Independence Test For High Dimensional Data Based On Regularized Canonical Correlation Coefficients

Yanrong Yang and Guangming Pan *

Abstract

This paper proposes a new statistic to test independence between two high dimensional random vectors $\mathbf{x} : p_1 \times 1$ and $\mathbf{y} : p_2 \times 1$. The proposed statistic is based on the sum of regularized sample canonical correlation coefficients of $\mathbf{x}$ and $\mathbf{y}$. The asymptotic distribution of the statistic under the null hypothesis is established as a corollary of general central limit theorems (CLT) for the linear statistics of classical and regularized sample canonical correlation coefficients when p_1 and p_2 are both comparable to the sample size n. As applications of the developed independence test, various types of dependent structures, such as factor models, ARCH models and a general uncorrelated but dependent case etc., are investigated by simulations. As an empirical application, cross-sectional dependence of daily stock returns of companies between different sections in New York Stock Exchange (NYSE) is detected by the proposed test.

Keywords: Canonical correlation coefficients; Independence test; Empirical spectral distribution; Large dimensional random matrix theory; Stieltjes transform; Central limit theorem.

1 Introduction

A prominent feature of data collection nowadays is that the number of variables is comparable with the sample size. 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, Guo and Hao (2012), Huang, Horowitz and Ma (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

*Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore 637371. This work is partially supported by the Ministry of Education, Singapore, under grant # ARC 14/11.

between the groups of variables. In other words, can we claim that the groups are mutually independent so that further statistical 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. Relying on the asymptotic theory of sample canonical correlation coefficients, this paper proposes a new statistic to test independence between two high dimensional random vectors.

Specifically, the aim is to test the hypothesis

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

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

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

where $\mathbf{\Sigma_{xx}}$ and $\mathbf{\Sigma_{yy}}$ are the population covariance matrices for $\mathbf{x}$ and $\mathbf{y}$ respectively and $\mathbf{\Sigma_{xy}}$ 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

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

About this point, one may refer to page 284 of Mardia (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 $\hat{\mathbf{\Sigma}}_{\mathbf{xx}}$, $\hat{\mathbf{\Sigma}}_{\mathbf{xy}}$ and $\hat{\mathbf{\Sigma}}_{\mathbf{yy}}$ 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 of sample covariance matrices to replace population covariance matrices to test (10.31) in some cases. We demonstrate such an example in Section 6.3.

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

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

where

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

We also call $\mathbf{A_{xx}}$, $\mathbf{A_{yy}}$ and $\mathbf{A_{xy}}$ 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 'curse 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), when $p_1 < n, p_2 < n$, one can see that $r_1^2, r_2^2, \ldots, r_{p_1}^2$ are the eigenvalues of the matrix

$$\mathbf{S_{xy}} = \mathbf{A_{xx}^{-1}}\mathbf{A_{xy}}\mathbf{A_{yy}^{-1}}\mathbf{A_{xy}^{T}}. \tag{1.5}$$

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

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

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 $\mathbf{A_{yy}^{-}}$ denotes the Moore-Penrose pseudoinverse matrix of $\mathbf{A_{yy}}$. Hence $\mathbf{T_{xy}}$ is well defined even in the case of $p_1, p_2 \geq n$. Moreover $\mathbf{T_{xy}}$ reduces to $\mathbf{S_{xy}}$ when p_1, p_2 are both smaller than n and $t = 0$.

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

$$\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \to \begin{pmatrix} \boldsymbol{\xi} \\ \boldsymbol{\eta} \end{pmatrix} = \begin{pmatrix} \mathcal{A}' & \mathbf{0} \\ \mathbf{0} & \mathcal{B}' \end{pmatrix}\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \tag{1.7}$$

such that

$$\begin{pmatrix} \mathcal{A}' & \mathbf{0} \\ \mathbf{0} & \mathcal{B}' \end{pmatrix}\begin{pmatrix} \mathbf{\Sigma_{xx}} & \mathbf{\Sigma_{xy}} \\ \mathbf{\Sigma_{yx}} & \mathbf{\Sigma_{yy}} \end{pmatrix}\begin{pmatrix} \mathcal{A} & \mathbf{0} \\ \mathbf{0} & \mathcal{B} \end{pmatrix} = \begin{pmatrix} \mathbf{I}_{p_1} & \mathcal{P} \\ \mathcal{P}' & \mathbf{I}_{p_2} \end{pmatrix}, \tag{1.8}$$

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

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

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

$$Var\begin{pmatrix} \boldsymbol{\xi} \\ \boldsymbol{\eta} \end{pmatrix} = \begin{pmatrix} \mathbf{I}_{p_1} & \mathcal{P} \\ \mathcal{P}' & \mathbf{I}_{p_2} \end{pmatrix}. \tag{1.9}$$

In view of this, if the joint distribution of $\mathbf{x}$ and $\mathbf{y}$ is Gaussian, independence between $\mathbf{x}$ and $\mathbf{y}$ is equivalent to asserting that the population canonical correlations all vanish: $\gamma_1 = \cdots = \gamma_{p_1} = 0$. Details can be referred to Chapter 11 of Fujikoshi et. (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 (10.31) 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 could be 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 order to apply it to conduct tests we further propose two modified statistics by either dividing the total samples into two groups or estimating the population covariance matrix of $\mathbf{x}$ in a framework of sparsity.

In addition to proposing a statistic for testing (10.31), another contribution of this paper 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 the CLT to 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 to that for $\mathbf{S_{xy}}$ except we first have to develop CLT of perturbed sample covariance matrices in the supplement material for establishing CLT of the matrix $\mathbf{T_{xy}}$ when the entries of $\mathbf{X}$ are Gaussian.

Here we would point out some works on canonical correlation coefficients under the high dimensional scenario. In the high dimensional case Wachter (1980) investigated the limit of the ESD 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.

The remainder of the paper is organized as follows. Section 2 proposes a new test statistic for (10.31) based on large dimensional random matrix theory and contains the main results. Two modified statistics are further provided in Section 3. Section 4 provides the powers of the test statistics. Two examples as statistical inference of independence test are explored in Section 5. Simulation results for several kinds of dependent structures are provided in Section 6. An empirical analysis of cross-sectional dependence of daily stock returns of companies from two different sections in New York Stock Exchange (NYSE) is investigated by the proposed independence test in Section 7. The proof of Theorem 1 is given in Appendix A in Section 8. Some useful lemmas and proofs of Theorems 2-7 are relegated to Appendix B while one theorem about the CLT of a sample covariance matrix plus a perturbation matrix is provided in the supplementary material.

2 Methodology and theory

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

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

Remark 1. *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 ESD of any $n \times n$ matrix $\mathbf{A}$ with real eigenvalues $\mu_1 \leq \mu_2 \leq \cdots \leq \mu_n$ by

$$F^{\mathbf{A}}(x) = \frac{1}{n}\#\{i : \mu_i \leq x\}, \tag{2.1}$$

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

$$\rho(x) = \frac{\sqrt{(x - L_1)(x + L_1)(L_2 - x)(L_2 + x)}}{\pi c_1 x(1 - x)(1 + x)}, \ x \in [L_1, L_2], \tag{2.2}$$

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 $L_1 = |\sqrt{c_2 - c_2c_1} - \sqrt{c_1 - c_1c_2}|$ and $L_2 = |\sqrt{c_2 - c_2c_1} + \sqrt{c_1 - c_1c_2}|$. Here the empirical measure of $r_1, r_2, \cdots, r_{p_1}$ is defined as in (10.54) with μ_i replaced by r_i.

Yang and Pan (2012) proved that (2.2) 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 $\{X_{ij}, 1 \leq i \leq p_1, 1 \leq j \leq n\}$ and $\{Y_{ij}, 1 \leq i \leq p_2, 1 \leq j \leq n\}$ 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 (2.2).*

Under Assumptions 2-4, instead of $F^{\mathbf{S}_{\mathbf{xy}}}$, we analyze the ESD, $F^{\mathbf{T}_{\mathbf{xy}}}$, of the regularized random matrix $\mathbf{T}_{\mathbf{xy}}$ given in (10.11). To this end, define the Stieltjes transform of any distribution function $G(x)$ by

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

where $\Im z$ denotes the imaginary part of the complex number z.

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

$$\mathbf{S}_1 = \frac{1}{p_2}\sum_{k=1}^{p_2} \mathbf{w}_k\mathbf{w}_k^T, \quad \mathbf{S}_{2t} = \frac{1}{n-p_2}\sum_{k=p_2+1}^{n} \mathbf{w}_k\mathbf{w}_k^T + t\frac{n}{n-p_2}\mathbf{\Sigma}_{\mathbf{xx}}^{-1}, \quad y_1 = \frac{c_1^{'}}{c_2^{'}}, \quad y_2 = \frac{c_1^{'}}{1-c_2^{'}}.$$

In the definition of $\mathbf{S}_{2t}$ we require $n > p_2$. 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

$$m_{y_2t}(z) = m_{H_t}\Big(z - \frac{1}{1 + y_2 m_{y_2t}(z)}\Big), \tag{2.3}$$

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 (1.4) in the supplement material). 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 spectral 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 zero eigenvalues and this leads to

$$\underline{m}_{\mathbf{y}}(z) = -\frac{1-y_1}{z} + y_1 m_{\mathbf{y}}(z). \tag{2.4}$$

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

Theorem 1. *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* $\tilde{F}(\frac{\lambda}{q(1-\lambda)})$ *with* $q = \frac{c_2^{'}}{1-c_2^{'}}$ *where* $\tilde{F}(\lambda)$ *is a nonrandom distribution and its Stieltjes transform* $m_{\mathbf{y}}(z)$ *is the unique solution in* $\mathbb{C}^+$ *to*

$$m_{\mathbf{y}}(z) = -\int \frac{dF_{y_2t}(1/\lambda)}{\lambda(1 - y_1 - y_1 z m_{\mathbf{y}}(z)) - z}. \tag{2.5}$$

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

$$m_{\tilde{G}}(z) = \int \frac{dH(\lambda)}{\lambda(1 - c_1^{\prime} - c_1^{\prime} z m_{\tilde{G}}(z)) - z}. \tag{2.6}$$

Remark 2. *Indeed, taking* $t = 0$ *in (2.5) 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 (2.2) when $\mathbf{x}$ and $\mathbf{y}$ are independent, while the limit of $F^{\mathbf{S_{xy}}}(x)$ could be different from (2.2) 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

$$\int \phi(x) dF^{\mathbf{S_{xy}}}(x) = \frac{1}{p_1} \sum_{i=1}^{p_1} \phi(r_i^2). \tag{2.7}$$

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

$$S_n = \int x dF^{\mathbf{S_{xy}}}(x) = \frac{1}{p_1} \sum_{i=1}^{p_1} r_i^2. \tag{2.8}$$

Indeed, extensive simulations based on Theorems 2 and 3 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 (10.31) with fixed dimensions is

$$MLR_n = \sum_{i=1}^{p_1} \log(1 - r_i^2) \tag{2.9}$$

(see Wilks (1935) and Aderson (1984)). That is, $\phi(x)$ in (10.29) 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 (2.2). 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 $\rho(x)$, L_2, can be equal to one so that some sample correlation coefficients r_i are close to one. For example $L_2 = 1$ when $c_1 = c_2 = 1/2$. 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 2-4, we substitute $\mathbf{T_{xy}}$ for $\mathbf{S_{xy}}$ and use the statistic

$$T_n = \int x dF^{\mathbf{T_{xy}}}(x). \tag{2.10}$$

We next establish the CLTs of the statistics (10.29) and (2.10). To this end, write

$$G^{(1)}_{p_1,p_2}(\lambda) = p_1 \big(F^{\mathbf{S_{xy}}}(\lambda) - F^{c_{1n},c_{2n}}(\lambda)\big), \tag{2.11}$$

and

$$G^{(2)}_{p_1,p_2}(\lambda) = p_1 \big(F^{\mathbf{T_{xy}}}(\lambda) - F^{c'_{1n},c'_{2n}}(\lambda)\big), \tag{2.12}$$

where $F^{c_{1n},c_{2n}}(\lambda)$ and $F^{c'_{1n},c'_{2n}}(\lambda)$ are obtained from $F^{c_1,c_2}(\lambda)$ and $F^{c'_1,c'_2}(\lambda)$ with c_1, c_2, c'_1, c'_2 and H replaced by $c_{1n} = \frac{p_1}{n}, c_{2n} = \frac{p_2}{n}$, $c'_{1n} = \frac{p_1}{n}, c'_{2n} = \frac{p_2}{n}$ and $F^{\mathbf{\Sigma_{xx}}}$ 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. The density of $F^{c_1,c_2}(\lambda)$ can be obtained from $\rho(x)$ in (2.2) while the density of $F^{c_1',c_2'}(\lambda)$ can be recovered from (2.5). We re-normalize (10.29) and (2.10) as

$$\int \phi(\lambda)dG^{(1)}_{p_1,p_2}(\lambda) := p_1\Big(\int \phi(\lambda)dF^{\mathbf{S_{xy}}}(\lambda) - \int \phi(\lambda)dF^{c_{1n},c_{2n}}(\lambda)\Big), \tag{2.13}$$

and

$$\int \phi(\lambda)dG^{(2)}_{p_1,p_2}(\lambda) := p_1\Big(\int \phi(\lambda)dF^{\mathbf{T_{xy}}}(\lambda) - \int \phi(\lambda)dF^{c_{1n}',c_{2n}'}(\lambda)\Big). \tag{2.14}$$

Also, let

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

$$a_2 = \frac{(1+h)^2}{(1-\bar{y}_2)^2},\ g_{\bar{y}_1,\bar{y}_2}(\lambda) = \frac{1-\bar{y}_2}{2\pi\lambda(\bar{y}_1+\bar{y}_2\lambda)}\sqrt{(a_2-\lambda)(\lambda-a_1)},\ a_1 < \lambda < a_2. \tag{2.15}$$

Theorem 2. *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*

$$EX_{11}^4 = 3. \tag{2.16}$$

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

$$\Big(\int \phi_1(\lambda)dG^{(1)}_{p_1,p_2}(\lambda), \ldots, \int \phi_s(\lambda)dG^{(1)}_{p_1,p_2}(\lambda)\Big) \tag{2.17}$$

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

$$EX_{\phi_i} = \lim_{r\downarrow 1}\frac{1}{4\pi i}\oint_{|\xi|=1} f_i\Big(\frac{1+h^2+2h\Re(\xi)}{(1-\bar{y}_2)^2}\Big)\Big[\frac{1}{\xi-r^{-1}} + \frac{1}{\xi+r^{-1}} - \frac{2}{\xi+\frac{\bar{y}_2}{h}}\Big]d\xi, \tag{2.18}$$

and covariance function

$$cov(X_{\phi_i}, X_{\phi_j}) = -\lim_{r\downarrow 1}\frac{1}{4\pi^2}\oint_{|\xi_1|=1}\oint_{|\xi_2|=1}\frac{f_i\big(\frac{1+h^2+2h\Re(\xi_1)}{(1-\bar{y}_2)^2}\big)f_j\big(\frac{1+h^2+2h\Re(\xi_2)}{(1-\bar{y}_2)^2}\big)}{(\xi_1-r\xi_2)^2}d\xi_1 d\xi_2, \tag{2.19}$$

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 3. *When $\phi(x) = x$, the mean of the limit distribution in Theorem 2 is 0 and the variance is $\frac{2h^2y_1^2y_2^2}{(y_1+y_2)^4}$. These are calculated in Example 4.2 of Zheng (2012). Moreover, the assumption (10.38) can be replaced by $EY_{11}^4 = 3$ since $\mathbf{X}$ and $\mathbf{Y}$ have an equal status in the matrix $\mathbf{S_{xy}}$.*

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

$$m_r(z) = \int \frac{dH_t(x)}{(x - z + \varpi(z))^r}, \quad \varpi(z) = \frac{1}{1 + y_2 m_{y_2 t}(z)}, \quad g(z) = \frac{y_2 (m_{y_2 t}(-\underline{m}_{\mathbf{y}}(z)))'}{(1 + y_2 m_{y_2 t}(-\underline{m}_{\mathbf{y}}(z)))^2}$$

$$s(z_1, z_2) = \frac{1}{1 + y_2 m_{y_2 t}(z_1)} - \frac{1}{1 + y_2 m_{y_2 t}(z_2)}, \quad h(z) = \frac{-\underline{m}_{\mathbf{y}}^2(z)}{1 - y_1 \underline{m}_{\mathbf{y}}^2(z) \int \frac{dF_{y_2 t}(x)}{(x + \underline{m}_{\mathbf{y}}(z))^2}},$$

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

Theorem 3. *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 (2.5). In addition to Assumptions 2-4, suppose that*

$$EX_{11}^4 = 3. \tag{2.20}$$

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

$$\Big(\int \phi_1(\lambda) dG_{p_1,p_2}^{(2)}(\lambda), \ldots, \int \phi_s(\lambda) dG_{p_1,p_2}^{(2)}(\lambda)\Big) \tag{2.21}$$

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

$$\begin{aligned} EX_{\phi_i} = &-\frac{1}{2\pi i} \oint_{\mathcal{C}} \phi_i\Big(\frac{qz}{1+qz}\Big)\Big(\frac{y_1 \int \underline{m}_{\mathbf{y}}(z)^3 x [x + \underline{m}_{\mathbf{y}}(z)]^{-3} dF_{y_2 t}(x)}{[1 - y_1 \int \underline{m}_{\mathbf{y}}(z)^2 (x + \underline{m}_{\mathbf{y}}(z))^{-2} dF_{y_2 t}(x)]^2} \\ &+ h(z) \frac{y_2 \varpi^2(-\underline{m}_{\mathbf{y}}(z)) m_3(-\underline{m}_{\mathbf{y}}(z)) + y_2^2 \varpi^4(-\underline{m}_{\mathbf{y}}(z)) m_{y_2 t}'(-\underline{m}_{\mathbf{y}}(z)) m_3(-\underline{m}_{\mathbf{y}}(z))}{1 - y_2 \varpi^2(-\underline{m}_{\mathbf{y}}(z)) m_2(-\underline{m}_{\mathbf{y}}(z))} \\ &- h(z) \frac{y_2^2 \varpi^3(-\underline{m}_{\mathbf{y}}(z)) m_{y_2 t}'(-\underline{m}_{\mathbf{y}}(z)) m_2(-\underline{m}_{\mathbf{y}}(z))}{1 - y_2 \varpi^2(-\underline{m}_{\mathbf{y}}(z)) m_2(-\underline{m}_{\mathbf{y}}(z))}\Big) dz \end{aligned} \tag{2.22}$$

and covariance

$$\begin{aligned} Cov(X_{\phi_i}, X_{\phi_j}) = &-\frac{1}{2\pi^2} \oint_{\mathcal{C}_1} \oint_{\mathcal{C}_2} \phi_i\Big(\frac{qz_1}{1+qz_1}\Big) \phi_j\Big(\frac{qz_2}{1+qz_2}\Big)\Big(\frac{\underline{m}_{\mathbf{y}}'(z_1) \underline{m}_{\mathbf{y}}'(z_2)}{(\underline{m}_{\mathbf{y}}(z_1) - \underline{m}_{\mathbf{y}}(z_2))^2} - \frac{1}{(z_1 - z_2)^2} \\ &- \frac{h(z_1) h(z_2)}{(-\underline{m}_{\mathbf{y}}(z_2) + \underline{m}_{\mathbf{y}}(z_1))^2} + \frac{h(z_1) h(z_2)[1 + g(z_1) + g(z_2) + g(z_1) g(z_2)]}{[-\underline{m}_{\mathbf{y}}(z_2) + \underline{m}_{\mathbf{y}}(z_1) + s(-\underline{m}_{\mathbf{y}}(z_1), -\underline{m}_{\mathbf{y}}(z_2))]^2}\Big) dz_1 dz_2. \end{aligned} \tag{2.23}$$

Here q is defined in Theorem 1. The contours in (2.22) and (2.23)(two in (2.23), which may be assumed to be nonoverlapping) are closed and are taken in the positive direction in the complex plain, each enclosing the support of $\widetilde{F}(\lambda)$.

b) If $c_2' \in [1, +\infty)$ ($p_2 \geq n$), (10.69) converges weakly to a Gaussian vector $(X_{\phi_1}, \ldots, X_{\phi_s})$ with mean

$$EX_{\phi_i} = -\frac{1}{2\pi i} \oint_{\mathcal{C}} \phi_i\Big(\frac{t^{-1} z}{1 + t^{-1} z}\Big) \frac{c_1' \int (1 + \lambda \underline{s}(z)^3)^{-3} \underline{s}(z)^3 \lambda^2 dH(\lambda)}{\big(1 - c_1' \int \underline{s}(z)^2 \lambda^2 (1 + \lambda \underline{s}(z))^{-2} dH(\lambda)\big)^2} dz \tag{2.24}$$

and

$$Cov(X_{\phi_i}, X_{\phi_j}) = -\frac{1}{2\pi^2} \oint_{\mathcal{C}_1} \oint_{\mathcal{C}_2} \phi_i\Big(\frac{t^{-1} z_1}{1 + t^{-1} z_1}\Big) \phi_i\Big(\frac{t^{-1} z_2}{1 + t^{-1} z_2}\Big) \frac{\underline{s}'(z_1) \underline{s}'(z_2)}{\big(\underline{s}(z_1) - \underline{s}(z_2)\big)^2} dz_1 dz_2, \tag{2.25}$$

where $\underline{s}(z)$ is Stieltjes transform of the LSD of the matrix $\frac{1}{n}\mathbf{W}^T\mathbf{\Sigma_{xx}}\mathbf{W}$. The contours in (2.24) and (2.25)(two in (2.25), which may be assumed to be nonoverlapping) are closed and are taken in the positive direction in the complex plain, each enclosing the support of $\widetilde{G}(\lambda)$.

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 2 and 3 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 the high power of their entries. See Section 6.5 for the further discussion.

Following Lytova and Pastur (2009) condition (10.38) can be removed. However it will significantly increase the length of this work and we will not pursue it here.

3 Test statistics

Note that the regularized statistic $\int \lambda dG^{(2)}_{p_1,p_2}(\lambda)$ in (2.14) (when $\phi(\lambda)=\lambda$) involves the unknown covariance matrix $\mathbf{\Sigma_{xx}}$ through $F^{c'_{1n},c'_{2n}}(\lambda)$. In order to apply it to conduct tests, one needs to estimate the unknown parameter. It is well known that estimating the population covariance matrix $\mathbf{\Sigma_{xx}}$ is very challenging unless it is sparse. El Karoui (2008) and Bai et al. (2010) proposed some approaches to estimate the limit of the ESD of $\mathbf{\Sigma_{xx}}$ or its moments. However the convergence rate is not fast enough to offset the order of p_1. Indeed, Theorem 1 of Bai et al. (2010) implies that the best possible convergence rate is $O_p(\frac{1}{n})$. In view of this, we provide two methods to deal with the problem. One is to estimate $\int \lambda dF^{c'_{1n},c'_{2n}}(\lambda)$ in a framework of sparsity while the other one is to eliminate this unknown parameter by dividing the samples into two groups.

3.1 Plug-in estimator under sparsity

When $c'_2<1$, it turns out that

$$\int \lambda dF^{c'_{1n},c'_{2n}}(\lambda)=\frac{p_2}{p_1}-\frac{p_2}{p_1}\frac{1}{1+c'_{1n}m_{nt}}, \tag{3.1}$$

where m_{nt} is a solution to the equation

$$m_{nt}=a_n-\frac{a_n t}{p_1}tr(a_n^{-1}\mathbf{\Sigma_{xx}}+t\mathbf{I})^{-1} \tag{3.2}$$

with $a_n=1+c'_{1n}m_{nt}$ (see the proof of Theorem 8). An estimator of m_{nt} is then proposed as $\hat{m}_{nt}$ which is a solution to the equation

$$\hat{m}_{nt}=\hat{a}_n-\frac{\hat{a}_n t}{p_1}tr(\hat{a}_n^{-1}\hat{\mathbf{\Sigma}}_{\mathbf{xx}}+t\mathbf{I})^{-1}, \tag{3.3}$$

with $\hat{a}_n=1+c'_{1n}\hat{m}_{nt}$. Here we use a thresholding estimator $\hat{\mathbf{\Sigma}}_{\mathbf{xx}}$ to estimate $\mathbf{\Sigma_{xx}}$, slightly different from that proposed by Bickel and Levina (2008). Specifically speaking, suppose that the underlying random variables $\{X_{ij}\}$ are mean zero and variable one. Then define $\hat{\mathbf{\Sigma}}_{\mathbf{xx}}$ to be a matrix whose diagonal entries are all one and the off diagonal entries are $\hat{\sigma}_{ij}I(|\hat{\sigma}_{ij}| \geq \ell)$ with $\ell = M\sqrt{\frac{\log p_1}{n}}$ and M being some appropriate constant (M will be selected by cross-validation). Here $\hat{\sigma}_{ij}$ denotes the entry at the (i,j)th position of sample covariance matrix $\frac{1}{n}\mathbf{X}\mathbf{X}^T$. Therefore the resulting test statistic is

$$p_1\Big(\int \lambda dF^{\mathbf{T}_{\mathbf{xy}}}(\lambda) - (\frac{p_2}{p_1} - \frac{p_2}{p_1}\frac{1}{1+c'_{1n}\hat{m}_{nt}})\Big). \tag{3.4}$$

When $p_2 \geq n$, it turns out that

$$\int \lambda dF^{c'_{1n},c'_{2n}}(\lambda) = 1 - tm_n^{(1t)}, \tag{3.5}$$

where $m_n^{(1t)}$ satisfies the equation

$$m_n^{(1t)} = \frac{1}{p_1}tr\Big((1 - c'_{1n} + c'_{1n}tm_n^{(1t)})\mathbf{\Sigma}_{\mathbf{xx}} + t\mathbf{I}\Big)^{-1}. \tag{3.6}$$

We then propose the resulting test statistic

$$p_1\Big(\int \lambda dF^{\mathbf{T}_{\mathbf{xy}}}(\lambda) - (1 - t\hat{m}_n^{(1t)})\Big) \tag{3.7}$$

where $\hat{m}_n^{(1t)}$ satisfies the equation

$$\hat{m}_n^{(1t)} = \frac{1}{p_1}tr\Big((1 - c'_{1n} + c'_{1n}t\hat{m}_n^{(1t)})\hat{\mathbf{\Sigma}}_{\mathbf{xx}} + t\mathbf{I}\Big)^{-1}. \tag{3.8}$$

Theorem 4. *In addition to Assumptions in Theorem 3, suppose that* $EX_{ij}^2 = 1$, $\sup_{i,j} E|X_{ij}|^{17} < \infty$ *for all* i *and* j *and that*

$$s_o(p_1)\Big(\frac{\log p_1}{n}\Big)^{(1-q)/2} \to 0, \tag{3.9}$$

where $\sum_{i\neq j}|\sigma_{ij}|^q = s_o(p_1)$ *with* $0 \leq q < 1$.

a) If $c'_2 < 1$, *then* $p_1\Big(\int \lambda dF^{\mathbf{T}_{\mathbf{xy}}}(\lambda) - (\frac{p_2}{p_1} - \frac{p_2}{p_1}\frac{1}{1+c'_{1n}\hat{m}_{nt}})\Big)$ *converges weakly to a normal distribution with the mean and variance given in (2.22) and (2.23) with* $\phi(\lambda) = \lambda$.

b) If $c'_2 \geq 1$, *then* $p_1\Big(\int \lambda dF^{\mathbf{T}_{\mathbf{xy}}}(\lambda) - (1 - t\hat{m}_n^{(1t)})\Big)$ *converges weakly to a normal distribution with the mean and variance given in Part (b) of Theorem 3 with* $\phi(\lambda) = \lambda$.

We demonstrate an example of sparse covariance matrix in the simulation parts, satisfying the sparse condition (3.9).

3.2 Strategy of dividing samples

If (3.9) is not satisfied, we then propose a strategy of dividing the total samples into two groups. Specifically speaking, we divide the n samples of $(\mathbf{x}, \mathbf{y})$ into two groups respectively, i.e.

$$Group\ 1 : \mathbf{X}^{(1)} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{[n/2]}), \quad \mathbf{Y}^{(1)} = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{[n/2]}) \tag{3.10}$$

and

$$\textit{Group 2}: \mathbf{X}^{(2)} = (\mathbf{x}_{[n/2]+1}, \mathbf{x}_{[n/2]+2}, \ldots, \mathbf{x}_n), \quad \mathbf{Y}^{(2)} = (\mathbf{y}_{[n/2]+1}, \mathbf{y}_{[n/2]+2}, \ldots, \mathbf{y}_n), \tag{3.11}$$

where $[n/2]$ is the largest integer not greater than $n/2$. When n is odd, we discard the last sample. However, if the above strategy of dividing samples into two groups is directly used, then the asymptotic means of the resulting statistic (the difference between the statistics in (2.14) obtained from two subsamples) are always zero in both null hypothesis and alternative hypothesis due to similarity of two groups so that the power of the test statistic is very low. This is also confirmed by simulations. Therefore we further propose its modified version as follows.

For $\mathbf{Y}^{(2)}$ in Group 2, we extract a sub-data $\widetilde{\mathbf{Y}}^{(2)}$, i.e.

$$\widetilde{\mathbf{Y}}^{(2)} = (\widetilde{\mathbf{y}}_{[n/2]+1}, \widetilde{\mathbf{y}}_{[n/2]+2}, \ldots, \widetilde{\mathbf{y}}_n),$$

where $\widetilde{\mathbf{y}}_j$ consists of the first $[p_2/2]$ components of $\mathbf{y}_j$, for all $j = [n/2]+1, [n/2]+2, \ldots, n$. We use $\widetilde{\mathbf{Y}}^{(2)}$ to form a new group

$$\textit{Modified Group 2}: \mathbf{X}^{(2)} = (\mathbf{x}_{[n/2]+1}, \mathbf{x}_{[n/2]+2}, \ldots, \mathbf{x}_n), \quad \widetilde{\mathbf{Y}}^{(2)} = (\widetilde{\mathbf{y}}_{[n/2]+1}, \widetilde{\mathbf{y}}_{[n/2]+2}, \ldots, \widetilde{\mathbf{y}}_n).$$

For Group 1, it follows from Theorem 3 that

$$\int \lambda d p_1 \Big(F^{\mathbf{T}^{(1)}_{\mathbf{xy}}}(\lambda) - F^{2c'_{1n}, 2c'_{2n}}(\lambda) \Big) \xrightarrow{d} Z_1, \tag{3.12}$$

where $\mathbf{T}^{(1)}_{\mathbf{xy}}$ is obtained from $\mathbf{T}_{\mathbf{xy}}$ with $\mathbf{X}$ and $\mathbf{Y}$ replaced by $\mathbf{X}^{(1)}$ and $\mathbf{Y}^{(1)}$ respectively and Z_1 is a normal random variable with mean and variance given in Theorem 3 with c'_1 and c'_2 replaced by $2c'_1$ and $2c'_2$ respectively and $\phi(\lambda) = \lambda$. Similarly, with Modified Group 2, by Theorem 3

$$\int \lambda d p_1 \Big(F^{\mathbf{T}^{(2)}_{\mathbf{xy}}}(\lambda) - F^{2c'_{1n}, c'_{2n}}(\lambda) \Big) \xrightarrow{d} Z_2, \tag{3.13}$$

where $\mathbf{T}^{(2)}_{\mathbf{xy}}$ is $\mathbf{T}_{\mathbf{xy}}$ with $\mathbf{X}$ and $\mathbf{Y}$ replaced by $\mathbf{X}^{(2)}$ and $\widetilde{\mathbf{Y}}^{(2)}$ respectively and Z_2 is a normal random variable with the mean and variance given in Theorem 3 with $\phi(\lambda) = \lambda$ and c'_1 replaced by $2c'_1$.

We next investigate the relation between $\int \lambda dF^{2c'_{1n}, 2c'_{2n}}(\lambda)$ and $\int \lambda dF^{2c'_{1n}, c'_{2n}}(\lambda)$, and then calculate some difference between the two statistics in (3.12) and (3.13) in order to eliminate the unknown parameters $\int \lambda dF^{2c'_{1n}, 2c'_{2n}}(\lambda)$ and $\int \lambda dF^{2c'_{1n}, c'_{2n}}(\lambda)$.

When $c'_2 < 1/2$ we have

$$\int \lambda dF^{2c'_{1n}, 2c'_{2n}}(\lambda) = \frac{p_2}{p_1} - \frac{p_2}{p_1} \frac{1}{1 + 2c'_{1n} \widetilde{m}_{nt}}, \tag{3.14}$$

where $\widetilde{m}_{nt}$ is obtained from m_{nt} satisfying (3.2) with c'_{1n} replaced by $2c'_{1n}$. On the other hand

$$\int \lambda dF^{2c'_{1n}, c'_{2n}}(\lambda) = \frac{p_2/2}{p_1} - \frac{p_2/2}{p_1} \frac{1}{1 + 2c'_{1n} \widetilde{m}_{nt}}. \tag{3.15}$$

It follows that

$$\int \lambda dF^{2c'_{1n}, 2c'_{2n}}(\lambda) = 2 \int \lambda dF^{2c'_{1n}, c'_{2n}}(\lambda). \tag{3.16}$$

When $[p_2/2] > [n/2]$, we have

$$\int \lambda dF^{2c'_{1n},2c'_{2n}}(\lambda) = \int \lambda dF^{2c'_{1n},c'_{2n}}(\lambda) = 1 - t\widetilde{m}_n^{(1t)}, \tag{3.17}$$

where $\widetilde{m}_n^{(1t)}$ is $m_n^{(1t)}$ satisfying (3.6) with c'_{1n} replaced by $2c'_{1n}$.

The last case is $[p_2/2] \le [n/2]$ and $c'_2 \ge 1/2$. For this case, if we still consider Group 1 and Modified Group 2, then

$$\int \lambda dF^{2c'_{1n},2c'_{2n}}(\lambda) = 1 - t\widetilde{m}_n^{(1t)}, \quad \int \lambda dF^{2c'_{1n},c'_{2n}}(\lambda) = \frac{[p_2/2]}{p_1} - \frac{[p_2/2]}{p_1}\frac{1}{1+2c'_{1n}\widetilde{m}_{nt}}.$$

From the above formulas it is difficult to figure out the relation between $\int \lambda dF^{2c'_{1n},2c'_{2n}}(\lambda)$ and $\int \lambda dF^{2c'_{1n},c'_{2n}}(\lambda)$ depending on the unknown parameter $\mathbf{\Sigma_{xx}}$. To overcome this difficulty, we also apply a 'sub-data' trick to Group 1. Specifically speaking, consider a modified Group 1 as follows.

$$Modified\ Group\ 1 : \mathbf{X}^{(1)} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{[n/2]}), \quad \dot{\mathbf{Y}}^{(1)} = (\dot{\mathbf{y}}_1, \dot{\mathbf{y}}_2, \ldots, \dot{\mathbf{y}}_{[n/2]}),$$

where $\dot{\mathbf{y}}_k$ consists of the last $[p_2/2]$ components of $\mathbf{y}_k$, i.e. the i-th component of $\dot{\mathbf{y}}_k$ is the $([p_2/2]+i)$-th component of $\mathbf{y}_k$, for all $i = 1, 2, \ldots, [p_2/2]$ and $k = 1, 2, \ldots, [n/2]$. For Modified Group 1, by Theorem 3, we have

$$\int \lambda dp_1\Big(F^{\widetilde{\mathbf{T}}^{(1)}_{\mathbf{xy}}}(\lambda) - F^{2c'_{1n},c'_{2n}}(\lambda)\Big) \xrightarrow{d} Z_3, \tag{3.18}$$

where $\widetilde{\mathbf{T}}^{(1)}_{\mathbf{xy}}$ is $\mathbf{T_{xy}}$ with $\mathbf{X}$ and $\mathbf{Y}$ replaced by $\mathbf{X}^{(1)}$ and $\dot{\mathbf{Y}}^{(1)}$ respectively; and Z_3 is a normal random variable with the mean and variance given in Theorem 3 with $\phi(\lambda) = \lambda$ and c'_1 replaced by $2c'_1$. Since the unknown parameters in (3.13) and (3.18) are the same the difference between (3.13) and (3.18) can be taken as the modified statistic.

The asymptotic distributions of the three resulting statistics are given in Theorem 5.

Theorem 5. *Suppose that Assumptions in Theorem 3 hold.*

a) If $c'_2 < 1/2$, the statistic $\int \lambda dF^{\mathbf{T}^{(1)}_{\mathbf{xy}}}(\lambda) - 2\int \lambda dF^{\mathbf{T}^{(2)}_{\mathbf{xy}}}(\lambda)$ converges weakly to a normal distribution with the mean $(\mu_1 - 2\mu_2)$ and variance $(\sigma_1^2 + 4\sigma_2^2)$, where μ_1 and σ_1^2 are given in (2.22) and (2.23) respectively with c'_1, c'_2 replaced by $2c'_1, 2c'_2$ respectively and $\phi(\lambda) = \lambda$; μ_2 and σ_2^2 are given in (2.22) and (2.23) respectively with c'_1 replaced by $2c'_1$ and $\phi(\lambda) = \lambda$.

b) If $c'_2 \ge 1$, the statistic $\int \lambda dF^{\mathbf{T}^{(1)}_{\mathbf{xy}}}(\lambda) - \int \lambda dF^{\mathbf{T}^{(2)}_{\mathbf{xy}}}(\lambda)$ converges weakly to a normal distribution with the mean zero and variance $2\sigma_3^2$, where σ_3^2 is given in (2.25) with c'_1 replaced by $2c'_1$ and $\phi(\lambda) = \lambda$.

c) If $1/2 \le c'_2 < 1$, the statistic $\int \lambda dF^{\widetilde{\mathbf{T}}^{(1)}_{\mathbf{xy}}}(\lambda) - \int \lambda dF^{\mathbf{T}^{(2)}_{\mathbf{xy}}}(\lambda)$ converges weakly to a normal distribution with mean zero and variance $2\sigma_4^2$, where σ_4^2 is given in (2.23) with c'_1 replaced by $2c'_1$ and $\phi(\lambda) = \lambda$.

Remark 4. *Unlike using Group 2 of (3.11) although the asymptotic means of the statistics in the cases (b) and (c) are zero under the null hypothesis, they are not necessarily equal to zero under the alternative hypothesis so that the power of the resulting test statistic becomes much better.*

Remark 5. *The asymptotic means and variances of the resulting statistics involve* $\mathbf{\Sigma_{xx}}$. *Estimators of the high dimensional covariance matrix* $\mathbf{\Sigma_{xx}}$ *have been developed in many literature, e.g. Bickel and Levina (2008) and Fan, Liao and Mincheva (2013), etc. We apply their approaches to estimate* $\mathbf{\Sigma_{xx}}$ *in the section of simulations. Such replacements do not affect the asymptotic distribution by Slutsky's theorem.*

4 The power under local alternatives

This section is to evaluate the power of S_n or T_n under a kind of local alternatives. Consider the alternative hypothesis

$$\mathbb{H}_1 : \mathbf{x} \textit{ and } \mathbf{y} \textit{ are dependent},$$

satisfying condition (4.1) below. Draw n samples from such alternatives $\mathbf{x}$ and $\mathbf{y}$ to form the respective analogues of (1.5) and (10.11) and denote them by $\mathbf{S}$ and $\mathbf{T}$ respectively. Suppose that the underlying random variables involved in $\mathbf{S_{xy}}, \mathbf{T_{xy}}$ and $\mathbf{S}, \mathbf{T}$ are in the same probability space (Ω, P).

Recall the definitions of $G^{(i)}_{p_1,p_2}, i = 1, 2$ in (10.65) and (10.66) and let $R^{(i)}_n = \int \lambda dG^{(i)}_{p_1,p_2}$.

Theorem 6. *In addition to assumptions in Theorem 2 or Theorem 3 suppose that for any* $M > 0$

$$P\Big(\Big|tr(\mathbf{S} - \mathbf{S_{xy}})\Big| \geq M\Big) \to 1, \quad P\Big(\Big|tr(\mathbf{T} - \mathbf{T_{xy}})\Big| \geq M\Big) \to 1, \tag{4.1}$$

Then

$$\lim_{n\to\infty} P(R^{(i)}_n > z^{(i)}_{1-\alpha} \textit{ or } R^{(i)}_n < z^{(i)}_{\alpha}|\mathbb{H}_1) = 1, \tag{4.2}$$

where $z^{(i)}_{1-\alpha}$ *and* $z^{(i)}_{\alpha}$ *are, respectively,* $(1-\alpha)$ *and* α *quantiles of the asymptotic distribution of the statistic* $R^{(i)}_n$ *under the null hypothesis.*

Remark 6. *For example one may take* $\mathbf{S} = (\mathbf{XLX}^T)^{-1}\mathbf{XLY}^T(\mathbf{YLY}^T)^{-1}\mathbf{YLX}^T$ *and* $\mathbf{S_{xy}} = (\mathbf{XX}^T)^{-1}\mathbf{XP_yX}^T$ *with* $\mathbf{L}$ *being a random matrix and* $\mathbf{P_y} = \mathbf{Y}^T(\mathbf{YY}^T)^{-1}\mathbf{Y}$. *Particularly, if* $\mathbf{L} = \mathbf{I} + \mathbf{ee}^T$ *with* $\mathbf{e} = x^2(1, 1, \cdots, 1)$ *and* x^2 *having finite moment, then under assumptions in Theorem 2 or Theorem 3 it can be proved that*

$$tr\Big(\mathbf{S} - \mathbf{S_{xy}}\Big) = O_p(n)$$

satisfying (4.1).

Next, we evaluate the powers of the modified statistics with the dividing-sample method. Draw n samples from alternatives $\mathbf{x}$ and $\mathbf{y}$ to form the respective analogues of $\mathbf{T}^{(i)}_{\mathbf{xy}}$, $i = 1, 2$, $\widetilde{\mathbf{T}}^{(1)}_{\mathbf{xy}}$ and denote them by $\mathbf{T}^{(i)}$, $i = 1, 2$, $\widetilde{\mathbf{T}}^{(1)}$ respectively. Let

$$J^{(1)}_n = \int \lambda dF^{\mathbf{T}^{(1)}}(\lambda) - 2\int \lambda dF^{\mathbf{T}^{(2)}}(\lambda),$$
$$J^{(2)}_n = \int \lambda dF^{\mathbf{T}^{(1)}}(\lambda) - \int \lambda dF^{\mathbf{T}^{(2)}}(\lambda),$$
$$J^{(3)}_n = \int \lambda dF^{\widetilde{\mathbf{T}}^{(1)}}(\lambda) - \int \lambda dF^{\mathbf{T}^{(2)}}(\lambda).$$

Theorem 7. *In addition to assumptions in Theorem 3, suppose that for any $M > 0$,*

$$P\Big(\big|tr(\mathbf{T}^{(1)}) - 2tr(\mathbf{T}^{(2)}) - \big(tr(\mathbf{T}^{(1)}_{\mathbf{xy}}) - 2tr(\mathbf{T}^{(2)}_{\mathbf{xy}})\big)\big| \geq M\Big) \to 1, \quad if \quad c_2^{'} < 1/2; \tag{4.3}$$

$$P\Big(\big|tr(\mathbf{T}^{(1)}) - tr(\mathbf{T}^{(2)}) - \big(tr(\mathbf{T}^{(1)}_{\mathbf{xy}}) - tr(\mathbf{T}^{(2)}_{\mathbf{xy}})\big)\big| \geq M\Big) \to 1, \quad if \quad c_2^{'} \geq 1; \tag{4.4}$$

$$P\Big(\big|tr(\widetilde{\mathbf{T}}^{(1)}) - tr(\mathbf{T}^{(2)}) - \big(tr(\widetilde{\mathbf{T}}^{(1)}_{\mathbf{xy}}) - tr(\mathbf{T}^{(2)}_{\mathbf{xy}})\big)\big| \geq M\Big) \to 1, \quad if \quad 1/2 \leq c_2^{'} < 1. \tag{4.5}$$

Then

$$\lim_{n\to\infty} P(J_n^{(i)} > z_{1-\alpha}^{(i)} \ or \ J_n^{(i)} < z_\alpha^{(i)} | \mathbb{H}_1) = 1 \quad i = 1, 2, 3,$$

where $z_{1-\alpha}^{(i)}$ and $z_\alpha^{(i)}$ are, respectively, $(1-\alpha)$ and α quantiles of the asymptotic distribution of the statistic $J_n^{(i)}$ under the null hypothesis, i=1,2,3.

5 Applications of CCA

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

5.1 Multivariate regression test with CCA

Consider the multivariate regression(MR) model as follows:

$$\mathbf{Y} = \mathbf{XB} + \mathbf{E}, \tag{5.1}$$

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 $\mathbf{A}_{\mathbf{xy}} = \frac{1}{n}\mathbf{X}^T\mathbf{Y}$ and $\mathbf{A}_{\mathbf{xx}} = \frac{1}{n}\mathbf{X}^T\mathbf{X}$. We have the least square estimate of $\mathbf{B}$

$$\hat{\mathbf{B}} = \mathbf{A}_{\mathbf{xx}}^{-1}\mathbf{A}_{\mathbf{xy}}. \tag{5.2}$$

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

$$\mathbb{H}_0 : \mathbf{B} = \mathbf{0}. \tag{5.3}$$

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

$$\Lambda = \frac{\det(\mathbf{E})}{\det(\mathbf{E} + \mathbf{H})} = \prod_{i=1}^{s}(1 + \lambda_i)^{-1}, \tag{5.4}$$

where

$$\mathbf{E} = \mathbf{Y}^T\big(\mathbf{I} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\big)\mathbf{Y} \tag{5.5}$$

and

$$\mathbf{H} = \hat{\mathbf{B}}^T(\mathbf{X}^T\mathbf{X})\hat{\mathbf{B}}; \tag{5.6}$$

and $\{\lambda_i : i = 1, \dots, s\}$ are the roots of $det(\mathbf{H} - \lambda\mathbf{E}) = 0$, $s = min(k, p)$. 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 $det(\mathbf{H}-\lambda\mathbf{E}) = 0$ becomes $det\Big(\mathbf{A_{yx}}\mathbf{A_{xx}^{-1}}\mathbf{A_{xy}} - \lambda(\mathbf{A_{yy}} - \mathbf{A_{yx}}\mathbf{A_{xx}^{-1}}\mathbf{A_{xy}})\Big) = 0$. From Theorem 2.6.8 of N.H.Timm (2001) we have $det(\mathbf{H} - \theta(\mathbf{H} + \mathbf{E})) = det\Big(\mathbf{A_{yx}}\mathbf{A_{xx}^{-1}}\mathbf{A_{xy}} - \theta\mathbf{A_{yy}}\Big) = 0$ so that

$$\Lambda = \prod_{i=1}^{s}(1 + \lambda_i)^{-1} = \prod_{i=1}^{s}(1 - \theta_i) = \frac{\det(\mathbf{A_{yy}} - \mathbf{A_{yx}}\mathbf{A_{xx}^{-1}}\mathbf{A_{xy}})}{\det(\mathbf{A_{yy}})}. \tag{5.7}$$

Evidently, the quantities $r_i^2 = \theta_i, i = 1, \dots, s$ are sample canonical correlation coefficients. Therefore the test statistic (10.10) can be rewritten as

$$\log\Lambda = \sum_{i=1}^{s}\log(1 - r_i^2). \tag{5.8}$$

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.

5.2 Testing for cointegration with CCA

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

$$\Delta\mathbf{y}_t = -\boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t,\ t = 1, \dots, T, \tag{5.9}$$

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 $I(0)$. Here $I(d)$ denotes integrated of order d.

The goal is to test

$$\mathbb{H}_0 : r = 0\ (\boldsymbol{\alpha} = \boldsymbol{\beta} = \mathbf{0});\ \textit{against}\ \mathbf{H}_1 : r > 0; \tag{5.10}$$

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

This cointegration test is equivalent to testing

$$\mathbb{H}_0 : \Delta\mathbf{y}_t\ \textit{is independent with}\ \Delta\mathbf{y}_{t-1};\ \textit{against}\ H_1 : \Delta\mathbf{y}_t\ \textit{is dependent with}\ \Delta\mathbf{y}_{t-1}. \tag{5.11}$$

In order to apply canonical correlation coefficients to cointegration test (5.10), we construct random matrices

$$\mathbf{X} = \big(\Delta\mathbf{y}_2, \Delta\mathbf{y}_4, \ldots, \Delta\mathbf{y}_{2t-2}, \Delta\mathbf{y}_{2t}, \ldots, \Delta\mathbf{y}_T\big), \tag{5.12}$$

$$\mathbf{Y} = \big(\Delta\mathbf{y}_1, \Delta\mathbf{y}_3, \ldots, \Delta\mathbf{y}_{2t-1}, \Delta\mathbf{y}_{2t+1}, \ldots, \Delta\mathbf{y}_{T-1}\big). \tag{5.13}$$

6 Simulation results

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

6.1 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

$$\hat{\alpha} = \frac{\{\sharp \ of \ S_n^H \geq z_{1-\alpha}\}}{K}, \tag{6.1}$$

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

The empirical power is calculated as

$$\hat{\beta} = \frac{\{\sharp \ of \ S_n^A \geq \hat{z}_{1-\alpha}\}}{K}, \tag{6.2}$$

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

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

6.2 Testing independence

Consider the data generating process

$$\mathbf{x} = \mathbf{\Sigma}_{\mathbf{xx}}^{1/2}\mathbf{w}, \quad \mathbf{y} = \mathbf{\Sigma}_{\mathbf{yy}}^{1/2}\mathbf{v}, \tag{6.3}$$

with

$$(a) \ \ \mathbf{\Sigma}_{\mathbf{xx}} = \mathbf{I}_{p_1}, \ \mathbf{\Sigma}_{\mathbf{yy}} = \mathbf{I}_{p_2}; \qquad (b) \ \ \mathbf{\Sigma}_{\mathbf{xx}} = (\sigma_{kh}^{SP})_{k,h=1}^{p_1}, \ \mathbf{\Sigma}_{\mathbf{yy}} = \mathbf{I}_{p_2},$$

$$(c) \ \ \mathbf{\Sigma}_{\mathbf{xx}} = (\sigma_{kh}^{AR})_{k,h=1}^{p_1}, \ \mathbf{\Sigma}_{\mathbf{yy}} = \mathbf{I}_{p_2}, \qquad (d) \ \ \mathbf{\Sigma}_{\mathbf{xx}} = \mathbf{B}^{'}cov(\mathbf{f}_t)\mathbf{B} + \mathbf{\Sigma}_{\mathbf{u}},$$

where

$$\sigma_{kh}^{AR} = \frac{\phi^{|k-h|}}{1-\phi^2}, \quad k, h = 1, 2, \ldots, p_1, \quad \phi = 0.8,$$

and $\sigma_{kh}^{SP} = 0$ except that

$$\sigma_{kk}^{SP} = 1, \quad k = 1, 2, \ldots, p_1; \quad \sigma_{1j}^{SP} = \sigma_{j1}^{SP} = \theta, \quad j = 2, 3, \ldots, [p_1^{1/3}], \quad \theta = 0.2.$$

Here $cov(\mathbf{f}_t)$ is an $r \times r$ identity matrix and $\mathbf{\Sigma_u}$ is a $p_1 \times p_1$ identity matrix and $\mathbf{B} = (\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_{p_1})$, where each $\mathbf{b}_i : r \times 1$ is generated independently from a normal distribution with covariance matrix being an $r \times r$ identity matrix and mean $\boldsymbol{\mu}_B$ consisting of all 1.

The empirical sizes of the proposed statistics S_n for cases (a) and (b) are listed in Table 1. Moreover, the empirical sizes for the re-normalized statistic MLR_n are included as comparison with S_n. Here the re-normalized statistic MLR_n means 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)-(d) are listed in Table 2. For GDP(a), we use the original statistic T_n; for GDP(b), the statistic in Theorem 8 is used; for GDP (c) and (d), the dividing-sample statistic in Theorem 5 is utilized.

From the results in Table 1 and 2, the proposed statistics S_n and T_n work well under Assumption 1 and 2 respectively.

Remark 7. *A banded type matrix in (c) and a sparse matrix in (b) are both estimated by the thresholding method in Bickel and Levina (2008). A low rank matrix plus a sparse matrix in (d) is estimated by combining principle component analysis and thresholding method originated in Fan, Liao and Mincheva (2013).*

6.3 Factor model dependence

We consider the factor model as follows:

$$\mathbf{x}_t = \Lambda_1 \mathbf{f}_t + \mathbf{u}_t, \quad \mathbf{y}_t = \Lambda_2 \mathbf{f}_t + \mathbf{v}_t, t = 1, 2, \ldots, n, \tag{6.4}$$

where Λ_1 and Λ_2 are $p_1 \times r$ and $p_2 \times r$ deterministic matrices respectively; all the components of Λ_1 are 0.2 and those of Λ_2 are 1.2. $\mathbf{f}_t, t = 1, 2, \ldots, n$ are $r \times 1$ random vectors with i.i.d standard Gaussian distributed elements and $\mathbf{u}_t$ and $\mathbf{v}_t, t = 1, 2, \ldots, n$ are independent random vectors whose elements are all standard 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. All the elements of Λ_2 and Λ_3 are generated independently from standard normal distribution in simulation. Table 3 and 4 illustrate the powers of the proposed statistics S_n and T_n respectively, as r increases from 1 to 4. For T_n, we use its modified version in Theorem 5. Results in these tables indicate that for one triple (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 we 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 $\mathbf{y}_t$ can be independent even if $r > 0$ 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 (10.11) are used.

6.4 Uncorrelated but dependent

The construction of (10.32) is based on the idea that the limit of $F^{\mathbf{S}_{\mathbf{xy}}}(x)$ could not be determined from (2.2) 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 $\mathbf{x}_t = (X_{1t}, X_{2t}, \ldots, X_{p_1 t})^T, t = 1, 2, \ldots, n$ be i.i.d normally distributed random vectors with zero means and unit variances. Define $\mathbf{y}_t = (Y_{1t}, Y_{2t}, \ldots, Y_{p_2 t})^T$, $t = 1, 2, \ldots, n$ by $Y_{it} = (X_{it}^{2k} - EX_{it}^{2k}), i = 1, 2, \ldots, \min(p_1, p_2)$ and if $p_1 < p_2$, we let $Y_{jt} = \varepsilon_{jt}, j = p_1 + 1, \ldots, p_2; t = 1, \ldots, n$, where $\varepsilon_{jt}, j = p_1 + 1, \ldots, p_2; t = 1, \ldots, n$ are i.i.d normal distributed random variables and independent with $\mathbf{x}_t$ and k is an positive integer.

Remark 8. *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 7 and Table 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 $k = 5$. For the statistic T_n, since the covariance matrix of $\mathbf{x}$ is an identity matrix, we use the original statistic T_n in Theorem 3.

6.5 ARCH type dependence

The statistic works in the above example because the limit of $F^{\mathbf{S}_{\mathbf{xy}}}$ can not be determined from (2.2) if $\mathbf{x}$ and $\mathbf{y}$ are uncorrelated. However the limit of $F^{\mathbf{S}_{\mathbf{xy}}}(x)$ might be the same as (2.2) when $\mathbf{x}$ and $\mathbf{y}$ are uncorrelated. We consider such an example as follows.

Consider two random vectors $\mathbf{x}_t = (X_{1t}, X_{2t}, \ldots, X_{p_1 t})$ and $\mathbf{y}_t = (Y_{1t}, Y_{2t}, \ldots, Y_{p_2 t})$ as follows:

$$Y_{it} = Z_{it}\sqrt{\alpha_0 + \alpha_1 X_{it}^2}, i = 1, 2, \ldots, \min(p_1, p_2); \tag{6.5}$$

$$if \quad p_1 < p_2, \quad Y_{jt} = Z_{jt}, j = p_1 + 1, \ldots, p_2, \tag{6.6}$$

where $\mathbf{z}_t = (Z_{1t}, Z_{2t}, \ldots, Z_{p_2 t})$ is a random vector consisting of i.i.d elements generated from Normal (0,1) and $\{\mathbf{z}_t : t = 1, \ldots, n\}$ are independent across t; $\mathbf{x}_t = (X_{1t}, X_{2t}, \ldots, X_{p_1 t})$ is also a random vector with i.i.d elements generated from Normal (0,1) and $\{\mathbf{x}_t : t = 1, \ldots, n\}$ are independent across t. Moreover, $\{\mathbf{z}_t : t = 1, \ldots, n$ are independent of $\{\mathbf{x}_t : t = 1, \ldots, n\}$.

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 $\mathbf{S}_{\mathbf{xy}}$, 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}.

Tables 5 and 6 list the powers of the proposed statistics S_n and T_n for testing model (6.5) in several cases, i.e. α_0 and α_1 take different values. For the statistic T_n, since the covariance matrix of $\mathbf{x}$ is an identity matrix, we use the original statistic T_n in Theorem 3. 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}.

7 Empirical applications

As an application of the proposed independence test, we test the cross-sectional dependence of daily stock returns of companies 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 by $\mathbf{x}_t = (x_{1t}, x_{2t}, \ldots, x_{p_1 t})^T$ and $\mathbf{y}_t = (y_{1t}, y_{2t}, \ldots, y_{p_2 t})^T$ respectively. We consider daily stock returns $\mathbf{r}_t^{\mathbf{x}} = (r_{1t}^{\mathbf{x}}, r_{2t}^{\mathbf{x}}, \ldots, r_{p_1 t}^{\mathbf{x}})$ and $\mathbf{r}_t^{\mathbf{y}} = (r_{1t}^{\mathbf{y}}, r_{2t}^{\mathbf{y}}, \ldots, r_{p_2 t}^{\mathbf{y}})$ with $r_{it}^{\mathbf{x}} = \log \frac{x_{it}}{x_{i,t-1}}$, $i = 1, 2, \ldots, p_1$ and $r_{jt}^{\mathbf{y}} = \log \frac{y_{jt}}{y_{j,t-1}}$, $j = 1, 2, \ldots, p_2$. The goal is to test the dependence between $\mathbf{r}_t^{\mathbf{x}}$ and $\mathbf{r}_t^{\mathbf{y}}$.

The proposed test S_n is applied to testing dependence of $\mathbf{r}_t^{\mathbf{x}}$ and $\mathbf{r}_t^{\mathbf{y}}$. 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 $\mathbf{X} = (\mathbf{r}_1^{\mathbf{x}}, \mathbf{r}_2^{\mathbf{x}}, \ldots, \mathbf{r}_{p_1}^{\mathbf{x}})$ and $\mathbf{Y} = (\mathbf{r}_1^{\mathbf{y}}, \mathbf{r}_2^{\mathbf{y}}, \ldots, \mathbf{r}_{p_2}^{\mathbf{y}})$, 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 daily stock returns 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 9, Table 10, and Table 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.

8 Appendix A: Proof of Theorem 1

Throughout this paper, 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.

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

$$\mathbf{B}_n = \tilde{\mathbf{P}}_{\mathbf{y}} \mathbf{P}_{t\mathbf{x}} \tilde{\mathbf{P}}_{\mathbf{y}}, \tag{8.1}$$

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 $p_2 > n$. 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 $\mathbf{T_{xy}}$ satisfy the equality

$$F^{\mathbf{B}_n}(x) = \frac{p_1}{n} F^{\mathbf{T_{xy}}}(x) + \frac{n - p_1}{n} I_{[0,+\infty)}(x). \tag{8.2}$$

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

$$\mathbf{X}^T = \mathbf{X}_1^T + \mathbf{X}_2^T, \tag{8.3}$$

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

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

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 $\tilde{\mathbf{P}}_{\mathbf{y}}$ 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

$$\tilde{\mathbf{P}}_{\mathbf{y}} = \tilde{\mathbf{P}}_{\mathbf{v}} = \mathbf{V}^T(\mathbf{V}\mathbf{V}^T)^{-}\mathbf{V}. \tag{8.4}$$

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

$$\begin{aligned}
& F^{\mathbf{B}_n}(x) = F^{\frac{1}{n}\mathbf{X}_1^T(\frac{1}{n}\mathbf{X}\mathbf{X}^T + t\mathbf{I})^{-1}\mathbf{X}_1}(x) \\
= \; & F^{\frac{1}{n}\mathbf{W}_1^T(\frac{1}{n}\mathbf{W}\mathbf{W}^T + t\mathbf{\Sigma}_{\mathbf{xx}}^{-1})^{-1}\mathbf{W}_1}(x) \\
= \; & \frac{p_1}{n} F^{\left(\frac{1}{n}\mathbf{W}_1\mathbf{W}_1^T + \frac{1}{n}\mathbf{W}_2\mathbf{W}_2^T + t\mathbf{\Sigma}_{\mathbf{xx}}^{-1} - (\frac{1}{n}\mathbf{W}_2\mathbf{W}_2^T + t\mathbf{\Sigma}_{\mathbf{xx}}^{-1})\right)(\frac{1}{n}\mathbf{W}_1\mathbf{W}_1^T + \frac{1}{n}\mathbf{W}_2\mathbf{W}_2^T + t\mathbf{\Sigma}_{\mathbf{xx}}^{-1})^{-1}}(x) + \frac{n - p_1}{n} I_{[0,+\infty)}(x) \\
= \; & \frac{p_1}{n} F^{\mathbf{I} - \left(\frac{1}{n}\mathbf{W}_1\mathbf{W}_1^T(\frac{1}{n}\mathbf{W}_2\mathbf{W}_2^T + t\mathbf{\Sigma}_{\mathbf{xx}}^{-1})^{-1} + \mathbf{I}\right)^{-1}}(x) + \frac{n - p_1}{n} I_{[0,+\infty)}(x).
\end{aligned} \tag{8.5}$$

This, together with (8.2), yields

$$F^{\mathbf{T_{xy}}}(x) = F^{\mathbf{I} - \left(\frac{1}{n}\mathbf{W}_1\mathbf{W}_1^T(\frac{1}{n}\mathbf{W}_2\mathbf{W}_2^T + t\mathbf{\Sigma}_{\mathbf{xx}}^{-1})^{-1} + \mathbf{I}\right)^{-1}}(x). \tag{8.6}$$

If $p_2 \geq n$, then $Rank(\tilde{\mathbf{P}}_{\mathbf{y}}) = tr\tilde{\mathbf{P}}_{\mathbf{y}} = tr\tilde{\mathbf{P}}_{\mathbf{v}} = n$ with probability one by the definition of Moore-Penrose pseudoinverse because $\mathbf{V}\mathbf{V}^T$ has $(p_2 - n)$ zero eigenvalues and from Theorem 1.1 of Rudelson and Vershynin (2010) with probability one

$$\frac{\lambda_{\min}(\mathbf{V}^T\mathbf{V})}{n} \geq \Big(\frac{\sqrt{p_2} - \sqrt{n-1}}{\sqrt{n}}\Big)^2 \frac{1}{n^2}. \tag{8.7}$$

It follows that with probability one

$$\tilde{\mathbf{P}}_{\mathbf{v}} = \mathbf{I}_n \tag{8.8}$$

so that $\mathbf{W}_1 = \mathbf{W}$ and $\mathbf{W}_2 = 0$. Hence $\frac{1}{n}\mathbf{W}_1\mathbf{W}_1^T(\frac{1}{n}\mathbf{W}_2\mathbf{W}_2^T + t\mathbf{\Sigma}_{\mathbf{xx}}^{-1})^{-1} = t^{-1}\frac{1}{n}\mathbf{W}\mathbf{W}^T\mathbf{\Sigma}_{\mathbf{xx}}$. 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 $Rank(\tilde{\mathbf{P}}_{\mathbf{y}}) = tr\tilde{\mathbf{P}}_{\mathbf{y}} = tr\tilde{\mathbf{P}}_{\mathbf{v}} = p_2$ with probability one by an inequality similar to (8.7). Therefore there exists a unitary matrix $\mathbf{U}$ such that with probability one

$$\mathbf{U}^*\tilde{\mathbf{P}}_{\mathbf{y}}\mathbf{U} = diag(1,\ldots,1,0,\ldots,0), \tag{8.9}$$

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

$$\frac{1}{n}\mathbf{W}_1\mathbf{W}_1^T(\frac{1}{n}\mathbf{W}_2\mathbf{W}_2^T + t\mathbf{\Sigma}_{\mathbf{xx}}^{-1})^{-1} \stackrel{d}{=} q_n\mathbf{S}_1\mathbf{S}_{2t}^{-1}, \tag{8.10}$$

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 $\mathbf{S}_1\mathbf{S}_{2t}^{-1}$. In view of (8.6) 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 $\mathbf{S}_{2t}$ is provided in Silverstein and Bai (1995). By taking $\mathbf{S} = \mathbf{S}_1$ and $\mathbf{T} = \mathbf{S}_{2t}^{-1}$, we see that (2.5) follows from (6.1.2) of Bai and Silverstein (2009).

As for the nonGaussian case, write

$$\mathbf{P}_{tx} = \frac{1}{n}\mathbf{X}^T(\frac{1}{n}\mathbf{X}\mathbf{X}^T + t\mathbf{I}_{p_1})^{-1}\mathbf{X} = \frac{1}{n}\mathbf{W}^T(\frac{1}{n}\mathbf{W}\mathbf{W}^T + t\mathbf{\Sigma}_{\mathbf{xx}}^{-1})^{-1}\mathbf{W}. \tag{8.11}$$

Then the proof of Theorem 1 of Yang and Pan (2012) 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 (2.5) of Yang and Pan (2012). In view of (8.4), to replace Gaussian entries in $\mathbf{V}$ by nonGaussian entries, as in (2.1) of Yang and Pan (2012), one can first prove that the Levy distance

$$L^3\Big(F^{\mathbf{P}_{tx}^{1/2}\tilde{\mathbf{P}}_{\mathbf{y}}\mathbf{P}_{tx}^{1/2}}, F^{\mathbf{P}_{t\mathbf{x}}^{1/2}\mathbf{P}_{u\mathbf{y}}\mathbf{P}_{t\mathbf{x}}^{1/2}}\Big) \le \frac{Mu^2}{n}tr(\frac{1}{n}\mathbf{V}\mathbf{V}^T + u\mathbf{I}_{p_2})^{-2} \le Mu^2 \xrightarrow{a.s.} 0, \text{ as } n\to\infty, \textit{ then } u \to 0$$

where $(\mathbf{P}_{t\mathbf{x}}^{1/2})^2 = \mathbf{P}_{t\mathbf{x}}$ and $\mathbf{P}_{u\mathbf{y}} = \frac{1}{n}\mathbf{V}^T(\frac{1}{n}\mathbf{V}\mathbf{V}^T + u\mathbf{I}_{p_2})^{-1}\mathbf{V}$, $u > 0$. Moreover, we see that conclusion (2.5) of Yang and Pan (2012) still holds if we replace $\mathbf{P}_{\mathbf{y}}$ and $\mathbf{P}_{t\mathbf{x}}$ there by $\mathbf{P}_{t\mathbf{x}}^{1/2}$ and $\mathbf{P}_{u\mathbf{y}}$ respectively and check on its argument carefully. Therefore (2.5) of Yang and Pan (2012) ensures that replacing Gaussian entries in $\mathbf{Y}$ by nonGaussian entries does not affect the LSD of $\mathbf{P}_{t\mathbf{x}}^{1/2}\mathbf{P}_{u\mathbf{y}}\mathbf{P}_{t\mathbf{x}}^{1/2}$ when the entries of $\mathbf{X}$ are nonGaussian. The proof is now complete.

9 Appendix B

This Appendix provides some useful lemmas and proofs of Theorems 2-7 in the paper. Throughout this Appendix, 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.

9.1 Some Useful Lemmas

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

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

Lemma 2 (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 $p \geq 2$,*

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

Lemma 3 (Duhamel formula). *Let $\mathbf{M}_1, \mathbf{M}_2$ be $n \times n$ matrices and $t \in \mathbb{R}$. Then we have*

$$e^{(\mathbf{M}_1+\mathbf{M}_2)t} = e^{\mathbf{M}_1 t} + \int_0^t e^{\mathbf{M}_1(t-s)}\mathbf{M}_2 e^{(\mathbf{M}_1+\mathbf{M}_2)s} ds. \tag{9.1}$$

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

$$\frac{d}{dt}e^{\mathbf{A}(t)} = \int_0^1 e^{s\mathbf{A}(t)}\mathbf{A}'(t)e^{(1-s)\mathbf{A}(t)} ds. \tag{9.2}$$

Lemma 4. *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 5. *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*

$$\frac{\partial g(\mathbf{U})}{\partial X_{ij}} = Tr[(\frac{\partial g(\mathbf{U})}{\partial \mathbf{U}})^T \frac{\partial \mathbf{U}}{\partial X_{ij}}]. \tag{9.3}$$

Lemma 6 (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*

$$E\{\xi_\ell\Phi(\xi)\} = E\{\xi_\ell^2\}E\{\Phi'_\ell(\xi)\}, \ \ell = 1, \ldots, p, \tag{9.4}$$

and

$$Var\{\Phi(\xi)\} \leq \sum_{\ell=1}^{p} E\{\xi_\ell^2\}E\{|\Phi'_\ell(\xi)|^2\}. \tag{9.5}$$

Lemma 7 (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*

$$E\{\xi\Phi(\xi)\} = \sum_{\ell=0}^{p} \frac{\kappa_{\ell+1}}{\ell!} E\{\Phi^{(\ell)}(\xi)\} + \varepsilon_p, \tag{9.6}$$

where the remainder term ε_p admits the bound

$$|\varepsilon_p| \leq C_p \int_0^1 E\left|\xi^{p+2}\Phi^{(p+1)}(\xi v)\right|(1-v)^p dv, \ C_p \leq \frac{1+(3+2p)^{p+2}}{(p+1)!}, \tag{9.7}$$

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

Lemma 8 (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*

$$\sum_{k=1}^{\nu} |\lambda_k - \delta_k|^2 \leq tr[(\mathbf{A} - \mathbf{B})(\mathbf{A} - \mathbf{B})^T], \tag{9.8}$$

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

9.2 Proof of Theorem 2

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). However, when the components of $\mathbf{X}$ are non-Gaussian distributed, the eigenvalues of the matrix $\mathbf{S_{xy}}$ do not have a relationship with those of an F-matrix any more. To overcome this difficulty, we employ an interpolation trick first adopted in Lytova and Pastur (2009) and extend the result to the nonGaussian 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.

9.2.1 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 $\tilde{\mathbf{P}}_\mathbf{y}$ 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

$$\mathbf{A}_n = \mathbf{P_y}\mathbf{P_x}\mathbf{P_y}. \tag{9.9}$$

Then we have

$$F^{\mathbf{A}_n}(x) = \frac{p_1}{n} F^{\mathbf{S_{xy}}}(x) + \frac{n-p_1}{n} I_{[0,+\infty)}(x). \tag{9.10}$$

Note that under Assumption 1 $Rank(\tilde{\mathbf{P}}_{\mathbf{y}}) = tr\tilde{\mathbf{P}}_{\mathbf{y}} = p_2$ 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 (8.10) in the paper, we obtain

$$\mathbf{X}_1\mathbf{X}_1^T \stackrel{d}{=} \sum_{k=1}^{p_2} \mathbf{x}_k\mathbf{x}_k^T,\ \mathbf{X}_2\mathbf{X}_2^T \stackrel{d}{=} \sum_{k=p_2+1}^{n} \mathbf{x}_k\mathbf{x}_k^T, \tag{9.11}$$

where $\mathbf{X}_1$ and $\mathbf{X}_2$ are similarly defined as in (8.3) in the paper with $\tilde{\mathbf{P}}_{\mathbf{y}}$ replaced by $\mathbf{P}_{\mathbf{y}}$. As in (8.6) in the paper we conclude that

$$F^{\mathbf{A}_n}(x) = F^{\mathbf{X}_1^T(\mathbf{X}\mathbf{X}^T)^{-1}\mathbf{X}_1}(x) = \frac{p_1}{n}F^{(\mathbf{X}_1\mathbf{X}_1^T+\mathbf{X}_2\mathbf{X}_2^T)^{-1}\mathbf{X}_1\mathbf{X}_1^T}(x) + \frac{n-p_1}{n}I_{[0,+\infty)}(x)$$
$$= \frac{p_1}{n}F^{\left(\mathbf{I}+\mathbf{X}_2\mathbf{X}_2^T(\mathbf{X}_1\mathbf{X}_1^T)^{-1}\right)^{-1}}(x) + \frac{n-p_1}{n}I_{[0,+\infty)}(x). \tag{9.12}$$

This, together with (10.9), yields

$$F^{\mathbf{S}_{\mathbf{xy}}}(x) = F^{\left(\mathbf{I}+\mathbf{X}_2\mathbf{X}_2^T(\mathbf{X}_1\mathbf{X}_1^T)^{-1}\right)^{-1}}(x). \tag{9.13}$$

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

$$F_{\bar{y}_1,\bar{y}_2}(dx) = g_{\bar{y}_1,\bar{y}_2}(x)I_{[a_1,a_2]}(x)dx + (1-\frac{1}{\bar{y}_1})I_{\{\bar{y}_1>1\}}\delta_0(dx), \tag{9.14}$$

where $g_{\bar{y}_1,\bar{y}_2}$ is given in (2.15) in the paper (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 $\mathbf{S}_{\mathbf{xy}}$ 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 (2.13) in the paper can be expressed as

$$\int \phi(\lambda)dG_{p_1,p_2}(\lambda) = \int \phi(\frac{1}{1+\frac{n-p_2}{p_2}\lambda})dp_1[F^{\frac{1}{n-p_2}\mathbf{X}_2\mathbf{X}_2^T(\frac{1}{p_2}\mathbf{X}_1\mathbf{X}_1^T)^{-1}}(\lambda) - F_{\bar{y}_{1n},\bar{y}_{2n}}(\lambda)], \tag{9.15}$$

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.

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

In view of (9.15), 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 2 for the Gaussian distribution ((2.20) in the paper holds in the Gaussian case).

9.2.2 The general case

We next consider the CLT for the general distribution by the interpolation trick. By (10.9), we have

$$\int \phi(\lambda)dG_{p_1,p_2}(\lambda) = n[\int \phi(\lambda)d\big(F_n^{\mathbf{P}_{\mathbf{y}}\mathbf{P}_{\mathbf{x}}\mathbf{P}_{\mathbf{y}}}(\lambda) - F_n^{yxy}(\lambda)\big)], \tag{9.16}$$

where $F_n^{yxy}(\lambda)$ is obtained from the limit, F^{yxy}, of $F^{\mathbf{P}_{\mathbf{y}}\mathbf{P}_{\mathbf{x}}\mathbf{P}_{\mathbf{y}}}$ 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

$$\widetilde{\mathbf{X}}_n = (\tilde{X}_{ij})_{p_1\times n},\ \breve{\mathbf{X}} = (\breve{X}_{ij})_{p_1\times n} \tag{9.17}$$

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

$$P\Big(\mathbf{P_x} \neq \breve{\mathbf{P}}_x\Big) \leq \sum_{i,j=1}^{p_1,n} P\Big(X_{ij} \neq \breve{X}_{ij}\Big) \leq \frac{K}{\varepsilon^4}EX_{11}^4 I_{(|X_{11}|>\sqrt{n}\varepsilon)} \to 0, \tag{9.18}$$

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 $\mathbf{C}_n = \mathbf{P_y P_x P_y}$ 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 8, we have, for each j=1,2,. . . ,s,

$$\begin{aligned}
&\Big|\int \phi_j(x)d\breve{G}_n(x) - \int \phi_j(x)d\widetilde{G}_n(x)\Big| \leq K\sum_{k=1}^n |\lambda_k^{\breve{\mathbf{C}}_n} - \lambda_k^{\widetilde{\mathbf{C}}_n}| \\
\leq\ & \sqrt{n}\Big(\sum_{k=1}^n |\lambda_k^{\breve{\mathbf{C}}_n} - \lambda_k^{\widetilde{\mathbf{C}}_n}|^2\Big)^{1/2} \leq \sqrt{n}\Big(tr(\breve{\mathbf{C}}_n - \widetilde{\mathbf{C}}_n)(\breve{\mathbf{C}}_n - \widetilde{\mathbf{C}}_n)^T\Big)^{1/2} \\
\leq\ & \sqrt{n}\Big(tr(\breve{\mathbf{P}}_x - \widetilde{\mathbf{P}}_x)(\breve{\mathbf{P}}_x - \widetilde{\mathbf{P}}_x)^T\Big)^{1/2},
\end{aligned} \tag{9.19}$$

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

$$(\sigma_{11}^{-1} - 1)^2 = o(n^{-2}), \qquad |E\breve{X}_{11}| = o(n^{-\frac{3}{2}}). \tag{9.20}$$

By the formula

$$\mathbf{A}^{-1} - \mathbf{B}^{-1} = \mathbf{A}^{-1}(\mathbf{B} - \mathbf{A})\mathbf{B}^{-1}, \tag{9.21}$$

we obtain

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

where

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

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 (9.20)

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

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

$$|X_{ij}| \le \sqrt{n}\varepsilon,\ EX_{ij} = 0,\ EX_{ij}^2 = 1. \tag{9.22}$$

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

$$\begin{aligned}&\mathbf{A}_n(s) = \mathbf{P_y}\mathbf{P_x}(s)\mathbf{P_y},\ \mathbf{X}(s) = s^{1/2}\mathbf{X} + (1-s)^{1/2}\hat{\mathbf{X}},\\ &\mathbf{P_x}(s) = \frac{1}{n}\mathbf{X}^T(s)\mathbf{H}^{-1}(s)\mathbf{X}(s), \quad \mathbf{H}^{-1}(s) = (\mathbf{H}(s))^{-1} = (\frac{1}{n}\mathbf{X}(s)\mathbf{X}^T(s))^{-1},\end{aligned}$$

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

$$\begin{aligned} e_n(s,x) &= exp\Big(ixTr\phi(\mathbf{A}_n(s))\Big),\ \ \mathbf{U}(t,s) = e^{it\mathbf{A}_n(s)}, \\ e_n^{\circ}(s,x) &= exp\Big(ix[Tr\phi(\mathbf{A}_n(s)) - n\int \phi(\lambda)dF_n^{yxy}(\lambda)]\Big).\end{aligned} \tag{9.23}$$

By the continuous theorem of characteristic functions and Subsection 9.2.1 it suffices to prove that

$$\hat{R}_n(x) = E\Big(e^{ix\mathcal{N}_n^{\circ}[\phi]}\Big) - E\Big(e^{ix\hat{\mathcal{N}_n^{\circ}}[\phi]}\Big) \to 0, \quad as\ n\to\infty, \tag{9.24}$$

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 $\chi(x):\mathbb{R}\to\mathbb{R}$:

$$\chi(x) = \begin{cases} 1, & |x| \le K_1 n^{-2} \\ 0, & |x| \ge 2K_1 n^{-2}, \end{cases} \tag{9.25}$$

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

To prove (9.24) we first claim that

$$\tilde{R}_n(x) = E\Big(e^{ix\mathcal{N}_n^{\circ}[\phi]}\Big) - E\Big(e^{ix\mathcal{N}_n^{\circ}[\phi]}\chi(\Im(m_n(in^{-2})))\Big) \to 0,\ as\ n\to\infty, \tag{9.26}$$

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

$$\Im(m_n(in^{-2})) = n^{-3}\sum_{i=1}^{n}\frac{1}{\tilde{\lambda}_i^2 + n^{-4}}, \tag{9.27}$$

we conclude that

$$\tilde{\lambda}_{p_1} > \frac{M_2}{n} \quad \text{if} \quad |\Im(m_n(in^{-2}))| \le M_1 n^{-2}, \tag{9.28}$$

where M_1 may be the same as or different from K_1 given in (9.25). From Theorem 9.13 of Bai and Silverstein (2009), under our truncation, we have, for any $x > 0$ and any integers $k \geq 2$,

$$P(\tilde{\lambda}_{p_1} \leq (1-\sqrt{c_1})^2 - x) = O(n^{-k}). \tag{9.29}$$

By (9.27) and taking an appropriate x we have

$$P\Big(|\Im(m_n(in^{-2}))| \leq K_1 n^{-2}\Big) \geq P\Big(\tilde{\lambda}_{p_1} > (1-\sqrt{c_1})^2 - x\Big) = 1 - O(n^{-k}). \tag{9.30}$$

This is equivalent to

$$P\Big(\chi(\Im(m_n(in^{-2}))) = 1\Big) = 1 - O(n^{-k}). \tag{9.31}$$

It follows that

$$\begin{aligned} |\tilde{R}_n(x)| &= |E\Big(e^{ix\mathcal{N}_n^{\circ}[\phi]}\big(1-\chi(\Im(m_n(in^{-2})))\big)\Big)| \\ &\leq P\Big(\chi(\Im(m_n(in^{-2}))) \neq 1\Big) = O(n^{-k}) \to 0, \ \ as\ n \to \infty. \end{aligned} \tag{9.32}$$

Thus (10.56) is true.

Evidently, (10.56) holds as well if $\mathbf{X}$ is replaced by its normal analogue, $\hat{\mathbf{X}}$. In view of (10.56), to prove (9.24), it suffices to prove that

$$R_n(x) = E\Big(e^{ix\mathcal{N}_n^{\circ}[\phi]}\chi\big(\Im(m_n(in^{-2}))\big)\Big) - E\Big(e^{ix\hat{\mathcal{N}}_n^{\circ}[\phi]}\chi\big(\Im(\hat{m}_n(in^{-2}))\big)\Big) \to 0, \quad as\ n \to \infty, \tag{9.33}$$

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 $(\lambda_{\min}^{-r}\chi(\Im(m_n^s(in^{-2})))$ 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 (9.28)-(9.31) 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 (9.28) and (10.50) we have, for any integer $r > 0$,

$$\begin{aligned} E\Big[\frac{\chi(\Im(m_n^s(in^{-2})))}{\lambda_{\min}^r}\Big] &\leq Mn^r \cdot P\big(\frac{M_2}{n} < \lambda_{\min} < (1-\sqrt{c_1})^2 - x\big) \\ &\quad + M\big((1-\sqrt{c_1})^2 - x\big)^{-r} = O(1). \end{aligned} \tag{9.34}$$

We now consider (9.33). In what follows, to simplify notation denote $\chi(\Im(m_n^s(in^{-2})))$ by χ_{ns}. By the inverse Fourier transform

$$\phi(\lambda) = \int e^{it\lambda}\hat{\phi}(t)dt, \tag{9.35}$$

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

$$\begin{aligned} & R_n(x) = \int_0^1 \frac{\partial}{\partial s}E\Big(e_n^{\circ}(s,x)\chi_{ns}\Big)ds \\ = \ & ixe^{-ixn\int\phi(\lambda)dF_n^{yxy}(\lambda)} \times \int_0^1 ds \int \hat{\phi}(\theta)\theta d\theta \cdot E\Big(Tr\mathbf{U}(\theta,s)\mathbf{P_y}\frac{\partial \mathbf{P_x}(s)}{\partial s}\mathbf{P_y}e_n(s,x)\chi_{ns}\Big) \\ & + \int_0^1 E\Big[e_n^{\circ}(s,x)\frac{\partial}{\partial s}\Big(\chi_{ns}\Big)\Big]ds. \end{aligned} \tag{9.36}$$

We next prove that the last term in (9.36) converges to zero.

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

$$\frac{\partial \mathbf{H}^{-1}(s)}{\partial s} = -\mathbf{H}^{-1}(s)\frac{\partial \mathbf{H}(s)}{\partial s}\mathbf{H}^{-1}(s), \tag{9.37}$$

and the chain rule of matrix derivatives, we have

$$\begin{aligned}\frac{\partial \mathbf{P}_{\mathbf{x}}(s)}{\partial s} &= \frac{1}{2n}\mathbf{X}_{ds}^T\mathbf{H}^{-1}(s)\mathbf{X}(s) + \frac{1}{2n}\mathbf{X}^T(s)\mathbf{H}^{-1}(s)\mathbf{X}_{ds} \\ &\quad -\frac{1}{2n^2}\mathbf{X}^T(s)\mathbf{H}^{-1}(s)[\mathbf{X}_{ds}\mathbf{X}^T(s) + \mathbf{X}(s)\mathbf{X}_{ds}^T]\mathbf{H}^{-1}(s)\mathbf{X}(s),\end{aligned} \tag{9.38}$$

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 (10.37) we obtain

$$D_{kj}(\mathbf{H}^{-1}(s)) = -\mathbf{H}^{-1}(s)\mathbf{W}_n(s,k,j)\mathbf{H}^{-1}(s), \qquad D_{kj}(\frac{1}{\sqrt{n}}\mathbf{X}(s)) = \mathbf{e}_k\mathbf{e}_j^T, \tag{9.39}$$

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 5, Lemma 4, (10.51), (10.16), (10.40) and (9.31), we have

$$\begin{aligned} &\Big|E\Big(D_{kj}\chi_{ns}\Big)\Big| = \Big|\frac{1}{n}E\Big(\chi_{ns}^{'}D_{kj}\Im\big(tr(\mathbf{H}(s)-in^{-2}\mathbf{I})^{-1}\big)\Big)\Big| \\ =& \Big|\frac{1}{n}E\Big[\chi_{ns}^{'}\Im\Big(Tr\big[(\mathbf{H}(s)-in^{-2}\mathbf{I})^{-2}\mathbf{W}_n(s,k,j)\big]\Big)\Big]\Big| \\ \le& Mn^7P(\chi_{ns}\neq 1) = O(n^{-k}), \quad \text{for any } \ k, \end{aligned} \tag{9.40}$$

where the last inequality uses the fact that $\chi_{ns}^{'}\neq 0$ occurs only when $K_1n^{-2}\le \Im(m_n^s(n^{-2}))\le 2K_1n^{-2}$. This ensures the last term in (9.36) converges to zero.

In view of (9.36), (9.38) and (9.40) we may write $R_n(x)$ as

$$R_n(x) = \frac{ixe^{-ixn\int \phi(\lambda)dF_n^{yxy}(\lambda)}}{2}\int_0^1 ds\int \hat{\phi}(\theta)\theta d\theta\sum_{i=1}^{2}[Q_n^{(i)} - V_n^{(i)}] + o(1), \tag{9.41}$$

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

$$\Phi_{kj}^{(1)} = \Phi_{kj}^{(1)}(X_{kjs}) = \big(\mathbf{H}^{-1}(s)\frac{1}{\sqrt{n}}\mathbf{X}(s)\mathbf{P}_{\mathbf{y}}\mathbf{U}(\theta,s)\mathbf{P}_{\mathbf{y}}\big)_{kj}e_n(s,x)\chi_{ns}, \quad X_{kjs} = s^{1/2}X_{kj} + (1-s)^{1/2}\hat{X}_{kj}; \tag{9.42}$$

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 (9.41) $\to 0$ 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 6 to the terms $V_n^{(1)}$ and $V_n^{(2)}$ respectively, we can obtain

$$V_n^{(1)} = \frac{1}{n} \sum_{j,k=1}^{n,p_1} E(D_{kj}\Phi_{kj}^{(1)}), \; V_n^{(2)} = \frac{1}{n} \sum_{j,k=1}^{n,p_1} E(D_{kj}\Phi_{kj}^{(2)}). \tag{9.43}$$

Similarly, by generalized stein's equation in Lemma 7 with $p = 3$, we have

$$Q_n^{(i)} = \sum_{\ell=0}^{3} T_{\ell\varepsilon}^{(i)} + \xi_3^{(i)}, \quad i = 1, 2; \tag{9.44}$$

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

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

where $\Phi_{kj}^{(i)}(vX_{kjs})$ is obtained from $\Phi_{kj}^{(i)}(X_{kjs})$ given in (9.42) 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 (9.23) and applying the Duhamel formula (1.2) in the paper and (10.40) we have

$$\begin{aligned} D_{kj}(\mathbf{U}(\theta, s)) &= \int_0^1 e^{it\theta\mathbf{A}_n(s)} D_{kj}\big(i\theta\mathbf{A}_n(s)\big) e^{i(1-t)\theta\mathbf{A}_n(s)} dt \\ &= i\int_0^{\theta} \mathbf{U}(\tau, s) D_{kj}\Big(\mathbf{A}_n(s)\Big)\mathbf{U}(\theta-\tau, s) d\tau \\ &= i\int_0^{\theta} \mathbf{U}(\tau, s)\mathbf{P_y}\mathbf{B}_{ns}\mathbf{P_y}\mathbf{U}(\theta-\tau, s) d\tau, \end{aligned} \tag{9.45}$$

where

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

It follows from (9.35), (9.45) and the chain rule of calculating matrix derivatives that

$$D_{kj}(e_n(s,x)) = -xe_n(s,x)\int \hat{\phi}(\theta)\theta Tr\Big[\mathbf{U}(\theta, s)\mathbf{P_y}\mathbf{B}_{ns}\mathbf{P_y}\Big] d\theta, \tag{9.46}$$

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 (9.42) and (10.40) we have

$$\begin{aligned} D_{kj}(\Phi^{(1)}_{kj}) &= -\mathbf{e}_k^T\mathbf{H}^{-1}(s)\mathbf{W}_n(s,k,j)\mathbf{Q}_{ns}\mathbf{U}(\theta,s)\mathbf{P}_\mathbf{y}\mathbf{e}_j e_n(s,x)\chi_{ns} \qquad (9.47)\\ &+ \mathbf{e}_k^T\mathbf{H}^{-1}(s)\mathbf{e}_k\mathbf{e}_j^T\mathbf{P}_\mathbf{y}\mathbf{U}(\theta,s)\mathbf{P}_\mathbf{y}\mathbf{e}_j e_n(s,x)\chi_{ns} + \mathbf{e}_k^T\mathbf{Q}_{ns}\Big(D_{kj}(\mathbf{U}(\theta,s))\Big)\mathbf{P}_\mathbf{y}\mathbf{e}_j e_n(s,x)\chi_{ns}\\ &+ \mathbf{e}_k^T\mathbf{Q}_{ns}\mathbf{U}(\theta,s)\mathbf{P}_\mathbf{y}\mathbf{e}_j\Big(D_{kj}(e_n(s,x))\Big)\chi_{ns} + \mathbf{e}_k^T\mathbf{Q}_{ns}\mathbf{U}(\theta,s)\mathbf{P}_\mathbf{y}\mathbf{e}_j e_n(s,x)D_{kj}(\chi_{ns}), \end{aligned}$$

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

Although there are many terms in the expansion of $D_{kj}(\Phi^{(1)}_{kj})$, from (9.47), (9.45) and (9.46) we see that each term must be products of some of the factors and their transposes below

$$\mathbf{e}_k^T\mathbf{H}^{-1}(s)\mathbf{e}_k,\ \mathbf{e}_k^T\mathbf{H}^{-1}(s)\frac{1}{\sqrt{n}}\mathbf{X}(s),\ \mathbf{e}_j^T\mathbf{P}_\mathbf{x}(s),\ \mathbf{P}_\mathbf{y},\ \mathbf{U}(\theta,s),\ \chi_{ns},\ \mathbf{e}_k,\ e_n(s,x),\ \mathbf{e}_j,\ D_{kj}(\chi_{ns}). \qquad (9.48)$$

By the facts that $|e_n(s,x)|\le 1$, $|\chi_{ns}|\le M$, $||\mathbf{P}_\mathbf{x}(s)|| = ||\mathbf{P}_\mathbf{y}|| = ||\mathbf{U}(\theta,s)|| = ||\mathbf{e}_k|| = ||\mathbf{e}_j|| = 1$ and (9.48), we conclude from (9.47) that

$$\begin{aligned} \Big|D_{kj}\Phi^{(1)}_{kj}\Big| &\le K||\lambda_{\min}||^{-r}||\mathbf{e}_k^T\mathbf{H}^{-1}(s)\frac{1}{\sqrt{n}}\mathbf{X}(s)||^d|\chi_{ns}| + K||\mathbf{e}_k^T\mathbf{H}^{-1}(s)\frac{1}{\sqrt{n}}\mathbf{X}(s)|||D_{kj}(\chi_{ns})|\\ &\le \frac{K}{\lambda_{\min}^{r+d/2}}|\chi_{ns}| + \frac{K}{\lambda_{\min}}|D_{kj}(\chi_{ns})|, \qquad (9.49) \end{aligned}$$

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 (9.40), (10.55) and (9.28) we see

$$E\Big(\frac{1}{\lambda^2_{\min}}|D_{kj}(\chi_{ns})|^2\Big) \le K, \qquad (9.50)$$

In view of (9.49), (9.52) and (10.55) we conclude that $E|D_{kj}\Phi^{(1)}_{kj}|^2$ is bounded.

We now claim that $E\big(D^\ell_{kj}\Phi^{(1)}_{kj}\big)^2$, $\ell = 2,3,4$ are bounded as well. Indeed, from (9.45) to (9.47) we see that each higher derivative of $E\big(D^1_{kj}\Phi^{(1)}_{kj}\big)$ 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 $D^\ell_{kj}(\chi_{ns})$. From (10.40)-(9.46) we see such derivatives must be formed by some of the factors listed in (9.48) as well as $D^\ell_{kj}(\chi_{ns})$. Here we would point out that the trace involved in (9.46) 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 (9.49), we have for $\ell = 2,3,4$

$$\begin{aligned} \Big|D^\ell_{kj}\Phi^{(\ell)}_{kj}\Big| &\le K||\lambda_{\min}||^{-r_1}||\mathbf{e}_k^T\mathbf{H}^{-1}(s)\frac{1}{\sqrt{n}}\mathbf{X}(s)||^{d_1}\sum_{m=0}^{\ell}|D^m_{kj}(\chi_{ns})|\\ &\le \frac{K}{\lambda_{\min}^{r_1+d_1/2}}\sum_{m=0}^{\ell}|D^m_{kj}(\chi_{ns})|, \qquad (9.51) \end{aligned}$$

where r_1, d_1 are some nonnegative integers, independent of n. Again, from the argument of (9.40), (10.55) and (9.28) one can verify that

$$E\Big(\frac{1}{\lambda_{\min}^{r_1+d_1/2}}|D_{kj}^{\ell}(\chi_{ns})|^2\Big) \le K. \tag{9.52}$$

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

$$E\Big|D_{kj}^{\ell}\Phi_{kj}^{(i)}\Big|^2 \le K, \quad \ell = 1,2,3,4, \quad i = 1,2. \tag{9.53}$$

Consider $\xi_3^{(i)}$ in (9.44) now. Define the event

$$B = \Big(\lambda_{\min} \ge (1-\sqrt{c_1})^2/2\Big). \tag{9.54}$$

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 (9.86) and (10.16) 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 (9.40) one may verify that

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

While (10.41) and (10.50) imply

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

It follows that

$$|\xi_3^{(i)}| \le \frac{K}{n^{5/2}}\sum_{k,j=1}^{n,p_1}\int_0^1 E\Big|X_{kjs}^5 D_{kj}^4\Phi_{kj}^{(i)}(vX_{kjs})\Big|dv \le K\varepsilon \to 0. \tag{9.55}$$

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

$$|\mu_{\ell,kj}^{\varepsilon} - \mu_{\ell,kj}| \le KE|X_{kj}|^{\ell}I(|X_{kj}| > \varepsilon\sqrt{n}) \le \frac{K}{(\sqrt{n}\varepsilon)^{4-\ell}}E|X_{kj}|^4 I(|X_{kj}| > \varepsilon\sqrt{n}). \tag{9.56}$$

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

$$\kappa_\ell = \sum_\lambda c_\lambda\mu_\lambda, \tag{9.57}$$

where the sum is over all additive partitions λ of the set $\{1,\ldots,\ell\}$, $\{c_\lambda\}$ are known coefficients and $\mu_\lambda=\prod_{\ell\in\lambda}\mu_\ell$. We then obtain from (9.56) and (9.57),

$$|\kappa^{\varepsilon}_{\ell,kj}-\kappa_{\ell,kj}|\le\frac{K}{(\sqrt{n}\varepsilon)^{4-\ell}}E|X_{kj}|^4I(|X_{kj}|>\varepsilon\sqrt{n}). \tag{9.58}$$

Recalling the definition of $T^{(1)}_{\ell\varepsilon}$ in (9.2.2), from (9.58) and (10.41) we may write

$$T^{(1)}_{\ell\varepsilon}=T^{(1)}_{\ell}+r^{(1)}_{\ell}, \tag{9.59}$$

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

$$|r^{(1)}_{\ell}| \;\le\; \frac{s^{(\ell-1)/2}}{\ell!n^{(\ell+1)/2}}\sum_{k,j=1}^{p_1,n}|\kappa^{\varepsilon}_{\ell+1,kj}-\kappa_{\ell+1,kj}||E(D^{\ell}_{kj}\Phi^{\varepsilon(i)}_{kj})|\le\frac{K\varepsilon^{\ell-4}}{\sqrt{n}} \tag{9.60}$$

and $T^{(1)}_{\ell}$ is the analogue of $T^{(1)}_{\ell\varepsilon}$ with $\kappa^{\varepsilon}_{\ell,kj}$ replaced by $\kappa_{\ell,kj}$. Note that $T^{(1)}_0=T^{(1)}_3=0,\ T^{(1)}_1=V^{(1)}_n$ because $\kappa_1=\kappa_4=0$. In view of Lemma 9 below, $T^{(1)}_2=o(1)$, and hence

$$Q^{(1)}_n=V^{(1)}_n+\xi^{(1)}_3+o(1). \tag{9.61}$$

With the same proof as above, we can obtain

$$Q^{(2)}_n=V^{(2)}_n+\xi^{(2)}_3+o(1). \tag{9.62}$$

This, together with (9.61), (9.55) and (9.41), completes the proof of this theorem.

Lemma 9. *Under the assumptions of Theorem 2,*

$$T^{(i)}_2=\frac{s^{\frac12}}{2n^{\frac32}}\sum_{j,k=1}^{n,p_1}\kappa^{\varepsilon}_{3,kj}E\Big(D^2_{kj}(\Phi^i_{kj})\Big)=o(1),\ i=1,2, \tag{9.63}$$

as $n\to\infty$.

By taking a further derivative of (9.47) we may obtain the expansion of $D^2_{kj}(\Phi^i_{kj})$. 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

$$C_1=(\mathbf{V}_n(s))_{kj},\ C_2=(\mathbf{V}_n(s)\frac{1}{\sqrt{n}}\mathbf{X}(s)\mathbf{H}^{-1}(s))_{kk},\ C_3=(\mathbf{P_x}(s)\mathbf{P_y}\mathbf{U}\mathbf{P_y})_{jj}, \tag{9.64}$$

$$C_4=(\mathbf{P_y}\mathbf{U}\mathbf{P_y})_{jj},\ C_5=e_n(s,x),\ C_6=(\mathbf{V}_n(s)\mathbf{P_x}(s))_{kj},\ C_7=(\mathbf{H}^{-1}(s))_{kk},\ C_{10}=(\mathbf{P_x}(s))_{jj},$$

$$C_8=(\mathbf{P_x}(s)\mathbf{P_y}\mathbf{U}\mathbf{P_y}\mathbf{P_x}(s))_{jj},\ C_9=(\frac{1}{\sqrt{n}}\mathbf{X}^T(s)\mathbf{H}^{-1}(s))_{jk},\ C_{11}=\chi_{ns},\ C_{12}=D^{\ell}_{kj}(\chi_{ns}),\ell=1,2,$$

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 $D^2_{kj}(s)\Phi^{(1)}_{kj}(s)$ 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 (9.46) and $D_{kj}((\mathbf{H}^{-1}(s))_{kk})$ includes C_9 from (10.40).

Thus, to prove (9.63), it suffices to estimate the following term

$$\frac{1}{n^{3/2}}\sum_{k,j=1}^{p_1,n} E\Big(C_i^{r_i+1}\prod_{h\in D,h\neq i} C_h^{r_h}\Big)=o(1),\ \ i=1,6,9, \tag{9.65}$$

where all $r_h, h\in D=\{1,\cdots,12\}$ are nonnegative integers, independent of n. As in (9.40) one may verify that (9.65) converges to zero if C_{12} is contained in (9.65). Below we consider only the case when C_{12} is not contained in (9.65) and as a result it must contain C_{11}.

We first prove (9.65) holds for the case when there are at least two of $C_i, i=1,6,9$ contained in the expectation sign of (9.65). Moreover for concreteness we consider the case when C_1 and C_6 are both contained in (9.65) 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 (10.55) and (9.49) we obtain

$$\begin{aligned}
&\Big|\frac{1}{n^{3/2}}\sum_{k,j=1}^{p_1,n} E\Big(C_1^{r_1+1}C_6^{r_6+1}C_{11}\prod_{h\in D_1} C_h^{r_h}\Big)\Big|\\
\leq\ & \frac{K}{n^{3/2}}E\Big(\sum_{k,j=1}^{p_1,n}|(\mathbf{V}_n(s))_{kj}|^{2(r_1+1)}\sum_{k,j=1}^{p_1,n}|(\mathbf{V}_n(s)\mathbf{P}_{\mathbf{x}}(s))_{kj}|^{2(r_6+1)}I(B)\Big)^{1/2}\\
&+\frac{Kn^{4+r_1+r_6+r_2+r_7}}{n^{3/2}}P\Big(\frac{M_2}{n}\leq\lambda_{\min}\leq\frac{(1-\sqrt{c_1})^2}{2}\Big)=O(\frac{1}{\sqrt{n}}),
\end{aligned} \tag{9.66}$$

where we also use the fact that recalling the definition of the event B in (9.54),

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

If there is only one of $C_i, i=1,6,9$ contained in (9.65) but its corresponding r_i being greater than zero, then repeating the argument of (9.66) ensures that (9.65) holds. We now consider the case when one of $C_i, i=1,6,9$ is contained in (9.65) but its corresponding r_i equals zero. For concreteness we consider C_1 contained in (9.65) and the remaining cases can be proved similarly. Let $D_2=\{2,\cdots,5,7,8,10\}$. By the Schwartz inequality

$$\begin{aligned}
&\Big|\frac{1}{n^{3/2}}\sum_{k,j=1}^{p_1,n} E\Big(C_1C_{11}\prod_{h\in D_2} C_h^{r_h}\Big)\Big|^2\\
\leq\ & \frac{K}{n^3}E\Big[\sum_{j=1}\big|C_3^{r_3}C_4^{r_4}C_8^{r_8}C_{10}^{r_{10}}\big|^2\sum_{j=1}\Big|\sum_{k}(\mathbf{V}_n(s))_{kj}C_2^{r_2}C_7^{r_7}\Big|^2I(B)\Big]\\
&+\frac{Kn^{8+2r_1+2r_2+2r_7}}{n^3}P\Big(\frac{M_2}{n}\leq\lambda_{\min}\leq\frac{(1-\sqrt{c_1})^2}{2}\Big)\\
\leq\ & \frac{K}{n^2}E\Big[\sum_{j=1}\sum_{k_1,k_2}(\mathbf{V}_n(s))_{k_1j}(\bar{\mathbf{V}}_n(s))_{k_2j}C_{2k_1k_1}^{r_2}C_{7k_1k_1}^{r_7}\bar{C}_{2k_2k_2}^{r_2}\bar{C}_{7k_2k_2}^{r_7}I(B)\Big] \qquad (9.67)\\
&+\frac{Kn^{8+2r_1+2r_2+2r_7}}{n^3}P\Big(\frac{M_2}{n}\leq\lambda_{\min}\leq\frac{(1-\sqrt{c_1})^2}{2}\Big)=O(\frac{1}{\sqrt{n}}), \qquad (9.68)
\end{aligned}$$

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 (9.67) we use the following fact that

$$(9.67) = \frac{K}{n^2}E\Big[\sum_{k_1,k_2}(\bar{\mathbf{V}}_n(s)\mathbf{V}_n^T(s))_{k_2k_1}C^{r_2}_{2k_1k_1}C^{r_7}_{7k_1k_1}\bar{C}^{r_2}_{2k_2k_2}\bar{C}^{r_7}_{7k_2k_2}I(B)\Big]$$

$$\le \frac{K}{n^2}E\Big[\sum_{k_1}|C^{r_2}_{2k_1k_1}C^{r_7}_{7k_1k_1}|^2\sum_{k_1}\Big|\sum_{k_2}(\bar{\mathbf{V}}_n(s)\mathbf{V}_n^T(s))_{k_2k_1}\bar{C}^{r_2}_{2k_2k_2}\bar{C}^{r_7}_{7k_2k_2}\Big|^2 I(B)\Big]^{1/2}$$

$$\le \frac{K}{n^{3/2}}E\Big[\sum_{k_1}\sum_{k_2,k_3}(\bar{\mathbf{V}}_n(s)\mathbf{V}_n^T(s))_{k_2k_1}(\mathbf{V}_n^*(s)\mathbf{V}_n(s))_{k_3k_1}\bar{C}^{r_2}_{2k_2k_2}\bar{C}^{r_7}_{7k_2k_2}C^{r_2}_{2k_3k_3}C^{r_7}_{7k_3k_3}I(B)\Big]^{1/2}$$

$$= \frac{K}{n^{3/2}}E\Big[\sum_{k_2,k_3}(\bar{\mathbf{V}}_n(s)(\mathbf{V}_n^T(s))^2\bar{\mathbf{V}}_n(s))_{k_2k_3}\bar{C}^{r_2}_{2k_2k_2}\bar{C}^{r_7}_{7k_2k_2}C^{r_2}_{2k_3k_3}C^{r_7}_{7k_3k_3}I(B)\Big]^{1/2} \le \frac{K}{\sqrt{n}},$$

where $\mathbf{V}_n^*(s)$ stands for the complex conjugate transpose of $\mathbf{V}_n(s)$. Therefore (9.65) holds for all cases and the proof of Lemma 9 is complete.

9.3 Proof of Theorem 3

9.3.1 The Gaussian case

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

We remind readers that we below use the same notations as those in Theorem 1. Recall $q_n = \frac{p_2}{n-p_2}$. From (8.6) in the paper, we can see that the statistic (2.10) in the paper can be expressed as

$$\int \phi(\lambda)dG^{(2)}_{p_1,p_2}(\lambda) = \int \phi(\frac{q_n\mu}{1+q_n\mu})dp_1[F^{\mathbf{S}_1\mathbf{S}_{2t}^{-1}}(\mu) - \tilde{F}_{y_{1n},y_{2n}}(\mu)], \tag{9.69}$$

where $\tilde{F}_{y_{1n},y_{2n}}(\mu)$ is obtained from $\tilde{F}_{y_1,y_2}(\mu)$, whose stieltjes transform is defined in (2.5) in the paper, 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 (9.69), 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:

$$\underline{m}_{\mathbf{n}}(z) = -\frac{1-y_{n1}}{z} + y_{n1}m_{\mathbf{n}}(z), \quad \underline{m}_{\mathbf{y}}(z) = -\frac{1-y_1}{z} + y_1m_{\mathbf{y}}(z); \tag{9.70}$$

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

$$z = -\frac{1}{\underline{m}_{\mathbf{y}}(z)} + \int \frac{y_1 dF_{y_2t}(x)}{x + \underline{m}_{\mathbf{y}}(z)}. \tag{9.71}$$

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)

$$z = -\frac{1}{\underline{m}_{\mathbf{y}_n}} + \int \frac{y_{n_1} dF_{y_{n_2t}}(x)}{x + \underline{m}_{\mathbf{y_n}}}. \tag{9.72}$$

Write

$$n_1[\underline{m}_{\mathbf{n}}(z) - \underline{m}_{\mathbf{y_n}}(z)] = n_1[\underline{m}_{\mathbf{n}}(z) - \underline{m}^{y_{n_1}, H_{n_2t}}(z)] + n_1[\underline{m}^{y_{n_1}, H_{n_2t}}(z) - \underline{m}_{\mathbf{y_n}}(z)], \tag{9.73}$$

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

$$z = -\frac{1}{\underline{m}^{\{y_{n_1}, H_{n_2t}\}}} + \int \frac{y_{n_1} dF_{n_2t}(x)}{x + \underline{m}^{\{y_{n_1}, H_{n_2t}\}}}. \tag{9.74}$$

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

$$n_1[\underline{m}_{\mathbf{n}} - \underline{m}^{\{y_{n_1}, H_{n_2t}\}}(z)]. \tag{9.75}$$

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 (9.75) 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

$$E(M_1(z)|\mathbf{W}_2) = \frac{y_1 \int \underline{m}(z)^3 x[x + \underline{m}(z)]^{-3} dF_{y_2t}(x)}{[1 - y_1 \int \underline{m}(z)^2 (x + \underline{m}(z))^{-2} dF_{y_2t}(x)]^2} \tag{9.76}$$

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

$$Cov(M_1(z_1), M_1(z_2)|\mathbf{W}_2) = 2\Big(\frac{\underline{m}'(z_1)\underline{m}'(z_2)}{(\underline{m}(z_1) - \underline{m}(z_2))^2} - \frac{1}{(z_1 - z_2)^2}\Big) \tag{9.77}$$

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

Step 2: Consider the limit distribution of

$$n_1[\underline{m}^{\{y_{n_1}, H_{n_2t}\}}(z) - \underline{m}_{\mathbf{y}_n}(z)]. \tag{9.78}$$

By the definition of the Stieltjes transform, rewrite the equations of (9.72) and (9.74) as

$$z = -\frac{1}{\underline{m}_{\mathbf{y}_n}} + y_{n_1} m_{y_{n_2t}}(-\underline{m}_{\mathbf{y}_n}), \ z = -\frac{1}{\underline{m}^{\{y_{n_1}, H_{n_2t}\}}} + y_{n_1} m_{n_2t}\big(-\underline{m}^{\{y_{n_1}, H_{n_2t}\}}\big). \tag{9.79}$$

Taking a difference of the above two identities we obtain

$$\begin{aligned}0 &= \frac{\underline{m}^{\{y_{n_1},H_{n_2t}\}}-\underline{m}_{\mathbf{y}_n}}{\underline{m}_{\mathbf{y}_n}\underline{m}^{\{y_{n_1},H_{n_2t}\}}}+y_{n1}[m_{n_2t}(-\underline{m}^{\{y_{n_1},H_{n_2t}\}})-m_{n_2t}(-\underline{m}_{\mathbf{y}_n})+m_{n_2t}(-\underline{m}_{\mathbf{y}_n})-m_{y_{n_2t}}(-\underline{m}_{\mathbf{y}_n})]\\ &= \frac{\underline{m}^{\{y_{n_1},H_{n_2t}\}}-\underline{m}_{\mathbf{y}_n}}{\underline{m}_{\mathbf{y}_n}\underline{m}^{\{y_{n_1},H_{n_2t}\}}}-y_{n_1}\int\frac{(\underline{m}^{\{y_{n_1},H_{n_2t}\}}-\underline{m}_{\mathbf{y}_n})dF_{n_2t}(x)}{(x+\underline{m}^{\{y_{n_1},H_{n_2t}\}})(x+\underline{m}_{\mathbf{y}_n})}+y_{n_1}[m_{n_2t}(-\underline{m}_{\mathbf{y}_n})-m_{y_{n_2t}}(-\underline{m}_{\mathbf{y}_n})].\end{aligned}$$

From the above equality, we can obtain

$$\begin{aligned}&n_1[\underline{m}^{\{y_{n_1},H_{n_2t}\}}(z)-\underline{m}_{\mathbf{y}_n}(z)]\\ =\;&-y_{n_1}\underline{m}_{\mathbf{y}_n}\underline{m}^{\{y_{n_1},H_{n_2t}\}}\frac{n_1[m_{n_2t}(-\underline{m}_{\mathbf{y}_n})-m_{y_{n_2t}}(-\underline{m}_{\mathbf{y}_n})]}{1-y_{n_1}\int\frac{\underline{m}_{\mathbf{y}_n}\underline{m}^{\{y_{n_1},H_{n_2t}\}}dF_{n_2t}(x)}{(x+\underline{m}_{\mathbf{y}_n})(x+\underline{m}^{\{y_{n_1},H_{n_2t}\}})}}.\end{aligned}\tag{9.80}$$

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 3. By Lemma 2 in the supplementary material, 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

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

and covariance

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

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 (9.80) 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

$$E(M_1(z) + M_3(z)) = \frac{y_1 \int \underline{m}(z)^3 x[x + \underline{m}(z)]^{-3} dF_{y_2 t}(x)}{[1 - y_1 \int \underline{m}(z)^2 (x + \underline{m}(z))^{-2} dF_{y_2 t}(x)]^2}$$
$$+h(z)\frac{y_2 \varpi^2(-\underline{m}(z)) m_3(-\underline{m}(z)) + y_2^2 \varpi^4(-\underline{m}(z)) m'_{y_2 t}(-\underline{m}(z)) m_3(-\underline{m}(z))}{1 - c\varpi^2(-\underline{m}(z)) m_2(-\underline{m}(z))}$$
$$-h(z)\frac{y_2^2 \varpi^3(-\underline{m}(z)) m'_{y_2 t}(-\underline{m}(z)) m_2(-\underline{m}(z))}{1 - y_2 \varpi^2(-\underline{m}(z)) m_2(-\underline{m}(z))} \tag{9.81}$$

and covariance function

$$Cov(M_1(z_1) + M_3(z_1), M_1(z_2) + M_3(z_2)) = 2\Big(\frac{\underline{m}'(z_1)\underline{m}'(z_2)}{(\underline{m}(z_1) - \underline{m}(z_2))^2} - \frac{1}{(z_1 - z_2)^2}\Big)$$
$$-\frac{2h(z_1)h(z_2)}{(-\underline{m}(z_2) + \underline{m}(z_1))^2} + \frac{h(z_1)h(z_2)2[1 + g(z_1) + g(z_2) + g(z_1)g(z_2)]}{[-\underline{m}(z_2) + \underline{m}(z_1) + s(-\underline{m}(z_1), -\underline{m}(z_2))]^2}. \tag{9.82}$$

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

$$\int f(x) dG^{(2)}_{p_1,p_2}(x) = -\frac{1}{2\pi i}\int f(z) m_G(z) dz. \tag{9.83}$$

Then

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

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

$$EX_{f_i} = -\frac{1}{2\pi i}\oint f_i(z) E(M_1(z) + M_3(z)) dz \tag{9.84}$$

and

$$Cov(X_{f_i}, X_{f_j}) = -\frac{1}{4\pi^2}\oint\oint f_i(z_1) f_j(z_2) Cov(M_1(z_1) + M_3(z_1), M_1(z_2) + M_3(z_2)) dz_1 dz_2. \tag{9.85}$$

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 Proof of Theorem 2 in Appendix B (replacing $\mathbf{P_x}$ there by $\mathbf{P}_{t\mathbf{x}}$ in (8.11) in the paper). We omit the details here.

9.4 The proof of Theorem 4

Since it is difficult to get an explicit expression for $\int \phi(\lambda) dF^{c'_{1n}, c'_{2n}}(\lambda)$ directly from (2.5) and (2.3) in the paper, we below develop its alternative expression when $\phi(\lambda) = \lambda$. In view of Theorems 1 and 3 it is enough to consider normal random variables when deriving such an expression below.

Consider the case when $c'_2 < 1$. As in (8.3)-(8.5) and the equalities above (8.10) we write

$$\int \lambda dF^{\mathbf{T_{xy}}}(\lambda) = \frac{1}{p_1} tr \mathbf{T_{xy}} = \frac{1}{p_1} tr \tilde{\mathbf{P}}_{\mathbf{y}} \mathbf{P}_{t\mathbf{x}} \tilde{\mathbf{P}}_{\mathbf{y}} = \frac{1}{p_1} tr \frac{1}{n} \mathbf{W}_1 \mathbf{W}_1^T \mathbf{A}^{-1}(t) \tag{9.86}$$

$$= \frac{1}{np_1}\sum_{k=1}^{p_2}\mathbf{w}_k^T\mathbf{A}^{-1}(t)\mathbf{w}_k = \frac{1}{np_1}\sum_{k=1}^{p_2}\frac{\mathbf{w}_k^T\mathbf{A}_k^{-1}(t)\mathbf{w}_k}{1+n^{-1}\mathbf{w}_k^T\mathbf{A}_k^{-1}(t)\mathbf{w}_k}$$

$$= \frac{p_2}{p_1} - \frac{1}{p_1}\sum_{k=1}^{p_2}\frac{1}{1+\frac{1}{n}tr\mathbf{A}_k^{-1}(t)} + o_p(1) = \frac{p_2}{p_1} - \frac{p_2}{p_1}\frac{1}{1+\frac{1}{n}tr\mathbf{A}^{-1}(t)} + o_p(1),$$

where $\mathbf{A}^{-1}(t) = (\frac{1}{n}\mathbf{W}_1\mathbf{W}_1^T + \frac{1}{n}\mathbf{W}_2\mathbf{W}_2^T + t\mathbf{\Sigma}_{\mathbf{xx}}^{-1})^{-1}$, $\mathbf{A}_k^{-1}(t) = (\frac{1}{n}\mathbf{W}_{1k}\mathbf{W}_{1k}^T + \frac{1}{n}\mathbf{W}_2\mathbf{W}_2^T + t\mathbf{\Sigma}_{\mathbf{xx}}^{-1})^{-1}$, $\mathbf{W}_{1k}\mathbf{W}_{1k}^T + \mathbf{w}_k\mathbf{w}_k^T = \mathbf{W}_1\mathbf{W}_1^T$ and the proof of the last two steps is straightforward. Moreover, denote the limit of $\frac{1}{p_1}tr\mathbf{A}^{-1}(t)$ by m_t. Then from (1.5) in the supplement material or using an argument similar to (4.5) in Silverstein and Bai (1995) we have

$$m_t = \int \frac{dH_1(\lambda)}{\lambda + \frac{1}{1+c_1' m_t}},$$

where $H_1(\lambda)$ stands for the limit of the ESD of $t\mathbf{\Sigma}_{\mathbf{xx}}^{-1}$. Replacing $H_1(\lambda)$ by $F^{t\mathbf{\Sigma}_{\mathbf{xx}}^{-1}}$ and c_1' by c_{1n}' yields

$$m_{nt} = \int \frac{dF^{t\mathbf{\Sigma}_{\mathbf{xx}}^{-1}}}{\lambda + a_n^{-1}} = \frac{1}{p_1}tr(t\mathbf{\Sigma}_{\mathbf{xx}}^{-1} + a_n^{-1}\mathbf{I})^{-1} = a_n - \frac{a_n t}{p_1}tr(a_n^{-1}\mathbf{\Sigma}_{\mathbf{xx}} + t\mathbf{I})^{-1} \tag{9.87}$$

with $a_n = 1 + c_{1n}' m_{nt}$. It follows that

$$\int \lambda dF^{c_{1n}', c_{2n}'}(\lambda) = \frac{p_2}{p_1} - \frac{p_2}{p_1}\frac{1}{1+c_{1n}' m_{nt}}. \tag{9.88}$$

In view of (10.40) it is enough to look at m_{nt}. Let $\mathbf{B}_{\mathbf{\Sigma}}^{-1} = (a_n^{-1}\mathbf{\Sigma}_{\mathbf{xx}} + t\mathbf{I})^{-1}$ and $\hat{\mathbf{B}}_{\mathbf{\Sigma}}^{-1} = ((1+c_{1n}'\hat{m}_{nt})^{-1}\hat{\mathbf{\Sigma}}_{\mathbf{xx}} + t\mathbf{I})^{-1}$. From (3.3) and (10.37) we have

$$p_1(1-c_{1n}')(m_{nt} - \hat{m}_{nt}) = -ta_n tr\mathbf{B}_{\mathbf{\Sigma}}^{-1} + t\hat{a}_n tr\hat{\mathbf{B}}_{\mathbf{\Sigma}}^{-1},$$

which further implies that

$$p_1 d_n(m_{nt} - \hat{m}_{nt}) = -ta_n^{-1}\hat{a}_n tr\hat{\mathbf{B}}_{\mathbf{\Sigma}}^{-1}\Big(\hat{\mathbf{\Sigma}}_{\mathbf{xx}} - \mathbf{\Sigma}_{\mathbf{xx}}\Big)\mathbf{B}_{\mathbf{\Sigma}}^{-1}, \tag{9.89}$$

where $d_n = 1 - c_{1n}' + c_{1n}' t\frac{1}{p_1}tr\mathbf{B}_{\mathbf{\Sigma}}^{-1} + \frac{c_{1n}' t\hat{m}_{nt}}{(1+c_{1n}'\hat{m}_{nt})(1+c_{1n}' m_{nt})}\frac{1}{p_1}tr\mathbf{B}_{\mathbf{\Sigma}}^{-1}\hat{\mathbf{\Sigma}}_{\mathbf{xx}}\hat{\mathbf{B}}_{\mathbf{\Sigma}}^{-1}$. It is straightforward to verify that $d_n > 1 - c_{1n}'$ and d_n converges to a nonzero number in probability when $c_1' = 1$.

Consider the term on the right hand of (9.89). Write

$$tr\hat{\mathbf{B}}_{\mathbf{\Sigma}}^{-1}\Big(\mathbf{\Sigma}_{\mathbf{xx}} - \hat{\mathbf{\Sigma}}_{\mathbf{xx}}\Big)\mathbf{B}_{\mathbf{\Sigma}}^{-1} = \sum_{i\neq j}^{p}\Big(\sigma_{ij} - \hat{\sigma}_{ij}I(|\hat{\sigma}_{ij}| \geq \ell)\Big)a_{ji}, \tag{9.90}$$

where $a_{ij} = \mathbf{e}_j^T\mathbf{B}_{\mathbf{\Sigma}}^{-1}\hat{\mathbf{B}}_{\mathbf{\Sigma}}^{-1}\mathbf{e}_i$. Note that $|a_{ij}| \leq \|\hat{\mathbf{B}}_{\mathbf{\Sigma}}^{-1}\|\|\mathbf{B}_{\mathbf{\Sigma}}^{-1}\|$, bounded in probability uniformly in i and j. Write

$$\hat{\sigma}_{ij} - \sigma_{ij} = \frac{1}{n}\sum_{k=1}^{n}X_{ik}X_{jk} - \sigma_{ij}.$$

Define $\tilde{X}_{ij} = X_{ij}I(|X_{ij}| \leq (n/\log n)^{1/4})$, $\check{X}_{ij} = \tilde{X}_{ij} - E\tilde{X}_{ij}$ and denote by $\tilde{\sigma}_{ij}, \check{\sigma}_{ij}$ the analogues of $\hat{\sigma}_{ij}$ with all entries $\{X_{ij}\}$ of $\mathbf{X}$ replaced by $\tilde{X}_{ij}$ and $\check{X}_{ij}$, respectively. Then one may verify that

$$P\Big(\max_{i\neq j}|\sigma_{ij} - \hat{\sigma}_{ij}| \neq \max_{i\neq j}|\sigma_{ij} - \tilde{\sigma}_{ij}|\Big) \leq \sum_{i,j}P(|X_{ij}| \geq (n/\log n)^{1/4}) = o(n^{-2}).$$

and

$$\Big|\max_{i\neq j}|\sigma_{ij}-\tilde{\sigma}_{ij}|-\max_{i\neq j}|\sigma_{ij}-\check{\sigma}_{ij}|\Big|\le\max_{i\neq j}|\frac{1}{n}\sum_{k=1}^{n}\tilde{X}_{ik}\tilde{X}_{jk}-\frac{1}{n}\sum_{k=1}^{n}\check{X}_{ik}\check{X}_{jk}|=o_p(n^{-2}),$$

by the fact that $EX_{11}(|X_{11}|\ge(n/\log n)^{1/4})=o(n^{-4}(\log n)^4)$. Thus it is enough to consider $\check{X}_{ij}$. By the Bernstein inequality one may obtain

$$P\Big(\max_{i\neq j}\Big|\frac{1}{\sqrt{n}}\sum_{k=1}^{n}(\check{X}_{ik}\check{X}_{jk}-\sigma_{ij})\Big|\ge C\sqrt{\log p}\Big)\le$$

$$P\Big(\max_{i\neq j}\Big|\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\big(\check{X}_{ik}\check{X}_{jk}-E(\check{X}_{ik}\check{X}_{jk})\big)\Big|\ge C\sqrt{\log p}\Big)+C(\log n)^{15/8}n^{-11/8}=O((\log n)^{15/8}n^{-11/8}).\tag{9.91}$$

In view of (9.91), mimicking the proof of Theorem 1 in Bickel and Levina (2008) (replacing $\max\limits_{i}$ in formulas (13)-(22) in their paper by $\sum\limits_{i}$), one may verify that

$$\sum_{i\neq j}^{p}\Big(\sigma_{ij}-\hat{\sigma}_{ij}I(|\hat{\sigma}_{ij}|\ge\ell)\Big)a_{ji}=O_p\Big(s_o(p)\Big(\frac{\log p}{n}\Big)^{(1-q)/2}\Big).$$

It follows that (9.90) tends to zero in probability. Hence Theorem 4 follows from Theorem 3 and Slutsky's theorem.

Consider the case when $p_2>n$. From (9.86) and the paragraph containing (8.7) in the paper, we see that

$$\int\lambda dF^{\mathbf{T}_{x,y}}(\lambda)=\frac{1}{p_1}tr\frac{1}{n}\mathbf{W}\mathbf{W}^T(\frac{1}{n}\mathbf{W}\mathbf{W}^T+t\mathbf{\Sigma}_{\mathbf{xx}}^{-1})^{-1}=1-\frac{t}{p_1}tr(\frac{1}{n}\mathbf{\Sigma}_{\mathbf{xx}}^{1/2}\mathbf{W}\mathbf{W}^T\mathbf{\Sigma}_{\mathbf{xx}}^{1/2}+t\mathbf{I})^{-1}\tag{9.92}$$

$$\xrightarrow{i.p.}1-tm^{(1t)},$$

where the last step uses formula (6.1.2) in Bai and Silverstein (2009) (or one may verify it directly) and $m^{(1t)}$ satisfies the equation

$$m^{(1t)}=\int\frac{dH(\lambda)}{\lambda(1-c_1'+c_1'tm^{(1t)})+t}.\tag{9.93}$$

It follows that

$$\int\lambda dF^{c_{1n}',c_{2n}'}(\lambda)=1-tm_n^{(1t)},\tag{9.94}$$

where $m_n^{(1t)}$ satisfies the equation

$$m_n^{(1t)}=\int\frac{dH_n(\lambda)}{\lambda(1-c_{1n}'+c_{1n}'tm_n^{(1t)})+t}=\frac{1}{p_1}tr\Big((1-c_{1n}'+c_{1n}'tm_n^{(1t)})\mathbf{\Sigma}_{\mathbf{xx}}+t\mathbf{I}\Big)^{-1}.$$

Note that

$$\hat{m}_n^{(1t)}=\frac{1}{p_1}tr\Big((1-c_{1n}'+c_{1n}'t\hat{m}_n^{(1t)})\hat{\mathbf{\Sigma}}_{\mathbf{xx}}+t\mathbf{I}\Big)^{-1}.\tag{9.95}$$

As in the case of $c_2'<1$, one may verify that

$$p_1(m_n^{(1t)}-\hat{m}_n^{(1t)})\xrightarrow{i.p.}0$$

so that

$$p_1\Big(\int\lambda dF^{c_{1n}',c_{2n}'}(\lambda)-(1-t\hat{m}_n^{(1t)})\Big)\xrightarrow{i.p.}0.$$

9.5 The proof of Theorem 5

We below only prove the case when $c_2^{'} < 1/2$ because the remaining cases are similar. For Group 1 and Modified Group 2, by part a) of Theorem 3, we have

$$\int \lambda dF^{\mathbf{T}_{\mathbf{xy}}^{(1)}}(\lambda) - \int \lambda dF^{2c_{1n}^{'},2c_{2n}^{'}}(\lambda) \xrightarrow{d} \mathcal{N}(\mu_1, \sigma_1^2), \tag{9.96}$$

and

$$\int \lambda dF^{\mathbf{T}_{\mathbf{xy}}^{(2)}}(\lambda) - \int \lambda dF^{2c_{1n}^{'},c_{2n}^{'}}(\lambda) \xrightarrow{d} \mathcal{N}(\mu_2, \sigma_2^2). \tag{9.97}$$

In view of (3.1) we have

$$\int \lambda dF^{2c_{1n}^{'},2c_{2n}^{'}}(\lambda) = 2\int \lambda dF^{2c_{1n}^{'},c_{2n}^{'}}(\lambda). \tag{9.98}$$

It follows from the independence between $\mathbf{T}_{\mathbf{xy}}^{(1)}$ and $\mathbf{T}_{\mathbf{xy}}^{(2)}$ that

$$\int \lambda dF^{\mathbf{T}_{\mathbf{xy}}^{(1)}}(\lambda) - 2\int \lambda dF^{\mathbf{T}_{\mathbf{xy}}^{(2)}}(\lambda) \xrightarrow{d} \mathcal{N}(\mu_1 - 2\mu_2, \sigma_1^2 + 4\sigma_2^2).$$

9.6 The proofs of Theorems 6 and 7

We only prove Theorem 6 because the proof of Theorem 7 is similar. Set

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

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

$$\begin{aligned} \beta_n &= P\Big(R_n^{(i)} > z_{1-\alpha} \ or \ R_n^{(i)} < z_\alpha \Big| \mathbb{H}_1\Big) \\ &= P\Big(D^{(i)} + R_n^{(i)0} > z_{1-\alpha} \ or \ D^{(i)} + R_n^{(i)0} < z_\alpha \Big| \mathbb{H}_1\Big) \\ &= P\Big(R_n^{(i)0} > z_{1-\alpha} - D^{(i)} \ or \ R_n^{(i)0} < z_\alpha - D^{(i)} \Big| \mathbb{H}_1\Big), \end{aligned} \tag{9.99}$$

where $R_n^{(i)0}$ stands for $R_n^{(i)}$ under $\mathbb{H}_0$. Under the condition (4.1), we have

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

10 Supplement material: CLT for a sample covariance matrix plus a perturbation

This supplement material is to provide the central limit theorem for linear spectral statistics, quantities of the form

$$\frac{1}{n}\sum_{j=1}^{n} f(\lambda_j) = \int f(x)dF^{\mathbf{B}_n}(x), \tag{10.1}$$

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

$$\mathbf{B}_n = \frac{1}{N}\mathbf{X}\mathbf{X}^* + \mathbf{T}_n. \tag{10.2}$$

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

$$F^{\mathbf{A}}(x) = \frac{1}{n}\#\{i : \mu_i \le x\}, \tag{10.3}$$

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 10 below for easy reference. The Stieltjes transform of any distribution function $G(x)$ is defined by

$$m_G(z) \equiv \int \frac{1}{\lambda - z}dG(\lambda), \quad \Im(z) \neq 0. \tag{10.4}$$

Lemma 10. *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

$$m^0(z) = m_H\Big(z - \frac{1}{1 + cm^0(z)}\Big), \tag{10.5}$$

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

Remark 9. *Indeed, Silverstein (1995) derives a more general equation than (10.5) 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 (10.5) 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 8, we introduce some notation. Set

$$G_n(x) = n[F^{\mathbf{B}_n}(x) - F_{c_n,H_n}(x)], \tag{10.6}$$

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

$$\begin{aligned} m_r(z) &= \int \frac{dH(x)}{(x - z + \varpi(z))^r}, \qquad \varpi(z) = \frac{1}{1 + cm^0(z)}, \\ s(z_1, z_2) &= \frac{1}{1 + cm^0(z_1)} - \frac{1}{1 + cm^0(z_2)}. \end{aligned} \tag{10.7}$$

where r is a positive integer.

The main result of this supplement material is Theorem 8.

Theorem 8. *Assume that*
(a) $\{X_{ij}, i \le n, j \le N\}$ *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 $f_1, \dots, f_k$ *be functions on* $\mathbb{R}$ *analytic on an open interval containing*

$$\Big[I_{(0,1)}(c)(1-\sqrt{c})^2 + \liminf_n \lambda_{min}^{\mathbf{T}_n}, (1+\sqrt{c})^2 + \limsup_n \lambda_{max}^{\mathbf{T}_n}\Big], \tag{10.8}$$

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

$$\Big(\int f_1(x) dG_n(x), \dots, \int f_k(x) dG_n(x)\Big) \tag{10.9}$$

forms a tight sequence in n.
(ii) If X_{11} *and* $\mathbf{T}_n$ *are real and* $EX_{11}^4 = 3$, *then (10.9) converges weakly to a Gaussian vector* $(X_{f_1}, \dots, X_{f_k})$ *with mean*

$$EX_f = \frac{1}{-2\pi i} \oint_{\mathcal{C}} f(z) \frac{c\varpi^2(z)m_3(z) + c^2\varpi^4(z)\big(m^0(z)\big)' m_3(z) - c^2\varpi^3(z)\big(m^0(z)\big)' m_2(z)}{1 - c\varpi^2(z)m_2(z)} dz \tag{10.10}$$

and covariance function

$$\begin{aligned} Cov(X_{f_i}, X_{f_j}) &= -\frac{1}{2\pi^2} \oint_{\mathcal{C}} \oint_{\mathcal{C}} f_i(z_1) f_j(z_2) [1 + \frac{c(m^0(z_1))'}{(1 + cm^0(z_1))^2} + \frac{c(m^0(z_2))'}{(1 + cm^0(z_2))^2} \\ &\quad + \frac{c(m^0(z_1))'}{(1 + cm^0(z_1))^2} \frac{c(m^0(z_2))'}{(1 + cm^0(z_2))^2}] \frac{1}{(z_2 - z_1 + s(z_1, z_2))^2} dz_1 dz_2. \end{aligned} \tag{10.11}$$

The contours in (10.10) and (10.11) 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 (10.11).*

Remark 10. *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 8 are the same, which is verified in the last part of the supplement material.*

10.1 Proof of Theorem 8

The proof of Theorem 8 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.

10.2 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

$$\delta_n \to 0, \quad \delta_n^{-4}\int_{\{|X_{11}|\geq \delta_n\sqrt{n}\}} |X_{11}|^4 \to 0. \tag{10.12}$$

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

$$\mathbf{B}_n = \mathbf{F}_n\mathbf{F}_n^*, \tag{10.13}$$

where

$$\mathbf{F}_n = (\mathbf{r}_1, \ldots, \mathbf{r}_N, \mathbf{t}_1, \ldots, \mathbf{t}_n) \tag{10.14}$$

with $\mathbf{r}_i = \frac{1}{N}\mathbf{X}_{\cdot i}$, $i = 1, \ldots, N$ and $\mathbf{X}_{\cdot i}$ 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 $j = 1, 2, \ldots, k$,

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

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.

10.3 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

$$\lambda_{max}^{\mathbf{B}_n} \le \lambda_{max}^{\mathbf{C}_n} + \lambda_{max}^{\mathbf{T}_n}, \quad \lambda_{min}^{\mathbf{B}_n} \ge \lambda_{min}^{\mathbf{C}_n} + \lambda_{min}^{\mathbf{T}_n}. \tag{10.15}$$

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

$$P(\lambda_{max}^{\mathbf{B}_n} \ge \eta) = o(n^{-\ell}), \quad P(\lambda_{min}^{\mathbf{B}_n} \le \theta) = o(n^{-\ell}), \tag{10.16}$$

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

$$f_\ell(x) = \frac{1}{2\pi i}\oint \frac{f_\ell(z)}{z-x} dz, \tag{10.17}$$

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

$$\sum_{\ell=1}^k a_\ell \int f_\ell(x) dG_n(x) = -\sum_{\ell=1}^k \frac{a_\ell}{2\pi i}\oint_{\mathcal{C}} f_\ell(z) M_n(z) dz, \tag{10.18}$$

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 (10.8). Let x_ℓ be any negative number if the left end point of (10.8) 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

$$\hat{M}_n(z) = \begin{cases} M_n(z), \quad if\ z \in \mathcal{C}_n^+ \cup \mathcal{C}_n^-; \\ \frac{nv+\varepsilon_n}{2\varepsilon_n} M_n(x_r + in^{-1}\varepsilon_n) + \frac{\varepsilon_n - nv}{2\varepsilon_n} M_n(x_r - in^{-1}\varepsilon_n), \\ \quad if\ x = x_r, v \in [-n^{-1}\varepsilon_n, n^{-1}\varepsilon_n]; \\ \frac{nv+\varepsilon_n}{2\varepsilon_n} M_n(x_l + in^{-1}\varepsilon_n) + \frac{\varepsilon_n - nv}{2\varepsilon_n} M_n(x_l - in^{-1}\varepsilon_n), \\ \quad if\ x = x_\ell > 0, v \in [-n^{-1}\varepsilon_n, n^{-1}\varepsilon_n]. \end{cases} \tag{10.19}$$

With probability one, for all n large,

$$\begin{aligned} & \Big| \oint_{\mathcal{C}} f(z)\big(M_n(z) - \hat{M}_n(z)\big)dz \Big| \\ \leq \; & K\varepsilon_n \Big(\Big| \max\big((1+\sqrt{c_n})^2 + \lambda_{max}^{\mathbf{T}_n}, \lambda_{max}^{\mathbf{B}_n}\big) - x_r \Big|^{-1} \Big) \\ & + \Big| \min\Big(I_{(0,1)}(c)(1-\sqrt{c})^2 + \lambda_{min}^{\mathbf{T}_n}, \lambda_{min}^{\mathbf{B}_n}\Big) - x_\ell \Big|^{-1} \to 0. \end{aligned} \tag{10.20}$$

In view of this and (10.18), as discussed in Bai and Silverstein (2004), it is enough to consider the limiting distribution of $\sum_{\ell=1}^k a_\ell \hat{M}_n(z_\ell)$.

10.4 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 11. *Under conditions (a)-(c) of Theorem 8, $\{\hat{M}_n(z)\}$ forms a tight sequence on $\mathcal{C}$. Moreover, if assumptions in (ii) or (iii) of Theorem 8 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),*

$$\mathbb{E}M(z) = \frac{c\varpi^2(z)m(z,3) + c^2\varpi^4(z)\big(m^0(z)\big)' m(z,3) - c^2\varpi^3(z)\big(m^0(z)\big)' m(z,2)}{1 - c\varpi^2(z)m(z,2)} \tag{10.21}$$

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

$$\begin{aligned} Cov(M(z_1), M(z_2)) &= -\frac{2}{(z_2 - z_1)^2} + 2[1 + \frac{c(m^0(z_1))'}{(1 + cm^0(z_1))^2} + \frac{c(m^0(z_2))'}{(1 + cm^0(z_2))^2} \\ &\quad + \frac{c(m^0(z_1))'}{(1 + cm^0(z_1))^2} \frac{c(m^0(z_2))'}{(1 + cm^0(z_2))^2}] \frac{1}{(z_2 - z_1 + s(z_1, z_2))^2}, \end{aligned} \quad (10.22)$$

while under the assumptions in (iii) $\mathbb{E}M(z) = 0$, and the covariance function similar to (10.22) is half of the right hand side of (10.22).

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*

$$\begin{aligned} &\Big|\mathbb{E}\Big(\prod_{k=1}^{p} \mathbf{r}_1^* \mathbf{A}_k \mathbf{r}_1 \prod_{\ell=1}^{q} (\mathbf{r}_1^* \mathbf{B}_\ell \mathbf{r}_1 - N^{-1} tr \mathbf{B}_\ell)\Big)\Big| \\ \le\ & K N^{-(1\wedge q)} \delta_n^{(2q-4)\vee 0} \prod_{k=1}^{p} ||\mathbf{A}_k|| \prod_{\ell=1}^{q} ||\mathbf{B}_\ell||, \quad p \ge 0, \quad q \ge 0. \end{aligned} \quad (10.23)$$

Proof. We now start the proof of Lemma 11. 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.

10.4.1 Step 1: Convergence of $M_n^{(1)}(z)$

Let $v_0 = \Im(z)$. To facilitate analysis we consider the case of $v_0 > 0$ only. We first introduce some notation as follows.

$$\mathbf{r}_j = \frac{1}{\sqrt{N}}\mathbf{X}_{\cdot j}, \ \mathbf{D}(z) = \mathbf{B}_n - z\mathbf{I}, \ \mathbf{D}_j(z) = \mathbf{D}(z) - \mathbf{r}_j\mathbf{r}_j^*, \ \gamma_j(z) = \mathbf{r}_j^*\mathbf{D}_j^{-1}(z)\mathbf{r}_j - \frac{1}{N}Etr\mathbf{D}_j^{-1}(z)$$

$$\varepsilon_j(z) = \mathbf{r}_j^*\mathbf{D}_j^{-1}(z)\mathbf{r}_j - \frac{1}{N}tr\mathbf{D}_j^{-1}(z), \ \ \delta_j(z) = \mathbf{r}_j^*\mathbf{D}_j^{-2}(z)\mathbf{r}_j - \frac{1}{N}tr\mathbf{D}_j^{-2}(z) = \frac{d}{dz}\varepsilon_j(z),$$

$$\beta_j(z) = \frac{1}{1 + \mathbf{r}_j^*\mathbf{D}_j^{-1}(z)\mathbf{r}_j}, \quad \beta_j^{tr}(z) = \frac{1}{1 + N^{-1}tr\mathbf{D}_j^{-1}(z)}, \quad b_n(z) = \frac{1}{1 + N^{-1}\mathbb{E}tr\mathbf{D}_1^{-1}(z)}.$$

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

$$\begin{aligned} n\Big(m_n(z) - \mathbb{E}m_n(z)\Big) &= tr\Big(\mathbf{D}^{-1}(z) - \mathbb{E}\mathbf{D}^{-1}(z)\Big) = \sum_{j=1}^{N} tr\mathbb{E}_j\mathbf{D}^{-1}(z) - tr\mathbb{E}_{j-1}\mathbf{D}^{-1}(z) \\ &= \sum_{j=1}^{N} tr\mathbb{E}_j\Big(\mathbf{D}^{-1}(z) - \mathbf{D}_j^{-1}(z)\Big) - tr\mathbb{E}_{j-1}\Big(\mathbf{D}^{-1}(z) - \mathbf{D}_j^{-1}(z)\Big) \\ &= -\sum_{j=1}^{N}(\mathbb{E}_j - \mathbb{E}_{j-1})\beta_j(z)\mathbf{r}_j^*\mathbf{D}_j^{-2}(z)\mathbf{r}_j, \end{aligned} \tag{10.24}$$

where the last equality uses

$$\mathbf{D}^{-1}(z) - \mathbf{D}_j^{-1}(z) = -\mathbf{D}_j^{-1}(z)\mathbf{r}_j\mathbf{r}_j^*\mathbf{D}_j^{-1}(z)\beta_j(z). \tag{10.25}$$

By the identity

$$\beta_j(z) = \beta_j^{tr}(z) - \beta_j(z)\beta_j^{tr}(z)\varepsilon_j(z) = \beta_j^{tr}(z) - (\beta_j^{tr}(z))^2\varepsilon_j(z) + (\beta_j^{tr}(z))^2\beta_j(z)\varepsilon_j^2(z), \tag{10.26}$$

we have

$$\begin{aligned} (\mathbb{E}_j - \mathbb{E}_{j-1})\beta_j(z)\mathbf{r}_j^*\mathbf{D}_j^{-2}(z)\mathbf{r}_j &= \mathbb{E}_j\Big(\beta_j^{tr}(z)\delta_j(z) - (\beta_j^{tr}(z))^2\varepsilon_j(z)\frac{1}{N}tr\mathbf{D}_j^{-2}(z)\Big) \\ &\quad -(\mathbb{E}_j - \mathbb{E}_{j-1})(\beta_j^{tr}(z))^2\Big(\varepsilon_j(z)\delta_j(z) - \beta_j(z)\mathbf{r}_j^*\mathbf{D}_j^{-2}(z)\mathbf{r}_j\varepsilon_j^2(z)\Big). \end{aligned}$$

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

$$Y_j(z) = \mathbb{E}_j\Big(\beta_j^{tr}(z)\delta_j(z) - (\beta_j^{tr}(z))^2\varepsilon_j(z)\frac{1}{N}tr\mathbf{D}_j^{-2}(z)\Big) = -\mathbb{E}_j\frac{d}{dz}\beta_j^{tr}(z)\varepsilon_j(z). \tag{10.27}$$

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

$$\sum_{j=1}^{N}\mathbb{E}_{j-1}\Big(Y_j(z_1)Y_j(z_2)\Big). \tag{10.28}$$

Consider the sum

$$\sum_{j=1}^{N}\mathbb{E}_{j-1}\Big(\mathbb{E}_j\big(\beta_j^{tr}(z_1)\varepsilon_j(z_1)\big)\mathbb{E}_j\big(\beta_j^{tr}(z_2)\varepsilon_j(z_2)\big)\Big). \tag{10.29}$$

Since

$$\frac{\partial^2}{\partial z_2\partial z_1}(10.29) = (10.28), \tag{10.30}$$

by the same arguments as those on page 571 of Bai and Silverstein (2004) we only need to show (10.29) 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

$$E|\frac{1}{N}tr\mathbf{D}_j^{-1}(z) - \frac{1}{N}Etr\mathbf{D}_j^{-1}(z)|^p \le KN^{-p/2}. \tag{10.31}$$

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

$$b_n(z_1)b_n(z_2)\sum_{j=1}^{N}\mathbb{E}_{j-1}\Big(\mathbb{E}_j\big(\varepsilon_j(z_1)\big)\mathbb{E}_j\big(\varepsilon_j(z_2)\big)\Big) \tag{10.32}$$

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

$$\begin{aligned} &\mathbb{E}(\mathbf{X}_{\cdot 1}^*\mathbf{A}\mathbf{X}_{\cdot 1} - tr\mathbf{A})(\mathbf{X}_{\cdot 1}^*\mathbf{B}\mathbf{X}_{\cdot 1} - tr\mathbf{B}) \\ = \ &(\mathbb{E}|X_{11}|^4 - |EX_{11}^2|^2 - 2)\sum_{i=1}^{n} a_{ii}b_{ii} + |EX_{11}^2|^2 tr\mathbf{A}\mathbf{B}^T + tr\mathbf{A}\mathbf{B} \end{aligned} \tag{10.33}$$

valid for $n \times n$ nonrandom matrices $\mathbf{A} = (a_{ij})$ and $\mathbf{B} = (b_{ij})$, (10.32) becomes

$$b_n(z_1)b_n(z_2)\frac{1}{N^2}\sum_{j=1}^{N}\Big(tr\mathbb{E}_j\big(\mathbf{D}_j^{-1}(z_1)\big)\mathbb{E}_j\big(\mathbf{D}_j^{-1}(z_2)\big) + o(1)A_n\Big), \tag{10.34}$$

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

$$b_n(z_1)b_n(z_2)\frac{1}{N^2}\sum_{j=1}^{N}tr\mathbb{E}_j\big(\mathbf{D}_j^{-1}(z_1)\big)\mathbb{E}_j\big(\mathbf{D}_j^{-1}(z_2)\big). \tag{10.35}$$

In the real case, namely $E|X_{11}|^4 = 3 + o(1)$, (10.32) should be double the limit of (10.35).

The next aim is to investigate (10.35). To this end, set $\mathbf{D}_{ij}(z) = \mathbf{D}(z) - \mathbf{r}_i\mathbf{r}_i^* - \mathbf{r}_j\mathbf{r}_j^*$,

$$\beta_{ij}(z) = \frac{1}{1 + \mathbf{r}_i^*\mathbf{D}_{ij}^{-1}(z)\mathbf{r}_i},\ b_1(z) = \frac{1}{1 + N^{-1}\mathbb{E}tr\mathbf{D}_{12}^{-1}(z)},\ \mathbf{H}^{-1}(z) = \big(z\mathbf{I} - \frac{N-1}{N}b_1(z)\mathbf{I} - \mathbf{T}_n\big)^{-1}.$$

Write

$$\mathbf{D}_j(z_1) + z_1\mathbf{I} - \frac{N-1}{N}b_1(z_1)\mathbf{I} - \mathbf{T}_n = \sum_{i\ne j}^{N}\mathbf{r}_i\mathbf{r}_i^* - \frac{N-1}{N}b_1(z_1)\mathbf{I}.$$

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

$$\mathbf{r}_i^*\mathbf{D}_j^{-1}(z_1) = \beta_{ij}(z_1)\mathbf{r}_i^*\mathbf{D}_{ij}^{-1}(z_1), \tag{10.36}$$

we conclude that

$$\begin{aligned}\mathbf{D}_j^{-1}(z_1) &= -\mathbf{H}^{-1}(z_1) + \sum_{i\neq j}^{N}\beta_{ij}(z_1)\mathbf{H}^{-1}(z_1)\mathbf{r}_i\mathbf{r}_i^*\mathbf{D}_{ij}^{-1}(z_1) - \frac{N-1}{N}b_1(z_1)\mathbf{H}^{-1}(z_1)\mathbf{D}_j^{-1}(z_1)\\ &= -\mathbf{H}^{-1}(z_1) + b_1(z_1)A(z_1) + B(z_1) + C(z_1),\end{aligned} \tag{10.37}$$

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,

$$\begin{aligned}&\Big|1 - \frac{1}{z\big(1+N^{-1}\mathbb{E}tr\mathbf{D}_{12}^{-1}(z)\big)} - \frac{t}{z}\Big|^{-1} = \Big|\frac{z\big(1+N^{-1}\mathbb{E}tr\mathbf{D}_{12}^{-1}(z)\big)}{(z-t)\big(1+N^{-1}\mathbb{E}tr\mathbf{D}_{12}^{-1}(z)\big)-1}\Big|\\ &\leq \frac{\big|z\big(1+N^{-1}\mathbb{E}tr\mathbf{D}_{12}^{-1}(z)\big)\big|}{\Im\Big[(z-t)\big(1+N^{-1}\mathbb{E}tr\mathbf{D}_{12}^{-1}(z)\big)\Big]} \leq \frac{|z|(1+n/(Nv_0))}{v_0},\end{aligned}$$

where the last inequality uses

$$\begin{aligned}&\Im\Big[(z-t)\big(1+N^{-1}\mathbb{E}tr\mathbf{D}_{12}^{-1}(z)\big)\Big] = v_0 + \Im\Big[(z-t)N^{-1}\mathbb{E}tr\mathbf{D}_{12}^{-1}(z)\Big]\\ =\ & v_0 + \Im\Big[(z-t)N^{-1}\sum_{i=1}^{n}\mathbb{E}\frac{1}{\lambda_i - z}\Big] = v_0 + \Im\Big[N^{-1}\sum_{i=1}^{n}\mathbb{E}\frac{(z-t)(\lambda_i - t - (\bar{z}-t))}{|\lambda_i - z|^2}\Big]\\ =\ & v_0 + N^{-1}\sum_{i=1}^{n}\mathbb{E}\frac{(\lambda_i - t)v_0}{|\lambda_i - z|^2} \geq v_0,\end{aligned}$$

with the fact that

$$\lambda_i \geq t, \forall i = 1, 2, \ldots, n,$$

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

$$\Big|\Big|\mathbf{H}^{-1}(z)\Big|\Big| = \Big|\Big|\Big(z\mathbf{I} - \frac{N-1}{N}b_1(z)\mathbf{I} - \mathbf{T}_n\Big)^{-1}\Big|\Big| \leq \frac{1+n/(Nv_0)}{v_0}. \tag{10.38}$$

Moreover from (10.31) and (10.23) we have

$$E|\gamma_j(z)|^p \leq KN^{-1}\delta_n^{2p-4}, \quad p \geq 2. \tag{10.39}$$

Therefore the discussions for (2.11)-(2.13) of Bai and Silverstein (2004) still work in our case. That is,

$$\mathbb{E}|tr\mathbf{B}(z_1)\mathbf{M}| \le K_{\mathbf{M}}KN^{1/2}, \quad \mathbb{E}|tr\mathbf{C}(z_1)\mathbf{M}| \le K_{\mathbf{M}}K, \tag{10.40}$$

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,

$$\mathbb{E}|tr\mathbf{A}(z_1)\mathbf{M}| \le K\|\mathbf{M}\|N^{1/2}, \tag{10.41}$$

where $\|\mathbf{M}\|$ denotes the spectral norm of a matrix.

Using an identity similar to (10.25) yields

$$tr\mathbb{E}_j\big(\mathbf{A}(z_1)\big)\mathbf{D}_j^{-1}(z_2) = A_1(z_1,z_2) + A_2(z_1,z_2) + A_3(z_1,z_2), \tag{10.42}$$

where

$$\begin{aligned} A_1(z_1,z_2) &= -tr\sum_{i<j}\mathbf{H}^{-1}(z_1)\mathbf{r}_i\mathbf{r}_i^*\mathbb{E}_j\big(\mathbf{D}_{ij}^{-1}(z_1)\big)\beta_{ij}(z_2)\mathbf{D}_{ij}^{-1}(z_2)\mathbf{r}_i\mathbf{r}_i^*\mathbf{D}_{ij}^{-1}(z_2) \\ =\ & -\sum_{i<j}\beta_{ij}(z_2)\mathbf{r}_i^*\mathbb{E}_j\big(\mathbf{D}_{ij}^{-1}(z_1)\big)\mathbf{D}_{ij}^{-1}(z_2)\mathbf{r}_i\mathbf{r}_i^*\mathbf{D}_{ij}^{-1}(z_2)\mathbf{H}^{-1}(z_1)\mathbf{r}_i, \\ A_2(z_1,z_2) &= -tr\sum_{i<j}\mathbf{H}^{-1}(z_1)N^{-1}\mathbb{E}_j\big(\mathbf{D}_{ij}^{-1}(z_1)\big)\big(\mathbf{D}_j^{-1}(z_2) - \mathbf{D}_{ij}^{-1}(z_2)\big), \\ A_3(z_1,z_2) &= tr\sum_{i<j}\mathbf{H}^{-1}(z_1)(\mathbf{r}_i\mathbf{r}_i^* - N^{-1}\mathbf{I})\mathbb{E}_j\big(\mathbf{D}_{ij}^{-1}(z_1)\big)\mathbf{D}_{ij}^{-1}(z_2). \end{aligned}$$

By arguments similar to (2.15) in Bai and Silverstein (2004), (10.38) and (10.23) 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

$$\mathbb{E}\Big|A_1(z_1,z_2) + \frac{j-1}{N^2}b_1(z_2)tr\Big(\mathbb{E}_j\big(\mathbf{D}_j^{-1}(z_1)\big)\mathbf{D}_j^{-1}(z_2)\Big)tr\mathbf{D}_j^{-1}(z_2)\mathbf{H}^{-1}(z_1)\Big| \le KN^{1/2}. \tag{10.43}$$

We conclude from (10.37)-(10.43) that

$$\begin{aligned} & tr\Big(\mathbb{E}_j\big(\mathbf{D}_j^{-1}(z_1)\big)\mathbf{D}_j^{-1}(z_2)\Big)\Big(1 + \frac{j-1}{N^2}b_1(z_1)b_1(z_2)tr\Big(\mathbf{D}_j^{-1}(z_2)\mathbf{H}^{-1}(z_1)\Big) \\ =\ & -tr\Big(\mathbf{H}^{-1}(z_1)\mathbf{D}_j^{-1}(z_2)\Big) + A_4(z_1,z_2), \end{aligned} \tag{10.44}$$

where

$$\mathbb{E}|A_4(z_1,z_2)| \le KN^{1/2}.$$

Applying the expression for $\mathbf{D}_j^{-1}(z_2)$ in (10.37), (10.40) and (10.41), we obtain

$$\begin{aligned} & tr\Big(\mathbb{E}_j\big(\mathbf{D}_j^{-1}(z_1)\big)\mathbf{D}_j^{-1}(z_2)\Big) \times \Big(1 - \frac{j-1}{N^2}b_1(z_1)b_1(z_2)tr\mathbf{H}^{-1}(z_1)\mathbf{H}^{-1}(z_2)\Big) \\ =\ & tr\Big(\mathbf{H}^{-1}(z_1)\mathbf{H}^{-1}(z_2)\Big) + A_5(z_1,z_2), \end{aligned} \tag{10.45}$$

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

$$\begin{aligned}
&\frac{1}{N} tr\Big(\mathbb{E}_j\big(\mathbf{D}_j^{-1}(z_1)\big)\mathbf{D}_j^{-1}(z_2)\Big) \\
&\times\Big(1 - \frac{j-1}{N} c_n \frac{1}{\big(1+cm_n^0(z_1)\big)\big(1+cm_n^0(z_2)\big)} \int \frac{dH_n(t)}{\big(z_2 - \frac{1}{1+cm_n^0(z_2)} - t\big)\big(z_1 - \frac{1}{1+cm_n^0(z_1)} - t\big)}\Big) \\
= \quad & c_n \int \frac{dH_n(t)}{\big(z_2 - \frac{1}{1+cm_n^0(z_2)} - t\big)\big(z_1 - \frac{1}{1+cm_n^0(z_1)} - t\big)} + A_6(z_1,z_2), \qquad (10.46)
\end{aligned}$$

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

$$|a_n(z_1,z_2)| < 1. \qquad (10.47)$$

Indeed, by the Cauchy-Schwartz inequality, we have

$$\begin{aligned}
&\Big|\frac{c_n}{\big(1+c_n m_n^0(z_1)\big)\big(1+c_n m_n^0(z_2)\big)} \int \frac{dH_n(t)}{\big(z_2 - \frac{1}{1+c_n m_n^0(z_2)} - t\big)\big(z_1 - \frac{1}{1+c_n m_n^0(z_1)} - t\big)}\Big| \\
\le \quad & \Big(\int \frac{c_n dH_n(t)}{|1+c_n m_n^0(z_1)|^2 |z_1 - \frac{1}{1+c_n m_n^0(z_1)} - t|^2}\Big)^{1/2} \Big(\int \frac{c_n dH_n(t)}{|1+c_n m_n^0(z_2)|^2 |z_2 - \frac{1}{1+c_n m_n^0(z_2)} - t|^2}\Big)^{1/2}. \qquad (10.48)
\end{aligned}$$

Note that $m_n^0(z)$ satisfies an equality similar to (10.5)

$$m_n^0(z) = \int \frac{dH_n(t)}{t - z + \frac{1}{1+c_n m_n^0(z)}}. \qquad (10.49)$$

Taking the imaginary part of the both sides of (10.49) leads to

$$\begin{aligned}
\Im\big(m_n^0(z)\big) &= \int \frac{\Im\Big(\overline{t - z + \frac{1}{1+c_n m_n^0(z)}}\Big) dH_n(t)}{|t - z + \frac{1}{1+c_n m_n^0(z)}|^2} \\
&= v_0 \int \frac{dH_n(t)}{|t - z + \frac{1}{1+c_n m_n^0(z)}|^2} + \frac{c_n \Im(m_n^0(z))}{|1+c_n m_n^0(z)|^2} \int \frac{dH_n(t)}{|t - z + \frac{1}{1+c_n m_n^0(z)}|^2}.
\end{aligned}$$

Dividing by $\Im\big(m_n^0(z)\big)$ on both sides, we have

$$\frac{c_n}{|1+c_n m_n^0(z)|^2} \int \frac{dH_n(t)}{|t - z + \frac{1}{1+c_n m_n^0(z)}|^2} = 1 - \frac{v_0}{\Im\big(m_n^0(z)\big)} \int \frac{dH_n(t)}{|t - z + \frac{1}{1+c_n m_n^0(z)}|^2} < 1.$$

This, together with (10.48), yields (10.47).

It follows from (10.46) and (10.47) that (10.35) 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

$$(10.35)\stackrel{i.p.}{\to} a(z_1,z_2)\int_0^1\frac{1}{1-ta(z_1,z_2)}dt=\int_0^{a(z_1,z_2)}\frac{1}{1-z}dz,$$

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 (10.5) and $s(z_1,z_2)=\frac{1}{1+cm^0(z_1)}-\frac{1}{1+cm^0(z_2)}$. Therefore the limit of (10.28) under the complex case is

$$\frac{\partial^2}{\partial z_2\partial z_1}\int_0^{a(z_1,z_2)}\frac{1}{1-z}dz=\frac{\partial}{\partial z_2}\Big(\frac{\partial a(z_1,z_2)/\partial z_1}{1-a(z_1,z_2)}\Big).$$

$$=\frac{\partial}{\partial z_2}\Big[\frac{s(z_1,z_2)+(z_1-z_2)\frac{c(m^0(z_1))'}{(1+cm^0(z_1))^2}}{(z_2-z_1+s(z_1,z_2))(z_2-z_1)}\Big]$$

$$=\frac{\partial}{\partial z_2}\Big[\frac{1}{z_2-z_1}-(1+\frac{c(m^0(z_1))'}{(1+cm^0(z_1))^2})\frac{1}{z_2-z_1+s(z_1,z_2)}\Big]$$

$$=-\frac{1}{(z_2-z_1)^2}+[1+\frac{c(m^0(z_1))'}{(1+cm^0(z_1))^2}+\frac{c(m^0(z_2))'}{(1+cm^0(z_2))^2}+\frac{c(m^0(z_1))'}{(1+cm^0(z_1))^2}\frac{c(m^0(z_2))'}{(1+cm^0(z_2))^2}]\frac{1}{(z_2-z_1+s(z_1,z_2))^2}.$$

10.4.2 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).

10.4.3 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

$$\mathbb{E}|\gamma_j|^p\le KN^{-1}\delta_n^{2p-4},\quad p\ge 2 \tag{10.50}$$

and

$$\mathbb{E}|\beta_1(z)|^p \leq K, \quad p \geq 1, \quad |b_n(z)| \leq K. \tag{10.51}$$

Similar to (3.1) and (3.2) in Bai and Silverstein (2004), by (10.16), we have for any positive p

$$\max\Big(\mathbb{E}||\mathbf{D}^{-1}(z)||^p, \mathbb{E}||\mathbf{D}_j^{-1}(z)||^p, \mathbb{E}||\mathbf{D}_{ij}^{-1}(z)||^p\Big) \leq K \tag{10.52}$$

and via (10.23) and (10.12)

$$\Big|\mathbb{E}\Big(a(v)\prod_{m=1}^{q}\big(\mathbf{r}_1^*\mathbf{B}_m(v)\mathbf{r}_1 - N^{-1}tr\mathbf{B}_m(v)\big)\Big)\Big| \leq KN^{-(1\wedge q)}\delta_n^{(2q-4)\vee 0}, \quad q \geq 0, \tag{10.53}$$

where the matrices $\mathbf{B}_m(v)$ are independent of $\mathbf{r}_1$ and

$$\max(|a(v)|, ||\mathbf{B}_m(v)||) \leq K\big(1 + n^s I(||\mathbf{B}_n|| \geq \eta_r \ or \ \lambda_{\min}^{\widetilde{\mathbf{B}}} \leq \eta_\ell)\big)$$

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 $n \to \infty$,

$$\sup_{z\in\mathcal{C}_n^+} |\mathbb{E}m_n(z) - m^0(z)| \to 0. \tag{10.54}$$

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

$$\mathbb{E}|tr\mathbf{D}_1^{-1}(z)\mathbf{M} - \mathbb{E}tr\mathbf{D}_1^{-1}(z)\mathbf{M}|^2 \leq K||\mathbf{M}||^2. \tag{10.55}$$

We next show

$$\sup_{z\in\mathcal{C}_n^+} ||\Big((\mathbb{E}\beta_1)\mathbf{I} - z\mathbf{I} + \mathbf{T}_n\Big)^{-1}|| < \infty. \tag{10.56}$$

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 (10.38) and (10.39).

Consider $x = x_\ell$ or x_r now. So $x \in S_{F_{c,H}}^c$, where $S_{F_{c,H}}^c$ 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 $x \in I \subset S_{F_{c,H}}^c$ 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 $D = \{z \in \mathbb{C} : \Im z > 0\}$. Let $m = m(z) = z - \frac{1}{1+cm^0(z)} \in D$ (for $z \in D$). From (10.5) we have

$$z(m) = m + \frac{1}{1 + cm_H(m)}. \tag{10.57}$$

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 (10.57) and (10.5). Therefore we have $z(m) \to x_0$ as $m \in D \to (x_0 - \frac{1}{1+cm_0})$.

(10.57) 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 $H(x)$

$$\inf_{z\in\mathcal{C}^0} |t_0 - z + \frac{1}{1+cm^0(z)}| > \eta \ \ and \ \ \sup_{z\in\mathcal{C}^0} |m^0(z)| < \kappa. \tag{10.58}$$

Also from (10.50), (10.51) and (10.54) we have

$$\sup_{z\in\mathcal{C}_n^+} |E\beta_1 - \frac{1}{1+cm^0(z)}| \to 0. \tag{10.59}$$

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

$$|\mu - t_0| < \eta/4. \tag{10.60}$$

We conclude from (10.60), (10.59) and (10.58) that

$$\inf_{z\in\mathcal{C}_\ell\cup\mathcal{C}_r} |\mu - z + E\beta_1| > \eta/2, \tag{10.61}$$

which ensures (10.56).

With $\mathbf{H}_1 = \mathbb{E}\beta_1(z)\mathbf{I} - z\mathbf{I} + \mathbf{T}_n$, write

$$\mathbf{D}(z) - \mathbf{H}_1 = \sum_{j=1}^N \mathbf{r}_j\mathbf{r}_j^* - (\mathbb{E}\beta_1(z))\mathbf{I}. \tag{10.62}$$

Postmultiplying $\mathbf{D}^{-1}(z)$ and premuliplying $\mathbf{H}_1^{-1}$ on the both sides, taking expectation and using an equality similar to (10.36) we get

$$\begin{aligned} \mathbf{H}_1^{-1} - \mathbb{E}\mathbf{D}^{-1}(z) &= \mathbf{H}_1^{-1}\mathbb{E}\Big[\Big(\sum_{j=1}^N \mathbf{r}_j\mathbf{r}_j^* - (\mathbb{E}\beta_1(z))\mathbf{I}\Big)\mathbf{D}^{-1}(z)\Big] \\ &= \mathbf{H}_1^{-1}\sum_{j=1}^N \mathbb{E}\Big(\mathbf{r}_j\beta_j(z)\mathbf{r}_j^*\mathbf{D}_j^{-1}(z)\Big) - \mathbf{H}_1^{-1}(\mathbb{E}\beta_1(z))\mathbb{E}\mathbf{D}^{-1}(z) \\ &= N\mathbb{E}\Big[\beta_1(z)\Big(\mathbf{H}_1^{-1}\mathbf{r}_1\mathbf{r}_1^*\mathbf{D}_1^{-1}(z) - \frac{1}{N}\mathbf{H}_1^{-1}\mathbb{E}\mathbf{D}^{-1}(z)\Big)\Big]. \end{aligned} \tag{10.63}$$

Taking trace on both sides, we have

$$\begin{aligned} & n\Big(\int \frac{dH_n(x)}{x - (z - \mathbb{E}\beta_1(z))} - \mathbb{E}m_n(z)\Big) \\ &= N\mathbb{E}\Big[\beta_1(z)\Big(\mathbf{r}_1^*\mathbf{D}_1^{-1}(z)\mathbf{H}_1^{-1}\mathbf{r}_1 - \frac{1}{N}tr\mathbf{H}_1^{-1}\mathbb{E}\mathbf{D}^{-1}(z)\Big)\Big]. \end{aligned} \tag{10.64}$$

When there is no confusion, we below drop z from $\beta_1(z), \gamma_1(z), b_n(z)$, etc. By (10.25), we have

$$\begin{aligned}
&\mathbb{E}tr\mathbf{H}_1^{-1}\mathbf{D}_1^{-1}(z) - \mathbb{E}tr\mathbf{H}_1^{-1}\mathbf{D}^{-1}(z) = \mathbb{E}\Big[\beta_1(z)tr\mathbf{H}_1^{-1}\mathbf{D}_1^{-1}(z)\mathbf{r}_1\mathbf{r}_1^*\mathbf{D}_1^{-1}(z)\Big] \\
= \;& b_n(z)\mathbb{E}\Big[(1-\beta_1\gamma_1)\mathbf{r}_1^*\mathbf{D}_1^{-1}(z)\mathbf{H}_1^{-1}\mathbf{D}_1^{-1}(z)\mathbf{r}_1\Big],
\end{aligned} \tag{10.65}$$

where the last equality uses $\beta_1 = b_n - \beta_1 b_n\gamma_1$. In view of (10.53), (10.50) and (10.56), we obtain

$$\Big|\mathbb{E}\beta_1(z)\gamma_1\mathbf{r}_1^*\mathbf{D}_1^{-1}(z)\mathbf{H}_1^{-1}\mathbf{D}_1^{-1}(z)\mathbf{r}_1\Big| \le KN^{-1}, \tag{10.66}$$

which implies that

$$\Big|(10.65) - N^{-1}b_n\mathbb{E}\mathbf{D}_1^{-1}(z)\mathbf{H}_1^{-1}\mathbf{D}_1^{-1}(z)\Big| \le KN^{-1}.$$

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 (10.50), (10.56) and (10.53) 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 (10.51), (10.50), (10.56) and (10.55) 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 (10.50), (10.51) 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

$$\begin{aligned}
& \mathbb{E}N\gamma_1\mathbf{r}_1^*\mathbf{D}_1^{-1}\mathbf{H}_1^{-1}\mathbf{r}_1 \\
= \;& N\mathbb{E}\Big[\Big(\mathbf{r}_1^*\mathbf{D}_1^{-1}\mathbf{r}_1 - N^{-1}tr\mathbf{D}_1^{-1}\Big)\Big(\mathbf{r}_1^*\mathbf{D}_1^{-1}\mathbf{H}_1^{-1}\mathbf{r}_1 - N^{-1}tr\mathbf{D}_1^{-1}\mathbf{H}_1^{-1}\Big)\Big] \\
& +N^{-1}Cov\Big(tr\mathbf{D}_1^{-1}, tr\mathbf{D}_1^{-1}(z)\mathbf{H}_1^{-1}\Big).
\end{aligned} \tag{10.67}$$

In view of (10.55), we see the second term above is $O(N^{-1})$. We conclude from (10.64)-(10.67) that

$$n\Big(\int \frac{dH_n(x)}{x-(z-\mathbb{E}\beta_1(z))} - \mathbb{E}m_n(z)\Big) = b_n^2(z)N^{-1}\mathbb{E}tr\mathbf{D}_1^{-1}\mathbf{H}_1^{-1}\mathbf{D}_1^{-1} \quad (10.68)$$

$$-b_n^2 N\mathbb{E}\Big[\Big(\mathbf{r}_1^*\mathbf{D}_1^{-1}\mathbf{r}_1 - N^{-1}tr\mathbf{D}_1^{-1}\Big)\Big(\mathbf{r}_1^*\mathbf{D}_1^{-1}\mathbf{H}_1^{-1}\mathbf{r}_1 - N^{-1}tr\mathbf{D}_1^{-1}\mathbf{H}_1^{-1}\Big)\Big] \quad (10.69)$$

$$+o(1). \quad (10.70)$$

Using (10.33) on (10.69) 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

$$n\Big(\int \frac{dH_n(x)}{x-(z-\mathbb{E}\beta_1(z))} - \mathbb{E}m_n(z)\Big) = -b_n^2 N^{-1}\mathbb{E}tr\mathbf{D}_1^{-1}\mathbf{H}_1^{-1}\mathbf{D}_1^{-1} + o(1). \quad (10.71)$$

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 (10.25),(10.51),(10.53) and (10.56) 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

$$N^{-1}\mathbb{E}tr\mathbf{D}^{-1}\mathbf{H}_1^{-1}\mathbf{D}^{-1}. \quad (10.72)$$

From (10.62), similar to (10.63) we have

$$\begin{aligned}\mathbf{D}^{-1}(z) &= \mathbf{H}_1^{-1} - \sum_{j=1}^{N}\beta_j\mathbf{H}_1^{-1}\mathbf{r}_j\mathbf{r}_j^*\mathbf{D}_j^{-1}(z) + \big(\mathbb{E}\beta_1\big)\mathbf{H}_1^{-1}\mathbf{D}^{-1}(z)\\ &= \mathbf{H}_1^{-1} + \big(\mathbb{E}\beta_1\big)A(z) + B(z) + C(z), \quad (10.73)\end{aligned}$$

where

$$A(z) = -\sum_{j=1}^{N}\mathbf{H}_1^{-1}\Big(\mathbf{r}_j\mathbf{r}_j^* - N^{-1}\mathbf{I}\Big)\mathbf{D}_j^{-1}(z),\ B(z) = -\sum_{j=1}^{N}\Big(\beta_j - \mathbb{E}\beta_1\Big)\mathbf{H}_1^{-1}\mathbf{r}_j\mathbf{r}_j^*\mathbf{D}_j^{-1}(z),$$

$$C(z) = -N^{-1}\big(\mathbb{E}\beta_1\big)\mathbf{H}_1^{-1}\sum_{j=1}^{N}\Big(\mathbf{D}_j^{-1}(z) - \mathbf{D}^{-1}(z)\Big) = -N^{-1}\big(\mathbb{E}\beta_1\big)\mathbf{H}_1^{-1}\sum_{j=1}^{N}\beta_j\mathbf{D}_j^{-1}(z)\mathbf{r}_j\mathbf{r}_j^*\mathbf{D}_j^{-1}(z).$$

It follows from (10.50) and (10.51) that

$$\mathbb{E}|\beta_1 - \mathbb{E}\beta_1|^2 \le KN^{-1}. \quad (10.74)$$

For any $n \times n$ matrix $\mathbf{M}$, by (10.52), (10.53), (10.51) and (10.74) we obtain

$$\begin{aligned}|N^{-1}\mathbb{E}tr\mathbf{B}(z)\mathbf{M}| &\leq K\big(\mathbb{E}|\beta_1(z)-\mathbb{E}\beta_1(z)|^2\big)^{1/2}\big(\mathbb{E}\big|\mathbf{r}_1^*\mathbf{r}_1||\mathbf{D}_1^{-1}\mathbf{M}||\big|^2\big)^{1/2}\\ &\leq KN^{-1/2}(\mathbb{E}||\mathbf{M}||^4)^{1/4} \qquad (10.75)\end{aligned}$$

and

$$\begin{aligned}|N^{-1}\mathbb{E}tr\mathbf{C}(z)\mathbf{M}| &\leq KN^{-1}\mathbb{E}|\beta_1(z)|\mathbf{r}_1^*\mathbf{r}_1||\mathbf{D}_1^{-1}(z)||^2||\mathbf{M}||\\ &\leq KN^{-1}(\mathbb{E}||\mathbf{M}||^2)^{1/2}. \qquad (10.76)\end{aligned}$$

For any $n \times n$ nonrandom matrix $\mathbf{M}$ with a bounded spectral norm, we write

$$tr\mathbf{A}(z)\mathbf{D}^{-1}(z)\mathbf{M} = A_1(z) + A_2(z) + A_3(z), \qquad (10.77)$$

where

$$\begin{aligned}A_1(z) &= -tr\sum_{j=1}^{N}\mathbf{H}_1^{-1}\mathbf{r}_j\mathbf{r}_j^*\mathbf{D}_j^{-1}(z)\Big(\mathbf{D}^{-1}(z)-\mathbf{D}_j^{-1}(z)\Big)\mathbf{M},\\ A_2(z) &= -tr\sum_{j=1}^{N}\mathbf{H}_1^{-1}\Big(\mathbf{r}_j\mathbf{r}_j^*\mathbf{D}_j^{-2}(z)-N^{-1}\mathbf{D}_j^{-2}(z)\Big)\mathbf{M},\\ A_3(z) &= -tr\sum_{j=1}^{N}\mathbf{H}_1^{-1}N^{-1}\mathbf{D}_j^{-1}(z)\Big(\mathbf{D}_j^{-1}(z)-\mathbf{D}^{-1}(z)\Big)\mathbf{M}.\end{aligned}$$

Obviously $\mathbb{E}A_2(z) = 0$ and similar to (10.76), we obtain

$$|\mathbb{E}N^{-1}A_3(z)| \leq KN^{-1}. \qquad (10.78)$$

From (10.50) and (10.51)

$$\mathbb{E}|\beta_1 - b_n|^2 \leq KN^{-1}. \qquad (10.79)$$

Using (10.53), (10.79) and (10.25) 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 (10.55) and (10.52), we have

$$\begin{aligned}&\Big|Cov\Big(N^{-1}tr\mathbf{D}^{-2}(z), N^{-1}tr\mathbf{D}^{-1}(z)\mathbf{M}\mathbf{H}_1^{-1}\Big)\Big|\\ \leq\ &\Big(\mathbb{E}|N^{-1}tr\mathbf{D}^{-2}(z)|^2\Big)^{1/2}N^{-1}\Big(\mathbb{E}\Big|tr\mathbf{D}^{-1}(z)\mathbf{M}\mathbf{H}_1^{-1} - \mathbb{E}tr\mathbf{D}^{-1}\mathbf{M}\mathbf{H}_1^{-1}\Big|^2\Big)^{1/2} \leq KN^{-1}.\end{aligned}$$

We thus have

$$\mathbb{E}N^{-1}A_1(z) = b_n\Big(\mathbb{E}N^{-1}tr\mathbf{D}^{-2}(z)\Big)\Big(\mathbb{E}N^{-1}tr\mathbf{D}^{-1}(z)\mathbf{M}\mathbf{H}_1^{-1}\Big) + o(1). \qquad (10.80)$$

Moreover, by (10.73), (10.75) and (10.76), we have

$$
\begin{aligned}
&\mathbb{E}N^{-1}tr\mathbf{D}^{-1}(z)\mathbf{H}_1^{-2} = N^{-1}tr\Big(\mathbf{H}_1^{-1}+\mathbb{E}B(z)+\mathbb{E}C(z)\Big)\mathbf{H}_1^{-2}\\
= \;& c_n\int\frac{dH_n(x)}{\big(x-z+\mathbb{E}\beta_1\big)^3}+o(1). \qquad (10.81)
\end{aligned}
$$

From (10.73)-(10.81) we conclude that

$$
\begin{aligned}
& N^{-1}\mathbb{E}tr\mathbf{D}^{-1}(z)\mathbf{H}_1^{-1}\mathbf{D}^{-1}(z)\\
= \;& \mathbb{E}N^{-1}tr\mathbf{D}^{-1}(z)\mathbf{H}_1^{-2}+b_n^2\Big(\mathbb{E}N^{-1}tr\mathbf{D}^{-2}(z)\Big)\Big(\mathbb{E}N^{-1}tr\mathbf{D}^{-1}(z)\mathbf{H}_1^{-2}\Big)+o(1)\\
= \;& c_n\int\frac{dH_n(x)}{\big(x-z+\mathbb{E}\beta_1\big)^3}+b_n^2c_n^2\mathbb{E}\int\frac{dF_n(x)}{(x-z)^2}\int\frac{dH_n(x)}{\big(x-z+\mathbb{E}\beta_1\big)^3}+o(1). \qquad (10.82)
\end{aligned}
$$

This, together with (10.71), (10.56), (10.54) and (10.59), leads to

$$
\begin{aligned}
& n\Big(\int\frac{dH_n(x)}{x-(z-\mathbb{E}\beta_1)}-\mathbb{E}m_n(z)\Big) \qquad (10.83)\\
= \;& -c_nb_n^2\int\frac{dH_n(x)}{\big(x-z+\mathbb{E}\beta_1\big)^3}-b_n^4c_n^2\mathbb{E}\int\frac{dF_n(x)}{(x-z)^2}\int\frac{dH_n(x)}{\big(x-z+\mathbb{E}\beta_1\big)^3}+o(1)\\
= \;& -c\varpi^2(z)\int\frac{dH(x)}{\big(x-z+\varpi(z)\big)^3}-\varpi^4(z)c^2\int\frac{dF_{c,H}(x)}{(x-z)^2}\int\frac{dH(x)}{\big(x-z+\varpi(z)\big)^3}+o(1),
\end{aligned}
$$

where the last step uses

$$
\sup_{z\in\mathcal{C}_n}\Big|\mathbb{E}\int\frac{dF_n(x)}{(x-z)^2}-\int\frac{F_{c,H}(x)}{(x-z)^2}\Big|\to 0,\quad as\quad n\to\infty, \qquad (10.84)
$$

which can be proved similarly to (4.1) in Bai and Silverstein (2004).

Let $\varpi_n(z)=1/(1+c_nm_n^0(z))$. By (10.49) we then write

$$
\begin{aligned}
& n\Big(\mathbb{E}m_n(z)-m_n^0(z)\Big)=n\Big(\mathbb{E}m_n(z)-\int\frac{dH_n(x)}{x-(z-\varpi_n(z))}\Big) \qquad (10.85)\\
= \;& n\Big(\mathbb{E}m_n(z)-\int\frac{dH_n(x)}{x-(z-\mathbb{E}\beta_1)}\Big)+n\Big(\int\frac{dH_n(x)}{x-(z-\mathbb{E}\beta_1)}-\int\frac{dH_n(x)}{x-(z-\varpi_n(z))}\Big)\\
= \;& n\Big(\mathbb{E}m_n(z)-\int\frac{dH_n(x)}{x-(z-\mathbb{E}\beta_1)}\Big)+n\Big(\varpi_n(z)-\mathbb{E}\beta_1\Big)\int\frac{dH_n(x)}{\big(x-(z-\varpi_n(z))\big)\big(x-(z-\mathbb{E}\beta_1)\big)}.
\end{aligned}
$$

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

$$
n\Big(\varpi_n(z)-\mathbb{E}\beta_1\Big)=n\Big(\varpi_n(z)-\mathbb{E}\beta^{tr}(z)\Big)+n\Big(\mathbb{E}\beta^{tr}(z)-\mathbb{E}\beta_1^{tr}(z)\Big)+n(E\beta_1^{tr}(z)-E\beta_1(z)). \qquad (10.86)
$$

First, by the fact that

$$
\beta^{tr}(z)=b(z)+\beta^{tr}(z)b(z)\big(c_nEm_n(z)-c_nm_n(z)\big) \qquad (10.87)
$$

we have

$$\begin{aligned} & n\Big(\varpi_n(z) - \mathbb{E}\beta^{tr}(z)\Big) = n\mathbb{E}\Big[\varpi_n(z)\beta^{tr}(z)(c_n m_n(z) - c_n m_n^0(z))\Big] \\ = \;& n\mathbb{E}\Big[\varpi_n(z)b(z)(c_n m_n(z) - c_n m_n^0(z))\Big] \\ & + n\mathbb{E}\Big[\varpi_n(z)b(z)\beta^{tr}(z)\Big(c_n m_n(z) - c_n m_n^0(z)\Big)\Big(\mathbb{E}c_n m_n(z) - c_n m_n(z)\Big)\Big] \\ = \;& n\varpi_n(z)b(z)\mathbb{E}\big(c_n m_n(z) - c_n m_n^0(z)\big) + o(1), \end{aligned} \tag{10.88}$$

where via (10.54), (10.55), (10.51) and (10.87)

$$\begin{aligned} & nc_n^2 b(z)\varpi_n(z)\mathbb{E}\Big[\beta^{tr}(z)\Big(m_n(z) - m_n^0(z)\Big)\Big(\mathbb{E}m_n(z) - m_n(z)\Big)\Big] \qquad (10.89) \\ = \;& nc_n^2 b(z)\varpi_n(z)\Big[\mathbb{E}\Big(\beta^{tr}(z)\big(Em_n(z) - m_n^0(z)\big)\big(\mathbb{E}m_n(z) - m_n(z)\big)\Big) \\ & - \mathbb{E}\Big(\beta^{tr}(z)\big(\mathbb{E}m_n(z) - m_n(z)\big)^2\Big)\Big] \\ = \;& o(1) + nc_n^2 b^2(z)\varpi_n(z)\Big[\mathbb{E}\Big(\mathbb{E}m_n(z) - m_n(z)\Big)^2 - \mathbb{E}\Big(\beta^{tr}(z)\big(\mathbb{E}m_n(z) - m_n(z)\big)^3\Big)\Big] = o(1), \end{aligned}$$

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 (10.86), by (10.25), we obtain

$$\begin{aligned} & n\Big(\mathbb{E}\beta^{tr}(z) - \mathbb{E}\beta_1^{tr}(z)\Big) = \frac{n}{N}\mathbb{E}\Big[\beta^{tr}(z)\beta_1^{tr}(z)tr\Big(\mathbf{D}_1^{-1}(z) - \mathbf{D}^{-1}(z)\Big)\Big] \\ = \;& \frac{n}{N}\mathbb{E}\Big[\beta^{tr}(z)\beta_1^{tr}(z)\mathbf{r}_1^*\mathbf{D}_1^{-2}(z)\mathbf{r}_1\beta_1(z)\Big] = c^2\varpi^3(z)\int\frac{dF_{c,H}(x)}{(x-z)^2} + o(1), \end{aligned} \tag{10.90}$$

where the last step uses (10.54), (10.55), (10.51), (10.53), (10.84) and (10.79).

As for the third term on the right side of (10.86) we conclude from (10.26) and (10.53) that

$$\begin{aligned} & n\Big(\mathbb{E}\beta_1^{tr}(z) - \mathbb{E}\beta_1(z)\Big) = -n\mathbb{E}\Big((\beta_j^{tr}(z))^2\beta_j(z)\varepsilon_j^2(z)\Big) \qquad (10.91) \\ = \;& -n\mathbb{E}\Big(\varepsilon_1^2(z)\big(\beta_1^{tr}(z)\big)^3\Big) + n\mathbb{E}\Big(\varepsilon_1^3(z)\beta_j(z)\big(\beta_1^{tr}(z)\big)^3\Big) = -n\mathbb{E}\Big(\varepsilon_1^2(z)\big(\beta_1^{tr}(z)\big)^3\Big) + o(1). \end{aligned}$$

Moreover by (10.53), (10.55), (10.33) and (10.54) 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

$$n\Big(\mathbb{E}\beta_1^{tr}(z) - \mathbb{E}\beta_1(z)\Big) \to -2c^2\varpi^3(z)\int\frac{dF_{c,H}(x)}{(x-z)^2}, \quad as\ N\to\infty, \tag{10.92}$$

while the limit is half of the above in the complex case.

Summarizing the above we conclude that

$$\begin{aligned} & n\Big(\varpi_n(z) - \mathbb{E}\beta_1(z)\Big) \qquad (10.93) \\ = \;& \begin{cases} c_n\varpi_n(z)b(z)n\mathbb{E}\big(m_n(z) - m_n^0(z)\big) - c^2\varpi^3(z)\int\frac{dF_{c,H}(x)}{(x-z)^2} + o(1) & \text{in the real case} \\ c_n\varpi_n(z)b(z)n\mathbb{E}\big(m_n(z) - m_n^0(z)\big) + o(1) & \text{in the complex case.} \end{cases} \end{aligned}$$

The proof for (10.47) 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 (10.85), (10.93), (10.56) and (10.54), yields

$$\begin{aligned}
&n\Big(\mathbb{E}m_n(z)-m_n^0(z)\Big)\\
&=\begin{cases}\frac{n\left(\mathbb{E}m_n(z)-\int\frac{dH_n(x)}{x-(z-\mathbb{E}\beta_1)}\right)-c^2\varpi^3(z)\int\frac{dF_{c,H}(x)}{(x-z)^2}\int\frac{dH_n(x)}{\left(x-(z-\varpi_n(z))\right)^2}}{1-c_n\varpi_n^2(z)\int\frac{dH_n(x)}{\left(x-(z-\varpi_n(z))\right)^2}}+o(1), & \text{in the real case}\\
o(1), & \text{in the complex case}\end{cases}\\
&\to\begin{cases}\frac{c\varpi^2(z)m_3(z)+c^2\varpi^4(z)\left(m^0(z)\right)'m_3(z)-c^2\varpi^3(z)\left(m^0(z)\right)'m_2(z)}{1-c\varpi^2(z)m_2(z)}, & \text{in the real case}\\
0, & \text{in the complex case}\end{cases}
\end{aligned}$$

where we use

$$m_r(z)=\int\frac{dH(x)}{(x-z+\varpi(z))^r},\quad \left(m^0(z)\right)'=\int\frac{dF_{c,H}(x)}{(x-z)^2}.$$

10.4.4 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).

□

10.5 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 8 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 (10.11) first. When $\mathbf{T}_n$ is a zero matrix, by (10.5) the Stieltjes transform $m^0(z)$ satisfies the following equation

$$m^0(z)=\frac{1}{1-z-c-cm^0(z)}. \tag{10.94}$$

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

$$\underline{m}^0(z)=-\frac{1-c}{z}+cm^0(z). \tag{10.95}$$

By (10.94) and (10.95), we have

$$\underline{m}^0(z)=-\frac{1}{zm^0(z)}-1. \tag{10.96}$$

Moreover, from (10.5)

$$\frac{1}{m^0(z)} = \frac{1}{1 + cm^0(z)} - z. \tag{10.97}$$

Combining (10.96) with (10.97), we get

$$z\underline{m}^0(z) = -\frac{1}{1 + cm^0(z)}. \tag{10.98}$$

We then conclude from (10.98) that

$$\frac{c\big(m^0(z)\big)'}{\big(1 + cm^0(z)\big)^2} = -\Big(\frac{1}{1 + cm^0(z)}\Big)' = \big(z\underline{m}^0(z)\big)'. \tag{10.99}$$

It follows that

$$\begin{aligned} & 1 + \frac{c(m^0(z_1))'}{(1 + cm^0(z_1))^2} + \frac{c(m^0(z_2))'}{(1 + cm^0(z_2))^2} + \frac{c(m^0(z_1))'}{(1 + cm^0(z_1))^2}\frac{c(m^0(z_2))'}{(1 + cm^0(z_2))^2} \\ = \; & 1 + \big(z_1\underline{m}^0(z_1)\big)' + \big(z_2\underline{m}^0(z_2)\big)' + \big(z_1\underline{m}^0(z_1)\big)'\big(z_1\underline{m}^0(z_1)\big)'. \end{aligned} \tag{10.100}$$

On the other hand, since (10.95) has an inverse (one may also see (1.2) in Bai and Silverstein (2004))

$$z = -\frac{1}{\underline{m}^0(z)} + \frac{c}{1 + \underline{m}^0(z)}, \tag{10.101}$$

we have

$$z\big(1 + \underline{m}^0(z)\big) = -1 + c - \frac{1}{\underline{m}^0(z)}. \tag{10.102}$$

From this, we have

$$\big(z\underline{m}^0(z)\big)' = \frac{\big(\underline{m}^0(z)\big)'}{\big(\underline{m}^0(z)\big)^2} - 1. \tag{10.103}$$

Thus by (10.98) and (10.102), we have

$$\begin{aligned} & z_2 - z_1 + s(z_1, z_2) = z_2\big(1 + \underline{m}^0(z_2)\big) - z_1\big(1 + \underline{m}^0(z_1)\big) \\ = \; & -\frac{1}{\underline{m}^0(z_2)} + \frac{1}{\underline{m}^0(z_1)} = \frac{\underline{m}^0(z_2) - \underline{m}^0(z_1)}{\underline{m}^0(z_1)\underline{m}^0(z_2)}. \end{aligned} \tag{10.104}$$

We then conclude from (10.100), (10.103) and (10.104) that

$$\begin{aligned} & [1 + \frac{c(m^0(z_1))'}{(1 + cm^0(z_1))^2} + \frac{c(m^0(z_2))'}{(1 + cm^0(z_2))^2} + \frac{c(m^0(z_1))'}{(1 + cm^0(z_1))^2}\frac{c(m^0(z_2))'}{(1 + cm^0(z_2))^2}]\frac{1}{(z_2 - z_1 + s(z_1, z_2))^2} \\ = \; & \Big[1 + \frac{(\underline{m}^0(z_1))'}{(\underline{m}^0(z_1))^2} - 1 + \frac{(\underline{m}^0(z_2))'}{(\underline{m}^0(z_2))^2} - 1 + \Big(\frac{(\underline{m}^0(z_1))'}{(\underline{m}^0(z_1))^2} - 1\Big)\Big(\frac{(\underline{m}^0(z_2))'}{(\underline{m}^0(z_2))^2} - 1\Big)\Big]\Big(\frac{\underline{m}^0(z_2) - \underline{m}^0(z_1)}{\underline{m}^0(z_1)\underline{m}^0(z_2)}\Big)^2 \\ = \; & \frac{(\underline{m}^0(z_1))'(\underline{m}^0(z_2))'}{(\underline{m}^0(z_1) - \underline{m}^0(z_2))^2}. \end{aligned} \tag{10.105}$$

In view of (10.105) we see that (1.7) in Bai and Silverstein (2004) and (10.11) 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 (10.10). When $\mathbf{T}_n = \mathbf{0}$, by (10.5), we get

$$m_r(z) = (m^0(z))^r. \tag{10.106}$$

Moreover we obtain from (10.96) and (10.98)

$$m^0(z) = -\frac{1}{z(\underline{m}^0(z)+1)}, \quad \varpi(z) = -z\underline{m}^0(z). \tag{10.107}$$

From (10.106) and (10.107), it follows that

$$\varpi^r(z)m_r(z) = \big(\underline{m}^0(z)\big)^r\big(1+\underline{m}^0(z)\big)^{-r}. \tag{10.108}$$

This ensures that $\mathbb{E}M(z)$ in (10.21) can be written as

$$\mathbb{E}M(z) = \frac{c(\underline{m}^0(z))^3(1+\underline{m}^0(z))^{-3}\Big(\frac{1}{\varpi(z)} + c\varpi(z)(m^0(z))' - c(m^0(z))'\frac{1}{m^0(z)}\Big)}{1-c(\underline{m}^0(z))^2(1+\underline{m}^0(z))^{-2}}. \tag{10.109}$$

Comparing (10.109) with (1.6) in Bai and Silverstein (2004), it is sufficient to prove that

$$\frac{1}{\varpi(z)} + c\varpi(z)(m^0(z))' - c(m^0(z))'\frac{1}{m^0(z)} = \frac{1}{1-c(\underline{m}^0(z))^2(1+\underline{m}^0(z))^{-2}}. \tag{10.110}$$

In view of (10.107) we have

$$\frac{1}{\varpi(z)} + c\varpi(z)(m^0(z))' - c(m^0(z))'\frac{1}{m^0(z)} = -\frac{1}{z\underline{m}^0(z)} + c\big(m^0(z)\big)'z. \tag{10.111}$$

Taking derivative with respect to z on the both sides of (10.95) we have

$$c\big(m^0(z)\big)'z = \big(\underline{m}^0(z)\big)'z - \frac{1-c}{z} = \frac{c\big(\underline{m}^0(z)\big)^2\big(1+\underline{m}^0(z)\big)^{-1} - \underline{m}^0(z)}{1-c\big(\underline{m}^0(z)\big)^2\big(1+\underline{m}^0(z)\big)^{-2}} - \frac{1-c}{z}, \tag{10.112}$$

the last step using the expression (10.101) for z.

In view of (10.110), (10.111) and (10.112) it is enough to show

$$-\frac{1}{z}\Big(\frac{1}{\underline{m}^0(z)} + 1 - c\Big) = \frac{1+\underline{m}^0(z) - c\big(\underline{m}^0(z)\big)^2\big(1+\underline{m}^0(z)\big)^{-1}}{1-c\big(\underline{m}^0(z)\big)^2\big(1+\underline{m}^0(z)\big)^{-2}}. \tag{10.113}$$

From (10.102) the left hand side of (10.113) 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 (10.113). The proof is completed.

References

T.W. Anderson, An introduction to multivariate statistical analysis, *New York: Wiley*, **2th edtion**(1984).

J. Bai, S. Ng, Determine the number of factors in approximate factor models, *Econometrica*, **70(1)**(2002), 191-221.

Z.D. Bai, J. Chen, J. Yao, On estimation of the population spectral distribution from a high-dimensional sample covariance matrix, *Aust. N. Z. J. Stat.*, **52**(2010), 423-437.

Z.D. Bai, H. Saranadasa, Effect of high dimension comparison of significance tests for a high dimensional two sample problem, *Statist. Sinica*, **6**(1996), 311-329.

Z.D. Bai, J.W. Silverstein, No eigenvalues outside the support of the limiting spectral distribution of large dimensional random matrices, *Ann. Probab.*, **26**(1998), 316-345.

Z.D. Bai, J.W. Silverstein, CLT for linear spectral statistics of large-dimensional sample covariance matrices, *Ann. Probab.*, **32(1A)**(2004), 553-605.

Z.D. Bai, J.W. Silverstein, Spectral analysis of large-dimensional random matrices, *Springer*, **2nd edition**(2009).

J. Baik and J. W. Silverstein, Eigenvalues of large sample covariance matrices of spiked population models, *J. Multivariate Anal.*, **97**(2006), 1382-1408.

P.J. Bickel, E. Levina, Regularized estimation of large covariance matrices, *Annal. Statist.*, **36(1)**(2008), 199-227.

P.J. Bickel, E. Levina, Covariance regularization by thresholding, *Annal. Statist.*, **36(6)**(2008), 2577-2604.

P. Billingsley, Convergence of probability measures. *Wiley*, **2nd edition**(1999).

M. Birke and H. Dette, A note on testing the covariance matrix for large dimension. *Statistics Probability Letters.*, **74** (2005) 281-289.

D.L. Burkholder, Distribution function inequalities for martingales, *Ann. Probab.*, **1** (1973)19-42.

J. Fan, and Y. Fan. High dimensional classification uning features annealed independence rules, *Annal. Statist.*, **36(6)**(2008), 2605-2637.

J. Fan, S. Guo, N. Hao, Variance estimation using refitted cross-validation in ultrahigh dimensional regression, *J. R. Statist. Soc. B*, **74**2012, 37-65.

J. Fan, Y. Liao, M. Mincheva, Large covariance estimation by thresholding principal orthogonal complements, *J. R. Statist. Soc. B*, **75(4)**2013, 603-680.

Y. Fujikoshi, V.V. Ulyanov, R. Shimizu, Multivariate statistics: high dimensional and large sample approximations, *Wiley*(2010).

J. Huang, J.L. Horowitz, S. Ma, Asymptotic properties of bridge estimators in sparse high-dimensional regression models, *Ann. Statist.*, **36(2)**(2008), 587-613.

I.M. Johnstone, Multivariate analysis and Jacobi ensembles: largest eigenvalues, tracy-widom limits and rates of convergence, *Ann. Stat.*, **36(6)**(2008),2638-2716.

D. Jonsson, Some limit theorems of the eigenvalues of sample convariance matrix, *J. Multivariate Anal.*, **12**(1982)1-38.

N. El Karoui, Spectrum estimation for large dimensional covariance matrices using random matrix theory, *Ann. Statist.*, **36**, 2757-2790.

A. Lytova, L. Pastur, Central limit theorem for linear eigenvalue statistics of random matrices with independent entries, *Ann. Probab.*, **37(5)**(2009), 1778-1840.

K.V. Mardia, J.T. Kent, and J.M. Bibby, Multivariate Analysis, *Academic Press*(1979), London. MR0560319.

J.W. Lindeberg, Eine neue herleitung des exponentialgesetzes in der wahrscheinlichkeitsrechnung, *Math. Z.*, **15**(1922)211–225.

V.A. Marčenko, L.A. Pastur, Distribution for some sets of random matrices, *Math. USSR-Sb.*, **1**(1967) 457-483.

G.M. Pan, Strong Convergence of The Empirical Distribution of Eigenvalues of Sample Covariance Matrices with A Perturbation Matrix, *J. Multivariate Anal.*, **101**(2010) 1330-1338.

G.M. Pan, W. Zhou, Central limit theorem for signal-to-interference ratio of reduced rank linear receiver, *Ann. Appl. Probab.*, **18(3)**(2008) 1232-1270.

M. Rudelson and R. Vershynin. The smallest singular value of a random rectangular matrix. Preprint (2011). http://www.math.missouri.edu/ rudelson/papers/rv-rectangular-matrices.pdf

J.W. Silverstein, Z.D. Bai, On the empirical distribution of eigenvalues of a class of large dimensional random matrices, *J. Multivariate Anal.*, **54**(1995), 175-192.

J.W. Silverstein, S.I. Choi, Analysis of the limiting spectral distribution of large dimensional random matrices. *J. Multivariate Anal.*, **54**(1995), 295-309.

N.H. Timm, Applied multivariate analysis, *Springer*, (2001).

K.W. Wachter, The limiting empirical measures of multiple discriminant ratios, *Ann. Stat.*, **8**(1980), 937-957.

Y.Q. Yin, Limiting spectral distribution for a class of random matrices, *J. Multivariate Anal.*, **20**(1986), 50-68.

Y.R. Yang, G.M. Pan, The convergence of the empirical distribution of canonical correlation coefficients, *E. Journal. Prob.* **17** (2012), 1-13.

S.R. Zheng, Central limit theorems for linear spectral statistics of large dimensional F-matrix, *Annales de l'Institut Henri Poincaré.* **48(2)**(2012), 444-476.

S.S. Wilks, On the independence of k sets of normally distributed statistical variables, *Econometria.* **3**(1935) 309-326.

Table 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 2: Empirical sizes of the proposed test T_n at 0.05 significance level for DGP(a)-DGP(d).

(p_1, p_2, n)	T_n DGP(a)	T_n DGP(b)	T_n DGP(c)	T_n DGP(d)
(100,50,80)	0.0569	0.0462	0.0622	0.0410
(140,70,120)	0.0573	0.0429	0.0582	0.0399
(180,90,150)	0.0577	0.0452	0.0470	0.0429
(200,100,170)	0.0552	0.0429	0.0467	0.0488
(240,120,180)	0.0581	0.0510	0.0533	0.0410
(280,140,250)	0.0571	0.0483	0.0518	0.0458
(320,160,270)	0.0521	0.0479	0.0550	0.0512
(360,180,290)	0.0529	0.0489	0.0530	0.0492
(400,190,300)	0.0542	0.0522	0.0481	0.0512
(440,220,330)	0.0557	0.0529	0.0469	0.0462
(480,240,350)	0.0531	0.0562	0.0471	0.0457

*The parameter t in the statistic T_n takes a value of 40. For GDP(a), we use the original statistic T_n in Theorem 3; for GDP(b), the statistic in Theorem 8 is used; for GDP (c) and (d), the dividing-sample statistic in Theorem 5 is utilized.

Table 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 (6.4). r is the number of factors.

Table 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
(100,50,80)	0.2460	0.5330	0.8220	0.9220
(140,70,120)	0.2750	0.6180	0.8090	0.9420
(180,90,150)	0.2990	0.5990	0.8340	0.9580
(200,100,170)	0.3120	0.6010	0.8440	0.9570
(240,120,180)	0.3540	0.6000	0.8710	0.9680
(280,140,250)	0.3220	0.5790	0.8920	0.9720
(320,160,270)	0.3630	0.5990	0.8500	0.9750
(360,180,290)	0.3240	0.6650	0.8390	0.9900
(400,200,310)	0.3790	0.6290	0.8900	0.9830
(440,220,330)	0.3740	0.6590	0.9000	0.9920
(480,240,350)	0.3690	0.6600	0.8890	0.9980

*The powers are under the alternative hypothesis that **x** and **y** satisfy the factor model (6.4). r is the number of factors. The parameter t in the statistic T_n takes value of 40. For T_n, we use its modified dividing-sample version in Theorem 5.

Table 5: Empirical powers of the proposed test S_n at 0.05 significance level for **x** and **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 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)	$(0.9, 0.1)$	$(0.8, 0.2)$	$(0.7, 0.3)$	$(0.6, 0.4)$	$(0.5, 0.5)$
(100,50,80)	0.5710	05830	0.7010	0.8220	0.9530
(140,70,120)	0.6290	0.7610	0.7990	0.8920	0.9670
(180,90,150)	0.7330	0.8420	0.9490	0.9810	1.0000
(200,100,170)	0.8020	0.8560	1.0000	1.0000	1.0000
(240,120,180)	0.8920	0.9620	1.0000	1.0000	1.0000
(280,140,250)	0.9370	0.9890	1.0000	1.0000	1.0000
(320,160,270)	0.9800	0.9970	1.0000	1.0000	1.0000
(360,180,290)	0.9870	0.9960	1.0000	1.0000	1.0000
(400,200,310)	0.9900	0.9990	1.0000	1.0000	1.0000
(440,220,330)	0.9960	1.0000	1.0000	1.0000	1.0000
(480,240,350)	0.9960	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 40. The original statistic T_n in Theorem 3 is used.

Table 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 $Y_{it} = X_{it}^{\omega} - EX_{it}^{\omega}, i = 1, 2, \ldots, p_1$ 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 8: Empirical powers of the proposed test T_n at 0.05 significance level for uncorrelated but dependent case.

(p_1, p_2, n)	$\omega = 4$	$\omega = 10$
(100,50,80)	0.7010	0.8520
(140,70,120)	0.6990	0.8730
(180,90,150)	0.7210	0.8880
(200,100,170)	0.7830	0.8930
(240,120,180)	0.8320	0.9250
(280,140,250)	0.8590	0.9750
(320,160,270)	0.8990	0.9840
(360,180,290)	0.9120	0.9900
(400,200,310)	0.9420	0.9960
(440,220,330)	0.9770	1.0000
(480,240,350)	0.9890	1.0000

*The powers are under the alternative hypothesis that $Y_{it} = X_{it}^{\omega} - EX_{it}^{\omega}, i = 1, 2, \ldots, p_1$ 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 40. The original statistic T_n in Theorem 3 is used.

Table 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.
$[0, 0.05]$	56	60
$[0.05, 0.1]$	22	20
$[0.1, 0.2]$	9	12
$[0.2, 0.3]$	2	5
$[0.3, 0.4]$	10	0
$[0.4, 0.5]$	1	3
$[0.6, 0.7]$	0	0
$[0.8, 0.9]$	0	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 daily stock returns during the period 1990.1.1 $-$ 2002.1.1. The number of repeated experiments is 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 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.
$[0, 0.05]$	76	84
$[0.05, 0.1]$	10	12
$[0.1, 0.2]$	4	2
$[0.2, 0.3]$	7	1
$[0.3, 0.4]$	0	1
$[0.4, 0.5]$	2	0
$[0.6, 0.7]$	1	0
$[0.8, 0.9]$	0	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 daily stock returns during the period 1990.1.1 − 2002.1.1. The number of repeated experiments is 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 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.
$[0, 0.05]$	90	92
$[0.05, 0.1]$	4	5
$[0.1, 0.2]$	5	1
$[0.2, 0.3]$	1	2
$[0.3, 0.4]$	0	0
$[0.4, 0.5]$	0	0
$[0.6, 0.7]$	0	0
$[0.8, 0.9]$	0	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 daily stock returns during the period 1990.1.1 − 2002.1.1. The number of repeated experiments is 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.