Statistica Sinica **23** (2013), 667-690
doi:http://dx.doi.org/10.5705/ss.2011.261

JACKKNIFE EMPIRICAL LIKELIHOOD TEST FOR EQUALITY OF TWO HIGH DIMENSIONAL MEANS

Ruodu Wang, Liang Peng and Yongcheng Qi

University of Waterloo, Georgia Institute of Technology and University of Minnesota–Duluth

Abstract: There is a long history of testing the equality of two multivariate means. A popular test is the Hotelling T^2, but in large dimensions it performs poorly due to the possible inconsistency of sample covariance estimation. Bai and Saranadasa (1996) and Chen and Qin (2010) proposed tests not involving the sample covariance, and derived asymptotic limits, which depend on whether the dimension is fixed or diverges, under a specific multivariate model. In this paper, we propose a jackknife empirical likelihood test that has a chi-square limit independent of the dimension. The conditions are much weaker than those needed in existing methods. A simulation study shows that the proposed new test has a very robust size across dimensions and has good power.

Key words and phrases: High dimensional mean, hypothesis test, Jackknife empirical likelihood.

1. Introduction

Let $X_1 = (X_{1,1}, \ldots, X_{1,d})^T, \ldots, X_{n_1} = (X_{n_1,1}, \ldots, X_{n_1,d})^T$ and $Y_1 = (Y_{1,1}, \ldots, Y_{1,d})^T, \ldots, Y_{n_2} = (Y_{n_2,1}, \ldots, Y_{n_2,d})^T$ be two independent random samples with means μ_1 and μ_2, respectively. The testing of $H_0 : \mu_1 = \mu_2$ against $H_a : \mu_1 \neq \mu_2$ for a fixed dimension d has a long history. When both X_1 and Y_1 are multivariate normal and with equal covariance, the well-known test is the Hotelling T^2 test based on

$$T^2 = \eta(\bar{\mathbf{X}} - \bar{\mathbf{Y}})^T A_n^{-1}(\bar{\mathbf{X}} - \bar{\mathbf{Y}}), \tag{1.1}$$

where $\eta = (n_1 + n_2 - 2)n_1 n_2/(n_1 + n_2)$, $\bar{\mathbf{X}} = (1/n_1)\sum_{i=1}^{n_1} X_i$, $\bar{\mathbf{Y}} = (1/n_2)\sum_{i=1}^{n_2} Y_i$ and $A_n = \sum_{i=1}^{n_1}(X_i - \bar{\mathbf{X}})(X_i - \bar{\mathbf{X}})^T + \sum_{i=1}^{n_2}(Y_i - \bar{\mathbf{Y}})(Y_i - \bar{\mathbf{Y}})^T$. However when $d = d(n_1, n_2) \to \infty$, the test performs poorly due to possible inconsistency of sample covariance estimation. When $d/(n_1 + n_2) \to c \in (0, 1)$, Bai and Saranadasa (1996) derived the asymptotic power of T^2. To overcome the restriction $c < 1$, they proposed

$$M_n = (\bar{\mathbf{X}} - \bar{\mathbf{Y}})^T(\bar{\mathbf{X}} - \bar{\mathbf{Y}}) - \eta^{-1}\mathrm{tr}(A_n)$$

as an alternative. Under a special multivariate model that did not assume multivariate normality but kept the condition of equal covariance, they derived the asymptotic limits as $d/n \to c > 0$. Recently Chen and Qin (2010) proposed using

$$CQ = \frac{\sum_{i \neq j}^{n_1} X_i^T X_j}{n_1(n_1-1)} + \frac{\sum_{i \neq j}^{n_2} Y_i^T Y_j}{n_2(n_2-1)} - 2\frac{\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} X_i^T Y_j}{n_1 n_2} \tag{1.2}$$

and allowed d to be of a possibly larger order than that in Bai and Saranadasa (1996). The asymptotic limit of CQ depends on whether the dimension is fixed or diverges, and results in either a normal or a chi-square limit, with special models for $\{X_i\}$ and $\{Y_i\}$ being employed. Another modification of the T^2 test is proposed by Srivastava and Du (2008) and Srivastava (2009) with the covariance matrix replaced by a diagonal matrix. Rates of convergence for high-dimensional means were studied by Kuelbs and Vidyashankar (2010). For nonasymptotic studies of high dimensional means, we refer to Arlot, Blanchard, and Roquain (2010a,b). Here, we are interested in finding a test that does not distinguish between fixed and divergent dimensions.

By noting that $\mu_1 = \mu_2$ is equivalent to $(\mu_1 - \mu_2)^T(\mu_1 - \mu_2) = 0$, one may think of applying an empirical likelihood test to the estimating equation $\mathrm{E}\{(X_{i_1} - Y_{j_1})^T(X_{i_2} - Y_{j_2})\} = 0$ for $i_1 \neq i_2$ and $j_1 \neq j_2$. If one applies the empirical likelihood method of Qin and Lawless (1994) to the samples $X_1, \ldots, X_{n_1}$ and $Y_1, \ldots, Y_{n_2}$, the empirical likelihood function is

$$\begin{aligned}\sup\Bigg\{\Big\{\prod_{i=1}^{n_1}(n_1 p_i)\Big\}\Big\{\prod_{j=1}^{n_2}(n_2 q_j)\Big\} : p_1 \geq 0, \ldots, p_{n_1} \geq 0, q_1 \geq 0, \ldots, q_{n_2} \geq 0,\\ \sum_{i=1}^{n_1} p_i = 1, \sum_{j=1}^{n_2} q_j = 1,\\ \sum_{i_1=1}^{n_1}\sum_{i_2 \neq i_1}\sum_{j_1=1}^{n_2}\sum_{j_2 \neq j_1}(p_{i_1}X_{i_1} - q_{j_1}Y_{j_1})^T(p_{i_2}X_{i_2} - q_{j_2}Y_{j_2}) = 0\Bigg\},\end{aligned}$$

but the estimating equation defines a nonlinear functional and, in general, one has to linearize it before applying the method. For more details on empirical likelihood methods, we refer to Owen (2001) and the review paper of Chen and Van Keilegom (2009). Recently, Jing, Yuan, and Zhou (2009) proposed a jackknife empirical likelihood method to construct confidence regions for nonlinear functionals with a particular focus on U-statistics. Using this idea, one needs to construct a jackknife sample based on the estimator $n_1^{-1}(n_1-1)^{-1}n_2^{-1}(n_2 - 1)^{-1}\sum_{i_1 \neq i_2}^{n_1}\sum_{j_1 \neq j_2}^{n_2}(X_{i_1} - Y_{j_1})^T(X_{i_2} - Y_{j_2})$, the statistic CQ given as (1.2). However, in order to have the jackknife empirical likelihood method work, one has to show that $\sqrt{n_1 n_2}\, n_1^{-1}(n_1-1)^{-1}n_2^{-1}(n_2-1)^{-1}\sum_{i_1 \neq i_2}^{n_1}\sum_{j_1 \neq j_2}^{n_2}(X_{i_1} - Y_{j_1})^T(X_{i_2} - Y_{j_2})$

has a normal limit when $\mu_1 = \mu_2$. If $n_1 = n_2 = n$, $d = 1$, and $\mu_1 = \mu_2$, it is easy to see that

$$\begin{aligned}
&n^{-1}(n-1)^{-2}\sum_{i_1\neq i_2}^{n}\sum_{j_1\neq j_2}^{n}(X_{i_1}-Y_{j_1})^T(X_{i_2}-Y_{j_2}) \\
&\quad = \frac{1}{n-1}\{\sum_{i=1}^{n}(X_i-Y_i)\}^2 - \frac{1}{n-1}\sum_{i=1}^{n}(X_i-Y_i)^2 + \frac{2}{n(n-1)}\sum_{i=1}^{n}X_i\sum_{j=1}^{n}Y_j \\
&\qquad -\frac{2}{(n-1)}\sum_{i=1}^{n}X_iY_i \\
&\quad \xrightarrow{d} (\chi_1^2-1)\mathrm{E}\,(X_1-Y_1)^2
\end{aligned}$$

as $n \to \infty$, where χ_1^2 denotes a random variable having a chi-square distribution with 1 degree of freedom. Obviously, the above limit is not normal. Hence a direct application of the jackknife empirical likelihood method to the statistic CQ does not lead to a chi-square limit.

In this paper, we formulate a jackknife empirical likelihood test for testing $H_0 : \mu_1 = \mu_2$ against $H_a : \mu_1 \neq \mu_2$ by dividing the samples into two parts. The proposed new test has no need to distinguish whether the dimension is fixed or goes to infinity. It turns out that the asymptotic limit of the new test under H_0 is a chi-square limit independent of the dimension, the conditions on p and the random vectors $\{X_i\}$ and $\{Y_j\}$ are weaker as well. A simulation study shows that the size of the new test is quite stable with respect to the dimension and that the proposed test has good power.

We organize this paper as follows. In Section 2, the new methodology and main results are given. Section 3 presents a simulation study and a data analysis. All proofs are in Section 4.

2. Methodology

Throughout, take $X_i = (X_{i,1}, \ldots, X_{i,d})^T$ and $Y_j = (Y_{j,1}, \ldots, Y_{j,d})^T$ for $i = 1, \ldots, n_1$, $j = 1, \ldots, n_2$ to be independent random samples with means μ_1 and μ_2, respectively. With $\min\{n_1, n_2\}$ going to infinity, we wish to test $H_0 : \mu_1 = \mu_2$ against $H_a : \mu_1 \neq \mu_2$. Since $\mu_1 = \mu_2$ is equivalent to $(\mu_1 - \mu_2)^T(\mu_1 - \mu_2) = 0$ and $\mathrm{E}\,(X_{i_1} - Y_{j_1})^T(X_{i_2} - Y_{j_2}) = (\mu_1 - \mu_2)^T(\mu_1 - \mu_2)$ for $i_1 \neq i_2$ and $j_1 \neq j_2$, we propose to apply the jackknife empirical likelihood method to this estimating equation, but a direct application of it fails to have a chi-square limit. Here we propose to split the samples into two groups, as follows.

Take $m_1 = [n_1/2]$, $m_2 = [n_2/2]$, $m = m_1 + m_2$, $\bar{X}_i = X_{i+m_1}$ for $i = 1, \ldots, m_1$, and $\bar{Y}_j = Y_{j+m_2}$ for $j = 1, \ldots, m_2$. First estimate $(\mu_1 - \mu_2)^T(\mu_1 - \mu_2)$ by

$$\frac{1}{m_1 m_2} \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} (X_i - Y_j)^T (\bar{X}_i - \bar{Y}_j). \tag{2.1}$$

This is less efficient than the statistic CQ, but it allows us to add more estimating equations and to employ the empirical likelihood method without estimating the asymptotic covariance. By noting that $\mathrm{E}\{(X_i - Y_j)^T(\bar{X}_i - \bar{Y}_j)\} = (\mu_1 - \mu_2)^T(\mu_1 - \mu_2) = ||\mu_1 - \mu_2||^2$ instead of $O(||\mu_1 - \mu_2||)$, one expects that a test based on (2.1) is as powerful for small values of $||\mu_1 - \mu_2||$; this is confirmed by a brief simulation study. In order to improve power, we propose to apply the jackknife empirical likelihood method in Jing, Yuan, and Zhou (2009) to (2.1) and a linear functional such as

$$\frac{1}{m_1 m_2} \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} \{\alpha^T (X_i - Y_j) + \alpha^T (\bar{X}_i - \bar{Y}_j)\} \tag{2.2}$$

rather than only (2.1), where $\alpha \in \mathbb{R}^d$ is a vector chosen based on prior information. Theoretically, when no additional information is available, any linear functional is a possible choice and $\alpha = (1, \ldots, 1)^T \in \mathbb{R}^d$ is a convenient one. Note that more linear functionals can be added at (2.2), with different choices of α, to further improve the power. In the simulation study we applied the jackknife empirical likelihood to (2.1) and (2.2) with $\alpha = (1, \ldots, 1) \in \mathbb{R}^d$, which resulted in a test with good power and quite robust size with respect to the dimension.

As in Jing, Yuan, and Zhou (2009), based on (2.1) and (2.2), we formulate the jackknife sample as $Z_k = (Z_{k,1}, Z_{k,2})^T$ for $k = 1, \ldots, m$, where

$$\begin{cases} Z_{k,1} = \dfrac{m_1 + m_2}{m_1 m_2} \displaystyle\sum_{i=1}^{m_1} \sum_{j=1}^{m_2} (X_i - Y_j)^T (\bar{X}_i - \bar{Y}_j) \\ \qquad - \dfrac{m_1 + m_2 - 1}{(m_1 - 1) m_2} \displaystyle\sum_{i \neq k, i=1}^{m_1} \sum_{j=1}^{m_2} (X_i - Y_j)^T (\bar{X}_i - \bar{Y}_j), \\ Z_{k,2} = \dfrac{m_1 + m_2}{m_1 m_2} \displaystyle\sum_{i=1}^{m_1} \sum_{j=1}^{m_2} \{\alpha^T (X_i - Y_j) + \alpha^T (\bar{X}_i - \bar{Y}_j)\} \\ \qquad - \dfrac{m_1 + m_2 - 1}{(m_1 - 1) m_2} \displaystyle\sum_{i \neq k, i=1}^{m_1} \sum_{j=1}^{m_2} \{\alpha^T (X_i - Y_j) + \alpha^T (\bar{X}_i - \bar{Y}_j)\}, \end{cases}$$

for $k = 1, \ldots, m_1$, and

$$\begin{cases} Z_{k,1} = \dfrac{m_1+m_2}{m_1 m_2}\displaystyle\sum_{i=1}^{m_1}\sum_{j=1}^{m_2}(X_i - Y_j)^T(\bar{X}_i - \bar{Y}_j) \\ \qquad -\dfrac{m_1+m_2-1}{m_1(m_2-1)}\displaystyle\sum_{i=1}^{m_1}\sum_{j\neq k-m_1, j=1}^{m_2}(X_i - Y_j)^T(\bar{X}_i - \bar{Y}_j), \\ Z_{k,2} = \dfrac{m_1+m_2}{m_1 m_2}\displaystyle\sum_{i=1}^{m_1}\sum_{j=1}^{m_2}\{\alpha^T(X_i - Y_j) + \alpha^T(\bar{X}_i - \bar{Y}_j)\} \\ \qquad -\dfrac{m_1+m_2-1}{m_1(m_2-1)}\displaystyle\sum_{i=1}^{m_1}\sum_{j\neq k-m_1, j=1}^{m_2}\{\alpha^T(X_i - Y_j) + \alpha^T(\bar{X}_i - \bar{Y}_j)\}, \end{cases}$$

for $k = m_1 + 1, \ldots, m$. Based on this sample, the jackknife empirical likelihood ratio function for testing $H_0 : \mu_1 = \mu_2$ is defined as

$$L_m = \sup\Big\{\prod_{i=1}^{m}(mp_i) : p_1 \geq 0, \ldots, p_m \geq 0, \sum_{i=1}^{m} p_i = 1, \sum_{i=1}^{m} p_i Z_i = (0,0)^T\Big\}.$$

By the Lagrange multiplier technique, we have $p_i = m^{-1}\{1 + \beta^T Z_i\}^{-1}$ for $i = 1, \ldots, m$, and

$$l_m := -2\log L_m = 2\sum_{i=1}^{m}\log\{1 + \beta^T Z_i\},$$

where β satisfies

$$\frac{1}{m}\sum_{i=1}^{m}\frac{Z_i}{1+\beta^T Z_i} = (0,0)^T. \tag{2.3}$$

Write $\Sigma = (\sigma_{i,j})_{1\leq i\leq d, 1\leq j\leq d} = \mathrm{E}\{(X_1 - \mu_1)(X_1 - \mu_1)^T\}$, the covariance matrix of X_1, and use $\lambda_1 \leq \cdots \leq \lambda_d$ to denote the d eigenvalues of the matrix Σ. Similarly, write $\bar{\Sigma} = (\bar{\sigma}_{i,j})_{1\leq i\leq d, 1\leq j\leq d} = \mathrm{E}\{(Y_1 - \mu_2)(Y_1 - \mu_2)^T\}$ and use $\bar{\lambda}_1 \leq \cdots \leq \bar{\lambda}_d$ to denote the d eigenvalues of the matrix $\bar{\Sigma}$. Also write

$$\rho_1 = \sum_{i,j=1}^{d}\sigma_{i,j}^2 = \mathbf{tr}(\Sigma^2),\ \rho_2 = \sum_{i,j=1}^{d}\bar{\sigma}_{i,j}^2 = \mathbf{tr}(\bar{\Sigma}^2),\ \tau_1 = 2\alpha^T\Sigma\alpha,\ \tau_2 = 2\alpha^T\bar{\Sigma}\alpha. \tag{2.4}$$

Here **tr** means trace for a matrix. Note that $\rho_1 = \mathrm{E}\,[(X_1 - \mu_1)^T(X_2 - \mu_1)]^2$, $\rho_2 = \mathrm{E}\,[(Y_1 - \mu_2)^T(Y_2 - \mu_2)]^2$, $\tau_1 = 2\mathrm{E}\,[\alpha^T(X_1 - \mu_1)]^2$ and $\tau_2 = 2\mathrm{E}\,[\alpha^T(Y_1 - \mu_2)]^2$, and these quantities may depend on n_1, n_2 since d does.

Theorem 1. *Suppose τ_1 and τ_2 in (2.4) are positive, and for some $\delta > 0$,*

$$\frac{\mathrm{E}\,|(X_1 - \mu_1)^T(\bar{X}_1 - \mu_1)|^{2+\delta}}{\rho_1^{(2+\delta)/2}} = o(m_1^{(\delta+\min(\delta,2))/4}), \tag{2.5}$$

$$\frac{\mathrm{E}\,|(Y_1-\mu_2)^T(\bar{Y}_1-\mu_2)|^{2+\delta}}{\rho_2^{(2+\delta)/2}} = o(m_2^{(\delta+\min(\delta,2))/4}), \tag{2.6}$$

$$\frac{\mathrm{E}\,|\alpha^T(X_1+\bar{X}_1-2\mu_1)|^{2+\delta}}{\tau_1^{(2+\delta)/2}} = o(m_1^{(\delta+\min(\delta,2))/4}), \tag{2.7}$$

$$\frac{\mathrm{E}\,|\alpha^T(Y_1+\bar{Y}_1-2\mu_2)|^{2+\delta}}{\tau_2^{(2+\delta)/2}} = o(m_2^{(\delta+\min(\delta,2))/4}). \tag{2.8}$$

Then, under $H_0 : \mu_1 = \mu_2$, l_m converges in distribution to a chi-square distribution with two degrees of freedom as $\min\{n_1, n_2\} \to \infty$.

Based on this result, one can test $H_0 : \mu_1 = \mu_2$ against $H_a : \mu_1 \neq \mu_2$ by rejecting H_0 when $l_m \geq \chi^2_{2,\gamma}$, where $\chi^2_{2,\gamma}$ denotes the $(1-\gamma)-$quantile of a chi-square distribution with two degrees of freedom and γ is the significant level.

Remark 1. In equations (2.5)−(2.8), the restrictions are put on $\mathrm{E}\,[|W|^{2+\delta}/(\mathrm{E}\,W^2)^{(2+\delta)/2}]$ for some random variables W; they are necessary for the CLT to hold for random arrays. Later we will see that the restrictions are satisfied by imposing some conditions on the higher-order moments or the dependence structure.

Remark 2. The proposed test has the following merits:

1. The limiting distribution is always chi-square, and does not require estimation of the asymptotic covariance.
2. It does not require a specific structure such as the one used in Bai and Saranadasa (1996) and Chen and Qin (2010).
3. With a higher-order moment condition or a special dependence structure of $\{X_i\}$ and $\{Y_i\}$, d can be very large.
4. There is no restriction imposed on the relation between n_1 and n_2 except that $\min\{n_1, n_2\} \to \infty$. Moreover, no assumptions are needed on ρ_1/ρ_2 or τ_1/τ_2. Hence the covariance matrices Σ_1 and Σ_2 can be arbitrary as long as τ_1 and τ_2 are positive, so that $\alpha^T X_1$ and $\alpha^T Y_1$ are non-degenerate.

We verify Theorem 1 by imposing conditions on the moments and dimension of the random vectors.

A1: $0 < \liminf_{n_1\to\infty} \lambda_1 \leq \limsup_{n_1\to\infty} \lambda_d < \infty$ and $0 < \liminf_{n_2\to\infty} \bar{\lambda}_1 \leq \limsup_{n_2\to\infty} \bar{\lambda}_d < \infty$;

A2: For some $\delta > 0$, $(1/d)\sum_{i=1}^d \mathrm{E}\,\{|X_{1,i}-\mu_{1,i}|^{2+\delta} + |Y_{1,i}-\mu_{2,i}|^{2+\delta}\} = O(1)$;

A3: $d = o(m^{[\delta+\min(\delta,2)]/[2(2+\delta)]})$.

Corollary 1. *If* $\min\{n_1, n_2\} \to \infty$ *and* **A1**$-$**A3** *hold, under* $H_0 : \mu_1 = \mu_2$, *Theorem* 1 *holds.*

Condition **A3** is a somewhat restrictive condition on the dimension d. Note that **A1** and **A2** are related only to the covariance matrices and some higher moments off the components of the random vectors: higher the moments, less the restriction on d. Condition **A3** can be removed for models with some special dependence structure. For comparisons, we prove the Wilks Theorem for the proposed jackknife empirical likelihood test under a model considered by Bai and Saranadasa (1996), Chen, Peng, and Qin (2009), and Chen and Qin (2010):

B (Factor model). $X_i = \Gamma_1 B_i + \mu_1$ for $i = 1, \ldots, n_1$, $Y_j = \Gamma_2 \bar{B}_j + \mu_2$ for $j = 1, \ldots, n_2$, where Γ_1, Γ_2 are $d \times k$ matrices with $\Gamma_1 \Gamma_1^T = \Sigma$, $\Gamma_2 \Gamma_2^T = \bar{\Sigma}$, $\{B_i = (B_{i,1}, \ldots, B_{i,k})^T\}_{i=1}^{n_1}$ and $\{\bar{B}_j = (\bar{B}_{j,1}, \ldots, \bar{B}_{j,k})^T\}_{i=1}^{n_2}$ are two independent random samples satisfying $\mathrm{E}\, B_i = \mathrm{E}\, \bar{B}_i = 0$, $\mathrm{Var}\,(B_i) = \mathrm{Var}\,(\bar{B}_i) = I_{k\times k}$, $\mathrm{E}\, B_{i,j}^4 = 3 + \xi_1 < \infty$, $\mathrm{E}\, \bar{B}_{i,j}^4 = 3 + \xi_2 < \infty$, $\mathrm{E} \prod_{l=1}^k B_{i,l}^{\nu_l} = \prod_{l=1}^k \mathrm{E}\, B_{i,l}^{\nu_l}$ and $\mathrm{E} \prod_{l=1}^k \bar{B}_{i,l}^{\nu_l} = \prod_{l=1}^k \mathrm{E}\, \bar{B}_{i,l}^{\nu_l}$ whenever $\nu_1 + \cdots + \nu_k = 4$ for distinct nonnegative integers ν_l's.

Theorem 2. *Suppose* τ_1 *and* τ_2 *in* (2.4) *are positive. Under* ***B*** *and* $H_0 : \mu_1 = \mu_2$, l_m *converges in distribution to a chi-square distribution with two degrees of freedom as* $\min\{n_1, n_2\} \to \infty$.

Remark 3. It can be seen from the proof of Theorem 2 that the assumptions $\mathrm{E}\, B_{i,j}^4 = 3 + \xi_1 < \infty$ and $\mathrm{E}\, \bar{B}_{i,j}^4 = 3 + \xi_2 < \infty$ in model **B** can be replaced by the much weaker conditions $\max_{1\le j\le k} \mathrm{E}\, B_{1,j}^4 = o(m)$ and $\max_{1\le j\le k} \mathrm{E}\, \bar{B}_{1,j}^4 = o(m)$. Unlike Bai and Saranadasa (1996) and Chen and Qin (2010), there is no restriction on d and k for our proposed method. The only constraint imposed on matrices Γ_1 and Γ_2 is that both $\alpha^T \Sigma \alpha$ and $\alpha^T \bar{\Sigma} \alpha$ are positive, which is very weak.

3. Simulation Study and Data Analysis

3.1. Simulation study

We investigated the finite sample behavior of the proposed jackknife empirical likelihood test (JEL) and compared it with the test statistic in (1.2) proposed by Chen and Qin (2010) in terms of both size and power.

Let $W_1, \ldots, W_d$ be iid $N(0,1)$ random variables, and let $\bar{W}_1, \ldots, \bar{W}_d$, independent of the $W_i's$ be iid $t(8)$ random variables. Put $X_{1,1} = W_1, X_{1,2} = W_1 + W_2, \ldots, X_{1,d} = W_{d-1} + W_d$, $Y_{1,1} = \bar{W}_1 + \mu_{2,1}, Y_{1,2} = \bar{W}_1 + \bar{W}_2 + \mu_{2,2}, \ldots, Y_{1,d} = \bar{W}_{d-1} + \bar{W}_d + \mu_{2,d}$, where $\mu_{2,i} = c_1$ if $i \le [c_2 d]$, and $\mu_{2,i} = 0$ if $i > [c_2 d]$. Thus, $100c_2\%$ of the components of Y_1 have a shifted mean compared to that of X_1.

Table 1. Sizes and powers of the proposed jackknife empirical likelihood test (JEL) and the test in Chen and Qin (2010) (CQ) are reported for the case of $(n_1, n_2) = (30, 30)$ at level 5%.

d	JEL $c_1 = 0$ $c_2 = 0.25$	CQ $c_1 = 0$ $c_2 = 0.25$	JEL $c_1 = 0.1$ $c_2 = 0.25$	CQ $c_1 = 0.1$ $c_2 = 0.25$	JEL $c_1 = 0.1$ $c_2 = 0.75$	CQ $c_1 = 0.1$ $c_2 = 0.75$
10	0.070	0.049	0.071	0.049	0.072	0.062
20	0.056	0.037	0.057	0.049	0.096	0.060
30	0.064	0.047	0.066	0.049	0.113	0.066
40	0.070	0.052	0.069	0.058	0.116	0.072
50	0.067	0.049	0.083	0.054	0.138	0.067
60	0.063	0.039	0.069	0.043	0.174	0.055
70	0.053	0.053	0.076	0.065	0.190	0.081
80	0.056	0.059	0.063	0.067	0.191	0.082
90	0.056	0.044	0.080	0.054	0.204	0.071
100	0.066	0.060	0.082	0.064	0.229	0.091
300	0.056	0.045	0.114	0.054	0.537	0.092
500	0.049	0.051	0.160	0.063	0.731	0.110

Since we test $H_0 : \mathrm{E}\, X_1 = \mathrm{E}\, Y_1$ against $H_a : \mathrm{E}\, X_1 \neq \mathrm{E}\, Y_1$, the case of $c_1 = 0$ denotes the size of tests. After drawing $1,000$ random samples of sizes $n_1 = 30, 100, 150$ from $X = (X_{1,1}, \ldots, X_{1,d})^T$ and independently drawing $1,000$ random samples of sizes $n_2 = 30, 100, 200$ from $Y = (Y_{1,1}, \ldots, Y_{1,d})^T$ with $d = 10, 20, \ldots, 100, 300, 500$, $c_1 = 0, 0.1$, and $c_2 = 0.25, 0.75$, we calculated the powers of the two tests mentioned above.

In Tables 1−3, we report the empirical sizes and powers for the proposed jackknife empirical likelihood test with $\alpha = (1, \ldots, 1)^T \in \mathbb{R}^d$ and the test in Chen and Qin (2010), at level 5%. Results for level 10% are similar. From the tables we observe that the sizes of both tests are comparable and quite stable with respect to the dimension d; the proposed jackknife empirical likelihood test is more powerful than the test in Chen and Qin (2010) for the case $c_2 = 0.75$, and the case when the data is sparse but d is large. Since equation (2.2) has nothing to do with sparsity, it is expected that the proposed jackknife empirical likelihood method is not powerful when the data is sparse. Hence, it would be of interest to connect sparsity with some estimating equations so as to improve the power of the proposed jackknife empirical likelihood test.

In conclusion, we have good evidence that the proposed jackknife empirical likelihood test has a very stable size with respect to the dimension and is powerful as well. Moreover, the new test is easy to compute, flexible enough to take other information into account, and works for both fixed and divergent dimension. In comparison with the test in Chen and Qin (2010), the new test has a comparable

Table 2. Sizes and powers of the proposed jackknife empirical likelihood test (JEL) and the test in Chen and Qin (2010) (CQ) are reported for the case of $(n_1, n_2) = (100, 100)$ at level 5%.

d	JEL $c_1 = 0$ $c_2 = 0.25$	CQ $c_1 = 0$ $c_2 = 0.25$	JEL $c_1 = 0.1$ $c_2 = 0.25$	CQ $c_1 = 0.1$ $c_2 = 0.25$	JEL $c_1 = 0.1$ $c_2 = 0.75$	CQ $c_1 = 0.1$ $c_2 = 0.75$
10	0.074	0.054	0.072	0.063	0.099	0.090
20	0.043	0.047	0.053	0.055	0.145	0.098
30	0.047	0.047	0.056	0.063	0.191	0.115
40	0.051	0.050	0.063	0.062	0.264	0.125
50	0.055	0.040	0.077	0.061	0.326	0.131
60	0.055	0.044	0.077	0.067	0.374	0.151
70	0.043	0.051	0.063	0.086	0.395	0.150
80	0.042	0.059	0.082	0.079	0.474	0.171
90	0.043	0.040	0.098	0.065	0.527	0.163
100	0.049	0.054	0.091	0.088	0.575	0.194
300	0.048	0.054	0.217	0.102	0.974	0.389
500	0.049	0.041	0.353	0.115	0.999	0.544

Table 3. Sizes and powers of the proposed jackknife empirical likelihood test (JEL) and the test in Chen and Qin (2010) (CQ) are reported for the case of $(n_1, n_2) = (150, 200)$ at level 5%.

d	JEL $c_1 = 0$ $c_2 = 0.25$	CQ $c_1 = 0$ $c_2 = 0.25$	JEL $c_1 = 0.1$ $c_2 = 0.25$	CQ $c_1 = 0.1$ $c_2 = 0.25$	JEL $c_1 = 0.1$ $c_2 = 0.75$	CQ $c_1 = 0.1$ $c_2 = 0.75$
10	0.048	0.054	0.054	0.062	0.129	0.116
20	0.055	0.042	0.078	0.075	0.237	0.166
30	0.052	0.054	0.079	0.081	0.330	0.207
40	0.039	0.035	0.070	0.068	0.430	0.212
50	0.039	0.048	0.071	0.094	0.480	0.231
60	0.047	0.051	0.092	0.095	0.598	0.273
70	0.046	0.051	0.086	0.107	0.658	0.309
80	0.042	0.047	0.113	0.109	0.753	0.327
90	0.046	0.043	0.148	0.098	0.781	0.346
100	0.048	0.059	0.141	0.117	0.821	0.365
300	0.044	0.040	0.370	0.163	1	0.703
500	0.047	0.045	0.555	0.235	1	0.899

size, but is more powerful when the data is not very sparse. Some further research on formulating sparsity into estimating equations will be pursued in future.

3.2. Data analysis

We applied the proposed method to the Colon data with 2,000 gene ex-

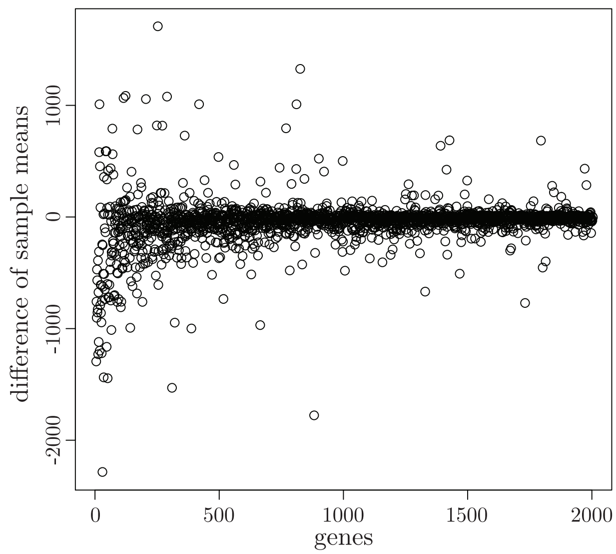

Figure 1. Colon data: differences of the sample means are plotted against each gene.

pression levels on 22 (n_1) normal colon tissues and 40 (n_2) tumor colon tissues. This data set is available from the R package 'plsgenomics', and has been anlayzed by Alon et al. (1999) and Srivastava and Du (2008). The p-values of three tests proposed by Srivastava and Du (2008) equal to 1.38e-06, 4.48e-11 and 0.00000, which clearly reject the null hypothesis that the tumor group has the same gene expression levels as the normal group. A direct application of the proposed jackknife empirical likelihood method and the CQ test for testing the equality of means gives p-values 1.36e-01 and 5.06e-09, respectively, contradictory results. Although the test in Chen and Qin (2010) and the three tests in Srivastava and Du (2008) clearly reject the null hypothesis, the p-values are significantly different. A closer look at the difference of sample means (see Figure 1) shows that some genes have a significant difference of sample means and a high variability; this may play an important role in the CQ test, and (2.2) with $\alpha = (1, \ldots, 1)^T \in \mathbb{R}^d$ for the proposed jackknife empirical likelihood method. To examine this effect, we applied the methods to those genes without the significant difference in sample means and the logarithms of the 2,000 gene expression levels.

First we applied the proposed jackknife empirical likelihood method and the CQ test to those genes satisfying $|n_1^{-1} \sum_{i=1}^{n_1} X_{i,j} - n_2^{-1} \sum_{i=1}^{n_2} Y_{i,j}| \le c_3$ for some given threshold $c_3 > 0$. In Table 4, we report the p-values for different c_3; this confirms the fact that some genes play an important role in rejecting the equality of means in the CQ test.

Table 4. Colon data: p-values for testing equal means of those genes with the absolute difference of sample means less than the threshold c_3.

c_3	number of genes	CQ	JEL
50	1158	2.94e-01	2.13e-01
100	1501	5.63e-01	2.82e-01
200	1742	7.21e-01	3.87e-01
500	1913	2.71e-02	3.75e-01
1000	1978	6.79e-05	3.40e-01
3000	2000	5.06e-09	1.36e-01

Figure 1 clearly shows that some genes have a large positive mean and some genes have a large negative mean, and the equation (2.2) with the simple $a = (1, 1, \ldots, 1)$ can not catch this characteristic. Here, we propose to replace (2.2) by

$$\frac{1}{m_1 m_2}\sum_{i=1}^{m_1}\sum_{j=1}^{m_2}(X_i - Y_j)^T\{I(\bar{X}_i - \bar{Y}_j > 0) - I(\bar{X}_i - \bar{Y}_j < 0)\}, \tag{3.1}$$

which results in the P-value 2.63e-03, and so we reject the null hypothesis that the tumor group has the same gene expression levels as the normal group. Note that $I(\bar{X}_i - \bar{Y}_i > 0)$ means $(I(\bar{X}_{i,1} - \bar{Y}_{i,1} > 0), \ldots, I(\bar{X}_{i,d} - \bar{Y}_{i,d} > 0))^T$. Although the derived theorems based on (2.1) and (2.2) can not be applied to (2.1) and (3.1) directly, results hold under some similar conditions by noting the boundedness of $I(\bar{X}_i - \bar{Y}_j > 0) - I(\bar{X}_i - \bar{Y}_j < 0)$.

It is well known that gene expression data are quite noisy and some transformation and normalization are needed before doing statistical analysis; see Chapter 6 of Lee (2004). Here we applied the CQ test and the proposed empirical likelihood methods based on (2.1) and (2.2), and (2.1) and (3.1) to testing the equality of means of the logarithms of the 2000 gene expression levels on normal colon tissues and tumor colon tissues, which give p-values 0.184, 0.206 and 0.148, respectively. We plot the differences of sample means of the logarithms in Figure 2, and note much less volatility than for the differences of sample means in Figure 1.

In summary, carefully choosing α in the empirical likelihood method is needed when it is applied to the colon data, which has a small sample size and a large variation. Simply choosing $\alpha = (1, \ldots, 1)^T$ in the empirical likelihood method gives a similar result as the test in Chen and Qin (2010) for testing the equal means of the logarithms of Colon data.

4. Proofs

In the proofs we use $||\cdot||$ to denote the L_2 norm of a vector or matrix. Under the null hypothesis $\mu_1 - \mu_2 = 0$, without loss of generality we take $\mu_1 = \mu_2 = 0$.

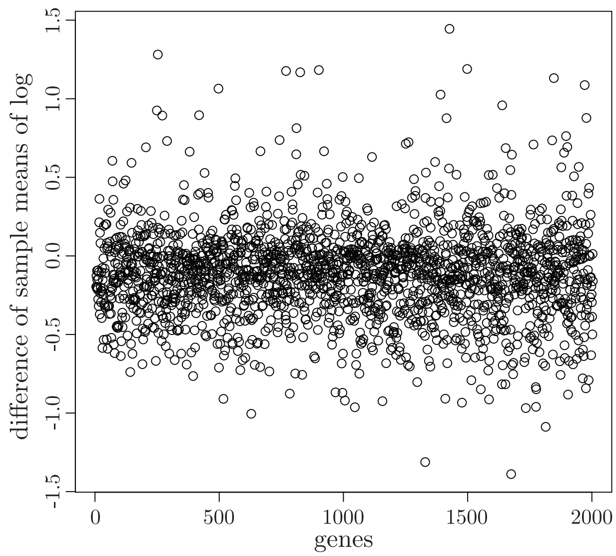

Figure 2. Colon data: differences of the sample means of logarithms of gene expression levels are plotted against each gene.

Write $u_{ij} = (X_i - Y_j)^T(\bar{X}_i - \bar{Y}_j)$ and $v_{ij} = \alpha^T(X_i - Y_j) + \alpha^T(\bar{X}_i - \bar{Y}_j)$ for $1 \le i \le m_1, 1 \le j \le m_2$. Then it is easily verified that for $1 \le i, k \le m_1, 1 \le j, l \le m_2$,

$$\begin{aligned} \mathrm{E}\,(u_{ij}) &= \mathrm{E}\,(v_{kl}) = \mathrm{E}\,(u_{ij}v_{kl}) = 0, \\ \mathrm{Var}\,(u_{kl}) &= \sum_{i,j=1}^{d} (\sigma_{i,j}^2 + \bar{\sigma}_{i,j}^2) = \rho_1 + \rho_2, \\ \mathrm{Var}\,(v_{kl}) &= 2\alpha^T(\Sigma + \bar{\Sigma})\alpha = \tau_1 + \tau_2. \end{aligned}$$

Lemma 1. *Under the conditions of Theorem* 1 *we have, as* $\min\{n_1, n_2\} \to \infty$,

$$\frac{1}{\sqrt{m_1}} \sum_{i=1}^{m_1} \frac{X_i^T \bar{X}_i}{\sqrt{\rho_1}} \xrightarrow{d} N(0,1), \tag{4.1}$$

$$\frac{1}{\sqrt{m_2}} \sum_{j=1}^{m_2} \frac{Y_j^T \bar{Y}_j}{\sqrt{\rho_2}} \xrightarrow{d} N(0,1), \tag{4.2}$$

$$\frac{1}{\sqrt{m_1}} \sum_{i=1}^{m_1} \frac{\alpha^T(X_i + \bar{X}_i)}{\sqrt{\tau_1}} \xrightarrow{d} N(0,1), \tag{4.3}$$

and

$$\frac{1}{\sqrt{m_2}} \sum_{j=1}^{m_2} \frac{\alpha^T(Y_j + \bar{Y}_j)}{\sqrt{\tau_2}} \xrightarrow{d} N(0,1). \tag{4.4}$$

Proof. Since $\mathrm{Var}\,(X_i^T \bar{X}_i) = \rho_1$ and $X_1^T \bar{X}_1, \ldots, X_{m_1}^T \bar{X}_{m_1}$ are i.i.d. for fixed m_1, (4.1) follows from (2.5) and the Lyapunov Central Limit Theorem. The rest can be shown in the same way.

From now on we write

$$\rho = \frac{m}{m_1}\rho_1 + \frac{m}{m_2}\rho_2 \quad \text{and} \quad \tau = \frac{m}{m_1}\tau_1 + \frac{m}{m_2}\tau_2.$$

Lemma 2. *Under the conditions of Theorem* 1 *we have, as* $\min\{n_1, n_2\} \to \infty$,

$$\frac{\sqrt{m}}{m_1 m_2 \sqrt{\rho}} \sum_{i=1}^{m_1} X_i^T \sum_{j=1}^{m_2} \bar{Y}_j \xrightarrow{p} 0, \tag{4.5}$$

$$\frac{1}{m_1\sqrt{\tau}} \sum_{i=1}^{m_1} \alpha^T X_i \xrightarrow{p} 0, \tag{4.6}$$

$$\frac{1}{m_2\sqrt{\tau}} \sum_{j=1}^{m_2} \alpha^T Y_j \xrightarrow{p} 0, \tag{4.7}$$

$$\frac{1}{m_1} \sum_{i=1}^{m_1} \frac{(X_i^T \bar{X}_i)^2}{\rho_1} \xrightarrow{p} 1, \tag{4.8}$$

$$\frac{1}{m_2} \sum_{j=1}^{m_2} \frac{(Y_j^T \bar{Y}_j)^2}{\rho_2} \xrightarrow{p} 1, \tag{4.9}$$

$$\frac{1}{m_1} \sum_{i=1}^{m_1} \frac{[\alpha^T (X_i + \bar{X}_i)]^2}{\tau_1} \xrightarrow{p} 1, \tag{4.10}$$

$$\frac{1}{m_2} \sum_{j=1}^{m_2} \frac{[\alpha^T (Y_j + \bar{Y}_j)]^2}{\tau_2} \xrightarrow{p} 1, \tag{4.11}$$

$$\frac{1}{m_1} \sum_{i=1}^{m_1} \frac{X_i^T \bar{X}_i [\alpha^T (X_i + \bar{X}_i)]}{\sqrt{\rho_1 \tau_1}} \xrightarrow{p} 0, \tag{4.12}$$

$$\frac{1}{m_2} \sum_{i=1}^{m_2} \frac{Y_j^T \bar{Y}_j [\alpha^T (Y_j + \bar{Y}_j)]}{\sqrt{\rho_2 \tau_2}} \xrightarrow{p} 0. \tag{4.13}$$

Proof. (4.5) follows from the fact that

$$\begin{aligned}
&\mathrm{Var}\,\Big(\frac{\sqrt{m}}{m_1 m_2 \sqrt{\rho}} \sum_{i=1}^{m_1} X_i^T \sum_{j=1}^{m_2} Y_j\Big) = \mathrm{E}\Big[\frac{m}{m_1^2 m_2^2 \rho}\Big(\sum_{i=1}^{m_1}\sum_{j=1}^{m_2} X_i^T Y_j\Big)^2\Big] \\
&= \mathrm{E}\Big[\frac{m}{m_1^2 m_2^2 \rho} \sum_{i=1}^{m_1}\sum_{j=1}^{m_2} (X_i^T Y_j)^2\Big] = \mathrm{E}\Big[\frac{m}{m_1 m_2 \rho}(X_1^T Y_1)^2\Big]
\end{aligned}$$

$$= \frac{m}{m_1 m_2 \rho} \sum_{i,j=1}^{d} \sigma_{ij} \bar{\sigma}_{ij} \le \frac{m}{m_1 + m_2} \frac{\rho_1 + \rho_2}{2\rho} \le \frac{1}{2}(\frac{1}{m_1} + \frac{1}{m_2}) = o(1).$$

In the same way, we can show (4.6) and (4.7).

To show (4.8), write $u_i = X_i^T \bar{X}_i$. We need to estimate $\mathrm{E}\,|\sum_{i=1}^{m_1} u_i^2 - m_1\rho_1|^{(2+\delta)/2}$. Note that $\rho_1 = \mathrm{E}\, u_1^2$. When $0 < \delta \le 2$, it follows from von Bahr and Esseen (1965) that

$$\mathrm{E}\Big|\sum_{i=1}^{m_1} u_i^2 - m_1\rho_1\Big|^{(2+\delta)/2} \le 2m_1 \mathrm{E}\,|u_1^2 - \mathrm{E}\,(u_1^2)|^{(2+\delta)/2} = O(m_1 \mathrm{E}\,|u_1|^{2+\delta}). \quad (4.14)$$

When $\delta > 2$, it follows from Dharmadhikari and Jogdeo (1969) that

$$\begin{aligned} \mathrm{E}\Big|\sum_{i=1}^{m_1} u_i^2 - m_1\rho_1\Big|^{(2+\delta)/2} &= O(m_1^{(2+\delta)/4} \mathrm{E}\,|u_1^2 - \mathrm{E}\,(u_1^2)|^{(2+\delta)/2}) \\ &= O(m_1^{(2+\delta)/4} \mathrm{E}\,|u_1|^{2+\delta}). \quad (4.15) \end{aligned}$$

Therefore, by (4.14), (4.15) and (2.5), we have for any $\varepsilon > 0$,

$$\begin{aligned} \mathbb{P}\Big(|\frac{\sum_{i=1}^{m_1} u_i^2}{m_1\rho_1} - 1| > \varepsilon\Big) &\le \varepsilon^{-(2+\delta)/2} \frac{\mathrm{E}\,|\sum_{i=1}^{m_1} u_i^2 - m\rho_1|^{(2+\delta)/2}}{(m_1\rho_1)^{(2+\delta)/2}} \\ &= O\Big(m_1^{-(\delta+\min(\delta,2))/4} \mathrm{E}\,|\frac{u_1}{\sqrt{\rho_1}}|^{2+\delta}\Big) = o(1), \end{aligned}$$

which implies (4.8). The rest can be shown in the same way.

Lemma 3. *Under the conditions of Theorem* 1 *we have, as* $\min\{n_1, n_2\} \to \infty$,

$$\frac{\sqrt{m}}{m_1 m_2} \sum_{j=1}^{m_2} \sum_{i=1}^{m_1} \begin{pmatrix} \frac{u_{ij}}{\sqrt{\rho}} \\ \frac{v_{ij}}{\sqrt{\tau}} \end{pmatrix} \xrightarrow{d} N(0, I_2), \quad (4.16)$$

$$\frac{m}{m_1^2 m_2^2 \rho} \sum_{k=1}^{m_1} \Big(\sum_{j=1}^{m_2} u_{kj}\Big)^2 - \frac{m\rho_1}{m_1\rho} \xrightarrow{p} 0, \quad (4.17)$$

$$\frac{m}{m_1^2 m_2^2 \rho} \sum_{k=1}^{m_2} \Big(\sum_{i=1}^{m_1} u_{ik}\Big)^2 - \frac{m\rho_2}{m_2\rho} \xrightarrow{p} 0, \quad (4.18)$$

$$\frac{m}{m_1^2 m_2^2 \tau} \sum_{k=1}^{m_1} \Big(\sum_{j=1}^{m_2} v_{kj}\Big)^2 - \frac{m\tau_1}{m_1\tau} \xrightarrow{p} 0, \quad (4.19)$$

$$\frac{m}{m_1^2 m_2^2 \tau} \sum_{k=1}^{m_2} \Big(\sum_{i=1}^{m_1} v_{ik}\Big)^2 - \frac{m\tau_2}{m_1\tau} \xrightarrow{p} 0, \quad (4.20)$$

$$\frac{m}{m_1^2 m_2^2 \sqrt{\rho\tau}} \sum_{k=1}^{m_1} \Big(\sum_{i=1}^{m_2} u_{ki} \sum_{j=1}^{m_2} v_{kj} \Big) \xrightarrow{p} 0, \tag{4.21}$$

$$\frac{m}{m_1^2 m_2^2 \sqrt{\rho\tau}} \sum_{k=1}^{m_2} \Big(\sum_{i=1}^{m_1} u_{ik} \sum_{j=1}^{m_1} v_{jk} \Big) \xrightarrow{p} 0, \tag{4.22}$$

where I_2 is the 2×2 identity matrix.

Proof. It follows from Lemma 2 that

$$\begin{aligned}
\frac{\sqrt{m}}{m_1 m_2} \sum_{j=1}^{m_2} \sum_{i=1}^{m_1} \frac{u_{ij}}{\sqrt{\rho}} &= \frac{\sqrt{m}}{m_1 m_2 \sqrt{\rho}} \sum_{j=1}^{m_2} \sum_{i=1}^{m_1} (X_i^T \bar{X}_i + Y_j^T \bar{Y}_j - X_i^T \bar{Y}_j - Y_j^T \bar{X}_i) \\
&= \frac{\sqrt{m}}{m_1 \sqrt{\rho}} \sum_{i=1}^{m_1} X_i^T \bar{X}_i + \frac{\sqrt{m}}{m_2 \sqrt{\rho}} \sum_{j=1}^{m_2} Y_j^T \bar{Y}_j - \frac{\sqrt{m}}{m_1 m_2 \sqrt{\rho}} \sum_{j=1}^{m_2} \sum_{i=1}^{m_1} (X_i^T \bar{Y}_j + Y_j^T \bar{X}_i) \\
&= \frac{\sqrt{m\rho_1}}{\sqrt{m_1 \rho}} \frac{1}{\sqrt{m_1}} \sum_{i=1}^{m_1} \frac{X_i^T \bar{X}_i}{\sqrt{\rho_1}} + \frac{\sqrt{m\rho_2}}{\sqrt{m_2 \rho}} \frac{1}{\sqrt{m_2}} \sum_{j=1}^{m_2} \frac{Y_j^T \bar{Y}_j}{\sqrt{\rho_2}} + o_p(1) \\
&= a_m A_m + b_m B_m + o_p(1),
\end{aligned}$$

where $a_m = \sqrt{m\rho_1}/\sqrt{m_1\rho}$, $b_m = \sqrt{m\rho_2}/\sqrt{m_2\rho}$, $A_m = (1/\sqrt{m_1}) \sum_{i=1}^{m_1} X_i^T \bar{X}_i / \sqrt{\rho_1} \xrightarrow{d} N(0,1)$ and $B_m = (1/\sqrt{m_2}) \sum_{j=1}^{m_2} Y_j^T \bar{Y}_j / \sqrt{\rho_2} \xrightarrow{d} N(0,1)$. Obviously $a_m^2 + b_m^2 = 1$ and A_m, B_m are independent. Denote the characteristic functions of A_m and B_m by Φ_m and Ψ_m, respectively. Then,

$$\begin{aligned}
\mathrm{E}\exp(it(a_m A_m + b_m B_m)) &= \mathrm{E}\exp(it a_m A_m)\mathrm{E}\exp(it b_m B_m) = \Phi_m(t a_m)\Psi_m(t b_m) \\
&= [\exp(-\frac{(t a_m)^2}{2}) + o(1)][\exp(-\frac{(t b_m)^2}{2}) + o(1)] \\
&= \exp(-\frac{t^2}{2}) + o(1),
\end{aligned}$$

i.e.,

$$\frac{\sqrt{m}}{m_1 m_2} \sum_{j=1}^{m_2} \sum_{i=1}^{m_1} \frac{u_{ij}}{\sqrt{\rho}} \xrightarrow{d} N(0,1). \tag{4.23}$$

Similarly, we have

$$\begin{aligned}
\frac{\sqrt{m}}{m_1 m_2} \sum_{j=1}^{m_2} \sum_{i=1}^{m_1} \frac{v_{ij}}{\sqrt{\tau}} = &\frac{\sqrt{m\tau_1}}{\sqrt{m_1 \tau}} \frac{1}{\sqrt{m_1}} \sum_{i=1}^{m_1} \frac{\alpha^T (X_i + \bar{X}_i)}{\sqrt{\tau_1}} \\
&- \frac{\sqrt{m\tau_2}}{\sqrt{m_2 \tau}} \frac{1}{\sqrt{m_2}} \sum_{j=1}^{m_2} \frac{\alpha^T (Y_j + \bar{Y}_j)}{\sqrt{\tau_2}} \xrightarrow{d} N(0,1).
\end{aligned}$$

Let a and b be two real numbers with $a^2+b^2\neq 0$. Note that

$$\begin{aligned}
&\frac{\sqrt{m}}{m_1m_2}\sum_{j=1}^{m_2}\sum_{i=1}^{m_1}(a\frac{u_{ij}}{\sqrt{\rho}}+b\frac{v_{ij}}{\sqrt{\tau}})\\
&=a\Big(\frac{\sqrt{m\rho_1}}{\sqrt{m_1\rho}}\frac{1}{\sqrt{m_1}}\sum_{i=1}^{m_1}\frac{X_i^T\bar{X}_i}{\sqrt{\rho_1}}-\frac{\sqrt{m\rho_2}}{\sqrt{m_2\rho}}\frac{1}{\sqrt{m_2}}\sum_{j=1}^{m_2}\frac{Y_j^T\bar{Y}_j}{\sqrt{\rho_2}}\Big)\\
&\quad+b\Big(\frac{\sqrt{m\tau_1}}{\sqrt{m_1\tau}}\frac{1}{\sqrt{m_1}}\sum_{i=1}^{m_1}\frac{\alpha^T(X_i+\bar{X}_i)}{\sqrt{\tau_1}}+\frac{\sqrt{m\tau_2}}{\sqrt{m_2\tau}}\frac{1}{\sqrt{m_2}}\sum_{j=1}^{m_2}\frac{\alpha^T(Y_j+\bar{Y}_j)}{\sqrt{\tau_2}}\Big)+o_p(1)\\
&=\Big(\frac{a\sqrt{m\rho_1}}{\sqrt{m_1\rho}}\frac{1}{\sqrt{m_1}}\sum_{i=1}^{m_1}\frac{X_i^T\bar{X}_i}{\sqrt{\rho_1}}+\frac{b\sqrt{m\tau_1}}{\sqrt{m_1\tau}}\frac{1}{\sqrt{m_1}}\sum_{i=1}^{m_1}\frac{\alpha^T(X_i+\bar{X}_i)}{\sqrt{\tau_1}}\Big)\\
&\quad+\Big(\frac{a\sqrt{m\rho_2}}{\sqrt{m_2\rho}}\frac{1}{\sqrt{m_2}}\sum_{j=1}^{m_2}\frac{Y_j^T\bar{Y}_j}{\sqrt{\rho_2}}-\frac{b\sqrt{m\tau_2}}{\sqrt{m_2\tau}}\frac{1}{\sqrt{m_2}}\sum_{j=1}^{m_2}\frac{\alpha^T(Y_j+\bar{Y}_j)}{\sqrt{\tau_2}}\Big)+o_p(1)\\
&=I_1+I_2+o_p(1).
\end{aligned}$$

Since $\sqrt{m\rho_1}/\sqrt{m_1\rho}$, $\sqrt{m\rho_2}/\sqrt{m_2\rho}$, $\sqrt{m\tau_1}/\sqrt{m_1\tau}$, $\sqrt{m\tau_2}/\sqrt{m_2\tau}$ are all bounded by one, it is easy to check that I_1 and I_2 satisfy the Lyapunov condition by (2.5)−(2.8). Therefore

$$\Big\{a^2\frac{m\rho_1}{m_1\rho}+b^2\frac{m\tau_1}{m_1\tau}\Big\}^{-1/2}I_1\xrightarrow{d}N(0,1)$$

and

$$\Big\{a^2\frac{m\rho_2}{m_2\rho}+b^2\frac{m\tau_2}{m_2\tau}\Big\}^{-1/2}I_2\xrightarrow{d}N(0,1).$$

Since the $X_i's$ are independent of the $Y_i's$, it follows from the arguments used in proving (4.23) that $I_1+I_2\xrightarrow{d}N(0,a^2+b^2)$, i.e., (4.16) holds.

To prove (4.17), we write

$$\begin{aligned}
&\frac{m}{m_1^2m_2^2\rho}\sum_{k=1}^{m_1}\Big(\sum_{j=1}^{m_2}u_{kj}\Big)^2=\frac{m}{m_1^2m_2^2\rho}\sum_{k=1}^{m_1}\Big(\sum_{j=1}^{m_2}(X_k^T\bar{X}_k+Y_j^T\bar{Y}_j-Y_j^T\bar{X}_k-X_k^T\bar{Y}_j)\Big)^2\\
&=\frac{m}{m_1^2\rho}\sum_{k=1}^{m_1}\Big(X_k^T\bar{X}_k+\frac{1}{m_2}\sum_{j=1}^{m_2}Y_j^T\bar{Y}_j-\frac{1}{m_2}\sum_{j=1}^{m_2}Y_j^T\bar{X}_k-X_k^T\frac{1}{m_2}\sum_{j=1}^{m_2}\bar{Y}_j\Big)^2. \qquad (4.24)
\end{aligned}$$

Since $m\rho_1/m_1\rho\le 1$, it follows from Lemma 2 that

$$\frac{m}{m_1^2\rho}\sum_{k=1}^{m_1}(X_k^T\bar{X}_k)^2-\frac{m\rho_1}{m_1\rho}\xrightarrow{p}0. \qquad (4.25)$$

By Lemma 1, we have

$$\frac{m}{m_1^2\rho}\sum_{k=1}^{m_1}\Big(\frac{1}{m_2}\sum_{j=1}^{m_2}Y_j^T\bar{Y}_j\Big)^2 = O_p\Big(\frac{m\rho_2}{m_1m_2\rho}\Big) = o_p(1). \tag{4.26}$$

A direct calculation shows that

$$\begin{aligned}
\mathrm{E}\Big\{\frac{1}{m_2}\sum_{j=1}^{m_2}Y_j^T\bar{X}_k\Big\}^2 &= \mathrm{E}\Big\{\Big(\frac{1}{m_2}\sum_{j=1}^{m_2}Y_j^T\Big)\bar{X}_k\bar{X}_k^T\Big(\frac{1}{m_2}\sum_{j=1}^{m_2}Y_j\Big)\Big\}\\
&= \mathrm{E}\,\mathbf{tr}\Big\{\frac{1}{m_2}\sum_{j=1}^{m_2}Y_j^T\bar{X}_k\bar{X}_k^T\Big(\frac{1}{m_2}\sum_{j=1}^{m_2}Y_j\Big)\Big\}\\
&= \mathrm{E}\,\mathbf{tr}\{\bar{X}_k\bar{X}_k^T\Big(\frac{1}{m_2}\sum_{j=1}^{m_2}Y_j\Big)\Big(\frac{1}{m_2}\sum_{i=1}^{m_2}Y_i^T)\Big\}\\
&= \mathbf{tr}\mathrm{E}\Big\{\bar{X}_k\bar{X}_k^T\Big(\frac{1}{m_2}\sum_{j=1}^{m_2}Y_j\Big)\Big(\frac{1}{m_2}\sum_{i=1}^{m_2}Y_i^T\Big)\Big\}\\
&= \mathbf{tr}\Big\{\Sigma\frac{1}{m_2}\bar{\Sigma}\Big\} = O\Big(\frac{\rho_1+\rho_2}{m_2}\Big)\\
&= O\Big(\frac{m_1\rho}{m_2m}\Big) + O\Big(\frac{\rho_2}{m_2}\Big)\\
&= o\Big(\frac{m_1\rho}{m}\Big),
\end{aligned}$$

which implies that

$$\frac{m}{m_1^2\rho}\sum_{k=1}^{m_1}\Big\{\frac{1}{m_2}\sum_{j=1}^{m_2}Y_j^T\bar{X}_k\Big\}^2 = o_p(1). \tag{4.27}$$

Similarly we have

$$\frac{m}{m_1^2\rho}\sum_{k=1}^{m_1}\Big\{X_k^T\frac{1}{m_2}\sum_{j=1}^{m_2}\bar{Y}_j\Big\}^2 = o_p(1). \tag{4.28}$$

It follows from (4.25) and (4.27) that

$$\begin{aligned}
&\Big|\frac{m}{m_1^2\rho}\sum_{k=1}^{m_1}(X_k^T\bar{X}_k)\Big(\frac{1}{m_2}\sum_{j=1}^{m_2}Y_j^T\bar{X}_k\Big)\Big|\\
&\le \Big\{\frac{m}{m_1^2\rho}\sum_{k=1}^{m_1}(X_k^T\bar{X}_k)^2\Big\}^{1/2}\Big\{\frac{m}{m_1^2\rho}\sum_{k=1}^{m_1}\Big(\frac{1}{m_2}\sum_{j=1}^{m_2}Y_j^T\bar{X}_k\Big)^2\Big\}^{1/2}\\
&= O_p(1)o_p(1) = o_p(1).
\end{aligned} \tag{4.29}$$

Similarly we can show that

$$\begin{cases} \dfrac{m}{m_1^2\rho}\displaystyle\sum_{k=1}^{m_1}(X_k^T\bar{X}_k)\Big(\frac{1}{m_2}\sum_{j=1}^{m_2}Y_j^T\bar{Y}_j\Big)=o_p(1) \\ \dfrac{m}{m_1^2\rho}\displaystyle\sum_{k=1}^{m_1}(X_k^T\bar{X}_k)(X_k^T\frac{1}{m_2}\sum_{j=1}^{m_2}\bar{Y}_j^T)=o_p(1) \\ \dfrac{m}{m_1^2\rho}\displaystyle\sum_{k=1}^{m_1}\Big(\frac{1}{m_2}\sum_{j=1}^{m_2}Y_j^T\bar{Y}_j\Big)\Big(\frac{1}{m_2}\sum_{i=1}^{m_2}Y_i^T\bar{X}_k\Big)=o_p(1) \\ \dfrac{m}{m_1^2\rho}\displaystyle\sum_{k=1}^{m_1}\Big(\frac{1}{m_2}\sum_{j=1}^{m_2}Y_j^T\bar{Y}_j\Big)\Big(X_k^T\frac{1}{m_2}\sum_{i=1}^{m_2}\bar{Y}_i\Big)=o_p(1) \\ \dfrac{m}{m_1^2\rho}\displaystyle\sum_{k=1}^{m_1}\Big(\frac{1}{m_2}\sum_{j=1}^{m_2}Y_j^T\bar{X}_k\Big)\Big(X_k^T\frac{1}{m_2}\sum_{i=1}^{m_2}Y_i\Big)=o_p(1). \end{cases} \tag{4.30}$$

Hence (4.17) follows from (4.24)–(4.30). The rest can be shown in the same way as was (4.17).

Lemma 4. *Under the conditions of Theorem* 1 *we have, as* $\min\{n_1,n_2\}\to\infty$,

$$\frac{1}{\sqrt{m}}\sum_{k=1}^{m}\begin{pmatrix}\frac{Z_{k,1}}{\sqrt{\rho}}\\ \frac{Z_{k,2}}{\sqrt{\tau}}\end{pmatrix}\xrightarrow{d}N(0,I_2), \tag{4.31}$$

$$\frac{1}{m\rho}\sum_{k=1}^{m}Z_{k,1}^2-1\xrightarrow{p}0, \tag{4.32}$$

$$\frac{1}{m\tau}\sum_{k=1}^{m}Z_{k,2}^2-1\xrightarrow{p}0, \tag{4.33}$$

$$\frac{1}{m\sqrt{\rho\tau}}\sum_{k=1}^{m}Z_{k,1}Z_{k,2}\xrightarrow{p}0. \tag{4.34}$$

Moreover,

$$\max_{1\le k\le m}|\frac{Z_{k,1}}{\sqrt{\rho}}|=o_p(m^{1/2})\quad\textit{and}\quad\max_{1\le k\le m}|\frac{Z_{k,2}}{\sqrt{\tau}}|=o_p(m^{1/2}). \tag{4.35}$$

Proof. Note that for $1\le k\le m_1$,

$$Z_{k,1}=\frac{-1}{(m_1-1)m_1}\sum_{j=1}^{m_2}\sum_{i=1}^{m_1}u_{ij}+\frac{m_1+m_2-1}{(m_1-1)m_2}\sum_{j=1}^{m_2}u_{kj},$$

$$Z_{k,2}=\frac{-1}{(m_1-1)m_1}\sum_{j=1}^{m_2}\sum_{i=1}^{m_1}v_{ij}+\frac{m_1+m_2-1}{(m_1-1)m_2}\sum_{j=1}^{m_2}v_{kj},$$

and for $m_1 + 1 \le k \le m$,

$$Z_{k,1} = \frac{-1}{(m_2-1)m_2}\sum_{j=1}^{m_2}\sum_{i=1}^{m_1} u_{ij} + \frac{m_1+m_2-1}{(m_2-1)m_1}\sum_{i=1}^{m_1} u_{i,k-m_1},$$

$$Z_{k,2} = \frac{-1}{(m_2-1)m_2}\sum_{j=1}^{m_2}\sum_{i=1}^{m_1} u_{ij} + \frac{m_1+m_2-1}{(m_2-1)m_1}\sum_{i=1}^{m_1} v_{i,k-m_1}.$$

Thus

$$\frac{1}{\sqrt{m}}\sum_{k=1}^{m}\frac{Z_{k,1}}{\sqrt{\rho}} = \frac{1}{\sqrt{m}}\Big(\frac{-1}{m_2-1} + \frac{-1}{m_1-1} + \frac{m_1+m_2-1}{(m_1-1)m_2} + \frac{m_1+m_2-1}{(m_2-1)m_1}\Big)\sum_{j=1}^{m_2}\sum_{i=1}^{m_1}\frac{u_{ij}}{\sqrt{\rho}}$$
$$= \frac{\sqrt{m}}{m_1 m_2}\sum_{j=1}^{m_2}\sum_{i=1}^{m_1}\frac{u_{ij}}{\sqrt{\rho}},$$

and similarly

$$\frac{1}{\sqrt{m}}\sum_{k=1}^{m}\frac{Z_{k,2}}{\sqrt{\tau}} = \frac{\sqrt{m}}{m_1 m_2}\sum_{j=1}^{m_2}\sum_{i=1}^{m_1}\frac{v_{ij}}{\sqrt{\tau}}.$$

This implies (4.31) by using Lemma 3.

It follows from Lemma 3 that

$$\frac{1}{m\rho}\sum_{k=1}^{m} Z_{k,1}^2$$
$$= \frac{1}{m\rho}\sum_{k=1}^{m_1}\Big(\frac{-1}{(m_1-1)m_1}\sum_{j=1}^{m_2}\sum_{i=1}^{m_1} u_{ij} + \frac{m_1+m_2-1}{(m_1-1)m_2}\sum_{j=1}^{m_2} u_{kj}\Big)^2$$
$$+ \frac{1}{m\rho}\sum_{k=1}^{m_2}\Big(\frac{-1}{(m_2-1)m_2}\sum_{j=1}^{m_2}\sum_{i=1}^{m_1} u_{ij} + \frac{m_1+m_2-1}{(m_2-1)m_1}\sum_{i=1}^{m_1} u_{ik}\Big)^2$$
$$= \Big\{\frac{1}{(m_1-1)\sqrt{m_1 m\rho}}\sum_{j=1}^{m_2}\sum_{i=1}^{m_1} u_{ij}\Big\}^2 + \frac{(m-1)^2}{(m_1-1)^2 m_2^2 m\rho}\sum_{k=1}^{m_1}\Big(\sum_{j=1}^{m_2} u_{kj}\Big)^2$$
$$-2\Big\{\Big(\frac{m-1}{m\rho(m_1-1)^2 m_1 m_2}\Big)^{1/2}\sum_{j=1}^{m_2}\sum_{i=1}^{m_1} u_{ij}\}^2 + \Big\{\frac{1}{(m_2-1)\sqrt{m_2 m\rho}}\sum_{j=1}^{m_2}\sum_{i=1}^{m_1} u_{ij}\Big\}^2$$
$$+ \frac{(m-1)^2}{m\rho(m_2-1)^2 m_1^2}\sum_{k=1}^{m_2}\Big(\sum_{i=1}^{m_1} u_{ik}\Big)^2 - 2\Big\{\Big(\frac{m-1}{m\rho(m_2-1)^2 m_2 m_1}\Big)^{1/2}\sum_{j=1}^{m_2}\sum_{i=1}^{m_1} u_{ij}\Big\}^2$$
$$= \Big\{O_p\Big(\frac{1}{m_1\sqrt{m_1 m}}\frac{m_1 m_2}{\sqrt{m}}\Big)\Big\}^2 + \frac{(m-1)^2 m_1^2}{(m_1-1)^2 m^2}\Big\{\frac{m\rho_1}{m_1\rho} + o_p(1)\Big\}$$
$$+ \Big\{O_p\Big(\frac{1}{m_1\sqrt{m_1 m_2}}\frac{m_1 m_2}{\sqrt{m}}\Big)\Big\}^2 + \Big\{O_p\Big(\frac{1}{m_2\sqrt{m_2 m}}\frac{m_1 m_2}{\sqrt{m}}\Big)\Big\}^2$$

$$
\begin{aligned}
&+\frac{(m-1)^2m_2^2}{m^2(m_2-1)^2}\Big\{\frac{m\rho_2}{m_2\rho}+o_p(1)\Big\}+\Big\{O_p\Big(\frac{1}{m_2\sqrt{m_2m_1}}\frac{m_1m_2}{\sqrt{m}}\Big)\Big\}^2\\
&=\frac{m\rho_1}{m_1\rho}+\frac{m\rho_2}{m_2\rho}+o_p(1)\\
&=1+o_p(1),
\end{aligned}
$$

so (4.32) holds. Similarly we can show (4.33) and (4.34).

Since $\mathrm{E}\,((\sum_{i=1}^{m_1}u_{ij})^2)=\mathrm{Var}\,(\sum_{i=1}^{m_1}u_{ij})=m_1(\rho_1+\rho_2)$, we have

$$
\mathrm{E}\Big(\max_{1\le j\le m_2}\Big(\sum_{i=1}^{m_1}u_{ij}\Big)^2\Big)\le\sum_{j=1}^{m_2}\mathrm{E}\Big(\Big(\sum_{i=1}^{m_1}u_{ij}\Big)^2\Big)=m_1m_2(\rho_1+\rho_2)
$$

which implies that

$$
\max_{1\le j\le m_2}\Big|\sum_{i=1}^{m_1}u_{ij}\Big|=O_p(\sqrt{m_2m_1(\rho_1+\rho_2)}).
$$

Similarly we have

$$
\max_{1\le 1\le m_1}\Big|\sum_{j=1}^{m_2}u_{ij}\Big|=O_p(\sqrt{m_2m_1(\rho_1+\rho_2)}).
$$

Hence by Lemma 3 and the expression for $Z_{k,1}$, we have

$$
\begin{aligned}
\max_{1\le k\le m}|\frac{Z_{k,1}}{\sqrt{\rho}}|&\le\frac{1}{(m_1-1)m_1}\Big|\sum_{i=1}^{m_1}\sum_{j=1}^{m_2}\frac{u_{ij}}{\sqrt{\rho}}\Big|+\max_{1\le k\le m_1}\Big|\frac{m-1}{(m_1-1)m_2}\sum_{j=1}^{m_2}\frac{u_{kj}}{\sqrt{\rho}}\Big|\\
&\quad+\frac{1}{(m_2-1)m_2}\Big|\sum_{i=1}^{m_1}\sum_{j=1}^{m_2}\frac{u_{ij}}{\sqrt{\rho}}\Big|+\max_{1\le k\le m_2}\Big|\frac{m-1}{(m_2-1)m_1}\sum_{j=1}^{m_1}\frac{u_{jk}}{\sqrt{\rho}}\Big|\\
&=o_p(1)+O_p\Big(\frac{m-1}{(m_1-1)m_2\sqrt{\rho}}\{m_1m_2(\rho_1+\rho_2)\}^{1/2}\Big)\\
&\quad+o_p(1)+O_p\Big(\frac{m-1}{(m_2-1)m_1\sqrt{\rho}}\{m_1m_2(\rho_1+\rho_2)\}^{1/2}\Big)\\
&=o_p(1)+O_p\Big(\frac{m^{1/2}}{(\min(m_1,m_2))^{1/2}}\Big)\\
&=o_p(m^{1/2}).
\end{aligned}
$$

Similarly we can show that

$$
\max_{1\le k\le m}\Big|\frac{Z_{k,2}}{\sqrt{\tau}}\Big|=o_p(m^{1/2}).
$$

Proof of Theorem 1. This follows from Lemma 4 and standard arguments in the empirical likelihood method (see Owen (1990)).

To show Corollary 1 and Theorem 2, we first prove two lemmas.

Lemma 5. $\mathbf{tr}(\Sigma^4) = O\big((\mathbf{tr}(\Sigma^2))^2\big)$, $\rho_1 = \sum_{j=1}^d \lambda_j^2$, *and* $2||\alpha||^2\lambda_1 \le \tau_1 \le 2||\alpha||^2\lambda_d$.

Proof. Since $\mathbf{tr}(\Sigma^j) = \sum_{i=1}^d \lambda_i^j$ for any positive integer j, the first equality follows immediately. The second equality follows since $\rho_1 = \mathbf{tr}(\Sigma^2)$. The third pair of inequalities on τ_1 come from the definition of τ_1.

Lemma 6. *For any* $\delta > 0$

$$\mathrm{E}\,|X_1^T\bar{X}_1|^{2+\delta} \le d^\delta\Big(\sum_{i=1}^d \mathrm{E}\,|X_{1,i}|^{2+\delta}\Big)^2,$$

$$\mathrm{E}\,|\alpha^T(X_1+\bar{X}_1)|^{2+\delta} \le 2^{4+\delta}||\alpha||^{2+\delta}d^{\delta/2}\sum_{i=1}^d \mathrm{E}\,|X_{1,i}|^{2+\delta}.$$

Proof. It follows from the Cauchy-Schwarz inequality that $|X_1^T\bar{X}_1|^2 \le ||X_1||^2||\bar{X}_1||^2$. Then by using the C_r inequality we conclude that

$$\begin{aligned}\mathrm{E}\,|X_1^T\bar{X}_1|^{2+\delta} &\le \mathrm{E}\Big(\sum_{i=1}^d X_{1,i}^2\Big)^{(2+\delta)/2}\mathrm{E}\Big(\sum_{i=1}^d \bar{X}_{1,i}^2\Big)^{(2+\delta)/2} \\ &= \Big(\mathrm{E}\Big(\sum_{i=1}^d X_{1,i}^2\Big)^{(2+\delta)/2}\Big)^2 \le \Big(d^{\delta/2}\sum_{i=1}^d \mathrm{E}\,|X_{1,i}|^{2+\delta}\Big)^2 \\ &= d^\delta\Big(\sum_{i=1}^d \mathrm{E}\,|X_{1,i}|^{2+\delta}\Big)^2.\end{aligned}$$

Similarly, from the C_r inequality we have

$$\begin{aligned}\mathrm{E}\,|\alpha^T(X_1+\bar{X}_1)|^{2+\delta} &\le 2^{4+\delta}\mathrm{E}\,|\alpha^TX_1|^{2+\delta} \le 2^{4+\delta}||\alpha||^{2+\delta}\mathrm{E}\,(||X_1||^{2+\delta}) \\ &= 2^{4+\delta}||\alpha||^{2+\delta}\mathrm{E}\Big|\sum_{i=1}^d |X_{1,i}|^2\Big|^{1+\delta/2} \\ &\le 2^{4+\delta}||\alpha||^{2+\delta}d^{\delta/2}\sum_{i=1}^d \mathrm{E}\,|X_{1,i}|^{2+\delta}.\end{aligned}$$

This completes the proof.

Proof of Corollary 1. Equations (2.5) and (2.7) follow from **A1**$-$**A3** by using Lemmas 5 and 6, as do (2.6) and (2.8), since we have the same assumptions on $\{X_i\}$ and $\{Y_j\}$.

Proof of Theorem 2. If suffices to verify conditions (2.5) and (2.7) with $\delta = 2$ in Theorem 1. Recall that $\mu_1 = \mu_2 = 0$, and note that $\mathrm{Var}\,(X_1) = \Sigma = \Gamma_1\Gamma_1^T$. With $\alpha^T\Gamma_1 = (a_1, \ldots, a_k)$ and $\Sigma' = \Gamma_1^T\Gamma_1 = (\sigma'_{j,l})_{1\le j,l\le k}$,

$$X_1^T\bar{X}_1 = \sum_{j=1}^k\sum_{l=1}^k \sigma'_{j,l}B_{1,j}B_{1+m_1,l},$$
$$\alpha^T(X_1+\bar{X}_1) = \sum_{j=1}^k a_j(B_{1,j}+B_{1+m_1,j}).$$

Set $\delta_{j_1,j_2,j_3,j_4} = \mathrm{E}\,(B_{1,j_1}B_{1,j_2}B_{1,j_3}B_{1,j_4})$. Then δ_{j_1,j_2,j_3,j_4} is $3+\xi_1$ if $j_1 = j_2 = j_3 = j_4$, is 1 if j_1, j_2, j_3 and j_4 form two different pairs of integers, and is zero otherwise. By Lemma 5, we have

$$\begin{aligned}
\mathrm{E}\,(X_1^T\bar{X}_1)^4 &= \sum_{j_1,j_2,j_3,j_4=1}^k\ \sum_{l_1,l_2,l_3,l_4=1}^k \sigma'_{j_1,l_1}\sigma'_{j_2,l_2}\sigma'_{j_3,l_3}\sigma'_{j_4,l_4}\delta_{j_1,j_2,j_3,j_4}\delta_{l_1,l_2,l_3,l_4}\\
&= O\Big(\big|\sum_{j_1\ne j_2}\sum_{l_1\ne l_2}\sigma'_{j_1,l_1}\sigma'_{j_1,l_2}\sigma'_{j_2,l_1}\sigma'_{j_2,l_2}\big|\Big) + O\Big(\sum_{j_1\ne j_2}\sum_{l=1}^k\sigma'^2_{j_1,l}\sigma'^2_{j_2,l}\Big)\\
&\quad + O\Big(\sum_{j=1}^k\sum_{l_1\ne l_2}\sigma'^2_{j,l_1}\sigma'^2_{j,l_2}\Big) + O\Big(\sum_{j=1}^k\sum_{l=1}^k\sigma'^4_{j,l}\Big)\\
&= O\Big(\big|\sum_{j_1=1}^k\sum_{j_2=1}^k\sum_{l_1=1}^k\sum_{l_2=1}^k\sigma'_{j_1,l_1}\sigma'_{j_1,l_2}\sigma'_{j_2,l_1}\sigma'_{j_2,l_2}\big|\Big)\\
&\quad + O\Big(\sum_{j_1=1}^k\sum_{j_2=1}^k\sum_{l=1}^k\sigma'^2_{j_1,l}\sigma'^2_{j_2,l}\Big) + O\Big(\sum_{j=1}^k\sum_{l=1}^k\sigma'^4_{j,l}\Big)\\
&= O\Big(\mathbf{tr}(\Sigma'^4)\Big) + O\Big(\sum_{j=1}^k\sum_{l=1}^k\sigma'^2_{j,l})^2\Big)\\
&= O\Big(\mathbf{tr}(\Sigma'^4)\Big) + O\Big((\mathbf{tr}(\Sigma'^2))^2\Big)\\
&= O\Big(\mathbf{tr}(\Sigma^4)\Big) + O\Big((\mathbf{tr}(\Sigma^2))^2\Big)\\
&= O(\rho_1^2),
\end{aligned}$$

so (2.5) holds with $\delta = 2$.

Similarly we have

$$\mathrm{E}\,(\alpha^T(X_1+\bar{X}_1))^4 \le 2^4\mathrm{E}\,\Big(\sum_{j=1}^k a_jB_{1,j}\Big)^4 = O\Big(\sum_{j_1,j_2=1}^k a_{j_1}^2a_{j_2}^2\Big) + O\Big(\sum_{j=1}^k a_j^4\Big)$$

$$= O\left(\left(\sum_{j=1}^{k} a_j^2\right)^2\right)$$
$$= O\left(\left(\alpha^T \Gamma_1 \Gamma_1^T \alpha\right)^2\right)$$
$$= O\left(\tau_1^2\right),$$

which yields (2.7) with $\delta = 2$. Equations (2.6) and (2.8) can be shown in the same way. Hence Theorem 2 follows from Theorem 1.

Acknowledgement

We thank an associate editor and two reviewers for helpful comments. Wang's research was supported by the Bob Price Fellowship in the School of Mathematics at Georgia Tech. Peng's research was supported by NSF Grant DMS-1005336, and Qi's research was supported by NSF Grant DMS-1005345.

References

Alon, U., Barkai, N., Notterman, D. A., Gish, K., Ybarra, S. Mack, D. and Levin A. J. (1999). Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 6745-6750.

Arlot, S., Blanchard, G. and Roquain, E. (2010a). Some nonasymptotic results on resampling in high dimension I: confidence regions. *Ann. Statist.* **38**, 51-82.

Arlot, S., Blanchard, G. and Roquain, E. (2010b). Some nonasymptotic results on resampling in high dimension II: multiple tests. *Ann. Statist.* **38**, 83-99.

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

Chen, S., Peng, L. and Qin, Y. (2009). Empirical likelihood methods for high dimension. *Biometrika* **96**, 711-722.

Chen, S. and Qin, Y. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *Ann. Statist.* **38**, 808-835.

Chen, S. and Van Keilegom, I. (2009). A review on empirical likelihood methods for regression. *Test* **18**, 415-447.

Dharmadhikari, S. W. and Jogdeo, K. (1969). Bounds on moments of certain random variables. *Ann. Math. Statist.* **40**, 1506-1508.

Jing, B. Y., Yuan, J. Q. and Zhou, W. (2009). Jackknife empirical likelihood. *J. Amer. Statist. Assoc.* **104**, 1224-1232.

Lee, M. L. (2004). *Analysis of Microarray Gene Expression Data.* Kluwer, Boston.

Kuelbs, J. and Vidyashankar, A. N. (2010). Asymptotic inference for high dimensional data. *Ann. Statist.* **38**, 836-869.

Owen, A. (1990). Empirical likelihood ratio confidence regions. *Ann. Statist.* **18**, 90-120.

Owen, A. (2001). *Empirical Likelihood.* Chapman and Hall/CRC, New York.

Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.* **22**, 300-325.

Srivastava, M. S. (2009). A test for the mean vector with fewer observations than the dimension under non-normality. *J. Multi. Anal.* **100**, 518-532.

Srivastava, M. S. and Du, M. (2008). A test for the mean vector with fewer observations than the dimension. *J. Multi. Anal.* **99**, 386-402.

von Bahr, B. and Esseen, C. G. (1965). Inequality for the rth absolute moment of a sum of random variables, $1 \leq r \leq 2$. *Ann. Math. Statist.* **36**, 299-393.

Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, ON N2L 3G1, Canada.

E-mail: ruodu.wang@uwaterloo.ca

School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332-0160, USA.

E-mail: peng@math.gatech.edu

Department of Mathematics and Statistics, University of Minnesota–Duluth, 1117 University Drive, Duluth, MN 55812, USA.

E-mail: yqi@d.umn.edu

(Received October 2011; accepted July 2012)