



On the test of covariance between two high-dimensional random vectors

Yongshuai Chen^{1,2} · Wenwen Guo² · Hengjian Cui²

Received: 4 March 2023 / Revised: 27 July 2023 / Published online: 7 October 2023

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract

We consider a problem of association test in high dimension. A new test statistic is proposed based on the covariance of random vectors and its asymptotic properties are derived under both the null hypothesis and the local alternatives. Furthermore power enhancement technique is utilized to boost the empirical power especially under sparse alternatives. We examine the finite-sample performances of the proposed test via Monte Carlo simulations, which show that the proposed test outperforms some existing procedures. An empirical analysis of a microarray data is demonstrated to detect the relationship between the genes.

Keywords Association test · High dimension · Covariance of random vectors · Power enhancement technique

1 Introduction

Traditional statistical methods for measuring the association between random vectors are generally based on coefficient (covariance). See Wilks (1935), Anderson (2003), Robert et al. (1985), Székely et al. (2007) and among others. Wilks (1935) introduced an effective likelihood ratio test (LRT) for block independence under Gaussian population and Anderson (2003) detailed LRT for the Gaussian population. RV correlation coefficient proposed by Escoufier (1973) was considered in Robert et al. (1985) to measure multivariate association between two sets of variables. Székely et al. (2007)

✉ Hengjian Cui
hjcui@bnu.edu.cn

Yongshuai Chen
yongshuaichen@cueb.edu.cn

Wenwen Guo
wwguo@cnu.edu.cn

¹ School of Statistics, Capital University of Economics and Business, Beijing, China

² School of Mathematical Sciences, Capital Normal University, Beijing, China

developed distance covariance and distance correlation, and provided an approach to the problem of testing the joint independence of random vectors. The asymptotic properties of these classical procedures aforementioned are established under the scheme that the sample size n tends to infinity and the dimensions (p and q) are fixed. This is the so-called “small p and q , large n ” paradigm.

However, high-dimensional data, such as microarray analysis, tumor classification and biomedical imaging, tend to have a dimension (p and (or) q) comparable to, or much larger than, the sample size n . This brings great challenges to these traditional methods. For example, the empirical power of the conventional test may be largely impacted by the increasing dimension, and even converges to the significance level α due to “the curse of dimensionality”, which means that the test cannot distinguish the null hypothesis from the alternatives. Therefore more and more statisticians are pursuing new methods to address the high-dimensional problems. To accommodate the large-dimensionality, Jiang et al. (2013) proposed the corrected likelihood ratio test and large-dimensional trace criterion to test the independence of two large sets of multivariate variables when the dimensions $p + q$ and the sample size n tend to infinity simultaneously. To make the RV coefficient be applicable for high-dimensional data, some test procedures are introduced in the following two papers. Srivastava and Reid (2012) proposed a new statistic based on the RV coefficient for testing the independence of two sub-vectors, and obtained its asymptotic properties under the scheme that $\min(p, q, n) \rightarrow \infty$, $p/(p + q) \rightarrow d_1 > 0$, $q/(p + q) \rightarrow d_2 > 0$ and $n = O((p + q)^\delta)$ for some constant $\delta > 0$. By constructing an unbiased estimator for the numerator of the RV coefficient, Li et al. (2017) considered the independence test under the assumption that only one random vector has a divergent dimension, that is, $\min(p, n) \rightarrow \infty$ and q is fixed. The asymptotic properties in these three papers (Jiang et al. 2013; Srivastava and Reid 2012; Li et al. 2017) are established under the assumption that the random vectors are multivariate normal distributed. Without normal constraint, Yang and Pan (2015) proposed a test statistic based on regularized canonical correlation coefficients, and they obtained the limiting distributions when both p and q are comparable to the sample size n . Discovering that the empirical distance correlation of the two vectors converges to one even though they are independent as dimensions tend to infinity, Székely and Rizzo (2013) extended the distance correlation with a modified version in high-dimensional settings, and obtained a distance correlation t-test for independence of random vectors in arbitrarily high dimension. Heller et al. (2012) presented a powerful test of association based on ranks of distances which is consistent against all alternatives and can be applied in any dimensions p and q even greater than n . On the basis of power enhancement technique introduced by Fan et al. (2015), Zheng et al. (2022) developed a powerful test on block-structured correlation of a high dimensional-random vector for sparse or non-sparse alternatives without normality assumption, and obtained the statistical properties under the asymptotic regime that $(p + q)/n \rightarrow y \in (0, \infty)$.

This paper aims to develop a new and powerful test on high-dimensional association under no strict distributional assumptions. To this end, we propose a U-statistic of order four based on the RV covariance introduced in Escoufier (1973). To boost the power especially under sparse alternatives a screening term is added. It is worth mentioning that the proposed test statistic is effective not only for non-sparse alternatives but also

for sparse alternatives. Four distinguishing features can be summarized as follows. First, although the proposed U-statistic is order four, it can be fast implemented with an estimated U-statistic of order two. Second, the asymptotic distributions of the U-statistic under both null hypothesis and local alternatives are derived in the scheme that $p + q$ tends to infinity. It is noteworthy that this scheme can be divided into two cases: one is that only p or q is much more than or comparable to the sample size n ; the other is that both p and q tend to infinity with n simultaneously. Third, the power enhancement technique can dramatically improve the performance of the proposed test, which is demonstrated by Monte Carlo simulations. Last but not least is some examples under specific structures are given to gain more insights on the regularity conditions.

The rest of this work is organized as follows. A new statistic for testing the association between two random vectors is proposed on the basis of RV covariance, and its asymptotic properties are established in Sect. 2. In Sect. 3 we investigate the properties of the screening term based on the power enhancement technique. Numerical studies and real data analysis are listed in Sect. 4 to examine the size and power of our proposed test. The Appendix is devoted to gather the technical proofs.

2 Association test in high dimension

In this part, our interest is to study and test the association between two random vectors $\mathbf{X} = (X_1, \dots, X_p)^\top \in \mathbb{R}^p$ and $\mathbf{Y} = (Y_1, \dots, Y_q)^\top \in \mathbb{R}^q$. Let $\Sigma_{\mathbf{XY}}$ denote the population covariance matrix of \mathbf{X} and \mathbf{Y} . Our association test hypothesis can be represented as follows:

$$H_0 : \Sigma_{\mathbf{XY}} = 0_{p \times q} \quad \text{VS} \quad H_1 : \Sigma_{\mathbf{XY}} \neq 0_{p \times q}.$$

It is worth noting that $\Sigma_{\mathbf{XY}} = 0_{p \times q}$ if and only if the two random vectors are uncorrelated. Escoufier (1973) defined $\text{tr}(\Sigma_{\mathbf{XY}}\Sigma_{\mathbf{YX}})$ as the ‘‘covariance’’ of two random vectors \mathbf{X} and \mathbf{Y} , where $\text{tr}(\cdot)$ denotes the trace operator. It is evident to see that $\text{tr}(\Sigma_{\mathbf{XY}}\Sigma_{\mathbf{YX}}) = 0$ is equivalent to $\Sigma_{\mathbf{XY}} = 0_{p \times q}$. This motivates us to utilize $\text{tr}(\Sigma_{\mathbf{XY}}\Sigma_{\mathbf{YX}})$ to quantify the discrepancy between $\Sigma_{\mathbf{XY}}$ and $0_{p \times q}$. Adopting the idea of \mathcal{U} -centring in Székely and Rizzo (2014) and Yao et al. (2018), we can construct an unbiased estimator, denoted by $T_{n,p,q}(\mathbf{X}, \mathbf{Y})$, of $\text{tr}(\Sigma_{\mathbf{XY}}\Sigma_{\mathbf{YX}})$. Suppose that $\mathbf{Z}_i = (\mathbf{X}_i, \mathbf{Y}_i)^\top \in \mathbb{R}^{p+q}$ are random samples of $\mathbf{Z} = (\mathbf{X}, \mathbf{Y})^\top$ in which $\mathbf{X}_i = (X_{i1}, \dots, X_{ip})^\top$ and $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{iq})^\top$. Then $T_{n,p,q}(\mathbf{X}, \mathbf{Y})$ is given by

$$T_{n,p,q}(\mathbf{X}, \mathbf{Y}) = \binom{n}{4}^{-1} \sum_{i < j < k < l} h(\mathbf{Z}_i, \mathbf{Z}_j, \mathbf{Z}_k, \mathbf{Z}_l),$$

where

$$h(\mathbf{Z}_1, \mathbf{Z}_2, \mathbf{Z}_3, \mathbf{Z}_4) = \frac{1}{4!} \sum_{(s,t,u,v)}^{(1,2,3,4)} \frac{1}{4} (\mathbf{X}_s - \mathbf{X}_t)^\top (\mathbf{X}_u - \mathbf{X}_v) (\mathbf{Y}_s - \mathbf{Y}_t)^\top (\mathbf{Y}_u - \mathbf{Y}_v)$$

and the summation is over all 24 permutations of the 4-tuples of indices (1, 2, 3, 4).

It is obvious to see that $T_{n,p,q}(\mathbf{X}, \mathbf{Y})$ is a U-statistic of order four and it is an unbiased estimator of $\text{tr}(\boldsymbol{\Sigma}_{\mathbf{XY}}\boldsymbol{\Sigma}_{\mathbf{YX}})$.

Remark 1 Tests for high-dimensional regression coefficients in Zhong and Chen (2011) and Cui et al. (2018) can be seen as a special case of our test hypothesis with respect to measuring the discrepancy. Consider a linear regression model

$$Y = \alpha + \mathbf{X}^\top \boldsymbol{\beta} + \varepsilon$$

where $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top \in \mathbb{R}^p$ is a p -dimensional vector of regression coefficients of interest and α is a nuisance intercept parameter. ε is random error with mean zero and variance σ^2 , and is independent of \mathbf{X} . Testing the high-dimensional regression coefficients simultaneously in the linear model can be formulated as follows:

$$H_0 : \boldsymbol{\beta} = 0_{p \times 1} \quad \text{VS} \quad H_1 : \boldsymbol{\beta} \neq 0_{p \times 1}.$$

It tests the overall significance of linear regression coefficients. Both (Zhong and Chen 2011) and (Cui et al. 2018) adopted $\boldsymbol{\beta}^\top \boldsymbol{\Sigma}_{\mathbf{X}}^{-2} \boldsymbol{\beta}$ as an effective measure of the difference between $\boldsymbol{\beta}$ and $0_{p \times 1}$. Simple calculation shows that $\text{tr}(\boldsymbol{\Sigma}_{\mathbf{XY}}\boldsymbol{\Sigma}_{\mathbf{YX}}) = \boldsymbol{\beta}^\top \boldsymbol{\Sigma}_{\mathbf{X}}^2 \boldsymbol{\beta}$.

Remark 2 $\text{tr}(\boldsymbol{\Sigma}_{\mathbf{XY}}\boldsymbol{\Sigma}_{\mathbf{YX}})$, the squared distance covariance $\text{dCov}^2(\mathbf{X}, \mathbf{Y})$ introduced by Székely et al. (2007) and the squared martingale difference divergence $\text{MDD}(Y|\mathbf{X})^2$ proposed in Shao and Zhang (2014) can be constructed in an analogous way. In fact, write

$$\begin{aligned} \tau(\mathbf{X}, \mathbf{Y}) &= \text{E}(\|\mathbf{X} - \mathbf{X}'\|^\alpha \|\mathbf{Y} - \mathbf{Y}'\|^\beta) + \text{E}(\|\mathbf{X} - \mathbf{X}''\|^\alpha) \text{E}(\|\mathbf{Y} - \mathbf{Y}''\|^\beta) \\ &\quad - 2\text{E}(\|\mathbf{X} - \mathbf{X}'\|^\alpha \|\mathbf{Y} - \mathbf{Y}''\|^\beta), \end{aligned}$$

where $\mathbf{Z}' = (\mathbf{X}', \mathbf{Y}')^\top$ and $\mathbf{Z}'' = (\mathbf{X}'', \mathbf{Y}'')^\top$ are independent copies of $\mathbf{Z} = (\mathbf{X}, \mathbf{Y})^\top$. When $\alpha = \beta = 1$, $\tau(\mathbf{X}, \mathbf{Y}) = \text{dCov}^2(\mathbf{X}, \mathbf{Y})$ and it characterizes independence of random vectors \mathbf{X} and \mathbf{Y} . $\tau(\mathbf{X}, \mathbf{Y}) = 2\text{MDD}(Y|\mathbf{X})^2$ as $\alpha = 1$ and $\beta = 2$, and it measures the departure of conditional mean independence between a scalar response variable Y and a vector predictor variable \mathbf{X} . When $\alpha = \beta = 2$, $\tau(\mathbf{X}, \mathbf{Y}) = 4\text{tr}(\boldsymbol{\Sigma}_{\mathbf{XY}}\boldsymbol{\Sigma}_{\mathbf{YX}})$. Furthermore, it is worth mentioning that $\text{tr}(\boldsymbol{\Sigma}_{\mathbf{XY}}\boldsymbol{\Sigma}_{\mathbf{YX}})$ can degenerate into the squared covariance of random variable X and Y , when $p = q = 1$. This suggests us consider not only the simultaneous measure between random vectors \mathbf{X} and \mathbf{Y} , but also the marginal quantization between their components X_i and $Y_j, i = 1, \dots, p, j = 1, \dots, q$ due to the curse of dimensionality.

Remark 3 According to the Székely and Rizzo (2014) and Yao et al. (2018), $T_{n,p,q}(\mathbf{X}, \mathbf{Y})$ can be fast implemented. Define the respective \mathcal{U} -centred versions of $a_{ij} = \mathbf{X}_i^\top \mathbf{X}_j$ and $b_{ij} = \mathbf{Y}_i^\top \mathbf{Y}_j$ as follows:

$$A_{ij} = a_{ij} - \frac{1}{n-2} \sum_{l \neq i} a_{il} - \frac{1}{n-2} \sum_{k \neq j} a_{kj} + \frac{1}{(n-1)(n-2)} \sum_{k \neq l} a_{kl},$$

$$B_{ij} = b_{ij} - \frac{1}{n-2} \sum_{l \neq i} b_{il} - \frac{1}{n-2} \sum_{k \neq j} b_{kj} + \frac{1}{(n-1)(n-2)} \sum_{k \neq l} b_{kl}.$$

Then $T_{n,p,q}(\mathbf{X}, \mathbf{Y})$ has the following reformulation:

$$T_{n,p,q}(\mathbf{X}, \mathbf{Y}) = \frac{1}{n(n-3)} \sum_{i \neq j} A_{ij} B_{ij}$$

which is shown to be a quick implementation in numerical simulations.

2.1 Asymptotic analysis of $T_{n,p,q}(\mathbf{X}, \mathbf{Y})$ under null hypothesis

In this subsection the asymptotic properties of $T_{n,p,q}(\mathbf{X}, \mathbf{Y})$ are investigated under some regularity assumptions. Some notations are introduced before studying the properties of the proposed statistic. Write $L_1(\mathbf{Z}, \mathbf{Z}') = E\{L(\mathbf{Z}, \mathbf{Z}')L(\mathbf{Z}', \mathbf{Z}') | (\mathbf{Z}, \mathbf{Z}')\}$, and $\zeta^2 = E\{L(\mathbf{Z}, \mathbf{Z}')^2\}$, where $L(\mathbf{Z}, \mathbf{Z}') = \mathbf{X}^\top \mathbf{X}' \mathbf{Y}^\top \mathbf{Y}'$. To obtain the asymptotic distribution of $T_{n,p,q}(\mathbf{X}, \mathbf{Y})$ we require the following technical assumption.

(A1) $E\{L_1(\mathbf{Z}, \mathbf{Z}')^2\} = o(\zeta^4)$, $E\{(\mathbf{X}^\top \mathbf{X}' \mathbf{Y}^\top \mathbf{Y}')^4\} = o(n\zeta^4)$, $E\{(\mathbf{X}^\top \mathbf{X}' \mathbf{Y}^\top \mathbf{Y}'')^4\} = o(n\zeta^4)$ and $E\{(\mathbf{X}^\top \mathbf{X}')^4\}E\{(\mathbf{Y}^\top \mathbf{Y}')^4\} = o(n\zeta^4)$.

The following theorem presents the limiting null distribution of $T_{n,p,q}(\mathbf{X}, \mathbf{Y})$.

Theorem 1 *Suppose assumption (A1) holds. Then under H_0 , as $(n, p + q) \rightarrow \infty$ we have that*

$$\frac{nT_{n,p,q}(\mathbf{X}, \mathbf{Y})}{\sqrt{2\zeta^2}} \xrightarrow{d} \mathcal{N}(0, 1),$$

and a ratio consistent estimator of ζ^2 is

$$\zeta_n^2 = \frac{1}{n(n-3)} \sum_{i \neq j} A_{ij}^2 B_{ij}^2,$$

where \xrightarrow{d} denotes convergence in distribution.

By referring to Zhang et al. (2018) and its supplementary material, we impose assumption (A1) to ensure the asymptotic normality of the degenerate U-statistic $T_{n,p,q}(\mathbf{X}, \mathbf{Y})$ under the null hypothesis. Meanwhile this assumption guarantees that ζ_n^2 is ratio-consistent. To further understand this assumption, the following proposition is established when $\mathbf{Z} = (\mathbf{X}, \mathbf{Y})^\top$ follows a multivariate normal distribution.

Proposition 1 *Suppose that $\mathbf{Z} = (\mathbf{X}, \mathbf{Y})^\top \sim \mathcal{N}(0_{(p+q) \times 1}, \mathbf{I}_{p+q})$. Then we have*

$$\begin{aligned} \zeta^2 &= pq, \quad E\{L_1(\mathbf{Z}, \mathbf{Z}')^2\} = pq, \\ E\{(\mathbf{X}^\top \mathbf{X}' \mathbf{Y}^\top \mathbf{Y}')^4\} &= 9pq(p+2)(q+2), \end{aligned}$$

$$E\{(\mathbf{X}^\top \mathbf{X}' \mathbf{Y}^\top \mathbf{Y}'')^4\} = 9pq(p+2)(q+2),$$

$$E\{(\mathbf{X}^\top \mathbf{X}')^4\}E\{(\mathbf{Y}^\top \mathbf{Y}')^4\} = 9pq(p+2)(q+2).$$

Proposition 1 may shed light on the assumption (A1). It indicates that when \mathbf{Z} follows a standard multivariate normal distribution, this condition holds automatically as long as n and $p + q$ tend to infinity, which implies that condition (A1) is imposed reasonably.

Remark 4 Theorem 1 states that the asymptotic null distribution of $T_{n,p,q}$ is normal, and we can reject H_0 at a significance level α if

$$nT_{n,p,q}(\mathbf{X}, \mathbf{Y}) \geq z_\alpha \sqrt{2\zeta_n^2},$$

where z_α denotes the upper α quantile of $\mathcal{N}(0, 1)$.

2.2 Asymptotic analysis of $T_{n,p,q}(\mathbf{X}, \mathbf{Y})$ under the local alternatives

In this subsection, we turn to the asymptotic analysis of $T_{n,p,q}(\mathbf{X}, \mathbf{Y})$ under the local alternatives.

The following assumption is required for theoretical study.

(A2) $E\{(\mathbf{X}^\top \Sigma_{\mathbf{X}\mathbf{Y}} \mathbf{Y})^2\} = o(n^{-1}\zeta^2)$ and $E\{(\mathbf{X}^\top \Sigma_{\mathbf{X}\mathbf{Y}} \mathbf{Y}')^2\} = o(\zeta^2)$.

Theorem 2 Suppose that assumptions (A1) and (A2) hold. Then as $(n, p + q) \rightarrow \infty$ we have

$$\frac{n\{T_{n,p,q}(\mathbf{X}, \mathbf{Y}) - \text{tr}(\Sigma_{\mathbf{X}\mathbf{Y}} \Sigma_{\mathbf{Y}\mathbf{X}})\}}{\sqrt{2\zeta^2}} \xrightarrow{d} \mathcal{N}(0, 1),$$

and ζ_n^2 is still a ratio-consistent estimator of ζ^2 .

Assumption (A2) characterizes the local alternative in the sense that the alternative is not too far away from the null hypothesis, and thus shows that our proposed statistic is a degenerate U-statistic. It is noteworthy that this assumption holds automatically under the null hypothesis and it is given in light of Zhang et al. (2018). In following we illustrate the assumptions (A1) and (A2) under linear regression model.

Proposition 2 Assume that $\mathbf{Y} = \mathbf{X}^\top \boldsymbol{\beta} + \varepsilon$ where \mathbf{X} and ε are independent with $\mathbf{X} \sim \mathcal{N}(0_{p \times 1}, \mathbf{I}_p)$ and $\varepsilon \sim \mathcal{N}(0, 1)$. Then we have

$$\begin{aligned} \zeta^2 &= (p + 8)\|\boldsymbol{\beta}\|^4 + 2(p + 2)\|\boldsymbol{\beta}\|^2 + p, \\ E\{(\mathbf{X}^\top \Sigma_{\mathbf{X}\mathbf{Y}} \mathbf{Y})^2\} &= 3\|\boldsymbol{\beta}\|^4 + \|\boldsymbol{\beta}\|^2, \\ E\{(\mathbf{X}^\top \Sigma_{\mathbf{X}\mathbf{Y}} \mathbf{Y}')^2\} &= \|\boldsymbol{\beta}\|^4 + \|\boldsymbol{\beta}\|^2, \\ E\{(\mathbf{X}^\top \mathbf{X}' \mathbf{Y}^\top \mathbf{Y}')^4\} &= O\{p^2(\|\boldsymbol{\beta}\|^2 + 1)^4\}, \\ E\{(\mathbf{X}^\top \mathbf{X}' \mathbf{Y}^\top \mathbf{Y}'')^4\} &= O\{p^2(\|\boldsymbol{\beta}\|^2 + 1)^4\}, \\ E\{(\mathbf{X}^\top \mathbf{X}')^4\}E\{(\mathbf{Y}^\top \mathbf{Y}')^4\} &= 27p(p + 2)(\|\boldsymbol{\beta}\|^2 + 1)^4 \end{aligned}$$

$$E\{L_1(\mathbf{Z}, \mathbf{Z}')^2\} = (3\|\boldsymbol{\beta}\|^2 + 1)^4 + (p - 1)(\|\boldsymbol{\beta}\|^2 + 1)^4.$$

Proposition 2 implies that assumption (A1) holds automatically as $(n, p) \rightarrow \infty$. Furthermore, assumption (A2) can be satisfied when $\|\boldsymbol{\beta}\|^2/(1 + \|\boldsymbol{\beta}\|^2) = o(p/n)$. Therefore, assumption (A2) can be viewed as local alternatives.

Remark 5 It can be inferred from theorem 2 that the asymptotic power of the proposed test under the local alternatives is

$$\Phi\left(-z_\alpha + \frac{n\text{tr}(\boldsymbol{\Sigma}_{\mathbf{X}\mathbf{Y}}\boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{X}})}{\sqrt{2\zeta^2}}\right),$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. It is clear to see that this power is dominantly affected by signal to noise ratio term $\eta(\boldsymbol{\Sigma}_{\mathbf{X}\mathbf{Y}}) = \text{tr}(\boldsymbol{\Sigma}_{\mathbf{X}\mathbf{Y}}\boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{X}})/\sqrt{2\zeta^2}$. Specially, the power tends to α when $\eta(\boldsymbol{\Sigma}_{\mathbf{X}\mathbf{Y}}) = o(n^{-1})$, which implies that the test fails to make a distinction between the null hypothesis and the local alternatives. Meanwhile, if $\eta(\boldsymbol{\Sigma}_{\mathbf{X}\mathbf{Y}})$ has a higher order of n^{-1} , the power converges to 1 and thus it is a consistent test.

3 Power enhancement technique

In Sect. 2, $T_{n,p,q}$ is constructed based on $\text{tr}(\boldsymbol{\Sigma}_{\mathbf{X}\mathbf{Y}}\boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{X}})$, and its asymptotic distributions are obtained under regularity conditions (A1)–(A2). It is worth noting that it is a simultaneous measure for association, and its power is adversely affected by the increasing dimensions, especially under the sparse alternatives. We adopt the power enhancement technique introduced by Fan et al. (2015), and utilize the marginal association to boost the empirical power in this case. Define

$$S = \{(i, j) : \rho^2(X_i, Y_j) \geq 2\delta_{n,p,q}, 1 \leq i \leq p, 1 \leq j \leq q\}$$

$$\mathcal{R}_n(X_i, Y_j) = \frac{T_{n,1,1}(X_i, Y_j)}{\sqrt{T_{n,1,1}(X_i, X_i)T_{n,1,1}(Y_j, Y_j)}}, 1 \leq i \leq p, 1 \leq j \leq q,$$

where $\delta_{n,p,q} = c \log(\log(n))/\log(\log(pq))(pq)^{1/4}(\log(n))^{3/4}/n$ and $\rho(X_i, Y_j)$ is the Pearson correlation coefficient.

The conditions below are imposed to derive the limiting properties of $\mathcal{R}_n(X_i, Y_j)$.
 (A3) $pq = O(n^\kappa)$, $0 < \kappa < 4$ and $E\{(\mathbf{X}^\top \mathbf{X}' \mathbf{Y}^\top \mathbf{Y}')^4\}$ exists.

(A4) $\max_{1 \leq i \leq p, 1 \leq j \leq q} \xi_{ij} = O((pq)^{-1/2})$, where $\xi_{ij} = \sigma^2(X_i, Y_j)\text{Var}\{(X_i - E(X_i))(Y_j - E(Y_j))\}$ and $\sigma(X_i, Y_j)$ is the covariance of random variables X_i and Y_j .

We present the following result regarding the asymptotic behavior of $\mathcal{R}_n(X_i, Y_j)$ under both the null hypothesis and the alternatives.

Theorem 3 *Suppose conditions (A3)–(A4) hold. Then we have that*

(1) *under H_0 , almost surely*

$$\max_{1 \leq i \leq p, 1 \leq j \leq q} |\mathcal{R}_n(X_i, Y_j)| = o(\delta_{n,p,q}), \quad n \rightarrow \infty.$$

(2) when $S \neq \emptyset$, almost surely

$$\max_{1 \leq i \leq p, 1 \leq j \leq q} |\mathcal{R}_n(X_i, Y_j)| \geq \delta_{n,p,q}, \quad n \rightarrow \infty.$$

Assumptions (A3)–(A4) are imposed based on the results of Section 5.3 in Serfling (1980). $pq = O(n^\kappa)$, $0 < \kappa < 4$ has an explicit relationship between p , q and n , and means $\delta_{n,p,q}$ converges to 0 as n tends to infinity. It is worth noting that assumption (A3) ensures an almost sure representation of the difference between U-statistic $T_{n,1,1}(X_i, Y_j)$ and its projection. As for condition (A4), it holds automatically under the null hypothesis and shows that ξ_{ij} is of order $(pq)^{-1/2}$ uniformly among pq components under sparse alternatives, which guarantees an almost sure representation for the projection of U-statistic $T_{n,1,1}(X_i, Y_j)$. Details can be seen in the proof of Theorem 3 in the Appendix.

Remark 6 In light of theorem 3, the screening term used for enhancing the empirical power can be constructed as follows:

$$T_{n,p,q}^0 = pq \mathbf{1} \left(\max_{1 \leq i \leq p, 1 \leq j \leq q} |\mathcal{R}_n(X_i, Y_j)| \geq \delta_{n,p,q} \right)$$

where $\mathbf{1}(\cdot)$ is an indicator function. It can be inferred from this theorem that $T_{n,p,q}^0$ is negligible under null hypothesis whereas it will diverge to infinity when $S \neq \emptyset$. Therefore the power of the test will be well enhanced if add the term $T_{n,p,q}^0$ owing to taking into account of more information from the alternative. Similar to the discussion in Fan et al. (2015), a general form of the test statistic can be proposed as follows:

$$\widehat{T}_{n,p,q} = T_{n,p,q} + T_{n,p,q}^0.$$

It is clear to see from Theorems 1 and 3 that the test statistic $\widehat{T}_{n,p,q}$ shares the same distribution as $T_{n,p,q}$, which implies that the proposed test reject H_0 at the significance level α if $\widehat{T}_{n,p,q} \geq z_\alpha$. Furthermore, it yields from Theorems 2 and 3 that the empirical power of $\widehat{T}_{n,p,q}$ can tend to 1 under some regularity conditions. This also indicates that the power enhancement technique can be utilized to boost the power as long as a test statistic has a limiting null distribution such as normal approximation.

Remark 7 Given that the kernel function of U-statistic $T_{n,1,1}(X_i, Y_j)$ has finite fourth moments, the convergence rate, $(\log(n))^{3/4}/n$, of difference between a U-statistic and its projection is obtained in Theorem 5.3.3 of Serfling (1980). $\max_{1 \leq i \leq p, 1 \leq j \leq q} |\mathcal{R}_n(X_i, Y_j)|$ refers to the maximum of pq marginal correlation, which indicates that the convergence rate should be multiplied by a function of pq . In light of the fact that the kernel function of $T_{n,1,1}(X_i, Y_j)$ has finite fourth moments, we choose $(pq)^{1/4}$. More details can be found in the proof of Theorems 3. Furthermore, it is worth noting that the choice of tuning parameter c will affect the performance of the proposed test in both empirical size and empirical power. To be specific, given the sample size and the dimension, larger c can lead to a higher probability of controlling the empirical size while smaller c has a greater chance of boosting the empirical

power, which indicates that the choice of c becomes a trade-off between controlling the empirical size and boosting the empirical power. In practical applications a series of values are given to identify which ones can meet the need of size control. Based on the selected values, we choose the smallest value as c , which can increase the empirical power. More details can be found from the Example 1 in the simulation studies.

4 Numerical studies

In this section we illustrate the proposed test procedure by investigating its finite sample performance through simulations and a real microarray gene data analysis. For the purpose of comparison, we also consider the following methods. The first one proposed in Zheng et al. (2022) adopts the Frobenius distance between covariance matrix. The second procedure is based on the trace of covariance matrix advocated in Li et al. (2017). The third one utilizes the modified distance correlation established in Székely and Rizzo (2013). The fourth one is built based on ranks of distances in Heller et al. (2012) and the remaining one is $T_{n,p,q}$ proposed in this paper. To be specific, we denote the method used in this section as follows:

- NEW1: the test based on $T_{n,p,q}$;
- NEW: the test based on $\widehat{T}_{n,p,q}$;
- FDS: the test of Zheng et al. (2022);
- TCM: the test of Li et al. (2017);
- MDC: the test of Székely and Rizzo (2013);
- HHG: the test of Heller et al. (2012).

To implement MDC and HHG test procedures, we adopt the `dcor.ttest` function in the energy package and the `hhg.test` function in the HHG package, respectively. We set the sample size $n = 100, 200$, and the dimension $p + q = 500$. The nominal significance level is fixed at $\alpha = 0.05$, and the number of independent replications is 1000. All simulation studies are conducted using R version 4.1.2.

Example 1 This example is designed to compare the finite sample performance of the test procedures. In this example, we assume that

$$\mathbf{Z} = (\mathbf{X}, \mathbf{Y})^\top = \boldsymbol{\Sigma}^{1/2} \mathbf{W},$$

where the components of $\mathbf{W} = (W_1, \dots, W_{p+q})^\top$ are i.i.d., and W_1 is from the standard normal distribution $\mathcal{N}(0, 1)$, the uniform distribution $U(-\sqrt{3}, \sqrt{3})$, the scaled student distribution $t(6)/\sqrt{6/4}$ and the scaled chi-square distribution $(\chi^2(6) - 6)/\sqrt{12}$ with 6 degrees of freedom, respectively. In this example, the dimension is $p = 1, 5, 10, 20, 50, 100$.

Recall from Remark 7, it is of importance to determine c under different choice of (n, p, q) . Thus we set $c = (0.1, 0.2, \dots, 2)$ and investigate their empirical size of our proposed test when $\boldsymbol{\Sigma} = \mathbf{I}_{p+q}$. Note that the choice of c should not be affected by the distribution of W_1 . Here we choose $W_1 \sim \mathcal{N}(0, 1)$. Figures 1 and 2 display the

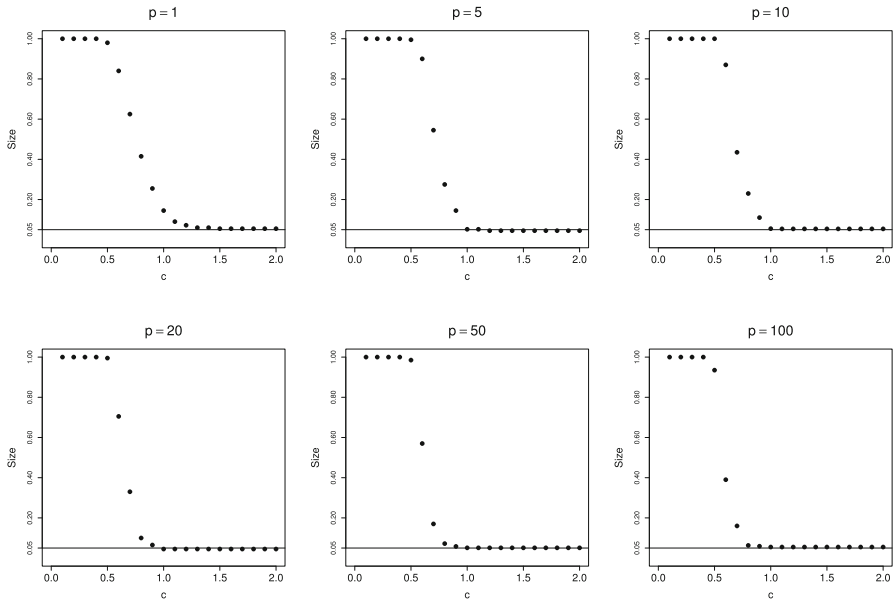


Fig. 1 The empirical size of the proposed test NEW under different values of c in Example 1 when $n = 100$

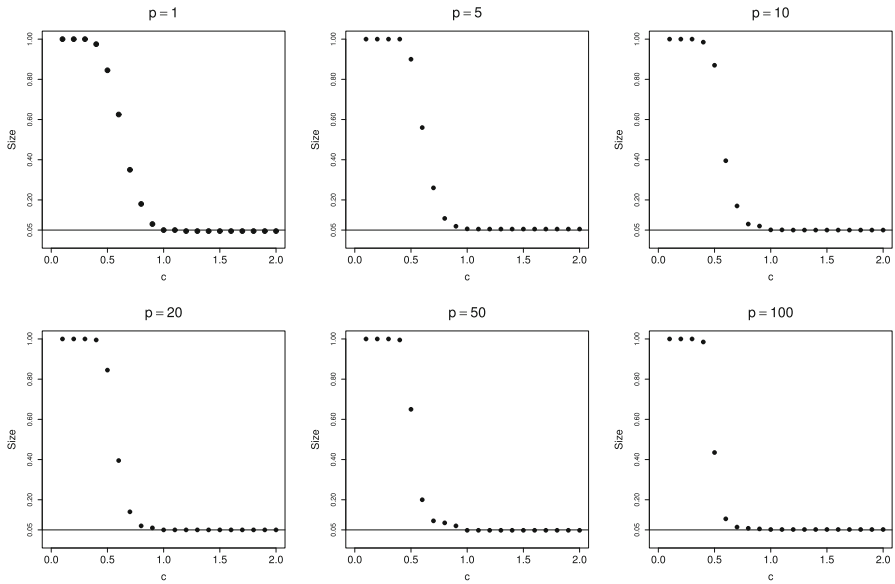


Fig. 2 The empirical size of the proposed test NEW under different values of c in Example 1 when $n = 200$

empirical size of the proposed test procedure NEW and show that $c = 1$ can be applied to all the settings except for the case $(n, p, q) = (100, 1, 499)$ in this Example. We choose $c = 1.5$ when $(n, p, q) = (100, 1, 499)$.

Based on the choice of c , we investigate the empirical size of these tests when $\Sigma = \mathbf{I}_{p+q}$. Table 1 presents the results and shows that all the empirical sizes are close to the nominal significance level $\alpha = 0.05$. Meanwhile, the fact that little difference existing between the empirical sizes of NEW and NEW1 indicates that the screening term $T_{n,p,q}^0$ has little effect on the size under the null hypothesis.

To examine the powers of these test procedures, we set $\Sigma = (0.5^{|i-j|})_{i,j=1}^{p+q}$. The simulation results are summarized in Table 2. From this table it is convenient to see that the performance of NEW and FDS outperform the remaining methods, especially when $n = 200$, which can be attributed to the adoption of the power enhancement technique. Furthermore, it is worth noting that the empirical power of each procedure as $n = 200$ is much higher than that under the setting $n = 100$, which shows the large sample theory.

Example 2 The power of the test procedures are evaluated via the model studied in Jiang et al. (2013). They define the populations \mathbf{X} and \mathbf{Y} as

$$\mathbf{X} = \mathbf{U}_1 + \gamma \mathbf{U}_2^p, \mathbf{Y} = \mathbf{U}_2 + \gamma \mathbf{U}_2$$

where $\mathbf{U}_1 = (U_{11}, \dots, U_{1p})^\top$ and $\mathbf{U}_2 = (U_{21}, \dots, U_{2q})^\top$ are independent, \mathbf{U}_2^p is a subset of \mathbf{U}_2 consisting of its first p variables, and the factor γ represents the degree of dependence. In this example $\gamma = 0.3$. We assume that $U_{11}, \dots, U_{1p}, U_{21}, \dots, U_{2q}$ are i.i.d., and follow the same distribution as W_1 in Example 1. The dimension is $p = 20, 50, 100$ when the sample size is $n = 100$. As for the sample size $n = 200$, the dimension is $p = 5, 10, 20$. From this model it is easy to calculate that the covariance matrices are respectively

$$\Sigma_{\mathbf{X}} = (1 + \gamma^2)\mathbf{I}_p, \Sigma_{\mathbf{Y}} = (1 + \gamma)^2\mathbf{I}_q, \Sigma_{\mathbf{XY}} = \gamma(1 + \gamma)(\mathbf{I}_{p_1}, O_{p,q-p}),$$

where $O_{p,q-p}$ denotes a $p \times (q - p)$ zero matrix. Table 3 displays the performance of these methods, and shows the empirical power of each procedure increases with the increasing dimension p , which implies that these methods are effective against the dense alternatives. At the same time to achieve the same performance for each of these test procedures, the dimension p under the setting that $n = 200$ can be much smaller than that of $n = 100$, which implies that the powers of these methods can tend to be 1 as long as the sample size n is sufficiently large.

Example 3 To illustrate the application of our proposed procedure in high-dimensional settings, we analyze a microarray data reported in Scheetz et al. (2006). In order to gain a broad perspective of gene regulation in the mammalian eye and to identify genetic variation relevant to human eye disease, 120 twelve-week-old male offspring were chosen for tissue harvesting from their eyes and microarray analysis. 18,976 probes on the array which was used to analyze the RNA from the eyes of these F2 animals

Table 1 Empirical sizes of the tests at the nominal significance level 5% in Example 1

n	p	NEW1	NEW	FDS	TCM	MDC	HHG	NEW1	NEW	FDS	TCM	MDC	HHG	
100	1	$\mathcal{N}(0, 1)$												
		5.6	5.6	5.3	5.6	5.2	4.8	5.4	5.5	5.0	5.7	5.6	5.0	
		5.4	5.7	4.9	6.0	5.8	5.3	4.7	5.0	4.3	4.7	4.2	4.3	
		5.6	5.8	4.9	5.4	5.0	4.6	4.9	5.4	4.4	5.2	5.2	5.5	
		4.3	5.2	5.0	5.4	5.1	4.6	5.0	5.5	4.7	5.2	4.0	5.1	
		5.6	5.9	4.4	5.0	4.9	4.5	4.3	4.5	4.2	4.2	4.4	4.4	
		5.0	5.8	4.3	6.0	6.1	5.1	5.0	5.2	6.0	6.8	6.5	4.6	
		$(\chi^2(6) - 6)/\sqrt{12}$												
		4.2	4.9	5.2	4.0	4.4	5.1	4.5	4.8	4.2	4.6	4.7	4.6	
		5.8	6.0	6.0	5.3	5.1	5.4	5.5	5.8	4.6	5.1	5.6	5.3	
200	1	$U(-\sqrt{3}, \sqrt{3})$												
		5.0	5.4	4.0	5.1	4.3	5.3	4.7	4.8	3.9	5.0	5.2	5.1	
		4.2	5.6	4.5	4.1	4.4	5.3	4.0	4.2	4.2	4.1	4.5	6.0	
		5.4	6.2	6.8	5.3	4.7	4.8	4.9	5.3	4.5	4.6	4.8	4.5	
		6.0	6.1	6.9	6.3	6.5	4.3	4.4	5.6	4.4	5.6	4.9	5.4	
		$\mathcal{N}(0, 1)$												
		4.2	4.2	4.8	4.7	4.3	5.5	4.7	4.7	4.6	4.5	5.6	4.6	
		5.2	5.4	5.0	5.3	5.0	4.1	5.1	5.4	4.6	4.8	4.0	5.0	
		5.7	5.8	5.9	5.9	5.3	5.6	4.5	6.1	4.4	4.7	4.4	5.6	
		5.1	5.6	5.2	4.8	5.0	4.7	6.0	6.3	6.2	6.5	5.9	5.1	
4.0	4.6	4.2	4.3	4.6	5.5	5.3	5.9	4.9	5.2	4.3	5.4			
5.0	5.0	5.3	5.0	4.3	5.0	4.3	4.5	4.4	5.0	4.8	5.0			
$(\chi^2(6) - 6)/\sqrt{12}$														
5.4	5.4	5.3	5.5	4.5	4.5	5.3	5.3	5.7	5.3	5.8	4.9			

Table 1 continued

n	p	NEW1	NEW	FDS	TCM	MDC	HHG	NEW1	NEW	FDS	TCM	MDC	HHG
		$\mathcal{N}(0, 1)$											
		$t(6)/\sqrt{6/4}$											
	5	5.0	5.4	4.8	4.9	4.6	5.8	4.7	5.9	4.9	4.8	4.3	4.8
	10	5.1	5.5	5.1	5.3	4.4	4.3	5.2	5.6	5.9	6.1	6.6	4.5
	20	4.1	5.2	4.3	4.2	4.6	5.0	4.3	4.9	4.5	3.9	4.3	5.0
	50	4.8	5.2	4.1	4.3	4.4	4.2	4.7	5.3	5.3	4.3	4.7	4.4
	100	4.2	4.5	5.4	5.6	5.8	4.0	5.2	5.6	5.5	5.2	4.1	4.5

All numbers reported in this table are multiplied by 100

Table 2 Empirical power(%) of the tests in Example 1

n	p	NEW1	NEW	FDS	TCM	MDC	HHG	NEW1	NEW	FDS	TCM	MDC	HHG
		$\mathcal{N}(0, 1)$											
100	1	19.3	78.8	61.6	18.6	15.9	5.2	21.3	81.9	56.4	20.7	17.9	5.9
	5	10.1	80.2	45.5	9.5	9.8	6.0	15.6	82.8	38.2	13.1	12.4	4.3
	10	7.9	74.6	39.8	7.3	7.8	5.5	14.3	75.6	32.8	12.4	12.1	4.2
	20	7.5	68.6	34.8	6.9	6.8	4.9	9.8	75.2	28.6	8.7	11.8	5.7
	50	6.5	65.7	31.2	6.4	6.7	4.7	8.6	57.5	15.9	6.8	8.9	5.3
	100	5.6	53.9	22.8	6.5	6.0	4.7	7.0	54.2	20.4	6.4	7.5	4.4
		$(\chi^2(6) - 6)/\sqrt{12}$											
	1	22.9	80.7	61.9	22.9	19.7	4.9	20.0	79.0	71.3	20.3	15.3	5.7
	5	9.6	82.4	40.7	9.9	9.8	5.6	12.3	81.6	58.4	11.4	10.3	5.4
	10	10.0	79.4	39.2	8.6	8.9	4.8	11.4	75.0	50.8	10.8	9.6	4.7
	20	8.5	71.6	41.5	8.8	7.6	5.7	9.4	70.8	35.8	8.0	9.0	4.9
	50	8.7	59.8	14.7	6.2	6.9	5.8	6.5	62.8	39.4	6.5	6.8	5.3
	100	8.2	53.8	13.8	5.6	6.8	3.9	6.8	55.6	34.8	6.2	6.5	6.1
		$\mathcal{N}(0, 1)$											
200	1	43.8	100	100	45.6	44.4	5.5	58.6	100	99.6	57.4	45.8	4.9
	5	21.4	100	99.6	22.5	17.9	4.2	17.6	100	96.5	16.4	17.4	5.5
	10	13.5	100	97.8	12.6	12.9	5.1	16.8	100	92.8	15.4	16.8	4.4
	20	11.5	100	97.3	12.1	11.5	3.5	15.8	100	89.4	12.8	11.5	5.2
	50	7.5	100	96.5	8.2	8.5	5.6	11.8	100	85.2	11.5	9.7	5.1
	100	4.8	98.6	90.4	6.5	4.9	5.4	7.5	98.7	84.5	8.6	6.5	5.3
		$(\chi^2(6) - 6)/\sqrt{12}$											
	1	49.6	100	98.5	46.8	36.9	5.6	54.6	100	100	52.2	46.8	5.0
	5	28.7	100	94.6	27.8	22.6	4.7	17.4	100	100	16.2	12.8	5.6

Table 2 continued

n	p	NEW1	NEW	FDS	TCM	MDC	HHG	NEW1	NEW	FDS	TCM	MDC	HHG
		$\mathcal{N}(0, 1)$											
		$t(6)/\sqrt{6/4}$											
	10	16.8	100	91.2	12.4	16.9	4.1	16.5	100	99.4	15.4	11.6	4.8
	20	14.5	100	92.5	12.5	12.1	3.6	11.8	100	96.2	10.8	9.6	6.6
	50	12.7	99.6	88.6	10.8	10.5	5.4	8.2	99.6	95.6	7.4	7.8	4.6
	100	10.7	98.5	85.9	8.6	9.8	5.3	6.5	98.2	93.8	6.2	4.5	6.2

All numbers reported in this table are multiplied by 100

Table 3 Empirical power(%) of the tests in Example 2

n	p	NEW1	NEW	FDS	TCM	MDC	HHG	NEW1	NEW	FDS	TCM	MDC	HHG	
		$\mathcal{N}(0, 1)$												
100	20	36.5	57.8	35.2	36.4	34.5	7.6	41.6	60.8	37.5	37.8	37.6	7.4	
	50	64.6	76.5	67.4	66.8	63.5	10.6	56.7	82.5	55.4	54.2	55.4	10.1	
	100	93.4	97.7	94.5	93.6	93.8	15.3	91.7	97.8	93.2	93.5	92.9	14.0	
		$W_1 \sim (\chi^2(6) - 6)/\sqrt{12}$												
	20	38.4	63.8	35.2	36.3	33.9	5.5	27.5	49.6	46.5	29.8	28.5	8.3	
	50	59.7	76.6	56.4	55.2	55.4	7.7	62.5	80.7	74.2	63.5	59.8	12.5	
	100	87.4	94.6	86.5	86.4	87.5	11.6	93.5	97.8	96.4	93.6	93.8	20.5	
		$W_1 \sim \mathcal{N}(0, 1)$												
200	5	28.4	90.6	65.4	28.3	28.5	5.5	30.6	93.4	65.8	31.8	27.6	5.7	
	10	51.6	97.4	83.2	50.4	48.7	7.8	50.8	96.5	71.4	48.8	54.9	8.0	
	20	68.9	99.2	96.3	67.5	67.8	9.6	77.8	100	93.7	77.6	78.9	9.0	
		$W_1 \sim (\chi^2(6) - 6)/\sqrt{12}$												
	5	32.6	90.5	57.6	39.7	34.9	5.9	39.8	88.5	77.5	27.8	22.6	8.5	
	10	52.4	96.6	72.1	51.9	51.4	7.0	47.2	98.4	93.6	45.8	40.9	8.3	
	20	79.8	99.2	88.7	78.4	77.8	7.2	73.8	98.9	98.4	73.5	70.8	14.6	

All numbers reported in this table are multiplied by 100

Table 4 P-value in Example 3

<i>n</i>	<i>p</i>	NEW1	NEW	FDS	TCM	MDC	HHG
120	18,975	0	0	0	0	0	0.005
	18,956	0.895	0.895	0.875	0.878	0.749	0.493

were detected sufficiently expressed and variable. More details of the experiment can be found in Scheetz et al. (2006).

Note that 1389163_at, one of the 18,976 sufficiently expressed probes, is from the gene TRIM32. This gene is found in Chiang et al. (2006) to cause an extremely heterogeneous human obesity syndrome known as Bardet-Biedl syndrome. The relationship between the probe 1389163_at and the remaining 18,975 ones is first investigated. The P-values of the three tests are listed in Table 4 and denoted by $p = 18,975$, which suggests that gene TRIM32 and the rest of gene exhibit some type of association. Huang et al. (2008) verified this situation and further studied the data set. They used the adaptive Lasso in sparse high-dimensional linear regression models and selected 19 genes whose expression are most correlated with that of gene TRIM32. Excluding these 19 probes, we take another test to check whether the probe 1389163_at is associated with the 18,956 ones or not. To remove the effects of these 19 genes on TRIM32, we replace the value of TRIM32 with the residual from a multiple linear regression on the response variable TRIM32 and these 19 genes being the predictors. Table 4 also presents the result when $p = 18,956$, which shows the absence of a linear relationship between the gene TRIM32 and the 18,956 ones.

Acknowledgements This work is supported by National Natural Science Foundation of China (No. 12031016, 11971324, 12371294, 11901406, 11471223), Beijing Natural Science Foundation (Z210003), R&D Program of Beijing Municipal Education Commission (KM202110028017), the Interdisciplinary Construction of Bioinformatics and Statistics, and the Academy for Multidisciplinary Studies, Capital Normal University, and New Young Teachers' Research Initiation Fund Project, Capital University of Economics and Business (XRZ2021046).

Appendix

Proof of Theorem 1 Without loss of generality, we can assume that $E(\mathbf{X}) = 0_{p \times 1}$ and $E(\mathbf{Y}) = 0_{q \times 1}$ hereafter. Define that $\xi_c = \text{Var}\{h_c(\mathbf{Z}_1, \dots, \mathbf{Z}_c)\}$, where $h_c(z_1, \dots, z_c) = E\{h(z_1, \dots, z_c, \mathbf{Z}_{c+1}, \dots, \mathbf{Z}_4)\}$ for $c = 1, 2, 3, 4$. Denote $G(\mathbf{X}, \mathbf{X}') = \mathbf{X}^\top \mathbf{X}'$ and $H(\mathbf{Y}, \mathbf{Y}') = \mathbf{Y}^\top \mathbf{Y}'$. Observe that $E\{L(\mathbf{Z}, \mathbf{Z}')\} = E\{L(\mathbf{Z}, \mathbf{Z}')|\mathbf{Z}\} = E\{L(\mathbf{Z}, \mathbf{Z}')|\mathbf{Z}'\} = 0$ under the null, and that $E\{G(\mathbf{X}, \mathbf{X}')|\mathbf{X}\} = E\{G(\mathbf{X}, \mathbf{X}')|\mathbf{X}'\} = E\{H(\mathbf{Y}, \mathbf{Y}')|\mathbf{Y}\} = E\{H(\mathbf{Y}, \mathbf{Y}')|\mathbf{Y}'\} = 0$. Then we can obtain under the null hypothesis that

$$\begin{aligned}
 h_1(\mathbf{Z}_1) &= 0, \\
 h_2(\mathbf{Z}_1, \mathbf{Z}_2) &= \frac{1}{6}G(\mathbf{X}_1, \mathbf{X}_2)H(\mathbf{Y}_1, \mathbf{Y}_2),
 \end{aligned}$$

$$\begin{aligned}
 h_3(\mathbf{Z}_1, \mathbf{Z}_2, \mathbf{Z}_3) &= \frac{1}{12} [\{ 2G(\mathbf{X}_1, \mathbf{X}_2) - G(\mathbf{X}_1, \mathbf{X}_3) - G(\mathbf{X}_2, \mathbf{X}_3) \} H(\mathbf{Y}_1, \mathbf{Y}_2) \\
 &\quad + \{ 2G(\mathbf{X}_1, \mathbf{X}_3) - G(\mathbf{X}_1, \mathbf{X}_2) - G(\mathbf{X}_2, \mathbf{X}_3) \} H(\mathbf{Y}_1, \mathbf{Y}_3) \\
 &\quad + \{ 2G(\mathbf{X}_2, \mathbf{X}_3) - G(\mathbf{X}_1, \mathbf{X}_2) - G(\mathbf{X}_1, \mathbf{X}_3) \} H(\mathbf{Y}_2, \mathbf{Y}_3)], \\
 h_4(\mathbf{Z}_1, \mathbf{Z}_2, \mathbf{Z}_3, \mathbf{Z}_4) &= \frac{1}{12} [\{ 2G(\mathbf{X}_1, \mathbf{X}_2) + 2G(\mathbf{X}_3, \mathbf{X}_4) - G(\mathbf{X}_1, \mathbf{X}_3) - G(\mathbf{X}_1, \mathbf{X}_4) \\
 &\quad - G(\mathbf{X}_2, \mathbf{X}_3) - G(\mathbf{X}_2, \mathbf{X}_4) \} \{ H(\mathbf{Y}_1, \mathbf{Y}_2) + H(\mathbf{Y}_3, \mathbf{Y}_4) \} \\
 &\quad + \{ 2G(\mathbf{X}_1, \mathbf{X}_3) + 2G(\mathbf{X}_2, \mathbf{X}_4) - G(\mathbf{X}_1, \mathbf{X}_2) - G(\mathbf{X}_1, \mathbf{X}_4) \\
 &\quad - G(\mathbf{X}_2, \mathbf{X}_3) - G(\mathbf{X}_3, \mathbf{X}_4) \} \{ H(\mathbf{Y}_1, \mathbf{Y}_3) + H(\mathbf{Y}_2, \mathbf{Y}_4) \} \\
 &\quad + \{ 2G(\mathbf{X}_1, \mathbf{X}_4) + 2G(\mathbf{X}_2, \mathbf{X}_3) - G(\mathbf{X}_1, \mathbf{X}_2) - G(\mathbf{X}_1, \mathbf{X}_3) \\
 &\quad - G(\mathbf{X}_2, \mathbf{X}_4) - G(\mathbf{X}_3, \mathbf{X}_4) \} \{ H(\mathbf{Y}_1, \mathbf{Y}_4) + H(\mathbf{Y}_2, \mathbf{Y}_3) \}].
 \end{aligned}$$

Under Assumption (A1) we have

$$\xi_1 = 0, \xi_2 = \frac{\zeta^2}{36}, \xi_3 = o(n\zeta^2), \xi_4 = o(n^2\zeta^2).$$

Applying the Hoeffding’ decomposition in Serfling (1980), we get that

$$\frac{nT_{n,p,q}(\mathbf{X}, \mathbf{Y})}{\sqrt{2\zeta^2}} = \frac{\sum_{1 \leq i < j \leq n} L(\mathbf{Z}_i, \mathbf{Z}_j)}{\sqrt{\binom{n}{2}\zeta^2}} + o_p(1).$$

Write $M_j = \sum_{i=1}^{j-1} L(\mathbf{Z}_i, \mathbf{Z}_j)$, $S_u = \sum_{j=2}^u \sum_{i=1}^{j-1} L(\mathbf{Z}_i, \mathbf{Z}_j) = \sum_{j=2}^u M_j$ and the filtration $\mathcal{F}_u = \sigma\{\mathbf{Z}_1, \dots, \mathbf{Z}_u\}$. Since $E\{L(\mathbf{Z}, \mathbf{Z}')\} = E\{L(\mathbf{Z}, \mathbf{Z}')|\mathbf{Z}\} = E[L(\mathbf{Z}, \mathbf{Z}')|\mathbf{Z}'] = 0$, then $E(S_u) = 0$ and for $u < v$,

$$\begin{aligned}
 E(S_v|\mathcal{F}_u) &= S_u + \sum_{j=u+1}^v \sum_{i=1}^u E\{L(\mathbf{Z}_i, \mathbf{Z}_j)|\mathbf{Z}_i\} + \sum_{j=u+2}^v \sum_{i=u+1}^{j-1} E\{L(\mathbf{Z}_i, \mathbf{Z}_j)\} \\
 &= S_u.
 \end{aligned}$$

This implies that S_u is adaptive to \mathcal{F}_u and thus it is a mean-zero martingale sequence.

To obtain the asymptotic normality of $\sum_{1 \leq i < j \leq n} L(\mathbf{Z}_i, \mathbf{Z}_j) / \sqrt{\binom{n}{2}\zeta^2}$, we only need to verify the two following conditions given by Corollary 3.1 in Hall and Heyde (1980).

Condition (1): the conditional variance

$$\frac{2}{n(n-1)\zeta^2} \sum_{j=2}^n E(M_j^2|\mathcal{F}_{j-1}) \xrightarrow{P} 1,$$

and Condition (2): the conditional Lindeberg condition, that is, for all $\varepsilon > 0$,

$$\frac{1}{n(n-1)\zeta^2} \sum_{j=2}^n \mathbb{E} \left[M_j^2 \mathbf{I} \{ |M_j| > \varepsilon \sqrt{n(n-1)\zeta^2} \} | \mathcal{F}_{j-1} \right] \xrightarrow{P} 0.$$

To prove condition (1), it is sufficient to prove that

$$\mathbb{E} \left\{ \frac{2}{n(n-1)\zeta^2} \sum_{j=2}^n \mathbb{E}(M_j^2 | \mathcal{F}_{j-1}) \right\} = 1,$$

and

$$\text{Var} \left\{ \frac{2}{n(n-1)\zeta^2} \sum_{j=2}^n \mathbb{E}(M_j^2 | \mathcal{F}_{j-1}) \right\} \rightarrow 0.$$

Notice again that $\mathbb{E}\{L(\mathbf{Z}, \mathbf{Z}')\} = \mathbb{E}\{L(\mathbf{Z}, \mathbf{Z}') | \mathbf{Z}\} = \mathbb{E}\{L(\mathbf{Z}, \mathbf{Z}') | \mathbf{Z}'\} = 0$. We then show that

$$\begin{aligned} & \mathbb{E} \left\{ \frac{2}{n(n-1)\zeta^2} \sum_{j=2}^n \mathbb{E}(M_j^2 | \mathcal{F}_{j-1}) \right\} \\ &= \frac{2}{n(n-1)\zeta^2} \sum_{j=2}^n \mathbb{E} \left\{ \sum_{i,i'=1}^{j-1} L(\mathbf{Z}_i, \mathbf{Z}_j) L(\mathbf{Z}_{i'}, \mathbf{Z}_j) \right\} \\ &= 1. \end{aligned}$$

Next we will prove that

$$\text{Var} \left\{ \frac{2}{n(n-1)\zeta^2} \sum_{j=2}^n \mathbb{E}(M_j^2 | \mathcal{F}_{j-1}) \right\} \rightarrow 0.$$

Recall that $L_1(\mathbf{Z}, \mathbf{Z}') = \mathbb{E}\{L(\mathbf{Z}, \mathbf{Z}'')L(\mathbf{Z}', \mathbf{Z}'') | (\mathbf{Z}, \mathbf{Z}')\}$. It is easy to check that

$$\mathbb{E}\{L_1(\mathbf{Z}, \mathbf{Z}')\} = \mathbb{E}\{L_1(\mathbf{Z}, \mathbf{Z}') | \mathbf{Z}\} = \mathbb{E}\{L_1(\mathbf{Z}, \mathbf{Z}') | \mathbf{Z}'\} = 0.$$

Then we have that

$$\begin{aligned} & \text{Var} \left\{ \sum_{j=2}^n \mathbb{E}(M_j^2 | \mathcal{F}_{j-1}) \right\} \\ &= \sum_{j=2}^n \text{Var} \left\{ \mathbb{E}(M_j^2 | \mathcal{F}_{j-1}) \right\} + 2 \sum_{2 \leq j < t \leq n} \text{Cov} \left\{ \mathbb{E}(M_j^2 | \mathcal{F}_{j-1}), \mathbb{E}(M_t^2 | \mathcal{F}_{t-1}) \right\} \end{aligned}$$

$$= \frac{n(n-1)(2n-1)}{6} \text{Var}\{L_1(\mathbf{Z}, \mathbf{Z})\} + \frac{n(n-1)^2(n-2)}{3} E\{L_1(\mathbf{Z}, \mathbf{Z}')^2\}.$$

When assumption (A1) is true, we thus obtain that

$$\text{Var} \left\{ \frac{2}{n(n-1)\zeta^2} \sum_{j=2}^n E(M_j^2 | \mathcal{F}_{j-1}) \right\} = \frac{4}{n^2(n-1)^2\zeta^4} \text{Var} \left\{ \sum_{j=2}^n E(M_j^2 | \mathcal{F}_{j-1}) \right\} \rightarrow 0,$$

which ensures condition (1).

Next condition (2) will be verified. Direct calculation shows that

$$\begin{aligned} \sum_{j=2}^n E(M_j^4) &= \sum_{j=2}^n \sum_{i_1, i_2, i_3, i_4=1}^{j-1} E\{L(\mathbf{Z}_{i_1}, \mathbf{Z}_j)L(\mathbf{Z}_{i_2}, \mathbf{Z}_j)L(\mathbf{Z}_{i_3}, \mathbf{Z}_j)L(\mathbf{Z}_{i_4}, \mathbf{Z}_j)\} \\ &= \frac{n(n-1)}{2} E\{L(\mathbf{Z}, \mathbf{Z}')^4\} + n(n-1)(n-2)E[L(\mathbf{Z}, \mathbf{Z}')^2L(\mathbf{Z}, \mathbf{Z}'')^2]. \end{aligned}$$

Again under assumption (A1), it is easy to show that

$$\frac{1}{n^4\zeta^4} \sum_{j=2}^n E(M_j^4) \rightarrow 0,$$

which implies that

$$\frac{1}{n^2(n-1)^2\zeta^4} \sum_{j=2}^n E(M_j^4 | \mathcal{F}_{j-1}) \xrightarrow{P} 0.$$

Notice that for all $\varepsilon > 0$,

$$\begin{aligned} &\frac{1}{n(n-1)\zeta^2} \sum_{j=2}^n E[M_j^2 \mathbf{1}\{|M_j| > \varepsilon\sqrt{n(n-1)\zeta^2}\} | \mathcal{F}_{j-1}] \\ &\leq \frac{1}{\varepsilon^2 n^2(n-1)^2\zeta^4} \sum_{j=2}^n E(M_j^4 | \mathcal{F}_{j-1}), \end{aligned}$$

which implies condition (2).

Therefore, we can show that

$$\frac{\sum_{1 \leq i < j \leq n} L(\mathbf{Z}_i, \mathbf{Z}_j)}{\sqrt{\binom{n}{2}\zeta^2}} \xrightarrow{d} \mathcal{N}(0, 1).$$

Using the Slutsky theorem we obtain that

$$\frac{nT_{n,p,q}((\mathbf{X}, \mathbf{Y}))}{\sqrt{2\zeta^2}} \xrightarrow{d} \mathcal{N}(0, 1). \tag{41}$$

In the following, we will show the consistency of ζ_n^2 . That is,

$$\frac{\zeta_n^2}{\zeta^2} \xrightarrow{P} 1. \tag{42}$$

To obtain this result, it is sufficient to verify $E(\zeta_n^2/\zeta^2) \rightarrow 1$ and $\text{Var}(\zeta_n^2/\zeta^2) \rightarrow 0$, respectively. Recall the definition of A_{ij} and B_{ij} , we can rewrite them in the following form:

$$\begin{aligned} A_{ij} &= a_{ij} - \frac{1}{n-2} \sum_{l \neq i} a_{il} - \frac{1}{n-2} \sum_{k \neq j} a_{kj} + \frac{1}{(n-1)(n-2)} \sum_{k \neq l} a_{kl}, \\ &= \frac{n-3}{n-1} \left\{ G(\mathbf{X}_i, \mathbf{X}_j) - \frac{1}{n-2} \sum_{l \notin \{i,j\}} G(\mathbf{X}_i, \mathbf{X}_l) - \frac{1}{n-2} \sum_{k \notin \{i,j\}} G(\mathbf{X}_k, \mathbf{X}_j) \right. \\ &\quad \left. + \frac{1}{(n-2)(n-3)} \sum_{\{k,l\} \cap \{i,j\} = \emptyset} G(\mathbf{X}_k, \mathbf{X}_l) \right\} \\ &:= \frac{n-3}{n-1} (I_{n,1} + I_{n,2} + I_{n,3} + I_{n,4}), \\ B_{ij} &= b_{ij} - \frac{1}{n-2} \sum_{l \neq i} b_{il} - \frac{1}{n-2} \sum_{k \neq j} b_{kj} + \frac{1}{(n-1)(n-2)} \sum_{k \neq l} b_{kl}, \\ &= \frac{n-3}{n-1} \left\{ H(\mathbf{Y}_i, \mathbf{Y}_j) - \frac{1}{n-2} \sum_{l \notin \{i,j\}} H(\mathbf{Y}_i, \mathbf{Y}_l) - \frac{1}{n-2} \sum_{k \notin \{i,j\}} H(\mathbf{Y}_k, \mathbf{Y}_j) \right. \\ &\quad \left. + \frac{1}{(n-2)(n-3)} \sum_{\{k,l\} \cap \{i,j\} = \emptyset} H(\mathbf{Y}_k, \mathbf{Y}_l) \right\} \\ &:= \frac{n-3}{n-1} (J_{n,1} + J_{n,2} + J_{n,3} + J_{n,4}). \end{aligned}$$

Notice that $E\{G(\mathbf{X}, \mathbf{X}')|\mathbf{X}\} = 0$ and $E\{H(\mathbf{Y}, \mathbf{Y}')|\mathbf{Y}\} = 0$, we have for $1 \leq i \neq j \leq n$,

$$\begin{aligned} &\frac{(n-1)^4}{(n-3)^4} E(A_{ij}^2 B_{ij}^2) \\ &= E\{(I_{n,1} + I_{n,2} + I_{n,3} + I_{n,4})^2 (J_{n,1} + J_{n,2} + J_{n,3} + J_{n,4})^2\} \\ &= E\{I_{n,1}^2 + 2I_{n,1}(I_{n,2} + I_{n,3} + I_{n,4}) + (I_{n,2} + I_{n,3} + I_{n,4})^2\} \end{aligned}$$

$$\begin{aligned} & \{J_{n,1}^2 + 2J_{n,1}(J_{n,2} + J_{n,3} + J_{n,4}) + (J_{n,2} + J_{n,3} + J_{n,4})^2\} \\ = & E(I_{n,1}^2 J_{n,1}^2) + 2E\{I_{n,1}^2 J_{n,1}(J_{n,2} + J_{n,3} + J_{n,4})\} + 2E\{J_{n,1}^2 I_{n,1}(I_{n,2} + I_{n,3} + I_{n,4})\} \\ & + E\{I_{n,1}^2 (J_{n,2} + J_{n,3} + J_{n,4})^2\} + E\{J_{n,1}^2 (I_{n,2} + I_{n,3} + I_{n,4})^2\} + E\{2I_{n,1}(I_{n,2} \\ & + I_{n,3} + I_{n,4}) + (I_{n,2} + I_{n,3} + I_{n,4})^2\} \\ & \{2J_{n,1}(J_{n,2} + J_{n,3} + J_{n,4}) + (J_{n,2} + J_{n,3} + J_{n,4})^2\}, \end{aligned}$$

which implies that

$$\begin{aligned} & \left| \frac{(n-1)^4}{(n-3)^4} E(A_{ij}^2 B_{ij}^2) - E(I_{n,1}^2 J_{n,1}^2) \right| \\ & \leq O\{E(I_{n,1}^2 J_{n,2}^2) + E(I_{n,1}^2 J_{n,3}^2) + E(I_{n,1}^2 J_{n,4}^2) \\ & \quad + E(I_{n,2}^2 J_{n,1}^2) + E(I_{n,2}^2 J_{n,2}^2) + E(I_{n,2}^2 J_{n,3}^2) + E(I_{n,2}^2 J_{n,4}^2) \\ & \quad + E(I_{n,3}^2 J_{n,1}^2) + E(I_{n,3}^2 J_{n,2}^2) + E(I_{n,3}^2 J_{n,3}^2) + E(I_{n,3}^2 J_{n,4}^2) \\ & \quad + E(I_{n,4}^2 J_{n,1}^2) + E(I_{n,4}^2 J_{n,2}^2) + E(I_{n,4}^2 J_{n,3}^2) + E(I_{n,4}^2 J_{n,4}^2)\}. \end{aligned}$$

By the Cauchy–Schwarz inequality, we have under condition (A1) that

$$\begin{aligned} E\{G(\mathbf{X}, \mathbf{X}')^2 H(\mathbf{Y}, \mathbf{Y}'')^2\} & \leq \sqrt{E\{G(\mathbf{X}, \mathbf{X}')^4 H(\mathbf{Y}, \mathbf{Y}'')^4\}} \\ & = o(n\zeta^2), \\ E\{G(\mathbf{X}, \mathbf{X}')^2\}E\{H(\mathbf{Y}, \mathbf{Y}'')^2\} & \leq \sqrt{E\{G(\mathbf{X}, \mathbf{X}')^4\}E\{H(\mathbf{Y}, \mathbf{Y}'')^4\}} \\ & = o(n\zeta^2). \end{aligned}$$

Then it is trivial to check that

$$\begin{aligned} E(I_{n,1}^2 J_{n,1}^2) & = \zeta^2, \\ E(I_{n,1}^2 J_{n,2}^2) & = \frac{1}{(n-2)^2} E \left[G(\mathbf{X}_i, \mathbf{X}_j)^2 \left\{ \sum_{l \notin \{i,j\}} H(\mathbf{Y}_i, \mathbf{Y}_l) \right. \right. \\ & \quad \left. \left. + \sum_{\{l,l'\} \cap \{i,j\} = \emptyset} H(\mathbf{Y}_i, \mathbf{Y}_l) H(\mathbf{Y}_i, \mathbf{Y}_{l'}) \right\} \right] \\ & = \frac{1}{(n-2)} E\{G(\mathbf{X}, \mathbf{X}')^2 H(\mathbf{Y}, \mathbf{Y}'')^2\} \\ & = o(\zeta^2), \\ E(I_{n,1}^2 J_{n,3}^2) & = E(I_{n,2}^2 J_{n,1}^2) = E(I_{n,3}^2 J_{n,1}^2) \\ & = \frac{1}{(n-2)} E\{G(\mathbf{X}, \mathbf{X}')^2 H(\mathbf{Y}, \mathbf{Y}'')^2\} \\ & = o(\zeta^2), \end{aligned}$$

$$\begin{aligned}
 E(I_{n,1}^2 J_{n,4}^2) &= \frac{1}{(n-2)^2(n-3)^2} E \left[G(\mathbf{X}_i, \mathbf{X}_j)^2 \left\{ \sum_{\{k,l\} \cap \{i,j\} = \emptyset} 2H(\mathbf{Y}_k, \mathbf{Y}_l)^2 \right. \right. \\
 &\quad + \sum_{\{k,l,l'\} \cap \{i,j\} = \emptyset} 4H(\mathbf{Y}_k, \mathbf{Y}_l)H(\mathbf{Y}_k, \mathbf{Y}_{l'}) \\
 &\quad \left. \left. + \sum_{\{k,k',l,l'\} \cap \{i,j\} = \emptyset} H(\mathbf{Y}_k, \mathbf{Y}_l)H(\mathbf{Y}_{k'}, \mathbf{Y}_{l'}) \right\} \right] \\
 &= \frac{2}{(n-2)(n-3)} E\{G(\mathbf{X}, \mathbf{X}')^2\} E\{H(\mathbf{Y}, \mathbf{Y}')^2\} \\
 &= o(\zeta^2),
 \end{aligned}$$

$$\begin{aligned}
 E(I_{n,4}^2 J_{n,1}^2) &= \frac{2}{(n-2)(n-3)} E\{G(\mathbf{X}, \mathbf{X}')^2\} E\{H(\mathbf{Y}, \mathbf{Y}')^2\} \\
 &= o(\zeta^2),
 \end{aligned}$$

$$\begin{aligned}
 E(I_{n,2}^2 J_{n,2}^2) &= \frac{1}{(n-2)^4} E \left[\left\{ \sum_{l \notin \{i,j\}} G(\mathbf{X}_i, \mathbf{X}_l)^2 + \sum_{\{l,l'\} \cap \{i,j\} = \emptyset} G(\mathbf{X}_i, \mathbf{X}_l)G(\mathbf{X}_i, \mathbf{X}_{l'}) \right\} \right. \\
 &\quad \left. \left\{ \sum_{l \notin \{i,j\}} H(\mathbf{Y}_i, \mathbf{Y}_l)^2 + \sum_{\{l,l'\} \cap \{i,j\} = \emptyset} H(\mathbf{Y}_i, \mathbf{Y}_l)H(\mathbf{Y}_i, \mathbf{Y}_{l'}) \right\} \right] \\
 &= \frac{1}{(n-2)^3} \zeta^2 + \frac{n-3}{(n-2)^3} E\{G(\mathbf{X}, \mathbf{X}')^2 H(\mathbf{Y}, \mathbf{Y}'')^2\} \\
 &\quad + \frac{2(n-3)}{(n-2)^3} E\{G(\mathbf{X}, \mathbf{X}')H(\mathbf{Y}, \mathbf{Y}')G(\mathbf{X}, \mathbf{X}'')H(\mathbf{Y}, \mathbf{Y}'')\} \\
 &\leq O \left\{ \frac{\zeta^2 + E\{G(\mathbf{X}, \mathbf{X}')^2 H(\mathbf{Y}, \mathbf{Y}'')^2\}}{n^2} \right\} \\
 &= o(\zeta^2),
 \end{aligned}$$

$$\begin{aligned}
 E(I_{n,3}^2 J_{n,3}^2) &= \frac{1}{(n-2)^3} \zeta^2 + \frac{n-3}{(n-2)^3} E\{G(\mathbf{X}, \mathbf{X}')^2 H(\mathbf{Y}, \mathbf{Y}'')^2\}, \\
 &\quad + \frac{2(n-3)}{(n-2)^3} E\{G(\mathbf{X}, \mathbf{X}')H(\mathbf{Y}, \mathbf{Y}')G(\mathbf{X}, \mathbf{X}'')H(\mathbf{Y}, \mathbf{Y}'')\} \\
 &\leq O \left\{ \frac{\zeta^2 + E\{G(\mathbf{X}, \mathbf{X}')^2 H(\mathbf{Y}, \mathbf{Y}'')^2\}}{n^2} \right\} \\
 &= o(\zeta^2),
 \end{aligned}$$

$$\begin{aligned}
 E(I_{n,2}^2 J_{n,3}^2) &= \frac{1}{(n-2)^4} E \left[\left\{ \sum_{l \notin \{i,j\}} G(\mathbf{X}_i, \mathbf{X}_l)^2 + \sum_{\{l,l'\} \cap \{i,j\} = \emptyset} G(\mathbf{X}_i, \mathbf{X}_l)G(\mathbf{X}_i, \mathbf{X}_{l'}) \right\} \right. \\
 &\quad \left. \left\{ \sum_{k \notin \{i,j\}} H(\mathbf{Y}_k, \mathbf{Y}_j)^2 + \sum_{\{k,k'\} \cap \{i,j\} = \emptyset} H(\mathbf{Y}_k, \mathbf{Y}_j)H(\mathbf{Y}_{k'}, \mathbf{Y}_j) \right\} \right] \\
 &= \frac{1}{(n-2)^3} E\{G(\mathbf{X}, \mathbf{X}')^2 H(\mathbf{Y}, \mathbf{Y}'')^2\} + \frac{n-3}{(n-2)^3} E\{G(\mathbf{X}, \mathbf{X}')^2\} E\{H(\mathbf{Y}, \mathbf{Y}')^2\}
 \end{aligned}$$

$$\begin{aligned}
 & + \frac{2(n-3)}{(n-2)^3} E\{G(\mathbf{X}, \mathbf{X}'')H(\mathbf{Y}, \mathbf{Y}''')G(\mathbf{X}', \mathbf{X}'')H(\mathbf{Y}', \mathbf{Y}''')\} \\
 & \leq O\left\{\frac{E\{G(\mathbf{X}, \mathbf{X}')^2H(\mathbf{Y}, \mathbf{Y}'')^2\} + E\{G(\mathbf{X}, \mathbf{X}')^2\}E\{H(\mathbf{Y}, \mathbf{Y}')^2\}}{n^2}\right\} \\
 & = o(\xi^2), \\
 E(I_{n,3}^2 J_{n,2}^2) & = \frac{1}{(n-2)^3} E\{G(\mathbf{X}, \mathbf{X}')^2H(\mathbf{Y}, \mathbf{Y}'')^2\} + \frac{n-3}{(n-2)^3} E\{G(\mathbf{X}, \mathbf{X}')^2\}E\{H(\mathbf{Y}, \mathbf{Y}')^2\} \\
 & + \frac{2(n-3)}{(n-2)^3} E\{G(\mathbf{X}, \mathbf{X}'')H(\mathbf{Y}, \mathbf{Y}''')G(\mathbf{X}', \mathbf{X}'')H(\mathbf{Y}', \mathbf{Y}''')\} \\
 & \leq O\left\{\frac{E\{G(\mathbf{X}, \mathbf{X}')^2H(\mathbf{Y}, \mathbf{Y}'')^2\} + E\{G(\mathbf{X}, \mathbf{X}')^2\}E\{H(\mathbf{Y}, \mathbf{Y}')^2\}}{n^2}\right\} \\
 & = o(\xi^2), \\
 E(I_{n,2}^2 J_{n,4}^2) & = \frac{1}{(n-2)^4(n-3)^2} E\left[\left\{\sum_{l \notin \{i,j\}} G(\mathbf{X}_i, \mathbf{X}_l)^2 + \sum_{\{l,l'\} \cap \{i,j\} = \emptyset} G(\mathbf{X}_i, \mathbf{X}_l)G(\mathbf{X}_i, \mathbf{X}_{l'})\right.\right. \\
 & \left.\left\{\sum_{\{k,l\} \cap \{i,j\} = \emptyset} 2H(\mathbf{Y}_k, \mathbf{Y}_l)^2 + \sum_{\{k,l,l'\} \cap \{i,j\} = \emptyset} 4H(\mathbf{Y}_k, \mathbf{Y}_l)H(\mathbf{Y}_k, \mathbf{Y}_{l'})\right.\right. \\
 & \left.\left. + \sum_{\{k,k',l,l'\} \cap \{i,j\} = \emptyset} H(\mathbf{Y}_k, \mathbf{Y}_l)H(\mathbf{Y}_{k'}, \mathbf{Y}_{l'})\right\}\right] \\
 & = \frac{1}{(n-2)^4(n-3)^2} \{4(n-2)(n-3)E(G(\mathbf{X}, \mathbf{X}')^2H(\mathbf{Y}, \mathbf{Y}'')^2) \\
 & \quad + 2(n-2)(n-3)(n-4)E(G(\mathbf{X}, \mathbf{X}')^2)E(H(\mathbf{Y}, \mathbf{Y}')^2) \\
 & \quad + 4(n-2)(n-3)E(G(\mathbf{X}, \mathbf{X}')G(\mathbf{X}, \mathbf{X}'')H(\mathbf{Y}', \mathbf{Y}'')^2) \\
 & \quad + 8(n-2)(n-3)(n-4)E\{G(\mathbf{X}, \mathbf{X}'')H(\mathbf{Y}, \mathbf{Y}''')G(\mathbf{X}', \mathbf{X}'')H(\mathbf{Y}', \mathbf{Y}''')\}\} \\
 & \leq O\left\{\frac{E(G(\mathbf{X}, \mathbf{X}')^2H(\mathbf{Y}, \mathbf{Y}'')^2)}{n^3}\right\} + O\left\{\frac{E(G(\mathbf{X}, \mathbf{X}')^2)E(H(\mathbf{Y}, \mathbf{Y}')^2)}{n^3}\right\} \\
 & = o(\xi^2),
 \end{aligned}$$

$$\begin{aligned}
 E(I_{n,3}^2 J_{n,4}^2) & = E(I_{n,4}^2 J_{n,2}^2) = E(I_{n,4}^2 J_{n,3}^2) \\
 & \leq O\left\{\frac{E(G(\mathbf{X}, \mathbf{X}')^2H(\mathbf{Y}, \mathbf{Y}'')^2)}{n^3}\right\} + O\left\{\frac{E(G(\mathbf{X}, \mathbf{X}')^2)E(H(\mathbf{Y}, \mathbf{Y}')^2)}{n^3}\right\} \\
 & = o(\xi^2),
 \end{aligned}$$

$$\begin{aligned}
 E(I_{n,4}^2 J_{n,4}^2) & = \frac{1}{(n-2)^4(n-3)^4} E\left[\left\{\sum_{\{k,l,l'\} \cap \{i,j\} = \emptyset} 4G(\mathbf{X}_k, \mathbf{X}_l)G(\mathbf{X}_k, \mathbf{X}_{l'})\right.\right. \\
 & \left.\left. + \sum_{\{k,l\} \cap \{i,j\} = \emptyset} 2G(\mathbf{X}_k, \mathbf{X}_l)^2 + \sum_{\{k,k',l,l'\} \cap \{i,j\} = \emptyset} G(\mathbf{X}_k, \mathbf{X}_l)G(\mathbf{X}_{k'}, \mathbf{X}_{l'})\right\}\right] \\
 & \left\{\sum_{\{k,l\} \cap \{i,j\} = \emptyset} 2H(\mathbf{Y}_k, \mathbf{Y}_l)^2 + \sum_{\{k,l,l'\} \cap \{i,j\} = \emptyset} 4H(\mathbf{Y}_k, \mathbf{Y}_l)H(\mathbf{Y}_k, \mathbf{Y}_{l'})\right\}
 \end{aligned}$$

$$\begin{aligned}
 & + \left. \sum_{\{k,k',l,l'\} \cap \{i,j\} = \emptyset} H(\mathbf{Y}_k, \mathbf{Y}_l) H(\mathbf{Y}_{k'}, \mathbf{Y}_{l'}) \right\} \\
 = & \frac{1}{(n-2)^4(n-3)^4} [O(n^2\zeta^2) + O\{n^3E(G(\mathbf{X}, \mathbf{X}')^2H(\mathbf{Y}, \mathbf{Y}'')^2)\} + O\{n^4E(G(\mathbf{X}, \mathbf{X}')^2)E(H(\mathbf{Y}, \mathbf{Y}')^2)\} + O\{n^3E(G(\mathbf{X}, \mathbf{X}')^2H(\mathbf{Y}, \mathbf{Y}'')H(\mathbf{Y}', \mathbf{Y}'''))\} \\
 & + O\{n^3E(H(\mathbf{Y}, \mathbf{Y}')^2G(\mathbf{X}, \mathbf{X}'')G(\mathbf{X}', \mathbf{X}'''))\} \\
 & + O\{n^3E(G(\mathbf{X}, \mathbf{X}')G(\mathbf{X}, \mathbf{X}'')H(\mathbf{Y}, \mathbf{Y}')H(\mathbf{Y}, \mathbf{Y}'''))\} \\
 & + O\{n^3E(G(\mathbf{X}, \mathbf{X}')G(\mathbf{X}, \mathbf{X}'')H(\mathbf{Y}, \mathbf{Y}')H(\mathbf{Y}', \mathbf{Y}'''))\} \\
 & + O\{n^4E(G(\mathbf{X}, \mathbf{X}')G(\mathbf{X}, \mathbf{X}'')H(\mathbf{Y}''', \mathbf{Y}')H(\mathbf{Y}''', \mathbf{Y}'''))\} \\
 & + O\{n^4E(G(\mathbf{X}, \mathbf{X}')G(\mathbf{X}'', \mathbf{X}''')H(\mathbf{Y}, \mathbf{Y}')H(\mathbf{Y}'', \mathbf{Y}'''))\} \\
 & + O\{n^4E(G(\mathbf{X}, \mathbf{X}')G(\mathbf{X}'', \mathbf{X}''')H(\mathbf{Y}, \mathbf{Y}'')H(\mathbf{Y}', \mathbf{Y}'''))\}] \\
 \leq & O\left\{\frac{\zeta^2}{n^4}\right\} + O\left\{\frac{E(G(\mathbf{X}, \mathbf{X}')^2H(\mathbf{Y}, \mathbf{Y}'')^2)}{n^4}\right\} + O\left\{\frac{E(G(\mathbf{X}, \mathbf{X}')^2)E(H(\mathbf{Y}, \mathbf{Y}')^2)}{n^4}\right\} \\
 = & o(\zeta^2).
 \end{aligned}$$

From the computations above we can conclude under assumption (A1) that

$$E\left(\frac{\zeta_n^2}{\zeta^2}\right) \rightarrow 1.$$

Using similar technique we also show under assumption (A1) that

$$\text{Var}\left(\frac{\zeta_n^2}{\zeta^2}\right) \rightarrow 0.$$

This proof is thus completed. □

Proof of Proposition 1. Note that $\mathbf{Z} = (\mathbf{X}, \mathbf{Y})^\top$ is from multivariate normal distribution $\mathcal{N}(0_{(p+q)\times 1}, \mathbf{I}_{(p+q)})$. Then we can conclude that \mathbf{X} and \mathbf{Y} are independent, $\mathbf{X} \sim \mathcal{N}(0_{p\times 1}, \mathbf{I}_p)$ and $\mathbf{Y} \sim \mathcal{N}(0_{q\times 1}, \mathbf{I}_q)$. Therefore we only need to compute ζ^2 , $E\{L_1(\mathbf{Z}, \mathbf{Z}')^2\}$ and $E\{(\mathbf{X}^\top \mathbf{X}' \mathbf{Y}^\top \mathbf{Y}')^4\}$ since $E\{(\mathbf{X}^\top \mathbf{X}' \mathbf{Y}^\top \mathbf{Y}')^4\} = E\{(\mathbf{X}^\top \mathbf{X}' \mathbf{Y}^\top \mathbf{Y}'')^4\} = E\{(\mathbf{X}^\top \mathbf{X}')^4\}E\{(\mathbf{Y}^\top \mathbf{Y}'')^4\}$. Direct calculations show that

$$\begin{aligned}
 \zeta^2 & = E\{\|\mathbf{X}\|^2\}E\{\|\mathbf{Y}\|^2\} = pq, \\
 E\{L_1(\mathbf{Z}, \mathbf{Z}')^2\} & = E\{(\mathbf{X}^\top \mathbf{X}')^2\}E\{(\mathbf{Y}^\top \mathbf{Y}')^2\} = pq, \\
 E\{(\mathbf{X}^\top \mathbf{X}' \mathbf{Y}^\top \mathbf{Y}')^4\} & = 9E\{\|\mathbf{X}\|^4\}E\{\|\mathbf{Y}\|^4\} = 9pq(p+2)(q+2).
 \end{aligned}$$

□

Proof of Theorem 2 Recall the expression of $h_4(\mathbf{Z}_1, \mathbf{Z}_2, \mathbf{Z}_3, \mathbf{Z}_4)$ in the proof of Theorem 1, and the fact that $E\{G(\mathbf{X}, \mathbf{X}')|\mathbf{X}\} = E\{G(\mathbf{X}, \mathbf{X}')|\mathbf{X}'\} = 0$ and $E\{H(\mathbf{Y}, \mathbf{Y}')|\mathbf{Y}\} = E\{H(\mathbf{Y}, \mathbf{Y}')|\mathbf{Y}'\} = 0$, we obtain the corresponding $h_1(\mathbf{Z}_1)$, $h_2(\mathbf{Z}_1, \mathbf{Z}_2)$ and $h_3(\mathbf{Z}_1,$

$\mathbf{Z}_2, \mathbf{Z}_3$) under the local alternative as follows.

$$\begin{aligned}
 h_1(\mathbf{Z}_1) &= \frac{1}{2}\{K(\mathbf{X}_1, \mathbf{Y}_1) + \text{tr}(\boldsymbol{\Sigma}_{\mathbf{XY}}\boldsymbol{\Sigma}_{\mathbf{YX}})\}, \\
 h_2(\mathbf{Z}_1, \mathbf{Z}_2) &= \frac{1}{6}\{G(\mathbf{X}_1, \mathbf{X}_2)H(\mathbf{Y}_1, \mathbf{Y}_2) + 2K(\mathbf{X}_1, \mathbf{Y}_1) + 2K(\mathbf{X}_2, \mathbf{Y}_2) \\
 &\quad - K(\mathbf{X}_2, \mathbf{Y}_1) - K(\mathbf{X}_1, \mathbf{Y}_2) + \text{tr}(\boldsymbol{\Sigma}_{\mathbf{XY}}\boldsymbol{\Sigma}_{\mathbf{YX}})\}, \\
 h_3(\mathbf{Z}_1, \mathbf{Z}_2, \mathbf{Z}_3) &= \frac{1}{12}[\{2G(\mathbf{X}_1, \mathbf{X}_2) - G(\mathbf{X}_1, \mathbf{X}_3) - G(\mathbf{X}_2, \mathbf{X}_3)\}H(\mathbf{Y}_1, \mathbf{Y}_2) \\
 &\quad + \{2G(\mathbf{X}_1, \mathbf{X}_3) - G(\mathbf{X}_1, \mathbf{X}_2) - G(\mathbf{X}_2, \mathbf{X}_3)\}H(\mathbf{Y}_1, \mathbf{Y}_3) \\
 &\quad + \{2G(\mathbf{X}_2, \mathbf{X}_3) - G(\mathbf{X}_1, \mathbf{X}_2) - G(\mathbf{X}_1, \mathbf{X}_3)\}H(\mathbf{Y}_2, \mathbf{Y}_3) \\
 &\quad + 2K(\mathbf{X}_1, \mathbf{Y}_1) - K(\mathbf{X}_2, \mathbf{Y}_1) - K(\mathbf{X}_3, \mathbf{Y}_1) + 2K(\mathbf{X}_2, \mathbf{Y}_2) \\
 &\quad - K(\mathbf{X}_1, \mathbf{Y}_2) - K(\mathbf{X}_3, \mathbf{Y}_2) + 2K(\mathbf{X}_3, \mathbf{Y}_3) - K(\mathbf{X}_1, \mathbf{Y}_3) \\
 &\quad - K(\mathbf{X}_2, \mathbf{Y}_3),
 \end{aligned}$$

where $K(\mathbf{X}_i, \mathbf{Y}_j) = \mathbf{X}_i^\top \boldsymbol{\Sigma}_{\mathbf{XY}} \mathbf{Y}_j$. Again using the Hoeffding decomposition in Serfling (1980). we can show under assumptions (A1) and (A2) that

$$\frac{n(T_{n,p,q}(\mathbf{X}, \mathbf{Y}) - \text{tr}(\boldsymbol{\Sigma}_{\mathbf{XY}}\boldsymbol{\Sigma}_{\mathbf{YX}}))}{\sqrt{2\xi^2}} = \frac{\sum_{1 \leq i < j \leq n} \tilde{L}(\mathbf{Z}_i, \mathbf{Z}_j)}{\sqrt{\binom{n}{2}\xi^2}} + o_p(1),$$

where $\tilde{L}(\mathbf{Z}_i, \mathbf{Z}_j) = L(\mathbf{Z}_i, \mathbf{Z}_j) - K(\mathbf{X}_i, \mathbf{Y}_i) - K(\mathbf{X}_j, \mathbf{Y}_j) + \text{tr}(\boldsymbol{\Sigma}_{\mathbf{XY}}\boldsymbol{\Sigma}_{\mathbf{YX}})$.

Note that $E(\tilde{L}(\mathbf{Z}, \mathbf{Z}')) = E(\tilde{L}(\mathbf{Z}, \mathbf{Z}') | \mathbf{Z}) = E(\tilde{L}(\mathbf{Z}, \mathbf{Z}') | \mathbf{Z}') = 0$. Using similar arguments by replacing $L(\mathbf{Z}_i, \mathbf{Z}_j)$ in the proof of Theorem 1 with $\tilde{L}(\mathbf{Z}_i, \mathbf{Z}_j)$, we can show that

$$\frac{\sum_{1 \leq i < j \leq n} \tilde{L}(\mathbf{Z}_i, \mathbf{Z}_j)}{\sqrt{\binom{n}{2}\xi^2}} \xrightarrow{d} \mathcal{N}(0, 1),$$

which yields that

$$\frac{n(T_{n,p,q}(\mathbf{X}, \mathbf{Y}) - \text{tr}(\boldsymbol{\Sigma}_{\mathbf{XY}}\boldsymbol{\Sigma}_{\mathbf{YX}}))}{\sqrt{2\xi^2}} \xrightarrow{d} \mathcal{N}(0, 1).$$

Applying similar technique in proving the ratio-consistency in theorem 1, it is easy to show that

$$\frac{\xi_n^2}{\xi^2} \xrightarrow{P} 1,$$

provided that conditions (A1) and (A2) hold. We thus complete this proof. □

Proof of Proposition 2. For $\|\beta\| \neq 0$, we assume that U is a $p \times p$ orthogonal matrix whose first column is $\beta/\|\beta\|$. Then $\mathbf{S} = (S_1, \dots, S_p)^\top = U^\top \mathbf{X} \sim \mathcal{N}(0_{p \times 1}, \mathbf{I}_p)$. Moreover, $\mathbf{S}' = (S'_1, \dots, S'_p)^\top = U^\top \mathbf{X}'$ and $\mathbf{S}'' = (S''_1, \dots, S''_p)^\top = U^\top \mathbf{X}''$ can be viewed as copies of \mathbf{S} . Note that $E(S_1) = 0$, $E(S_1^2) = 1$ and $E(S_1^4) = 3$. Elemental calculations show that

$$\begin{aligned} \zeta^2 &= E\{L(\mathbf{Z}, \mathbf{Z}')^2\} \\ &= E\{(\mathbf{X}^\top \mathbf{X}')^2 (\mathbf{X}^\top \beta + \varepsilon)^2 (\mathbf{X}'^\top \beta + \varepsilon')^2\} \\ &= E\{(\mathbf{X}^\top \mathbf{X}')^2 ((\mathbf{X}^\top \beta)^2 + 1) ((\mathbf{X}'^\top \beta)^2 + 1)\} \\ &= E\{(\mathbf{X}^\top \mathbf{X}')^2 ((\mathbf{X}^\top \beta)^2 (\mathbf{X}'^\top \beta)^2 + (\mathbf{X}^\top \beta)^2 + (\mathbf{X}'^\top \beta)^2 + 1)\} \\ &= E\{(\mathbf{S}^\top \mathbf{S}')^2 (\|\beta\|^4 S_1^2 S_1'^2 + \|\beta\|^2 S_1^2 + \|\beta\|^2 S_1'^2 + 1)\} \\ &= (p + 8)\|\beta\|^4 + 2(p + 2)\|\beta\|^2 + p, \\ E\{(\mathbf{X}^\top \mathbf{X}' \mathbf{Y}^\top \mathbf{Y}')^4\} &= E\{(\mathbf{X}^\top \mathbf{X}')^4 (\mathbf{X}^\top \beta + \varepsilon)^4 (\mathbf{X}'^\top \beta + \varepsilon')^4\} \\ &= E\{(\mathbf{X}^\top \mathbf{X}')^4 ((\mathbf{X}^\top \beta)^4 + 6(\mathbf{X}^\top \beta)^2 + 3) ((\mathbf{X}'^\top \beta)^4 + 6(\mathbf{X}'^\top \beta)^2 + 3)\} \\ &= E\{(\mathbf{S}^\top \mathbf{S}')^4 (\|\beta\|^4 S_1^4 + 6\|\beta\|^2 S_1^2 + 3) (\|\beta\|^4 S_1'^4 + 6\|\beta\|^2 S_1'^2 + 3)\} \\ &= O\{p^2 (\|\beta\|^2 + 1)^4\}, \\ E\{(\mathbf{X}^\top \mathbf{X}' \mathbf{Y}^\top \mathbf{Y}'')^4\} &= E\{(\mathbf{X}^\top \mathbf{X}')^4 (\mathbf{X}^\top \beta + \varepsilon)^4 (\mathbf{X}''^\top \beta + \varepsilon'')^4\} \\ &= 9(\|\beta\|^2 + 1)^2 E\{\|\mathbf{X}\|^4 ((\mathbf{X}^\top \beta)^4 + 6(\mathbf{X}^\top \beta)^2 + 3)\} \\ &= 9(\|\beta\|^2 + 1)^2 E\{\|\mathbf{S}\|^4 (\|\beta\|^4 S_1^4 + 6\|\beta\|^2 S_1^2 + 3)\} \\ &= O\{p^2 (\|\beta\|^2 + 1)^4\}, \\ E\{(\mathbf{X}^\top \mathbf{X}')^4\} E\{(\mathbf{Y}^\top \mathbf{Y}')^4\} &= E\{(\mathbf{X}^\top \mathbf{X}')^4\} E\{(\mathbf{X}^\top \beta + \varepsilon)^4\} E\{(\mathbf{X}'^\top \beta + \varepsilon')^4\} \\ &= 27(\|\beta\|^2 + 1)^4 E\{\|\mathbf{X}\|^4\} \\ &= 27p(p + 2)(\|\beta\|^2 + 1)^4. \end{aligned}$$

We then calculate $E\{(\mathbf{X}^\top \Sigma_{\mathbf{X}\mathbf{Y}} \mathbf{Y})^2\}$ and $E\{K(\mathbf{X}^\top \Sigma_{\mathbf{X}\mathbf{Y}} \mathbf{Y}')^2\}$. It is easy to show that $\mathbf{X}^\top \Sigma_{\mathbf{X}\mathbf{Y}} \mathbf{Y} = \mathbf{X}^\top \beta \mathbf{Y}$ and $\mathbf{X}^\top \Sigma_{\mathbf{X}\mathbf{Y}} \mathbf{Y}' = \mathbf{X}^\top \beta \mathbf{Y}'$. Then we can obtain that

$$\begin{aligned} E\{(\mathbf{X}^\top \Sigma_{\mathbf{X}\mathbf{Y}} \mathbf{Y})^2\} &= E\{(\mathbf{X}^\top \beta)^2 ((\mathbf{X}^\top \beta)^2 + 2\varepsilon \mathbf{X}^\top \beta + \varepsilon^2)\} \\ &= E\{(\mathbf{X}^\top \beta)^4\} + E\{(\mathbf{X}^\top \beta)^2\} \\ &= \|\beta\|^4 E(S_1^4) + \|\beta\|^2 E(S_1^2) \\ &= 3\|\beta\|^4 + \|\beta\|^2, \\ E\{(\mathbf{X}^\top \Sigma_{\mathbf{X}\mathbf{Y}} \mathbf{Y}')^2\} &= E\{(\mathbf{X}^\top \beta)^2\} E\{(\mathbf{X}'^\top \beta)^2 + 2\varepsilon' \mathbf{X}'^\top \beta + \varepsilon'^2\} \\ &= E\{(\mathbf{X}^\top \beta)^2\} E\{(\mathbf{X}'^\top \beta)^2 + 1\} \\ &= \|\beta\|^2 E(S_1^2) E(\|\beta\|^2 S_1'^2 + 1) \\ &= \|\beta\|^4 + \|\beta\|^2. \end{aligned}$$

Next we consider $E\{L_1(\mathbf{Z}, \mathbf{Z}')^2\}$. Careful calculations show that

$$\begin{aligned} E\{L_1(\mathbf{Z}, \mathbf{Z}')^2\} &= E[\{E(L(\mathbf{Z}, \mathbf{Z}'')L(\mathbf{Z}', \mathbf{Z}'')|(\mathbf{Z}, \mathbf{Z}'))\}^2] \\ &= E[\{E(\mathbf{Y}\mathbf{Y}'\mathbf{X}''^\top\mathbf{X}\mathbf{X}'^\top\mathbf{X}''\mathbf{Y}''^2|(\mathbf{X}, \mathbf{Y}, \mathbf{X}', \mathbf{Y}'))\}^2] \\ &= E[\{E(\mathbf{Y}\mathbf{Y}'\mathbf{X}''^\top\mathbf{X}\mathbf{X}'^\top\mathbf{X}''((\mathbf{X}''^\top\boldsymbol{\beta})^2 + 1)|(\mathbf{X}, \mathbf{Y}, \mathbf{X}', \mathbf{Y}'))\}^2] \\ &= E\left[(\|\boldsymbol{\beta}\|S_1 + \varepsilon)^2(\|\boldsymbol{\beta}\|S'_1 + \varepsilon')^2\right. \\ &\quad \left.\left\{(3\|\boldsymbol{\beta}\|^2 + 1)S_1S'_1 + (\|\boldsymbol{\beta}\|^2 + 1)\sum_{i=2}^p S_iS'_i\right\}^2\right] \\ &= E\left[(\|\boldsymbol{\beta}\|^4S_1^2S'^2_1 + \|\boldsymbol{\beta}\|^2S_1^2 + \|\boldsymbol{\beta}\|^2S'^2_1 + 1)\right. \\ &\quad \left.\left\{(3\|\boldsymbol{\beta}\|^2 + 1)S_1S'_1 + (\|\boldsymbol{\beta}\|^2 + 1)\sum_{i=2}^p S_iS'_i\right\}^2\right] \\ &= (3\|\boldsymbol{\beta}\|^2 + 1)^4 + (p - 1)(\|\boldsymbol{\beta}\|^2 + 1)^4. \end{aligned}$$

Thus this proof is completed. □

Proof of Theorem 3 (1) Let $\lambda_{n,p_n,q_n} = n(p_nq_n)^{-1/4}(\log n)^{-3/4}$. It suffices to show that, for any $\varepsilon > 0$, almost surely $\lambda_{n,p_n,q_n} \max_{1 \leq i \leq p_n, 1 \leq j \leq q_n} |\mathcal{R}_n(X_i, Y_j)| < \varepsilon$ for all n sufficiently large, that is,

$$\Pr\left(\lambda_{n,p_n,q_n} \max_{1 \leq i \leq p_n, 1 \leq j \leq q_n} |\mathcal{R}_n(X_i, Y_j)| > \varepsilon \text{ for infinitely many } n\right) = 0.$$

Applying the Borel–Cantelli lemma, it suffices to show that

$$\sum_{k=0}^{\infty} \Pr\left(\max_{2^k \leq n \leq 2^{k+1}} \left(\lambda_{n,p_n,q_n} \max_{1 \leq i \leq p_n, 1 \leq j \leq q_n} |\mathcal{R}_n(X_i, Y_j)|\right) > \varepsilon\right) < \infty.$$

Notice that $T_{n,1,1}(X_i, X_i)$ and $T_{n,1,1}(Y_j, Y_j)$ are U-statistics of order 4. By the Theorem 5.4.C in Serfling (1980), it is easy to check that for each $i, 1 \leq i \leq p, 1 \leq j \leq q$, almost surely

$$\frac{T_{n,1,1}(X_i, X_i)}{\sigma^2(X_i, X_i)} = 1 + O\left(\frac{1}{\sqrt{n}}\right), \quad \frac{T_{n,1,1}(Y_j, Y_j)}{\sigma^2(Y_j, Y_j)} = 1 + O\left(\frac{1}{\sqrt{n}}\right),$$

provided the assumption (A3) holds. In addition, under H_0 , $E\{G(X_i, X'_i)H(Y_j, Y'_j)\} = K(X_i, Y_j) = 0$. We can obtain from a standard result in section 32 of Loève (1978) and Markov’s inequality that

$$\begin{aligned} & \sum_{k=0}^{\infty} \Pr \left(\max_{2^k \leq n \leq 2^{k+1}} (\lambda_{n, p_n, q_n} \max_{1 \leq i \leq p_n, 1 \leq j \leq q_n} |\mathcal{R}_n(X_i, Y_j)|) > \varepsilon \right) \\ & \leq \sum_{k=0}^{\infty} \sum_{i=1}^{p_{2^{k+1}}} \sum_{j=1}^{q_{2^{k+1}}} \Pr \left(\max_{2^k \leq n \leq 2^{k+1}} (\lambda_{n, p_n, q_n} |\mathcal{R}_n(X_i, Y_j)|) > \varepsilon \right) \\ & \leq \sum_{k=0}^{\infty} \sum_{i=1}^{p_{2^{k+1}}} \sum_{j=1}^{q_{2^{k+1}}} \varepsilon^{-4} \frac{2^{4(k+1)}}{(k+1)^3 p_{2^{k+1}} q_{2^{k+1}} (\log 2)^3} \frac{E(|T_{2^k, 1, 1}(X_i, Y_j)|^4)}{\sigma^4(X_i, X_i) \sigma^4(Y_j, Y_j)} \\ & = \sum_{k=0}^{\infty} O((k+1)^{-3}), \end{aligned}$$

where the last equality follows from Theorem 5.3.2 in Serfling (1980). Thus the result in (1) can be obtained.

(2) We need to show that almost surely

$$\Pr \left(\max_{1 \leq i \leq p, 1 \leq j \leq q} |\mathcal{R}_n(X_i, Y_j)| < \delta_{n, p, q} | S \neq \emptyset \right) \rightarrow 0, \quad n \rightarrow \infty. \tag{43}$$

To obtain the result, it is sufficient to prove that

$$\Pr \left(\max_{1 \leq i \leq p, 1 \leq j \leq q} \frac{|T_{n, 1, 1}(X_i, Y_j) - \sigma^2(X_i, Y_j)|}{\sqrt{\sigma^2(X_i, X_i) \sigma^2(Y_j, Y_j)}} > \delta_{n, p, q} | S \neq \emptyset \right) \rightarrow 0,$$

since

$$\begin{aligned} & \Pr \left(\max_{1 \leq i \leq p, 1 \leq j \leq q} |\mathcal{R}_n(X_i, Y_j)| < \delta_{n, p, q} | S \neq \emptyset \right) \\ & \leq \Pr \left(\max_{1 \leq i \leq p, 1 \leq j \leq q} \frac{|T_{n, 1, 1}(X_i, Y_j) - \sigma^2(X_i, Y_j)|}{\sqrt{\sigma^2(X_i, X_i) \sigma^2(Y_j, Y_j)}} > \delta_{n, p, q} | S \neq \emptyset \right). \end{aligned}$$

Write the projection of U-statistic $T_{n, 1, 1}(X_i, Y_j)$ as

$$\widehat{T}_{n, 1, 1}(X_i, Y_j) = \sum_{t=1}^n E\{T_{n, 1, 1}(X_i, Y_j) | (X_{ti}, Y_{tj})\} - (n-1)\sigma^2(X_i, Y_j).$$

Thus we obtain that

$$\begin{aligned}
 & T_{n,1,1}(X_i, Y_j) - \sigma^2(X_i, Y_j) \\
 &= \frac{4}{n} \sum_{t=1}^n \left\{ K(X_{ti}, Y_{tj}) - \sigma^2(X_i, Y_j) \right\} + (T_{n,1,1}(X_i, Y_j) - \widehat{T}_{n,1,1}(X_i, Y_j)),
 \end{aligned}$$

where $K(X_{ti}, Y_{tj}) = \sigma(X_i, Y_j)X_{ti}Y_{tj}$. Similar to prove (1), it is easy to check that almost surely

$$\max_{1 \leq i \leq p, 1 \leq j \leq q} \frac{|T_{n,1,1}(X_i, Y_j) - \widehat{T}_{n,1,1}(X_i, Y_j)|}{\sqrt{\sigma^2(X_i, X_i)\sigma^2(Y_j, Y_j)}} = o(\delta_{n,p,q}), \quad n \rightarrow \infty.$$

Hence, we only need to show that

$$\Pr \left(\max_{1 \leq i \leq p, 1 \leq j \leq q} \left| \frac{4}{n} \sum_{t=1}^n \frac{K(X_{ti}, Y_{tj}) - \sigma^2(X_i, Y_j)}{\sqrt{\sigma^2(X_i, X_i)\sigma^2(Y_j, Y_j)}} \right| > \delta_{n,p,q} \mid S \neq \emptyset \right) \rightarrow 0.$$

Similarly, we have that

$$\begin{aligned}
 & \sum_{k=0}^{\infty} \Pr \left(\max_{2^k \leq n \leq 2^{k+1}} (\lambda_{n,p_n,q_n} \max_{1 \leq i \leq p_n, 1 \leq j \leq q_n} \left| \frac{4}{n} \sum_{t=1}^n \frac{K(X_{ti}, Y_{tj}) - \sigma^2(X_i, Y_j)}{\sqrt{\sigma^2(X_i, X_i)\sigma^2(Y_j, Y_j)}} \right|) > \varepsilon \right) \\
 & \leq \sum_{k=0}^{\infty} \sum_{i=1}^{p_{2^{k+1}}} \sum_{j=1}^{q_{2^{k+1}}} \varepsilon^{-2} \frac{2^{2(k+1)}}{(k+1)^{3/2}(p_{2^{k+1}}q_{2^{k+1}})^{1/2}(\log 2)^{3/2}} \max_{1 \leq i \leq p_{2^{k+1}}, 1 \leq j \leq q_{2^{k+1}}} \xi_{ij} \\
 & = \sum_{k=0}^{\infty} O((k+1)^{-3/2}),
 \end{aligned}$$

where the last equality is true when (A4) holds. Therefore, we have almost surely that

$$\max_{1 \leq i \leq p, 1 \leq j \leq q} \frac{|T_{n,1,1}(X_i, Y_j) - \sigma^2(X_i, Y_j)|}{\sqrt{\sigma^2(X_i, X_i)\sigma^2(Y_j, Y_j)}} = o(\delta_{n,p,q}), \quad n \rightarrow \infty.$$

This proof is thus completed. □

References

Anderson T (2003) An introduction to multivariate statistical analysis, 3rd edn. Wiley, New York
 Chiang AP, Beck JS, Yen HJ, Tayeh MK, Scheetz TE, Swiderski RE, Sheffield VC (2006) Homozygosity mapping with SNP arrays identifies TRIM32, an E3 ubiquitin ligase, as a BardetCiebl syndrome gene (BBS11). Proc Natl Acad Sci 103(16):6287–6292

- Cui H, Guo W, Zhong W (2018) Test for high-dimensional regression coefficients using refitted cross-validation variance estimation. *Ann Stat* 46(3):958–988
- Escoufier Y (1973) Le traitement des variables vectorielles. *Biometrics* 29(4):751–760
- Fan J, Liao Y, Yao J (2015) Power enhancement in high-dimensional cross-sectional tests. *Econometrica* 83(4):1497–1541
- Hall P, Heyde CC (1980) Martingale limit theory and its application. Academic Press, New York
- Heller R, Heller Y, Gorfine M (2012) A consistent multivariate test of association based on ranks of distances. *Biometrika* 100(2):503–510
- Huang J, Ma S, Zhang CH (2008) Adaptive Lasso for sparse high-dimensional regression models. *Stat Sin* 18:1603–1618
- Jiang D, Bai Z, Zheng S (2013) Testing the independence of sets of large-dimensional variables. *Sci China Math* 56(1):135–147
- Li W, Chen J, Yao J (2017) Testing the independence of two random vectors where only one dimension is large. *Stat* 51(1):141–153
- Loève M (1978) Probability theory II, 4th edn. Springer, New York
- Robert P, Cléroux R, Ranger N (1985) Some results on vector correlation. *Comput Stat Data Anal* 3:25–32
- Scheetz TE, Kim KYA, Swiderski RE, Philp AR, Braun TA, Knudtson KL, Sheffield VC (2006) Regulation of gene expression in the mammalian eye and its relevance to eye disease. *Proc Natl Acad Sci* 103(39):14429–14434
- Serfling RJ (1980) Approximation theorems of mathematical statistics. Wiley, New York
- Shao X, Zhang J (2014) Martingale difference correlation and its use in high-dimensional variable screening. *J Am Stat Assoc* 109(507):1302–1318
- Srivastava MS, Reid N (2012) Testing the structure of the covariance matrix with fewer observations than the dimension. *J Multivar Anal* 112:156–171
- Székely GJ, Rizzo ML (2013) The distance correlation t-test of independence in high dimension. *J Multivar Anal* 117:193–213
- Székely GJ, Rizzo ML (2014) Partial distance correlation with methods for dissimilarities. *Ann Stat* 42(6):2382–2412
- Székely GJ, Rizzo ML, Bakirov NK (2007) Measuring and testing independence by correlation of distances. *Ann Stat* 35(6):2769–2794
- Taskinen S, Oja H, Randles RH (2005) Multivariate nonparametric tests of independence. *J Am Stat Assoc* 100(471):916–925
- Wilks SS (1935) On the independence of k sets of normally distributed statistical variables. *Econometrica* 3(3):309–326
- Yang Y, Pan G (2015) Independence test for high dimensional data based on regularized canonical correlation coefficients. *Ann Stat* 43(2):467–500
- Yao S, Zhang X, Shao X (2018) Testing mutual independence in high dimension via distance covariance. *J R Stat Soc B* 80(3):455–480
- Zhang X, Yao S, Shao X (2018) Conditional mean and quantile dependence testing in high dimension. *Ann Stat* 46(1):219–246
- Zheng S, He X, Guo J (2022) Hypothesis testing for block-structured correlation. *Stat Sin* 32(2):719–735
- Zhong PS, Chen SX (2011) Tests for high-dimensional regression coefficients with factorial designs. *J Am Stat Assoc* 106(493):260–274

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.