Tests for high dimensional data based on means, spatial signs and spatial ranks

Anirvan Chakraborty and Probal Chaudhuri

Theoretical Statistics and Mathematics Unit,
Indian Statistical Institute
203, B. T. Road, Kolkata - 700108, INDIA.
emails: vanchak@gmail.com, probal@isical.ac.in

Abstract

Tests based on sample mean vectors and sample spatial signs have been studied in the recent literature for high dimensional data with the dimension larger than the sample size. For suitable sequences of alternatives, we show that the powers of the mean based tests and the tests based on spatial signs and ranks tend to be same as the data dimension grows 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 setup. 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 grows 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.

Keywords: ARMA processes, heavy tailed distributions, permutation tests, ρ-mixing, randomly scaled ρ-mixing, spherical distributions, stationary sequences

1 Introduction

For univariate data, nonparametric tests based on signs and ranks are well-known competitors of tests based on sample means like the t-test. These nonparametric tests have 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 Puri and Sen (1971), Oja (2010) and Hettmansperger and McKean (2011)), they do not have the distribution-free property in general, 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 et al. (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 $\mathbb{R}^d$. For testing $H_0 : E(\mathbf{X}) = E(\mathbf{Y})$ against the alternative $H_A : E(\mathbf{X}) \neq E(\mathbf{Y})$ 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

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

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

respectively, where $(p)_q = p(p-1)\ldots(p-q+1)$ for integers $p \geq 1$ and $1 \leq q < p$.

Well known multivariate spatial sign and rank based tests (see Möttönen and Oja (1995), Möttönen et al. (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 et al. (2015) proposed a one sample test of the mean vector based on spatial signs given by

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

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 et al. (2015), and it is given by

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

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

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

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 tests 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 grows 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 subsection 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 grow 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 et al. (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 to grow to infinity. In Section 4, we demonstrate the performances of the tests based on sample means and spatial signs and ranks using some real 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 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 Appendix – I.

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 $\rho(d) = \sup_{k\geq 1} \sup_{f\in\mathcal{F}_k, g\in\mathcal{F}_{d+k}} |Corr(f,g)|$ converges to zero as $d \to \infty$. Here, $\rho(\cdot)$ is called the ρ-mixing* *coefficient of $\mathcal{X}$, and $\mathcal{F}_k$ denotes the σ-field generated by measurable square integrable functions of $(X_1, X_2, \ldots, X_k)$ for $k \geq 1$.*

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 the 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.

Denote $\mu = \mu_2 - \mu_1$, $\sigma_1^2 = Var(X_1) > 0$, $\sigma_2^2 = Var(Y_1) > 0$, $\Sigma_1 = Disp(\mathbf{X})$, and $\Sigma_2 = Disp(\mathbf{Y})$, 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 $\mu'(\Sigma_1 + \Sigma_2)\mu = o(tr(\Sigma_1^2 + \Sigma_2^2))$ as $d \to \infty$.

Examples of ρ-mixing sequences include m-dependent sequences, stationary ARMA(p,q) processes with 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)). 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 $d \to \infty$, where $\lambda_1 < \lambda_2 < \ldots < \lambda_d$ 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 p. 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, p. 811), e.g., distributions with independent coordinates, moving average processes and more generally m-dependent sequences as well as 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 et al. (2013) studied a two sample mean based 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. (2014), 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, those authors required the existence of sixteenth order moments. Cai et al. (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 et al. (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. (2015) studied the properties of their test under spherical Gaussian distributions, which are special cases of the ρ-mixing models considered here.

Theorem 2.1. *Suppose that conditions (C1)–(C3) are satisfied. Define,* $\Gamma_1 = 2tr(\Sigma_1^2)/(m)_2 + 2tr(\Sigma_2^2)/(n)_2 + 4tr(\Sigma_1\Sigma_2)/(mn)$. *Then, each of* $[d(\sigma_1^2+\sigma_2^2)T_{WMW} - ||\mu||^2]/{\Gamma_1}^{1/2}$ *and* $(T_{CQ}^{(2)} - ||\mu||^2)/{\Gamma_1}^{1/2}$ *converges weakly to a standard Gaussian variable as* $d \to \infty$ *for every fixed* $m, n \geq 1$.

When the null hypothesis $H_0 : \mu = \mathbf{0}$ is true, the above theorem yields the asymptotic null distributions of T_{WMW} and $T_{CQ}^{(2)}$ as $d \to \infty$. Let us observe that the asymptotic distribution of $T_{CQ}^{(2)}$ obtained in the above theorem as $d \to \infty$ is the same as that obtained by Chen and Qin (2010) in their Theorem 1 when both $d, n \to \infty$. These authors used an assumption similar to that in the second part of condition (C3) for deriving the asymptotic distribution of their test statistic, when both d and n are large (see (3.4) in p. 812 in Chen and Qin (2010)).

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 $T_{CQ}^{(2)}$ for high dimensional data. Let $\beta_{T_{WMW}}(\mu)$ and $\beta_{T_{CQ}^{(2)}}(\mu)$ be the powers of these two tests at a given level of significance.

Theorem 2.2. *Suppose that conditions (C1)–(C3) are satisfied, and assume* $\lim_{d\to\infty} ||\mu||^2/\Gamma_1^{1/2} = c$ *for some* $c \in [0, \infty]$. *Then,* $\lim_{d\to\infty} \beta_{T_{WMW}}(\mu) = \lim_{d\to\infty} \beta_{T_{CQ}^{(2)}}(\mu) = \beta$ *for every fixed* $m, n \geq 1$, *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 $m, n \geq 1$. 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 $T_{CQ}^{(2)}$ 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 $T_{CQ}^{(2)}$ under the ρ-mixing setup, we can use their limiting null distributions obtained from Theorem 2.1 after plugging-in the following unbiased 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
$$\widehat{\mathrm{tr}(\Sigma_1^2)} = \frac{1}{4(m)_4}\sum_{\substack{i_1,i_2,i_3,i_4\\ all\ distinct}} [(\mathbf{X}_{i_1}-\mathbf{X}_{i_2})'(\mathbf{X}_{i_3}-\mathbf{X}_{i_4})]^2,$$

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

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

Also, $\widehat{\sigma_1^2} = [d(m-1)]^{-1}\sum_{k=1}^d\sum_{i=1}^m (X_{ik}-\overline{X}_k)^2$, 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 unlike the estimator proposed by Chen and Qin (2010, p. 815). Moreover, for all simulated datasets and real datasets considered later, the empirical sizes and powers of the test based on $T_{CQ}^{(2)}$ 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 $T_{CQ}^{(2)}$, we have considered the $AR(1)$ models with correlation 0.7 having Gaussian and $t(5)$ innovations. The sample sizes are $m = n = 20$, and $\mu = (c, 0, 0, \ldots, 0)$ with $c = 1.5, 3, 4.5, 6, 7.5$ for $d = 100, 200, 400, 800, 1600$, respectively. The sizes and the powers of the tests based on T_{WMW} and $T_{CQ}^{(2)}$ 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 the models. It is seen from Figure 1 that the powers of these two tests are similar for all data dimensions considered under both the models. The power curves are so close that they are overlaid on each other.

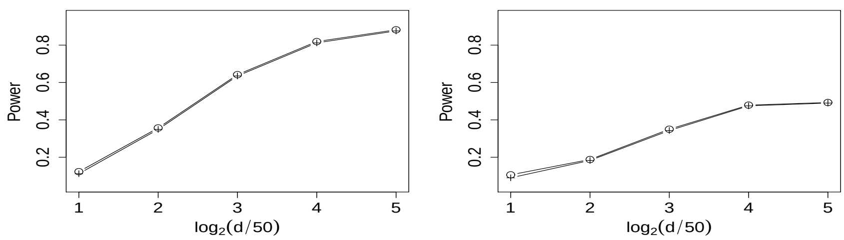

Figure 1: Powers of the tests at nominal 5% level based on T_{WMW} (- + - curves) and $T_{CQ}^{(2)}$ (- ∘ - curves) for the $AR(1)$ model with Gaussian innovation (left panel) and $t(5)$ innovation (right panel). The two power curves are overlaid on each other in both the plots.

2.2 Asymptotic behaviours of one sample tests under ρ-mixing

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 gives the asymptotic distributions of T_S, T_{SR} and $T_{CQ}^{(1)}$ and compares their asymptotic powers, when the data dimension is large. Denote $\beta_{T_S}(\mu)$, $\beta_{T_{SR}}(\mu)$ and $\beta_{T_{CQ}^{(1)}}(\mu)$ to be the powers of the tests based on T_S, T_{SR} and $T_{CQ}^{(1)}$ at a given level of significance, when the alternative hypothesis $H_A : \mu = \mathbf{0}$ 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* $\mu'\Sigma\mu = o(tr(\Sigma^2))$ *as* $d \to \infty$*, where* $\Sigma = Disp(\mathbf{X})$.

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). Define* $\Gamma_2 = 2tr(\Sigma^2)/(n)_2$*, and* $\sigma^2 = Var(X_1)$*, where* $\mathbf{X} = (X_1, X_2, \ldots, X_d)$.

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

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

We get the limiting null distributions of T_S, T_{SR} and $T_{CQ}^{(1)}$ when $\mu = \mathbf{0}$ in the above theorem. When both the data dimension and the sample size grow to infinity, Wang et al. (2015) proved that the test based on T_S is asymptotically as powerful as the test based on $T_{CQ}^{(1)}$ 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 $T_{CQ}^{(1)}$ stated in part (b) of our Theorem 2.3 holds for any sample size and for many non-spherical distributions.

Remark 2.1. *In both the one and the two sample problems, when our* ρ*-mixing model for the data holds, the equality of the limiting powers of the tests based on sample means and the tests based on spatial signs and ranks, 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, where 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 non-degenerate 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. For instance, the infinite sequence of random variables associated with the multivariate spherical t distribution has this property. In fact, by Theorem 1.31 in Kallenberg (2005), it follows that any rotatable sequence $\mathcal{X}$, i.e., 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 non-negative random variable independent of $\mathcal{R}$. 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 non-negative 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 et al. (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 et al. (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 $T_{CQ}^{(2)}$ 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 Disp$(\mathbf{V})$ and 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 $(dT_{WMW} - ||\mu||^2 S_1)/S_2^{1/2}$ and $(T_{CQ}^{(2)} - ||\mu||^2)/S_3^{1/2}$ converges weakly to a standard Gaussian variable as $d \to \infty$ for every $m, n \ge 1$. Consequently, for every fixed $m, n \ge 1$, the distributions of T_{WMW} and $T_{CQ}^{(2)}$ 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, and $||\mu||^2/d^{1/2}$ tends to a finite non-negative limit as $d \to \infty$. Then, there exist real numbers ψ_1 and ψ_2 such that $\lim_{m,n\to\infty} \lim_{d\to\infty} P\{(dT_{WMW} - ||\mu||^2\psi_1)/\psi_2^{1/2} \le x\} = \lim_{m,n\to\infty} \lim_{d\to\infty} P\{(T_{CQ}^{(2)} - ||\mu||^2)/\Gamma_1^{1/2} \le x\} = \Phi(x)$ 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 $T_{CQ}^{(2)}$ 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 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 $T_{CQ}^{(2)}$ 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 $\beta_{T_{WMW}}(\mu)$ and $\beta_{T_{CQ}^{(2)}}(\mu)$ denote the powers of the tests based on T_{WMW} and $T_{CQ}^{(2)}$ 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,* $\lim_{m,n\to\infty} \lim_{d\to\infty} \beta_{T_{WMW}}(\mu) > \lim_{m,n\to\infty} \lim_{d\to\infty} \beta_{T_{CQ}^{(2)}}(\mu)$.

If $\lim_{m,n\to\infty} \lim_{d\to\infty} ||\mu||^2/\Gamma_1^{1/2}$ equals zero (respectively, infinity), then the asymptotic powers of the tests based on T_{WMW} and $T_{CQ}^{(2)}$ in the setup of Theorem 3.2 coincide, and they are both equal to the nominal level (respectively, 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 $T_{CQ}^{(2)}$ 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 $T_{CQ}^{(2)}$ 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. (2015) 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 $T_{CQ}^{(2)}$ 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 subsection 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 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 $T_{CQ}^{(2)}$ adequately approximates their true distributions. Also, the test based on T_{WMW} significantly outperforms the test based on $T_{CQ}^{(2)}$, which conforms with the result in Theorem 3.2.

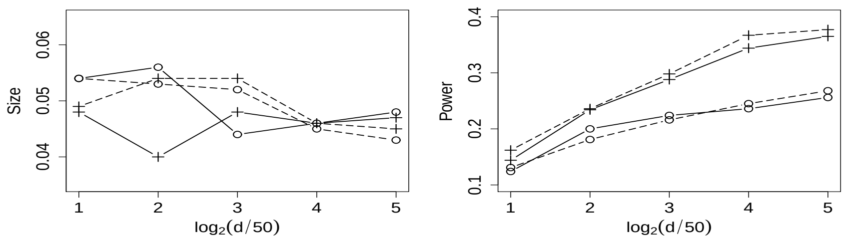

Figure 2: Empirical sizes and powers of the tests based on T_{WMW} (+) and $T_{CQ}^{(2)}$ ($\circ$) at nominal 5% level for the spherical $t(5)$ distribution using the permutation implementation (solid curves) and the true implementation (dashed curves).

3.2 Asymptotic behaviours of one sample tests under stronger dependence

We will now study the asymptotic distributions of the one sample tests considered in subsection 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 $T_{CQ}^{(1)}$ and yields their asymptotic powers. As earlier, we can write $\mathbf{X}_i = \mu + \mathbf{V}_i/P_i$, $1 \leq i \leq n$. Also, $\beta_{T_S}(\mu)$, $\beta_{T_{SR}}(\mu)$ and $\beta_{T_{CQ}^{(1)}}(\mu)$ denote the powers of the tests based on T_S, T_{SR} and $T_{CQ}^{(1)}$ 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 $Disp(\widetilde{\mathbf{V}})$.*
(a) There exist $\Gamma_3 > 0$ and random variables Z_k, $1 \leq k \leq 4$, which are functions of the P_i's, such that each of $(dT_S - ||\mu||^2 Z_1)/\Gamma_3^{1/2}$, $(dT_{SR} - 2||\mu||^2 Z_2)/(2Z_3^{1/2})$ and $(T_{CQ}^{(1)} - ||\mu||^2)/Z_4^{1/2}$ converges weakly to a standard Gaussian variable as $d \to \infty$ for each $m, n \geq 1$. Consequently, for each fixed $m, n \geq 1$, the distributions of T_S, T_{SR} and $T_{CQ}^{(1)}$ 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 non-negative limit as $d \to \infty$. Define $\sigma^2 = Var(X_1)$. There exist real numbers θ_k, $1 \leq k \leq 3$ such that $\lim_{n\to\infty} \lim_{d\to\infty} P\{(d\sigma^2 T_S - ||\mu||^2\theta_1)/\Gamma_2^{1/2} \leq x\} = \lim_{n\to\infty} \lim_{d\to\infty} P\{(d\sigma^2 T_{SR} - 2||\mu||^2\theta_2)/(2\theta_3^{1/2}) \leq x) = \lim_{n\to\infty} \lim_{d\to\infty} P\{(T_{CQ}^{(1)} - ||\mu||^2)/\Gamma_2^{1/2} \leq x) = \Phi(x)$ for all $x \in \mathbb{R}$.

Here, 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 $\lim_{n\to\infty}\lim_{d\to\infty}\beta_{T_S}(\mu)>\lim_{n\to\infty}\lim_{d\to\infty}\beta_{T_{CQ}^{(1)}}(\mu)$. We also have $\lim_{n\to\infty}\lim_{d\to\infty}\beta_{T_{SR}}(\mu)>\lim_{n\to\infty}\lim_{d\to\infty}\beta_{T_{CQ}^{(1)}}(\mu)$.

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), i.e., its limiting null distribution is Gaussian irrespective of whether the sample size grows 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 $T_{CQ}^{(1)}$ is the same as that obtained in Theorem 3.2 under the ρ-mixing setup, and it coincides with the asymptotic distribution of $T_{CQ}^{(1)}$ obtained by Chen and Qin (2010). For the spherical t distribution, which is a distribution included in our RSRM models, Wang et al. (2015) derived the asymptotic distribution of $T_{CQ}^{(1)}$ 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 (respectively, infinity), then the asymptotic powers of the tests based on T_S, T_{SR} and $T_{CQ}^{(1)}$ coincide, and they are all equal to the nominal level (respectively, 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 non-negative 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 $T_{CQ}^{(2)}$ if we first let the dimension and then the sample sizes grow 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 sample sizes if only the dimension grows 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 $T_{CQ}^{(1)}$ if we first let the dimension and then the sample size grow 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 sample size if only the dimension grows to infinity.*

4 Analysis of real data

We now investigate the performances of the two sample tests based on T_{WMW} and $T_{CQ}^{(2)}$ on some real datasets, when they are implemented in two different ways, namely, as in the ρ-mixing setup described in subsection 2.2, and using their permutation distributions. 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, which is obtained from `http://datam.i2r.a-star.edu.sg/datasets/krbd/ColonTumor/ColonTumor.zip` and 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. To estimate the sizes of the tests based on T_{WMW} and $T_{CQ}^{(2)}$ 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, respectively. These choices are made to ensure that the resulting datasets remain high dimensional, and 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 the four datasets. Further, the powers of the tests based on T_{WMW} and $T_{CQ}^{(2)}$ 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 $T_{CQ}^{(2)}$ for the Colon data and the Sonar data.

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

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

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 et al. (2013) and Gregory et al. (2014) by T_{SKK} and T_{GCBL}, respectively. For the $AR(1)$ models in subsection 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 the size of this test converges to one as the dimension and the sample sizes grow to infinity at a certain rate for a class of models, which include these $AR(1)$ models. Under the spherical $t(5)$ model in subsection 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} (respectively, $T_{CQ}^{(2)}$) under all the above models (respectively, $AR(1)$ models) but they outperform the test based on $T_{CQ}^{(2)}$ under the spherical $t(5)$ model. The readers are referred to Appendix – II for more details.

Cai et al. (2014) showed that their test has better power than other tests based on sum of squares of coordinatewise mean difference or coordinatewise t statistics, when the mean shift has only a few non-zero coordinates. However, we observed that this test becomes significantly less powerful than the tests based on T_{WMW} and $T_{CQ}^{(2)}$, when the mean shifts in the models considered in subsections 2.1 and 3.1 are distributed equally among all the coordinates. Moreover, the size of the test in Cai et al. (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 et al. (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 Appendix – II.

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 test based on $T_{CQ}^{(2)}$ would be asymptotically incorrect. To compare the performance of the tests based on T_{WMW} and $T_{CQ}^{(2)}$ 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 $T_{CQ}^{(2)}$ for all values of d (see Appendix – II).

Appendix – I

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

$$\begin{aligned} ||\mathbf{X}-\mathbf{Y}||^2 &= [||\mathbf{X}||^2 + ||\mathbf{Y}-\mu||^2 - 2\mathbf{X}'(\mathbf{Y}-\mu) + 2\mu'(\mathbf{X}-\mathbf{Y}+\mu) + ||\mu||^2] \\ &= \sum_{k=1}^{d}[V_k^2 + W_k^2 - 2V_kW_k] + 2\mu'(\mathbf{V}-\mathbf{W}) + ||\mu||^2, \end{aligned} \tag{5.1}$$

It follows from Bradley (2005, Theorem 5.2(b)) that for any function $h : \mathbb{R}^2 \to \mathbb{R}$, the sequence $(h(V_k, W_k) : k \geq 1)$ is ρ-mixing with its mixing coefficient bounded by $\max\{\rho_1(\cdot), \rho_2(\cdot)\}$. This fact, (5.1) above 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

$$||\mathbf{X}-\mathbf{Y}||^2/d - (\sigma_1^2 + \sigma_2^2) = o(d^{-1/2+\epsilon}) \tag{5.2}$$

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

$$\begin{aligned} T_{WMW} &= \frac{1}{(m)_2(n)_2} \sum_{i_1 \neq i_2} \sum_{j_1 \neq j_2} \frac{(\mathbf{X}_{i_1}-\mathbf{Y}_{j_1})'(\mathbf{X}_{i_2}-\mathbf{Y}_{j_2})}{d(\sigma_1^2+\sigma_2^2)} \\ &\quad + \frac{1}{(m)_2(n)_2} \sum_{i_1 \neq i_2} \sum_{j_1 \neq j_2} \left[\frac{(\mathbf{X}_{i_1}-\mathbf{Y}_{j_1})'(\mathbf{X}_{i_2}-\mathbf{Y}_{j_2})}{d(\sigma_1^2+\sigma_2^2)} \times \left\{ \frac{d(\sigma_1^2+\sigma_2^2)}{||\mathbf{X}_{i_1}-\mathbf{Y}_{j_1}||\;||\mathbf{X}_{i_2}-\mathbf{Y}_{j_2}||} - 1 \right\} \right] \\ &= (T_{CQ}^{(2)} + T_{WMW}^{(2)})/\{d(\sigma_1^2+\sigma_2^2)\}, \end{aligned} \tag{5.3}$$

where $T^{(2)}_{CQ} = [(m)_2(n)_2]^{-1}\sum_{i_1\neq i_2}\sum_{j_1\neq j_2}(\mathbf{X}_{i_1}-\mathbf{Y}_{j_1})'(\mathbf{X}_{i_2}-\mathbf{Y}_{j_2})$ as defined in the Introduction, and

$$T^{(2)}_{WMW} = \frac{1}{(m)_2(n)_2}\sum_{i_1\neq i_2}\sum_{j_1\neq j_2}\left[\frac{(\mathbf{X}_{i_1}-\mathbf{Y}_{j_1})'(\mathbf{X}_{i_2}-\mathbf{Y}_{j_2})}{d(\sigma_1^2+\sigma_2^2)}\times\left\{\frac{d(\sigma_1^2+\sigma_2^2)}{||\mathbf{X}_{i_1}-\mathbf{Y}_{j_1}||\,||\mathbf{X}_{i_2}-\mathbf{Y}_{j_2}||}-1\right\}\right].$$

So, $E(T^{(2)}_{CQ}) = ||\mu||^2$. Further, it follows from Chen and Qin (2010, p. 825) that $Var(T^{(2)}_{CQ}) = \Gamma_1 + 4\mu'\Sigma_1\mu/m + 4\mu'\Sigma_2\mu/n$, where $\Gamma_1 = 2\text{tr}(\Sigma_1^2)/(m)_2 + 2\text{tr}(\Sigma_2^2)/(n)_2 + 4\text{tr}(\Sigma_1\Sigma_2)/(mn)$ is defined in the statement of the theorem. Note that $(\mu'\Sigma_1\mu/m) + (\mu'\Sigma_2\mu/n) \leq \mu'(\Sigma_1+\Sigma_2)\mu/\min(m,n)$. Also, the denominator of each of the three terms in Γ_1 is less than $(N)_2$, where $N = \max(m,n)$. This implies that $\Gamma_1 \geq [2\text{tr}(\Sigma_1^2) + 2\text{tr}(\Sigma_2^2) + 4\text{tr}(\Sigma_1\Sigma_2)]/(N)_2 = 2\text{tr}[(\Sigma_1+\Sigma_2)^2]/(N)_2$. These facts and Assumption (C3) imply that $Var(T^{(2)}_{CQ}) = \Gamma_1(1+o(1))$ as $d\to\infty$. Further,

$$\begin{aligned} T^{(2)}_{CQ} - ||\mu||^2 &= \frac{1}{(m)_2(n)_2}\sum_{i_1\neq i_2}\sum_{j_1\neq j_2}(\mathbf{X}_{i_1}-\mathbf{Y}_{j_1}+\mu)'(\mathbf{X}_{i_2}-\mathbf{Y}_{j_2}+\mu) - \frac{2}{mn}\sum_{i,j}\mu'(\mathbf{X}_i-\mathbf{Y}_j+\mu) \\ &= T_1 - T_2, \end{aligned}$$

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 $Var(T_2) = 4\mu'[(\Sigma_1/m)+(\Sigma_2/n)]\mu$. So, using the inequality $\Gamma_1 \geq 2\text{tr}[(\Sigma_1+\Sigma_2)^2]/(N)_2$, 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

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

So, $E(T_1) = 0$ and $Var(T_1) = \Gamma_1$. This follows from computations similar to those used in deriving $Var(T^{(2)}_{CQ})$ 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 $m,n\geq 1$. This and the fact that $T^{(2)}_2$ converges to zero *in probability* as $d\to\infty$ for each fixed $m,n\geq 1$ together imply that

$$(T^{(2)}_{CQ} - ||\mu||^2)/\Gamma_1^{1/2} \xrightarrow{\mathcal{L}} N(0,1) \tag{5.4}$$

as $d\to\infty$ for each fixed $m,n\geq 1$. Next, let us write

$$\begin{aligned} T^{(2)}_{WMW}/\Gamma_1^{1/2} &= \frac{1}{(m)_2(n)_2}\sum_{i_1\neq i_2}\sum_{j_1\neq j_2}\left[\frac{(\mathbf{X}_{i_1}-\mathbf{Y}_{j_1})'(\mathbf{X}_{i_2}-\mathbf{Y}_{j_2}) - ||\mu||^2}{\Gamma_1^{1/2}}\times\right. \\ &\qquad\left.\left\{\frac{d(\sigma_1^2+\sigma_2^2)}{||\mathbf{X}_{i_1}-\mathbf{Y}_{j_1}||\,||\mathbf{X}_{i_2}-\mathbf{Y}_{j_2}||}-1\right\}\right] \\ &\quad + \frac{||\mu||^2}{(m)_2(n)_2\Gamma_1^{1/2}}\sum_{i_1\neq i_2}\sum_{j_1\neq j_2}\left\{\frac{d(\sigma_1^2+\sigma_2^2)}{||\mathbf{X}_{i_1}-\mathbf{Y}_{j_1}||\,||\mathbf{X}_{i_2}-\mathbf{Y}_{j_2}||}-1\right\} \\ &= T^{(3)}_{WMW} + T^{(4)}_{WMW}, \end{aligned} \tag{5.5}$$

where

$$T_{WMW}^{(3)} = \frac{1}{(m)_2(n)_2} \sum_{i_1 \neq i_2} \sum_{j_1 \neq j_2} \left[\frac{(\mathbf{X}_{i_1} - \mathbf{Y}_{j_1})'(\mathbf{X}_{i_2} - \mathbf{Y}_{j_2}) - ||\mu||^2}{\Gamma_1^{1/2}} \times \right.$$
$$\left. \left\{ \frac{d(\sigma_1^2 + \sigma_2^2)}{||\mathbf{X}_{i_1} - \mathbf{Y}_{j_1}||\ ||\mathbf{X}_{i_2} - \mathbf{Y}_{j_2}||} - 1 \right\} \right] \quad \text{and}$$
$$T_{WMW}^{(4)} = \frac{||\mu||^2}{(m)_2(n)_2\Gamma_1^{1/2}} \sum_{i_1 \neq i_2} \sum_{j_1 \neq j_2} \left\{ \frac{d(\sigma_1^2 + \sigma_2^2)}{||\mathbf{X}_{i_1} - \mathbf{Y}_{j_1}||\ ||\mathbf{X}_{i_2} - \mathbf{Y}_{j_2}||} - 1 \right\}.$$

As mentioned earlier, $\Gamma_1 \geq 2\text{tr}[(\Sigma_1 + \Sigma_2)^2]/(N)_2$. Also, from the stationarity of the sequences $\mathcal{X}$ and $\mathcal{Y}$ and using the Cauchy-Schwarz inequality, it follows that $\text{tr}[(\Sigma_1+\Sigma_2)^2] \geq d(\sigma_1^2+\sigma_2^2)^2$. These facts along with (5.2) and Assumption (C3) imply that each term inside the double summation appearing in the definition of $T_{WMW}^{(4)}$ converges to zero *in probability* as $d \to \infty$ for each fixed $m, n \geq 1$. So, $T_{WMW}^{(4)}$ converges to zero *in probability* as $d \to \infty$ for each fixed $m, n \geq 1$.

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 $T_{WMW}^{(3)}$. It follows from (5.2) 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

$$(\mathbf{X}_{i_1} - \mathbf{Y}_{j_1})'(\mathbf{X}_{i_2} - \mathbf{Y}_{j_2}) - ||\mu||^2 = (\mathbf{X}_{i_1} - \mathbf{Y}_{j_1} + \mu)'(\mathbf{X}_{i_2} - \mathbf{Y}_{j_2} + \mu) \tag{5.6}$$
$$- \mu'(\mathbf{X}_{i_1} - \mathbf{Y}_{j_1} + \mu) - \mu'(\mathbf{X}_{i_2} - \mathbf{Y}_{j_2} + \mu).$$

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 (5.6) is asymptotically Gaussian with zero mean and variance $2\text{tr}[(\Sigma_1+\Sigma_2)^2]$ 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 (5.6) after scaling by $\Gamma_1^{1/2}$ converge to zero *in probability* as $d \to \infty$. So, the left hand side of (5.6) after scaling by $\Gamma_1^{1/2}$ converges *weakly* to a Gaussian distribution as $d \to \infty$. Thus, $T_{WMW}^{(3)}$ converges to zero *in probability* as $d \to \infty$ for each fixed $m, n \geq 1$. This and the fact that $T_{WMW}^{(4)}$ converges to zero *in probability* as $d \to \infty$ together imply that $T_{WMW}^{(2)}/\Gamma_1^{1/2}$ converges to zero *in probability* as $d \to \infty$ for each fixed $m, n \geq 1$. Combining this fact with (5.3) and (5.4) yields

$$\{d(\sigma_1^2 + \sigma_2^2)T_{WMW} - ||\mu||^2\}/\Gamma_1^{1/2} \xrightarrow{\mathcal{L}} N(0,1) \tag{5.7}$$

as $d \to \infty$ for each fixed $m, n \geq 1$. □

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

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

and

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

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 $T_{CQ}^{(2)}$ follows from (5.4) and (5.7). 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 $T_{CQ}^{(1)}$ 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

$$\begin{aligned}
T_{SR} &= \frac{1}{(n)_4} \sum_{i_1\neq i_2\neq i_3\neq i_4} \frac{(\mathbf{X}_{i_1} + \mathbf{X}_{i_2})'(\mathbf{X}_{i_3} + \mathbf{X}_{i_4})}{2d\sigma^2} \\
&\quad + \frac{1}{(n)_4} \sum_{i_1\neq i_2\neq i_3\neq i_4} \left[\frac{(\mathbf{X}_{i_1} + \mathbf{X}_{i_2})'(\mathbf{X}_{i_3} + \mathbf{X}_{i_4})}{2d\sigma^2} \times \left\{\frac{2d\sigma^2}{||\mathbf{X}_{i_1} + \mathbf{X}_{i_2}||\ ||\mathbf{X}_{i_3} + \mathbf{X}_4||} - 1\right\}\right] \\
&= \frac{2}{(n)_2} \sum_{i_1\neq i_2} \frac{\mathbf{X}_{i_1}'\mathbf{X}_{i_2}}{d\sigma^2} + \frac{1}{(n)_4} \sum_{i_1\neq i_2\neq i_3\neq i_4} \left[\frac{(\mathbf{X}_{i_1} + \mathbf{X}_{i_2})'(\mathbf{X}_{i_3} + \mathbf{X}_{i_4})}{2d\sigma^2} \times \right. \qquad (5.8) \\
&\qquad\qquad \left. \left\{\frac{2d\sigma^2}{||\mathbf{X}_{i_1} + \mathbf{X}_{i_2}||\ ||\mathbf{X}_{i_3} + \mathbf{X}_4||} - 1\right\}\right]
\end{aligned}$$

The first term in (5.8) equals $2T_{CQ}^{(1)}/(d\sigma^2)$. Using Assumption (C4), it can be shown that $E(T_{CQ}^{(1)}) = ||\mu||^2$ and $Var(T_{CQ}^{(1)}) = \Gamma_2(1 + o(1))$ as $d \to \infty$. Using arguments similar to those in the proof of Theorem 2.1, we have the *weak convergence* of $(T_{CQ}^{(1)} - ||\mu||^2)/\psi_2^{1/2}$ to a standard Gaussian distribution. Further, the second term in (5.8) after scaling by $\Gamma_2^{1/2}$ converges to zero *in probability* as $d \to \infty$ for each $n \geq 1$. The previous two statements together imply that $(d\sigma^2 T_{SR} - 2||\mu||^2)/\psi_2^{1/2}$ converges *weakly* to a $N(0, 4)$ distribution as $d \to \infty$ for each $n \geq 1$.

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

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 $\mathbf{V} = (V_1, V_2, \ldots, V_d)'$ and $\mathbf{W} = (W_1, W_2, \ldots, W_d)'$. Denote $\Sigma_V = Disp(\mathbf{V})$, $\Sigma_W = Disp(\mathbf{W})$, $\sigma_V^2 = Var(V_1)$ and $\sigma_W^2 = Var(V_2)$.

(a) We will first derive the asymptotic distribution of T_{WMW}. Using similar arguments as those used in proving (5.1), we get

$$||\mathbf{X} - \mathbf{Y}||^2 = \sum_{k=1}^{d} \left[\frac{V_k^2}{P^2} + \frac{W_k^2}{Q^2} - \frac{2V_kW_k}{PQ}\right] + 2\mu'\left(\frac{\mathbf{V}}{P} - \frac{\mathbf{W}}{Q}\right) + ||\mu||^2. \qquad (5.9)$$

Consider the event $E = \{||\mathbf{X}-\mathbf{Y}||^2/d-(\sigma_V^2/P^2+\sigma_W^2/Q^2) = o(d^{-1/2+\epsilon})$ 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 (5.9) above along with the assumptions in the theorem and Theorem 8.2.2 in Lin and Lu (1996), we get that for any given $\epsilon \in (0, 1/2)$,

$$Pr(E|P,Q) = 1 \tag{5.10}$$

for *almost every* P and Q. Now,

$$\begin{aligned} T_{WMW} &= \frac{1}{(m)_2(n)_2}\sum_{i_1\neq i_2}\sum_{j_1\neq j_2}\frac{(\mathbf{X}_{i_1}-\mathbf{Y}_{j_1})'(\mathbf{X}_{i_2}-\mathbf{Y}_{j_2})}{d(\sigma_V^2P_{i_1}^{-2}+\sigma_W^2Q_{j_1}^{-2})^{1/2}(\sigma_V^2P_{i_2}^{-2}+\sigma_W^2Q_{j_2}^{-2})^{1/2}} \\ &\quad + \frac{1}{(m)_2(n)_2}\sum_{i_1\neq i_2}\sum_{j_1\neq j_2}\left[\frac{(\mathbf{X}_{i_1}-\mathbf{Y}_{j_1})'(\mathbf{X}_{i_2}-\mathbf{Y}_{j_2})}{d(\sigma_V^2P_{i_1}^{-2}+\sigma_W^2Q_{j_1}^{-2})^{1/2}(\sigma_V^2P_{i_2}^{-2}+\sigma_W^2Q_{j_2}^{-2})^{1/2}}\times\right. \\ &\quad \left.\left\{\frac{d(\sigma_V^2P_{i_1}^{-2}+\sigma_W^2Q_{j_1}^{-2})^{1/2}(\sigma_V^2P_{i_2}^{-2}+\sigma_W^2Q_{j_2}^{-2})^{1/2}}{||\mathbf{X}_{i_1}-\mathbf{Y}_{j_1}||\ ||\mathbf{X}_{i_2}-\mathbf{Y}_{j_2}||}-1\right\}\right]. \\ &= (T_{WMW}^{(1)}+T_{WMW}^{(2)})/d, \end{aligned} \tag{5.11}$$

where

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

and

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

Some straightforward algebra yields

$$\begin{aligned} T_{WMW}^{(1)} &= \frac{1}{d(m)_2(n)_2}\left\{\sum_{i_1\neq i_2}A_{i_1,i_2}\mathbf{X}_{i_1}'\mathbf{X}_{i_2} - 2\sum_{i,j}C_{i,j}\mathbf{X}_i'\mathbf{Y}_j + \sum_{j_1\neq j_2}B_{j_1,j_2}\mathbf{Y}_{j_1}'\mathbf{Y}_{j_2}\right\} \\ &= \frac{1}{d(m)_2(n)_2}\left\{\sum_{i_1\neq i_2}[P_{i_1}P_{i_2}]^{-1}A_{i_1,i_2}\mathbf{V}_{i_1}'\mathbf{V}_{i_2} - 2\sum_{i,j}[P_iQ_j]^{-1}C_{i,j}\mathbf{V}_i'(Q_j\mu+\mathbf{W}_j)\right. \\ &\quad \left. + \sum_{j_1\neq j_2}[Q_{j_1}Q_{j_2}]^{-1}B_{j_1,j_2}(Q_{j_1}\mu+\mathbf{W}_{j_1})'(Q_{j_2}\mu+\mathbf{W}_{j_2})\right\}, \end{aligned} \tag{5.12}$$

where $A_{i_1,i_2} = \sum_{j_1\neq j_2}(\sigma_V^2P_{i_1}^{-2}+\sigma_W^2Q_{j_2}^{-2})^{-1/2}(\sigma_V^2P_{i_2}^{-2}+\sigma_W^2Q_{j_1}^{-2})^{-1/2}$, $B_{j_1,j_2} = \sum_{i_1\neq i_2}(\sigma_V^2P_{i_1}^{-2}+\sigma_W^2Q_{j_2}^{-2})^{-1/2}(\sigma_V^2P_{i_2}^{-2}+\sigma_W^2Q_{j_1}^{-2})^{-1/2}$, $C_{i_1,j_1} = \sum_{i_2\neq i_1}\sum_{j_2\neq j_1}(\sigma_V^2P_{i_1}^{-2}+\sigma_W^2Q_{j_2}^{-2})^{-1/2}(\sigma_V^2P_{i_2}^{-2}+\sigma_W^2Q_{j_1}^{-2})^{-1/2}$.

Define

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

It follows from the expression of $T_{WMW}^{(1)}$ in (5.12) that $E(dT_{WMW}^{(1)} | P_i, Q_j, 1 \leq i \leq m, 1 \leq j \leq n) = ||\mu||^2 S_1$. Further, it can be shown that

$$\begin{aligned}
& Var(dT_{WMW}^{(1)} | P_i, Q_j, 1 \leq i \leq m, 1 \leq j \leq n) \\
& = S_2 + \frac{4}{[(m)_2(n)_2]^2} \sum_{i, j_1 \neq j_2} \{\mu' \Sigma_V \mu\} P_i^{-2} C_{i,j_1} C_{i,j_2} \\
& \quad + \frac{1}{[(m)_2(n)_2]^2} \sum_{\substack{j_1, j_2, j_3 \\ all \ distinct}} \left[\{\mu' \Sigma_W \mu\} (P_{j_1}^{-2} + P_{j_2}^{-2}) B_{j_1,j_2} (2B_{j_1,j_2} + B_{j_1,j_3} + B_{j_2,j_3}) \right], \\
& = S_2 + \frac{1}{[(m)_2(n)_2]^2} \{L_1 \mu' \Sigma_V \mu + L_2 \mu' \Sigma_W \mu\}, \qquad (5.13)
\end{aligned}$$

where $L_1 = 4 \sum_{i, j_1 \neq j_2} P_i^{-2} C_{i,j_1} C_{i,j_2}$ and $L_2 = \sum (P_{j_1}^{-2} + P_{j_2}^{-2}) B_{j_1,j_2} (2B_{j_1,j_2} + B_{j_1,j_3} + B_{j_2,j_3})$. Here, the latter summation is taken over distinct indices j_1, j_2 and j_3. Also,

$$\begin{aligned}
S_2 &= \frac{1}{[(m)_2(n)_2]^2} \left\{ 2 \sum_{i_1 \neq i_2} [P_{i_1} P_{i_2}]^{-2} A_{i_1,i_2}^2 \mathrm{tr}(\Sigma_V^2) \right. \\
& \quad \left. + \ 2 \sum_{j_1 \neq j_2} [Q_{j_1} Q_{j_2}]^{-2} B_{j_1,j_2}^2 \mathrm{tr}(\Sigma_W^2) + 4 \sum_{i,j} [P_i Q_j]^{-2} C_{i,j}^2 \mathrm{tr}(\Sigma_V \Sigma_W) \right\}, \\
&= \{L_3 \mathrm{tr}(\Sigma_V^2) + L_4 \mathrm{tr}(\Sigma_W^2) + 2L_5 \mathrm{tr}(\Sigma_V \Sigma_W)\} / [(m)_2(n)_2]^2,
\end{aligned}$$

where $L_3 = 2 \sum_{i_1 \neq i_2} [P_{i_1} P_{i_2}]^{-2} A_{i_1,i_2}^2$, $L_4 = 2 \sum_{j_1 \neq j_2} [Q_{j_1} Q_{j_2}]^{-2} B_{j_1,j_2}^2$, and $L_5 = 2 \sum_{i,j} [P_i Q_j]^{-2} C_{i,j}^2$. Note that $(L_1 \mu' \Sigma_V \mu + L_2 \mu' \Sigma_W \mu) \leq \max\{L_1, L_2\} \mu' (\Sigma_V + \Sigma_W) \mu$. Also, $S_2 \geq [(m)_2(n)_2]^{-2} \min\{L_3, L_4, L_5\} [\mathrm{tr}(\Sigma_V^2) + \mathrm{tr}(\Sigma_W^2) + 2\mathrm{tr}(\Sigma_V \Sigma_W)] = [(m)_2(n)_2]^{-2} \min\{L_3, L_4, L_5\} \mathrm{tr}[(\Sigma_V + \Sigma_W)^2]$. These facts along with (5.13) and Assumption (C3) imply that $Var(dT_{WMW}^{(1)} | P_i, Q_j, 1 \leq i \leq m, 1 \leq j \leq n) = S_2(1 + o(1))$ as $d \to \infty$. Now,

$$\begin{aligned}
& (dT_{WMW}^{(1)} - ||\mu||^2 S_1) / S_2^{1/2} \\
& = \left[\frac{1}{(m)_2(n)_2} \sum_{i_1 \neq i_2} \sum_{j_1 \neq j_2} \frac{(\mathbf{X}_{i_1} - \mathbf{Y}_{j_1} + \mu)' (\mathbf{X}_{i_2} - \mathbf{Y}_{j_2} + \mu)}{d(\sigma_V^2 P_{i_1}^{-2} + \sigma_W^2 Q_{j_1}^{-2})^{1/2} (\sigma_V^2 P_{i_2}^{-2} + \sigma_W^2 Q_{j_2}^{-2})^{1/2}} \right. \\
& \quad \left. - \frac{2}{(m)_2(n)_2} \sum_{i,j} C_{i,j} \mu' (\mathbf{X}_i - \mathbf{Y}_j + \mu) \right] / S_2^{1/2} \\
& = (\widetilde{T}_{WMW}^{(1)} - \widetilde{T}_{WMW}^{(2)}) / S_2^{1/2}, \qquad (5.14)
\end{aligned}$$

where

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

and $\widetilde{T}^{(2)}_{WMW} = 2[(m)_2(n)_2]^{-1}\sum_{i,j} C_{i,j}\mu'(\mathbf{X}_i - \mathbf{Y}_j + \mu)$. It can be shown that $E(\widetilde{T}^{(2)}_{WMW}|P_i, Q_j, 1 \le i \le m, 1 \le j \le n) = 0$ and

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

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

Next note that

$$\widetilde{T}^{(1)}_{WMW}$$
$$= \frac{1}{(m)_2(n)_2}\sum_{i_1\neq i_2}\sum_{j_1\neq j_2}\left\{\frac{(\mathbf{V}_{i_1}/P_{i_1} - \mathbf{W}_{j_1}/Q_{j_1})'(\mathbf{V}_{i_2}/P_{i_2} - \mathbf{W}_{j_2}/Q_{j_2})}{d(\sigma_V^2P_{i_1}^{-2} + \sigma_W^2Q_{j_1}^{-2})^{1/2}(\sigma_V^2P_{i_2}^{-2} + \sigma_W^2Q_{j_2}^{-2})^{1/2}}\right\}$$
$$= \frac{1}{(m)_2(n)_2}\sum_{k=1}^{d}\sum_{i_1\neq i_2}\sum_{j_1\neq j_2}\left\{\frac{(Q_{j_1}V_{i_1k} - P_{i_1}W_{j_1k})(Q_{j_2}V_{i_2k} - P_{i_2}W_{j_2k})}{dP_{i_1}P_{i_2}Q_{j_1}Q_{j_2}(\sigma_V^2P_{i_1}^{-2} + \sigma_W^2Q_{j_1}^{-2})^{1/2}(\sigma_V^2P_{i_2}^{-2} + \sigma_W^2Q_{j_2}^{-2})^{1/2}}\right\}.$$

It is easy to see that $E(\widetilde{T}^{(1)}_{WMW}|P_i, Q_j, 1 \le i \le m, 1 \le j \le n) = 0$. Further, from algebraic computations similar to those used earlier in deriving $Var(dT^{(1)}_{WMW}|P_i, Q_j, 1 \le i \le m, 1 \le j \le n)$, it can be shown that $Var(\widetilde{T}^{(1)}_{WMW}|P_i, Q_j, 1 \le i \le m, 1 \le j \le n) = S_2$. Thus, by Theorem 4.0.1 in Lin and Lu (1996) and Assumption (C4), the conditional distribution of $\widetilde{T}^{(1)}_{WMW}/S_2^{1/2}$ given the P_i's and the Q_j's converges to a standard Gaussian distribution as $d \to \infty$. This fact along with (5.14) and the fact that conditionally on the P_i's and the Q_j's, $\widetilde{T}^{(2)}_{WMW}/S_2^{1/2}$ converges to zero *in probability* as $d \to \infty$ yield

$$\lim_{d\to\infty} P\{(dT^{(1)}_{WMW} - ||\mu||^2S_1)/S_2^{1/2} \le x\} = \Phi(x). \tag{5.15}$$

Next, let us write

$$
\begin{aligned}
&T^{(2)}_{WMW} \\
&= \frac{1}{(m)_2(n)_2} \sum_{i_1 \neq i_2} \sum_{j_1 \neq j_2} \left[\frac{(\mathbf{X}_{i_1} - \mathbf{Y}_{j_1})'(\mathbf{X}_{i_2} - \mathbf{Y}_{j_2}) - ||\mu||^2}{(\sigma_V^2 P_{i_1}^{-2} + \sigma_W^2 Q_{j_1}^{-2})^{1/2} (\sigma_V^2 P_{i_2}^{-2} + \sigma_W^2 Q_{j_2}^{-2})^{1/2}} \times \right. \\
&\qquad \left. \left\{ \frac{d(\sigma_V^2 P_{i_1}^{-2} + \sigma_W^2 Q_{j_1}^{-2})^{1/2} (\sigma_V^2 P_{i_2}^{-2} + \sigma_W^2 Q_{j_2}^{-2})^{1/2}}{||\mathbf{X}_{i_1} - \mathbf{Y}_{j_1}||\ ||\mathbf{X}_{i_2} - \mathbf{Y}_{j_2}||} - 1 \right\} \right] \\
&\quad + \frac{||\mu||^2}{(m)_2(n)_2} \sum_{i_1 \neq i_2} \sum_{j_1 \neq j_2} \left[\frac{1}{(\sigma_V^2 P_{i_1}^{-2} + \sigma_W^2 Q_{j_1}^{-2})^{1/2} (\sigma_V^2 P_{i_2}^{-2} + \sigma_W^2 Q_{j_2}^{-2})^{1/2}} \times \right. \\
&\qquad \left. \left\{ \frac{d(\sigma_V^2 P_{i_1}^{-2} + \sigma_W^2 Q_{j_1}^{-2})^{1/2} (\sigma_V^2 P_{i_2}^{-2} + \sigma_W^2 Q_{j_2}^{-2})^{1/2}}{||\mathbf{X}_{i_1} - \mathbf{Y}_{j_1}||\ ||\mathbf{X}_{i_2} - \mathbf{Y}_{j_2}||} - 1 \right\} \right] \\
&= \widetilde{T}^{(3)}_{WMW} + \widetilde{T}^{(4)}_{WMW},
\end{aligned} \tag{5.16}
$$

where

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

As mentioned earlier, $S_2 \geq [(m)_2(n)_2]^{-2} \min\{L_3, L_4, L_5\} \mathrm{tr}[(\Sigma_V + \Sigma_W)^2]$. Moreover, the stationarity of the sequences $\mathcal{X}$ and $\mathcal{Y}$ and the Cauchy-Schwarz inequality imply that $\mathrm{tr}[(\Sigma_V + \Sigma_W)^2] \geq d(\sigma_V^2 + \sigma_W^2)^2$. These facts along with (5.10) and Assumption (C3) imply that conditionally on the P_i's and the Q_j's, each term inside the double sum appearing in $\widetilde{T}^{(4)}_{WMW}$ above is $o_P(S_2^{1/2})$ as $d \to \infty$. So, $\widetilde{T}^{(4)}_{WMW}/S_2^{1/2}$ 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 $\widetilde{T}^{(3)}_{WMW}$. It follows from (5.10) 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 $d \to \infty$. Also, note that

$$
\begin{aligned}
(\mathbf{X}_{i_1} - \mathbf{Y}_{j_1})'(\mathbf{X}_{i_2} - \mathbf{Y}_{j_2}) - ||\mu||^2 &= (\mathbf{X}_{i_1} - \mathbf{Y}_{j_1} + \mu)'(\mathbf{X}_{i_2} - \mathbf{Y}_{j_2} + \mu) \qquad (5.17) \\
&\quad - \mu'(\mathbf{X}_{i_1} - \mathbf{Y}_{j_1} + \mu) - \mu'(\mathbf{X}_{i_2} - \mathbf{Y}_{j_2} + \mu) \\
&= \sum_{k=1}^{d} \left\{ \frac{(Q_{j_1} V_{i_1 k} - P_{i_1} W_{j_1 k})(Q_{j_2} V_{i_2 k} - P_{i_2} W_{j_2 k})}{P_{i_1} P_{i_2} Q_{j_1} Q_{j_2}} \right\} \qquad (5.18) \\
&\quad - \mu'(Q_{j_1} \mathbf{V}_{i_1} - P_{i_1} \mathbf{W}_{j_1})\mu/(P_{i_1} Q_{j_1}) - \mu'(Q_{j_2} \mathbf{V}_{i_2} - P_{i_2} \mathbf{W}_{j_2})\mu/(P_{i_2} Q_{j_2}).
\end{aligned}
$$

It is easy to show that the conditional expectation of the first term in (5.18) given the P_i's and the Q_j's is zero, and its conditional variance is $v_{i_1 i_2 j_1 j_2} = [P_{i_1} P_{i_2}]^{-2}\mathrm{tr}(\Sigma_V^2) + [Q_{j_1} Q_{j_2}]^{-2}\mathrm{tr}(\Sigma_W^2) + \{[P_{i_1} Q_{j_2}]^{-2} + [P_{i_2} Q_{j_1}]^{-2}\}\mathrm{tr}(\Sigma_V \Sigma_W)$. So, $v_{i_1 i_2 j_1 j_2} = O(\mathrm{tr}[(\Sigma_V + \Sigma_W)^2])$. Hence, using the fact that $S_2 \geq [(m)_2 (n)_2]^{-2} \min\{L_3, L_4, L_5\}\mathrm{tr}[(\Sigma_V + \Sigma_W)^2]$ and Chebyshev's inequality, it follows that the first term in (5.18) 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 $\widetilde{T}^{(2)}_{WMW}$ earlier, we get that the second and the third terms in (5.18) after scaling by $S_2^{1/2}$ converge to zero *in probability* as $d \to \infty$. So, the left hand side of the equation (5.17) after scaling by $S_2^{1/2}$ is bounded *in probability*, conditional on the P_i's and the Q_j's, as $d \to \infty$. Thus, $\widetilde{T}^{(3)}_{WMW}/S_2^{1/2}$ converges to zero *in probability* as $d \to \infty$. This along with (5.16) and the fact that $\widetilde{T}^{(4)}_{WMW}/S_2^{1/2}$ converges to zero *in probability* as $d \to \infty$ together imply that $T^{(2)}_{WMW}/S_2^{1/2}$ converges to zero *in probability* as $d \to \infty$. Combining this fact with (5.15) and (5.11), we get $\lim_{d\to\infty} P\{(dT_{WMW} - ||\mu||^2 S_1)/S_2^{1/2} \leq x | P_i, Q_j, 1 \leq i \leq m, 1 \leq j \leq n\} = \Phi(x)$ for all $x \in \mathbb{R}$ and for each $m, n \geq 1$. Consequently,

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

for all $x \in \mathbb{R}$ and for each $m, n \geq 1$.

We now derive the asymptotic distribution of $T^{(2)}_{CQ}$. As in the proof of Theorem 2.1, $T^{(2)}_{CQ} = T_1 - T_2$. In the setup of the present theorem, $T_1 = [(m)_2 (n)_2]^{-1} \sum_{i_1 \neq i_2} \sum_{j_1 \neq j_2} (\mathbf{V}_{i_1}/P_{i_1} - \mathbf{W}_{j_1}/Q_{j_1})'(\mathbf{V}_{i_2}/P_{i_2} - \mathbf{W}_{j_2}/Q_{j_2})'$ and $T_2 = 2(mn)^{-1} \sum_{i,j} \mu'(\mathbf{V}_i/P_i - \mathbf{W}_j/Q_j)$. So, $E(T_1 | P_i, Q_j, 1 \leq i \leq m, 1 \leq j \leq n) = 0$. Further, from algebraic computations similar to those used to derive the variance of $T^{(2)}_{CQ}$ in the proof of Theorem 2.1, it follows that

$$Var(T_1 | P_i, Q_j, 1 \leq i \leq m, 1 \leq j \leq n) = \frac{1}{[(m)_2 (n)_2]^2} \left\{ 2 \sum_{i_1 \neq i_2} [P_{i_1} P_{i_2}]^{-2} \mathrm{tr}(\Sigma_V^2) \right.$$
$$\left. + 2 \sum_{j_1 \neq j_2} [Q_{j_1} Q_{j_2}]^{-2} \mathrm{tr}(\Sigma_W^2) + 4 \sum_{i,j} [P_i Q_j]^{-2} \mathrm{tr}(\Sigma_V \Sigma_W) \right\}.$$

Define $S_3 = Var(T_1 | P_i, Q_j, 1 \leq i \leq m, 1 \leq j \leq n)$. Also, $E(T_2 | P_i, Q_j, 1 \leq i \leq m, 1 \leq j \leq n) = 0$, and $Var(T_2 | P_i, Q_j, 1 \leq i \leq m, 1 \leq j \leq n) = o(S_3)$ 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 $\widetilde{T}^{(1)}_{WMW}$ 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 $m, n \geq 1$. Combining these facts, we have

$$\lim_{d\to\infty} P\{(T^{(2)}_{CQ} - ||\mu||^2/S_3^{1/2} \leq x\} = \Phi(x)$$

for all $x \in \mathbb{R}$ and all $m, n \geq 1$.

(b) Note that S_1 is a real valued 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 $\psi_2 = 2\mathrm{tr}(\Sigma_V^2)\psi_{21}/(m)_2 + 2\mathrm{tr}(\Sigma_W^2)\psi_{22}/(n)_2 + 4\mathrm{tr}(\Sigma_V\Sigma_W)\psi_{23}/(mn)$. Recall that $S_2 = 2\mathrm{tr}(\Sigma_V^2)S_{21}/(m)_2 + 2\mathrm{tr}(\Sigma_W^2)S_{22}/(n)_2 + 4\mathrm{tr}(\Sigma_V\Sigma_W)S_{23}/(mn)$. 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 $\mathrm{tr}(\Sigma_V^2)$, $\mathrm{tr}(\Sigma_W^2)$ and $\mathrm{tr}(\Sigma_V\Sigma_W)$ equals a constant multiple of d plus a remainder term, which is $o(d)$ as $d \to \infty$. Thus, for each fixed $m, n \geq 1$, there exist constants A_1, A_2 and A_3 such that with *probability one*

$$\lim_{d\to\infty} \frac{\psi_2}{S_2} = \frac{2\psi_{21}A_1/(m)_2 + 2\psi_{22}A_2/(n)_2 + 4\psi_{23}A_3/(mn)}{2S_{21}A_1/(m)_2 + 2S_{22}A_2/(n)_2 + 4S_{23}A_3/(mn)} \tag{5.19}$$

We denote the right hand side of (5.19) by $R_{m,n}$. Further, the assumption in the theorem and arguments preceding (5.19) imply that $||\mu||^2/\psi_2^{1/2}$ converges to a finite non-negative limit b^2 (say) as $d \to \infty$. Now,

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

where the last equality above follows since $R_{m,n}$ converges to one and $S_1 - \psi_1$ converges to zero *almost surely* as $m, n \to \infty$.

Note that $[(m)_2\{(n)_2\}^2]^{-1}\sum_{i_1\neq i_2}[P_{i_1}P_{i_2}]^{-2}$, $[(n)_2\{(m)_2\}^2]^{-1}\sum_{j_1\neq j_2}[Q_{j_1}Q_{j_2}]^{-2}$ 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 $\Sigma_1 = Disp(\mathbf{X}) = \Sigma_V E(P^{-2})$ and $\Sigma_2 = Disp(\mathbf{Y}) = \Sigma_W E(Q^{-2})$. So, arguing as in the case of S_2 above, we get that S_3/Γ_1

converges *in probability* to one as first $d \to \infty$ and then $m, n \to \infty$. Thus, it follows that $\lim_{m,n\to\infty} \lim_{d\to\infty} P\{(T_{CQ}^{(2)} - ||\mu||^2)/\Gamma_1^{1/2} \le x\} = \Phi(x)$ 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 $\psi_2 = [\sigma_V^2 E(P^{-2})]^{-2} E^2\{Q_1Q_2/[(P_1^2+Q_1^2)^{1/2}(P_1^2+Q_2^2)^{1/2}]\}\Gamma_1$. Here, ψ_1 and ψ_2 are as in the proof of Theorem 3.1. Since $\lim_{m,n\to\infty}\lim_{d\to\infty} ||\mu||^2/\Gamma_2^{1/2} = c$ for some $c \in (0,\infty)$, we have $\lim_{m,n\to\infty}\lim_{d\to\infty} \beta_{T_{CQ}^{(2)}}(\mu) = \Phi(-\zeta_\alpha + c)$, and

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

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 $\lim_{m,n\to\infty}\lim_{d\to\infty}\beta_{T_{WMW}}(\mu) > \lim_{m,n\to\infty}\lim_{d\to\infty}\beta_{T_{CQ}^{(2)}}(\mu)$. $\square$

Proof of Theorem 3.3. (a) Let us consider the conditional distribution of T_{SR} given $P_1, P_2, \ldots, P_n$. By definition,

$$T_{SR} = \frac{1}{(n)_4} \sum_{\substack{i_1,i_2,i_3,i_4 \\ all\ distinct}} \frac{(P_{i_2}\mathbf{V}_{i_1} + P_{i_1}\mathbf{V}_{i_2} + 2\mu P_{i_1}P_{i_2})'(P_{i_4}\mathbf{V}_{i_3} + P_{i_3}\mathbf{V}_{i_4} + 2\mu P_{i_3}P_{i_4})}{||P_{i_2}\mathbf{V}_{i_1} + P_{i_1}\mathbf{V}_{i_2} + 2\mu P_{i_1}P_{i_2}||\ ||P_{i_4}\mathbf{V}_{i_3} + P_{i_3}\mathbf{V}_{i_4} + 2\mu P_{i_3}P_{i_4}||}.$$

Consider the event $F = \{d^{-1}||P_2\mathbf{V}_1 + P_1\mathbf{V}_2 + 2\mu P_1P_2||^2 - \sigma_V^2(P_1^2+P_2^2) = o(d^{-1/2+\epsilon})$ as $d \to \infty\}$. From Theorem 8.2.2 in Lin and Lu (1996) and the assumptions in the theorem, it follows that for any given $\epsilon \in (0, 1/2)$, $Pr(F|P_1, P_2) = 1$ for *almost every* P_1 and P_2. Let us rewrite T_{SR} as

$$\begin{aligned}
&T_{SR} \\
&= \frac{1}{(n)_4}\sum_{i_1\neq i_2\neq i_3\neq i_4} \frac{(P_{i_2}\mathbf{V}_{i_1} + P_{i_1}\mathbf{V}_{i_2} + 2\mu P_{i_1}P_{i_2})'(P_{i_4}\mathbf{V}_{i_3} + P_{i_3}\mathbf{V}_{i_4} + 2\mu P_{i_3}P_{i_4})}{d\sigma_V^2(P_{i_1}^2+P_{i_2}^2)^{1/2}(P_{i_3}^2+P_{i_4}^2)^{1/2}} \\
&+ \frac{1}{(n)_4}\sum_{i_1\neq i_2\neq i_3\neq i_4}\left[\frac{(P_{i_2}\mathbf{V}_{i_1} + P_{i_1}\mathbf{V}_{i_2} + 2\mu P_{i_1}P_{i_2})'(P_{i_4}\mathbf{V}_{i_3} + P_{i_3}\mathbf{V}_{i_4} + 2\mu P_{i_3}P_{i_4})}{d\sigma_V^2(P_{i_1}^2+P_{i_2}^2)^{1/2}(P_{i_3}^2+P_{i_4}^2)^{1/2}}\right. \\
&\times \left.\left\{\frac{d\sigma_V^2(P_{i_1}^2+P_{i_2}^2)^{1/2}(P_{i_3}^2+P_{i_4}^2)^{1/2}}{||P_{i_2}\mathbf{V}_{i_1} + P_{i_1}\mathbf{V}_{i_2} + 2\mu P_{i_1}P_{i_2}||\ ||P_{i_4}\mathbf{V}_{i_3} + P_{i_3}\mathbf{V}_{i_4} + 2\mu P_{i_3}P_{i_4}||} - 1\right\}\right] \\
&\qquad = (T_{SR}^{(1)} + T_{SR}^{(2)})/d,
\end{aligned} \tag{5.20}$$

where

$$T_{SR}^{(1)} = \frac{1}{(n)_4}\sum_{i_1\neq i_2\neq i_3\neq i_4} \frac{(P_{i_2}\mathbf{V}_{i_1} + P_{i_1}\mathbf{V}_{i_2} + 2\mu P_{i_1}P_{i_2})'(P_{i_4}\mathbf{V}_{i_3} + P_{i_3}\mathbf{V}_{i_4} + 2\mu P_{i_3}P_{i_4})}{\sigma_V^2(P_{i_1}^2+P_{i_2}^2)^{1/2}(P_{i_3}^2+P_{i_4}^2)^{1/2}}$$

and

$$T_{SR}^{(2)} = \frac{1}{(n)_4} \sum_{i_1 \neq i_2 \neq i_3 \neq i_4} \left[\frac{(P_{i_2}\mathbf{V}_{i_1} + P_{i_1}\mathbf{V}_{i_2} + 2\mu P_{i_1}P_{i_2})'(P_{i_4}\mathbf{V}_{i_3} + P_{i_3}\mathbf{V}_{i_4} + 2\mu P_{i_3}P_{i_4})}{\sigma_V^2 (P_{i_1}^2 + P_{i_2}^2)^{1/2}(P_{i_3}^2 + P_{i_4}^2)^{1/2}} \right.$$
$$\left. \times \left\{ \frac{d\sigma_V^2 (P_{i_1}^2 + P_{i_2}^2)^{1/2}(P_{i_3}^2 + P_{i_4}^2)^{1/2}}{||P_{i_2}\mathbf{V}_{i_1} + P_{i_1}\mathbf{V}_{i_2} + 2\mu P_{i_1}P_{i_2}||\ ||P_{i_4}\mathbf{V}_{i_3} + P_{i_3}\mathbf{V}_{i_4} + 2\mu P_{i_3}P_{i_4}||} - 1 \right\} \right].$$

Recall that a similar decomposition of T_{SR} was obtained in the proof of Theorem 2.3. The proof of the asymptotic Gaussianity of T_{SR} follows from the ideas used to prove the asymptotic Gaussianity of T_{WMW} in Theorem 3.1, and the details are provided in Appendix – II. Z_2 and Z_3 appearing in the asymptotic Gaussian distribution of T_{SR} are given by $Z_2 = 2[(n)_4\sigma_V^2]^{-1} \sum_{i_1 \neq i_2} \widetilde{U}_{i_1,i_2} P_{i_1} P_{i_2}$ and $Z_3 = 8\mathrm{tr}(\Sigma_V^2)[(n)_4\sigma_V^2]^{-2} \sum_{i_1 \neq i_2} \widetilde{U}_{i_1,i_2}^2$, where $\widetilde{U}_{i_1,i_2} = \sum_{i_3 \neq i_4 \neq i_1 \neq i_2} P_{i_3}P_{i_4}/[(P_{i_1}^2 + P_{i_3}^2)^{1/2}(P_{i_2}^2 + P_{i_4}^2)^{1/2}]$.

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 $Z_1 = [(n)_2\sigma_V^2]^{-1} \sum_{i_1 \neq i_2} P_{i_1}P_{i_2}$ and $\Gamma_3 = 2\mathrm{tr}(\Sigma_V^2)/[(n)_2\sigma_V^4]$.

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

(b) Observe that Z_1, Z_2, $(n)_4 Z_3$ and $(n)_2 Z_4$ are real-valued V-statistics, whose kernels have finite expectations by the assumption in part (b) of the theorem. So, 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 $\Sigma = Disp(\mathbf{X}) = \Sigma_V E(P^{-2})$ and $\sigma^2 = Var(X_1) = \sigma_V^2 E(P^{-2})$. 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,

$$\begin{aligned}
&\lim_{n\to\infty} \lim_{d\to\infty} \beta_{T_S}(\mu) = \Phi(-\zeta_\alpha + cE^2(P_1)E(P_1^{-2})),\\
&\lim_{n\to\infty} \lim_{d\to\infty} \beta_{T_{SR}}(\mu) = \Phi(-\zeta_\alpha + \frac{cE^2\{P_1P_2/(P_1^2 + P_2^2)^{1/2}\}E(P_1^{-2})}{E\{P_2P_3/[(P_1^2 + P_2^2)^{1/2}(P_1^2 + P_3^2)^{1/2}]\}}),\\
&\lim_{n\to\infty} \lim_{d\to\infty} \beta_{T_{CQ}^{(1)}}(\mu) = \Phi(-\zeta_\alpha + c).
\end{aligned}$$

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, $\lim_{n\to\infty} \lim_{d\to\infty} \beta_{T_S}(\mu) > \lim_{n\to\infty} \lim_{d\to\infty} \beta_{T_{CQ}^{(1)}}(\mu)$. The proof of the other part of the theorem is similar to the proof of Theorem 3.2. □

Appendix – II

Additional mathematical details related to the proof of part (a) of Theorem 3.3

Here, we provide more details related to the derivations of the asymptotic distributions of T_{SR}, T_S and $T_{CQ}^{(1)}$ under the assumptions of Theorem 3.3. Recall that $T_{SR} = (T_{SR}^{(1)} + T_{SR}^{(2)})/d$, where $T_{SR}^{(1)}$ and $T_{SR}^{(2)}$ are defined in the proof of Theorem 3.3 in Appendix – I. Define $\widetilde{U}_{i_1,i_2} = \sum_{i_3 \neq i_4 \neq i_1 \neq i_2} P_{i_3} P_{i_4} / [(P_{i_1}^2 + P_{i_3}^2)^{1/2} (P_{i_2}^2 + P_{i_4}^2)^{1/2}]$. Then, by the definition of $T_{SR}^{(1)}$, we have

$$T_{SR}^{(1)} \quad = \quad \frac{4}{(n)_4 \sigma_V^2} \sum_{i_1 \neq i_2} \left\{ \widetilde{U}_{i_1,i_2} \mathbf{W}_{i_1}' \mathbf{W}_{i_2} + 2\widetilde{U}_{i_1,i_2} P_{i_2} \mu' \mathbf{W}_{i_1} + ||\mu||^2 \widetilde{U}_{i_1,i_2} P_{i_1} P_{i_2} \right\}.$$

It follows easily that $E(T_{SR}^{(1)} | P_i, 1 \leq i \leq n) = 4||\mu||^2 [(n)_4 \sigma_V^2]^{-1} \sum_{i_1 \neq i_2} \widetilde{U}_{i_1,i_2} P_{i_1} P_{i_2}$. Set $Z_2 = 2[(n)_4 \sigma_V^2]^{-1} \sum_{i_1 \neq i_2} \widetilde{U}_{i_1,i_2} P_{i_1} P_{i_2}$. Further, it can be shown using the assumptions in the theorem that $Var(T_{SR}^{(1)} | P_i, 1 \leq i \leq n) = 32\text{tr}(\Sigma_V^2)[(n)_4 \sigma_V^2]^{-2} \sum_{i_1 \neq i_2} \widetilde{U}_{i_1,i_2}^2 (1 + o(1))$ as $d \to \infty$. Let $Z_3 = 8\text{tr}(\Sigma_V^2)[(n)_4 \sigma_V^2]^{-2} \sum_{i_1 \neq i_2} \widetilde{U}_{i_1,i_2}^2$. So, using arguments similar to those used to prove the asymptotic Gaussianity of $T_{WMW}^{(1)}$ in the proof of Theorem 3.1, it follows that $(T_{SR}^{(1)} - 2||\mu||^2 Z_2)/(2Z_3^{1/2})$ converges *weakly* to a standard Gaussian distribution as $d \to \infty$ for each $n \geq 1$. Moreover, using arguments similar to those used to prove the convergence *in probability* of $T_{WMW}^{(2)}$ in the proof of Theorem 3.1, it follows that $T_{SR}^{(2)}/Z_3^{1/2}$ converges to zero *in probability* as $d \to \infty$ for each $n \geq 1$. This fact along with the equation $T_{SR} = (T_{SR}^{(1)} + T_{SR}^{(2)})/d$ and the asymptotic Gaussianity of $T_{SR}^{(1)}$ yields

$$\lim_{d \to \infty} P\{(dT_{SR} - 2||\mu||^2 Z_2)/(2Z_3^{1/2}) \leq x\} = \Phi(x)$$

for all $x \in \mathbb{R}$ and each $n \geq 1$. Here, Φ is the standard Gaussian cumulative distribution function.

Using very similar arguments as above, we get that

$$\lim_{d \to \infty} P\{(dT_S - ||\mu||^2 Z_1)/\Gamma_3^{1/2} \leq x\} = \Phi(x)$$

for all $x \in \mathbb{R}$ and each $n \geq 1$, where $Z_1 = [(n)_2 \sigma_V^2]^{-1} \sum_{i_1 \neq i_2} P_{i_1} P_{i_2}$, and $\Gamma_3 = 2\text{tr}(\Sigma_V^2)/[(n)_2 \sigma_V^4]$.

Next, consider the conditional distribution of $T_{CQ}^{(1)}$ given the P_i's, and note that

$$\begin{aligned} T_{CQ}^{(1)} &= \frac{1}{(n)_2} \sum_{i_1 \neq i_2} \frac{(\mathbf{V}_{i_1} + \mu P_{i_1})'(\mathbf{V}_{i_2} + \mu P_{i_2})}{P_{i_1} P_{i_2}} \\ &= \frac{1}{(n)_2} \sum_{i_1 \neq i_2} \frac{\mathbf{V}_{i_1}' \mathbf{V}_{i_2}}{P_{i_1} P_{i_2}} + \frac{2}{n} \sum_i \frac{\mu' \mathbf{V}_i}{P_i} + ||\mu||^2. \end{aligned}$$

So, $E(T_{CQ}^{(1)} | P_i, 1 \leq i \leq n) = ||\mu||^2$, and $Var(T_{CQ}^{(1)} | P_i, 1 \leq i \leq n) = Z_4(1 + o(1))$ as $d \to \infty$, where $Z_4 = 2\text{tr}(\Sigma_V^2)[(n)_2]^{-2} \sum_{i_1 \neq i_2} [P_{i_1} P_{i_2}]^{-2}$. Using the assumptions in the theorem, it follows that conditional on the P_i's, $\sum_i \mu' \mathbf{V}_i / P_i = o_P(Z_4^{1/2})$ as $d \to \infty$. Thus, we get

$$\lim_{d \to \infty} P\{(T_{CQ}^{(1)} - ||\mu||^2)/Z_4^{1/2} \leq x\} = \Phi(x)$$

for all $x \in \mathbb{R}$ and each $n \geq 1$.

Detailed results of the simulation study done in Section 4

Here, we present the results on the sizes and the powers of the tests based on T_{SKK} (Srivastava et al., 2013) and T_{GCBL} (Gregory et al., 2014) discussed in Section 4. We also present the sizes and the powers of the test in Cai et al. (2014) for which the test statistic is denoted by T_{CLX}. Table 2 reports the sizes of these tests implemented using the asymptotic approximations given in their original papers under the models considered in subsections 2.1 and 3.1 of our paper. We also report the sizes of the tests implemented using the permutation distributions of these test statistics.

Table 2: Sizes of the tests based on T_{SKK}, T_{GCBL} and T_{CLX} under some simulated models

Test	d	$AR(1)$ with Gaussian innovation	$AR(1)$ with $t(5)$ innovation	spherical $t(5)$ distribution
T_{SKK}–original	100	0.06	0.064	0.011
	200	0.068	0.06	0.001
	400	0.071	0.072	0
	800	0.089	0.089	0
	1600	0.101	0.089	0
T_{SKK}–permutation	100	0.045	0.039	0.043
	200	0.047	0.048	0.044
	400	0.054	0.043	0.039
	800	0.048	0.052	0.049
	1600	0.042	0.054	0.051
T_{GCBL}-original	100	0.077	0.071	0.137
	200	0.075	0.078	0.148
	400	0.086	0.081	0.141
	800	0.125	0.134	0.152
	1600	0.164	0.152	0.185
T_{GCBL}–permutation	100	0.042	0.048	0.046
	200	0.051	0.042	0.044
	400	0.05	0.056	0.038
	800	0.046	0.047	0.042
	1600	0.055	0.047	0.039
T_{CLX}–original	100	0.082	0.075	0.076
	200	0.101	0.114	0.093
	400	0.136	0.147	0.105
	800	0.167	0.184	0.131

Note that we could not implement the test based on T_{CLX} using its permutation distribution because the test procedure uses a computationally intensive optimization. For the same reason, we could not implement this test for $d = 1600$ under any of the above models using the asymptotic distribution given in Cai et al. (2014). Recall that we have discussed the sizes of the tests based on T_{WMW} and $T_{CQ}^{(2)}$ for the above models in detail in subsections 2.1 and 3.1.

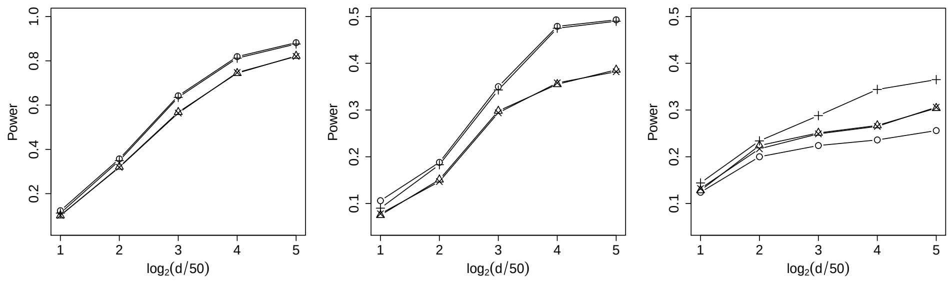

Figure 3: Powers of the tests at nominal 5% level based on T_{WMW} (- $+$ - curves), $T_{CQ}^{(2)}$ (- $\circ$ - curves), T_{SKK} (- $\times$ - curves) and T_{GCBL} (- $\triangle$ - curves) for the $AR(1)$ model with Gaussian innovation (left panel), the $AR(1)$ model with $t(5)$ innovation (middle panel) and the spherical $t(5)$ distribution (right panel).

In Figure 3, we give the plots of the empirical powers of the tests based on T_{SKK} and T_{GCBL}, when they are implemented using their permutation distributions. Each plot in Figure 3 also includes the empirical powers of the tests based on T_{WMW} and $T_{CQ}^{(2)}$. The power curves for these two tests are so close that they are overlaid on each other in the left and the middle plots. Similarly, the power curves corresponding to the tests based on T_{SKK} and T_{GCBL} are overlaid on each other in all the plots.

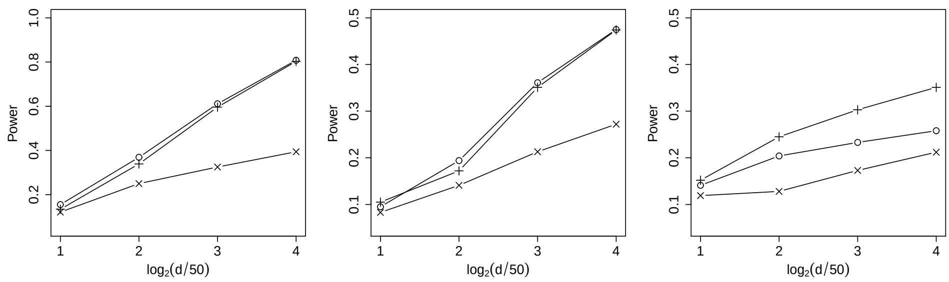

Figure 4: Powers of the tests at nominal 5% level based on T_{WMW} (- $+$ - curves), $T_{CQ}^{(2)}$ (- $\circ$ - curves) and T_{CLX} (- $\times$ - curves) for the $AR(1)$ model with Gaussian innovation (left panel), the $AR(1)$ model with $t(5)$ innovation (middle panel) and the spherical $t(5)$ distribution (right panel).

Figure 4 gives the plots of the empirical powers of the tests based on T_{WMW}, $T_{CQ}^{(2)}$ and T_{CLX}, when the mean shifts in the models considered in subsections 2.1 and 3.1 are distributed equally among all the coordinates. Once again, the power curves corresponding to the tests based on T_{WMW} and $T_{CQ}^{(2)}$ are sufficiently close making the curves overlaid on each other in the left and the middle plots.

In Figure 5, we give the sizes and the powers of the tests based on T_{WMW} and $T_{CQ}^{(2)}$ for the multivariate Gaussian distribution with dispersion matrix $(1-\beta)I_d+\beta\mathbf{1}_d\mathbf{1}_d'$ with $\beta=0.7$ considered in Section 4.

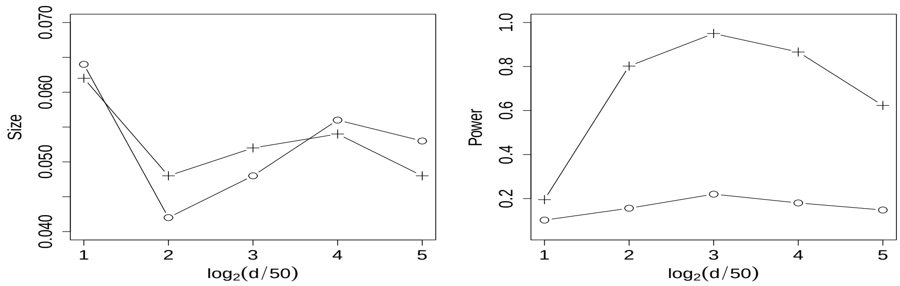

Figure 5: Powers of the tests at nominal 5% level based on T_{WMW} (- + - curves) and $T_{CQ}^{(2)}$ (- ∘ - curves) for the multivariate Gaussian distribution with dispersion matrix $(1-\beta)I_d+\beta\mathbf{1}_d\mathbf{1}_d'$ with $\beta=0.7$.

References

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

Bradley, R. C. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probability Surveys*, 2:107–144. Update of, and a supplement to, the 1986 original.

Cai, T. T., Liu, W., and Xia, Y. (2014). Two-sample test of high dimensional means under dependence. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, 76(2):349–372.

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

Choi, K. and Marden, J. (1997). An approach to multivariate rank tests in multivariate analysis of variance. *Journal of the American Statistical Association*, 92(440):1581–1590.

Fan, J. and Lin, S.-K. (1998). Test of significance when data are curves. *Journal of the American Statistical Association*, 93(443):1007–1021.

Feng, L., Zou, C., Wang, Z., and Zhu, L. (2015). Two sample Behrens-Fisher problem for high-dimensional data. *Statistica Sinica*. To appear.

Gregory, K. B., Carroll, R. J., Baladandayuthapani, V., and Lahiri, S. N. (2014). A two-sample test for equality of means in high dimension. *Journal of the American Statistical Association*.

Hettmansperger, T. P. and McKean, J. W. (2011). *Robust nonparametric statistical methods*, volume 119 of *Monographs on Statistics and Applied Probability*. CRC Press, Boca Raton, FL, second edition.

Ibragimov, I. A. and Linnik, Y. V. (1971). *Independent and stationary sequences of random variables*. Wolters-Noordhoff Publishing, Groningen. With a supplementary chapter by I. A. Ibragimov and V. V. Petrov, Translation from the Russian edited by J. F. C. Kingman.

Kallenberg, O. (2005). *Probabilistic symmetries and invariance principles*. Probability and its Applications (New York). Springer, New York.

Katayama, S. and Kano, Y. (2014). A new test on high-dimensional mean vector without any assumption on population covariance matrix. *Communications in Statistics - Theory and Methods*, 43(24):5290–5304.

Kolmogorov, A. N. and Rozanov, J. A. (1960). On a strong mixing condition for stationary Gaussian processes. *Akademija Nauk SSSR. Teorija Verojatnosteĭ i ee Primenenija*, 5:222–227.

Lin, Z. and Lu, C. (1996). *Limit theory for mixing dependent random variables*, volume 378 of *Mathematics and its Applications*. Kluwer Academic Publishers, Dordrecht; Science Press Beijing, New York.

Marden, J. I. (1999). Multivariate rank tests. In *Multivariate analysis, design of experiments, and survey sampling*, volume 159 of *Statist. Textbooks Monogr.*, pages 401–432. Dekker, New York.

Möttönen, J. and Oja, H. (1995). Multivariate spatial sign and rank methods. *Journal of Nonparametric Statistics*, 5(2):201–213.

Möttönen, J., Oja, H., and Tienari, J. (1997). On the efficiency of multivariate spatial sign and rank tests. *The Annals of Statistics*, 25(2):542–552.

Oja, H. (2010). *Multivariate nonparametric methods with R*, volume 199 of *Lecture Notes in Statistics*. Springer, New York. An approach based on spatial signs and ranks.

Puri, M. L. and Sen, P. K. (1971). *Nonparametric methods in multivariate analysis*. John Wiley & Sons, Inc., New York-London-Sydney.

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

Wang, L., Peng, B., and Li, R. (2015). A high-dimensional nonparametric multivariate test for mean vector. *Journal of the American Statistical Association*.

Wei, S., Lee, C., Wichers, L., and Marron, J. (2015). Direction-projection-permutation for high dimensional hypothesis tests. *Journal of Computational and Graphical Statistics*. To appear.