An approximate randomization test for high-dimensional two-sample Behrens-Fisher problem under arbitrary covariances

Rui Wang and Wangli Xu*

Center for Applied Statistics and School of Statistics, Renmin University of China, Beijing 100872, China

Tuesday 23rd August, 2022

Abstract

This paper is concerned with the problem of comparing the population means of two groups of independent observations. An approximate randomization test procedure based on the test statistic of Chen and Qin (2010) is proposed. The asymptotic behavior of the test statistic as well as the randomized statistic is studied under weak conditions. In our theoretical framework, observations are not assumed to be identically distributed even within groups. No condition on the eigenstructure of the covariance matrices is imposed. And the sample sizes of the two groups are allowed to be unbalanced. Under general conditions, all possible asymptotic distributions of the test statistic are obtained. We derive the asymptotic level and local power of the approximate randomization test procedure. Our theoretical results show that the proposed test procedure can adapt to all possible asymptotic distributions of the test statistic and always has correct test level asymptotically. Also, the proposed test procedure has good power behavior. Our numerical experiments show that the proposed test procedure has favorable performance compared with several alternative test procedures.

Key words: Behrens-Fisher problem; High-dimensional data; Randomization test; Lindeberg principle.

1 Introduction

Two-sample mean testing is a fundamental problem in statistics with an enormous range of applications. In modern statistical applications, high-dimensional data, where the data dimension may be much larger than the sample size, is ubiquitous. However, most classical two-sample mean tests are designed for low-dimensional data, and may not be feasible, or may have suboptimal power, for high-dimensional data; see, e.g., Bai and Saranadasa (1996). In recent years, the study of high-dimensional two-sample mean tests has attracted increasing attention.

Suppose that $X_{k,i}$, $i = 1, \ldots, n_k$, $k = 1, 2$, are independent p-dimensional random vectors with $\mathrm{E}(X_{k,i}) = \mu_k$, $\mathrm{var}(X_{k,i}) = \mathbf{\Sigma}_{k,i}$. The hypothesis of interest is

$$\mathcal{H}_0 : \mu_1 = \mu_2 \quad \text{v.s.} \quad \mathcal{H}_1 : \mu_1 \neq \mu_2. \tag{1}$$

*Corresponding author
Email address: wlxu@ruc.edu.cn

Denote by $\bar{\mathbf{\Sigma}}_k = n_k^{-1} \sum_{i=1}^{n_k} \mathbf{\Sigma}_{k,i}$ the average covariance matrix within group k, $k = 1, 2$. Most existing methods on two-sample mean tests assumed that the observations within groups are identically distributed. In this case, $\mathbf{\Sigma}_{k,i} = \bar{\mathbf{\Sigma}}_k$, $i = 1, \ldots, n_k$, $k = 1, 2$, and the considered testing problem reduces to the well-known Behrens-Fisher problem. In this paper, we consider the more general setting where the observations are allowed to have different distributions even within groups. Also, we consider the case where the data is possibly high-dimensional, that is, the data dimension p may be much larger than the sample size n_k, $k = 1, 2$.

Let $\bar{X}_k = n_k^{-1} \sum_{i=1}^{n_k} X_{k,i}$ and $\mathbf{S}_k = (n_k - 1)^{-1} \sum_{i=1}^{n_k} (X_{k,i} - \bar{X}_k)(X_{k,i} - \bar{X}_k)^\intercal$ denote the sample mean vector and the sample covariance matrix of group k, respectively, $k = 1, 2$. Denote $n = n_1 + n_2$. A classical test statistic for hypothesis (1) is Hotelling's T^2 statistic, defined as

$$\frac{n_1 n_2}{n} (\bar{X}_1 - \bar{X}_2)^\intercal \mathbf{S}^{-1} (\bar{X}_1 - \bar{X}_2),$$

where $\mathbf{S} = (n-2)^{-1}\{(n_1 - 1)\mathbf{S}_1 + (n_2 - 1)\mathbf{S}_2\}$ is the pooled sample covariance matrix. When $p > n - 2$, the matrix $\mathbf{S}$ is not invertible, and consequently Hotelling's T^2 statistic is not well-defined. In a seminal work, Bai and Saranadasa (1996) proposed the test statistic

$$T_{\mathrm{BS}}(\mathbf{X}_1, \mathbf{X}_2) = \|\bar{X}_1 - \bar{X}_2\|^2 - \frac{n}{n_1 n_2} \operatorname{tr}(\mathbf{S}),$$

where $\mathbf{X}_k = (X_{k,1}, \ldots, X_{k,n_k})^\intercal$ is the data matrix of group k, $k = 1, 2$. The test statistic $T_{\mathrm{BS}}(\mathbf{X}_1, \mathbf{X}_2)$ is well-defined for arbitrary p. Bai and Saranadasa (1996) assumed that the observations within groups are identically distributed and $\bar{\mathbf{\Sigma}}_1 = \bar{\mathbf{\Sigma}}_2$. The main term of $T_{\mathrm{BS}}(\mathbf{X}_1, \mathbf{X}_2)$ is $\|\bar{X}_1 - \bar{X}_2\|^2$. Chen and Qin (2010) removed terms $X_{k,i}^\intercal X_{k,i}$ from $\|\bar{X}_1 - \bar{X}_2\|^2$ and proposed the test statistic

$$T_{\mathrm{CQ}}(\mathbf{X}_1, \mathbf{X}_2) = \sum_{k=1}^{2} \frac{2 \sum_{i=1}^{n_k} \sum_{j=i+1}^{n_k} X_{k,i}^\intercal X_{k,j}}{n_k(n_k - 1)} - \frac{2 \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} X_{1,i}^\intercal X_{2,j}}{n_1 n_2}.$$

The test method of Chen and Qin (2010) can be applied when $\bar{\mathbf{\Sigma}}_1 \neq \bar{\mathbf{\Sigma}}_2$. Both Bai and Saranadasa (1996) and Chen and Qin (2010) used the asymptotical normality of the statistics to determine the critical value of the test. However, the asymptotical normality only holds for a restricted class of covariance structures. For example, Chen and Qin (2010) assumed that

$$\operatorname{tr}(\bar{\mathbf{\Sigma}}_k^4) = o\left[[\operatorname{tr}\{(\bar{\mathbf{\Sigma}}_1 + \bar{\mathbf{\Sigma}}_2)^2\}]^2\right], \quad k = 1, 2. \tag{2}$$

However, (2) may not hold when $\bar{\mathbf{\Sigma}}_k$ has a few eigenvalues significantly larger than the others, excluding some important senarios in practice. For example, (2) is often violated when the variables are affected by some common factors; see, e.g., Fan et al. (2021) and the references therein. If the condition (2) is violated, the test procedures of Bai and Saranadasa (1996) and Chen and Qin (2010) may have incorrect test level; see, e.g., Wang and Xu (2019). For years, how to construct test procedures that are valid under general covariances has been an important open problem in the field of high-dimensional hypothesis testing; see the review paper of Hu and Bai (2016).

Intuitively, one may resort to the bootstrap method to control the test level under general covariances. Surprisingly, as shown by our numerical experiments, the empirical bootstrap method does not work well for $T_{\mathrm{CQ}}(\mathbf{X}_1, \mathbf{X}_2)$. Also, the wild bootstrap method, which is popular in high-dimensional statistics, does not work well either. We will give heuristic arguments to understand this phenomenon in Section 3. Recently, several methods were proposed to give a better control of the test level of $T_{\mathrm{CQ}}(\mathbf{X}_1, \mathbf{X}_2)$ and related test statistics. In the setting of $\bar{\mathbf{\Sigma}}_1 = \bar{\mathbf{\Sigma}}_2$, Zhang et al. (2020) considered the statistic $\|\bar{X}_1 - \bar{X}_2\|^2$ and proposed to use the Welch-Satterthwaite χ^2-approximation to determine the critical value. Subsequently, Zhang et al. (2021) extended the test procedure of Zhang et al. (2020) to the two-sample Behrens-Fisher problem, and Zhang and Zhu (2021) considered a χ^2-approximation of $T_{\text{CQ}}(\mathbf{X}_1, \mathbf{X}_2)$. However, the test methods of Zhang et al. (2020), Zhang et al. (2021) and Zhang and Zhu (2021) still require certain conditions on the eigenstructure of $\bar{\mathbf{\Sigma}}_k$, $k = 1, 2$. In another work, Wu et al. (2018) investigated the general distributional behavior of $\|\bar{X}_1\|^2$ for one-sample mean testing problem. They proposed a half-sampling procedure to determine the critical value of the test statistic. A generalization of the method of Wu et al. (2018) to the statistc $\|\bar{X}_1 - \bar{X}_2\|^2$ for the two-sample Behrens-Fisher problem when both n_1 and n_2 are even numbers was presented in Chapter 2 of Lou (2020). Unfortunately, Wu et al. (2018) and Lou (2020) did not include detailed proofs of their theoretical results. Recently, Wang and Xu (2019) considered a randomization test based on the test statistic of Chen and Qin (2010) for the one-sample mean testing problem. Their randomization test has exact test level if the distributions of the observations are symmetric and has asymptotically correct level in the general setting. Although many efforts have been made, no existing test procedure based on $T_{\text{CQ}}(\mathbf{X}_1, \mathbf{X}_2)$ is valid for arbitrary $\bar{\mathbf{\Sigma}}_k$ with rigorous theoretical guarantees.

The statistics $T_{\text{BS}}(\mathbf{X}_1, \mathbf{X}_2)$ and $T_{\text{CQ}}(\mathbf{X}_1, \mathbf{X}_2)$ are based on sum-of-squares. Some researchers investigated variants of sum-of-squares statistics that are scalar-invariant; see, e.g., Srivastava and Du (2008), Srivastava et al. (2013) and Feng et al. (2015). These test methods also impose some nontrivial assumptions on the eigenstructure of the covariance matrices. There is another line of research initiated by Cai et al. (2013) which utilizes extreme value type statistics to test hypothesis (1). Chang et al. (2017) considered a parametric bootstrap method to determine the critical value of the extreme value type statistics. The method of Chang et al. (2017) allows for general covariance structures. Recently, Xue and Yao (2020) investigated a wild bootstrap method. The test procedure in Xue and Yao (2020) does not require that the observations within groups are identically distributed. The theoretical analyses in Chang et al. (2017) and Xue and Yao (2020) are based on recent results about high-dimensional bootstrap methods developed by Chernozhukov et al. (2013) and Chernozhukov et al. (2017). While their theoretical framework can be applied to extreme value type statistics, it can not be applied to sum-of-squares type statistics like $T_{\text{CQ}}(\mathbf{X}_1, \mathbf{X}_2)$. Indeed, we shall see that the wild bootstrap method does not work well for $T_{\text{CQ}}(\mathbf{X}_1, \mathbf{X}_2)$.

In the present paper, we propose a new test procedure based on $T_{\text{CQ}}(\mathbf{X}_1, \mathbf{X}_2)$. In the proposed test procedure, we use a new randomization method to determine the critical value of the test statistic. The proposed randomization method is motivated by the randomization method proposed in Fisher (1935), Section 21. While this randomization method is widely used in one-sample mean testing problem, it can not be directly applied to two-sample Behrens-Fisher problem. In comparison, the proposed randomization method has satisfactory performance for two-sample Behrens-Fisher problem. We investigate the asymptotic properties of the proposed test procedure and rigorously prove that it has correct test level asymptotically under fairly weak conditions. Compared with existing test procedures based on sum-of-squares, the present work has the following favorable features. First, the proposed test procedure has correct test level asymptotically without any restriction on $\bar{\mathbf{\Sigma}}_k$, $k = 1, 2$. As a consequence, the proposed test procedure serves as a rigorous solution to the open problem that how to construct a valid test based on $T_{\text{CQ}}(\mathbf{X}_1, \mathbf{X}_2)$ under general covariances. Second, the proposed test procedure is valid even if the observations within groups are not identically distributed. To the best of our knowledge, the only existing method that can work in such a general setting is the test procedure in Xue and Yao (2020). However, the test procedure in Xue and Yao (2020) is based on extreme value statistic, which is different from the present work. Third, our theoretical results are valid for arbitrary n_1 and n_2 as long as $\min(n_1, n_2) \to \infty$. In comparison, all the above mentioned methods only considered the balanced data, that is, the sample sizes n_1 and n_2 are of the same order. We also derive the asymptotic local power of the proposed method. It shows that the asymptotic local power of the proposed test procedure is the same as that of the oracle test procedure where the distribution of the test statistic is known. A key tool for the proofs of our theoretical results is a universality property of the generalized quadratic forms. This result may be of independent interest. We also conduct numerical experiments to compare the proposed test procedure with some existing methods. Our theoretical and numerical results show that the proposed test procedure has good performance in both level and power.

2 Asymptotics of $T_{\mathrm{CQ}}(\mathbf{X}_1, \mathbf{X}_2)$

In this section, we investigate the asymptotic behavior of $T_{\mathrm{CQ}}(\mathbf{X}_1, \mathbf{X}_2)$ under general conditions. We begin with some notations. For two random variables ξ and ζ, let $\mathcal{L}(\xi)$ denote the distribution of ξ, and $\mathcal{L}(\xi \mid \zeta)$ denote the conditional distribution of ξ given ζ. For two probability measures ν_1 and ν_2 on $\mathbb{R}$, let $\nu_1 - \nu_2$ be the signed measure such that for any Borel set $\mathcal{A}$, $(\nu_1 - \nu_2)(\mathcal{A}) = \nu_1(\mathcal{A}) - \nu_2(\mathcal{A})$. Let $\mathscr{C}_b^3(\mathbb{R})$ denote the class of bounded functions with bounded and continuous derivatives up to order 3. It is known that a sequence of random variables $\{\xi_n\}_{n=1}^{\infty}$ converges weakly to a random variable ξ if and only if for every $f \in \mathscr{C}_b^3(\mathbb{R})$, $\mathrm{E} f(\xi_n) \to \mathrm{E} f(\xi)$; see, e.g., Pollard (1984), Chapter III, Theorem 12. We use this property to give a metrization of the weak convergence in $\mathbb{R}$. For a function $f \in \mathscr{C}_b^3(\mathbb{R})$, let $f^{(i)}$ denote the ith derivative of f, $i = 1, 2, 3$. For a finite signed measure ν on $\mathbb{R}$, we define the norm $\|\nu\|_3$ as $\sup_f \left|\int_{\mathbb{R}^d} f(\mathbf{x})\, \nu(\mathrm{d}\mathbf{x})\right|$, where the supremum is taken over all $f \in \mathscr{C}_b^3(\mathbb{R})$ such that $\sup_{\mathbf{x} \in \mathbb{R}} |f(x)| \leq 1$ and $\sup_{\mathbf{x} \in \mathbb{R}} |f^{(\ell)}(x)| \leq 1$, $\ell = 1, 2, 3$. It is straightforward to verify that $\|\cdot\|_3$ is indeed a norm for finite signed measures. Also, a sequence of probability measures $\{\nu_n\}_{n=1}^{\infty}$ converges weakly to a probability measure ν if and only if $\|\nu_n - \nu\|_3 \to 0$; see, e.g., Dudley (2002), Corollary 11.3.4.

The test procedure of Chen and Qin (2010) determines the critical value based on the asymptotic normality of $T_{\mathrm{CQ}}(\mathbf{X}_1, \mathbf{X}_2)$ under the null hypothesis. Define $Y_{k,i} = X_{k,i} - \mu_k$, $i = 1, \ldots, n_k$, $k = 1, 2$. Then $\mathrm{E}(Y_{k,i}) = \mathbf{0}_p$. Under the null hypothesis, we have $T_{\mathrm{CQ}}(\mathbf{X}_1, \mathbf{X}_2) = T_{\mathrm{CQ}}(\mathbf{Y}_1, \mathbf{Y}_2)$ where $\mathbf{Y}_k = (Y_{k,1}, \ldots, Y_{k,n_k})^\intercal$, $k = 1, 2$. Denote by $\sigma_{T,n}^2$ the variance of $T_{\mathrm{CQ}}(\mathbf{Y}_1, \mathbf{Y}_2)$. Under certain conditions, Chen and Qin (2010) proved that $T_{\mathrm{CQ}}(\mathbf{Y}_1, \mathbf{Y}_2)/\sigma_{T,n}$ converges weakly to $\mathcal{N}(0, 1)$. They reject the null hypothesis if $T_{\mathrm{CQ}}(\mathbf{X}_1, \mathbf{X}_2)/\hat{\sigma}_{T,n}$ is greater than the $1 - \alpha$ quantile of $\mathcal{N}(0, 1)$ where $\hat{\sigma}_{T,n}$ is an estimate of $\sigma_{T,n}$ and $\alpha \in (0, 1)$ is the test level. However, we shall see that normal distribution is just one of the possible asymptotic distributions of $T_{\mathrm{CQ}}(\mathbf{Y}_1, \mathbf{Y}_2)$ in the general setting.

Now we list the main conditions imposed by Chen and Qin (2010) to prove the asymptotic normality of $T_{\mathrm{CQ}}(\mathbf{Y}_1, \mathbf{Y}_2)$. First, Chen and Qin (2010) imposed the condition (2) on the eigenstructure of $\bar{\mathbf{\Sigma}}_k$, $k = 1, 2$. Second, Chen and Qin (2010) assumed that as $n \to \infty$,

$$n_1/n \to b \in (0, 1). \tag{3}$$

That is, Chen and Qin (2010) assumed the sample sizes in two groups are balanced. This condition is commonly adopted by existing test procedures for hypothesis (1). Third, Chen and Qin (2010) assumed the general multivariate model

$$Y_{k,i} = \mathbf{\Gamma}_k Z_{k,i}, \quad i = 1, \ldots, n_k, \quad k = 1, 2, \tag{4}$$

where $\mathbf{\Gamma}_k$ is a $p \times m$ matrix for some $m \geq p$, $k = 1, 2$, and $\{Z_{k,i}\}_{i=1}^{n_k}$ are m-dimensional independent and identically distributed random vectors such that $\mathrm{E}(Z_{k,i}) = \mathbf{0}_m$, $\mathrm{var}(Z_{k,i}) = \mathbf{I}_m$, and the elements of $Z_{k,i} = (z_{k,i,1}, \ldots, z_{k,i,m})^\intercal$ satisfy $\mathrm{E}(z_{k,i,j}^4) = 3 + \Delta < \infty$, and $\mathrm{E}(z_{k,i,\ell_1}^{\alpha_1} z_{k,i,\ell_2}^{\alpha_2} \cdots z_{k,i,\ell_q}^{\alpha_q}) = \mathrm{E}(z_{k,i,\ell_1}^{\alpha_1})\, \mathrm{E}(z_{k,i,\ell_2}^{\alpha_2}) \cdots \mathrm{E}(z_{k,i,\ell_q}^{\alpha_q})$ for a positive integer q such that $\sum_{\ell=1}^{q} \alpha_\ell \leq 8$ and distinct $\ell_1, \ldots, \ell_q \in$

$\{1,\ldots,m\}$. We note that the general multivariate model (4) assumes that the observations within group k are identically distributed, $k = 1, 2$. Also, the m elements $z_{k,i,1},\ldots,z_{k,i,m}$ of $Z_{k,i}$ have finite moments up to order 8 and behave as if they are independent.

As we have noted, the condition (2) can be violated when the variables are affected by some common factors. The condition (3) is assumed by most existing test methods. However, this condition is not reasonable if the sample sizes are unbalanced. The general multivariate model (4) is commonly adopted by many high-dimensional test methods; see, e.g., Chen and Qin (2010), Zhang et al. (2020) and Zhang et al. (2021). However, the conditions on $Z_{k,i}$ may be difficult to verify and are not generally satisfied by the elliptical symmetric distributions. Also, the model (4) is not valid if the observations within groups are not identically distributed. Thus, we would like to investigate the asymptotic behavior of $T_{\mathrm{CQ}}(\mathbf{X}_1,\mathbf{X}_2)$ beyond the conditions (2), (3) and (4). We consider the asymptotic setting where $n\to\infty$ and all quantities except for absolute constants are indexed by n, a subscript we often suppress. We make the following assumption on n_1 and n_2.

Assumption 1. *Suppose as* $n\to\infty$, $\min(n_1,n_2)\to\infty$.

Assumption 1 only requires that both n_1 and n_2 tend to infinity, which allows for the unbalanced sample sizes. This relaxes the condition (3). We make the following assumption on the distributions of $Y_{k,i}$, $i = 1,\ldots,n_k$, $k = 1, 2$.

Assumption 2. *Assume there exists an absolute constant* $\tau\geq 3$ *such that for any* $p\times p$ *positive semi-definite matrix* $\mathbf{B}$,

$$\mathrm{E}\left\{(Y_{k,i}^{\intercal}\mathbf{B}Y_{k,i})^2\right\}\leq\tau\left\{\mathrm{E}(Y_{k,i}^{\intercal}\mathbf{B}Y_{k,i})\right\}^2<\infty,\quad i=1,\ldots,n_k,\quad k=1,2.$$

Intuitively, Assumption 2 requires that the fourth moments of $Y_{k,i}$ are of the same order as the squared second moments of $Y_{k,i}$. We shall see that this assumption is fairly weak. However, the above inequality is required to be satisfied for all positive semi-definite matrix $\mathbf{B}$, which may not be straightforward to check in some cases. The following lemma gives a sufficient condition of Assumption 2.

Lemma 1. *Suppose* $Y_{k,i}=\mathbf{\Gamma}_{k,i}Z_{k,i}$, $i=1,\ldots,n_k$, $k=1,2$, *where* $\mathbf{\Gamma}_{k,i}$ *is an arbitrary* $p\times m_{k,i}$ *matrix and* $m_{k,i}$ *is an arbitrary positive integer. Suppose* $\mathrm{E}(Z_{k,i})=\mathbf{0}_{m_{k,i}}$, $\mathrm{var}(Z_{k,i})=\mathbf{I}_{m_{k,i}}$, *and the elements of* $Z_{k,i}=(z_{k,i,1},\ldots,z_{k,i,m_{k,i}})^{\intercal}$ *satisfy* $\mathrm{E}(z_{k,i,j}^4)\leq C<\infty$ *where* C *is an absolute constant. Suppose for any distinct* $\ell_1,\ell_2,\ell_3,\ell_4\in\{1,\ldots,m_{k,i}\}$,

$$\mathrm{E}(z_{k,i,\ell_1}z_{k,i,\ell_2}z_{k,i,\ell_3}z_{k,i,\ell_4})=0,\quad \mathrm{E}(z_{k,i,\ell_1}z_{k,i,\ell_2}z_{k,i,\ell_3}^2)=0,\quad \mathrm{E}(z_{k,i,\ell_1}z_{k,i,\ell_2}^3)=0. \tag{5}$$

Then Assumption 2 holds with $\tau=3C$.

It can be seen that the conditions of Lemma 1 is strictly weaker than the multivariate model (4). In fact, Lemma 1 does not require $m_{k,i}\geq p$, nor does it require the finiteness of 8th moments of $z_{k,i,j}$. Also, the moment conditions in (5) are much weaker than that required by the multivariate model (4). In addition to the multivariate model (4), the conditions of Lemma 1 also allow for an important class of elliptical symmetric distributions. In fact, if $Y_{k,i}$ has an elliptical symmetric distribution, then $Y_{k,i}$ can be written as $Y_{k,i}=\eta_{k,i}\mathbf{\Gamma}_{k,i}U_{k,i}$, where $\mathbf{\Gamma}_{k,i}$ is a $p\times m_{k,i}$ matrix, $U_{k,i}$ is a random vector distributed uniformly on the unit sphere in $\mathbb{R}^{m_{k,i}}$, and $\eta_{k,i}$ is a nonnegative random variable which is independent of $U_{k,i}$; see, e.g., Fang et al. (1990), Theorem 2.5. Suppose there is an absolute constant C such that $\mathrm{E}(\eta_{k,i}^4)\leq C\{\mathrm{E}(\eta_{k,i}^2)\}^2<\infty$. Then from the symmetric property of $U_{k,i}$ and the independence of $\eta_{k,i}$ and $U_{k,i}$, the conditions of Lemma

1 hold. In comparison, the multivariate model (4) does not allow for elliptical symmetric distributions in general.

Most existing test methods for the hypothesis (1) assume that the observations within groups are identically distributed. In this case, Assumptions 1 and 2 are all we need, and we completely avoid an assumption on the eigenstucture of $\bar{\mathbf{\Sigma}}_k$ like (2). In the general setting, $\mathbf{\Sigma}_{k,i}$ may be different within groups, and we make the following assumption to avoid the case in which there exist certain observations with significantly larger variance than the others.

Assumption 3. *Suppose that as $n \to \infty$,*

$$\frac{1}{n_k^2} \sum_{i=1}^{n_k} \operatorname{tr}(\mathbf{\Sigma}_{k,i}^2) = o\left\{\operatorname{tr}(\bar{\mathbf{\Sigma}}_k^2)\right\}, \quad k = 1, 2.$$

If the covariance matrices within groups are equal, i.e., $\mathbf{\Sigma}_{k,i} = \bar{\mathbf{\Sigma}}_k$, $i = 1, \ldots, n_k$, $k = 1, 2$, then Assumption 3 holds for arbitrary $\bar{\mathbf{\Sigma}}_k$, $k = 1, 2$, provided $\min(n_1, n_2) \to \infty$ as $n \to \infty$. In this view, Assumption 3 imposes no restriction on $\bar{\mathbf{\Sigma}}_k$, $k = 1, 2$.

Define $\mathbf{\Psi}_n = n_1^{-1}\bar{\mathbf{\Sigma}}_1 + n_2^{-1}\bar{\mathbf{\Sigma}}_2$. Let $\boldsymbol{\xi}_p$ be a p-dimensional standard normal random vector. We have the following theorem.

Theorem 1. *Suppose Assumptions 1, 2 and 3 hold, and $\sigma_{T,n}^2 > 0$ for all n. Then as $n \to \infty$,*

$$\left\| \mathcal{L}\left\{\frac{T_{\mathrm{CQ}}(\mathbf{Y}_1, \mathbf{Y}_2)}{\sigma_{T,n}}\right\} - \mathcal{L}\left[\frac{\boldsymbol{\xi}_p^\intercal \mathbf{\Psi}_n \boldsymbol{\xi}_p - \operatorname{tr}(\mathbf{\Psi}_n)}{\{2\operatorname{tr}(\mathbf{\Psi}_n^2)\}^{1/2}}\right] \right\|_3 \to 0.$$

Theorem 1 characterizes the general distributional behavior of $T_{\mathrm{CQ}}(\mathbf{Y}_1, \mathbf{Y}_2)$. It implies that the distributions of $T_{\mathrm{CQ}}(\mathbf{Y}_1, \mathbf{Y}_2)/\sigma_{T,n}$ and $\{\boldsymbol{\xi}_p^\intercal \mathbf{\Psi}_n \boldsymbol{\xi}_p - \operatorname{tr}(\mathbf{\Psi}_n)\}/\{2\operatorname{tr}(\mathbf{\Psi}_n^2)\}^{1/2}$ are equivalent asymptotically. To gain further insights on the distributional behavior of $T_{\mathrm{CQ}}(\mathbf{Y}_1, \mathbf{Y}_2)$, we would like to derive the asymptotic distributions of $T_{\mathrm{CQ}}(\mathbf{Y}_1, \mathbf{Y}_2)/\sigma_{T,n}$. However, Theorem 1 implies that $T_{\mathrm{CQ}}(\mathbf{Y}_1, \mathbf{Y}_2)/\sigma_{T,n}$ may not converge weakly in general. Nevertheless, $\mathcal{L}\{T_{\mathrm{CQ}}(\mathbf{Y}_1, \mathbf{Y}_2)/\sigma_{T,n}\}$ is uniformly tight and we can use Theorem 1 to derive all possible asympotitc distributions of $T_{\mathrm{CQ}}(\mathbf{Y}_1, \mathbf{Y}_2)/\sigma_{T,n}$.

Corollary 1. *Suppose the conditions of Theorem 1 hold. Then $\mathcal{L}\left\{T_{\mathrm{CQ}}(\mathbf{Y}_1, \mathbf{Y}_2)/\sigma_{T,n}\right\}$ is uniformly tight and all possible asymptotic distributions of $T_{\mathrm{CQ}}(\mathbf{Y}_1, \mathbf{Y}_2)/\sigma_{T,n}$ are give by*

$$\mathcal{L}\left\{(1 - \sum_{i=1}^{\infty} \kappa_i^2)^{1/2}\xi_0 + 2^{-1/2}\sum_{i=1}^{\infty} \kappa_i(\xi_i^2 - 1)\right\}, \tag{6}$$

where $\{\xi_i\}_{i=0}^{\infty}$ is a sequence of independent standard normal random variables, $\{\kappa_i\}_{i=1}^{\infty}$ is a sequence of positive numbers such that $\sum_{i=1}^{\infty} \kappa_i^2 \in [0, 1]$ and $\sum_{i=1}^{\infty} \kappa_i(\xi_i^2 - 1)$ is the almost sure limit of $\sum_{i=1}^{r} \kappa_i(\xi_i^2 - 1)$ as $r \to \infty$.

Remark 1. From Lévy's equivalence theorem and three-series theorem (see, e.g., Dudley (2002), Theorem 9.7.1 and Theorem 9.7.3), the series $\sum_{i=1}^{\infty} \kappa_i(\xi_i^2 - 1)$ converges almost surely and weakly. Hence the distribution (6) is well defined.

Corollary 1 gives a full characteristic of the possible asymptotic distributions of $T_{\mathrm{CQ}}(\mathbf{Y}_1, \mathbf{Y}_2)/\sigma_{T,n}$. In general, these possible asymptotic distributions are weighted sums of an independent normal random variable and centered χ^2 random variables. From the proof of Corollary 1, the parameters $\{\kappa_i\}_{i=1}^{\infty}$ are in fact the limit of the eigenvalues of the matrix $\mathbf{\Psi}_n/\{\operatorname{tr}(\mathbf{\Psi}_n^2)\}^{1/2}$ along a subsequence of $\{n\}$. If $\sum_{i=1}^{\infty}\kappa_i^2 = 0$, then (6) becomes the standard normal distribution, and the test procedure of Chen and Qin (2010) has correct level asymptotically. On the other hand, if $\sum_{i=1}^{\infty}\kappa_i^2 = 1$, then (6) becomes the distribution of a weighted sum of independent centered χ^2 random variables. This case was considered in Zhang et al. (2021). However, these two settings are just special cases among all possible asymptotic distributions where $\sum_{i=1}^{\infty}\kappa_i^2 \in [0,1]$. In general, the distribution (6) relies on the nuisance paramters $\{\kappa_i\}_{i=1}^{\infty}$. To construct a test procedure based on the asymptotic distributions, one needs to estimate these nuisance parameters consistently. Unfortunately, the estimation of the eigenvalues of high-dimensional covariance matrices may be a highly nontrivial task; see, e.g., Kong and Valiant (2017) and the references therein. Hence in general, it may not be a good choice to construct test procedures based on the asymptotic distributions.

3 Test procedure

An intuitive idea to control the test level of $T_{\mathrm{CQ}}(\mathbf{X}_1, \mathbf{X}_2)$ in the general setting is to use the bootstrap method. Surprisingly, the empirical bootstrap method and wild bootstrap method may not work for $T_{\mathrm{CQ}}(\mathbf{X}_1, \mathbf{X}_2)$. This phenomenon will be shown by our numerical experiments. For now, we give a heuristic argument to understand this phenomenon. First we consider the empirical bootstrap method. Suppose the resampled observations $\{X_{k,i}^*\}_{i=1}^{n_k}$ are uniformly sampled from $\{X_{k,i} - \bar{X}_k\}_{i=1}^{n_k}$ with replacement, $k = 1, 2$. Denote $\mathbf{X}_k^* = (X_{k,1}^*, \ldots, X_{k,n_k}^*)^\intercal$, $k = 1, 2$. The empirical bootstrap method uses the conditional distribution $\mathcal{L}\{T_{\mathrm{CQ}}(\mathbf{X}_1^*, \mathbf{X}_2^*) \mid \mathbf{X}_1, \mathbf{X}_2\}$ to approximate the null distribution of $T_{\mathrm{CQ}}(\mathbf{X}_1, \mathbf{X}_2)$. If this bootstrap method can work, one may expect that the first two moments of $\mathcal{L}\{T_{\mathrm{CQ}}(\mathbf{X}_1^*, \mathbf{X}_2^*) \mid \mathbf{X}_1, \mathbf{X}_2\}$ can approximately match the first two moments of $T_{\mathrm{CQ}}(\mathbf{X}_1, \mathbf{X}_2)$ under the null hypothesis. Under the null hypothesis, we have $\mathrm{E}\{T_{\mathrm{CQ}}(\mathbf{X}_1, \mathbf{X}_2)\} = 0$. Lemma S.5 implies that under Assumptions 1 and 3 and under the null hypothesis,

$$\operatorname{var}\{T_{\mathrm{CQ}}(\mathbf{X}_1, \mathbf{X}_2)\} = \sigma_{T,n}^2 = \{1 + o(1)\} 2 \operatorname{tr}(\mathbf{\Psi}_n^2).$$

On the othen hand, it is straightforward to show that $\mathrm{E}\{T_{\mathrm{CQ}}(\mathbf{X}_1^*, \mathbf{X}_2^*) \mid \mathbf{X}_1, \mathbf{X}_2\} = 0$. Also, under Assumptions 1 and 3,

$$\operatorname{var}\{T_{\mathrm{CQ}}(\mathbf{X}_1^*, \mathbf{X}_2^*) \mid \mathbf{X}_1, \mathbf{X}_2\} = \{1 + o_p(1)\} 2 \operatorname{tr}\{(n_1^{-1}\mathbf{S}_1 + n_2^{-1}\mathbf{S}_2)^2\}.$$

Unfortunately, $\operatorname{tr}\{(n_1^{-1}\mathbf{S}_1 + n_2^{-1}\mathbf{S}_2)^2\}$ is not a ratio-consistent estimator of $\operatorname{tr}(\mathbf{\Psi}_n^2)$ even in the settings where $X_{k,i}$ is normally distributed, the covariance matrices $\mathbf{\Sigma}_{k,i}$, $i = 1, \ldots, n_k$, $k = 1, 2$, are all equal and $n_1 = n_2$; see, e.g., Bai and Saranadasa (1996) and Zhou and Guo (2017). Consequently, the empirical bootstrap method may not be valid for $T_{\mathrm{CQ}}(\mathbf{X}_1, \mathbf{X}_2)$.

We turn to the wild bootstrap method. Recently, the wild bootstrap method has been widely used for extreme value type statistics in the high-dimensional setting; see, e.g., Chernozhukov et al. (2013), Chernozhukov et al. (2017), Xue and Yao (2020) and Deng and Zhang (2020). For the wild bootstrap method, the resampled observations are defined as $X_{k,i}^* = \varepsilon_{k,i}(X_{k,i} - \bar{X}_k)$, $i = 1, \ldots, n_k$, $k = 1, 2$, where $\{\varepsilon_{k,i}\}$ are independent and identically distributed random variables with $\mathrm{E}(\varepsilon_{k,i}) = 0$ and $\operatorname{var}(\varepsilon_{k,i}) = 1$, and are independent of the original data $\mathbf{X}_1$, $\mathbf{X}_2$. We have $\mathrm{E}\{T_{\mathrm{CQ}}(\mathbf{X}_1^*, \mathbf{X}_2^*) \mid \mathbf{X}_1, \mathbf{X}_2\} = 0$. With some tedious but straightforward derivations, it can be seen that under Assumptions 1 and 3,

$$\mathrm{E}\{\operatorname{var}(T_{\mathrm{CQ}}(\mathbf{X}_1^*, \mathbf{X}_2^*) \mid \mathbf{X}_1, \mathbf{X}_2)\} = \{1 + o(1)\} 2 \operatorname{tr}(\mathbf{\Psi}_n^2) + b,$$

where b is the bias term and satisfies

$$
\begin{aligned}
b = &\{1+o(1)\}\sum_{k=1}^{2}\sum_{i=1}^{n_k}\frac{4}{n_k^5}\,\mathrm{E}\{(Y_{k,i}^\intercal Y_{k,i})^2\} - \{1+o(1)\}\sum_{k=1}^{2}\frac{2}{n_k^4}\{\operatorname{tr}(\bar{\mathbf{\Sigma}}_k)\}^2 \\
\geq &\{1+o(1)\}\sum_{k=1}^{2}\sum_{i=1}^{n_k}\frac{2}{n_k^5}\{\operatorname{tr}(\mathbf{\Sigma}_{k,i})\}^2.
\end{aligned}
$$

The above inequality implies that the bias term b may not be negligible compared with $2\operatorname{tr}(\mathbf{\Psi}_n^2)$. For example, if $\mathbf{\Sigma}_{k,i} = \mathbf{I}_p$, $i = 1,\ldots,n_k$, $k = 1,2$, then we have $2\operatorname{tr}(\mathbf{\Psi}_n^2) = 2(n_1^{-1}+n_2^{-1})^2 p$ and $b \geq \{1+o(1)\}2(n_1^{-4}+n_2^{-4})p^2$. In this case, the bias term b is not negligible provided $n^2 = O(p)$. Thus, the wild bootstrap method may not be valid for $T_{\mathrm{CQ}}(\mathbf{X}_1,\mathbf{X}_2)$ either.

We have seen that the methods based on asymptotic distributions and the bootstrap methods may not work well for the test statistic $T_{\mathrm{CQ}}(\mathbf{X}_1,\mathbf{X}_2)$. These phenomenons imply that it is highly nontrivial to construct a valid test procedure based on $T_{\mathrm{CQ}}(\mathbf{X}_1,\mathbf{X}_2)$. To construct a valid test procedure, we resort to the idea of the randomization test, a powerful tool in statistical hypothesis testing. The randomization test is an old idea and can at least date back to Fisher (1935), Section 21; see Hoeffding (1952), Lehmann and Romano (2005), Section 15.2, Zhu (2005) and Hemerik and Goeman (2018) for general frameworks and extensions of the randomization tests. The original randomization method considered in Fisher (1935), Section 21 can be abstracted into the following general form. Suppose $\xi_1,\ldots,\xi_n$ are independent p-dimensional random vectors such that $\mathcal{L}(\xi_i) = \mathcal{L}(-\xi_i)$. Let $T(\xi_1,\ldots,\xi_n)$ be any statistics taking values in $\mathbb{R}$. Suppose $\epsilon_1,\ldots,\epsilon_n$ are independent and identically distributed Rademacher random variables, i.e., $\operatorname{pr}(\epsilon_i = 1) = \operatorname{pr}(\epsilon_i = -1) = 1/2$, and are independent of $\xi_1,\ldots,\xi_n$. Define the conditional cumulative distribution function $\hat{F}(\cdot)$ as $\hat{F}(x) = \operatorname{pr}\{T(\epsilon_1\xi_1,\ldots,\epsilon_n\xi_n) \leq x \mid \xi_1,\ldots,\xi_n\}$. Then from the theory of randomization test (see, e.g., Lehmann and Romano (2005), Section 15.2), for any $\alpha \in (0,1)$,

$$
\operatorname{pr}\left\{T(\xi_1,\ldots,\xi_n) > \hat{F}^{-1}(1-\alpha)\right\} \leq \alpha,
$$

where for any right continuous cumulative distribution function $F(\cdot)$ on $\mathbb{R}$, $F^{-1}(q) = \min\{x \in \mathbb{R} : F(x) \geq q\}$ for $q \in (0,1)$. Also, under mild conditions, the difference between the above probability and α is negligible.

To apply Fisher's randomization method to specific problems, the key is to construct random variables ξ_i such that $\mathcal{L}(\xi_i) = \mathcal{L}(-\xi_i)$ under the null hypothesis. This randomization method can be directly applied to one-sample mean testing problem, as did in Wang and Xu (2019). However, it can not be readily applied to the testing problem (1). In fact, the mean vector μ_k of $X_{k,i}$ is unknown under the null hypothesis, and consequently, one can not expect that $\mathcal{L}(X_{k,i}) = \mathcal{L}(-X_{k,i})$ holds under the null hypothesis. As a result, $X_{k,i}$ can not serve as ξ_i in Fisher's randomization method.

We observe that the difference $X_{k,i} - X_{k,i+1}$ has zero means and hence is free of the mean vector μ_k. Also, if $\mathcal{L}(X_{k,i}) = \mathcal{L}(X_{k,i+1})$, then $\mathcal{L}(X_{k,i} - X_{k,i+1}) = \mathcal{L}(X_{k,i+1} - X_{k,i})$. These facts imply that Fisher's randomization method may be applied to the random vectors $\tilde{X}_{k,i} = (X_{k,2i} - X_{k,2i-1})/2$, $i = 1,\ldots,m_k$, $k = 1,2$ where $m_k = \lfloor n_k/2 \rfloor$, $k = 1,2$. Define

$$
T_{\mathrm{CQ}}(\tilde{\mathbf{X}}_1,\tilde{\mathbf{X}}_2) = \sum_{k=1}^{2}\frac{2\sum_{i=1}^{m_k}\sum_{j=i+1}^{m_k}\tilde{X}_{k,i}^\intercal\tilde{X}_{k,j}}{m_k(m_k-1)} - \frac{2\sum_{i=1}^{m_1}\sum_{j=1}^{m_2}\tilde{X}_{1,i}^\intercal\tilde{X}_{2,j}}{m_1 m_2},
$$

where $\tilde{\mathbf{X}}_k = (\tilde{X}_{k,1},\ldots,\tilde{X}_{k,m_k})^\intercal$, $k = 1,2$. Let $E = (\epsilon_{1,1},\ldots,\epsilon_{1,m_1},\epsilon_{2,1},\ldots,\epsilon_{2,m_2})^\intercal$ where $\{\epsilon_{k,i}\}$ are independent and identically distributed Rademacher random variables which are independent of $\tilde{\mathbf{X}}_1$, $\tilde{\mathbf{X}}_2$.

Define randomized statistic

$$T_{\mathrm{CQ}}(E;\tilde{\mathbf{X}}_1,\tilde{\mathbf{X}}_2)=\sum_{k=1}^{2}\frac{2\sum_{i=1}^{m_k}\sum_{j=i+1}^{m_k}\epsilon_{k,i}\epsilon_{k,j}\tilde{X}_{k,i}^{\intercal}\tilde{X}_{k,j}}{m_k(m_k-1)}-\frac{2\sum_{i=1}^{m_1}\sum_{j=1}^{m_2}\epsilon_{1,i}\epsilon_{2,j}\tilde{X}_{1,i}^{\intercal}\tilde{X}_{2,j}}{m_1m_2}.$$

Define the conditional distribution function $\hat{F}_{\mathrm{CQ}}(x)=\mathrm{pr}\left\{T_{\mathrm{CQ}}(E;\tilde{\mathbf{X}}_1,\tilde{\mathbf{X}}_2)\leq x\mid\tilde{\mathbf{X}}_1,\tilde{\mathbf{X}}_2\right\}$. From Fisher's randomization method, if $\mathcal{L}(\tilde{X}_{k,i})=\mathcal{L}(-\tilde{X}_{k,i})$, then for $\alpha\in(0,1)$, $\mathrm{pr}\left\{T_{\mathrm{CQ}}(\tilde{\mathbf{X}}_1,\tilde{\mathbf{X}}_2)>\hat{F}_{\mathrm{CQ}}^{-1}(1-\alpha)\right\}\leq\alpha$. It can be seen that $T_{\mathrm{CQ}}(\tilde{\mathbf{X}}_1,\tilde{\mathbf{X}}_2)$ and $T_{\mathrm{CQ}}(\mathbf{X}_1,\mathbf{X}_2)$ take a similar form. Also, under the null hypothesis, $\mathrm{E}\{T_{\mathrm{CQ}}(\tilde{\mathbf{X}}_1,\tilde{\mathbf{X}}_2)\}=\mathrm{E}\{T_{\mathrm{CQ}}(\mathbf{X}_1,\mathbf{X}_2)\}=0$ and $\mathrm{var}\{T_{\mathrm{CQ}}(\tilde{\mathbf{X}}_1,\tilde{\mathbf{X}}_2)\}=\{1+o(1)\}\,\mathrm{var}\{T_{\mathrm{CQ}}(\mathbf{X}_1,\mathbf{X}_2)\}$. Thus, it may be expected that $\mathcal{L}\{T_{\mathrm{CQ}}(\mathbf{X}_1,\mathbf{X}_2)\}\approx\mathcal{L}\{T_{\mathrm{CQ}}(\tilde{\mathbf{X}}_1,\tilde{\mathbf{X}}_2)\}$ under the null hypothesis. On the other hand, the classical results of Hoeffding (1952) on randomization tests give the insight that the randomness of $\hat{F}_{\mathrm{CQ}}^{-1}(1-\alpha)$ may be negligible for large samples. From the above insights, it can be expected that under the null hypothesis, for $\alpha\in(0,1)$,

$$\mathrm{pr}\left\{T_{\mathrm{CQ}}(\mathbf{X}_1,\mathbf{X}_2)>\hat{F}_{\mathrm{CQ}}^{-1}(1-\alpha)\right\}\approx\mathrm{pr}\left\{T_{\mathrm{CQ}}(\tilde{\mathbf{X}}_1,\tilde{\mathbf{X}}_2)>\hat{F}_{\mathrm{CQ}}^{-1}(1-\alpha)\right\}\approx\alpha.$$

Motivated by the above heuristics, we propose a new test procedure which rejects the null hypothesis if

$$T_{\mathrm{CQ}}(\mathbf{X}_1,\mathbf{X}_2)>\hat{F}_{\mathrm{CQ}}^{-1}(1-\alpha).$$

While the assumption $\mathcal{L}(\tilde{X}_{k,i})=\mathcal{L}(-\tilde{X}_{k,i})$ is used in the above heuristic arguments, it will not be assumed in our theoretical analysis. This generality may not be surprising. Indeed, for low-dimensional testing problems, it is known that such symmetry conditions can often be relaxed for randomization tests; see, e.g., Romano (1990), Chung and Romano (2013) and Canay et al. (2017). In the proposed procedure, the conditional distribution $\mathcal{L}\{T_{\mathrm{CQ}}(E;\tilde{\mathbf{X}}_1,\tilde{\mathbf{X}}_2)\mid\tilde{\mathbf{X}}_1,\tilde{\mathbf{X}}_2\}$ is used to approximate the null distribution of $T_{\mathrm{CQ}}(\mathbf{X}_1,\mathbf{X}_2)$. We have $\mathrm{E}\{T_{\mathrm{CQ}}(E;\tilde{\mathbf{X}}_1,\tilde{\mathbf{X}}_2)\mid\tilde{\mathbf{X}}_1,\tilde{\mathbf{X}}_2\}=0$. From Lemma S.10, under Assumptions 1-3, we have

$$\mathrm{var}\{T_{\mathrm{CQ}}(E;\tilde{\mathbf{X}}_1,\tilde{\mathbf{X}}_2)\mid\tilde{\mathbf{X}}_1,\tilde{\mathbf{X}}_2\}=\{1+o_P(1)\}\sigma_{T,n}^2.$$

That is, the first two moments of $\mathcal{L}\{T_{\mathrm{CQ}}(E;\tilde{\mathbf{X}}_1,\tilde{\mathbf{X}}_2)\mid\tilde{\mathbf{X}}_1,\tilde{\mathbf{X}}_2\}$ can match those of $\mathcal{L}\{T_{\mathrm{CQ}}(\mathbf{X}_1,\mathbf{X}_2)\}$. As we have seen, this favorable property is not shared by the empirical bootstrap method and wild bootstrap method.

We should emphasis that the proposed test procedure is only inspired by the randomization test, and is not a randomization test in itself. As a consequence, it can not be expected that the proposed test can have an exact control of the test level. In fact, even if the observations are 1-dimensional and normally distributed, the exact control of the test level for Behrens-Fisher problem is not trivial; see Linnik (1966) and the references therein. Nevertheless, we shall show that the proposed test procedure can control the test level asymptotically under Assumptions 1-3.

The conditional distribution $\mathcal{L}\{T_{\mathrm{CQ}}(E;\tilde{\mathbf{X}}_1,\tilde{\mathbf{X}}_2)\mid\tilde{\mathbf{X}}_1,\tilde{\mathbf{X}}_2\}$ is a discrete distribution uniformly distributed on $2^{m_1+m_2}$ values. Hence it is not feasible to compute the exact quantile of $\hat{F}_{\mathrm{CQ}}(\cdot)$. In practice, one can use a finite sample from $\mathcal{L}\{T_{\mathrm{CQ}}(E;\tilde{\mathbf{X}}_1,\tilde{\mathbf{X}}_2)\mid\tilde{\mathbf{X}}_1,\tilde{\mathbf{X}}_2\}$ to approximate the p-value of the proposed test procedure; see, e.g., Lehmann and Romano (2005), Chapter 15. More specifically, given data, we can independently sample $E^{(i)}$ and compute $T_{\mathrm{CQ}}^{(i)}=T_{\mathrm{CQ}}(E^{(i)};\tilde{\mathbf{X}}_1,\tilde{\mathbf{X}}_2)$, $i=1,\ldots,B$, where B is a sufficiently large number. Then the null hypothesis is rejected if

$$\frac{1}{B+1}\left[1+\sum_{i=1}^{B}\mathbf{1}_{\{T_{\mathrm{CQ}}^{(i)}\geq T_{\mathrm{CQ}}(\mathbf{X}_1,\mathbf{X}_2)\}}\right]\leq\alpha.$$

In the above procedure, one needs to compute the original statistic $T_{\mathrm{CQ}}(\mathbf{X}_1, \mathbf{X}_2)$ and the randomized statistics $T_{\mathrm{CQ}}^{(1)}, \ldots, T_{\mathrm{CQ}}^{(B)}$. The direct computation of the original statistic $T_{\mathrm{CQ}}(\mathbf{X}_1, \mathbf{X}_2)$ costs $O(n^2 p)$ time. During the computation of $T_{\mathrm{CQ}}(\mathbf{X}_1, \mathbf{X}_2)$, we can cache the inner products $\tilde{X}_{k,i}^\intercal \tilde{X}_{k,j}$, $1 \leq i < j \leq m_k$, $k = 1, 2$, and $\tilde{X}_{1,i}^\intercal \tilde{X}_{2,j}$, $i = 1, \ldots, m_1$, $j = 1, \ldots, m_2$. Then the computation of $T_{\mathrm{CQ}}^{(i)}$ only requires $O(n^2)$ time. In total, the computation of the proposed test procedure can be completed within $O\{n^2(p + B)\}$ time.

Now we rigorously investigate the theoretical properties of the proposed test procedure. From Theorem 1, the distribution of $T_{\mathrm{CQ}}(\mathbf{X}_1, \mathbf{X}_2)$ is asymptotically equivalent to the distribution of a quadratic form in normal random variables. Now we show that the conditional distribution $\mathcal{L}\{T_{\mathrm{CQ}}(E; \tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2) \mid \tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2\}$ is equivalent to the same quadratic form. In fact, our result is more general and includes the case that the elements of E are generated from the standard normal distribution.

Theorem 2. *Suppose the conditions of Theorem 1 hold. Let* $E^* = (\epsilon_{1,1}^*, \ldots, \epsilon_{1,m_1}^*, \epsilon_{2,1}^*, \ldots, \epsilon_{2,m_2}^*)^\intercal$*, where* $\{\epsilon_{k,i}^*\}$ *are independent and identically distributed random variables, and* $\epsilon_{1,1}^*$ *is a standard normal random variable or a Rademacher random variable. Then as* $n \to \infty$*,*

$$\left\| \mathcal{L}\left\{ \frac{T_{\mathrm{CQ}}(E^*; \tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2)}{\sigma_{T,n}} \mid \tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2 \right\} - \mathcal{L}\left[\frac{\boldsymbol{\xi}_p^\intercal \boldsymbol{\Psi}_n \boldsymbol{\xi}_p - \operatorname{tr}(\boldsymbol{\Psi}_n)}{\{2 \operatorname{tr}(\boldsymbol{\Psi}_n^2)\}^{1/2}} \right] \right\|_3 \xrightarrow{P} 0.$$

From Theorems 1 and 2, the conditional distribution $\mathcal{L}\{T_{\mathrm{CQ}}(E; \tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2) \mid \tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2\}$ is asymptotically equivalent to the distribution of $T_{\mathrm{CQ}}(\mathbf{X}_1, \mathbf{X}_2)$ under the null hypothesis. These two theorems allow us to derive the asymptotic level and local power of the proposed test procedure. Let $G_n(\cdot)$ denote the cumulative distribution function of $\{\boldsymbol{\xi}_p^\intercal \boldsymbol{\Psi}_n \boldsymbol{\xi}_p - \operatorname{tr}(\boldsymbol{\Psi}_n)\}/\{2 \operatorname{tr}(\boldsymbol{\Psi}_n^2)\}^{1/2}$.

Corollary 2. *Suppose the conditions of Theorem 1 hold,* $\alpha \in (0, 1)$ *is an absolute constant and as* $n \to \infty$*,* $(\mu_1 - \mu_2)^\intercal \boldsymbol{\Psi}_n (\mu_1 - \mu_2) = o\{\operatorname{tr}(\boldsymbol{\Psi}_n^2)\}$*. Then as* $n \to \infty$*,*

$$\operatorname{pr}\left\{ T_{\mathrm{CQ}}(\mathbf{X}_1, \mathbf{X}_2) > \hat{F}_{\mathrm{CQ}}^{-1}(1 - \alpha) \right\} = 1 - G_n\left[G_n^{-1}(1 - \alpha) - \frac{\|\mu_1 - \mu_2\|^2}{\{2 \operatorname{tr}(\boldsymbol{\Psi}_n^2)\}^{1/2}} \right] + o(1).$$

Corollary 2 implies that under the conditions of Theorem 1, the proposed test procedure has correct test level asymptotically. In particular, Corollary 2 provides a rigorous theoretical guarantee of the validity of the proposed test procedure for two-sample Behrens-Fisher problem with arbitrary $\bar{\boldsymbol{\Sigma}}_k$, $i = 1, 2$. Furthermore, the proposed test procedure is still valid when the observations are not identically distributed within groups and the sample sizes are unbalanced. To the best of our knowledge, the proposed test procedure is the only one that is guaranteed to be valid in such a general setting. Corollary 2 also gives the asymptotic power of the proposed test procedure under the local alternative hypotheses. If $T_{\mathrm{CQ}}(\mathbf{X}_1, \mathbf{X}_2)$ is asymptotically normally distributed, that is, G_n converges weakly to the cumulative distribution function of the standard normal distribution, then the proposed test procedure has the same local asymptotic power as the test procedure of Chen and Qin (2010). In general, the proposed test procedure has the same local asymptotic power as the oracle test procedure which rejects the null hypothesis when $T_{\mathrm{CQ}}(\mathbf{X}_1, \mathbf{X}_2)$ is greater than the $1 - \alpha$ quantile of $\mathcal{L}\{T_{\mathrm{CQ}}(\mathbf{X}_1, \mathbf{X}_2)\}$. Thus, the proposed test procedure has good power behavior.

4 Simulations

In this section, we conduct simulations to examine the performance of the proposed test procedure and compare it with 9 alternative test procedures. The first 4 competing test procedures are based on $T_{\mathrm{CQ}}(\mathbf{X}_1, \mathbf{X}_2)$, including the original test procedure of Chen and Qin (2010) which is based on the asymptotic normality of $T_{\mathrm{CQ}}(\mathbf{X}_1, \mathbf{X}_2)$, the test procedure based on the empirical bootstrap method, the wild bootstrap method described in Section 3 where $\{\varepsilon_{k,i}\}$ are Rademacher random variables, and the χ^2-approximation method in Zhang and Zhu (2021). The next 2 competing test procedures are based on the statistic $\|\bar{X}_1 - \bar{X}_2\|^2$, including the χ^2-approximation method in Zhang et al. (2021) and the half-sampling method in Lou (2020). The last 3 competing test procedures are scalar-invariant tests of Srivastava and Du (2008), Srivastava et al. (2013) and Feng et al. (2015).

In our simulations, the nominal test level is $\alpha = 0.05$. For the proposed method and competing resampling methods, the resampling number is $B = 1,000$. The reported empirical sizes and powers are computed based on $10,000$ independent replications. We consider the following data generation models for $\{Y_{k,i}\}$.

- Model I: $Y_{k,i} \sim \mathcal{N}(\mathbf{0}_p, \mathbf{I}_p)$, $i = 1, \ldots, n_k$, $k = 1, 2$.
- Model II: $Y_{k,i} \sim \mathcal{N}(\mathbf{0}_p, \bar{\mathbf{\Sigma}}_k)$, $i = 1, \ldots, n_k$, $k = 1, 2$, where $\bar{\mathbf{\Sigma}}_k = \mathbf{V}_k \mathbf{\Lambda} \mathbf{V}_k^\intercal + \mathbf{I}_p$ with

$$\mathbf{\Lambda} = \begin{pmatrix} 1 & 0 \\ 0 & 1/2 \end{pmatrix}, \quad \mathbf{V}_1 = \begin{pmatrix} \mathbf{1}_{p/4} & \mathbf{1}_{p/4} \\ \mathbf{1}_{p/4} & -\mathbf{1}_{p/4} \\ \mathbf{1}_{p/4} & \mathbf{1}_{p/4} \\ \mathbf{1}_{p/4} & -\mathbf{1}_{p/4} \end{pmatrix}, \quad \mathbf{V}_2 = \begin{pmatrix} \mathbf{1}_{p/4} & \mathbf{1}_{p/4} \\ \mathbf{1}_{p/4} & -\mathbf{1}_{p/4} \\ -\mathbf{1}_{p/4} & -\mathbf{1}_{p/4} \\ -\mathbf{1}_{p/4} & \mathbf{1}_{p/4} \end{pmatrix}.$$

- Model III: $Y_{k,i} = \mathbf{\Gamma}_{k,i} Z_{k,i}$, $i = 1, \ldots, n_k$, $k = 1, 2$, where $Z_{k,i} = (z_{k,i,1}, \ldots, z_{k,i,p})^\intercal$, and $\{z_{k,i,j}\}$ are independent standardized χ^2 random variables with 1 degree of freedom, that is, $z_{k,i,1} \sim \{\chi^2(1) - 1\}/\sqrt{2}$. For $i = 1, \ldots, n_k/2$, $\mathbf{\Gamma}_{k,i} = \{k \operatorname{diag}(1, 2, \ldots, p)\}^{1/2}$, and for $i = n_k/2 + 1, \ldots, n_k$, $\mathbf{\Gamma}_{k,i} = \{k \operatorname{diag}(p, p-1, \ldots, 1)\}^{1/2}$.
- Model IV: The jth element of $Y_{k,i}$ is $y_{k,i,j} = \sum_{\ell=0}^{5} 1.01^{j+\ell-1} z_{k,i,j+\ell}$, $j = 1, \ldots, p$, $i = 1, \ldots, n_k$, $k = 1, 2$, where $\{z_{k,i,j}\}$ are independent standardized χ^2 random variables with 1 degree of freedom.

For Model I, observations are simply normal random vectors with identity covariance matrix. For Model II, the variables are correlated, $\bar{\mathbf{\Sigma}}_1 \neq \bar{\mathbf{\Sigma}}_2$ and the condition (2) is not satisfied. For Model III, $\mathbf{\Sigma}_{k,1}, \ldots, \mathbf{\Sigma}_{k,n_k}$ are not equal and the observations have skewed distributions. For Model IV, the variables are correlated and their variances are different.

In Section S.1, we give quantile-quantile plots to examine the correctness of Theorem 1 and Corollary 1 of the proposed test statistic. The results show that the distribution approximation in Theorem 1 is quite accurate, and the asymptotic distributions in Corollary 1 are reasonable even for finite sample size.

Now we consider the simulations of empirical sizes. We take $\mu_1 = \mu_2 = \mathbf{0}_p$. Table 1 lists the empirical sizes of various test procedures. It can be seen that the test procedure of Chen and Qin (2010) tends to have inflated empirical sizes, especially for Models II and IV. The empirical bootstrap method does not work well for Models I, III and IV. While the wild bootstrap method has a better performance than the empirical bootstrap method, it is overly conservative for Models I and III when n is small. The performance of bootstrap methods confirm our heuristic arguments in Section 3. It is interesting that the empirical bootstrap method has relatively good performance for Model II. In fact, in Model II, $\bar{\mathbf{\Sigma}}_k$ has a low-rank structure, and therefore, the test statistic $T_{\mathrm{CQ}}(\mathbf{X}_1, \mathbf{X}_2)$ may have a similar behavior as in the low-dimensional setting. It is known that the empirical bootstrap method has good performance in the low-dimensional setting. Hence it is reasonable that the empirical bootstrap method works well for Model II. The χ^2-approximation methods of Zhang et al. (2021) and Zhang and Zhu (2021) have reasonable performance under Model I, II and IV,

Table 1: Empirical sizes (multiplied by 100) of test procedures

Model	n_1	n_2	p	NEW	CQ	EB	WB	ZZ	ZZGZ	LOU	SD	SKK	FZWZ
I	8	12	300	4.95	5.42	0.00	1.32	4.92	5.46	6.13	3.77	32.8	3.20
	16	24	300	5.19	5.94	0.00	3.88	5.41	5.72	5.98	4.37	14.4	5.08
	32	48	300	5.00	5.65	0.00	4.79	5.21	5.33	5.42	4.21	8.36	5.11
	8	12	600	5.07	5.52	0.00	0.45	4.99	5.73	6.22	2.62	48.2	2.55
	16	24	600	4.85	5.44	0.00	2.25	5.07	5.40	5.57	3.05	17.5	4.21
	32	48	600	5.23	5.50	0.00	4.22	5.21	5.35	5.40	3.94	9.48	5.14
II	8	12	300	8.99	10.1	5.91	8.56	8.23	9.87	9.14	3.31	5.94	7.78
	16	24	300	6.65	8.15	5.54	6.70	6.41	7.39	6.70	2.12	3.52	7.37
	32	48	300	5.46	7.45	5.14	5.56	5.34	6.25	5.57	1.85	2.52	7.04
	8	12	600	9.02	10.4	6.10	8.80	8.44	9.85	9.48	2.64	4.77	8.02
	16	24	600	6.10	7.83	4.79	5.95	5.60	6.84	6.06	1.64	2.60	6.72
	32	48	600	5.27	7.09	4.71	5.20	5.07	5.87	5.20	1.18	1.78	6.65
III	8	12	300	5.23	5.50	0.00	1.30	1.63	2.18	6.14	0.17	14.1	0.00
	16	24	300	5.31	5.63	0.00	3.78	2.89	3.22	5.71	0.22	8.30	0.00
	32	48	300	5.56	6.09	0.01	5.10	3.94	4.16	5.81	0.16	6.79	0.13
	8	12	600	5.58	5.77	0.00	0.58	2.04	2.46	6.38	0.00	19.5	0.00
	16	24	600	5.01	5.28	0.00	2.39	2.64	2.94	5.46	0.00	10.5	0.00
	32	48	600	5.00	5.32	0.00	4.20	3.64	3.82	5.12	0.02	7.11	0.03
IV	8	12	300	5.90	7.27	1.33	5.92	5.22	6.15	6.64	3.05	11.2	2.05
	16	24	300	5.21	6.73	2.29	5.49	4.63	5.35	5.48	3.25	6.90	3.53
	32	48	300	5.03	6.76	3.17	5.58	4.92	5.50	5.38	4.02	5.82	5.21
	8	12	600	5.97	7.49	1.29	6.03	5.20	6.17	6.75	2.24	14.4	1.48
	16	24	600	5.60	7.19	2.37	5.99	5.06	5.91	5.94	2.52	7.62	3.13
	32	48	600	5.50	7.11	3.59	5.90	5.16	5.90	5.85	3.66	5.86	4.46

NEW, the proposed test procedure; CQ, the test of Chen and Qin (2010); EB, the empirical bootstrap method based on $T_{\mathrm{CQ}}(\mathbf{X}_1, \mathbf{X}_2)$; WB, the wild bootstrap method based on $T_{\mathrm{CQ}}(\mathbf{X}_1, \mathbf{X}_2)$; ZZ, the test of Zhang and Zhu (2021); ZZGZ, the test of Zhang et al. (2021); LOU, the test of Lou (2020); SD, the test of Srivastava and Du (2008); SKK, the test of Srivastava et al. (2013); FZWZ, the test of Feng et al. (2015).

but are overly conservative for Model III. The half-sampling method of Lou (2020) has inflated empirical sizes, especially for small n. Compared with the test procedures based on $\|\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2\|^2$ or $T_{\mathrm{CQ}}(\mathbf{X}_1, \mathbf{X}_2)$, the scalar-invariant tests of Srivastava and Du (2008), Srivastava et al. (2013) and Feng et al. (2015) have relatively poor performance, especially when n is small. For Models I, III and IV, the empirical sizes of the proposed test procedure are quite close to the nominal test level. For Model II, all test procedures based on $\|\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2\|^2$ or $T_{\mathrm{CQ}}(\mathbf{X}_1, \mathbf{X}_2)$ have inflated empirical sizes. For this model, the proposed test procedure outperforms the test procedure of Chen and Qin (2010). Also, as n increases, the empirical sizes of the proposed test procedure tends to the nominal test level. For $n_1 = 32$, $n_2 = 48$, the proposed test procedure has reasonable performance in all settings. Overall, the proposed test procedure has reasonably good performance in terms of empirical sizes, which confirms our theoretical results.

Now we consider the empirical powers of various test procedures. In view of the expression of the asymptotic power given in Corollary 2, we define the signal-to-noise ratio $\beta = \|\mu_1 - \mu_2\|^2 / \{2\,\mathrm{tr}(\mathbf{\Psi}_n^2)\}^{1/2}$. We take $\mu_1 = \mathbf{0}_p$ and $\mu_2 = c\mathbf{1}_p$ where c is chosen such that β reaches given values of signal-to-noise ratio. Table 2 lists the empirical powers of various test procedures when $p = 300$. It can be seen that for Model IV where the variables have different variance scale, the scalar-invariant tests have better performance

Table 2: Empirical powers (multiplied by 100) of test procedures for $p = 300$

Model	n_1	n_2	β	NEW	CQ	EB	WB	ZZ	ZZGZ	LOU	SD	SKK	FZWZ
I	8	12	1	24.4	26.4	0.00	10.6	23.5	26.7	28.0	19.9	65.9	16.7
	16	24	1	25.3	27.1	0.01	20.7	26.0	26.7	27.2	21.5	43.7	23.2
	32	48	1	25.3	27.2	0.37	24.2	25.8	26.2	26.5	22.8	33.7	25.6
	8	12	2	55.5	59.1	0.0	35.1	53.8	59.6	61.0	48.2	88.7	43.4
	16	24	2	57.7	60.5	0.22	51.5	58.9	59.9	60.7	52.1	75.5	54.5
	32	48	2	58.5	60.5	3.88	57.3	58.8	59.3	59.6	55.5	66.3	58.2
II	8	12	1	28.1	31.7	24.0	28.3	27.8	31.0	29.5	16.4	21.8	26.4
	16	24	1	25.5	30.0	24.2	26.6	26.0	28.0	26.6	14.9	17.9	27.8
	32	48	1	24.5	29.3	23.7	25.0	24.5	26.7	25.3	13.7	16.1	28.4
	8	12	2	46.7	50.9	42.4	47.5	46.8	50.3	48.7	31.8	38.7	45.0
	16	24	2	44.3	50.6	43.1	46.2	45.4	48.5	46.5	29.6	34.6	48.1
	32	48	2	44.0	50.0	43.1	44.6	44.2	46.6	44.9	29.4	32.9	48.8
III	8	12	1	24.8	25.9	0.00	11.1	12.7	14.7	27.3	0.01	34.0	0.00
	16	24	1	25.5	27.2	0.02	20.8	17.5	18.7	27.1	0.00	28.2	0.01
	32	48	1	25.1	26.8	0.49	24.2	20.7	21.3	26.0	0.03	24.4	1.22
	8	12	2	57.9	60.7	0.00	36.9	39.4	43.9	62.0	0.00	80.1	0.00
	16	24	2	57.9	60.2	0.29	52.5	47.8	49.6	60.0	0.10	69.3	0.87
	32	48	2	59.7	61.7	5.37	58.6	54.0	54.8	60.6	0.36	63.2	12.4
IV	8	12	1	22.3	26.9	7.83	22.6	20.6	23.3	25.1	100	100	100
	16	24	1	21.4	25.9	12.2	22.8	20.2	22.3	22.7	100	100	100
	32	48	1	21.3	26.4	15.7	22.5	20.8	22.6	22.2	100	100	100
	8	12	2	49.9	57.0	24.2	51.2	47.7	51.6	54.3	100	100	100
	16	24	2	49.5	56.5	33.1	51.7	47.7	51.2	51.9	100	100	100
	32	48	2	49.4	56.5	39.6	51.1	48.4	51.0	50.5	100	100	100

NEW, the proposed test procedure; CQ, the test of Chen and Qin (2010); EB, the empirical bootstrap method based on $T_{\text{CQ}}(\mathbf{X}_1, \mathbf{X}_2)$; WB, the wild bootstrap method based on $T_{\text{CQ}}(\mathbf{X}_1, \mathbf{X}_2)$; ZZ, the test of Zhang and Zhu (2021); ZZGZ, the test of Zhang et al. (2021); LOU, the test of Lou (2020); SD, the test of Srivastava and Du (2008); SKK, the test of Srivastava et al. (2013); FZWZ, the test of Feng et al. (2015).

than other tests. However, for Model III where $\mathbf{\Sigma}_{k,1}, \ldots, \mathbf{\Sigma}_{k,n_k}$ are not identical and the observations have skewed distributions, the scalar-invariant tests have relatively low powers. The proposed test procedure has a reasonable power behavior among the test procedures based on $\|\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2\|^2$ or $T_{\text{CQ}}(\mathbf{X}_1, \mathbf{X}_2)$. We have seen that some competing tests do not have a good control of test level. To get rid of the effect of distorted test level on the power, we present the receiver operating characteristic curves of the test procedures in Section S.1. It shows that for a given test level, all test procedures based on $\|\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2\|^2$ or $T_{\text{CQ}}(\mathbf{X}_1, \mathbf{X}_2)$ have quite similar power behavior. In summary, the proposed test procedure has promising performance of empirical sizes, and has no power loss compared with existing tests based on $\|\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2\|^2$ or $T_{\text{CQ}}(\mathbf{X}_1, \mathbf{X}_2)$.

5 Real-data example

In this section, we apply the proposed test procedure to the gene expression dataset released by Alon et al. (1999). This dataset consists of the gene expression levels of $n_1 = 22$ normal and $n_2 = 40$ tumor colon tissue samples. It contains the expression of $p = 2,000$ genes with highest minimal intensity across the $n = 62$ tissues. We would like to test if the normal and tumor colon tissue samples have the same average

Table 3: p-values of test procedures for the real data

NEW	CQ	EB	WB	ZZ	ZZGZ	LOU	SD	SKK	FZWZ
9.99e-4	1.89e-10	9.99e-4	9.99e-4	7.74e-4	7.47e-5	0	0.27	0.18	4.78e-3

NEW, the proposed test procedure; CQ, the test of Chen and Qin (2010); EB, the empirical bootstrap method based on $T_{\text{CQ}}(\mathbf{X}_1, \mathbf{X}_2)$; WB, the wild bootstrap method based on $T_{\text{CQ}}(\mathbf{X}_1, \mathbf{X}_2)$; ZZ, the test of Zhang and Zhu (2021); ZZGZ, the test of Zhang et al. (2021); LOU, the test of Lou (2020); SD, the test of Srivastava and Du (2008); SKK, the test of Srivastava et al. (2013); FZWZ, the test of Feng et al. (2015).

Table 4: Empirical sizes (multiplied by 100) of test procedures for the resampled real datasets

NEW	CQ	EB	WB	ZZ	ZZGZ	LOU	SD	SKK	FZWZ
5.43	7.28	5.22	5.84	5.19	5.98	5.71	0.60	0.79	7.03

NEW, the proposed test procedure; CQ, the test of Chen and Qin (2010); EB, the empirical bootstrap method based on $T_{\text{CQ}}(\mathbf{X}_1, \mathbf{X}_2)$; WB, the wild bootstrap method based on $T_{\text{CQ}}(\mathbf{X}_1, \mathbf{X}_2)$; ZZ, the test of Zhang and Zhu (2021); ZZGZ, the test of Zhang et al. (2021); LOU, the test of Lou (2020); SD, the test of Srivastava and Du (2008); SKK, the test of Srivastava et al. (2013); FZWZ, the test of Feng et al. (2015).

gene expression levels. Table 3 lists the p-values of various test procedures. With $\alpha = 0.05$, all but the test procedures of Srivastava and Du (2008) and Srivastava et al. (2013) reject the null hypothesis, claiming that the average gene expression levels of normal and tumor colon tissue samples are significantly different.

We would also like to examine the empirical sizes of various test procedures on the gene expression data. To mimick the null distribution of the gene expression data, we generate resampled datasets as follows: the resampled observations $\{X_{1,i}^*\}_{i=1}^{22}$ are uniformly sampled from $\{X_{1,i} - \bar{X}_1\}_{i=1}^{22}$ with replacement, and $\{X_{2,i}^*\}_{i=1}^{40}$ are uniformly sampled from $\{X_{2,i} - \bar{X}_2\}_{i=1}^{40}$ with replacement. We conduct various test procedures with $\alpha = 0.05$ on the resampled observations $\{X_{k,i}^*\}$. The above procedure is independently replicated for $10,000$ times to compute the empirical sizes. The results are listed in Table 4. It can be seen that the test procedures of Srivastava and Du (2008) and Srivastava et al. (2013) are overly conservative. Hence the p-values of these two test procedures for gene expression data may not be reliable. The test procedures of Chen and Qin (2010) and Feng et al. (2015) are a little inflated. In comparison, the remaining test procedures, including the proposed test procedure, have a good control of the test level for the resampled gene expression data. This implies that the p-value of the proposed test procedure for the gene expression data is reliable.

Acknowledgements

The authors thank the editor, associate editor and three reviewers for their valuable comments and suggestions. This work was supported by Beijing Natural Science Foundation (No Z200001), National Natural Science Foundation of China (No 11971478). Wangli Xu serves as the corresponding author of the present paper.

References

Alon, U., Barkai, N., Notterman, D. A., Gish, K., Ybarra, S., Mack, D., and Levine, A. J. (1999). Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proceedings of the National Academy of Sciences*, 96(12):6745–6750.

Bai, Z. and Saranadasa, H. (1996). Effect of high dimension: by an example of a two sample problem. *Statistica Sinica*, 6(2):311–329.

Cai, T., Liu, W., and Xia, Y. (2013). Two-sample test of high dimensional means under dependence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(2):349–372.

Canay, I. A., Romano, J. P., and Shaikh, A. M. (2017). Randomization tests under an approximate symmetry assumption. *Econometrica*, 85(3):1013–1030.

Chang, J., Zheng, C., Zhou, W.-X., and Zhou, W. (2017). Simulation-based hypothesis testing of high dimensional means under covariance heterogeneity. *Biometrics*, 73(4):1300–1310.

Chatterjee, S. (2006). A generalization of the lindeberg principle. *The Annals of Probability*, 34(6):2061–2076.

Chen, S. X. and Qin, Y. L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *The Annals of Statistics*, 38(2):808–835.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *The Annals of Statistics*, 41(6):2786–2819.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2017). Central limit theorems and bootstrap in high dimensions. *The Annals of Probability*, 45(4):2309–2352.

Chung, E. and Romano, J. P. (2013). Exact and asymptotically robust permutation tests. *The Annals of Statistics*, 41(2):484–507.

Cohn, D. L. (2013). *Measure Theory*. Birkhäuser, New York, 2nd edition.

de Jong, P. (1987). A central limit theorem for generalized quadratic forms. *Probability Theory & Related Fields*, 75(2):261–277.

Deng, H. and Zhang, C.-H. (2020). Beyond Gaussian approximation: bootstrap for maxima of sums of independent random vectors. *The Annals of Statistics*, 48(6):3643–3671.

Döbler, C. and Peccati, G. (2017). Quantitative de jong theorems in any dimension. *Electronic Journal of Probability*, 22(0).

Dudley, R. M. (2002). *Real Analysis and Probability*. Cambridge University Press.

Fan, J., Wang, K., Zhong, Y., and Zhu, Z. (2021). Robust high-dimensional factor models with applications to statistical machine learning. *Statistical Science*, 36(2):303–327.

Fang, K. T., Kotz, S., and Ng, K. W. (1990). *Symmetric multivariate and related distributions*. Chapman and Hall, Ltd., London.

Feng, L., Zou, C., Wang, Z., and Zhu, L. (2015). Two-sample Behrens-Fisher problem for high-dimensional data. *Statistica Sinica*, 25(4):1297–1312.

Fisher, R. A. (1935). *The design of experiments*. Oliver and Boyd, Edinburgh, 1 edition.

Hemerik, J. and Goeman, J. (2018). Exact testing with random permutations. *TEST*, 27(4):811–825.

Hoeffding, W. (1952). The large-sample power of tests based on permutations of observations. *The Annals of Mathematical Statistics*, 23(2):169–192.

Hu, J. and Bai, Z. (2016). A review of 20 years of naive tests of significance for high-dimensional mean vectors and covariance matrices. *Science China. Mathematics*, 59(12):2281–2300.

Kong, W. and Valiant, G. (2017). Spectrum estimation from samples. *The Annals of Statistics*, 45(5):2218–2247.

Lehmann, E. L. and Romano, J. P. (2005). *Testing Statistical Hypotheses*. Springer, New York, 3rd edition.

Linnik, J. V. (1966). Latest investigations on Behrens-Fisher problem. *Sankhyā. Series A*, 28:15–24.

Lou, Z. (2020). *High Dimensional Inference Based on Quadratic Forms*. Thesis (Ph.D.)–The University of Chicago.

Mossel, E., O'Donnell, R., and Oleszkiewicz, K. (2010). Noise stability of functions with low influences: invariance and optimality. *Annals of Mathematics. Second Series*, 171(1):295–341.

Nourdin, I., Peccati, G., and Reinert, G. (2010). Invariance principles for homogeneous sums: universality of Gaussian Wiener chaos. *The Annals of Probability*, 38(5):1947–1985.

Pollard, D. (1984). *Convergence of stochastic processes*. Springer, New York, 1st edition.

Romano, J. P. (1990). On the behavior of randomization tests without a group invariance assumption. *Journal of the American Statistical Association*, 85(411):686–692.

Simon, B. (2015). *Real analysis*. A Comprehensive Course in Analysis, Part 1. American Mathematical Society, Providence, RI. With a 68 page companion booklet.

Srivastava, M. S. and Du, M. (2008). A test for the mean vector with fewer observations than the dimension. *Journal of Multivariate Analysis*, 99(3):386–402.

Srivastava, M. S., Katayama, S., and Kano, Y. (2013). A two sample test in high dimensional data. *Journal of Multivariate Analysis*, 114(Supplement C):349–358.

Tao, T. (2012). *Topics in random matrix theory*, volume 132 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI.

Tropp, J. A. (2015). An introduction to matrix concentration inequalities. *Foundations and Trends® in Machine Learning*, 8(1-2):1–230.

Wang, R. and Xu, X. (2019). A feasible high dimensional randomization test for the mean vector. *Journal of Statistical Planning and Inference*, 199:160–178.

Wu, W. B., Lou, Z., and Han, Y. (2018). Hypothesis testing for high-dimensional data. In *Handbook of big data analytics*, pages 203–224.

Xu, M., Zhang, D., and Wu, W. B. (2019). Pearson's chi-squared statistics: approximation theory and beyond. *Biometrika*, 106(3):716–723.

Xue, K. and Yao, F. (2020). Distribution and correlation-free two-sample test of high-dimensional means. *The Annals of Statistics*, 48(3):1304–1328.

Zhang, J.-T., Guo, J., Zhou, B., and Cheng, M.-Y. (2020). A simple two-sample test in high dimensions based on L^2-norm. *Journal of the American Statistical Association*, 115(530):1011–1027.

Zhang, J.-T., Zhou, B., Guo, J., and Zhu, T. (2021). Two-sample Behrens-Fisher problems for high-dimensional data: a normal reference approach. *Journal of Statistical Planning and Inference*, 213:142–161.

Zhang, J.-T. and Zhu, T. (2021). A further study on chen-qin's test for two-sample behrens-fisher problems for high-dimensional data. Department of Statistics and Applied Probability, National University of Singapore.

Zhou, B. and Guo, J. (2017). A note on the unbiased estimator of Σ^2. *Statistics & Probability Letters*, 129:141–146.

Zhu, L. (2005). *Nonparametric Monte Carlo tests and their applications*. Springer, New York.

S.1 Additional numerical results

In this section, we present additional numerical results. The experimental setting is as described in the main text.

We would like to use quantile-quantile plots to examine the correctness of Theorem 1 and Corollary 1. First we consider the correctness of Theorem 1. Theorem 1 implies that the distribution of $T_{\mathrm{CQ}(\mathbf{Y}_1,\mathbf{Y}_2)}/\sigma_{T,n}$ can be approximated by that of $\{\boldsymbol{\xi}_p^\intercal \boldsymbol{\Psi}_n \boldsymbol{\xi}_p - \operatorname{tr}(\boldsymbol{\Psi}_n)\}/\{2\operatorname{tr}(\boldsymbol{\Psi}_n^2)\}^{1/2}$. Fig. 1 illustrates the plots of the empirical quantiles of $T_{\mathrm{CQ}(\mathbf{Y}_1,\mathbf{Y}_2)}/\sigma_{T,n}$ against that of $\{\boldsymbol{\xi}_p^\intercal \boldsymbol{\Psi}_n \boldsymbol{\xi}_p - \operatorname{tr}(\boldsymbol{\Psi}_n)\}/\{2\operatorname{tr}(\boldsymbol{\Psi}_n^2)\}^{1/2}$ under Models I-IV described in the main text with $n_1 = 16$, $n_2 = 24$, $p = 300$. The empirical quantiles of $T_{\mathrm{CQ}}(\mathbf{Y}_1,\mathbf{Y}_2)/\sigma_{T,n}$ are obtained by $10,000$ replications. The results imply that the distribution approximation in Theorem 1 is quite accurate for finite sample size.

Now we consider the correctness of Corollary 1. Corollary 1 claims that the general asymptotic distributions of $T_{\mathrm{CQ}}(\mathbf{Y}_1,\mathbf{Y}_2)/\sigma_{T,n}$ are weighted sums of independent normal random variable and centered χ^2 random variables. In Corollay 1, the parameters $\{\kappa_i\}_{i=1}^{\infty}$ relies on the limits of the eigenvalues of $\boldsymbol{\Psi}_n/\{\operatorname{tr}(\boldsymbol{\Psi}_n^2)\}^{1/2}$ along a subsequence of $\{n\}$. To cover different senarios of asymptotic distributions, we consider the following model. Suppose $Y_{k,i} \sim \mathcal{N}\{\mathbf{0}_p, \gamma \mathbf{1}_p \mathbf{1}_p^\intercal + (1-\gamma)\mathbf{I}_p\}$, $i = 1,\ldots,n_k$, $k = 1,2$, where $\gamma \in [0,1]$. In this case, the eigenvalues of $\boldsymbol{\Psi}_n/\{\operatorname{tr}(\boldsymbol{\Psi}_n^2)\}^{1/2}$ are

$$\frac{p\gamma + 1 - \gamma}{\{(p\gamma + 1 - \gamma)^2 + (p-1)(1-\gamma)^2\}^{1/2}} \quad \text{and} \quad \frac{1-\gamma}{\{(p\gamma + 1 - \gamma)^2 + (p-1)(1-\gamma)^2\}^{1/2}},$$

with multiplicities 1 and $p-1$, respectively. We assume $p \to \infty$ as $n \to \infty$. We consider four choices of γ. First, we consider $\gamma = 0$. In this case, $\kappa_i = 0$, $i = 1, 2, \ldots$, and the asymptotic distribution of

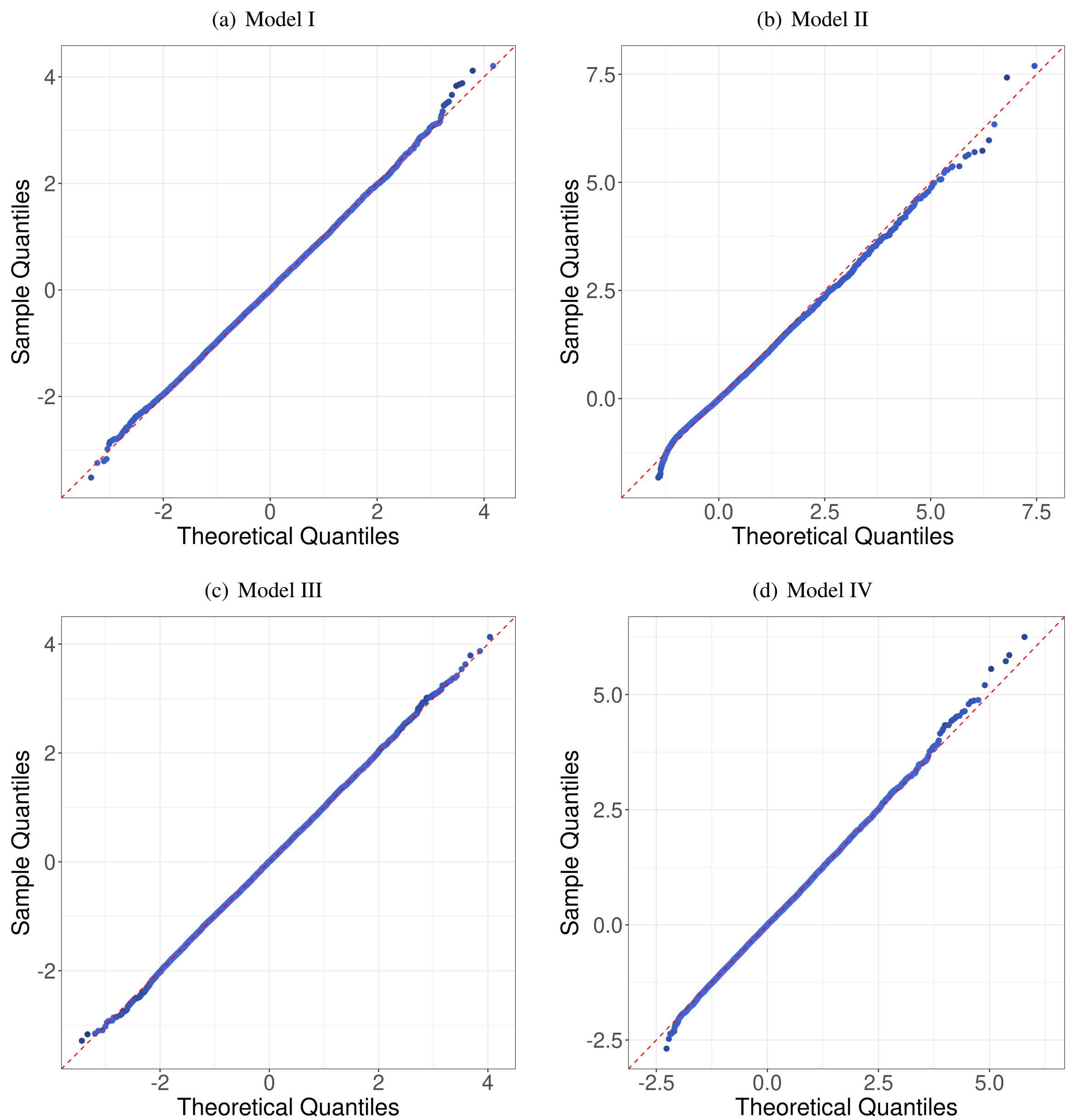

Figure 1: Plots of the empirical quantiles of $T_{\mathrm{CQ}(\mathbf{Y}_1,\mathbf{Y}_2)}/\sigma_{T,n}$ against that of $(\boldsymbol{\xi}_p^\intercal \boldsymbol{\Psi}_n \boldsymbol{\xi}_p - \mathrm{tr}(\boldsymbol{\Psi}_n))/\{2\,\mathrm{tr}(\boldsymbol{\Psi}_n^2)\}^{1/2}$. $n_1 = 16$, $n_2 = 24$, $p = 300$.

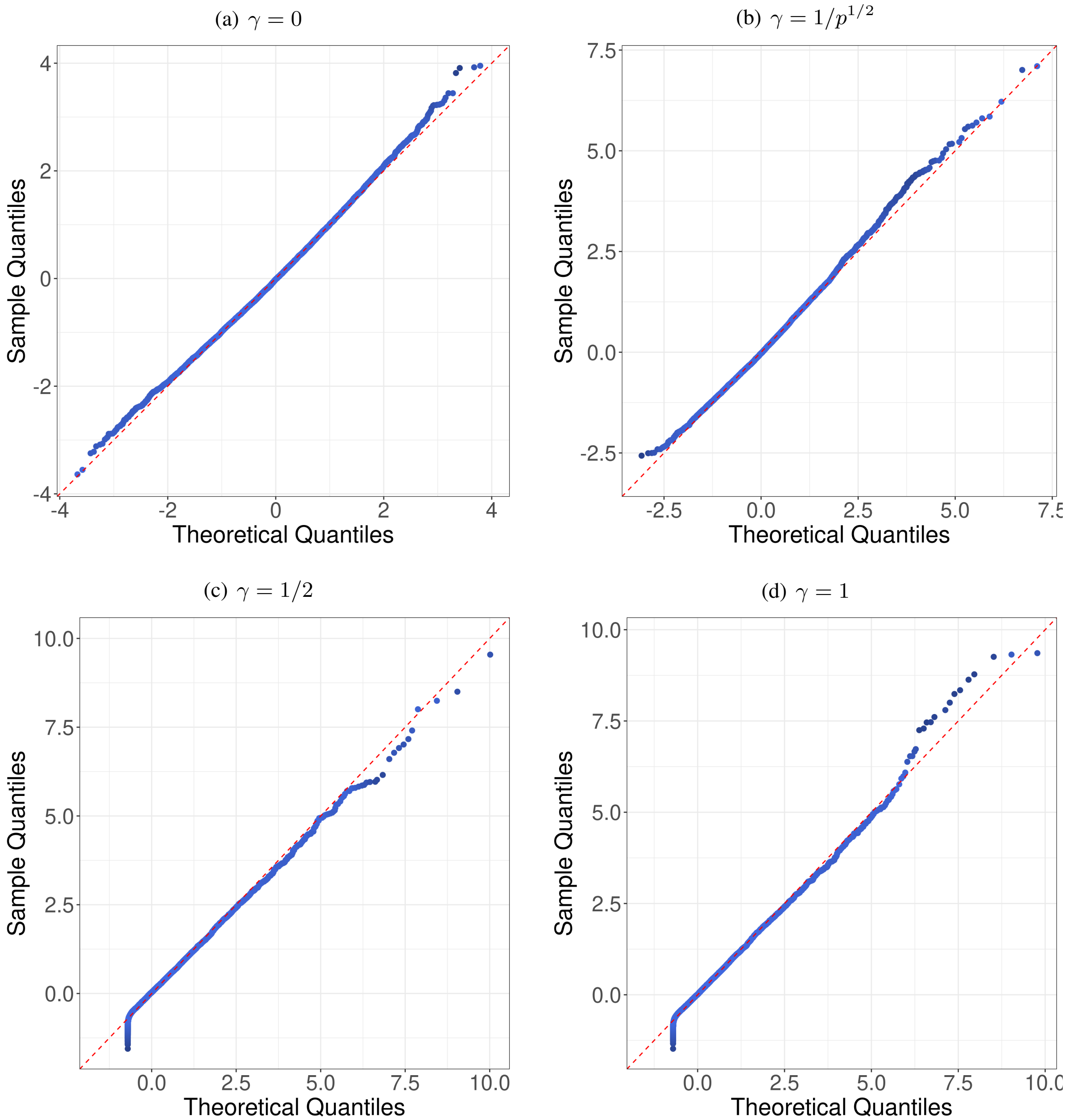

Figure 2: Plots of the empirical quantiles of $T_{\mathrm{CQ}}(\mathbf{Y}_1, \mathbf{Y}_2)/\sigma_{T,n}$ against that of the asymptotic distribution in (6). $n_1 = 16$, $n_2 = 24$, $p = 300$.

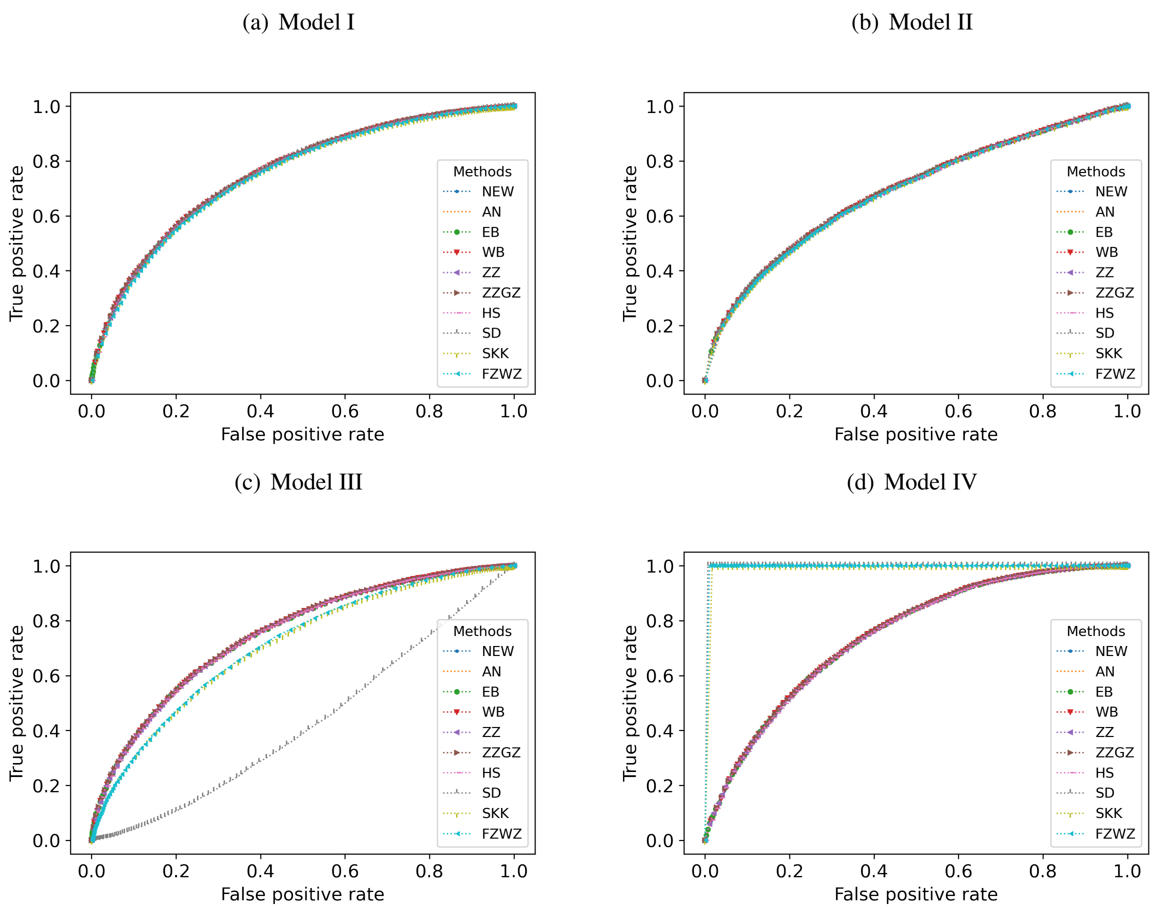

Figure 3: Receiver operating characteristic curve of various test procedures. NEW, the proposed test procedure; CQ, the test procedure of Chen and Qin (2010); EB, the empirical bootstrap method based on $T_{\mathrm{CQ}}(\mathbf{X}_1, \mathbf{X}_2)$; WB, the wild bootstrap method based on $T_{\mathrm{CQ}}(\mathbf{X}_1, \mathbf{X}_2)$; ZZ, the test procedure of Zhang and Zhu (2021); ZZGZ, the test procedure of Zhang et al. (2021); LOU, the test procedure of Lou (2020); SD, the test procedure of Srivastava and Du (2008); SKK, the test procedure of Srivastava et al. (2013); FZWZ, the test procedure of Feng et al. (2015).

$T_{\rm CQ}(\mathbf{Y}_1,\mathbf{Y}_2)/\sigma_{T,n}$ is the standard normal distribution. Second, we consider $\gamma = 1/p^{1/2}$. In this case, $\kappa_1 = 1/\sqrt{2}$, and $\kappa_i = 0$, $i = 2, 3, \ldots$, and the asymptotic distribution of $T_{\rm CQ}(\mathbf{Y}_1,\mathbf{Y}_2)/\sigma_{T,n}$ is $\mathcal{N}(0,1)/\sqrt{2} + \{\chi^2(1) - 1\}/2$. In the third and fourth cases, we consider $\gamma = 1/2$ and 1, respectively. In these two cases, $\kappa_1 = 1$, and $\kappa_i = 0$, $i = 2, 3, \ldots$, and the asymptotic distribution of $T_{\rm CQ}(\mathbf{Y}_1,\mathbf{Y}_2)/\sigma_{T,n}$ is the standardized χ^2 distribution with 1 degree of freedom. Fig. 2 illustrates the plots of the empirical quantiles of $T_{\rm CQ}(\mathbf{Y}_1,\mathbf{Y}_2)/\sigma_{T,n}$ against that of the asymptotic distribution in (6) for various values of γ. The empirical quantiles of $T_{\rm CQ}(\mathbf{Y}_1,\mathbf{Y}_2)/\sigma_{T,n}$ are obtained by $10,000$ replications. It can be seen that the distribution of $T_{\rm CQ}(\mathbf{Y}_1,\mathbf{Y}_2)/\sigma_{T,n}$ can be well approximated by the asymptotic distributions given in Corollary 1. This verifies the conclusion of Corollary 1. We note that the approximation $\{\boldsymbol{\xi}_p^\intercal \boldsymbol{\Psi}_n \boldsymbol{\xi}_p - \operatorname{tr}(\boldsymbol{\Psi}_n)\}/\{2\operatorname{tr}(\boldsymbol{\Psi}_n^2)\}^{1/2}$ in Theorem 1 is slightly better than the distribution approximations in Corollary 1. This phenomenon is reasonable since the distributions in Corollary 1 are in fact the asymptotic distributions of $\{\boldsymbol{\xi}_p^\intercal \boldsymbol{\Psi}_n \boldsymbol{\xi}_p - \operatorname{tr}(\boldsymbol{\Psi}_n)\}/\{2\operatorname{tr}(\boldsymbol{\Psi}_n^2)\}^{1/2}$ in Theorem 1, and hence may have larger approximation error than $\{\boldsymbol{\xi}_p^\intercal \boldsymbol{\Psi}_n \boldsymbol{\xi}_p - \operatorname{tr}(\boldsymbol{\Psi}_n)\}/\{2\operatorname{tr}(\boldsymbol{\Psi}_n^2)\}^{1/2}$.

We have seen that many competing test procedures do not have a good control of test level. To get rid of the effect of distorted test level, we plot the receiver operating characteristic curve of the test procedures. Fig. 3 illustrates the receiver operating characteristic curve of various test procedures with $n_1 = 16$, $n_2 = 24$ and $p = 300$. It can be seen that for Models I and II, all test procedures have similar power behavior. For Model III, the scalar-invariant tests are less powerful than other tests. For Model IV, the scalar-invariant tests are more powerful than other tests. These results show that various test procedures based on $\|\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2\|^2$ or $T_{\rm CQ}(\mathbf{X}_1,\mathbf{X}_2)$ may not have essential difference in power, and their performances are largely driven by the test level.

S.2 Universality of generalized quadratic forms

In this section, we investigate the universality property of generalized quadratic forms, which is the key tool to study the distributional behavior of the proposed test procedure. The result in this section is also interesting in its own right.

Suppose $\xi_1, \ldots, \xi_n$ are independent random elements taking values in a Polish space $\mathcal{X}$. We consider the generalized quadratic form

$$W(\xi_1, \ldots, \xi_n) = \sum_{1 \le i < j \le n} w_{i,j}(\xi_i, \xi_j),$$

where $w_{i,j}(\cdot,\cdot) : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is measurable with respect to the product σ-algebra on $\mathcal{X} \times \mathcal{X}$, $1 \le i < j \le n$. The generalized quadratic form includes the statistic $T_{\rm CQ}(\mathbf{Y}_1,\mathbf{Y}_2)$ as a special case. To see this, consider $\xi_i = Y_{1,i}$, $i = 1, \ldots, n_1$ and $\xi_j = Y_{2,j-n_1}$, $j = n_1 + 1, \ldots, n$. Let

$$w_{i,j}(\xi_i, \xi_j) = \begin{cases} \frac{2\xi_i^\intercal \xi_j}{n_1(n_1-1)} & \text{for } 1 \le i < j \le n_1, \\ \frac{-2\xi_i^\intercal \xi_j}{n_1 n_2} & \text{for } 1 \le i \le n_1 \text{ and } n_1 + 1 \le j \le n, \\ \frac{2\xi_i^\intercal \xi_j}{n_2(n_2-1)} & \text{for } n_1 + 1 \le i < j \le n. \end{cases}$$

In this case, the generalized quadratic form $W(\xi_1, \ldots, \xi_n)$ becomes the statistic $T_{\rm CQ}(\mathbf{Y}_1,\mathbf{Y}_2)$. Similarly, conditioning on $\tilde{\mathbf{X}}_1$ and $\tilde{\mathbf{X}}_2$, the randomized statistic $T_{\rm CQ}(E; \tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2)$ is a special case of the generalized quadratic form with $\xi_i = \epsilon_{1,i}$, $i = 1, \ldots, m_1$ and $\xi_j = \epsilon_{2,j-m_1}$, $j = m_1 + 1, \ldots, m_1 + m_2$. Hence it is meaningful to investigate the general behavior of the generalized quadratic form.

The asymptotic normality of the generalized quadratic forms was studied by de Jong (1987) via martingale central limit theorem and by Döbler and Peccati (2017) via Stein's method. However, we are interested in the general setting in which $W(\xi_1, \ldots, \xi_n)$ may not be asymptotically normally distributed.

Therefore, compared with the asymptotic normality, we are more interested in the *universality* property of $W(\xi_1, \ldots, \xi_n)$; i.e., the distributional behavior of $W(\xi_1, \ldots, \xi_n)$ does not rely on the particular distribution of $\xi_1, \ldots, \xi_n$ asymptotically. In this regard, many achievements have been made for the universality of $W(\xi_1, \ldots, \xi_n)$ for special form of $w_{i,j}(\xi_i, \xi_j)$; see, e.g., Mossel et al. (2010), Nourdin et al. (2010), Xu et al. (2019) and the references therein. However, these results can not be used to deal with $T_{\mathrm{CQ}}(\mathbf{Y}_1, \mathbf{Y}_2)$ in our setting. In fact, the results in Mossel et al. (2010) and Nourdin et al. (2010) can not be readily applied to $T_{\mathrm{CQ}}(\mathbf{Y}_1, \mathbf{Y}_2)$ while the result in Xu et al. (2019) can only be applied to identically distributed observations. To the best of our knowledge, the universality of the generalized quadratic forms was never considered in the literature. We shall derive a universality property of the generalized quadratic forms using Lindeberg principle, an old and powerful technique; see, e.g., Chatterjee (2006), Mossel et al. (2010) for more about Lindeberg principle.

Assumption S.1. *Suppose $\xi_1, \ldots, \xi_n$ are independent random elements taking values in a Polish space $\mathcal{X}$. Assume the following conditions hold for all $1 \le i < j \le n$:*

(a) $\mathrm{E}\{w_{i,j}(\xi_i, \xi_j)^4\} < \infty$.

(b) For all $\mathbf{a} \in \mathcal{X}$, $\mathrm{E}\left\{w_{i,j}(\xi_i, \mathbf{a})\right\} = \mathrm{E}\left\{w_{i,j}(\mathbf{a}, \xi_j)\right\} = 0$.

Define $\sigma_{i,j}^2 = \mathrm{E}\{w_{i,j}(\xi_i, \xi_j)^2\}$. Under Assumption S.1, we have $\mathrm{E}\{W(\xi_1, \ldots, \xi_n)\} = 0$ and $\mathrm{var}\{W(\xi_1, \ldots, \xi_n)\} = \sum_{1=1}^n \sum_{j=i+1}^n \sigma_{i,j}^2$. We would like to give explicit bound for the difference between the distributions of $W(\xi_1, \ldots, \xi_n)$ and $W(\eta_1, \ldots, \eta_n)$ for a general class of random vectors $\eta_1, \ldots, \eta_n$. We impose the following conditions on $\eta_1, \ldots, \eta_n$.

Assumption S.2. *Suppose $\eta_1, \ldots, \eta_n$ are independent random elements taking values in $\mathcal{X}$ and are independent of $\xi_1, \ldots, \xi_n$. Assume the following conditions hold for all $1 \le i < j \le n$:*

(a) $\mathrm{E}\{w_{i,j}(\xi_i, \eta_j)^4\} < \infty$, $\mathrm{E}\{w_{i,j}(\eta_i, \xi_j)^4\} < \infty$ *and* $\mathrm{E}\{w_{i,j}(\eta_i, \eta_j)^4\} < \infty$.

(b) For all $\mathbf{a} \in \mathcal{X}$, $\mathrm{E}\{w_{i,j}(\eta_i, \mathbf{a})\} = \mathrm{E}\{w_{i,j}(\mathbf{a}, \eta_j)\} = 0$.

(c) For any $\mathbf{a}, \mathbf{b} \in \mathcal{X}$,

$$
\begin{aligned}
\mathrm{E}\{w_{i,k}(\mathbf{a}, \xi_k) w_{j,k}(\mathbf{b}, \xi_k)\} &= \mathrm{E}\{w_{i,k}(\mathbf{a}, \eta_k) w_{j,k}(\mathbf{b}, \eta_k)\}, \quad \text{for} \quad 1 \le i \le j < k \le n, \\
\mathrm{E}\{w_{i,j}(\mathbf{a}, \xi_j) w_{j,k}(\xi_j, \mathbf{b})\} &= \mathrm{E}\{w_{i,j}(\mathbf{a}, \eta_j) w_{j,k}(\eta_j, \mathbf{b})\}, \quad \text{for} \quad 1 \le i < j < k \le n, \\
\mathrm{E}\{w_{i,j}(\xi_i, \mathbf{a}) w_{i,k}(\xi_i, \mathbf{b})\} &= \mathrm{E}\{w_{i,j}(\eta_i, \mathbf{a}) w_{i,k}(\eta_i, \mathbf{b})\}, \quad \text{for} \quad 1 \le i < j \le k \le n.
\end{aligned}
$$

We claim that under Assumptions S.1 and S.2, there exists a nonnegative C (which possibly depends on n) such that for all $1 \le i < j \le n$,

$$
\max\left[\mathrm{E}\{w_{i,j}(\xi_i, \xi_j)^4\}, \mathrm{E}\{w_{i,j}(\xi_i, \eta_j)^4\}, \mathrm{E}\{w_{i,j}(\eta_i, \xi_j)^4\}, \mathrm{E}\{w_{i,j}(\eta_i, \eta_j)^4\}\right] \le C\sigma_{i,j}^4. \tag{S.1}
$$

In fact, by Assumption S.1, (a) and Assumption S.2, (a), the left hand side of (S.1) is finite. Also, if $\sigma_{i,j}^4 = 0$, that is, $\mathrm{E}\{w_{i,j}(\xi_i, \xi_j)^2\} = 0$, then from (c) of Assumption S.2,

$$
0 = \mathrm{E}\{w_{i,j}(\xi_i, \xi_j)^2\} = \mathrm{E}\{w_{i,j}(\xi_i, \eta_j)^2\} = \mathrm{E}\{w_{i,j}(\eta_i, \xi_j)^2\} = \mathrm{E}\{w_{i,j}(\eta_i, \eta_j)^2\}.
$$

It follows that $w_{i,j}(\xi_i, \xi_j) = w_{i,j}(\xi_i, \eta_j) = w_{i,j}(\eta_i, \xi_j) = w_{i,j}(\eta_i, \eta_j) = 0$ almost surely. In this case, the left hand side of (S.1) is also 0. Hence our claim is valid. Let ρ_n denote the minimum nonnegative C such that (S.1) holds for all $1 \le i < j \le n$, It will turn out that the difference between the distributions of $W(\xi_1, \ldots, \xi_n)$ and $W(\eta_1, \ldots, \eta_n)$ relies on ρ_n.

As in Mossel et al. (2010), we define the *influence* of ξ_i on $W(\xi_1,\dots,\xi_n)$ as

$$\mathrm{Inf}_i = \mathrm{E}\left\{\mathrm{var}(W(\xi_1,\dots,\xi_n) \mid \xi_1,\dots,\xi_{i-1},\xi_{i+1},\dots,\xi_n)\right\}.$$

It can be seen that $\mathrm{Inf}_i = \sum_{j=1}^{i-1}\sigma_{j,i}^2 + \sum_{j=i+1}^{n}\sigma_{i,j}^2$.

The following theorem provides a universality property of $W(\xi_1,\dots,\xi_n)$.

Theorem S.1. *Under Assumptions S.1 and S.2, we have*

$$\|\mathcal{L}\{W(\xi_1,\dots,\xi_n)\} - \mathcal{L}\{W(\eta_1,\dots,\eta_n)\}\|_3 \le \frac{\rho_n^{3/4}}{3^{1/4}}\sum_{i=1}^{n}\mathrm{Inf}_i^{3/2}.$$

From Theorem S.1, the distance between $W(\xi_1,\dots,\xi_n)$ and $W(\eta_1,\dots,\eta_n)$ is bounded by a function of ρ_n and the influences. Suppose as $n\to\infty$, ρ_n is bounded and $\sum_{i=1}^{n}\mathrm{Inf}_i^{3/2}$ tends to 0. Then Theorem S.1 implies that $W(\xi_1,\dots,\xi_n)$ and $W(\eta_1,\dots,\eta_n)$ share the same possible asymptotic dsitribution. That is, the distribution of $W(\xi_1,\dots,\xi_n)$ enjoys a universality property. In the proof of Theorem 1 and Theorem 2, we apply Theorem S.1 to $T_{\mathrm{CQ}}(\mathbf{Y}_1,\mathbf{Y}_2)$ and $T_{\mathrm{CQ}}(E;\tilde{\mathbf{X}}_1,\tilde{\mathbf{X}}_2)$, respectively, and consider normally distributed η_i, $i=1,\dots,n$. With this technique, the distributional behaviors of $T_{\mathrm{CQ}}(\mathbf{Y}_1,\mathbf{Y}_2)$ and $T_{\mathrm{CQ}}(E;\tilde{\mathbf{X}}_1,\tilde{\mathbf{X}}_2)$ are reduced to the circumstances where the observations are normally distributed.

of Theorem S.1. For $k=1,\dots,n+1$, define

$$W_k = W(\eta_1,\dots,\eta_{k-1},\xi_k,\dots,\xi_n).$$

Then $W_1 = W(\xi_1,\dots,\xi_n)$ and $W_{n+1} = W(\eta_1,\dots,\eta_n)$. Fix an $f\in\mathscr{C}_b^3(\mathbb{R})$. We have

$$|\mathrm{E} f(W(\xi_1,\dots,\xi_n)) - \mathrm{E} f(W(\eta_1,\dots,\eta_n))| \le \sum_{k=1}^{n}\left|\mathrm{E}\left\{f(W_k)-f(W_{k+1})\right\}\right|. \tag{S.2}$$

Define

$$W_{k,0} = \sum_{1\le i<j\le k-1} w_{i,j}(\eta_i,\eta_j) + \sum_{k+1\le i<j\le n} w_{i,j}(\xi_i,\xi_j) + \sum_{1\le i\le k-1}\sum_{k+1\le j\le n} w_{i,j}(\eta_i,\xi_j).$$

Note that $W_{k,0}$ only relies on $\eta_1,\dots,\eta_{k-1},\xi_{k+1},\dots,\xi_n$. It can be seen that

$$\begin{aligned} W_k =& W_{k,0} + \sum_{i=1}^{k-1} w_{i,k}(\eta_i,\xi_k) + \sum_{j=k+1}^{n} w_{k,j}(\xi_k,\xi_j),\\ W_{k+1} =& W_{k,0} + \sum_{i=1}^{k-1} w_{i,k}(\eta_i,\eta_k) + \sum_{j=k+1}^{n} w_{k,j}(\eta_k,\xi_j). \end{aligned}$$

From Taylor's theorem,

$$\left|f(W_k) - f(W_{k,0}) - \sum_{i=1}^{2}\frac{1}{i!}(W_k - W_{k,0})^i f^{(i)}(W_{k,0})\right| \le \frac{\sup_{x\in\mathbb{R}}|f^{(3)}(x)|}{6}\left|W_k - W_{k,0}\right|^3, \tag{S.3}$$

$$\left|f(W_{k+1}) - f(W_{k,0}) - \sum_{i=1}^{2}\frac{1}{i!}(W_{k+1} - W_{k,0})^i f^{(i)}(W_{k,0})\right| \le \frac{\sup_{x\in\mathbb{R}}|f^{(3)}(x)|}{6}\left|W_{k+1} - W_{k,0}\right|^3. \tag{S.4}$$

Now we show that

$$\mathrm{E}\left\{\sum_{i=1}^{2}\frac{1}{i!}(W_k - W_{k,0})^i f^{(i)}(W_{k,0})\right\} = \mathrm{E}\left\{\sum_{i=1}^{2}\frac{1}{i!}(W_{k+1} - W_{k,0})^i f^{(i)}(W_{k,0})\right\}. \tag{S.5}$$

By conditioning on $\eta_1,\ldots,\eta_{k-1},\xi_{k+1},\ldots,\xi_n$, it can be seen that (S.5) holds provided that for $k = 1,\ldots,n$ and $\ell = 1, 2$,

$$\mathrm{E}\{(W_k - W_{k,0})^\ell \mid \eta_1,\ldots,\eta_{k-1},\xi_{k+1},\ldots,\xi_n\} = \mathrm{E}\{(W_{k+1} - W_{k,0})^\ell \mid \eta_1,\ldots,\eta_{k-1},\xi_{k+1},\ldots,\xi_n\}.$$

For the case of $\ell = 1$, we have

$$\mathrm{E}\{W_k - W_{k,0} \mid \eta_1,\ldots,\eta_{k-1},\xi_{k+1},\ldots,\xi_n\} = \sum_{i=1}^{k-1}\mathrm{E}\{w_{i,k}(\eta_i,\xi_k) \mid \eta_i\} + \sum_{j=k+1}^{n}\mathrm{E}\{w_{k,j}(\xi_k,\xi_j) \mid \xi_j\},$$

which equals 0 by (b) of Assumption S.1. Similarly, we have

$$\mathrm{E}\{W_{k+1} - W_{k,0} \mid \eta_1,\ldots,\eta_{k-1},\xi_{k+1},\ldots,\xi_n\} = 0.$$

Now we deal with the case of $\ell = 2$. From (c) of Assumption S.2, we have

$$\begin{aligned}
&\mathrm{E}\{(W_k - W_{k,0})^2 \mid \eta_1,\ldots,\eta_{k-1},\xi_{k+1},\ldots,\xi_n\}\\
=&\sum_{i_1=1}^{k-1}\sum_{i_2=1}^{k-1}\mathrm{E}\{w_{i_1,k}(\eta_{i_1},\xi_k)w_{i_2,k}(\eta_{i_2},\xi_k) \mid \eta_{i_1},\eta_{i_2}\} + \sum_{j_1=k+1}^{n}\sum_{j_2=k+1}^{n}\mathrm{E}\{w_{k,j_1}(\xi_k,\xi_{j_1})w_{k,j_2}(\xi_k,\xi_{j_2}) \mid \xi_{j_1},\xi_{j_2}\}\\
&+ 2\sum_{i=1}^{k-1}\sum_{j=k+1}^{n}\mathrm{E}\{w_{i,k}(\eta_i,\xi_k)w_{k,j}(\xi_k,\xi_j) \mid \eta_i,\xi_j\}\\
=&\sum_{i_1=1}^{k-1}\sum_{i_2=1}^{k-1}\mathrm{E}\{w_{i_1,k}(\eta_{i_1},\eta_k)w_{i_2,k}(\eta_{i_2},\eta_k) \mid \eta_{i_1},\eta_{i_2}\} + \sum_{j_1=k+1}^{n}\sum_{j_2=k+1}^{n}\mathrm{E}\{w_{k,j_1}(\eta_k,\xi_{j_1})w_{k,j_2}(\eta_k,\xi_{j_2}) \mid \xi_{j_1},\xi_{j_2}\}\\
&+ 2\sum_{i=1}^{k-1}\sum_{j=k+1}^{n}\mathrm{E}\{w_{i,k}(\eta_i,\eta_k)w_{k,j}(\eta_k,\xi_j) \mid \eta_i,\xi_j\}\\
=&\,\mathrm{E}\{(W_{k+1} - W_{k,0})^2 \mid \eta_1,\ldots,\eta_{k-1},\xi_{k+1},\ldots,\xi_n\}.
\end{aligned}$$

Therefore, (S.5) holds. It follows from (S.3), (S.4) and (S.5) that

$$\begin{aligned}
|\mathrm{E}\,f(W_{k+1}) - \mathrm{E}\,f(W_k)| \le&\frac{\sup_{x\in\mathbb{R}}|f^{(3)}(x)|}{6}\left(\mathrm{E}\,|W_k - W_{k,0}|^3 + \mathrm{E}\,|W_{k+1} - W_{k,0}|^3\right)\\
\le&\frac{\sup_{x\in\mathbb{R}}|f^{(3)}(x)|}{6}\left[\left[\mathrm{E}\left\{(W_k - W_{k,0})^4\right\}\right]^{3/4} + \left[\mathrm{E}\left\{(W_{k+1} - W_{k,0})^4\right\}\right]^{3/4}\right].
\end{aligned}$$

Combining (S.2) and the above inequality leads to

$$\begin{aligned}
&|\mathrm{E}\,f(W(\eta_1,\ldots,\eta_n)) - \mathrm{E}\,f(W(\xi_1,\ldots,\xi_n))|\\
\le&\frac{\sup_{x\in\mathbb{R}}|f^{(3)}(x)|}{6}\sum_{k=1}^{n}\left[\left[\mathrm{E}\left\{(W_k - W_{k,0})^4\right\}\right]^{3/4} + \left[\mathrm{E}\left\{(W_{k+1} - W_{k,0})^4\right\}\right]^{3/4}\right].
\end{aligned}$$

Now we derive upper bounds for $\mathrm{E}\{(W_k - W_{k,0})^4\}$ and $\mathrm{E}\{(W_{k+1} - W_{k,0})^4\}$. We have

$$\begin{aligned}
\mathrm{E}\left\{(W_k - W_{k,0})^4\right\} =& \mathrm{E}\left[\Big\{\sum_{i=1}^{k-1} w_{i,k}(\eta_i,\xi_k)\Big\}^4\right] + \mathrm{E}\left[\Big\{\sum_{j=k+1}^{n} w_{k,j}(\xi_k,\xi_j)\Big\}^4\right] \\
&+ 6\,\mathrm{E}\left[\Big\{\sum_{i=1}^{k-1} w_{i,k}(\eta_i,\xi_k)\Big\}^2\Big\{\sum_{j=k+1}^{n} w_{k,j}(\xi_k,\xi_j)\Big\}^2\right] \\
=& \sum_{i=1}^{k-1}\mathrm{E}\left\{w_{i,k}(\eta_i,\xi_k)^4\right\} + 6\sum_{i_1=1}^{k-1}\sum_{i_2=i_1+1}^{k-1}\mathrm{E}\left\{w_{i_1,k}(\eta_{i_1},\xi_k)^2 w_{i_2,k}(\eta_{i_2},\xi_k)^2\right\} \\
&+ \sum_{j=k+1}^{n}\mathrm{E}\left\{w_{k,j}(\xi_k,\xi_j)^4\right\} + 6\sum_{j_1=k+1}^{n}\sum_{j_2=j_1+1}^{n}\mathrm{E}\left\{w_{k,j_1}(\xi_k,\xi_{j_1})^2 w_{k,j_2}(\xi_k,\xi_{j_2})^2\right\} \\
&+ 6\sum_{i=1}^{k-1}\sum_{j=k+1}^{n}\mathrm{E}\left\{w_{i,k}(\eta_i,\xi_k)^2 w_{k,j}(\xi_k,\xi_j)^2\right\}.
\end{aligned}$$

From the above equality and the definition of ρ_n, we have

$$\begin{aligned}
\mathrm{E}\left\{(W_k - W_{k,0})^4\right\} \le& \rho_n\Bigg(\sum_{i=1}^{k-1}\sigma_{i,k}^4 + 6\sum_{i_1=1}^{k-1}\sum_{i_2=i_1+1}^{k-1}\sigma_{i_1,k}^2\sigma_{i_2,k}^2 \\
&+ \sum_{j=k+1}^{n}\sigma_{k,j}^4 + 6\sum_{j_1=k+1}^{n}\sum_{j_2=j_1+1}^{n}\sigma_{k,j_1}^2\sigma_{k,j_2}^2 + 6\sum_{i=1}^{k-1}\sum_{j=k+1}^{n}\sigma_{i,k}^2\sigma_{k,j}^2\Bigg) \\
\le& 3\rho_n\left(\sum_{i=1}^{k-1}\sigma_{i,k}^2 + \sum_{j=k+1}^{n}\sigma_{k,j}^2\right)^2 \\
=& 3\rho_n \mathrm{Inf}_k^2.
\end{aligned}$$

Similarly, $\mathrm{E}\left\{(W_{k+1} - W_{k,0})^4\right\} \le 3\rho_n\mathrm{Inf}_k^2$. Thus,

$$|\mathrm{E}\, f(W(\eta_1,\dots,\eta_n)) - \mathrm{E}\, f(W(\xi_1,\dots,\xi_n))| \le \sup_{x\in\mathbb{R}}|f^{(3)}(x)|\frac{\rho_n^{3/4}}{3^{1/4}}\sum_{i=1}^{n}\mathrm{Inf}_i^{3/2}.$$

This completes the proof. $\square$

S.3 Technical results

We begin with some notations that will be used throughout our proofs of main results. For a matrix $\mathbf{A}$, let $\|\mathbf{A}\|_F$ denote the Frobenious norm of $\mathbf{A}$. For a matrix $\mathbf{A} \in \mathbb{R}^{p\times p}$, let $\lambda_i(\mathbf{A})$ denote the ith largest eigenvalue of $\mathbf{A}$. We adopt the convention that $\lambda_i(\mathbf{A}) = 0$ for $i > p$.

The following lemma collects some trace inequalities that will be used in our proofs of main results. To make our proofs of main results concise, the application of these trace inequalities are often tacit.

Lemma S.1. *The following trace inequalities hold:*

(i) For any matrices $\mathbf{A} \in \mathbb{R}^{p\times q}$, $\mathbf{B} \in \mathbb{R}^{q\times r}$, $\operatorname{tr}(\mathbf{A}\mathbf{B}) \le \{\operatorname{tr}(\mathbf{A}\mathbf{A}^\intercal)\operatorname{tr}(\mathbf{B}\mathbf{B}^\intercal)\}^{1/2}$.

(ii) *For any positive semi-definite matrices* $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{p\times p}$, $0 \le \operatorname{tr}(\mathbf{AB}) \le \lambda_1(\mathbf{A}) \operatorname{tr}(\mathbf{B})$.

(iii) *Suppose* f *is an increasing function from* $\mathbb{R}$ *to* $\mathbb{R}$, *and* $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{p\times p}$ *are symmetric. Then* $\operatorname{tr}(f(\mathbf{A})) \le \operatorname{tr}(f(\mathbf{B}))$.

(iv) *For any symmetric matrices* $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{p\times p}$, $\operatorname{tr}\{(\mathbf{AB})^2\} \le \operatorname{tr}(\mathbf{A}^2\mathbf{B}^2)$.

Remark 2. Lemma S.1, (i) and (ii) are standard. We refer to Tropp (2015), Proposition 8.3.3 for a proof of Lemma S.1, (iii). We refer to Tao (2012), inequality (3.18) for a proof of Lemma S.1, (iv).

Lemma S.2. *Suppose* $\mathbf{A}_1, \dots, \mathbf{A}_n$ *are* $p \times p$ *positive semi-definite matrices. Let* $\bar{\mathbf{A}} = n^{-1}\sum_{i=1}^n \mathbf{A}_i$. *Then*

$$\left|\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n\sum_{\ell=1}^n \operatorname{tr}(\mathbf{A}_i\mathbf{A}_j\mathbf{A}_k\mathbf{A}_\ell)\mathbf{1}_{\{i,j,k,\ell \text{ are distinct}\}}\right|$$
$$\le n^4 \operatorname{tr}\left\{(\bar{\mathbf{A}})^4\right\} + 10n^2 \operatorname{tr}\left\{(\bar{\mathbf{A}})^2\right\}\sum_{i=1}^n \operatorname{tr}\left(\mathbf{A}_i^2\right) + 13\left\{\sum_{i=1}^n \operatorname{tr}(\mathbf{A}_i^2)\right\}^2.$$

Proof. For distinct i, j, k, we have $\mathbf{1}_{\{\ell\notin\{i,j,k\}\}} = 1 - \mathbf{1}_{\{\ell=i\}} - \mathbf{1}_{\{\ell=j\}} - \mathbf{1}_{\{\ell=k\}}$. Hence

$$\begin{aligned}
&\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n\sum_{\ell=1}^n \operatorname{tr}(\mathbf{A}_i\mathbf{A}_j\mathbf{A}_k\mathbf{A}_\ell)\mathbf{1}_{\{i,j,k,\ell \text{ are distinct}\}}\\
=&n\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n \operatorname{tr}\left\{\mathbf{A}_i\mathbf{A}_j\mathbf{A}_k\bar{\mathbf{A}}\right\}\mathbf{1}_{\{i,j,k \text{ are distinct}\}} - 2\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n \operatorname{tr}(\mathbf{A}_i^2\mathbf{A}_j\mathbf{A}_k)\mathbf{1}_{\{i,j,k \text{ are distinct}\}}\\
&-\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n \operatorname{tr}(\mathbf{A}_i\mathbf{A}_j\mathbf{A}_k\mathbf{A}_j)\mathbf{1}_{\{i,j,k \text{ are distinct}\}}.
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
&\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n \operatorname{tr}\left\{\mathbf{A}_i\mathbf{A}_j\mathbf{A}_k\bar{\mathbf{A}}\right\}\mathbf{1}_{\{i,j,k \text{ are distinct}\}}\\
=&n\sum_{i=1}^n\sum_{j=1}^n \operatorname{tr}\left\{\mathbf{A}_i\mathbf{A}_j(\bar{\mathbf{A}})^2\right\}\mathbf{1}_{\{i\neq j\}} - \sum_{i=1}^n\sum_{j=1}^n \operatorname{tr}\left\{\mathbf{A}_i\mathbf{A}_j\mathbf{A}_i\bar{\mathbf{A}}\right\}\mathbf{1}_{\{i\neq j\}} - \sum_{i=1}^n\sum_{j=1}^n \operatorname{tr}\left\{\mathbf{A}_i\mathbf{A}_j^2\bar{\mathbf{A}}\right\}\mathbf{1}_{\{i\neq j\}}\\
=&n^3 \operatorname{tr}\left\{(\bar{\mathbf{A}})^4\right\} - 2n\sum_{i=1}^n \operatorname{tr}\left\{\mathbf{A}_i^2(\bar{\mathbf{A}})^2\right\} - n\sum_{i=1}^n \operatorname{tr}\left\{(\mathbf{A}_i\bar{\mathbf{A}})^2\right\} + 2\sum_{i=1}^n \operatorname{tr}\left(\mathbf{A}_i^3\bar{\mathbf{A}}\right).
\end{aligned}$$

And

$$\begin{aligned}
&\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n \operatorname{tr}(\mathbf{A}_i^2\mathbf{A}_j\mathbf{A}_k)\mathbf{1}_{\{i,j,k \text{ are distinct}\}}\\
=&n\sum_{i=1}^n\sum_{j=1}^n \operatorname{tr}(\mathbf{A}_i^2\mathbf{A}_j\bar{\mathbf{A}})\mathbf{1}_{\{i\neq j\}} - \sum_{i=1}^n\sum_{j=1}^n \operatorname{tr}(\mathbf{A}_i^3\mathbf{A}_j)\mathbf{1}_{\{i\neq j\}} - \sum_{i=1}^n\sum_{j=1}^n \operatorname{tr}(\mathbf{A}_i^2\mathbf{A}_j^2)\mathbf{1}_{\{i\neq j\}}\\
=&n^2\sum_{i=1}^n \operatorname{tr}\{\mathbf{A}_i^2(\bar{\mathbf{A}})^2\} - 2n\sum_{i=1}^n \operatorname{tr}(\mathbf{A}_i^3\bar{\mathbf{A}}) + 2\sum_{i=1}^n \operatorname{tr}(\mathbf{A}_i^4) - \sum_{i=1}^n\sum_{j=1}^n \operatorname{tr}(\mathbf{A}_i^2\mathbf{A}_j^2).
\end{aligned}$$

And

$$\begin{aligned}
&\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\operatorname{tr}(\mathbf{A}_i\mathbf{A}_j\mathbf{A}_k\mathbf{A}_j)\mathbf{1}_{\{i,j,k \text{ are distinct}\}}\\
=&n\sum_{i=1}^{n}\sum_{j=1}^{n}\operatorname{tr}(\mathbf{A}_i\mathbf{A}_j\bar{\mathbf{A}}\mathbf{A}_j)\mathbf{1}_{\{i\neq j\}}-\sum_{i=1}^{n}\sum_{j=1}^{n}\operatorname{tr}\{(\mathbf{A}_i\mathbf{A}_j)^2\}\mathbf{1}_{\{i\neq j\}}-\sum_{i=1}^{n}\sum_{j=1}^{n}\operatorname{tr}(\mathbf{A}_i\mathbf{A}_j^3)\mathbf{1}_{\{i\neq j\}}\\
=&n^2\sum_{i=1}^{n}\operatorname{tr}\{(\mathbf{A}_i\bar{\mathbf{A}})^2\}-2n\sum_{i=1}^{n}\operatorname{tr}(\mathbf{A}_i^3\bar{\mathbf{A}})-\sum_{i=1}^{n}\sum_{j=1}^{n}\operatorname{tr}\{(\mathbf{A}_i\mathbf{A}_j)^2\}+2\sum_{i=1}^{n}\operatorname{tr}(\mathbf{A}_i^4).
\end{aligned}$$

It follows that

$$\begin{aligned}
&\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{\ell=1}^{n}\operatorname{tr}(\mathbf{A}_i\mathbf{A}_j\mathbf{A}_k\mathbf{A}_\ell)\mathbf{1}_{\{i,j,k,\ell \text{ are distinct}\}}\\
=&n^4\operatorname{tr}\left\{(\bar{\mathbf{A}})^4\right\}-4n^2\sum_{i=1}^{n}\operatorname{tr}\left\{\mathbf{A}_i^2(\bar{\mathbf{A}})^2\right\}-2n^2\sum_{i=1}^{n}\operatorname{tr}\left\{(\mathbf{A}_i\bar{\mathbf{A}})^2\right\}+8n\sum_{i=1}^{n}\operatorname{tr}\left(\mathbf{A}_i^3\bar{\mathbf{A}}\right)\\
&-6\sum_{i=1}^{n}\operatorname{tr}(\mathbf{A}_i^4)+2\sum_{i=1}^{n}\sum_{j=1}^{n}\operatorname{tr}(\mathbf{A}_i^2\mathbf{A}_j^2)+\sum_{i=1}^{n}\sum_{j=1}^{n}\operatorname{tr}\{(\mathbf{A}_i\mathbf{A}_j)^2\}.
\end{aligned}$$

Thus, we have

$$\begin{aligned}
&\left|\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{\ell=1}^{n}\operatorname{tr}(\mathbf{A}_i\mathbf{A}_j\mathbf{A}_k\mathbf{A}_\ell)\mathbf{1}_{\{i,j,k,\ell \text{ are distinct}\}}\right|\\
\leq&n^4\operatorname{tr}\left\{(\bar{\mathbf{A}})^4\right\}+6n^2\sum_{i=1}^{n}\operatorname{tr}\left\{\mathbf{A}_i^2(\bar{\mathbf{A}})^2\right\}+8n\left\{\sum_{i=1}^{n}\operatorname{tr}(\mathbf{A}_i^4)\right\}^{1/2}\left\{\sum_{i=1}^{n}\operatorname{tr}\{\mathbf{A}_i^2(\bar{\mathbf{A}})^2\}\right\}^{1/2}+9\left\{\sum_{i=1}^{n}\operatorname{tr}(\mathbf{A}_i^2)\right\}^2\\
\leq&n^4\operatorname{tr}\left\{(\bar{\mathbf{A}})^4\right\}+10n^2\sum_{i=1}^{n}\operatorname{tr}\left\{\mathbf{A}_i^2(\bar{\mathbf{A}})^2\right\}+13\left\{\sum_{i=1}^{n}\operatorname{tr}(\mathbf{A}_i^2)\right\}^2\\
\leq&n^4\operatorname{tr}\left\{(\bar{\mathbf{A}})^4\right\}+10n^2\operatorname{tr}\left\{(\bar{\mathbf{A}})^2\right\}\sum_{i=1}^{n}\operatorname{tr}\left(\mathbf{A}_i^2\right)+13\left\{\sum_{i=1}^{n}\operatorname{tr}(\mathbf{A}_i^2)\right\}^2.
\end{aligned}$$

This completes the proof.

□

Lemma S.3. *Suppose $\mathbf{A}_1,\ldots,\mathbf{A}_{n_1}$ and $\mathbf{B}_1,\ldots,\mathbf{B}_{n_2}$ are $p\times p$ positive semi-definite matrices. Let $\bar{\mathbf{A}}=n_1^{-1}\sum_{i=1}^{n_1}\mathbf{A}_i$ and $\bar{\mathbf{B}}=n_2^{-1}\sum_{i=1}^{n_2}\mathbf{B}_i$. Then*

$$\begin{aligned}
&\left|\sum_{i=1}^{n_1}\sum_{j=1}^{n_1}\sum_{k=1}^{n_2}\sum_{\ell=1}^{n_2}\operatorname{tr}(\mathbf{A}_i\mathbf{B}_k\mathbf{A}_j\mathbf{B}_\ell)\mathbf{1}_{\{i\neq j\}}\mathbf{1}_{\{k\neq\ell\}}\right|\\
\leq&\frac{1}{2}n_2^4\operatorname{tr}\{(\bar{\mathbf{A}})^4\}+\frac{1}{2}n_1^4\operatorname{tr}\{(\bar{\mathbf{B}})^4\}+n_2^2\operatorname{tr}\{(\bar{\mathbf{B}})^2\}\sum_{i=1}^{n_1}\operatorname{tr}\{\mathbf{A}_i^2\}+n_1^2\operatorname{tr}\{(\bar{\mathbf{A}})^2\}\sum_{i=1}^{n_2}\operatorname{tr}\{(\mathbf{B}_i)^2\}+\left\{\sum_{i=1}^{n_1}\operatorname{tr}(\mathbf{A}_i^2)\right\}\left\{\sum_{i=1}^{n_2}\operatorname{tr}(\mathbf{B}_i^2)\right\}.
\end{aligned}$$

Proof. Note that

$$\mathbf{1}_{\{i\neq j\}}\mathbf{1}_{\{k\neq\ell\}}=(1-\mathbf{1}_{\{i=j\}})(1-\mathbf{1}_{\{k=\ell\}})=1-\mathbf{1}_{\{i=j\}}-\mathbf{1}_{\{k=\ell\}}+\mathbf{1}_{\{i=j,k=\ell\}}.$$

Hence we have

$$
\begin{aligned}
&\sum_{i=1}^{n_1}\sum_{j=1}^{n_1}\sum_{k=1}^{n_2}\sum_{\ell=1}^{n_2}\operatorname{tr}(\mathbf{A}_i\mathbf{B}_k\mathbf{A}_j\mathbf{B}_\ell)\mathbf{1}_{\{i\neq j\}}\mathbf{1}_{\{k\neq \ell\}}\\
=&n_1^2n_2^2\operatorname{tr}\{(\bar{\mathbf{A}}\bar{\mathbf{B}})^2\}-n_2^2\sum_{i=1}^{n_1}\operatorname{tr}\{(\mathbf{A}_i\bar{\mathbf{B}})^2\}-n_1^2\sum_{i=1}^{n_2}\operatorname{tr}\{(\bar{\mathbf{A}}\mathbf{B}_i)^2\}+\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\operatorname{tr}\{(\mathbf{A}_i\mathbf{B}_j)^2\}.
\end{aligned}
$$

It follows that

$$
\begin{aligned}
&\left|\sum_{i=1}^{n_1}\sum_{j=1}^{n_1}\sum_{k=1}^{n_2}\sum_{\ell=1}^{n_2}\operatorname{tr}(\mathbf{A}_i\mathbf{B}_k\mathbf{A}_j\mathbf{B}_\ell)\mathbf{1}_{\{i\neq j\}}\mathbf{1}_{\{k\neq \ell\}}\right|\\
\leq&\frac{1}{2}n_2^4\operatorname{tr}\{(\bar{\mathbf{A}})^4\}+\frac{1}{2}n_1^4\operatorname{tr}\{(\bar{\mathbf{B}})^4\}+n_2^2\operatorname{tr}\{(\bar{\mathbf{B}})^2\}\sum_{i=1}^{n_1}\operatorname{tr}\{\mathbf{A}_i^2\}+n_1^2\operatorname{tr}\{(\bar{\mathbf{A}})^2\}\sum_{i=1}^{n_2}\operatorname{tr}\{(\mathbf{B}_i)^2\}+\left\{\sum_{i=1}^{n_1}\operatorname{tr}(\mathbf{A}_i^2)\right\}\left\{\sum_{i=1}^{n_2}\operatorname{tr}(\mathbf{B}_i^2)\right\}.
\end{aligned}
$$

This completes the proof. □

Lemma S.4. *Suppose ξ and η are two random variables. Then*

$$
\|\mathcal{L}(\xi+\eta)-\mathcal{L}(\xi)\|_3\leq\left\{\operatorname{E}(\eta^2)\right\}^{1/2}.
$$

Proof. For $f\in\mathscr{C}_b^3(\mathbb{R})$, Taylor's theorem implies that

$$
|\operatorname{E}f(\xi+\eta)-\operatorname{E}f(\xi)|\leq\sup_{x\in\mathbb{R}}|f^{(1)}(x)|\operatorname{E}|\eta|\leq\sup_{x\in\mathbb{R}}|f^{(1)}(x)|\left\{\operatorname{E}(\eta^2)\right\}^{1/2}.
$$

Then the conclusion follows from the definition of the norm $\|\cdot\|_3$. □

Lemma S.5. *Under Assumptions 1 and 3, as $n\to\infty$,*

$$
\sigma_{T,n}^2=(1+o(1))2\operatorname{tr}(\mathbf{\Psi}_n^2).
$$

Proof. We have

$$
\begin{aligned}
\sigma_{T,n}^2=&\sum_{k=1}^{2}\frac{2}{(n_k-1)^2}\left\{\operatorname{tr}(\bar{\mathbf{\Sigma}}_k^2)-\frac{1}{n_k^2}\sum_{i=1}^{n_k}\operatorname{tr}(\mathbf{\Sigma}_{k,i}^2)\right\}+\frac{4}{n_1n_2}\operatorname{tr}(\bar{\mathbf{\Sigma}}_1\bar{\mathbf{\Sigma}}_2)\\
=&(1+o(1))\left\{\sum_{k=1}^{2}\frac{2}{(n_k-1)^2}\operatorname{tr}(\bar{\mathbf{\Sigma}}_k^2)+\frac{4}{n_1n_2}\operatorname{tr}(\bar{\mathbf{\Sigma}}_1\bar{\mathbf{\Sigma}}_2)\right\}\\
=&(1+o(1))2\operatorname{tr}\left\{(n_1^{-1}\bar{\mathbf{\Sigma}}_1+n_2^{-1}\bar{\mathbf{\Sigma}}_2)^2\right\},
\end{aligned}
$$

where the second equality follows from Assumption 3 and the third equality follows from Assumption 1. The conclusion follows. □

Lemma S.6. *Suppose $\boldsymbol{\zeta}_n$ is a d_n-dimensional standard normal random vector where $\{d_n\}$ is an arbitrary sequence of positive integers. Suppose $\mathbf{A}_n$ is a $d_n\times d_n$ symmetric matrix and $\mathbf{B}_n$ is an $r\times d_n$ matrix where r is a fixed positive integer. Furthermore, suppose*

$$
\limsup_{n\to\infty}\operatorname{tr}(\mathbf{A}_n^2)\in[0,+\infty),\quad\lim_{n\to\infty}\operatorname{tr}(\mathbf{A}_n^4)=0,\quad\limsup_{n\to\infty}\|\mathbf{B}_n\mathbf{B}_n^\intercal\|_F\in[0,+\infty).
$$

Let $\{c_n\}$ be a sequence of positive numbers such that $|c_n - \{2\operatorname{tr}(\mathbf{A}_n^2)\}^{1/2}| \to 0$. Let $\{\mathbf{D}_n\}$ be a sequence of $r \times r$ matrices such that $\|\mathbf{D}_n - \mathbf{B}_n\mathbf{B}_n^\intercal\|_F \to 0$. Then

$$\|\mathcal{L}\left(\boldsymbol{\zeta}_n^\intercal \mathbf{A}_n \boldsymbol{\zeta}_n - \operatorname{tr}(\mathbf{A}_n) + \boldsymbol{\zeta}_n^\intercal \mathbf{B}_n^\intercal \mathbf{B}_n \boldsymbol{\zeta}_n - \operatorname{tr}(\mathbf{B}_n^\intercal \mathbf{B}_n)\right) - \mathcal{L}\left(c_n \xi_0 + \boldsymbol{\xi}_r^\intercal \mathbf{D}_n \boldsymbol{\xi}_r - \operatorname{tr}(\mathbf{D}_n)\right)\|_3 \to 0,$$

where $\boldsymbol{\xi}_r$ is an r-dimensional standard normal random vector and ξ_0 is a standard normal random variable which is independent of $\boldsymbol{\xi}_r$.

Proof. The conclusion holds if and only if for any subsequence of $\{n\}$, there is a further subsequence along which the conclusion holds. Using this subsequence trick, we only need to prove that the conclusion holds for a subsequence of $\{n\}$. By taking a subsequence, we can assume without loss of generality that

$$\lim_{n\to\infty} \{2\operatorname{tr}(\mathbf{A}_n^2)\}^{1/2} = \gamma, \quad \lim_{n\to\infty} \operatorname{tr}(\mathbf{A}_n^4) = 0, \quad \lim_{n\to\infty} \mathbf{B}_n \mathbf{B}_n^\intercal = \boldsymbol{\Omega},$$

where $\gamma \geq 0$ and $\boldsymbol{\Omega}$ is an $r \times r$ positive semi-definite matrix. Then we have $\lim_{n\to\infty} c_n = \gamma$ and $\lim_{n\to\infty} \mathbf{D}_n = \boldsymbol{\Omega}$. Consequently, $\mathcal{L}\left(c_n\xi_0 + \boldsymbol{\xi}_r^\intercal \mathbf{D}_n \boldsymbol{\xi}_r - \operatorname{tr}(\mathbf{D}_n)\right)$ converges weakly to $\mathcal{L}\left(\gamma\xi_0 + \boldsymbol{\xi}_r^\intercal \boldsymbol{\Omega} \boldsymbol{\xi}_r - \operatorname{tr}(\boldsymbol{\Omega})\right)$. Hence we only need to prove that

$$\|\mathcal{L}\left(\boldsymbol{\zeta}_n^\intercal \mathbf{A}_n \boldsymbol{\zeta}_n - \operatorname{tr}(\mathbf{A}_n) + \boldsymbol{\zeta}_n^\intercal \mathbf{B}_n^\intercal \mathbf{B}_n \boldsymbol{\zeta}_n - \operatorname{tr}(\mathbf{B}_n^\intercal \mathbf{B}_n)\right) - \mathcal{L}\left(\gamma \xi_0 + \boldsymbol{\xi}_r^\intercal \boldsymbol{\Omega} \boldsymbol{\xi}_r - \operatorname{tr}(\boldsymbol{\Omega})\right)\|_3 \to 0.$$

To prove this, it suffices to prove that

$$\begin{pmatrix} \boldsymbol{\zeta}_n^\intercal \mathbf{A}_n \boldsymbol{\zeta}_n - \operatorname{tr}(\mathbf{A}_n) \\ \mathbf{B}_n \boldsymbol{\zeta}_n \end{pmatrix} \rightsquigarrow \mathcal{N}\left(\mathbf{0}_{r+1}, \begin{pmatrix} \gamma^2 & 0 \\ 0 & \boldsymbol{\Omega} \end{pmatrix}\right),$$

where "$\rightsquigarrow$" means weak convergence.

Denote by $\mathbf{A}_n = \mathbf{P}_n \boldsymbol{\Lambda}_n \mathbf{P}_n^\intercal$ the spectral decomposition of $\mathbf{A}_n$ where $\boldsymbol{\Lambda}_n = \operatorname{diag}\{\lambda_1(\mathbf{A}_n), \ldots, \lambda_{d_n}(\mathbf{A}_n)\}$ and $\mathbf{P}_n \in \mathbb{R}$ is an $d_n \times d_n$ orthogonal matrix. Define $\boldsymbol{\zeta}_n^* = \mathbf{P}_n^\intercal \boldsymbol{\zeta}_n$. Then $\boldsymbol{\zeta}_n^*$ is also a d_n-dimensional standard normal random vector and $\boldsymbol{\zeta}_n^\intercal \mathbf{A}_n \boldsymbol{\zeta}_n - \operatorname{tr}(\mathbf{A}_n) = \boldsymbol{\zeta}_n^{*\intercal} \boldsymbol{\Lambda}_n \boldsymbol{\zeta}_n^* - \operatorname{tr}(\boldsymbol{\Lambda}_n)$, $\mathbf{B}_n \boldsymbol{\zeta}_n = \mathbf{B}_n^* \boldsymbol{\zeta}_n^*$ where $\mathbf{B}_n^* = \mathbf{B}_n \mathbf{P}_n$.

Fix $a \in \mathbb{R}$ and $\mathbf{b} \in \mathbb{R}^r$. Let $\tilde{b}_{n,i}$ be the ith element of $\mathbf{B}_n^{*\intercal}\mathbf{b}$, $i = 1, \ldots, d_n$. The characteristic function of the random vector $(\boldsymbol{\zeta}_n^\intercal \mathbf{A}_n \boldsymbol{\zeta}_n - \operatorname{tr}(\mathbf{A}_n), (\mathbf{B}_n \boldsymbol{\zeta}_n)^\intercal)^\intercal$ at $(a, \mathbf{b}^\intercal)^\intercal$ is

$$\begin{aligned}
&\operatorname{E}\exp\left\{i\left(a(\boldsymbol{\zeta}_n^\intercal \mathbf{A}_n \boldsymbol{\zeta}_n - \operatorname{tr}(\mathbf{A}_n)) - \mathbf{b}^\intercal \mathbf{B}_n \boldsymbol{\zeta}_n\right)\right\} \\
=&\operatorname{E}\exp\left\{i\left(a(\boldsymbol{\zeta}_n^{*\intercal} \boldsymbol{\Lambda}_n \boldsymbol{\zeta}_n^* - \operatorname{tr}(\boldsymbol{\Lambda}_n)) - \mathbf{b}^\intercal \mathbf{B}_n^* \boldsymbol{\zeta}_n^*\right)\right\} \\
=&\exp\left\{-\frac{1}{2}\sum_{j=1}^{d_n} \log(1 - 2ia\lambda_j(\mathbf{A}_n)) - \frac{1}{2}\sum_{j=1}^{d_n} \frac{\tilde{b}_{n,j}^2}{1 - 2ia\lambda_j(\mathbf{A}_n)} - ia\operatorname{tr}(\mathbf{A}_n)\right\},
\end{aligned}$$

where the last equality can be obtained from the characteristic function of noncentral χ^2 random variable, and for log functions, we put the cut along $(-\infty, 0]$, that is, $\log(z) = \log(|z|) + i\arg(z)$ with $-\pi < \arg(z) < \pi$.

The condition $\lim_{n\to\infty} \operatorname{tr}(\mathbf{A}_n^4) = 0$ implies that $\lambda_1(\mathbf{A}_n) \to 0$. Consequently, Taylor's theorem implies that

$$\sum_{j=1}^{d_n} \log(1 - 2ia\lambda_j(\mathbf{A}_n)) = \sum_{j=1}^{d_n} \left\{-2ia\lambda_j(\mathbf{A}_n) - \frac{1}{2}\left(2ia\lambda_j(\mathbf{A}_n)\right)^2(1 + e_{n,j})\right\},$$

where $\lim_{n\to\infty} \max_{j\in\{1,\ldots,d_n\}} |e_{n,j}| \to 0$. It follows that

$$\sum_{j=1}^{d_n} \log(1 - 2ia\lambda_j(\mathbf{A}_n)) = -2ia\operatorname{tr}(\mathbf{A}_n) + (1 + o(1))2a^2\operatorname{tr}(\mathbf{A}_n^2).$$

On the other hand,

$$\sum_{j=1}^{d_n} \tilde{b}_{n,j}^2/(1-2ia\lambda_j(\mathbf{A}_n)) = (1+o(1))\sum_{j=1}^{n} \tilde{b}_{n,j}^2 = (1+o(1))\mathbf{b}^\intercal \mathbf{B}_n \mathbf{B}_n^\intercal \mathbf{b}.$$

Thus,

$$\begin{aligned}\mathrm{E}\left[\exp\left\{i\left(a(\boldsymbol{\zeta}_n^\intercal \mathbf{A}_n \boldsymbol{\zeta}_n - \operatorname{tr}(\mathbf{A}_n)) - \mathbf{b}^\intercal \mathbf{B}_n \boldsymbol{\zeta}_n\right)\right\}\right] &= \exp\left\{-(1+o(1))\frac{1}{2}\left(2a^2 \operatorname{tr}(\mathbf{A}_n^2) + \mathbf{b}^\intercal \mathbf{B}_n \mathbf{B}_n^\intercal \mathbf{b}\right)\right\} \\ &\to \exp\left\{-\frac{1}{2}\left(\gamma^2 a^2 + \mathbf{b}^\intercal \boldsymbol{\Omega} \mathbf{b}\right)\right\},\end{aligned}$$

where $\exp\left\{-\frac{1}{2}\left(\gamma^2 a^2 + \mathbf{b}^\intercal \boldsymbol{\Omega}\mathbf{b}\right)\right\}$ is the characteristic function of the distribution

$$\mathcal{N}\left(\mathbf{0}_{r+1}, \begin{pmatrix} \gamma^2 & 0 \\ 0 & \boldsymbol{\Omega} \end{pmatrix}\right).$$

This completes the proof. □

Lemma S.7. *Suppose $\boldsymbol{\Psi}_n$ is a $p \times p$ symmetric matrix where p is a function of n.*

$$\lim_{n\to\infty} \frac{\lambda_i(\boldsymbol{\Psi}_n)}{\{\operatorname{tr}(\boldsymbol{\Psi}_n^2)\}^{1/2}} = \kappa_i, \quad i = 1, 2, \ldots.$$

Let $\boldsymbol{\xi}_p$ be a p-dimensional standard normal random vector and $\{\xi_i\}_{i=0}^\infty$ a sequence of independent standard normal random variables. Then as $n \to \infty$,

$$\left\| \mathcal{L}\left(\frac{\boldsymbol{\xi}_p^\intercal \boldsymbol{\Psi}_n \boldsymbol{\xi}_p - \operatorname{tr}(\boldsymbol{\Psi}_n)}{\{2\operatorname{tr}(\boldsymbol{\Psi}_n^2)\}^{1/2}}\right) - \mathcal{L}\left((1-\sum_{i=1}^\infty \kappa_i^2)^{1/2}\xi_0 + 2^{-1/2}\sum_{i=1}^\infty \kappa_i(\xi_i^2 - 1)\right)\right\|_3 \to 0.$$

Proof. Let $\kappa_{i,n} = \lambda_i(\boldsymbol{\Psi}_n)/\{\operatorname{tr}(\boldsymbol{\Psi}_n^2)\}^{1/2}$. Then we have

$$\mathcal{L}\left(\frac{\boldsymbol{\xi}_p^\intercal \boldsymbol{\Psi}_n \boldsymbol{\xi}_p - \operatorname{tr}(\boldsymbol{\Psi}_n)}{\{2\operatorname{tr}(\boldsymbol{\Psi}_n^2)\}^{1/2}}\right) = \mathcal{L}\left(2^{-1/2}\sum_{i=1}^p \kappa_{i,n}(\zeta_i^2 - 1)\right),$$

where $\{\zeta_i\}_{i=1}^\infty$ is a sequence of independent standard normal random variables which are independent of $\{\xi_i\}_{i=0}^\infty$. From Fatou's lemma, we have $\sum_{i=1}^\infty \kappa_i^2 \le \lim_{n\to\infty}\sum_{i=1}^\infty \kappa_{i,n}^2 = 1$. It follows from Lévy's equivalence theorem and three-series theorem (see, e.g., Dudley (2002), Theorem 9.7.1 and Theorem 9.7.3) that $\sum_{i=1}^r \kappa_i(\zeta_i^2 - 1)$ converges weakly to $\sum_{i=1}^\infty \kappa_i(\zeta_i^2 - 1)$ as $r \to \infty$. Hence for any $\delta > 0$, there is a positive integer r such that

$$\left\| \mathcal{L}\left((1-\sum_{i=1}^\infty \kappa_i^2)^{1/2}\xi_0 + 2^{-1/2}\sum_{i=1}^\infty \kappa_i(\zeta_i^2 - 1)\right) - \mathcal{L}\left((1-\sum_{i=1}^r \kappa_i^2)^{1/2}\xi_0 + 2^{-1/2}\sum_{i=1}^r \kappa_i(\zeta_i^2 - 1)\right)\right\|_3 \le \delta. \quad \text{(S.6)}$$

By taking a possibly larger r, we can also assume that $\kappa_{r+1} \le \delta$. Now we fix such an r and apply Lindeberg principle. For $j = r+1, \ldots, p$, define

$$V_{j,0} = 2^{-1/2}\sum_{i=1}^{j-1} \kappa_{i,n}(\zeta_i^2 - 1) + \sum_{i=j+1}^{p} \kappa_{i,n}\xi_i.$$

Then we have $V_{j+1,0} + \kappa_{j+1,n}\xi_{j+1} = V_{j,0} + 2^{-1/2}\kappa_{j,n}(\zeta_j^2 - 1)$, $j = r+1, \ldots, p-1$, and

$$V_{r+1,0} + \kappa_{r+1,n}\xi_{r+1} = 2^{-1/2}\sum_{i=1}^{r}\kappa_{i,n}(\zeta_i^2 - 1) + \sum_{i=r+1}^{p}\kappa_{i,n}\xi_i,$$
$$V_{p,0} + 2^{-1/2}\kappa_{p,n}(\zeta_p^2 - 1) = 2^{-1/2}\sum_{i=1}^{p}\kappa_{i,n}(\zeta_i^2 - 1).$$

For $f \in \mathscr{C}_b^3(\mathbb{R})$, we have

$$\begin{aligned}
&\left\| f\left(2^{-1/2}\sum_{i=1}^{p}\kappa_{i,n}(\zeta_i^2 - 1)\right) - f\left(2^{-1/2}\sum_{i=1}^{r}\kappa_{i,n}(\zeta_i^2 - 1) + \sum_{i=r+1}^{p}\kappa_{i,n}\xi_i\right)\right\|_3 \\
=& \left|\mathrm{E}\, f(V_{p,0} + 2^{-1/2}\kappa_{p,n}(\zeta_p^2 - 1)) - \mathrm{E}\, f(V_{r+1,0} + \kappa_{r+1,n}\xi_{r+1})\right| \\
=& \left|\sum_{j=r+1}^{p}(\mathrm{E}\, f(V_{j,0} + 2^{-1/2}\kappa_{j,n}(\zeta_j^2 - 1)) - \mathrm{E}\, f(V_{j,0} + \kappa_{j,n}\xi_j))\right| \\
\leq& \sum_{j=r+1}^{p}\left|\mathrm{E}\, f(V_{j,0} + 2^{-1/2}\kappa_{j,n}(\zeta_j^2 - 1)) - \mathrm{E}\, f(V_{j,0} + \kappa_{j,n}\xi_j)\right|. \qquad \text{(S.7)}
\end{aligned}$$

For $j = r+1, \ldots, p$, we have

$$\left| f\left(V_{j,0} + 2^{-1/2}\kappa_{j,n}(\zeta_j^2 - 1)\right) - f(V_{j,0}) - \sum_{i=1}^{2}\frac{1}{i!}\kappa_{j,n}^i\left\{2^{-1/2}(\zeta_j^2 - 1)\right\}^i f^{(i)}(V_{j,0})\right| \leq \frac{1}{6}\kappa_{j,n}^3\left|2^{-1/2}(\zeta_j^2 - 1)\right|^3\sup_{x\in\mathbb{R}}|f^{(3)}(x)|,$$
$$\left| f(V_{j,0} + \kappa_{j,n}\xi_j) - f(V_{j,0}) - \sum_{i=1}^{2}\frac{1}{i!}\kappa_{j,n}^i\xi_j^i f^{(i)}(V_{j,0})\right| \leq \frac{1}{6}\kappa_{j,n}^3|\xi_j|^3\sup_{x\in\mathbb{R}}|f^{(3)}(x)|.$$

But

$$\mathrm{E}(\xi_j) = \mathrm{E}\left\{2^{-1/2}(\zeta_j^2 - 1)\right\} = 0, \quad \mathrm{E}(\xi_j^2) = \mathrm{E}\left[\left\{2^{-1/2}(\zeta_j^2 - 1)\right\}^2\right] = 1.$$

Thus, for $j = r+1, \ldots, p$, we have

$$\left|\mathrm{E}\, f(V_{j,0} + 2^{-1/2}\kappa_{j,n}(\zeta_j^2 - 1)) - \mathrm{E}\, f(V_{j,0} + \kappa_{j,n}\xi_j)\right| \leq \frac{1}{6}\kappa_{j,n}^3\,\mathrm{E}\left\{|2^{-1/2}(\zeta_1^2 - 1)|^3 + |\xi_1|^3\right\}\sup_{x\in\mathbb{R}}|f^{(3)}(x)|. \quad \text{(S.8)}$$

From (S.7) and (S.8), we have

$$\begin{aligned}
&\left\| f\left(2^{-1/2}\sum_{i=1}^{p}\kappa_{i,n}(\zeta_i^2 - 1)\right) - f\left(2^{-1/2}\sum_{i=1}^{r}\kappa_{i,n}(\zeta_i^2 - 1) + \sum_{i=r+1}^{p}\kappa_{i,n}\xi_i\right)\right\|_3 \\
\leq& \frac{1}{6}\sum_{j=r+1}^{p}\kappa_{j,n}^3\,\mathrm{E}\left\{|2^{-1/2}(\zeta_1^2 - 1)|^3 + |\xi_1|^3\right\}\sup_{x\in\mathbb{R}}|f^{(3)}(x)| \\
\leq& \frac{\kappa_{r+1,n}}{6}\,\mathrm{E}\left\{|2^{-1/2}(\zeta_1^2 - 1)|^3 + |\xi_1|^3\right\}\sup_{x\in\mathbb{R}}|f^{(3)}(x)|,
\end{aligned}$$

where the last inequality holds since $\sum_{j=r+1}^{p}\kappa_{j,n}^3 \leq \kappa_{r+1,n}\sum_{j=r+1}^{p}\kappa_{j,n}^2$ and $\sum_{j=1}^{p}\kappa_{j,n}^2 = 1$. Note that

$\lim_{n\to\infty} \kappa_{r+1,n} = \kappa_{r+1} \le \delta$. Hence we have

$$\limsup_{n\to\infty} \left\| \mathcal{L}\left(2^{-1/2}\sum_{i=1}^{p} \kappa_{i,n}(\zeta_i^2 - 1)\right) - \mathcal{L}\left(2^{-1/2}\sum_{i=1}^{r} \kappa_{i,n}(\zeta_i^2 - 1) + \sum_{i=r+1}^{p} \kappa_{i,n}\xi_i\right)\right\|_3$$
$$\le \frac{\delta}{6} \operatorname{E}\left\{|2^{-1/2}(\zeta_1^2 - 1)|^3 + |\xi_1|^3\right\}. \tag{S.9}$$

Note that $\sum_{i=r+1}^{p} \kappa_{i,n}\xi_i$ has the same distribution as $(1 - \sum_{i=1}^{r} \kappa_{i,n}^2)^{1/2}\xi_0$. Then from Lemma S.4,

$$\left\| \mathcal{L}\left(2^{-1/2}\sum_{i=1}^{r} \kappa_{i,n}(\zeta_i^2 - 1) + \sum_{i=r+1}^{p} \kappa_{i,n}\xi_i\right) - \mathcal{L}\left((1 - \sum_{i=1}^{r} \kappa_i^2)^{1/2}\xi_0 + 2^{-1/2}\sum_{i=1}^{r} \kappa_i(\zeta_i^2 - 1)\right)\right\|_3$$
$$\le \left|(1 - \sum_{i=1}^{r} \kappa_{i,n}^2)^{1/2} - (1 - \sum_{i=1}^{r} \kappa_i^2)^{1/2}\right| \left\{\operatorname{E}(\xi_0^2)\right\}^{1/2} + \left(\sum_{i=1}^{r} |\kappa_{i,n} - \kappa_i|\right) 2^{-1/2}\left[\operatorname{E}\left\{(\zeta_1^2 - 1)^2\right\}\right]^{1/2}, \tag{S.10}$$

where the right hand side tends to 0 as $n \to \infty$. From (S.6), (S.9) and (S.10), we have

$$\limsup_{n\to\infty} \left\| \mathcal{L}\left(2^{-1/2}\sum_{i=1}^{p} \kappa_{i,n}(\zeta_i^2 - 1)\right) - \mathcal{L}\left((1 - \sum_{i=1}^{\infty} \kappa_i^2)^{1/2}\xi_0 + 2^{-1/2}\sum_{i=1}^{\infty} \kappa_i(\zeta_i^2 - 1)\right)\right\|_3$$
$$\le \left[\frac{1}{6} \operatorname{E}\left\{|\xi_1|^3 + |2^{-1/2}(\zeta_1^2 - 1)|^3\right\} + 1\right] \delta.$$

But δ is an arbitrary positive real number. Hence the above limit must be 0. This completes the proof. □

S.4 Proof of Lemma 1

In this section, we provide the proof of Lemma 1.

Fix an arbitrary $p \times p$ positive semi-definite matrix $\mathbf{B}$. Let $\mathbf{G} = \mathbf{\Gamma}_{k,i}^\intercal \mathbf{B} \mathbf{\Gamma}_{k,i}$. We only need to show that

$$\operatorname{E}\{(Z_{k,i}^\intercal \mathbf{G} Z_{k,i})^2\} \le 3C\{\operatorname{tr}(\mathbf{G})\}^2.$$

Let $g_{i,j}$ denote the (i,j)th element of $\mathbf{G}$. Then

$$\begin{aligned}(Z_{k,i}^\intercal \mathbf{G} Z_{k,i})^2 =& \Big(2\sum_{j_1=1}^{m_{k,i}} \sum_{j_2=j_1+1}^{m_{k,i}} g_{j_1,j_2} z_{k,i,j_1} z_{k,i,j_2} + \sum_{j_3=1}^{m_{k,i}} g_{j_3,j_3} z_{k,i,j_3}^2\Big)^2\\ =& 4\Big(\sum_{j_1=1}^{m_{k,i}} \sum_{j_2=j_1+1}^{m_{k,i}} g_{j_1,j_2} z_{k,i,j_1} z_{k,i,j_2}\Big)^2 + \Big(\sum_{j_3=1}^{m_{k,i}} g_{j_3,j_3} z_{k,i,j_3}^2\Big)^2\\ &+ 4\Big(\sum_{j_1=1}^{m_{k,i}} \sum_{j_2=j_1+1}^{m_{k,i}} g_{j_1,j_2} z_{k,i,j_1} z_{k,i,j_2}\Big)\Big(\sum_{j_3=1}^{m_{k,i}} g_{j_3,j_3} z_{k,i,j_3}^2\Big).\end{aligned}$$

From the condition (5), the expectation of the cross term is zero, and

$$\begin{aligned}\operatorname{E}\Big\{\Big(\sum_{j_1=1}^{m_{k,i}} \sum_{j_2=j_1+1}^{m_{k,i}} g_{j_1,j_2} z_{k,i,j_1} z_{k,i,j_2}\Big)^2\Big\} =& \sum_{j_1=1}^{m_{k,i}} \sum_{j_2=j_1+1}^{m_{k,i}} g_{j_1,j_2}^2 \operatorname{E}(z_{k,i,j_1}^2 z_{k,i,j_2}^2)\\ \le& (C/2)\operatorname{tr}(\mathbf{G}^2)\\ \le& (C/2)\{\operatorname{tr}(\mathbf{G})\}^2,\end{aligned}$$

where the last inequality holds since $\mathbf{G}$ is positive semi-definite. On the other hand

$$\mathrm{E}\Big\{\Big(\sum_{j_3=1}^{m_{k,i}} g_{j_3,j_3} z_{k,i,j_3}^2\Big)^2\Big\} \le C\Big(\sum_{j_3=1}^{m_{k,i}} g_{j_3,j_3}\Big)^2 = C\{\mathrm{tr}(\mathbf{G})\}^2.$$

Hence the conclusion holds.

S.5 Proof of Theorem 1

In this section, we provide the proof of Theorem 1.

Let $Y_{k,i}^*$, $i = 1, \ldots, n_k$, $k = 1, 2$, be independent p-dimensional random vectors with $Y_{k,i}^* \sim \mathcal{N}(\mathbf{0}_p, \mathbf{\Sigma}_{k,i})$. Let $\mathbf{Y}_k^* = (Y_{k,1}, \ldots, Y_{k,n_k})^\intercal$, $k = 1, 2$. First we apply Theorem S.1 to prove that

$$\left\| \mathcal{L}\left(\frac{T_{\mathrm{CQ}}(\mathbf{Y}_1, \mathbf{Y}_2)}{\sigma_{T,n}}\right) - \mathcal{L}\left(\frac{T_{\mathrm{CQ}}(\mathbf{Y}_1^*, \mathbf{Y}_2^*)}{\sigma_{T,n}}\right) \right\|_3 \to 0. \tag{S.11}$$

Let

$$\xi_i = \begin{cases} Y_{1,i} & \text{for } i = 1, \ldots, n_1, \\ Y_{2,i-n_1} & \text{for } i = n_1 + 1, \ldots, n, \end{cases} \quad \text{and} \quad \eta_i = \begin{cases} Y_{1,i}^* & \text{for } i = 1, \ldots, n_1, \\ Y_{2,i-n_1}^* & \text{for } i = n_1 + 1, \ldots, n. \end{cases}$$

Define

$$w_{i,j}(\mathbf{a}, \mathbf{b}) = \begin{cases} \frac{2\mathbf{a}^\intercal \mathbf{b}}{n_1(n_1-1)\sigma_{T,n}} & \text{for } 1 \le i < j \le n_1, \\ \frac{-2\mathbf{a}^\intercal \mathbf{b}}{n_1 n_2 \sigma_{T,n}} & \text{for } 1 \le i \le n_1 \text{ and } n_1 + 1 \le j \le n, \\ \frac{2\mathbf{a}^\intercal \mathbf{b}}{n_2(n_2-1)\sigma_{T,n}} & \text{for } n_1 + 1 \le i < j \le n. \end{cases}$$

With the above definitions, we have $W(\xi_1, \ldots, \xi_n) = T_{\mathrm{CQ}}(\mathbf{Y}_1, \mathbf{Y}_2)/\sigma_{T,n}$ and $W(\eta_1, \ldots, \eta_n) = T_{\mathrm{CQ}}(\mathbf{Y}_1^*, \mathbf{Y}_2^*)/\sigma_{T,n}$, where the function $W(\cdot, \ldots, \cdot)$ is defined in Section S.2.

It follows from Assumption 2 and the fact that $Y_{k,i}^*$ has the same first two moments as $Y_{k,i}$ that Assumptions S.1 and S.2 hold. By direct calculation, we have

$$\sigma_{i,j}^2 = \begin{cases} \frac{4\,\mathrm{tr}(\mathbf{\Sigma}_{1,i}\mathbf{\Sigma}_{1,j})}{n_1^2(n_1-1)^2\sigma_{T,n}^2} & \text{for } 1 \le i < j \le n_1, \\ \frac{4\,\mathrm{tr}(\mathbf{\Sigma}_{1,i}\mathbf{\Sigma}_{2,j-n_1})}{n_1^2 n_2^2 \sigma_{T,n}^2} & \text{for } 1 \le i \le n_1 \text{ and } n_1 + 1 \le j \le n, \\ \frac{4\,\mathrm{tr}(\mathbf{\Sigma}_{2,i-n_1}\mathbf{\Sigma}_{2,j-n_1})}{n_2^2(n_2-1)^2\sigma_{T,n}^2} & \text{for } n_1 + 1 \le i < j \le n. \end{cases}$$

Consequently,

$$\mathrm{Inf}_i = \begin{cases} \frac{4\,\mathrm{tr}(\bar{\mathbf{\Sigma}}_1\mathbf{\Sigma}_{1,i})}{n_1(n_1-1)^2\sigma_{T,n}^2} - \frac{4\,\mathrm{tr}(\mathbf{\Sigma}_{1,i}^2)}{n_1^2(n_1-1)^2\sigma_{T,n}^2} + \frac{4\,\mathrm{tr}(\bar{\mathbf{\Sigma}}_2\mathbf{\Sigma}_{1,i})}{n_1^2 n_2 \sigma_{T,n}^2} & \text{for } 1 \le i \le n_1, \\ \frac{4\,\mathrm{tr}(\bar{\mathbf{\Sigma}}_2\mathbf{\Sigma}_{2,i-n_1})}{n_2(n_2-1)^2\sigma_{T,n}^2} - \frac{4\,\mathrm{tr}(\mathbf{\Sigma}_{2,i-n_1}^2)}{n_2^2(n_2-1)^2\sigma_{T,n}^2} + \frac{4\,\mathrm{tr}(\bar{\mathbf{\Sigma}}_1\mathbf{\Sigma}_{2,i-n_1})}{n_1 n_2^2 \sigma_{T,n}^2} & \text{for } n_1 + 1 \le i \le n. \end{cases}$$

We have

$$\sum_{i=1}^n \mathrm{Inf}_i^{3/2} \le \left(\max_{i \in \{1,\ldots,n\}} \mathrm{Inf}_i\right)^{1/2} \left(\sum_{i=1}^n \mathrm{Inf}_i\right) \le \left(\sum_{i=1}^n \mathrm{Inf}_i^2\right)^{1/4} \left(\sum_{i=1}^n \mathrm{Inf}_i\right).$$

It can be seen that $\sum_{i=1}^{n} \operatorname{Inf}_i = 2$. On the other hand, from Assumption 3,

$$
\begin{aligned}
\sum_{i=1}^{n} \operatorname{Inf}_i^2 \leq & \sum_{i=1}^{n_1} \left\{ \frac{32 \operatorname{tr}(\bar{\mathbf{\Sigma}}_1^2) \operatorname{tr}(\mathbf{\Sigma}_{1,i}^2)}{n_1^2 (n_1-1)^4 \sigma_{T,n}^4} + \frac{32 \operatorname{tr}(\bar{\mathbf{\Sigma}}_2^2) \operatorname{tr}(\mathbf{\Sigma}_{1,i}^2)}{n_1^4 n_2^2 \sigma_{T,n}^4} \right\} \\
& + \sum_{i=1}^{n_2} \left\{ \frac{32 \operatorname{tr}(\bar{\mathbf{\Sigma}}_2^2) \operatorname{tr}(\mathbf{\Sigma}_{2,i}^2)}{n_2^2 (n_2-1)^4 \sigma_{T,n}^4} + \frac{32 \operatorname{tr}(\bar{\mathbf{\Sigma}}_1^2) \operatorname{tr}(\mathbf{\Sigma}_{2,i}^2)}{n_1^2 n_2^4 \sigma_{T,n}^4} \right\} \\
= & o\left(\frac{\left\{\operatorname{tr}(\bar{\mathbf{\Sigma}}_1^2)\right\}^2}{(n_1-1)^4 \sigma_{T,n}^4} + \frac{\left\{\operatorname{tr}(\bar{\mathbf{\Sigma}}_2^2)\right\}^2}{(n_2-1)^4 \sigma_{T,n}^4} + \frac{\operatorname{tr}(\bar{\mathbf{\Sigma}}_1^2) \operatorname{tr}(\bar{\mathbf{\Sigma}}_2^2)}{n_1^2 n_2^2 \sigma_{T,n}^4} \right) \\
= & o\left(\frac{\left\{\operatorname{tr}(\bar{\mathbf{\Sigma}}_1^2)\right\}^2}{(n_1-1)^4 \sigma_{T,n}^4} + \frac{\left\{\operatorname{tr}(\bar{\mathbf{\Sigma}}_2^2)\right\}^2}{(n_2-1)^4 \sigma_{T,n}^4} \right) \\
= & o(1).
\end{aligned}
$$

It follows that $\sum_{i=1}^{n} \operatorname{Inf}_i^{3/2} \to 0$.

On the other hand, from Assumption 2, for all $1 \leq i < j \leq n$,

$$
\max\left(\mathrm{E}\{w_{i,j}(\xi_i, \xi_j)^4\}, \mathrm{E}\{w_{i,j}(\xi_i, \eta_j)^4\}, \mathrm{E}\{w_{i,j}(\eta_i, \xi_j)^4\}, \mathrm{E}\{w_{i,j}(\eta_i, \eta_j)^4\} \right) \leq \tau^2 \sigma_{i,j}^4.
$$

Hence ρ_n is upper bounded by the absolute constant τ^2. Thus, (S.11) holds.

Now we deal with the distribution of $T_{\mathrm{CQ}}(\mathbf{Y}_1^*, \mathbf{Y}_2^*)/\sigma_{T,n}$. We have

$$
\begin{aligned}
T_{\mathrm{CQ}}(\mathbf{Y}_1^*, \mathbf{Y}_2^*) = & \|\bar{Y}_1^* - \bar{Y}_2^*\|^2 - \operatorname{tr}(\mathbf{\Psi}_n) + \sum_{k=1}^{2} \frac{1}{n_k - 1} \left\{ \|\bar{Y}_k^*\|^2 - \frac{1}{n_k} \operatorname{tr}(\bar{\mathbf{\Sigma}}_k) \right\} \\
& - \sum_{k=1}^{2} \frac{1}{n_k(n_k-1)} \sum_{i=1}^{n_k} \left\{ \|Y_{k,i}^*\|^2 - \operatorname{tr}(\mathbf{\Sigma}_{k,i}) \right\},
\end{aligned}
$$

where $\bar{Y}_k^* = n_k^{-1} \sum_{i=1}^{n_k} Y_{k,i}$, $k = 1, 2$. From Lemma S.4,

$$
\begin{aligned}
& \left\| \mathcal{L}\left(\frac{T_{\mathrm{CQ}}(\mathbf{Y}_1^*, , \mathbf{Y}_2^*)}{\sigma_{T,n}} \right) - \mathcal{L}\left(\frac{\|\bar{Y}_1^* - \bar{Y}_2^*\|^2 - \operatorname{tr}(\mathbf{\Psi}_n)}{\sigma_{T,n}} \right) \right\|_3 \\
\leq & \frac{1}{\sigma_{T,n}} \sum_{k=1}^{2} \frac{1}{n_k - 1} \left[\mathrm{E}\left\{ \left(\|\bar{Y}_k^*\|^2 - \frac{1}{n_k} \operatorname{tr}(\bar{\mathbf{\Sigma}}_k) \right)^2 \right\} \right]^{1/2} \\
& + \frac{1}{\sigma_{T,n}} \sum_{k=1}^{2} \frac{1}{n_k(n_k-1)} \left[\mathrm{E}\left\{ \left(\sum_{i=1}^{n_k} \left(\|Y_{k,i}^*\|^2 - \operatorname{tr}(\mathbf{\Sigma}_{k,i}) \right) \right)^2 \right\} \right]^{1/2} \\
= & \frac{1}{\sigma_{T,n}} \sum_{k=1}^{2} \frac{1}{n_k(n_k-1)} \left\{ 2 \operatorname{tr}(\bar{\mathbf{\Sigma}}_k^2) \right\}^{1/2} + \frac{1}{\sigma_{T,n}} \sum_{k=1}^{2} \frac{1}{n_k(n_k-1)} \left\{ 2 \sum_{i=1}^{n_k} \operatorname{tr}(\mathbf{\Sigma}_{k,i}^2) \right\}^{1/2} \\
= & o(1),
\end{aligned} \tag{S.12}
$$

where the last equality follows from Assumption 3.

Note that $\bar{Y}_1^* - \bar{Y}_2^* \sim \mathcal{N}(\mathbf{0}_p, \mathbf{\Psi}_n)$. Then from Lemma S.4,

$$\begin{aligned}
&\left\| \mathcal{L}\left(\frac{\|\bar{Y}_1^* - \bar{Y}_2^*\|^2 - \operatorname{tr}(\mathbf{\Psi}_n)}{\sigma_{T,n}} \right) - \mathcal{L}\left(\frac{\|\bar{Y}_1^* - \bar{Y}_2^*\|^2 - \operatorname{tr}(\mathbf{\Psi}_n)}{\{2\operatorname{tr}(\mathbf{\Psi}_n^2)\}^{1/2}} \right) \right\|_3 \\
\le& \left| \frac{\left\{2\operatorname{tr}\left(\mathbf{\Psi}_n^2\right)\right\}^{1/2}}{\sigma_{T,n}} - 1 \right| \left[\mathrm{E}\left[\frac{\|\bar{Y}_1^* - \bar{Y}_2^*\|^2 - \operatorname{tr}(\mathbf{\Psi}_n)}{\{2\operatorname{tr}\left(\mathbf{\Psi}_n^2\right)\}^{1/2}} \right]^2 \right]^{1/2} \\
=& \left| \frac{\left\{2\operatorname{tr}\left(\mathbf{\Psi}_n^2\right)\right\}^{1/2}}{\sigma_{T,n}} - 1 \right| \\
\to& 0,
\end{aligned} \tag{S.13}$$

where the last equality follows from Lemma S.5. Then the conclusion follows from (S.11), (S.12) and (S.13).

S.6 Proof of Corollary 1

In this section, we provide the proof of Corollary 1.

Since $T_{\mathrm{CQ}}(\mathbf{Y}_1, \mathbf{Y}_2)/\sigma_{T,n}$ has zero mean and unit variance, the distribution of $T_{\mathrm{CQ}}(\mathbf{Y}_1, \mathbf{Y}_2)/\sigma_{T,n}$ is uniformly tight. From Theorem 1, $T_{\mathrm{CQ}}(\mathbf{Y}_1, \mathbf{Y}_2)/\sigma_{T,n}$ and $(\boldsymbol{\xi}_p^\intercal \mathbf{\Psi}_n \boldsymbol{\xi}_p - \operatorname{tr}(\mathbf{\Psi}_n))/\{2\operatorname{tr}(\mathbf{\Psi}_n^2)\}^{1/2}$ share the same possible asymptotic distributions. Hence we only need to find all possible asymptotic distributions of $(\boldsymbol{\xi}_p^\intercal \mathbf{\Psi}_n \boldsymbol{\xi}_p - \operatorname{tr}(\mathbf{\Psi}_n))/\{2\operatorname{tr}(\mathbf{\Psi}_n^2)\}^{1/2}$.

Let ν be a possible asymptotic distribution of $(\boldsymbol{\xi}_p^\intercal \mathbf{\Psi}_n \boldsymbol{\xi}_p - \operatorname{tr}(\mathbf{\Psi}_n))/\{2\operatorname{tr}(\mathbf{\Psi}_n^2)\}^{1/2}$. We show that ν can be represented in the form of (6). Note that there is a subsequence of $\{n\}$ along which $(\boldsymbol{\xi}_p^\intercal \mathbf{\Psi}_n \boldsymbol{\xi}_p - \operatorname{tr}(\mathbf{\Psi}_n))/\{2\operatorname{tr}(\mathbf{\Psi}_n^2)\}^{1/2}$ converges weakly to ν. Denote $\kappa_{i,n} = \lambda_i(\mathbf{\Psi}_n)/\{\operatorname{tr}(\mathbf{\Psi}_n^2)\}^{1/2}$, $i = 1, 2, \ldots$. By Cantor's diagonalization trick (see, e.g., Simon (2015), Section 1.5), there exists a further subsequence along which $\lim_{n\to\infty} \kappa_{i,n} = \kappa_i$, $i = 1, 2, \ldots$, where $\kappa_1, \kappa_2, \ldots$ are real numbers in $[0, 1]$. Then Lemma S.7 implies that along this further subsequence,

$$\left\| \mathcal{L}\left(\frac{\boldsymbol{\xi}_p^\intercal \mathbf{\Psi}_n \boldsymbol{\xi}_p - \operatorname{tr}(\mathbf{\Psi}_n)}{\{2\operatorname{tr}(\mathbf{\Psi}_n^2)\}^{1/2}} \right) - \mathcal{L}\left((1 - \sum_{i=1}^\infty \kappa_i^2)^{1/2} \xi_0 + 2^{-1/2} \sum_{i=1}^\infty \kappa_i (\zeta_i^2 - 1) \right) \right\|_3 \to 0.$$

Thus, $\nu = \mathcal{L}\left((1 - \sum_{i=1}^\infty \kappa_i^2)^{1/2} \xi_0 + 2^{-1/2} \sum_{i=1}^\infty \kappa_i (\zeta_i^2 - 1)\right)$.

Now we prove that for any sequence of positve numbers $\{\kappa_i\}_{i=1}^\infty$ such that $\sum_{i=1}^\infty \kappa_i^2 \in [0, 1]$, there exits a sequence $\{\mathbf{\Psi}_n\}_{n=1}^\infty$ such that $\mathcal{L}\left((1 - \sum_{i=1}^\infty \kappa_i^2)^{1/2} \xi_0 + 2^{-1/2} \sum_{i=1}^\infty \kappa_i (\zeta_i^2 - 1)\right)$ is the asymptotic distribution of $(\boldsymbol{\xi}_p^\intercal \mathbf{\Psi}_n \boldsymbol{\xi}_p - \operatorname{tr}(\mathbf{\Psi}_n))/\{2\operatorname{tr}(\mathbf{\Psi}_n^2)\}^{1/2}$. To construct the sequence $\{\mathbf{\Psi}_n\}_{n=1}^\infty$, we take $p = n^2$ and let $\mathbf{\Psi}_n = \operatorname{diag}(\kappa_{1,n}, \ldots, \kappa_{n^2,n})$, where $\kappa_{i,n} = \kappa_i$ for $i \in \{1, \ldots, n\}$ and $\kappa_{i,n} = \{(1 - \sum_{i=1}^n \kappa_i^2)/(n^2 - n)\}^{1/2}$ for $i \in \{n+1, \ldots, n^2\}$. Then $\sum_{i=1}^p \kappa_{i,n}^2 = 1$ and $\lim_{n\to\infty} \kappa_{i,n} = \kappa_i$, $i = 1, 2, \ldots$. It is straightforward to show that $(\boldsymbol{\xi}_p^\intercal \mathbf{\Psi}_n \boldsymbol{\xi}_p - \operatorname{tr}(\mathbf{\Psi}_n))/\{2\operatorname{tr}(\mathbf{\Psi}_n^2)\}^{1/2}$ converges weakly to $\mathcal{L}\left((1 - \sum_{i=1}^\infty \kappa_i^2)^{1/2} \xi_0 + 2^{-1/2} \sum_{i=1}^\infty \kappa_i (\zeta_i^2 - 1)\right)$. This completes the proof.

S.7 Proof of Theorem 2

In this section, we provide the proof of Theorem 2.

If the conclusion holds for the case that $\epsilon_{1,1}^*$ is a standard normal random variable, then Lemma S.11 implies that it also holds for the case of Rademacher random variable. Hence without loss of generality, we assume that $\epsilon_{1,1}^*$ is a standard normal random variable.

We begin with some basic notations and facts that are useful for our proof. Denote by $\boldsymbol{\Psi}_n = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\intercal$ the spectral decomposition of $\boldsymbol{\Psi}_n$ where $\boldsymbol{\Lambda} = \operatorname{diag}\{\lambda_1(\boldsymbol{\Psi}_n), \ldots, \lambda_p(\boldsymbol{\Psi}_n)\}$ and $\mathbf{U}$ is an orthogonal matrix. Let $\mathbf{U}_m \in \mathbb{R}^{p\times m}$ be the first m columns of $\mathbf{U}$ and $\tilde{\mathbf{U}}_m \in \mathbb{R}^{p\times(p-m)}$ be the last $p-r$ columns of $\mathbf{U}$. Then we have $\mathbf{U}_m^\intercal \tilde{\mathbf{U}}_m = \mathbf{O}_{m\times(p-m)}$ and $\mathbf{U}_m\mathbf{U}_m^\intercal + \tilde{\mathbf{U}}_m\tilde{\mathbf{U}}_m^\intercal = \mathbf{I}_p$. For $k = 1, 2$, we have

$$\lambda_1(\tilde{\mathbf{U}}_m^\intercal \bar{\boldsymbol{\Sigma}}_k \tilde{\mathbf{U}}_m) \leq n_k \lambda_1(\tilde{\mathbf{U}}_m^\intercal \boldsymbol{\Psi}_n \tilde{\mathbf{U}}_m) = n_k \lambda_{m+1}(\boldsymbol{\Psi}_n). \tag{S.14}$$

To prove the conclusion, we apply the subsequence trick. That is, for any subsequence of $\{n\}$, we prove that there is a further subsequence along which the conclusion holds. For any subsequence of $\{n\}$, by Cantor's diagonalization trick (see, e.g., Simon (2015), Section 1.5), there exists a further subsequence along which

$$\frac{\lambda_i(\boldsymbol{\Psi}_n)}{\{\operatorname{tr}(\boldsymbol{\Psi}_n^2)\}^{1/2}} \to \kappa_i, \quad i = 1, 2, \ldots. \tag{S.15}$$

Thus, we only need to prove the conclusion with the additional condition (S.15). Without loss of generality, we assume that (S.15) holds for the full sequence $\{n\}$.

From Fatou's lemma, we have $\sum_{i=1}^\infty \kappa_i^2 \leq 1$. Now we claim that there exists a sequence $\{r_n^*\}$ of non-decreasing integers which tends to infinity such that

$$\frac{\sum_{i=1}^{r_n^*} \lambda_i^2(\boldsymbol{\Psi}_n)}{\operatorname{tr}(\boldsymbol{\Psi}_n^2)} \to \sum_{i=1}^\infty \kappa_i^2. \tag{S.16}$$

We defer the proof of this fact to Lemma S.8 in Section S.9.

Fix a positive integer r. Then for large n, we have $r_n^* > r$. Let $\check{\mathbf{U}}_r$ be the $(r+1)$th to the r_n^*th columns of $\mathbf{U}$. Then $\check{\mathbf{U}}_r \in \mathbb{R}^{p\times(r_n^*-r)}$ is a column orthogonal matrix such that $\check{\mathbf{U}}_r\check{\mathbf{U}}_r^\intercal = \mathbf{U}_{r_n^*}\mathbf{U}_{r_n^*}^\intercal - \mathbf{U}_r\mathbf{U}_r^\intercal$. We can decompose the identity matrix into the sum of three mutually orthogonal projection matrices as $\mathbf{I}_p = \mathbf{U}_r\mathbf{U}_r^\intercal + \check{\mathbf{U}}_r\check{\mathbf{U}}_r^\intercal + \tilde{\mathbf{U}}_{r_n^*}\tilde{\mathbf{U}}_{r_n^*}^\intercal$. Then we have $T_{\mathrm{CQ}}(E^*; \tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2) = T_{1,r} + T_{2,r} + T_{3,r}$, where

$$T_{1,r} = T_{\mathrm{CQ}}(E^*; \tilde{\mathbf{X}}_1\mathbf{U}_r, \tilde{\mathbf{X}}_2\mathbf{U}_r), \quad T_{2,r} = T_{\mathrm{CQ}}(E^*; \tilde{\mathbf{X}}_1\check{\mathbf{U}}_r, \tilde{\mathbf{X}}_2\check{\mathbf{U}}_r), \quad T_{3,r} = T_{\mathrm{CQ}}(E^*; \tilde{\mathbf{X}}_1\tilde{\mathbf{U}}_{r_n^*}, \tilde{\mathbf{X}}_2\tilde{\mathbf{U}}_{r_n^*}).$$

Let $\{\xi_i\}_{i=0}^\infty$ be a sequence of independent standard normal random variables. We claim that

$$\left\| \mathcal{L}\left(\frac{T_{1,r} + T_{3,r}}{\sigma_{T,n}} \mid \tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2 \right) - \mathcal{L}\left((1 - \sum_{i=1}^\infty \kappa_i^2)^{1/2}\xi_0 + 2^{-1/2}\sum_{i=1}^r \kappa_i(\xi_i^2 - 1) \right) \right\|_3 \xrightarrow{P} 0. \tag{S.17}$$

The above quantity is a continuous function of $\tilde{\mathbf{X}}_1$ and $\tilde{\mathbf{X}}_2$ and is thus measurable. We defer the proof of (S.17) to Lemma S.12 in Section S.9. It is known that for random variables in $\mathbb{R}$, convergence in probability is metrizable; see, e.g., Dudley (2002), Section 9.2. As a standard property of metric space, (S.17) also holds when r is replaced by certain r_n with $r_n \to \infty$ and $r_n \leq r_n^*$. Also note that Lévy's equivalence theorem and three-series theorem (see, e.g., Dudley (2002), Theorem 9.7.1 and Theorem 9.7.3) implies that $\sum_{i=1}^{r_n} \kappa_i(\xi_i^2 - 1)$ converges weakly to $\sum_{i=1}^\infty \kappa_i(\xi_i^2 - 1)$. Thus,

$$\left\| \mathcal{L}\left(\frac{T_{1,r_n} + T_{3,r_n}}{\sigma_{T,n}} \mid \tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2 \right) - \mathcal{L}\left((1 - \sum_{i=1}^\infty \kappa_i^2)^{1/2}\xi_0 + 2^{-1/2}\sum_{i=1}^\infty \kappa_i(\xi_i^2 - 1) \right) \right\|_3 \xrightarrow{P} 0. \tag{S.18}$$

Now we deal with T_{2,r_n}. From Lemma S.10, we have

$$\mathrm{E}\left(\frac{T_{2,r_n}^2}{\sigma_{T,n}^2} \mid \tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2 \right) = \frac{2\operatorname{tr}\left\{ \left(\check{\mathbf{U}}_{r_n}^\intercal \boldsymbol{\Psi}_n \check{\mathbf{U}}_{r_n} \right)^2 \right\}}{\sigma_{T,n}^2} + o_P(1) = \sum_{i=1}^{r_n^*} \frac{\lambda_i^2(\boldsymbol{\Psi}_n)}{\operatorname{tr}(\boldsymbol{\Psi}_n^2)} - \sum_{i=1}^{r_n} \frac{\lambda_i^2(\boldsymbol{\Psi}_n)}{\operatorname{tr}(\boldsymbol{\Psi}_n^2)} + o_P(1).$$

From Fatou's lemma,

$$\liminf_{n\to\infty} \sum_{i=1}^{r_n} \frac{\lambda_i^2(\mathbf{\Psi}_n)}{\operatorname{tr}(\mathbf{\Psi}_n^2)} \geq \sum_{i=1}^{\infty} \liminf_{n\to\infty} \frac{\lambda_i^2(\mathbf{\Psi}_n)}{\operatorname{tr}(\mathbf{\Psi}_n^2)} = \sum_{i=1}^{\infty} \kappa_i^2.$$

Combining the above inequality and (S.16) leads to

$$\mathrm{E}\left(\frac{T_{2,r_n}^2}{\sigma_{T,n}^2} \mid \tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2 \right) = o_P(1).$$

Then from Lemma S.4,

$$\left\| \mathcal{L}\left(\frac{T_{\mathrm{CQ}}(E^*; \tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2)}{\sigma_{T,n}} \mid \tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2 \right) - \mathcal{L}\left(\frac{T_{1,r_n} + T_{3,r_n}}{\sigma_{T,n}} \right) \right\|_3 \xrightarrow{P} 0.$$

Combining the above equality and (S.18) leads to

$$\left\| \mathcal{L}\left(\frac{T_{\mathrm{CQ}}(E^*; \tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2)}{\sigma_{T,n}} \mid \tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2 \right) - \mathcal{L}\left((1 - \sum_{i=1}^{\infty} \kappa_i^2)^{1/2} \xi_0 + 2^{-1/2} \sum_{i=1}^{\infty} \kappa_i (\xi_i^2 - 1) \right) \right\|_3 \xrightarrow{P} 0. \tag{S.19}$$

Then the conclusion follows from (S.19) and Lemma S.7.

S.8 Proof of Corollary 2

Using the subsequence trick, it suffices to prove the conclusion for a subsequence of $\{n\}$. Then from Corollary 1, we can assume without loss of generality that

$$\left\| \mathcal{L}\left(\frac{\boldsymbol{\xi}_p^\intercal \mathbf{\Psi}_n \boldsymbol{\xi}_p - \operatorname{tr}(\mathbf{\Psi}_n)}{\{2 \operatorname{tr}(\mathbf{\Psi}_n^2)\}^{1/2}} \right) - \mathcal{L}\left((1 - \sum_{i=1}^{\infty} \kappa_i^2)^{1/2} \xi_0 + 2^{-1/2} \sum_{i=1}^{\infty} \kappa_i (\xi_i^2 - 1) \right) \right\|_3 \to 0, \tag{S.20}$$

where $\{\xi_i\}_{i=0}^{\infty}$ is a sequence of independent standard normal random variables, $\{\kappa_i\}_{i=1}^{\infty}$ is a sequence of positive numbers such that $\sum_{i=1}^{\infty} \kappa_i^2 \in [0,1]$. Let $\tilde{F}(\cdot)$ denote the cumulative distribution function of $(1 - \sum_{i=1}^{\infty} \kappa_i^2)^{1/2}\xi_0 + 2^{-1/2} \sum_{i=1}^{\infty} \kappa_i(\xi_i^2 - 1)$. We claim that $\tilde{F}(\cdot)$ is continuous and strictly increasing on the interval $\{x \in \mathbb{R} : \tilde{F}(x) > 0\}$. We defer the proof of this fact to Lemma S.13 in Section S.9. This fact, combined with (S.20), leads to

$$\sup_{x\in\mathbb{R}} |G_n(x) - \tilde{F}(x)| = o(1), \quad G_n^{-1}(1-\alpha) = \tilde{F}^{-1}(1-\alpha) + o(1). \tag{S.21}$$

Furthermore, in view of Theorem 2 and (S.20), we have

$$\frac{\hat{F}_{\mathrm{CQ}}^{-1}(1-\alpha)}{\sigma_{T,n}} = \tilde{F}^{-1}(1-\alpha) + o_P(1). \tag{S.22}$$

We have

$$\begin{aligned} &\operatorname{pr}\left\{ T_{\mathrm{CQ}}(\mathbf{X}_1, \mathbf{X}_2) > \hat{F}_{\mathrm{CQ}}^{-1}(1-\alpha) \right\} \\ =&\operatorname{pr}\left\{ \frac{T_{\mathrm{CQ}}(\mathbf{X}_1, \mathbf{X}_2) - \|\mu_1 - \mu_2\|^2}{\sigma_{T,n}} + \tilde{F}^{-1}(1-\alpha) - \frac{\hat{F}_{\mathrm{CQ}}^{-1}(1-\alpha)}{\sigma_{T,n}} > \tilde{F}^{-1}(1-\alpha) - \frac{\|\mu_1 - \mu_2\|^2}{\sigma_{T,n}} \right\} \end{aligned}$$

Note that

$$\frac{T_{\mathrm{CQ}}(\mathbf{X}_1,\mathbf{X}_2)-\|\mu_1-\mu_2\|^2}{\sigma_{T,n}}=\frac{T_{\mathrm{CQ}}(\mathbf{Y}_1,\mathbf{Y}_2)}{\sigma_{T,n}}+\frac{2(\mu_1-\mu_2)^\intercal(\bar{Y}_1-\bar{Y}_2)}{\sigma_{T,n}}=\frac{T_{\mathrm{CQ}}(\mathbf{Y}_1,\mathbf{Y}_2)}{\sigma_{T,n}}+o_P(1),$$

where the last equality holds since

$$\mathrm{var}\left(\frac{2(\mu_1-\mu_2)^\intercal(\bar{Y}_1-\bar{Y}_2)}{\sigma_{T,n}}\right)=(1+o(1))\frac{2(\mu_1-\mu_2)^\intercal\mathbf{\Psi}_n(\mu_1-\mu_2)}{\mathrm{tr}(\mathbf{\Psi}_n^2)}=o(1).$$

Then it follows from Theorem 1, equality (S.22) and the fact that $\tilde{F}(\cdot)$ is continuous that

$$\sup_{x\in\mathbb{R}}\left|\mathrm{pr}\left(\frac{T_{\mathrm{CQ}}(\mathbf{X}_1,\mathbf{X}_2)-\|\mu_1-\mu_2\|^2}{\sigma_{T,n}}+\tilde{F}^{-1}(1-\alpha)-\frac{\hat{F}_{\mathrm{CQ}}^{-1}(1-\alpha)}{\sigma_{T,n}}\le x\right)-\tilde{F}(x)\right|=o(1).$$

Therefore,

$$\begin{aligned}\mathrm{pr}\left\{T_{\mathrm{CQ}}(\mathbf{X}_1,\mathbf{X}_2)>\hat{F}_{\mathrm{CQ}}^{-1}(1-\alpha)\right\}=&1-\tilde{F}\left(\tilde{F}^{-1}(1-\alpha)-\frac{\|\mu_1-\mu_2\|^2}{\sigma_{T,n}}\right)+o(1)\\=&1-G_n\left(G_n^{-1}(1-\alpha)-\frac{\|\mu_1-\mu_2\|^2}{\{2\,\mathrm{tr}(\mathbf{\Psi}_n^2)\}^{1/2}}\right)+o(1),\end{aligned}$$

where the last equality follows from (S.21). This completes the proof.

S.9 Deferred proofs

In this section, we provide proofs of some intermediate results in our proofs of main results. Some results in this section are also used in the main text.

Lemma S.8. *Suppose the conditions of Theorem 2 hold. Furthermore, suppose the condition* (S.15) *holds. Then there exists a sequence* $\{r_n^*\}$ *of non-decreasing integers which tends to infinity such that* (S.16) *holds.*

Proof. For any fixed positive integer m, we have

$$\frac{\sum_{i=1}^m\lambda_i^2(\mathbf{\Psi}_n)}{\mathrm{tr}(\mathbf{\Psi}_n^2)}\to\sum_{i=1}^m\kappa_i^2.$$

Therefore, there exists an n_m such that for any $n>n_m$,

$$\left|\frac{\sum_{i=1}^m\lambda_i^2(\mathbf{\Psi}_n)}{\mathrm{tr}(\mathbf{\Psi}_n^2)}-\sum_{i=1}^m\kappa_i^2\right|<\frac{1}{m}.$$

We can without loss of generality and assume $n_1<n_2<\cdots$ since otherwise we can enlarge some n_m. Define $r_n^*=m$ for $n_m<n\le n_{m+1}$, $m=1,2,\ldots$ and $r_n^*=1$ for $n\le n_1$. By definition, r_n^* is non-decreasing and $\lim_{n\to\infty}r_n^*=\infty$. Also, for any $n>n_1$,

$$\left|\frac{\sum_{i=1}^{r_n^*}\lambda_i^2(\mathbf{\Psi}_n)}{\mathrm{tr}(\mathbf{\Psi}_n^2)}-\sum_{i=1}^{r_n^*}\kappa_i^2\right|<\frac{1}{r_n^*}.$$

Thus,

$$\lim_{n\to\infty}\left|\frac{\sum_{i=1}^{r_n^*}\lambda_i^2(\mathbf{\Psi}_n)}{\mathrm{tr}(\mathbf{\Psi}_n^2)}-\sum_{i=1}^{r_n^*}\kappa_i^2\right|=0.$$

The conclusion follows from the above limit and the fact that $\sum_{i=1}^{r_n^*}\kappa_i^2\to\sum_{i=1}^{\infty}\kappa_i^2$. □

Lemma S.9. *Suppose Assumption 2 holds. Then for any $p \times p$ positive semi-definite matrix $\mathbf{B}$ and $k = 1, 2$,*

$$\mathrm{E}\{(\tilde{X}_{k,i}^{\intercal}\mathbf{B}\tilde{X}_{k,i})^2\} \le \tau\{\mathrm{E}(\tilde{X}_{k,i}^{\intercal}\mathbf{B}\tilde{X}_{k,i})\}^2.$$

Proof. We have

$$\begin{aligned}\mathrm{E}\{(\tilde{X}_{k,i}^{\intercal}\mathbf{B}\tilde{X}_{k,i})^2\} =&\frac{1}{16}\,\mathrm{E}\{(Y_{k,2i}^{\intercal}\mathbf{B}Y_{k,2i})^2\} + \frac{1}{16}\,\mathrm{E}\{(Y_{k,2i-1}^{\intercal}\mathbf{B}Y_{k,2i-1})^2\}\\ &+ \frac{1}{8}\,\mathrm{E}\{(Y_{k,2i}^{\intercal}\mathbf{B}Y_{k,2i})(Y_{k,2i-1}^{\intercal}\mathbf{B}Y_{k,2i-1})\} + \frac{1}{4}\,\mathrm{E}\{(Y_{k,2i}^{\intercal}\mathbf{B}Y_{k,2i-1})^2\}\\ \le&\frac{\tau}{16}\{\mathrm{tr}(\mathbf{B}\boldsymbol{\Sigma}_{k,2i})\}^2 + \frac{\tau}{16}\{\mathrm{tr}(\mathbf{B}\boldsymbol{\Sigma}_{k,2i-1})\}^2\\ &+ \frac{1}{8}\,\mathrm{tr}(\mathbf{B}\boldsymbol{\Sigma}_{k,2i})\,\mathrm{tr}(\mathbf{B}\boldsymbol{\Sigma}_{k,2i-1}) + \frac{1}{4}\,\mathrm{tr}(\mathbf{B}\boldsymbol{\Sigma}_{k,2i}\mathbf{B}\boldsymbol{\Sigma}_{k,2i-1}),\end{aligned}$$

where the last inequality follows from Assumption 2. Note that

$$\begin{aligned}\mathrm{tr}(\mathbf{B}\boldsymbol{\Sigma}_{k,2i}\mathbf{B}\boldsymbol{\Sigma}_{k,2i-1}) &= \mathrm{tr}\{(\mathbf{B}^{1/2}\boldsymbol{\Sigma}_{k,2i}\mathbf{B}^{1/2})(\mathbf{B}^{1/2}\boldsymbol{\Sigma}_{k,2i-1}\mathbf{B}^{1/2})\}\\ &\le \mathrm{tr}(\mathbf{B}^{1/2}\boldsymbol{\Sigma}_{k,2i}\mathbf{B}^{1/2})\,\mathrm{tr}(\mathbf{B}^{1/2}\boldsymbol{\Sigma}_{k,2i-1}\mathbf{B}^{1/2})\\ &= \mathrm{tr}(\mathbf{B}\boldsymbol{\Sigma}_{k,2i})\,\mathrm{tr}(\mathbf{B}\boldsymbol{\Sigma}_{k,2i-1}).\end{aligned}$$

It follows from the above two inequalities and the condition $\tau \ge 3$ that

$$\mathrm{E}\{(\tilde{X}_{k,i}^{\intercal}\mathbf{B}\tilde{X}_{k,i})^2\} \le \frac{\tau}{16}\{\mathrm{tr}(\mathbf{B}\boldsymbol{\Sigma}_{k,2i}) + \mathrm{tr}(\mathbf{B}\boldsymbol{\Sigma}_{k,2i-1})\}^2 = \tau\{\mathrm{E}(\tilde{X}_{k,i}^{\intercal}\mathbf{B}\tilde{X}_{k,i})\}^2.$$

This completes the proof. □

Lemma S.10. *Suppose Assumptions 1, 2 and 3 hold, and $\sigma_{T,n}^2 > 0$ for all n. Let $\{\mathbf{B}_n\}$ be a sequence of matrices where $\mathbf{B}_n \in \mathbb{R}^{p\times m_n}$ is column orthogonal and the column number $m_n \le p$. Let $E^* = (\epsilon_{1,1}^*, \ldots, \epsilon_{1,m_1}^*, \epsilon_{2,1}^*, \ldots, \epsilon_{2,m_2}^*)^{\intercal}$, where $\epsilon_{k,i}^*$, $i = 1, \ldots, m_k$, $k = 1, 2$, are independent random variables with $\mathrm{E}(\epsilon_{k,i}^*) = 0$ and $\mathrm{var}(\epsilon_{k,i}^*) = 1$. Then as $n \to \infty$,*

$$\frac{\mathrm{var}\{T_{\mathrm{CQ}}(E^*; \tilde{\mathbf{X}}_1\mathbf{B}_n, \tilde{\mathbf{X}}_2\mathbf{B}_n) \mid \tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2\}}{\sigma_{T,n}^2} = \frac{2\,\mathrm{tr}\{(\mathbf{B}_n^{\intercal}\boldsymbol{\Psi}_n\mathbf{B}_n)^2\}}{\sigma_{T,n}^2} + o_P(1).$$

Proof. We have

$$\begin{aligned}&\mathrm{var}\{T_{\mathrm{CQ}}(E^*; \tilde{\mathbf{X}}_1\mathbf{B}_n, \tilde{\mathbf{X}}_2\mathbf{B}_n) \mid \tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2\}\\ =&\sum_{k=1}^{2} \frac{4\sum_{i=1}^{m_k}\sum_{j=i+1}^{m_k}(\tilde{X}_{k,i}^{\intercal}\mathbf{B}_n\mathbf{B}_n^{\intercal}\tilde{X}_{k,j})^2}{m_k^2(m_k-1)^2} + \frac{4\sum_{i=1}^{m_1}\sum_{j=1}^{m_2}(\tilde{X}_{1,i}^{\intercal}\mathbf{B}_n\mathbf{B}_n^{\intercal}\tilde{X}_{2,j})^2}{m_1^2m_2^2}.\end{aligned}$$

Fisrt we deal with $\sum_{i=1}^{m_k}\sum_{j=i+1}^{m_k}(\tilde{X}_{k,i}^{\intercal}\mathbf{B}_n\mathbf{B}_n^{\intercal}\tilde{X}_{k,j})^2$, $k = 1, 2$. We have

$$\begin{aligned}&\mathrm{E}\left\{\sum_{i=1}^{m_k}\sum_{j=i+1}^{m_k}(\tilde{X}_{k,i}^{\intercal}\mathbf{B}_n\mathbf{B}_n^{\intercal}\tilde{X}_{k,j})^2\right\}\\ =&\frac{1}{16}\sum_{i=1}^{m_k}\sum_{j=i+1}^{m_k}\mathrm{tr}(\mathbf{B}_n^{\intercal}(\boldsymbol{\Sigma}_{k,2j-1} + \boldsymbol{\Sigma}_{k,2j})\mathbf{B}_n\mathbf{B}_n^{\intercal}(\boldsymbol{\Sigma}_{k,2i-1} + \boldsymbol{\Sigma}_{k,2i})\mathbf{B}_n)\\ =&\frac{1}{32}\left(\sum_{i=1}^{2m_k}\sum_{j=1}^{2m_k}\mathrm{tr}(\mathbf{B}_n^{\intercal}\boldsymbol{\Sigma}_{k,j}\mathbf{B}_n\mathbf{B}_n^{\intercal}\boldsymbol{\Sigma}_{k,i}\mathbf{B}_n) - \sum_{i=1}^{m_k}\mathrm{tr}\{(\mathbf{B}_n^{\intercal}(\boldsymbol{\Sigma}_{k,2i-1} + \boldsymbol{\Sigma}_{k,2i})\mathbf{B}_n)^2\}\right).\end{aligned}$$

Hence

$$
\begin{aligned}
&\left|\mathrm{E}\left\{\sum_{i=1}^{m_k}\sum_{j=i+1}^{m_k}(\tilde{X}_{k,i}^{\intercal}\mathbf{B}_n\mathbf{B}_n^{\intercal}\tilde{X}_{k,j})^2\right\}-\frac{1}{32}n_k^2\operatorname{tr}\{(\mathbf{B}_n^{\intercal}\bar{\mathbf{\Sigma}}_k\mathbf{B}_n)^2\}\right|\\
\leq&\frac{1}{16}n_k\operatorname{tr}(\mathbf{B}_n^{\intercal}\mathbf{\Sigma}_{k,n_k}\mathbf{B}_n\mathbf{B}_n^{\intercal}\bar{\mathbf{\Sigma}}_k\mathbf{B}_n)+\frac{1}{16}\sum_{i=1}^{n_k}\operatorname{tr}\{(\mathbf{B}_n^{\intercal}\mathbf{\Sigma}_{k,i}\mathbf{B}_n)^2\}\\
\leq&\frac{1}{16}\left\{n_k^2\operatorname{tr}(\bar{\mathbf{\Sigma}}_k^2)\operatorname{tr}(\mathbf{\Sigma}_{k,n_k}^2)\right\}^{1/2}+\frac{1}{16}\sum_{i=1}^{n_k}\operatorname{tr}(\mathbf{\Sigma}_{k,i}^2)\\
=&o\left(n_k^2\operatorname{tr}(\bar{\mathbf{\Sigma}}_k^2)\right),
\end{aligned}
\tag{S.23}
$$

where the last equality follows from Assumption 3.

Now we compute the variance of $\sum_{i=1}^{m_k}\sum_{j=i+1}^{m_k}(\tilde{X}_{k,i}^{\intercal}\mathbf{B}_n\mathbf{B}_n^{\intercal}\tilde{X}_{k,j})^2$. Note that

$$
\begin{aligned}
&\left[\mathrm{E}\left\{\sum_{i=1}^{m_k}\sum_{j=i+1}^{m_k}(\tilde{X}_{k,i}^{\intercal}\mathbf{B}_n\mathbf{B}_n^{\intercal}\tilde{X}_{k,j})^2\right\}\right]^2\\
=&\sum_{i=1}^{m_k}\sum_{j=i+1}^{m_k}[\mathrm{E}\{(\tilde{X}_{k,i}^{\intercal}\mathbf{B}_n\mathbf{B}_n^{\intercal}\tilde{X}_{k,j})^2\}]^2+2\sum_{i=1}^{m_k}\sum_{j=i+1}^{m_k}\sum_{\ell=j+1}^{m_k}\mathrm{E}\{(\tilde{X}_{k,i}^{\intercal}\mathbf{B}_n\mathbf{B}_n^{\intercal}\tilde{X}_{k,j})^2\}\,\mathrm{E}\{(\tilde{X}_{k,i}^{\intercal}\mathbf{B}_n\mathbf{B}_n^{\intercal}\tilde{X}_{k,\ell})^2\}\\
&+2\sum_{i=1}^{m_k}\sum_{j=i+1}^{m_k}\sum_{\ell=j+1}^{m_k}\mathrm{E}\{(\tilde{X}_{k,i}^{\intercal}\mathbf{B}_n\mathbf{B}_n^{\intercal}\tilde{X}_{k,j})^2\}\,\mathrm{E}\{(\tilde{X}_{k,j}^{\intercal}\mathbf{B}_n\mathbf{B}_n^{\intercal}\tilde{X}_{k,\ell})^2\}\\
&+2\sum_{i=1}^{m_k}\sum_{j=i+1}^{m_k}\sum_{\ell=j+1}^{m_k}\mathrm{E}\{(\tilde{X}_{k,i}^{\intercal}\mathbf{B}_n\mathbf{B}_n^{\intercal}\tilde{X}_{k,\ell})^2\}\,\mathrm{E}\{(\tilde{X}_{k,j}^{\intercal}\mathbf{B}_n\mathbf{B}_n^{\intercal}\tilde{X}_{k,\ell})^2\}\\
&+2\sum_{i=1}^{m_k}\sum_{j=i+1}^{m_k}\sum_{\ell=j+1}^{m_k}\sum_{r=\ell+1}^{m_k}\mathrm{E}\{(\tilde{X}_{k,i}^{\intercal}\mathbf{B}_n\mathbf{B}_n^{\intercal}\tilde{X}_{k,j})^2\}\,\mathrm{E}\{(\tilde{X}_{k,\ell}^{\intercal}\mathbf{B}_n\mathbf{B}_n^{\intercal}\tilde{X}_{k,r})^2\}\\
&+2\sum_{i=1}^{m_k}\sum_{j=i+1}^{m_k}\sum_{\ell=j+1}^{m_k}\sum_{r=\ell+1}^{m_k}\mathrm{E}\{(\tilde{X}_{k,i}^{\intercal}\mathbf{B}_n\mathbf{B}_n^{\intercal}\tilde{X}_{k,\ell})^2\}\,\mathrm{E}\{(\tilde{X}_{k,j}^{\intercal}\mathbf{B}_n\mathbf{B}_n^{\intercal}\tilde{X}_{k,r})^2\}\\
&+2\sum_{i=1}^{m_k}\sum_{j=i+1}^{m_k}\sum_{\ell=j+1}^{m_k}\sum_{r=\ell+1}^{m_k}\mathrm{E}\{(\tilde{X}_{k,i}^{\intercal}\mathbf{B}_n\mathbf{B}_n^{\intercal}\tilde{X}_{k,r})^2\}\,\mathrm{E}\{(\tilde{X}_{k,j}^{\intercal}\mathbf{B}_n\mathbf{B}_n^{\intercal}\tilde{X}_{k,\ell})^2\}.
\end{aligned}
$$

We denote the above 7 terms by $C_1, \ldots, C_7$. On the other hand,

$$
\begin{aligned}
&\left\{\sum_{i=1}^{m_k}\sum_{j=i+1}^{m_k}(\tilde{X}_{k,i}^\intercal \mathbf{B}_n\mathbf{B}_n^\intercal \tilde{X}_{k,j})^2\right\}^2\\
=&\sum_{i=1}^{m_k}\sum_{j=i+1}^{m_k}(\tilde{X}_{k,i}^\intercal \mathbf{B}_n\mathbf{B}_n^\intercal \tilde{X}_{k,j})^4+2\sum_{i=1}^{m_k}\sum_{j=i+1}^{m_k}\sum_{\ell=j+1}^{m_k}(\tilde{X}_{k,i}^\intercal \mathbf{B}_n\mathbf{B}_n^\intercal \tilde{X}_{k,j})^2(\tilde{X}_{k,i}^\intercal \mathbf{B}_n\mathbf{B}_n^\intercal \tilde{X}_{k,\ell})^2\\
&+2\sum_{i=1}^{m_k}\sum_{j=i+1}^{m_k}\sum_{\ell=j+1}^{m_k}(\tilde{X}_{k,i}^\intercal \mathbf{B}_n\mathbf{B}_n^\intercal \tilde{X}_{k,j})^2(\tilde{X}_{k,j}^\intercal \mathbf{B}_n\mathbf{B}_n^\intercal \tilde{X}_{k,\ell})^2\\
&+2\sum_{i=1}^{m_k}\sum_{j=i+1}^{m_k}\sum_{\ell=j+1}^{m_k}(\tilde{X}_{k,i}^\intercal \mathbf{B}_n\mathbf{B}_n^\intercal \tilde{X}_{k,\ell})^2(\tilde{X}_{k,j}^\intercal \mathbf{B}_n\mathbf{B}_n^\intercal \tilde{X}_{k,\ell})^2\\
&+2\sum_{i=1}^{m_k}\sum_{j=i+1}^{m_k}\sum_{\ell=j+1}^{m_k}\sum_{r=\ell+1}^{m_k}(\tilde{X}_{k,i}^\intercal \mathbf{B}_n\mathbf{B}_n^\intercal \tilde{X}_{k,j})^2(\tilde{X}_{k,\ell}^\intercal \mathbf{B}_n\mathbf{B}_n^\intercal \tilde{X}_{k,r})^2\\
&+2\sum_{i=1}^{m_k}\sum_{j=i+1}^{m_k}\sum_{\ell=j+1}^{m_k}\sum_{r=\ell+1}^{m_k}(\tilde{X}_{k,i}^\intercal \mathbf{B}_n\mathbf{B}_n^\intercal \tilde{X}_{k,\ell})^2(\tilde{X}_{k,j}^\intercal \mathbf{B}_n\mathbf{B}_n^\intercal \tilde{X}_{k,r})^2\\
&+2\sum_{i=1}^{m_k}\sum_{j=i+1}^{m_k}\sum_{\ell=j+1}^{m_k}\sum_{r=\ell+1}^{m_k}(\tilde{X}_{k,i}^\intercal \mathbf{B}_n\mathbf{B}_n^\intercal \tilde{X}_{k,r})^2(\tilde{X}_{k,j}^\intercal \mathbf{B}_n\mathbf{B}_n^\intercal \tilde{X}_{k,\ell})^2.
\end{aligned}
$$

We denote the above 7 terms by $T_1, \ldots, T_7$. It can be seen that for $i = 5, 6, 7$, $\mathrm{E}(T_i) = C_i$. Thus,

$$
\operatorname{var}\left\{\sum_{i=1}^{m_k}\sum_{j=i+1}^{m_k}(\tilde{X}_{k,i}^\intercal \mathbf{B}_n\mathbf{B}_n^\intercal \tilde{X}_{k,j})^2\right\}=\sum_{i=1}^{4}\mathrm{E}(T_i)-\sum_{i=1}^{4}C_i\le\sum_{i=1}^{4}\mathrm{E}(T_i).
$$

Note that for $k = 1, 2$ and $i, j, \ell \in \{1, \ldots, n_k\}$,

$$
\mathrm{E}\left\{(\tilde{X}_{k,i}^\intercal \mathbf{B}_n\mathbf{B}_n^\intercal \tilde{X}_{k,j})^2(\tilde{X}_{k,i}^\intercal \mathbf{B}_n\mathbf{B}_n^\intercal \tilde{X}_{k,\ell})^2\right\}\le\left[\mathrm{E}\left\{(\tilde{X}_{k,i}^\intercal \mathbf{B}_n\mathbf{B}_n^\intercal \tilde{X}_{k,j})^4\right\}\mathrm{E}\left\{(\tilde{X}_{k,i}^\intercal \mathbf{B}_n\mathbf{B}_n^\intercal \tilde{X}_{k,\ell})^4\right\}\right]^{1/2}.
$$

Consequently,

$$
\operatorname{var}\left\{\sum_{i=1}^{m_k}\sum_{j=i+1}^{m_k}(\tilde{X}_{k,i}^\intercal \mathbf{B}_n\mathbf{B}_n^\intercal \tilde{X}_{k,j})^2\right\}\le\sum_{i=1}^{4}\mathrm{E}(T_i)\le\sum_{i=1}^{m_k}\left[\sum_{j=1}^{m_k}\left[\mathrm{E}\left\{(\tilde{X}_{k,i}^\intercal \mathbf{B}_n\mathbf{B}_n^\intercal \tilde{X}_{k,j})^4\right\}\right]^{1/2}\mathbf{1}_{\{j\neq i\}}\right]^2.
$$

From Lemma S.9, for $k = 1, 2$ and distinct $i, j \in \{1, \ldots, n_k\}$,

$$
\begin{aligned}
\mathrm{E}\{(\tilde{X}_{k,i}^\intercal \mathbf{B}_n\mathbf{B}_n^\intercal \tilde{X}_{k,j})^4\}\le&\mathrm{E}\{(\tilde{X}_{k,j}^\intercal \mathbf{B}_n\mathbf{B}_n^\intercal \tilde{X}_{k,i}\tilde{X}_{k,i}^\intercal \mathbf{B}_n\mathbf{B}_n^\intercal \tilde{X}_{k,j})^2\}\\
\le&\frac{\tau^2}{256}\Big\{\operatorname{tr}(\mathbf{B}_n^\intercal\mathbf{\Sigma}_{k,2i-1}\mathbf{B}_n\mathbf{B}_n^\intercal\mathbf{\Sigma}_{k,2j-1}\mathbf{B}_n)+\operatorname{tr}(\mathbf{B}_n^\intercal\mathbf{\Sigma}_{k,2i-1}\mathbf{B}_n\mathbf{B}_n^\intercal\mathbf{\Sigma}_{k,2j}\mathbf{B}_n)\\
&+\operatorname{tr}(\mathbf{B}_n^\intercal\mathbf{\Sigma}_{k,2i}\mathbf{B}_n\mathbf{B}_n^\intercal\mathbf{\Sigma}_{k,2j-1}\mathbf{B}_n)+\operatorname{tr}(\mathbf{B}_n^\intercal\mathbf{\Sigma}_{k,2i}\mathbf{B}_n\mathbf{B}_n^\intercal\mathbf{\Sigma}_{k,2j}\mathbf{B}_n)\Big\}^2.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\operatorname{var}\left\{\sum_{i=1}^{m_k}\sum_{j=i+1}^{m_k}(\tilde{X}_{k,i}^{\intercal}\mathbf{B}_n\mathbf{B}_n^{\intercal}\tilde{X}_{k,j})^2\right\} \leq & \frac{\tau^2}{128}n_k^2\sum_{i=1}^{n_k}\left\{\operatorname{tr}(\mathbf{B}_n^{\intercal}\boldsymbol{\Sigma}_{k,i}\mathbf{B}_n\mathbf{B}_n^{\intercal}\bar{\boldsymbol{\Sigma}}_k\mathbf{B}_n)\right\}^2 \\
\leq & \frac{\tau^2}{128}n_k^2\sum_{i=1}^{n_k}\operatorname{tr}\{(\mathbf{B}_n^{\intercal}\boldsymbol{\Sigma}_{k,i}\mathbf{B}_n)^2\}\operatorname{tr}\{(\mathbf{B}_n^{\intercal}\bar{\boldsymbol{\Sigma}}_k\mathbf{B}_n)^2\} \\
= & o\left(n_k^4\left(\operatorname{tr}(\bar{\boldsymbol{\Sigma}}_k^2)\right)^2\right),
\end{aligned}
$$

where the last equality follows from Assumption 3. Combining the above bound and (S.23) leads to

$$
\sum_{i=1}^{m_k}\sum_{j=i+1}^{m_k}(\tilde{X}_{k,i}^{\intercal}\mathbf{B}_n\mathbf{B}_n^{\intercal}\tilde{X}_{k,j})^2 = \frac{1}{32}n_k^2\operatorname{tr}\{(\mathbf{B}_n^{\intercal}\bar{\boldsymbol{\Sigma}}_k\mathbf{B}_n)^2\} + o\left(n_k^2\operatorname{tr}(\bar{\boldsymbol{\Sigma}}_k^2)\right). \tag{S.24}
$$

Now we deal with $\sum_{i=1}^{m_1}\sum_{j=1}^{m_2}(\tilde{X}_{1,i}^{\intercal}\mathbf{B}_n\mathbf{B}_n^{\intercal}\tilde{X}_{2,j})^2$. We have

$$
\operatorname{E}\left(\sum_{i=1}^{m_1}\sum_{j=1}^{m_2}(\tilde{X}_{1,i}^{\intercal}\mathbf{B}_n\mathbf{B}_n^{\intercal}\tilde{X}_{2,j})^2\right) = \frac{1}{16}\sum_{i=1}^{m_1}\sum_{j=1}^{m_2}\operatorname{tr}(\mathbf{B}_n^{\intercal}(\boldsymbol{\Sigma}_{1,2i-1}+\boldsymbol{\Sigma}_{1,2i})\mathbf{B}_n\mathbf{B}_n^{\intercal}(\boldsymbol{\Sigma}_{2,2j-1}+\boldsymbol{\Sigma}_{2,2j})\mathbf{B}_n).
$$

Hence from Assumption 3,

$$
\begin{aligned}
&\left|\operatorname{E}\left(\sum_{i=1}^{m_1}\sum_{j=1}^{m_2}(\tilde{X}_{1,i}^{\intercal}\mathbf{B}_n\mathbf{B}_n^{\intercal}\tilde{X}_{2,j})^2\right) - \frac{1}{16}n_1n_2\operatorname{tr}(\mathbf{B}_n^{\intercal}\bar{\boldsymbol{\Sigma}}_1\mathbf{B}_n\mathbf{B}_n^{\intercal}\bar{\boldsymbol{\Sigma}}_2\mathbf{B}_n)\right| \\
\leq & \frac{1}{16}n_2\operatorname{tr}(\mathbf{B}_n^{\intercal}\boldsymbol{\Sigma}_{1,n_1}\mathbf{B}_n\mathbf{B}_n^{\intercal}\bar{\boldsymbol{\Sigma}}_2\mathbf{B}_n) + \frac{1}{16}n_1\operatorname{tr}(\mathbf{B}_n^{\intercal}\boldsymbol{\Sigma}_{2,n_2}\mathbf{B}_n\mathbf{B}_n^{\intercal}\bar{\boldsymbol{\Sigma}}_1\mathbf{B}_n) \\
\leq & \frac{1}{16}\left\{n_2^2\operatorname{tr}\{(\mathbf{B}_n^{\intercal}\boldsymbol{\Sigma}_{1,n_1}\mathbf{B}_n)^2\}\operatorname{tr}\{(\mathbf{B}_n^{\intercal}\bar{\boldsymbol{\Sigma}}_2\mathbf{B}_n)^2\}\right\}^{1/2} + \frac{1}{16}\left\{n_1^2\operatorname{tr}\{(\mathbf{B}_n^{\intercal}\boldsymbol{\Sigma}_{2,n_2}\mathbf{B}_n)^2\}\operatorname{tr}\{(\mathbf{B}_n^{\intercal}\bar{\boldsymbol{\Sigma}}_1\mathbf{B}_n)^2\}\right\}^{1/2} \\
= & o\left[\left\{n_1^2n_2^2\operatorname{tr}(\bar{\boldsymbol{\Sigma}}_1^2)\operatorname{tr}(\bar{\boldsymbol{\Sigma}}_2^2)\right\}^{1/2}\right].
\end{aligned}
$$

Now we compute the variance of $\sum_{i=1}^{m_1}\sum_{j=1}^{m_2}(\tilde{X}_{1,i}^{\intercal}\mathbf{B}_n\mathbf{B}_n^{\intercal}\tilde{X}_{2,j})^2$. Note that

$$
\begin{aligned}
&\left(\sum_{i=1}^{m_1}\sum_{j=1}^{m_2}(\tilde{X}_{1,i}^{\intercal}\mathbf{B}_n\mathbf{B}_n^{\intercal}\tilde{X}_{2,j})^2\right)^2 \\
= & \sum_{i=1}^{m_1}\sum_{j=1}^{m_2}(\tilde{X}_{1,i}^{\intercal}\mathbf{B}_n\mathbf{B}_n^{\intercal}\tilde{X}_{2,j})^4 + 2\sum_{i=1}^{m_1}\sum_{\ell=1}^{m_2}\sum_{r=\ell+1}^{m_2}(\tilde{X}_{1,i}^{\intercal}\mathbf{B}_n\mathbf{B}_n^{\intercal}\tilde{X}_{2,\ell})^2(\tilde{X}_{1,i}^{\intercal}\mathbf{B}_n\mathbf{B}_n^{\intercal}\tilde{X}_{2,r})^2 \\
& + 2\sum_{i=1}^{m_1}\sum_{j=i+1}^{m_1}\sum_{\ell=1}^{m_2}(\tilde{X}_{1,i}^{\intercal}\mathbf{B}_n\mathbf{B}_n^{\intercal}\tilde{X}_{2,\ell})^2(\tilde{X}_{1,j}^{\intercal}\mathbf{B}_n\mathbf{B}_n^{\intercal}\tilde{X}_{2,\ell})^2 \\
& + 4\sum_{i=1}^{m_1}\sum_{j=i+1}^{m_1}\sum_{\ell=1}^{m_2}\sum_{r=\ell+1}^{m_2}(\tilde{X}_{1,i}^{\intercal}\mathbf{B}_n\mathbf{B}_n^{\intercal}\tilde{X}_{2,\ell})^2(\tilde{X}_{1,j}^{\intercal}\mathbf{B}_n\mathbf{B}_n^{\intercal}\tilde{X}_{2,r})^2.
\end{aligned}
$$

Hence

$$
\begin{aligned}
&\operatorname{var}\left(\sum_{i=1}^{m_1}\sum_{j=1}^{m_2}(\tilde{X}_{1,i}^{\intercal}\mathbf{B}_n\mathbf{B}_n^{\intercal}\tilde{X}_{2,j})^2\right)\\
\leq&\sum_{i=1}^{m_1}\sum_{j=1}^{m_2}\mathrm{E}\{(\tilde{X}_{1,i}^{\intercal}\mathbf{B}_n\mathbf{B}_n^{\intercal}\tilde{X}_{2,j})^4\}+2\sum_{i=1}^{m_1}\sum_{\ell=1}^{m_2}\sum_{r=\ell+1}^{m_2}\mathrm{E}\{(\tilde{X}_{1,i}^{\intercal}\mathbf{B}_n\mathbf{B}_n^{\intercal}\tilde{X}_{2,\ell})^2(\tilde{X}_{1,i}^{\intercal}\mathbf{B}_n\mathbf{B}_n^{\intercal}\tilde{X}_{2,r})^2\}\\
&+2\sum_{i=1}^{m_1}\sum_{j=i+1}^{m_1}\sum_{\ell=1}^{m_2}\mathrm{E}\{(\tilde{X}_{1,i}^{\intercal}\mathbf{B}_n\mathbf{B}_n^{\intercal}\tilde{X}_{2,\ell})^2(\tilde{X}_{1,j}^{\intercal}\mathbf{B}_n\mathbf{B}_n^{\intercal}\tilde{X}_{2,\ell})^2\}\\
\leq&\sum_{i=1}^{m_1}\left[\sum_{j=1}^{m_2}\left[\mathrm{E}\left\{(\tilde{X}_{1,i}^{\intercal}\mathbf{B}_n\mathbf{B}_n^{\intercal}\tilde{X}_{2,j})^4\right\}\right]^{1/2}\right]^2+\sum_{j=1}^{m_2}\left[\sum_{i=1}^{m_1}\left[\mathrm{E}\left\{(\tilde{X}_{1,i}^{\intercal}\mathbf{B}_n\mathbf{B}_n^{\intercal}\tilde{X}_{2,j})^4\right\}\right]^{1/2}\right]^2\\
\leq&\frac{\tau^2}{128}n_2^2\sum_{i=1}^{n_1}\left\{\operatorname{tr}(\mathbf{B}_n^{\intercal}\mathbf{\Sigma}_{1,i}\mathbf{B}_n\mathbf{B}_n^{\intercal}\bar{\mathbf{\Sigma}}_2\mathbf{B}_n)\right\}^2+\frac{\tau^2}{128}n_1^2\sum_{j=1}^{n_2}\left\{\operatorname{tr}(\mathbf{B}_n^{\intercal}\bar{\mathbf{\Sigma}}_1\mathbf{B}_n\mathbf{B}_n^{\intercal}\mathbf{\Sigma}_{2,j}\mathbf{B}_n)\right\}^2\\
\leq&\frac{\tau^2}{128}n_2^2\sum_{i=1}^{n_1}\operatorname{tr}(\mathbf{\Sigma}_{1,i}^2)\operatorname{tr}(\bar{\mathbf{\Sigma}}_2^2)+\frac{\tau^2}{128}n_1^2\sum_{j=1}^{n_2}\operatorname{tr}(\mathbf{\Sigma}_{2,i}^2)\operatorname{tr}(\bar{\mathbf{\Sigma}}_1^2)\\
=&o\left(n_1^2n_2^2\operatorname{tr}(\bar{\mathbf{\Sigma}}_1^2)\operatorname{tr}(\bar{\mathbf{\Sigma}}_2^2)\right).
\end{aligned}
$$

Thus,

$$
\sum_{i=1}^{m_1}\sum_{j=1}^{m_2}(\tilde{X}_{1,i}^{\intercal}\mathbf{B}_n\mathbf{B}_n^{\intercal}\tilde{X}_{2,j})^2=\frac{1}{16}n_1n_2\operatorname{tr}(\mathbf{B}_n^{\intercal}\bar{\mathbf{\Sigma}}_1\mathbf{B}_n\mathbf{B}_n^{\intercal}\bar{\mathbf{\Sigma}}_2\mathbf{B}_n)+o_P\left[\left\{n_1^2n_2^2\operatorname{tr}(\bar{\mathbf{\Sigma}}_1^2)\operatorname{tr}(\bar{\mathbf{\Sigma}}_2^2)\right\}^{1/2}\right].
$$

It follows from (S.24) and the above equality that

$$
\begin{aligned}
&\sum_{k=1}^{2}\frac{4\sum_{i=1}^{m_k}\sum_{j=i+1}^{m_k}(\tilde{X}_{k,i}^{\intercal}\mathbf{B}_n\mathbf{B}_n^{\intercal}\tilde{X}_{k,j})^2}{m_k^2(m_k-1)^2}+\frac{4\sum_{i=1}^{m_1}\sum_{j=1}^{m_2}(\tilde{X}_{1,i}^{\intercal}\mathbf{B}_n\mathbf{B}_n^{\intercal}\tilde{X}_{2,j})^2}{m_1^2m_2^2}\\
=&(1+o(1))\left[\sum_{k=1}^{2}\frac{2\operatorname{tr}\{(\mathbf{B}_n^{\intercal}\bar{\mathbf{\Sigma}}_k\mathbf{B}_n)^2\}}{n_k^2}+\frac{4\operatorname{tr}(\mathbf{B}_n^{\intercal}\bar{\mathbf{\Sigma}}_1\mathbf{B}_n\mathbf{B}_n^{\intercal}\bar{\mathbf{\Sigma}}_2\mathbf{B}_n)}{n_1n_2}\right]\\
&+o_P\left[\sum_{k=1}^{2}\frac{\operatorname{tr}(\bar{\mathbf{\Sigma}}_k^2)}{n_k^2}+\frac{\left\{\operatorname{tr}(\bar{\mathbf{\Sigma}}_1^2)\operatorname{tr}(\bar{\mathbf{\Sigma}}_2^2)\right\}^{1/2}}{n_1n_2}\right].
\end{aligned}
$$

Then the conclusion follows. □

Lemma S.11. *Suppose the conditions of Theorem 2 hold. Let* $E=(\epsilon_{1,1},\ldots,\epsilon_{1,m_1},\epsilon_{2,1},\ldots,\epsilon_{2,m_2})^{\intercal}$*, where* $\epsilon_{k,i}$*,* $i=1,\ldots,m_k$*,* $k=1,2$*, are independent Rademacher random variables. Let* $E_k^*=(\epsilon_{1,1}^*,\ldots,\epsilon_{1,m_1}^*,\epsilon_{2,1}^*,\ldots,\epsilon_{2,m_2}^*)^{\intercal}$*, where* $\epsilon_{k,i}^*$*,* $i=1,\ldots,m_k$*,* $k=1,2$*, are independent standard normal random variables. Then as* $n\to\infty$*,*

$$
\left\|\mathcal{L}\left(\frac{T_{\mathrm{CQ}}(E^*;\tilde{\mathbf{X}}_1,\tilde{\mathbf{X}}_2)}{\sigma_{T,n}}\mid\tilde{\mathbf{X}}_1,\tilde{\mathbf{X}}_2\right)-\mathcal{L}\left(\frac{T_{\mathrm{CQ}}(E;\tilde{\mathbf{X}}_1,\tilde{\mathbf{X}}_2)}{\sigma_{T,n}}\mid\tilde{\mathbf{X}}_1,\tilde{\mathbf{X}}_2\right)\right\|_3\xrightarrow{P}0.
$$

Proof. We apply Theorem S.1 conditioning on $\tilde{\mathbf{X}}_1$ and $\tilde{\mathbf{X}}_2$. Define

$$
\xi_i=\begin{cases}\epsilon_{1,i} & \text{for } i=1,\ldots,m_1,\\ \epsilon_{2,i-m_1} & \text{for } i=m_1+1,\ldots,m_1+m_2,\end{cases}\qquad\text{and}\qquad \eta_i=\begin{cases}\epsilon_{1,i}^* & \text{for } i=1,\ldots,m_1,\\ \epsilon_{2,i-m_1}^* & \text{for } i=m_1+1,\ldots,m_1+m_2.\end{cases}
$$

Define

$$w_{i,j}(a,b) = \begin{cases} \frac{2ab\tilde{X}_{1,i}^\intercal \tilde{X}_{1,j}}{m_1(m_1-1)\sigma_{T,n}} & \text{for } 1 \le i < j \le m_1, \\ \frac{-2ab\tilde{X}_{1,i}^\intercal \tilde{X}_{2,j-m_1}}{m_1 m_2 \sigma_{T,n}} & \text{for } 1 \le i \le m_1, m_1+1 \le j \le m_1+m_2, \\ \frac{2ab\tilde{X}_{2,i-m_1}^\intercal \tilde{X}_{2,j-m_1}}{m_2(m_2-1)\sigma_{T,n}} & \text{for } m_1+1 \le i < j \le m_1+m_2. \end{cases}$$

With the above definitions, we have $W(\xi_1,\ldots,\xi_n) = T_{\mathrm{CQ}}(E;\tilde{\mathbf{X}}_1,\tilde{\mathbf{X}}_2)/\sigma_{T,n}$ and $W(\eta_1,\ldots,\eta_n) = T_{\mathrm{CQ}}(E^*;\tilde{\mathbf{X}}_1,\tilde{\mathbf{X}}_2)/\sigma_{T,n}$. It can be easily seen that Assumptions S.1 and S.2 hold. By direct calculation, we have

$$\sigma_{i,j}^2 = \begin{cases} \frac{4(\tilde{X}_{1,i}^\intercal \tilde{X}_{1,j})^2}{m_1^2(m_1-1)^2\sigma_{T,n}^2} & \text{for } 1 \le i < j \le m_1, \\ \frac{4(\tilde{X}_{1,i}^\intercal \tilde{X}_{2,j-m_1})^2}{m_1^2 m_2^2 \sigma_{T,n}^2} & \text{for } 1 \le i \le m_1, m_1+1 \le j \le m_1+m_2, \\ \frac{4(\tilde{X}_{2,i-m_1}^\intercal \tilde{X}_{2,j-m_1})^2}{m_2^2(m_2-1)^2\sigma_{T,n}^2} & \text{for } m_1+1 \le i < j \le m_1+m_2. \end{cases}$$

Hence

$$\mathrm{Inf}_i = \begin{cases} \frac{4\sum_{j=1}^{m_1}(\tilde{X}_{1,i}^\intercal \tilde{X}_{1,j})^2 \mathbf{1}_{\{j\neq i\}}}{m_1^2(m_1-1)^2\sigma_{T,n}^2} + \frac{4\sum_{j=1}^{m_2}(\tilde{X}_{1,i}^\intercal \tilde{X}_{2,j})^2}{m_1^2 m_2^2\sigma_{T,n}^2} & \text{for } 1 \le i \le m_1, \\ \frac{4\sum_{j=1}^{m_2}(\tilde{X}_{2,i-m_1}^\intercal \tilde{X}_{2,j})^2 \mathbf{1}_{\{j\neq i\}}}{m_2^2(m_2-1)^2\sigma_{T,n}^2} + \frac{4\sum_{j=1}^{m_1}(\tilde{X}_{1,j}^\intercal \tilde{X}_{2,i-m_1})^2}{m_1^2 m_2^2\sigma_{T,n}^2} & \text{for } m_1+1 \le i \le m_1+m_2. \end{cases}$$

It can be easily seen that $\rho_n = 9$ for the above defined random variables. From Theorem S.1, it suffices to prove that $\sum_{i=1}^{m_1+m_2} \mathrm{Inf}_i^{3/2} \xrightarrow{P} 0$. We have

$$\sum_{i=1}^{m_1+m_2} \mathrm{Inf}_i^{3/2} \le \left(\max_{i\in\{1,\ldots,m_1+m_2\}} \mathrm{Inf}_i\right)^{1/2} \left(\sum_{i=1}^{m_1+m_2} \mathrm{Inf}_i\right) \le \left(\sum_{i=1}^{m_1+m_2} \mathrm{Inf}_i^2\right)^{1/4} \left(\sum_{i=1}^{m_1+m_2} \mathrm{Inf}_i\right).$$

But

$$\begin{aligned}\sum_{i=1}^{m_1+m_2} \mathrm{Inf}_i =& \sum_{k=1}^{2} \frac{4\sum_{i=1}^{m_k}\sum_{j=1}^{m_k}(\tilde{X}_{k,i}^\intercal \tilde{X}_{k,j})^2 \mathbf{1}_{\{j\neq i\}}}{m_k^2(m_k-1)^2\sigma_{T,n}^2} + \frac{8\sum_{i=1}^{m_1}\sum_{j=1}^{m_2}(\tilde{X}_{2,i}^\intercal \tilde{X}_{1,j})^2}{m_1^2 m_2^2 \sigma_{T,n}^2} \\ =& 2\frac{\mathrm{var}\{T_{\mathrm{CQ}}(E^*;\tilde{\mathbf{X}}_1,\tilde{\mathbf{X}}_2) \mid \tilde{\mathbf{X}}_1,\tilde{\mathbf{X}}_2\}}{\sigma_{T,n}^2} \\ =& 2 + o_P(1),\end{aligned}$$

where the last equality follows from Lemma S.10. Hence it suffices to prove that $\sum_{i=1}^{m_1+m_2} \mathrm{Inf}_i^2 \xrightarrow{P} 0$. We have

$$\begin{aligned}\mathrm{E}\left(\sum_{i=1}^{m_1} \mathrm{Inf}_i^2\right) \le& \frac{32}{m_1^4(m_1-1)^4\sigma_{T,n}^4} \mathrm{E}\left[\sum_{i=1}^{m_1}\left\{\sum_{j=1}^{m_1}(\tilde{X}_{1,i}^\intercal \tilde{X}_{1,j})^2 \mathbf{1}_{\{j\neq i\}}\right\}^2\right] \\ &+ \frac{32}{m_1^4 m_2^4 \sigma_{T,n}^4} \mathrm{E}\left[\sum_{i=1}^{m_1}\left\{\sum_{j=1}^{m_2}(\tilde{X}_{1,i}^\intercal \tilde{X}_{2,j})^2\right\}^2\right].\end{aligned}$$

We have

$$
\begin{aligned}
&\mathrm{E}\left[\sum_{i=1}^{m_1}\left\{\sum_{j=1}^{m_1}(\tilde{X}_{1,i}^\intercal \tilde{X}_{1,j})^2\mathbf{1}_{\{j\neq i\}}\right\}^2\right]\\
\leq&\sum_{i=1}^{m_1}\sum_{j=1}^{m_1}\sum_{\ell=1}^{m_1}\left[\mathrm{E}\left\{(\tilde{X}_{1,i}^\intercal \tilde{X}_{1,j})^4\right\}\mathrm{E}\left\{(\tilde{X}_{1,i}^\intercal \tilde{X}_{1,\ell})^4\right\}\right]^{1/2}\mathbf{1}_{\{j\neq i\}}\mathbf{1}_{\{\ell\neq i\}}\\
\leq&\frac{\tau^2}{256}\sum_{i=1}^{m_1}\sum_{j=1}^{m_1}\sum_{\ell=1}^{m_1}\left\{\sum_{i'=2i-1}^{2i}\sum_{j'=2j-1}^{2j}\operatorname{tr}(\mathbf{\Sigma}_{1,i'}\mathbf{\Sigma}_{1,j'})\right\}\left\{\sum_{i'=2i-1}^{2i}\sum_{\ell'=2\ell-1}^{2\ell}\operatorname{tr}(\mathbf{\Sigma}_{1,i'}\mathbf{\Sigma}_{1,\ell'})\right\}\\
\leq&\frac{\tau^2}{256}n_1^2\sum_{i=1}^{m_1}\left\{\sum_{i'=2i-1}^{2i}\operatorname{tr}(\mathbf{\Sigma}_{1,i'}\bar{\mathbf{\Sigma}}_1)\right\}^2\\
\leq&\frac{\tau^2}{128}n_1^2\sum_{i=1}^{n_1}\operatorname{tr}(\mathbf{\Sigma}_{1,i}^2)\operatorname{tr}(\bar{\mathbf{\Sigma}}_1^2)\\
=&o\left[n_1^4\left\{\operatorname{tr}(\bar{\mathbf{\Sigma}}_1^2)\right\}^2\right],
\end{aligned}
$$

where the second inequality follows from Lemma S.9 and the last equality follows from Assumption 3. On the other hand,

$$
\begin{aligned}
\mathrm{E}\left[\sum_{i=1}^{m_1}\left\{\sum_{j=1}^{m_2}(\tilde{X}_{1,i}^\intercal \tilde{X}_{2,j})^2\right\}^2\right]\leq&\sum_{i=1}^{m_1}\sum_{j=1}^{m_2}\sum_{\ell=1}^{m_2}\left[\mathrm{E}\left\{(\tilde{X}_{1,i}^\intercal \tilde{X}_{2,j})^4\right\}\mathrm{E}\left\{(\tilde{X}_{1,i}^\intercal \tilde{X}_{2,\ell})^4\right\}\right]^{1/2}\\
\leq&\frac{\tau^2}{128}n_2^2\sum_{i=1}^{n_1}\operatorname{tr}(\mathbf{\Sigma}_{1,i}^2)\operatorname{tr}(\bar{\mathbf{\Sigma}}_2^2)\\
=&o\left\{n_1^2n_2^2\operatorname{tr}(\bar{\mathbf{\Sigma}}_1^2)\operatorname{tr}(\bar{\mathbf{\Sigma}}_2^2)\right\}.
\end{aligned}
$$

Thus,

$$
\sum_{i=1}^{m_1}\operatorname{Inf}_i^2=o_P\left[\frac{\left\{\operatorname{tr}(\bar{\mathbf{\Sigma}}_1^2)\right\}^2}{n_1^4\sigma_{T,n}^4}+\frac{\operatorname{tr}(\bar{\mathbf{\Sigma}}_1^2)\operatorname{tr}(\bar{\mathbf{\Sigma}}_2^2)}{n_1^2n_2^2\sigma_{T,n}^4}\right]=o_P\left[\frac{\left\{\operatorname{tr}(\bar{\mathbf{\Sigma}}_1^2)\right\}^2}{n_1^4\sigma_{T,n}^4}+\frac{\left\{\operatorname{tr}(\bar{\mathbf{\Sigma}}_2^2)\right\}^2}{n_2^4\sigma_{T,n}^4}\right]=o_P(1).
$$

Similarly,

$$
\sum_{i=m_1+1}^{m_1+m_2}\operatorname{Inf}_i^2=o_P(1).
$$

This completes the proof. □

Lemma S.12. *Suppose the conditions of Theorem 2 hold. Furthermore, suppose the condition* (S.15) *holds. Then* (S.17) *holds.*

Proof. Let $E_1^*=(e_{1,1}^*,\ldots,e_{1,m_1}^*)^\intercal$ and $E_2^*=(e_{2,1}^*,\ldots,e_{2,m_2}^*)^\intercal$. Define

$$
T_{1,r}^*=\left\|\frac{1}{m_1}\mathbf{U}_r^\intercal\tilde{\mathbf{X}}_1^\intercal E_1^*-\frac{1}{m_2}\mathbf{U}_r^\intercal\tilde{\mathbf{X}}_2^\intercal E_2^*\right\|^2-\frac{1}{m_1^2}\|\tilde{\mathbf{X}}_1\mathbf{U}_r\|_F^2-\frac{1}{m_2^2}\|\tilde{\mathbf{X}}_2\mathbf{U}_r\|_F^2.
$$

From Lemma S.4 it suffices to prove

$$\mathrm{E}\left\{\frac{(T_{1,r}-T_{1,r}^*)^2}{\sigma_{T,n}^2}\mid \tilde{\mathbf{X}}_1,\tilde{\mathbf{X}}_2\right\}\xrightarrow{P}0, \tag{S.25}$$

and

$$\left\|\mathcal{L}\left(\frac{T_{1,r}^*+T_{3,r}}{\sigma_{T,n}}\mid \tilde{\mathbf{X}}_1,\tilde{\mathbf{X}}_2\right)-\mathcal{L}\left((1-\sum_{i=1}^{\infty}\kappa_i^2)^{1/2}\xi_0+2^{-1/2}\sum_{i=1}^{r}\kappa_i(\xi_i^2-1)\right)\right\|_3\xrightarrow{P}0. \tag{S.26}$$

To prove (S.25), we only need to show that $\mathrm{E}\{(T_{1,r}-T_{1,r}^*)^2\}=o(\sigma_{T,n}^2)$ where the expectation is unconditional. It is straightforward to show that

$$T_{1,r}-T_{1,r}^*=\sum_{k=1}^{2}\frac{\left\|\mathbf{U}_r^\intercal\tilde{\mathbf{X}}_k^\intercal E_k^*\right\|^2-\left\|\tilde{\mathbf{X}}_k\mathbf{U}_r\right\|_F^2}{m_k^2(m_k-1)}-\sum_{k=1}^{2}\sum_{i=1}^{m_k}\frac{(\epsilon_{k,i}^{*2}-1)\left\|\mathbf{U}_r^\intercal\tilde{X}_{k,i}\right\|^2}{m_k(m_k-1)}.$$

For $k=1,2$, we have

$$\mathrm{E}\left\{\left(\sum_{i=1}^{m_k}\frac{(\epsilon_{k,i}^{*2}-1)\left\|\mathbf{U}_r^\intercal\tilde{X}_{k,i}\right\|^2}{m_k(m_k-1)}\right)^2\right\}=\frac{2}{m_k^2(m_k-1)^2}\sum_{i=1}^{m_k}\mathrm{E}\{(\tilde{X}_{k,i}^\intercal\mathbf{U}_r\mathbf{U}_r^\intercal\tilde{X}_{k,i})^2\}.$$

But

$$\sum_{i=1}^{m_k}\mathrm{E}\{(\tilde{X}_{k,i}^\intercal\mathbf{U}_r\mathbf{U}_r^\intercal\tilde{X}_{k,i})^2\}\le\frac{\tau}{8}\sum_{i=1}^{n_k}\{\mathrm{tr}(\mathbf{U}_r^\intercal\mathbf{\Sigma}_{k,i}\mathbf{U}_r)\}^2\le\frac{\tau r}{8}\sum_{i=1}^{n_k}\mathrm{tr}\{(\mathbf{U}_r^\intercal\mathbf{\Sigma}_{k,i}\mathbf{U}_r)^2\}, \tag{S.27}$$

where the first inequality follows from Lemma S.9 and the second inequality follows from Cauchy-Schwarz inequality. Note that $\mathrm{tr}\{(\mathbf{U}_r^\intercal\mathbf{\Sigma}_{k,i}\mathbf{U}_r)^2\}\le\mathrm{tr}(\mathbf{\Sigma}_{k,i}^2)$. Thus,

$$\mathrm{E}\left\{\left(\sum_{i=1}^{m_k}\frac{(\epsilon_{k,i}^{*2}-1)\left\|\mathbf{U}_r^\intercal\tilde{X}_{k,i}\right\|^2}{m_k(m_k-1)}\right)^2\right\}=O\left(\frac{1}{n_k^4}\sum_{i=1}^{n_k}\mathrm{tr}(\mathbf{\Sigma}_{k,i}^2)\right)=o\left(\frac{1}{n_k^2}\mathrm{tr}(\bar{\mathbf{\Sigma}}_k^2)\right)=o(\sigma_{T,n}^2),$$

where the second equality follows from Assumption 3. On the other hand, for $k=1,2$, we have

$$\mathrm{E}\left\{\left(\frac{\left\|\mathbf{U}_r^\intercal\tilde{\mathbf{X}}_k^\intercal E_k^*\right\|^2-\|\tilde{\mathbf{X}}_k\mathbf{U}_r\|_F^2}{m_k^2(m_k-1)}\right)^2\right\}=\frac{2}{m_k^4(m_k-1)^2}\mathrm{E}\left[\mathrm{tr}\left\{(\mathbf{U}_r^\intercal\tilde{\mathbf{X}}_k^\intercal\tilde{\mathbf{X}}_k\mathbf{U}_r)^2\right\}\right].$$

But

$$\begin{aligned}\mathrm{E}\left[\mathrm{tr}\left\{(\mathbf{U}_r^\intercal\tilde{\mathbf{X}}_k^\intercal\tilde{\mathbf{X}}_k\mathbf{U}_r)^2\right\}\right]=&\sum_{i=1}^{m_k}\mathrm{E}\left[\mathrm{tr}\left\{(\mathbf{U}_r^\intercal\tilde{X}_{k,i}\tilde{X}_{k,i}^\intercal\mathbf{U}_r)^2\right\}\right]\\&+\frac{1}{8}\sum_{i=1}^{m_k}\sum_{j=i+1}^{m_k}\mathrm{tr}\left\{(\mathbf{U}_r^\intercal(\mathbf{\Sigma}_{k,2i-1}+\mathbf{\Sigma}_{k,2i})\mathbf{U}_r)(\mathbf{U}_r^\intercal(\mathbf{\Sigma}_{k,2j-1}+\mathbf{\Sigma}_{k,2j})\mathbf{U}_r)\right\}\\\le&\frac{\tau r}{8}\sum_{i=1}^{n_k}\mathrm{tr}\{(\mathbf{U}_r^\intercal\mathbf{\Sigma}_{k,i}\mathbf{U}_r)^2\}+\frac{1}{8}\sum_{i=1}^{n_k}\sum_{j=i+1}^{n_k}\mathrm{tr}\{(\mathbf{U}_r^\intercal\mathbf{\Sigma}_{k,i}\mathbf{U}_r)(\mathbf{U}_r^\intercal\mathbf{\Sigma}_{k,j}\mathbf{U}_r)\}\\\le&\frac{\tau r}{8}n_k^2\,\mathrm{tr}\{(\mathbf{U}_r^\intercal\bar{\mathbf{\Sigma}}_k\mathbf{U}_r)^2\}\\\le&\frac{\tau r}{8}n_k^2\,\mathrm{tr}(\bar{\mathbf{\Sigma}}_k^2),\end{aligned}$$

where the first inequality follows from (S.27). It follows that

$$\mathrm{E}\left\{\left(\frac{\left\|\mathbf{U}_r^\intercal\tilde{\mathbf{X}}_k^\intercal E_k^*\right\|^2-\|\tilde{\mathbf{X}}_k\mathbf{U}_r\|_F^2}{m_k^2(m_k-1)}\right)^2\right\}=o\left(\sigma_{T,n}^2\right).$$

Thus, (S.25) holds.

Now we apply Lemma S.6 to prove (S.26). In Lemma S.6, we take $\boldsymbol{\zeta}_n=E^*$, and

$$\mathbf{A}_n=\begin{pmatrix}\mathbf{A}_{1,1} & \mathbf{A}_{1,2}\\ \mathbf{A}_{1,2}^\intercal & \mathbf{A}_{2,2}\end{pmatrix},\quad \mathbf{B}_n=\sigma_{T,n}^{-1/2}\left(\frac{1}{m_1}\mathbf{U}_r^\intercal\tilde{\mathbf{X}}_1^\intercal,\quad -\frac{1}{m_2}\mathbf{U}_r^\intercal\tilde{\mathbf{X}}_2^\intercal\right),$$

where

$$\mathbf{A}_{1,1}=\frac{1}{\sigma_{T,n}m_1(m_1-1)}\begin{pmatrix}0 & \tilde{X}_{1,1}^\intercal\tilde{\mathbf{U}}_{r_n^*}\tilde{\mathbf{U}}_{r_n^*}^\intercal\tilde{X}_{1,2} & \cdots & \tilde{X}_{1,1}^\intercal\tilde{\mathbf{U}}_{r_n^*}\tilde{\mathbf{U}}_{r_n^*}^\intercal\tilde{X}_{1,m_1}\\ \tilde{X}_{1,2}^\intercal\tilde{\mathbf{U}}_{r_n^*}\tilde{\mathbf{U}}_{r_n^*}^\intercal\tilde{X}_{1,1} & 0 & \cdots & \tilde{X}_{1,2}^\intercal\tilde{\mathbf{U}}_{r_n^*}\tilde{\mathbf{U}}_{r_n^*}^\intercal\tilde{X}_{1,m_1}\\ \vdots & \vdots & \ddots & \vdots\\ \tilde{X}_{1,m_1}^\intercal\tilde{\mathbf{U}}_{r_n^*}\tilde{\mathbf{U}}_{r_n^*}^\intercal\tilde{X}_{1,1} & \tilde{X}_{1,m_1}^\intercal\tilde{\mathbf{U}}_{r_n^*}\tilde{\mathbf{U}}_{r_n^*}^\intercal\tilde{X}_{1,2} & \cdots & 0\end{pmatrix},$$

$$\mathbf{A}_{1,2}=-\frac{1}{\sigma_{T,n}m_1m_2}\begin{pmatrix}\tilde{X}_{1,1}^\intercal\tilde{\mathbf{U}}_{r_n^*}\tilde{\mathbf{U}}_{r_n^*}^\intercal\tilde{X}_{2,1} & \tilde{X}_{1,1}^\intercal\tilde{\mathbf{U}}_{r_n^*}\tilde{\mathbf{U}}_{r_n^*}^\intercal\tilde{X}_{2,2} & \cdots & \tilde{X}_{1,1}^\intercal\tilde{\mathbf{U}}_{r_n^*}\tilde{\mathbf{U}}_{r_n^*}^\intercal\tilde{X}_{2,m_2}\\ \tilde{X}_{1,2}^\intercal\tilde{\mathbf{U}}_{r_n^*}\tilde{\mathbf{U}}_{r_n^*}^\intercal\tilde{X}_{2,1} & \tilde{X}_{1,2}^\intercal\tilde{\mathbf{U}}_{r_n^*}\tilde{\mathbf{U}}_{r_n^*}^\intercal\tilde{X}_{2,2} & \cdots & \tilde{X}_{1,2}^\intercal\tilde{\mathbf{U}}_{r_n^*}\tilde{\mathbf{U}}_{r_n^*}^\intercal\tilde{X}_{2,m_2}\\ \vdots & \vdots & \ddots & \vdots\\ \tilde{X}_{1,m_1}^\intercal\tilde{\mathbf{U}}_{r_n^*}\tilde{\mathbf{U}}_{r_n^*}^\intercal\tilde{X}_{2,1} & \tilde{X}_{1,m_1}^\intercal\tilde{\mathbf{U}}_{r_n^*}\tilde{\mathbf{U}}_{r_n^*}^\intercal\tilde{X}_{2,2} & \cdots & \tilde{X}_{1,m_1}^\intercal\tilde{\mathbf{U}}_{r_n^*}\tilde{\mathbf{U}}_{r_n^*}^\intercal\tilde{X}_{2,m_2}\end{pmatrix},$$

$$\mathbf{A}_{2,2}=\frac{1}{\sigma_{T,n}m_2(m_2-1)}\begin{pmatrix}0 & \tilde{X}_{2,1}^\intercal\tilde{\mathbf{U}}_{r_n^*}\tilde{\mathbf{U}}_{r_n^*}^\intercal\tilde{X}_{2,2} & \cdots & \tilde{X}_{2,1}^\intercal\tilde{\mathbf{U}}_{r_n^*}\tilde{\mathbf{U}}_{r_n^*}^\intercal\tilde{X}_{2,m_2}\\ \tilde{X}_{2,2}^\intercal\tilde{\mathbf{U}}_{r_n^*}\tilde{\mathbf{U}}_{r_n^*}^\intercal\tilde{X}_{2,1} & 0 & \cdots & \tilde{X}_{2,2}^\intercal\tilde{\mathbf{U}}_{r_n^*}\tilde{\mathbf{U}}_{r_n^*}^\intercal\tilde{X}_{2,m_2}\\ \vdots & \vdots & \ddots & \vdots\\ \tilde{X}_{2,m_2}^\intercal\tilde{\mathbf{U}}_{r_n^*}\tilde{\mathbf{U}}_{r_n^*}^\intercal\tilde{X}_{2,1} & \tilde{X}_{2,m_2}^\intercal\tilde{\mathbf{U}}_{r_n^*}\tilde{\mathbf{U}}_{r_n^*}^\intercal\tilde{X}_{2,2} & \cdots & 0\end{pmatrix}.$$

Then we have

$$\boldsymbol{\zeta}_n^\intercal\mathbf{A}_n\boldsymbol{\zeta}_n-\operatorname{tr}(\mathbf{A}_n)=\frac{T_{\mathrm{CQ}}(E^*;\tilde{\mathbf{X}}_1\tilde{\mathbf{U}}_{r_n^*},\tilde{\mathbf{X}}_2\tilde{\mathbf{U}}_{r_n^*})}{\sigma_{T,n}}=\frac{T_{3,r}}{\sigma_{T,n}},$$

and

$$\boldsymbol{\zeta}_n^\intercal\mathbf{B}_n^\intercal\mathbf{B}_n\boldsymbol{\zeta}_n-\operatorname{tr}(\mathbf{B}_n^\intercal\mathbf{B}_n)=\frac{T_{1,r}^*}{\sigma_{T,n}}.$$

From Lemma S.10,

$$\begin{aligned}2\operatorname{tr}(\mathbf{A}_n^2)=&\frac{\operatorname{var}\{T_{\mathrm{CQ}}(E^*;\tilde{\mathbf{X}}_1\tilde{\mathbf{U}}_{r_n^*},\tilde{\mathbf{X}}_2\tilde{\mathbf{U}}_{r_n^*})\mid\tilde{\mathbf{X}}_1,\tilde{\mathbf{X}}_2\}}{\sigma_{T,n}^2}\\ =&\frac{\operatorname{tr}\{(\tilde{\mathbf{U}}_{r_n^*}^\intercal\boldsymbol{\Psi}_n\tilde{\mathbf{U}}_{r_n^*})^2\}}{\operatorname{tr}(\boldsymbol{\Psi}_n^2)}+o_P(1)\\ =&1-\frac{\sum_{i=1}^{r_n^*}\lambda_i^2(\boldsymbol{\Psi}_n)}{\operatorname{tr}\left(\boldsymbol{\Psi}_n^2\right)}+o_P(1)\\ =&1-\sum_{i=1}^{\infty}\kappa_i^2+o_P(1),\end{aligned}\tag{S.28}$$

where the last equality follows from (S.16). On the other hand,

$$\begin{aligned}\operatorname{tr}(\mathbf{A}_n^4) \leq& 8\left\{\operatorname{tr}\left\{\begin{pmatrix}\mathbf{A}_{1,1} & \mathbf{O}_{n_1\times n_2}\\ \mathbf{O}_{n_2\times n_1} & \mathbf{A}_{2,2}\end{pmatrix}^4\right\} + \operatorname{tr}\left\{\begin{pmatrix}\mathbf{O}_{n_1\times n_1} & \mathbf{A}_{1,2}\\ \mathbf{A}_{1,2}^\intercal & \mathbf{O}_{n_2\times n_2}\end{pmatrix}^4\right\}\right\}\\ =& 8\operatorname{tr}(\mathbf{A}_{1,1}^4) + 8\operatorname{tr}(\mathbf{A}_{2,2}^4) + 16\operatorname{tr}\{(\mathbf{A}_{1,2}\mathbf{A}_{1,2}^\intercal)^2\}.\end{aligned}$$

For $i,j = 1,\ldots,m_1$, let $w_{i,j} = \tilde{X}_{1,i}^\intercal \tilde{\mathbf{U}}_{r_n^*}\tilde{\mathbf{U}}_{r_n^*}^\intercal \tilde{X}_{1,j}$. We have

$$\sigma_{T,n}^4 m_1^4(m_1-1)^4\operatorname{tr}(\mathbf{A}_{1,1}^4) = \sum_{i=1}^{m_1}\sum_{j=1}^{m_1}\sum_{k=1}^{m_1}\sum_{\ell=1}^{m_1} w_{i,j}w_{j,k}w_{k,\ell}w_{\ell,i}\mathbf{1}_{\{i\neq j\}}\mathbf{1}_{\{j\neq k\}}\mathbf{1}_{\{k\neq \ell\}}\mathbf{1}_{\{\ell\neq i\}}.$$

We split the above sum into the following four cases: $k=i,\ell=j$; $k=i,\ell\neq j$; $k\neq i,\ell=j$; $k\neq i,\ell\neq j$. The second and the third cases result in the same sum. Then we have

$$\begin{aligned}\sigma_{T,n}^4 m_1^4(m_1-1)^4\operatorname{tr}(\mathbf{A}_{1,1}^4) =& \sum_{i=1}^{m_1}\sum_{j=1}^{m_1} w_{i,j}^4\mathbf{1}_{\{i\neq j\}} + 2\sum_{i=1}^{m_1}\sum_{j=1}^{m_1}\sum_{k=1}^{m_1} w_{i,j}^2 w_{i,k}^2\mathbf{1}_{\{i,j,k \text{ are distinct}\}}\\ &+\sum_{i=1}^{m_1}\sum_{j=1}^{m_1}\sum_{k=1}^{m_1}\sum_{\ell=1}^{m_1} w_{i,j}w_{j,k}w_{k,\ell}w_{\ell,i}\mathbf{1}_{\{i,j,k,\ell \text{ are distinct}\}}.\end{aligned}\tag{S.29}$$

First we deal with the first two terms of (S.29). For distinct $i,j\in\{1,\ldots,m_1\}$, two applications of Lemma S.9 yield

$$\begin{aligned}\operatorname{E}(w_{i,j}^4) =& \operatorname{E}\left(\tilde{X}_{1,i}^\intercal\tilde{\mathbf{U}}_{r_n^*}\tilde{\mathbf{U}}_{r_n^*}^\intercal\tilde{X}_{1,j}\tilde{X}_{1,j}^\intercal\tilde{\mathbf{U}}_{r_n^*}\tilde{\mathbf{U}}_{r_n^*}^\intercal\tilde{X}_{1,i}\right)^2\\ \leq& \frac{\tau^2}{256}\Big\{\operatorname{tr}(\tilde{\mathbf{U}}_{r_n^*}^\intercal\mathbf{\Sigma}_{1,2i-1}\tilde{\mathbf{U}}_{r_n^*}\tilde{\mathbf{U}}_{r_n^*}^\intercal\mathbf{\Sigma}_{1,2j-1}\tilde{\mathbf{U}}_{r_n^*}) + \operatorname{tr}(\tilde{\mathbf{U}}_{r_n^*}^\intercal\mathbf{\Sigma}_{1,2i-1}\tilde{\mathbf{U}}_{r_n^*}\tilde{\mathbf{U}}_{r_n^*}^\intercal\mathbf{\Sigma}_{1,2j}\tilde{\mathbf{U}}_{r_n^*})\\ &+\operatorname{tr}(\tilde{\mathbf{U}}_{r_n^*}^\intercal\mathbf{\Sigma}_{1,2i}\tilde{\mathbf{U}}_{r_n^*}\tilde{\mathbf{U}}_{r_n^*}^\intercal\mathbf{\Sigma}_{1,2j-1}\tilde{\mathbf{U}}_{r_n^*}) + \operatorname{tr}(\tilde{\mathbf{U}}_{r_n^*}^\intercal\mathbf{\Sigma}_{1,2i}\tilde{\mathbf{U}}_{r_n^*}\tilde{\mathbf{U}}_{r_n^*}^\intercal\mathbf{\Sigma}_{1,2j}\tilde{\mathbf{U}}_{r_n^*})\Big\}^2.\end{aligned}$$

Consequently,

$$\begin{aligned}&\operatorname{E}\left\{\sum_{i=1}^{m_1}\sum_{j=1}^{m_1} w_{i,j}^4\mathbf{1}_{\{i\neq j\}} + 2\sum_{i=1}^{m_1}\sum_{j=1}^{m_1}\sum_{k=1}^{m_1} w_{i,j}^2w_{i,k}^2\mathbf{1}_{\{i,j,k \text{ are distinct}\}}\right\}\\ \leq& 2\sum_{i=1}^{m_1}\left[\sum_{j=1}^{m_1}\{\operatorname{E}(w_{i,j}^4)\}^{1/2}\mathbf{1}_{\{i\neq j\}}\right]^2\\ \leq& \frac{\tau^2}{128}n_1^2\sum_{i=1}^{m_1}\left\{\operatorname{tr}(\tilde{\mathbf{U}}_{r_n^*}^\intercal\mathbf{\Sigma}_{1,2i-1}\tilde{\mathbf{U}}_{r_n^*}\tilde{\mathbf{U}}_{r_n^*}^\intercal\bar{\mathbf{\Sigma}}_1\tilde{\mathbf{U}}_{r_n^*}) + \operatorname{tr}(\tilde{\mathbf{U}}_{r_n^*}^\intercal\mathbf{\Sigma}_{1,2i}\tilde{\mathbf{U}}_{r_n^*}\tilde{\mathbf{U}}_{r_n^*}^\intercal\bar{\mathbf{\Sigma}}_1\tilde{\mathbf{U}}_{r_n^*})\right\}^2\\ \leq& \frac{\tau^2}{64}n_1^2\sum_{i=1}^{n_1}\left\{\operatorname{tr}(\tilde{\mathbf{U}}_{r_n^*}^\intercal\mathbf{\Sigma}_{1,i}\tilde{\mathbf{U}}_{r_n^*}\tilde{\mathbf{U}}_{r_n^*}^\intercal\bar{\mathbf{\Sigma}}_1\tilde{\mathbf{U}}_{r_n^*})\right\}^2\\ \leq& \frac{\tau^2}{64}n_1^2\operatorname{tr}(\bar{\mathbf{\Sigma}}_1^2)\sum_{i=1}^{n_1}\operatorname{tr}(\mathbf{\Sigma}_{1,i}^2)\\ =& o\left[n_1^4\left\{\operatorname{tr}(\bar{\mathbf{\Sigma}}_1^2)\right\}^2\right],\end{aligned}\tag{S.30}$$

where the last equality follows from Assumption 3. Now we deal with the third term of (S.29). For distinct

$i, j, k, \ell \in \{1, \ldots, m_1\}$, we have

$$\begin{aligned}&\mathrm{E}(w_{i,j}w_{j,k}w_{k,\ell}w_{\ell,i})\\=&\frac{1}{256}\operatorname{tr}\left\{\tilde{\mathbf{U}}_{r_n^*}^\intercal\left(\sum_{i^\dagger=2i-1}^{2i}\mathbf{\Sigma}_{1,i^\dagger}\right)\tilde{\mathbf{U}}_{r_n^*}\tilde{\mathbf{U}}_{r_n^*}^\intercal\left(\sum_{j^\dagger=2j-1}^{2j}\mathbf{\Sigma}_{1,j^\dagger}\right)\tilde{\mathbf{U}}_{r_n^*}\tilde{\mathbf{U}}_{r_n^*}^\intercal\left(\sum_{k^\dagger=2k-1}^{2k}\mathbf{\Sigma}_{1,k^\dagger}\right)\tilde{\mathbf{U}}_{r_n^*}\tilde{\mathbf{U}}_{r_n^*}^\intercal\left(\sum_{\ell^\dagger=2\ell-1}^{2\ell}\mathbf{\Sigma}_{1,\ell^\dagger}\right)\tilde{\mathbf{U}}_{r_n^*}\right\}.\end{aligned}$$

Then from Lemma S.2, we have

$$\begin{aligned}&\mathrm{E}\left\{\sum_{i=1}^{m_1}\sum_{j=1}^{m_1}\sum_{k=1}^{m_1}\sum_{\ell=1}^{m_1}w_{i,j}w_{j,k}w_{k,\ell}w_{\ell,i}\mathbf{1}_{\{i,j,k,\ell\text{ are distinct}\}}\right\}\\=&O\left[n_1^4\operatorname{tr}\left\{(\tilde{\mathbf{U}}_{r_n^*}^\intercal\bar{\mathbf{\Sigma}}_1\tilde{\mathbf{U}}_{r_n^*})^4\right\}+n_1^2\operatorname{tr}\left(\bar{\mathbf{\Sigma}}_1^2\right)\sum_{i=1}^{n_1}\operatorname{tr}\left(\mathbf{\Sigma}_{1,i}^2\right)+\left\{\sum_{i=1}^{n_1}\operatorname{tr}(\mathbf{\Sigma}_{1,i}^2)\right\}^2\right]\\=&O\left[n_1^6\left\{\lambda_{r_n^*+1}(\mathbf{\Psi}_n)\right\}^2\operatorname{tr}(\bar{\mathbf{\Sigma}}_1^2)\right]+o\left[n_1^4\{\operatorname{tr}(\bar{\mathbf{\Sigma}}_1^2)\}^2\right],\end{aligned}$$

where the last equality follows from (S.14) and Assumption 3. It follows from (S.30) and the above bound that

$$\operatorname{tr}(\mathbf{A}_{1,1}^4)=O_P\left\{\frac{\left\{\lambda_{r_n^*+1}\left(\mathbf{\Psi}_n\right)\right\}^2}{\sigma_{T,n}^2}\frac{\operatorname{tr}(\bar{\mathbf{\Sigma}}_1^2)}{n_1^2\sigma_{T,n}^2}\right\}+o_P\left[\left\{\frac{\operatorname{tr}(\bar{\mathbf{\Sigma}}_1^2)}{n_1^2\sigma_{T,n}^2}\right\}^2\right]=o_P(1).$$

Similarly, we have $\operatorname{tr}(\mathbf{A}_{2,2}^4)=o_P(1)$.

Now we deal with $\operatorname{tr}\{(\mathbf{A}_{1,2}\mathbf{A}_{1,2}^\intercal)^2\}$. For $i=1,\ldots,m_1$, $j=1,\ldots,m_2$, let $w_{i,j}^*=\tilde{X}_{1,i}^\intercal\tilde{\mathbf{U}}_{r_n^*}\tilde{\mathbf{U}}_{r_n^*}^\intercal\tilde{X}_{2,j}$. Then

$$\begin{aligned}\sigma_{T,n}^4m_1^4m_2^4\operatorname{tr}\{(\mathbf{A}_{1,2}\mathbf{A}_{1,2}^\intercal)^2\}=&\sum_{i=1}^{m_1}\sum_{k=1}^{m_2}w_{i,k}^{*4}+\sum_{i=1}^{m_1}\sum_{j=1}^{m_1}\sum_{k=1}^{m_2}w_{i,k}^{*2}w_{j,k}^{*2}\mathbf{1}_{\{i\neq j\}}+\sum_{i=1}^{m_1}\sum_{k=1}^{m_2}\sum_{\ell=1}^{m_2}w_{i,k}^{*2}w_{i,\ell}^{*2}\mathbf{1}_{\{k\neq\ell\}}\\&+\sum_{i=1}^{m_1}\sum_{j=1}^{m_1}\sum_{k=1}^{m_2}\sum_{\ell=1}^{m_2}w_{i,k}^*w_{j,k}^*w_{i,\ell}^*w_{j,\ell}^*\mathbf{1}_{\{i\neq j\}}\mathbf{1}_{\{k\neq\ell\}}.\end{aligned}\tag{S.31}$$

First we deal with the first three terms of (S.31). From Lemma S.9, for $i=1,\ldots,m_1$ and $j=1,\ldots,m_2$,

$$\begin{aligned}\mathrm{E}(w_{i,j}^{*4})=&\mathrm{E}\{(\tilde{X}_{1,i}^\intercal\tilde{\mathbf{U}}_{r_n^*}\tilde{\mathbf{U}}_{r_n^*}^\intercal\tilde{X}_{2,j}\tilde{X}_{2,j}^\intercal\tilde{\mathbf{U}}_{r_n^*}\tilde{\mathbf{U}}_{r_n^*}^\intercal\tilde{X}_{1,i})^2\}\\\leq&\frac{\tau^2}{256}\Big\{\operatorname{tr}(\tilde{\mathbf{U}}_{r_n^*}^\intercal\mathbf{\Sigma}_{1,2i-1}\tilde{\mathbf{U}}_{r_n^*}\tilde{\mathbf{U}}_{r_n^*}^\intercal\mathbf{\Sigma}_{2,2j-1}\tilde{\mathbf{U}}_{r_n^*})+\operatorname{tr}(\tilde{\mathbf{U}}_{r_n^*}^\intercal\mathbf{\Sigma}_{1,2i-1}\tilde{\mathbf{U}}_{r_n^*}\tilde{\mathbf{U}}_{r_n^*}^\intercal\mathbf{\Sigma}_{2,2j}\tilde{\mathbf{U}}_{r_n^*})\\&+\operatorname{tr}(\tilde{\mathbf{U}}_{r_n^*}^\intercal\mathbf{\Sigma}_{1,2i}\tilde{\mathbf{U}}_{r_n^*}\tilde{\mathbf{U}}_{r_n^*}^\intercal\mathbf{\Sigma}_{2,2j-1}\tilde{\mathbf{U}}_{r_n^*})+\operatorname{tr}(\tilde{\mathbf{U}}_{r_n^*}^\intercal\mathbf{\Sigma}_{1,2i}\tilde{\mathbf{U}}_{r_n^*}\tilde{\mathbf{U}}_{r_n^*}^\intercal\mathbf{\Sigma}_{2,2j}\tilde{\mathbf{U}}_{r_n^*})\Big\}^2.\end{aligned}$$

Consequently,

$$
\begin{aligned}
&\mathrm{E}\left\{\sum_{i=1}^{m_1}\sum_{k=1}^{m_2} w_{i,k}^{*4} + \sum_{i=1}^{m_1}\sum_{j=1}^{m_1}\sum_{k=1}^{m_2} w_{i,k}^{*2}w_{j,k}^{*2}\mathbf{1}_{\{i\neq j\}} + \sum_{i=1}^{m_1}\sum_{k=1}^{m_2}\sum_{\ell=1}^{m_2} w_{i,k}^{*2}w_{i,\ell}^{*2}\mathbf{1}_{\{k\neq \ell\}}\right\}\\
\leq&\sum_{i=1}^{m_1}\left[\sum_{j=1}^{m_2}\left\{\mathrm{E}(w_{i,j}^{*4})\right\}^{1/2}\right]^2 + \sum_{j=1}^{m_2}\left[\sum_{i=1}^{m_1}\left\{\mathrm{E}(w_{i,j}^{*4})\right\}^{1/2}\right]^2\\
\leq&\frac{\tau^2 n_2^2}{128}\sum_{i=1}^{n_1}\left\{\mathrm{tr}(\tilde{\mathbf{U}}_{r_n^*}^\intercal \mathbf{\Sigma}_{1,i}\tilde{\mathbf{U}}_{r_n^*}\tilde{\mathbf{U}}_{r_n^*}^\intercal\bar{\mathbf{\Sigma}}_2\tilde{\mathbf{U}}_{r_n^*})\right\}^2 + \frac{\tau^2 n_1^2}{128}\sum_{j=1}^{n_2}\left\{\mathrm{tr}(\tilde{\mathbf{U}}_{r_n^*}^\intercal \bar{\mathbf{\Sigma}}_1\tilde{\mathbf{U}}_{r_n^*}\tilde{\mathbf{U}}_{r_n^*}^\intercal\mathbf{\Sigma}_{2,j}\tilde{\mathbf{U}}_{r_n^*})\right\}^2\\
\leq&\frac{\tau^2 n_2^2}{128}\,\mathrm{tr}(\bar{\mathbf{\Sigma}}_2^2)\sum_{i=1}^{n_1}\mathrm{tr}(\mathbf{\Sigma}_{1,i}^2) + \frac{\tau^2 n_1^2}{128}\,\mathrm{tr}(\bar{\mathbf{\Sigma}}_1^2)\sum_{j=1}^{n_2}\mathrm{tr}(\mathbf{\Sigma}_{2,j}^2)\\
=&o\left(n_1^2 n_2^2\,\mathrm{tr}(\bar{\mathbf{\Sigma}}_1^2)\,\mathrm{tr}(\bar{\mathbf{\Sigma}}_2^2)\right)\\
=&o\left[n_1^4 n_2^4\left\{\mathrm{tr}(\mathbf{\Psi}_n^2)\right\}^2\right],
\end{aligned}
\tag{S.32}
$$

where the second last equality follows from Assumption 3. Now we deal with the fourth term of (S.31). For distinct $i, j \in \{1, \ldots, m_1\}$ and distinct $k, \ell \in \{1, \ldots, m_2\}$, we have

$$
\begin{aligned}
&\mathrm{E}(w_{i,k}^* w_{j,k}^* w_{i,\ell}^* w_{j,\ell}^*)\\
=&\frac{1}{256}\,\mathrm{tr}\left\{\tilde{\mathbf{U}}_{r_n^*}^\intercal\left(\sum_{i^\dagger=2i-1}^{2i}\mathbf{\Sigma}_{1,i^\dagger}\right)\tilde{\mathbf{U}}_{r_n^*}\tilde{\mathbf{U}}_{r_n^*}^\intercal\left(\sum_{k^\dagger=2k-1}^{2k}\mathbf{\Sigma}_{2,k^\dagger}\right)\tilde{\mathbf{U}}_{r_n^*}\tilde{\mathbf{U}}_{r_n^*}^\intercal\left(\sum_{j^\dagger=2j-1}^{2j}\mathbf{\Sigma}_{1,j^\dagger}\right)\tilde{\mathbf{U}}_{r_n^*}\tilde{\mathbf{U}}_{r_n^*}^\intercal\left(\sum_{\ell^\dagger=2\ell-1}^{2\ell}\mathbf{\Sigma}_{2,\ell^\dagger}\right)\tilde{\mathbf{U}}_{r_n^*}\right\}.
\end{aligned}
$$

Then from Lemma S.3, we have

$$
\begin{aligned}
&\mathrm{E}\left(\sum_{i=1}^{m_1}\sum_{j=1}^{m_1}\sum_{k=1}^{m_2}\sum_{\ell=1}^{m_2} w_{i,k}^* w_{j,k}^* w_{i,\ell}^* w_{j,\ell}^*\mathbf{1}_{\{i\neq j\}}\mathbf{1}_{\{k\neq\ell\}}\right)\\
=&O\Bigg[n_2^4\,\mathrm{tr}\left\{\left(\tilde{\mathbf{U}}_{r_n^*}^\intercal\bar{\mathbf{\Sigma}}_1\tilde{\mathbf{U}}_{r_n^*}\right)^4\right\} + n_1^4\,\mathrm{tr}\left\{\left(\tilde{\mathbf{U}}_{r_n^*}^\intercal\bar{\mathbf{\Sigma}}_2\tilde{\mathbf{U}}_{r_n^*}\right)^4\right\}\\
&\quad + n_2^2\,\mathrm{tr}(\bar{\mathbf{\Sigma}}_2^2)\sum_{i=1}^{n_1}\mathrm{tr}(\mathbf{\Sigma}_{1,i}^2) + n_1^2\,\mathrm{tr}(\bar{\mathbf{\Sigma}}_1^2)\sum_{i=1}^{n_2}\mathrm{tr}(\mathbf{\Sigma}_{2,i}^2) + \left\{\sum_{i=1}^{n_1}\mathrm{tr}(\mathbf{\Sigma}_{1,i}^2)\right\}\left\{\sum_{i=1}^{n_2}\mathrm{tr}(\mathbf{\Sigma}_{2,i}^2)\right\}\Bigg]\\
=&O\left[n_1^4 n_2^4\left\{\lambda_{r_n^*+1}(\mathbf{\Psi}_n)\right\}^2\mathrm{tr}\left(\mathbf{\Psi}_n^2\right)\right] + o\left[n_1^4 n_2^4\left\{\mathrm{tr}(\mathbf{\Psi}_n^2)\right\}^2\right],
\end{aligned}
$$

where the last equality follows from (S.14) and Assumption 3. It follows from the above inequality and (S.32) that

$$
\mathrm{tr}\{(\mathbf{A}_{1,2}\mathbf{A}_{1,2}^\intercal)^2\} = O_P\left[\frac{\left\{\lambda_{r_n^*+1}(\mathbf{\Psi}_n)\right\}^2}{\sigma_{T,n}^2}\right] + o_P(1) = o_P(1).
$$

Thus, we have

$$
\mathrm{tr}(\mathbf{A}^4) = o_P(1). \tag{S.33}
$$

Now we deal with $\mathbf{B}_n\mathbf{B}_n^\intercal$. We have

$$
\mathbf{B}_n\mathbf{B}_n^\intercal = \sigma_{T,n}^{-1}\left(\frac{1}{m_1^2}\mathbf{U}_r^\intercal\tilde{\mathbf{X}}_1^\intercal\tilde{\mathbf{X}}_1\mathbf{U}_r + \frac{1}{m_2^2}\mathbf{U}_r^\intercal\tilde{\mathbf{X}}_2^\intercal\tilde{\mathbf{X}}_2\mathbf{U}_r\right).
$$

For $k=1,2$, we have

$$\mathrm{E}(\mathbf{U}_r^\intercal \tilde{\mathbf{X}}_k^\intercal \tilde{\mathbf{X}}_k \mathbf{U}_r) = \frac{1}{4}\sum_{i=1}^{m_k} \mathbf{U}_r^\intercal (\mathbf{\Sigma}_{k,2i-1} + \mathbf{\Sigma}_{k,2i})\mathbf{U}_r.$$

Consequently,

$$\begin{aligned}
\mathrm{E}\left\| \mathbf{U}_r^\intercal \tilde{\mathbf{X}}_k^\intercal \tilde{\mathbf{X}}_k \mathbf{U}_r - \frac{n_k}{4}\mathbf{U}_r^\intercal \bar{\mathbf{\Sigma}}_k \mathbf{U}_r \right\|_F^2 \leq & \sum_{i=1}^{m_k} \mathrm{E}\left\| \mathbf{U}_r^\intercal \tilde{X}_{k,i}\tilde{X}_{k,i}^\intercal \mathbf{U}_r - \frac{1}{4}\mathbf{U}_r^\intercal (\mathbf{\Sigma}_{k,2i-1} + \mathbf{\Sigma}_{k,2i})\mathbf{U}_r \right\|_F^2 \\
& + \frac{1}{16}\|\mathbf{U}_r^\intercal \mathbf{\Sigma}_{k,n_k}\mathbf{U}_r\|_F^2 \\
\leq & \sum_{i=1}^{m_k} \mathrm{E}\left\| \mathbf{U}_r^\intercal \tilde{X}_{k,i}\right\|^4 + \frac{1}{16}\operatorname{tr}(\mathbf{\Sigma}_{k,n_k}^2) \\
\leq & \tau r \sum_{i=1}^{n_k} \operatorname{tr}(\mathbf{\Sigma}_{k,i}^2) \\
= & o\left\{ n_k^2 \operatorname{tr}(\bar{\mathbf{\Sigma}}_k^2)\right\}.
\end{aligned}$$

where the third inequality follows from (S.27) and the last equality follows from Assumption 3. Thus,

$$\mathbf{U}_r^\intercal \tilde{\mathbf{X}}_k^\intercal \tilde{\mathbf{X}}_k \mathbf{U}_r = \frac{n_k}{4}\mathbf{U}_r^\intercal \bar{\mathbf{\Sigma}}_k \mathbf{U}_r + o_p\left[n_k \left\{\operatorname{tr}(\bar{\mathbf{\Sigma}}_k^2)\right\}^{1/2}\right].$$

It follows that

$$\mathbf{B}_n \mathbf{B}_n^\intercal = \sigma_{T,n}^{-1}\mathbf{U}_r^\intercal \mathbf{\Psi}_n \mathbf{U}_r + o_P(1) = 2^{-1/2}\operatorname{diag}(\kappa_1,\ldots,\kappa_r) + o_P(1). \tag{S.34}$$

Note that we have proved that (S.28), (S.33) and (S.34) hold in probability. Then for every subsequence of $\{n\}$, there is a further subsequence along which these three equalities hold almost surely, and consequently (S.26) holds almost surely by Lemma S.6. That is, (S.26) holds in probability. This completes the proof. □

Lemma S.13. *Suppose $\{\xi_i\}_{i=0}^\infty$ is a sequence of independent standard normal random variables and $\{\kappa_i\}_{i=1}^\infty$ is a sequence of positive numbers such that $\sum_{i=1}^\infty \kappa_i^2 \in [0,1]$. Then the cumulative distribution function $F(\cdot)$ of $(1-\sum_{i=1}^\infty \kappa_i^2)^{1/2}\xi_0 + 2^{-1/2}\sum_{i=1}^\infty \kappa_i(\xi_i^2-1)$ is continuous and strictly increasing on the interval $\{x\in\mathbb{R} : F(x)>0\}$.*

Proof. If $\kappa_i=0$, $i=1,2,\ldots$, then the conclusion holds since $(1-\sum_{i=1}^\infty \kappa_i^2)^{1/2}\xi_0 + 2^{-1/2}\sum_{i=1}^\infty \kappa_i(\xi_i^2-1)=\xi_0$ is a standard normal random variable. Otherwise, we can assume ithout loss of generality that $\kappa_1>0$. Let $\zeta = (1-\sum_{i=1}^\infty \kappa_i^2)^{1/2}\xi_0 - 2^{-1/2}\kappa_1 + 2^{-1/2}\sum_{i=2}^\infty \kappa_i(\xi_i^2-1)$. Then $(1-\sum_{i=1}^\infty \kappa_i^2)^{1/2}\xi_0 + 2^{-1/2}\sum_{i=1}^\infty \kappa_i(\xi_i^2-1) = 2^{-1/2}\kappa_1\xi_1^2+\zeta$. Let $f_1(\cdot)$ denote the probability density function of $2^{-1/2}\kappa_1\xi_1^2$. Let $F_\zeta(\cdot)$ denote the cumulative distribution function of ζ. Then $2^{-1/2}\kappa_1\xi_1^2+\zeta$ has density function

$$f(x) = \int_{-\infty}^{+\infty} f_1(x-t)\,\mathrm{d}F_\zeta(t).$$

As a result, $F(\cdot)$ is continuous.

Now we prove that $F(\cdot)$ is strictly increasing on the interval $\{x\in\mathbb{R} : F(x)>0\}$. Let c be a point in the support of $\mathcal{L}(\zeta)$. For any real number a such that $a>c$ and for any $\delta>0$,

$$\begin{aligned}
&\operatorname{pr}\left\{2^{-1/2}\kappa_1\xi_1^2+\zeta \in (a, a+\delta)\right\} \\
\geq & \operatorname{pr}\left\{2^{-1/2}\kappa_1\xi_1^2 \in (a-c+\delta/4, a-c+3\delta/4)\right\}\operatorname{pr}\left\{\zeta \in (c-\delta/4, c+\delta/4)\right\}
\end{aligned}$$

Since c is in the support of $\mathcal{L}(\zeta)$, we have $\mathrm{pr}\,(c - \delta/4 < \zeta < c + \delta/4) > 0$; see, e.g., Cohn (2013), Section 7.4. Thus, $\mathrm{pr}\,\left\{2^{-1/2}\kappa_1\xi_1^2 + \zeta \in (a, a+\delta)\right\} > 0$. Therefore, $F(\cdot)$ is strictly increasing on the interval $(c, +\infty)$. Then the conclusion follows from the fact that c is an arbitrary point in the support of $\mathcal{L}(\zeta)$. $\square$