G is defined as

$$m_G(z) = \int \frac{1}{\lambda - z} dG(\lambda), \quad \mathrm{Im} z > 0 \tag{5.4}$$

and G can be recovered by the inversion formula

$$G\{[x_1, x_2]\} = \frac{1}{\pi}\lim_{\varepsilon \to 0+}\int_{x_1}^{x_2} \mathcal{I}m(m_G(x + i\varepsilon))dx,$$

where x_1 and x_2 are continuity points of G.

Moreover, equation (5.3) takes a simpler form when $F^{c,H}$ is replaced by

$$\underline{F}^{c,H} = (1-c)I_{[0,\infty]} + cF^{c,H}, \tag{5.5}$$

which is the limiting ESD of $\underline{\mathbf{A}}_n = \frac{1}{n}\mathbf{X}\mathbf{X}^*$. Its Stieltjes transform

$$\underline{m}(z) = -\frac{1-c}{z} + cm(z) \tag{5.6}$$

has an inverse

$$z = z(\underline{m}) = \frac{1}{\underline{m}} + c\int \frac{x}{1+x\underline{m}} dH(x). \tag{5.7}$$

The construction of our test statistic relies on the following observation: the limit of the ESD of the sample covariance matrix $\mathbf{A}_n$ is the M-P law by (5.2) when $\mathbf{x}_1, \cdots, \mathbf{x}_p$ are independent and satisfy Assumptions 1 and 2, while the limit of the ESD is determined from (5.3) when there is some correlation among $\mathbf{x}_1, \cdots, \mathbf{x}_p$ with the covariance matrix $\mathbf{T}_p$ different from the identity matrix. Moreover, preliminary investigations indicate that when $\mathbf{x}_1, \cdots, \mathbf{x}_p$ are only uncorrelated (without any further assumptions), the limit of the ESD of $\mathbf{A}_n$ is not the M-P law (see Ryan and Debbah (2009)). From this point, any deviation of the limit of the ESD from the M-P law is evidence of dependence. So, these motivate us to use the ESD of $\mathbf{A}_n$, $F^{\mathbf{A}_n}(x)$, as a test statistic. However, there is no central limit theorem for $(F^{\mathbf{A}_n}(x) - F^c(x))$, as argued by Bai and Silverstein (2004). Therefore, instead, we consider the characteristic function of $F^{\mathbf{A}_n}(x)$.

The characteristic function of $F^{\mathbf{A}_n}(x)$ is

$$s_n(t) \triangleq \int e^{itx} dF^{\mathbf{A}_n}(x) = \frac{1}{p}\sum_{i=1}^{p} e^{it\lambda_i}, \tag{5.8}$$

where $\lambda_i, i = 1, \ldots, p$ are eigenvalues of the sample covariance matrix $\mathbf{A}_n$.

Our test statistic is then proposed as follows:

$$M_n = \int_{T_1}^{T_2} |s_n(t) - s(t)|^2 dU(t), \tag{5.9}$$

where $s(t) := s(t, c_n)$ is the characteristic function of $F^{c_n}(x)$, obtained from the M-P law $F^c(x)$ with c being replaced by $c_n = p/n$, and $U(t)$ is a weight function with its support on a compact interval, say $[T_1, T_2]$.

To develop the asymptotic distribution of M_n under a local alternative, the following assumption is needed.

Assumption 3. *Let $\boldsymbol{T}_p$ be a $p \times p$ random Hermitian nonnegative definite matrix with a bounded spectral norm. Let $\mathbf{y}_j^T = \mathbf{w}_j^T \mathbf{T}_p^{1/2}$, where $\mathbf{T}_p^{1/2}$ is the $p \times p$ Hermitian matrix that satisfies $(\mathbf{T}_p^{1/2})^2 = \mathbf{T}_p$ and $\mathbf{w}_j^T = (W_{j1}, \cdots, W_{jp}), j = 1, \ldots, n$ are i.i.d random vectors, in which $W_{ji}, j \leq n, i \leq p$ are i.i.d with mean zero, variance one and finite fourth moment.*

The empirical spectral distribution $F^{\boldsymbol{T}_p}$ of $\boldsymbol{T}_p$ converges weakly to a distribution H on $[0, \infty)$ as $n \to \infty$; all the diagonal elements of the matrix $\mathbf{T}_p$ are equal to 1.

Note that under Assumption 3, $\mathbf{A}_n$ becomes $\mathbf{T}_p^{1/2}\mathbf{W}^*\mathbf{W}\mathbf{T}_p^{1/2}$, where $\mathbf{W} = (\mathbf{w}_1, \cdots, \mathbf{w}_n)^T$. The assumption that all the diagonal elements of $\mathbf{T}_p$ are equal to 1 is used to guarantee that $EX_{ji}^2 = 1$. Under Assumption 3, when $\mathbf{T}_p = \mathbf{I}_p$, the random vectors $\mathbf{x}_1, \ldots, \mathbf{x}_p$ are independent and when $\mathbf{T}_p \neq \mathbf{I}_p$, they are not independent. For convenience, we name this dependent structure as 'linear dependence'.

To develop the asymptotic distribution of the test statistic, write

$$G_n(x) = p[F^{\mathbf{A}_n}(x) - F^{c_n}(x)]. \tag{5.10}$$

Then, $p(s_n(t) - s(t))$ can be decomposed as sum of the stochastic part and the non–stochastic part as follows:

$$
\begin{aligned}
p(s_n(t) - s(t)) &= \int e^{itx} dG_n(x) \\
&= \int e^{itx} d\left(p\left[F^{\mathbf{A}_n}(x) - F^{c_n,H_n}(x)\right]\right) + \int e^{itx} d\left(p\left[F^{c_n,H_n}(x) - F^{c_n}(x)\right]\right)
\end{aligned}
\tag{5.11}
$$

where F^{c_n,H_n} is obtained from $F^{c,H}$ with c and H replaced by $c_n = p/n$ and $H_n = F^{\mathbf{T}_p}$.

To simplify the statements of the following theorems, we introduce some notation here:

$$
\begin{aligned}
\delta_1(t) &= \lim_{n\to\infty} \int cos(tx) dp\Big(F^{c_n,H_n}(x) - F^{c_n}(x)\Big), \\
\delta_2(t) &= \lim_{n\to\infty} \int sin(tx) dp\Big(F^{c_n,H_n}(x) - F^{c_n}(x)\Big),
\end{aligned}
\tag{5.12}
$$

$$
\begin{aligned}
\rho_1(t) &= \oint_\gamma cos(tz) \frac{c\underline{m}^3(z)h_2(z)}{1 - c\int \underline{m}^2(z)\tau^2(1+\tau\underline{m}(z))^{-2}dH(\tau)} dz, \\
\rho_2(t_j,t_h) &= \oint_{\gamma_1}\oint_{\gamma_2} \frac{cos(t_jz_1)sin(t_hz_2)}{(\underline{m}(z_1)-\underline{m}(z_2))^2} \frac{d}{dz_2}\underline{m}(z_2) \frac{d}{dz_1}\underline{m}(z_1) dz_1dz_2, \\
\rho_3(t_j,t_h) &= \oint_{\gamma_1}\oint_{\gamma_2} cos(t_jz_1)sin(t_hz_2) \frac{d^2}{dz_1dz_2}[\underline{m}(z_1)\underline{m}(z_2)h_1(z_1,z_2)]dz_1dz_2,
\end{aligned}
$$

$$
\begin{aligned}
E(V^{re}(t)) &= \frac{1}{2\pi i}\oint_\gamma cos(tz) \frac{c\int \underline{m}^3(z)\tau^2(1+\tau\underline{m}(z))^{-3}dH(\tau)}{(1 - c\int \underline{m}^2(z)\tau^2(1+\tau\underline{m}(z))^{-2}dH(\tau))^2} dz \\
&\quad - \frac{EX_{11}^4 - 3}{2\pi i}\rho_1(t),
\end{aligned}
\tag{5.13}
$$

$$
Cov\Big(V^{re}(t_j), Z^{re}(t_h)\Big) = -\frac{1}{2\pi^2}\rho_2(t_j,t_h) - \frac{c(EX_{11}^4 - 3)}{4\pi^2}\rho_3(t_j,t_h), \tag{5.14}
$$

$$
E(V^{im}(t)) = -\frac{E|X_{11}|^4 - 2}{2\pi i}\rho_1(t), \tag{5.15}
$$

and

$$
Cov\Big(V^{im}(t_j), Z^{im}(t_h)\Big) = -\frac{1}{4\pi^2}\rho_2(t_j,t_h) - \frac{c(E|X_{11}|^4 - 2)}{4\pi^2}\rho_3(t_j,t_h). \tag{5.16}
$$

We now state the first theorem of this chapter.

Theorem 7. *In addition to Assumptions 1 and 2, let the fourth moment of each X_{ji} be finite.*

1) Suppose that Assumption 3 and the following conditions hold:

$$\frac{1}{n}\sum_{i=1}^{n}\mathbf{e}_i^*\boldsymbol{T}_p^{1/2}(\underline{m}(z_1)\boldsymbol{T}_p+\boldsymbol{I})^{-1}\boldsymbol{T}_p^{1/2}\mathbf{e}_i\mathbf{e}_i^*\boldsymbol{T}_p^{1/2}(\underline{m}(z_2)\boldsymbol{T}_p+\boldsymbol{I})^{-1}\boldsymbol{T}_p^{1/2}\mathbf{e}_i \to h_1(z_1,z_2) \tag{5.17}$$

and

$$\frac{1}{n}\sum_{i=1}^{n}\mathbf{e}_i^*\boldsymbol{T}_p^{1/2}(\underline{m}(z)\boldsymbol{T}_p+\boldsymbol{I})^{-1}\boldsymbol{T}_p^{1/2}\mathbf{e}_i\mathbf{e}_i^*\boldsymbol{T}_p^{1/2}(\underline{m}(z)\boldsymbol{T}_p+\boldsymbol{I})^{-2}\boldsymbol{T}_p^{1/2}\mathbf{e}_i \to h_2(z), \tag{5.18}$$

where $\mathbf{e}_i^*$ is the n-dimensional row vector with the i-th element being 1 and others 0. Then the scaled proposed test statistic p^2M_n converges in distribution to a random variable R of the form

$$R=\int_{T_1}^{T_2}(|V(t)+\delta_1(t)|^2+|Z(t)+\delta_2(t)|^2)dU(t), \tag{5.19}$$

where $(V(t),Z(t))$ is a Gaussian vector and $\delta_j(t),\ j=1,2$ are defined in (5.12).

Denote $(V(t),Z(t))$ by $(V^{re}(t),Z^{re}(t))$ when X_{11} is real and the mean of $V^{re}(t)$ is specified in (5.13). Replacing $cos(tz)$ in $E(V^{re}(t))$ by $sin(t_jz)$ yields the expression of $E(Z^{re}(t))$; the covariance of $(V^{re}(t),Z^{re}(t))$ is given in (5.14).

Denote $(V(t),Z(t))$ as $(V^{im}(t),Z^{im}(t))$ when X_{11} is complex with $EX_{11}^2=0$ and then the mean of $V^{im}(t)$ is specified in (5.15). Similarly replacing $cos(tz)$ in $E(V^{im}(t))$ by $sin(t_jz)$ yields the expression of $E(Z^{im}(t))$ in this case; the covariance of $(V^{im}(t),Z^{im}(t))$ is given in (5.16).

In both cases, the definitions of $Cov(V(t_j),V(t_h))$ and $Cov(Z(t_j),Z(t_h))$ are similar to that of $Cov(V(t_j),Z(t_h))$ except replacing $cos(t_jz)sin(t_hz)$ by $cos(t_jz)cos(t_hz)$ and $sin(t_jz)sin(t_hz)$ respectively. The contours γ, γ_1 and

γ_2 *above are all closed and are taken in the positive direction in the complex plane, each enclosing the support of* $F^{c,H}$. *Also* γ_1 *and* γ_2 *are disjoint.*

2) Under the null hypothesis $\mathbf{H}_0$ *the scaled statistic* $p^2 M_n$ *then converges in distribution to*

$$R_0 = \int_{T_1}^{T_2} (|\tilde{V}(t)|^2 + |\tilde{Z}(t)|^2) dU(t), \qquad (5.20)$$

where the distribution of $(\tilde{V}(t), \tilde{Z}(t))$ *can be obtained from that of* $(V(t), Z(t))$ *with* $H(\tau)$ *being the degenerate distribution at the point* 1, $\underline{m}(z)$ *being the Stieltjes transform of the M-P law,* $h_1(z_1, z_2) = \frac{1}{(\underline{m}(z_1)+1)(\underline{m}(z_2)+1)}$ *and* $h_2(z) = \frac{1}{(\underline{m}(z)+1)^3}$.

Remark 10. *Assumption 1 assumes that all the entries of* $\mathbf{x}_i$ *are identically distributed. It is of practical interest to consider removing the identical distribution condition. Instead of assuming identically distributed entries for* $\mathbf{x}_i$, *we need only to impose the following additional assumptions: for any* $k = 1, \ldots, p; j = 1, \ldots, n$, $E[X_{jk}] = 0$, $E\left[X_{jk}^2\right] = 1$, $\sup_{j,k} E\left[X_{jk}^4\right] < \infty$ *and for any* $\eta > 0$,

$$\frac{1}{\eta^4 np} \sum_{j=1}^{n} \sum_{k=1}^{p} E\left[|X_{jk}|^4 I_{(|X_{jk}| \geq \eta\sqrt{n})}\right] \to 0, \text{ as } n \to \infty. \qquad (5.21)$$

A careful checking on the arguments of Theorem 1.1 *of Bai and Silverstein (2004) and Theorem* 1.4 *of Pan and Zhou (2008) indicates that Lemma 20 listed in the Appendix still holds and hence Theorem 7 holds under (5.21). For the expressions of the mean and covariance of the asymptotic random vector, we substitute the fourth moment* $E|W_{11}|^4$ *in Lemma 20 with the average of all the fourth moments of all the entries, i.e.* $\frac{\sum_{j=1,k=1}^{n,p} E|W_{jk}|^4}{np}$.

Here we would like to point out that there is no need to impose conditions (5.17) and (5.18) when establishing the asymptotic distribution of the test statistic under the null hypothesis (the second part of the above theorem). Moreover, conditions (5.17) and (5.18) can be removed if $E\left[W_{11}^4\right] = 3$ in the real–number case or if $E\left[W_{11}^2\right] = 2$ in the complex–number case (see Bai and Silverstein (2004)). The first part of the above theorem is concerned with asymptotic distributions of the test statistic under a local alternative hypothesis, i.e., Assumption 3. With respect to Assumption 3, we would like to make the following comments, which are useful in the subsequent application section.

If $\mathbf{y}_j^T = \tilde{\mathbf{w}}_j^T \mathbf{C}$, where $\mathbf{C}$ is any $q \times p$ nonrandom matrix and $\tilde{\mathbf{w}}_j, j = 1, \ldots, n$ are i.i.d. $q \times 1$ random vectors with their respective entries being i.i.d random variables, then Theorem 7 is still applicable. This is because $\frac{1}{n}\mathbf{X}\mathbf{X}^*$ in this case becomes $\frac{1}{n}\tilde{\mathbf{W}}\mathbf{C}\mathbf{C}^*\tilde{\mathbf{W}}^*$ and the nonnegative definitive matrix $\mathbf{C}\mathbf{C}^*$ can be decomposed into $\mathbf{C}\mathbf{C}^* = \mathbf{T}_q^{1/2}\mathbf{T}_q^{1/2}$, where $\mathbf{T}_q^{1/2}$ is a $q \times q$ Hermitian matrix and $\tilde{\mathbf{W}} = (\tilde{\mathbf{w}}_1, \ldots, \tilde{\mathbf{w}}_n)$. Note that the eigenvalues of $\frac{1}{n}\mathbf{X}\mathbf{X}^*$ differ from those of $\frac{1}{n}\mathbf{T}_q^{1/2}\mathbf{W}^*\mathbf{W}\mathbf{T}_q^{1/2}$ by $|p - q|$ zeros. Thus, we may instead resort to CLT of $\frac{1}{n}\mathbf{T}_q^{1/2}\tilde{\mathbf{W}}^*\tilde{\mathbf{W}}\mathbf{T}_q^{1/2}$.

We can evaluate the power of the statistic M_n for a class of local alternatives, although it is difficult to establish an asymptotic distribution for the test statistic under such a class of local alternatives.

Theorem 8. *Let the following hold in probability,*

$$\limsup_{n\to\infty} p \int e^{itx} d(F_{\mathbf{H}_0}^{\mathbf{A}_n} - F_{\mathbf{H}_1}^{\mathbf{A}_n}) = \infty, \tag{5.22}$$

where $F_{\mathbf{H}_0}^{\mathbf{A}_n}$ *stands for the ESD of* $\mathbf{A}_n$ *under* $\mathbf{H}_0$ *(satisfying Assumptions 1 and 2) and* $F_{\mathbf{H}_1}^{\mathbf{A}_n}$ *is the ESD of* $\mathbf{A}_n$ *under* $\mathbf{H}_1$*. Then*

$$\lim_{n\to\infty} P(p^2 M_n > \gamma_\alpha | \mathbf{H}_1) = 1,$$

where γ_α *is the critical value of* $n^2 M_n$ *under* $\mathbf{H}_0$ *(determined by* R *in Theorem 7) corresponding to the significance level* α*.*

Remark 11. *Note that if* $F^{\mathbf{A}_n}_{\mathbf{H}_0}$ *and* $F^{\mathbf{A}_n}_{\mathbf{H}_1}$ *have different limits in probability, then* $\int e^{itx} d(F^{\mathbf{A}_n}_{\mathbf{H}_0} - F^{\mathbf{A}_n}_{\mathbf{H}_1})$ *converges in probability to a nonzero constant depending* t *by Levy's continuity theorem. This ensures (5.22) is true. Most of the examples given in the subsequent sections satisfy (5.22).*

5.2 Applications to Multiple MA(1), AR(1) and Spatial Cross–sectional dependent Structures

This section is to explore some applications of the proposed test. In the last section, we have discussed the case where dependent vectors can be expressed as linear combinations of independent random vectors, i.e. $\mathbf{y}_j = \mathbf{w}_j\mathbf{C}$. Although this chapter mainly focuses on the analysis of cross–sectional dependence, many other dependent structures incurred by time series also satisfy the dependent structure developed in the last section. As an illustration of this point, we provide MA(1) and AR(1) models here. For cross-sectional dependence of interest in panel data analysis, we give examples of some spatial models that can be discussed in a way similar to what has been done in Section 2.

Example 3.1. Consider a multiple moving average model of order 1(MA(1)) of the form:

$$\mathbf{v}_t = \mathbf{z}_t + \psi \mathbf{z}_{t-1}, t = 1, \ldots, p, \tag{5.23}$$

where $|\psi| < \infty$; $\mathbf{z}_t = (Z_{1t}, \ldots, Z_{nt})^T$ is an n-dimensional random vector with i.i.d. elements, each of which has zero mean and unit variance; and $\mathbf{v}_t = (V_{1t}, \ldots, V_{nt})^T$. Denote by $\hat{\mathbf{V}}_j^T$ and $\hat{\mathbf{Z}}_j^T$ respectively the j-th rows of $\mathbf{V} = (V_{jt})_{n\times p}$ and $\mathbf{Z} = (Z_{jt})_{n\times(p+1)}$.

For each $j = 1, \ldots, n$, the MA(1) model (5.23) can be written as

$$\hat{\mathbf{V}}_j^T = \hat{\mathbf{Z}}_j^T \mathbf{C}, \tag{5.24}$$

where

$$\mathbf{C} = \begin{pmatrix} \psi & 0 & 0 & \cdots & 0 & 0 \\ 1 & \psi & 0 & \cdots & 0 & 0 \\ 0 & 1 & \psi & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & \psi \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{pmatrix}_{(p+1)\times p} . \tag{5.25}$$

From Assumption 3, the preceding subsection and Theorem 7, our test is able to capture the dependence of $\mathbf{v}_1, \cdots, \mathbf{v}_p$ as n and p go to infinity in the same order.

Example 3.2. Consider a vector autoregressive model of order 1 (VAR(1)) of the form:

$$\mathbf{v}_t = \phi \mathbf{v}_{t-1} + \mathbf{z}_t, \ t = 1, \ldots, p, \tag{5.26}$$

where $\mathbf{v}_0 = \frac{1}{\sqrt{1-\phi^2}} \boldsymbol{\epsilon}$ with $\boldsymbol{\epsilon}$ being an n-dimensional random vector with i.i.d elements whose means are zero and variances are unit, $|\phi| < 1$; for any $t = 1, \ldots, p$, $\mathbf{z}_t = (Z_{1t}, \ldots, Z_{nt})^T$ is an n-dimensional random vector with i.i.d. elements, each of which has zero mean and unit variance; and $\mathbf{v}_t = (V_{1t}, \ldots, V_{nt})^T$. The choice of $\frac{1}{\sqrt{1-\phi^2}}$ in the definition of $\mathbf{v}_0$ is to ensure that the first two moments of $\{\mathbf{v}_t\}_{t=0}^p$ are well defined. In order to writing this VAR(1) model into the linear dependent structure proposed in the last section, we add $\mathbf{z}_0 = \boldsymbol{\epsilon}$ to the sequence $\{\mathbf{z}_t\}_{t=1}^p$ and thus $\{\mathbf{z}_t\}_{t=0}^p$ has the same length with $\{\mathbf{v}_t\}_{t=0}^p$. Denote the j-th rows of $\mathbf{V} = (V_{jt})_{n\times(p+1)}$ and $\mathbf{Z} = (Z_{jt})_{n\times(p+1)}$ by $\hat{\mathbf{V}}_j^T$ and $\hat{\mathbf{Z}}_j^T$ respectively.

For each $j = 1, \ldots, n$, the AR(1) model (5.26) can be written as

$$\hat{\mathbf{V}}_j^T = \hat{\mathbf{Z}}_j^T \mathbf{D}, \tag{5.27}$$

where

$$\mathbf{D} = \begin{pmatrix} 1 & -\phi & (-\phi)^2 & \cdots & (-\phi)^{p-2} & (-\phi)^{p-1}/\sqrt{1-\phi^2} \\ 0 & 1 & -\phi & \cdots & (-\phi)^{p-3} & (-\phi)^{p-2}/\sqrt{1-\phi^2} \\ 0 & 0 & 1 & \cdots & (-\phi)^{p-4} & (-\phi)^{p-3}/\sqrt{1-\phi^2} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -\phi/\sqrt{1-\phi^2} \\ 0 & 0 & 0 & \cdots & 0 & 1/\sqrt{1-\phi^2} \end{pmatrix}. \tag{5.28}$$

By Theorem 7, we can apply the proposed test M_n to this AR(1) model as well.

Example 3.3. We now consider a panel data case. Let $\{v_{ji} : i = 1, \ldots, p; j = 1, \ldots, n\}$ be the error components involved in a panel data model. They may be cross–sectionally correlated. In panel data analysis, it is of interest to consider the cross–sectional independence hypothesis, i.e.

$$\mathbf{H}_{00} : \ Cov(v_{ji}, v_{jh}) = 0 \ for \ all \ j = 1, \ldots, n \ and \ all \ i \neq h;$$

against

$$\mathbf{H}_{11} : \ Cov(v_{ji}, v_{jh}) \neq 0 \ for \ some \ j \ and \ some \ i \neq h. \tag{5.29}$$

Under the assumption that $\{v_{ji} : i = 1, \ldots, p; j = 1, \ldots, n\}$ are normally distributed, this hypothesis is equivalent to the independence hypothesis that

$$\mathbf{H}_0 : \ \mathbf{x}_1, \ldots, \mathbf{x}_p \ are \ independent; \ \ against \ \ \mathbf{H}_1 : \ \mathbf{x}_1, \ldots, \mathbf{x}_p \ are \ not \ independent, \tag{5.30}$$

where $\mathbf{x}_i = (v_{1i}, \ldots, v_{ni})^T$, $i = 1, \ldots, p$.

Modern panel data literature has mainly adopted two different approaches to model error cross–sectional dependence: the spatial approach and the factor-structure approach. For the spatial approach, there are three popular spatial models: the Spatial Moving Average (SMA), Spatial AutoRegressive (SAR) and Spatial Error Components (SEC) processes. They are defined as follows:

$$SMA: \ v_{ji} = \sum_{k=1}^{p} \omega_{ik}\varepsilon_{jk} + \varepsilon_{ji}, \tag{5.31}$$

$$SAR: \ v_{ji} = \sum_{k=1}^{p} \omega_{ik}v_{jk} + \varepsilon_{ji}, \tag{5.32}$$

$$SEC: \ v_{ji} = \sum_{k=1}^{p} \omega_{ik}\xi_{jk} + \varepsilon_{ji}, \tag{5.33}$$

where ω_{ik} is the i-specific spatial weight attached to individual k; $\{\varepsilon_{ji} : i = 1, \ldots, p; j = 1, \ldots, n\}$ and $\{\xi_{ji} : i = 1, \ldots, p; j = 1, \ldots, n\}$ are two sets with i.i.d. random components with zero mean and unit variance; moreover, $\{\xi_{ji} : i = 1, \ldots, p; j = 1, \ldots, n\}$ are uncorrelated with $\{\varepsilon_{ji}, i = 1, \ldots, p; j = 1, \ldots, n\}$.

Denote the j-th row of $\mathbf{V} = (v_{ji})_{n\times p}$, $\boldsymbol{\varepsilon} = (\varepsilon_{ji})_{n\times p}$ and $\boldsymbol{\xi} = (\xi_{ji})_{n\times p}$ by $\hat{\mathbf{v}}_j^T$, $\hat{\boldsymbol{\varepsilon}}_j^T$ and $\hat{\boldsymbol{\xi}}_j^T$ respectively. Set $\boldsymbol{\omega} = (\omega_{ik})_{p\times p}$.

Model SMA (5.31) may be rewritten as $\hat{\mathbf{v}}_j^T = \hat{\boldsymbol{\varepsilon}}_j^T(\boldsymbol{\omega}^T + \mathbf{I}_p), \forall j = 1, \ldots, n$ and hence $\mathbf{T}_p = (\boldsymbol{\omega} + \mathbf{I}_p)(\boldsymbol{\omega}^T + \mathbf{I}_p)$. For model SAR (5.32), assume that $\boldsymbol{\omega} - \mathbf{I}_p$ is invertible. We then write $\hat{\mathbf{v}}_j^T = \hat{\boldsymbol{\varepsilon}}_j^T(\boldsymbol{\omega}^T - \mathbf{I}_p)^{-1}, \forall j = 1, \ldots, n$. Hence $\mathbf{T}_p = (\boldsymbol{\omega} - \mathbf{I}_p)^{-1}(\boldsymbol{\omega}^T - \mathbf{I}_p)^{-1}$. Therefore the test statistic M_n can be used to identify whether $\mathbf{x}_1, \cdots, \mathbf{x}_p$ of models (5.31) and (5.32) are independent. Hence it can capture the cross–sectional dependence for the SMA model and SAR model

As for the SEC model, whether the statistic M_n can detect the dependence of the SEC model relies on the properties of the sample covariance matrix in the form of

$$\mathbf{B}_n = \frac{1}{n}(\boldsymbol{\omega}\boldsymbol{\xi} + \boldsymbol{\varepsilon})(\boldsymbol{\omega}\boldsymbol{\xi} + \boldsymbol{\varepsilon})^T, \tag{5.34}$$

where $\boldsymbol{\xi} = (\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_p)^T$ and $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_p)^T$.

Under the null hypothesis $\mathbf{H}_0$, Dozier and Silverstein (2007) provides the limit of the ESD of the matrix $\mathbf{B}_n$ whose Stieljes transform is

$$\hat{m}(z) = \int \frac{d\hat{H}(x)}{\frac{x}{1+c\hat{m}(z)} - (1 + c\hat{m}(z))z + 1 - c}, \tag{5.35}$$

where $\hat{H}(x)$ is the limit of $F^{\frac{1}{n}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T}$.

With this result, we know that the limit of the ESD of the matrix $\mathbf{B}_n$ is not the M-P law so that condition (5.22) is satisfied. By Theorem 8, the proposed test M_n can be applied to capture the dependence of the SEC model (5.33).

5.3 A general panel data model

It is well known that there are two common used cross-sectional dependent structures in panel data: spatial structures and factor models. As stated in the last section, our developed dependent structure $\mathbf{y}_j = \mathbf{w}_j\mathbf{C}$ covers some spatial structures in panel data. In this section, we consider a simple factor model which is utilised to describe cross–sectional dependence. A new asymptotic theory is established as a consequence of our discussion.

Note that the proposed test is based on the idea that the limits of ESDs under the null and local alternative hypotheses are different. Yet, it may be the case that there exists some dependence among the set of vectors of

$\mathbf{x}_1, \cdots, \mathbf{x}_p$ but the limit of the ESD associated with such vectors is the M-P law. Then a natural question is whether the statistic M_n works in this case. We below investigate the panel data model as an example.

Consider a panel data model of the form

$$v_{ij} = \varepsilon_{ij} + \frac{1}{\sqrt{p}} u_i, \ i = 1, \ldots, p; \ j = 1, \ldots, n, \tag{5.36}$$

where $\{\varepsilon_{ij}, i = 1, \ldots, p; j = 1, \ldots, n\}$ is a sequence of i.i.d. real random variables with $E\varepsilon_{11} = 0$ and $E\varepsilon_{11}^2 = 1$, and $\{u_i, i = 1, \ldots, p\}$ are real random variables, and independent of $\{\varepsilon_{ij}, i = 1, \ldots, p; j = 1, \ldots, n\}$.

For any $i = 1, \ldots, p$, set

$$\mathbf{x}_i = (v_{i1}, \ldots, v_{in})^T. \tag{5.37}$$

The aim of this section is to test the null hypothesis specified in (5.30) for model (5.36).

Model (5.36) can be written as

$$\mathbf{X} = \boldsymbol{\varepsilon} + \mathbf{u}\mathbf{e}^T, \tag{5.38}$$

where $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_p)^T$, $\mathbf{u} = (\frac{1}{\sqrt{p}} u_1, \ldots, \frac{1}{\sqrt{p}} u_p)^T$ and $\mathbf{e}$ is $p \times 1$ vector with all elements being one.

Consider the sample covariance matrix

$$\mathbf{S}_n = \frac{1}{n}\mathbf{X}\mathbf{X}^T = \frac{1}{n}(\boldsymbol{\varepsilon} + \mathbf{u}\mathbf{e}^T)(\boldsymbol{\varepsilon} + \mathbf{u}\mathbf{e}^T)^T. \tag{5.39}$$

By the rank inequality (see Lemma 3.5 of Yin (1986)) and the fact that $rank(\mathbf{u}\mathbf{e}^T) \le 1$, it can be concluded that the limit of the ESD of the matrix $\mathbf{S}$ is the same as that of the matrix $\frac{1}{n}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T$, i.e. the M-P law. Even so, we still would like to use the proposed statistic M_n to test the null hypothesis of mutual independence. However, this model does not satisfy

Assumption 1 because the elements of each vector $\mathbf{x}_i$ are not independent and they include the common random factor u_i, and Theorem 7 thus can not be directly applicable to this model. Therefore, we need to develop a new asymptotic theory for the proposed statistic M_n for this model.

Theorem 9. *Consider model (5.36) and let Assumption 2 hold. Additionally, suppose that $\{\varepsilon_{ij}\}$ are i.i.d with mean zero, variance one and finite fourth moment and that*

$$E\|\mathbf{u}\|^4 < \infty \quad \textit{and} \quad \frac{1}{p^2} E\left[\sum_{i\neq j}^{p} (u_i^2 - \bar{u})(u_j^2 - \bar{u})\right] \to 0 \quad \textit{as } n \to \infty, \tag{5.40}$$

where $\bar{u}$ is a positive constant number.

Then, the proposed test statistic $p^2 M_n$ converges in distribution to the random variable R_2 given by

$$R_2 = \int_{t_1}^{t_2} \left(|W(t)|^2 + |Q(t)|^2\right) dU(t), \tag{5.41}$$

where $(W(t), Q(t))$ is a Gaussian vector whose mean and covariance are specified as follows:

$$\begin{aligned} EW(t) = &-\frac{c}{2\pi i}\oint_\gamma cos(t_j z)\frac{\underline{m}^3(z)(1+\underline{m}(z))}{((1+\underline{m}(z))^2 - c\underline{m}^2(z))^2}dz \\ &- \frac{c(EY_{11}^4 - 3)}{2\pi i}\oint_\gamma cos(t_j z)\frac{\underline{m}^3(z)}{(1+\underline{m}(z))^2 - c\underline{m}^2(z)}dz \\ &+ \frac{c}{2\pi i}\oint_\gamma cos(t_j z)\frac{\underline{m}(z)}{z((1+\underline{m}(z))^2 - c\underline{m}^2(z))}dz \\ &- \frac{1}{2\pi i}\oint_\gamma cos(t_j z)\frac{\frac{c\underline{m}(z)}{z[(1+\underline{m}(z))^2 - c\underline{m}^2(z)]} + \bar{u}\int\frac{1}{(\lambda - z)^2}dF^{MP}(\lambda)}{\bar{u}z\underline{m}^2(z) - 1}dz; \end{aligned} \tag{5.42}$$

and

$$\begin{aligned} Cov(W(t_j), Q(t_h)) = &-\frac{1}{2\pi^2}\oint_{\gamma_1}\oint_{\gamma_2}\frac{cos(t_j z_1)sin(t_h z_2)}{(\underline{m}(z_1) - \underline{m}(z_2))^2}\frac{d}{dz_1}\underline{m}(z_1)\frac{d}{dz_2}\underline{m}(z_2)dz_1dz_2 \\ &-\frac{c(EY_{11}^4 - 3)}{4\pi^2}\oint_{\gamma_1}\oint_{\gamma_2} cos(t_j z_1)cos(t_h z_2)\frac{d^2}{dz_1dz_2}[\frac{\underline{m}(z_1)\underline{m}(z_2)}{(1+\underline{m}(z_1))(1+\underline{m}(z_2))}]dz_1dz_2. \end{aligned}$$

(5.43)

Replacing $cos(t_j z)$ *in* $E[W(t_j)]$ *by* $sin(t_j z)$ *yields the expression of* $E[Q(t_j)]$. *The covariances* $\mathrm{Cov}(W(t_j), W(t_h))$ *and* $Cov(Q(t_j), Q(t_h))$ *are similar except replacing* $sin(t_h z)$ *and* $cos(t_j z)$ *by* $cos(t_h z)$ *and* $sin(t_j z)$ *respectively. The contours in (5.42) and (5.43) both enclose the interval* $[(1-\sqrt{c})^2 + 2c\bar{u}, (1+\sqrt{c})^2 + 2c\bar{u}]$. *Moreover, the contours* γ_1 *and* γ_2 *are disjoint.*

Remark 12. *When* $u_1, \cdots, u_p$ *are independent and hence* $\mathbf{v}_1, \cdots, \mathbf{v}_p$ *are independent, condition (5.40) holds automatically.*

In view of Theorem 9, we see that the proposed test statistic M_n still works mainly due to the involvement of the last term on the right–hand side of (5.11). Section 5 below employs the proposed test to evaluate the finite–sample performance and the practical applicability of the proposed test.

5.4 Small sample simulation studies

This section provides some simulated examples to evaluate the finite sample performance of the proposed test. In addition, we also compare the performance of the proposed test with that of a likelihood ratio test proposed by Anderson (1984). Simulations are used to evaluate both the empirical sizes and powers of the proposed test. To show the wide applicability and efficiency of our test, two kinds of dependent structures considered in Sections 3 and 4, such as multiple MA(1) and AR(1) model, SMA and the general panel data model, are investigated.

5.4.1 Empirical sizes and empirical power values

First we introduce the method of calculating empirical sizes and power values. Let $z_{\frac{1}{2}\alpha}$ and $z_{1-\frac{1}{2}\alpha}$ be the $100(\frac{1}{2}\alpha)\%$ and $100(1-\frac{1}{2}\alpha)\%$ quantiles of the asymptotic null distribution of the test statistic M_n respectively. With K replications of the data set simulated under the null hypothesis, we calculate the empirical size as

$$\hat{\alpha} = \frac{\{\sharp\ of\ M_n^H \geq z_{1-\frac{1}{2}\alpha}\ or\ M_n^H \leq z_{\frac{1}{2}\alpha}\}}{K}, \tag{5.44}$$

where M_n^H represents the values of the test statistic M_n based on the data simulated under the null hypothesis.

In our simulation, we choose $K = 1000$ as the number of repeated simulations. The significance level is $\alpha = 0.05$. Since the asymptotic null distribution of the test statistic is not a classical distribution, we need to estimate the quantiles $z_{\frac{1}{2}\alpha}$ and $z_{1-\frac{1}{2}\alpha}$. Naturally, we do as follows: generate K replications of the asymptotic distributed random variable and then select the $(K\frac{1}{2}\alpha)$-th smallest value $\hat{z}_{\frac{1}{2}\alpha}$ and $(K\frac{1}{2}\alpha)$-th largest value $\hat{z}_{1-\frac{1}{2}\alpha}$ as the estimated $100(\frac{1}{2}\alpha)\%$ and $100(1-\frac{1}{2}\alpha)\%$ quantiles of the asymptotic distributed random variable.

With the estimated critical points $\hat{z}_{\frac{1}{2}\alpha}$ and $\hat{z}_{1-\frac{1}{2}\alpha}$ under the null hypothesis, the empirical power is calculated as

$$\hat{\beta} = \frac{\{\sharp\ of\ M_n^A \geq \hat{z}_{1-\frac{1}{2}\alpha}\ or\ M_n^A \leq \hat{z}_{\frac{1}{2}\alpha}\}}{K}, \tag{5.45}$$

where M_n^A represents the values of the test statistic M_n based on the data simulated under the alternative hypothesis.

5.4.2 Comparisons with the classical likelihood ratio test

For the proposed test, we generate n numbers of p–dimensional independent and identical distributed random vectors $\{\mathbf{y}_j\}_{j=1}^n$, each with the mean vector $\mathbf{0}_p$ and the covariance matrix Σ. Under the null hypothesis, $\{\mathbf{y}_j\}_{j=1}^n$ are generated in two scenarios:

1. Each $\mathbf{w}_j$ is a p-dimensional normal random vector with the mean vector $\mathbf{0}_p$ and the covariance matrix $\Sigma = \mathbf{I}_p$; $\forall j = 1, \ldots n$, $\mathbf{y}_j = \mathbf{T}_p \mathbf{w}_j$ with $\mathbf{T}_p = \mathbf{I}_p$;

2. Each $\mathbf{w}_j$ consists of i.i.d. random variables with standardized Gamma(4,2) distribution, so they have zero means and unit variances; $\forall j = 1, \ldots n$, $\mathbf{y}_j = \mathbf{T}_p \mathbf{w}_j$ with $\mathbf{T}_p = \mathbf{I}_p$.

Under the alternative hypothesis, we consider the case:
$\mathbf{T}_p^{1/2} = (\sqrt{0.95}\mathbf{I}_p, \sqrt{0.05}\mathbf{1}_p)$, where $\mathbf{1}_p$ is a p-dimensional vector with 1 as entries. In this case, the population covariance matrix of $\mathbf{y}_j$ is $\Sigma = 0.95\mathbf{I}_p + 0.05\mathbf{1}_p\mathbf{1}_p^{'}$, which is called the compound symmetric covariance matrix.

For the normally distributed data, the fourth moment of each element is $E\left[X_{11}^4\right] = 3$; for standardized Gamma(4,2) distributional data, $E\left[X_{11}^4\right] = 4.5$. Anderson (1984) provides a likelihood ratio criterion (LRT) to test independence for a fixed number of fixed dimensional normal distributed random vectors. We compare it with the proposed test.

Under the null hypothesis, the distribution of L is the distribution of $L_2 L_3 \cdots L_p$, where $L_2, \ldots, L_p$ are independently distributed with each L_k having the distribution of $U_{m,(k-1)m,n-1-(k-1)m}$. Furthermore, for any $k = 2, \ldots, p$, as $n \to \infty$, $-(n - \frac{3}{2} - \frac{km}{2}) \log\left(U_{m,(k-1)m,n-1-(k-1)m}\right)$ has a χ^2

distribution of $(k-1)m^2$ degrees of freedom (See section 8.5 of Anderson (1984)).

From the construction of the LRT test, we can see that the LRT utilises additional n observations of the random vectors $\mathbf{x}_1, \ldots, \mathbf{x}_p$ under investigation, while the proposed test does not need this information. However, we can choose $m = 1$ and apply LRT to independence test of the random vectors $\tilde{\mathbf{x}}_1, \ldots, \tilde{\mathbf{x}}_p$, where for any $i = 1, \ldots, p$, the elements of the vector $\tilde{\mathbf{x}}_i$ consist of its n observations. Hence the LRT test can test independence for p numbers of random vectors with dimension n by choosing $m = 1$.

Tables 5.1 and 5.3 show the empirical sizes and empirical power values of our proposed test and the LRT test for the normally distributed random vectors respectively. From Table 5.1 and Table 5.3, we can see that the LRT test does not work when p and n are both large while the proposed test possesses good performance when p and n are both large and increase at the same order. The LRT test is only applicable to the case where p is fixed and n increases from $n = 5$ to $n = 100$. From Table 5.3, it can be seen that the LRT fails when p is large at the same order as that of n. When the difference between p and n is large, the sizes and the power values of the proposed test become worse. This is because our test is proposed under the restriction that p and n are required to increase and vary at the same order. The proposed test also works well for gamma random vectors while the LRT test is not applicable to gamma case, since, in theory, LRT test is originally proposed only for normal random vectors. Tables 5.4 and 5.5 provide the empirical sizes and power values of the proposed test for the gamma case. In our simulation, we choose $p, n = 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100$ for the proposed test and the LRT test. The significant level α is chosen as 0.05. In each case, we run $K = 1000$ repeated simulations. Our simulation results for the empirical powers show that the proposed test can well test independence for both normal and gamma vectors.

As comparison, without using the asymptotic distribution under the null hypothesis, we use the bootstrap method to obtain a bootstrap critical value (bcv) in each case to estimate the empirical sizes for normal and gamma distributed data. The results are listed in Table 5.2. From the table, we can see that the bootstrap sizes are better than those estimated from the null asymptotic distribution.

5.4.3 Multiple MA(1), AR(1) and SMA model

Consider multiple MA(1) model

$$\mathbf{v}_t = \mathbf{z}_t + \psi \mathbf{z}_{t-1}, t = 1, \ldots, p. \tag{5.46}$$

We choose $\psi = 0.5$ and $\mathbf{z}_t \overset{i.i.d}{\sim} normal(\mathbf{0}, \mathbf{I}_n), \forall t = 0, 1, \ldots, p$. The simulation results in Table 5.6 show that the proposed test performs well for this model.

Consider multiple AR(1)

$$\mathbf{v}_t = \phi \mathbf{v}_{t-1} + \mathbf{z}_t, \ t = 1, \ldots, p, \tag{5.47}$$

where $\mathbf{v}_0 = \frac{1}{\sqrt{1-\phi^2}} \mathbf{z}_0$ and $\mathbf{z}_t \overset{i.i.d}{\sim} normal(\mathbf{0}, \mathbf{I}_n), \forall t = 0, 1, \ldots, p$. Let $\phi = 0.5$. The empirical powers for this model are provided in Table 5.7. As n and p increase in the same order, the empirical power tends to 1.

As for the Spatial Moving Average (SMA) model, i.e.

$$v_{ji} = \sum_{k=1}^{p} \omega_{ik} \varepsilon_{jk} + \varepsilon_{ji}, \tag{5.48}$$

we generate $\varepsilon_{jk} \overset{i.i.d}{\sim} normal(1, 1), \forall j = 1, \ldots, n; k = 1, \ldots, p$. Apply the proposed statistic M_n for the sample matrix $\frac{1}{n} \mathbf{V}^* \mathbf{V}$ and the empirical powers are illustrated in Table 5.8. These power values show that M_n performs well for capturing the cross–sectional dependence for SMA model.

5.4.4 The general panel data model

We examine the finite sample performance of the proposed test for the general panel data model (5.36), i.e.

$$v_{ij} = \varepsilon_{ij} + \frac{1}{\sqrt{p}} u_i, \; i = 1, \ldots, p; \; j = 1, \ldots, n, \tag{5.49}$$

where $\{\varepsilon_{ij}, i = 1, \ldots, p; j = 1, \ldots, n\}$ is a sequence of i.i.d. random variables and $E\varepsilon_{11} = 0$, $E\varepsilon_{11}^2 = 1$; $\{u_i, i = 1, \ldots, p\}$ are independent of $\{\varepsilon_{ij}, i = 1, \ldots, p; j = 1, \ldots, n\}$.

Under the null hypothesis, we generate $u_i \overset{i.i.d}{\sim} normal(1,1)$, $i = 1, \ldots, p$ and under the alternative hypothesis, we experiment with $\mathbf{u} = (\frac{1}{\sqrt{p}} u_1, \frac{1}{\sqrt{p}} u_2, \ldots, \frac{1}{\sqrt{p}} u_p) \sim \frac{1}{\sqrt{p}} N(\mathbf{1}_p, \Sigma)$, where $\Sigma = \mathbf{T}\mathbf{T}^T$ and $\mathbf{T}$ is a $p \times p$ matrix whose elements are generated $t_{ik} \overset{i.i.d.}{\sim} U(0,1)$, $i, k = 1, \ldots, p$.

The simulation results including empirical sizes and power values in Table 5.9 show that the proposed test can capture the dependence for the general panel data model (5.36).

5.4.5 Some time series and Vandermonde matrix

Dependent structures of a set of random vectors are often described by non-zero correlations among them, such as the linear dependent structure developed in Section 3. However, there are some data which are not independent but uncorrelated. We consider three such examples and test their dependence by the proposed test.

5.4.5.1 Nonlinear MA model

Consider nonlinear MA models of the form

$$R_{tj} = Z_{t-1,j} Z_{t-2,j} (Z_{t-2,j} + Z_{tj} + 1), \; t = 1, \ldots, p; \; j = 1, \ldots, n; \tag{5.50}$$

where $\mathbf{z}_t = (Z_{t1}, \ldots, Z_{tn})$ is an n-dimensional random vector with i.i.d. elements, each of which has zero mean and unit variance; and $\mathbf{r}_t = (R_{t1}, \ldots, R_{tn})$. For any $j = 1, \ldots, n$, the correlation matrix of $(R_{j1}, R_{j2}, \ldots, R_{jp})$ is a diagonal matrix. This model is provided by Kuan and Lee (2004) which tests the martingale difference hypothesis. Our proposed independence test can be applied to this nonlinear MA model, and the powers in Table 10 show that this test performs well for this model.

This result also implies that the limit of the ESD of the nonlinear MA model (5.50) is not the M-P law since the proposed test statistic is established on the characteristic function of the M-P law.

5.4.5.2 Multiple ARCH(1) model

Consider the multiple autoregressive conditional heteroscedastic(ARCH(1)) model:

$$W_{tj} = Z_{tj} \sqrt{\alpha_0 + \alpha_1 W_{t-1,j}^2}, \; t = 1, \ldots, p; \; j = 1, \ldots, n; \tag{5.51}$$

where $\mathbf{z}_t = (Z_{t1}, \ldots, Z_{tn})$ is an n-dimensional random vector with i.i.d. elements, each of which has zero mean and unit variance; and $\boldsymbol{\omega}_t = (W_{t1}, \ldots, W_{tn})$.

For each $j = 1, \ldots, n$, ARCH(1) model $(W_{1j}, W_{2j}, \ldots, W_{pj})$ is a martingale difference sequence. ARCH(1) model has many applications in financial analysis. There exists no theoretical results stating that the limit of the ESD of the sample covariance matrix for ARCH(1) model is the M-P Law. A rigorous study is under investigation. For the ARCH(1) model, the proposed test can not capture the dependence of $(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \ldots, \boldsymbol{\omega}_p)$ directly, but we can test the dependence of $(\boldsymbol{\omega}_1^2, \boldsymbol{\omega}_2^2, \ldots, \boldsymbol{\omega}_p^2)$. Since this test can tell us that $(\boldsymbol{\omega}_1^2, \boldsymbol{\omega}_2^2, \ldots, \boldsymbol{\omega}_p^2)$ are not independent, naturally it can be concluded that $(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \ldots, \boldsymbol{\omega}_p)$ are not independent either. Here we take $\alpha_0 = 0.9$ and $\alpha_1 = 0.1$. Table 5.11 shows the power values of our test for testing dependence of ARCH(1) model.

5.4.5.3 Vandermonde matrix

Consider the $n \times p$ vandermonde matrix $\mathbf{V}$ of the form

$$\mathbf{V} = \frac{1}{\sqrt{n}} \begin{pmatrix} 1 & 1 & \cdots & 1 \\ e^{-i\omega_1} & e^{-i\omega_2} & \cdots & e^{-i\omega_p} \\ \vdots & \vdots & \ddots & \vdots \\ e^{-i(n-1)\omega_1} & e^{-i(n-1)\omega_2} & \ldots & e^{-i(n-1)\omega_p} \end{pmatrix} \tag{5.52}$$

where $\omega_i, i = 1, \ldots, p$ are called phased distributions and are assumed i.i.d on $[0, 2\pi)$. Then the entries of $\mathbf{V}$ lie on the unit circle. Obviously, all the entries of the rows of $\mathbf{V}$ are not independent while the columns are independent. Denote the sample covariance matrix of $\mathbf{V}$ by $\mathbf{D} = \mathbf{V}^H\mathbf{V}$.

Vandermonde matrices play an important role in signal processing and wireless applications such as direction of arrival estimation, pre-coding or sparse sampling theory, etc.. $\emptyset$. Ryan and Debbah (2009) established that as both n, p go to ∞ with their ratio being a positive constant, the limiting spectral distribution of $\mathbf{D} = \mathbf{V}^H\mathbf{V}$ is not the M-P law. From Theorem 8 we see that the proposed test could capture the dependence structure of the rows of the matrix $\mathbf{V}$. It is easy to see that, for any $k = 1, \ldots, n-1$ and $j = 1, \ldots, p$, $E(e^{-ik\omega_j})^2 = 0$ and $E|e^{-ik\omega_j}|^4 = 1$. The empirical power values in Table 5.12 show that the proposed test works well in detecting dependence of Vandermonde matrices.

5.5 Conclusions

This chapter has established a general test for testing independence among a large number of high dimensional random vectors based on the characteristic function of the empirical spectral distribution of the sample covariance matrix of the random vectors. This test can capture various kinds of dependent structures, e.g. MA(1), AR(1) model, nonlinear MA(1) model, ARCH(1) model and the general panel data model established in the simulation section. The conventional method (LRT proposed by Anderson (1984)) utilizes the correlated relationship between random vectors to capture their dependence. This idea is only efficient for normal distributed data. It may be an inappropriate tool for non-Gaussian distributed data, such as martingale difference sequences (e.g.ARCH(1) model), nonlinear MA(1) model, the Vandermonde matrix, etc., which possess dependent but uncorrelated structures. The proposed test is not restricted to normally distributed data. In general, the proposed test is proposed for testing independence among a large number of high dimensional random vectors.

5.6 Appendix

5.6.1 Some useful lemmas

Lemma 16 (Theorem 8.1 of Billingsley (1999)). *Let P_n and P be probability measures on (C, φ). If the finite dimensional distributions of P_n converge weakly to those of P, and if $\{P_n\}$ is tight, then $P_n \Rightarrow P$.*

Lemma 17 (Theorem 12.3 of Billingsley (1999)). *The sequence $\{X_n\}$ is tight if it satisfies these two conditions*

(I) The sequence $\{X_n(0)\}$ is tight.

(II) There exists constants $\gamma \geq 0$, $\alpha > 1$, and a nondecreasing, continuous function F on $[0,1]$ such that

$$E\{|X_n(t_2) - X_n(t_1)|^\gamma\} \leq |F(t_2) - F(t_1)|^\alpha \tag{5.53}$$

holds for all t_1, t_2, and n.

Lemma 18 (Continuous Theorem)**.** *Let X_n and X be random elements defined on a metric space S. Suppose $g : S \to S^{'}$ has a set of discontinuous points D_g such that $P(X \in D_g) = 0$. Then*

$$X_n \xrightarrow{d} X \Rightarrow g(X_n) \xrightarrow{d} g(X). \tag{5.54}$$

Lemma 19 (Complex mean value theorem (see Lemma 2.4 of Guo and Higham (2006)))**.** *Let Ω be an open convex set in $\mathbb{C}$. If $f : \Omega \to \mathbb{C}$ is an analytic function and a, b are distinct points in Ω, then there exist points u, v on $L(a,b)$ such that*

$$Re(\frac{f(a) - f(b)}{a - b}) = Re(f^{'}(u)), \ Im(\frac{f(a) - f(b)}{a - b}) = Im(f^{'}(v)), \tag{5.55}$$

where $Re(z)$ and $Im(z)$ are the real and imaginary parts of z respectively; and $L(a,b) \triangleq \{a + t(b-a) : t \in (0,1)\}$.

5.6.2 Proofs of the main theorems

This section provides the proofs of three main theorems. Lemmas 20 and 21 involved in the respective proofs of Theorem 7 and Theorem 9 are listed and proved in Section 7.3 below.

Proof of Theorem 7. Let t belong to a closed interval $I = [T_1, T_2]$. To finish Theorem 7, in view of Lemma 16 and Lemma 20, it suffices to prove the tightness of $\{\big(\phi_n(t), \psi_n(t)\big) : t \in I\}$. Thus it suffices to prove the tightness of $p(s_n(t) - s(t))$. Repeating the same truncation and centralization steps as those in Bai and Silverstein (2004), we may assume that

$$|X_{ij}| < \delta_n \sqrt{n}, \ EX_{ij} = 0, \ E|X_{ij}|^2 = 1, \ E|X_{ij}|^4 < \infty. \tag{5.56}$$

Set $M_n(z) = n[m_{F^{\mathbf{A}_n}}(z) - m_{F^{c_n,H_n}}(z)]$. By the Cauchy theorem

$$f(x) = -\frac{1}{2\pi i}\oint \frac{f(z)}{z-x}dz, \tag{5.57}$$

we have, with probability one, for all n large,

$$\int e^{itx}dp(F^{\mathbf{A}_n}(x) - F^{c_n,H_n}(x)) = -\frac{1}{2\pi i}\oint_{\mathcal{C}} e^{itz}M_n(z)dz. \tag{5.58}$$

The contour $\mathcal{C}$ involved in the above integral is specified as follows. Let

$$\mathcal{C}_u = \{x + iv_0 : x \in [x_l, x_r]\}, \tag{5.59}$$

where $v_0 > 0$, x_r is any number greater than $\limsup_n \lambda_{max}(\mathbf{T}_n)(1+\sqrt{c})^2$, x_l is any negative number if $c \geq 1$ and otherwise choose $x_l \in (0, \limsup_n \lambda_{min}(\mathbf{T}_n)(1-\sqrt{c})^2)$. Then the contour $\mathcal{C}$ is defined by the union of $\mathcal{C}_+$ and its symmetric part $\mathcal{C}_-$ with respect to the x-axis, where

$$\mathcal{C}_+ = \{x_l + iv : v \in [0, v_0]\} \cup \mathcal{C}_u \cup \{x_r + iv : v \in [0, v_0]\}. \tag{5.60}$$

From Lemma 1 and the argument regarding equivalence in probability of $M_n(z)$ and its truncation version given in Page 563 in Bai and Silverstein (2004) and Lemma 3 we have

$$\oint_{\mathcal{C}} |M_n(z)||dz| \xrightarrow{D} \oint_{\mathcal{C}} |M(z)||dz|, \tag{5.61}$$

where $M(z)$ is a Gaussian process, the limit of $M_n(z)$.

We conclude from Lemma 19 that for any $\delta > 0$

$$\begin{aligned}
&\sup_{|t_1-t_2|<\delta, t_1,t_2\in I} \Big|\oint_{\mathcal{C}} (e^{it_1 z} - e^{it_2 z})M_n(z)dz\Big| \\
\leq\ & \sup_{|t_1-t_2|<\delta, t_1,t_2\in I} \Big|\oint_{\mathcal{C}} \sqrt{\big(Re(ize^{it_3 z})\big)^2 + \big(Im(ize^{it_4 z})\big)^2}\,\delta|M_n(z)||dz|\Big| \\
\leq\ & K\delta\Big|\oint_{\mathcal{C}} |M_n(z)||dz|\Big| \xrightarrow{D} K\delta\Big|\oint_{\mathcal{C}} |M(z)||dz|\Big|, \quad \text{as} \quad n\to\infty,
\end{aligned} \tag{5.62}$$

where t_3 and t_4 lies in the interval $[T_1, T_2]$, the last inequality uses (5.61) and the fact that $Re(ize^{it_3 z})$, $Im(ize^{it_4 z})$ are bounded on the contour $\mathcal{C}$ and K (and in the sequel) is a constant number which may be different from line to line.

By (5.62), we have for any $\varepsilon > 0$,

$$P\Big(\sup_{|t_1-t_2|<\delta, t_1,t_2\in[0,1]} \Big|\oint_{\mathcal{C}} (e^{it_1 z} - e^{it_2 z}) M_n(z)dz\Big| \geq \varepsilon\Big) \leq P\Big(K\delta\Big|\oint_{\mathcal{C}} |M_n(z)||dz|\Big| \geq \varepsilon\Big) \tag{5.63}$$

and

$$\lim_{\delta\to 0}\limsup_{n\to\infty} P\Big(K\delta\Big|\oint_{\mathcal{C}} |M_n(z)||dz|\Big| \geq \varepsilon\Big) = \lim_{\delta\to 0} P\Big(K\delta\Big|\oint_{\mathcal{C}} |M(z)||dz|\Big| \geq \varepsilon\Big) = 0. \tag{5.64}$$

Hence (5.63) and (5.64) imply that

$$\lim_{\delta\to 0}\limsup_{n\to\infty} P\Big(\sup_{|t_1-t_2|<\delta, t_1,t_2\in I} \Big|\oint_{\mathcal{C}} (e^{it_1 z} - e^{it_2 z}) M_n(z)dz\Big| \geq \varepsilon\Big) = 0. \tag{5.65}$$

By Theorem 7.3 of Billingsley (1999), $\int e^{itx} dp(F^{\mathbf{A}_n}(x) - F^{c_n,H_n}(x))$ is tight. Moreover from the assumption we see that $\int e^{itx} dp(F^{c_n,H_n}(x) - F^{c_n}(x))$ is tight by Lemma 4.

□

Proof of Theorem 8. Consider $p^2 M_n$ under the alternative $\mathbf{H}_1$ and rewrite it as follows

$$p^2 M_n = p^2 \int_{T_1}^{T_2} |s_n(t) - s(t)|^2 dU(t) = \int_{T_1}^{T_2} \Big[M_n^{\cos,\mathbf{H}_1}(t)\Big]^2 dU(t) + \int_{T_1}^{T_2} \Big[M_n^{\sin,\mathbf{H}_1}(t)\Big]^2 dU(t),$$

where

$$M_n^{\cos,\mathbf{H}_1}(t) = \int \cos(tx) dp(F_{\mathbf{H}_1}^{\mathbf{A}_n}(x) - F^{c_n}(x)),\ M_n^{\sin,\mathbf{H}_1}(t) = \int \sin(tx) dp(F_{\mathbf{H}_1}^{\mathbf{A}_n}(x) - F^{c_n}(x)).$$

We may further write

$$\Big[M_n^{\cos,\mathbf{H}_1}(t)\Big]^2 = \Big[M_n^{\cos,\mathbf{H}_0}(t)\Big]^2 + \Big[M_n^{\cos,\mathbf{H}_1,\mathbf{H}_0}(t)\Big]^2 + 2\Big[M_n^{\cos,\mathbf{H}_1,\mathbf{H}_0}(t)\Big]\Big[M_n^{\cos,\mathbf{H}_0}(t)\Big],$$

where $M_n^{\cos,\mathbf{H}_0}(t)$ is obtained from $M_n^{\cos,\mathbf{H}_1}(t)$ with $F_{\mathbf{H}_1}^{\mathbf{A}_n}(x)$ replaced by $F_{\mathbf{H}_0}^{\mathbf{A}_n}(x)$ and

$$M_n^{\cos,\mathbf{H}_1,\mathbf{H}_0}(t) = \int \cos(tx) dp(F_{\mathbf{H}_1}^{\mathbf{A}_n}(x) - F_{\mathbf{H}_0}^{\mathbf{A}_n}(x)).$$

By Holder's inequality we obtain

$$\Big|\int_{T_1}^{T_2}\Big[M_n^{\cos,\mathbf{H}_1,\mathbf{H}_0}(t)\Big]\Big[M_n^{\cos,\mathbf{H}_0}(t)\Big]dU(t)\Big|^2\le\int_{T_1}^{T_2}\Big[M_n^{\cos,\mathbf{H}_1,\mathbf{H}_0}(t)\Big]^2dU(t)\int_{T_1}^{T_2}\Big[M_n^{\cos,\mathbf{H}_0}(t)\Big]^2dU(t)\Big|.$$

This, together with Lemma 20 (see Section 7.3 below) and the proof of Theorem 1, implies that

$$\int_{T_1}^{T_2}\Big[M_n^{\cos,\mathbf{H}_1,\mathbf{H}_0}(t)\Big]\Big[M_n^{\cos,\mathbf{H}_0}(t)\Big]dU(t)=o_p\Big(\int_{T_1}^{T_2}(M_n^{\cos,\mathbf{H}_1,\mathbf{H}_0}(t))^2+(M_n^{\sin,\mathbf{H}_1,\mathbf{H}_0}(t))^2dU(t)\Big),$$

where

$$M_n^{\sin,\mathbf{H}_1,\mathbf{H}_0}(t)=\int\sin(tx)dp(F_{\mathbf{H}_1}^{\mathbf{A}_n}(x)-F_{\mathbf{H}_0}^{\mathbf{A}_n}(x)).$$

Similarly

$$\Big[M_n^{\sin,\mathbf{H}_1}(t)\Big]^2=\Big[M_n^{\sin,\mathbf{H}_0}(t)\Big]^2+\Big[M_n^{\sin,\mathbf{H}_1,\mathbf{H}_0}(t)\Big]^2+2\Big[M_n^{\sin,\mathbf{H}_1,\mathbf{H}_0}(t)\Big]\Big[M_n^{\sin,\mathbf{H}_0}(t)\Big]$$

and

$$\int_{T_1}^{T_2}\Big[M_n^{\sin,\mathbf{H}_1,\mathbf{H}_0}(t)\Big]\Big[M_n^{\sin,\mathbf{H}_0}(t)\Big]dU(t)=o_p\Big(\int_{T_1}^{T_2}(M_n^{\sin,\mathbf{H}_1,\mathbf{H}_0}(t))^2+(M_n^{\sin,\mathbf{H}_1,\mathbf{H}_0}(t))^2dU(t)\Big),$$

where $M_n^{\sin,\mathbf{H}_0}(t)$ is similarly defined. Note that

$$\int_{T_1}^{T_2}(M_n^{\sin,\mathbf{H}_1,\mathbf{H}_0}(t))^2+(M_n^{\sin,\mathbf{H}_1,\mathbf{H}_0}(t))^2dU(t)=\int_{T_1}^{T_2}\Big|\int e^{itx}dp(F_{\mathbf{H}_1}^{\mathbf{A}_n}(x)-F_{\mathbf{H}_0}^{\mathbf{A}_n}(x))\Big|^2dU(t).$$

Summarizing the above we have obtained

$$\begin{aligned}p^2M_n=&\int_{T_1}^{T_2}(\Big[M_n^{\cos,\mathbf{H}_0}(t)\Big]^2+\Big[M_n^{\sin,\mathbf{H}_0}(t)\Big]^2)dU(t)+\int_{T_1}^{T_2}\Big|\int e^{itx}dp(F_{\mathbf{H}_1}^{\mathbf{A}_n}(x)-F_{\mathbf{H}_0}^{\mathbf{A}_n}(x))\Big|^2dU(t)\\&+o_p\Big(\int_{T_1}^{T_2}\Big|\int e^{itx}dp(F_{\mathbf{H}_1}^{\mathbf{A}_n}(x)-F_{\mathbf{H}_0}^{\mathbf{A}_n}(x))\Big|^2dU(t)\Big).\end{aligned}$$

Thus, Theorem 8 follows from Theorem 1 and condition (5.22).

□

Proof of Theorem 9. As in the proof of Theorem 7, in view of Lemma 21 (see Section 7.3 below), it suffices to prove the tightness of $\{p(s_n(t)-s(t)):t\in I\}$. As before, write

$$p(s_n(t)-s(t))\;=\;p\int e^{itx}d[F^{\mathbf{S}_n}(x)-F^{c_n}(x)]$$

$$= \quad -\frac{1}{2\pi i}\oint_{\gamma} e^{itz}(tr(\mathbf{S}_n - z\mathbf{I}_p)^{-1} - pm_{c_n}(z))dz, \quad (5.66)$$

where the contour γ is specified in Lemma 21.

From the formula (5.98), we have

$$tr\mathcal{D}_n^{-1}(z) = tr\mathbf{D}_n^{-1}(z) + \frac{\bar{\varepsilon}^T\mathbf{D}_n^{-2}(z)\bar{\varepsilon}}{1-\bar{\varepsilon}^T\mathbf{D}_n^{-1}(z)\bar{\varepsilon}}. \quad (5.67)$$

This, together with (5.97), yields

$$tr\mathbf{S}_n^{-1}(z) - pm_{c_n}(z) = tr\mathbf{D}_n^{-1}(z) - pm_{c_n}(z) + \frac{\bar{\varepsilon}^T\mathbf{D}_n^{-2}(z)\bar{\varepsilon}}{1-\bar{\varepsilon}^T\mathbf{D}_n^{-1}(z)\bar{\varepsilon}} - \frac{\bar{\mathbf{v}}^T\mathcal{D}_n^{-2}(z)\bar{\mathbf{v}}}{1+\bar{\mathbf{v}}^T\mathcal{D}_n^{-1}(z)\bar{\mathbf{v}}}. \quad (5.68)$$

By (5.94) and noting that $M_n(z) = tr\mathbf{D}_n^{-1}(z) - pm_{c_n}(z)$, it is sufficient to prove the tightness of the following three terms:

$$g_{n1}(t) = -\frac{1}{2\pi i}\oint_{\gamma} e^{itz}M_n(z)dz, \quad (5.69)$$

$$g_{n2}(t) = -\frac{1}{2\pi i}\oint_{\gamma} e^{itz}\frac{\bar{\varepsilon}^T\mathbf{D}_n^{-2}(z)\bar{\varepsilon}}{1-\bar{\varepsilon}^T\mathbf{D}_n^{-1}(z)\bar{\varepsilon}}dz, \quad (5.70)$$

$$g_{n3}(t) = -\frac{1}{2\pi i}\oint_{\gamma} e^{itz}\frac{\bar{\mathbf{v}}^T\mathcal{D}_n^{-2}(z)\bar{\mathbf{v}}}{1+\bar{\mathbf{v}}^T\mathcal{D}_n^{-1}(z)\bar{\mathbf{v}}}dz, \quad (5.71)$$

The tightness of $\{g_{n1}(t) : t \in I = [T_1, T_2]\}$ has been proved in Theorem 7. Next, via the same method adopted by Theorem 7, we prove the tightness of $\{g_{ni}(t) : t \in I = [T_1, T_2]\}$, $i = 2, 3$ as follows.

By (5.102), (5.103) and Slutsky's theorem, we have

$$\sup_{z\in\gamma}\Big|\frac{\bar{\varepsilon}^T\mathbf{D}_n^{-2}(z)\bar{\varepsilon}}{1-\bar{\varepsilon}^T\mathbf{D}_n^{-1}(z)\bar{\varepsilon}} + \frac{c\underline{m}(z)}{z\big((1+\underline{m}(z))^2 - c\underline{m}^2(z)\big)}\Big| \xrightarrow{i.p.} 0. \quad (5.72)$$

We conclude from (5.72), (5.123) and Lemma 3 that, as $n \to \infty$,

$$\oint_{\gamma}\Big|\frac{\bar{\varepsilon}^T\mathbf{D}_n^{-2}(z)\bar{\varepsilon}}{1-\bar{\varepsilon}^T\mathbf{D}_n^{-1}(z)\bar{\varepsilon}}\Big||dz| \xrightarrow{a.s.} \oint_{\gamma}\Big|\frac{c\underline{m}(z)}{z\big((1+\underline{m}(z))^2 - c\underline{m}^2(z)\big)}\Big||dz| \quad (5.73)$$

and

$$\oint_{\gamma}\Big|\frac{\bar{\mathbf{v}}^T\mathcal{D}_n^{-2}(z)\bar{\mathbf{v}}}{1+\bar{\mathbf{v}}^T\mathcal{D}_n^{-1}(z)\bar{\mathbf{v}}}\Big||dz| \xrightarrow{a.s.} \oint_{\gamma}\Big|\frac{\frac{c\underline{m}(z)}{z[(1+\underline{m}(z))^2 - c\underline{m}^2(z)]} + \bar{u}\int\frac{1}{(\lambda-z)^2}dF^{MP}(\lambda)}{\bar{u}z\underline{m}^2(z) - 1}\Big||dz|.$$

$$(5.74)$$

By (5.73), (5.74) and the same proof as (5.62) to (5.65), the tightness of $\{g_{ni}(t) : t \in I = [T_1, T_2]\}$, $i = 2, 3$ can be derived.

□

5.6.3 Proofs of Lemma 20 and Lemma 21

Bai and Silverstein (2004) established the remarkable central limit theorem for functional of eigenvalues of $\mathbf{A}_n$ under the additional assumption that $E|X_{11}|^4 = 3$ while Pan and Zhou (2008) provided a supplement to this theorem by eliminating the condition to some extent. From Theorem 4 of Pan and Zhou (2008) and (5.11) we can directly obtain the following lemma.

Lemma 20. *Under Assumptions 2 and 3, we have, for any positive integer k,*

$$\Big(\int cos(t_1x)dG_n(x), \ldots, \int cos(t_kx)dG_n(x), \int sin(t_1x)dG_n(x), \ldots, \int sin(t_kx)dG_n(x)\Big) \tag{5.75}$$

converges in distribution to Gaussian vectors $(V_1 + \delta_1(t_1), \ldots, V_k + \delta_1(t_k), Z_1 + \delta_2(t_1), \ldots, Z_k + \delta_2(t_k))$, where $\delta_1(t)$, $\delta_2(t)$ are, respectively, defined as

$$\delta_1(t) = \lim_{n\to\infty} \int cos(tx)dp(F^{c_n,H_n}(x) - F^{c_n}(x)), \tag{5.76}$$

$$\delta_2(t) = \lim_{n\to\infty} \int sin(tx)dp(F^{c_n,H_n}(x) - F^{c_n}(x)). \tag{5.77}$$

The means and covariances of V_j and Z_j are specified as follows:

If X_{11} is real, then for any $j = 1, \ldots, k$,

$$\begin{aligned} EV_j &= \frac{1}{2\pi i}\oint_\gamma cos(t_jz)\frac{c\int \underline{m}^3(z)\tau^2(1+\tau\underline{m}(z))^{-3}dH(\tau)}{(1-c\int \underline{m}^2(z)\tau^2(1+\tau\underline{m}(z))^{-2}dH(\tau))^2}dz \\ &\quad -\frac{EX_{11}^4-3}{2\pi i}\oint_\gamma cos(t_jz)\frac{c\underline{m}^3(z)h_2(z)}{1-c\int \underline{m}^2(z)\tau^2(1+\tau\underline{m}(z))^{-2}dH(\tau)}dz. \end{aligned}$$

Replacing $cos(t_jz)$ in EV_j by $sin(t_jz)$ yields the expression of EZ_j. For any $j, h = 1, \ldots, k$,

$$Cov(V_j, Z_h) = -\frac{1}{2\pi^2}\oint_{\gamma_1}\oint_{\gamma_2}\frac{cos(t_jz_1)sin(t_hz_2)}{(\underline{m}(z_1)-\underline{m}(z_2))^2}\frac{d}{dz_2}\underline{m}(z_2)\frac{d}{dz_1}\underline{m}(z_1)dz_1dz_2$$

$$-\frac{c(EX_{11}^4-3)}{4\pi^2}\oint_{\gamma_1}\oint_{\gamma_2} cos(t_j z_1)sin(t_h z_2)\frac{d^2}{dz_1 dz_2}[\underline{m}(z_1)\underline{m}(z_2)h_1(z_1,z_2)]dz_1 dz_2.$$

If X_{11} is complex with $EX_{11}^2=0$, then

$$EV_j=-\frac{E|X_{11}|^4-2}{2\pi i}\oint_{\gamma} cos(t_j z)\frac{c\underline{m}^3(z)h_2(z)}{1-c\int \underline{m}^2(z)\tau^2 dH(\tau)/(1+\tau\underline{m}(z))^2}dz,$$

and the covariance

$$Cov(V_j,Z_h)=-\frac{1}{4\pi^2}\oint_{\gamma_1}\oint_{\gamma_2}\frac{cos(t_j z_1)sin(t_h z_2)}{(\underline{m}(z_1)-\underline{m}(z_2))^2}\frac{d}{dz_2}\underline{m}(z_2)\frac{d}{dz_1}\underline{m}(z_1)dz_1 dz_2$$
$$-\frac{c(E|X_{11}|^4-2)}{4\pi^2}\oint_{\gamma_1}\oint_{\gamma_2} cos(t_j z_1)sin(t_h z_2)\frac{d^2}{dz_1 dz_2}[\underline{m}(z_1)\underline{m}(z_2)h_1(z_1,z_2)]dz_1 dz_2.$$

The contours γ, γ_1 and γ_2 above are all closed and are taken in the positive direction in the complex plane, each enclosing the support of $F^{c,H}$. Also γ_1 and γ_2 are disjoint.

The covariance structures $Cov(V_j,V_h)$ and $Cov(Z_j,Z_h)$ are similar to $Cov(V_j,Z_h)$ except replacing $(cos(t_j z), sin(t_h z))$ by $(cos(t_j z), cos(t_h z))$ and $(sin(t_j z), sin(t_h z))$ respectively.

Remark 13. *When $\mathbf{T}_n=\mathbf{I}$, the mean and variance of the asymptotic Gaussian distribution for power functions $f(x)=x^r$, $\forall r\in\mathbb{Z}^+$ is calculated in Pan and Zhou (2008) and Bai and Silverstein (2004). Hence the corresponding means and covariances for $f_1(x)=sintx$ and $f_2(x)=costx$ can be derived by Taylor series of $sintx$ and $costx$.*

While Lemma 21 below is only used in part of the proof of Theorem 3, it is of some general interest and its proof is also not trivial. We thus include both the statement of this lemma and its proof in this chapter.

Write

$$H_n(x)=p[F^{\mathbf{S}_n}(x)-F^{c_n}(x)], \tag{5.78}$$

where $\mathbf{S}_n$ is defined in (5.39).

Lemma 21. *Under the assumptions of Theorem 9, we have for any positive integer k,*

$$\Big(\int cos(t_1x)dH_n(x), \ldots, \int cos(t_kx)dH_n(x), \int sin(t_1x)dH_n(x), \ldots, \int sin(t_kx)dH_n(x)\Big) \tag{5.79}$$

converges in distribution to a Gaussian vector $(W_1, \ldots, W_k, Q_1, \ldots, Q_k)$ *whose mean and covariance are specified as follows:*

$$\begin{aligned} EW_j &= -\frac{c}{2\pi i}\oint_\gamma cos(t_jz)\frac{\underline{m}^3(z)(1+\underline{m}(z))}{((1+\underline{m}(z))^2 - c\underline{m}^2(z))^2}dz \\ &- \frac{c(EY_{11}^4-3)}{2\pi i}\oint_\gamma cos(t_jz)\frac{\underline{m}^3(z)}{(1+\underline{m}(z))^2 - c\underline{m}^2(z)}dz \\ &+ \frac{c}{2\pi i}\oint_\gamma cos(t_jz)\frac{\underline{m}(z)}{z((1+\underline{m}(z))^2 - c\underline{m}^2(z))}dz \\ &- \frac{1}{2\pi i}\oint_\gamma cos(t_jz)\frac{\frac{c\underline{m}(z)}{z[(1+\underline{m}(z))^2 - c\underline{m}^2(z)]} + \bar{u}\int\frac{1}{(\lambda-z)^2}dF^{MP}(\lambda)}{\bar{u}z\underline{m}^2(z)-1}dz; \end{aligned} \tag{5.80}$$

and

$$\begin{aligned} Cov(W_j, Q_h) &= -\frac{1}{2\pi^2}\oint_{\gamma_1}\oint_{\gamma_2}\frac{cos(t_jz_1)sin(t_hz_2)}{(\underline{m}(z_1)-\underline{m}(z_2))^2}\frac{d}{dz_1}\underline{m}(z_1)\frac{d}{dz_2}\underline{m}(z_2)dz_1dz_2 \\ &- \frac{c(EY_{11}^4-3)}{4\pi^2}\oint_{\gamma_1}\oint_{\gamma_2}cos(t_jz_1)cos(t_hz_2)\frac{d^2}{dz_1dz_2}[\frac{\underline{m}(z_1)\underline{m}(z_2)}{(1+\underline{m}(z_1))(1+\underline{m}(z_2))}]dz_1dz_2. \end{aligned} \tag{5.81}$$

Replacing $cos(t_jz)$ *in* EW_j *by* $sin(t_jz)$ *yields the expression of* EQ_j. *The covariances* $Cov(W_j, W_h)$ *and* $Cov(Q_j, Q_h)$ *are similar except replacing* $sin(t_hz)$ *and* $cos(t_jz)$ *by* $cos(t_hz)$ *and* $sin(t_jz)$ *respectively. The contours in (5.80) and (5.81) both enclose the interval* $[(1-\sqrt{c})^2 + 2c\bar{u}, (1+\sqrt{c})^2 + 2c\bar{u}]$. *Moreover, the contours* γ_1 *and* γ_2 *are disjoint.*

Proof of Lemma 21. Repeating the same truncation and centralization steps as those in Bai and Silverstein (2004), we may assume that

$$|\varepsilon_{ij}| < \delta_n\sqrt{n}, \ E\varepsilon_{ij} = 0, \ E|\varepsilon_{ij}|^2 = 1, \ E|\varepsilon_{ij}|^4 < \infty. \tag{5.82}$$

For the panel data model (5.36), let

$$\mathbf{v}_j = (v_{1j}, \ldots, v_{pj})^T, \ \boldsymbol{\varepsilon}_j = (\boldsymbol{\varepsilon}_{1j}, \ldots, \boldsymbol{\varepsilon}_{pj})^T, \ \mathbf{u} = (\frac{1}{\sqrt{p}}u_1, \ldots, \frac{1}{\sqrt{p}}u_p)^T, \ j = 1, \ldots, n. \tag{5.83}$$

The model can be written in the vector form as

$$\mathbf{v}_j = \boldsymbol{\varepsilon}_j + \mathbf{u}, \ j = 1, \ldots, n. \tag{5.84}$$

We then define the sample covariance matrix by $\mathbf{S}_n = \frac{1}{n}\sum_{j=1}^{n} \mathbf{v}_j \mathbf{v}_j^T$. Moreover write

$$\bar{\mathbf{v}} = \frac{1}{n}\sum_{j=1}^{n} \mathbf{v}_j, \ \bar{\boldsymbol{\varepsilon}} = \frac{1}{n}\sum_{j=1}^{n} \boldsymbol{\varepsilon}_j, \tag{5.85}$$

and

$$\mathbf{D}_n = \frac{1}{n}\sum_{j=1}^{n} \boldsymbol{\varepsilon}_j \boldsymbol{\varepsilon}_j^T, \quad \mathcal{S}_n = \frac{1}{n}\sum_{j=1}^{n} (\mathbf{v}_j - \bar{\mathbf{v}})(\mathbf{v}_j - \bar{\mathbf{v}})^T, \quad \mathcal{D}_n = \frac{1}{n}\sum_{j=1}^{n} (\boldsymbol{\varepsilon}_j - \bar{\boldsymbol{\varepsilon}})(\boldsymbol{\varepsilon}_j - \bar{\boldsymbol{\varepsilon}})^T. \tag{5.86}$$

Note that $\mathcal{S}_n = \mathcal{D}_n$. The sample covariance matrix $\mathbf{S}_n$ can be then expressed as

$$\mathbf{S}_n = \mathcal{S}_n + \bar{\mathbf{v}}\bar{\mathbf{v}}^T = \mathcal{D}_n + \bar{\mathbf{v}}\bar{\mathbf{v}}^T. \tag{5.87}$$

By (5.40) and the Burkholder inequality we have

$$E|\mathbf{u}^T \bar{\boldsymbol{\varepsilon}}|^4 \le \frac{1}{n^4} E|\sum_{j=1}^{n} \mathbf{u}^T \boldsymbol{\varepsilon}_j|^4 \le \frac{K}{n^4} E|\sum_{j=1}^{n} \mathbf{u}^T \mathbf{u}|^2 + \frac{K}{n^4}\sum_{j=1}^{n} E|\mathbf{u}^T \boldsymbol{\varepsilon}_j|^4 = O(\frac{1}{n^2}),$$

which, together with Borel-Cantelli's Lemma, implies that

$$\mathbf{u}^T \bar{\boldsymbol{\varepsilon}} \xrightarrow{a.s.} 0.$$

Also, condition (5.40) ensures that

$$\mathbf{u}^T \mathbf{u} \to \bar{u}. \tag{5.88}$$

Therefore by (2.25) of Pan (2012) and (5.40), we have, as $n \to \infty$,

$$\lambda_{max}(\bar{\mathbf{v}}\bar{\mathbf{v}}^T) = \bar{\mathbf{v}}^T\bar{\mathbf{v}} = \bar{\varepsilon}^T\bar{\varepsilon} + \mathbf{u}^T\mathbf{u} + 2\mathbf{u}^T\bar{\varepsilon} \xrightarrow{a.s.} c + \bar{u}, \quad as\ n \to \infty. \tag{5.89}$$

Furthermore, Jiang (2004) proved that

$$\lambda_{max}(\mathcal{D}_n) \to (1+\sqrt{c})^2,\ a.s.\ as\ n \to \infty \tag{5.90}$$

and Xiao and Zhou (2010) proved that, when $c \le 1$

$$\lambda_{min}(\mathcal{D}_n) \to (1-\sqrt{c})^2,\ a.s.\ as\ n \to \infty. \tag{5.91}$$

By (5.89) (5.90) and (5.91), the maximal and minimal eigenvalues of $\mathbf{S}_n$ satisfy with probability one

$$\limsup_{n\to\infty} \lambda_{max}(\mathbf{S}_n) \le c + \bar{u} + (1+\sqrt{c})^2, \tag{5.92}$$

and

$$\liminf_{n\to\infty} \lambda_{min}(\mathbf{S}_n) \ge (1-\sqrt{c})^2. \tag{5.93}$$

As in the proof of Theorem 7, we obtain from Cauchy's formula, with probability one, for n large,

$$\begin{aligned} p\int f(x)d[F^{\mathbf{S}_n}(x) - F^{c_n}(x)] &= \frac{p}{2\pi i}\int\oint_\gamma \frac{f(z)}{z-x}dzd[F^{\mathbf{S}_n}(x) - F^{c_n}(x)] \\ = & \frac{p}{2\pi i}\oint_\gamma f(z)dz\int\frac{1}{z-x}d[F^{\mathbf{S}_n}(x) - F^{c_n}(x)] \\ = & -\frac{1}{2\pi i}\oint_\gamma f(z)(tr(\mathbf{S}_n - z\mathbf{I}_p)^{-1} - pm_{c_n}(z))dz, \end{aligned} \tag{5.94}$$

where $m_{c_n}(z)$ is obtained from $m(z)$ with c replaced by c_n. The contour γ is specified as follows: Let $v_0 > 0$ be arbitrary and set $\gamma_\mu = \{\mu + iv_0, \mu \in [\mu_\ell, \mu_r]\}$, where $\mu_r > c + \bar{u} + (1+\sqrt{c})^2$ and $0 < \mu_\ell < I_{(0,1)}(c)(1-\sqrt{c})^2$ or μ_ℓ is any negative number if $c \ge 1$. Then define

$$\gamma^+ = \{\mu_\ell + iv : v \in [0, v_0]\} \cup \gamma_\mu \cup \{\mu_r + iv : v \in [0, v_0]\} \tag{5.95}$$

and let γ^- be the symmetric part of γ^+ about the real axis. Then set $\gamma = \gamma^+ \cup \gamma^-$.

Set

$$\begin{aligned} &\mathcal{S}_n^{-1}(z) = (\mathcal{S}_n - z\mathbf{I}_p)^{-1}, \ \mathbf{S}_n^{-1}(z) = (\mathbf{S}_n - z\mathbf{I}_p)^{-1}, \\ &\mathcal{D}_n^{-1}(z) = (\mathcal{D}_n - z\mathbf{I}_p)^{-1}, \ \mathbf{D}_n^{-1}(z) = (\mathbf{D}_n - z\mathbf{I}_p)^{-1}. \end{aligned} \tag{5.96}$$

Then we have

$$tr\mathbf{S}_n^{-1}(z) - pm_{c_n}(z) = \big(tr\mathcal{D}_n^{-1}(z) - pm_{c_n}(z)\big) - \frac{\bar{\mathbf{v}}^T\mathcal{D}_n^{-2}(z)\bar{\mathbf{v}}}{1+\bar{\mathbf{v}}^T\mathcal{D}_n^{-1}(z)\bar{\mathbf{v}}}, \tag{5.97}$$

where we have used the identity

$$(\mathbf{C}+\mathbf{r}\mathbf{r}^T)^{-1} = \mathbf{C}^{-1} - \frac{\mathbf{C}^{-1}\mathbf{r}\mathbf{r}^T\mathbf{C}^{-1}}{1+\mathbf{r}^T\mathbf{C}^{-1}\mathbf{r}}, \tag{5.98}$$

where $\mathbf{C}$ and $(\mathbf{C}+\mathbf{r}\mathbf{r}^T)$ are both invertible; and $\mathbf{r} \in \mathbb{R}^p$. The first term on the right hand of (5.97) was investigated in Pan (2012). In what follows we consider the second term on the right hand of (5.97).

One may verify that

$$(\mathbf{C}+q\mathbf{r}\mathbf{v}^T)^{-1} = \frac{\mathbf{C}^{-1}}{1+q\mathbf{v}^T\mathbf{C}^{-1}\mathbf{r}}, \tag{5.99}$$

where $\mathbf{C}$ and $(\mathbf{C}+q\mathbf{r}\mathbf{v}^T)$ are both invertible, q is a scalar and $\mathbf{r}, \mathbf{v} \in \mathbb{R}^p$. This, together with (5.84) and (5.98), yields

$$\begin{aligned} \bar{\mathbf{v}}^T\mathcal{D}_n^{-1}(z)\bar{\mathbf{v}} &= \bar{\varepsilon}^T\mathcal{D}_n^{-1}(z)\bar{\varepsilon} + 2\mathbf{u}^T\mathcal{D}_n^{-1}(z)\bar{\varepsilon} + \mathbf{u}^T\mathcal{D}_n^{-1}(z)\mathbf{u} \\ &= \frac{\bar{\varepsilon}^T\mathbf{D}_n^{-1}(z)\bar{\varepsilon}}{1-\bar{\varepsilon}^T\mathbf{D}_n^{-1}(z)\bar{\varepsilon}} + 2\frac{\mathbf{u}^T\mathbf{D}_n^{-1}(z)\bar{\varepsilon}}{1-\bar{\varepsilon}\mathbf{D}_n^{-1}(z)\bar{\varepsilon}} + \mathbf{u}^T\mathcal{D}_n^{-1}(z)\mathbf{u} \end{aligned} \tag{5.100}$$

and

$$\begin{aligned} &\bar{\mathbf{v}}^T\mathcal{D}_n^{-2}(z)\bar{\mathbf{v}} = \bar{\varepsilon}^T\mathcal{D}_n^{-2}(z)\bar{\varepsilon} + 2\mathbf{u}^T\mathcal{D}_n^{-2}(z)\bar{\varepsilon} + \mathbf{u}^T\mathcal{D}_n^{-2}(z)\mathbf{u} \\ = &\frac{\bar{\varepsilon}^T\mathbf{D}_n^{-2}(z)\bar{\varepsilon}}{(1-\bar{\varepsilon}\mathbf{D}_n^{-1}(z)\bar{\varepsilon})^2} + \frac{2\mathbf{u}^T\mathbf{D}_n^{-2}(z)\bar{\varepsilon}}{1-\bar{\varepsilon}^T\mathbf{D}_n^{-1}(z)\bar{\varepsilon}} + \frac{2\mathbf{u}^T\mathbf{D}_n^{-1}(z)\bar{\varepsilon}\bar{\varepsilon}^T\mathbf{D}_n^{-2}(z)\bar{\varepsilon}}{(1-\bar{\varepsilon}^T\mathbf{D}_n^{-1}(z)\bar{\varepsilon})^2} + \mathbf{u}^T\mathcal{D}^{-2}(z)\mathbf{u}. \end{aligned} \tag{5.101}$$

It is proved in Section 2.5 and (4.3) of Pan (2012) that as $n \to \infty$,

$$\sup_{z\in\gamma}\left|\bar{\varepsilon}^T\mathbf{D}_n^{-2}(z)\bar{\varepsilon} - \frac{c\underline{m}^2(z)}{(1+\underline{m}(z))^2 - c\underline{m}^2(z)}\right| \xrightarrow{i.p.} 0; \tag{5.102}$$

$$\sup_{z\in\gamma}\left|\bar{\boldsymbol{\varepsilon}}^T\mathbf{D}_n^{-1}(z)\bar{\boldsymbol{\varepsilon}}-(1+z\underline{m}(z))\right|\xrightarrow{i.p.}0; \tag{5.103}$$

and

$$\sup_{z\in\gamma}\left|\mathbf{u}^T\mathbf{D}_n^{-1}(z)\bar{\boldsymbol{\varepsilon}}\right|\xrightarrow{i.p.}0, \tag{5.104}$$

(where we also use an argument similar to (2.28) of Pan (2012)). By (3.4) and (4.3) in Pan (2012)), and (5.40), we have as $n\to\infty$,

$$\sup_{z\in\gamma}\left|\mathbf{u}^T\mathcal{D}_n^{-1}(z)\mathbf{u}-\bar{u}\underline{m}(z)\right|\xrightarrow{i.p.}0. \tag{5.105}$$

The next aim is to prove that

$$\sup_{z\in\gamma}\left|\mathbf{u}^T\mathcal{D}_n^{-2}(z)\mathbf{u}-\bar{u}\int\frac{1}{(\lambda-z)^2}dF^{MP}(\lambda)\right|\xrightarrow{i.p.}0 \tag{5.106}$$

and that

$$\sup_{z\in\gamma}\left|\mathbf{u}^T\mathbf{D}_n^{-2}(z)\bar{\boldsymbol{\varepsilon}}\right|\xrightarrow{i.p.}0. \tag{5.107}$$

Consider (5.106) first. By the formula (5.98), we have an expansion

$$\begin{aligned}\mathbf{u}^T\mathcal{D}_n^{-2}(z)\mathbf{u} &= \mathbf{u}^T\mathbf{D}_n^{-2}(z)\mathbf{u}+\frac{2\mathbf{u}^T\mathbf{D}_n^{-2}(z)\bar{\boldsymbol{\varepsilon}}\bar{\boldsymbol{\varepsilon}}^T\mathbf{D}_n^{-1}(z)\mathbf{u}}{1-\bar{\boldsymbol{\varepsilon}}^T\mathbf{D}_n^{-1}(z)\bar{\boldsymbol{\varepsilon}}}\\ &\quad+\frac{\mathbf{u}^T\mathbf{D}_n^{-1}(z)\bar{\boldsymbol{\varepsilon}}\bar{\boldsymbol{\varepsilon}}^T\mathbf{D}_n^{-2}(z)\bar{\boldsymbol{\varepsilon}}\bar{\boldsymbol{\varepsilon}}^T\mathbf{D}_n^{-1}(z)\mathbf{u}}{(1-\bar{\boldsymbol{\varepsilon}}^T\mathbf{D}_n^{-1}(z)\bar{\boldsymbol{\varepsilon}})^2}.\end{aligned}$$

For any given $z\in\gamma$, we conclude from Theorem 1 of Pan (2012) and Helly-Bray's theorem that

$$\mathbf{u}^T\mathcal{D}_n^{-2}(z)\mathbf{u}-\bar{u}\int\frac{1}{(\lambda-z)^2}dF^{MP}(\lambda)\xrightarrow{i.p.}0\quad\text{as}\quad n\to\infty. \tag{5.108}$$

By the expansion of $\mathbf{u}^T\mathcal{D}_n^{-2}(z)\mathbf{u}$ and (5.102)-(5.104), to prove (5.106), it suffices to prove the tightness of $\left\{K_n^{(1)}(z)=\mathbf{u}^T\mathbf{D}_n^{-2}(z)\mathbf{u}-\bar{u}\int\frac{1}{(\lambda-z)^2}dF^{MP}(\lambda), z\in\gamma\right\}$ and $\left\{\mathbf{u}^T\mathbf{D}_n^{-2}(z)\bar{\boldsymbol{\varepsilon}}, z\in\gamma\right\}$.

To this end, as in Bai and Silverstein (2004), below introduce the truncated version of $\mathbf{u}^T\mathbf{D}_n^{-2}(z)\mathbf{u}$. Define $\gamma_r=\{\mu_r+iv: v\in[n^{-1}\rho_n, v_0]\}$,

$$\gamma_\ell=\begin{cases}\{\mu_\ell+iv: v\in[n^{-1}\rho_n, v_0]\}, & \mu_\ell>0,\\ \{\mu_\ell+iv: v\in[0, v_0]\}, & \mu_\ell<0,\end{cases} \tag{5.109}$$

where

$$\rho_n \downarrow 0, \quad \rho_n \geq n^{-\theta}, \; for\ some\ \theta \in (0,1). \tag{5.110}$$

Let $\gamma_n^+ = \gamma_\ell \cup \gamma_\mu \cup \gamma_r$ and γ_n^- denote the symmetric part of γ_n^+ with respect to the real axis. We then define the truncated process $\mathbf{u}^T\widehat{\mathbf{D}_n^{-2}(z)}\mathbf{u}$ of the process $\mathbf{u}^T\mathbf{D}_n^{-2}(z)\mathbf{u}$ for $z = \alpha + iv$ by

$$\mathbf{u}^T\widehat{\mathbf{D}_n^{-2}(z)}\mathbf{u} = \begin{cases} \mathbf{u}^T\mathbf{D}_n^{-2}(z)\mathbf{u} & z \in \gamma_n = \gamma_n^+ \cup \gamma_n^-, \\ \frac{nv+\rho_n}{2\rho_n}\mathbf{u}^T\mathbf{D}_n^{-2}(z_{r_1})\mathbf{u} + \frac{\rho_n - nv}{2\rho_n}\mathbf{u}^T\mathbf{D}_n^{-2}(z_{r_2})\mathbf{u} & \mu = \mu_r, v \in I, \\ \frac{nv+\rho_n}{2\rho_n}\mathbf{u}^T\mathbf{D}_n^{-2}(z_{\ell_1})\mathbf{u} + \frac{\rho_n - nv}{2\rho_n}\mathbf{u}^T\mathbf{D}_n^{-2}(z_{\ell_2})\mathbf{u} & \mu = \mu_\ell > 0, v \in I, \end{cases} \tag{5.111}$$

where $z_{r_1} = \mu_r + in^{-1}\rho_n$, $z_{r_2} = \mu_r - in^{-1}\rho_n$, $z_{\ell_1} = \mu_\ell + in^{-1}\rho_n$, $z_{\ell_2} = \mu_\ell - in^{-1}\rho_n$ and $I = [-n^{-1}\rho_n, n^{-1}\rho_n]$. We then have

$$\sup_{z\in\gamma}\left|\mathbf{u}^T\widehat{\mathbf{D}_n^{-2}(z)}\mathbf{u} - \mathbf{u}^T\mathbf{D}_n^{-2}(z)\mathbf{u}\right| \leq K\rho_n\|\mathbf{u}\|^2\left(\frac{1}{|\lambda_{\max}(\mathbf{D}_n) - \mu_r|} + \frac{1}{|\lambda_{\min}(\mathbf{D}_n) - \mu_l|}\right) \xrightarrow{i.p.} 0. \tag{5.112}$$

It is proved in Section 3 of Bai and Silverstein (2004) that, for any positive integer k and $z \in \gamma_n^+ \cup \gamma_n^-$,

$$max\big(E||D_n^{-1}(z)||^k\big) \leq K. \tag{5.113}$$

It follows from independence between $\mathbf{u}$ and $\boldsymbol{\varepsilon}_j, j = 1, \cdots, n$ that

$$\begin{aligned} & E\left|\mathbf{u}^T\mathbf{D}_n^{-2}(z)\mathbf{u} - \bar{u}\int \frac{1}{(\lambda - z)^2}dF^{MP}(\lambda)\right| \\ \leq\ & E|\mathbf{u}^T\mathbf{D}_n^{-2}(z)\mathbf{u}| + \left|\bar{u}\int \frac{1}{(\lambda - z)^2}dF^{MP}(\lambda)\right| \\ \leq\ & E||\mathbf{u}^T||^2E||\mathbf{D}_n^{-2}(z)||^2 + K \leq K, \end{aligned} \tag{5.114}$$

which ensures Condition (1) of Lemma 17. Similarly, we can derive $E|\mathbf{u}^T\mathbf{D}_n^{-2}(z)\bar{\boldsymbol{\varepsilon}}|^2 \leq K$.

Next, we prove condition (2) of Lemma 17, i.e.

$$\sup_{n, z_1, z_2 \in \gamma_n^+ \cup \gamma_n^-} \frac{E|K_n^{(i)}(z_1) - K_n^{(i)}(z_2)|^2}{|z_1 - z_2|^2} < \infty, \quad i = 1, 2. \tag{5.115}$$

Note that

$$\mathbf{A}^{-1} - \mathbf{B}^{-1} = \mathbf{A}^{-1}(\mathbf{B} - \mathbf{A})\mathbf{B}^{-1}, \tag{5.116}$$

where $\mathbf{A}$ and $\mathbf{B}$ are any two nonsingular matrices. We then conclude that

$$\mathbf{D}_n^{-2}(z_1) - \mathbf{D}_n^{-2}(z_2) = (z_1 - z_2)\mathbf{D}_n^{-2}(z_1)\mathbf{D}_n^{-1}(z_2) + (z_1 - z_2)\mathbf{D}_n^{-1}(z_1)\mathbf{D}_n^{-2}(z_2). \tag{5.117}$$

Then

$$\begin{aligned} \frac{K_n^{(1)}(z_1) - K_n^{(1)}(z_2)}{z_1 - z_2} &= \mathbf{u}^T\mathbf{D}_n^{-2}(z_1)\mathbf{D}_n^{-1}(z_2)\mathbf{u} + \mathbf{u}^T\mathbf{D}_n^{-1}(z_1)\mathbf{D}_n^{-2}(z_2)\mathbf{u} \\ &\quad -\bar{u}\int \frac{(\lambda - z_1) + (\lambda - z_2)}{(\lambda - z_1)^2(\lambda - z_2)^2} dF^{MP}(\lambda). \end{aligned} \tag{5.118}$$

As in (5.114), we can obtain

$$E|\mathbf{u}^T\mathbf{D}_n^{-2}(z_1)\mathbf{D}_n^{-1}(z_2)\mathbf{u}|^2 \le K, \quad E|\mathbf{u}^T\mathbf{D}_n^{-1}(z_1)\mathbf{D}_n^{-2}(z_2)\mathbf{u}|^2 \le K. \tag{5.119}$$

Since $f(\lambda) = \frac{(\lambda - z_1) + (\lambda - z_2)}{(\lambda - z_1)^2(\lambda - z_2)^2}$ is a continuous function when $z_1, z_2 \in \gamma$, the integral $\int f(\lambda)dF^{MP}(\lambda)$ is bounded. This, together with (5.119), implies

$$\sup_{n, z_1, z_2 \in \gamma} \frac{E|K_n^{(1)}(z_1) - K_n^{(1)}(z_2)|^2}{|z_1 - z_2|^2} < \infty. \tag{5.120}$$

By (2.38) in Pan (2012) and an argument similar to (5.117)-(5.120) we may verify that $\mathbf{u}^T\mathbf{D}_n^{-2}(z)\bar{\boldsymbol{\varepsilon}}$ is tight for $z \in \gamma$. Summarizing the above we obtain (5.106).

Consider (5.107) now. From the last paragraph we see that it is enough to consider the pointwise convergence of $\mathbf{u}^T\mathbf{D}_n^{-2}(z)\bar{\boldsymbol{\varepsilon}}$. As in (5.111) and (5.112) we may define the truncated process $\mathbf{u}^T\widehat{\mathbf{D}_n^{-2}}(z)\bar{\boldsymbol{\varepsilon}}$ of the process $\mathbf{u}^T\mathbf{D}_n^{-2}(z)\bar{\boldsymbol{\varepsilon}}$ and then prove that their difference tends to zero in probability. As in (4.3) of Pan (2012) one may prove that for given $z \in \gamma_n^+$ $\mathbf{u}^T\mathbf{D}_n^{-2}(z)\bar{\boldsymbol{\varepsilon}} \xrightarrow{i.p.} 0$.

From (5.100) to (5.107) we have

$$\sup_{z \in \gamma}\Big|\bar{\mathbf{v}}^T\mathcal{D}_n^{-1}(z)\bar{\mathbf{v}} - \big(\frac{1 + z\underline{m}(z)}{-z\underline{m}(z)} + \bar{u}\underline{m}(z)\big)\Big| \xrightarrow{i.p.} 0, \tag{5.121}$$

and

$$\sup_{z\in\gamma}\Big|\bar{\mathbf{v}}^T\mathcal{D}_n^{-2}(z)\bar{\mathbf{v}}-\big(\frac{c}{z^2[(1+\underline{m}(z))^2-c\underline{m}^2(z)]}+\bar{u}\int\frac{1}{(\lambda-z)^2}dF^{MP}(\lambda)\big)\Big|\xrightarrow{i.p.}0. \tag{5.122}$$

We then conclude from Slutsky's theorem that

$$\sup_{z\in\gamma}\Big|\frac{\bar{\mathbf{v}}^T\mathcal{D}_n^{-2}(z)\bar{\mathbf{v}}}{1+\bar{\mathbf{v}}^T\mathcal{D}_n^{-1}(z)\bar{\mathbf{v}}}-\frac{\frac{c\underline{m}(z)}{z[(1+\underline{m}(z))^2-c\underline{m}^2(z)]}+\bar{u}\int\frac{1}{(\lambda-z)^2}dF^{MP}(\lambda)}{\bar{u}z\underline{m}^2(z)-1}\Big|\xrightarrow{i.p.}0. \tag{5.123}$$

The arguments of Theorem 1 of Pan (2012) show that the truncation version of $\big(tr\mathcal{D}_n^{-1}(z)-pm_{c_n}(z)\big)$ converges in distribution to a two-dimensional Gaussian process and that the difference between $\big(tr\mathcal{D}_n^{-1}(z)-pm_{c_n}(z)\big)$ and its truncation version goes to zero in probability (see Page 563 of Bai and Silverstein (2004) and (2.28) of Pan (2012)). Theorem 3 then follows from (5.123), (5.94) and (5.97), Slutsky's theorem and Lemma 3 (one may refer to Page 563 of Bai and Silverstein (2004)). $\square$

Table 5.1: *Empirical sizes and power values of the proposed test at the 5% significant level for standardized normally distributed random vectors.*

	p										
n	5	10	20	30	40	50	60	70	80	90	100
					Empirical sizes						
5	0.069	0.064	0.027	0.039	0.034	0.052	0.035	0.033	0.037	0.043	0.047
10	0.049	0.059	0.048	0.046	0.055	0.044	0.043	0.048	0.048	0.038	0.046
20	0.045	0.042	0.047	0.051	0.052	0.052	0.051	0.041	0.042	0.046	0.041
30	0.034	0.035	0.052	0.053	0.068	0.063	0.050	0.049	0.047	0.040	0.046
40	0.055	0.050	0.052	0.061	0.056	0.038	0.042	0.064	0.054	0.044	0.066
50	0.038	0.041	0.048	0.044	0.050	0.052	0.050	0.058	0.046	0.053	0.041
60	0.064	0.061	0.043	0.049	0.047	0.049	0.050	0.049	0.059	0.046	0.048
70	0.047	0.042	0.051	0.039	0.052	0.055	0.058	0.044	0.037	0.038	0.049
80	0.028	0.032	0.033	0.041	0.052	0.054	0.047	0.042	0.047	0.046	0.040
90	0.042	0.037	0.032	0.048	0.054	0.044	0.053	0.041	0.037	0.051	0.042
100	0.037	0.046	0.036	0.035	0.045	0.029	0.045	0.056	0.048	0.047	0.056
					Empirical power values						
5	0.044	0.053	0.072	0.070	0.065	0.048	0.053	0.063	0.057	0.053	0.080
10	0.054	0.065	0.046	0.051	0.052	0.049	0.049	0.037	0.049	0.060	0.062
20	0.056	0.066	0.091	0.059	0.062	0.051	0.045	0.075	0.078	0.078	0.063
30	0.051	0.062	0.054	0.068	0.051	0.081	0.092	0.087	0.115	0.128	0.148
40	0.041	0.036	0.052	0.055	0.126	0.183	0.144	0.208	0.209	0.276	0.254
50	0.068	0.051	0.054	0.066	0.151	0.262	0.329	0.420	0.474	0.432	0.574
60	0.046	0.027	0.055	0.098	0.154	0.310	0.455	0.592	0.596	0.733	0.715
70	0.050	0.044	0.018	0.110	0.211	0.379	0.565	0.747	0.793	0.846	0.841
80	0.053	0.046	0.036	0.102	0.233	0.439	0.687	0.892	0.906	0.944	0.977
90	0.028	0.040	0.054	0.122	0.226	0.484	0.752	0.878	0.963	0.980	0.999

*The power values are under the alternative hypothesis that the population covariance matrix is $\Sigma = 0.95I_p + 0.05\mathbf{1}_p\mathbf{1}_p^{'}$.

Table 5.2: *Bootstrap sizes and power values of the proposed test at the 5% significant level for standardized normal and gamma(4,2) random vectors respectively.*

						p					
n	5	10	20	30	40	50	60	70	80	90	100
					Normal Data						
5	0.044	0.045	0.047	0.045	0.040	0.045	0.046	0.045	0.043	0.042	0.040
10	0.042	0.047	0.047	0.044	0.042	0.043	0.042	0.042	0.040	0.043	0.045
20	0.040	0.045	0.045	0.047	0.046	0.046	0.047	0.045	0.045	0.042	0.043
30	0.042	0.045	0.049	0.053	0.052	0.052	0.051	0.048	0.047	0.046	0.048
40	0.040	0.044	0.050	0.052	0.056	0.048	0.042	0.052	0.054	0.044	0.052
50	0.039	0.040	0.049	0.044	0.050	0.052	0.050	0.058	0.046	0.053	0.048
60	0.037	0.041	0.047	0.049	0.047	0.049	0.050	0.049	0.059	0.046	0.049
70	0.032	0.040	0.048	0.049	0.052	0.055	0.053	0.046	0.048	0.047	0.049
80	0.031	0.039	0.043	0.046	0.052	0.054	0.047	0.042	0.047	0.046	0.047
90	0.030	0.040	0.041	0.048	0.051	0.044	0.053	0.041	0.047	0.051	0.048
100	0.033	0.038	0.040	0.042	0.044	0.043	0.046	0.048	0.051	0.049	0.051
					Gamma Data						
5	0.065	0.064	0.060	0.065	0.063	0.065	0.069	0.066	0.073	0.072	0.070
10	0.068	0.062	0.060	0.063	0.061	0.058	0.062	0.058	0.060	0.063	0.065
20	0.069	0.061	0.059	0.058	0.062	0.056	0.057	0.055	0.058	0.057	0.060
30	0.062	0.059	0.062	0.057	0.058	0.055	0.054	0.053	0.055	0.056	0.058
40	0.063	0.065	0.061	0.056	0.057	0.055	0.055	0.053	0.056	0.054	0.054
50	0.064	0.063	0.063	0.059	0.058	0.054	0.054	0.054	0.056	0.053	0.052
60	0.069	0.071	0.064	0.060	0.054	0.052	0.049	0.050	0.055	0.052	0.053
70	0.073	0.072	0.067	0.063	0.055	0.050	0.053	0.051	0.048	0.050	0.050
80	0.074	0.074	0.067	0.064	0.058	0.052	0.049	0.052	0.049	0.049	0.050
90	0.072	0.070	0.068	0.067	0.058	0.052	0.050	0.051	0.049	0.051	0.049
100	0.076	0.072	0.071	0.067	0.059	0.053	0.052	0.049	0.051	0.049	0.050

Table 5.3: *Empirical sizes and power values of the LRT at the* 5% *significant level for standardized normally distributed random vectors.*

	p										
n	5	10	20	30	40	50	60	70	80	90	100
						Empirical sizes					
5	0.005	0.475	0.495	0	0	0	0	0	0	0	0
10	0.030	0.115	0.504	0.454	0	0	0	0	0	0	0
20	0.047	0.030	0.522	1	0.500	0	0	0	0	0	0
30	0.040	0.032	0.076	0.873	1	1	0	0	0	0	0
40	0.044	0.034	0.057	0.148	0.982	1	1	1	0	0	0
50	0.057	0.034	0.044	0.062	0.299	1	1	1	1	1	0
60	0.049	0.053	0.040	0.065	0.101	0.578	1	1	1	1	1
70	0.038	0.060	0.049	0.054	0.108	0.229	0.784	1	1	1	1
80	0.055	0.040	0.043	0.072	0.079	0.147	0.442	0.950	1	1	1
90	0.054	0.061	0.047	0.033	0.048	0.129	0.233	0.691	0.991	1	1
100	0.055	0.048	0.066	0.069	0.076	0.085	0.185	0.402	0.844	0.997	1
						Empirical power values					
5	0.005	0.504	0.505	0	0	0	0	0	0	0	0
10	0.033	0.125	0.506	0.485	0	0	0	0	0	0	0
20	0.051	0.049	0.587	1	0.496	0	0	0	0	0	0
30	0.056	0.062	0.121	0.912	1	1	0	0	0	0	0
40	0.061	0.077	0.139	0.311	0.990	1	1	1	0	0	0
50	0.092	0.090	0.150	0.206	0.576	1	1	1	1	1	0
60	0.088	0.142	0.220	0.266	0.420	0.849	1	1	1	1	1
70	0.102	0.170	0.222	0.320	0.463	0.691	0.984	1	1	1	1
80	0.133	0.169	0.246	0.392	0.478	0.666	0.883	0.994	1	1	1
90	0.122	0.190	0.331	0.404	0.526	0.666	0.819	0.983	1	1	1
100	0.153	0.232	0.411	0.550	0.658	0.713	0.842	0.952	0.995	1	1

*The power values are under the alternative hypothesis that the population covariance matrix is $\Sigma = 0.95 I_p + 0.05 \mathbf{1}_p \mathbf{1}_p^{'}$.

Table 5.4: *Empirical sizes of the proposed test at the 5% significance level for standardized gamma random vectors.*

					p						
n	5	10	20	30	40	50	60	70	80	90	100
					Empirical sizes						
5	0.089	0.078	0.068	0.059	0.060	0.063	0.061	0.047	0.050	0.061	0.044
10	0.066	0.075	0.068	0.054	0.046	0.051	0.047	0.048	0.050	0.042	0.052
20	0.057	0.058	0.072	0.062	0.053	0.054	0.053	0.052	0.049	0.051	0.048
30	0.069	0.069	0.069	0.076	0.058	0.056	0.054	0.047	0.067	0.066	0.054
40	0.061	0.049	0.058	0.040	0.065	0.048	0.063	0.065	0.068	0.047	0.067
50	0.053	0.054	0.055	0.057	0.059	0.048	0.067	0.066	0.043	0.059	0.053
60	0.059	0.052	0.057	0.060	0.052	0.067	0.058	0.064	0.064	0.061	0.069
70	0.044	0.050	0.064	0.055	0.071	0.054	0.067	0.064	0.051	0.077	0.048
80	0.045	0.050	0.061	0.043	0.071	0.055	0.071	0.053	0.056	0.070	0.060
90	0.041	0.067	0.034	0.056	0.049	0.054	0.050	0.060	0.047	0.060	0.058
100	0.070	0.045	0.059	0.055	0.047	0.062	0.069	0.057	0.056	0.060	0.061

Table 5.5: *Empirical power values of the proposed test at the 5% significance level for standardized gamma random vectors.*

				p			
n	5	10	20	30	40	50	60
				Empirical powers			
5	0.334	0.575	0.853	0.944	0.983	0.989	0.998
10	0.513	0.838	0.979	0.999	1	1	1
20	0.721	0.970	0.999	1	1	1	1
30	0.834	0.998	1	1	1	1	1
40	0.914	1	1	1	1	1	1
50	0.952	1	1	1	1	1	1
60	0.991	1	1	1	1	1	1

*The power values are under the alternative hypothesis that the population covariance matrix is $\Sigma = 0.95 I_p + 0.05 \mathbf{1}_p \mathbf{1}_p^{'}$.

Table 5.6: *Empirical power values of the proposed test at the 5% significance level for MA(1) model.*

	p						
n	5	10	20	30	40	50	60
5	0.096	0.097	0.097	0.111	0.210	0.214	0.198
10	0.082	0.090	0.173	0.227	0.374	0.715	0.722
20	0.099	0.165	0.400	0.597	0.683	0.822	0.951
30	0.067	0.121	0.611	0.733	0.803	0.986	1
40	0.091	0.321	0.653	0.938	0.968	1	1
50	0.139	0.416	0.910	0.956	0.998	1	1
60	0.117	0.412	0.918	0.994	1	1	1

Table 5.7: *Empirical power values of the proposed test at the 5% significance level for AR(1) model.*

	p					
n	5	10	20	30	40	50
5	0.130	0.094	0.264	0.253	0.181	0.169
10	0.167	0.289	0.482	0.640	0.668	0.609
20	0.230	0.544	0.878	0.271	0.954	0.994
30	0.205	0.602	0.993	0.999	0.671	0.936
40	0.344	0.916	0.998	1	1	0.982
50	0.541	0.984	1	1	1	1

Table 5.8: *Empirical power values of the proposed test at the 5% significance level for SMA(1) model.*

	p				
n	5	10	20	30	40
5	0.330	0.431	0.782	0.999	1
10	0.647	1	1	1	1
20	0.967	1	1	1	1
30	0.962	1	1	1	1
40	0.998	1	1	1	1

Table 5.9: *Empirical sizes and power values of the proposed test at the 5% significance level for the general panel data model.*

	p										
n	5	10	20	30	40	50	60	70	80	90	100
Empirical sizes											
5	0.038	0.045	0.047	0.057	0.058	0.067	0.069	0.069	0.072	0.076	0.074
10	0.041	0.042	0.046	0.049	0.056	0.050	0.065	0.047	0.068	0.063	0.069
20	0.035	0.042	0.049	0.056	0.052	0.046	0.063	0.043	0.069	0.055	0.057
30	0.043	0.049	0.043	0.059	0.048	0.068	0.059	0.057	0.040	0.055	0.047
40	0.048	0.052	0.043	0.060	0.057	0.046	0.049	0.054	0.046	0.058	0.061
50	0.057	0.048	0.046	0.058	0.052	0.055	0.048	0.049	0.050	0.040	0.041
60	0.058	0.056	0.055	0.048	0.047	0.045	0.053	0.066	0.058	0.049	0.050
70	0.062	0.060	0.059	0.056	0.049	0.057	0.049	0.068	0.052	0.036	0.043
80	0.071	0.063	0.067	0.047	0.048	0.058	0.056	0.044	0.059	0.057	0.055
90	0.065	0.068	0.065	0.048	0.053	0.048	0.056	0.048	0.048	0.066	0.060
100	0.037	0.046	0.036	0.035	0.045	0.043	0.045	0.056	0.048	0.047	0.055
Empirical power values											
5	0.150	0.238	0.345	0.417	0.484	0.529	0.549	0.615	0.611	0.668	0.692
10	0.125	0.247	0.452	0.526	0.568	0.633	0.669	0.737	0.765	0.751	0.800
20	0.206	0.343	0.493	0.615	0.673	0.752	0.788	0.813	0.860	0.864	0.876
30	0.111	0.404	0.535	0.684	0.757	0.756	0.855	0.875	0.882	0.909	0.953
40	0.308	0.393	0.605	0.698	0.786	0.820	0.878	0.898	0.944	0.959	0.953
50	0.207	0.450	0.603	0.718	0.815	0.889	0.923	0.938	0.966	0.973	0.980
60	0.268	0.430	0.594	0.780	0.826	0.918	0.913	0.926	0.974	0.976	0.984
70	0.144	0.434	0.649	0.798	0.888	0.883	0.944	0.968	0.971	0.982	0.996
80	0.171	0.454	0.678	0.796	0.872	0.921	0.938	0.967	0.989	0.992	0.995
90	0.204	0.431	0.683	0.834	0.874	0.916	0.963	0.985	0.985	0.994	0.994
100	0.291	0.398	0.687	0.836	0.884	0.931	0.973	0.987	0.992	0.994	1

Table 5.10: *Empirical power values of the proposed test at the 5% significance level for nonlinear MA model.*

	p									
n	5	20	30	40	50	60	70	80	90	100
5	0.033	0.004	0.008	0.005	0.007	0.008	0.007	0.009	0.014	0.070
20	0.804	0.703	0.614	0.581	0.511	0.447	0.340	0.306	0.257	0.216
30	0.854	0.777	0.779	0.780	0.740	0.662	0.662	0.597	0.579	0.555
40	0.878	0.856	0.856	0.845	0.825	0.779	0.770	0.772	0.702	0.698
50	0.884	0.868	0.884	0.864	0.888	0.860	0.875	0.869	0.828	0.820
60	0.892	0.882	0.904	0.920	0.923	0.933	0.900	0.892	0.892	0.882
70	0.896	0.906	0.927	0.934	0.921	0.952	0.925	0.943	0.917	0.926
80	0.936	0.925	0.921	0.952	0.950	0.943	0.943	0.958	0.953	0.936
90	0.922	0.939	0.935	0.958	0.959	0.955	0.982	0.962	0.957	0.954
100	0.926	0.920	0.937	0.952	0.964	0.970	0.975	0.978	0.970	0.965

Table 5.11: *Empirical power values of the proposed test at the 5% significance level for ARCH(1) model.*

	p										
n	5	10	20	30	40	50	60	70	80	90	100
5	0.148	0.093	0.073	0.076	0.058	0.099	0.113	0.088	0.106	0.112	0.141
10	0.454	0.399	0.315	0.215	0.128	0.121	0.112	0.121	0.093	0.095	0.092
20	0.401	0.428	0.508	0.464	0.444	0.468	0.436	0.338	0.356	0.337	0.302
30	0.272	0.364	0.616	0.712	0.753	0.759	0.793	0.743	0.638	0.639	0.598
40	0.232	0.357	0.572	0.823	0.790	0.874	0.915	0.885	0.876	0.855	0.860
50	0.222	0.339	0.622	0.757	0.891	0.969	0.957	0.975	0.967	0.957	0.975
60	0.216	0.448	0.617	0.862	0.901	0.976	0.983	0.987	0.992	0.997	0.997
70	0.200	0.339	0.592	0.864	0.931	0.982	0.993	0.998	0.997	0.998	0.996
80	0.194	0.376	0.566	0.824	0.950	0.967	0.997	1	1	0.999	1
90	0.184	0.456	0.721	0.839	0.960	0.995	0.999	1	0.999	1	1
100	0.128	0.306	0.802	0.859	0.934	0.992	1	1	1	1	1

Table 5.12: *Empirical power values of the proposed test at the 5% significance level for Vandermonde Matrix.*

	p										
n	10	20	30	40	50	60	70	80	90	100	120
10	0.180	0.197	0.192	0.169	0.202	0.211	0.190	0.185	0.171	0.177	0.240
20	0.309	0.332	0.356	0.327	0.291	0.303	0.301	0.295	0.321	0.243	0.478
30	0.324	0.433	0.473	0.413	0.461	0.408	0.445	0.395	0.368	0.397	0.606
40	0.458	0.512	0.527	0.546	0.533	0.490	0.518	0.498	0.450	0.457	0.655
50	0.593	0.437	0.540	0.571	0.614	0.569	0.577	0.566	0.565	0.537	0.764
60	0.504	0.538	0.551	0.567	0.616	0.662	0.588	0.581	0.572	0.607	0.744
70	0.548	0.526	0.560	0.627	0.668	0.641	0.694	0.707	0.641	0.678	0.741
80	0.550	0.545	0.580	0.633	0.712	0.719	0.693	0.768	0.729	0.749	0.805
90	0.589	0.544	0.596	0.667	0.695	0.712	0.743	0.754	0.738	0.728	0.807
100	0.464	0.549	0.610	0.645	0.704	0.757	0.772	0.751	0.752	0.808	0.928
120	0.633	0.660	0.736	0.737	0.759	0.855	0.854	0.909	0.960	0.999	1

Chapter 6

Independence Test For Covariance Stationary Time Series

6.1 Preliminary

The observed n random vectors $\mathbf{x}_i = (X_{1i}, X_{2i}, \ldots, X_{pi})'$ with $i = 1, 2, \ldots, n$ are grouped into a matrix $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$. Denote the sample covariance matrix by

$$\mathbf{S} = \frac{1}{n}\mathbf{X}\mathbf{X}^T. \tag{6.1}$$

The goal is to do the following independence hypothesis test

$$\mathbb{H}_0 : \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n \textit{ are independent}; \textit{ against } \mathbb{H}_1 : \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n \textit{ are dependent}.$$

Throughout the chapter, we consider two types of high dimensional random vectors $\mathbf{x}_i$. The first type $\mathbf{x}_i$ is stationary time series specified as follows.

Assumption 4. *The n time series can be expressed as*

$$X_{jt} = \sum_{k=0}^{\infty} b_k \xi_{j-k,t}, \ j = 1, \ldots, p; \ t = 1, \ldots, n, \tag{6.2}$$

where for any $t = 1, 2, \dots, n$, $\{\xi_{i,t}\}_{i=-\infty}^{\infty}$ *is an independent and identically distributed (i.i.d) sequence with mean zero and variance one;* $\{b_k\}_{k=0}^{\infty}$ *is a sequence of real numbers satisfying* $\sum_{k=0}^{\infty} |b_k| < \infty$.

This assumption covers many classical covariance stationary time series, for example, the autoregressive (AR), moving average (MA), and autoregressive and moving average(ARMA) time series of finite orders, etc.. In addition to ensuring stationary, the condition $\sum_{k=0}^{\infty} |b_k| < \infty$ is imposed to also guarantee that the spectral norm of the population covariance matrix $\mathbf{T}_1$ of each time series under investigation is bounded, as will be seen from the proof.

The second type $\mathbf{x}_i$ is linearly generated by $\mathbf{y}_i$ whose components are independent, as defined below.

Assumption 5. *Let* $\mathbf{x}_i = \mathbf{T}_1^{1/2}\mathbf{y}_i$ *with* $\mathbf{y}_i = (Y_{1i}, \cdots, Y_{pi})^T$ *and* $\mathbf{T}_1^{1/2}$ *being a Hermitian square root of the nonrandom nonnegative definite Hermitian matrix* $\mathbf{T}_1$. *For each* $i = 1, \dots, n$, $Y_{1i}, \cdots, Y_{pi}$ *are i.i.d with mean zero and variance one.*

Assumption 6. *Let* p *be some function of* n. *Assume that* n *and* p *tend to infinity in the same order, i.e.*

$$c := \lim_{n\to\infty} \frac{p}{n} \in (0, +\infty).$$

When $\{\xi_{i,t}\}$ are normally distributed, Assumption 4 is a special case of Assumption 5. Indeed, it is clear that each X_{jt} is Gaussian distributed and each $\mathbf{x}_i$ is multivariate Gaussian distribution, whose covariance matrix is a Toeplitz matrix, if $\{\xi_{i,t}\}$ are normally distributed. Then $\mathbf{x}_i$ in Assumption 4 can be written as a form of $\mathbf{T}_1^{1/2}\mathbf{y}_i$ as well. Here, to save notation, we still use $\mathbf{T}_1$ as a covariance matrix of $\mathbf{x}_i$ although it is a Toeplitz matrix.

Therefore in this case the sample covariance matrices $\mathbf{S}$ associated with Assumptions 4 and 5 have a unified expression

$$\frac{1}{n}\mathbf{T}_1^{1/2}\mathbf{Y}\mathbf{Y}^T\mathbf{T}_1^{1/2}, \tag{6.3}$$

where $\mathbf{Y} = (\mathbf{y}_1, \cdots, \mathbf{y}_n)$.

Denote the sample covariance matrix in the form of (6.3) by $\mathbf{S}_1$. We are now interested in its limiting spectral distribution (LSD) which is the limit of the empirical spectral distribution(ESD) $F^{\mathbf{S}_1}(x)$. Here for any $\mathbf{A}$ of size $p \times p$ with real eigenvalues, its ESD is defined by

$$F^{\mathbf{A}}(x) = \frac{1}{p}\sum_{j=1}^{p} I(\mu_j \le x),$$

where $\mu_1 \le \mu_2 \le \cdots \le \mu_p$ are eigenvalues of the matrix $\mathbf{A}$. A common way to find the LSD is to first establish an equation of its Stieltjes transform, which is defined as, for any cumulative distribution function (CDF) $G(x)$,

$$m_G(z) = \int \frac{1}{\lambda - z} dG(\lambda), \ Im(z) \neq 0.$$

It can be then recovered by the Frobenius-Perron formula inversion formula

$$G\{[a,b]\} = \frac{1}{\pi}\lim_{\eta\to 0^+}\int_a^b Im\Big(m_G(\zeta + i\eta)\Big)d\zeta, \tag{6.4}$$

where a, b are points of continuity of $G(x)$.

Silverstein's result (1995) shows that the LSD of $\mathbf{S}_1$ in (6.3) is $F_{c,H}(x)$ whose Stieltjes transform is the unique solution to

$$m(z) = \int \frac{1}{\tau\Big(1 - c - czm(z)\Big) - z} dH(\tau), \tag{6.5}$$

in the set $\{m \in \mathbb{C} : -\frac{1-c}{z} + cm \in \mathbb{C}^+\}$ if $F^{\mathbf{T}_1} \to H(\tau)$. This also yields the LSD of the sample covariance matrix $\mathbf{S}$ for linear stationary processes with the Gaussian entries because the condition that $F^{\mathbf{T}_1} \to H(\tau)$ holds automatically in the case of linear stationary time series. An alternative expression of (6.5) for stationary time series will be given in the next section by using its spectral density.

To propose a statistic to test the hypothesis $\mathbb{H}_0$ based on the feature of $F_{c,H}(x)$, we consider an alternative that the sample covariance matrix $\mathbf{S}$ takes the form of

$$\frac{1}{n}\mathbf{T}_1^{1/2}\mathbf{Y}\mathbf{T}_2\mathbf{Y}^T\mathbf{T}_1^{1/2}, \tag{6.6}$$

where $\mathbf{T}_2$ is an $n \times n$ deterministic Hermitian matrix. Hence the dependence of the n time series is described by the matrix $\mathbf{T}_2$.

Denote the sample covariance matrix in the form of (6.6) by $\mathbf{S}_2$. Zhang (2006) provides the LSD of the matrix $\mathbf{S}_2$ different from (6.5). For easy reference, we state this result in the following lemma.

Lemma 22. *In addition to Assumptions 5 and 6, we assume that as $n \to \infty$, the ESDs of $\mathbf{T}_1$ and $\mathbf{T}_2$, denoted by $F^{\mathbf{T}_1}$ and $F^{\mathbf{T}_2}$ respectively, converge weakly to two probability functions, H_1 and H_2, respectively. Then the ESD of the matrix $\mathbf{S}_2$ converges weakly to a non-random CDF F_{c,H_1,H_2} with probability one, for which if $H_1 \equiv 1_{[0,+\infty)}$ or $H_2 \equiv 1_{[0,+\infty)}$, then $F_{c,H_1,H_2} \equiv 1_{[0,+\infty)}$; otherwise if for each $z \in \mathbb{C}^+$,*

$$\begin{cases} s(z) = -z^{-1}(1-c) - z^{-1}c\int \frac{1}{1+q(z)x}dH_2(x) \\ s(z) = -z^{-1}\int \frac{1}{1+p(z)y}dH_1(y) \\ s(z) = -z^{-1} - p(z)q(z) \end{cases} \tag{6.7}$$

is viewed as a system of equations for the complex vector $(s(z), p(z), q(z))$, then the Stieltjes transform of F_{c,H_1,H_2}, denoted by $m_{F_{c,H_1,H_2}}(z)$, together with two other functions, denoted by $g_1(z)$ and $g_2(z)$, both of which are *analytic on* $\mathbb{C}^+$*, will satisfy that* $\big(m_{F_{c,H_1,H_2}}(z), g_1(z), g_2(z)\big)$ *is the unique solution to (6.7) in the set*

$$\Big\{\big(s(z), p(z), q(z)\big) : Im\big(s(z)\big) > 0, Im\big(zp(z)\big) > 0, Im\big(q(z)\big) > 0\Big\}.$$

From (6.5) and (6.7), we see that the LSD of the matrix $\mathbf{S}_1$ is different from that of $\mathbf{S}_2$ since the latter one depends on the spectral distribution of the matrix $\mathbf{T}_2$ which is an identity matrix under the null hypothesis $\mathbb{H}_0$. Based on the observation, a natural idea is to utilize the difference between the LSDs of $\mathbf{S}$ under $\mathbb{H}_0$ and $\mathbb{H}_1$ to distinguish independence from dependence.

To this end let

$$G_n(\lambda) = p\Big(F^{\mathbf{S}}(\lambda) - F_{c_n,H_n}(\lambda)\Big) \tag{6.8}$$

and consider the linear spectral statistic of $\mathbf{S}$:

$$M_n = \int f(\lambda) dG_n(\lambda), \tag{6.9}$$

where $F_{c_n,H_n}(\lambda)$ is obtained from the LSD $F_{c,H}(\lambda)$ of $\mathbf{S}$ under $\mathbf{H}_0$ and Assumptions 4 or 5 with c and H replaced by $c_n = p/n$ and H_n respectively; $H_n = F^{\mathbf{T}_1}$ and $f(\lambda)$ is a smooth function. Roughly speaking, the difference between the LSDs of $\mathbf{S}$ under $\mathbb{H}_0$ and $\mathbb{H}_1$ is reflected in behaviour of M_n. Indeed, if we rewrite the statistic M_n as

$$p\Big[\int f(\lambda)d\Big(F^{\mathbf{S}}(\lambda) - F_{c_n,H_n,\mathbb{H}_1}(\lambda)\Big)\Big] + p\Big[\int f(\lambda)d\Big(F_{c_n,H_n,\mathbb{H}_1}(\lambda) - F_{c_n,H_n}(\lambda)\Big)\Big], \tag{6.10}$$

where $F_{c_n,H_n,\mathbb{H}_1}(\lambda)$ denotes the LSD of $\mathbf{S}$ under the alternative hypothesis $\mathbb{H}_1$, then we see that the last term of (6.10) captures the difference between the LSDs of $\mathbf{S}$ under $\mathbb{H}_0$ and $\mathbb{H}_1$, not to mention the first term of (6.10). One typical example of $F_{c_n,H_n,\mathbb{H}_1}(\lambda)$ could be F_{c,H_1,H_2} in Lemma 22.

Central limit theorems (CLT) of M_n corresponding to Assumptions 4 and 5 will be given in the next section. Based on it we then propose our test statistic.

6.2 Main theorems and the test statistic

6.2.1 Covariance stationary time series

The aim of this subsection is to establish the LSD of $\mathbf{S}$ and CLT of the linear spectral statistic M_n under the null hypothesis $\mathbb{H}_0$ and Assumption 4. Below we first present the LSD of $\mathbf{S}$.

Theorem 10. *Under Assumptions 4 and 6 and the null hypothesis $\mathbb{H}_0$, with probability one, the ESD $F^{\mathbf{S}}(x)$ converges to a nonrandom distribution function $F_{c,\phi}(x)$ whose Stieltjes transform $m_\phi(z)$ satisfies*

$$z = -\frac{1}{m_\phi(z)} + \frac{1}{2\pi}\int_0^{2\pi} \frac{1}{cm_\phi(z) + \big(\phi(\lambda)\big)^{-1}} d\lambda, \tag{6.11}$$

where $\phi(\lambda)$ denotes the spectral density of $\mathbf{x}_t$

$$\phi(\lambda) = \sum_{k=-\infty}^{\infty} \phi_k e^{ik\lambda}, \ \lambda \in [0, 2\pi),$$

with $\phi_k = Cov(X_{jt}, X_{j+k,t})$.

Remark 14. *This weakens the finite fourth moment condition imposed in Yao (2012). In addition we would point out that (6.11) is just an alternative expression of (6.5) in terms of the spectral density of $\mathbf{x}_i$. Therefore we use $F_{c,\phi}(x)$ to denote $F_{c,H}(x)$ in the case of stationary time series.*

From (6.11), we see that the Stieltjes transform $m_\phi(z)$ does not have an explicit expression. In practice, we can adopt a numerical method to calculate it which is provided in Yao (2012). For easy reference, we state it below:

Algorithm of calculating $m_\phi(z)$: Choose an initial value $m_\phi^{(0)}(z) = u+i\varepsilon$, where $z = x+i\varepsilon$ with x a given value and ε a small enough number. Iterate the following mapping below for $k \geq 0$:

$$\frac{1}{m_\phi(z)} = -z + A(m_\phi(z)), \tag{6.12}$$

where

$$A(m_\phi(z)) = \frac{1}{2\pi}\int_0^{2\pi} \frac{1}{cm_\phi(z) + \phi^{-1}(\lambda)} d\lambda,$$

until convergence. Let $m_\phi^{(K)}(z)$ be the final value.

We next develop CLT of M_n, which, we believe, is new in the literature. Recall the definition of $G_n(\lambda)$ in (6.8).

Theorem 11. *In addition to Assumptions 4 and 6, we suppose that* $E\xi_{j-k,t}^4 = 3$. *Let* $f_1, f_2, \ldots, f_h$ *be functions analytic on an open region containing the support of* F_{c_n,H_n}. *Then the random vector*

$$\Big(\int f_1(\lambda)dG_n(\lambda), \int f_2(\lambda)dG_n(\lambda), \ldots, \int f_h(\lambda)dG_n(\lambda)\Big) \tag{6.13}$$

converges in distribution to a Gaussian random vector $\Big(X_{f_1}, X_{f_2}, \ldots, X_{f_h}\Big)$ *with mean function for* $\ell = 1, 2, \cdots, h$,

$$EX_{f_\ell} = -\frac{1}{2\pi i}\oint_{\mathcal{C}} f_\ell(z) \frac{\frac{1}{2\pi}\int_0^{2\pi} c\underline{m}_\phi^3(z)\phi^2(\lambda)\big(1+\phi(\lambda)\underline{m}_\phi(z)\big)^{-3} d\lambda}{\Big(1 - c\frac{1}{2\pi}\int_0^{2\pi} \underline{m}_\phi^2(z)\phi^2(\lambda)\big(1+\phi(\lambda)\underline{m}_\phi(z)\big)^{-2} d\lambda\Big)^2} dz$$

and covariance element for $\ell, r = 1, 2, \cdots, h$,

$$Cov(X_{f_\ell}, X_{f_r}) = -\frac{1}{2\pi^2}\oint_{\mathcal{C}_1}\oint_{\mathcal{C}_2} \frac{f_\ell(z_1)f_r(z_2)}{\big(\underline{m}_\phi(z_1) - \underline{m}_\phi(z_2)\big)^2} \frac{d\underline{m}_\phi(z_1)}{dz_1}\frac{d\underline{m}_\phi(z_2)}{dz_2} dz_1 dz_2. \tag{6.14}$$

The contours $\mathcal{C}$ above are closed and are taken in the positive direction in the complex plane, each enclosing the support of $F_{c,\phi}(\lambda)$ and $\underline{m}_{\phi}(z)$ is the Stieltjes transform of the LSD of the matrix $\underline{\mathbf{S}} = \frac{1}{n}\mathbf{X}^T\mathbf{X}$.

Here $\underline{m}_{\phi}(z)$ can be obtained from $m_{\phi}(z)$ of (6.11) because the spectra of $\underline{\mathbf{S}}$ differs from that of $\mathbf{S}$ by $|n-p|$ zeros.

6.2.2 Linear independent structures

This subsection is to consider $\mathbf{x}_i$ satisfying Assumption 5.

The CLT of the linear spectral statistic M_n defined in (6.9) has been reported in Theorem 9.10 of Bai and Silverstein (2009). For easy reference, we list it below.

Proposition 3. *In addition to Assumptions 5 and 6 suppose that $EY_{11}^4 = 3$ and $\|\mathbf{T}_1\|$, the spectral norm of $\mathbf{T}_1$, is bounded and $F^{\mathbf{T}_1}$ converges weakly to $H(x)$. Then the random vector (6.13) converges in distribution to a Gaussian vector with mean*

$$EX_f = -\frac{1}{2\pi i}\oint_{\mathcal{C}} f(z)\frac{c\int \frac{\underline{m}^3(z)t^2dH(t)}{(1+t\underline{m}(z))^3}}{\left(1 - c\int \frac{\underline{m}^2(z)t^2dH(t)}{(1+t\underline{m}(z))^2}\right)^2}dz \tag{6.15}$$

and covariance function being the same as (6.14) with $\underline{m}_{\phi}(z)$ replaced by $\underline{m}(z)$. Here $\underline{m}(z)$, which can be obtained from $m(z)$ in (6.5), is the Stieltjes transform of the LSD of the matrix $\underline{\mathbf{S}} = \frac{1}{n}\mathbf{X}^T\mathbf{X}$.

When $\mathbf{T}_1$ becomes the identity matrix, $H(t)$ becomes a degenerate distribution at point 1 and we do not need to assume that $EY_{11}^4 = 3$ in this case. Theorem 1.4 of Pan and Zhou (2008) gives CLT for the random vector (6.13). We list it below.

Proposition 4. *In addition to Assumptions 5 and 6 suppose that $EY_{11}^4 < \infty$. Then the random vector (6.13) converges in distribution to a Gaussian vector with mean*

$$EX_f = -\frac{1}{2\pi i}\oint_{\mathcal{C}} f(z)\frac{c\frac{\underline{m}^3(z)}{(1+\underline{m}(z))^3}}{\left(1-c\frac{\underline{m}^2(z)}{(1+t\underline{m}(z))^2}\right)^2}dz - \frac{c(EX_{11}^4-3)}{2\pi^2}\oint_{\mathcal{C}} f(z)\frac{\frac{\underline{m}^3(z)}{(1+\underline{m}(z))^3}}{1-c\frac{\underline{m}^2(z)}{(1+t\underline{m}(z))}}dz \tag{6.16}$$

and covariance

$$\begin{aligned}Cov(X_{f_l}, X_{f_r}) &= -\frac{1}{\pi^2}\oint_{\mathcal{C}_1}\oint_{\mathcal{C}_2}\frac{f_l(z_1)f_r(z_2)}{(\underline{m}(z_1)-\underline{m}(z_2))^2}\frac{d}{dz_2}\underline{m}(z_2)\frac{d}{dz_1}\underline{m}(z_1)dz_1dz_2\\ &-\frac{c(EX_{11}^4-3)}{2\pi^2}\oint_{\mathcal{C}_1}\oint_{\mathcal{C}_2} f_l(z_1)f_r(z_2)\frac{d}{dz_1}[\frac{1}{1+\underline{m}(z_1)}]\frac{d}{dz_2}[\frac{1}{1+\underline{m}(z_2)}]dz_1dz_2.\end{aligned} \tag{6.17}$$

6.2.3 Test statistic

There are two questions to be addressed before proposing a test statistic based on Theorem 11, Propositions 3 and 4. The first one is the choice of the test function $f(\lambda)$ associated with M_n in (6.9). The second one is that the mean of the asymptotic distribution of M_n, which includes the spectral density $\phi(\lambda)$ of time series $\mathbf{x}_i$ or $H(x)$ associated with linear independence structures, is often unknown in practice no matter what $f(\lambda)$ is.

For the first question, we choose two simple test functions $f_1(\lambda) = \lambda$ and $f_2(\lambda) = \lambda^2$ for simplicity and consider their linear combination. To overcome the second difficulty, we divide n time series into two groups, each of which contains $[n/2]$ time series, where $[n/2]$ is the largest integer smaller than $n/2$. By Theorem 11 or Proposition 3 we have

$$\Big(\int xdG_n^{(i)}(x), \int x^2dG_n^{(i)}(x)\Big) \xrightarrow{d} \Big(X_x^{(i)}, X_{x^2}^{(i)}\Big),\ as\ n\to\infty,\ i=1,2 \tag{6.18}$$

where $G_n^{(i)}(x) = p\Big(F^{\mathbf{S}^{(i)}}(x) - F_{c_{n^{(i)}},H_{n^{(i)}}}(x)\Big)$ with $c_{n^{(i)}} = p/[n/2], H_{n^{(i)}} = H_n$, $F_{c_{n^{(i)}},H_{n^{(i)}}}(x)$ is the analogue of F_{c_n,H_n} but corresponding to $\mathbf{S}^{(i)} = \frac{1}{[n/2]}\mathbf{X}^{(i)}\mathbf{X}^{(i)\prime}$ and $\mathbf{X}^{(i)}$ consisting of the i-th group of the divided time series, $i = 1, 2$ ($\mathbf{X} = (\mathbf{X}^{(1)}, \mathbf{X}^{(2)})$ when n is even). Here $\Big(X_x^{(i)}, X_{x^2}^{(i)}\Big)$ is the limiting distribution corresponding to the i-th group time series. Since the statistics on the left hand side of (6.18) for the two groups of time series are independent under $\mathbf{H}_0$, we calculate the difference of the two statistics and obtain

$$\Big(\int x d\widetilde{G}_n(x), \int x^2 d\widetilde{G}_n(x)\Big) \xrightarrow{d} \Big(\widetilde{X}_x, \widetilde{X}_{x^2}\Big), \ as\ n \to \infty, \tag{6.19}$$

where

$$\widetilde{G}_n(x) = G_n^{(1)}(x) - G_n^{(2)}(x) = p\Big(F^{\mathbf{S}^{(1)}}(x) - F^{\mathbf{S}^{(2)}}(x)\Big), \tag{6.20}$$

and $\widetilde{X}_x = X_x^{(1)} - X_x^{(2)}$, $\widetilde{X}_{x^2} = X_{x^2}^{(1)} - X_{x^2}^{(2)}$.

It follows from Theorem 11 that $(\widetilde{X}_x, \widetilde{X}_{x^2})$ is bivariate normal with mean 0 and covariance matrix $\widetilde{\mathbf{\Omega}}$, where $\widetilde{\mathbf{\Omega}} = 2\mathbf{\Omega}$ and $\mathbf{\Omega} = (\omega_{gh})_{2\times 2}$ is the asymptotic covariance matrix of $(\int x dG_n^{(i)}(x), \int x^2 dG_n^{(i)}(x))$ given by

$$\omega_{gh} = -\frac{1}{\pi^2} \oint_{\mathcal{C}_1} \oint_{\mathcal{C}_2} \frac{f_g(z_1) f_h(z_2)}{(\underline{m}_\phi(z_1) - \underline{m}_\phi(z_2))^2} \frac{d}{dz_2}\underline{m}_\phi(z_2) \frac{d}{dz_1}\underline{m}_\phi(z_1) dz_1 dz_2. \tag{6.21}$$

Note that (6.20) does not involve any unknown parameters. Therefore, we propose the following testing statistic for $\mathbb{H}_0$:

$$L_n = \Big(\int x d\widetilde{G}_n(x), \int x^2 d\widetilde{G}_n(x)\Big)\widetilde{\mathbf{\Omega}}^{-1}\begin{pmatrix} \int x d\widetilde{G}_n(x) \\ \int x^2 d\widetilde{G}_n(x) \end{pmatrix}. \tag{6.22}$$

As for the linear independence structures, the statistic L_n is the same except that $\underline{m}_\phi(z)$ in (6.21) should be replaced by the Stieltjes transform $\underline{m}(z)$ given in Proposition 3.

The following theorem is a direct application of Theorem 11 or Proposition 3.

Theorem 12. *Under the assumptions in Theorem 11 or in Proposition 3, the test statistic L_n converges in distribution to $\chi^2(2)$, which denotes the chi–squared random variable with the degree of freedom being 2.*

Remark 15. *The proposed statistic L_n contains the inverse covariance matrix $\widetilde{\mathbf{\Omega}}^{-1}$ and this matrix contains the unknown parameter $\underline{m}_\phi(z)$. This parameter can be estimated either by the algorithm provided above, or the sample Stieltjes transform $\underline{m}_n(z) = \frac{1}{p}tr(\mathbf{X}'\mathbf{X} - z\mathbf{I}_n)^{-1}$. Furthermore, the asymptotic distribution is still χ^2 after plugging in the estimator of $\underline{m}_\phi(z)$ by the Slutsky theorem. In view of this the proposed statistic L_n is easy to implement.*

Remark 16. *Traditionally, the method of dividing total samples into two parts is to use one part to do test and the other part to estimate unknown parameters. However, the strategy of dividing total samples into two parts here serves as a different purpose, eliminating the unknown term involved in the linear spectral statistic M_n. Indeed, we make use of the full strength of all observations, because if the first group and the second group are not independent or there is dependence among each group, then (6.19) is not true.*

We also considered a Bootstrap method as follows. By a parametric bootstrap we may redraw a sample $\mathbf{x}^* = \{\mathbf{x}_1^*, ..., \mathbf{x}_n^*\}$ from the p-variate normal distribution with mean zero and the population covariance matrix $\mathbf{S}$ defined in (6.1). Then consider the bootstrap linear spectral statistic

$$\int f(x)dG_n^*(x), \tag{6.23}$$

where $G_n^*(x) = p\Big[F^{\mathbf{S}_3}(x) - F_{c_n,F^{\mathbf{S}}}(x)\Big]$ and $\mathbf{S}_3 = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i^*(\mathbf{x}_i^*)^T$. We can further construct a statistic like (6.22) by replacing $\widetilde{G}_n(x)$ with $G_n^*(x)$. More-

over its asymptotic distribution can be directly obtained from Theorem 11 or Proposition 3.

However simulations show that the bootstrap statistic is not as powerful as the one proposed based on the strategy of dividing observations. The key reason is that the independence assumption under $\mathbb{H}_0$ *is reflected in* $F^{\mathbf{S}}$ *and its limit only such that the difference* $p(F^{\mathbf{S}} - F_{c_n,H_n})$ *is not used. As a consequence it can not identify the alternatives whose limit is the same as the one determined by* (6.6) *such as* $\frac{1}{n}\mathbf{X}\mathbf{T}_3\mathbf{X}^T$ *with* $T_3 = \mathbf{I} + \mathbf{e}\mathbf{e}^T$ *(all components of* $\mathbf{e}$ *are one).*

6.2.4 The power under local alternatives

This section is to investigate the power for some local alternatives. The first interesting example (local alternative) is that $\mathbf{x}_1, \cdots, \mathbf{x}_n$ satisfy Assumption 5 but $\mathbf{T}_1$ there is assumed to be random, independent of $\{Y_{ij}\}$. Evidently, $\mathbf{x}_1, \cdots, \mathbf{x}_n$ are not independent in this case. Yet, Silverstein's result (1995) indicates that (6.5) still holds if $\{Y_{ij}\}$ are independent and independent of $\mathbf{T}_1$. This indicates that there may be the cases where the LSD of sample covariance matrix is also determined by (6.5) even when $\mathbf{x}_1, \cdots, \mathbf{x}_n$ are not independent. A nature concern is whether the statistic L_n works in this case. We now consider the case when the random $\mathbf{T}_1$ is the inverse of another sample covariance matrix ($\mathbf{S}$ becomes the F matrix in this case). It is then proved in Theorem 3.1 of Zheng (2012) that L_n has a central limit theorem different from that for independent $\mathbf{x}_1, \cdots, \mathbf{x}_p$. The difference is caused by randomness of $\mathbf{T}_1$ and one may refer to (6.32) in Step 2 of Zheng's proof.

Although it is difficult to provide a central limi theorem for the statistic L_n for the general alternative hypothesis $\mathbb{H}_1$, we can still evaluate the power of L_n for a class of local alternatives. Specifically speaking, we consider a kind of local alternative with a sample covariance matrix in the form of $\mathbf{X}\mathbf{T}_2\mathbf{X}^T$, as in (6.6). Set

$$R_j^{(i)} = p\int x^i d\Big(F_{\mathbb{H}_1}^{\mathbf{S}^{(j)}}(x) - F_{\mathbb{H}_0}^{\mathbf{S}^{(j)}}(x)\Big),\ i = 1, 2;\ j = 1, 2; \tag{6.24}$$

where $F_{\mathbb{H}_0}^{\mathbf{S}^{(j)}}$ stands for the ESD of $\mathbf{S}^{(j)}$ under $\mathbb{H}_0$ while $F_{\mathbb{H}_1}^{\mathbf{S}^{(j)}}$ is the ESD of $\mathbf{S}^{(j)}$ under $\mathbb{H}_1$.

Theorem 13. *In addition to assumptions in Theorem 11 or Theorem 3, suppose that in probability*

$$\lim_{n\to\infty}\left|R_j^{(i)}\right| \to \infty, \quad \textit{for any } i, j. \tag{6.25}$$

Then

$$\lim_{n\to\infty} P(L_n > \gamma_{1-\alpha}|\mathbb{H}_1) = 1,$$

where $\gamma_{1-\alpha}$ is the critical value of χ^2 under $\mathbb{H}_0$ corresponding to the significance level α.

Remark 17. *Suppose that each column of $\mathbf{X}$ satisfies either Assumption 4 or Assumption 5 and all columns are independent. Condition (6.25) is equivalent to requiring*

$$tr\Big(\mathbf{X}^{(j)}\mathbf{T}^{(j)}(\mathbf{X}^{(j)})^T\Big)^i - tr\Big(\mathbf{X}^{(j)}(\mathbf{X}^{(j)})^T\Big)^i \to \infty, \quad \textit{for any } i, j \tag{6.26}$$

in probability, where $\mathbf{X}^{(j)}\mathbf{T}^{(j)}(\mathbf{X}^{(j)})^T$ denotes the sample covariance matrix of the jth group of the observations under the alternative $\mathbb{H}_1$ with $\mathbf{T}^{(j)}$ characterizing the dependence among observations, while $\mathbf{X}^{(j)}(\mathbf{X}^{(j)})^T$ stands for the sample covariance matrix of the jth group of the observations under the null hypothesis H_0.

If

$$\mathbf{T}^{(j)} = \mathbf{I} + \mathbf{e}\mathbf{e}^T,$$

where the elements of the vector $\mathbf{e}$ *are all equal to one, then it is straightforward to verify that (6.26) is true. Moreover, most of the examples given in the subsequent section satisfy (6.26).*

6.3 Simulation results

This section provides some simulated examples to show the finite sample performance of the proposed test statistic L_n. To show the efficiency of our test, some classical time series models, such as MA(1), AR(1) and ARMA(1,1) processes, are considered. As for the dependent structures, we consider some dependent structures described by MA(1) model, AR(1) model, ARMA(1,1) model and factor model. The factor model is commonly used to illustrate cross-sectional dependence in cross-sectional panel data analysis.

6.3.1 Empirical sizes and empirical powers

First we introduce the method of calculating empirical sizes and empirical powers. Since the asymptotic distribution of the proposed test statistic L_n is a classical distribution, i.e. χ^2 distribution of degree 2, the empirical sizes and powers are easy to calculate. Let $z_{1-\frac{1}{2}\alpha}$ be the $100(1-\frac{1}{2}\alpha)\%$ quantile of the asymptotic null distribution $\chi^2(2)$ of the test statistic L_n. With K replications of the data set simulated under the null hypothesis, we calculate the empirical size as

$$\hat{\alpha} = \frac{\{\sharp\ of\ L_n^H \geq z_{1-\frac{1}{2}\alpha}\}}{K}, \tag{6.27}$$

where L_n^H represents the value of the test statistic L_n based on the data simulated under the null hypothesis.

In our simulation, we choose $K = 1000$ as the number of repeated simulations. The significance level is $\alpha = 0.05$. Since the asymptotic null distribution of the test statistic is a classical distribution, the quantile $z_{1-\frac{1}{2}\alpha}$ is easy to know. Similarly, the empirical power is calculated as

$$\hat{\beta} = \frac{\{\sharp\ of\ L_n^A \geq z_{1-\frac{1}{2}\alpha}\}}{K}, \tag{6.28}$$

where L_n^A represents the value of the test statistic L_n based on the data simulated under the alternative hypothesis.

6.3.2 Testing independence

In order to derive independent stationary time series $\{\mathbf{x}_i = (X_{1i}, X_{2i}, \ldots, X_{pi})' : i = 1, \ldots, n\}$, we generate data from the following three data generating processes (DGPs):

$$DGP1: \ X_{ji} = Z_{ji} + \theta_1 Z_{j-1,i}, \ j = 1, 2, \ldots, p; \ i = 1, 2, \ldots, n; \tag{6.29}$$

$$DGP2: \ X_{ji} = \phi_1 X_{j-1,i} + Z_{ji}, \ j = 1, 2, \ldots, p; \ i = 1, 2, \ldots, n; \tag{6.30}$$

$$DGP3: \ X_{ji} - \phi_1 X_{j-1,i} = Z_{ji} + \theta_1 Z_{j-1,i}, \ j = 1, 2, \ldots, p; \ i = 1, 2, \ldots, n, \tag{6.31}$$

where $\{X_{0i}, Z_{ji} : j = 1, 2, \ldots, p; i = 1, 2, \ldots, n\} \sim i.i.d\ N(0, 1)$. For each DGP, we generate $p + 100$ observations and then discard the first 100 data in order to mitigate the impact of the initial values.

With these simulated data, we apply the proposed statistic L_n and calculate the empirical sizes. Table 6.1, Table 6.3 and Table 6.5 establish the empirical sizes for the three DGPs under different pairs of (p, n). The results show that our statistic L_n works well under the null hypothesis $\mathbb{H}_0$.

Additionally, their empirical sizes from the bootstrap method proposed in Remark 16 are illustrated in Table 6.2, Table 6.4 and Table 6.6 respectively.

6.3.3 Testing dependence

6.3.3.1 Three types of correlated structures

In this section, we test four dependent structures with the proposed test and provide the powers under each case. As in the last part of this section, we first generate data $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$ under DGP 1. To describe the cross-sectional dependence between $\mathbf{x}_{i_1}$ and $\mathbf{x}_{i_2}$, $\forall i_1 \neq i_2$, we generate new data $\widetilde{\mathbf{X}} = \mathbf{X}\mathbf{T}$, where $\mathbf{T}$ is a $p \times p$ Hermitian matrix which is the square root of a covariance matrix. $\mathbf{T}$ is constructed by the following three methods.

1. MA(1) type covariance matrix $\mathbf{\Sigma}_{MA} = (\sigma_{kh}^{MA})_{k,h=1}^p$:

$$\sigma_{kh}^{(MA)} = \begin{cases} (1+\theta^2), & k = h; \\ \theta, & |k-h| = 1; \\ 0, & |k-h| > 1. \end{cases} \tag{6.32}$$

   Under this case, $\mathbf{T} = \mathbf{\Sigma}_{MA}^{1/2}$.

2. AR(1) type covariance matrix $\mathbf{\Sigma}_{AR} = (\sigma_{kh}^{(AR)})_{k,h=1}^p$:

$$\sigma_{kh}^{(AR)} = \frac{\phi^{|k-h|}}{1-\phi^2}. \tag{6.33}$$

   Under this case, $\mathbf{T} = \mathbf{\Sigma}_{AR}^{1/2}$.

3. ARMA(1,1) type covariance matrix $\mathbf{\Sigma}_{ARMA} = (\sigma_{kh}^{(ARMA)})_{k,h=1}^p$:

$$\sigma_{kh}^{(ARMA)} = \begin{cases} 1 + \frac{(\phi+\theta)^2}{1-\phi^2}, & k = h; \\ \phi + \theta + \frac{(\phi+\theta)^2\phi}{1-\phi^2}, & |k-h| = 1; \\ \phi^{|k-h|-1}(\phi + \theta + \frac{(\phi+\theta)^2\phi}{1-\phi^2}), & |k-h| \geq 2. \end{cases} \tag{6.34}$$

Under this case, $\mathbf{T} = \mathbf{\Sigma}_{ARMA}^{1/2}$.

The powers under the three cases are illustrated in Table 6.7, Table 6.8 and Table 6.9. The true parameters are taken as $\phi = 0.8$ and $\theta = 0.2$. It can be seen that the powers are near 1 as n and p tend to infinity in the same order.

6.3.3.2 Factor model dependence

We consider a data generating process which comes from a dynamic factor model, which is always used to describe cross-sectional dependence.

$$X_{ji} = \boldsymbol{\lambda}' \mathbf{f}_j + \varepsilon_{ji},\ i = 1, 2, \ldots, n,\ j = 1, 2, \ldots, p, \tag{6.35}$$

with

$$\mathbf{f}_j = \mathbf{z}_j + \theta \mathbf{z}_{j-1},\ i = 1, 2, \ldots, n,\ j = 1, 2, \ldots, p, \tag{6.36}$$

where $\boldsymbol{\lambda}$ is an $r \times 1$ deterministic vector whose elements are called factor loadings; $\mathbf{f}_j$ is an $r \times 1$ random vector generated from (6.36), whose elements are called factors and the cross-section dependence between $\mathbf{x}_{i_1}$ and $\mathbf{x}_{i_2}$ are caused by the common factors $\mathbf{f}_j$. $\{\mathbf{z}_j : j = 1, 2, \ldots, p\} \sim i.i.d\ N(\mathbf{0}_r, \mathbf{I}_r)$ where $\mathbf{0}_r$ is an $r \times 1$ vector with elements 0 and $\mathbf{I}_r$ is an $r \times r$ identity matrix. $\{\varepsilon_{ji} : j = 1, 2, \ldots, p; i = 1, 2, \ldots, n\} \sim i.i.d\ N(0, 1)$ are idiosyncratic errors.

First, we generate the factor loadings in the vector $\boldsymbol{\lambda}$ from $N(4, 1)$ before generating data from (6.35) and (6.36). After generating the data, we can apply the proposed test statistic L_n to the data and the empirical powers are shown in Table 6.10. From this table, we can see that the powers increase as the number of factors r increases. This is reasonable in the sense that more factors should bring in stronger dependence.

6.3.3.3 Common random dependence

We consider a special dependent structure which is caused by a common random part. The data generating process is as follows.

$$\mathbf{x}_i = \mathbf{A}\mathbf{y}_i, \quad i = 1, 2, \ldots, n, \tag{6.37}$$

where $\mathbf{A}$ is a $p \times p$ random matrix whose components are i.i.d standard normal random variables; and $\mathbf{y}_i$, $i = 1, 2, \ldots, n$ are independent $p \times 1$ random vectors, whose components are assumed to be i.i.d standard normal random variables.

Therefore the random vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ are dependent due to the common random part $\mathbf{A}$. The empirical powers are listed in Table 6.11. From the table, we can see that the proposed statistic L_n is powerful to capture this kind of dependence.

6.3.3.4 ARCH type dependence

It is known that dependent relations may be linear dependence or nonlinear dependence. The examples above are all linear dependent structures. In this section, we will present a nonlinear dependent structure.

Let us consider an autoregressive conditional heteroskedasticity (ARCH) model of the form:

$$X_{ji} = Z_{ji}\sqrt{\alpha_0 + \alpha_1 X_{j,i-1}^2}, i = 1, 2, \ldots, n;\ j = 1, 2, \ldots, p; \tag{6.38}$$

where $\{Z_{ji} : j = 1, 2, \ldots, p; i = 1, 2, \ldots, n\}$ are white noise error terms with zero mean and unit variance. Here we take $\alpha_0, \alpha_1 \in (0, 1)$ and $3\alpha_1^2 < 1$, since the fourth moment of the elements of X_{ji} exists.

From this model, the sequences $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ are dependent but uncorrelated. Moreover, this sequence is a multiple martingale difference sequence. The components of each vector $\mathbf{x}_i$ are independent here. This simplified assumption is imposed because the asymptotic theory is established for covariance time series under the assumption that the fourth moment equals 3 while the asymptotic theorem is provided for random vectors with i.i.d. components without this restriction.

Simulation results indicate that the proposed test statistic L_n can not detect this type of dependence between $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$. Nevertheless, if we replace the elements X_{jt} by X_{jt}^2, then our statistic L_n can capture the dependence of this type. This efficiency is due to the correlation between the high powers of $\{X_{jt} : t = 1, 2, \ldots, n\}$.

Table 6.12 lists the powers of the proposed statistics L_n testing model (6.38) in several cases, i.e. α_0 and α_1 take different values. From the table, we can find the phenomenon that as α_1 increases, the powers also increase. This is consistent with our intuition that larger α_1 brings about larger correlation between $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$.

6.4 Conclusion

This chapter provides a novel approach for independence test among a large number random vectors including covariance stationary time series of length p by using the empirical spectral distribution of the sample covariance matrix of the grouped time series under investigation. This test can capture various kinds of dependent structures, e.g. MA(1) model, AR(1) model, ARCH(1) model and the dynamic factor model established in the simulation section. The conventional method(LRT proposed by Anderson (1984)) utilized the correlated relationship between random vectors with i.i.d components to capture their dependence, instead of covariance stationary time series. Hong (1996) proposed a test statistic based on correlation functions to investigate independence between two covariance stationary time series. On the one hand, this idea is only efficient for normal distributed data. It may be an inappropriate tool for non-Gaussian distributed data, such as martingale difference sequences (e.g.ARCH(1) model), nonlinear MA(1) model etc., which possess dependent but uncorrelated structures. On the other hand, his method is only applicable to independence test for finite number of covariance stationary time series. Then the proposed test is more advantageous in these two points. The simulation results and an empirical application to cross-sectional independence test for stock prices in S&P500 highlight this approach.

6.5 Appendix

First, we present some lemmas and technical facts used in the proofs of the main theorems.

6.5.1 Useful lemmas

We would point out that (3.52) can be obtained from the proof of Lytova and Pastur (2009).

Our proof utilizes the generalized Fourier transform as follows:

Lemma 23 (Proposition 2.1 of Lytova and Pastur (2009)). *Let $g : \mathbb{R}_+ \to \mathbb{C}$ be locally Lipshitzian and such that for some $\delta > 0$*

$$\sup_{t\geq 0} e^{-\delta t}|g(t)| < \infty$$

and let $\widetilde{g} : \{z \in \mathbb{C} : Im(z) < -\delta\} \to \mathbb{C}$ be its generalized Fourier transform

$$\widetilde{g}(z) = i^{-1}\int_0^\infty e^{-izt}g(t)dt.$$

The inversion formula is given by

$$g(t) = \frac{i}{2\pi} \int_L e^{izt} \widetilde{g}(z) dz, t \geq 0,$$

where $L = (-\infty - i\varepsilon, \infty - i\varepsilon)$*,* $\varepsilon > \delta$*, and the principal value of the integral at infinity is used.*

Denote the correspondence between functions and their generalized Fourier transforms as $g \leftrightarrow \widetilde{g}$. Then we have

$$\begin{aligned} & g^{'}(t) \leftrightarrow i\big(g(+0) + z\widetilde{g}(z)\big); \int_0^t g(\tau) d\tau \leftrightarrow (iz)^{-1} \widetilde{g}(z); \\ & \int_0^t g_1(t-\tau) g_2(\tau) d\tau := (g_1 * g_2)(t) \leftrightarrow i\widetilde{g}_1(z)\widetilde{g}_2(z). \end{aligned} \tag{6.39}$$

Furthermore, we introduce a simple fact about exponential matrices below.

Lemma 24 (Duhamel formula)**.** *Let* $\mathbf{W}_1, \mathbf{W}_2$ *be* $n \times n$ *matrices and* $t \in \mathbb{R}$*. Then we have*

$$\mathbf{e}^{(\mathbf{W}_1 + \mathbf{W}_2)t} = \mathbf{e}^{\mathbf{W}_1 t} + \int_0^t \mathbf{e}^{\mathbf{W}_1 (t-s)} \mathbf{W}_2 e^{(\mathbf{W}_1 + \mathbf{W}_2)s} ds. \tag{6.40}$$

Moreover, if $\big(W_{ij}(t)\big)_{1 \leq i,j \leq n}$ *is a matrix-valued function of* $t \in \mathbb{R}$ *that is* C^∞ *in the sense that each matrix element* $W_{ij}(t)$ *is* C^∞*. Then*

$$\frac{d}{dt} \mathbf{e}^{\mathbf{W}(t)} = \int_0^1 \mathbf{e}^{s\mathbf{W}(t)} \mathbf{W}^{'}(t) \mathbf{e}^{(1-s)\mathbf{W}(t)} ds, \tag{6.41}$$

where $\mathbf{W}^{'}(t)$ *is an* $n \times n$ *matrix with elements being the derivatives of the corresponding elements of* $\mathbf{W}(t)$*.*

Proof of Theorem 10: Since

$$E\Big(\int \lambda dF^{\mathbf{S}}(\lambda)\Big) = E\Big(\frac{1}{p} tr(\frac{1}{n}\mathbf{X}\mathbf{X}^{'})\Big) = \sum_{k=0}^{\infty} b_k^2,$$

the sequence $E\{F^{\mathbf{S}}(\lambda)\}$ is tight. By Theorem B.9 of Bai and Silverstein (2009), the proof of Theorem 10 is complete if we can verify the following two steps:

1. For any fixed $z \in \mathcal{C}^+$, $m_n(z) - Em_n(z) \to 0$, a.s. as $n \to \infty$, where $m_n(z) = \frac{1}{p} tr\mathbf{G}^{-1}(z)$ with $\mathbf{G}^{-1}(z) = (\mathbf{S} - z\mathbf{I}_p)^{-1}$ and $\mathbf{I}_p$ being a $p \times p$ identity matrix.

2. For any fixed $z \in \mathcal{C}^+$, $Em_n(z) \to m_\phi(z)$, as $n \to \infty$, where $m_\phi(z) = \int \frac{1}{\lambda - z} dF_{c,\phi}(\lambda)$.

The first step is omitted here, since it is similar to the proof on page 54 of Bai and Silverstein (2009).

We will finish the second step by comparing $Em_n(z)$ for the Gaussian case and nonGaussian case: as $n \to \infty$

$$Em_n(z) - E\hat{m}_n(z) \to 0, \tag{6.42}$$

$$E\hat{m}_n(z) \to m_\phi(z), \tag{6.43}$$

where $\hat{m}_n(z)$ is obtained from $m_n(z)$ with the elements $X_{jt} = \sum_{k=0}^{\infty} b_k \xi_{j-k,t}$ replaced by $\hat{X}_{jt} = \sum_{k=0}^{\infty} b_k \hat{\xi}_{j-k,t}$. Here $\{\hat{\xi}_{j-k,t}\}$ are i.i.d Gaussian random variables with mean zero and variance one and $\{\hat{\xi}_{j-k,t}\}$ are independent of $\{\xi_{j-k,t}\}$. (6.43) obviously holds by Yao (2012).

Let $Im(z) = v > 0$ and below we will frequently use the fact that $|\hat{m}_n(z)|$ and $|m_n(z)|$ are both bounded by $1/v$ without mention. We now consider (6.42) and start with the truncation of underlying random variables. Define

$$\mathbf{S}^\tau = \frac{1}{n}\mathbf{X}^\tau(\mathbf{X}^\tau)^T,\ \mathbf{X}^\tau = (X_{jt}^\tau)_{p\times n}, \tag{6.44}$$

$$X_{jt}^\tau = \sum_{k=0}^{\infty} b_k \xi_{j-k,t}^\tau,\ \xi_{j-k,t}^\tau = \xi_{j-k,t} I(|\xi_{j-k,t}| \le \tau\sqrt{n}), \tag{6.45}$$

where $\tau = \tau_n$ is a positive sequence satisfying

$$\tau \to 0, \quad \frac{1}{\tau} E(|\xi_{11}|^2 I(|\xi_{11}| > \tau\sqrt{n})) \to 0. \tag{6.46}$$

We claim that for every $\tau > 0$,

$$\lim_{n\to\infty} \Big| Em_n(z) - Em_n^\tau(z) \Big| = 0, \tag{6.47}$$

where $m_n^\tau(z) = \frac{1}{p} tr \mathbf{G}_\tau^{-1}(z)$ with $\mathbf{G}_\tau^{-1}(z) = \frac{1}{p} tr(\mathbf{S}^\tau - z\mathbf{I}_p)^{-1}$. In fact, we have

$$\begin{aligned}
& \Big| Em_n(z) - Em_n^\tau(z) \Big| \\
\le\ & \Big| \frac{1}{p\sqrt{n}} \sum_{j,t=1}^{p,n} E\Big(\big(\mathbf{G}_\tau^{-1}(z)\mathbf{G}^{-1}(z)\frac{1}{\sqrt{n}}\mathbf{X}\big)_{jt} \big(X_{jt} - X_{jt}^\tau\big)\Big) \Big| \\
& + \Big| \frac{1}{p\sqrt{n}} \sum_{j,t=1}^{p,n} E\Big(\big(X_{jt} - X_{jt}^\tau\big)\big(\mathbf{G}^{-1}(z)\mathbf{G}_\tau^{-1}(z)\frac{1}{\sqrt{n}}\mathbf{X}\big)_{jt}\Big) \Big| \\
\le\ & \frac{Cnp}{p\sqrt{n}} \sum_{k=0}^{\infty} |b_k| E|\xi_{11}| I(|\xi_{11}| \ge \tau\sqrt{n}) \le \frac{CE|\xi_{11}|^2 I(|\xi_{11}| \ge \tau\sqrt{n})}{\tau} \sum_{k=0}^{\infty} |b_k| \to 0,
\end{aligned}$$

where the first inequality uses the resolvent identity

$$(\mathbf{A} - z\mathbf{I}_p)^{-1} - (\mathbf{B} - z\mathbf{I}_p)^{-1} = -(\mathbf{A} - z\mathbf{I}_p)^{-1}(\mathbf{A} - \mathbf{B})(\mathbf{B} - z\mathbf{I}_p)^{-1},$$

holding for any Hermitian matrices $\mathbf{A}$ and $\mathbf{B}$ and the second inequality uses

$$\begin{aligned}
\Big| \Big(\mathbf{G}_\tau^{-1}(z)\mathbf{G}^{-1}(z)\frac{1}{\sqrt{n}}\mathbf{X}\Big)_{jt} \Big| &\le \frac{1}{v} ||\mathbf{G}^{-1}(z)\frac{1}{\sqrt{n}}\mathbf{X}|| = \frac{1}{v} ||\mathbf{G}^{-1}(z)\frac{1}{n}\mathbf{X}\mathbf{X}^T\mathbf{G}^{-1}(z)||^{1/2} \\
&\le \frac{1}{v} ||\mathbf{G}^{-1}(z)||^{1/2} + \frac{1}{v}|z|^{1/2} ||\mathbf{G}^{-1}(z)|| \le C. \qquad (6.48)
\end{aligned}$$

Here $\| \cdot \|$ denotes the spectral norm of a matrix. Also throughout the chapter we use C to denote constants which may change from line to line.

In view of (6.47) it is sufficient to prove that$|Em_n^\tau(z) - E\hat{m}_n(z)| \to 0,\ as\, n \to \infty$. However for simplicity below we still use notation $m_n(z), \mathbf{X}, X_{jt}, \xi_{j-k,t}$ instead of using $m_n^\tau(z), \mathbf{X}^\tau, X_{jt}^\tau, \xi_{j-k,t}^\tau$ and prove (6.42). But one should keep in mind that $|\xi_{j-k,t}| \le \tau\sqrt{n}$.

We next prove (6.42) by an interpolation technique first introduced in Lytova and Pastur (2009). To this end define the interpolation matrix

$$\mathbf{S}(s) = \frac{1}{n}\mathbf{X}(s)\mathbf{X}^T(s), \mathbf{X}(s) = \Big(X_{\theta,t}(s)\Big) = s^{1/2}\mathbf{X} + (1-s)^{1/2}\hat{\mathbf{X}},\ s \in [0,1] \tag{6.49}$$

and

$$\mathbf{G}^{-k}(s,z) = \Big(\mathbf{S}(s) - z\mathbf{I}_p\Big)^{-k},\ m_n(s,z) = \frac{1}{p}tr\mathbf{G}^{-1}(s,z), \quad k = 1,2.$$

Write $\Phi_{jt}(s) = \Big(\mathbf{G}^{-2}(s,z)\frac{1}{\sqrt{n}}\mathbf{X}(s)\Big)_{jt}$. We then have

$$Em_n(z) - E\hat{m}_n(z) = \int_0^1 \frac{\partial}{\partial s}Em_n(s,z)ds =$$

$$-\frac{1}{p}\int_0^1 \sum_{j,t=1}^{p,n} E\Big[s^{-1/2}\frac{1}{\sqrt{n}}X_{jt}\Phi_{jt}(s)\Big]ds + \frac{1}{p}\int_0^1 \sum_{j,t=1}^{p,n} E\Big[(1-s)^{-1/2}\frac{1}{\sqrt{n}}\hat{X}_{jt}\Phi_{jt}(s)\Big]ds, \tag{6.50}$$

where we have used the formula below

$$\frac{\partial \mathbf{G}^{-1}(s,z)}{\partial s} = -\mathbf{G}^{-1}(s,z)\frac{\partial \mathbf{S}(s)}{\partial s}\mathbf{G}^{-1}(s,z).$$

Consider the second term in (6.50) first. Since $\hat{X}_{jt} = \sum_{k=0}^{\infty} b_k\hat{\xi}_{j-k,t}$ we have

$$E\Big(\frac{1}{\sqrt{n}}\hat{X}_{jt}\Phi_{jt}(s)\Big) = \sum_{k=0}^{\infty} b_k E\Big(\frac{1}{\sqrt{n}}\hat{\xi}_{j-k,t}\Phi_{jt}(s)\Big). \tag{6.51}$$

Applying Lemma 10 to each summand in (6.51) we have

$$(1-s)^{-1/2}\sum_{k=0}^{\infty} b_k E(\frac{1}{\sqrt{n}}\hat{\xi}_{j-k,t}\Phi_{jt}(s)) = \frac{1}{n}\sum_{k=0}^{\infty} b_k \sum_{\theta=j-k}^{p} b_{\theta-j+k}E\Big(D_{\theta,t}(\Phi_{jt}(s))\Big), \tag{6.52}$$

where the partial derivative $D_{\theta,t} = \partial/\partial(\frac{1}{\sqrt{n}}X_{\theta t}(s))$ and we used the fact that

$$\frac{\partial \hat{X}_{\theta t}}{\partial \hat{\xi}_{j-k,t}} = b_{\theta-j+k}, \quad \frac{\partial X_{\theta t}(s)}{\partial \hat{X}_{\theta t}} = (1-s)^{1/2}.$$

Consider the first term in (6.50) now. As before, applying the fact that $X_{jt} = \sum_{k=0}^{\infty} b_k \xi_{j-k,t}$ and Lemma 11 to each summand of the first term in (6.50), we obtain

$$E\Big(s^{-1/2}\frac{1}{\sqrt{n}}X_{jt}\Phi_{jt}(s)\Big) = s^{-1/2}\sum_{k=0}^{\infty} b_k E\Big(\frac{1}{\sqrt{n}}\xi_{j-k,t}\Phi_{jt}(s)\Big) \tag{6.53}$$
$$= s^{-1/2}\frac{1}{\sqrt{n}}\sum_{k=0}^{\infty} b_k \kappa_{1,\tau} E\Phi_{jt}(s) + s^{-1/2}\frac{1}{n}\sum_{k=0}^{\infty} b_k\kappa_{2,\tau}\sum_{\zeta=j-k}^{p} b_{\zeta-j+k}E\Big(D_{\zeta,t}(\Phi_{jt}(s))\Big) + \varepsilon_1,$$

where $\kappa_{i,\tau}$ denotes the ith cumulant of the variable $\xi_{j-k,t}$ with $i = 1, 2$,

$$|\varepsilon_1| \leq \frac{C_1 s^{-1/2}}{n^{3/2}}\sum_{k=0}^{\infty}|b_k| E\Big(|\xi_{j-k,t}|^3 \sup_{|\xi_{j-k,t}|\leq\tau\sqrt{n}} |\widetilde{D}^2_{j-k,t}(\Phi_{jt}(s))|\Big),$$

with

$$\widetilde{D}^2_{j-k,t}(\Phi_{jt}(s)) = \widetilde{D}_{j-k,t}\Big(\sum_{\zeta=j-k}^{p}\frac{\partial\Phi_{jt}(s)}{\partial\frac{1}{\sqrt{n}}X_{\zeta,t}(s)}\frac{\partial\frac{1}{\sqrt{n}}X_{\zeta,t}(s)}{\partial\frac{1}{\sqrt{n}}X_{\zeta,t}}\frac{\partial\frac{1}{\sqrt{n}}X_{\zeta,t}}{\partial\frac{1}{\sqrt{n}}\xi_{j-k,t}}\Big)$$
$$= s\sum_{\zeta=j-k}^{p}\sum_{\gamma=j-k}^{p} b_{\zeta-j+k}b_{\gamma-j+k}D_{\zeta,t}\Big(D_{\gamma,t}(\Phi_{jt}(s))\Big),$$

where $\widetilde{D}_{j-k,t} = \partial/\partial\frac{1}{\sqrt{n}}\xi_{j-k,t}$. Here we would point out that checking the argument of Lemma 11 in Lytova and Pastur (2009) shows that $\sup_{t\in R}$ in (3.52) can be replaced by $\sup_{|\xi_{j-k,t}|\leq\tau\sqrt{n}}$ in the remainder ε_1 due to the truncation step.

We conclude from (6.50)-(6.53) that

$$Em_n(z) - E\hat{m}_n(z) = -\int_0^1 \Big[\frac{s^{-1/2}}{pn^{1/2}}\sum_{k=0}^{\infty} b_k\sum_{j,t=1}^{p,n}\kappa_{1,\tau}E\Phi_{jt}(s) + \frac{1}{p}\sum_{j,t=1}^{p,n}\varepsilon_1$$
$$+\frac{s^{-1/2}}{np}\sum_{k=0}^{\infty} b_k\sum_{j,t=1}^{p,n}(\kappa_{2,\tau}-1)\sum_{\zeta=j-k}^{p} b_{\zeta-j+k}E\Big(D_{\zeta,t}(\Phi_{jt}(s))\Big)\Big]ds. \tag{6.54}$$

The next aim is to prove that each of the three integrands goes to zero as n tends to infinity. To this end, first let $\mu_{\ell,\tau}(\mu_\ell)$ and $\kappa_{\ell,\tau}(\kappa_\ell)$ be the

ℓth moment and cumulant of the truncated ξ_{jt} and the untruncated (ξ_{jt}) respectively. Then

$$|\mu_{\ell,\tau} - \mu_\ell| \le CE\Big(|\xi_{11}|^\ell I(|\xi_{11}| > \tau\sqrt{n})\Big).$$

As a result we have

$$|\kappa_{\ell,\tau} - \kappa_\ell| \le CE\Big(|\xi_{11}|^\ell I(|\xi_{11}| > \tau\sqrt{n})\Big) \le \frac{C}{(\tau\sqrt{n})^{2-\ell}} E(|\xi_{11}|^2 I(|\xi_{11}| > \tau\sqrt{n}). \tag{6.55}$$

This result uses the fact that cumulants can be expressed by moments as follows

$$\kappa_j = \sum_\lambda c_\lambda \mu_\lambda,$$

where the sum is over all additive partitions λ of the set $\{1, \dots, j\}$, $\{c_\ell : \ell \in \lambda\}$ are known coefficients and $\mu_\lambda = \prod_{\ell\in\lambda} \mu_\ell$.

Second we provide the upper bound of $\Phi_{jt}(s)$, $D_{\gamma,t}(\Phi_{jt}(s))$ and $D_{\zeta,t}\Big(D_{\gamma,t}(\Phi_{jt}(s))\Big)$. For simplicity, we introduce more new notation.

$$\begin{aligned}
&\mathbf{I}(\zeta,\gamma) = \mathbf{e}_\gamma \mathbf{e}_\zeta^T + \mathbf{e}_\zeta \mathbf{e}_\gamma^T, \quad \mathbf{W}(\gamma,t) = \mathbf{e}_\gamma \mathbf{e}_t^T \frac{1}{\sqrt{n}} \mathbf{X}^T(s) + \frac{1}{\sqrt{n}} \mathbf{X}(s) \mathbf{e}_t \mathbf{e}_\gamma^T,\\
&\mathbf{J}_1(\zeta) = \mathbf{G}^{-1}(s,z)\mathbf{W}(\zeta,t)\mathbf{G}^{-2}(s,z), \quad \mathbf{J}_2(\gamma,\zeta) = \mathbf{G}^{-1}(s,z)\mathbf{I}(\gamma,\zeta)\mathbf{G}^{-2}(s,z)\\
&\mathbf{J}_3(\gamma,\zeta) = \mathbf{G}^{-1}(s,z)\mathbf{W}(\gamma,t)\mathbf{G}^{-2}(s,z)\mathbf{W}(\zeta,t)\mathbf{G}^{-1}(s,z),\\
&\mathbf{J}_4(\zeta,\gamma) = \mathbf{G}^{-1}(s,z)\mathbf{W}(\zeta,t)\mathbf{G}^{-1}(s,z)\mathbf{W}(\gamma,t)\mathbf{G}^{-2}(s,z),
\end{aligned}$$

where $\mathbf{e}_\gamma$ and $\mathbf{e}_j$ are $p \times 1$ unit vectors with the γ-th and j-th elements being 1 respectively and others being zeros; and $\mathbf{e}_t$ is $n \times 1$ a unit vector with the t-th element being 1 and others being zeros. With these notation by a simple but tedious calculation we obtain

$$D_{\gamma,t}(\Phi_{jt}(s)) \;=\; -\mathbf{e}_j^T \mathbf{G}^{-2}(s,z)\mathbf{e}_\gamma + \mathbf{e}_j^T \mathbf{J}_1(\gamma) \frac{1}{\sqrt{n}} \mathbf{X}(s)\mathbf{e}_t + \mathbf{e}_j^T \mathbf{J}_1^T(\gamma) \frac{1}{\sqrt{n}} \mathbf{X}(s)\mathbf{e}_t$$

and

$$D_{\zeta,t}\Big(D_{\gamma,t}(\Phi_{jt}(s))\Big) = \mathbf{e}_j^T\mathbf{J}_1(\zeta)\mathbf{e}_\gamma + \mathbf{e}_j^T\mathbf{J}_1^T(\zeta)\mathbf{e}_\gamma - \mathbf{e}_j^T\mathbf{J}_4(\zeta,\gamma)\frac{1}{\sqrt{n}}\mathbf{X}(s)\mathbf{e}_t$$
$$-\mathbf{e}_j^T\mathbf{J}_4(\gamma,\zeta)\frac{1}{\sqrt{n}}\mathbf{X}(s)\mathbf{e}_t - \mathbf{e}_j^T\mathbf{J}_3(\gamma,\zeta)\frac{1}{\sqrt{n}}\mathbf{X}(s)\mathbf{e}_t + \mathbf{e}_j^T\mathbf{J}_1(\gamma)\mathbf{e}_\zeta - \mathbf{e}_j^T\mathbf{J}_1^T(\gamma)\mathbf{e}_\zeta$$
$$-\mathbf{e}_j^T\mathbf{J}_3(\zeta,\gamma)\frac{1}{\sqrt{n}}\mathbf{X}(s)\mathbf{e}_t - \mathbf{e}_j^T\mathbf{J}_4^T(\gamma,\zeta)\frac{1}{\sqrt{n}}\mathbf{X}(s)\mathbf{e}_t - \mathbf{e}_j^T\mathbf{J}_4^T(\zeta,\gamma)\frac{1}{\sqrt{n}}\mathbf{X}(s)\mathbf{e}_t$$
$$+\mathbf{e}_j^T\mathbf{J}_2(\gamma,\zeta)\frac{1}{\sqrt{n}}\mathbf{X}(s)\mathbf{e}_t + \mathbf{e}_j^T\mathbf{J}_2^T(\gamma,\zeta)\frac{1}{\sqrt{n}}\mathbf{X}(s)\mathbf{e}_t.$$

From the expansions of $\Phi_{jt}(s), D_{\gamma,t}(\Phi_{jt}(s))$ and $D_{\zeta,t}\Big(D_{\gamma,t}(\Phi_{jt}(s))\Big)$ we see that all the terms in such expansions include only three factors below:

$$D_1 = \Big(\frac{1}{\sqrt{n}}\mathbf{X}^T(s)\mathbf{G}^{-\ell}(s,z)\frac{1}{\sqrt{n}}\mathbf{X}(s)\Big)_{tt},\ D_2 = \Big(\mathbf{G}^{-\ell}(s,z)\frac{1}{\sqrt{n}}\mathbf{X}(s)\Big)_{kt},$$
$$D_3 = \mathbf{G}^{-\ell}(s,z)_{kk'},\quad \ell = 1,2, k, k' = j,\zeta,\ or\ \gamma.$$

These three factors turn out to be bounded, as seen below.

Obviously $|D_3| \leq v^{-\ell}$. Similar to (6.48) using

$$\mathbf{G}^{-1}(z)\frac{1}{n}\mathbf{X}(s)\mathbf{X}^T(s) = I + z\mathbf{G}^{-1}(s,z). \tag{6.56}$$

one may verify that

$$|D_2| \leq \frac{1}{v^{\ell-1}}\|\mathbf{G}^{-1}(s,z)\frac{1}{\sqrt{n}}\mathbf{X}(s)\| \leq C,\ j = 1,2$$

and

$$|D_1| \leq \|\frac{1}{\sqrt{n}}\mathbf{X}^T(s)\mathbf{G}^{-\ell}(s,z)\frac{1}{\sqrt{n}}\mathbf{X}(s)\| = \|\ \mathbf{G}^{-\ell}(s,z)\frac{1}{n}\mathbf{X}(s)\mathbf{X}^T(s)\ \| \leq C.$$

Therefore $\Phi_{jt}(s)$ and the two derivatives $D_{\gamma,t}(\Phi_{jt}(s))$, $D_{\zeta,t}\Big(D_{\gamma,t}(\Phi_{jt}(s))\Big)$ are bounded. This, together with (6.55) and (6.46), yields

$$\Big|\frac{s^{-1/2}}{pn^{1/2}}\sum_{k=0}^{\infty} b_k \sum_{j,t=1}^{p,n}\kappa_{1,\tau}E\Phi_{jt}\Big| \leq \frac{C}{\tau}E(|\xi_{11}|^2 I(|\xi_{11}| > \tau\sqrt{n})) \to 0$$

and

$$\Big|\frac{1}{np}\sum_{k=0}^{\infty} b_k \sum_{j,t=1}^{p,n}(\kappa_{2,\tau}-1)\sum_{\zeta=j-k}^{p} b_{\zeta-j+k}E\Big(D_{\zeta,t}(s)\Phi_{jt}(s)\Big)\Big|$$
$$\leq \quad CE(|\xi_{11}|^2 I(|\xi_{11}|>\tau\sqrt{n}))\to 0.$$

Moreover since $E|\xi_{jt}|^3\leq \tau\sqrt{n}$ and (6.46) we have

$$\Big|\frac{1}{p}\sum_{j,t=1}^{p,n}\varepsilon_1\Big|\leq C\tau\to 0,\ as\ n\to\infty.$$

These, together with (6.54), yield (6.42). The proof of this theorem is complete. □

Proof of Theorem 11: The strategy of the proof is the same as that in Lytova and Pastur (2009). That is, we first establish CLT for the case when $\{\xi_{j-k,t}\}$ are i.i.d $N(0,1)$ and then generalize it to the general distributions.

When $\{\xi_{j-k,t}\}$ are i.i.d $N(0,1)$, as stated in Section 2, under $\mathbf{H}_0$, the matrix $\mathbf{S}$ can be written in the form that $\mathbf{S}=\frac{1}{n}\mathbf{T}_1^{1/2}\mathbf{X}\mathbf{X}^T\mathbf{T}_1^{1/2}$ so that Theorem 9.10 of Bai and Silverstein (2009) is applicable. The asymptotic variance of Theorem 11 is the same as that in Bai and Silverstein (2009) while the asymptotic mean is obtained from that in Bai and Silverstein (2009) and the facts that (See Yao (2012) and Gray (2009))

$$\lim_{p\to\infty}\frac{1}{p}\sum_{k=1}^{p}f(\sigma_k)=\int_0^{\infty}f(x)dH(x)=\frac{1}{2\pi}\int_0^{2\pi}f(\phi(\lambda))d\lambda.$$

However to apply Bai and Silverstein (2009), we have to make sure that the spectral norm of the population covariance matrix $\mathbf{T}_1$ of each time series is bounded. We claim that this is ensured by the condition $\sum|b_j|<\infty$. In fact, let $\sigma_k=Cov(X_{jt},X_{j+k,t})$. By the expression (2.2) of the time series and a change of variables we have

$$\sum_{k=0}^{\infty}|\sigma_k|=\sum_{k=0}^{\infty}|Cov(\sum_{k_1=0}^{\infty}b_{k_1}\xi_{j-k_1,t},\sum_{k_2=0}^{\infty}b_{k_2}\xi_{j+k-k_2,t})|$$

$$= \sum_{k=0}^{\infty} |\sum_{k_1=0}^{\infty} b_{k_1} b_{k_1+k}| < (\sum_{k=0}^{\infty} |b_k|)^2 < \infty. \tag{6.57}$$

By Lemma 4.1 of Gray (2009) and (6.57) we conclude that

$$||\mathbf{T}_1|| \leq 4 \sum_{k=0}^{\infty} |\sigma_k| < \infty. \tag{6.58}$$

We next adopt an interpolation trick and compare the CLT of the general case with that of the Gaussian case. Recall the definition of $G_n(\lambda)$ in (2.8). Let

$$\mathcal{N}_n^{\circ}[f] = \int f(\lambda) dG_n(\lambda), \quad \mathcal{N}_n[f] = \int f(\lambda) dp F^{\mathbf{S}}(\lambda).$$

Define $\widehat{\mathcal{N}}_n^{\circ}[f]$ and $\widehat{\mathcal{N}}_n[f]$ to be obtained from $\mathcal{N}_n^{\circ}[f]$ and $\mathcal{N}_n[f]$ respectively, with the entries $X_{jt} = \sum_{k=0}^{\infty} b_k \xi_{j-k,t}$ replaced by $\hat{X}_{jt} = \sum_{k=0}^{\infty} b_k \hat{\xi}_{j-k,t}$ where $\{\hat{\xi}_{j-k,t}\}$ are i.i.d. N(0,1) and independent of $\{\xi_{j-k,t}\}$. By the continuous theorem of characteristic functions, it suffices to show that

$$R_n(x) := E\Big(e^{ix\mathcal{N}_n^{\circ}[f]}\Big) - E\Big(e^{ix\widehat{\mathcal{N}}_n^{\circ}[f]}\Big) \to 0, \ as \ n \to \infty. \tag{6.59}$$

Since the integrand function f admits the Fourier transform

$$\hat{f}(\theta) = \frac{1}{2\pi} \int e^{-i\theta\lambda} f(\lambda) d\lambda,$$

the Fourier inversion formula is

$$f(\lambda) = \int e^{i\theta\lambda} \hat{f}(\theta) d\theta. \tag{6.60}$$

Then the statistic $\mathcal{N}_n[f]$ can be written as

$$\mathcal{N}_n[f] = \int \hat{f}(\theta) u_n(\theta) d\theta,$$

where

$$u_n(\theta) = Tr\mathbf{U}(\theta), \ \mathbf{U}(\theta) = \mathbf{e}^{i\theta\mathbf{S}}. \tag{6.61}$$

By (6.60) we obtain

$$f^{'}(\mathbf{S}) = i\int \hat{f}(\theta)\theta\mathbf{U}(\theta)d\theta. \tag{6.62}$$

We still use the same truncation as that in (6.44) (and use the same notation) but this time τ satisfies (see formula (9.7.7) of Bai and Silverstein (2009))

$$\tau \to 0, \quad \tau^{-4}E|\xi_{j-k,t}|^4 I(\xi_{j-k,t}| > \tau\sqrt{n}) \to 0. \tag{6.63}$$

Note that

$$P\{\mathbf{X} \neq \mathbf{X}^{\tau}\} \leq \sum_{j,t=1}^{p,n} P\{X_{jt} \neq X_{jt}^{\tau}\} \leq \frac{1}{\tau^4 n^2}\sum_{j,t=1}^{p,n}\sum_{k=0}^{\infty} b_k E|\xi_{j-k,t}|^4 I(\xi_{j-k,t}| > \tau\sqrt{n}) \to 0.$$

In view of this it is enough to prove that

$$E\Big(e^{ix\mathcal{N}_{n\tau}^{\circ}[f]}\Big) - E\Big(e^{ix\widehat{\mathcal{N}}_{n}^{\circ}[f]}\Big) \to 0,\ as\ n \to \infty, \tag{6.64}$$

where $\mathcal{N}_{n\tau}^{\circ}[f]$ is obtained from $\mathcal{N}_{n}^{\circ}[f]$ with $\mathbf{X}$ replaced by $\mathbf{X}^{\tau}$.

As in the proof of Theorem 10 we still use notation $\xi_{j-k,t}, \mathbf{X}, \mathcal{N}_n[f]$ rather than $\xi_{j-k,t}^{\tau}, \mathbf{X}^{\tau}, \mathcal{N}_{n\tau}^{\circ}[f]$ and below prove (6.59). Recall the interpolation matrix defined in (6.49) and furthermore define

$$e_n(s,x) = \exp\Big(ixTrf\big(\mathbf{S}(s)\big)\Big), \quad \mathbf{U}(s,\theta) = (U_{jk}) = \mathbf{e}^{i\theta\mathbf{S}(s)}.$$

By (6.62) we have

$$\begin{aligned} R_n(x) &= a_n\int_0^1 \frac{\partial}{\partial s}E\Big(e_n(s,x)\Big)ds \\ &= ixa_n\int_0^1 E\Big[e_n(s,x)Tr\Big(f^{'}(\mathbf{S}(s))\big(s^{-1/2}\frac{1}{\sqrt{n}}\mathbf{X} - (1-s)^{-1/2}\frac{1}{\sqrt{n}}\hat{\mathbf{X}}\big)\frac{1}{\sqrt{n}}\mathbf{X}^{\tau'}(s)\Big)\Big]ds \\ &= -xa_n\int_0^1 ds\int \theta\hat{f}(\theta)(D_n - B_n)d\theta, \end{aligned} \tag{6.65}$$

where $a_n = \exp(-ix\int fdpF_{c_n,\phi_n})$ and

$$D_n = \frac{1}{\sqrt{ns}}\sum_{j,t=1}^{p,n} E\Big(X_{jt}\Psi_{jt}(s)\Big),\ B_n = \frac{1}{\sqrt{n(1-s)}}\sum_{j,t=1}^{p,n} E\Big(\hat{X}_{jt}\Psi_{jt}(s)\Big),$$

with

$$\Psi_{jt}(s) = e_n(s,x)\Big(\mathbf{U}(s,\theta)\frac{1}{\sqrt{n}}\mathbf{X}(s)\Big)_{jt}.$$

By Lemma 10 and a calculation similar to (6.51), (6.52) and (6.53) we obtain

$$B_n = \frac{1}{n}\sum_{j,t=1}^{p,n}\sum_{k=0}^{\infty} b_k \sum_{k^{(1)}=j-k}^{p} b_{k^{(1)}-j+k} E\Big(D_{k^{(1)}t}(\Psi_{jt}(s))\Big), \tag{6.66}$$

where $D_{k^{(1)}t} = \partial/\partial\frac{1}{\sqrt{n}}X_{k^{(1)}t}$.

Also, by Lemma 11 with $q=3$ we have

$$D_n = \sum_{\ell=0}^{3} T_{\ell\tau} + \varepsilon_3, \tag{6.67}$$

where

$$T_{0\tau} = \frac{s^{-1/2}}{\sqrt{n}}\sum_{j,t=1}^{p,n}\kappa_{1,\tau}\sum_{k=0}^{\infty} E\Psi_{jt}(s),$$

$$\begin{aligned} T_{\ell\tau} &= \frac{s^{(\ell-1)/2}}{\ell! n^{(\ell+1)/2}}\sum_{j,t=1}^{p,n}\kappa_{\ell+1,\tau}\sum_{k=0}^{\infty} b_k \sum_{k^{(\ell)},k^{(\ell-1)},\ldots,k^{(1)}}^{p} b_{k^{(\ell)}-j+k} b_{k^{(\ell-1)}-j+k}\cdots b_{k^{(1)}-j+k} \\ &\quad \cdot E\Big(D_{k^{(\ell)}t}D_{k^{(\ell-1)}t}\cdots D_{k^{(1)}t}\Psi_{jt}(s)\Big),\ \ell = 1,2,3; \end{aligned}$$

and

$$\begin{aligned} |\varepsilon_3| &\le \frac{Cs^2}{n^{5/2}}\sum_{j,t=1}^{p,n}\sum_{k=0}^{\infty}|b_k|\sum_{k^{(4)},\ldots,k^{(1)}=j-k}^{p}|b_{k^{(4)}-j+k}|\cdots|b_{k^{(1)}-j+k}| \\ &\quad \cdot\int_0^1 E\Big[|\xi_{j-k,t}|^5 D_{k^{(4)}t}\cdots D_{k^{(1)}t}\Psi_{jt}(s)\big|_{\xi_{j-k,t}=v\xi_{j-k,t}}\Big](1-v)^3 dv \end{aligned} \tag{6.68}$$

where $\Psi_{jt}(s)\big|_{\xi_{j-k,t}=v\xi_{j-k,t}}$ means that $\xi_{j-k,t}$ involved in $\Psi_{jt}(s)$ is replaced by $v\xi_{j-k,t}$ and $\kappa_{\ell,\tau}$ is the ℓth cumulant of $\xi_{j-k,t}$.

Next, we provide the upper bounds of derivatives:

$$D_{k^{(\ell)}t}D_{k^{(\ell-1)}t}\cdots D_{k^{(1)}t}\Psi_{jt}(s),\ \ell = 0,1,2,3,4.$$

Let $\mathbf{Y}(s) = (Y_{rt}(s)) = \frac{1}{\sqrt{n}}\mathbf{X}(s)$. Applying the Duhamel formula of Lemma 23 to the entries, $U_{jk^{(\ell)}}$, of $\mathbf{U}(s,\theta)$ we have

$$D_{\beta\alpha}(U_{jk^{(\ell)}}) = i\Big[\Big((\mathbf{UY}(s))_{j\alpha} * U_{\beta,k^{(\ell)}}\Big)(\theta) + \Big((\mathbf{UY}(s))_{k^{(\ell)}\alpha} * U_{j\beta}\Big)(\theta)\Big], \tag{6.69}$$

where the convolution $*$ is defined in (6.39). Here and below we use $\mathbf{U}$ to denote $\mathbf{U}(s,\theta)$ when there is no confusion. In view of (6.69) and the fact that $\mathbf{I}_p = \sum_{r=1}^{p} \mathbf{e}_r\mathbf{e}_r^{'}$ we have

$$\begin{aligned} & D_{k^{(\ell)}t}(\mathbf{UY}(s))_{jt} = D_{k^{(\ell)}t}\Big(\sum_{r=1}^{p} Y_{rt}(s)U_{rj}\Big) \\ = \ & U_{k^{(\ell)}j} + i\Big[\Big((\mathbf{Y}^T(s)\mathbf{UY}(s))_{tt} * U_{k^{(\ell)}j}\Big)(\theta) + \Big((\mathbf{UY}(s))_{jt} * (\mathbf{UY}(s))_{k^{(\ell)}t}\Big)(\theta)\Big], \end{aligned} \tag{6.70}$$

$$\begin{aligned} & D_{k^{(d)}t}(\mathbf{Y}^T(s)\mathbf{UY}(s))_{tt} = D_{k^{(d)}t}\Big(\sum_{r=1}^{p}(\mathbf{UY}(s))_{rt}\mathbf{Y}_{rt}(s)\Big) \\ = \ & 2(\mathbf{UY}(s))_{k^{(d)}t} + 2i\Big((\mathbf{Y}^{\tau'}(s)\mathbf{UY}(s))_{tt} * (\mathbf{UY}(s))_{k^{(d)}t}\Big)(\theta), \end{aligned} \tag{6.71}$$

and by (6.62)

$$D_{k^{(\ell)}t}(e_n(s,x)) = -2xe_n(s,x)\int \theta\hat{f}(\theta)(\mathbf{UY}(s))_{k^{(\ell)}t}d\theta, \tag{6.72}$$

where $\ell, d = 1,2,3,4$.

Since $\sum_{t=1}^{n}|U_{\alpha t}|^2 = 1$ and $||\mathbf{U}|| = 1$, from Hölder's inequality, we obtain

$$|(\mathbf{UY}(s))_{jt}| \le \Big(\sum_{r=1}^{p}(Y_{rt}(s))^2\Big)^{1/2}, \ |(\mathbf{Y}^T(s)\mathbf{UY}(s))_{tt}| \le \sum_{r=1}^{p}(Y_{rt}(s))^2. \tag{6.73}$$

Recalling the definition of $\Psi_{jt}(s)$ and repeatedly using (6.69)-(6.73) one can verify that

$$\Big|D_{k^{(\ell)}t}D_{k^{(\ell-1)}t}\cdots D_{k^{(1)}t}\Psi_{jt}(s)\Big| \le C + C\Big(\sum_{r=1}^{p}(Y_{rt}(s))^2\Big)^{(\ell+1)/2}, \ \ell = 0,1,2,3,4. \tag{6.74}$$

For example see (6.83) below for the expansion of $D_{k^{(1)}t}\Psi_{jt}(s)$. Moreover it is straightforward to check that $\ell = 0, 1, 2, 3$, $E\big(\sum_{r=1}^{p}(Y_{rt}(s))^2\big)^{(\ell+1)/2}$ is bounded by the fact that $n^2E|Y_{rt}(s)|^4 = E|X_{rt}(s)|^4 < \infty$. We then conclude that

$$E\Big|D_{k^{(\ell)}t}D_{k^{(\ell-1)}t}\cdots D_{k^{(1)}t}\Psi_{jt}(s)\Big| \le C_\ell,\ \ell = 0, 1, 2, 3. \tag{6.75}$$

However, to prove $\varepsilon_3 \to 0$, (6.74) for the case $\ell = 4$ is not enough for our purpose since

$$E|X_{rt}(s)|^5 \le C\tau\sqrt{n}, \tag{6.76}$$

not bounded. To offset this $\sqrt{n}$, one key observation is that from (6.69)-(6.72) we see that each term in the expansion of $D_{k^{(\ell)}t}D_{k^{(\ell-1)}t}\cdots D_{k^{(1)}t}\Psi_{jt}(s)$ is a product or a convolution of some of the following factors

$$(\mathbf{UY}(s))_{h_1t},\ (\mathbf{U})_{h_2h_3},\ (\mathbf{Y}^T(s)\mathbf{UY}(s))_{tt},\ e_n(s,x),$$

where h_i can be j or any $k^{(\ell)}$, $\ell = 1, \cdots, 4$. Let m_1 and m_2 be the total number of factors of types of $(\mathbf{UY}(s))_{h_1t}$ and $(\mathbf{Y}^T(s)\mathbf{UY}(s))_{tt}$ appearing in each term of the expansion, respectively. Then from (6.69)-(6.72) and (6.83) below we see that $(m_1 + 2m_2) \le 5$ (this explains (6.74) to some extent). Consider the case when $(m_1 + 2m_2) = 5$ first. In this case from (6.69)-(6.72) and (6.83) below we see that at least one $(\mathbf{UY}(s))_{h_1t}$ must be contained in the expansion. We below show how to handle such terms by demonstrating one example and all other cases can be similarly proved. Consider the term

$$(\mathbf{UY}(s))_{jt}(\mathbf{U})_{k^{(2)}k^{(3)}}(\mathbf{U})_{k^{(4)}k^{(1)}}(\mathbf{Y}^T(s)\mathbf{UY}(s))^2_{tt} \tag{6.77}$$

($m_1 = 1$ and $m_2 = 2$ in this case). Then for (6.68), it can be estimated as follows

$$\frac{1}{n^{5/2}}\sum_{j,t=1}^{p,n}\int_0^1 E\Big[|\xi_{j-k,t}|^5\Big((\mathbf{UY}(s))_{jt}(\mathbf{U})_{k^{(2)}k^{(3)}}(\mathbf{U})_{k^{(4)}k^{(1)}}(\mathbf{Y}^T(s)\mathbf{UY}(s))^2_{tt}\Big)\big|_{\xi_{j-k,t}=v\xi_{j-k,t}}\Big]$$

$$
\begin{aligned}
&\times(1-v)^3 dv\\
= &\ \frac{1}{n^{5/2}}\sum_{j,t=1}^{p,n}\int_0^1 E\Big[|\eta|^5\Big(((\mathbf{U}\mathbf{Y}(s))_{jt}(\mathbf{U})_{k^{(2)}k^{(3)}}(\mathbf{U})_{k^{(4)}k^{(1)}}(\mathbf{Y}^T(s)\mathbf{U}\mathbf{Y}(s))_{tt}^2\Big)\big|_{\xi_{j-k,t}=v\eta}\Big]\\
&\times(1-v)^3 dv\\
\le &\ \frac{1}{n^{5/2}}\sum_{t=1}^{n}\int_0^1 E\Big[|\eta|^5\Big(\sum_{j=1}^{p}|(\mathbf{U}\mathbf{Y}(s))_{jt}||(\mathbf{Y}^T(s)\mathbf{U}\mathbf{Y}(s))_{tt}|\Big)\big|_{\xi_{j-k,t}=v\eta}\Big](1-v)^3 dv\\
\le &\ \frac{1}{n^{5/2}}\sum_{t=1}^{n}\int_0^1 E\Big[|\eta|^5\sqrt{p}\Big((\sum_{r=1}^{p}Y_{rt}^2)^{5/2}\Big)\big|_{\xi_{j-k,t}=v\eta}\Big](1-v)^3 dv, \qquad (6.78)
\end{aligned}
$$

where η has the same distribution as $\{\xi_{r-k,t}\}$ and is independent of them, and satisfies $|\eta| \le \tau\sqrt{n}$; the first inequality uses the fact that $|(\mathbf{U})_{h_1h_2}| \le 1$; and the second inequality uses the second inequality of (6.73) and the following estimation

$$
\begin{aligned}
&\sum_{j=1}^{p}|(\mathbf{U}\mathbf{Y}(s))_{jt}| \le \sqrt{p}\Big(\sum_{j=1}^{p}|(\mathbf{U}\mathbf{Y}(s))_{jt}|^2\Big)^{1/2}\\
= &\ \sqrt{p}\Big(\sum_{j=1}^{p}\mathbf{e}_t^T\mathbf{Y}^T(s)\mathbf{U}^T\mathbf{e}_j\mathbf{e}_j^T\bar{\mathbf{U}}\mathbf{Y}(s)\mathbf{e}_t\Big)^{1/2}\\
= &\ \sqrt{p}\Big(\mathbf{e}_t^T\mathbf{Y}^T(s)\mathbf{Y}(s)\mathbf{e}_t\Big)^{1/2} = \sqrt{p}\Big(\sum_{r=1}^{p}Y_{rt}^2(s)\Big)^{1/2}, \qquad (6.79)
\end{aligned}
$$

where the second equality uses the fact that $\mathbf{U}$ is a symmetric unitary matrix. Moreover, since for any $h = 1, 2, \ldots, p$, the coefficient of $v\eta$ in the expansion of $Y_{rt}\big|_{\xi_{j-k,t}=v\eta}$ is b_{r-j+k} when $\xi_{j-k,t}$ is replaced by $v\eta$, we have

$$
\begin{aligned}
\Big(\sum_{r=1}^{p}Y_{rt}^2(s)\big|_{\xi_{j-k,t}=v\eta}\Big)^{m/2} &\le \frac{C}{n^{m/2}}\Big((\sum_{r=1}^{p}|b_{r-h+k}|)^m(\tau\sqrt{n})^m + (\sum_{r=1}^{p}\widetilde{X}_{rt}^2(s))^{m/2}\Big)\\
&\le C + \frac{C}{n}\sum_{r=1}^{p}\widetilde{X}_{rt}^m(s),\ 2 \le m \le 5, \qquad (6.80)
\end{aligned}
$$

where $\widetilde{X}_{rt}(s)$ is $X_{rt}(s) = s^{1/2}\sum_{\ell=0}^{\infty}b_\ell\xi_{r-\ell,t} + (1-s)^{1/2}\hat{X}_{rt}$ without the factor $\xi_{j-k,t} = v\eta$; and the last inequality utilizes the condition that $\sum_{\ell=0}^{\infty}|b_\ell| <$

∞. Note that $\widetilde{X}_{rt}$ is independent of η. This, together with (6.80) and the fact$E(|\widetilde{X}_{rt}|^5(s)) \le C\tau\sqrt{n}$, implies that

$$(6.78) \le C\tau \to 0.$$

If $(\mathbf{UY}(s))_{jt}$ in (6.77) is replaced by any $(\mathbf{UY}(s))_{k^{(i)}t}, i = 1, 2, 3, 4$, then an estimate similar to (6.78) also holds by exchanging the order of summation as follows

$$\sum_{j=1}^{p} \sum_{k=0}^{\infty} \sum_{k^{(i)}=j-k}^{p} b_{k^{(i)}-j+k} (\mathbf{UY}(s))_{k^{(i)}t} = \sum_{k=0}^{\infty} \sum_{k^{(i)}=1-k}^{p} (\mathbf{UY}(s))_{k^{(i)}t} \sum_{j=1}^{k^{(i)}+k} b_{k^{(i)}-j+k}. \tag{6.81}$$

Next consider the case when $(m_1 + m_2) \le 4$. By (6.74) and (6.80), one may verify that

$$\begin{aligned} &\frac{1}{n^{5/2}} \sum_{j,t=1}^{p,n} \int_0^1 E\Big[|\xi_{j-k,t}|^5 \Big((\mathbf{UY}(s))_{h_1 t}^{m_1} (\mathbf{U})_{k^{(2)}k^{(3)}}^{m_3} (\mathbf{U})_{k^{(4)}k^{(1)}}^{m_4} \\ &\qquad (\mathbf{Y}^T(s)\mathbf{UY}(s))_{tt}^{m_2} \Big) \big|_{\xi_{j-k,t}=v\xi_{j-k,t}} \Big] (1-v)^3 dv \\ &\le C\tau \to 0, \end{aligned}$$

where $m_i \ge 0, i = 3, 4$. Summarizing the above we may conclude that

$$|\varepsilon_3| \le C\tau \to 0. \tag{6.82}$$

Recall the definition of $T_{\ell\tau}$ in (6.67). Denote the analogues of $T_{\ell\tau}$ by T_ℓ with the truncated matrix $\mathbf{X}(s)$ replaced by the initial matrix $\mathbf{X}(s)$. Then write

$$T_{\ell\tau} = T_\ell + r_\ell, \quad \ell = 0, 1, 2, 3,$$

where

$$|r_\ell| \;\; \le \;\; \frac{s^{(\ell-1)/2}}{\ell! n^{(\ell+1)/2}} \sum_{j,t=1}^{p,n} |\kappa_{(\ell+1),\tau} - \kappa_{\ell+1}| \Big| \sum_{k=0}^{\infty} |b_k|$$

$$\begin{aligned}
&\sum_{k^{(\ell)},k^{(\ell-1)},\ldots,k^{(1)}}^{p} |b_{k^{(\ell)}-j+k}b_{k^{(\ell-1)}-j+k}\cdots b_{k^{(1)}-j+k}|E\Big(D_{k^{(\ell)}t}D_{k^{(\ell-1)}t}\cdots D_{k^{(1)}t}\Psi_{jt}(s)\Big)\Big| \\
\leq\ & \frac{C}{(\tau)^{3-\ell}}E\Big(|\xi_{11}|^4\cdot I(|\xi_{11}|>\tau\sqrt{n})\Big)\to 0,
\end{aligned}$$

where the last step uses (6.63), (6.75) and an estimate similar to (6.55).

By Lemma 25 below, (6.67), (6.82) and the facts that $T_0 = T_3 = 0$ (because $\kappa_1 = \kappa_4 = 0$) and that $T_1 = B_n$ (see (6.66)) we see

$$D_n = B_n + o(1).$$

This, together with (6.65), ensures (6.64) by the facts that $|a_n| = 1$ and that the function f is an analytic function. The proof of theorem is complete. □

Lemma 25.

$$\begin{aligned}
T_2 &= \frac{s^{1/2}\kappa_3}{2n^{3/2}}\sum_{j,t=1}^{p,n}\sum_{k=0}^{\infty}b_k\sum_{k^{(2)},k^{(1)}=j-k}^{p} b_{k^{(2)}-j+k}b_{k^{(1)}-j+k}E\Big(D_{k^{(2)}t}(s)D_{k^{(1)}t}(s)\Psi_{jt}(s)\Big) \\
&= o(1),
\end{aligned}$$

as $n\to\infty$.

Proof. It follows from (6.69)-(6.72) that the expansion of $D_{k^{(1)}t}\Psi_{jt}(s)$ is

$$\begin{aligned}
D_{k^{(1)}t}\Psi_{jt}(s) &= e_n(s,x)\Big[-2x\int\theta\hat f(\theta)\big(\mathbf{UY}(s)\big)_{k^{(1)}t}d\theta\big(\mathbf{UY}(s)\big)_{jt} + U_{k^{(1)}j} \\
&\quad +i\big(\mathbf{Y}'(s)\mathbf{UY}(s)\big)_{tt} * U_{k^{(1)}j} + i\big(\mathbf{UY}(s)\big)_{jt} * \big(\mathbf{UY}(s)\big)_{k^{(1)}t}\Big].
\end{aligned} \tag{6.83}$$

By (6.69)-(6.72) we can further obtain the expansion of $D_{k^{(2)}t}D_{k^{(1)}t}\Psi_{jt}(s)$. Since such an expansion is complicated we do not list it here. However each term of the expansion is a constant multiple of one of the following forms

$$A_1 = \big(\mathbf{UY}(s)\big)_{h_1t}\circ U_{h_2h_3}e_n(s,x),$$

$$A_2 = \big(\mathbf{UY}(s)\big)_{h_1t} \circ \big(\mathbf{Y}^T(s)\mathbf{UY}(s)\big)_{tt} \circ U_{h_2h_3}e_n(s,x),$$
$$A_3 = \big(\mathbf{UY}(s)\big)_{k^{(2)}t} \circ \big(\mathbf{UY}(s)\big)_{k^{(1)}t} \circ \big(\mathbf{UY}(s)\big)_{jt}e_n(s,x),$$

where "$\circ$" denotes a product or a convolution; $h_i = k^{(2)}, k^{(1)}$ or j with $i = 1,2,3$ and $h_1 \neq h_2 \neq h_3$. In view of this it then suffices to prove that

$$T_{2i} = \frac{1}{n^{3/2}} \sum_{j,t=1}^{p,n} \sum_{k=0}^{\infty} b_k \sum_{k^{(1)},k^{(2)}=j-k}^{p} b_{k^{(1)}-j+k} b_{k^{(2)}-j+k} EA_i = o(1), \quad i = 1,2,3.$$

Without loss of generality, we below consider $h_1 = j$, $h_2 = k^{(1)}$ and $h_3 = k^{(2)}$ only, otherwise one may first exchange the order of the summation as in (6.81) when necessary and then proceed as follows. Consider T_{22}. Note that the fact that $\mathbf{UY}(s)\mathbf{Y}^T(s) = \mathbf{Y}(s)\mathbf{Y}^T(s)\mathbf{U}$. A simple calculation then yields

$$\begin{aligned} E\Big[\sum_{t=1}^{n} |\sum_{j=1}^{p} \big(\mathbf{UY}(s)\big)_{jt}|^2\Big] &= E\Big[\sum_{j_1,j_2=1}^{p} \big(\mathbf{UY}(s)\mathbf{Y}^T(s)\bar{\mathbf{U}}^T\big)_{j_1j_2}\Big] \\ = \ E\Big[\sum_{j_1,j_2=1}^{p} \big(\mathbf{Y}(s)\mathbf{Y}^T(s)\big)_{j_1j_2}\Big] &= O(n). \end{aligned} \tag{6.84}$$

By the Schwartz inequality, (6.73) and (6.84), we have

$$\begin{aligned} |T_{22}|^2 &\le \frac{C}{n^3} E\Big[\sum_{t=1}^{n} |\big(\mathbf{Y}^T(s)\mathbf{UY}(s)\big)_{tt}|^2\Big] E\Big[\sum_{t=1}^{n} |\sum_{j=1}^{p} \big(\mathbf{UY}(s)\big)_{jt}|^2\Big] \\ &\le \frac{C}{n^2} E\Big[\sum_{t=1}^{n} \big(\sum_{r=1}^{p} Y_{rt}^2(s)\big)^2\Big] = O(\frac{1}{n}). \end{aligned} \tag{6.85}$$

This argument also works for T_{21} and T_{23} and we ignore the details here. Therefore

$$T_2 = O(\frac{1}{\sqrt{n}}).$$

□

Proof of Theorem 4. Set

$$X_n^{(i)} = \int x^i d\widetilde{G}_n(x), \quad i = 1, 2; \qquad \widetilde{\mathbf{\Omega}}^{-1} = \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix}.$$

Furthermore, under $\mathbb{H}_1$, $X_n^{(i)}$, $i = 1, 2$, can be written as

$$X_n^{(i)} = Y_1^{(i)} + Y_2^{(i)}, \tag{6.86}$$

where

$$Y_1^{(i)} = p \int x^i d\Big(F_{\mathbb{H}_1}^{\mathbf{S}^{(1)}}(x) - F_{\mathbb{H}_0}^{\mathbf{S}^{(1)}}(x)\Big) + p \int x^i d\Big(F_{\mathbb{H}_0}^{\mathbf{S}^{(2)}}(x) - F_{\mathbb{H}_1}^{\mathbf{S}^{(2)}}(x)\Big)$$

and

$$Y_2^{(i)} = p \int x^i d\Big(F_{\mathbb{H}_0}^{\mathbf{S}^{(1)}}(x) - F_{\mathbb{H}_0}^{\mathbf{S}^{(2)}}(x)\Big).$$

From (6.86) we have

$$a_{11}(X_n^{(1)})^2 + a_{22}(X_n^{(2)})^2 + 2a_{12}X_n^{(1)}X_n^{(2)} = W_1 + W_2 + W_3,$$

where

$$W_1 = a_{11}(Y_2^{(1)})^2 + a_{22}(Y_2^{(2)})^2 + 2a_{12}Y_2^{(1)}Y_2^{(2)},$$
$$W_2 = a_{11}(Y_1^{(1)})^2 + a_{22}(Y_1^{(2)})^2 + 2a_{12}Y_1^{(1)}Y_1^{(2)}$$

and

$$W_3 = 2a_{11}Y_1^{(1)}Y_2^{(1)} + 2a_{22}Y_1^{(2)}Y_2^{(2)} + 2a_{12}[Y_2^{(1)}Y_1^{(2)} + Y_1^{(1)}Y_2^{(2)}].$$

Note that W_1 converges in distribution to $\chi^2(2)$ by Theorem 11 or Proposition 1. Also $Y_2^{(i)}, i = 1, 2$ converge in distribution to Gaussian distribution by Theorem 11 or Proposition 1. We next prove that $W_2 \to \infty$ in probability while $W_3 = o_p(W_2)$. By Assumption (3.15) $Y_1^{(1)} \to \infty$ or $Y_1^{(2)} \to \infty$ in probability (we would point out that $Y_1^{(i)} \geq 0$). If $Y_1^{(1)} \to \infty$ and

$\limsup Y_1^{(2)} < \infty$ in probability, then $W_2 \to +\infty$ in probability. It is then easy to verify that $W_3 = o_p(W_2)$. This argument also applies to the case when $Y_1^{(2)} \to \infty$ and $\limsup Y_1^{(1)} < \infty$ in probability. If $Y_1^{(1)} \to \infty$ and $Y_1^{(2)} \to \infty$ in probability then by Holder's inequality

$$W_2 \geq 2(\sqrt{a_{11}}\sqrt{a_{22}} + a_{12})Y_1^{(1)}Y_1^{(2)} \to +\infty$$

in probability, because

$$det(\widetilde{\mathbf{\Omega}}^{-1}) = a_{11}a_{22} - a_{12}^2 > 0.$$

It is then easy to verify that $W_3 = o_p(W_2)$ in this case.

In view of the above we conclude from the definition of L_n that

$$\begin{aligned} P(L_n > \gamma_{1-\alpha}|\mathbb{H}_1) &= P\Big((X_n^{(1)}, X_n^{(2)})\widetilde{\mathbf{\Omega}}^{-1}\begin{pmatrix} X_n^{(1)} \\ X_n^{(2)} \end{pmatrix} > \gamma_{1-\alpha}\Big|\mathbb{H}_1\Big) \\ &= P\Big((a_{11}(X_n^{(1)})^2 + a_{22}(X_n^{(2)})^2 + 2a_{12}X_n^{(1)}X_n^{(2)} > \gamma_{1-\alpha}\Big|\mathbb{H}_1\Big) \\ &= P\Big(W_1 + W_2 + W_3 > \sqrt{\gamma_{1-\alpha}}\Big|\mathbb{H}_1\Big) \to 1, \quad as \quad n \to \infty. \qquad (6.87) \end{aligned}$$

$\square$

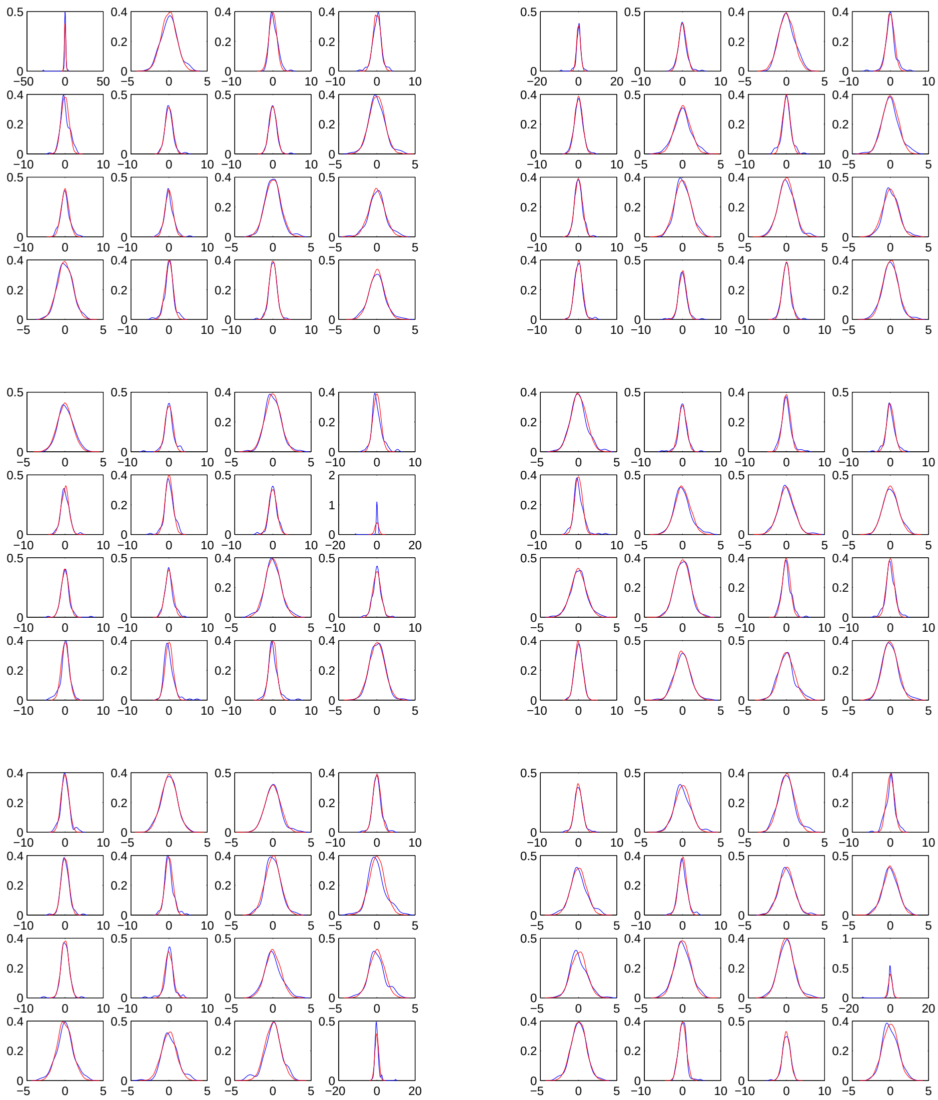
Table 6.1: *Empirical sizes of the proposed test L_n at significant level* 0.05 *for n time series generated from DGP 1 with $\theta_1 = 0.8$ in model (6.29).*

					p					
n	50	100	150	200	250	300	350	400	450	500
					Empirical sizes					
50	0.023	0.022	0.027	0.021	0.019	0.021	0.023	0.018	0.022	0.020
100	0.037	0.040	0.035	0.039	0.036	0.031	0.034	0.030	0.029	0.025
150	0.042	0.040	0.039	0.043	0.041	0.038	0.039	0.034	0.034	0.031
200	0.039	0.043	0.046	0.045	0.043	0.045	0.042	0.041	0.038	0.040
250	0.040	0.045	0.048	0.044	0.045	0.044	0.042	0.040	0.041	0.041
300	0.037	0.041	0.049	0.054	0.052	0.048	0.043	0.041	0.039	0.045
350	0.041	0.048	0.053	0.049	0.055	0.052	0.049	0.047	0.045	0.047
400	0.038	0.041	0.047	0.052	0.052	0.048	0.053	0.051	0.046	0.046
450	0.035	0.037	0.040	0.046	0.050	0.055	0.059	0.060	0.058	0.055
500	0.032	0.035	0.035	0.040	0.047	0.052	0.054	0.057	0.054	0.058

Table 6.2: *Bootstrap sizes of the proposed test L_n at significant level* 0.05 *for n time series generated from DGP 1 with $\theta_1 = 0.8$ in model (6.29).*

					p					
n	50	100	150	200	250	300	350	400	450	500
					Empirical sizes					
50	0.039	0.041	0.038	0.037	0.040	0.037	0.039	0.035	0.037	0.037
100	0.043	0.042	0.040	0.039	0.039	0.040	0.042	0.040	0.046	0.041
150	0.046	0.042	0.048	0.043	0.045	0.047	0.040	0.049	0.045	0.041
200	0.042	0.047	0.043	0.046	0.049	0.051	0.048	0.048	0.046	0.050
250	0.052	0.050	0.046	0.054	0.048	0.051	0.050	0.049	0.052	0.055
300	0.048	0.053	0.055	0.052	0.056	0.057	0.053	0.051	0.049	0.054
350	0.046	0.054	0.052	0.054	0.050	0.051	0.050	0.048	0.053	0.054
400	0.043	0.048	0.046	0.051	0.054	0.051	0.052	0.054	0.049	0.052
450	0.046	0.052	0.048	0.049	0.053	0.050	0.054	0.055	0.053	0.052
500	0.042	0.046	0.045	0.047	0.050	0.053	0.055	0.052	0.055	0.054

Figure 6.1: *Graphs of smoothed density function of the transformed data vs standard normal distribution*



*These graphs contain the empirical density functions of the transformed data for all 96 stocks used in our empirical application. The blue line is the smoothed density function of the transformed data for one stock and the red graph is standard normal density function.

Table 6.3: *Empirical sizes of the proposed test L_n at significant level 0.05 for n time series generated from DGP 2.*

					p					
n	50	100	150	200	250	300	350	400	450	500
					Empirical sizes					
50	0.037	0.035	0.030	0.031	0.034	0.029	0.032	0.030	0.030	0.028
100	0.040	0.043	0.045	0.042	0.040	0.037	0.040	0.035	0.035	0.032
150	0.042	0.043	0.048	0.043	0.045	0.040	0.040	0.039	0.037	0.037
200	0.041	0.046	0.051	0.045	0.049	0.053	0.047	0.043	0.040	0.041
250	0.044	0.049	0.053	0.051	0.047	0.052	0.044	0.047	0.045	0.045
300	0.040	0.042	0.046	0.046	0.050	0.055	0.047	0.044	0.046	0.048
350	0.038	0.046	0.049	0.053	0.054	0.051	0.053	0.046	0.045	0.046
400	0.039	0.041	0.043	0.047	0.055	0.051	0.058	0.056	0.051	0.053
450	0.037	0.039	0.039	0.043	0.049	0.053	0.055	0.048	0.050	0.048
500	0.037	0.035	0.042	0.047	0.045	0.055	0.047	0.054	0.052	0.055

*The data are simulated from model (6.30). $\phi_1 = 0.2$.

Table 6.4: *Bootstrap sizes of the proposed test L_n at significant level 0.05 for n time series generated from DGP 2.*

					p					
n	50	100	150	200	250	300	350	400	450	500
					Empirical sizes					
50	0.041	0.039	0.040	0.042	0.038	0.037	0.035	0.038	0.036	0.038
100	0.044	0.047	0.043	0.040	0.042	0.041	0.040	0.045	0.039	0.040
150	0.046	0.048	0.049	0.051	0.048	0.047	0.053	0.055	0.052	0.047
200	0.043	0.050	0.053	0.049	0.052	0.048	0.049	0.054	0.051	0.047
250	0.045	0.052	0.054	0.050	0.050	0.054	0.051	0.048	0.047	0.052
300	0.044	0.048	0.051	0.047	0.048	0.050	0.046	0.053	0.052	0.047
350	0.046	0.051	0.047	0.054	0.052	0.050	0.051	0.051	0.052	0.049
400	0.042	0.047	0.052	0.049	0.051	0.050	0.055	0.050	0.054	0.052
450	0.045	0.049	0.053	0.053	0.050	0.051	0.054	0.049	0.049	0.051
500	0.042	0.035	0.045	0.049	0.047	0.050	0.051	0.047	0.050	0.053

*The data are simulated from model (6.30). $\phi_1 = 0.2$.

Table 6.5: *Empirical sizes of the proposed test L_n at significant level* 0.05 *for n time series generated from DGP 3.*

					p					
n	50	100	150	200	250	300	350	400	450	500
					Empirical sizes					
50	0.036	0.039	0.035	0.030	0.035	0.031	0.036	0.035	0.030	0.029
100	0.040	0.042	0.040	0.039	0.042	0.045	0.040	0.037	0.039	0.034
150	0.043	0.048	0.051	0.045	0.047	0.051	0.043	0.042	0.039	0.040
200	0.040	0.046	0.055	0.052	0.047	0.055	0.049	0.045	0.044	0.044
250	0.039	0.041	0.048	0.053	0.051	0.057	0.053	0.055	0.058	0.059
300	0.042	0.045	0.045	0.050	0.048	0.055	0.054	0.048	0.052	0.056
350	0.037	0.042	0.047	0.052	0.050	0.051	0.047	0.045	0.048	0.052
400	0.035	0.045	0.052	0.047	0.051	0.048	0.053	0.054	0.052	0.050
450	0.038	0.041	0.045	0.046	0.045	0.047	0.049	0.052	0.050	0.048
500	0.039	0.043	0.047	0.052	0.048	0.048	0.051	0.055	0.046	0.051

*The data are simulated from model (6.31). $\theta_1 = 0.8$ and $\phi_1 = 0.2$.

Table 6.6: *Bootstrap sizes of the proposed test L_n at significant level* 0.05 *for n time series generated from DGP 3.*

					p					
n	50	100	150	200	250	300	350	400	450	500
					Empirical sizes					
50	0.040	0.042	0.045	0.037	0.035	0.034	0.032	0.036	0.034	0.038
100	0.043	0.045	0.048	0.046	0.039	0.042	0.046	0.047	0.044	0.044
150	0.047	0.052	0.049	0.055	0.057	0.050	0.046	0.045	0.040	0.043
200	0.045	0.048	0.053	0.050	0.051	0.052	0.055	0.055	0.056	0.052
250	0.047	0.048	0.051	0.055	0.054	0.056	0.056	0.052	0.054	0.053
300	0.045	0.049	0.050	0.053	0.050	0.053	0.058	0.054	0.053	0.055
350	0.048	0.053	0.057	0.053	0.052	0.050	0.049	0.047	0.053	0.055
400	0.045	0.048	0.050	0.055	0.051	0.054	0.053	0.056	0.055	0.056
450	0.042	0.044	0.051	0.048	0.053	0.054	0.050	0.051	0.053	0.053
500	0.045	0.048	0.050	0.054	0.050	0.049	0.050	0.053	0.053	0.054

*The data are simulated from model (6.31). $\theta_1 = 0.8$ and $\phi_1 = 0.2$.

Table 6.7: *Empirical powers of the proposed test L_n at significant level* 0.05 *for n time series with MA(1) type dependent structure.*

			p			
n	50	100	200	300	350	400
			Empirical sizes			
50	0.210	0.279	0.429	0.445	0.505	0.614
100	0.469	0.513	0.725	0.779	0.794	0.805
200	0.712	0.793	0.814	0.889	0.903	0.921
300	0.787	0.899	0.932	0.921	0.945	0.962
350	0.823	0.956	0.983	0.972	0.989	0.994
400	0.921	0.993	0.994	0.999	1.000.	0.999

*The data are simulated from model (6.32). Each time series $\mathbf{x}_i$ is generated from DGP 3 with $\theta_1 = 0.8$ and $\phi_1 = 0.2$. In (6.32), we take $\theta = 0.8$.

Table 6.8: *Empirical powers of the proposed test L_n at significant level* 0.05 *for n time series with AR(1) type dependent structure.*

			p			
n	50	100	200	300	350	400
			Empirical sizes			
50	0.656	0.720	0.714	0.801	0.823	0.842
100	0.792	0.824	0.846	0.891	0.907	0.917
200	0.858	0.889	0.922	0.926	0.954	0.985
300	0.901	0.935	0.958	0.982	0.992	0.0.993
350	0.892.	0.970	0.992	0.995	0.999	0.999
400	0.941	0.989	0.999	1.000	1.000	1.000

*The data are simulated from model (6.33). Each time series $\mathbf{x}_i$ is generated from DGP 3 with $\theta_1 = 0.8$ and $\phi_1 = 0.2$. In (6.33), we take $\phi = 0.2$.

Table 6.9: *Empirical powers of the proposed test L_n at significant level* 0.05 *for n time series with ARMA(1,1) type dependent structure.*

	p					
n	50	100	200	300	350	400
			Empirical sizes			
50	0.592	0.613	0.654	0.719	0.746	0.758
100	0.713	0.748	0.855	0.891	0.904	0.909
200	0.776	0.833	0.892	0.903	0.955	0.968
300	0.856	0.901	0.963	0.981	0.982	0.993
350	0.902	0.946	0.980	0.999	0.998	1.000
400	0.933	0.951	0.991	1.000	1.000	1.000

*The data are simulated from model (6.34). Each time series $\mathbf{x}_i$ is generated from DGP 3 with $\theta_1 = 0.8$ and $\phi_1 = 0.2$. In (6.34), we take $\theta = 0.8$ and $\phi = 0.2$.

Table 6.10: *Empirical powers of the proposed test L_n at* 0.05 *significance level for the dynamic factor model.*

(p, n)	r=1	r=2	r=3	r=4
(50,50)	0.342	0.553	0.889	0.950
(50,100)	0.358	0.622	0.949	0.968
(100,100)	0.403	0.685	0.972	0.984
(200,100)	0.526	0.741	0.983	0.998
(300,200)	0.557	0.763	0.987	1.000
(200,300)	0.637	0.785	0.983	0.999
(100,200)	0.656	0.791	0.988	0.999
(200,400)	0.671	0.785	0.990	0.999
(400,200)	0.685	0.768	0.991	1.000
(100,300)	0.682	0.784	0.980	1.000
(300,100)	0.701	0.782	0.989	1.000

*The data are simulated from model (6.35) and (6.36).

Table 6.11: *Empirical powers of the proposed test L_n at significant level* 0.05 *for n random vectors with common random dependence.*

	p					
n	50	70	90	110	130	150
			Empirical sizes			
50	0.894	0.920	0.923	0.942	0.966	0.959
70	0.910	0.948	0.955	0.975	0.980	0.995
90	0.960	0.958	0.969	0.984	0.989	0.999
110	0.941	0.956	0.984	0.992	0.994	1.000
130	0.930	0.972	0.990	0.995	0.999	1.000
150	0.952	0.980	0.989	1.000	1.000	1.000

*The data are simulated from model (6.37).

Table 6.12: *Empirical powers of the proposed test L_n at* 0.05 *significance level for ARCH(1) dependent type.*

(p, n)	$(0.9, 0.1)$	$(0.8, 0.2)$	$(0.7, 0.3)$	$(0.6, 0.4)$	$(0.5, 0.5)$
(50,50)	0.257	0.396	0.425	0.605	0.732
(50,100)	0.597	0.879	0.890	0.899	0.998
(100,200)	0.727	0.978	0.997	0.998	0.999
(200,200)	0.738	0.990	0.999	1.000	1.000
(200,300)	0.828	0.992	0.998	1.000	1.000
(200,400)	0.887	0.997	1.000	1.000	1.000
(300,400)	0.906	1.000	1.000	1.000	1.000
(400,400)	0.922	1.000	1.000	1.000	1.000

*The data are simulated from model (6.38).

Table 6.13: L_n *under various scenarios for* 5 *randomly selected samples*

(n,p)	5%critical values	1	2	3	4	5
(60,30)	[0, 5.99]	395.44	462.76	481.85	443.79	481.46
(70,35)	[0, 5.99]	595.84	642.31	620.96	592.63	632.87
(90,40)	[0, 5.99]	902.55	928.89	1318.6	1173.9	914.25

*The critical values are the corresponding quantiles of the limiting distribution $\chi^2(2)$ of the statistic L_n for $(n,p) = (60,30), (70,35), (90,40)$ respectively.

Chapter 7

Discussion and Future Research

7.1 Conclusion

This research work develops some independence tests for high dimensional data by the tool of large dimensional random matrix theory. Two types of independence tests are considered.

For the independence test between two high dimensional random vectors, we propose linear spectral statistics of classical and regularized canonical correlation matrices respectively. Moreover, the LSD's and CLT's for these matrices are developed by discussing the Gaussian case and the general case respectively.

Regard to the independence test between a large number of high dimensional random vectors, we have talked about three cases. When the components of each random vector are i.i.d., we propose a linear spectral statistic by using the characteristic function of the ESD of the sample covariance matrix. When each random vector has a linear dependent structure or is a covariance stationary process, the first two moments of the ESD of the sample covariance matrix are utilized to do the independence test. As an independent contribution in random matrix theory, the LSD and CLT are developed for the sample covariance matrix whose columns are independent covariance stationary processes.

For each independence test, some simulation results are provided to show the effectiveness of our proposed test statistics.

7.2 Future Research

In high dimensional data analysis, more complicated data appear rather than just i.i.d. or stationary processes. We will use large dimensional random matrix theory to investigate more practical data which appear frequently in economic or finance, such as co-integrated time series, etc.

As this research work focuses on independence test for high dimensional data, some other problems arise once the null hypothesis is rejected. As more and more sections are grouped together, the appearance of cross-sectional dependence is quite natural and common. In view of this, measuring the degree of cross-sectional dependence is more important than testing its presence. A natural question is how to model dependence between a large number of random vectors? Modeling and Estimation of dependence are one of our main future research work.

Bibliography

Anderson, T. “An introduction to multivariate statistical analysis.” *New York: Wiley*, 2th edtion (1984).

Bai, J. and Ng, S. “Determine the number of factors in approximate factor models.” *Econometrica*, 70(1):191–221 (2002).

Bai, Z. and Saranadasa, H. “Effect of high dimension comparison of significance tests for a high dimensional two sample problem.” *Statist. Sinica*, 6:311–329 (1996).

Bai, Z. and Silverstein, J. “No eigenvalues outside the support of the limiting spectral distribution of large dimensional random matrices.” *Ann. Probab.*, 26:316–345 (1998).

—. “CLT for linear spectral statistics of large-dimensional sample covariance matrices.” *Ann. Probab.*, 32(1A):553–605 (2004).

—. “Spectral analysis of large dimensional random matrices.” *Springer*, 2nd edition (2009).

Billingsley, P. “Convergence of probability measures.” *Wiley*, 2nd edition (1999).

Birke, M. and Dette, H. "A note on testing the covariance matrix for large dimension." *Statistics Probability Letters.*, 74:281–289 (2005).

Burkholder, D. "Distribution function inequalities for martingales." *Ann. Probab.*, 1(1):19–42 (1973).

Chatterjee, S. "A generalization of the Lindeberg principle." *Ann. Probab.*, 34(6):2061–2076 (2006).

Chen, J., Gao, J., and Li, D. "A new diagnostic test for cross–sectional uncorrelatedness in nonparametric panel data models." *Econometric Theory*, 28:1144–1163 (2012).

Dozier, R. and Silverstein, J. "On the empirical distribution of eigenvalues of large dimensional information-plus-noise type matrices." *J. Multivariate Anal.*, 98(4):678–694 (2007).

∅. Ryan and Debbah, M. "Asymptotic behaviour of random vandermonde matrices with entries on the unit circle." *IEEE Trans. on Information Theory*, 55(7):3115–3148 (2009).

Fan, J. and Fan, Y. "High dimensional classification uning features annealed independence rules." *Annal. Statist.*, 36(6):2605–2637 (2008).

Fan, J., Guo, S., and Hao, N. "Variance estimation using refitted cross-validation in ultrahigh dimensional regression." *J. R. Statist. Soc. B*, 74:37–65 (2012).

Fan, J. and Lv, J. "A selective overview of variable selection in high dimensional feature space." *Statist. Sinica*, 20(1):101–148 (2010).

Frees, E. "Assessing Cross-sectional Correlation in Panel Data." *Journal of Econometrics*, 69:393–414 (1995).

Fujikoshi, Y., Ulyanov, V., and Shimizu, R. "Multivariate statistics: high dimensional and large sample approximations." *Wiley*, 2nd edition (2010).

Geweke, J. "A comparison of tests of independence of two covariance stationary time series." *J. Am. Statist. Assoc.*, 76:363–373 (1993).

Gray, R. "Toeplitz and Circulant Matrices: A review." *http://ee.stanford.edu/ gray/toeplitz.pdf* (2009).

Guo, C. and Higham, N. "A Schur-Newton method for the matrix p-th root and its inverse." *SIAM Journal on Matrix Analysis and Applications*, 28(3):788–804 (2006).

Haugh, L. "Checking the independence of two covariance stationary time series: a univariate residual cross correlation approach." *J. Am. Statist. Assoc.*, 71:378–385 (1976).

Hoeffding, W. "Probability inequalities for sums of bounded random variables." *J. Amer. Statist. Assoc.*, 58(301):13–30 (1963).

Hong, Y. "Testing for independence between two covariance stationary time series." *Biometrika*, 83(3):615–625 (1996).

—. "Hypothesis testing in time series via the empirical characteristic function: a generalized spectral density approach." *J. Am. Statist. Ass.*, 94(448):1201–1220 (1999).

—. "Generalized spectral tests for serial dependence." *J. R. Statist. Soc. B*, 62(3):557–574 (2000).

—. "Asymptotic distribution theory for nonparametric entropy measures of serial dependence." *Econometrica*, 73(3):837–901 (2005).

Hsiao, C., Pesaran, M., and Pick, A. "Diagnostic tests of cross–sectional independence for limited dependent variable panel data models." *Oxford Bulletin of Economics and Statistics* (2009).

Huang, J., Horowitz, J., and Ma, S. "Asymptotic properties of bridge estimators in sparse high-dimensional regression models." *Ann. Statist.*, 36(2):587–613 (2008).

Jiang, T. "The limiting distributions of eigenvalues of sample correlation matrices." *Sankhya*, 66(1):35–48 (2004).

Johnstone, I. "Multivariate analysis and Jacobi ensembles: largest eigenvalue, tracy-widom limits and rates of convergence." *Ann. Statist.*, 36(6):2638–2716 (2008).

Jonsson, D. "Some limit theorems of the eigenvalues of a sample convariance matrix." *J. Multivariate Anal.*, 12(1):1–38 (1982).

Kuan, C. and Lee, W. "A new test for the martingale difference hypothesis." *The Annals of Applied Statistics*, 8(4):1–26 (2004).

Lindeberg, J. "Eine neue herleitung des exponentialgesetzes in der wahrscheinlichkeitsrechnung." *Math. Z.*, 15:211–225 (1922).

Lytova, A. and Pastur, L. "Central limit theorem for linear eigenvalue statistics of random matrices with independent entries." *Ann. Probab.*, 37(5):1778–1840 (2009).

Marcenko, V. and Pastur, L. "Distribution of eigenvalues for some sets of random matrices." *Math. USSR-Sb.*, 1(4):457–483 (1967).

Mardia, K., Kent, J., and Bibby, J. "Multivariate Analysis." *Academic Press*, 1st edition (1979).

Mundlak, Y. "On the pooling of time series and cross section data." *Econometrica*, 46:69–85 (1978).

Pan, G. "Strong convergence of the empirical distribution of eigenvalues of sample covariance matrices with a perturbation matrix." *J. Multivariate Anal.*, 101:1330–1338 (2010).

—. "Comparison between two types of large sample covariance matrices." *Annales de l'Institut Henri Poincaré*, accepted (2012).

Pan, G. and Zhou, W. "Central limit theorem for signal-to-interference ratio of reduced rank linear receiver." *Ann. Appl. Probab.*, 18(3):1232–1270 (2008).

Rudelson, M. and Vershynin, R. "The smallest singular value of a random rectangular matrix." *http://www.math.missouri.edu/ rudelson/papers/rv-rectangular-matrices.pdf* (2011).

Sarafidis, V., Yamagata, T., and Robertson, D. "A Test of cross section dependence for a linear dynamic panel model with regressors." *Journal of Econometrics*, 148(2):149–161 (2009).

Silverstein, J. and Bai, Z. "On the empirical distribution of eigenvalues of a class of large dimensional random matrices." *J. Multivariate Anal.*, 54:175–192 (1995).

Silverstein, J. and Choi, S. "Analysis of the limiting spectral distribution of large dimensional random matrices." *J. Multivariate Anal.*, 54:295–309 (1995).

Timm, N. "Applied multivariate analysis." *Springer* (2001).

Wachter, K. "The limiting empirical measure of multiple discriminant ratios." *Ann. Statist.*, 8(5):937–957 (1980).

Wilks, S. "On the independence of k sets of normally distributed statistical variables." *Econometrica*, 3:309–326 (1935).

Xiao, H. and Zhou, W. "Almost sure limit of the smallest eigenvalue of some sample correlation matrices." *Journal of Theoretical Probability*, 23(1):1–20 (2010).

Yao, J. "A note on a MarcenkoPastur type theorem for time series." *Statist. and Probab. Letters.*, 82(1):22–28 (2012).

Yin, Y. "Limiting spectral distribution for a class of random matrices." *J. Multivariate Anal.*, 20(1):50–68 (1986).

Zhang, L. "Spectral analysis of large dimensional random matrices." *National University of Singapore PHD Thesis* (2006).

Zheng, S. "Central limit theorems for linear spectral statistics of large dimensional F-matrix." *Annales de l'Institut Henri Poincaré*, 48(2):444–476 (2012).