The Annals of Statistics
2017, Vol. 45, No. 2, 771–799
DOI: 10.1214/16-AOS1467

TESTS FOR HIGH-DIMENSIONAL DATA BASED ON MEANS, SPATIAL SIGNS AND SPATIAL RANKS

BY ANIRVAN CHAKRABORTY AND PROBAL CHAUDHURI

Indian Statistical Institute

Tests based on mean vectors and spatial signs and ranks for a zero mean in one-sample problems and for the equality of means in two-sample problems have been studied in the recent literature for high-dimensional data with the dimension larger than the sample size. For the above testing problems, we show that under suitable sequences of alternatives, the powers of the mean-based tests and the tests based on spatial signs and ranks tend to be same as the data dimension tends to infinity for any sample size when the coordinate variables satisfy appropriate mixing conditions. Further, their limiting powers do not depend on the heaviness of the tails of the distributions. This is in striking contrast to the asymptotic results obtained in the classical multivariate setting. On the other hand, we show that in the presence of stronger dependence among the coordinate variables, the spatial-sign- and rank-based tests for high-dimensional data can be asymptotically more powerful than the mean-based tests if, in addition to the data dimension, the sample size also tends to infinity. The sizes of some mean-based tests for high-dimensional data studied in the recent literature are observed to be significantly different from their nominal levels. This is due to the inadequacy of the asymptotic approximations used for the distributions of those test statistics. However, our asymptotic approximations for the tests based on spatial signs and ranks are observed to work well when the tests are applied on a variety of simulated and real datasets.

1. Introduction. Let $X = \mu_1 + V$ and $Y = \mu_2 + W$ be independent random variables with $E(V) = E(W) = 0$. Tests based on sample means like the t-test for testing the one-sample and the two-sample hypotheses $H_0 : \mu_1 = 0$ and $H_0 : \mu_1 = \mu_2$, respectively, assume that V and W have Gaussian distributions. Nonparametric competitors of the t-test for the same hypotheses that are based on signs and ranks do not require the assumption of Gaussianity and can be carried out if V and W are assumed to have only symmetric distributions. These nonparametric tests have the distribution-free property and they are asymptotically more efficient than the mean-based tests for non-Gaussian distributions having heavy tails. Although various extensions of these nonparametric tests have been proposed for multivariate data [see, e.g., Puri and Sen (1971), Oja (2010) and Hettmansperger and McKean (2011)], they do not have the distribution-free property in general

Received May 2015; revised March 2016.
MSC2010 subject classifications. Primary 62H15, 62G10; secondary 60G10, 62E20.
Key words and phrases. ARMA processes, heavy tailed distributions, permutation tests, ρ-mixing, randomly scaled ρ-mixing, spherical distributions, stationary sequences.

and they are often implemented using their permutation distributions. However, like their univariate counterparts, they are usually asymptotically more efficient than the mean-based Hotelling's T^2 test for multivariate non-Gaussian distributions with heavy tails [see Choi and Marden (1997), Möttönen, Oja and Tienari (1997), Marden (1999) and Oja (2010)].

For high-dimensional data, where the data dimension is larger than the sample size, Hotelling's T^2 test is not applicable due to the singularity of the sample dispersion matrix. Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_m$ and $\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_n$ be i.i.d. copies of independent random vectors $\mathbf{X}$ and $\mathbf{Y}$ in $\mathbb{R}^d$ with $\mu_1 = E(\mathbf{X})$ and $\mu_2 = E(\mathbf{Y})$. For testing $H_0 : \mu_1 = \mu_2$ against the alternative $H_A : \mu_1 \neq \mu_2$ for two high-dimensional observations $\mathbf{X}$ and $\mathbf{Y}$, Bai and Saranadasa (1996) proposed a test based on $\|\overline{\mathbf{X}} - \overline{\mathbf{Y}}\|^2$, where $\overline{\mathbf{X}}$ and $\overline{\mathbf{Y}}$ are the sample means of the two samples. Chen and Qin (2010) proposed a test statistic after removing the terms $\sum_{i=1}^m \|\mathbf{X}_i\|^2$ and $\sum_{j=1}^n \|\mathbf{Y}_j\|^2$ appearing in the expansion of $\|\overline{\mathbf{X}} - \overline{\mathbf{Y}}\|^2$, which makes the resulting statistic an unbiased estimator of $\|E(\mathbf{X} - \mathbf{Y})\|^2$. The one-sample and the two-sample statistics of Chen and Qin (2010) based on sample means are

$$T_{\mathrm{CQ}}^{(1)} = \frac{1}{(m)_2} \sum_{\substack{i_1, i_2 = 1, \\ i_1 \neq i_2}}^{m} \mathbf{X}_{i_1}' \mathbf{X}_{i_2} \quad \text{and} \tag{1.1}$$

$$T_{\mathrm{CQ}}^{(2)} = \frac{1}{(m)_2 (n)_2} \sum_{\substack{i_1, i_2 = 1, \\ i_1 \neq i_2}}^{m} \sum_{\substack{j_1, j_2 = 1, \\ j_1 \neq j_2}}^{n} (\mathbf{X}_{i_1} - \mathbf{Y}_{j_1})' (\mathbf{X}_{i_2} - \mathbf{Y}_{j_2}), \tag{1.2}$$

respectively, where $(p)_q = p(p-1)\cdots(p-q+1)$ for $1 \leq q < p$.

Multivariate spatial-sign- and rank-based tests [see, e.g., Möttönen and Oja (1995), Möttönen, Oja and Tienari (1997), Choi and Marden (1997), Marden (1999) and Oja (2010)] also involve inverses of dispersion matrices computed from the sample which become singular when the data dimension exceeds the sample size. Wang, Peng and Li (2015) proposed a one-sample test of the mean vector based on spatial signs given by

$$T_S = \frac{1}{(m)_2} \sum_{\substack{i_1, i_2 = 1, \\ i_1 \neq i_2}}^{m} S(\mathbf{X}_{i_1})' S(\mathbf{X}_{i_2}),$$

where $S(\mathbf{x}) = \mathbf{x}/\|\mathbf{x}\|$ denotes the spatial sign of any $\mathbf{x} \in \mathbb{R}^d$. A natural high-dimensional version of the one-sample spatial signed rank statistic can be defined using the idea of Wang, Peng and Li (2015), and it is given by

$$T_{\mathrm{SR}} = \frac{1}{(m)_4} \sum_{\substack{i_1, i_2, i_3, i_4 \\ \text{all distinct}}} S(\mathbf{X}_{i_1} + \mathbf{X}_{i_2})' S(\mathbf{X}_{i_3} + \mathbf{X}_{i_4}).$$

Similarly, a two-sample spatial rank statistic can be defined as

$$T_{\mathrm{WMW}} = \frac{1}{(m)_2(n)_2} \sum_{\substack{i_1,i_2=1,\\ i_1\neq i_2}}^{m} \sum_{\substack{j_1,j_2=1,\\ j_1\neq j_2}}^{n} S(\mathbf{Y}_{j_1} - \mathbf{X}_{i_1})' S(\mathbf{Y}_{j_2} - \mathbf{X}_{i_2}).$$

Note that T_S, T_{SR} and T_{WMW} are unbiased estimators of $\|E\{S(\mathbf{X}_1)\}\|^2$, $\|E\{S(\mathbf{X}_1 + \mathbf{X}_2)\}\|^2$ and $\|E\{S(\mathbf{X} - \mathbf{Y})\}\|^2$, respectively.

In this article, we study the behaviours of different one-sample and two-sample tests for mean(s) based on sample means, spatial signs and ranks under various probability models for high-dimensional data. In Section 2, we prove that under appropriate mixing conditions on the coordinate variables and suitable sequences of alternatives, the limiting powers of the spatial rank-based test and the mean-based tests are the same as the data dimension tends to infinity. This is true for all sample sizes and irrespective of the heaviness of the tails of the underlying distributions. Analogous results hold for the one-sample spatial-sign- and signed-rank-based tests and the mean-based tests, and those are presented in Section 2.1. These results are in striking contrast to the asymptotic results obtained in the traditional multivariate setup, where the data dimension is fixed and the sample sizes tend to infinity. In such a setup, the multivariate spatial-sign- and rank-based tests are asymptotically less efficient than Hotelling's T^2 test for Gaussian distributions, and they are more efficient than the T^2 test for non-Gaussian distributions with heavy tails [see Möttönen, Oja and Tienari (1997), Choi and Marden (1997), Marden (1999) and Oja (2010)]. Recall that for multivariate Gaussian data, the Hotelling's T^2 test is actually the likelihood ratio test and the most powerful invariant test. In Section 3, we prove that in the presence of some stronger dependence among the coordinate variables, the limiting powers of the spatial-sign- and rank-based tests can be more than those of their competitors based on sample means if we first let the data dimension and then the sample size tend to infinity. Thus, it follows from Sections 2 and 3 that for a large class of well-known and widely used models for high-dimensional data, the tests based on spatial signs and ranks are either asymptotically as powerful or more powerful compared to some of the mean-based tests studied in the literature. In Section 4, we demonstrate the finite sample performances of these tests using some real datasets. It is found that the above mentioned superiority of the spatial-sign- and rank-based tests over the mean-based tests also holds for those datasets. In Section 5, we discuss the performances of these tests in comparison with some other mean-based tests for high-dimensional data available in the recent literature. It is found that the sizes of some of the mean-based tests are significantly different from their nominal sizes due to the inadequacy of the asymptotic approximations used for the distributions of the corresponding test statistics. The proofs of all the theorems are presented in the Appendix.

2. Asymptotic behaviours of different tests under ρ-mixing. Let $\mathcal{X} = (X_1, X_2, \ldots)$ be an infinite sequence of random variables defined over a probability space $(\Omega, \mathcal{A}, P)$.

DEFINITION 2.1 [Kolmogorov and Rozanov (1960)]. A sequence $\mathcal{X}$ is said to be ρ-mixing if $\rho(r) = \sup_{k\geq 1} \sup\{|\operatorname{Corr}(f, g)| : f \in \mathcal{F}_1^k, g \in \mathcal{F}_{r+k}^{\infty}\}$ converges to zero as $r \to \infty$. Here, $\rho(\cdot)$ is called the ρ-mixing coefficient of $\mathcal{X}$ and $\mathcal{F}_a^b$ denotes the σ-field generated by measurable square integrable functions of $(X_s : a \leq s \leq b, s \in \mathbb{N})$ for $1 \leq a \leq b \leq \infty$.

We refer to Lin and Lu (1996) and Bradley (2005) for further details about ρ-mixing sequences. Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_m$ and $\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_n$ be i.i.d. copies of independent random vectors $\mathbf{X}$ and $\mathbf{Y}$ in $\mathbb{R}^d$. We assume the following conditions.

(C1) $\mathbf{X} = \mu_1 + \mathbf{V}$ and $\mathbf{Y} = \mu_2 + \mathbf{W}$ for some $\mu_1, \mu_2 \in \mathbb{R}^d$, where $\mathbf{V}$ and $\mathbf{W}$ are vectors formed by the first d coordinates of mutually independent zero mean, strictly stationary, and ρ-mixing sequences $\mathcal{V} = (V_1, V_2, \ldots)$ and $\mathcal{W} = (W_1, W_2, \ldots)$ satisfying $E(V_1^4) < \infty$ and $E(W_1^4) < \infty$.

(C2) The ρ-mixing coefficients $\rho_1(\cdot)$ and $\rho_2(\cdot)$ of $\mathcal{V}$ and $\mathcal{W}$ satisfy $\sum_{k=1}^{\infty} \rho_1(2^k) < \infty$ and $\sum_{k=1}^{\infty} \rho_2(2^k) < \infty$, respectively.

Let $\mu = \mu_2 - \mu_1$, $\sigma_1^2 = \operatorname{Var}(X_1) > 0$, $\sigma_2^2 = \operatorname{Var}(Y_1) > 0$, $\Sigma_1 = \operatorname{Disp}(\mathbf{X})$, and $\Sigma_2 = \operatorname{Disp}(\mathbf{Y})$, where $\mathbf{X} = (X_1, X_2, \ldots, X_d)$ and $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_d)$.

(C3) $\|\mu\|^2/d^{1/2+\epsilon} \to 0$ for some $\epsilon > 0$ and $\mu'(\Sigma_1 + \Sigma_2)\mu = o(\operatorname{tr}(\Sigma_1 + \Sigma_2)^2)$ as $d \to \infty$.

Examples of ρ-mixing sequences are m-dependent sequences, stationary ARMA(p, q) processes with a white noise innovation process [see Lin and Lu (1996), Theorem 1.1.2], and hidden Markov models whose underlying generator sequences are stationary, Gaussian and geometrically ergodic Markov chains [see Bradley (2005), Theorem 3.7]. Stationary ARMA processes are well-known models for time series data, and hidden Markov models are used in varied fields like computational biology, econometrics, speech recognition, etc. [see Cappé, Moulines and Rydén (2005) for an exposition]. For all of the above models, condition (C2) holds. Condition (C3) is trivially true under the null hypothesis $H_0 : \mu = \mathbf{0}$. Note that when Σ_1 and Σ_2 are identity matrices, the second part of condition (C3) is automatically true if its first part holds. In general, the second part of condition (C3) holds if in addition to the first part, we have $\lambda_d^{-1} \sum_{k=1}^{d} \lambda_k^2 = O(d^{1/2+\epsilon})$ as $d \to \infty$ for some $\epsilon > 0$, where $\lambda_1 < \lambda_2 < \cdots < \lambda_d$ are the eigenvalues of $\Sigma_1 + \Sigma_2$.

Chen and Qin (2010) worked in a setup where $\mathbf{X}$ and $\mathbf{Y}$ are affine transformations of certain zero mean random vectors whose coordinates are "pseudo-independent" [see (3.2) in page 811 in that paper]. The distributional assumptions in (C1) and (C2) cover many distributions that satisfy the model assumptions stated in (3.1) in Chen and Qin (2010), page 811, for example, distributions with independent coordinates, moving average processes and more generally m-dependent sequences as well as stationary autoregressive processes. Fan and Lin (1998) considered the problem of testing equality of two mean curves for functional data, and they modelled the data as a finite-dimensional one, where the data dimension is larger than the sample size. A class of probability models considered by them are stationary linear Gaussian processes, many of which satisfy the model assumptions considered above. Srivastava, Katayama and Kano (2013) studied a two-sample test based on the sum of squares of the coordinatewise t-statistics and studied its properties assuming multivariate Gaussianity of the data, which includes many distributions satisfying Assumptions (C1) and (C2). A closely related test was proposed by Gregory et al. (2015), and they studied its properties under α-mixing [see Lin and Lu (1996)] conditions on the data, which is weaker than the ρ-mixing setup considered above. However, these authors required the existence of sixteenth-order moments. Cai, Liu and Xia (2014) proposed a mean-based test for detecting sparse alternatives and studied its properties primarily under the assumption of multivariate Gaussianity of the data. Feng et al. (2015) proposed a modification of the test in Srivastava, Katayama and Kano (2013) and they worked in a setup similar to that considered by Chen and Qin (2010). Thus, as in the case of the latter paper, many probability distributions included in the setup considered by Feng et al. (2015) satisfy the ρ-mixing assumptions described here. Wei et al. (2016) studied the properties of their test under spherical Gaussian distributions, which are special cases of the ρ-mixing models considered in this paper.

We now state our main results. To proceed further, define

$$\Gamma_1 = 2\,\mathrm{tr}(\Sigma_1^2)/(m)_2 + 2\,\mathrm{tr}(\Sigma_2^2)/(n)_2 + 4\,\mathrm{tr}(\Sigma_1\Sigma_2)/(mn). \tag{2.1}$$

THEOREM 2.1. *Suppose that conditions* (C1)–(C3) *are satisfied. Then each of* $[d(\sigma_1^2+\sigma_2^2)T_{\mathrm{WMW}} - \|\mu\|^2]/\Gamma_1^{1/2}$ *and* $(T_{\mathrm{CQ}}^{(2)} - \|\mu\|^2)/\Gamma_1^{1/2}$ *converges weakly to a standard Gaussian variable as* $d \to \infty$ *for every fixed* $m, n \geq 2$.

When the null hypothesis $H_0 : \mu = \mathbf{0}$ is true, the above theorem yields the asymptotic null distributions of T_{WMW} and $T_{\mathrm{CQ}}^{(2)}$ as $d \to \infty$ while m and n are held fixed.

Chen and Qin (2010) obtained the asymptotic distribution of $T_{\mathrm{CQ}}^{(2)}$ under a different set of conditions and when both d and n tend to infinity. However, the asymptotic distribution derived by them is the same as that obtained in Theorem 2.1.

When the alternative hypothesis $H_A : \mu \neq \mathbf{0}$ is true, the next theorem compares the asymptotic powers of the tests based on T_{WMW} and $T_{\mathrm{CQ}}^{(2)}$ for high-dimensional data. Let $\beta_{T_{\mathrm{WMW}}}(\mu)$ and $\beta_{T_{\mathrm{CQ}}^{(2)}}(\mu)$ be the powers of these two tests at a given level of significance.

THEOREM 2.2. *Suppose that the conditions* (C1)–(C3) *hold and that* $\lim_{d\to\infty} \|\mu\|^2/\Gamma_1^{1/2} = c$, *for some* $c \in [0,\infty]$. *Then* $\lim_{d\to\infty} \beta_{T_{\mathrm{WMW}}}(\mu) = \lim_{d\to\infty} \beta_{T^{(2)}_{\mathrm{CQ}}}(\mu) = \beta$ *for every fixed* $m, n \geq 2$, *where* $\beta = \alpha$, $\beta = 1$, *or* $\beta \in (\alpha, 1)$ *according as* $c = 0$, $c = \infty$, *or* $c \in (0,\infty)$, *respectively. Here*, α *is the level of significance of the test.*

The above theorem implies that the asymptotic powers of the mean-based and the spatial-rank-based tests are the same as $d \to \infty$ for each fixed $m, n \geq 2$. If Σ_1 and Σ_2 equal the $d \times d$ identity matrix, and d is large, we get different powers of the tests based on T_{WMW} and $T^{(2)}_{\mathrm{CQ}}$ according as $\|\mu\|/d^{1/4}$ converges to zero, infinity or some $c \in (0,\infty)$.

2.1. *Empirical study using some ρ-mixing models*. For implementing the tests based on T_{WMW} and $T^{(2)}_{\mathrm{CQ}}$ under the ρ-mixing setup, we can use their limiting null distributions obtained from Theorem 2.1 after plugging in the following estimators of the parameters involved.

$$\widehat{\Gamma_1} = \frac{2}{(m)_2}\widehat{\mathrm{tr}(\Sigma_1^2)} + \frac{2}{(n)_2}\widehat{\mathrm{tr}(\Sigma_2^2)} + \frac{4}{mn}\widehat{\mathrm{tr}(\Sigma_1\Sigma_2)},$$

where

$$\widehat{\mathrm{tr}(\Sigma_1^2)} = \frac{1}{4(m)_4} \sum_{\substack{i_1,i_2,i_3,i_4\\ \text{all distinct}}} \big[(\mathbf{X}_{i_1} - \mathbf{X}_{i_2})'(\mathbf{X}_{i_3} - \mathbf{X}_{i_4})\big]^2,$$

$$\widehat{\mathrm{tr}(\Sigma_2^2)} = \frac{1}{4(n)_4} \sum_{\substack{j_1,j_2,j_3,j_4\\ \text{all distinct}}} \big[(\mathbf{Y}_{j_1} - \mathbf{Y}_{j_2})'(\mathbf{Y}_{j_3} - \mathbf{Y}_{j_4})\big]^2 \quad \text{and}$$

$$\widehat{\mathrm{tr}(\Sigma_1\Sigma_2)} = \frac{1}{4(m)_2(n)_2} \sum_{i_1\neq i_2}\sum_{j_1\neq j_2} \big[(\mathbf{X}_{i_1} - \mathbf{X}_{i_2})'(\mathbf{Y}_{j_1} - \mathbf{Y}_{j_2})\big]^2.$$

Further, we define $\widehat{\sigma_1^2} = [d(m-1)]^{-1}\sum_{k=1}^{d}\sum_{i=1}^{m}(X_{ik} - \overline{X}_k)^2$, where $\overline{X}_k = d^{-1}\sum_{i=1}^{m} X_{ik}$ with $\mathbf{X}_i = (X_{i1}, X_{i2}, \ldots, X_{id})$, $1 \leq i \leq m$ and $\widehat{\sigma_2^2} = [d(n-1)]^{-1}\sum_{k=1}^{d}\sum_{j=1}^{n}(Y_{jk} - \overline{Y}_k)^2$, where $\overline{Y}_k = d^{-1}\sum_{j=1}^{n} Y_{jk}$ with $\mathbf{Y}_j = (Y_{j1}, Y_{j2}, \ldots, Y_{jd})$, $1 \leq j \leq n$. Note that $\widehat{\Gamma}_1$ is invariant under location transformations and is also computationally inexpensive unlike the estimator proposed by Chen and Qin (2010), page 815. Further, it is a U-statistic for estimating Γ_1 and is an unbiased estimator. Moreover, for all simulated datasets and real datasets considered later, the empirical sizes and powers of the test based on $T^{(2)}_{\mathrm{CQ}}$ implemented as above are similar to those of the original two-sample test in Chen and Qin (2010).

To compare the performances of the tests based on T_{WMW} and $T^{(2)}_{\mathrm{CQ}}$, we have considered the following models. Let $\mathbf{X} = (X_1, X_2, \ldots, X_d)$ be such that the X_k's

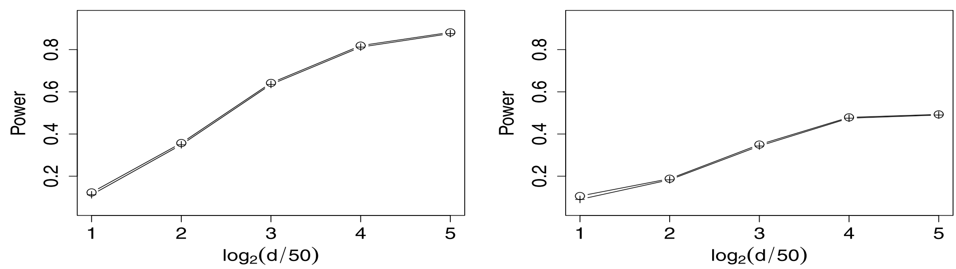

FIG. 1. *Powers of the tests at nominal* 5% *level based on* T_{WMW} (- + - *curves*) *and* $T_{\mathrm{CQ}}^{(2)}$ (- ∘ - *curves*) *for the* AR(1) *model with Gaussian innovation* (*left panel*) *and* $t(5)$ *innovation* (*right panel*). *The two power curves are overlaid on each other in both the plots.*

are consecutive random observations from a stationary AR(1) model with correlation 0.7, i.e., $X_k = 0.7X_{k-1} + \varepsilon_k$. We have considered both Gaussian and $t(5)$ distributions for the innovation process $\{\varepsilon_k\}$. For simulating from the distribution of $\mathbf{X}$, we used the in-built code "arima.sim" in the R software. It generates a stationary time series by allowing the process to evolve over a sufficiently large number of time points after the initialization with random observations from the innovation distribution. Let $\mathbf{Y} = \mu + \widetilde{\mathbf{X}}$, where $\widetilde{\mathbf{X}}$ has the same distribution as that of $\mathbf{X}$, and $\mu = (c, 0, 0, \dots, 0)$ with $c = 1.5, 3, 4.5, 6, 7.5$ for $d = 100, 200, 400, 800, 1600$, respectively. The sample sizes chosen are $m = n = 20$, and the sizes and the powers of the tests based on T_{WMW} and $T_{\mathrm{CQ}}^{(2)}$ are averaged over 1000 Monte Carlo simulations. We found that the sizes of the tests are not significantly different from the nominal 5% level for both of the models. It is seen from Figure 1 that the powers of these two tests are similar for all data dimensions considered under both of the models. The power curves are so close that they are overlaid on each other.

2.2. *Asymptotic behaviours of one-sample tests under ρ-mixing.* Let $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ be i.i.d. copies of a random vector $\mathbf{X} \in \mathbb{R}^d$. The following theorem gives the asymptotic distributions of T_S, T_{SR} and $T_{\mathrm{CQ}}^{(1)}$ and compares their asymptotic powers when the data dimension is large. Denote as $\beta_{T_S}(\mu)$, $\beta_{T_{\mathrm{SR}}}(\mu)$ and $\beta_{T_{\mathrm{CQ}}^{(1)}}(\mu)$ the powers of the tests based on T_S, T_{SR} and $T_{\mathrm{CQ}}^{(1)}$ at a given level of significance when the alternative hypothesis $H_A : \mu \neq \mathbf{0}$ is true. Let us assume the following condition, which is the one-sample version of condition (C3).

(C4) $\|\mu\|^2/d^{1/2+\epsilon} \to 0$ for some $\epsilon > 0$ and $\mu'\Sigma\mu = o(\mathrm{tr}(\Sigma^2))$ as $d \to \infty$, where $\Sigma = \mathrm{Disp}(\mathbf{X})$.

Define $\sigma^2 = \mathrm{Var}(X_1)$, where $\mathbf{X} = (X_1, X_2, \dots, X_d)$, and define

$$\Gamma_2 = 2\,\mathrm{tr}(\Sigma^2)/(n)_2. \tag{2.2}$$

THEOREM 2.3. *Let* $\mathbf{X} = \mu + \mathbf{V}$, *where* $\mathbf{V}$ *is the vector formed by the first* d *coordinates of the infinite sequence* $\mathcal{V}$ *satisfying conditions* (C1) *and* (C2) *and* μ *satisfies condition* (C4).

(a) *Each of* $(d\sigma^2 T_S - \|\mu\|^2)/\Gamma_2^{1/2}$, $(d\sigma^2 T_{\mathrm{SR}} - 2\|\mu\|^2)/(2\Gamma_2^{1/2})$ *and* $(T_{\mathrm{CQ}}^{(1)} - \|\mu\|^2)/\Gamma_2^{1/2}$ *converges weakly to a standard Gaussian variable as* $d \to \infty$ *for every fixed* $n \geq 4$.

(b) *Assume* $\lim_{d\to\infty} \|\mu\|^2/\Gamma_2^{1/2} = c$ *for some* $c \in [0, \infty]$. *Then* $\lim_{d\to\infty} \beta_{T_S}(\mu) = \lim_{d\to\infty} \beta_{T_{\mathrm{SR}}}(\mu) = \lim_{d\to\infty} \beta_{T_{\mathrm{CQ}}^{(1)}}(\mu) = \beta$ *for every fixed* $n \geq 4$, *where* $\beta = \alpha$ *or* $\beta = 1$ *or* $\beta \in (\alpha, 1)$ *according as* $c = 0$ *or* $c = \infty$ *or* $c \in (0, \infty)$.

We get the limiting null distributions of T_S, T_{SR} and $T_{\mathrm{CQ}}^{(1)}$ when $\mu = \mathbf{0}$ in the above theorem. When both the data dimension and the sample size tend to infinity, Wang, Peng and Li (2015) proved that the test based on T_S is asymptotically (when both $n, d \to \infty$) as powerful as the test based on $T_{\mathrm{CQ}}^{(1)}$ for spherical Gaussian distributions, which is a distribution included in our ρ-mixing model. The equality of the asymptotic powers of the tests based on T_S and $T_{\mathrm{CQ}}^{(1)}$ stated in part (b) of our Theorem 2.3 holds for any sample size and for many nonspherical distributions.

REMARK 2.1. In both the one- and the two-sample problems, under the ρ-mixing model, the equality of the limiting powers of the tests based on sample means and the tests based on spatial signs and ranks holds when the data dimension is large. This is true for any sample size and irrespective of whether the coordinate variables have Gaussian or some other heavy tailed distributions.

3. Asymptotic behaviours of different tests under stronger dependence. We now consider another class of probability models for high-dimensional data under which there is stronger dependence among the coordinate variables than what we have considered in the previous section.

DEFINITION 3.1. Consider an infinite sequence $\mathcal{X}$ defined over a probability space $(\Omega, \mathcal{A}, P)$. We say that $\mathcal{X}$ is a randomly scaled ρ-mixing sequence (RSRM sequence, say) if there exist a zero mean ρ-mixing sequence $\mathcal{R}$ and a positive nondegenerate random variable U defined on $(\Omega, \mathcal{A}, P)$ such that $\mathcal{X} = \mathcal{R}/U$.

The RSRM property is satisfied by many important probability models for high-dimensional data. It follows from Theorem 1.31 in Kallenberg (2005) that any rotatable or spherically symmetric sequence $\mathcal{X}$, that is, a sequence for which all finite-dimensional marginals are spherically symmetric, can be viewed as a RSRM sequence. Here, $\mathcal{R}$ can be taken as a sequence of i.i.d. standard Gaussian variables and U as a nonnegative random variable independent of $\mathcal{R}$. An example of a rotatable sequence is the infinite sequence of random variables associated with the multivariate spherical t distribution. More generally, if every finite dimensional marginal of a sequence $\mathcal{X}$ is elliptically symmetric, then $\mathcal{X} = \mathcal{R}/U$ with probability one, where $\mathcal{R}$ is a sequence of zero-mean Gaussian variables, and U is a nonnegative random variable independent of $\mathcal{R}$. In this case, $\mathcal{X}$ has the RSRM property if the Gaussian sequence $\mathcal{R}$ is a ρ-mixing sequence. Let us mention here that Wang, Peng and Li (2015) primarily worked under the setup of elliptically symmetric models, and from the above discussion it follows that this class includes many distributions that have the RSRM property. Cai, Liu and Xia (2014) also considered different classes of non-Gaussian models, and many of them have the RSRM property.

For deriving the asymptotic distributions of T_{WMW} and $T^{(2)}_{\text{CQ}}$ under the RSRM model, we assume the following.

(C5) $\mathbf{X} = \mu_1 + \widetilde{\mathbf{V}}$ and $\mathbf{Y} = \mu_2 + \widetilde{\mathbf{W}}$ for some $\mu_1, \mu_2 \in \mathbb{R}^d$, where $\widetilde{\mathbf{V}}$ and $\widetilde{\mathbf{W}}$ are vectors formed by the first d coordinates of RSRM sequences $\widetilde{\mathcal{V}}$ and $\widetilde{\mathcal{W}}$. Let $\widetilde{\mathbf{V}} = \mathbf{V}/P$ and $\widetilde{\mathbf{W}} = \mathbf{W}/Q$, where $\mathcal{V}$ and $\mathcal{W}$ are independent ρ-mixing sequences satisfying (C1) and (C2) and P and Q are independent positive random variables.

As earlier, let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_m$ and $\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_n$ be i.i.d. copies of independent random vectors $\mathbf{X}$ and $\mathbf{Y}$ in $\mathbb{R}^d$. Then we can write $\mathbf{X}_i = \mu_1 + \mathbf{V}_i/P_i$, $1 \le i \le m$, and $\mathbf{Y}_j = \mu_2 + \mathbf{W}_j/Q_j$, $1 \le j \le n$.

THEOREM 3.1. *Assume that* (C5) *holds and* $\mu = \mu_2 - \mu_1$ *satisfies condition* (C3) *with* Σ_1 *and* Σ_2 *in that condition replaced by* $\operatorname{Disp}(\mathbf{V})$ *and* $\operatorname{Disp}(\mathbf{W})$, *respectively.*

(a) *There exist random variables* S_1, S_2 *and* S_3 *that are functions of the* P_i*'s and the* Q_j*'s such that each of* $(dT_{\text{WMW}} - \|\mu\|^2 S_1)/S_2^{1/2}$ *and* $(T^{(2)}_{\text{CQ}} - \|\mu\|^2)/S_3^{1/2}$ *converges weakly to a standard Gaussian variable as* $d \to \infty$ *for every* $m, n \ge 2$. *Consequently, for every fixed* $m, n \ge 2$, *the distributions of* T_{WMW} *and* $T^{(2)}_{\text{CQ}}$ *can be approximated by location and scale mixtures of Gaussian distributions when the data dimension is large.*

(b) *Assume further that all of* $E(P)$, $E(Q)$, $E(P^{-2})$ *and* $E(Q^{-2})$ *are finite. Suppose that* $\|\mu\|^2/d^{1/2}$ *tends to a finite nonnegative limit* $b^2_{m,n}$ *as* $d \to \infty$, *where* $b^2_{m,n} = o((m+n)^{-1/2})$ *and* $m/(m+n) \to \gamma \in (0,1)$ *as* $m, n \to \infty$. *Then there exist* $\psi_1, \psi_2 \in \mathbb{R}$ *such that* $\lim_{m,n\to\infty} \lim_{d\to\infty} P\{(dT_{\text{WMW}} - \|\mu\|^2\psi_1)/\psi_2^{1/2} \le x\} = \lim_{m,n\to\infty} \lim_{d\to\infty} P\{(T^{(2)}_{\text{CQ}} - \|\mu\|^2)/\Gamma_1^{1/2} \le x\} = \Phi(x)$ *for all* $x \in \mathbb{R}$. *Here,* Φ *is the cumulative distribution function of standard Gaussian distribution, and* Γ_1 *is as defined in Theorem* 2.1.

Unlike the setup considered in Section 2, where the coordinate variables are ρ-mixing, here the distributions of T_{WMW} and $T^{(2)}_{\text{CQ}}$ cannot be approximated by Gaussian distributions when m and n are small even if d is large. However, if the sample sizes are also large in addition to the data dimension, we can approximate the distributions of these statistics by Gaussian distributions. It is easy to see that many probability models with the RSRM property do not satisfy the model assumptions in (3.1) in Chen and Qin (2010). Nevertheless, the asymptotic distribution of $T_{\mathrm{CQ}}^{(2)}$ obtained from part (b) of Theorem 3.1 coincides with that obtained in Theorem 1 in Chen and Qin (2010). Further, it also coincides with the Gaussian distribution obtained under the ρ-mixing model in Theorem 2.1.

Let $\beta_{T_{\mathrm{WMW}}}(\mu)$ and $\beta_{T_{\mathrm{CQ}}^{(2)}}(\mu)$ denote the powers of the tests based on T_{WMW} and $T_{\mathrm{CQ}}^{(2)}$ under the alternative hypothesis $H_A : \mu \neq \mathbf{0}$ at a given level of significance. The next theorem gives a comparison of the asymptotic powers of these tests.

THEOREM 3.2. *Assume that* $\mathbf{Y}$ *has the same distribution as* $\mathbf{X} + \mu$. *Suppose that all the conditions assumed in Theorem* 3.1 *hold. Also, assume that* $\lim_{m,n\to\infty} \lim_{d\to\infty} \|\mu\|^2/\Gamma_1^{1/2} = c$ *for some* $c \in (0, \infty)$. *Then* $\lim_{m,n\to\infty} \lim_{d\to\infty} \beta_{T_{\mathrm{WMW}}}(\mu) > \lim_{m,n\to\infty} \lim_{d\to\infty} \beta_{T_{\mathrm{CQ}}^{(2)}}(\mu)$.

If $\lim_{m,n\to\infty} \lim_{d\to\infty} \|\mu\|^2/\Gamma_1^{1/2}$ equals zero (resp., infinity), then the asymptotic powers of the tests based on T_{WMW} and $T_{\mathrm{CQ}}^{(2)}$ in the setup of Theorem 3.2 coincide, and they are both equal to the nominal level (resp., equal to one). Theorem 3.2 shows that for appropriate sequences of alternatives, the test based on T_{WMW} is more powerful than the test based on $T_{\mathrm{CQ}}^{(2)}$ for a large class of distributions including many spherical non-Gaussian distributions when the data dimension as well as the sample sizes are large. Note that if $\mathbf{X}$ and $\mathbf{Y}$ have spherically symmetric distributions, then the conditions on μ in Theorems 3.1 and 3.2 hold if $\lim_{m,n\to\infty} \lim_{d\to\infty} (m+n)\|\mu\|^2/d^{1/2} = c' \in (0, \infty)$ and $\lim_{m,n\to\infty} m/(m+n) = \gamma \in (0, 1)$.

3.1. *Empirical study using some RSRM models.* The limiting null distribution of T_{WMW} obtainable from Theorem 3.1 cannot be used to implement this test because the parameters appearing in its limiting distribution cannot be estimated from the data. To compare the performances of the tests based on T_{WMW} and $T_{\mathrm{CQ}}^{(2)}$ for data from the spherical $t(5)$ distribution, we implemented these tests using their permutation distributions. Such an implementation has also been used by Wei et al. (2016) for their test. Though it is not possible to implement the test based on T_{WMW} using its true asymptotic distribution in practice, we can do it for a simulation study, where the distributions and the associated parameters are known. On the other hand, since the true asymptotic null distribution of $T_{\mathrm{CQ}}^{(2)}$ for RSRM models coincides with its asymptotic null distribution in the ρ-mixing setup, the implementation of this test can be done in the same way as described in Section 2.1. We have chosen $m = n = 20$ and $\mu = (c, 0, 0, \ldots, 0)$ with $c = 1, 1.5, 2, 2.5, 3$ for $d = 100, 200, 400, 800, 1600$, respectively. Figure 2 shows that the sizes and the

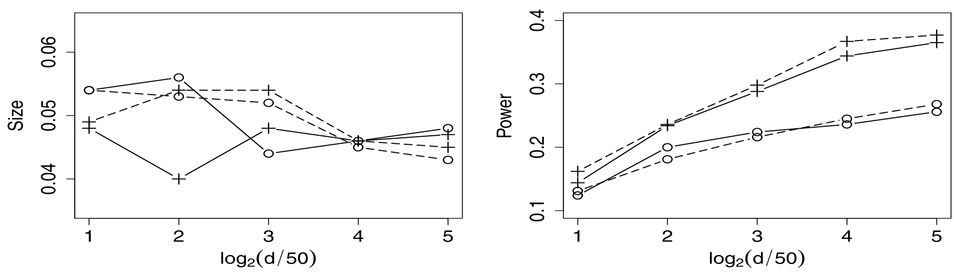

FIG. 2. *Empirical sizes and powers of the tests based on* T_{WMW} $(+)$ *and* $T_{\mathrm{CQ}}^{(2)}$ $(\circ)$ *at nominal* 5% *level for the spherical* $t(5)$ *distribution using the permutation implementation* (*solid curves*) *and the true implementation* (*dashed curves*).

powers of these tests obtained by using the permutation implementation are not significantly different from the sizes and the powers of the tests implemented using their true asymptotic distributions. The permutation distributions of T_{WMW} and $T_{\mathrm{CQ}}^{(2)}$ adequately approximate their true distributions. Also, the test based on T_{WMW} significantly outperforms the test based on $T_{\mathrm{CQ}}^{(2)}$, which conforms with the result in Theorem 3.2.

3.2. *Asymptotic behaviours of one-sample tests under stronger dependence*. We now study the asymptotic distributions of the one-sample tests considered in Section 2.1 under the RSRM model. Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be i.i.d. copies of a random vector $\mathbf{X} \in \mathbb{R}^d$. The following theorem summarizes the asymptotic distributions of T_S, T_{SR} and $T_{\mathrm{CQ}}^{(1)}$ and yields their asymptotic powers. As earlier, we can write $\mathbf{X}_i = \mu + \mathbf{V}_i/P_i$, $1 \le i \le n$. Also, $\beta_{T_S}(\mu)$, $\beta_{T_{\mathrm{SR}}}(\mu)$ and $\beta_{T_{\mathrm{CQ}}^{(1)}}(\mu)$ denote the powers of the tests based on T_S, T_{SR} and $T_{\mathrm{CQ}}^{(1)}$ at a given level of significance when the alternative hypothesis $H_A : \mu \neq \mathbf{0}$ is true.

THEOREM 3.3. *Let* $\mathbf{X} = \mu + \widetilde{\mathbf{V}}$, *where* $\widetilde{\mathbf{V}}$ *is the vector formed by the first* d *coordinates of the sequence* $\widetilde{\mathcal{V}}$ *satisfying condition* (C5), *and* μ *satisfies condition* (C4) *with* Σ *in that condition replaced by* $\mathrm{Disp}(\mathbf{V})$.

(a) *There exist* $\Gamma_3 > 0$ *and random variables* Z_k, $1 \le k \le 4$, *which are functions of the* P_i*'s, such that each of* $(dT_S - \|\mu\|^2 Z_1)/\Gamma_3^{1/2}$, $(dT_{\mathrm{SR}} - 2\|\mu\|^2 Z_2)/(2Z_3^{1/2})$ *and* $(T_{\mathrm{CQ}}^{(1)} - \|\mu\|^2)/Z_4^{1/2}$ *converges weakly to a standard Gaussian variable as* $d \to \infty$ *for each* $n \ge 4$. *Consequently, for each fixed* $n \ge 4$, *the distributions of* T_S, T_{SR} *and* $T_{\mathrm{CQ}}^{(1)}$ *are given by location and scale mixtures of Gaussian distributions when the data dimension is large*.

(b) *Also, assume that both* $E(P)$ *and* $E(P^{-2})$ *are finite and* $\|\mu\|^2/d^{1/2}$ *tends to a finite nonnegative limit* c_n^2 *as* $d \to \infty$, *where* $c_n^2 = o(n^{-1/2})$ *as* $n \to \infty$. *There exist* $\theta_1, \theta_2, \theta_3 \in \mathbb{R}$ *such that* $\lim_{n\to\infty} \lim_{d\to\infty} P\{(d\sigma^2 T_S -$

$\|\mu\|^2\theta_1)/\Gamma_2^{1/2} \le x\} = \lim_{n\to\infty}\lim_{d\to\infty} P\{(d\sigma^2 T_{\mathrm{SR}} - 2\|\mu\|^2\theta_2)/(2\theta_3^{1/2}) \le x\} = \lim_{n\to\infty}\lim_{d\to\infty} P\{(T_{\mathrm{CQ}}^{(1)} - \|\mu\|^2)/\Gamma_2^{1/2} \le x\} = \Phi(x)$ *for all* $x \in \mathbb{R}$. *Here,* $\sigma^2 = \mathrm{Var}(X_1)$, Φ *denotes the cumulative distribution function of a standard Gaussian distribution, and* Γ_2 *is as defined in Theorem* 2.3.

(c) *Further, if we let* $\lim_{n\to\infty}\lim_{d\to\infty}\|\mu\|^2/\Gamma_2^{1/2} = c$, *where* $c \in (0,\infty)$, *we have* $\lim_{n\to\infty}\lim_{d\to\infty}\beta_{T_S}(\mu) > \lim_{n\to\infty}\lim_{d\to\infty}\beta_{T_{\mathrm{CQ}}^{(1)}}(\mu)$. *We also have* $\lim_{n\to\infty}\lim_{d\to\infty}\beta_{T_{\mathrm{SR}}}(\mu) > \lim_{n\to\infty}\lim_{d\to\infty}\beta_{T_{\mathrm{CQ}}^{(1)}}(\mu)$.

It is seen from the proof of part (a) of Theorem 3.3 that if $E(P^{-2}) < \infty$, we have $\Gamma_3 = \sigma^{-4}\Gamma_2$. In this case, we get the same limiting null distributions of T_S from parts (a) and (b), that is, its limiting null distribution is Gaussian irrespective of whether the sample size tends to infinity or not. Further, this limiting null distribution under the RSRM model is the same as that obtained under the ρ-mixing model in part (a) of Theorem 3.1. This is because the spatial sign $S(\mathbf{x}) = \mathbf{x}/\|\mathbf{x}\|$, and thus T_S, remain invariant under homogeneous positive scale transformations of the coordinate variables.

Note that the asymptotic distribution of $T_{\mathrm{CQ}}^{(1)}$ is the same as that obtained in Theorem 3.2 under the ρ-mixing setup and it coincides with the asymptotic distribution of $T_{\mathrm{CQ}}^{(1)}$ obtained by Chen and Qin (2010). For the spherical t distribution, which is a distribution included in our RSRM models, Wang, Peng and Li (2015) derived the asymptotic distribution of $T_{\mathrm{CQ}}^{(1)}$ and proved that the test based on T_S is asymptotically more powerful than the former test. In the setup of Theorem 3.3, if $\lim_{n\to\infty}\lim_{d\to\infty}\|\mu\|^2/\Gamma_2^{1/2}$ equals zero (resp., infinity), then the asymptotic powers of the tests based on T_S, T_{SR} and $T_{\mathrm{CQ}}^{(1)}$ coincide and they are all equal to the nominal level (resp., equal to one).

REMARK 3.1. Suppose that in a two-sample problem, $\mathbf{Y}$ is distributed as $\mathbf{X} + \mu$, where $\mathbf{X}$ is the vector formed by the first d coordinates of a zero mean spherically symmetric or rotatable infinite sequence $\mathcal{X}$. Then it follows from Theorem 1.31 in Kallenberg (2005) that $\mathbf{X} = \mathbf{V}/P$, where $\mathbf{V}$ is a standard spherical Gaussian vector and P is a nonnegative random variable independent of $\mathbf{V}$. Suppose that $\lim_{m,n\to\infty}\lim_{d\to\infty}(m+n)\|\mu\|^2/d^{1/2} = c' \in (0,\infty)$ and $\lim_{m,n\to\infty} m/(m+n) = \gamma \in (0,1)$. Also, assume that both $E(P)$ and $E(P^{-2})$ are finite and positive. Then it follows from Theorems 2.2 and 3.2 that the test based on T_{WMW} is asymptotically at least as powerful as the test based on $T_{\mathrm{CQ}}^{(2)}$ if we first let the dimension and then the sample sizes tend to infinity. Further, their asymptotic powers are equal if and only if $\mathbf{X}$ has a spherical Gaussian distribution. In fact, in this case, their asymptotic powers are the same for any $m, n \ge 2$ if only the dimension tends to infinity.

REMARK 3.2. Suppose that in a one-sample problem, we have $\mathbf{X} = \mu + \widetilde{\mathbf{V}}$, where $\widetilde{\mathbf{V}}$ is the vector formed by the first d coordinates of a spherically symmetric infinite sequence. Assume that $\lim_{n\to\infty} \lim_{d\to\infty} n\|\mu\|^2/d^{1/2} = c' \in (0, \infty)$. Also, let both $E(P)$ and $E(P^{-2})$ be finite. Then it follows from Theorems 2.3 and 3.3 that the tests based on T_S and T_{SR} are asymptotically at least as powerful as the test based on $T_{\mathrm{CQ}}^{(1)}$ if we first let the dimension and then the sample size tend to infinity. Further, the asymptotic powers of all three tests are equal if and only if the distribution of $\mathbf{X}$ is spherical Gaussian. In fact, in this case, their asymptotic powers are the same for any $n \geq 4$ if only the dimension tends to infinity.

4. Analysis of real data. In Sections 2 and 3, we observed that the spatial-sign and rank-based tests are at least as powerful as the mean-based tests for a large class of models. We now investigate whether such a phenomenon occurs when we analyze real data. For that, we implement different tests on some real benchmark datasets in order to see their performances. Two datasets are obtained from http://www.cs.ucr.edu/eamonn/time_series_data, and the first of them is the ECG Data, which contains 69 normal ECG curves and 31 ECG curves of patients with a particular heart disease, and each curve is measured at 96 time points. The second data is the Gun Data, which contains the readings along the horizontal axis of the centroid of the right-hand during two action sequences, namely, gun-draw and gun-point with 24 samples and 26 samples, respectively. Each action sequence is recorded at 150 time points. The third data is the Colon Data obtained from http://datam.i2r.a-star.edu.sg/datasets/krbd/ColonTumor/ColonTumor.zip. This dataset contains the expression levels of 2000 genes from 40 tumor tissue and 22 normal tissue. The fourth data is the Sonar Data obtained from http://archive.ics.uci.edu/ml/datasets.html, which contains sonar signals emitted from 111 metal cylinder samples and 97 rock samples, and each signal is recorded at 60 wavelengths.

The tests based on T_{WMW} and $T_{\mathrm{CQ}}^{(2)}$ will be implemented first using the critical values as obtained in the ρ-mixing setup described in Section 2. Since we do not know the underlying probability distributions for the real datasets, we also implement them using the permutation distributions of these test statistics. Further, unlike what we have done in our simulation studies, where the underlying probability distributions were known, we use the following subsampling technique to generate random samples from a given real dataset to evaluate the sizes and the powers of the tests. To estimate the sizes of the tests based on T_{WMW} and $T_{\mathrm{CQ}}^{(2)}$ for each data, we selected two random subsamples 1000 times from one class in that data and computed the proportion of rejections for each test. The same procedure is now repeated for the other class and the two values obtained for each test are averaged. For evaluating the powers of these tests, we selected 1000 random subsamples each from the two classes and computed the proportions of rejections for the tests. The size of each subsample is 20%, 40%, 40% and 20% of the original sample size for the ECG Data, the Gun Data, the Colon Data and the Sonar Data,

TABLE 1
Sizes and powers of the tests based on T_{WMW} and $T_{\mathrm{CQ}}^{(2)}$ at nominal 5% level for some real data

Data ⟶	**ECG**		**Gun**		**Colon**		**Sonar**	
	Size	**Power**	**Size**	**Power**	**Size**	**Power**	**Size**	**Power**
				Implementation as in the ρ-mixing setup				
T_{WMW}	0.052	0.593	0.052	0.501	0.056	0.747	0.036	0.507
$T_{\mathrm{CQ}}^{(2)}$	0.063	0.601	0.058	0.500	0.063	0.641	0.058	0.432
				Permutation implementation				
T_{WMW}	0.057	0.643	0.055	0.472	0.055	0.723	0.043	0.519
$T_{\mathrm{CQ}}^{(2)}$	0.057	0.624	0.052	0.442	0.060	0.596	0.038	0.360

respectively. These choices are made to ensure that the resulting datasets remain high-dimensional and that the powers of the tests are neither too close to the nominal 5% level nor to one. For computing the permutation distributions of the test statistics, we have used 500 random permutations of the two subsamples.

Table 1 shows that the sizes as well as the powers of the tests for the two implementations are not significantly different. However, the permutation implementation required almost ten times more computing time. Moreover, the sizes of the tests are close to the nominal 5% level for all four datasets. Further, the powers of the tests based on T_{WMW} and $T_{\mathrm{CQ}}^{(2)}$ are not significantly different for the ECG data and the Gun data. However, the test based on T_{WMW} is significantly more powerful than the test based on $T_{\mathrm{CQ}}^{(2)}$ for the Colon data and the Sonar data. Hence, the behaviours of the tests when applied to the real datasets are similar to their behaviours in the simulation studies in Sections 2.1 and 3.1 despite the fact that here we do not know whether the mixing or the RSRM models are valid. We are not much concerned about the validity of these models for the real datasets. In fact, there is some evidence that the ECG data and the Colon data may not satisfy either of these models. This indicates that these models are only sufficient but not necessary for the behaviours of the tests that we have observed in the asymptotic analyses and the simulation studies in Sections 2 and 3.

5. Concluding remarks and discussion. We now consider the performances of some other mean-based tests studied in the literature and discussed in Section 2 on some simulated datasets. We denote the test statistics associated with the tests in Srivastava, Katayama and Kano (2013) and Gregory et al. (2015) by T_{SKK} and T_{GCBL}, respectively. All the sizes and the powers referred to in this section are the empirical sizes and the empirical powers for the different tests under the various models. For the AR(1) models in Section 2.1, we found that the size of the test based on T_{SKK} increases with d and becomes significantly larger than the nominal

5% level for $d \geq 400$. Feng et al. (2015) proved that this test rejects the null hypothesis with probability one as the dimension and the sample sizes tend to infinity at a certain rate for a class of models which includes these AR(1) models. Under the spherical $t(5)$ model in Section 3.1, the size of the test based on T_{SKK} is significantly less than the nominal level for all values of d considered and decreases to zero as d increases. The size of the test based on T_{GCBL} is significantly larger than the nominal level for all values of d considered under the AR(1) models as well as the spherical $t(5)$ model. It seems that the estimates of the critical values for the tests based on T_{SKK} and T_{GCBL} are adversely affected if the sample size is much smaller than the dimension as in our simulation study. On the other hand, we found that permutation implementations of these tests correct their sizes under all of the above models. Even then, these tests are significantly less powerful than the test based on T_{WMW} (resp., $T_{\text{CQ}}^{(2)}$) under all the above models [resp., AR(1) models], but they outperform the test based on $T_{\text{CQ}}^{(2)}$ under the spherical $t(5)$ model. The readers are referred to the Supplementary Material [Chakraborty and Chaudhuri (2016)] for more details.

Cai, Liu and Xia (2014) showed that their test has better power than other tests based on sums of squares of coordinatewise mean differences or coordinatewise t-statistics when the mean shift has only a few nonzero coordinates. However, we observed that this test becomes significantly less powerful than the tests based on T_{WMW} and $T_{\text{CQ}}^{(2)}$, when the mean shifts in the models considered in Sections 2.1 and 3.1 are distributed equally among all the coordinates. Moreover, the size of the test in Cai, Liu and Xia (2014) increases with d and becomes significantly larger than the nominal level for $d \geq 400$ under all of the above models. It seems that the asymptotic extreme value distribution of this statistic is not adequate if the data dimension is much larger than the sample size. Since the test in Cai, Liu and Xia (2014) involves a computationally intensive optimization involving sample dispersion matrices, we could not implement this test using the permutation approach. The detailed results of the simulation study are provided in the Supplementary Material [Chakraborty and Chaudhuri (2016)].

Multivariate Gaussian distributions with dispersion matrices of the form $(1 - \beta)I_d + \beta \mathbf{1}_d \mathbf{1}_d'$ for some $\beta \in (0, 1)$, where $\mathbf{1}_d$ denotes the d-dimensional vector of one's, are neither ρ-mixing nor have the RSRM property. Recently, Katayama and Kano (2014) mentioned that for such probability models for high-dimensional data, the size of the test based on $T_{\text{CQ}}^{(2)}$ would be asymptotically incorrect. To compare the performances of the tests based on T_{WMW} and $T_{\text{CQ}}^{(2)}$ for such models, we have chosen $\beta = 0.7$, $m = n = 20$ and used the permutation implementations of these tests. The mean shifts chosen are $\mu = (c, 0, 0, \ldots, 0)$ with $c = 2.5, 5, 7.5, 10, 12.5$ for $d = 100, 200, 400, 800, 1600$, respectively. We found that the test based on T_{WMW} significantly outperforms the test based on $T_{\text{CQ}}^{(2)}$ for all values of d (see the Supplementary Material [Chakraborty and Chaudhuri (2016)]).

APPENDIX: MATHEMATICAL DETAILS

PROOF OF THEOREM 2.1. Without any loss of generality, we can take $E(\mathbf{X}_1) = \mathbf{0}$. Let us write $\mathbf{X}_i = (X_{i1}, X_{i2}, \dots, X_{id})'$, $1 \le i \le m$, and $\mathbf{Y}_j = (Y_{j1}, Y_{j2}, \dots, Y_{jd})'$, $1 \le j \le n$. First note that

$$\begin{aligned} &\|\mathbf{X} - \mathbf{Y}\|^2 \\ &\quad = \bigl[\|\mathbf{X}\|^2 + \|\mathbf{Y} - \mu\|^2 - 2\mathbf{X}'(\mathbf{Y} - \mu) + 2\mu'(\mathbf{X} - \mathbf{Y} + \mu) + \|\mu\|^2\bigr] \\ &\quad = \sum_{k=1}^{d}\bigl[V_k^2 + W_k^2 - 2V_k W_k\bigr] + 2\mu'(\mathbf{V} - \mathbf{W}) + \|\mu\|^2. \end{aligned} \tag{A.1}$$

It follows from Bradley [(2005), Theorem 5.2(b)], that for any function $h : \mathbb{R}^2 \to \mathbb{R}$, the sequence $(h(V_k, W_k) : k \ge 1)$ is ρ-mixing with its mixing coefficient bounded by $\max\{\rho_1(\cdot), \rho_2(\cdot)\}$. This fact combined with (A.1) along with Assumptions (C1)–(C3) and Theorem 8.2.2 in Lin and Lu (1996) imply that for any given $\epsilon \in (0, 1/2)$, we have

$$\|\mathbf{X} - \mathbf{Y}\|^2/d - \bigl(\sigma_1^2 + \sigma_2^2\bigr) = o\bigl(d^{-1/2+\epsilon}\bigr) \tag{A.2}$$

as $d \to \infty$ *almost surely*. Now,

$$\begin{aligned} T_{\mathrm{WMW}} &= \frac{1}{(m)_2(n)_2}\sum_{i_1 \ne i_2}\sum_{j_1 \ne j_2}\frac{(\mathbf{X}_{i_1} - \mathbf{Y}_{j_1})'(\mathbf{X}_{i_2} - \mathbf{Y}_{j_2})}{d(\sigma_1^2 + \sigma_2^2)} \\ &\quad + \frac{1}{(m)_2(n)_2}\sum_{i_1 \ne i_2}\sum_{j_1 \ne j_2}\biggl[\frac{(\mathbf{X}_{i_1} - \mathbf{Y}_{j_1})'(\mathbf{X}_{i_2} - \mathbf{Y}_{j_2})}{d(\sigma_1^2 + \sigma_2^2)} \\ &\quad \times \biggl\{\frac{d(\sigma_1^2 + \sigma_2^2)}{\|\mathbf{X}_{i_1} - \mathbf{Y}_{j_1}\|\|\mathbf{X}_{i_2} - \mathbf{Y}_{j_2}\|} - 1\biggr\}\biggr] \\ &= \bigl(T_{\mathrm{CQ}}^{(2)} + T_{\mathrm{WMW}}^{(2)}\bigr)/\bigl\{d\bigl(\sigma_1^2 + \sigma_2^2\bigr)\bigr\}, \end{aligned} \tag{A.3}$$

where $T_{\mathrm{CQ}}^{(2)}$ is given by (1.2), and

$$\begin{aligned} T_{\mathrm{WMW}}^{(2)} &= \frac{1}{(m)_2(n)_2}\sum_{i_1 \ne i_2}\sum_{j_1 \ne j_2}\biggl[\frac{(\mathbf{X}_{i_1} - \mathbf{Y}_{j_1})'(\mathbf{X}_{i_2} - \mathbf{Y}_{j_2})}{d(\sigma_1^2 + \sigma_2^2)} \\ &\quad \times \biggl\{\frac{d(\sigma_1^2 + \sigma_2^2)}{\|\mathbf{X}_{i_1} - \mathbf{Y}_{j_1}\|\|\mathbf{X}_{i_2} - \mathbf{Y}_{j_2}\|} - 1\biggr\}\biggr]. \end{aligned}$$

So, $E(T_{\mathrm{CQ}}^{(2)}) = \|\mu\|^2$. Further, it follows from Chen and Qin (2010), page 825, that $\mathrm{Var}(T_{\mathrm{CQ}}^{(2)}) = \Gamma_1 + 4\mu'\Sigma_1\mu/m + 4\mu'\Sigma_2\mu/n$, where Γ_1 is given by (2.1). Note that $(\mu'\Sigma_1\mu/m) + (\mu'\Sigma_2\mu/n) \le \mu'(\Sigma_1 + \Sigma_2)\mu/\min(m, n)$. Further,

$$\Gamma_1 \ge 2\,\mathrm{tr}\bigl[(\Sigma_1 + \Sigma_2)^2\bigr]/\bigl(\max(m, n)\bigr)_2. \tag{A.4}$$

These facts and Assumption (C3) imply that $\mathrm{Var}(T_{\mathrm{CQ}}^{(2)}) = \Gamma_1(1 + o(1))$ as $d \to \infty$. Further,

$$\begin{aligned}
&\big(T_{\mathrm{CQ}}^{(2)} - \|\mu\|^2\big)\\
&\quad = \frac{1}{(m)_2(n)_2}\sum_{i_1\neq i_2}\sum_{j_1\neq j_2}(\mathbf{X}_{i_1} - \mathbf{Y}_{j_1} + \mu)'(\mathbf{X}_{i_2} - \mathbf{Y}_{j_2} + \mu)\\
&\qquad - \frac{2}{mn}\sum_{i,j}\mu'(\mathbf{X}_i - \mathbf{Y}_j + \mu) = T_1 - T_2,
\end{aligned}$$

where $T_1 = [(m)_2(n)_2]^{-1}\sum_{i_1\neq i_2}\sum_{j_1\neq j_2}(\mathbf{X}_{i_1} - \mathbf{Y}_{j_1} + \mu)'(\mathbf{X}_{i_2} - \mathbf{Y}_{j_2} + \mu)$ and $T_2 = 2(mn)^{-1}\sum_{i,j}\mu'(\mathbf{X}_i - \mathbf{Y}_j + \mu)$. It is easy to verify that $E(T_2) = 0$ and $\mathrm{Var}(T_2) = 4\mu'[(\Sigma_1/m) + (\Sigma_2/n)]\mu$. So, using (A.4), Assumption (C3) and Chebyshev's inequality, it follows that $T_2/\Gamma_1^{1/2}$ converges to zero *in probability* as $d \to \infty$. Note that

$$T_1 = \frac{1}{(m)_2(n)_2}\sum_{k=1}^{d}\bigg\{\sum_{i_1\neq i_2}\sum_{j_1\neq j_2}(V_{i_1k} - W_{j_1k})(V_{i_2k} - W_{j_2k})\bigg\}.$$

So, $E(T_1) = 0$ and $\mathrm{Var}(T_1) = \Gamma_1$. This follows from computations similar to those used in deriving $\mathrm{Var}(T_{\mathrm{CQ}}^{(2)})$ earlier. Thus, by Theorem 4.0.1 in Lin and Lu (1996) and Assumptions (C1) and (C2), we have the *weak convergence* of $T_1/\Gamma_1^{1/2}$ to a standard Gaussian distribution as $d \to \infty$ for each fixed $m, n \geq 2$. This and the fact that $T_2^{(2)}$ converges to zero *in probability* as $d \to \infty$ for each fixed $m, n \geq 2$ together imply that

$$\big(T_{\mathrm{CQ}}^{(2)} - \|\mu\|^2\big)/\Gamma_1^{1/2} \xrightarrow{\mathcal{L}} N(0,1) \tag{A.5}$$

as $d \to \infty$ for each fixed $m, n \geq 2$. Next, let us write

$$\begin{aligned}
&T_{\mathrm{WMW}}^{(2)}/\Gamma_1^{1/2}\\
&\quad = \frac{1}{(m)_2(n)_2}\sum_{i_1\neq i_2}\sum_{j_1\neq j_2}\bigg[\frac{(\mathbf{X}_{i_1} - \mathbf{Y}_{j_1})'(\mathbf{X}_{i_2} - \mathbf{Y}_{j_2}) - \|\mu\|^2}{\Gamma_1^{1/2}}\\
&\qquad \times\bigg\{\frac{d(\sigma_1^2 + \sigma_2^2)}{\|\mathbf{X}_{i_1} - \mathbf{Y}_{j_1}\|\|\mathbf{X}_{i_2} - \mathbf{Y}_{j_2}\|} - 1\bigg\}\bigg]\\
&\qquad + \frac{\|\mu\|^2}{(m)_2(n)_2\Gamma_1^{1/2}}\sum_{i_1\neq i_2}\sum_{j_1\neq j_2}\bigg\{\frac{d(\sigma_1^2 + \sigma_2^2)}{\|\mathbf{X}_{i_1} - \mathbf{Y}_{j_1}\|\|\mathbf{X}_{i_2} - \mathbf{Y}_{j_2}\|} - 1\bigg\}\\
&\quad = T_{\mathrm{WMW}}^{(3)} + T_{\mathrm{WMW}}^{(4)},
\end{aligned} \tag{A.6}$$

where

$$T_{\mathrm{WMW}}^{(3)} = \frac{1}{(m)_2(n)_2} \sum_{i_1 \neq i_2} \sum_{j_1 \neq j_2} \left[\frac{(\mathbf{X}_{i_1} - \mathbf{Y}_{j_1})'(\mathbf{X}_{i_2} - \mathbf{Y}_{j_2}) - \|\mu\|^2}{\Gamma_1^{1/2}} \times \left\{ \frac{d(\sigma_1^2 + \sigma_2^2)}{\|\mathbf{X}_{i_1} - \mathbf{Y}_{j_1}\| \|\mathbf{X}_{i_2} - \mathbf{Y}_{j_2}\|} - 1 \right\} \right]$$

and

$$T_{\mathrm{WMW}}^{(4)} = \frac{\|\mu\|^2}{(m)_2(n)_2\Gamma_1^{1/2}} \sum_{i_1 \neq i_2} \sum_{j_1 \neq j_2} \left\{ \frac{d(\sigma_1^2 + \sigma_2^2)}{\|\mathbf{X}_{i_1} - \mathbf{Y}_{j_1}\| \|\mathbf{X}_{i_2} - \mathbf{Y}_{j_2}\|} - 1 \right\}.$$

From the stationarity of the sequences $\mathcal{X}$ and $\mathcal{Y}$ and using the Cauchy–Schwarz inequality, it follows that $\mathrm{tr}[(\Sigma_1 + \Sigma_2)^2] \geq d(\sigma_1^2 + \sigma_2^2)^2$. This fact along with (A.4), (A.2) and Assumption (C3) imply that $T_{\mathrm{WMW}}^{(4)}$ converges to zero *in probability* as $d \to \infty$ for each fixed $m, n \geq 2$.

Next, fix any $i_1 \neq i_2$ and $j_1 \neq j_2$ and consider the corresponding term inside the double summation appearing in the definition of $T_{\mathrm{WMW}}^{(3)}$. It follows from (A.2) that $d(\sigma_1^2 + \sigma_2^2)/\{\|\mathbf{X}_{i_1} - \mathbf{Y}_{j_1}\| \|\mathbf{X}_{i_2} - \mathbf{Y}_{j_2}\|\} - 1$ converges to zero *in probability* as $d \to \infty$. Also, note that

$$\begin{aligned} &(\mathbf{X}_{i_1} - \mathbf{Y}_{j_1})'(\mathbf{X}_{i_2} - \mathbf{Y}_{j_2}) - \|\mu\|^2 \\ &\quad = (\mathbf{X}_{i_1} - \mathbf{Y}_{j_1} + \mu)'(\mathbf{X}_{i_2} - \mathbf{Y}_{j_2} + \mu) \\ &\qquad - \mu'(\mathbf{X}_{i_1} - \mathbf{Y}_{j_1} + \mu) - \mu'(\mathbf{X}_{i_2} - \mathbf{Y}_{j_2} + \mu). \end{aligned} \tag{A.7}$$

Using arguments similar to those used to prove the asymptotic normality of $T_1^{(2)}$ and using Theorem 4.0.1 in Lin and Lu (1996), it follows that the first term in the right-hand side of (A.7) is asymptotically Gaussian with zero mean and variance $2\,\mathrm{tr}[(\Sigma_1 + \Sigma_2)^2]$ as $d \to \infty$. Using Assumption (C3) and Chebyshev's inequality, it follows that the second and the third terms in the right-hand side of (A.7) after dividing by $\Gamma_1^{1/2}$ converge to zero *in probability* as $d \to \infty$. So, the left-hand side of (A.7) after dividing by $\Gamma_1^{1/2}$ converges *weakly* to a Gaussian distribution as $d \to \infty$. Thus, $T_{\mathrm{WMW}}^{(3)}$ converges to zero *in probability* as $d \to \infty$ for each fixed $m, n \geq 2$. This and the fact that $T_{\mathrm{WMW}}^{(4)}$ converges to zero *in probability* as $d \to \infty$ together imply that $T_{\mathrm{WMW}}^{(2)}/\Gamma_1^{1/2}$ converges to zero *in probability* as $d \to \infty$ for each fixed $m, n \geq 2$. Combining this fact with (A.3) and (A.5) yields

$$\{d(\sigma_1^2 + \sigma_2^2)T_{\mathrm{WMW}} - \|\mu\|^2\}/\Gamma_1^{1/2} \xrightarrow{\mathcal{L}} N(0, 1) \tag{A.8}$$

as $d \to \infty$ for each fixed $m, n \geq 2$. □

PROOF OF THEOREM 2.2. Let ζ_α be the $(1-\alpha)$-quantile of the standard Gaussian distribution. Note that

$$\begin{aligned}\beta_{T_{\mathrm{WMW}}}(\mu) &= P\{d(\sigma_1^2+\sigma_2^2)T_{\mathrm{WMW}}/\Gamma_1^{1/2} > \zeta_\alpha\}\\ &= P\{[d(\sigma_1^2+\sigma_2^2)T_{\mathrm{WMW}} - \|\mu\|^2]/\Gamma_1^{1/2} > \zeta_\alpha - \|\mu\|^2/\Gamma_1^{1/2}\}\end{aligned}$$

and

$$\beta_{T_{\mathrm{CQ}}^{(2)}}(\mu) = P\{T_{\mathrm{CQ}}^{(2)}/\Gamma_1^{1/2} > \zeta_\alpha\} = P\{(T_{\mathrm{CQ}}^{(2)} - \|\mu\|^2)/\Gamma_1^{1/2} > \zeta_\alpha - \|\mu\|^2/\Gamma_1^{1/2}\},$$

where the probabilities are computed under the alternative hypothesis. Since $\lim_{d\to\infty}\|\mu\|^2/\Gamma_1^{1/2}$ exists, the equality of the asymptotic powers of the tests based on T_{WMW} and $T_{\mathrm{CQ}}^{(2)}$ follows from (A.5) and (A.8). Moreover, their common value is $\Phi(-\zeta_\alpha + \lim_{d\to\infty}\|\mu\|^2/\Gamma_1^{1/2}) = \Phi(-\zeta_\alpha + c)$, which follows from the expressions of their powers and their asymptotic Gaussian distributions proved in Theorem 2.1. The last part of the present theorem now follows easily. $\square$

PROOF OF THEOREM 2.3. (a) We will derive the asymptotic distribution of T_{SR} and $T_{\mathrm{CQ}}^{(1)}$ only since the derivation of that of T_S is simpler and follows from similar arguments. Using the assumptions in the theorem and the arguments similar to those in the proof of Theorem 2.1, we have

$$\begin{aligned}T_{\mathrm{SR}} &= \frac{1}{(n)_4}\sum_{\substack{i_1,i_2,i_3,i_4\\ \text{all distinct}}}\frac{(\mathbf{X}_{i_1}+\mathbf{X}_{i_2})'(\mathbf{X}_{i_3}+\mathbf{X}_{i_4})}{2d\sigma^2}\\ &\quad + \frac{1}{(n)_4}\sum_{\substack{i_1,i_2,i_3,i_4\\ \text{all distinct}}}\biggl[\frac{(\mathbf{X}_{i_1}+\mathbf{X}_{i_2})'(\mathbf{X}_{i_3}+\mathbf{X}_{i_4})}{2d\sigma^2}\\ &\quad \times\biggl\{\frac{2d\sigma^2}{\|\mathbf{X}_{i_1}+\mathbf{X}_{i_2}\|\|\mathbf{X}_{i_3}+\mathbf{X}_4\|}-1\biggr\}\biggr]\\ &= \frac{2}{(n)_2}\sum_{i_1\neq i_2}\frac{\mathbf{X}_{i_1}'\mathbf{X}_{i_2}}{d\sigma^2} + \frac{1}{(n)_4}\sum_{\substack{i_1,i_2,i_3,i_4\\ \text{all distinct}}}\biggl[\frac{(\mathbf{X}_{i_1}+\mathbf{X}_{i_2})'(\mathbf{X}_{i_3}+\mathbf{X}_{i_4})}{2d\sigma^2}\\ &\quad \times\biggl\{\frac{2d\sigma^2}{\|\mathbf{X}_{i_1}+\mathbf{X}_{i_2}\|\|\mathbf{X}_{i_3}+\mathbf{X}_4\|}-1\biggr\}\biggr].\end{aligned}\tag{A.9}$$

It follows from (1.1) that the first term in (A.9) equals $2T_{\mathrm{CQ}}^{(1)}/(d\sigma^2)$. Using Assumption (C4), it can be shown that $E(T_{\mathrm{CQ}}^{(1)}) = \|\mu\|^2$ and $\mathrm{Var}(T_{\mathrm{CQ}}^{(1)}) = \Gamma_2(1+o(1))$ as $d\to\infty$, where Γ_2 is defined in (2.2). Using arguments similar to those in the proof of Theorem 2.1, we have the *weak convergence* of $(T_{\mathrm{CQ}}^{(1)} - \|\mu\|^2)/\psi_2^{1/2}$ to a standard Gaussian distribution. Further, the second term in (A.9) after dividing by $\Gamma_2^{1/2}$ converges to zero *in probability* as $d \to \infty$ for each $n \geq 4$. The previous two statements together imply that $(d\sigma^2 T_{\mathrm{SR}} - 2\|\mu\|^2)/\psi_2^{1/2}$ converges *weakly* to a $N(0, 4)$ distribution as $d \to \infty$ for each $n \geq 4$.

(b) The proof of this part of the theorem follows from arguments similar to those used in the proof of Theorem 2.2. □

PROOF OF THEOREM 3.1. Without any loss of generality, we can take $\mu_1 = \mathbf{0}$, so that $\mu = \mu_2$. Let us write $\mathbf{X}_i = \widetilde{\mathbf{V}}_i$ and $\mathbf{Y}_j = \mu + \widetilde{\mathbf{W}}_j$, where $\widetilde{\mathbf{V}}_i = \mathbf{V}_i/P_i$ and $\widetilde{\mathbf{W}}_j = \mathbf{W}_j/Q_j$ for $1 \leq i \leq m$ and $i \leq j \leq n$. Let $\mathbf{V} = (V_1, V_2, \dots, V_d)'$ and $\mathbf{W} = (W_1, W_2, \dots, W_d)'$. Denote $\Sigma_V = \mathrm{Disp}(\mathbf{V})$, $\Sigma_W = \mathrm{Disp}(\mathbf{W})$, $\sigma_V^2 = \mathrm{Var}(V_1)$ and $\sigma_W^2 = \mathrm{Var}(V_2)$.

(a) We will first derive the asymptotic distribution of T_{WMW}. Using arguments similar to those used in proving (A.1), we get

$$\|\mathbf{X} - \mathbf{Y}\|^2 = \sum_{k=1}^{d} \left[\frac{V_k^2}{P^2} + \frac{W_k^2}{Q^2} - \frac{2V_k W_k}{PQ}\right] + 2\mu'\left(\frac{\mathbf{V}}{P} - \frac{\mathbf{W}}{Q}\right) + \|\mu\|^2. \tag{A.10}$$

Consider the event $E = \{\|\mathbf{X} - \mathbf{Y}\|^2/d - (\sigma_V^2/P^2 + \sigma_W^2/Q^2) = o(d^{-1/2+\epsilon}) \text{ as } d \to \infty\}$. It follows from Bradley [(2005), Theorem 5.2(b)], that for any function $h : \mathbb{R}^2 \to \mathbb{R}$, the sequence $(h(V_k, W_k) : k \geq 1)$ is ρ-mixing with its mixing coefficient bounded by $\max\{\rho_1(\cdot), \rho_2(\cdot)\}$. Using this fact and (A.10) above along with the assumptions in the theorem and Theorem 8.2.2 in Lin and Lu (1996), we get that for any given $\epsilon \in (0, 1/2)$,

$$\Pr(E | P, Q) = 1 \tag{A.11}$$

for *almost every* P and Q. Now,

$$\begin{aligned}
T_{\mathrm{WMW}} &= \frac{1}{(m)_2(n)_2} \sum_{i_1 \neq i_2} \sum_{j_1 \neq j_2} \frac{(\mathbf{X}_{i_1} - \mathbf{Y}_{j_1})'(\mathbf{X}_{i_2} - \mathbf{Y}_{j_2})}{d(\sigma_V^2 P_{i_1}^{-2} + \sigma_W^2 Q_{j_1}^{-2})^{1/2}(\sigma_V^2 P_{i_2}^{-2} + \sigma_W^2 Q_{j_2}^{-2})^{1/2}} \\
&\quad + \frac{1}{(m)_2(n)_2} \sum_{i_1 \neq i_2} \sum_{j_1 \neq j_2} \Bigg[\frac{(\mathbf{X}_{i_1} - \mathbf{Y}_{j_1})'(\mathbf{X}_{i_2} - \mathbf{Y}_{j_2})}{d(\sigma_V^2 P_{i_1}^{-2} + \sigma_W^2 Q_{j_1}^{-2})^{1/2}(\sigma_V^2 P_{i_2}^{-2} + \sigma_W^2 Q_{j_2}^{-2})^{1/2}} \\
&\quad \times \left\{\frac{d(\sigma_V^2 P_{i_1}^{-2} + \sigma_W^2 Q_{j_1}^{-2})^{1/2}(\sigma_V^2 P_{i_2}^{-2} + \sigma_W^2 Q_{j_2}^{-2})^{1/2}}{\|\mathbf{X}_{i_1} - \mathbf{Y}_{j_1}\|\|\mathbf{X}_{i_2} - \mathbf{Y}_{j_2}\|} - 1\right\}\Bigg] \\
&= \left(T_{\mathrm{WMW}}^{(1)} + T_{\mathrm{WMW}}^{(2)}\right)/d,
\end{aligned} \tag{A.12}$$

where

$$T_{\mathrm{WMW}}^{(1)} = \frac{1}{(m)_2(n)_2} \sum_{i_1 \neq i_2} \sum_{j_1 \neq j_2} \frac{(\mathbf{X}_{i_1} - \mathbf{Y}_{j_1})'(\mathbf{X}_{i_2} - \mathbf{Y}_{j_2})}{d(\sigma_V^2 P_{i_1}^{-2} + \sigma_W^2 Q_{j_1}^{-2})^{1/2}(\sigma_V^2 P_{i_2}^{-2} + \sigma_W^2 Q_{j_2}^{-2})^{1/2}},$$

$$T^{(2)}_{\mathrm{WMW}} = \frac{1}{(m)_2(n)_2}\sum_{i_1\neq i_2}\sum_{j_1\neq j_2}\biggl[\frac{(\mathbf{X}_{i_1}-\mathbf{Y}_{j_1})'(\mathbf{X}_{i_2}-\mathbf{Y}_{j_2})}{d(\sigma_V^2P_{i_1}^{-2}+\sigma_W^2Q_{j_1}^{-2})^{1/2}(\sigma_V^2P_{i_2}^{-2}+\sigma_W^2Q_{j_2}^{-2})^{1/2}}$$
$$\times\biggl\{\frac{d(\sigma_V^2P_{i_1}^{-2}+\sigma_W^2Q_{j_1}^{-2})^{1/2}(\sigma_V^2P_{i_2}^{-2}+\sigma_W^2Q_{j_2}^{-2})^{1/2}}{\|\mathbf{X}_{i_1}-\mathbf{Y}_{j_1}\|\|\mathbf{X}_{i_2}-\mathbf{Y}_{j_2}\|}-1\biggr\}\biggr].$$

Let us define $A_{i_1,i_2}=\sum_{j_1\neq j_2}(\sigma_V^2P_{i_1}^{-2}+\sigma_W^2Q_{j_2}^{-2})^{-1/2}(\sigma_V^2P_{i_2}^{-2}+\sigma_W^2Q_{j_1}^{-2})^{-1/2}$, $B_{j_1,j_2}=\sum_{i_1\neq i_2}(\sigma_V^2P_{i_1}^{-2}+\sigma_W^2Q_{j_2}^{-2})^{-1/2}(\sigma_V^2P_{i_2}^{-2}+\sigma_W^2Q_{j_1}^{-2})^{-1/2}$, and $C_{i_1,j_1}=\sum_{i_2\neq i_1}\sum_{j_2\neq j_1}(\sigma_V^2P_{i_1}^{-2}+\sigma_W^2Q_{j_2}^{-2})^{-1/2}(\sigma_V^2P_{i_2}^{-2}+\sigma_W^2Q_{j_1}^{-2})^{-1/2}$. Some straightforward algebra yields

$$\begin{aligned}T^{(1)}_{\mathrm{WMW}} &= \frac{1}{d(m)_2(n)_2}\biggl\{\sum_{i_1\neq i_2}A_{i_1,i_2}\mathbf{X}_{i_1}'\mathbf{X}_{i_2}-2\sum_{i,j}C_{i,j}\mathbf{X}_i'\mathbf{Y}_j\\ &\qquad+\sum_{j_1\neq j_2}B_{j_1,j_2}\mathbf{Y}_{j_1}'\mathbf{Y}_{j_2}\biggr\}\\ &= \frac{1}{d(m)_2(n)_2}\biggl\{\sum_{i_1\neq i_2}\frac{A_{i_1,i_2}}{P_{i_1}P_{i_2}}\mathbf{V}_{i_1}'\mathbf{V}_{i_2}-2\sum_{i,j}\frac{C_{i,j}}{P_iQ_j}\mathbf{V}_i'(Q_j\mu+\mathbf{W}_j)\\ &\qquad+\sum_{j_1\neq j_2}\frac{B_{j_1,j_2}}{Q_{j_1}Q_{j_2}}(Q_{j_1}\mu+\mathbf{W}_{j_1})'(Q_{j_2}\mu+\mathbf{W}_{j_2})\biggr\}.\end{aligned}\tag{A.13}$$

Now, $E(dT^{(1)}_{\mathrm{WMW}}|P_i,Q_j,1\le i\le m,1\le j\le n)=\|\mu\|^2S_1$, where

$$S_1=\frac{1}{(m)_2(n)_2}\sum_{i_1\neq i_2}\sum_{j_1\neq j_2}\bigl(\sigma_V^2P_{i_1}^{-2}+\sigma_V^2Q_{j_2}^{-2}\bigr)^{-1/2}\bigl(\sigma_V^2P_{i_2}^{-2}+\sigma_W^2Q_{j_1}^{-2}\bigr)^{-1/2}.$$

Define $L_1=4\sum_{i,j_1\neq j_2}P_i^{-2}C_{i,j_1}C_{i,j_2}$ and $L_2=\sum(P_{j_1}^{-2}+P_{j_2}^{-2})B_{j_1,j_2}(2B_{j_1,j_2}+B_{j_1,j_3}+B_{j_2,j_3})$ with the latter summation taken over distinct indices j_1, j_2 and j_3. Denote $L_3=2\sum_{i_1\neq i_2}[P_{i_1}P_{i_2}]^{-2}A^2_{i_1,i_2}$, $L_4=2\sum_{j_1\neq j_2}[Q_{j_1}Q_{j_2}]^{-2}B^2_{j_1,j_2}$, and $L_5=2\sum_{i,j}[P_iQ_j]^{-2}C^2_{i,j}$. It can be shown that

$$\begin{aligned}&\mathrm{Var}\bigl(dT^{(1)}_{\mathrm{WMW}}|P_i,Q_j,1\le i\le m,1\le j\le n\bigr)\\ &\quad= S_2+\frac{4\mu'\Sigma_V\mu}{[(m)_2(n)_2]^2}\sum_{i,j_1\neq j_2}P_i^{-2}C_{i,j_1}C_{i,j_2}\\ &\qquad+\frac{\mu'\Sigma_W\mu}{[(m)_2(n)_2]^2}\sum_{\substack{j_1,j_2,j_3\\ \text{all distinct}}}(P_{j_1}^{-2}+P_{j_2}^{-2})B_{j_1,j_2}(2B_{j_1,j_2}+B_{j_1,j_3}+B_{j_2,j_3})\\ &\quad= S_2+\frac{1}{[(m)_2(n)_2]^2}\{L_1\mu'\Sigma_V\mu+L_2\mu'\Sigma_W\mu\},\end{aligned}\tag{A.14}$$

where

$$S_2 = \frac{1}{[(m)_2(n)_2]^2}\Bigl\{2\sum_{i_1\neq i_2}[P_{i_1}P_{i_2}]^{-2}A^2_{i_1,i_2}\operatorname{tr}(\Sigma_V^2)$$
$$+2\sum_{j_1\neq j_2}[Q_{j_1}Q_{j_2}]^{-2}B^2_{j_1,j_2}\operatorname{tr}(\Sigma_W^2)+4\sum_{i,j}[P_iQ_j]^{-2}C^2_{i,j}\operatorname{tr}(\Sigma_V\Sigma_W)\Bigr\}$$
$$=\{L_3\operatorname{tr}(\Sigma_V^2)+L_4\operatorname{tr}(\Sigma_W^2)+2L_5\operatorname{tr}(\Sigma_V\Sigma_W)\}/[(m)_2(n)_2]^2.$$

Note that $(L_1\mu'\Sigma_V\mu+L_2\mu'\Sigma_W\mu)\le \max\{L_1,L_2\}\mu'(\Sigma_V+\Sigma_W)\mu$. Also,

$$S_2 \ge [(m)_2(n)_2]^{-2}\min\{L_3,L_4,L_5\}\operatorname{tr}[(\Sigma_V+\Sigma_W)^2]. \tag{A.15}$$

These facts along with (A.14) and Assumption (C3) imply that $\operatorname{Var}(dT^{(1)}_{\mathrm{WMW}}| P_i, Q_j, 1\le i\le m, 1\le j\le n) = S_2(1+o(1))$ as $d\to\infty$. Now,

$$\bigl(dT^{(1)}_{\mathrm{WMW}}-\|\mu\|^2S_1\bigr)/S_2^{1/2}$$
$$=\Bigl[\frac{1}{(m)_2(n)_2}\sum_{i_1\neq i_2}\sum_{j_1\neq j_2}\frac{(\mathbf{X}_{i_1}-\mathbf{Y}_{j_1}+\mu)'(\mathbf{X}_{i_2}-\mathbf{Y}_{j_2}+\mu)}{d(\sigma_V^2P_{i_1}^{-2}+\sigma_W^2Q_{j_1}^{-2})^{1/2}(\sigma_V^2P_{i_2}^{-2}+\sigma_W^2Q_{j_2}^{-2})^{1/2}}$$
$$-\frac{2}{(m)_2(n)_2}\sum_{i,j}C_{i,j}\mu'(\mathbf{X}_i-\mathbf{Y}_j+\mu)\Bigr]\Big/S_2^{1/2}$$
$$=(\widetilde{T}^{(1)}_{\mathrm{WMW}}-\widetilde{T}^{(2)}_{\mathrm{WMW}})/S_2^{1/2}, \tag{A.16}$$

where $\widetilde{T}^{(2)}_{\mathrm{WMW}} = 2[(m)_2(n)_2]^{-1}\sum_{i,j}C_{i,j}\mu'(\mathbf{X}_i-\mathbf{Y}_j+\mu)$ and

$$\widetilde{T}^{(1)}_{\mathrm{WMW}}=\frac{1}{(m)_2(n)_2}\sum_{i_1\neq i_2}\sum_{j_1\neq j_2}\frac{(\mathbf{X}_{i_1}-\mathbf{Y}_{j_1}+\mu)'(\mathbf{X}_{i_2}-\mathbf{Y}_{j_2}+\mu)}{d(\sigma_V^2P_{i_1}^{-2}+\sigma_W^2Q_{j_1}^{-2})^{1/2}(\sigma_V^2P_{i_2}^{-2}+\sigma_W^2Q_{j_2}^{-2})^{1/2}}.$$

It can be shown that $E(\widetilde{T}^{(2)}_{\mathrm{WMW}}|P_i,Q_j,1\le i\le m,1\le j\le n)=0$ and

$$\operatorname{Var}(\widetilde{T}^{(2)}_{\mathrm{WMW}}|P_i,Q_j,1\le i\le m,1\le j\le n)$$
$$=4[(m)_2(n)_2]^{-2}$$
$$\times\Bigl\{\sum_{i,j_1\neq j_2}C_{i,j_1}C_{i,j_2}P_i^{-2}\mu'\Sigma_V\mu+\sum_{i_1\neq i_2,j}C_{i_1,j}C_{i_2,j}Q_j^{-2}\mu'\Sigma_W\mu$$
$$+\sum_{i,j}C^2_{i,j}\mu'(\Sigma_V/P_i^2+\Sigma_W/Q_j^2)\mu\Bigr\}.$$

So, using Assumption (C3) and arguments similar to those used earlier to show that $\operatorname{Var}(dT^{(1)}_{\mathrm{WMW}}|P_i,Q_j,1\le i\le m,1\le j\le n)=S_2(1+o(1))$ as $d\to\infty$, we get that $\operatorname{Var}(\widetilde{T}^{(2)}_{\mathrm{WMW}}|P_i,Q_j,1\le i\le m,1\le j\le n)=o(S_2)$ as $d\to\infty$. Thus, Chebyshev's inequality implies that $\widetilde{T}^{(2)}_{\mathrm{WMW}}/S_2^{1/2}$ converges to zero *in probability* as $d \to \infty$.

Next, note that

$$\begin{aligned}
&\widetilde{T}^{(1)}_{\mathrm{WMW}} \\
&\quad = \frac{1}{(m)_2(n)_2} \sum_{i_1 \neq i_2} \sum_{j_1 \neq j_2} \frac{(\mathbf{V}_{i_1}/P_{i_1} - \mathbf{W}_{j_1}/Q_{j_1})'(\mathbf{V}_{i_2}/P_{i_2} - \mathbf{W}_{j_2}/Q_{j_2})}{d(\sigma_V^2 P_{i_1}^{-2} + \sigma_W^2 Q_{j_1}^{-2})^{1/2}(\sigma_V^2 P_{i_2}^{-2} + \sigma_W^2 Q_{j_2}^{-2})^{1/2}} \\
&\quad = \frac{1}{(m)_2(n)_2} \\
&\qquad \times \sum_{k=1}^{d} \sum_{\substack{i_1 \neq i_2 \\ j_1 \neq j_2}} \frac{(Q_{j_1} V_{i_1 k} - P_{i_1} W_{j_1 k})(Q_{j_2} V_{i_2 k} - P_{i_2} W_{j_2 k})/(P_{i_1} P_{i_2} Q_{j_1} Q_{j_2})}{d(\sigma_V^2 P_{i_1}^{-2} + \sigma_W^2 Q_{j_1}^{-2})^{1/2}(\sigma_V^2 P_{i_2}^{-2} + \sigma_W^2 Q_{j_2}^{-2})^{1/2}}.
\end{aligned}$$

It is easy to see that $E(\widetilde{T}^{(1)}_{\mathrm{WMW}} | P_i, Q_j, 1 \leq i \leq m, 1 \leq j \leq n) = 0$. Further, from calculations similar to those used earlier in deriving $\mathrm{Var}(dT^{(1)}_{\mathrm{WMW}} | P_i, Q_j, 1 \leq i \leq m, 1 \leq j \leq n)$, it can be shown that $\mathrm{Var}(\widetilde{T}^{(1)}_{\mathrm{WMW}} | P_i, Q_j, 1 \leq i \leq m, 1 \leq j \leq n) = S_2$. Thus, by Theorem 4.0.1 in Lin and Lu (1996) and Assumption (C4), the conditional distribution of $\widetilde{T}^{(1)}_{\mathrm{WMW}}/S_2^{1/2}$ given the P_i's and the Q_j's converges to a standard Gaussian distribution as $d \to \infty$. This fact along with (A.16) and the fact that conditionally on the P_i's and the Q_j's, $\widetilde{T}^{(2)}_{\mathrm{WMW}}/S_2^{1/2}$ converges to zero *in probability* as $d \to \infty$ yield

$$\lim_{d \to \infty} P\{(dT^{(1)}_{\mathrm{WMW}} - \|\mu\|^2 S_1)/S_2^{1/2} \leq x\} = \Phi(x). \tag{A.17}$$

Next, let us write

$$\begin{aligned}
&T^{(2)}_{\mathrm{WMW}} \\
&\quad = \frac{1}{(m)_2(n)_2} \sum_{i_1 \neq i_2} \sum_{j_1 \neq j_2} \Bigg[\frac{(\mathbf{X}_{i_1} - \mathbf{Y}_{j_1})'(\mathbf{X}_{i_2} - \mathbf{Y}_{j_2}) - \|\mu\|^2}{(\sigma_V^2 P_{i_1}^{-2} + \sigma_W^2 Q_{j_1}^{-2})^{1/2}(\sigma_V^2 P_{i_2}^{-2} + \sigma_W^2 Q_{j_2}^{-2})^{1/2}} \\
&\qquad \times \left\{ \frac{d(\sigma_V^2 P_{i_1}^{-2} + \sigma_W^2 Q_{j_1}^{-2})^{1/2}(\sigma_V^2 P_{i_2}^{-2} + \sigma_W^2 Q_{j_2}^{-2})^{1/2}}{\|\mathbf{X}_{i_1} - \mathbf{Y}_{j_1}\| \|\mathbf{X}_{i_2} - \mathbf{Y}_{j_2}\|} - 1 \right\} \Bigg] \\
&\qquad + \frac{\|\mu\|^2}{(m)_2(n)_2} \sum_{i_1 \neq i_2} \sum_{j_1 \neq j_2} \Bigg[\frac{1}{(\sigma_V^2 P_{i_1}^{-2} + \sigma_W^2 Q_{j_1}^{-2})^{1/2}(\sigma_V^2 P_{i_2}^{-2} + \sigma_W^2 Q_{j_2}^{-2})^{1/2}} \\
&\qquad \times \left\{ \frac{d(\sigma_V^2 P_{i_1}^{-2} + \sigma_W^2 Q_{j_1}^{-2})^{1/2}(\sigma_V^2 P_{i_2}^{-2} + \sigma_W^2 Q_{j_2}^{-2})^{1/2}}{\|\mathbf{X}_{i_1} - \mathbf{Y}_{j_1}\| \|\mathbf{X}_{i_2} - \mathbf{Y}_{j_2}\|} - 1 \right\} \Bigg] \\
&\quad = \widetilde{T}^{(3)}_{\mathrm{WMW}} + \widetilde{T}^{(4)}_{\mathrm{WMW}},
\end{aligned} \tag{A.18}$$

where

$$\widetilde{T}^{(3)}_{\mathrm{WMW}} = \frac{1}{(m)_2(n)_2}\sum_{i_1\neq i_2}\sum_{j_1\neq j_2}\Bigg[\frac{(\mathbf{X}_{i_1}-\mathbf{Y}_{j_1})'(\mathbf{X}_{i_2}-\mathbf{Y}_{j_2})-\|\mu\|^2}{(\sigma_V^2 P_{i_1}^{-2}+\sigma_W^2 Q_{j_1}^{-2})^{1/2}(\sigma_V^2 P_{i_2}^{-2}+\sigma_W^2 Q_{j_2}^{-2})^{1/2}} \times\left\{\frac{d(\sigma_V^2 P_{i_1}^{-2}+\sigma_W^2 Q_{j_1}^{-2})^{1/2}(\sigma_V^2 P_{i_2}^{-2}+\sigma_W^2 Q_{j_2}^{-2})^{1/2}}{\|\mathbf{X}_{i_1}-\mathbf{Y}_{j_1}\|\|\mathbf{X}_{i_2}-\mathbf{Y}_{j_2}\|}-1\right\}\Bigg],$$

$$\widetilde{T}^{(4)}_{\mathrm{WMW}} = \frac{1}{(m)_2(n)_2}\sum_{i_1\neq i_2}\sum_{j_1\neq j_2}\Bigg[\frac{\|\mu\|^2}{(\sigma_V^2 P_{i_1}^{-2}+\sigma_W^2 Q_{j_1}^{-2})^{1/2}(\sigma_V^2 P_{i_2}^{-2}+\sigma_W^2 Q_{j_2}^{-2})^{1/2}} \times\left\{\frac{d(\sigma_V^2 P_{i_1}^{-2}+\sigma_W^2 Q_{j_1}^{-2})^{1/2}(\sigma_V^2 P_{i_2}^{-2}+\sigma_W^2 Q_{j_2}^{-2})^{1/2}}{\|\mathbf{X}_{i_1}-\mathbf{Y}_{j_1}\|\|\mathbf{X}_{i_2}-\mathbf{Y}_{j_2}\|}-1\right\}\Bigg].$$

The stationarity of the sequences $\mathcal{V}$ and $\mathcal{W}$ and the Cauchy–Schwarz inequality imply that $\mathrm{tr}[(\Sigma_V+\Sigma_W)^2]\geq d(\sigma_V^2+\sigma_W^2)^2$. This along with (A.15), (A.11) and Assumption (C3) imply that conditionally on the P_i's and the Q_j's, each term inside the double sum appearing in $\widetilde{T}^{(4)}_{\mathrm{WMW}}$ above is $o_P(S_2^{1/2})$ as $d\to\infty$. So, $\widetilde{T}^{(4)}_{\mathrm{WMW}}/S_2^{1/2}$ converges to zero *in probability* as $d\to\infty$.

Next, fix any $i_1\neq i_2$ and $j_1\neq j_2$ and consider the corresponding term inside the double summation appearing in the expression of $\widetilde{T}^{(3)}_{\mathrm{WMW}}$. It follows from (A.11) that $d(\sigma_V^2 P_{i_1}^{-2}+\sigma_W^2 Q_{j_1}^{-2})^{1/2}(\sigma_V^2 P_{i_2}^{-2}+\sigma_W^2 Q_{j_2}^{-2})^{1/2}/[\|\mathbf{X}_{i_1}-\mathbf{Y}_{j_1}\|\|\mathbf{X}_{i_2}-\mathbf{Y}_{j_2}\|]-1$ converges to zero *in probability* as $d\to\infty$. Note that

$$\begin{aligned}
&(\mathbf{X}_{i_1}-\mathbf{Y}_{j_1})'(\mathbf{X}_{i_2}-\mathbf{Y}_{j_2})-\|\mu\|^2\\
&\quad=(\mathbf{X}_{i_1}-\mathbf{Y}_{j_1}+\mu)'(\mathbf{X}_{i_2}-\mathbf{Y}_{j_2}+\mu) \qquad\text{(A.19)}\\
&\qquad-\mu'(\mathbf{X}_{i_1}-\mathbf{Y}_{j_1}+\mu)-\mu'(\mathbf{X}_{i_2}-\mathbf{Y}_{j_2}+\mu)\\
&\quad=\sum_{k=1}^{d}\left\{\frac{(Q_{j_1}V_{i_1k}-P_{i_1}W_{j_1k})(Q_{j_2}V_{i_2k}-P_{i_2}W_{j_2k})}{P_{i_1}P_{i_2}Q_{j_1}Q_{j_2}}\right\}\\
&\qquad-\mu'(Q_{j_1}\mathbf{V}_{i_1}-P_{i_1}\mathbf{W}_{j_1})\mu/(P_{i_1}Q_{j_1})\\
&\qquad-\mu'(Q_{j_2}\mathbf{V}_{i_2}-P_{i_2}\mathbf{W}_{j_2})\mu/(P_{i_2}Q_{j_2}). \qquad\text{(A.20)}
\end{aligned}$$

It is easy to show that the conditional expectation of the first term in (A.20) given the P_i's and the Q_j's is zero, and its conditional variance is $v_{i_1i_2j_1j_2}=[P_{i_1}P_{i_2}]^{-2}\mathrm{tr}(\Sigma_V^2)+[Q_{j_1}Q_{j_2}]^{-2}\mathrm{tr}(\Sigma_W^2)+\{[P_{i_1}Q_{j_2}]^{-2}+[P_{i_2}Q_{j_1}]^{-2}\}\mathrm{tr}(\Sigma_V\Sigma_W)$. So, $v_{i_1i_2j_1j_2}=O(\mathrm{tr}[(\Sigma_V+\Sigma_W)^2])$. Hence, using (A.15) and Chebyshev's inequality, it follows that the first term in (A.20) after scaling by $S_2^{1/2}$ is bounded *in probability*, conditional on the P_i's and the Q_j's, as $d\to\infty$. Using Assumption (C3), Chebyshev's inequality and arguments similar to those used to prove the convergence *in probability* to zero of $\widetilde{T}^{(2)}_{\mathrm{WMW}}$ earlier, we get that the second and the third terms in (A.20) after dividing by $S_2^{1/2}$ converge to zero *in probability* as $d \to \infty$. So, the left-hand side of the equation (A.19) after dividing by $S_2^{1/2}$ is bounded *in probability*, conditional on the P_i's and the Q_j's, as $d \to \infty$. Thus, $\widetilde{T}_{\mathrm{WMW}}^{(3)}/S_2^{1/2}$ converges to zero *in probability* as $d \to \infty$. This along with (A.18) and the fact that $\widetilde{T}_{\mathrm{WMW}}^{(4)}/S_2^{1/2}$ converges to zero *in probability* as $d \to \infty$ together imply that $T_{\mathrm{WMW}}^{(2)}/S_2^{1/2}$ converges to zero *in probability* as $d \to \infty$. Combining this fact with (A.17) and (A.12), we get $\lim_{d\to\infty} P\{(dT_{\mathrm{WMW}} - \|\mu\|^2 S_1)/S_2^{1/2} \le x | P_i, Q_j, 1 \le i \le m, 1 \le j \le n\} = \Phi(x)$ for all $x \in \mathbb{R}$ and for each $m, n \ge 2$. Consequently,

$$\lim_{d\to\infty} P\{(dT_{\mathrm{WMW}} - \|\mu\|^2 S_1)/S_2^{1/2} \le x\} = \Phi(x)$$

for all $x \in \mathbb{R}$ and for each $m, n \ge 2$.

We now derive the asymptotic distribution of $T_{\mathrm{CQ}}^{(2)}$. As in the proof of Theorem 2.1, $T_{\mathrm{CQ}}^{(2)} = T_1 - T_2$. In the setup of the present theorem, $T_1 = [(m)_2(n)_2]^{-1} \sum_{i_1 \ne i_2} \sum_{j_1 \ne j_2} (\mathbf{V}_{i_1}/P_{i_1} - \mathbf{W}_{j_1}/Q_{j_1})'(\mathbf{V}_{i_2}/P_{i_2} - \mathbf{W}_{j_2}/Q_{j_2})'$ and $T_2 = 2(mn)^{-1} \sum_{i,j} \mu'(\mathbf{V}_i/P_i - \mathbf{W}_j/Q_j)$. So, $E(T_1 | P_i, Q_j, 1 \le i \le m, 1 \le j \le n) = 0$. Further, from algebraic computations similar to those used to derive the variance of $T_{\mathrm{CQ}}^{(2)}$ in the proof of Theorem 2.1, it follows that

$$\begin{aligned}
&\mathrm{Var}(T_1 | P_i, Q_j, 1 \le i \le m, 1 \le j \le n) \\
&\quad = \frac{1}{[(m)_2(n)_2]^2} \biggl\{ 2 \sum_{i_1 \ne i_2} [P_{i_1} P_{i_2}]^{-2} \mathrm{tr}(\Sigma_V^2) \\
&\qquad + 2 \sum_{j_1 \ne j_2} [Q_{j_1} Q_{j_2}]^{-2} \mathrm{tr}(\Sigma_W^2) + 4 \sum_{i,j} [P_i Q_j]^{-2} \mathrm{tr}(\Sigma_V \Sigma_W) \biggr\}.
\end{aligned}$$

Define $S_3 = \mathrm{Var}(T_1 | P_i, Q_j, 1 \le i \le m, 1 \le j \le n)$. Also, $E(T_2 | P_i, Q_j, 1 \le i \le m, 1 \le j \le n) = 0$, and $\mathrm{Var}(T_2 | P_i, Q_j, 1 \le i \le m, 1 \le j \le n) = o(S_3)$ as $d \to \infty$ using the assumptions in the theorem. Thus, $T_2/S_3^{1/2}$ converges *in probability* to zero as $d \to \infty$. Further, using arguments similar to those used to prove the asymptotic Gaussianity of $\widetilde{T}_{\mathrm{WMW}}^{(1)}$ above, it follows that the conditional distribution of $T_1/S_3^{1/2}$ given the P_i's and the Q_j's converges *weakly* to a standard Gaussian distribution as $d \to \infty$ for all $m, n \ge 2$. Combining these facts, we have

$$\lim_{d\to\infty} P\{(T_{\mathrm{CQ}}^{(2)} - \|\mu\|^2)/S_3^{1/2} \le x\} = \Phi(x)$$

for all $x \in \mathbb{R}$ and all $m, n \ge 2$.

(b) Note that S_1 is a V-statistic whose kernel $(\sigma_V^2 P_{i_1}^{-2} + \sigma_V^2 Q_{j_2}^{-2})^{-1/2}(\sigma_V^2 P_{i_2}^{-2} + \sigma_W^2 Q_{j_1}^{-2})^{-1/2}$ has finite expectation $\psi_1 = E^2\{PQ/(\sigma_V^2 Q^2 + \sigma_W^2 P^2)^{1/2}\}$, by the assumption in the theorem. Thus, it follows that S_1 converges *almost surely* to ψ_1. Define $S_{21} = [(m)_2\{(n)_2\}^2]^{-1} L_3$, $S_{22} = [\{(m)_2\}^2 (n)_2]^{-1} L_4$ and $S_{23} =$

$[mn(m-1)^2(n-1)^2]^{-1}L_5$. Each of S_{21}, S_{22} and S_{23} is a real valued V-statistic whose kernel is bounded, and thus has finite expectation. So, there exist ψ_{21}, ψ_{22} and ψ_{23} depending only on the distributions of P and Q such that S_{21}, S_{22} and S_{23} converge *almost surely* to ψ_{21}, ψ_{22} and ψ_{23}, respectively. Here, $\psi_{21} = E^2\{Q_1Q_2/[(\sigma_V^2 Q_2^2 + \sigma_W^2 P_1^2)(\sigma_V^2 Q_2^2 + \sigma_W^2 P_1^2)]^{1/2}\}$, $\psi_{22} = E^2\{P_1P_2/[(\sigma_V^2 Q_1^2 + \sigma_W^2 P_1^2)(\sigma_V^2 Q_1^2 + \sigma_W^2 P_2^2)]^{1/2}\}$, and $\psi_{23} = [\psi_{21}\psi_{22}]^{1/2}$. Define $\psi_2 = 2\operatorname{tr}(\Sigma_V^2)\psi_{21}/(m)_2 + 2\operatorname{tr}(\Sigma_W^2)\psi_{22}/(n)_2 + 4\operatorname{tr}(\Sigma_V\Sigma_W)\psi_{23}/(mn)$. Recall that $S_2 = 2\operatorname{tr}(\Sigma_V^2)S_{21}/(m)_2 + 2\operatorname{tr}(\Sigma_W^2)S_{22}/(n)_2 + 4\operatorname{tr}(\Sigma_V\Sigma_W)S_{23}/(mn)$. Conditions (C1) and (C2) along with Theorem 2.1.5 in Lin and Lu (1996) imply that both $\mathcal{V}$ and $\mathcal{W}$ possess continuous spectral densities. Now, the proof of Theorem 18.2.1 in Ibragimov and Linnik (1971) implies that each of $\operatorname{tr}(\Sigma_V^2)$, $\operatorname{tr}(\Sigma_W^2)$ and $\operatorname{tr}(\Sigma_V\Sigma_W)$ equals a constant multiple of d plus a remainder term, which is $o(d)$ as $d\to\infty$. Thus, for each fixed $m,n\ge 2$, there exist constants A_1, A_2 and A_3 such that with *probability one*

$$\lim_{d\to\infty}\frac{\psi_2}{S_2} = \frac{2\psi_{21}A_1/(m)_2 + 2\psi_{22}A_2/(n)_2 + 4\psi_{23}A_3/(mn)}{2S_{21}A_1/(m)_2 + 2S_{22}A_2/(n)_2 + 4S_{23}A_3/(mn)}. \tag{A.21}$$

We denote the right-hand side of (A.21) by $R_{m,n}$. Further, the assumption in the theorem and arguments preceding (A.21) imply that $\|\mu\|^2/\psi_2^{1/2}$ converges to a finite nonnegative limit $\widetilde{b}_{m,n}^2$ (say) as $d\to\infty$, where $\widetilde{b}_{m,n}^2 = o((m+n)^{1/2})$ as $m,n\to\infty$. Now,

$$\begin{aligned}
&\lim_{d\to\infty} P\left\{\frac{dT_{\mathrm{WMW}} - \|\mu\|^2\psi_1}{\psi_2^{1/2}} \le x\right\}\\
&\quad = \lim_{d\to\infty} P\left\{\frac{dT_{\mathrm{WMW}} - \|\mu\|^2 S_1}{S_2^{1/2}} \le \frac{x\psi_2^{1/2}}{S_2^{1/2}} - \frac{\|\mu\|^2(S_1-\psi_1)}{S_2^{1/2}}\right\}\\
&\quad = E\left[\lim_{d\to\infty} P\left\{\frac{dT_{\mathrm{WMW}} - \|\mu\|^2 S_1}{S_2^{1/2}} \le \frac{x\psi_2^{1/2}}{S_2^{1/2}} - \frac{\|\mu\|^2(S_1-\psi_1)}{S_2^{1/2}}\Big| P_i's,\, Q_j's\right\}\right]\\
&\quad = E\left[\Phi\left(\lim_{d\to\infty}\frac{\psi_2^{1/2}}{S_2^{1/2}}\left\{x - (S_1-\psi_1)\lim_{d\to\infty}\frac{\|\mu\|^2}{\psi_2^{1/2}}\right\}\right)\right]\\
&\quad = E\left[\Phi\left(R_{m,n}\{x - (S_1-\psi_1)\widetilde{b}_{m,n}^2\}\right)\right].
\end{aligned}$$

The above expectation converges to $\Phi(x)$ if we now let $m,n\to\infty$ because $R_{m,n}$ converges to one, $\widetilde{b}_{m,n}^2 = o((m+n)^{1/2})$, and $(m+n)^{1/2}(S_1-\psi_1)$ converges *weakly* to a standard Gaussian distribution as $m,n\to\infty$.

Now, $[(m)_2\{(n)_2\}^2]^{-1}\sum_{i_1\ne i_2}[P_{i_1}P_{i_2}]^{-2}$, $[(n)_2\{(m)_2\}^2]^{-1}\sum_{j_1\ne j_2}[Q_{j_1}Q_{j_2}]^{-2}$ and $[mn(m-1)^2(n-1)^2]^{-1}\sum_{i,j}[P_iQ_j]^{-2}$ appearing in the expression of S_3 converge to $E^2(P^{-2})$, $E^2(Q^{-2})$ and $E(P^{-2})E(Q^{-2})$, respectively, as $m,n\to\infty$. Also note that $\Sigma_1 = \operatorname{Disp}(\mathbf{X}) = \Sigma_V E(P^{-2})$ and $\Sigma_2 = \operatorname{Disp}(\mathbf{Y}) = \Sigma_W E(Q^{-2})$.

So, arguing as in the case of S_2 above, we get that S_3/Γ_1 converges *in probability* to one if we first let $d \to \infty$ and then $m, n \to \infty$. Thus, $\lim_{m,n\to\infty} \lim_{d\to\infty} P \times \{(T_{\mathrm{CQ}}^{(2)} - \|\mu\|^2)/\Gamma_1^{1/2} \le x\} = \Phi(x)$ for all $x \in \mathbb{R}$. $\square$

PROOF OF THEOREM 3.2. Since $\mathbf{Y}$ is distributed as $\mathbf{X} + \mu$, we have $\psi_1 = \sigma_V^{-2} E^2\{PQ/(P^2 + Q^2)^{1/2}\}$ and $\psi_2 = [\sigma_V^2 E(P^{-2})]^{-2} E^2\{Q_1 Q_2/[(P_1^2 + Q_1^2)^{1/2}(P_1^2 + Q_2^2)^{1/2}]\}\Gamma_1$. Here, ψ_1 and ψ_2 are as in the proof of Theorem 3.1. Since $\lim_{m,n\to\infty} \lim_{d\to\infty} \|\mu\|^2/\Gamma_1^{1/2} = c$ for some $c \in (0, \infty)$, we have $\lim_{m,n\to\infty} \lim_{d\to\infty} \beta_{T_{\mathrm{CQ}}^{(2)}}(\mu) = \Phi(-\zeta_\alpha + c)$, and

$$\lim_{m,n\to\infty} \lim_{d\to\infty} \beta_{T_{\mathrm{WMW}}}(\mu) = \Phi\left(-\zeta_\alpha + \frac{cE(P^{-2})E^2\{PQ/(P^2+Q^2)^{1/2}\}}{E\{Q_1Q_2/[(P_1^2+Q_1^2)^{1/2}(P_1^2+Q_2^2)^{1/2}]\}}\right).$$

Now, $E^2\{Q_1Q_2/[(P_1^2+Q_1^2)^{1/2}(P_1^2+Q_2^2)^{1/2}]\} = E[E^2\{Q_1/(P_1^2+Q_1^2)^{1/2}|P_1\}] < E[E\{Q_1^2/(P_1^2 + Q_1^2)|P_1\}] = E\{Q_1^2/(P_1^2 + Q_1^2)\} = 1/2$. Here, the inequality can be obtained using Jensen's inequality. Further, $E^2\{PQ/(P^2 + Q^2)^{1/2}\} > E^{-2}\{(P^2+Q^2)^{1/2}/PQ\} > E^{-1}\{(P^2+Q^2)/P^2Q^2\} = 1/\{E(P^{-2}) + E(Q^{-2})\} = [2E(P^{-2})]^{-1}$. Here, the inequalities follow from Cauchy–Schwarz inequality. Combining the previous two inequalities, we get $\lim_{m,n\to\infty} \lim_{d\to\infty} \beta_{T_{\mathrm{WMW}}}(\mu) > \lim_{m,n\to\infty} \lim_{d\to\infty} \beta_{T_{\mathrm{CQ}}^{(2)}}(\mu)$. $\square$

PROOF OF THEOREM 3.3. (a) The proof of the asymptotic Gaussianity of T_{SR} is provided in the Supplementary Material [Chakraborty and Chaudhuri (2016)]. Also, Z_2 and Z_3 appearing in its asymptotic distribution are $Z_2 = 2[(n)_4\sigma_V^2]^{-1} \sum_{i_1\ne i_2} \widetilde{U}_{i_1,i_2} P_{i_1} P_{i_2}$ and $Z_3 = 8\,\mathrm{tr}(\Sigma_V^2)[(n)_4\sigma_V^2]^{-2} \sum_{i_1\ne i_2} \widetilde{U}_{i_1,i_2}^2$, where $\widetilde{U}_{i_1,i_2} = \sum P_{i_3}P_{i_4}/[(P_{i_1}^2 + P_{i_3}^2)^{1/2}(P_{i_2}^2 + P_{i_4}^2)^{1/2}]$ and the summation is taken over indices (i_3, i_4) satisfying $i_3 \ne i_4$ and $(i_3, i_4) \ne (i_1, i_2)$.

The proof of the asymptotic Gaussianity of T_S will follow from arguments similar to those used to prove the asymptotic Gaussianity of T_{SR}, and we skip the details. Z_1 and Γ_3 in the asymptotic distribution of T_S are given by $Z_1 = [(n)_2\sigma_V^2]^{-1} \sum_{i_1\ne i_2} P_{i_1}P_{i_2}$ and $\Gamma_3 = 2\,\mathrm{tr}(\Sigma_V^2)/[(n)_2\sigma_V^4]$.

The proof of the asymptotic Gaussianity of $T_{\mathrm{CQ}}^{(1)}$ is also provided in the Supplementary Material [Chakraborty and Chaudhuri (2016)], and Z_4 appearing in its asymptotic distribution is given by $Z_4 = 2\,\mathrm{tr}(\Sigma_V^2)[(n)_2]^{-2} \sum_{i_1\ne i_2} [P_{i_1}P_{i_2}]^{-2}$.

(b) Since Z_1, Z_2, $(n)_4 Z_3$ and $(n)_2 Z_4$ are real-valued V-statistics with finite expectations by the assumptions in part (b) of the theorem, they converge *almost surely* as $n \to \infty$. The corresponding limits are $\theta_1 = E^2(P_1)/\sigma_V^2$, $\theta_2 = E^2\{P_1P_2/(P_1^2 + P_2^2)^{1/2}\}/\sigma_V^2$, $\theta_3 = \mathrm{tr}(\Sigma_V^2)E^2\{P_2P_3/(P_1^2 + P_2^2)^{1/2}(P_1^2 + P_3^2)^{1/2}\}/\sigma_V^4$ and $\theta_4 = 2\,\mathrm{tr}(\Sigma_V^2)E^2(P_1^{-2})$. Note that since $E(P^{-2})$ is finite, we have $\Sigma = \mathrm{Disp}(\mathbf{X}) = \Sigma_V E(P^{-2})$ and $\sigma^2 = \mathrm{Var}(X_1) = \sigma_V^2 E(P^{-2})$. So, $\theta_4 = 2\,\mathrm{tr}(\Sigma^2)$.

Arguments similar to those used in the proof of part (b) of Theorem 3.1 complete the proof of part (b) of the present theorem.

(c) Suppose that $\lim_{n\to\infty}\lim_{d\to\infty}\|\mu\|^2/\Gamma_2^{1/2}=c$ for some $c\in(0,\infty)$. Then

$$\lim_{n\to\infty}\lim_{d\to\infty}\beta_{T_S}(\mu)=\Phi\bigl(-\zeta_\alpha+cE^2(P_1)E\bigl(P_1^{-2}\bigr)\bigr),$$

$$\lim_{n\to\infty}\lim_{d\to\infty}\beta_{T_{\mathrm{SR}}}(\mu)=\Phi\biggl(-\zeta_\alpha+\frac{cE^2\{P_1P_2/(P_1^2+P_2^2)^{1/2}\}E(P_1^{-2})}{E\{P_2P_3/[(P_1^2+P_2^2)^{1/2}(P_1^2+P_3^2)^{1/2}]\}}\biggr),$$

$$\lim_{n\to\infty}\lim_{d\to\infty}\beta_{T_{\mathrm{CQ}}^{(1)}}(\mu)=\Phi(-\zeta_\alpha+c).$$

Now, from Jensen's inequality, we have $E^2(P_1)>E^{-2}(P_1^{-1})>E^{-1}(P_1^{-2})$, which implies that $E^2(P_1)E(P_1^{-2})>1$. Thus, $\lim_{n\to\infty}\lim_{d\to\infty}\beta_{T_S}(\mu)>\lim_{n\to\infty}\lim_{d\to\infty}\beta_{T_{\mathrm{CQ}}^{(1)}}(\mu)$. The proof of the other part of the theorem is similar to the proof of Theorem 3.2. □

SUPPLEMENTARY MATERIAL

Supplement to "Tests for high-dimensional data based on means, spatial signs and spatial ranks" (DOI: 10.1214/16-AOS1467SUPP; .pdf). This supplemental article contains additional mathematical details related to the proof of part (a) of Theorem 3.3 and the detailed results of the simulation study done in Section 5 of the paper.

REFERENCES

BAI, Z. and SARANADASA, H. (1996). Effect of high dimension: By an example of a two sample problem. *Statist. Sinica* **6** 311–329. MR1399305

BRADLEY, R. C. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probab. Surv.* **2** 107–144. MR2178042

CAI, T. T., LIU, W. and XIA, Y. (2014). Two-sample test of high dimensional means under dependence. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 349–372. MR3164870

CAPPÉ, O., MOULINES, E. and RYDÉN, T. (2005). *Inference in Hidden Markov Models*. Springer, New York. MR2159833

CHAKRABORTY, A. and CHAUDHURI, P. (2016). Supplement to "Tests for high-dimensional data based on means, spatial signs and spatial ranks." DOI:10.1214/16-AOS1467SUPP.

CHEN, S. X. and QIN, Y.-L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *Ann. Statist.* **38** 808–835. MR2604697

CHOI, K. and MARDEN, J. (1997). An approach to multivariate rank tests in multivariate analysis of variance. *J. Amer. Statist. Assoc.* **92** 1581–1590. MR1615267

FAN, J. and LIN, S.-K. (1998). Test of significance when data are curves. *J. Amer. Statist. Assoc.* **93** 1007–1021. MR1649196

FENG, L., ZOU, C., WANG, Z. and ZHU, L. (2015). Two-sample Behrens–Fisher problem for high-dimensional data. *Statist. Sinica* **25** 1297–1312. MR3409068

GREGORY, K. B., CARROLL, R. J., BALADANDAYUTHAPANI, V. and LAHIRI, S. N. (2015). A two-sample test for equality of means in high dimension. *J. Amer. Statist. Assoc.* **110** 837–849. MR3367268

HETTMANSPERGER, T. P. and MCKEAN, J. W. (2011). *Robust Nonparametric Statistical Methods*, 2nd ed. *Monographs on Statistics and Applied Probability* **119**. CRC Press, Boca Raton, FL. MR2779026

IBRAGIMOV, I. A. and LINNIK, Y. V. (1971). *Independent and Stationary Sequences of Random Variables*. Wolters-Noordhoff Publishing, Groningen. MR0322926

KALLENBERG, O. (2005). *Probabilistic Symmetries and Invariance Principles*. Springer, New York. MR2161313

KATAYAMA, S. and KANO, Y. (2014). A new test on high-dimensional mean vector without any assumption on population covariance matrix. *Comm*. *Statist*. *Theory Methods* **43** 5290–5304. MR3282878

KOLMOGOROV, A. N. and ROZANOV, J. A. (1960). On a strong mixing condition for stationary Gaussian processes. *Teor. Verojatnost*. *i Primenen*. **5** 222–227. MR0133175

LIN, Z. and LU, C. (1996). *Limit Theory for Mixing Dependent Random Variables*. *Mathematics and Its Applications* **378**. Kluwer, Dordrecht. MR1486580

MARDEN, J. I. (1999). Multivariate rank tests. In *Multivariate Analysis*, *Design of Experiments*, *and Survey Sampling*. *Statist*. *Textbooks Monogr.* **159** 401–432. Dekker, New York. MR1719102

MÖTTÖNEN, J. and OJA, H. (1995). Multivariate spatial sign and rank methods. *J*. *Nonparametr. Stat*. **5** 201–213. MR1346895

MÖTTÖNEN, J., OJA, H. and TIENARI, J. (1997). On the efficiency of multivariate spatial sign and rank tests. *Ann*. *Statist*. **25** 542–552. MR1439313

OJA, H. (2010). *Multivariate Nonparametric Methods with R*: *An Approach Based on Spatial Signs and Ranks*. *Lecture Notes in Statistics* **199**. Springer, New York. MR2598854

PURI, M. L. and SEN, P. K. (1971). *Nonparametric Methods in Multivariate Analysis*. Wiley, New York. MR0298844

SRIVASTAVA, M. S., KATAYAMA, S. and KANO, Y. (2013). A two sample test in high dimensional data. *J*. *Multivariate Anal*. **114** 349–358. MR2993891

WANG, L., PENG, B. and LI, R. (2015). A high-dimensional nonparametric multivariate test for mean vector. *J*. *Amer. Statist*. *Assoc*. **110** 1658–1669. MR3449062

WEI, S., LEE, C., WICHERS, L. and MARRON, J. S. (2016). Direction-projection-permutation for high-dimensional hypothesis tests. *J*. *Comput*. *Graph*. *Statist*. **25** 549–569. MR3499694

THEORETICAL STATISTICS AND MATHEMATICS UNIT
INDIAN STATISTICAL INSTITUTE
203, B. T. ROAD,
KOLKATA—700 108
INDIA
E-MAIL: vanchak@gmail.com
probal@isical.ac.in