The Annals of Statistics
2015, Vol. 43, No. 2, 467–500
DOI: 10.1214/14-AOS1284

INDEPENDENCE TEST FOR HIGH DIMENSIONAL DATA BASED ON REGULARIZED CANONICAL CORRELATION COEFFICIENTS

BY YANRONG YANG AND GUANGMING PAN[1]

Monash University and Nanyang Technological University

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 the New York Stock Exchange (NYSE) is detected by the proposed test.

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 [1]). There have been a substantial body of research work dealing with high dimensional data, for example, [2, 4, 7, 9, 10, 12], etc.

The importance of the independence assumption for inference arises in many aspects of multivariate analysis. For example, it is often the case in multivariate analysis that a number of variables can be rationally classified into several mutually exclusive categories. When variables can be grouped in such a way, a natural question is whether there is any significant relationship between the groups of variables. In other words, can we claim that the groups are mutually independent so that further 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, [20] used the likelihood ratio statistic to test independence

Received May 2012; revised June 2014.

[1]Supported by Ministry of Education in Singapore under Tier 2 Grant ARC 14/11 and Tier 1 Grant RG25/14.

MSC2010 subject classifications. 60K35.

Key words and phrases. Canonical correlation coefficients, central limit theorem, large dimensional random matrix theory, independence test, linear spectral statistics.

for k sets of normal distributed random variables and one may also refer to Chapter 12 of [1] 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

$$\begin{aligned} &\mathbb{H}_0 : \mathbf{x} \text{ and } \mathbf{y} \text{ are independent;} \quad \text{against} \\ &\mathbb{H}_1 : \mathbf{x} \text{ and } \mathbf{y} \text{ are dependent,} \end{aligned} \tag{1.1}$$

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

It is well known that canonical correlation analysis (CCA) deals with the correlation structure between two random vectors (see Chapter 12 of [1]). 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 $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n) = (X_{ij})_{p_1 \times n}$ and $p_2 \times n$ random matrix $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_n) = (Y_{ij})_{p_2 \times n}$, 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

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

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 $\gamma_1, \dots, \gamma_{p_1}$ can be found.

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

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

Regarding this point, one may refer to page 284 of [15]. 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 $\rho_1, \rho_2, \dots, \rho_{p_1}$ 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

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

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

However, it is inappropriate to use these types of sample covariance matrices to replace population covariance matrices to test (1.1) in some cases. We demonstrate such an example in Section 6.3.

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

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

where

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

We also call $\mathbf{A}_{\mathbf{xx}}$, $\mathbf{A}_{\mathbf{yy}}$ and $\mathbf{A}_{\mathbf{xy}}$ sample covariance matrices, as in the random matrix community. However, whichever sample covariance matrices are used they are not consistent estimators of population covariance matrices, which is called the "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 [20] and [1]) 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 $r_1^2, r_2^2, \dots, r_{p_1}^2$ are the eigenvalues of the matrix

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

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

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

where $\mathbf{A}_{t\mathbf{x}}^{-1} = (\frac{1}{n}\mathbf{X}\mathbf{X}^T + t\mathbf{I}_{p_1})^{-1}$, t is a positive constant number and $\mathbf{I}_{p_1}$ is a $p_1 \times p_1$ identity matrix, and $\mathbf{A}_{\mathbf{yy}}^{-}$ denotes the Moore–Penrose pseudoinverse matrix of $\mathbf{A}_{\mathbf{yy}}$. Hence, $\mathbf{T}_{\mathbf{xy}}$ is well defined even in the case of $p_1, p_2 \geq n$. Moreover, $\mathbf{T}_{\mathbf{xy}}$ reduces to $\mathbf{S}_{\mathbf{xy}}$ 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 $\mathcal{P} = (\mathcal{P}_1, \mathbf{0})$, $\mathcal{P}_1 = \operatorname{diag}(\gamma_1, \ldots, \gamma_{p_1})$ 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

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

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 [11]. A natural criteria for this test should be $\sum_{i=1}^{p_1} \gamma_i^2$.

As pointed out, r_i is not a consistent estimator of the corresponding population version γ_i in the high dimensional case. However, fortunately, the classical sample canonical correlation coefficients $r_1, r_2, \ldots, r_{p_1}$ or its regularized analogous still contain important information so that hypothesis testing for (1.1) is possible although the classical likelihood ratio statistic does not work well in the high dimensional case. This is due to the fact that the limits of the empirical spectral distribution (ESD) of $r_1, \ldots, r_{p_1}$ 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, that is, $\sum_{i=1}^{p_1} r_i^2$. 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 (1.1), 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, that is, 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 [22]. 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 non-Gaussian 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 [23] 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, [19] investigated the limit of the ESD of the classical sample canonical correlation coefficients $r_1, r_2, \dots, r_{p_1}$ and [13] 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 (1.1) based on large dimensional random matrix theory and contains the main results. Two modified statistics are further provided in Section 3 and estimators for some unknown parameters in the asymptotic mean and variance for the asymptotic distribution are also proposed. 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. Some useful lemmas and proofs of all theorems and Propositions 4–5 are relegated to Appendix A while one theorem about the CLT of a sample covariance matrix plus a perturbation matrix is provided in Appendix B. All appendices are given in the supplementary material [23].

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 $\mathbf{X} = \mathbf{\Sigma_{xx}^{1/2}}\mathbf{W}$ and $\mathbf{Y} = \mathbf{\Sigma_{yy}^{1/2}}\mathbf{V}$, where $\mathbf{W} = (\mathbf{w}_1, \dots, \mathbf{w}_n) = (W_{ij})_{i,j=1}^{p_1,n}$ consists of i.i.d. real random variables $\{W_{ij}\}$ with $EW_{11} = 0$ and $E|W_{11}|^2 = 1$; $\mathbf{V} = (\mathbf{v}_1, \dots, \mathbf{v}_n) = (V_{ij})_{i,j=1}^{p_2,n}$ consists of i.i.d. real random variables with $EV_{11} = 0$ and $E|V_{11}|^2 = 1$; $\mathbf{\Sigma_{xx}^{1/2}}$ and

$\mathbf{\Sigma}_{\mathbf{yy}}^{1/2}$ are Hermitian square roots of positive definite matrices $\mathbf{\Sigma}_{\mathbf{xx}}$ and $\mathbf{\Sigma}_{\mathbf{yy}}$, respectively, so that $(\mathbf{\Sigma}_{\mathbf{xx}}^{1/2})^2 = \mathbf{\Sigma}_{\mathbf{xx}}$ and $(\mathbf{\Sigma}_{\mathbf{yy}}^{1/2})^2 = \mathbf{\Sigma}_{\mathbf{yy}}$.

ASSUMPTION 4. $F^{\mathbf{\Sigma}_{\mathbf{xx}}} \xrightarrow{D} H$, a proper cumulative distribution function.

REMARK 1. By the definition of the matrix $\mathbf{S}_{\mathbf{xy}}$, 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 $\mu_1 \le \mu_2 \le \cdots \le \mu_n$ by

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

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, [19] proved that the empirical measure of the classical sample canonical correlation coefficients $r_1, r_2, \dots, r_{p_1}$ converges in probability to a fixed distribution whose density is given by

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

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 $r_1, r_2, \dots, r_{p_1}$ is defined as in (2.1) with μ_i replaced by r_i.

[21] 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* $\{X_{ij}, 1 \le i \le p_1, 1 \le j \le n\}$ *and* $\{Y_{ij}, 1 \le i \le p_2, 1 \le j \le n\}$ *are independent. Then the empirical measure of* $r_1, r_2, \dots, r_{p_1}$ *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 (1.6). To this end, define the Stieltjes transform of any distribution function $G(x)$ by

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

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

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

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

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 [17] and [16], $m_{y_2t}(z)$ is the unique solution in $\mathbb{C}^+$ to

$$m_{y_2t}(z) = m_{H_t}\left(z - \frac{1}{1+y_2m_{y_2t}(z)}\right), \tag{2.3}$$

where $m_{H_t}(z)$ denotes the Stieltjes transform of the LSD of the matrix $t\frac{n}{n-p_2}\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}$ (one may also see (1.4) in the supplement material [23]). 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 $\mathbf{W}_1^T\mathbf{S}_{2t}^{-1}\mathbf{W}_1$ are denoted by $\underline{m}_{\mathbf{n}}(z)$ and $\underline{m}_{\mathbf{y}}(z)$, respectively, where $\mathbf{W}_1 = (\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_{p_2})$. Observe that the spectral of $\mathbf{S}_1\mathbf{S}_{2t}^{-1}$ and $\mathbf{W}_1^T\mathbf{S}_{2t}^{-1}\mathbf{W}_1$ are the same except zero eigenvalues and this leads to

$$\underline{m}_{\mathbf{y}}(z) = -\frac{1-y_1}{z} + y_1m_{\mathbf{y}}(z). \tag{2.4}$$

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

THEOREM 2.1. *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* $\widetilde{F}(\frac{\lambda}{q(1-\lambda)})$ *with* $q = \frac{c_2'}{1-c_2'}$ *where* $\widetilde{F}(\lambda)$ *is a nonrandom distribution and its Stieltjes transform* $m_{\mathbf{y}}(z)$ *is the unique solution in* $\mathbb{C}^+$ *to*

$$m_{\mathbf{y}}(z) = -\int \frac{dF_{y_2t}(1/\lambda)}{\lambda(1-y_1-y_1zm_{\mathbf{y}}(z)) - z}. \tag{2.5}$$

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

$$m_{\widetilde{G}}(z) = \int \frac{dH(\lambda)}{\lambda(1-c_1'-c_1'zm_{\widetilde{G}}(z)) - z}. \tag{2.6}$$

REMARK 2. Indeed, taking $t=0$ in (2.5) recovers [19]'s result (one may refer to the result of F matrix in [5]).

Let us now introduce the test statistic. Under Assumptions 1 and 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 2.2 and 2.3 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 (1.1) with fixed dimensions is

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

(see [20] and [1]). That is, $\phi(x)$ in (2.7) takes $\log(1-x)$. Note that the density $\rho(x)$ has atom size of $\max(0,1-(1-c_2)/c_1)$ at unity by (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 $c_1+c_2\le 1$, 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 (2.7) 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 renormalize (2.7) and (2.10) as

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

and

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

Also, let

$$\begin{aligned}
&\bar{y}_1:=\frac{c_1}{1-c_2}\in(0,+\infty),\qquad \bar{y}_2:=\frac{c_1}{c_2}\in(0,1),\\
&h=\sqrt{\bar{y}_1+\bar{y}_2-\bar{y}_1\bar{y}_2},\qquad a_1=\frac{(1-h)^2}{(1-\bar{y}_2)^2},\qquad a_2=\frac{(1+h)^2}{(1-\bar{y}_2)^2},\\
&g_{\bar{y}_1,\bar{y}_2}(\lambda)=\frac{1-\bar{y}_2}{2\pi\lambda(\bar{y}_1+\bar{y}_2\lambda)}\sqrt{(a_2-\lambda)(\lambda-a_1)},\qquad a_1<\lambda<a_2.
\end{aligned} \tag{2.15}$$

THEOREM 2.2. *Let $\phi_1,\dots,\phi_s$ 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*

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

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

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

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

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

and covariance function

$$
\begin{aligned}
&\operatorname{cov}(X_{\phi_i}, X_{\phi_j}) \\
&\quad = -\lim_{r\downarrow 1} \frac{1}{4\pi^2} \oint_{|\xi_1|=1} \oint_{|\xi_2|=1} \biggl(\biggl(f_i\biggl(\frac{1+h^2+2h\Re(\xi_1)}{(1-\bar{y}_2)^2} \biggr) \\
&\qquad\qquad \times f_j\biggl(\frac{1+h^2+2h\Re(\xi_2)}{(1-\bar{y}_2)^2} \biggr) \biggr) \\
&\qquad\qquad \Big/ (\xi_1 - r\xi_2)^2 \biggr) \, d\xi_1 \, d\xi_2,
\end{aligned}
\tag{2.19}
$$

where $f_i(\lambda) = \phi_i(\frac{1}{1+((1-c_2)/c_2)\lambda})$, $\Re$ *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 2.2 is 0 and the variance is $\frac{2h^2 y_1^2 y_2^2}{(y_1+y_2)^4}$. These are calculated in Example 4.2 of [22]. Moreover, assumption (2.16) 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

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

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

THEOREM 2.3. *Let* $\phi_1, \ldots, \phi_s$ *be functions analytic in an open region in the complex plane containing the support of the LSD* $\widetilde{F}(\lambda)$ *whose Stieltjes transform is* (2.5). *In addition to Assumptions* 2–4, *suppose that the spectra norm of* $\mathbf{\Sigma_{xx}}$ *is bounded and*

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

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

$$\left(\int \phi_1(\lambda)\, dG^{(2)}_{p_1,p_2}(\lambda), \ldots, \int \phi_s(\lambda)\, dG^{(2)}_{p_1,p_2}(\lambda)\right) \tag{2.21}$$

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

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

and covariance

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

Here, q *is defined in Theorem* 2.1. *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* $c_2' \in [1, +\infty)$ $(p_2 \geq n)$, (2.21) *converges weakly to a Gaussian vector* $(X_{\phi_1}, \dots, X_{\phi_s})$ *with mean*

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

and

$$\begin{aligned} \operatorname{cov}(X_{\phi_i}, X_{\phi_j}) &= -\frac{1}{2\pi^2} \oint_{\mathcal{C}_1} \oint_{\mathcal{C}_2} \phi_i\left(\frac{t^{-1}z_1}{1+t^{-1}z_1}\right) \\ &\quad \times \phi_i\left(\frac{t^{-1}z_2}{1+t^{-1}z_2}\right) \frac{\underline{s}'(z_1)\underline{s}'(z_2)}{(\underline{s}(z_1) - \underline{s}(z_2))^2} \, dz_1 \, dz_2, \end{aligned} \tag{2.25}$$

where $\underline{s}(z)$ *is Stieltjes transform of the LSD of the matrix* $\frac{1}{n}\mathbf{W}^T \boldsymbol{\Sigma}_{\mathbf{xx}} \mathbf{W}$. *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 **x** and **y** by CCA is based on the fact that the uncorrelatedness between **x** and **y** is equivalent to independence between them when the random vector of size $(p_1 + p_2)$ consisting of the components of **x** and **y** is a Gaussian random vector. See Wilks [20] and Anderson [1]. For non-Gaussian random vectors **x** and **y**, uncorrelatedness is not equivalent to independence. CCA may fail in this case. Yet, since Theorems 2.2 and 2.3 hold for non-Gaussian random vectors **x** and **y** CCA can be still utilized to capture dependent but uncorrelated **x** and **y** such as ARCH type of dependence by considering the high power of their entries. See Section 6.5 for further discussion.

Following [14], condition (2.16) 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 $\boldsymbol{\Sigma}_{\mathbf{xx}}$ 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 $\boldsymbol{\Sigma}_{\mathbf{xx}}$ is very challenging unless it is sparse. [8] and [3] proposed some approaches to estimate the limit of the ESD of $\boldsymbol{\Sigma}_{\mathbf{xx}}$ or its moments. However, the convergence rate is not fast enough to offset the order of p_1. Indeed, Theorem 1 of [3] 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

$$m_{nt} = a_n - \frac{a_n t}{p_1} \operatorname{tr}\bigl(a_n^{-1} \boldsymbol{\Sigma}_{\mathbf{xx}} + t\mathbf{I}\bigr)^{-1} \tag{3.2}$$

with $a_n = 1 + c'_{1n} m_{nt}$ (see the proof of Theorem 3.1). An estimator of m_{nt} is then proposed as $\hat{m}_{nt}$ which is a solution to the equation

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

with $\hat{a}_n = 1 + c'_{1n} \hat{m}_{nt}$. Here, we use a thresholding estimator $\hat{\boldsymbol{\Sigma}}_{\mathbf{xx}}$ to estimate $\boldsymbol{\Sigma}_{\mathbf{xx}}$, slightly different from that proposed by [6]. Specifically speaking, suppose that the underlying random variables $\{X_{ij}\}$ are mean zero and variable one. Then define $\hat{\boldsymbol{\Sigma}}_{\mathbf{xx}}$ 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

$$p_1\biggl(\int \lambda \, dF^{\mathbf{T_{xy}}}(\lambda) - \biggl(\frac{p_2}{p_1} - \frac{p_2}{p_1} \frac{1}{1 + c'_{1n} \hat{m}_{nt}}\biggr)\biggr). \tag{3.4}$$

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

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

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

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

We then propose the resulting test statistic

$$p_1\biggl(\int \lambda \, dF^{\mathbf{T_{xy}}}(\lambda) - \bigl(1 - t\hat{m}_n^{(1t)}\bigr)\biggr), \tag{3.7}$$

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

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

THEOREM 3.1. *In addition to assumptions in Theorem 2.3, suppose that* $EX_{ij}^2 = 1$, $\sup_{i,j} E|X_{ij}|^{17} < \infty$ *for all* i *and* j *and that*

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

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

(a) *If* $c_2' < 1$, *then* $p_1(\int \lambda \, dF^{\mathbf{T}_{\mathbf{xy}}}(\lambda) - (\frac{p_2}{p_1} - \frac{p_2}{p_1} \frac{1}{1+c_{1n}' \hat{m}_{nt}}))$ *converges weakly to a normal distribution with the mean and variance given in* (2.22) *and* (2.23) *with* $\phi(\lambda) = \lambda$.

(b) *If* $c_2' \ge 1$, *then* $p_1(\int \lambda \, dF^{\mathbf{T}_{\mathbf{xy}}}(\lambda) - (1 - t\hat{m}_n^{(1t)}))$ *converges weakly to a normal distribution with the mean and variance given in part* (b) *of Theorem* 2.3 *with* $\phi(\lambda) = \lambda$.

We demonstrate an example of sparse covariance matrices 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, that is,

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

and

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

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 $\widetilde{\mathbf{Y}}^{(2)}$, that is,

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

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

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

For Group 1, it follows from Theorem 2.3 that

$$\int \lambda\, dp_1\big(F^{\mathbf{T}_{\mathbf{xy}}^{(1)}}(\lambda) - F^{2c'_{1n},2c'_{2n}}(\lambda)\big) \xrightarrow{d} Z_1, \tag{3.12}$$

where $\mathbf{T}_{\mathbf{xy}}^{(1)}$ 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 2.3 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 2.3

$$\int \lambda\, dp_1\big(F^{\mathbf{T}_{\mathbf{xy}}^{(2)}}(\lambda) - F^{2c'_{1n},c'_{2n}}(\lambda)\big) \xrightarrow{d} Z_2, \tag{3.13}$$

where $\mathbf{T}_{\mathbf{xy}}^{(2)}$ 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 2.3 with $\phi(\lambda) = \lambda$ and c'_1 replaced by $2c'_1$.

We next investigate the relation between

$$\int \lambda\, dF^{2c'_{1n},2c'_{2n}}(\lambda) \quad \text{and} \quad \int \lambda\, dF^{2c'_{1n},c'_{2n}}(\lambda),$$

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

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

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

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

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

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

where $\tilde{m}_n^{(1t)}$ 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

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

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

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

where $\dot{\mathbf{y}}_k$ consists of the last $[p_2/2]$ components of $\mathbf{y}_k$, that is, the ith component of $\dot{\mathbf{y}}_k$ is the $([p_2/2]+i)$th component of $\mathbf{y}_k$, for all $i = 1, 2, \dots, [p_2/2]$ and $k = 1, 2, \dots, [n/2]$. For Modified Group 1, by Theorem 2.3, 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 2.3 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 3.2.

THEOREM 3.2. *Suppose that assumptions in Theorem* 2.3 *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* $c'_2 \geq 1$, *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* $1/2 \leq c'_2 < 1$, *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.

Theorem 3.2 proposes test statistics which do not involve the unknown parameter H, that is, the LSD of the matrix $\mathbf{\Sigma_{xx}}$. However, their asymptotic means and asymptotic variances contain some terms involving the unknown parameter H. We below provide consistent estimators for such terms appearing in (2.22)–(2.25) in

Theorem 2.3, which, together with Slutsky's theorem and the dominant convergence theorem, are enough for applications.

We use an estimator developed by [8] for H. For easy reference, we briefly state his estimator $\hat{H}_{p_1}$ for H in the following proposition.

PROPOSITION 2 (Theorem 2 of [8]). *In addition to Assumptions 2–4, suppose that the spectra norm of $\boldsymbol{\Sigma}_{\mathbf{xx}}$ is bounded. Let $J_1, J_2, \dots$ be a sequence of integers tending to ∞. Let $z_0 \in \mathbb{C}^+$ and $r \in \mathbb{R}^+$ be such that $\mathbf{B}(z_0, r) \subset \mathbb{C}^+$, where $\mathbf{B}(z_0, r)$ denotes the closed ball of center z_0 and radius r. Let $z_1, z_2, \dots$ be a sequence of complex variables with a limit point, all contained in $\mathbf{B}(z_0, r)$. Let $\hat{H}_{p_1}$ be the solution of*

$$\hat{H}_{p_1} = \arg\min_{G} \max_{j \le J_n} \left| \frac{1}{\hat{\underline{s}}_n(z_j)} + z_j - \frac{p_1}{n} \int \frac{\lambda\, dG(\lambda)}{1 + \lambda \hat{\underline{s}}_n(z_j)} \right|, \tag{3.19}$$

where G is a probability measure and $\hat{\underline{s}}_n(z)$ is the Stieltjes transform of the ESD of the matrix $\underline{\mathbf{A}}_{\mathbf{xx}} = \frac{1}{n}\mathbf{X}^T\mathbf{X}$. Then we have

$$\hat{H}_{p_1} \Rightarrow H \qquad a.s.$$

REMARK 5. The estimator $\hat{H}_{p_1}$ given in (3.19) is proposed based on the Marčenko–Pastur equation which links the Stieltjes transform of the empirical spectral distribution of the sample covariance matrix to an integral against the population spectral distribution, that is, the LSD H satisfies the Marčenko–Pastur equation

$$-\frac{1}{\underline{s}(z)} = z - c \int \frac{\lambda\, dH(\lambda)}{1 + \lambda \underline{s}(z)} \qquad \forall z \in \mathbb{C}^+, \tag{3.20}$$

where $\underline{s}(z)$ is the limit of $\hat{\underline{s}}_n(z) = \frac{1}{n}\operatorname{tr}(\underline{\mathbf{A}}_{\mathbf{xx}} - z\mathbf{I}_n)^{-1}$.

El Karoui [8] developed an algorithm for the estimator $\hat{H}_{p_1}$ in (3.19) and we state it below.

(1) *A "basis pursuit" for measure space*. Instead of searching among all possible probability measures, the search space is restricted to mixtures of certain classes of probability measures in order to deal with (3.19). In other words, first select a "dictionary" of probability measures on the real line and then decompose the estimator on this dictionary, searching for the best coefficients. Hence, the problem can be formulated as

finding the best possible weights

$$\{\hat{w}_1, \dots, \hat{w}_K\} \qquad \text{with } d\hat{H}_{p_1} = \sum_{i=1}^{K} \hat{w}_i\, dM_i, \sum_{i=1}^{K} \hat{w}_i = 1, w_i \ge 0,$$

where the M_i's are the measures in the dictionary. A "probability measures" dictionary is given as follows:

1. Point masses $\{\delta_{t_k}(x)\}_{k=1}^{K}$, where $\{t_k\}_{k=1}^{K}$ is a grid of points.
2. Probability measures that are uniform on an interval: in this case, $dM_i^{(1)}(x) = I_{[a_i,b_i]}(x)\,dx/(b_i - a_i)$.
3. Probability measures that have a linearly increasing (or decreasing) density on an interval $[d_i, h_i]$ and density 0 elsewhere. So, for the increasing case, $dM_i^{(2)}(x) = I_{[d_i,h_i]}(x) \cdot 2(x - d_i)/((h_i - d_i)^2)\,dx$, and density 0 elsewhere.

(2) *A convex optimization problem*. Let

$$e_j = \frac{1}{\hat{\underline{s}}_n(z_j)} + z_j - \frac{p_1}{n}\int \frac{\lambda\, dG(\lambda)}{1 + \lambda\hat{\underline{s}}_n(z_j)}, \qquad j = 1, 2, \ldots, J_n,$$

where $G(\cdot)$ has the form of

$$G(x) = \sum_{i=1}^{K_1} w_i \delta_{t_i}(x) + \sum_{i=K_1+1}^{K_2} w_i\, dM_i^{(1)}(x) + \sum_{i=K_2+1}^{K} w_i\, dM_i^{(2)}(x),$$

with all points $t_k, k = 1, 2, \ldots, K_1$, intervals $[a_i, b_i], i = K_1 + 1, \ldots, K_2$ and intervals $[d_j, h_j]$, $j = K_2 + 1, \ldots, K$ being in the interval $[\ell_{p_1}, \ell_1]$. Here, ℓ_{p_1} and ℓ_1 are, respectively, the smallest and largest eigenvalues of the sample covariance matrix $\mathbf{A}_{\mathbf{xx}} = \frac{1}{n}\mathbf{X}\mathbf{X}^T$. Moreover, $1 < K_1 < K_2 < K$.

The "translation" of the problem (3.19) into a convex optimization problem is

$$\min_{w_1,\ldots,w_K,u} u$$

$$\forall j = 1, \ldots, J_n, \qquad -u \le \Re(e_j) \le u,$$

$$\forall j = 1, \ldots, J_n, \qquad -u \le \Im(e_j) \le u$$

$$\text{subject to} \qquad \sum_{i=1}^{K} w_i = 1 \text{ and } w_i \ge 0, \forall i = 1, 2, \ldots, K.$$

The following proposition provides consistency of the proposed algorithm above.

PROPOSITION 3 (Corollary 1 of [8]). *Assume the same assumptions as in Proposition* 2. *Call* $\hat{H}_{p_1}$ *the solution of* (3.19), *where the optimization is now over measures which are sums of atoms*, *the locations of which are restricted to belong to a grid* (*depending on* n) *whose step size is going to* 0 *as* $n \to \infty$. *Then*

$$\hat{H}_{p_1} \Rightarrow H \qquad a.s.$$

Then the estimator $\hat{H}_{p_1}$ derived from the algorithm has the form of

$$d\hat{H}_{p_1}(x) = \sum_{i=1}^{K_1} \hat{w}_i \delta_{t_i}(x) + \sum_{i=K_1+1}^{K_2} \hat{w}_i\, dM_i^{(1)}(x) + \sum_{i=K_2+1}^{K} \hat{w}_i\, dM_i^{(2)}(x), \tag{3.21}$$

where $\{\hat{w}_i, i = 1, 2, \dots, K\}$ is the solution of the convex optimization problem above. In practice, we follow the implementation details in Appendix of [8] to derive the estimator $\hat{H}_{p_1}(x)$.

Once the estimator (3.21) of the LSD H is available, we are in a position to provide estimators for $m_H(z)$ and $m_{y_{2t}}(z)$, the Stieltjes transforms of H and the LSD of the matrix $\mathbf{S}_{2t}$, respectively.

PROPOSITION 4. *For any $z \in \mathcal{C}_1$, the contour specified in Theorem 2.3, let*

$$\begin{aligned}\hat{m}_H(z) = &\sum_{i=1}^{K_1} \frac{\hat{w}_i}{t_i - z} + \sum_{i=K_1+1}^{K_2} \frac{\hat{w}_i}{b_i - a_i} \log \frac{b_i - z}{a_i - z} \\ &+ \sum_{i=K_2+1}^{K} \frac{2\hat{w}_i}{(h_i - d_i)^2} \left(h_i - d_i + (z - d_i) \log \frac{h_i - z}{d_i - z} \right),\end{aligned} \tag{3.22}$$

where log *stands for the corresponding principal branch. The estimator $\hat{m}_{y_{2t}}(z)$ satisfies the equation*

$$\hat{m}_{y_{2t}}(z) = \hat{m}_{H_t}\left(z - \frac{1}{1 + y_{2n}\hat{m}_{y_{2t}}(z)} \right), \tag{3.23}$$

where $\hat{m}_{H_t}(z)$ is

$$\hat{m}_{H_t}(z) = -\frac{1}{z} - \frac{t}{(1 - c'_{2n})z^2} \hat{m}_H\left(\frac{t}{(1 - c'_{2n})z} \right). \tag{3.24}$$

Then $\hat{m}_H(z)$ in (3.22) and $\hat{m}_{y_{2t}}(z)$ in (3.23) are consistent estimators of $m_H(z)$ and $m_{y_{2t}}(z)$, respectively.

REMARK 6. As to choices of intervals $[a_i, b_i] : i = K_1 + 1, K_1 + 2, \dots, K_2$; and $[h_j, d_j] : j = K_2 + 1, K_2 + 2, \dots, K$ in (3.21), we follow the implementation details provided in Appendix of [8]. Furthermore, choice of $(z_j, \underline{\hat{s}}_n(z_j))$ in the convex optimization problem, choice of interval to estimate $H(\cdot)$, and choice of dictionary are all provided in the Appendix of [8].

With consistent estimators for $m_H(z)$ and $m_{y_{2t}}(z)$ in Proposition 4, we may further provide consistent estimators for all terms appearing in (2.22)–(2.25).

PROPOSITION 5. 1. *The estimator for $m_{\mathbf{y}}(z)$ is*

$$\hat{m}_{\mathbf{y}}(z) = \frac{q_n}{(1 + q_n z)^2} \left(m_{\mathbf{T_{xy}}}\left(\frac{q_n z}{1 + q_n z} \right) - 1 - q_n z \right),$$

where $y_{1n} = \frac{c'_{1n}}{c'_{2n}}$, $y_{2n} = \frac{c'_{1n}}{1 - c'_{2n}}$, $q_n = \frac{c'_{2n}}{1 - c'_{2n}}$ and $m_{\mathbf{T_{xy}}}(z)$ is the Stieltjes transform of the matrix $\mathbf{T_{xy}}$.

2. *The estimators for* $\varpi(z)$ *and* $s(z_1, z_2)$ *are, respectively,*

$$\hat{\varpi}(z) = \frac{1}{1 + y_{2n}\hat{m}_{y_{2t}}(z)},$$

$$\hat{s}(z_1, z_2) = \frac{1}{1 + y_{2n}\hat{m}_{y_{2t}}(z_1)} - \frac{1}{1 + y_{2n}\hat{m}_{y_{2t}}(z_2)}.$$

3. *The estimators for* $m_r(z)$ *with* $r = 2, 3$ *are*

$$\hat{m}_2(z) = \hat{m}_{H_t}^{(1)}\big(z - \hat{\varpi}(z)\big), \qquad \hat{m}_3(z) = \tfrac{1}{2}\hat{m}_{H_t}^{(2)}\big(z - \hat{\varpi}(z)\big),$$

respectively, where $\hat{m}_{H_t}^{(j)}(z)$ *is the* j*th derivative of* $\hat{m}_{H_t}(z)$ *with respect to* z *with* $j = 1, 2$.

4. *The estimators for* $g(z)$ *and* $h(z)$ *are*

$$\hat{g}(z) = \frac{y_{2n}(\hat{m}_{y_{2t}}^{(1)}(-\underline{\hat{m}}_{\mathbf{y}}(z)))}{(1 + y_{2n}\hat{m}_{y_{2t}}(-\underline{\hat{m}}_{\mathbf{y}}(z)))^2}, \qquad \hat{h}(z) = \frac{-\underline{\hat{m}}_{\mathbf{y}}^2(z)}{1 - y_{1n}\underline{\hat{m}}_{\mathbf{y}}^2(z)\hat{m}_{y_{2t}}^{(1)}(-\underline{\hat{m}}_{\mathbf{y}}(z))},$$

respectively, where $\underline{\hat{m}}_{\mathbf{y}}(z) = -\frac{1-y_{1n}}{z} + y_{1n}\hat{m}_{\mathbf{y}}(z)$.

5. *The estimators for* $\varpi_1 = \int \underline{m}_{\mathbf{y}}^3(z)x[x + \underline{m}_{\mathbf{y}}(z)]^{-3}\,dF_{y_{2t}}(x)$ *and* $\varpi_2 = \int \underline{m}_{\mathbf{y}}^2(z)[x + \underline{m}_{\mathbf{y}}(z)]^{-2}\,dF_{y_{2t}}(x)$ *are*

$$\hat{\varpi}_1 = \underline{\hat{m}}_{\mathbf{y}}^3(z)\hat{m}_{y_{2t}}^{(1)}\big(-\underline{\hat{m}}_{\mathbf{y}}(z)\big) - \tfrac{1}{2}\underline{\hat{m}}_{\mathbf{y}}^4(z)\hat{m}_{y_{2t}}^{(2)}\big(-\underline{\hat{m}}_{\mathbf{y}}(z)\big),$$

$$\hat{\varpi}_2 = \underline{\hat{m}}_{\mathbf{y}}^2(z)\hat{m}_{y_{2t}}^{(1)}\big(-\underline{\hat{m}}_{\mathbf{y}}(z)\big),$$

respectively.

6. *The estimators for* $\varpi_3 = \int [1 + \lambda \underline{s}^3(z)]^{-3}\underline{s}(z)^3\lambda^2\,dH(\lambda)$ *and* $\varpi_4 = \int \underline{s}^2(z) \times \lambda^2[1 + \lambda\underline{s}(z)]^{-2}\,dH(\lambda)$ *are*

$$\begin{aligned}\hat{\varpi}_3 &= \frac{1}{[\underline{\hat{s}}_n(z)]^6}\hat{m}_H\left(-\frac{1}{[\underline{\hat{s}}_n(z)]^3}\right) - \frac{2}{[\underline{\hat{s}}_n(z)]^9}\hat{m}_H^{(1)}\left(-\frac{1}{\underline{\hat{s}}_n(z)}\right)\\ &\quad + \frac{1}{2[\underline{\hat{s}}_n(z)]^{12}}\hat{m}_H^{(2)}\left(-\frac{1}{\underline{\hat{s}}_n(z)}\right),\\ \hat{\varpi}_4 &= 1 - \frac{2}{\underline{\hat{s}}_n(z)}\hat{m}_H\left(-\frac{1}{\underline{\hat{s}}_n(z)}\right) + \frac{1}{[\underline{\hat{s}}_n(z)]^2}\hat{m}_H^{(1)}\left(-\frac{1}{\underline{\hat{s}}_n(z)}\right),\end{aligned}$$

respectively, where $\underline{\hat{s}}_n(z)$ *is defined as* $\underline{\hat{s}}_n(z) = \frac{1}{n}\operatorname{tr}(\underline{\mathbf{A}}_{\mathbf{xx}} - z\mathbf{I}_n)^{-1}$ *with* $\underline{\mathbf{A}}_{\mathbf{xx}} = \frac{1}{n}\mathbf{X}^T\mathbf{X}$.

All estimators listed above are consistent for the corresponding unknown parameters.

The proofs of Propositions 4 and 5 are provided in Appendix A of [23].

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

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

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 (1.6) 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 (2.11) and (2.12), and let $R^{(i)}_n = \int \lambda \, dG^{(i)}_{p_1,p_2}$.

THEOREM 4.1. *In addition to assumptions in Theorems 2.2 or 2.3 suppose that for any $M > 0$*

$$P\big(\big|\operatorname{tr}(\mathbf{S} - \mathbf{S_{xy}})\big| \ge M\big) \to 1, \qquad P\big(\big|\operatorname{tr}(\mathbf{T} - \mathbf{T_{xy}})\big| \ge M\big) \to 1. \tag{4.1}$$

Then

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

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 7. For example, one may take $\mathbf{S} = (\mathbf{XLX}^T)^{-1}\mathbf{XLY}^T(\mathbf{YLY}^T)^{-1} \times \mathbf{YLX}^T$ 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 $\mathbf{e} = x^2(1, 1, \dots, 1)$ and x^2 having finite moment, then under assumptions in Theorems 2.2 or 2.3 it can be proved that

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

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 4.2. *In addition to assumptions in Theorem 2.3, suppose that for any $M > 0$,*

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

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

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

Then

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

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

$$\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{p_1}]_{n\times p_1}, \qquad \mathbf{X} = [\mathbf{1}_n, \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{p_2}]_{n\times p_2},$$
$$\mathbf{B} = [\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \dots, \boldsymbol{\beta}_{p_1}]_{p_2\times p_1}, \qquad \mathbf{E} = [\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{p_1}]_{n\times p_1},$$

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

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

and

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

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, $\mathbf{H} = \mathbf{A}_{\mathbf{yx}}\mathbf{A}_{\mathbf{xx}}^{-1}\mathbf{A}_{\mathbf{xy}}$ and $\mathbf{E} = \mathbf{A}_{\mathbf{yy}} - \mathbf{A}_{\mathbf{yx}}\mathbf{A}_{\mathbf{xx}}^{-1}\mathbf{A}_{\mathbf{xy}}$, so that $\det(\mathbf{H} - \lambda\mathbf{E}) = 0$ becomes $\det(\mathbf{A}_{\mathbf{yx}}\mathbf{A}_{\mathbf{xx}}^{-1}\mathbf{A}_{\mathbf{xy}} - \lambda(\mathbf{A}_{\mathbf{yy}} - \mathbf{A}_{\mathbf{yx}}\mathbf{A}_{\mathbf{xx}}^{-1}\mathbf{A}_{\mathbf{xy}})) = 0$. From Theorem 2.6.8 of [18], we have $\det(\mathbf{H} - \theta(\mathbf{H}+\mathbf{E})) = \det(\mathbf{A}_{\mathbf{yx}}\mathbf{A}_{\mathbf{xx}}^{-1}\mathbf{A}_{\mathbf{xy}} - \theta\mathbf{A}_{\mathbf{yy}}) = 0$ so that

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

Evidently, the quantities $r_i^2 = \theta_i$, $i = 1, \dots, s$ are sample canonical correlation coefficients. Therefore, the test statistic (5.4) can be rewritten as

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

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

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

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

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

that is, whether there exists cointegration relationships among the elements of the time series $\{\mathbf{y}_t\}$.

This cointegration test is equivalent to testing:

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

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

$$\mathbf{X} = (\Delta\mathbf{y}_2, \Delta\mathbf{y}_4, \ldots, \Delta\mathbf{y}_{2t-2}, \Delta\mathbf{y}_{2t}, \ldots, \Delta\mathbf{y}_T), \tag{5.11}$$

$$\mathbf{Y} = (\Delta\mathbf{y}_1, \Delta\mathbf{y}_3, \ldots, \Delta\mathbf{y}_{2t-1}, \Delta\mathbf{y}_{2t+1}, \ldots, \Delta\mathbf{y}_{T-1}). \tag{5.12}$$

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

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

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

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

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

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

with

(a) $\boldsymbol{\Sigma}_{\mathbf{xx}} = \mathbf{I}_{p_1}, \qquad \boldsymbol{\Sigma}_{\mathbf{yy}} = \mathbf{I}_{p_2};$

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

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

(d) $\boldsymbol{\Sigma}_{\mathbf{xx}} = \mathbf{B}' \operatorname{cov}(\mathbf{f}_t)\mathbf{B} + \boldsymbol{\Sigma}_{\mathbf{u}}, \qquad \boldsymbol{\Sigma}_{\mathbf{yy}} = \mathbf{I}_{p_2},$

where

$$\sigma_{kh}^{SP} = \begin{cases} 1, & k = h; \\ \theta, & k = 1; h = 2, 3, \ldots, [p_1^{1/3}]; \text{ or } h = 1; k = 2, 3, \ldots, [p_1^{1/3}]; \\ 0, & \text{others} \end{cases}$$

with $\theta = 0.2$ and

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

Here, $\mathbf{B} = \frac{1}{\sqrt{p_1}}(\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_{p_1})$ is a deterministic matrix. In the simulation, 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. $\mathrm{cov}(\mathbf{f}_t)$ is also an $r \times r$ identity matrix and $\boldsymbol{\Sigma}_{\mathbf{u}}$ is a $p_1 \times p_1$ identity matrix.

The empirical sizes of the proposed statistics S_n for data generating processes (DGPs) (a) and (b) are listed in Table 1. Moreover, the empirical sizes for the renor-

TABLE 1
Empirical sizes of the proposed test S_n and the renormalized 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	0.0642	0.0602
(140, 70, 120)	0.0573	0.0429	0.0619	0.0600
(180, 90, 150)	0.0577	0.0452	0.0623	0.0583
(200, 100, 170)	0.0552	0.0429	0.0594	0.0592
(240, 120, 180)	0.0581	0.0510	0.0602	0.0608
(280, 140, 250)	0.0571	0.0483	0.0592	0.0584
(320, 160, 270)	0.0521	0.0479	0.0603	0.0549
(360, 180, 290)	0.0529	0.0489	0.0574	0.0569
(400, 190, 300)	0.0542	0.0522	0.0589	0.0579
(440, 220, 330)	0.0557	0.0529	0.0542	0.0581
(480, 240, 350)	0.0531	0.0562	0.0579	0.0569

The parameter t in the statistic T_n takes a value of 40. For DGP (a), we use the original statistic T_n in Theorem 2.3; for DGP (b), the statistic in Theorem 3.1 is used; for DGPs (c) and (d), the dividing-sample statistic in Theorem 3.2 is utilized.

malized statistic MLR_n are included as comparison with S_n. Here the renormalized 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 DGPs (a)–(d) are listed in Table 2. For DGP (a), we use the original statistic T_n; for DGP (b), the statistic in Theorem 3.1 is used; for DGPs (c) and (d), the dividing-sample statistic in Theorem 3.2 is utilized. For Theorem 3.2, we follow the implementation details in Appendix of [8] to estimate the LSD H for the matrix $\mathbf{\Sigma_{xx}}$.

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

6.3. *Factor model dependence.* We consider the factor model as follows:

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

where $\Lambda_1 = \frac{1}{\sqrt{p_1}}(\boldsymbol{\lambda}_1^{(1)}, \boldsymbol{\lambda}_2^{(1)}, \dots, \boldsymbol{\lambda}_r^{(1)})$ and $\Lambda_2 = \frac{1}{\sqrt{p_2}}(\boldsymbol{\lambda}_1^{(2)}, \boldsymbol{\lambda}_2^{(2)}, \dots, \boldsymbol{\lambda}_r^{(2)})$ are $p_1 \times r$ and $p_2 \times r$ deterministic matrices, respectively. In the simulation, all the components of $\boldsymbol{\lambda}_k^{(j)} : k = 1, 2, \dots, r;\ j = 1, 2$ are generated from a normal distribution with mean being 0.8 and variance being 1. $\mathbf{f}_t, t = 1, 2, \dots, n$ are $r \times 1$ random vectors with i.i.d. standard Gaussian distributed elements and $\mathbf{u}_t$ and $\mathbf{v}_t, t = 1, 2, \dots, n$ 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. Tables 3 and 4 illustrate

TABLE 3
Empirical powers of the proposed test S_n at 0.05 significance level for factor models

(p_1, p_2, n)	$r = 3$	$r = 5$	$r = 7$	$r = 10$
(10, 20, 40)	0.3750	0.5200	0.5910	0.9320
(30, 60, 120)	0.3070	0.6240	0.8450	0.9330
(50, 100, 200)	0.3090	0.6700	0.7980	0.9700
(70, 140, 280)	0.3520	0.6470	0.8330	0.9850
(90, 180, 360)	0.3670	0.6720	0.8230	0.9880
(110, 220, 440)	0.3570	0.6690	0.8490	0.9850
(130, 260, 520)	0.3440	0.6390	0.8510	0.9960
(150, 300, 600)	0.3780	0.6440	0.8370	0.9990
(170, 340, 680)	0.3580	0.6580	0.8590	1.0000
(190, 380, 760)	0.3490	0.6620	0.8720	1.0000
(210, 420, 840)	0.3460	0.6790	0.8890	1.0000
(230, 460, 920)	0.3800	0.6930	0.8770	1.0000
(250, 500, 1000)	0.3470	0.6890	0.8940	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.

the powers of the proposed statistics S_n and T_n, respectively, as r increases from 3 to 10. For T_n, we use its modified version in Theorem 3.2. 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

TABLE 4
Empirical powers of the proposed test T_n at 0.05 significance level for factor models

(p_1, p_2, n)	$r = 3$	$r = 5$	$r = 7$	$r = 10$
(100, 50, 80)	0.3680	0.6380	0.7330	0.9470
(140, 70, 120)	0.3380	0.6440	0.8690	0.9520
(180, 90, 150)	0.3290	0.6190	0.8890	0.9740
(200, 100, 170)	0.3410	0.6270	0.8920	0.9820
(240, 120, 180)	0.3340	0.6290	0.8840	0.9790
(280, 140, 250)	0.3570	0.6480	0.8730	0.9870
(320, 160, 270)	0.3490	0.7120	0.8890	0.9940
(360, 180, 290)	0.3690	0.6890	0.8930	0.9920
(400, 200, 310)	0.3830	0.7080	0.9030	0.9980
(440, 220, 330)	0.3920	0.7040	0.8930	1.0000
(480, 240, 350)	0.3970	0.6990	0.9110	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. The parameter t in the statistic T_n takes value of 40. For T_n, we use its modified dividing-sample version in Theorem 3.2.

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 (1.6) are used.

6.4. *Uncorrelated but dependent*. The construction of (2.8) is based on the idea that the limit of $F^{\mathbf{S_{xy}}}(x)$ 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 $\mathbf{x}_t = (X_{1t}, X_{2t}, \dots, X_{p_1 t})^T, t = 1, 2, \dots, n$ be i.i.d. normally distributed random vectors with zero means and unit variances. Define $\mathbf{y}_t = (Y_{1t}, Y_{2t}, \dots, Y_{p_2 t})^T$, $t = 1, 2, \dots, n$ by $Y_{it} = (X_{it}^{2k} - EX_{it}^{2k}), i = 1, 2, \dots, \min(p_1, p_2)$ and if $p_1 < p_2$, we let $Y_{jt} = \varepsilon_{jt}, j = p_1 + 1, \dots, p_2; t = 1, \dots, n$, where $\varepsilon_{jt}, j = p_1 + 1, \dots, p_2; t = 1, \dots, n$ are i.i.d. normal distributed random variables and independent with $\mathbf{x}_t$ and k is an positive integer.

REMARK 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 $\operatorname{cov}(X_{it}, Y_{it}) = EX_{it}^{2k+1} - EX_{it}EX_{it}^{2k} = 0$. Simulation results in Tables 5 and 6 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 2.3.

6.5. *ARCH type dependence*. The statistic works in the above example because the limit of $F^{\mathbf{S_{xy}}}$ cannot be determined from (2.2) if $\mathbf{x}$ and $\mathbf{y}$ are uncorrelated. However, the limit of $F^{\mathbf{S_{xy}}}(x)$ 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 $\mathbf{x}_t = (X_{1t}, X_{2t}, \dots, X_{p_1 t})$ and $\mathbf{y}_t = (Y_{1t}, Y_{2t}, \dots, Y_{p_2 t})$ as follows:

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

$$\text{if } p_1 < p_2, \qquad Y_{jt} = Z_{jt}, \qquad j = p_1 + 1, \dots, p_2, \tag{6.6}$$

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

TABLE 5
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 $Y_{it} = X_{it}^{\omega} - EX_{it}^{\omega}, i = 1, 2, \dots, p_1$ and $Y_{jt} = \varepsilon_{jt}, j = p_1 + 1, \dots, p_2$; $t = 1, \dots, n$, where $\varepsilon_{jt}, j = p_1 + 1, \dots, p_2$; $t = 1, \dots, n$ are standard normal distributed and independent with X_{it} and $\omega = 4, 10$.

TABLE 6
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)	0.7240	0.8690
(140, 70, 120)	0.7940	0.8890
(180, 90, 150)	0.7830	0.8940
(200, 100, 170)	0.7910	0.9340
(240, 120, 180)	0.8420	0.9290
(280, 140, 250)	0.8680	0.9580
(320, 160, 270)	0.9010	0.9820
(360, 180, 290)	0.9190	0.9940
(400, 200, 310)	0.9530	0.9990
(440, 220, 330)	0.9820	1.0000
(480, 240, 350)	0.9940	1.0000

The powers are under the alternative hypothesis that $Y_{it} = X_{it}^{\omega} - EX_{it}^{\omega}, i = 1, 2, \dots, p_1$ and $Y_{jt} = \varepsilon_{jt}, j = p_1 + 1, \dots, p_2$; $t = 1, \dots, n$, where $\varepsilon_{jt}, j = p_1 + 1, \dots, p_2$; $t = 1, \dots, n$ are 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 2.3 is used.

TABLE 7
Empirical powers of the proposed test S_n at 0.05 significance level for $\mathbf{x}$ and $\mathbf{y}$ with ARCH(1) dependent type

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

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

Normal (0, 1) and $\{\mathbf{x}_t : t = 1, \dots, n\}$ are independent across t. Moreover, $\{\mathbf{z}_t : t = 1, \dots, n\}$ are independent of $\{\mathbf{x}_t : t = 1, \dots, n\}$.

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

Tables 7 and 8 list the powers of the proposed statistics S_n and T_n for testing model (6.5) in several cases, that is, 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 2.3. 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.

TABLE 8
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)	0.6020	0.6180	0.7270	0.8930	0.9660
(140, 70, 120)	0.6370	0.7890	0.8020	0.8990	0.9820
(180, 90, 150)	0.7490	0.8280	0.9090	0.9920	1.0000
(200, 100, 170)	0.8130	0.8730	0.9930	1.0000	1.0000
(240, 120, 180)	0.8920	0.9720	0.9950	1.0000	1.0000
(280, 140, 250)	0.9470	0.9870	1.0000	1.0000	1.0000
(320, 160, 270)	0.9900	0.9980	1.0000	1.0000	1.0000
(360, 180, 290)	0.9910	0.9940	1.0000	1.0000	1.0000
(400, 200, 310)	0.9890	0.9950	1.0000	1.0000	1.0000
(440, 220, 330)	0.9920	1.0000	1.0000	1.0000	1.0000
(480, 240, 350)	0.9980	0.9970	1.0000	1.0000	1.0000

The powers are under the alternative hypothesis that $Y_{it} = Z_{it}\sqrt{\alpha_0 + \alpha_1 X_{it}^2}, i = 1, 2, \dots, p_1; Y_{jt} = Z_{jt}, j = p_1 + 1, \dots, p_2$. 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 2.3 is used.

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 $\mathbf{x}_t = (x_{1t}, x_{2t}, \dots, x_{p_1 t})^T$ and $\mathbf{y}_t = (y_{1t}, y_{2t}, \dots, y_{p_2 t})^T$, respectively. We consider daily stock returns $\mathbf{r}_t^{\mathbf{x}} = (r_{1t}^{\mathbf{x}}, r_{2t}^{\mathbf{x}}, \dots, r_{p_1 t}^{\mathbf{x}})$ and $\mathbf{r}_t^{\mathbf{y}} = (r_{1t}^{\mathbf{y}}, r_{2t}^{\mathbf{y}}, \dots, r_{p_2 t}^{\mathbf{y}})$ with $r_{it}^{\mathbf{x}} = \log \frac{x_{it}}{x_{i,t-1}}$, $i = 1, 2, \dots, p_1$ and $r_{jt}^{\mathbf{y}} = \log \frac{y_{jt}}{y_{j,t-1}}$, $j = 1, 2, \dots, p_2$. 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 $\mathbf{X} = (\mathbf{r}_1^{\mathbf{x}}, \mathbf{r}_2^{\mathbf{x}}, \dots, \mathbf{r}_{p_1}^{\mathbf{x}})$ and $\mathbf{Y} = (\mathbf{r}_1^{\mathbf{y}}, \mathbf{r}_2^{\mathbf{y}}, \dots, \mathbf{r}_{p_2}^{\mathbf{y}})$, 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, that is, basic industry section and capital goods section, public utility section and capital goods section, finance section and healthcare section. From Tables 9, 10 and 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.

TABLE 9
P-values for (p_1, p_2) companies from basic industry section and capital goods section of NYSE

P-values: P-value interval	No. of exp. (p_1, p_2, n) (10, 15, 20)	No. of exp. (p_1, p_2, n) (15, 20, 25)
[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 2000.1.1–2002.1.1. 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-value interval	No. of exp. (p_1, p_2, n) (10, 15, 20)	No. of exp. (p_1, p_2, n) (15, 20, 25)
[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: public utility section and capital goods section, each of which has n daily stock returns during the period 2000.1.1–2002.1.1. 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

	No. of exp.	
P-values: P-value interval	**(p_1, p_2, n) (10, 15, 20)**	**(p_1, p_2, n) (15, 20, 25)**
$[0, 0.05]$	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: finance section and healthcare section, each of which has n daily stock returns during the period 2000.1.1–2002.1.1. 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.

8. Acknowledgement. The authors would like to thank the Editor, an Associate Editor and the referees for their constructive comments and suggestions which significantly improved this paper.

SUPPLEMENTARY MATERIAL

Supplement to "Independence test for high dimensional data based on regularized canonical correlation coefficients" (DOI: 10.1214/14-AOS1284SUPP; .pdf). The supplementary material is divided into Appendices A and B. Some useful lemmas, and proofs of all theorems and Proposition 4–5 are given in Appendix A while one theorem related to CLT of a sample covariance matrix plus a perturbation matrix is provided in Appendix B.

REFERENCES

[1] ANDERSON, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*, 2nd ed. Wiley, New York. MR0771294

[2] BAI, J. and NG, S. (2002). Determining the number of factors in approximate factor models. *Econometrica* **70** 191–221. MR1926259

[3] BAI, Z., CHEN, J. and YAO, J. (2010). On estimation of the population spectral distribution from a high-dimensional sample covariance matrix. *Aust. N. Z. J. Stat.* **52** 423–437. MR2791528

[4] BAI, Z. and SARANADASA, H. (1996). Effect of high dimension: By an example of a two sample problem. *Statist. Sinica* **6** 311–329. MR1399305
[5] BAI, Z. and SILVERSTEIN, J. W. (2010). *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. Springer, New York. MR2567175
[6] BICKEL, P. J. and LEVINA, E. (2008). Regularized estimation of large covariance matrices. *Ann. Statist.* **36** 199–227. MR2387969
[7] BIRKE, M. and DETTE, H. (2005). A note on testing the covariance matrix for large dimension. *Statist. Probab. Lett.* **74** 281–289. MR2189467
[8] EL KAROUI, N. (2008). Spectrum estimation for large dimensional covariance matrices using random matrix theory. *Ann. Statist.* **36** 2757–2790. MR2485012
[9] FAN, J. and FAN, Y. (2008). High-dimensional classification using features annealed independence rules. *Ann. Statist.* **36** 2605–2637. MR2485009
[10] FAN, J., GUO, S. and HAO, N. (2012). Variance estimation using refitted cross-validation in ultrahigh dimensional regression. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **74** 37–65. MR2885839
[11] FUJIKOSHI, Y., ULYANOV, V. V. and SHIMIZU, R. (2010). *Multivariate Statistics: High-Dimensional and Large-Sample Approximations*. Wiley, Hoboken, NJ. MR2640807
[12] HUANG, J., HOROWITZ, J. L. and MA, S. (2008). Asymptotic properties of bridge estimators in sparse high-dimensional regression models. *Ann. Statist.* **36** 587–613. MR2396808
[13] JOHNSTONE, I. M. (2008). Multivariate analysis and Jacobi ensembles: Largest eigenvalue, Tracy–Widom limits and rates of convergence. *Ann. Statist.* **36** 2638–2716. MR2485010
[14] LYTOVA, A. and PASTUR, L. (2009). Central limit theorem for linear eigenvalue statistics of random matrices with independent entries. *Ann. Probab.* **37** 1778–1840. MR2561434
[15] MARDIA, K. V., KENT, J. T. and BIBBY, J. M. (1979). *Multivariate Analysis*. Academic Press, San Diego. MR0560319
[16] PAN, G. (2010). Strong convergence of the empirical distribution of eigenvalues of sample covariance matrices with a perturbation matrix. *J. Multivariate Anal.* **101** 1330–1338. MR2609495
[17] SILVERSTEIN, J. W. and BAI, Z. D. (1995). On the empirical distribution of eigenvalues of a class of large-dimensional random matrices. *J. Multivariate Anal.* **54** 175–192. MR1345534
[18] TIMM, N. H. (2002). *Applied Multivariate Analysis*. Springer, New York. MR1908225
[19] WACHTER, K. W. (1980). The limiting empirical measure of multiple discriminant ratios. *Ann. Statist.* **8** 937–957. MR0585695
[20] WILKS, S. S. (1935). On the independence of k sets of normally distributed statistical variables. *Econometrica* **3** 309–326.
[21] YANG, Y. and PAN, G. (2012). The convergence of the empirical distribution of canonical correlation coefficients. *Electron. J. Probab.* **17** no. 64, 13. MR2968671
[22] ZHENG, S. (2012). Central limit theorems for linear spectral statistics of large dimensional F-matrices. *Ann. Inst. Henri Poincaré Probab. Stat.* **48** 444–476. MR2954263
[23] YANG, Y. and PAN, G. (2014). Supplement to "Independence test for high dimensional data based on regularized canonical correlation coefficients." DOI:10.1214/14-AOS1284SUPP.

DEPARTMENT OF ECONOMETRICS
AND BUSINESS STATISTICS
MONASH UNIVERSITY
CAULFIELD EAST, VICTORIA 3145
AUSTRALIA
E-MAIL: yanrong.yang@monash.edu

SCHOOL OF PHYSICAL
AND MATHEMATICAL SCIENCES
NANYANG TECHNOLOGICAL UNIVERSITY
SINGAPORE
E-MAIL: GMPAN@ntu.edu.sg