Biometrika (2020), **107**, 3, *pp.* 591–607
Printed in Great Britain

doi: 10.1093/biomet/asaa005
Advance Access publication 19 March 2020

Empirical likelihood test for a large-dimensional mean vector

BY XIA CUI

School of Economics and Statistics, Guangzhou University, 230 Wai Huan Xi Road, Guangzhou 510006, China

cuixia@gzhu.edu.cn

RUNZE LI

Department of Statistics, Pennsylvania State University, University Park, Pennsylvania 16802, U.S.A.

rzli@psu.edu

GUANGREN YANG

Department of Statistics, School of Economics, Jinan University, 601 Huangpu W. Ave., Guangzhou 510632, China

tygr@jnu.edu.cn

AND WANG ZHOU

Department of Statistics and Applied Probability, National University of Singapore, 21 Lower Kent Ridge Road, 117546 Singapore

stazw@nus.edu.sg

SUMMARY

This paper is concerned with empirical likelihood inference on the population mean when the dimension p and the sample size n satisfy $p/n \to c \in [1, \infty)$. As shown in Tsao (2004), the empirical likelihood method fails with high probability when $p/n > 1/2$ because the convex hull of the n observations in $\mathbb{R}^p$ becomes too small to cover the true mean value. Moreover, when $p > n$, the sample covariance matrix becomes singular, and this results in the breakdown of the first sandwich approximation for the log empirical likelihood ratio. To deal with these two challenges, we propose a new strategy of adding two artificial data points to the observed data. We establish the asymptotic normality of the proposed empirical likelihood ratio test. The proposed test statistic does not involve the inverse of the sample covariance matrix. Furthermore, its form is explicit, so the test can easily be carried out with low computational cost. Our numerical comparison shows that the proposed test outperforms some existing tests for high-dimensional mean vectors in terms of power. We also illustrate the proposed procedure with an empirical analysis of stock data.

Some key words: Empirical likelihood test; One-sample problem; Random matrix theory.

1. Introduction

This paper studies how to use the empirical likelihood method to test whether a high-dimensional mean vector takes a specific value. The empirical likelihood method was proposed by Owen (1988, 1990) and is a powerful nonparametric statistical tool. Suppose that $\{x_1, \ldots, x_n\}$ is an independent and identically distributed random sample from a p-dimensional population with mean μ. The empirical likelihood function for μ is defined as

$$R_n(\mu) = \max\left(\prod_{i=1}^{n} n\omega_i : \omega_i \geqslant 0, \sum_{i=1}^{n} \omega_i = 1, \sum_{i=1}^{n} \omega_i x_i = \mu\right).$$

The asymptotic behaviour of $R_n(\mu)$ has been studied in the setting of fixed and finite p by Owen (1990) and in the settings with $p = o(n^{1/3})$ and $p = o(n^{1/2})$ by Hjort et al. (2009) and Chen et al. (2009), respectively.

There are some challenges in extending the empirical likelihood method for statistical inference of a large-dimensional mean vector when $c_n = p/n \to c > 0$, i.e., in the case where p and n are of the same order. The first difficulty is that the corresponding optimization problem of the empirical likelihood method becomes problematic because, as shown in Tsao (2004), when $c_n > 1/2$ there is a positive probability of the event that the mean vector μ falls outside the convex hull of the n data points. As a result, the empirical likelihood for a p-dimensional population breaks down with positive probability when $c_n > 1/2$. To deal with this problem, which is the main obstacle related to the convex hull problem, Chen et al. (2008) proposed adding an artificial data point to the n observations. Their proposal pushes the mean vector into the convex hull of the $n+1$ data points. However, the sample mean of the $n+1$ data points is different from the sample mean of the n actual observations. As a result, the empirical likelihood method cannot be used for statistical inference on the population mean. To deal with this issue, Emerson & Owen (2009) proposed adding two artificial data points to the n observations, and Chen et al. (2015) further showed that the empirical likelihood is still valid when $c_n \to c \in (0, 1)$ as n goes to infinity. Their proposal is not applicable in the $c_n \geqslant 1$ case because the two artificial data points rely on the inverse of the sample covariance matrix, which becomes singular when $p \geqslant n$. Indeed, the singularity of the sample covariance matrix is the second obstacle in extending the empirical likelihood method to the setting with $c_n \geqslant 1$.

This paper aims to develop an empirical likelihood statistic for the mean vector μ in the setting of $c_n \to c \in [1, \infty)$. To deal with the optimization problem and the singularity of the sample covariance matrix, we propose a new strategy of adding two artificial data points to the n actual observations so that the true mean vector falls in the convex hull of $n+2$ data points, the sample mean of the $n+2$ data points equals the sample mean of the n actual observations, and the two artificial data points do not involve the inverse of the sample covariance matrix. The proposed empirical likelihood ratio test statistic has several appealing properties. First, we are able to derive a closed-form expression for the empirical likelihood statistic, so it can be implemented without involving numerical optimization. This is a very useful property because the empirical likelihood method is typically computationally expensive. Second, the proposed empirical likelihood ratio test statistic does not involve the inverse of the sample covariance matrix. Therefore, it can be implemented in both of the settings $p \geqslant n$ and $p < n$. We also study the asymptotic behaviour of the proposed empirical likelihood ratio test and derive its asymptotic distribution under the null hypothesis, and under a local alternative hypothesis by using probability tools on the concentration properties of certain quadratic forms. Under the local alternative hypothesis, we show that our new test yields a significant power gain over tests proposed by Bai & Sarandasa (1996), Chen & Qin (2010) and Wang et al. (2015). We further construct an empirical likelihood ratio test statistic for linear hypotheses about the population mean and derive its limiting null distribution. Simulation results indicate that the proposed tests perform well.

2. Large-dimensional empirical likelihood method

2.1. *Notation and setting*

Suppose that $\{x_i : i = 1, \ldots, n\}$ is a random sample from a p-dimensional population x with mean μ and covariance matrix Σ. Throughout this paper, we write $\bar{x} = \sum_{i=1}^{n} x_i/n$ and $S = (n-1)^{-1} \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})^{\mathrm{T}}$ for the sample mean and sample covariance matrix, respectively. This section develops an empirical likelihood test for

$$H_0 : \mu = \mu_0 \quad \text{versus} \quad H_{\mathrm{a}} : \mu \neq \mu_0 \tag{1}$$

in the case where p is large. Throughout the paper we impose the following assumptions.

Assumption 1. We can represent x as

$$x = \mu + \Sigma^{1/2} z, \tag{2}$$

where $z = (z_1, \ldots, z_p)^{\mathrm{T}}$ has independent and identically distributed elements with mean 0 and variance 1 such that $E(|z_j|^6) < \infty$.

Assumption 2. Writing $c_n = p/n$, we have $c_n \to c \in [1, \infty)$.

We refer to the model defined in Assumption 1 as the independent component model. It is easy to verify that the multivariate normal distribution satisfies Assumption 1. As a natural extension of the multivariate normal distribution, the independent component model is a typical assumption in the literature on statistical inference on mean vectors (Bai & Sarandasa, 1996; Chen & Qin, 2010) and has received attention in signal processing and machine learning (Hyvärinen et al., 2001) as well. For ease of notation and without loss of generality, we focus on the case where z is a p-dimensional vector in Assumption 1. The proposed procedure can be directly extended to the model $x = \mu + \Gamma z$, where z is q-dimensional random vector with $q > p$ and Γ is a $p \times q$ constant matrix of rank p. Under this model, $\Sigma = \mathrm{cov}(x) = \Gamma\Gamma^{\mathrm{T}}$ is of full rank.

2.2. *Empirical likelihood method*

The traditional empirical likelihood method becomes inapplicable when $p > n$, because the zero vector falls outside the convex hull of $\{x_i - \mu : i = 1, \ldots, n\}$ with positive probability (Tsao, 2004). Furthermore, the singularity of the sample covariance matrix of x results in the breakdown of the first sandwich approximation for the log empirical likelihood ratio. Thus, it is challenging to construct an empirical likelihood test for (1).

To deal with the convex hull issue, we propose adding the following two pseudo-observations:

$$x_{n+1} = \mu_0 - a_n(\bar{x} - \mu_0), \quad x_{n+2} = \mu_0 + (2 + a_n)(\bar{x} - \mu_0). \tag{3}$$

Introducing x_{n+1} ensures that the zero vector will be contained in the convex hull of $\{x_i - \mu : i = 1, \ldots, n, n+1\}$. The second point x_{n+2} is introduced to maintain the original sample mean since $\sum_{i=1}^{n+2} x_i/(n+2) = \bar{x}$. The factor a_n plays a key role in achieving the desirable theoretical properties of the empirical likelihood-based statistic. We propose

$$a_n = \frac{l_n}{\{\|\bar{x} - \mu_0\|^2 + k_n|\alpha^{\mathrm{T}}(\bar{x} - \mu_0)|^2\}^{1/2}}, \tag{4}$$

where a_n depends on positive constants k_n and l_n, and α is a p-dimensional constant vector satisfying $\|\alpha\| = 1$.

Our proposal in (3) is motivated by the proposal of Chen et al. (2015), but these two approaches differ in the choice of a_n. Chen et al. (2015) proposed taking a_n proportional to $\{(\bar{x}-\mu_0)^{\mathrm{T}} S^{-1}(\bar{x}-\mu_0)\}^{1/2}/\|\bar{x} - \mu_0\|^2$ and showed that their test is asymptotically equivalent to Hotelling's T^2 test. Clearly, the method of Chen et al. (2015) becomes infeasible when $c_n \geqslant 1$ due to the singularity of S. As demonstrated in Bai & Sarandasa (1996), Hotelling's T^2 may suffer from lower power when c_n approaches 1 from below. As demonstrated in Theorem 1, inclusion of the term $k_n|\alpha^{\mathrm{T}}(\bar{x}-\mu_0)|^2$ in (4) is necessary for the empirical likelihood ratio test for (1) to achieve a higher local power.

For the new dataset $\{x_i : i = 1, \ldots, n, n+1, n+2\}$, the empirical likelihood ratio is

$$R(\mu_0, k_n) = \max\left\{\prod_{i=1}^{n+2}(n+2)\omega_i : \omega_i \geqslant 0,\ \sum_{i=1}^{n+2}\omega_i = 1,\ \sum_{i=1}^{n+2}\omega_i(x_i - \mu_0) = 0\right\},$$

where we write $R(\mu_0, k_n)$ to emphasize its dependence on k_n, and the log empirical likelihood statistic is $W(\mu_0, k_n) = -2\log R(\mu_0, k_n)$.

Surprisingly, there is an explicit formula for $W(\mu_0, k_n)$ when $p \geqslant n$. Specifically, let $\xi_n = (n+2)/(1+a_n)$; then we can show that a closed-form expression for $W(\mu_0, k_n)$ is

$$\begin{aligned} W(\mu_0, k_n) = -2\Bigg(& n\log\left[1 + \frac{1}{n}\left\{1 - \left(1 + \frac{n}{n+2}\xi_n^2\right)^{1/2}\right\}\right] \\ &+ \log\left\{\frac{1}{2} + \frac{\xi_n}{2} + \frac{1}{2}\left(1 + \frac{n}{n+2}\xi_n^2\right)^{1/2}\right\} \\ &+ \log\left\{\frac{1}{2} - \frac{\xi_n}{2} + \frac{1}{2}\left(1 + \frac{n}{n+2}\xi_n^2\right)^{1/2}\right\}\Bigg). \end{aligned} \tag{5}$$

The proof of (5) is given in the Appendix. Clearly, $W(\mu_0, k_n)$ does not involve S^{-1} and can easily be calculated without involving numerical optimization.

Remark 1. The traditional empirical likelihood cannot be defined when $p \geqslant n$ because the matrix $(x_1, \ldots, x_n)$ is of full rank n. This implies that the solution to $\sum_{i=1}^{n}\omega_i(x_i - \mu_0) = 0$ is $\omega_1 = \cdots = \omega_n = 0$. However, we provide a feasible empirical likelihood in (5) for the $p \geqslant n$ case. The proof of (5) shows that the probability weight of each observation is the same, but nonzero. It is interesting to study data-driven probability weights in the case where $p \geqslant n$.

2.3. *Empirical likelihood-based test statistic*

As shown in Lemma A1 in the Appendix, the asymptotic mean and variance of $2nl_n^2(n+2)^{-2}W(\mu_0, k_n)$ are $\mathrm{tr}(\Omega)$ and $2\,\mathrm{tr}(\Omega^2)$, respectively, where $\Omega = \Sigma^{1/2}(I_p + k_n\alpha\alpha^{\mathrm{T}})\Sigma^{1/2}$

and I_p is the $p \times p$ identity matrix. Let $\hat{\mathrm{tr}}(\Omega^s)$ denote an estimator of $\mathrm{tr}(\Omega^s)$ for $s = 1$ and 2. Define an empirical likelihood-based test statistic as

$$T_n = \{2\,\hat{\mathrm{tr}}(\Omega^2)\}^{-1/2}\left\{\frac{2nl_n^2}{(n+2)^2}\,W(\mu_0, k_n) - \hat{\mathrm{tr}}(\Omega)\right\}. \tag{6}$$

Next we introduce an estimator of $\mathrm{tr}(\Omega^s)$ for $s = 1$ and 2. Note that

$$\mathrm{tr}(\Omega) = \mathrm{tr}(\Sigma) + k_n\alpha^{\mathrm{T}}\Sigma\alpha, \quad \mathrm{tr}(\Omega^2) = \mathrm{tr}(\Sigma^2) + 2k_n\alpha^{\mathrm{T}}\Sigma^2\alpha + k_n^2(\alpha^{\mathrm{T}}\Sigma\alpha)^2.$$

Let $P_n^r = n!/(n-r)!$. Following Chen et al. (2010),

$$\hat{\mathrm{tr}}(\Sigma) = \frac{1}{n}\sum_{i=1}^{n} x_i^{\mathrm{T}}x_i - \frac{1}{P_n^2}\sum_{i \neq j} x_i^{\mathrm{T}}x_j, \tag{7}$$

$$\hat{\mathrm{tr}}(\Sigma^2) = \frac{1}{P_n^2}\sum_{i \neq j}(x_i^{\mathrm{T}}x_j)^2 - \frac{2}{P_n^3}\sum_{i,j,k}^{*} x_i^{\mathrm{T}}x_j x_j^{\mathrm{T}}x_k + \frac{1}{P_n^4}\sum_{i,j,k,l}^{*} x_i^{\mathrm{T}}x_j x_k^{\mathrm{T}}x_l \tag{8}$$

are unbiased estimators of $\mathrm{tr}(\Sigma)$ and $\mathrm{tr}(\Sigma^2)$, respectively, where $\sum^*$ denotes summation over pairwise different indices; for example, $\sum_{i,j,k}^*$ means summation over $\{(i,j,k) : i \neq j, j \neq k, k \neq i\}$.

By using (7) and (8), $\mathrm{tr}(\Omega)$ and $\mathrm{tr}(\Omega^2)$ can be estimated by the following two estimators:

$$\hat{\mathrm{tr}}(\Omega) = \hat{\mathrm{tr}}(\Sigma) + k_n\alpha^{\mathrm{T}}S\alpha, \quad \hat{\mathrm{tr}}(\Omega^2) = \hat{\mathrm{tr}}(\Sigma^2) + 2k_n\alpha^{\mathrm{T}}S^2\alpha + k_n^2(\alpha^{\mathrm{T}}S\alpha)^2.$$

Under the assumptions of Theorem 1, $\hat{\mathrm{tr}}(\Omega^2)/\mathrm{tr}(\Omega^2) - 1 \to 0$ in probability as $n \to \infty$. This result is sufficient for us to derive the central limit theorem for T_n.

Theorem 1 below establishes the asymptotic distribution of T_n under H_0 and under the local alternative

$$H_{\mathrm{a}} : \mu = \mu_0 + n^{-1/2}\delta u, \quad |\delta| \leqslant C, \ \|u\| = 1, \tag{9}$$

where C is a constant independent of p.

THEOREM 1. *Suppose that Assumptions 1 and 2 hold and that there exist two positive constants c_0 and C_0 such that $c_0 \leqslant \lambda_{\min}(\Sigma) \leqslant \lambda_{\max}(\Sigma) < C_0$, where $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ stand for the smallest and largest eigenvalues of a matrix A, respectively. Further assume that Σ satisfies either of the following two conditions:*

Condition 1. Σ is a diagonal matrix;

Condition 2. there exists a constant $b_1^2 > 0$ such that $1 - \mathrm{tr}(\Sigma \circ \Sigma)/\mathrm{tr}(\Sigma^2) \geqslant b_1^2$, where $C \circ C = (c_{ij}b_{ij})$ for a matrix $C = (c_{ij})$.

Assume additionally that as $n \to \infty$, $l_n^{-1}n^{5/4} \to 0$, $k_n \to \infty$ and $k_n/\sqrt{p} \to 0$. Suppose that α is a p-dimensional vector with $\|\alpha\| = 1$. Then:

(i) *under H_0, $T_n \to N(0,1)$ in distribution as $n \to \infty$;*

(ii) *under* H_{a} *in* (9),

$$T_n - \frac{n\|\mu - \mu_0\|^2 + nk_n|\alpha^{\mathrm{T}}(\mu - \mu_0)|^2}{\{2\operatorname{tr}(\Omega^2)\}^{1/2}} \to N(0,1)$$

in distribution.

The proof of Theorem 1 is given in the Appendix. If the covariance matrix Σ satisfies Condition 1, it leads to a trivial situation where all components of x are independent. Let us consider the autoregressive correlation matrix $\Sigma = (\rho^{|i-j|})_{p\times p}$ for some $\rho \in (-1,1)$. It can be shown after some algebra that $\operatorname{tr}(\Sigma^2) = p/(1-\rho^2) + \rho^2(\rho^{2p}-1)/(1-\rho^2)^2$, $\operatorname{tr}(\Sigma \circ \Sigma) = p$ and

$$1 - \operatorname{tr}(\Sigma \circ \Sigma)/\operatorname{tr}(\Sigma^2) = \frac{\rho^2 p + \rho^2(\rho^{2p}-1)/(1-\rho^2)}{p + \rho^2(\rho^{2p}-1)/(1-\rho^2)} \geqslant \rho^2.$$

Hence Σ satisfies Condition 2. Let $\sigma_i^2 = \operatorname{var}(x_i)$ and consider another covariance matrix

$$\Sigma = (\sigma_i\sigma_j\rho^{|i-j|})_{p\times p}.$$

If the values of $\{\sigma_1^2,\ldots,\sigma_p^2\}$ are uniformly bounded away from zero and infinity, then using the above result for the correlation matrix it can be shown that Condition 2 is also satisfied. In Theorem 1, we require $l_n \to \infty$ along with the condition that $n^{5/4}/l_n \to 0$. In practice, we can choose $l_n = n^{5/4}\log(n)$.

Accordingly, the asymptotic power of T_n under local alternative (9) is

$$\beta_{\mathrm{EL}}(\mu) = \Phi\left[-z_{1-\alpha_0} + \frac{(1 + k_n|\alpha^{\mathrm{T}}u|^2)\delta^2}{\{2\operatorname{tr}(\Omega^2)\}^{1/2}}\right], \tag{10}$$

where $\Phi(x)$ denotes the standard normal cumulative distribution function and $z_{1-\alpha_0}$ its $(1-\alpha_0)$-quantile. Notice that all the eigenvalues of Σ are between two positive absolute constants c_0 and C_0, and that $\operatorname{tr}(\Sigma) = O(p)$ and $\operatorname{tr}(\Sigma^2) = O(p)$. Thus the term $\delta^2/\{2\operatorname{tr}(\Sigma^2)\}^{1/2}$ is of order $O(p^{-1/2})$, and the order of $(1 + k_n|\alpha^{\mathrm{T}}u|^2)\delta^2/\{2\operatorname{tr}(\Omega^2)\}^{1/2}$ in (10) is $O(p^{-1/2} + p^{-1/2}k_n|\alpha^{\mathrm{T}}u|^2)$, which reduces to $O(p^{-1/2}k_n)$ when $|\alpha^{\mathrm{T}}u| \neq 0$. In this situation, the test statistic T_n is much more powerful than T_{BS}, the test statistic proposed by Bai & Sarandasa (1996). However, when $|\alpha^{\mathrm{T}}u| = 0$ or $k_n = 0$, T_n and T_{BS} have the same asymptotic power.

Remark 2. In practice, it is important to find the α in (6) such that the corresponding asymptotic power in (10) attains its maximum value. Some calculations give that

$$\beta_{\mathrm{EL}}(\mu) = \Phi\left[-z_{1-\alpha_0} + p^{-1/2}\{(1 + k_n|\alpha^{\mathrm{T}}u|^2)\delta^2\}\{1 + o(1)\}\right].$$

Therefore one can choose $\alpha = u$ to achieve the approximate optimal test power. One feasible choice is $\alpha = 1_p/\sqrt{p}$, obtained by treating all elements of u as being equal. Alternatively, one could choose α based on prior knowledge.

Peng et al. (2014) proposed a large-dimensional empirical likelihood for testing a mean vector when $p = o[n^{\{\delta+\min(\delta,2)\}/\{2(2+\delta)\}}]$, which means $p < n^{1/2}$. They split the sample into two parts and applied the empirical likelihood method to two equations, $\sum_{i=1}^{n/2}(x_{1i} - \mu_0)^{\mathrm{T}}(x_{2i} - \mu_0) = 0$ and $\sum_{i=1}^{n/2} 1_p^{\mathrm{T}}(x_{1i} + x_{2i} - 2\mu_0) = 0$, where x_1 and x_2 are independent and identically distributed random vectors with mean μ and 1_p denotes a p-dimensional vector with all elements equal to 1. Peng et al. (2014) used the second equation to enhance test power, while we use a weighted linear combination of $\|\bar{x}-\mu_0\|^2$ and $|\alpha^{\mathrm{T}}(\bar{x}-\mu_0)|^2$. An interesting connection is that the term $1^{\mathrm{T}}(\bar{x}-\mu_0)$ is the key to enhancing test power.

2.4. *Test of the linear hypothesis*

We now study the linear hypothesis about μ. Consider

$$H_0 : F\mu = \mu_0 \quad \text{versus} \quad H_{\mathrm{a}} : F\mu \neq \mu_0,$$

where F is a $q \times p$ $(q > n)$ matrix not depending on the data or μ and of full rank $q \leqslant p$.

To test the linear hypothesis $H_0 : F\mu = \mu_0$, we can use the empirical likelihood given in (6) by taking

$$\tilde{a}_n = \frac{l_n}{\{\|F\bar{x}-\mu_0\|^2 + k_n|\gamma^{\mathrm{T}}(F\bar{x}-\mu_0)|^2\}^{1/2}},$$

where γ is a known q-dimensional vector with $\|\gamma\| = 1$. The resulting log empirical likelihood statistic is denoted by $W^F(\mu_0, k_n)$. We have the following analogue of the empirical likelihood test statistic (6):

$$T_n^F = \{2\,\hat{\mathrm{tr}}(\Lambda^2)\}^{-1/2}\left\{\frac{2nl_n^2}{(n+2)^2}\,W^F(\mu_0,k_n) - \hat{\mathrm{tr}}(\Lambda)\right\},$$

with $\Lambda = \Gamma^{\mathrm{T}}F^{\mathrm{T}}(I_q + k_n\gamma\gamma^{\mathrm{T}})F\Gamma$. Here

$$\hat{\mathrm{tr}}(\Lambda) = \mathrm{tr}\{S_nF^{\mathrm{T}}(I_q + k_n\gamma\gamma^{\mathrm{T}})F\}, \quad \hat{\mathrm{tr}}(\Lambda^2) = \mathrm{tr}[\{S_nF^{\mathrm{T}}(I_q + k_n\gamma\gamma^{\mathrm{T}})F\}^2],$$

which are estimators of $\mathrm{tr}(\Lambda)$ and $\mathrm{tr}(\Lambda^2)$, respectively, as $\lambda_{\max}(F^{\mathrm{T}}F) = O(p^2)$.

If the assumptions of Theorem 1 hold, then from the proof of Theorem 1 we have that under H_0, $T_n^F \to N(0,1)$ in distribution as $n \to \infty$, which can be used to test the hypothesis in the F matrix.

3. Numerical studies

3.1. *Preliminaries*

This section is devoted to assessing the finite-sample performance of the proposed empirical likelihood-based test. We first conduct Monte Carlo simulations to compare the performance of the proposed empirical likelihood based-test with the test of Chen & Qin (2010), denoted by T_{CQ}, and the test of Wang et al. (2015), denoted by T_{WPL}. To save space, in this section we present simulation results only for H_0; results for the linear hypothesis and the real-data example are reported in the Supplementary Material.

In our simulations, the data were generated from $x = \mu + \Sigma^{1/2}z$, where $z = (z_1,\ldots,z_p)^{\mathrm{T}}$ has independent and identically distributed elements with mean 0 and variance 1. We consider three scenarios for z_j: (I) $z_j \sim N(0,1)$; (II) $z_j \sim \{\mathrm{Ga}(4,2) - 2\}$, where $\mathrm{Ga}(4,2)$ denotes the gamma distribution with shape parameter 4 and rate parameter 2; and (III) $z_j \sim (3/5)^{1/2}t(5)$, where $t(5)$ is the t distribution with five degrees of freedom. Throughout the simulations we set $\Sigma = (\sigma_{ij})$ with $\sigma_{ij} = 0.5^{|i-j|}$.

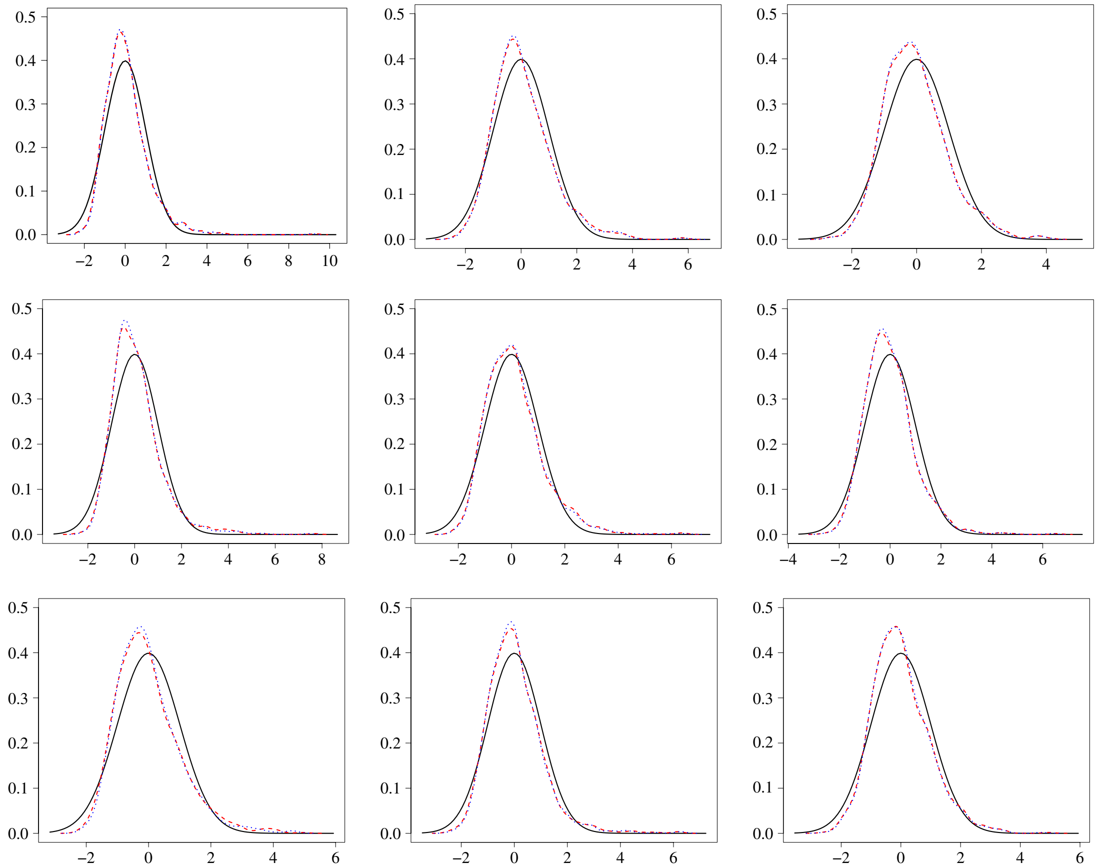

Fig. 1. Estimated density functions of T_n and T_n^0 under H_0: the solid, dotted and dashed curves are the estimated density curves of $N(0,1)$, T_n and T_n^0, respectively; the top, middle and bottom rows correspond to $z_j \sim N(0,1)$, $z_j \sim \mathrm{Ga}(4,2)-2$ and $z_j \sim (3/5)^{1/2}t(5)$, respectively; and the left, middle and right columns correspond to $(n, p/n) = (200, 1.5)$, $(400, 1.5)$ and $(800, 1.5)$, respectively.

3.2. *Limiting null distribution*

We first examine the impact of the plug-in estimate of the asymptotic mean and variance of $W_n(\mu_0, k_n)$ given in § 2.2, and study how close the finite-sample distribution of T_n is to the standard normal distribution. To this end, we define

$$T_n^0 = \{2\,\mathrm{tr}(\Omega^2)\}^{-1/2}\left\{\frac{2nl_n^2}{(n+2)^2}\,W(\mu_0, k_n) - \mathrm{tr}(\Omega)\right\},$$

which is the pre-plug-in version of T_n with $\mu_0 = 0$. We take $l_n = n^{5/4}\log(n)$, $k_n = (p/\log p)^{1/2}$ and $\alpha = (1,\ldots,1)^{\mathrm{T}}/\sqrt{p}$ in this simulation. We consider the sample sizes $(n,p) = (200, 300)$, $(400, 600)$ and $(800, 1200)$, such that $c_n = p/n = 1.5$. For each case we conduct 1000 simulations, based on which we obtain the kernel density estimates of T_n and T_n^0. Figure 1 shows the curves of the kernel density estimates along with the density curve of $N(0,1)$. It can be seen that the estimated density curves of T_n and T_n^0 are very close. This implies that the estimation error of the asymptotic mean and variance of $W_n(\mu_0, k_n)$ does not have a significant impact on the asymptotic distribution of T_n. The tails of these estimated density curves are close to those of $N(0,1)$. This means that the percentile of the limiting null distribution can serve as the critical value of the

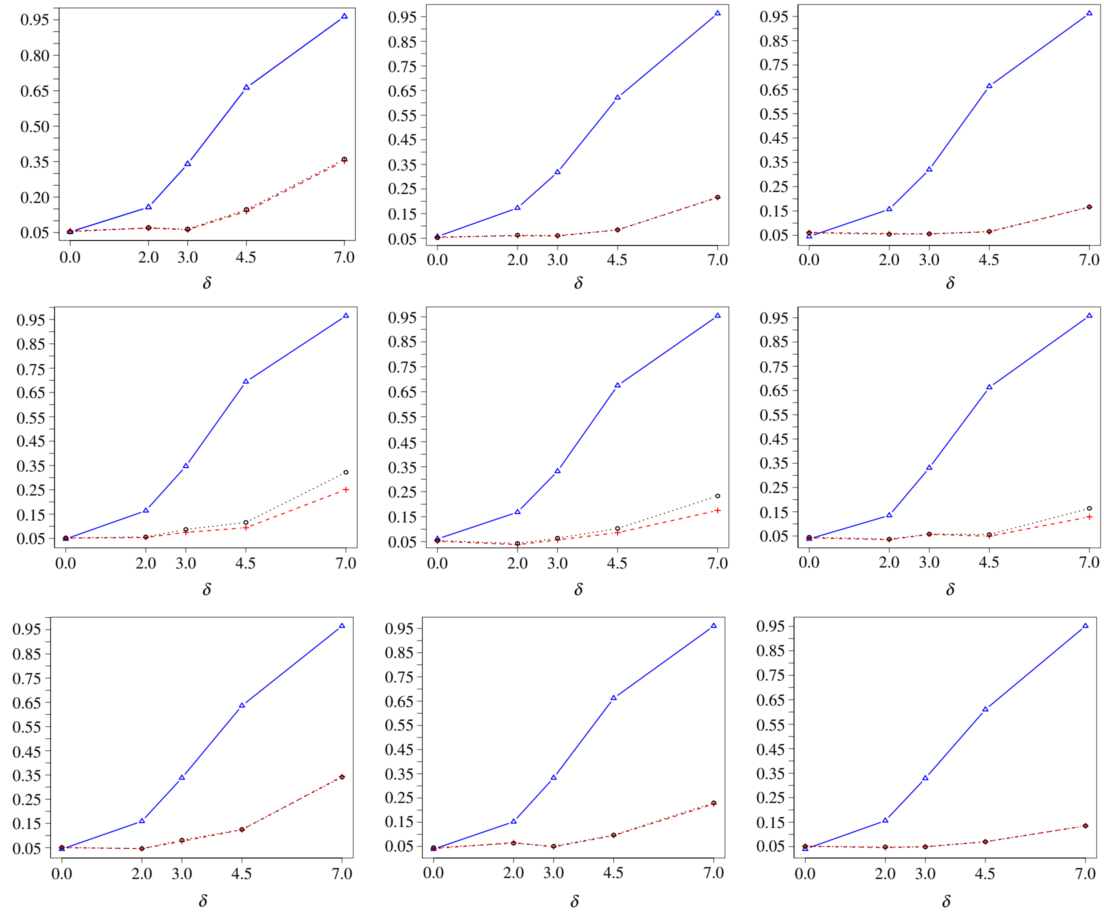

Fig. 2. Empirical power functions of T_n, T_{CQ} and T_{WPL} with $p/n = 1.5$: the solid, dotted and dashed curves are the empirical power curves of T_n, T_{CQ} and T_{WPL}, respectively; the top, middle and bottom rows correspond to $z_j \sim N(0, 1)$, $z_j \sim \mathrm{Ga}(4, 2) - 2$ and $z_j \sim (3/5)^{1/2} t(5)$, respectively; and the left, middle and right columns correspond to $n = 200$, 400 and 800, respectively.

proposed test. The overall patterns of the density plots are similar across different combinations of n and p, indicating that T_n is robust to the population dimensionality when c_n remains the same.

3.3. *Type I error rate and local power*

We examine the Type I error rate and local power of the proposed test. We compare the performance of the proposed empirical likelihood-based test with the test of Chen & Qin (2010), denoted by T_{CQ}, and the test of Wang et al. (2015), denoted by T_{WPL}. We set $\mu = \delta(2, 1, \ldots, 1)^{\mathrm{T}}/\sqrt{n}$ with $\delta = 0, 2, 3, 4.5, 7$. Thus we can examine the Type I error rate when $\delta = 0$ and the power when $\delta \neq 0$. As in § 3.2 we set $l_n = n^{5/4} \log n$, $k_n = (p/\log p)^{1/2}$ and $\alpha = (1, \ldots, 1)^{\mathrm{T}}/\sqrt{p}$. Figure 2 plots the rejection rates over 1000 simulations, i.e., the empirical power curves of T_n, T_{CQ} and T_{WPL} under the $p/n = 1.5$ setting. All three tests are observed to maintain the Type I error rate quite well. When $z_j \sim N(0, 1)$, the power curves are shown in the top row of Fig. 2, from which it can be seen that (i) T_n has more power than T_{CQ} and T_{WPL} for this local alternative; and (ii) the power curves of T_{CQ} and T_{WPL} are very close, with no visible difference between them. This is expected from the theoretical analyses of Chen & Qin (2010) and Wang et al. (2015). The

Table 1. *Empirical power of T_n with the choices* (c1), (c2), *and* (c3), *where $\alpha = e_i$ when the ith element of $|\bar{x} - \mu_0|$ is the largest, α is a random direction with $\|\alpha\| = 1$, and $\alpha = 1_p/\sqrt{p}$, respectively, while keeping $l_n = n^{5/4}\log n$ and $k_n = (p/\log p)^{1/2}$*

	$z_j \sim N(0,1)$			$z_j \sim \mathrm{Ga}(4,2) - 2$			$z_j \sim (3/5)^{1/2}t(5)$		
δ	(c1)	(c2)	(c3)	(c1)	(c2)	(c3)	(c1)	(c2)	(c3)
0	0.169	0.023	0.062	0.199	0.026	0.036	0.186	0.032	0.048
2	0.194	0.112	0.153	0.211	0.142	0.164	0.199	0.135	0.166
3	0.237	0.242	0.331	0.251	0.262	0.343	0.233	0.258	0.308
4.5	0.305	0.517	0.616	0.309	0.528	0.644	0.341	0.556	0.646
7	0.521	0.922	0.958	0.538	0.924	0.961	0.506	0.912	0.956

middle row of Fig. 2 displays the power curves for $z_j \sim \mathrm{Ga}(4,2) - 2$ and indicates that T_n is more powerful than T_{CQ}, which is more powerful than T_{WPL}. The bottom row of Fig. 2 shows the power curves for $z_j \sim (3/5)^{1/2}t(5)$; these exhibit the same pattern as in the top and middle rows, indicating that T_n is more powerful than the other two tests. In the Supplementary Material we report results of the simulation under the setting of $p/n = 1.2$.

3.4. *Sensitivity to α, l_n and k_n*

We conduct some simulations to examine how sensitive the test is to the choices of α, l_n and k_n. We mainly focus on the setting of $n = 400$ and $p/n = 1.5$. First, we keep l_n and k_n fixed as in § 3.3 and change α. That is, we consider the following choices:

(c1) $l_n = n^{5/4}\log n$, $k_n = (p/\log p)^{1/2}$ and $\alpha = e_i$ when the ith element of $|\bar{x} - \mu_0|$ is the largest;
(c2) $l_n = n^{5/4}\log n$, $k_n = (p/\log p)^{1/2}$ and α is a uniformly distributed random direction with $\|\alpha\| = 1$;
(c3) $l_n = n^{5/4}\log n$, $k_n = (p/\log p)^{1/2}$ and $\alpha = 1_p/\sqrt{p}$.

The choices of l_n, k_n and α in (c3) are the same as in § 3.3. Table 1 shows the results, which indicate that the choice (c1) fails to maintain the Type I error rate because it depends on the sample. The choices of α in (c2) and (c3) maintain the Type I error rate well, with the α in (c3) performing slightly better than that in (c2).

Next, we examine the impact of different choices of l_n, by considering the following scenarios:

(d1) $l_n = 0.5n^{5/4}\log(n)$, $k_n = (p/\log p)^{1/2}$ and $\alpha = 1_p/\sqrt{p}$;
(d2) $l_n = 2n^{5/4}\log(n)$, $k_n = (p/\log p)^{1/2}$ and $\alpha = 1_p/\sqrt{p}$.

Table 2 shows that the results are robust with respect to the choice of l_n.

Third, we keep α and l_n fixed as in § 2.3 and take the following:

(e1) $l_n = n^{5/4}\log n$, $k_n = 0.5(p/\log p)^{1/2}$ and $\alpha = 1_p/\sqrt{p}$;
(e2) $l_n = n^{5/4}\log n$, $k_n = 2(p/\log p)^{1/2}$ and $\alpha = 1_p/\sqrt{p}$.

The results are given in Table 3, from which it can be seen that the performance of the proposed test is robust to these choices of k_n.

Acknowledgement

The authors thank the editor, associate editor and two reviewers for their constructive comments, as well as Ms A. Applegate for her help with the writing. This work was supported by

Table 2. *Empirical power of T_n with the choices* (d1) *and* (d2)*, where $l_n = 0.5n^{5/4}\log n$ and $l_n = 2n^{5/4}\log n$, respectively, while keeping $\alpha = 1_p/\sqrt{p}$ and $k_n = (p/\log p)^{1/2}$*

	$z_j \sim N(0,1)$			$z_j \sim \text{Ga}(4,2)-2$			$z_j \sim (3/5)^{1/2}t(5)$		
δ	(d1)	(d2)	(c3)	(d1)	(d2)	(c3)	(d1)	(d2)	(c3)
0	0.041	0.033	0.062	0.058	0.049	0.036	0.046	0.051	0.048
2	0.163	0.147	0.153	0.169	0.149	0.164	0.156	0.156	0.166
3	0.326	0.316	0.331	0.330	0.341	0.343	0.336	0.375	0.308
4.5	0.632	0.626	0.616	0.656	0.632	0.644	0.656	0.656	0.646
7	0.957	0.951	0.958	0.976	0.953	0.961	0.957	0.966	0.956

Table 3. *Empirical power of T_n with the choices* (e1) *and* (e2)*, where $k_n = 0.5(p/\log p)^{1/2}$ and $k_n = 2(p/\log p)^{1/2}$, respectively, while keeping $\alpha = 1_p/\sqrt{p}$ and $l_n = n^{5/4}\log n$*

	$z_j \sim N(0,1)$			$z_j \sim \text{Ga}(4,2)-2$			$z_j \sim (3/5)^{1/2}t(5)$		
δ	e1	e2	c3	e1	e2	c3	e1	e2	c3
0	0.042	0.051	0.062	0.038	0.028	0.036	0.053	0.047	0.048
2	0.131	0.198	0.153	0.118	0.168	0.164	0.120	0.174	0.166
3	0.243	0.382	0.331	0.232	0.378	0.343	0.226	0.387	0.308
4.5	0.522	0.708	0.616	0.510	0.717	0.644	0.502	0.689	0.646
7	0.918	0.971	0.958	0.895	0.976	0.961	0.897	0.980	0.956

the National Natural Science Foundation of China, the U.S. National Institutes of Health and National Science Foundation, the National Social Science Foundation of China, Fundamental Research Funds for the Central University, and the Ministry of Education Tier One grant at the National University of Singapore.

Supplementary material

Supplementary material available at *Biometrika* online includes the proof of Lemma A2, additional simulation results and a real-data example.

Appendix

Technical proofs

Proof of equation (5). Write the condition $\sum_{i=1}^{n+2}\omega_i(x_i-\mu_0)=0_p$ as

$$0=\sum_{i=1}^{n}\omega_i(x_i-\mu_0)-a_n\omega_{n+1}(\bar{x}-\mu_0)+\omega_{n+2}(2+a_n)(\bar{x}-\mu_0)$$
$$=\sum_{i=1}^{n}\{\omega_i-a_n\omega_{n+1}/n+(2+a_n)\omega_{n+2}/n\}(x_i-\mu_0)=\Sigma^{1/2}Zw,$$

where $Z=(z_1,\ldots,z_n)$ is a $p\times n$ matrix and w is an $n\times 1$ vector whose ith element is $\omega_i-a_n\omega_{n+1}/n+(2+a_n)\omega_{n+2}/n$. From Bai & Yin (1993), if $p/n\to c>1$ then we know that $Z^{\mathrm{T}}Z/p$ is an $n\times n$ matrix whose smallest eigenvalue converges to $(1-c^{-1/2})^2$ almost surely. That is, $\lambda_{\min}(Z^{\mathrm{T}}Z/p)\to(1-c^{-1/2})^2$ almost surely. Hence

$$\lambda_{\min}(Z^{\mathrm{T}}\Sigma Z/p)>\lambda_{\min}(\Sigma)\,\lambda_{\min}(Z^{\mathrm{T}}Z/p)>0$$

almost surely. If $p = n$, by Remark 4 in Pan & Zhou (2010) we know that $\lambda_{\min}(Z^{\mathrm{T}}Z/p) > 0$ almost surely. So $Z^{\mathrm{T}}\Sigma Z/p$ is nonsingular almost surely. Multiplying $0 = \Sigma^{1/2}Zw$ by $(Z^{\mathrm{T}}\Sigma Z/p)^{-1}Z^{\mathrm{T}}\Sigma^{1/2}/p$ on both sides of the equality gives $w = 0_n$. This implies that

$$\omega_i = \frac{a_n}{n}\omega_{n+1} - \frac{a_n+2}{n}\omega_{n+2} \quad (i = 1, \ldots, n).$$

On the other hand, from the condition $\sum_{i=1}^{n+2}\omega_i = 1$ we obtain

$$\begin{cases} \omega_i = \dfrac{a_n}{n(1+a_n)} - \dfrac{2}{n}\omega_{n+2} \quad (i = 1, \ldots, n), \\ \omega_{n+1} = \dfrac{1}{1+a_n} + \omega_{n+2}. \end{cases}$$

Hence it follows that

$$\begin{aligned} W(\mu_0, k_n) = -2\max_{\omega_{n+2}}\Bigg[& n\log\left\{\frac{(n+2)a_n}{n(1+a_n)} - \frac{2(n+2)\omega_{n+2}}{n}\right\} \\ & + \log\left\{\frac{n+2}{1+a_n} + (n+2)\omega_{n+2}\right\} + \log\{(n+2)\omega_{n+2}\}\Bigg]. \end{aligned}$$

It can be shown that the solution to the above maximization problem is

$$\omega_0 = \frac{1}{2(n+2)}\left\{1 - \xi_n + \left(1 + \frac{n}{n+2}\xi_n^2\right)^{1/2}\right\},$$

where $\xi_n = (n+2)/(1+a_n)$. Hence, the explicit expression for $W(\mu_0, k_n)$ is

$$\begin{aligned} W(\mu_0, k_n) = -2\Bigg(& n\log\left[1 + \frac{1}{n}\left\{1 - \left(1+\frac{n}{n+2}\xi_n^2\right)^{1/2}\right\}\right] + \log\left\{\frac{1}{2} + \frac{\xi_n}{2} + \frac{1}{2}\left(1+\frac{n}{n+2}\xi_n^2\right)^{1/2}\right\} \\ & + \log\left\{\frac{1}{2} - \frac{\xi_n}{2} + \frac{1}{2}\left(1 + \frac{n}{n+2}\xi_n^2\right)^{1/2}\right\}\Bigg), \end{aligned}$$

and thus the proof of (5) is completed. □

We need the following two lemmas to prove Theorem 1. To facilitate the proof, the lemmas start with the situation in which a_n is set to $l_n/\|\bar{x} - \mu_0\|$.

LEMMA A1. *Suppose that Condition 1 or Condition 2 of Theorem 1 holds, $a_n = l_n/\|\bar{x} - \mu_0\|$ and $\xi_n = (n+2)/(1+a_n)$. Then*

$$\frac{2nl_n^2(n+2)^{-2}W(\mu_0, k_n) - \mathrm{tr}(\Sigma)}{\{2\,\mathrm{tr}(\Sigma^2)\}^{1/2}} \to N(0, 1)$$

in distribution as $n \to \infty$.

Proof. Condition 1 implies that all components of x are independent; the proof is trivial and is omitted. Suppose that Condition 2 holds. It follows from the assumption $c_0 \leqslant \lambda_{\min}(\Sigma) \leqslant \lambda_{\max}(\Sigma) \leqslant C_0$ that

$c_0 p \leqslant \mathrm{tr}(\Sigma) \leqslant C_0 p$ and $c_0^2 p \leqslant \mathrm{tr}(\Sigma^2) \leqslant C_0^2 p$. Hence $\mathrm{tr}(\Sigma) = O(p)$ and $\mathrm{tr}(\Sigma^2) = O(p)$. We next show that $\xi_n^2 = o_{\mathrm{p}}(1)$. By the definition of ξ_n, we have that

$$\begin{aligned}\xi_n^2 &\leqslant \frac{(n+2)^2}{l_n^2}\|\bar{x} - \mu_0\|^2 \\ &\leqslant \frac{(n+2)^2\{2\,\mathrm{tr}(\Sigma^2)\}^{1/2}}{nl_n^2}\left|\frac{n\|\bar{x} - \mu_0\|^2 - \mathrm{tr}(\Sigma)}{\{2\,\mathrm{tr}(\Sigma^2)\}^{1/2}}\right| + \frac{(n+2)^2\,\mathrm{tr}(\Sigma)}{nl_n^2} = o_{\mathrm{p}}(1),\end{aligned}$$

where the last equality holds because $\{n\|\bar{x} - \mu_0\|^2 - \mathrm{tr}(\Sigma)\}/\{2\,\mathrm{tr}(\Sigma^2)\}^{1/2} = O_{\mathrm{p}}(1)$, $p/n = O(1)$ and $n^{5/4}/l_n = o(1)$. Therefore, by Taylor's expansion,

$$\left(1 + \frac{n}{n+2}\xi_n^2\right)^{1/2} = 1 + \eta_n \tag{A1}$$

and

$$\eta_n = \frac{n}{2(n+2)}\xi_n^2 - \frac{n^2}{8(n+2)^2}(1+h_1)^{-3/2}\xi_n^4, \quad 0 \leqslant h_1 \leqslant \xi_n^2. \tag{A2}$$

The proof is then divided into three steps.

Step 1. We first show that

$$W(\mu_0, k_n) = \xi_n^2/2 + O_{\mathrm{p}}(\xi_n^4). \tag{A3}$$

Define $L_1 = -2n\log(1 - n^{-1}\eta_n)$, $L_2 = -2\log(1 + \xi_n/2 + \eta_n/2)$ and $L_3 = -2\log(1 - \xi_n/2 + \eta_n/2)$. Then, substituting (A1) into (5) gives

$$W(\mu_0, k_n) = L_1 + L_2 + L_3. \tag{A4}$$

Using Taylor's expansion for L_j, we obtain

$$\begin{aligned}L_1 &= 2\eta_n + L_1^R, \\ L_2 &= -(\xi_n + \eta_n) + \frac{1}{4}(\xi_n + \eta_n)^2 - \frac{1}{12}(\xi_n + \eta_n)^3 + L_2^R, \\ L_3 &= -(-\xi_n + \eta_n) + \frac{1}{4}(-\xi_n + \eta_n)^2 - \frac{1}{12}(-\xi_n + \eta_n)^3 + L_3^R,\end{aligned}$$

where

$$\begin{aligned}L_1^R &= -\frac{\eta_n^2}{n(h_2-1)^2}, \quad |h_2| \leqslant \left|\frac{\eta_n}{n}\right|, \\ L_2^R &= \frac{(\xi_n + \eta_n)^4}{32(1+h_3)^4}, \quad |h_3| \leqslant \left|\frac{\xi_n + \eta_n}{2}\right|, \\ L_3^R &= \frac{(-\xi_n + \eta_n)^4}{32(1+h_4)^4}, \quad |h_4| \leqslant \left|\frac{-\xi_n + \eta_n}{2}\right|.\end{aligned}$$

Then, by (A2), $\xi_n^2 = o_{\mathrm{p}}(1)$ and the fact that $\eta_n = O_{\mathrm{p}}(\xi_n^2) = o_{\mathrm{p}}(1)$, we have

$$L_1^R = O_{\mathrm{p}}\left(\frac{\xi_n^4}{n}\right), \quad L_2^R = O_{\mathrm{p}}(\xi_n^4), \quad L_3^R = O_{\mathrm{p}}(\xi_n^4). \tag{A5}$$

It follows from (A4) and (A5) that

$$W(\mu_0, k_n) = \xi_n^2/2 + \eta_n^2/2 - \xi_n^2\eta_n/6 - \eta_n^3/2 + O_{\rm p}(\xi_n^4) = \xi_n^2/2 + O_{\rm p}(\xi_n^4).$$

Step 2. We show that

$$\{2\,{\rm tr}(\Sigma^2)\}^{-1/2}\left\{\frac{nl_n^2}{(n+2)^2}\,\xi_n^2 - {\rm tr}(\Sigma)\right\} \to N(0,1) \tag{A6}$$

in distribution. Note that

$$\frac{nl_n^2}{(n+2)^2}\,\xi_n^2 = n\|\bar{x}-\mu_0\|^2 + \left\{\frac{l_n^2}{(l_n+\|\bar{x}-\mu_0\|)^2} - 1\right\} n\|\bar{x}-\mu_0\|^2.$$

According to Lemma S1, it suffices to prove

$$\{2\,{\rm tr}(\Sigma^2)\}^{-1/2}\left\{\frac{l_n^2}{(l_n+\|\bar{x}-\mu_0\|)^2} - 1\right\} n\|\bar{x}-\mu_0\|^2 = o_{\rm p}(1).$$

From Lemma S1, $\|\bar{x}-\mu_0\| = O_{\rm p}(1)$ as $n\to\infty$. So

$$\left|\frac{l_n^2}{(l_n+\|\bar{x}-\mu_0\|)^2} - 1\right| = O_{\rm p}(l_n^{-1}).$$

Since ${\rm tr}(\Sigma^2) \geqslant c_0^2 p$ and $l_n^{-1}n^{5/4}\to 0$,

$$\{2\,{\rm tr}(\Sigma^2)\}^{-1/2}\left\{\frac{l_n^2}{(l_n+\|\bar{x}-\mu\|)^2} - 1\right\} n\|\bar{x}-\mu_0\|^2 = O_{\rm p}\{n/(pl_n)\} = o_{\rm p}(1).$$

Step 3. We show that

$$\{2\,{\rm tr}(\Sigma^2)\}^{-1/2}\left\{\frac{2nl_n^2}{(n+2)^2}W(\mu_0,k_n) - {\rm tr}(\Sigma)\right\} \to N(0,1)$$

in distribution. From (A3) it follows after some calculation that

$$\{2\,{\rm tr}(\Sigma^2)\}^{-1/2}\left\{\frac{2nl_n^2}{(n+2)^2}\,W(\mu_0,k_n) - {\rm tr}(\Sigma)\right\} = \{2\,{\rm tr}(\Sigma^2)\}^{-1/2}\left\{\frac{nl_n^2}{(n+2)^2}\,\xi_n^2 - {\rm tr}(\Sigma)\right\} + W^R,$$

where $W^R = O_{\rm p}[nl_n^2(n+2)^{-2}\{{\rm tr}(\Sigma^2)\}^{-1/2}\xi_n^4]$. To prove (A6), it suffices to show that

$$W^R = o_{\rm p}(1).$$

Since $\xi_n = (n+2)/(1+a_n)$,

$$\begin{aligned}
\frac{nl_n^2}{(n+2)^2\{2\,{\rm tr}(\Sigma^2)\}^{1/2}}\,\xi_n^4 &\leqslant \frac{nl_n^2}{(n+2)^2\{2\,{\rm tr}(\Sigma^2)\}^{1/2}}\,\frac{(n+2)^4}{a_n^4}\\
&= \frac{\{2\,{\rm tr}(\Sigma^2)\}^{1/2}(n+2)^2}{nl_n^2}\left[\frac{n\|\bar{x}-\mu\|^2}{\{2\,{\rm tr}(\Sigma^2)\}^{1/2}}\right]^2\\
&\leqslant \frac{\{2\,{\rm tr}(\Sigma^2)\}^{1/2}(n+2)^2}{nl_n^2}\left[\left|\frac{n\|\bar{x}-\mu\|^2-{\rm tr}(\Sigma)}{\{2\,{\rm tr}(\Sigma^2)\}^{1/2}}\right| + \frac{{\rm tr}(\Sigma)}{\{2\,{\rm tr}(\Sigma^2)\}^{1/2}}\right]^2\\
&= O_{\rm p}\left(\frac{n^{5/2}}{l_n^2}\right) = o_{\rm p}(1),
\end{aligned}$$

where the second-to-last equality follows from $\mathrm{tr}(\Sigma) = O(p)$, $\mathrm{tr}(\Sigma^2) = O(p)$ and $p/n = O(1)$, and the last equality follows from the condition that $n^{5/4}/l_n = o(1)$. This analysis implies that $W^R = o_{\mathrm{p}}(1)$.

The proof of Lemma A1 is thus completed. □

LEMMA A2. *Under the assumptions of Lemma A1 , it follows that*

$$\frac{2nl_n^2(n+2)^{-2}W(\mu_0, k_n) - \hat{\mathrm{tr}}(\Sigma)}{\{2\,\hat{\mathrm{tr}}(\Sigma^2)\}^{1/2}} \to N(0,1)$$

in distribution as $n \to \infty$.

The proof of Lemma A2 is given in the Supplementary Material.

Proof of Theorem 1. (i) Recall the notation $\mathrm{tr}(\Omega)$, $\mathrm{tr}(\Omega^2)$, $\hat{\mathrm{tr}}(\Omega)$ and $\hat{\mathrm{tr}}(\Omega^2)$ defined in § 2.3 and the assumption that $c_0 \leqslant \lambda_{\min}(\Sigma) \leqslant \lambda_{\max}(\Sigma) \leqslant C_0$. It follows that

$$c_0 p + k_n c_0 \leqslant \mathrm{tr}(\Omega) \leqslant C_0 p + C_0 k_n, \quad c_0^2 p + c_0^2 k_n + c_0^2 k_n^2 \leqslant \mathrm{tr}(\Omega^2) \leqslant C_0^2 p + C_0^2 k_n + C_0^2 k_n^2,$$

since $\|\alpha\| = 1$.

Since $k_n = o(\surd p)$, we have $\mathrm{tr}(\Omega) = O(p)$ and $\mathrm{tr}(\Omega^2) = O(p)$. Furthermore, from Assumptions 1 and 2 it follows that $\alpha^{\mathrm{T}} S\alpha - \alpha^{\mathrm{T}}\Sigma\alpha = o_{\mathrm{p}}(1)$ under $\|\alpha\| = 1$. Under the assumption that $c_0 \leqslant \lambda_{\min}(\Sigma) \leqslant \lambda_{\max}(\Sigma) \leqslant C_0$, we can show that if $\|\alpha\| = 1$, then $\alpha^{\mathrm{T}}\Sigma^2\alpha = O(1)$ and $\alpha^{\mathrm{T}} S^2\alpha = O_{\mathrm{p}}(1)$.

In the proof of Lemma A2 we showed that $\hat{\mathrm{tr}}(\Sigma) - \mathrm{tr}(\Sigma) = O_{\mathrm{p}}(1)$ and $\hat{\mathrm{tr}}(\Sigma^2) - \mathrm{tr}(\Sigma^2) = O_{\mathrm{p}}(1)$. Hence

$$\hat{\mathrm{tr}}(\Omega) - \mathrm{tr}(\Omega) = O_{\mathrm{p}}(1) + o_{\mathrm{p}}(k_n), \quad \hat{\mathrm{tr}}(\Omega^2) - \mathrm{tr}(\Omega^2) = O_{\mathrm{p}}(1) + O_{\mathrm{p}}(k_n) + o_{\mathrm{p}}(k_n^2). \tag{A7}$$

Using (A7) and the same strategy as in the proof of Lemma A2, we obtain

$$T_n = \left\{2\,\mathrm{tr}(\Omega^2)\right\}^{-1/2}\left\{\frac{2nl_n^2}{(n+2)^2}\,W(\mu_0,k_n) - \mathrm{tr}(\Sigma)\right\} + \left[\frac{\{2\,\mathrm{tr}(\Omega^2)\}^{1/2}}{\{2\,\hat{\mathrm{tr}}(\Omega^2)\}^{1/2}} - 1\right]\frac{2nl_n^2(n+2)^{-2}W(\mu_0,k_n) - \mathrm{tr}(\Omega)}{\{2\,\mathrm{tr}(\Omega^2)\}^{1/2}} + o_{\mathrm{p}}(1)$$

since $k_n = o(\surd p)$. Because $\{2\,\mathrm{tr}(\Omega^2)\}^{1/2}/\{2\,\hat{\mathrm{tr}}(\Omega^2)\}^{1/2} - 1 = o_{\mathrm{p}}(1)$, it is sufficient to prove that

$$\{2\,\mathrm{tr}(\Omega^2)\}^{-1/2}\left\{\frac{2nl_n^2}{(n+2)^2}\,W(\mu_0,k_n) - \mathrm{tr}(\Omega)\right\} \to N(0,1)$$

in distribution. We first show that $\xi_n = o_{\mathrm{p}}(1)$, which follows from the calculation

$$\xi_n^2 \leqslant \frac{(n+2)^2}{l_n^2}\left\{\|\bar{x} - \mu_0\|^2 + k_n|\alpha^{\mathrm{T}}(\bar{x} - \mu_0)|^2\right\} = o_{\mathrm{p}}(1)$$

since $|\alpha^{\mathrm{T}}(\bar{x} - \mu_0)| = O_{\mathrm{p}}(n^{-1/2})$. In fact, $\alpha^{\mathrm{T}}(\bar{x} - \mu_0) = O_{\mathrm{p}}([\mathrm{var}\{\alpha^{\mathrm{T}}(\bar{x} - \mu_0)\}]^{1/2}) = O_{\mathrm{p}}([\alpha^{\mathrm{T}}E\{(\bar{x} - \mu_0) \times (\bar{x} - \mu_0)^{\mathrm{T}}\}\alpha]^{1/2}) = O_{\mathrm{p}}\{(\alpha^{\mathrm{T}}n^{-1}\Sigma\alpha)^{1/2}\} = O_{\mathrm{p}}(n^{-1/2})$, $k_n = o(\surd p)$ and $n/l_n = o(1)$. By the same argument as in Step 1 of the proof of Lemma A1, we have that

$$W(\mu_0, k_n) = \xi_n^2 + O_{\mathrm{p}}(\xi_n^4).$$

Similarly to Step 3 of the proof of Lemma A1, it can be shown that under the assumption $n^{5/4}/l_n = o(1)$,

$$\frac{nl_n^2}{(n+2)^2}\,\xi_n^2 = o_{\rm p}(\surd p).$$

Then

$$\frac{2nl_n^2}{(n+2)^2}\,W(\mu_0,k_n) = \frac{nl_n^2}{(n+2)^2}\,\xi_n^2 + o_{\rm p}(\surd p).$$

So we only need to prove that

$$\{2\,{\rm tr}(\Omega^2)\}^{-1/2}\left\{\frac{nl_n^2}{(n+2)^2}\,\xi_n^2 - {\rm tr}(\Omega)\right\} \to N(0,1) \qquad \text{(A8)}$$

in distribution. The term on the left-hand side is

$$\{2\,{\rm tr}(\Omega^2)\}^{-1/2}\left\{n(\|\bar x-\mu_0\|^2 + k_n|\alpha^{\rm T}(\bar x-\mu_0)|^2) - {\rm tr}(\Omega)\right\} + o_{\rm p}(1),$$

since $\|\bar x-\mu_0\|^2 + k_n|\alpha^{\rm T}(\bar x-\mu_0)|^2 = O_{\rm p}(1)$ and $n^{5/4}/l_n = o(1)$. Note that $\lambda_{\max}(\Omega)/\{{\rm tr}(\Omega^2)\}^{1/2} = O(k_n/\surd p) = o(1)$ and $1-{\rm tr}(\Omega\circ\Omega)/{\rm tr}(\Omega^2) \geqslant b_1^2$. Thus (A8) follows from Lemma S1 upon replacing Σ by Ω.

(ii) Under the local alternative $\mu = \mu_0 + n^{-1/2}\delta u$ with $|\delta| \leqslant C$ and $\|u\| = 1$, we have the decomposition

$$\begin{aligned} &T_n - \frac{n\|\mu-\mu_0\|^2 + nk_n|\alpha^{\rm T}(\mu-\mu_0)|^2}{\{2\,{\rm tr}(\Omega^2)\}^{1/2}} \\ &= T_n^{\rm a} + \left[\frac{\{2\,{\rm tr}(\Omega^2)\}^{1/2}}{\{2\,\hat{\rm tr}(\Omega^2)\}^{1/2}} - 1\right]T_n^{\rm a} + \frac{{\rm tr}(\Omega) - \hat{\rm tr}(\Omega^2)}{\{2\,\hat{\rm tr}(\Omega^2)\}^{1/2}} + \frac{n\|\mu-\mu_0\|^2 + nk_n|\alpha^{\rm T}(\mu-\mu_0)|^2}{\{2\,\hat{\rm tr}(\Omega^2)\}^{1/2}}, \end{aligned}$$

where

$$T_n^{\rm a} = \frac{\{2nl_n^2/(n+2)^2\}W(\mu_0,k_n) - {\rm tr}(\Omega) - n\{\|\mu-\mu_0\|^2 + k_n|\alpha^{\rm T}(\mu-\mu_0)|^2\}}{\{2\,{\rm tr}(\Omega^2)\}^{1/2}}.$$

Here the third and fourth terms are of order $o_{\rm p}(1)$ since $\hat{\rm tr}(\Omega) - {\rm tr}(\Omega) = O_{\rm p}(1) + o_{\rm p}(k_n)$, $\hat{\rm tr}(\Omega^2) - {\rm tr}(\Omega^2) = O_{\rm p}(1) + o_{\rm p}(k_n^2)$, $n|\alpha^{\rm T}(\mu-\mu_0)|^2 \leqslant 1$ and $k_n = o(\surd p)$.

Next, we use the same techniques as in the proof of Lemma A1 to show that $T_n^{\rm a} \to N(0,1)$ in distribution. Under the local alternative $\mu = \mu_0 + n^{-1/2}\delta u$,

$$\xi_n^2 \leqslant \frac{2(n+2)^2}{l_n^2}\{\|\bar x-\mu_0\|^2 + k_n|\alpha^{\rm T}(\bar x-\mu_0)|^2\} + \frac{2(n+2)^2}{l_n^2}(\delta^2/n + k_n\delta^2|\alpha^{\rm T}u|^2/n) = o_{\rm p}(1).$$

The last equality follows from $\|\bar x-\mu_0\|^2 + k_n|\alpha^{\rm T}(\bar x-\mu_0)|^2 = O_{\rm p}(1)$ and $n^{5/4}/l_n = o(1)$. As in Step 1 of the proof of Lemma A1, $T_n^{\rm a}$ can be decomposed as

$$T_n^{\rm a} = \frac{\{nl_n^2/(n+2)^2\}\xi_n^2 - {\rm tr}(\Omega) - n\{\|\mu-\mu_0\|^2 + k_n|\alpha^{\rm T}(\mu-\mu_0)|^2\}}{\{2\,{\rm tr}(\Omega^2)\}^{1/2}} + O_{\rm p}\left[\frac{nl_n^2}{(n+2)^2\{{\rm tr}(\Omega^2)\}^{1/2}}\xi_n^4\right].$$

Here the second term is of order $o_{\rm p}(1)$ because

$$\frac{nl_n^2}{(n+2)^2\{{\rm tr}(\Omega^2)\}^{1/2}}\,\xi_n^4 \leqslant \frac{n(n+2)^2}{l_n^2\{{\rm tr}(\Omega^2)\}^{1/2}}\{\|\bar x-\mu_0\|^2 + k_n|\alpha^{\rm T}(\bar x-\mu_0)|^2\}^2 = o_{\rm p}(1).$$

For the first term we have that

$$\frac{nl_n^2}{(n+2)^2}\,\xi_n^2 - \mathrm{tr}(\Omega) - n\{\|\mu-\mu_0\|^2 + k_n|\alpha^{\mathrm T}(\mu-\mu_0)|^2\}$$
$$= n\{\|\bar x-\mu_0\|^2 + k_n|\alpha^{\mathrm T}(\bar x-\mu_0)|^2\} - \mathrm{tr}(\Omega)$$
$$+\; 2n\{(\bar x-\mu)^{\mathrm T}(\mu-\mu_0) + k_n\alpha^{\mathrm T}(\bar x-\mu)(\mu-\mu_0)^{\mathrm T}\alpha\}$$
$$+\left(\frac{l_n^2}{[l_n+\{\|\bar x-\mu_0\|^2 + k_n|\alpha^{\mathrm T}(\bar x-\mu_0)|^2\}^{1/2}]^2} - 1\right) n\big\{\|\bar x-\mu_0\|^2 + k_n|\alpha^{\mathrm T}(\bar x-\mu_0)|^2\big\}$$
$$= n\{\|\bar x-\mu_0\|^2 + k_n|\alpha^{\mathrm T}(\bar x-\mu_0)|^2\} - \mathrm{tr}(\Omega) + o_{\mathrm p}(\surd n).$$

This follows from the fact that the second summand in the first equality is $o_{\mathrm p}(\surd n)$ since

$$E\{n(\bar x-\mu_0)^{\mathrm T}(\mu-\mu_0)\}^2 = n(\mu-\mu_0)^{\mathrm T}\Gamma\Gamma^{\mathrm T}(\mu-\mu_0) = o(n),$$
$$E\{nk_n\alpha^{\mathrm T}(\bar x-\mu_0)\alpha^{\mathrm T}(\mu-\mu_0)\}^2 = k_n n\{\alpha^{\mathrm T}(\mu-\mu_0)\}^2\alpha^{\mathrm T}\Gamma\Gamma^{\mathrm T}\alpha = o(n),$$

while the last summand is $o_{\mathrm p}(\surd n)$ since $\|\bar x-\mu_0\|^2 + k_n|\alpha^{\mathrm T}(\bar x-\mu_0)|^2 = O_{\mathrm p}(1)$ and $n^{5/4}/l_n = o(1)$. □

References

Bai, Z. & Sarandasa, H. (1996). Effect of high dimension: By an example of a two sample problem. *Statist. Sinica* **6**, 311–29.

Bai, Z. D. & Yin, Y. Q. (1993). Limit of the smallest eigenvalue of a large dimensional sample covariance matrix. *Ann. Prob.* **21**, 1275–94.

Chen, B. B., Pan, G. M. & Zhou, W. (2015). Large dimensional empirical likelihood. *Statist. Sinica* **25**, 1659–77.

Chen, J., Variyath, A. M. & Abraham, B. (2008). Adjusted empirical likelihood and its properties. *J. Comp. Graph. Statist.* **17**, 426–43.

Chen, S. X., Peng, L. & Qin, Y. L. (2009). Effects of data dimension on empirical likelihood. *Biometrika* **96**, 711–22.

Chen, S. X. & Qin, Y. L. (2010). A two-sample test for high-dimensional data with application to gene-set testing. *Ann. Statist.* **38**, 808–35.

Chen, S. X., Zhang, L.-X. & Zhong, P.-S. (2010). Tests for high-dimensional covariance matrices. *J. Am. Statist. Assoc.* **105**, 810–9.

Emerson, S. & Owen, A. (2009). Calibration of the empirical likelihood method for a vector mean. *Electron. J. Statist.* **3**, 1161–92.

Hjort, N. L., McKeague, I. W. & Van Keilegom, I. (2009). Extending the scope of empirical likelihood. *Ann. Statist.* **37**, 1079–111.

Hyvärinen, A., Karhunen, J. & Oja, E., eds. (2001). *Independent Component Analysis*. New York: Wiley.

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75**, 237–49.

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

Pan, G. & Zhou, W. (2010). Circular law, extreme singular values and potential theory. *J. Mult. Anal.* **101**, 645–56.

Peng, L., Qi, Y. & Wang, F. (2014). Test for a mean vector with fixed or divergent dimension. *Statist. Sci.* **29**, 113–27.

Tsao, M. (2004). Bounds on coverage probabilities of the empirical likelihood ratio confidence regions. *Ann. Statist.* **32**, 1215–21.

Wang, L., Peng, B. & Li, R. (2015). A high-dimensional nonparametric multivariate test for mean vector. *J. Am. Statist. Assoc.* **110**, 1658–69.

[*Received on* 16 *January* 2019. *Editorial decision on* 27 *September* 2019]