

# DISTRIBUTION AND CORRELATION-FREE TWO-SAMPLE TEST OF HIGH-DIMENSIONAL MEANS

BY KAIJIE XUE<sup>1</sup> AND FANG YAO<sup>2</sup>

<sup>1</sup>School of Statistics and Data Science, Nankai University, [kaijie@nankai.edu.cn](mailto:kaijie@nankai.edu.cn)

<sup>2</sup>Department of Probability and Statistics, School of Mathematical Sciences, Center for Statistical Science, Peking University, [fyao@math.pku.edu.cn](mailto:fyao@math.pku.edu.cn)

We consider a two-sample test for high-dimensional means that can be either distributional or correlation-free, depending on the weak condition on the moment details of the elements and the spectral limit theorem (Ann. Probab. **45** (2017) 2309–2352) and a practically efficient procedure for the spectral gap test. It is a new type of a two-sample test. In this article, the spectral test is extended to the case where the elements are identically distributed random vectors, which are allowed to have different distributions and arbitrary correlations. Further, the new feature is that the weaker moment details than existing methods, allow a certain high-dimensional limit theorem behavior. The final generalization, data dimensionality is allowed to be exponentially high, the umbrella of the general condition. Simulated and real data examples have demonstrated favorable empirical performance.

**1. Introduction.** Two-sample test for high-dimensional means has attracted a great deal of attention in statistical applications, including [2, 5, 10, 12, 19, 24, 26, 29] and [21], among others. In this article, we tackle this problem with the theoretical advantage brought by a high-dimensional two-sample central limit theorem. Based on this, we consider a new type of testing procedure, called distributional or correlation-free (DCF) two-sample mean test, which can be either distributional or correlation-free and generally has competitive performance.

We denote two sample  $X^n = \{X_1, \dots, X_n\}$  and  $Y^m = \{Y_1, \dots, Y_m\}$  as vectors, where  $X^n$  is a collection of  $n$  independent (not necessarily identically distributed) random vectors in  $\mathbb{R}^p$ , with  $X_i = (X_{i1}, \dots, X_{ip})'$  and  $E(X_i) = \mu^X = (\mu_1^X, \dots, \mu_p^X)'$ ,  $i = 1, \dots, n$ , and  $Y^m$  is defined similarly with  $E(Y_i) = \mu^Y = (\mu_1^Y, \dots, \mu_p^Y)'$  for all  $i = 1, \dots, m$ . The normalized sums  $S_n^X$  and  $S_m^Y$  are defined by  $S_n^X = n^{-1/2} \sum_{i=1}^n X_i = (S_{n1}^X, \dots, S_{np}^X)'$  and  $S_m^Y = m^{-1/2} \sum_{i=1}^m Y_i = (S_{m1}^Y, \dots, S_{mp}^Y)'$ , respectively. Note that we allow the means to be arbitrary, and each sample has a common mean. The hypothesis testing is

$$H_0: \mu^X = \mu^Y \quad \dots \quad H_a: \mu^X \neq \mu^Y,$$

and the considered two-sample DCF mean test is such that we reject  $H_0: \mu^X = \mu^Y$  at significance level  $\alpha \in (0, 1)$ , provided that

$$T_n = \|S_n^X - n^{1/2}m^{-1/2}S_m^Y\|_\infty \geq c_B(\alpha),$$

where  $T_n = \|S_n^X - n^{1/2}m^{-1/2}S_m^Y\|_\infty$  is the test statistic that is defined as the maximum of the sample mean difference, and  $c_B(\alpha)$  that is a certain critical value of the test. It is worth noting that  $c_B(\alpha)$  is a

Received October 2018; revised January 2019.

MSC2010 subject classifications. 62H05, 62F05.

Key words and phrases. High-dimensional central limit theorem, Kolmogorov–Smirnov test, spectral test.

computationally tractable and efficiently distributed (i.i.d.) random variable that depends on the data, hence the explicit calculation is described after (6). Note that the complexity of the proposed method is  $O\{n(p+N)\}$ , more efficient than  $O(Nnp)$  that is required by the standard method. The implementation of  $T_n$  will illustrate the theoretical results and numerical efficiency of the proposed method.

We emphasize that the main contributions of the design are as follows. First, we prove that the proposed algorithm is efficient in terms of the number of iterations. Theorem 3.5. We begin with deriving the initial value of the parameter  $\alpha$  for the optimal limit theorem and discuss the digression of the maximum value of the high dimension [9], where we derive the relationship between the parameter  $n/(n+m)$  and the convergence rate of the algorithm. Then, we prove that for  $(c_1, c_2)$ ,  $0 < c_1 \leq c_2 < 1$ , as  $n, m \rightarrow \infty$ . Further, Theorem 3.6 and 3.7 provide the conditions for the DCF method. If  $\alpha$  and  $\beta$  are all  $\alpha \in (0, 1)$ . Theorem 3.8 of the proposed test is established. Theorem 4 that establishes the asymptotic efficiency of the proposed method. Moreover, the asymptotic efficiency of the proposed test is established. Theorem 5 demonstrates the relationship between the proposed test and the existing tests.

The  $\mathbf{c}$  is estimated itself as a fitting method by all  $\mathbf{v}_i$  i.g.f.c. -i.i.d. random vectors in both samples. The distribution-free feature is that, under the umbrella of the mild assumption that the meta data tail extends of the coordinate, there is the restriction that the distribution of the random vectors. In contrast, existing literature considers the random vectors in this sample to be i.i.d. [3–6], and the method for the restriction of the coordinate to full  $\mathbf{v}_i$  a certain type of distribution, such as Gaussian or b-Gaussian [26, 29]. This feature of the  $\mathbf{c}$  is estimated free from making an assumption that the  $\mathbf{c}$  is b-Gaussian, which is desirable as distributions for real data are flexible and do not necessarily follow a Gaussian distribution. Another key feature is correlation-free in the sense that it does not depend on the meta data tail extending of the coordinate. Because, in this case, the  $\mathbf{c}$  is a common to all  $\mathbf{v}_i$  in this sample correlation matrix, but all meta data coordinates, such as the trace [5], mixing coordinates [21] or bounded eigenvalue problem [3]. It is worth noting that as an assumption that the meta data tail extends of the coordinate in random vectors are all weaker than the existing literature, for example, [3, 11] and [21] assumed a common fixed or bounded norm meta, [5] and [19] all used a statistic of the meta data to give a bound on the correlation assumption.

We also note that the considered test is sensitive to the behavior of the far left tail of the alternative (a mild deviation is bounded  $\mu^X - \mu^Y$  in Theorem 5), with either a significant correlation dimension, while the  $\chi^2$  test is sensitive to the local behavior of the distribution [26], or a significant correlation [21], as both [3]. Last, we note that the data dimension can be extremely high relative to the sample size of the mixture of each mild alternative. This is all favorable compared to the  $\chi^2$  test, as [3, 5] and [21] all need high dimension for detection of local deviations either the distribution (e.g., b-Gaussian) or the correlation structure (as both) as a trade off.

We conclude the Introduction by giving a table of the main results of the paper. The table is organized as follows. The first column contains the title of the section. The second column contains the title of the subsection. The third column contains the title of the theorem. The fourth column contains the title of the lemma. The fifth column contains the title of the proposition. The sixth column contains the title of the corollary. The seventh column contains the title of the definition. The eighth column contains the title of the example. The ninth column contains the title of the remark. The tenth column contains the title of the note. The eleventh column contains the title of the reference. The twelfth column contains the title of the appendix. The thirteenth column contains the title of the bibliography. The fourteenth column contains the title of the index. The fifteenth column contains the title of the glossary. The sixteenth column contains the title of the list of figures. The seventeenth column contains the title of the list of tables. The eighteenth column contains the title of the list of equations. The nineteenth column contains the title of the list of symbols. The twentieth column contains the title of the list of abbreviations. The twenty-first column contains the title of the list of acronyms. The twenty-second column contains the title of the list of initialisms. The twenty-third column contains the title of the list of contractions. The twenty-fourth column contains the title of the list of colloquialisms. The twenty-fifth column contains the title of the list of slang. The twenty-sixth column contains the title of the list of idioms. The twenty-seventh column contains the title of the list of proverbs. The twenty-eighth column contains the title of the list of sayings. The twenty-ninth column contains the title of the list of maxims. The thirtieth column contains the title of the list of aphorisms. The thirty-first column contains the title of the list of epigrams. The thirty-second column contains the title of the list of epigrams. The thirty-third column contains the title of the list of epigrams. The thirty-fourth column contains the title of the list of epigrams. The thirty-fifth column contains the title of the list of epigrams. The thirty-sixth column contains the title of the list of epigrams. The thirty-seventh column contains the title of the list of epigrams. The thirty-eighth column contains the title of the list of epigrams. The thirty-ninth column contains the title of the list of epigrams. The fortieth column contains the title of the list of epigrams. The forty-first column contains the title of the list of epigrams. The forty-second column contains the title of the list of epigrams. The forty-third column contains the title of the list of epigrams. The forty-fourth column contains the title of the list of epigrams. The forty-fifth column contains the title of the list of epigrams. The forty-sixth column contains the title of the list of epigrams. The forty-seventh column contains the title of the list of epigrams. The forty-eighth column contains the title of the list of epigrams. The forty-ninth column contains the title of the list of epigrams. The fiftieth column contains the title of the list of epigrams. The fifty-first column contains the title of the list of epigrams. The fifty-second column contains the title of the list of epigrams. The fifty-third column contains the title of the list of epigrams. The fifty-fourth column contains the title of the list of epigrams. The fifty-fifth column contains the title of the list of epigrams. The fifty-sixth column contains the title of the list of epigrams. The fifty-seventh column contains the title of the list of epigrams. The fifty-eighth column contains the title of the list of epigrams. The fifty-ninth column contains the title of the list of epigrams. The sixtieth column contains the title of the list of epigrams. The sixty-first column contains the title of the list of epigrams. The sixty-second column contains the title of the list of epigrams. The sixty-third column contains the title of the list of epigrams. The sixty-fourth column contains the title of the list of epigrams. The sixty-fifth column contains the title of the list of epigrams. The sixty-sixth column contains the title of the list of epigrams. The sixty-seventh column contains the title of the list of epigrams. The sixty-eighth column contains the title of the list of epigrams. The sixty-ninth column contains the title of the list of epigrams. The seventieth column contains the title of the list of epigrams. The seventy-first column contains the title of the list of epigrams. The seventy-second column contains the title of the list of epigrams. The seventy-third column contains the title of the list of epigrams. The seventy-fourth column contains the title of the list of epigrams. The seventy-fifth column contains the title of the list of epigrams. The seventy-sixth column contains the title of the list of epigrams. The seventy-seventh column contains the title of the list of epigrams. The seventy-eighth column contains the title of the list of epigrams. The seventy-ninth column contains the title of the list of epigrams. The eightieth column contains the title of the list of epigrams. The eighty-first column contains the title of the list of epigrams. The eighty-second column contains the title of the list of epigrams. The eighty-third column contains the title of the list of epigrams. The eighty-fourth column contains the title of the list of epigrams. The eighty-fifth column contains the title of the list of epigrams. The eighty-sixth column contains the title of the list of epigrams. The eighty-seventh column contains the title of the list of epigrams. The eighty-eighth column contains the title of the list of epigrams. The eighty-ninth column contains the title of the list of epigrams. The ninetieth column contains the title of the list of epigrams. The ninety-first column contains the title of the list of epigrams. The ninety-second column contains the title of the list of epigrams. The ninety-third column contains the title of the list of epigrams. The ninety-fourth column contains the title of the list of epigrams. The ninety-fifth column contains the title of the list of epigrams. The ninety-sixth column contains the title of the list of epigrams. The ninety-seventh column contains the title of the list of epigrams. The ninety-eighth column contains the title of the list of epigrams. The ninety-ninth column contains the title of the list of epigrams. The hundredth column contains the title of the list of epigrams.

In this Section, we first give the definition of the  $\mathbb{R}^p$ -valued random variable  $X$  and the  $\mathbb{R}^p$ -valued random variable  $Y$ . We collect the auxiliary lemma and the  $\mathbb{R}^p$ -valued random variable  $X$  and the  $\mathbb{R}^p$ -valued random variable  $Y$ . Theorem 3.5 in the Appendix, and delegate the  $\mathbb{R}^p$ -valued random variable  $X$  and the  $\mathbb{R}^p$ -valued random variable  $Y$  to the Appendix. Theorem 1.2, Corollary 1 and the auxiliary lemma to the Appendix. Material [22] for the  $\mathbb{R}^p$ -valued random variable  $X$  and the  $\mathbb{R}^p$ -valued random variable  $Y$ .

## 2. Two-sample central limit theorem and multiplier bootstrap in high dimensions.

In this section, we first give the definition of the  $\mathbb{R}^p$ -valued random variable  $X$  and the  $\mathbb{R}^p$ -valued random variable  $Y$ . We collect the auxiliary lemma and the  $\mathbb{R}^p$ -valued random variable  $X$  and the  $\mathbb{R}^p$ -valued random variable  $Y$ . Theorem 3.5 in the Appendix, and delegate the  $\mathbb{R}^p$ -valued random variable  $X$  and the  $\mathbb{R}^p$ -valued random variable  $Y$  to the Appendix. Theorem 1.2, Corollary 1 and the auxiliary lemma to the Appendix. Material [22] for the  $\mathbb{R}^p$ -valued random variable  $X$  and the  $\mathbb{R}^p$ -valued random variable  $Y$ .

We first list the main results. Let  $X = (X_1, \dots, X_p) \in \mathbb{R}^p$  and  $Y = (Y_1, \dots, Y_p) \in \mathbb{R}^p$ , we say  $X \leq Y$  if  $X_j \leq Y_j$  for all  $j = 1, \dots, p$ . For  $a = (a_1, \dots, a_p) \in \mathbb{R}^p$  and  $b = (b_1, \dots, b_p) \in \mathbb{R}^p$ , we say  $a \vee b = \max\{a, b\}$  and  $a \wedge b = \min\{a, b\}$ . For  $a = (a_1, \dots, a_p) \in \mathbb{R}^p$  and  $b = (b_1, \dots, b_p) \in \mathbb{R}^p$ , we say  $a \lesssim b$  if  $a_n \leq Cb_n$  for all  $n$  and  $C > 0$ , and  $a_n \sim b_n$  if  $a_n \lesssim b_n$  and  $b_n \lesssim a_n$ . For a matrix  $A = (a_{ij})$ , we say  $\|A\|_\infty = \max_{i,j} |a_{ij}|$ . For a function  $f: \mathbb{R} \rightarrow \mathbb{R}$ , we say  $\|f\|_\infty = \sup_{z \in \mathbb{R}} |f(z)|$ . For a function  $g: \mathbb{R}^p \rightarrow \mathbb{R}$ , we say  $\partial_j \partial_k \partial_l g = g_{jkl}$ . For  $\alpha > 0$ , we say the function  $\psi_\alpha(x) = \exp(-x^\alpha)$  for  $x \in [0, \infty)$ , the function  $\psi_\alpha$  is a random variable  $X$ , we say

$$(1) \quad \|X\|_{\psi_\alpha} = \inf\{\lambda > 0 : E\{\psi_\alpha(|X|/\lambda)\} \leq 1\},$$

which is a Orlicz norm for  $\alpha \in [1, \infty)$  and a  $\psi_\alpha$ -norm for  $\alpha \in (0, 1)$ .

Let  $F^n = \{F_1, \dots, F_n\}$  and  $G^m = \{G_1, \dots, G_m\}$  be  $\mathbb{R}^p$ -valued random variables such that  $F_i = (F_{i1}, \dots, F_{ip})'$  and  $F_i \sim N_p(\mu^X, E\{(X_i - \mu^X)(X_i - \mu^X)'\})$  for all  $i = 1, \dots, n$ , which denote a Gaussian random variable  $X^n$ . Likewise, let  $G_j = (G_{j1}, \dots, G_{jp})'$  and  $G_j \sim N_p(\mu^Y, E\{(Y_j - \mu^Y)(Y_j - \mu^Y)'\})$  for all  $j = 1, \dots, m$  denote a Gaussian random variable  $Y^m$ . The  $X^n$ ,  $Y^m$ ,  $F^n$  and  $G^m$  are assumed to be independent of each other. Then, we denote the  $\mathbb{R}^p$ -valued random variables  $S_n^X, S_n^F, S_m^Y$  and  $S_m^G$  by  $S_n^X = n^{-1/2} \sum_{i=1}^n X_i = (S_{n1}^X, \dots, S_{np}^X)'$ ,  $S_n^F = n^{-1/2} \sum_{i=1}^n F_i = (S_{n1}^F, \dots, S_{np}^F)'$ ,  $S_m^Y = m^{-1/2} \sum_{j=1}^m Y_j = (S_{m1}^Y, \dots, S_{mp}^Y)'$  and  $S_m^G = m^{-1/2} \sum_{j=1}^m G_j = (S_{m1}^G, \dots, S_{mp}^G)'$ , where  $S_n^X$  and  $S_m^G$  are Gaussian random variables for  $S_n^X$  and  $S_m^Y$ , respectively. Last, denote a  $\mathbb{R}^p$ -valued random variable  $e^{n+m} = \{e_1, \dots, e_{n+m}\}$  that is independent of  $X^n, F^n, Y^m$  and  $G^m$ .

**2.1. Two-sample central limit theorem in high dimensions.** In this section, we first give the definition of the  $\mathbb{R}^p$ -valued random variable  $X$  and the  $\mathbb{R}^p$ -valued random variable  $Y$ . We collect the auxiliary lemma and the  $\mathbb{R}^p$ -valued random variable  $X$  and the  $\mathbb{R}^p$ -valued random variable  $Y$ . Theorem 3.5 in the Appendix, and delegate the  $\mathbb{R}^p$ -valued random variable  $X$  and the  $\mathbb{R}^p$ -valued random variable  $Y$  to the Appendix. Theorem 1.2, Corollary 1 and the auxiliary lemma to the Appendix. Material [22] for the  $\mathbb{R}^p$ -valued random variable  $X$  and the  $\mathbb{R}^p$ -valued random variable  $Y$ .

$$L_n^X = \max_{1 \leq j \leq p} \sum_{i=1}^n E(|X_{ij} - \mu_j^X|^3)/n, \quad L_m^Y = \max_{1 \leq j \leq p} \sum_{i=1}^m E(|Y_{ij} - \mu_j^Y|^3)/m.$$

We denote the kernel  $\rho_{n,m}^{**}$  by

$$(2) \quad \rho_{n,m}^{**} = \int_{A \in \mathcal{A}^{\mathbb{R}^p}} |P(S_n^X - n^{1/2} \mu^X + \delta_{n,m} S_m^Y - \delta_{n,m} m^{1/2} \mu^Y \in A) - P(S_n^F - n^{1/2} \mu^X + \delta_{n,m} S_m^G - \delta_{n,m} m^{1/2} \mu^Y \in A)|,$$

where  $P(S_n^X - n^{1/2} \mu^X + \delta_{n,m} S_m^Y - \delta_{n,m} m^{1/2} \mu^Y \in A)$  is the probability of the event  $A$  for  $S_n^X - n^{1/2} \mu^X + \delta_{n,m} S_m^Y - \delta_{n,m} m^{1/2} \mu^Y \in A$  and  $P(S_n^F - n^{1/2} \mu^X + \delta_{n,m} S_m^G - \delta_{n,m} m^{1/2} \mu^Y \in A)$  is the probability of the event  $A$  for  $S_n^F - n^{1/2} \mu^X + \delta_{n,m} S_m^G - \delta_{n,m} m^{1/2} \mu^Y \in A$ . We say  $\rho_{n,m}^{**}$  measures the difference between the two distributions.

hyperrectangle  $A \in \mathcal{A}^{\text{Re}}$ . Note that  $\mathcal{A}^{\text{Re}}$  is the class of all hyperrectangle in  $\mathbb{R}^p$  of the form  $\{w \in \mathbb{R}^p : a_j \leq w_j \leq b_j \text{ for all } j = 1, \dots, p\}$  with  $-\infty \leq a_j \leq b_j \leq \infty$  for all  $j = 1, \dots, p$ . By a mixing measure condition, Theorem 1 gives a more explicit bound  $\rho_{n,m}^{**}$  compared to Lemma 4.

**THEOREM 1.** *For any sequence of constants  $\delta_{n,m}$ , assume we have the following conditions (a)–(e):*

- (a) *There exist universal constants  $\delta_1 > \delta_2 > 0$  such that  $\delta_2 < |\delta_{n,m}| < \delta_1$ .*  
 (b) *There exists a universal constant  $b > 0$  such that*

$$\min_{1 \leq j \leq p} E\{(S_{nj}^X - n^{1/2}\mu_j^X + \delta_{n,m}S_{mj}^Y - \delta_{n,m}n^{1/2}\mu_j^Y)^2\} \geq b.$$

- (c) *There exists a sequence of constants  $B_{n,m} \geq 1$  such that  $L_n^X \leq B_{n,m}$  and  $L_m^Y \leq B_{n,m}$ .*  
 (d) *The sequence of constants  $B_{n,m}$  defined in (c) also satisfies*

$$\max_{1 \leq i \leq n} \max_{1 \leq j \leq p} E\{\alpha(|X_{ij} - \mu_j^X|/B_{n,m})\} \leq 2,$$

$$\max_{1 \leq i \leq m} \max_{1 \leq j \leq p} E\{\alpha(|Y_{ij} - \mu_j^Y|/B_{n,m})\} \leq 2.$$

- (e) *There exists a universal constant  $c_1 > 0$  such that*

$$(B_{n,m})^2\{1 - g(pn)\}^7/n \leq c_1, \quad (B_{n,m})^2\{1 - g(pm)\}^7/m \leq c_1.$$

Then we have the following property, where  $\rho_{n,m}^{**}$  is defined in (2):

$$\rho_{n,m}^{**} \leq K_3[(B_{n,m})^2\{1 - g(pn)\}^7/n]^{1/6} + [(B_{n,m})^2\{1 - g(pm)\}^7/m]^{1/6},$$

for a universal constant  $K_3 > 0$ .

Condition (a) (c) controls the moment structure of the coordinate, and (d) controls the tail structure. If all of (a) and (b) that the moment on average are bounded below, a famous theorem, hence all interesting statistics of the moment test converge. This is weaker than the classical central limit theorem, if sample size is bounded all moment [3, 11, 21]. Condition (c) implies that the moment on average has a bound  $B_{n,m}$  that can diverge if it is not too quickly correlated, then the more flexibility that the literature that demand either a fixed or bounded a certain correlation to converge both. To appreciate this, letting  $B_{n,m} \sim n^{1/3}$ , we note that all the variance of the coordinate are all bounded if sample size  $B_{n,m}^{2/3} \sim n^{2/9} \rightarrow \infty$  due to condition (c), while the correlation is needed. As a consequence, if we have a common covariance matrix, and  $\Sigma = (\Sigma_{jk})_{1 \leq j,k \leq p}$  with each  $\Sigma_{jk} = n^{2/9}\rho^{1\{j \neq k\}}$  for some constant  $\rho \in (0, 1)$ , then the trace condition in [5] implies that  $p = o(1)$ . Compared with a fixed or bounded the tail of the coordinate [3, 21], condition (d) allows for if sample size goes to infinity  $B_{n,m} \rightarrow \infty$ . Condition (e) indicates that the data dimension  $p$  can grow exponentially in  $n$ , provided that  $B_{n,m}$  is finite and stable. The condition allows the behavior of the so-called distributional dependence feature.

**2.2. Two-sample multiplier bootstrap in high dimensions.** Denote the  $k$ th sample is  $\rho_{n,m}^{**}$  (2) denoting the Gaussian assumption, it limits the applicability of the central limit theorem for inference. The idea is to adapt a multivariate bootstrap to approximate the Gaussian assumption, and a sufficient assumption is established. Denote

$$\Sigma^X = n^{-1} \sum_{i=1}^n E\{(\text{838 (X53(Denote))T7 Tc 9.8629 0 0 9.8629 149.922 44.058.3ng})$$

where  $\bar{X} = n^{-1} \sum_{i=1}^n X_i = (\bar{X}_1, \dots, \bar{X}_p)'$ . Analogously, denote  $\Sigma^Y$ ,  $\hat{\Sigma}^Y$  and  $\bar{Y}$ . Note that the multipliers  $\{e_1, \dots, e_{n+m}\}$  are i.i.d. standard normal random variables independent of the data, and the

$$(3) \quad S_n^{eX} = n^{-1/2} \sum_{i=1}^n e_i (X_i - \bar{X}), \quad S_m^{eY} = m^{-1/2} \sum_{i=1}^m e_{i+n} (Y_i - \bar{Y}),$$

and it is clear that  $E_e(S_n^{eX} S_n^{eX'}) = \hat{\Sigma}^X$  and  $E_e(S_n^{eY} S_n^{eY'}) = \hat{\Sigma}^Y$ , where  $E_e(\cdot)$  means the expectation with respect to  $e^{n+m}$ . Therefore, we define  $\delta_{n,m}$  that depends on  $n$  and  $m$ , and denote the following probability

$$(4) \quad \rho_{n,m}^{MB} = \inf_{A \in \mathcal{A}^{\text{Re}}} |P_e(S_n^{eX} + \delta_{n,m} S_m^{eY} \in A) - P(S_n^F - n^{1/2} \mu^X + \delta_{n,m} S_m^G - \delta_{n,m} m^{1/2} \mu^Y \in A)|,$$

where  $P_e(\cdot)$  means the probability with respect to  $e^{n+m}$ , and  $P_e(S_n^{eX} + \delta_{n,m} S_m^{eY} \in A)$  and  $P(S_n^F - n^{1/2} \mu^X + \delta_{n,m} S_m^G - \delta_{n,m} m^{1/2} \mu^Y \in A)$ . In addition,  $\rho_{n,m}^{MB}$  can be defined as a measure of the difference between the two probabilities for all hypotheses  $A \in \mathcal{A}^{\text{Re}}$ . The following lemma is the main result of this paper.

**THEOREM 2.** *For any sequence of constants  $\delta_{n,m}$ , assume we have the following conditions (a)–(e),*

- (a) *There exists a universal constant  $\delta_1 > 0$  such that  $|\delta_{n,m}| < \delta_1$ .*
- (b) *There exists a universal constant  $b > 0$  such that*

$$\inf_{1 \leq j \leq p} E\{(S_{nj}^X - n^{1/2} \mu_j^X + \delta_{n,m} S_{mj}^Y - \delta_{n,m} m^{1/2} \mu_j^Y)^2\} \geq b.$$

- (c) *There exists a sequence of constants  $B_{n,m} \geq 1$  such that*

$$\max_{1 \leq j \leq p} \sum_{i=1}^n E\{(X_{ij} - \mu_j^X)^4\}/n \leq B_{n,m}^2,$$

$$\max_{1 \leq j \leq p} \sum_{i=1}^m E\{(Y_{ij} - \mu_j^Y)^4\}/m \leq B_{n,m}^2.$$

- (d) *The sequence of constants  $B_{n,m}$  defined in (c) also satisfies*

$$\max_{1 \leq i \leq n} \max_{1 \leq j \leq p} E\{\alpha \cdot (|X_{ij} - \mu_j^X|/B_{n,m})\} \leq 2,$$

$$\max_{1 \leq i \leq m} \max_{1 \leq j \leq p} E\{\alpha \cdot (|Y_{ij} - \mu_j^Y|/B_{n,m})\} \leq 2.$$

- (e) *There exists a sequence of constants  $\alpha_{n,m} \in (0, e^{-1})$  such that*

$$B_{n,m}^2 \leq g^5(pn) \leq g^2(1/\alpha_{n,m})/n \leq 1,$$

$$B_{n,m}^2 \leq g^5(pm) \leq g^2(1/\alpha_{n,m})/m \leq 1.$$

Then there exists a universal constant  $c^* > 0$  such that with probability at least  $1 - \gamma_{n,m}$  where

$$\begin{aligned} \gamma_{n,m} = & (\alpha_{n,m})^{1/g(pn)/3} + 3(\alpha_{n,m})^{1/g^{1/2}(pn)/c^*} + (\alpha_{n,m})^{1/g(pm)/3} \\ & + 3(\alpha_{n,m})^{1/g^{1/2}(pm)/c^*} + (\alpha_{n,m})^{1/g^3(pn)/6} + 3(\alpha_{n,m})^{1/g^3(pn)/c^*} \\ & + (\alpha_{n,m})^{1/g^3(pm)/6} + 3(\alpha_{n,m})^{1/g^3(pm)/c^*}, \end{aligned}$$

we have the following property, where  $\rho_{n,m}^{MB}$  is defined in (4),

$$\rho_{n,m}^{MB} \lesssim \{B_{n,m}^2 \log^5(pn) \log^2(1/\alpha_{n,m})/n\}^{1/6} \\ + \{B_{n,m}^2 \log^5(pm) \log^2(1/\alpha_{n,m})/m\}^{1/6}.$$

Condition (a)–(c) state the moment conditions for the coordinate, condition (d) characterizes the tail condition and condition (e) characterizes the decay of  $p$ . The condition has the desirable feature that the Theorem 1, which allows for a general distribution of the data and the moment conditions. Moreover, by combining Theorem 2 with the Bernstein inequality (i.e., Lemma 6), we can deduce Corollary 1 below, which facilitates the derivation of the main result in Theorem 3.

**COROLLARY 1.** *For any sequence of constants  $\delta_{n,m}$ , assume the conditions (a)–(e) in Theorem 2 hold. Also suppose that the condition (f) holds as follows:*

(f) *The sequence of constants  $\gamma_{n,m}$  defined in Theorem 2 also satisfies*

$$\sum_n \sum_m \gamma_{n,m} < \infty.$$

*Then with probability one, we have the following property, where  $\rho_{n,m}^{MB}$  is defined in (4),*

$$\rho_{n,m}^{MB} \lesssim \{B_{n,m}^2 \log^5(pn) \log^2(1/\alpha_{n,m})/n\}^{1/6} \\ + \{B_{n,m}^2 \log^5(pm) \log^2(1/\alpha_{n,m})/m\}^{1/6}.$$

**3. Two-sample mean test in high dimensions.** In this section, based on the theoretical results from the preceding section, we state the main result, Theorem 3, which gives a consistent test for the mean difference  $(\mu^X - \mu^Y)$  and, ideally, the DCF test is considered. We state that the theoretical guarantee holds for all  $\alpha \in (0, 1)$  with probability one.

**THEOREM 3.** *Assume we have the following conditions (a)–(e):*

- (a)  $n/(n+m) \in (c_1, c_2)$ , for some universal constants  $0 < c_1 < c_2 < 1$ .  
 (b) *There exists a universal constant  $b > 0$  such that*

$$\min_{1 \leq j \leq p} [E\{(S_{nj}^X - n^{1/2}\mu_j^X)^2\} + E\{(S_{mj}^Y - m^{1/2}\mu_j^Y)^2\}] \geq b.$$

- (c) *There exists a sequence of constants  $B_{n,m} \geq 1$  such that*

$$\max_{1 \leq j \leq p} \sum_{i=1}^n E(|X_{ij} - \mu_j^X|^{k+2})/n \leq B_{n,m}^k, \\ \max_{1 \leq j \leq p} \sum_{i=1}^m E(|Y_{ij} - \mu_j^Y|^{k+2})/m \leq B_{n,m}^k,$$

for all  $k = 1, 2$ .

- (d) *The sequence of constants  $B_{n,m}$  defined in (c) also satisfies*

$$\max_{1 \leq i \leq n} \max_{1 \leq j \leq p} E\{\exp(-|X_{ij} - \mu_j^X|/B_{n,m})\} \leq 2, \\ \max_{1 \leq i \leq m} \max_{1 \leq j \leq p} E\{\exp(-|Y_{ij} - \mu_j^Y|/B_{n,m})\} \leq 2.$$

- (e)  $B_{n,m}^2 \log^7(pn)/n \rightarrow 0$  as  $n \rightarrow \infty$ .

Then with probability one, the Kolmogorov distance between the distributions of the quantity  $\|S_n^X - n^{1/2}m^{-1/2}S_m^Y - n^{1/2}(\mu^X - \mu^Y)\|_\infty$  and the quantity  $\|S_n^{eX} - n^{1/2}m^{-1/2}S_m^{eY}\|_\infty$  satisfies

$$\begin{aligned} & \lim_{t \geq 0} \left| P(\|S_n^X - n^{1/2}m^{-1/2}S_m^Y - n^{1/2}(\mu^X - \mu^Y)\|_\infty \leq t) \right. \\ & \quad \left. - P_e(\|S_n^{eX} - n^{1/2}m^{-1/2}S_m^{eY}\| \right. \end{aligned}$$

It is easy to see that the complexity of the DCF test is of the order  $O\{n(p+N)\}$ , compared with  $O(Nnp)$  that is required by all demanded bootstrap sampling methods.

According to (6), the test  $P_{\mathbb{F}_n}(\mu^X - \mu^Y)$  is effective for the test case before mentioned as

$$(7) \quad P_{\mathbb{F}_n}(\mu^X - \mu^Y) = P\{\|S_n^X - n^{1/2}m^{-1/2}S_m^Y\|_{\infty} \geq c_B(\alpha) \mid \mu^X - \mu^Y\}.$$

Therefore, if the test  $P_{\mathbb{F}_n}(\mu^X - \mu^Y)$  of the DCF test, the expression (7) is directly applicable to the distribution of  $(S_n^X - n^{1/2}m^{-1/2}S_m^Y)$  in  $\mathbb{R}^p$ . Motivated by Theorem 3, we consider a theoretical lower bound for a simulation of  $P_{\mathbb{F}_n}(\mu^X - \mu^Y)$ , based on a different method to add random variables  $e^{*n+m} = \{e_1^*, \dots, e_{n+m}^*\}$  independent of  $e^{n+m}$  that are used to calculate  $c_B(\alpha)$ ,

$$(8) \quad \begin{aligned} P_{\mathbb{F}_n}(\mu^X - \mu^Y) &= P_{e^*}\{\|S_n^{e^*X} - n^{1/2}m^{-1/2}S_m^{e^*Y} + n^{1/2}(\mu^X - \mu^Y)\|_{\infty} \geq c_B(\alpha)\}, \end{aligned}$$

where  $S_n^{e^*X}$  and  $S_m^{e^*Y}$  are added in (3) with  $e^{*n+m}$  instead of  $e^{n+m}$ , and  $P_{e^*}(\cdot)$  means the probability with respect to  $e^{*n+m}$ . The following theorem indicates the asymptotic behavior of  $P_{\mathbb{F}_n}(\mu^X - \mu^Y)$  and  $P_{\mathbb{F}_n}(\mu^X - \mu^Y)$  under the same condition as in Theorem 3.

**THEOREM 4.** Assume the conditions (a)–(e) in Theorem 3 hold, then for any  $\mu^X - \mu^Y \in \mathbb{R}^p$ , we have with probability one,

$$|P_{\mathbb{F}_n}(\mu^X - \mu^Y) - P_{\mathbb{F}_n}(\mu^X - \mu^Y)| \lesssim \{B_{n,m}^2 \log(pn)/n\}^{1/6}.$$

Based on the condition in Theorem 4, it is worth mentioning that either a strict correlation restriction is required, as mentioned in [3], or a weaker condition is sufficient. To appreciate this, the asymptotic behavior of the test statistic is considered by condition (f) in addition to the theorem below.

**THEOREM 5.** Assume the conditions (a)–(e) in Theorem 3 and that

(f)  $\mathcal{F}_{n,m,p} = \{\mu^X \in \mathbb{R}^p, \mu^Y \in \mathbb{R}^p : \|\mu^X - \mu^Y\|_{\infty} \geq K_s \{B_{n,m} \log(pn)/n\}^{1/2}\}$ , for a sufficiently large universal constant  $K_s > 0$ .

Then for any  $\mu^X - \mu^Y \in \mathcal{F}_{n,m,p}$ , we have with probability tending to one,

$$P_{\mathbb{F}_n}(\mu^X - \mu^Y) \rightarrow 1 \quad \text{as } n \rightarrow \infty.$$

The set  $\mathcal{F}_{n,m,p}$  in (f) implies a lower bound on the distance between  $\mu^X$  and  $\mu^Y$ , which is comparable to the asymptotic  $\max_i |\delta_i / \sigma_{i,i}^{1/2}| \geq \{2\beta \log(p)/n\}^{1/2}$  in Theorem 2 in [3]. The latter is in fact a special case of condition (f) where the eigenvalue  $B_{n,m}$  is constant. It is worth mentioning that the asymptotic behavior of the test statistic depends on either a strict correlation assumption or the context of the theorem. In contrast, Theorem 2 in [3] requires a lower bound on the eigenvalue of the covariance matrix  $\text{diag}(\Sigma)^{-1/2} \Sigma \text{diag}(\Sigma)^{-1/2}$  in  $\mathbb{R}^p$  and is based on a different asymptotic result. The essential reason is that the DCF is effective for a broader range of alternatives. We conclude this by noting that the test of the DCF test is based on  $L_2$ -norms and can be fitted to both the test statistic, which is defined as follows.



mal i ai i  $\check{X}_i$  a d  $\check{Y}_i'$  b i de e de t a d hea -tailed i ai  $(5/3)^{-1/2}t(5)$   
i h mea e a d i i a i a ce , e f e r e d i a c m l e t e l e l a e d a d hea -tailed e t  
i g. - The i t h e t t i g i a l a a l g . t t h e t h i r d , h i l e i d e e d e t a d k e d i a t i  
t i  $8^{-1/2}\{\chi^2(4) - 4\}$  i h mea e a d i i a i a ce e e d , d e t e d b c m l e t e l  
e l a e d a d k e d e t t i g. -

We conduct the following tests and calculate the rejection rates for the empirical  $\alpha$  at different significance levels  $\delta$  and  $\beta$  in each setting as described above, based on 1000 Monte Carlo simulations. The empirical rejection rates are shown in Table 12. For illustration, we depict the empirical  $\alpha$  for all settings in Figure 1. We also display the multiplier bootstrap  $\alpha$  estimators based on the identified effect size  $N = 10^4$ , which agree well with the empirical  $\alpha$  of the DCF test adjusted for the theoretical asymptotics. Theorem 4. We see that the empirical effect size DCF test agrees well with the nominal level 0.05 in all settings. By comparison, the CQ test is a liberal test, and the CL and XL tests have overestimated the effect size in all settings.

Regarding  $\alpha$  &  $\beta$ : since  $\delta = 0.1$  and  $\delta = 0.15$ , the DCF test still dominates the fixed-growth test at all levels of  $\beta$ . When the growth rate is  $\delta = 0.2$ , there is still a setting I indicate that the DCF test is better than the fixed-growth test, except for the CQ test where  $\beta \geq 80\%$  (and decrease as  $\alpha$ ). Although the  $\alpha$  &  $\beta$  CQ test increases about half of DCF test at  $\beta = 80\%$ , the gain is relatively small. The best test has high  $\alpha$ . Similar patterns are observed in Settings II, III, V, VI, with  $\delta = 0.25$  for  $\beta$ ; a gain between  $80\%$  and  $83\%$ , and Setting III, IV, with  $\delta = 0.3$  for  $\beta$  at  $80\%$  and  $90\%$ , respectively. This means that in all cases, the  $\alpha$  &  $\beta$  test in Figure 1. It also tested the DCF test dominates the CL ( $L_\infty$ ) and XL (combined) tests if  $\alpha$  is high enough and all levels of  $\delta$  and  $\beta$ . To summarize, except for the standard increase of  $\alpha$  &  $\beta$  CQ test or decrease as  $\alpha$ , the DCF test is better than the fixed-growth test across all levels of  $\delta$  and  $\beta$ . As it levels off,  $\beta$  for alternative with varied magnitude and sign. Moreover, the gain is stable in the situation that the data become more complex, for example, highly balanced or heavy-tailed skewed distribution.

We further examine alternative with common fixed effects, see, e.g., [10], and the common level related setting, defined by Setting VII, here we let  $\mu^Y = \delta(1, \dots, 1_{[\beta p]}, 0'_{p-[\beta p]})'$ . Note that the empirical size of the test is 5(11)-26621(18360)

TABLE I  
Rejection proportions (%) calculated for four testing methods at different signal strength levels of  $\delta$  and sparsity levels of  $\beta$  based on 1000 Monte Carlo runs, where  $\beta = 0$  corresponds to the null hypothesis  $\beta = 1$  to the fully dense alternative, and  $(n, m, p) = (200, 300, 1000)$

Setti g I: i.i.d. e . al c																				
Te t	$\delta = 0.1$				$\delta = 0.15$				$\delta = 0.2$				$\delta = 0.25$				$\delta = 0.3$			
	DCF	CL	XL	CQ	DCF	CL	XL	CQ	DCF	CL	XL	CQ	DCF	CL	XL	CQ	DCF	CL	XL	CQ
$\beta = 0$	4.20	2.40	3.90	5.80	4.30	2.30	2.40	3.60	4.50	2.80	3.70	6.00	4.60	2.70	2.20	3.80	5.00	3.10	3.80	6.10
$\beta = 0.02$	5.00	3.20	2.50	3.40	7.50	4.80	3.70	3.50	15.4	10.5	6.50	3.90	31.7	23.3	14.6	4.40	59.0	47.9	32.6	4.90
$\beta = 0.04$	5.80	3.70	2.80	3.60	10.0	6.20	4.30	3.90	20.6	14.2	8.80	4.70	40.6	30.8	20.0	5.10	72.0	58.9	41.5	5.30
$\beta = 0.2$	9.90	6.50	3.90	4.50	22.7	15.9	9.10	5.30	48.7	37.3	23.7	7.40	84.5	72.4	52.0	11.6	99.3	97.1	87.2	23.4
$\beta = 0.4$	13.9	9.40	5.30	5.20	35.3	25.4	14.4	7.80	68.8	57.1	37.9	16.5	96.8	91.1	72.7	42.5	100	100	97.7	96.9
$\beta = 0.6$	17.8	11.8	6.70	5.60	45.8	33.7	20.3	12.8	82.7	71.8	51.1	39.9	99.6	97.2	86.8	99.1	100	100	100	100
$\beta = 0.8$	22.4	13.8	9.00	8.30	55.5	40.1	24.4	23.1	91.3	81.7	61.5	91.7	100	99.2	95.7	100	100	100	100	100
$\beta = 1$	26.5	17.9	10.9	10.7	64.5	48.1	30.6	39.5	95.0	88.5	70.1	100	100	99.6	100	100	100	100	100	100

Setti g II: i.i.d. . e . al c																				
Te t	$\delta = 0.1$				$\delta = 0.15$				$\delta = 0.2$				$\delta = 0.25$				$\delta = 0.3$			
	DCF	CL	XL	CQ	DCF	CL	XL	CQ	DCF	CL	XL	CQ	DCF	CL	XL	CQ	DCF	CL	XL	CQ
$\beta = 0$	4.90	1.80	3.70	6.10	5.20	1.30	2.20	3.80	5.00	1.60	3.60	6.00	4.80	1.20	3.50	6.30	5.00	1.90	3.90	6.20
$\beta = 0.02$	4.70	1.00	2.40	3.80	6.60	1.40	2.70	4.10	10.7	2.60	2.90	4.10	19.1	6.70	4.80	4.40	33.3	14.4	8.80	4.50
$\beta = 0.04$	5.80	1.30	2.50	4.10	7.90	1.80	2.80	4.30	12.5	3.50	3.40	4.50	24.7	9.30	6.00	4.60	42.5	20.3	12.2	5.00
$\beta = 0.2$	8.10	1.90	2.70	4.60	15.0	4.40	3.80	4.90	30.9	11.2	7.20	6.40	57.6	26.5	16.3	8.40	86.8	52.1	33.9	11.8
$\beta = 0.4$	10.6	2.80	3.10	5.70	22.4	7.20	5.70	6.50	47.3	19.6	11.6	10.0	78.7	43.2	26.6	19.1	97.5	74.1	53.2	45.7
$\beta = 0.6$	13.5	3.30	3.80	6.70	29.2	9.60	6.70	8.40	59.0	26.5	17.1	18.7	90.5	56.2	36.7	54.4	99.8	88.1	70.1	99.6
$\beta = 0.8$	16.4	4.60	4.50	7.40	37.4	11.9	8.60	12.6	70.9	32.9	21.4	39.6	95.6	67.0	47.0	F = 4	1	T	f	3

TABLE 1  
(Continued)

Test	Setting III: common level selected																			
	$\delta = 0.1$				$\delta = 0.15$				$\delta = 0.2$				$\delta = 0.25$				$\delta = 0.3$			
	DCF	CL	XL	CQ	DCF	CL	XL	CQ	DCF	CL	XL	CQ	DCF	CL	XL	CQ	DCF	CL	XL	CQ
$\beta = 0$	4.70	2.00	3.90	6.30	4.50	1.70	2.30	3.50	4.80	1.90	3.70	6.10	4.60	2.20	2.80	3.90	5.10	2.10	3.80	6.20
$\beta = 0.02$	4.90	2.10	3.20	4.40	6.50	2.70	3.50	5.30	9.40	4.30	4.00	5.60	13.6	7.80	6.20	5.70	24.9	12.9	10.1	5.90
$\beta = 0.04$	5.60	2.40	3.50	4.70	7.60	3.40	4.20	5.40	12.1	6.00	5.00	5.80	19.1	10.8	8.80	6.00	32.8	19.1	13.8	6.50
$\beta = 0.2$	7.50	3.80	4.30	5.80	12.1	6.00	5.60	6.60	23.9	12.5	8.90	7.50	44.2	26.3	16.6	9.30	71.6	50.2	32.1	14.1
$\beta = 0.4$	9.40	3.90	4.50	6.30	18.4	9.00	8.00	7.60	35.8	19.9	12.7	11.7	62.3	40.8	26.4	18.5	89.3	69.9	48.6	31.5
$\beta = 0.6$	11.5	4.90	6.20	6.80	24.0	10.8	8.90	9.50	48.0	28.2	18.2	17.8	76.8	55.3	37.0	35.7	96.5	83.8	64.6	83.1
$\beta = 0.8$	13.6	6.40	6.60	7.00	30.3	13.5	11.7	12.7	57.3	36.4	23.4	28.5	86.7	65.0	45.1	81.2	98.5	91.6	77.4	100
$\beta = 0.83$	14.3	7.10	6.80	7.50	31.0	14.6	11.8	13.1	58.0	37.6	23.9	30.8	87.6	66.1	46.1	88.0	98.9	92.6	79.2	100
$\beta = 1$	16.6	8.50	7.40	8.00	35.0	17.2	13.9	17.3	65.6	42.8	28.3	48.2	90.8	75.7	56.0	99.9	99.2	95.5	95.7	100

TABLE 2  
Rejection proportions (%) calculated for four testing methods at different signal strength levels of  $\delta$  and sparsity levels of  $\beta$  based on 1000 Monte Carlo runs, where  $\beta = 0$  corresponds to the null hypothesis  $\beta = 1$  to the fully dense alternative,  $(n, m, p) = (100, 400, 1000)$  for Setting IV, and  $(n, m, p) = (200, 300, 1000)$  for Settings V and VI

Setti g IV: c m level :elaæda d high l e al am le i e																				
Te t	ð = 0.1				ð = 0.15				ð = 0.2				ð = 0.25				ð = 0.3			
	DCF	CL	XL	CQ	DCF	CL	XL	CQ	DCF	CL	XL	CQ	DCF	CL	XL	CQ	DCF	CL	XL	CQ
β = 0	4.70	0.800	3.90	6.80	4.90	0.900	3.80	6.30	5.20	0.700	3.90	6.10	4.50	0.600	3.50	6.00	4.90	0.500	3.40	6.10
β = 0.02	5.20	1.10	2.90	4.70	5.90	1.00	3.60	5.60	6.70	1.40	4.60	5.80	8.90	2.40	5.00	5.80	13.2	4.20	6.20	5.90
β = 0.04	5.40	1.20	3.00	4.80	6.30	1.30	4.50	5.70	7.80	1.90	5.00	6.00	11.2	3.30	5.60	6.10	17.6	5.70	7.10	6.20
β = 0.2	6.60	1.30	3.30	5.40	9.20	2.20	5.10	5.80	14.9	3.90	5.70	6.20	25.3	8.70	7.00	7.50	42.8	16.5	11.8	8.80
β = 0.4	7.80	2.00	4.30	5.50	12.4	3.40	5.20	6.10	22.3	6.60	7.10	8.60	38.2	13.0	9.70	10.7	61.3	24.8	17.0	15.8
β = 0.6	9.10	2.40	4.60	5.80	16.1	3.80	5.50	7.90	29.5	10.0	9.20	10.8	49.9	19.3	14.3	17.6	75.3	33.7	21.9	34.2
β = 0.8	10.5	2.50	4.70	6.10	19.9	5.20	6.70	9.20	36.9	12.7	10.9	14.5	60.1	24.0	19.3	32.2	84.9	46.6	33.6	78.2
β = 0.9	11.3	2.80	4.80	6.40	21.9	5.40	7.10	9.90	39.5	13.3	12.6	17.7	64.6	26.6	21.6	43.8	88.0	48.6	35.3	94.0
β = 1	12.1	2.90	5.30	7.30	23.4	5.90	7.30	11.0	42.0	14.6	12.8	21.7	68.6	29.6	24.5	59.0	90.9	53.1	41.9	99.4
Setti g V: c m level :elaæda d hea -tailed																				
Te t	ð = 0.1				ð = 0.15				ð = 0.2				ð = 0.25				ð = 0.3			
	DCF	CL	XL	CQ	DCF	CL	XL	CQ	DCF	CL	XL	CQ	DCF	CL	XL	CQ	DCF	CL	XL	CQ
β = 0	4.20	2.20	3.80	6.20	5.20	2.50	3.90	6.10	4.70	1.90	2.90	6.00	4.30	2.00	1.70	3.90	4.50	2.30	2.00	3.70
β = 0.02	5.50	2.10	3.70	5.40	6.40	2.50	3.90	5.50	9.50	4.40	4.60	6.10	15.3	7.40	6.30	6.10	25.5	15.0	10.3	6.20
β = 0.04	6.20	2.30	3.80	5.50	7.20	3.60	4.20	6.00	12.6	6.60	5.80	6.20	18.9	9.80	7.00	6.50	33.3	20.7	13.0	7.10
β = 0.2	7.50	3.60	4.00	5.80	12.4	6.80	6.50	7.30	23.5	13.0	9.60	8.90	45.6	27.6	17.9	11.3	71.7	52.6	33.8	14.1
β = 0.4	9.50	4.20	4.40	5.90	18.1	9.00	8.30	8.90	35.9	21.3	14.0	12.7	64.4	43.2	26.9	18.5	90.3	73.4	52.0	33.7
β = 0.6	11.5	5.10	4.50	6.00	23.8	12.6	10.1	11.7	46.7	29.2	19.4	17.8	77.5	55.9	37.4	38.9	97.4	86.5	65.6	88.2
β = 0.8	13.7	7.30	6.20	8.80	29.4	16.0	12.3	14.1	56.5	36.9	24.9	28.9	87.4	69.1	48.3	81.4	99.2	93.6	80.0	100
β = 0.83	14.1	7.50	6.30	9.20	30.6	17.3	13.0	15.2	58.1	38.1	26.0	32.0	88.1	70.1	49.5	87.5	99.3	94.1	82.1	100
β = 1	16.1	8.90	7.40	9.40	34.9	18.9	15.0	17.2	64.5	44.6	30.5	52.2	91.6	75.1	56.6	99.8	99.7	96.5	96.0	100



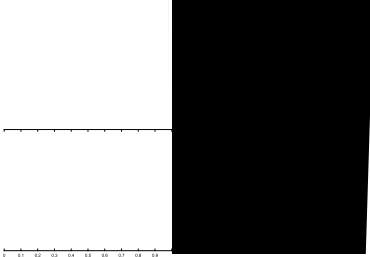


TABLE 3  
Shown are the results of four tests based the original dataset, the bootstrapped samples and the random permutations

$p$ -value of the test based the data set				
Test	DCF	CL	XL	CQ
$p$ -value	0.006	0.1708	0.093	0.0955
Rejection rate (%) of the test at $n = 500$ bootstrapped data set				
Test	DCF	CL	XL	CQ
Rejection rate	82	65.8	65	58
Rejection rate (%) of the test at $n = 500$ random data set				
Test	DCF	CL	XL	CQ
Rejection rate	4.6	1.8	3.4	7.4

500 bootstrapped data sets are given in Table 3, which shows that the highest rejection rate is achieved by DCF at 82%. This is in line with the smaller standard  $p$ -value given by the DCF test based on the data set itself. We also examined 500 random data sets of the whole data set (i.e., mixing groups to eliminate the group difference) and conducted the test on each examined data set. From Table 3, we see that the rejection rate of the DCF test (0.046) is close to the nominal level  $\alpha = 0.05$ , while the other three tests differ considerably.

# APPENDIX

We first recall the auxiliary lemma that are key for deriving the main theorem. The first is due to Lemma 1, for a  $\beta > 0$  and  $y \in \mathbb{R}^p$ , we define a function  $F_\beta(w)$  as

$$F_\beta(w) = \beta^{-1} \mathbb{1} \left[ \sum_{j=1}^p \alpha_j \{ \beta(w_j - y_j) \} \right], \quad w \in \mathbb{R}^p,$$

which attains the constant

$$0 \leq F_\beta(w) - \max_{1 \leq j \leq p} (w_j - y_j) \leq \beta^{-1} \mathbb{1} \{ p \},$$

for every  $w \in \mathbb{R}^p$  by (1) in [8]. In addition, we let  $\varphi_0 : \mathbb{R} \rightarrow [0, 1]$  be a real valued function such that  $\varphi_0$  is thrice continuous and differentiable and  $\varphi_0(z) = 1$  for  $z \leq 0$  and  $\varphi_0(z) = 0$  for  $z \geq 1$ . For a  $\phi \geq 1$ , define a function  $\varphi(z) = \varphi_0(\phi z)$ ,  $z \in \mathbb{R}$ . Then, for a  $\phi \geq 1$  and  $y \in \mathbb{R}^p$ , define  $\beta = \phi \mathbb{1} \{ p \}$  and define a function  $\kappa : \mathbb{R}^p \rightarrow [0, 1]$  as

$$(9) \quad \kappa(w) = \varphi_0(\phi F_{\phi \mathbb{1} \{ p \}}(w)) = \varphi(F_\beta(w)), \quad w \in \mathbb{R}^p.$$

Lemma 1 is devoted to characterizing the constant of the function  $\kappa$  defined in (9), which can be also referred to Lemma A.5 and A.6 in [7].

LEMMA 1. For any  $\phi \geq 1$  and  $y \in \mathbb{R}^p$ , we denote  $\beta = \phi \mathbb{1} \{ p \}$ , then the function  $\kappa$  defined in (9) has the following properties, where  $\kappa_{jkl}$  denotes  $\partial_j \partial_k \partial_l \kappa$ . For any  $j, k, l = 1, \dots, p$ , there exists a nonnegative function  $Q_{jkl}$  such that:

- (1)  $|\kappa_{jkl}(w)| \leq Q_{jkl}(w)$  for all  $w \in \mathbb{R}^p$ ,
- (2)  $\sum_{j=1}^p \sum_{k=1}^p \sum_{l=1}^p Q_{jkl}(w) \lesssim (\phi^3 + \phi^2\beta + \phi\beta^2) \lesssim \phi\beta^2$  for all  $w \in \mathbb{R}^p$ ,
- (3)  $Q_{jkl}(w) \lesssim Q_{jkl}(w + \tilde{w}) \lesssim Q_{jkl}(w)$  for all  $w \in \mathbb{R}^p$  and  $\tilde{w} \in \{w^* \in \mathbb{R}^p : \max_{1 \leq j \leq p} |w_j^*| \beta \leq 1\}$ .

Take Lemma 2, and combine it with Lemma 5.1 in [9], for a given  $\epsilon$  of constant  $\delta_{n,m}$  that depends on  $n$  and  $m$ , we define the  $\rho_{n,m}$  by

$$\begin{aligned} \rho_{n,m} = & \int_{v \in [0,1]} \int_{y \in \mathbb{R}^p} |P\{v^{1/2}(S_n^X - n^{1/2}\mu^X + \delta_{n,m}S_m^Y - \delta_{n,m}m^{1/2}\mu^Y) \\ & + (1-v)^{1/2}(S_n^F - n^{1/2}\mu^X + \delta_{n,m}S_m^G - \delta_{n,m}m^{1/2}\mu^Y) \leq y\} \\ & - P(S_n^F - n^{1/2}\mu^X + \delta_{n,m}S_m^G - \delta_{n,m}m^{1/2}\mu^Y \leq y)|. \end{aligned}$$

(10)

Lemma 2 : Under the above conditions, the sequence  $\rho_{n,m}$  converges to 0.

LEMMA 2. For any  $\phi_1, \phi_2 \geq 1$  and any sequence of constants  $\delta_{n,m}$ , assume the following condition (a) holds,

- (a) There exists a universal constant  $b > 0$  such that

$$\min_{1 \leq j \leq n} E\{(S_{nj}^X - n^{1/2}\mu_j^X + \delta_{n,m}S_{mj}^Y - \delta_{n,m}m^{1/2}\mu_j^Y)^2\} \geq b.$$

---

Then we have

$$\rho_{n,m} \lesssim n^{-1/2} \phi_1^2 (1 + p) 21$$

Then we have

$$\begin{aligned} \rho_{n,m}^* &\leq K^* [n^{-1/2} \phi_1^2 (1 \text{ g } p)^2 \{ \phi_1 L_n^X \rho_{n,m}^* + L_n^X (1 \text{ g } p)^{1/2} + \phi_1 M_n(\phi_1) \} \\ &\quad + m^{-1/2} \phi_2^2 (1 \text{ g } p)^2 |\delta_{n,m}|^3 \{ \phi_2 L_m^Y \rho_{n,m}^* + L_m^Y (1 \text{ g } p)^{1/2} + \phi_2 M_m^*(\phi_2) \} \\ &\quad + (\min \{ \phi_1, \phi_2 \})^{-1} (1 \text{ g } p)^{1/2}], \end{aligned}$$

up to a universal constant  $K^* > 0$  that depends only on  $b$ , where  $\rho_{n,m}^*$  is defined in (11).

Before stating the next lemma, fix a  $\phi \geq 1$ , we define  $M_n(\phi) = M_n^X(\phi) + M_n^F(\phi)$ , hence  $M_n^X(\phi)$  and  $M_n^F(\phi)$  are given as follows, respectively,

$$\begin{aligned} n^{-1} \sum_{i=1}^n E \left[ \max_{1 \leq j \leq p} |X_{ij} - \mu_j^X|^3 1 \left\{ \max_{1 \leq j \leq p} |X_{ij} - \mu_j^X| > n^{1/2} / (4\phi \text{ g } p) \right\} \right], \\ n^{-1} \sum_{i=1}^n E \left[ \max_{1 \leq j \leq p} |F_{ij} - \mu_j^F|^3 1 \left\{ \max_{1 \leq j \leq p} |F_{ij} - \mu_j^F| > n^{1/2} / (4\phi \text{ g } p) \right\} \right], \end{aligned}$$

similar to the adopted in [9]. Likewise, fix a  $\phi \geq 1$  and define  $M_m(\phi) = M_m^Y(\phi) + M_m^G(\phi)$  with  $M_m^Y(\phi)$  and  $M_m^G(\phi)$  as follows, respectively,

$$\begin{aligned} m^{-1} \sum_{i=1}^m E \left[ \max_{1 \leq j \leq p} |Y_{ij} - \mu_j^Y|^3 1 \left\{ \max_{1 \leq j \leq p} |Y_{ij} - \mu_j^Y| > m^{1/2} / (4|\delta_{n,m}| \phi \text{ g } p) \right\} \right], \\ m^{-1} \sum_{i=1}^m E \left[ \max_{1 \leq j \leq p} |G_{ij} - \mu_j^G|^3 1 \left\{ \max_{1 \leq j \leq p} |G_{ij} - \mu_j^G| > m^{1/2} / (4|\delta_{n,m}| \phi \text{ g } p) \right\} \right]. \end{aligned}$$

Recalling the definition of  $\rho_{n,m}^{**}$  in (2), Lemma 4 gives a abstract condition on  $\rho_{n,m}^{**}$  denoted mild condition as follows.

LEMMA 4. For any sequence of constants  $\delta_{n,m}$ , assume we have the following conditions (a)–(b):

(a) There exists a universal constant  $b > 0$  such that

$$\min_{1 \leq j \leq p} E \{ (S_{nj}^X - n^{1/2} \mu_j^X + \delta_{n,m} S_{mj}^Y - \delta_{n,m} m^{1/2} \mu_j^Y)^2 \} \geq b.$$

(b) There exist two sequences of constants  $\bar{L}_n^*$  and  $\bar{L}_m^{**}$  such that we have  $\bar{L}_n^* \geq L_n^X$  and  $\bar{L}_m^{**} \geq L_m^Y$ , respectively. Moreover, we also have

$$\begin{aligned} \phi_n^* &= K_1 \{ (\bar{L}_n^*)^2 (1 \text{ g } p)^4 / n \}^{-1/6} \geq 2, \\ \phi_m^{**} &= K_1 \{ (\bar{L}_m^{**})^2 (1 \text{ g } p)^4 |\delta_{n,m}|^6 / m \}^{-1/6} \geq 2, \end{aligned}$$

for a universal constant  $K_1 \in (0, (K^* \vee 2)^{-1}]$ , where the positive constant  $K^*$  that depends on  $n$  as defined in Lemma 3 in the Appendix.

Then we have the following property, where  $\rho_{n,m}^{**}$  is defined in (2),

$$\begin{aligned} \rho_{n,m}^{**} &\leq K_2 \{ \{ (\bar{L}_n^*)^2 (1 \text{ g } p)^7 / n \}^{1/6} + \{ M_n(\phi_n^*) / \bar{L}_n^* \} \\ &\quad + \{ (\bar{L}_m^{**})^2 (1 \text{ g } p)^7 |\delta_{n,m}|^6 / m \}^{1/6} + \{ M_m^*(\phi_m^{**}) / \bar{L}_m^{**} \} \}, \end{aligned}$$

for a universal constant  $K_2 > 0$  that depends only on  $b$ .

Take due Lemma 5, find a sequence of constants  $\delta_{n,m}$  that depend on  $n$  and  $m$ , define a self-adjoint  $\hat{\Delta}_{n,m} = \|\hat{\Sigma}^X - \Sigma^X + \delta_{n,m}^2(\hat{\Sigma}^Y - \Sigma^Y)\|_\infty$ . Lemma 5 belongs to the abstract and  $\rho_{n,m}^{MB}$  defined in (4).

LEMMA 5. For any sequence of constants  $\delta_{n,m}$ , assume we have the following condition (a):

(a) There exists a universal constant  $b > 0$  such that

$$\min_{1 \leq j \leq p} E\{(S_{nj}^X - n^{1/2}\mu_j^X + \delta_{n,m}S_{mj}^Y - \delta_{n,m}m^{1/2}\mu_j^Y)^2\} \geq b.$$

Then for any sequence of constants  $\bar{\Delta}_{n,m} > 0$ , on the event  $\{\hat{\Delta}_{n,m} \leq \bar{\Delta}_{n,m}\}$ , we have the following property, where  $\rho_{n,m}^{MB}$  is defined in (4),

$$\rho_{n,m}^{MB} \lesssim (\bar{\Delta}_{n,m})^{1/3}(\log p)^{2/3}.$$

Let  $\mathbb{P}_\mathbb{P}$  be the probability measure on  $\mathcal{B}$  defined in Lemma 6.

LEMMA 6. Let  $\{A_{n,m} : n \geq 1, m \geq 1, (n, m) \in A\}$  be a sequence of events in the sample space  $\Omega$ , where  $A$  is the set of all possible combinations  $(n, m)$ , which has the form  $A = \{(n, m) : n \geq 1, m \in \sigma(n)\}$  where  $\sigma(n)$  is a set of positive integers determined by  $n$ , possibly the empty set. Assume the following condition (a):

(a)  $\sum_{n=1}^\infty \sum_{m \in \sigma(n)} P(A_{n,m}) < \infty$ .

Then we have the following property:

$$P\left(\bigcap_{k_1=1}^\infty \bigcap_{k_2=1}^\infty \bigcup_{n=k_1}^\infty \bigcup_{m \in \sigma(k_2) \cap \sigma(n)} A_{n,m}\right) = 0,$$

where  $\sigma(k_2) = \{k : k \in \mathbb{Z}, k \geq k_2\}$ .

Note that if  $m \in \sigma(n) = \emptyset$ , we just delete the self-adjoint  $A_{n,m}$  and  $A_{n,m}^c$  and ignore all the corresponding additivity, and the same also holds for  $P(A_{n,m})$  and  $P(A_{n,m}^c)$  and the corresponding deduction.

Before proceeding, we mention that the definition of Theorem 1.2 is essentially the same as the definition in [9], but we mention technical details in the appendix Lemma 4.5 to address the challenge in giving the formal definition. The definition of Corollary 1.1 is based on Theorem 1.1 and the abstract version of the Bell-Cattell lemma (Lemma 6) that is a corollary of the so-called  $\mathbb{P}_\mathbb{P}$ .

Theorem 3.5 regarding the DCF is a special case of the more general case. The corresponding self-adjointness of Theorem 3.5 belongs to the self-adjointness of Theorem 1.2, Corollary 1.1 and the auxiliary lemma are delegated to the following Lemma 5.1. Material for the appendix.

PROOF OF THEOREM 3. First of all, we define a sequence of constants  $\delta_{n,m}$  by

(12) 
$$\delta_{n,m} = -n^{1/2}m^{-1/2}.$$

Together with condition (a), it can be deduced that

(13) 
$$\delta_2 < |\delta_{n,m}| < \delta_1,$$

where  $\delta_1 = \{c_2/(1 - c_2)\}^{1/2} > 0$  and  $\delta_2 = \{c_1/(1 - c_1)\}^{1/2} >$

PROOF OF THEOREM 4. Given  $(\mu^X - \mu^Y)_{\mathbb{V}}$  e h a e

$$\begin{aligned}
 & P_{\mathbb{V}} \mathfrak{E}^*(\mu^X - \mu^Y) \\
 &= P_{e^*} \{ \|S_n^{e^*X} - n^{1/2}m^{-1/2}S_m^{e^*Y} + n^{1/2}(\mu^X - \mu^Y)\|_{\infty} \geq c_B(\alpha) \} \\
 &= 1 - P_{e^*} \{ \|S_n^{e^*X} - n^{1/2}m^{-1/2}S_m^{e^*Y} + n^{1/2}(\mu^X - \mu^Y)\|_{\infty} < c_B(\alpha) \} \\
 &= 1 - P_{e^*} \{ -n^{1/2}(\mu^X - \mu^Y) - c_B(\alpha) < S_n^{e^*X} - n^{1/2}m^{-1/2}S_m^{e^*Y} < \\
 &\quad -n^{1/2}(\mu^X - \mu^Y) + c_B(\alpha) \} \\
 &= 1 - P_{e^*} \{ -n^{1/2}(\mu^X - \mu^Y) - c_B(\alpha) < S_n^{e^*X} - n^{1/2}m^{-1/2}S_m^{e^*Y} < \\
 &\quad -n^{1/2}(\mu^X - \mu^Y) + c_B(\alpha) \} \\
 &\quad + P \{ -n^{1/2}(\mu^X - \mu^Y) - c_B(\alpha) < S_n^X - n^{1/2}m^{-1/2}S_m^Y \\
 &\quad - n^{1/2}(\mu^X - \mu^Y) < -n^{1/2}(\mu^X - \mu^Y) + c_B(\alpha) \} \\
 (22) \quad &\quad - P \{ -n^{1/2}(\mu^X - \mu^Y) - c_B(\alpha) < S_n^X - n^{1/2}m^{-1/2}S_m^Y \\
 &\quad - n^{1/2}(\mu^X - \mu^Y) < -n^{1/2}(\mu^X - \mu^Y) + c_B(\alpha) \} \\
 &\geq 1 - \sum_{A \in \mathcal{A}^{\text{Re}}} |P(\|S_n^X - n^{1/2}m^{-1/2}S_m^Y \\
 &\quad - n^{1/2}(\mu^X - \mu^Y)\|_{\infty} \in A) - P_{e^*}(\|S_n^{e^*X} - n^{1/2}m^{-1/2}S_m^{e^*Y}\|_{\infty} \in A)| \\
 &\quad - P\{\|S_n^X - n^{1/2}m^{-1/2}S_m^Y\|_{\infty} < c_B(\alpha)\} \\
 &= P_{\mathbb{V}} \mathfrak{E}(\mu^X - \mu^Y) \\
 &\quad - \sum_{A \in \mathcal{A}^{\text{Re}}} |P(\|S_n^X - n^{1/2}m^{-1/2}S_m^Y - n^{1/2}(\mu^X - \mu^Y)\|_{\infty} \in A) \\
 &\quad - P_{e^*}(\|S_n^{e^*X} - n^{1/2}m^{-1/2}S_m^{e^*Y}\|_{\infty} \in A)|.
 \end{aligned}$$

Like i e, given  $(\mu^X - \mu^Y)_{\mathbb{V}}$  e h a e

$$\begin{aligned}
 & P_{\mathbb{V}} \mathfrak{E}(\mu^X - \mu^Y) \\
 &= P\{\|S_n^X - n^{1/2}m^{-1/2}S_m^Y\|_{\infty} \geq c_B(\alpha)\} \\
 &= 1 - P\{\|S_n^X - n^{1/2}m^{-1/2}S_m^Y\|_{\infty} < c_B(\alpha)\} \\
 &= 1 - P\{-c_B(\alpha) < S_n^X - n^{1/2}m^{-1/2}S_m^Y < c_B(\alpha)\} \\
 &= 1 + P_{e^*} \{ -n^{1/2}(\mu^X - \mu^Y) - c_B(\alpha) < S_n^{e^*X} - n^{1/2}m^{-1/2}S_m^{e^*Y} < \\
 &\quad -n^{1/2}(\mu^X - \mu^Y) + c_B(\alpha) \} - P\{ -n^{1/2}(\mu^X - \mu^Y) - c_B(\alpha) \\
 (23) \quad &\quad < S_n^X - n^{1/2}m^{-1/2}S_m^Y - n^{1/2}(\mu^X - \mu^Y) < -n^{1/2}(\mu^X - \mu^Y) + c_B(\alpha) \} \\
 &\quad - P_{e^*} \{ -n^{1/2}(\mu^X - \mu^Y) - c_B(\alpha) < S_n^{e^*X} - n^{1/2}m^{-1/2}S_m^{e^*Y} \\
 &\quad < -n^{1/2}(\mu^X - \mu^Y) + c_B(\alpha) \} \\
 &\geq 1 - \sum_{A \in \mathcal{A}^{\text{Re}}} |P(\|S_n^X - n^{1/2}m^{-1/2}S_m^Y - n^{1/2}(\mu^X - \mu^Y)\|_{\infty} \in A) \\
 &\quad - P_{e^*}(\|S_n^{e^*X} - n^{1/2}m^{-1/2}S_m^{e^*Y}\|_{\infty} \in A)|
 \end{aligned}$$

$$\begin{aligned}
& - P_{e^*} \{ \|S_n^{e^*X} - n^{1/2}m^{-1/2}S_m^{e^*Y} + n^{1/2}(\mu^X - \mu^Y)\|_\infty < c_B(\alpha) \} \\
= & P_{\mathbb{V}} \{ \mathfrak{A}^*(\mu^X - \mu^Y) \\
& - \big| P(\|S_n^X - n^{1/2}m^{-1/2}S_m^Y - n^{1/2}(\mu^X - \mu^Y)\|_\infty \in A) \\
& - P_{e^*}(\|S_n^{e^*X} - n^{1/2}m^{-1/2}S_m^{e^*Y}\|_\infty \in A) \big|.
\end{aligned}$$

Putting (22) and (23) together indicate that

$$\begin{aligned}
(24) \quad & |P_{\mathbb{V}} \{ \mathfrak{A}^*(\mu^X - \mu^Y) - P_{\mathbb{V}} \{ \mathfrak{A}(\mu^X - \mu^Y) \} \\
& \leq \big| P(\|S_n^X - n^{1/2}m^{-1/2}S_m^Y - n^{1/2}(\mu^X - \mu^Y)\|_\infty \in A) \\
& - P_{e^*}(\|S_n^{e^*X} - n^{1/2}m^{-1/2}S_m^{e^*Y}\|_\infty \in A) \big|.
\end{aligned}$$

Moreover, by similar argument as in the proof of Theorem 3, we can show that with probability one,

$$\begin{aligned}
(25) \quad & \big| P(\|S_n^X - n^{1/2}m^{-1/2}S_m^Y - n^{1/2}(\mu^X - \mu^Y)\|_\infty \in A) \\
& - P_{e^*}(\|S_n^{e^*X} - n^{1/2}m^{-1/2}S_m^{e^*Y}\|_\infty \in A) \big| \\
& \lesssim \{B_{n,m}^2 \log^7(pn)/n\}^{1/6}.
\end{aligned}$$

Finally, by combining (24) with (25), for a  $\mu^X - \mu^Y \in \mathbb{R}^p$ , we have that with probability one,

$$|P_{\mathbb{V}} \{ \mathfrak{A}^*(\mu^X - \mu^Y) - P_{\mathbb{V}} \{ \mathfrak{A}(\mu^X - \mu^Y) \} | \lesssim \{B_{n,m}^2 \log^7(pn)/n\}^{1/6},$$

which complete the proof.  $\square$

PROOF OF THEOREM 5. First of all, the basic part of (8) and the triangle inequality, it is clear that

$$\begin{aligned}
(26) \quad & P_{\mathbb{V}} \{ \mathfrak{A}^*(\mu^X - \mu^Y) \geq P_{e^*} \{ \|S_n^{e^*X} - n^{1/2}m^{-1/2}S_m^{e^*Y}\|_\infty \\
& \leq \|n^{1/2}(\mu^X - \mu^Y)\|_\infty - c_B(\alpha) \}.
\end{aligned}$$

At this point, with the above fact, we denote  $\{e_j : j \leq p\}$  as the standard basis for  $\mathbb{R}^p$ . Then it follows from independence and central limit theorem that for a  $t \geq 0$ ,

$$\begin{aligned}
(27) \quad & P_{e^*} \{ \|S_n^{e^*X} - n^{1/2}m^{-1/2}S_m^{e^*Y}\|_\infty \geq t \} \\
& \leq \sum_{j=1}^p P_{e^*} \{ |S_{nj}^{e^*X} - n^{1/2}m^{-1/2}S_{mj}^{e^*Y}| \geq t \} \\
& \leq \sum_{j=1}^p 2\alpha \left[ -t^2 / \{2e'_j(\hat{\Sigma}^X + nm^{-1}\hat{\Sigma}^Y)e_j\} \right] \\
& \leq 2p\alpha \left( -t^2 / \left[ 2 \max_{j \leq p} \{e'_j(\hat{\Sigma}^X + nm^{-1}\hat{\Sigma}^Y)e_j\} \right] \right).
\end{aligned}$$

By plugging  $t = c_B(\alpha)$  into (27), it follows from the definition of  $c_B(\alpha)$  that

$$\begin{aligned}
(28) \quad & c_B(\alpha) \leq \left[ 2 \log(2p/\alpha) \max_{j \leq p} \{e'_j(\hat{\Sigma}^X + nm^{-1}\hat{\Sigma}^Y)e_j\} \right]^{1/2} \\
& \leq \left[ 4 \log(pn) \max_{j \leq p} \{e'_j(\hat{\Sigma}^X + nm^{-1}\hat{\Sigma}^Y)e_j\} \right]^{1/2},
\end{aligned}$$

for the following large  $n$ . To bound the  $\max_{j \leq p} \{e'_j(\hat{\Sigma}^X + nm^{-1}\hat{\Sigma}^Y)e_j\}$ , it is sufficient to show that

$$\begin{aligned} & \max_{j \leq p} \{e'_j(\hat{\Sigma}^X + nm^{-1}\hat{\Sigma}^Y)e_j\} \\ (29) \quad &= \|\hat{\Sigma}^X + nm^{-1}\hat{\Sigma}^Y\|_{\infty} \\ &\leq \|\hat{\Sigma}^X - \Sigma^X + nm^{-1}(\hat{\Sigma}^Y - \Sigma^Y)\|_{\infty} + \|\Sigma^X + nm^{-1}\Sigma^Y\|_{\infty}. \end{aligned}$$

For the term  $\|\hat{\Sigma}^X - \Sigma^X + nm^{-1}(\hat{\Sigma}^Y - \Sigma^Y)\|_{\infty}$ , in equality (53) and (54) from the Supplementary Material together with (12), (17) and condition (a) we find that there exist a universal constant  $c_1 > 0$  such that

$$(30) \quad \|\hat{\Sigma}^X - \Sigma^X + nm^{-1}(\hat{\Sigma}^Y - \Sigma^Y)\|_{\infty} \leq c_1 \{B_{n,m}^2 \log^3(pn)/n\}^{1/2},$$

with probability going to one. Regarding the term  $\|\Sigma^X + nm^{-1}\Sigma^Y\|_{\infty}$ , we have

$$\begin{aligned} & \|\Sigma^X + nm^{-1}\Sigma^Y\|_{\infty} \\ &\leq \|\Sigma^X\|_{\infty} + nm^{-1}\|\Sigma^Y\|_{\infty} \leq \|\Sigma^X\|_{\infty} + c_2\|\Sigma^Y\|_{\infty} \\ &= \max_{1 \leq j \leq p} \sum_{i=1}^n E\{(X_{ij} - \mu_j^X)^2\}/n + c_2 \max_{1 \leq j \leq p} \sum_{i=1}^m E\{(Y_{ij} - \mu_j^Y)^2\}/m \\ (31) \quad &\leq \max_{1 \leq j \leq p} \sum_{i=1}^n [E\{(X_{ij} - \mu_j^X)^4\}]^{1/2}/n \\ &\quad + c_2 \max_{1 \leq j \leq p} \sum_{i=1}^m [E\{(Y_{ij} - \mu_j^Y)^4\}]^{1/2}/m \\ &\leq \left[ \max_{1 \leq j \leq p} \sum_{i=1}^n E\{(X_{ij} - \mu_j^X)^4\}/n \right]^{1/2} \\ &\quad + c_2 \left[ \max_{1 \leq j \leq p} \sum_{i=1}^m E\{(Y_{ij} - \mu_j^Y)^4\}/m \right]^{1/2} \\ &\leq c_3 B_{n,m}, \end{aligned}$$

for some universal constant  $c_2, c_3 > 0$ . Hence the second inequality in (a), the third inequality based on Jensen's inequality, the first inequality holds from the Cauchy-Schwarz inequality and the last inequality follows from condition (c). Therefore, combining (30), (31), (e) with (29), it can be deduced that there exist a universal constant  $c_4 > 0$  such that

$$(32) \quad \max_{j \leq p} \{e'_j(\hat{\Sigma}^X + nm^{-1}\hat{\Sigma}^Y)e_j\} \leq c_4 B_{n,m},$$

with probability going to one. Together with (28), it can be verified that

$$(33) \quad c_B(\alpha) \leq \{4c_4 B_{n,m} \log(pn)\}^{1/2},$$

with probability going to one. Next, we set the constant  $K_s$  in (f) as  $K_s = 4c_4^{1/2}$ , and in the following we prove (f) and (33) that

$$(34) \quad \|n^{1/2}(\mu^X - \mu^Y)\|_{\infty} - c_B(\alpha) \geq \{4c_4 B_{n,m} \log(pn)\}^{1/2},$$

with probability  $\delta$ . Hence, it can be deduced that with probability  $\delta$ ,  
 $e$ ,

$$\begin{aligned} & P_{\mathbb{P}} \left( \mathfrak{E}^*(\mu^X - \mu^Y) \right) \\ & \geq P_{e^*} \left[ \|S_n^{e^*X} - n^{1/2} m^{-1/2} S_m^{e^*Y}\|_{\infty} \leq \{4c_4 B_{n,m} \log(pn)\}^{1/2} \right] \\ & = 1 - P_{e^*} \left[ \|S_n^{e^*X} - n^{1/2} m^{-1/2} S_m^{e^*Y}\|_{\infty} \geq \{4c_4 B_{n,m} \log(pn)\}^{1/2} \right] \\ & \geq 1 - 2p \mathfrak{E} \left( -4c_4 B_{n,m} \log(pn) / \left[ 2 \max_{j \leq p} \{e'_j (\hat{\Sigma}^X + nm^{-1} \hat{\Sigma}^Y) e_j\} \right] \right) \end{aligned}$$

- [9] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2017). Central limit theorems and bootstrap for high dimensional data. *Ann. Probab.* **45** 2309–2352. [MR3693963](#) [https://doi.org/10.1214/16-AOP1113](#)
- [10] FENG, L., ZOU, C., WANG, Z. and ZHU, L. (2015).  $T_{\sqrt{n}}$ -asymptotically efficient Fisher information for high-dimensional data. *Statist. Sinica* **25** 1297–1312. [MR3409068](#)
- [11] GREGORY, K. B., CARROLL, R. J., BALADANDAYUTHAPANI, V. and LAHIRI, S. N. (2015). Asymptotically efficient inference for high dimensional data. *J. Amer. Statist. Assoc.* **110** 837–849. [MR3367268](#) [https://doi.org/10.1080/01621459.2014.934826](#)
- [12] HU, J., BAI, Z., WANG, C. and WANG, W. (2017). On testing the asymptotically efficient inference. *Ann. Inst. Statist. Math.* **69** 365–387. [MR3611524](#) [https://doi.org/10.1007/s10463-015-0543-8](#)
- [13] HUSSAIN, L., AZIZ, W., NADEEM, S. A., SHAH, S. A. and MAJID, A. (2015). Electroencephalogram (EEG) analysis of functional connectivity in glioma patients. *J. Multidiscip. Eng. Sci. Technol.* **1** 3159–0040.
- [14] PARK, J. and AYYALA, D. N. (2013). Asymptotic efficiency of large dimensional and small sample size. *J. Statist. Plann. Inference* **143** 929–943. [MR3011304](#) [https://doi.org/10.1016/j.jspi.2012.11.001](#)
- [15] SHEN, Y. and LIN, Z. (2015). Asymptotic efficiency of the mean of large  $p$ -small  $n$  problem. *Comput. Statist. Data Anal.* **89** 25–38. [MR3349665](#) [https://doi.org/10.1016/j.csda.2015.03.004](#)
- [16] SRIVASTAVA, M. S. (2007). Multivariate test for a large high dimensional data. *J. Japan Statist. Soc.* **37** 53–86. [MR2392485](#) [https://doi.org/10.14490/jjs.37.53](#)
- [17] SRIVASTAVA, M. S. (2009). Asymptotic efficiency of the Fisher information for the dimensionality. *J. Multivariate Anal.* **100** 518–532. [MR2483435](#) [https://doi.org/10.1016/j.jmva.2008.06.006](#)
- [18] SRIVASTAVA, M. S. and DU, M. (2008). Asymptotic efficiency of the Fisher information for the dimensionality. *J. Multivariate Anal.* **99** 386–402. [MR2396970](#) [https://doi.org/10.1016/j.jmva.2006.11.002](#)
- [19] SRIVASTAVA, M. S. and KUBOKAWA, T. (2013). Testing multivariate analysis of variance for high dimensional data. *J. Multivariate Anal.* **115** 204–216. [MR3004555](#) [https://doi.org/10.1016/j.jmva.2012.10.011](#)
- [20] WANG, L., PENG, B. and LI, R. (2015). A high-dimensional asymptotic multivariate test for mean. *J. Amer. Statist. Assoc.* **110** 1658–1669. [MR3449062](#) [https://doi.org/10.1080/01621459.2014.988215](#)
- [21] XU, G., LIN, L., WEI, P. and PAN, W. (2016). Asymptotic efficiency of high-dimensional mean. *Biometrika* **103** 609–624. [MR3551787](#) [https://doi.org/10.1093/biomet/asw029](#)
- [22] XUE, K. and YAO, F. (2019). Sufficient Dimensionality and Correlation-free asymptotically efficient high-dimensional mean. [https://doi.org/10.1214/19-AOS1848SUPP](#)
- [23] YAGI, A. and SEO, T. (2014). Asymptotic efficiency of the dimensionality test for three-dimensional time series data. *Amer. J. Math. Management Sci.* **33** 161–175.
- [24] YAMADA, T. and HIMENO, T. (2015). Testing homogeneity of mean for heterogeneous data in high-dimensional data. *J. Multivariate Anal.* **139** 7–27. [MR3349477](#) [https://doi.org/10.1016/j.jmva.2015.02.005](#)
- [25] ZHANG, J. and PAN, M. (2016). A high-dimensional asymptotically efficient mean regression. *Comput. Statist. Data Anal.* **97** 87–97. [MR3447038](#) [https://doi.org/10.1016/j.csda.2015.12.004](#)
- [26] ZHANG, X. (2015). Testing high dimensional mean difference. Preprint. Available at [arXiv:1509.08444](#) 2.
- [27] ZHAO, J. (2017). Asymptotic efficiency of large dimensional and small sample size. *Comm. Statist. Simulation Comput.* **46** 6115–6128. [MR3740770](#) [https://doi.org/10.1080/03610918.2016.1197244](#)
- [28] ZHONG, P.-S., CHEN, S. X. and XU, M. (2013). Testing alternative hypothesis for high-dimensional mean difference and clustering. *Ann. Statist.* **41** 2820–2851. [MR3161449](#) [https://doi.org/10.1214/13-AOS1168](#)
- [29] ZHU, Y. and BRADIC, J. (2016).  $T_{\sqrt{n}}$ -asymptotically efficient high-dimensional linear model. Preprint. Available at [arXiv:1610.04580](#) 1.