Statistical Science
2014, Vol. 29, No. 1, 113–127
DOI: 10.1214/13-STS425

Test for a Mean Vector with Fixed or Divergent Dimension

Liang Peng, Yongcheng Qi and Fang Wang

Abstract. It has been a long history in testing whether a mean vector with a fixed dimension has a specified value. Some well-known tests include the Hotelling T^2-test and the empirical likelihood ratio test proposed by Owen [*Biometrika* **75** (1988) 237–249; *Ann. Statist.* **18** (1990) 90–120]. Recently, Hotelling T^2-test has been modified to work for a high-dimensional mean, and the empirical likelihood method for a mean has been shown to be valid when the dimension of the mean vector goes to infinity. However, the asymptotic distributions of these tests depend on whether the dimension of the mean vector is fixed or goes to infinity. In this paper, we propose to split the sample into two parts and then to apply the empirical likelihood method to two equations instead of d equations, where d is the dimension of the underlying random vector. The asymptotic distribution of the new test is independent of the dimension of the mean vector. A simulation study shows that the new test has a very stable size with respect to the dimension of the mean vector, and is much more powerful than the modified Hotelling T^2-test.

Key words and phrases: Empirical likelihood, high-dimensional mean, test.

1. INTRODUCTION

Suppose $X_1 = (X_{1,1}, \ldots, X_{1,d})^T, \ldots, X_n = (X_{n,1}, \ldots, X_{n,d})^T$ are independent random vectors having common distribution function F with mean μ and covariance matrix Σ. It has been a long history to test $H_0 : \mu = \mu_0$ against $H_a : \mu \neq \mu_0$ for a given μ_0. When the dimension d is fixed, a traditional test is the so-called Hotelling T^2-test defined as

$$T^2 = (\bar{X}_n - \mu_0)^T \cdot \left\{ \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)(X_i - \bar{X}_n)^T \right\}^{-1} \cdot (\bar{X}_n - \mu_0),$$

where $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$. Another commonly used one is the empirical likelihood ratio test proposed by Owen (1988, 1990). More specifically, by defining the empirical likelihood function as

$$L(\mu) = \sup \left\{ \prod_{i=1}^n (np_i) : p_1 \geq 0, \ldots, p_n \geq 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i X_i = \mu \right\}, \tag{1}$$

Owen (1988, 1990) showed that the Wilks theorem holds under some regularity conditions, that is, $-2 \log L(\mu_0)$ converges in distribution to a chi-square limit with d degrees of freedom, where μ_0 denotes the true value of the mean of X_i. Therefore, based on the chi-square limit, one can construct a confidence region for μ or test $H_0 : \mu = \mu_0$ against $H_a : \mu \neq \mu_0$.

Without assuming a family of distributions for the data, the empirical likelihood ratio statistics can be defined to share similar properties as the likelihood ratio for parametric distributions. For instance, the empirical likelihood method produces confidence regions whose shape and orientation are determined entirely by

Liang Peng is Professor, School of Mathematics, Georgia Institute of Technology, Atlanta, Georgia 30332-0160, USA (e-mail: peng@math.gatech.edu). Yongcheng Qi is Professor, Department of Mathematics and Statistics, University of Minnesota Duluth, 1117 University Drive, Duluth, Minnesota 55812, USA (e-mail: yqi@d.umn.edu). Fang Wang is Associate Professor, School of Mathematical Sciences, Capital Normal University, Beijing 100048, PR China (e-mail: fang72_wang@126.com).

the data. In comparison with the normal approximation method and the bootstrap method for constructing confidence regions, the empirical likelihood method does not require a pivotal quantity, and it has better small sample performance (see Hall and La Scala, 1990). For more details on empirical likelihood methods, we refer to Owen (2001) and the recent review paper of Chen and Van Keilegom (2009).

Motivated by applications in neuroimaging and bioinformatics studies, some tests for a mean vector with divergent dimension have been proposed in the literature. It is known that, as the dimension is large, the calculation of the inverse of the sample covariance matrix in Hotelling T^2-test statistic becomes problematic and the sample covariance matrix may diverge when $d/n \to c > 0$; see Yin, Bai and Krishnaiah (1988). Moreover, Hotelling T^2-test is valid only when $d < n$. In order to allow $d > n$, one may remove the sample matrix in Hotelling's T^2-test statistic and avoid the singularity of the sample covariance. This is exactly what has been done in Bai and Saranadasa (1996) and Chen and Qin (2010) for the two-sample test problem. The one-sample analogues of the two-sample test statistics in Bai and Saranadasa (1996) and Chen and Qin (2010) lead to the following test statistics:

$$M_n = (\bar{X}_n - \mu_0)^T(\bar{X}_n - \mu_0) - n^{-1}\operatorname{tr}\left(\frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)(X_i - \bar{X}_n)^T\right)$$

and

$$F_n = n^{-1}(n-1)^{-1}\sum_{i \neq j}^n (X_i - \mu_0)^T(X_j - \mu_0),$$

respectively, where tr means the trace of a matrix. It is easy to check that

$$\begin{aligned}
M_n &= \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n (X_i - \mu_0)^T(X_j - \mu_0) \\
&\quad - \frac{1}{n(n-1)} \cdot \operatorname{tr}\left\{\sum_{i=1}^n (X_i - \mu_0)(X_i - \mu_0)^T - n^{-1}\sum_{i=1}^n\sum_{j=1}^n (X_i - \mu_0)(X_j - \mu_0)^T\right\} \\
&= \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n (X_i - \mu_0)^T(X_j - \mu_0) \\
&\quad - \frac{1}{n(n-1)} \cdot \left\{\sum_{i=1}^n (X_i - \mu_0)^T(X_i - \mu_0) - n^{-1}\sum_{i=1}^n\sum_{j=1}^n (X_i - \mu_0)^T(X_j - \mu_0)\right\} \\
&= \frac{1}{n(n-1)}\sum_{i=1}^n\sum_{j=1}^n (X_i - \mu_0)^T(X_j - \mu_0) - \frac{1}{n(n-1)}\sum_{i=1}^n (X_i - \mu_0)^T(X_i - \mu_0) \\
&= F_n.
\end{aligned}$$

That is, the test in Bai and Saranadasa (1996) is the same as that in Chen and Qin (2010) when the one-dimensional data is concerned. As mentioned in the end of Section 3 of Chen and Qin (2010), the asymptotic behavior of F_n depends on whether d is fixed or goes to infinity. Alternatively, Srivastava and Du (2008) and Srivastava (2009) proposed to replace the covariance matrix in Hotelling T^2-test statistic by a diagonal matrix. Rates of convergence for the high-dimensional mean are studied by Kuelbs and Vidyashankar (2010). For nonasymptotic studies, we refer to Arlot, Blanchard and Roquain (2010a, 2010b).

Although it is known that the empirical likelihood method performs worse when the dimension d is large and the sample size n is not large enough, Hjort, McKeague and Van Keilegom (2009) and Chen, Peng and Qin (2009) showed that the empirical likelihood method for a fixed-dimensional mean is still valid when $d = d(n) \to \infty$ as $n \to \infty$. More specifically, they showed under some regularity conditions that $(2d)^{-1/2}\{-2\log L(\mu_0) - d\}$ converges in distribution to a standard normal if $d \to \infty$ as $n \to \infty$. That is, the limiting distribution differs when the dimension of the mean vector is fixed or diverges.

Now, the question is whether there exists a way to test $H_0 : \mu = \mu_0$ against $H_a : \mu \neq \mu_0$ without distinguishing the dimension of μ is finite or goes to infinity. Motivated by the tests in Bai and Saranadasa (1996) and Chen and Qin (2010), we propose to apply the empirical likelihood method to the equation $E\{(X_1 - \mu_0)^T(X_2 - \mu_0)\} = 0$ instead of $EX_1 = \mu_0$ for testing $H_0 : \mu = \mu_0$ against $H_a : \mu \neq \mu_0$, where X_1 and X_2 are independent and identically distributed random vectors with mean μ. Although the equation $E\{(X_1 - \mu_0)^T(X_2 - \mu_0)\} = 0$ is equivalent to $H_0 : \mu = \mu_0$, a test only based on the equation $E\{(X_1-\mu_0)^T(X_2-\mu_0)\}=0$ has a poorer power than a test based on $EX_1=\mu_0$. The reason is that $E\{(X_1-\mu_0)^T(X_2-\mu_0)\}/d=\delta^2/n$ instead of the standard order $1/\sqrt{n}$ when $EX_1=\mu_0+\delta\mathbf{1}_d/\sqrt{n}$, where $\mathbf{1}_d=(1,\dots,1)^T$ is a d-dimensional vector. To overcome this issue so as to improve the test power, we propose to add one more linear equation. More specifically, we propose to consider the following two equations:

$$E\{(X_1-\mu_0)^T(X_2-\mu_0)\}=0$$

and

$$E\{\mathbf{1}_d^T(X_1+X_2-2\mu_0)\}=0.$$

It is easy to see that $E\{\mathbf{1}_d^T(X_1+X_2-2\mu_0)\}/d=O(1/\sqrt{n})$ rather than $O(1/n)$ when $EX_1=\mu_0+\delta\mathbf{1}_d/\sqrt{n}$. The first equation ensures the consistency of the proposed test, and the second equation enhances the power in detecting a deviation. It turns out that the empirical likelihood method based on the above two equations works for either fixed d or divergent d. This differs from the results in Bai and Saranadasa (1996) and Chen and Qin (2010). More interestingly, the new method allows one to easily include more independent equations if such equations characterize the departure from the null hypothesis and are available. On the other hand, when the number of equations becomes large, the minimization in the empirical likelihood method turns out to be nontrivial.

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

2. METHODOLOGY

Assume $X_1=(X_{1,1},\dots,X_{1,d})^T,\dots,X_n=(X_{n,1},\dots,X_{n,d})^T$ are independent and identically distributed random vectors having common distribution function F with mean μ and covariance matrix Σ. For testing $H_0:\mu=\mu_0$ against $H_a:\mu\neq\mu_0$ for a given μ_0, we propose to apply the empirical likelihood method to the equations $E\{(X_1-\mu_0)^T(X_2-\mu_0)\}=0$ and $E\{\mathbf{1}_d^T(X_1+X_2-2\mu_0)\}=0$. In order to have two independent samples, we simply split the first $m=[n/2]$ observations into a subsample and the second m observations into another subsample, and put $Y_i(\mu)=(u_i(\mu),v_i(\mu))^T$, where

$$u_i(\mu)=(X_i-\mu)^T(X_{i+m}-\mu),$$
$$v_i(\mu)=\mathbf{1}_d^T(X_i+X_{i+m}-2\mu)\quad\text{for } i=1,\dots,m.$$

Hence, $\{Y_i(\mu), 1\le i\le m\}$ are i.i.d. bivariate random vectors. Define the empirical likelihood function as

$$\tilde{L}(\mu)=\sup\left\{\prod_{i=1}^m(mp_i):p_1\ge0,\dots,p_m\ge0,\ \sum_{i=1}^m p_i=1,\sum_{i=1}^m p_iY_i(\mu)=0\right\}. \tag{2}$$

By the Lagrange multiplier technique, we have $p_i=m^{-1}\{1+\beta^TY_i(\mu)\}^{-1}$ for $i=1,\dots,m$ and $\tilde{l}(\mu)=-2\log\tilde{L}(\mu)=2\sum_{i=1}^m\log\{1+\beta^TY_i(\mu)\}$, where $\beta=\beta(\mu)=(\beta_1(\mu),\beta_2(\mu))^T$ satisfies

$$\frac{1}{m}\sum_{i=1}^m\frac{Y_i(\mu)}{1+\beta^TY_i(\mu)}=0. \tag{3}$$

Write $\Sigma=(\sigma_{i,j})_{1\le i\le d,1\le j\le d}=E\{(X_1-\mu)(X_1-\mu)^T\}$, the covariance matrix of X_1, and use $\lambda_1\le\cdots\le\lambda_d$ to denote the d eigenvalues of the matrix Σ. Note that λ_i's may depend on n when d depends on n.

First we show the Wilks theorem under very general conditions.

THEOREM 1. *Assume $\sum_{i=1}^d\sum_{j=1}^d\sigma_{i,j}>0$ and for some $\delta>0$*

$$\frac{E|u_1(\mu)|^{2+\delta}}{(\sum_{i=1}^d\sum_{j=1}^d\sigma_{i,j}^2)^{(2+\delta)/2}}=o\big(n^{(\delta+\min(\delta,2))/4}\big) \tag{4}$$

and

$$\frac{E|v_1(\mu)|^{2+\delta}}{(\sum_{i=1}^d\sum_{j=1}^d\sigma_{i,j})^{(2+\delta)/2}}=o\big(n^{(\delta+\min(\delta,2))/4}\big). \tag{5}$$

Then under $H_0:\mu=\mu_0$, $\tilde{l}(\mu_0)$ converges in distribution to a chi-square limit with two degrees of freedom as $n\to\infty$.

Based on the above theorem, one can test $H_0:\mu=\mu_0$ against $H_a:\mu\neq\mu_0$. A test with level α is to reject H_0 when $\tilde{l}(\mu_0)>\xi_{1-\alpha}$, where $\xi_{1-\alpha}$ is the $(1-\alpha)$th quantile of a chi-square limit with two degrees of freedom.

Note that the proposed method works as well if one is interested in testing the difference of two mean vectors based on paired data. However, it is not applicable to the two-sample case with different sample sizes.

Next we verify Theorem 1 by imposing conditions on the moments and dimension of the random vector:

(A1): $0<C_1\le\liminf_{n\to\infty}\lambda_1\le\limsup_{n\to\infty}\lambda_d\le C_2<\infty$ for some constants C_1 and C_2;

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

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

COROLLARY 1. *Assume conditions* (A1)–(A3) *hold*. *Then conditions* (4) *and* (5) *are satisfied and*, *thus*, *Theorem* 1 *holds*.

Condition (A3) is a somewhat restrictive condition for the dimension d. Note that conditions (A1) and (A2) are related only to the covariance matrix and higher moments on the components of the random vectors. Condition (A3) can be removed for models with some special dependence structures. For comparisons, we prove the Wilks theorem for the proposed empirical likelihood method under the following model B considered by Bai and Saranadasa (1996), Chen, Peng and Qin (2009) and Chen and Qin (2010):

Model B. $X_i = \Gamma Z_i + \mu$ for $i = 1, \dots, n$, where Γ is a $d \times k$ matrix with $\Gamma\Gamma^T = \Sigma = (\sigma_{i,j})$ and $Z_i = (Z_{i,1}, \dots, Z_{i,k})^T$ are i.i.d. random k-vectors with $EZ_i = 0$, $\mathrm{Var}(Z_i) = I_{k\times k}$, $EZ_{i,j}^4 = 3 + \Delta < \infty$ and $E\prod_{l=1}^k Z_{i,l}^{\nu_l} = \prod_{l=1}^k EZ_{i,l}^{\nu_l}$ whenever $\nu_1 + \cdots + \nu_k = 4$ for nonnegative integers ν_l's.

THEOREM 2. *Assume* $\sum_{i=1}^d \sum_{j=1}^d \sigma_{i,j} > 0$. *Then under model B and* $H_0: \mu = \mu_0$, $\tilde{l}(\mu_0)$ *converges in distribution to a chi-square limit with two degrees of freedom as* $n \to \infty$.

THEOREM 3. *Assume* $\sum_{i=1}^d \sum_{j=1}^d \sigma_{i,j} > 0$ *and put*

$$\tau = \frac{m\|\mu_0 - \mu\|^4}{\sum_{i=1}^d \sum_{j=1}^d \sigma_{i,j}^2} + \frac{2m(\mathbf{1}_d^T(\mu_0 - \mu))^2}{\sum_{i=1}^d \sum_{j=1}^d \sigma_{i,j}}.$$

Then under model B and $H_a: \mu \neq \mu_0$, *we have*

$$P(\tilde{l}(\mu_0) > \xi_{1-\alpha}) = P(\chi_{2,\tau}^2 > \xi_{1-\alpha}) + o(1) \tag{6}$$

as $n \to \infty$, *where* $\xi_{1-\alpha}$ *denotes the* $(1-\alpha)$*th quantile of a chi-square limit with two degrees of freedom, and* $\chi_{2,\tau}^2$ *denotes a noncentral chi-square random variable with two degrees of freedom and noncentrality parameter* τ.

REMARK 1. It can be seen from the proof of Theorem 2 that assumption $EZ_{i,j}^4 = 3 + \Delta < \infty$ in model B can be replaced by the much weaker condition $\max_{1\le j\le k} EZ_{1,j}^4 = o(m)$.

REMARK 2. Unlike Bai and Saranadasa (1996) and Chen and Qin (2010), there is no restriction on d and k for our proposed method in Theorem 2. The only constraint imposed on matrix Γ is also very weak, that is, $\sum_{i=1}^d \sum_{j=1}^d \sigma_{i,j} > 0$ or, equivalently, $\sum_{i=1}^d X_{1,i}$ is a nondegenerate random variable.

REMARK 3. We notice that conditions (4) and (5) in Theorem 1 impose some restriction on d implicitly. Whether such a restriction can be relaxed or removed depends on how sharp the moments in the right-hand sides of (4) and (5) can be estimated. In Corollary 1, since we do not assume any dependence structure among the components of X_1, the best order of d allowed in (A3) is less than $n^{1/2}$ even for bounded X_1. On the other hand, since model B assumes that the components of X_1 are linear combinations of some orthogonal random variables, conditions (4) and (5) become trivial and, consequently, the restriction on d is removed in Theorem 2.

REMARK 4. When the test in Bai and Saranadasa (1996) is applied to model B for one sample, its power is

$$\Phi\left(-\xi_{1-\alpha}^* + \frac{n\|\mu_0 - \mu\|^2}{\sqrt{2\sum_{i,j=1}^d \sigma_{i,j}^2}}\right) + o(1), \tag{7}$$

where $\Phi(x)$ denotes the standard normal distribution function and $\xi_{1-\alpha}^*$ denotes its $(1-\alpha)$th quantile; see Theorem 4.1 of Bai and Saranadasa (1996). Under model B, assumption (A1), $d \to \infty$, and $\mu = \mu_0 + a_n\bar{\mu}$ for $a_n \neq 0$ and $\|\bar{\mu}\| = 1$, it follows from Lemma 1 in Section 4 that τ in Theorem 3 has the order $\Delta_1 = \frac{na_n^4}{d} + \frac{n(\mathbf{1}_d^T\bar{\mu})^2 a_n^2}{d}$, and the order of $\frac{n\|\mu_0-\mu\|^2}{\sqrt{2\sum_{i,j=1}^d \sigma_{i,j}^2}}$ in (7) is $\Delta_2 = \frac{na_n^2}{\sqrt{d}}$. When both $\Delta_1 \to \infty$ and $\Delta_2 \to \infty$, the power of both tests goes to one. Due to the o(1) term in Theorem 3 and (7), one cannot claim which power goes to one faster in this case. When $0 < \liminf \Delta_2 \le \limsup \Delta_2 < \infty$ and $\frac{(\mathbf{1}_d^T\bar{\mu})^2}{\sqrt{d}} \to \infty$, the test in Bai and Saranadasa (1996) has a power bounded from one, but the proposed new test has a power tending to one, that is, the proposed empirical likelihood test is much more powerful than the test in Bai and Saranadasa (1996) in this situation. However, when $0 < \liminf \Delta_2 \le \limsup \Delta_2 < \infty$ and $\frac{(\mathbf{1}_d^T\bar{\mu})^2}{\sqrt{d}} \to 0$, the proposed empirical likelihood test is much less powerful. In this case, a different linear functional $c^T(X_i + X_{i+m} - \mu_0)$ has to be employed to replace $\mathbf{1}_d^T(X_i + X_{i+m} - \mu_0)$ so as to improve the test power, where c is a d-dimensional constant. When $\mathbf{1}_d^T(X_i + X_{i+m} - \mu_0)$ is replaced by any new functional $c^T(X_i + X_{i+m} - \mu_0)$ in the definition of the empirical likelihood $\tilde{L}(\mu)$ given in (2), similar results to Theorems 1, 2 and 3 can also be derived easily. Moreover, the above Δ_1 becomes $\frac{na_n^4}{d} + \frac{n(c^T\bar{\mu})^2 a_n^2}{d}$. Therefore, when $\frac{(\mathbf{1}_d^T\bar{\mu})^2}{\sqrt{d}} \to 0$, one can choose c such that $\liminf \frac{(c^T\bar{\mu})^2}{\sqrt{d}} > 0$ so as to improve the power. However, as discussed in the Introduction, it remains open on how to find such c or the optimal linear functionals.

3. SIMULATION STUDY

We investigate the finite sample behavior of the proposed empirical likelihood method (NELM) and compare it with the Hotelling's T^2-test (HT) and the test statistic M_n in Bai and Saranadasa (1996) (BS) in terms of size and power. A simulation reveals that the standard empirical likelihood method (OELM) in Owen (1990) has a size much larger than the nominal level when $d > 20$ and, thus, it makes no sense to compare these two empirical likelihood methods.

Let $W_1, \dots, W_d$ be independent and identically distributed random variables with distribution function either the standard normal [notation $N(0, 1)$] or t distribution with 6 degrees of freedom [notation $t(6)$]. Consider the following two models:

Model 1: $X_{1,1} = W_1 + \delta/\sqrt{n}$, $X_{1,2} = W_1 + W_2 + \delta/\sqrt{n}, \dots, X_{1,d} = W_{d-1} + W_d + \delta/\sqrt{n}$.

Model 2:

$$(X_{1,1}, \dots, X_{1,d})^T \sim N(\delta \mathbf{1}_d/\sqrt{n}, (0.5^{|i-j|})_{1\le i,j\le d}),$$

where $\delta \in R$ and n is the sample size. The question is to test $H_0 : \mu = 0$ against $H_a : \mu \neq 0$. Hence, the case of $\delta = 0$ denotes the size of tests. It is easy to check that these two models are a special case of model B in Section 2. For example, model 1 corresponds to model B with

$$\Gamma = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 1 & \cdots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ 0 & 0 & 0 & \cdots & 1 & 1 \end{pmatrix}$$

and $\sum_{i=1}^d \sum_{j=1}^d \sigma_{i,j} = 4d - 3$, $\sum_{i=1}^d \sum_{j=1}^d \sigma_{i,j}^2 = 6d - 5$. Hence, Theorem 2 holds for model 1 without restriction on the dimension d, and τ in Theorem 3 equals $\frac{md^2\delta^4}{(6d-5)n^2} + \frac{2md^2\delta^2}{(4d-3)n}$. Moreover, Δ_1 and Δ_2 defined in Remark 4 are $d\delta^2$ and $\sqrt{d}\delta^2$, respectively, as $d \to \infty$. Hence, theoretically the proposed empirical likelihood method is much more powerful when $\sqrt{d}\delta^2$ is bounded away from infinity. When δ is fixed, both tests have a power tending to one. In this case Theorem 3 and equation (7) in Remark 4 cannot be used to claim which test is more powerful theoretically, but the simulation results below show that the empirical likelihood method is still more powerful. Similarly, we can verify that Theorem 2 holds for model 2 without restriction on the dimension as well and $\tau = \frac{md}{n^2} \frac{\delta^4}{5/3-8(1-0.5^{2d})d^{-1}/9} + \frac{md}{n} \frac{2\delta^2}{3-4(1-0.5^d)d^{-1}}$ in Theorem 3. Using Remark 4, we conclude that the proposed empirical likelihood method for model 2 is more powerful than the test in Bai and Saranadasa (1996) when $\sqrt{d}\delta^2$ is bounded away from infinity and $d \to \infty$. When δ is fixed and $d \to \infty$, both tests have a power tending to one and theoretical comparison does not exist. However, the following simulation results show that the proposed empirical likelihood method is more powerful.

By drawing 10,000 random samples of sample size $n = 100$ and 300 from $X = (X_{1,1}, \dots, X_{1,d})^T$ with $d = 5, 10, 15, \dots, 200$ and $\delta = 0, 0.1, 0.5$, we calculate the empirical sizes and powers of those tests mentioned above.

In Figure 1 we plot the empirical sizes (i.e., $\delta = 0$) of these tests against $d = 5, 10, \dots, 200$ at a nominal level 0.05. Note that the Hotelling's T^2-test only works for $d < n$. As we see, the size of the proposed empirical likelihood method is slightly larger than the nominal level and less accurate than the other two tests when $n = 100$, but it becomes close to the nominal level and comparable to the other two tests when $n = 300$.

In Figures 2 and 3 the powers for $\delta = 0.1$ and 0.5 are plotted against $d = 5, 10, \dots, 200$ at level 0.05. These figures clearly show that the proposed empirical likelihood method is much more powerful than others especially when d becomes relatively large.

In conclusion, the proposed empirical likelihood method has a stable size with respect to the dimension and a large power, and performs well for all considered d.

4. PROOFS

In the proofs we use $\|\cdot\|$ to denote the L_2 norm of a vector or matrix. Without loss of generality, we assume $\mu_0 = 0$. Write $u_i = u_i(0)$ and $v_i = v_i(0)$ for $1 \le i \le m$. Then it is easily verified that

$$E(u_1) = E(v_1) = E(u_1 v_1) = 0,$$

$$\mathrm{Var}(u_1) = \sum_{i,j=1}^d \sigma_{i,j}^2 =: \pi_{11}$$

and

$$\mathrm{Var}(v_1) = 2 \sum_{i,j=1}^d \sigma_{i,j} =: \pi_{22}.$$

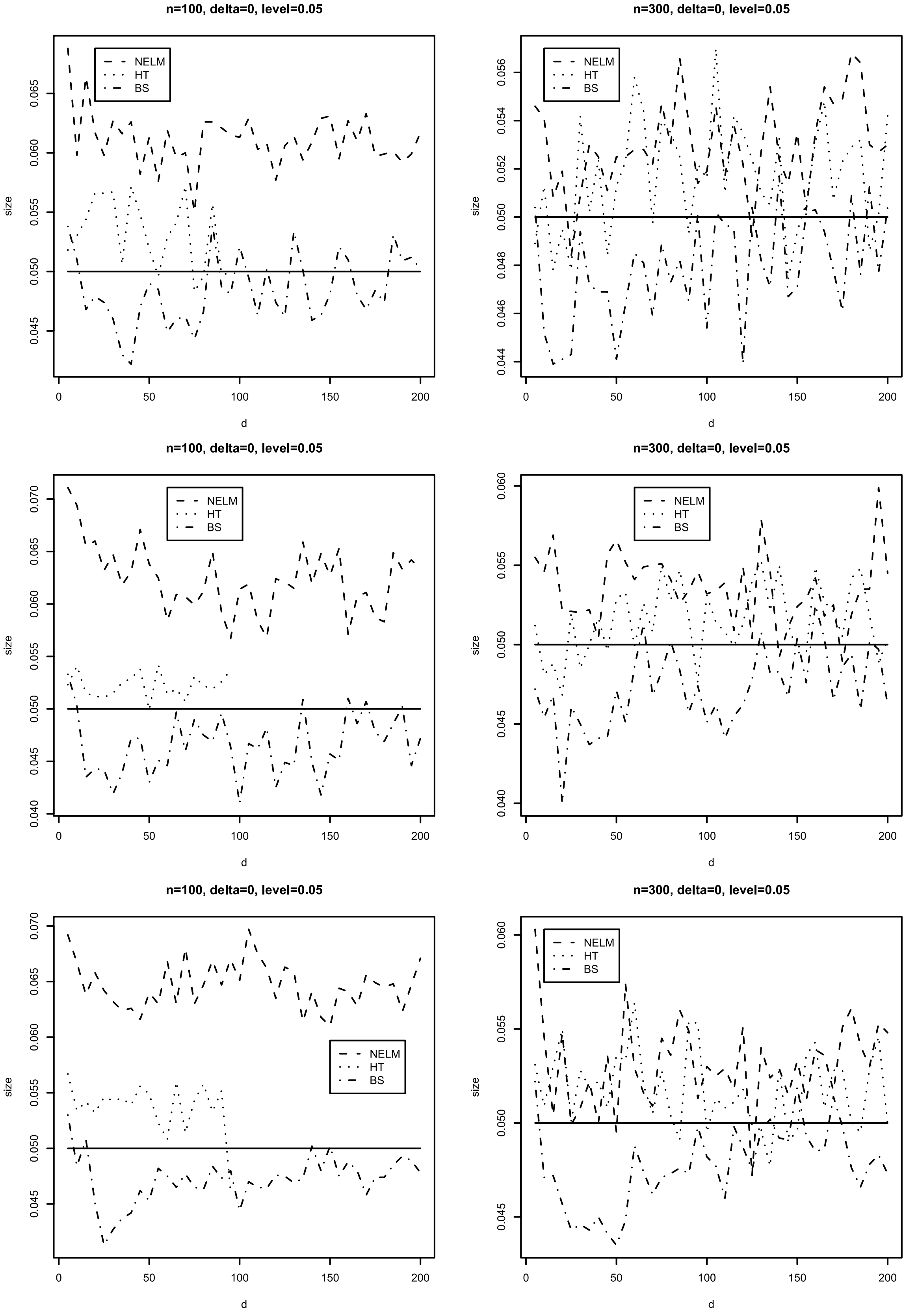

FIG. 1. *Sizes of tests are plotted against $d = 5, 10, \ldots, 200$ for $\delta = 0$ and level 0.05. The upper, middle and lower panels represent model 1 with $W_i \sim N(0, 1)$, model 1 with $W_i \in t_6$ and model 2, respectively. Solid line is the nominal level.*

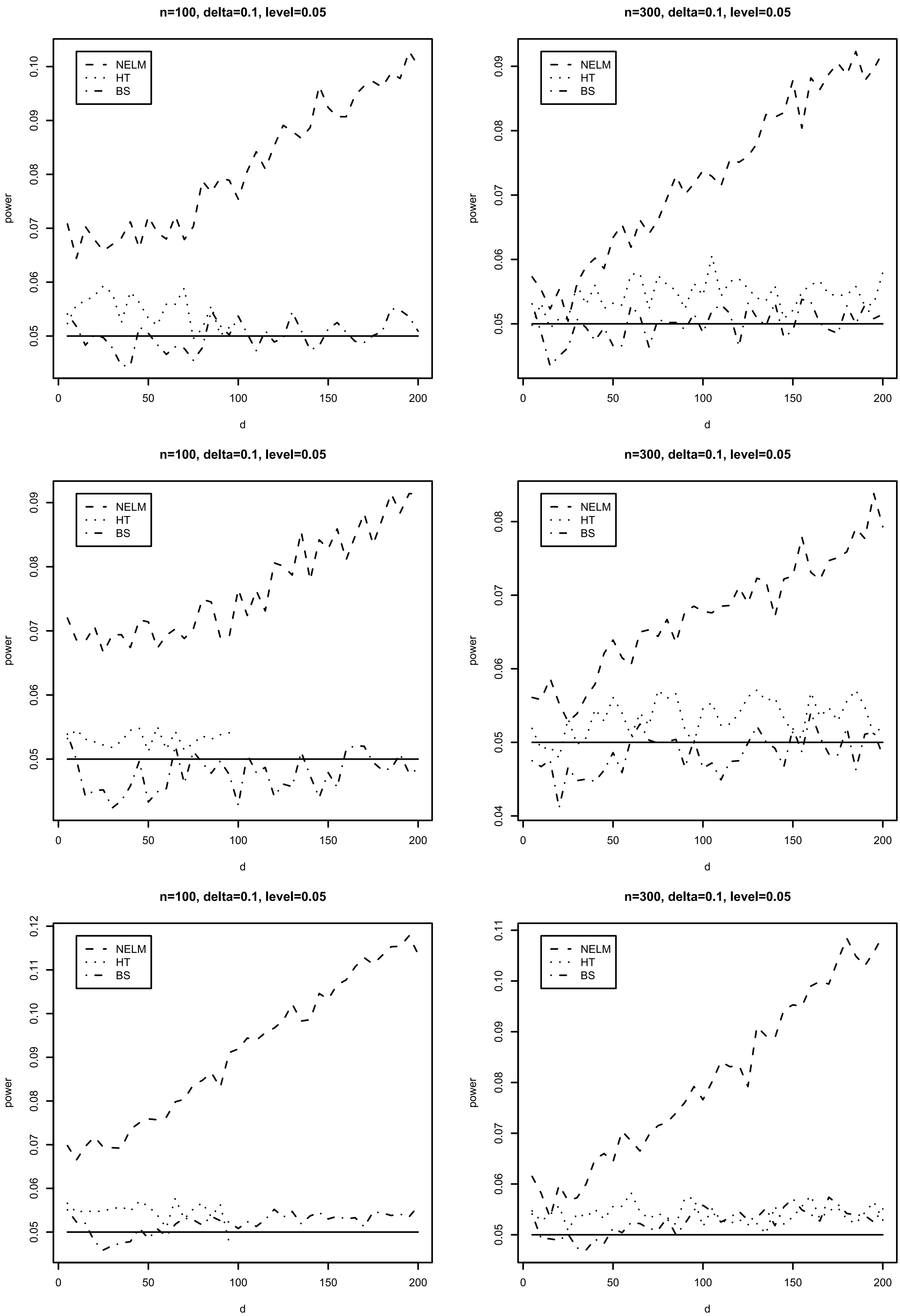

FIG. 2. *Powers of tests are plotted against* $d = 5, 10, \dots, 200$ *for* $\delta = 0.1$ *and level* 0.05. *The upper, middle and lower panels represent model* 1 *with* $W_i \sim N(0, 1)$, *model* 1 *with* $W_i \in t_6$ *and model* 2, *respectively. Solid line is the nominal level.*

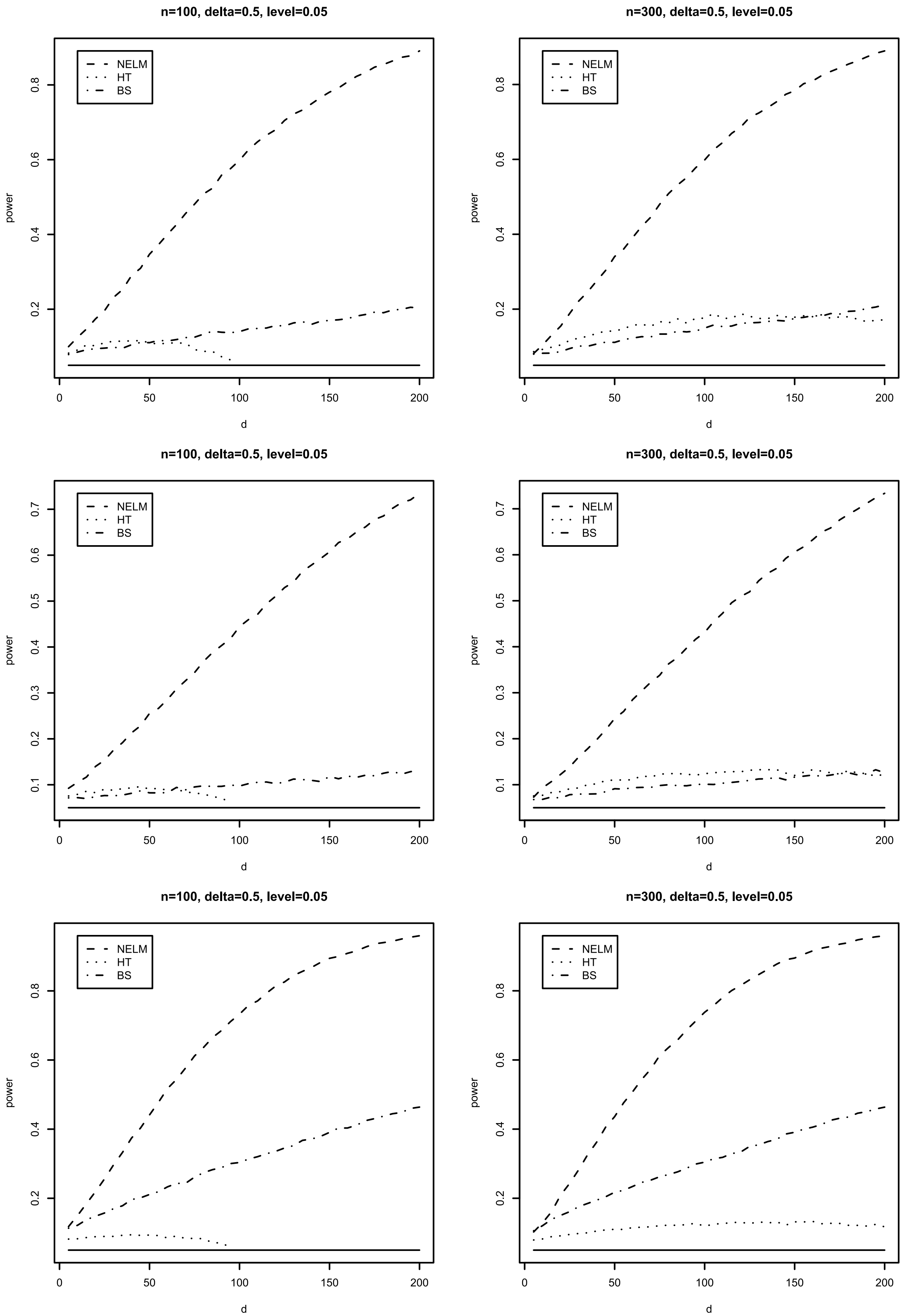

FIG. 3. *Powers of tests are plotted against* $d = 5, 10, \ldots, 200$ *for* $\delta = 0.5$ *and level* 0.05. *The upper, middle and lower panels represent model* 1 *with* $W_i \sim N(0, 1)$, *model* 1 *with* $W_i \in t_6$ *and model* 2, *respectively. Solid line is the nominal level.*

LEMMA 1.

$$\operatorname{tr}(\Sigma^4) = O\big((\mathbf{tr}(\Sigma^2))^2\big),$$

$$\pi_{11} = \sum_{j=1}^{d} \lambda_j^2$$

and

$$2d\lambda_1 \le \pi_{22} \le 2d\lambda_d.$$

PROOF. Since $\operatorname{tr}(\Sigma^j) = \sum_{i=1}^d \lambda_i^j$ for any positive integer j, the first equality follows immediately. The second equality follows since $\pi_{11} = \operatorname{tr}(\Sigma^2)$. The third inequalities on π_{22} can be proved easily. The proof of the lemma is complete. □

LEMMA 2. *Assume conditions* (4) *and* (5) *hold. Then*

$$\frac{1}{\sqrt{m}} \sum_{i=1}^{m} \begin{pmatrix} \frac{u_i}{\sqrt{\pi_{11}}} \\ \frac{v_i}{\sqrt{\pi_{22}}} \end{pmatrix} \xrightarrow{d} N(0, I_2), \tag{8}$$

$$\frac{\sum_{i=1}^m u_i^2}{m\pi_{11}} - 1 \xrightarrow{p} 0, \tag{9}$$

$$\frac{\sum_{i=1}^m v_i^2}{m\pi_{22}} - 1 \xrightarrow{p} 0, \tag{10}$$

$$\frac{\sum_{i=1}^m u_i v_i}{m\sqrt{\pi_{11}\pi_{22}}} \xrightarrow{p} 0, \tag{11}$$

where I_2 is a 2×2 identity matrix. Moreover, we have

$$\begin{aligned} &\max_{1\le i\le m}\left|\frac{u_i}{\sqrt{\pi_{11}}}\right| = o_p(m^{1/2}) \quad \textit{and} \\ &\max_{1\le i\le m}\left|\frac{v_i}{\sqrt{\pi_{22}}}\right| = o_p(m^{1/2}). \end{aligned} \tag{12}$$

PROOF. Note that u_1 and v_1 are uncorrelated. To show (8), we need to prove that for any constants a and b with $a^2 + b^2 \ne 0$,

$$\frac{1}{\sqrt{m}} \sum_{i=1}^{m} \left(a\frac{u_i}{\sqrt{\pi_{11}}} + b\frac{v_i}{\sqrt{\pi_{22}}}\right) \xrightarrow{d} N(0, a^2 + b^2).$$

Therefore, we shall verify the Lindeberg condition, which is a consequence of the Lyapunov condition as follows:

$$\begin{aligned} &\frac{1}{m^{(2+\delta)/2}} \sum_{i=1}^{m} E\left|a\frac{u_i}{\sqrt{\pi_{11}}} + b\frac{v_i}{\sqrt{\pi_{22}}}\right|^{2+\delta} \\ &\quad = \frac{1}{m^{\delta/2}} E\left|a\frac{u_1}{\sqrt{\pi_{11}}} + b\frac{v_1}{\sqrt{\pi_{22}}}\right|^{2+\delta} \\ &\quad \le \frac{(2|a|)^{2+\delta}}{m^{\delta/2}} E\left|\frac{u_1}{\sqrt{\pi_{11}}}\right|^{2+\delta} + \frac{(2|b|)^{2+\delta}}{m^{\delta/2}} E\left|\frac{v_1}{\sqrt{\pi_{22}}}\right|^{2+\delta} \\ &\quad \to 0 \end{aligned} \tag{13}$$

from conditions (4) and (5).

To show (9), we need to estimate $E|\sum_{i=1}^m u_i^2 - m\pi_{11}|^{(2+\delta)/2}$. We have from von Bahr and Esseen (1965) that

$$\begin{aligned} &E\left|\sum_{i=1}^{m} u_i^2 - m\pi_{11}\right|^{(2+\delta)/2} \\ &\quad \le 2mE|u_1^2 - E(u_1^2)|^{(2+\delta)/2} \\ &\quad = O(mE|u_1|^{2+\delta}), \end{aligned} \tag{14}$$

if $0 < \delta \le 2$, and from Dharmadhikari and Jogdeo (1969) that

$$\begin{aligned} &E\left|\sum_{i=1}^{m} u_i^2 - m\pi_{11}\right|^{(2+\delta)/2} \\ &\quad \le Cm^{(2+\delta)/4}E|u_1^2 - E(u_1^2)|^{(2+\delta)/2} \\ &\quad = O(m^{(2+\delta)/4}E|u_1|^{2+\delta}), \end{aligned} \tag{15}$$

if $\delta > 2$. Therefore, by (14), (15) and (4) we have for any $\varepsilon > 0$

$$\begin{aligned} &P\left(\left|\frac{\sum_{i=1}^m u_i^2}{m\pi_{11}} - 1\right| > \varepsilon\right) \\ &\quad \le \varepsilon^{-(2+\delta)/2} \frac{E|\sum_{i=1}^m u_i^2 - m\pi_{11}|^{(2+\delta)/2}}{(m\pi_{11})^{(2+\delta)/2}} \\ &\quad = O\left(m^{-(\delta+\min(\delta,2))/4} E\left|\frac{u_1}{\sqrt{\pi_{11}}}\right|^{2+\delta}\right) \\ &\quad = o(1), \end{aligned}$$

which implies (9). Similarly, we can show (10) and (11). Equation (12) follows from the Lyapunov condition (13) by letting $a = 1$ and $b = 0$ or $a = 0$ and $b = 1$. This completes the proof of the lemma. □

LEMMA 3. *For any $\delta > 0$*

$$E|u_1|^{2+\delta} \le d^{\delta}\left(\sum_{i=1}^{d} E|X_{1,i}|^{2+\delta}\right)^2$$

and

$$E|v_1|^{2+\delta} \le 2^{4+\delta} d^{1+\delta} \sum_{i=1}^{d} E|X_{1,i}|^{2+\delta}.$$

PROOF. It follows from the Cauchy–Schwarz inequality that

$$|u_1|^2 \le \|X_1\|^2 \|X_{m+1}\|^2.$$

Using the C_r inequality that $E|\sum_{i=1}^d Z_i|^r \le d^{r-1} \times \sum_{i=1}^d E|Z_i|^r$ for any random variables $Z_1, \dots, Z_d$ and positive constant $r > 1$, we conclude that

$$\begin{aligned}
E|u_1|^{2+\delta} &\le E\left(\sum_{i=1}^d X_{1,i}^2\right)^{(2+\delta)/2} E\left(\sum_{i=1}^d X_{m+1,i}^2\right)^{(2+\delta)/2} \\
&= \left(E\left(\sum_{i=1}^d X_{1,i}^2\right)^{(2+\delta)/2}\right)^2 \\
&\le \left(d^{\delta/2}\sum_{i=1}^d E|X_{1,i}|^{2+\delta}\right)^2 \\
&= d^{\delta}\left(\sum_{i=1}^d E|X_{1,i}|^{2+\delta}\right)^2.
\end{aligned}$$

Similarly, from the C_r inequality we have

$$\begin{aligned}
E|v_1|^{2+\delta} &\le 2^{4+\delta} E\left(\sum_{i=1}^d |X_{1,i}|\right)^{2+\delta} \\
&\le 2^{4+\delta} d^{1+\delta} \sum_{i=1}^d E|X_{1,i}|^{2+\delta}.
\end{aligned}$$

This completes the proof. □

PROOF OF THEOREM 1. Set $u_i' = u_i/\sqrt{\pi_{11}}$, $v_i' = v_i/\sqrt{\pi_{22}}$ and $Y_i' = (u_i', v_i')^T$ for $i = 1, \dots, m$. Then it is easy to see that

$$\tilde{l}(0) = -2\log\tilde{L}(0) = 2\sum_{i=1}^m \log\{1 + \rho^T Y_i'\},$$

where $\rho = (\rho_1, \rho_2)^T$ solves

$$\frac{1}{m}\sum_{i=1}^m \frac{Y_i'}{1+\rho^T Y_i'} = 0. \tag{16}$$

It follows from Lemma 2 that

$$\frac{1}{\sqrt{m}}\sum_{i=1}^m Y_i' \xrightarrow{d} N(0, I_2), \tag{17}$$

$$\left\|\frac{1}{m}\sum_{i=1}^m Y_i'(Y_i')^T - I_2\right\| \xrightarrow{p} 0, \tag{18}$$

$$\max_{1\le i\le m}\|Y_i'\| = o_p(m^{1/2}). \tag{19}$$

Similar to the proof of (2.14) in Owen (1990), we can show $\|\rho\| = O_p(m^{-1/2})$. Then it follows from (19) that

$$\max_{1\le i\le m}\left\|\frac{\rho^T Y_i'}{1+\rho^T Y_i'}\right\| = o_p(1).$$

Therefore, we have from (16) that

$$\begin{aligned}
0 &= \frac{1}{m}\sum_{i=1}^m \frac{\rho^T Y_i'}{1+\rho^T Y_i'} \\
&= \frac{1}{m}\sum_{i=1}^m \rho^T Y_i'\left(1 - \rho^T Y_i' + \frac{(\rho^T Y_i')^2}{1+\rho^T Y_i'}\right) \\
&= \frac{1}{m}\sum_{i=1}^m \rho^T Y_i' - \frac{1}{m}\sum_{i=1}^m (\rho^T Y_i')^2 + \frac{1}{m}\sum_{i=1}^m \frac{(\rho^T Y_i')^3}{1+\rho^T Y_i'} \\
&= \frac{1}{m}\sum_{i=1}^m \rho^T Y_i' - \frac{(1+o_p(1))}{m}\sum_{i=1}^m (\rho^T Y_i')^2,
\end{aligned}$$

which implies

$$\frac{1}{m}\sum_{i=1}^m \rho^T Y_i' = \frac{(1+o_p(1))}{m}\sum_{i=1}^m (\rho^T Y_i')^2. \tag{20}$$

By using (16) and (18) we obtain

$$\begin{aligned}
0 &= \frac{1}{m}\sum_{i=1}^m \frac{Y_i'}{1+\rho^T Y_i'} \\
&= \frac{1}{m}\sum_{i=1}^m Y_i'\left(1 - (Y_i')^T\rho + \frac{(\rho^T Y_i')^2}{1+\rho^T Y_i'}\right) \\
&= \frac{1}{m}\sum_{i=1}^m Y_i' - \frac{1}{m}\sum_{i=1}^m Y_i'(Y_i')^T\rho + \frac{1}{m}\sum_{i=1}^n \frac{Y_i'(\rho^T Y_i')^2}{1+\rho^T Y_i'} \\
&= \frac{1}{m}\sum_{i=1}^m Y_i' - \frac{1}{m}\sum_{i=1}^m Y_i'(Y_i')^T\rho \\
&\quad + O_p\left(\max_{1\le i\le m}\left\|\frac{Y_i'}{1+\rho^T Y_i'}\right\|\frac{1}{m}\sum_{i=1}^n (\rho^T Y_i')^2\right) \\
&= \frac{1}{m}\sum_{i=1}^m Y_i' - \frac{1}{m}\sum_{i=1}^m Y_i'(Y_i')^T\rho \\
&\quad + o_p\left(m^{1/2}\rho^T\left(\frac{1}{m}\sum_{i=1}^n Y_i'(Y_i')^T\right)\rho\right) \\
&= \frac{1}{m}\sum_{i=1}^m Y_i' - \frac{1}{m}\sum_{i=1}^m Y_i'(Y_i')^T\rho + o_p(m^{-1/2}),
\end{aligned}$$

which implies

$$\rho=\left(\frac{1}{m}\sum_{i=1}^{m}Y_i'(Y_i')^T\right)^{-1}\frac{1}{m}\sum_{i=1}^{m}Y_i'+o_p(m^{-1/2}). \tag{21}$$

Finally, by using Taylor's expansion, (20), (21), (17) and (18), we obtain

$$\begin{aligned}\tilde{l}(0)&=2\sum_{i=1}^{m}\rho^TY_i'-(1+o_p(1))\sum_{i=1}^{m}(\rho^TY_i')^2\\&=(1+o_p(1))\rho^T\left(\sum_{i=1}^{m}Y_i'(Y_i')^T\right)\rho\\&=(1+o_p(1))\left(\frac{1}{\sqrt{m}}\sum_{i=1}^{m}Y_i'\right)^T\\&\quad\cdot\left(\frac{1}{m}\sum_{i=1}^{m}Y_i'(Y_i')^T\right)^{-1}\frac{1}{\sqrt{m}}\sum_{i=1}^{m}Y_i'\\&\quad+o_p(1)\\&\xrightarrow{d}\chi_2^2\quad\text{as } n\to\infty.\end{aligned}$$

This completes the proof of Theorem 1. □

PROOF OF COROLLARY 1. Equations (4) and (5) follow from conditions (A1)–(A3) by using Lemmas 1 and 3. □

PROOF OF THEOREM 2. It suffices to verify conditions (4) and (5) with $\delta=2$ in Theorem 1. As before, assume $\mu_0=0$. Write $\Gamma=(\gamma_{i,j})_{1\le i\le d,1\le j\le k}$. Then $\mathrm{Var}(X_1)=\Sigma=\Gamma\Gamma^T$. Denote $\mathbf{1}_d^T\Gamma=(a_1,\dots,a_k)$ and $\Sigma'=\Gamma^T\Gamma=(\sigma'_{j,l})_{1\le j,l\le k}$. Then $v_1=v_1(0)=\sum_{j=1}^{k}a_j(Z_{1,j}+Z_{1+m,j})$ and

$$u_1=u_1(0)=\sum_{j=1}^{k}\sum_{l=1}^{k}\sigma'_{j,l}Z_{1,j}Z_{1+m,l}.$$

Set $\delta_{j_1,j_2,j_3,j_4}=E(Z_{1,j_1}Z_{1,j_2}Z_{1,j_3}Z_{1,j_4})$. Then

$$\delta_{j_1,j_2,j_3,j_4}=3+\Delta,$$

if $j_1=j_2=j_3=j_4$, 1 if j_1, j_2, j_3 and j_4 form two different pairs of integers, and zero otherwise. It follows from Lemma 1 that

$$\begin{aligned}Eu_1^4&=\sum_{j_1,j_2,j_3,j_4=1}^{k}\sum_{l_1,l_2,l_3,l_4=1}^{k}\sigma'_{j_1,l_1}\sigma'_{j_2,l_2}\sigma'_{j_3,l_3}\sigma'_{j_4,l_4}\\&\qquad\qquad\cdot\delta_{j_1,j_2,j_3,j_4}\delta_{l_1,l_2,l_3,l_4}\\&=O\left(\left|\sum_{j_1\ne j_2}\sum_{l_1\ne l_2}\sigma'_{j_1,l_1}\sigma'_{j_1,l_2}\sigma'_{j_2,l_1}\sigma'_{j_2,l_2}\right|\right)\\&\quad+O\left(\sum_{j_1\ne j_2}\sum_{l=1}^{k}\sigma'^2_{j_1,l}\sigma'^2_{j_2,l}\right)\\&\quad+O\left(\sum_{j=1}^{k}\sum_{l_1\ne l_2}\sigma'^2_{j,l_1}\sigma'^2_{j,l_2}\right)+O\left(\sum_{j=1}^{k}\sum_{l=1}^{k}\sigma'^4_{j,l}\right)\\&=O\left(\left|\sum_{j_1=1}^{k}\sum_{j_2=1}^{k}\sum_{l_1=1}^{k}\sum_{l_2=1}^{k}\sigma'_{j_1,l_1}\sigma'_{j_1,l_2}\sigma'_{j_2,l_1}\sigma'_{j_2,l_2}\right|\right)\\&\quad+O\left(\sum_{j_1=1}^{k}\sum_{j_2=1}^{k}\sum_{l=1}^{k}\sigma'^2_{j_1,l}\sigma'^2_{j_2,l}\right)\\&\quad+O\left(\sum_{j=1}^{k}\sum_{l=1}^{k}\sigma'^4_{j,l}\right)\\&=O(\mathrm{tr}(\Sigma'^4))+O\left(\left(\sum_{j=1}^{k}\sum_{l=1}^{k}\sigma'^2_{j,l}\right)^2\right)\\&=O(\mathrm{tr}(\Sigma'^4))+O((\mathrm{tr}(\Sigma'^2))^2)\\&=O(\mathrm{tr}(\Sigma^4))+O((\mathrm{tr}(\Sigma^2))^2)\\&=o(m(\mathrm{tr}(\Sigma^2))^2),\end{aligned}$$

that is, (4) holds with $\delta=2$.

Similarly, we have

$$\begin{aligned}Ev_1^4&\le 2^4E\left(\sum_{j=1}^{k}a_jZ_{1,j}\right)^4\\&=O\left(\sum_{j_1,j_2=1}^{k}a_{j_1}^2a_{j_2}^2\right)+O\left(\sum_{j=1}^{k}a_j^4\right)\\&=O\left(\left(\sum_{j=1}^{k}a_j^2\right)^2\right)\\&=O((\mathbf{1}_d^T\Gamma\Gamma^T\mathbf{1}_d)^2)\\&=O\left(\left(\sum_{i=1}^{d}\sum_{j=1}^{d}\sigma_{i,j}\right)^2\right),\end{aligned}$$

which yields (5) with $\delta=2$. The proof is complete. □

PROOF OF THEOREM 3. We continue to use the notation in the proof of Theorem 1. Define

$$\rho_{n1}=\frac{(\mu_0-\mu)^T(\mu_0-\mu)}{\sqrt{\pi_{11}}},$$

$$\rho_{n2}=\frac{\mathbf{1}_d^T(2\mu-2\mu_0)}{\sqrt{\pi_{22}}}.$$

Then $\tau = m\rho_{n1}^2 + m\rho_{n2}^2$.

Notice that the true value for the mean of X_1 is μ under the alternative hypothesis. Since for $1 \le i \le m$

$$u_i(\mu_0) = u_i(\mu) + (\mu - \mu_0)^T(\mu - \mu_0) + (\mu - \mu_0)^T(X_i + X_{i+m} - 2\mu) \tag{22}$$

and

$$v_i(\mu_0) = v_i(\mu) + \mathbf{1}_d^T(2\mu - 2\mu_0),$$

we have

$$Y_i' = \begin{pmatrix} u_i(\mu)/\sqrt{\pi_{11}} \\ v_i(\mu)/\sqrt{\pi_{22}} \end{pmatrix} + \begin{pmatrix} \rho_{n1} \\ \rho_{n2} \end{pmatrix} + \begin{pmatrix} s_i(\mu) \\ 0 \end{pmatrix},$$

where $s_i(\mu) = (\mu - \mu_0)^T(X_i + X_{i+m} - 2\mu)/\sqrt{\pi_{11}}$ and $Y_i' = \binom{u_i(\mu_0)/\sqrt{\pi_{11}}}{v_i(\mu_0)/\sqrt{\pi_{22}}}$ as defined in the proof of Theorem 1.

First we consider the case of $\tau = o(m)$. Since $\tau = o(m)$ implies that $\rho_{n1} = o(1)$ and $\rho_{n2} = o(1)$, it follows from Lemma 1 that

$$\begin{aligned} E(s_1^2(\mu)) &= O\left(\frac{1}{\pi_{11}}(\mu - \mu_0)^T\Sigma(\mu - \mu_0)\right) \\ &= O\left(\frac{\lambda_d}{\pi_{11}}(\mu - \mu_0)^T(\mu - \mu_0)\right) \\ &= O\left(\frac{\lambda_d}{\sqrt{\pi_{11}}}\rho_{n1}\right) \to 0, \end{aligned} \tag{23}$$

which implies

$$\frac{\sum_{i=1}^m s_i^2(\mu)}{m} \xrightarrow{p} 0$$

and

$$\frac{\max_{1\le i\le m}|s_i(\mu)|}{\sqrt{m}} \le \sqrt{\frac{\sum_{i=1}^m s_i^2(\mu)}{m}} \xrightarrow{p} 0$$

as $m \to \infty$. Hence, we conclude that

$$\begin{aligned} V_n := \begin{pmatrix} V_{n1} \\ V_{n2} \end{pmatrix} &= \frac{1}{\sqrt{m}}\sum_{i=1}^m \left\{Y_i' - \begin{pmatrix} \rho_{n1} \\ \rho_{n2} \end{pmatrix}\right\} \\ &\xrightarrow{d} N(0, I_2), \end{aligned} \tag{24}$$

and both (18) and (19) hold when $\tau = o(m)$. Following the proof of Theorem 1, we can show that

$$\begin{aligned} \tilde{l}(\mu_0) &= (1 + o_p(1))\left(\frac{1}{\sqrt{m}}\sum_{i=1}^m Y_i'\right)^T \\ &\quad \cdot \left(\frac{1}{m}\sum_{i=1}^m Y_i'(Y_i')^T\right)^{-1}\frac{1}{\sqrt{m}}\sum_{i=1}^m Y_i' + o_p(1) \\ &= (V_{n1} + \sqrt{m}\rho_{n1})^2(1 + o_p(1)) \\ &\quad + (V_{n2} + \sqrt{m}\rho_{n2})^2(1 + o_p(1)) + o_p(1), \end{aligned} \tag{25}$$

when $\tau = o(m)$.

If the limit of τ, say, τ_0, is finite, then it follows from (24) and (25) that $\tilde{l}(\mu_0)$ converges in distribution to a noncentral chi-square distribution with two degrees of freedom and noncentrality parameter τ_0 and, consequently, (6) holds. If τ goes to infinity, the limit of the right-hand side of (6) is 1. By (25), together with the elementary inequality $(a+b)^2 \ge \frac{a^2}{2} - b^2$, we have that

$$\begin{aligned} \tilde{l}(\mu_0) &\ge \left(\frac{m\rho_{n1}^2}{2} - V_{n1}^2\right)(1 + o_p(1)) \\ &\quad + \left(\frac{m\rho_{n2}^2}{2} - V_{n2}^2\right)(1 + o_p(1)) + o_p(1) \\ &= \frac{\tau}{2}(1 + o_p(1)) \\ &\quad - (V_{n1}^2 + V_{n2}^2)(1 + o_p(1)) + o_p(1) \\ &\xrightarrow{p} \infty, \end{aligned} \tag{26}$$

which implies that the limit of the left-hand side of (6) is also 1. Thus, (6) also holds when $\tau = o(m)$.

Next we consider the case of $\liminf \rho_{n2}^2 > 0$. Since $\sum_{i=1}^m p_i Y_i(\mu_0) = 0$ implies that $\sum_{i=1}^m p_i v_i(\mu_0) = 0$, we have

$$\begin{aligned} \tilde{L}(\mu_0) &\le \sup\left\{\prod_{i=1}^m (mp_i) : p_1 \ge 0, \dots, p_m \ge 0, \right. \\ &\qquad \left. \sum_{i=1}^m p_i = 1, \sum_{i=1}^m p_i v_i(\mu_0) = 0\right\} \\ &= \sup\left\{\prod_{i=1}^m (mp_i) : p_1 \ge 0, \dots, p_m \ge 0, \right. \\ &\qquad \left. \sum_{i=1}^m p_i = 1, \sum_{i=1}^m p_i \frac{v_i(\mu_0)}{\sqrt{\pi_{22}}} = 0\right\}. \end{aligned} \tag{27}$$

Define

$$\begin{aligned} L^*(\theta) = \sup\left\{\prod_{i=1}^m (mp_i) : p_1 \ge 0, \dots, p_m \ge 0, \right. \\ \left. \sum_{i=1}^m p_i = 1, \sum_{i=1}^m p_i \left(\frac{v_i(\mu_0)}{\sqrt{\pi_{22}}} - \rho_{n2}\right) = \theta\right\}. \end{aligned}$$

Put $\theta^* = \frac{1}{m}\sum_{i=1}^m (\frac{v_i(\mu_0)}{\sqrt{\pi_{22}}} - \rho_{n2})$. Then

$$-2\log L^*(\theta^*) = 0. \tag{28}$$

Since $E\{v_i(\mu_0)/\sqrt{\pi_{22}} - \rho_{n2}\} = E\{v_i(\mu)/\sqrt{\pi_{22}}\} = 0$ and $E\{v_i(\mu_0)/\sqrt{\pi_{22}} - \rho_{n2}\}^2 = 1$ under $H_a: \mu \ne \mu_0$, we have by using Chebyshev's inequality that

$$P(|\theta^*| > m^{-2/5}) \to 0. \tag{29}$$

Using $E\{v_i(\mu_0)/\sqrt{\pi_{22}} - \rho_{n2}\}^2 = 1$, similar to the proof of (26), we can show that

$$-2\log L^*(\theta_1^*) \xrightarrow{p} \infty \quad \text{and} \quad -2\log L^*(\theta_2^*) \xrightarrow{p} \infty,$$

where $\theta_1^* = m^{-1/4}$ and $\theta_2^* = -m^{-1/4}$, which satisfy $m(\theta_1^*)^2 = o(m)$ and $m(\theta_2^*)^2 = o(m)$. To help understand this better, we first notice that $\tilde{L}(\mu_0)$ can be rewritten as follows:

$$\begin{aligned}\tilde{L}(\mu_0) &= \sup\Biggl\{\prod_{i=1}^m (mp_i) : p_1 \geq 0, \ldots, p_m \geq 0, \\ &\qquad\qquad \sum_{i=1}^m p_i = 1, \sum_{i=1}^m p_i Y_i' = 0\Biggr\} \\ &= \sup\Biggl\{\prod_{i=1}^m (mp_i) : p_1 \geq 0, \ldots, p_m \geq 0, \\ &\qquad \sum_{i=1}^m p_i = 1, \\ &\qquad \sum_{i=1}^m p_i\left(Y_i' - \begin{pmatrix}\rho_{n1}\\ \rho_{n2}\end{pmatrix}\right) = -\begin{pmatrix}\rho_{n1}\\ \rho_{n2}\end{pmatrix}\Biggr\}.\end{aligned}$$

In equation (26), it is the quantity $m\|-\binom{\rho_{n1}}{\rho_{n2}}\|^2 = \tau$ that determines whether $\tilde{l}(\mu_0)$ diverges. As a one-dimensional analogue of $\tilde{L}(\mu_0)$, for any sequence θ_n, if $\theta_n = o(1)$, $-2\log L^*(\theta_n)$ can be expanded as in (25) via replacing Y_i' by $v_i(\mu_0)/\sqrt{\pi_{22}} - \rho_{n2}$, and replacing $-\binom{\rho_{n1}}{\rho_{n2}}$ by θ_n. And if, further, $m\theta_n^2 \to \infty$, $-2\log L^*(\theta_n)$ goes to infinity in probability, just like (26). Obviously, with the choices of $\theta_n = \pm m^{-1/4}$, conditions $\theta_n = o(1)$ and $m\theta_n^2 \to \infty$ are satisfied.

It follows from Hall and La Scala (1990) that the set $\{\theta : -2\log L^*(\theta) \leq c\} =: I_c$ is convex for any c. Take $c = c_n = \min\{-2\log L^*(\theta_1^*), -2\log L^*(\theta_2^*)\}/2$. If $-\rho_{n2}$ belongs to the above convex set, then it follows from (28) that $-a\rho_{n2} + (1-a)\theta^*$ belongs to that convex set for any $a \in [0, 1]$ or, equivalently, any number between $-\rho_{n2}$ and θ^* belongs to I_{c_n}. Recall that we assume $\liminf \rho_{n2}^2 > 0$. Assume n is large such that $m^{-1/4} < |\rho_{n2}|$. Under the condition $|\theta^*| \leq m^{-2/5}$, if $\rho_{n2} > 0$, then $-\rho_{n2} < \theta_2^* = -m^{-1/4} < \theta^*$, and if $\rho_{n2} < 0$, then $\theta^* < \theta_1^* = m^{-1/4} < -\rho_{n2}$. Therefore, if $|\theta^*| \leq m^{-2/5}$ and $-\rho_{n2} \in I_{c_n}$, at least one of θ_1^* and θ_2^* belongs to I_{c_n}. Precisely, we have, as n goes to infinity,

$$\begin{aligned}P\bigl(|\theta^*| &\leq m^{-2/5}, -\rho_{n2} \in I_{c_n}\bigr) \\ &\leq P\bigl(\theta_1^* \in I_{c_n} \text{ or } \theta_2^* \in I_{c_n}\bigr) \\ &= P\bigl(\min\{-2\log L^*(\theta_1^*), -2\log L^*(\theta_2^*)\} \leq c_n\bigr) \\ &= P\bigl(\min\{-2\log L^*(\theta_1^*), -2\log L^*(\theta_2^*)\} = 0\bigr) \\ &\to 0,\end{aligned}$$

which, together with (29), implies

$$\begin{aligned}P\bigl(-2\log L^*(-\rho_{n2}) > c_n\bigr) &= P(-\rho_{n2} \notin I_{c_n}) \\ &\geq 1 - P\bigl(|\theta^*| \leq m^{-2/5}, -\rho_{n2} \in I_{c_n}\bigr) \\ &\quad - P\bigl(|\theta^*| > m^{-2/5}\bigr) \\ &\to 1\end{aligned}$$

and, therefore,

$$-2\log L^*(-\rho_{n2}) \xrightarrow{p} \infty \tag{30}$$

since $c_n \xrightarrow{p} \infty$. Hence, combining with (27), we have

$$\begin{aligned}P\bigl(-2\log \tilde{L}(\mu_0) > \xi_{1-\alpha}\bigr) &\geq P\bigl(-2\log L^*(-\rho_{n2}) > \xi_{1-\alpha}\bigr) \\ &\to 1,\end{aligned}$$

when $\liminf \rho_{n2}^2 > 0$.

Next we consider the case of $\liminf \rho_{n1} > 0$. Define $\pi_{33} = E\{(\mu - \mu_0)^T(X_1 + X_{1+m} - 2\mu)\}^2$ and $\rho_{n3} = \frac{(\mu_0-\mu)^T(\mu_0-\mu)}{\sqrt{\pi_{11}+\pi_{33}}}$. As before, we have

$$\begin{aligned}\tilde{L}(\mu_0) &\leq \sup\Biggl\{\prod_{i=1}^m (mp_i) : p_1 \geq 0, \ldots, p_m \geq 0, \\ &\qquad\qquad \sum_{i=1}^m p_i = 1, \sum_{i=1}^m p_i u_i(\mu_0) = 0\Biggr\} \\ &= \sup\Biggl\{\prod_{i=1}^m (mp_i) : p_1 \geq 0, \ldots, p_m \geq 0, \\ &\qquad \sum_{i=1}^m p_i = 1, \sum_{i=1}^m p_i \frac{u_i(\mu_0)}{\sqrt{\pi_{11}+\pi_{33}}} = 0\Biggr\}.\end{aligned} \tag{31}$$

Define

$$\begin{aligned}L^{**}(\theta) = \sup\Biggl\{&\prod_{i=1}^m (mp_i) : p_1 \geq 0, \ldots, p_m \geq 0, \\ &\sum_{i=1}^m p_i = 1, \\ &\sum_{i=1}^m p_i\left(\frac{u_i(\mu_0)}{\sqrt{\pi_{11}+\pi_{33}}} - \rho_{n3}\right) = \theta\Biggr\}.\end{aligned}$$

Since $u_1(\mu)$ and $(\mu - \mu_0)^T(X_1 + X_{1+m} - 2\mu)$ are two uncorrelated variables with zero means, we have

$\mathrm{Var}(u_1(\mu)+(\mu-\mu_0)^T(X_1+X_{1+m}-2\mu))=\pi_{11}+\pi_{33}$. As we have shown in the proof of Theorem 2, $E|u_1(\mu)|^4=o(m\pi_{11}^2)$. Following the same lines for estimating $E(v_1^4)$ in the end of the proof of Theorem 2, we have

$$E\{(\mu-\mu_0)^T(X_1+X_{1+m}-2\mu)\}^4=O(\pi_{33}^2).$$

Then it follows that

$$\begin{aligned} &E\{u_1(\mu)+(\mu-\mu_0)^T(X_1+X_{1+m}-2\mu)\}^4\\ &\quad\le 8\bigl(E|u_1(\mu)|^4\\ &\qquad+E\{(\mu-\mu_0)^T(X_1+X_{1+m}-2\mu)\}^4\bigr)\\ &\quad=o\bigl(m(\pi_{11}+\pi_{33})^2\bigr).\end{aligned}$$

From (22),

$$\begin{aligned}&\frac{u_i(\mu_0)}{\sqrt{\pi_{11}+\pi_{33}}}-\rho_{n3}\\ &\quad=\frac{u_i(\mu)+(\mu-\mu_0)^T(X_i+X_{i+m}-2\mu)}{\sqrt{\pi_{11}+\pi_{33}}}\end{aligned}$$

and, thus, we have

$$\begin{aligned}&E\biggl(\frac{u_i(\mu_0)}{\sqrt{\pi_{11}+\pi_{33}}}-\rho_{n3}\biggr)^4\\ &\quad=\frac{E(u_i(\mu)+(\mu-\mu_0)^T(X_i+X_{i+m}-2\mu))^4}{(\pi_{11}+\pi_{33})^2}\\ &\quad=o(m).\end{aligned}$$

This ensures the validity of the Wilks theorem for $-2\log L^{**}(0)$, that is, $-2\log L^{**}(0)$ converges in distribution to a chi-square distribution with one degree of freedom. Note that in Theorem 1, two similar conditions, (4) and (5), are imposed to obtain the Wilks theorem for the log-empirical likelihood statistic for two-dimensional mean vectors. Similar to the proof of (26), we can show that

$$-2\log L^{**}(\theta_1^*)\xrightarrow{p}\infty\quad\text{and}\quad -2\log L^{**}(\theta_2^*)\xrightarrow{p}\infty,$$

where $\theta_1^*=m^{-1/4}$ and $\theta_2^*=-m^{-1/4}$, which satisfy $m(\theta_1^*)^2=o(m)$ and $m(\theta_2^*)^2=o(m)$.

Put $\theta^{**}=\frac{1}{m}\sum_{i=1}^m(\frac{u_i(\mu_0)}{\sqrt{\pi_{11}+\pi_{33}}}-\rho_{n3})$. Then

$$-2\log L^{**}(\theta^{**})=0. \tag{32}$$

Since

$$\begin{aligned}&E\{u_i(\mu_0)/\sqrt{\pi_{11}+\pi_{33}}-\rho_{n3}\}\\ &\quad=E\biggl\{\frac{u_i(\mu)+(\mu-\mu_0)^T(X_i+X_{i+m}-2\mu)}{\sqrt{\pi_{11}+\pi_{33}}}\biggr\}\\ &\quad=0\end{aligned}$$

and

$$\begin{aligned}&E\biggl\{\frac{u_i(\mu_0)}{\sqrt{\pi_{11}+\pi_{33}}}-\rho_{n3}\biggr\}^2\\ &\quad=E\biggl\{\frac{u_i(\mu)+(\mu-\mu_0)^T(X_i+X_{i+m}-2\mu)}{\sqrt{\pi_{11}+\pi_{33}}}\biggr\}^2\\ &\quad=1\end{aligned}$$

under $H_a:\mu\ne\mu_0$, we have from Chebyshev's inequality that

$$P\bigl(|\theta^{**}|>m^{-2/5}\bigr)\to 0. \tag{33}$$

By (23), we have $\pi_{33}/\pi_{11}=Es_1^2(\mu)=O(\rho_{n1})$, which implies that there exists a constant $M>0$ such that

$$\begin{aligned}\rho_{n3}/m^{-1/4}&=m^{1/4}\rho_{n1}\frac{\sqrt{\pi_{11}}}{\sqrt{\pi_{11}+\pi_{33}}}\\ &\ge m^{1/4}\rho_{n1}\{1+M\rho_{n1}\}^{-1/2}\\ &\to\infty\end{aligned}$$

since $\liminf\rho_{n1}>0$.

Using (32), (33) and the same arguments in proving (30), we have

$$-2\log L^{**}(-\rho_{n3})\xrightarrow{p}\infty.$$

Hence, combining with (31), we have

$$\begin{aligned}&P\bigl(-2\log\tilde{L}(\mu_0)>\xi_{1-\alpha}\bigr)\\ &\quad\ge P\bigl(-2\log L^{**}(-\rho_{n3})>\xi_{1-\alpha}\bigr)\\ &\quad\to 1,\end{aligned}$$

when $\liminf\rho_{n1}^2>0$. Therefore, (6) holds when $\liminf\rho_{n1}>0$. This completes the proof of Theorem 3. □

ACKNOWLEDGMENTS

We thank the Editor, one Associate Editor and two reviewers for their helpful comments. Peng's research was supported by NSF Grant DMS-10-05336, Qi's research was supported by NSF Grant DMS-10-05345 and Wang's research was supported by NSFC Grant No. 11271033, Foundation of Beijing Education Bureau Grant No. KM201110028003.

REFERENCES

ARLOT, S., BLANCHARD, G. and ROQUAIN, E. (2010a). Some nonasymptotic results on resampling in high dimension. I. Confidence regions. *Ann. Statist.* **38** 51–82. MR2589316

ARLOT, S., BLANCHARD, G. and ROQUAIN, E. (2010b). Some nonasymptotic results on resampling in high dimension. II. Multiple tests. *Ann. Statist.* **38** 83–99. MR2589317

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

CHEN, S. X., PENG, L. and QIN, Y. (2009). Empirical likelihood methods for high dimension. *Biometrika* **96** 711–722.

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

CHEN, S. X. and VAN KEILEGOM, I. (2009). A review on empirical likelihood methods for regression. *TEST* **18** 415–447. MR2566404

DHARMADHIKARI, S. W. and JOGDEO, K. (1969). Bounds on moments of certain random variables. *Ann. Math. Statist.* **40** 1506–1509. MR0243587

HALL, P. and LA SCALA, B. (1990). Methodology and algorithms of empirical likelihood. *Internat. Statist. Rev.* **58** 109–127.

HJORT, N. L., MCKEAGUE, I. W. and VAN KEILEGOM, I. (2009). Extending the scope of empirical likelihood. *Ann. Statist.* **37** 1079–1111. MR2509068

KUELBS, J. and VIDYASHANKAR, A. N. (2010). Asymptotic inference for high-dimensional data. *Ann. Statist.* **38** 836–869. MR2604698

OWEN, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75** 237–249. MR0946049

OWEN, A. (1990). Empirical likelihood ratio confidence regions. *Ann. Statist.* **18** 90–120. MR1041387

OWEN, A. (2001). *Empirical Likelihood*. Chapman & Hall/CRC, Boca Raton, FL.

SRIVASTAVA, M. S. (2009). A test for the mean vector with fewer observations than the dimension under non-normality. *J. Multivariate Anal.* **100** 518–532. MR2483435

SRIVASTAVA, M. S. and DU, M. (2008). A test for the mean vector with fewer observations than the dimension. *J. Multivariate Anal.* **99** 386–402. MR2396970

VON BAHR, B. and ESSEEN, C. G. (1965). Inequality for the rth absolute moment of a sum of random variables, $1 \le r \le 2$. *Ann. Math. Statist.* **36** 299–393.

YIN, Y. Q., BAI, Z. D. and KRISHNAIAH, P. R. (1988). On the limit of the largest eigenvalue of the large-dimensional sample covariance matrix. *Probab. Theory Related Fields* **78** 509–521. MR0950344