COMMUNICATIONS IN STATISTICS—THEORY AND METHODS
2017, VOL. 46, NO. 17, 8479–8492
https://doi.org/10.1080/03610926.2016.1183785

Taylor & Francis
Taylor & Francis Group

Check for updates

The likelihood ratio test for high-dimensional linear regression model

Junshan Xie and Nannan Xiao

School of Mathematics and Statistics, Henan University, Kaifeng, P.R. China

ABSTRACT
The paper considers a significance test of regression variables in the high-dimensional linear regression model when the dimension of the regression variables p, together with the sample size n, tends to infinity. Under two sightly different cases, we proved that the likelihood ratio test statistic will converge in distribution to a Gaussian random variable, and the explicit expressions of the asymptotical mean and covariance are also obtained. The simulations demonstrate that our high-dimensional likelihood ratio test method outperforms those using the traditional methods in analyzing high-dimensional data.

ARTICLE HISTORY
Received 6 August 2015
Accepted 18 April 2016

KEYWORDS
High-dimensional data; likelihood ratio test; multivariate regression.

MATHEMATICS SUBJECT CLASSIFICATION
Primary 62H15; Secondary 62H10

1. Introduction

Linear regression analysis is one of the powerful and widely used techniques in statistics, the main goal of the regression analysis is to establish a linear relationship between the observation variable or vector Y and the regression variables or vectors X. A typical multivariate linear regression model in statistics is

$$Y_i = \mathbf{B}X_i + \varepsilon_i, \quad i = 1, 2, \ldots, n \tag{1.1}$$

where Y_i is a $p \times 1$ random observation vector, X_i is a known q-dimension regression variable, $\mathbf{B}$ is a $p \times q$ regression coefficient matrix, and $\{\varepsilon_i\}$ is a sequence of i.i.d. Gaussian noise $N_p(0, \mathbf{\Sigma})$ with an unknown covariance matrix $\mathbf{\Sigma}$. We assume that the $q \times n$ matrix $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ has rank q, and assume $n \geq p + q$ to ensure that the maximum-likelihood estimation of $\mathbf{\Sigma}$ is non singular.

One of the interesting problems in multivariate regression analysis is to consider whether the dimension of the regression variables X_i can be reduced, then a hypothesis test is meaningful to consider whether some of the regression variables are significant. As the significance of regression variables X_i can be determined by the elements of the regression coefficient matrix $\mathbf{B}$, then in order to test the significance of the first r variables in the regression matrix $\mathbf{X}$, we can partition the matrix $\mathbf{B}$ as

$$\mathbf{B} = (\mathbf{B_1}, \ \mathbf{B_2})$$

where $\mathbf{B_1}$ is $p \times r$ and $\mathbf{B_2}$ is $p \times (q - r)$, and it will be translated into the test of the null hypothesis

$$H_0 : \mathbf{B_1} = \mathbf{0} \tag{1.2}$$

CONTACT Nannan Xiao xjsnwpu@163.com School of Mathematics and Statistics, Henan University, Kaifeng 475000, P.R. China.

© 2017 Taylor & Francis Group, LLC

Downloaded by [Australian National University] at 00:21 02 October 2017

against $H_1 : \mathbf{B_1} \neq \mathbf{0}$. If the null hypothesis is accepted, then the dimension of the regression variables can be reduced.

For the sample $Y_1, Y_2, \ldots, Y_n$, we can denote two matrices A and C to be

$$A_{q\times q} = \sum_{i=1}^{n} X_i X_i', \; C_{p\times q} = \sum_{i=1}^{n} Y_i X_i'$$

Under the alternative hypothesis, a result in Sec. 8.3.1 of Anderson (1984) shows that the maximum-likelihood function for the parameter matrices $\mathbf{B}$, $\mathbf{\Sigma}$ is

$$L_1 := \max_{\mathbf{B},\mathbf{\Sigma}} L(\mathbf{B}, \mathbf{\Sigma}) = (2\pi e)^{-\frac{1}{2}pn} |\hat{\mathbf{\Sigma}}_1|^{-\frac{1}{2}n}$$

where $\hat{\mathbf{\Sigma}}_1 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{\mathbf{B}}X_i)(Y_i - \hat{\mathbf{B}}X_i)'$ and $\hat{\mathbf{B}} = CA^{-1}$ are the maximum-likelihood estimation of the population covariance matrices $\mathbf{\Sigma}$ and the regression matrix $\mathbf{B}$, respectively.

To find the maximum-likelihood function under the null hypothesis, we will partition the vector X_i as $X_i' = (X_{i1}', X_{i2}')$, where the vectors X_{i1} and X_{i2} are constructed by the first r elements and the last $q - r$ elements of X_i respectively. Then the matrices C and A can be partitioned in the manner corresponding to the partitioning of X_i as

$$C = (C_1, C_2), \; A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

where C_1 is $p \times r$, C_2 is $p \times (q - r)$, A_{11} is $r \times r$ and A_{22} is $(q - r) \times (q - r)$. The maximum-likelihood function can be expressed as

$$L_0 := \max_{\mathbf{B_2},\mathbf{\Sigma}} L(\mathbf{B}, \mathbf{\Sigma}) = (2\pi e)^{-\frac{1}{2}pn} |\hat{\mathbf{\Sigma}}_0|^{-\frac{1}{2}n}$$

where $\hat{\mathbf{\Sigma}}_0 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{\mathbf{B}}_2X_{i2})(Y_i - \hat{\mathbf{B}}_2X_{i2})'$ and $\hat{\mathbf{B}}_2 = \sum_{i=1}^{n} Y_i X_{i2}' A_{22}^{-1} = C_2 A_{22}^{-1}$ are the maximum-likelihood estimation of $\mathbf{\Sigma}$ and $\mathbf{B_2}$, respectively.

Then the likelihood ratio statistic for testing H_0 can be defined as

$$\Lambda_n = \frac{L_0}{L_1} = \frac{|\hat{\mathbf{\Sigma}}_0|^{-\frac{1}{2}n}}{|\hat{\mathbf{\Sigma}}_1|^{-\frac{1}{2}n}} := W_n^{\frac{n}{2}} \tag{1.3}$$

The distribution of the test statistic W_n can be studied through its moments. When the null hypothesis $H_0 : \mathbf{B_1} = \mathbf{0}$ is true, a result appeared in Sec. 10.5 of Muirhead (1982) or Sec. 8.4 of Anderson (1984) reveals that

$$EW_n^k = \frac{\Gamma_p\left[\frac{1}{2}(n-q)+k\right]}{\Gamma_p\left[\frac{1}{2}(n-q)\right]} \frac{\Gamma_p\left[\frac{1}{2}(n-q+r)\right]}{\Gamma_p\left[\frac{1}{2}(n-q+r)+k\right]} \tag{1.4}$$

where $k > -\frac{1}{2}(n + 1 - p - q)$ and $\Gamma_p(\cdot)$ is the p-dimensional Gamma function, whose definition is recalled in Sec. 4.

When the null hypothesis is true, for the fixed numbers n and p, the likelihood ratio statistic W_n has the same distribution as a product of independent Beta random variables. However, it is not a simple matter to actually find the density of W_n, then we cannot compute the critical value λ_0 such that $P(W_n < \lambda_0) = \alpha_0$, where α_0 is the size of the test, and it has little significance in the hypothesis test.

When n is large and p, r are fixed numbers, a classical result reveals that when $n \to \infty$, the statistic $-n \log W_n$ can be approximated by χ^2_{pr}, and we will call it as classical LRT method. A more precise result due to Box (1949), who give an explicit expansion of the distribution

Downloaded by [Australian National University] at 00:21 02 October 2017

Downloaded by [Australian National University] at 00:21 02 October 2017

function of $-N \log W_n$, where $N = n - q - \frac{1}{2}(p - r + 1)$, as $N \to \infty$:

$$P(-N \log W_n \le x) = P\left(\chi^2_{pr} \le x\right) + \frac{\gamma}{N^2}\left[P\left(\chi^2_{pr+4} \le x\right) - P\left(\chi^2_{pr} \le x\right)\right] + O(N^{-3}) \tag{1.5}$$

where $\gamma = \frac{pr(p^2+r^2-5)}{48}$. It can be referenced from Sec. 10.5.3 of Muirhead (1982) or Sec. 8.5.2 of Anderson (1984).

Along with the rapid development and widesdpread application of computer technology, high-dimensional data are increasingly encountered in many areas. In the fields of wireless communication, finance, genomics, and economics, we often encounter the data that both the dimension p and the sample size n are very large; this is often called ”large p, large n” phenomenon in the literature. As many traditional multivariate statistical methods and theories are established under the assumption that the dimension p is fixed, they may not necessarily work in those cases. Bai and Saranadasa (1996) discussed that the performance of the classical methods, which deal with the multivariate statistical problems no matter when the parameters n, p are fixed or the parameter n is large and p is fixed, will become very poor when facing the high-dimension statistical problems with both p and n tend to infinity. Then the analysis of such high-dimensional data calls for new statistical methodologies and theories, and it has one of the most popular topics in modern statistics; some discussions can be referred to Donoho et al. (2000), Cai and Shen (2010) and Fujikoshi et al. (2010).

For the linear regression analysis, the corresponding high-dimensional problem is also existed, and many efforts have been devoted to solve this problem. One major method to deal with high-dimensional regression model is variable selection method, which mainly adopts some penalty functions and turn the problem into a penalized optimization problem. See Tibshirani (1996), Fan and Li (2001), Fan and Lv (2008), and Wang (2009). The other method is hypothesis testing. In the setting of “$p \to \infty, n \to \infty$”, Goeman et al. (2011) formulated an Empirical Bayes test of the regression coefficients. Zhong and Chen (2011) considered a test for all high-dimensional regression coefficients for both simple random and factorial designs. Wang and Cui (2013, 2015) investigated the generalized F-test and a new U-statistics test for the part of regression coefficients. In addition, Feng et al. (2013) considered a robust score test on the regression coefficients under the that assumption of the dimension p is much larger than the sample size n. Under a similar situation, Wen et al. (2014) proposed a new procedure for testing the significance of a subset of regression coefficients. Zang et al. (2016) further established a jackknife empirical likelihood test for the coefficients of high-dimensional linear model.

As the fact that looking for the significance at the level of clusters of variables is more practical than that of the level of individual variables, then the test on high-dimensional regression coefficient matrix has become one of the most interesting topics. There is a remarkable result on the high-dimensional regression coefficient matrix test due to Bai et al. (2013), who also focus on the case that the dimension p and the sample size n tend to infinity with the same speed. By using the central limit theorem for linear spectral statistics of large random matrices, they developed a correction to the classical likelihood ratio statistic to make it tends to a Gaussian random variable, and their result can be stated as follows:

Lemma 1.1. *For the hypothesis test (1.2), when $n \to \infty$, we assume that $p \to \infty$, $y'_{n1} = \frac{p}{n-q} \to y'_1 \in (0, 1)$ and $y'_{n2} = \frac{p}{r} \to y'_2 \in (0, 1)$. Then under $H_0 : \mathbf{B}_1 = \mathbf{0}$, we have*

$$T_n = \frac{-\log W_n - p\theta_n - \mu_0}{\sigma_0} \xrightarrow{d} N(0, 1)$$

Downloaded by [Australian National University] at 00:21 02 October 2017

Here

$$\mu_0 = \frac{1}{2}\log\frac{(c^2-d^2)h^2}{(ch-y_1'd)^2}, \qquad \sigma_0^2 = 2\log\left(\frac{c^2}{c^2-d^2}\right)$$
$$\theta_n = \frac{y_{n1}'-1}{y_{n1}'}\log c_n + \frac{y_{n2}'-1}{y_{n2}'}\log(c_n-d_nh_n) + \frac{y_{n1}'+y_{n2}'}{y_{n1}'y_{n2}'}\log\left(\frac{c_nh_n-d_ny_{n1}'}{h_n}\right)$$

where

$$h = \sqrt{y_1'+y_2'-y_1'y_2'}, \qquad a_0, b_0 = \frac{(1\mp h)^2}{(1-y_1')^2}, \qquad c, d = \frac{1}{2}\left[\sqrt{1+\frac{y_1'}{y_2'}b_0} \pm \sqrt{1+\frac{y_1'}{y_2'}a_0}\right]$$

$$h_n = \sqrt{y_{n1}'+y_{n2}'-y_{n1}'y_{n2}'},\ a_n, b_n = \frac{(1\mp h_n)^2}{(1-y_{n1}')^2},\ c_n, d_n = \frac{1}{2}\left[\sqrt{1+\frac{y_{n1}'}{y_{n2}'}b_n} \pm \sqrt{1+\frac{y_{n1}'}{y_{n2}'}a_n}\right]$$

The paper will also consider the similar high-dimensional regression coefficient matrix test, and a high-dimensional likelihood ratio test method for (1.2) will be proposed. We will show that the high-dimensional likelihood ratio test statistic also converges in distribution to a Gaussian distributed random variable.

The rest of this paper is organized as follows. We present the main results in Sec. 2, and the asymptotic distribution of the proposed test statistics are established under two sightly different cases. In Sec. 3, we will compare the performance of four procedures for testing (1.2) through simulations. The proofs of our results are listed in Sec. 4.

2. Main results

In this paper, we follow a completely different approach introduced by Jiang and Yang (2013) to revisit the same high-dimensional regression problem as $n \to \infty$ and $p \to \infty$. In particular, by means of the asymptotic expansion of the multivariate Gamma function, the asymptotic Gaussian distribution of the likelihood ratio statistic in the high-dimension case will also obtained. Now we state the main result of the paper as follows.

Theorem 2.1. *Assume that $p = p_n$, $q = q_n$, and $r = r_n$ are three series of positive integers depending on n such that $p < n - q$ for all $n \geq 3$ and $r \geq p$. When $n \to \infty$, we assume that $p \to \infty$, and it holds that*

$$\frac{p}{n_1} \to y_1 \in (0, 1]; \ \frac{p}{n_2} \to y_2 \in (0, 1)$$

where $n_1 = n - q$ and $n_2 = n - (q - r)$. Define W_n as in (1.3). Then under $H_0 : \mathbf{B}_1 = \mathbf{0}$, we have

$$\frac{\log W_n - \mu_n}{\sigma_n} \xrightarrow{d} N(0, 1)$$

where

$$\sigma_n^2 = 2\left(\gamma_{n1}^2 - \gamma_{n2}^2\right), \quad \gamma_{ni}^2 = -\log\left(1 - \frac{p}{n_i}\right), i = 1, 2$$
$$\mu_n = p\log\frac{n_1}{n_2} - (p - n_1 + 0.5)\gamma_{n1}^2 + (p - n_2 + 0.5)\gamma_{n2}^2$$

Remark 2.1. A typical testing $H_0 : \mathbf{B_1} = \mathbf{0}$ is considered in the paper, the more general form of testing $H_0 : \mathbf{B_1} = \mathbf{B_1}^*$ against $H_1 : \mathbf{B_1} \neq \mathbf{B_1}^*$, where $\mathbf{B_1}^*$ is a specified matrix, can be discussed by the same procedure, and the two types of testing problem are essentially equivalent. In fact, the likelihood ratio criterion for testing on $\mathbf{B_1}$ is invariant with respect to any non singular linear transformations.

Remark 2.2. Our result is similar to that of Bai et al. (2013) , i.e., the asymptotical normal distribution of the test statistic $\log W_n$ is obtained under the assumptions that $n \to \infty$ and $p \to \infty$, while their result is proved by the method of large dimensional random matrix theory and our result will take a completely different approach. Although the expressions of the asymptotic mean and variance in two results are different, the asymptotic mean and variance in Bai et al. (2013) are deduced by some complicate integrals of complex valued function, and ours are obtained by the fundamental Taylor expansion of Gamma function. It is not easy to find the equivalence of the form of the two results. In fact, the numerical experiments in Sec. 3 show that when we take some special values of the parameters p, q, r, and n, expect some very slightly difference, the values of asymptotic mean and variance in two results are basically the same, indicating that the two results are consistent.

Furthermore, under a sightly different case, we will also consider the problem by another method. Before the result, we will introduce a notation which will appears in the result. For two sequences of numbers $\{a_n\}$ and $\{b_n\}$, the notation $O(a_n) \asymp O(b_n)$ as $n \to \infty$ means $limsup(\frac{a_n}{b_n}) < \infty$ and $limsup(\frac{b_n}{a_n}) < \infty$. The result can be stated as follows.

Theorem 2.2. *Assume that $p = p_n$, $q = q_n$, and $r = r_n$ are three series of positive integers depending on n such that $p < n - q$ for all $n \geq 3$ and $r \geq p$. When $n \to \infty$, we assume that $p \to \infty$. $n_1 = n - q$, $n_2 = n - (q - r)$, and $O(n_1 + 1 - p) \asymp O(n_2 + 1 - p) \asymp O(n_2 - n_1)$. Define W_n as in (1.3). Then under $H_0 : \mathbf{B}_1 = \mathbf{0}$, as $n \to \infty$, we have*

$$\frac{\log W_n - \widetilde{\mu}_n}{\widetilde{\sigma}_n} \xrightarrow{d} N(0, 1)$$

where

$$\widetilde{\mu}_n = \sum_{i=1}^{p} \left[\left(\frac{1}{n_2 + 1 - i} - \frac{1}{n_1 + 1 - i} \right) - \log \frac{n_2 + 1 - i}{n_1 + 1 - i} \right]$$

$$\widetilde{\sigma}_n^2 = \sum_{i=1}^{p} \left(\frac{2}{n_1 + 1 - i} - \frac{2}{n_2 + 1 - i} \right)$$

The two results are all focused on the case when the dimension variables p, together with sample size n, tend to infinity in high-dimensional regression model. Although the assumptions have sightly difference, the results of Theorems 2.1 and 2.2 are consistent, and this can be seen from the following corollary.

Corollary 2.1. *The result of Theorems 2.1 and 2.2 are consistent. In particular, under the assumptions of Theorem 2.2, the asymptotic mean $\widetilde{\mu}_n$ and variance $\widetilde{\sigma}_n^2$ in Theorem 2.2 are equivalent to μ_n and σ_n^2 in Theorem 2.1, respectively. That is, when $n \to \infty$, we have*

$$\widetilde{\mu}_n = \mu_n + o(1)$$

and

$$\widetilde{\sigma}_n^2 = \sigma_n^2 + o(1)$$

Downloaded by [Australian National University] at 00:21 02 October 2017

Remark 2.3. Compared with the assumptions on parameters p, n_1, and n_2 in Theorems 2.1 and 2.2, we can see that two kinds of assumptions cannot cover each other. They are complementary and both are of interest. On the one hand, Theorems 2.1 does not work for all assumptions in 2.2. For example, take

$$p = \sqrt[3]{n}, \quad q = n - \sqrt[3]{n} - \sqrt{n}, \quad r = \sqrt{n}$$

respectively.

We have

$$n_1 = n - q = \sqrt[3]{n} + \sqrt{n}, \qquad n_2 = n - (q - r) = \sqrt[3]{n} + 2\sqrt{n}$$
$$n_1 + 1 - p = \sqrt{n} + 1, \quad n_2 + 1 - p = 2\sqrt{n} + 1, \quad n_2 - n_1 = \sqrt{n}$$

As a result, we have

$$\frac{p}{n_1} \to 0, \quad \frac{p}{n_2} \to 0, \quad O(n_1 + 1 - p) \asymp O(n_2 + 1 - p) \asymp O(n_2 - n_1)$$

This satisfies the assumptions of 2.2, but does not satisfy the assumptions of Theorems 2.1.

On the other hand, since 2.2 has an additional assumption that $O(n_1 + 1 - p) \asymp O(n_2 + 1 - p) \asymp O(n_2 - n_1)$, 2.2 also does not work under all assumptions in Theorems 2.1. In particular, if taking

$$p = \sqrt{n} - 1, \quad q = n - \sqrt{n}, \quad r = 2\sqrt{n}$$

respectively, we have

$$n_1 = n - q = \sqrt{n}, \qquad n_2 = n - (q - r) = 3\sqrt{n}$$
$$n_1 + 1 - p = 2, \quad n_2 + 1 - p = 2\sqrt{n} + 2, \quad n_2 - n_1 = 2\sqrt{n}$$

As a result,
we have

$$\frac{p}{n_1} \to 1 \in (0, 1], \quad \frac{p}{n_2} \to \frac{1}{3} \in (0, 1)$$

but it does not satisfy the assumption

$$O(n_1 + 1 - p) \asymp O(n_2 + 1 - p) \asymp O(n_2 - n_1)$$

which indicates that it satisfies the assumptions of Theorems 2.1, but does not satisfy the assumptions of 2.2.

3. Simulations

In order to test the significance of our method introduced in Theorems 2.1, we will give some simulation results to compare the significance of four procedures on testing the problem (1.2). The classical LRT method (LRT), the Box's method (BOX) related to (1.5), the random matrix theory method (RMT) by Bai et al., and our high-dimensional LRT method (HLRT) will be examined.

To study the size of these tests, we simulate data for the model

$$Y_i = \mathbf{B_2} z_{2i} + \varepsilon_i, \quad i = 1, 2, \ldots, n$$

where B_2 is a $p \times (q - r)$ real matrices with each elements are 1. z_{2i} is $(q - r) \times 1$ random vectors, and the i.i.d. entries of z_{2i} are drawn from the distribution $N(1, 0.25)$. The errors ε_i have a multivariate Gaussian distribution $N(0, I_p)$.

Downloaded by [Australian National University] at 00:21 02 October 2017

Table 1. Empirical sizes of the proposed tests.

(p,n,q,r)	LRT	BOX	RMT	HLRT
(2,10,6,4)	0.5760	0.0630	0.4610	0.4605
(2,10,7,4)	0.7441	0.0840	0.2692	0.2686
(2,100,4,3)	0.0570	0.0491	0.0420	0.0418
(2,300,4,3)	0.0585	0.0522	0.0390	0.0393
(2,500,4,3)	0.0530	0.0520	0.0683	0.0678
(5,20,10,6)	0.8942	0.0770	0.0640	0.0639
(5,20,13,6)	0.9930	0.1334	0.0691	0.0694
(10,100,30,20)	0.9085	0.0630	0.0412	0.0408
(15,100,50,30)	1	0.1147	0.0482	0.0478
(25,100,70,50)	1	0.9983	0.0504	0.0501
(28,100,70,50)	1	1	0.0523	0.0527
(30,300,80,40)	1	0.0694	0.0510	0.0513
(50,500,100,60)	1	0.0662	0.0507	0.0494
(80,300,120,100)	1	0.8260	0.0462	0.0459
(100,500,200,120)	1	0.5232	0.0490	0.0489
(120,300,170,140)	1	1	0.0491	0.0491
(128,300,170,140)	1	1	0.0521	0.0528
(180,500,300,200)	1	1	0.0500	0.0501
(198,500,300,200)	1	1	0.0531	0.0537

For different values of (p, n, q, r), we will take the same significance level $\alpha = 0.05$, and 10000 independent replications will be carried out. The empirical sizes of the four proposed tests are summarized in Table 1.

As for the empirical powers of the four proposed tests, we will consider model (1.1) in a special case where $\mathbf{B}_1$ is a $p \times r$ real matrices with each elements are $\frac{1}{8}$, $\mathbf{B}_2$ is a $p \times (q - r)$ real matric with each elements are 1. The i.i.d. entries of $q \times 1$ random vectors z_i are drawn from the distribution $N(1, 0.25)$. The error data ε_i also come from $N(0, I_p)$. We will also run 10,000 independent replications. The powers of the tests are presented in Table 2.

From Tables 1 and 2, we can see that when the parameters p and r are small and the sample size n is large, the sizes of the traditional LRT method are always around the set level 0.05, meaning that the traditional LRT method is well performed. When the parameters p and r are large, both the traditional LRT method and the BOX method have much higher sizes

Table 2. Empirical powers of the proposed tests.

(p,n,q,r)	LRT	BOX	RMT	HLRT
(2,10,6,4)	0.6201	0.0856	0.0630	0.0633
(2,10,7,4)	0.7834	0.1022	0.0579	0.0579
(2,100,4,3)	0.5216	0.4880	0.4777	0.4779
(2,300,4,3)	0.9694	0.9675	0.9651	0.9652
(2,500,4,3)	0.9993	0.9993	0.9991	0.9993
(5,20,10,6)	0.9374	0.1320	0.0782	0.0780
(5,20,13,6)	0.9962	0.1713	0.0698	0.0697
(10,100,30,20)	0.9958	0.9939	0.9847	0.9845
(15,100,50,30)	0.9989	0.9639	0.8474	0.8477
(25,100,70,50)	0.9998	1	0.9940	0.9939
(28,100,70,50)	1	1	0.9774	0.9775
(30,300,80,40)	1	1	1	1
(50,500,100,60)	1	1	1	1
(80,300,120,100)	1	1	1	1
(100,500,200,120)	1	1	1	1
(120,300,170,140)	1	1	0.9858	0.9857
(128,300,170,140)	1	1	0.9820	0.9823
(180,500,300,200)	1	1	0.9972	0.9975
(198,500,300,200)	1	1	0.9868	0.9869

Downloaded by [Australian National University] at 00:21 02 October 2017

than the set level 0.05 in all considered cases. In particular, when both p and n are large, our simulations show that the traditional LRT method and the BOX method always reject H_0, leading to a 100% alpha error. However, as mentioned in Remark 2.2, the empirical sizes and the empirical powers of the HLRT method are always basically the same as that of the RMT method. The difference values are always no more than the order of 10^{-4} and they have very little difference in the above numerical simulations, meaning that the HLRT method using Theorems 2.1 performs as good as the RMT method, and their sizes are very close to 0.05 all the time. On the power side, our simulations show that all of the four methods have high powers. In particular, when p becomes large, their powers reach or close to 100%. In some cases, the powers of the LRT method and Box method are better than those of the RMT method and HLRT method. However, the failure of the test (100% alpha error) is the reason why the powers of the traditional LRT method and the BOX method go up to 100% in these cases. In a word, the proposed HLRT method using Theorem 1.1 outperforms the traditional LRT method and the BOX method, and it performs as efficient as the RMT method.

Downloaded by [Australian National University] at 00:21 02 October 2017

4. Proofs of the main results

To prove the result of Theorems 2.1, we will first present the continuity theorem of the moment generating function, which is one of the most important tools to prove the weak convergence of random variable series. The following result can be seen from Theorem 9.5 in Gut (2005).

Lemma 4.1. *Let $\{\xi_n\}$ be a sequence of random variables such that the moment generating function $\psi_{\xi_n}(t) = Ee^{t\xi_n}$ exists for $|t| < b$, for some $b > 0$, and for all n. Suppose further that ξ is a random variable whose moment generating function, $\psi_\xi(t) = Ee^{t\xi}$, exists for $|t| \le b_1 < b$ for some $b_1 > 0$, and when $n \to \infty$,*

$$\psi_{\xi_n}(t) \to \psi_\xi(t)$$

for all $|t| \le b_1$. Then we have

$$\xi_n \xrightarrow{d} \xi$$

In the following proofs, the asymptotic expansion of multivariate Gamma function $\Gamma_p(\alpha)$ will play an important role. To introduce the result, we will first recall the definition of the multivariate Gamma function.

Definition 4.1. The multivariate Gamma function, denoted by $\Gamma_p(\alpha)$, is defined as

$$\Gamma_p(\alpha) = \int_{A>0} \exp\{tr(-A)\}(detA)^{\alpha-\frac{p+1}{2}}(dA) \tag{4.1}$$

for $\alpha \in \mathbb{C}$ with $Re(\alpha) > \frac{p-1}{2}$, where the integral is over the space of positive definite symmetric $p \times p$ matrices $\{A : A > 0\}$.

Note that when $p = 1$, (4.1) just becomes the usual definition of a Gamma function $\Gamma_1(\alpha) = \Gamma(\alpha) = \int_0^\infty \exp(-x)x^{\alpha-1}dx$ with $\alpha \in \mathbb{C}$ and $Re(\alpha) > 0$. The multivariate Gamma function can be expressed as a product of ordinary Gamma function, i.e.,

$$\Gamma_p(\alpha) = \pi^{\frac{p(p-1)}{4}} \prod_{i=1}^{p} \Gamma\left[\alpha - \frac{1}{2}(i-1)\right] \; for \; Re(\alpha) > \frac{1}{2}(p-1) \tag{4.2}$$

see Theorem 2.1.12 in Muirhead (1982).

There is a remarkable work due by Jiang and Yang (2013), who obtained the asymptotic expansion of Multivariate Gamma function. The result can be stated as follows.

Lemma 4.2. *Let $n > p = p_n$ and $\gamma_n = [-\log(1-\frac{p}{n})]^{\frac{1}{2}}$. Assume $\frac{p}{n} \to y \in (0,1]$ and $t = t_n = O(\frac{1}{\gamma_n})$ as $n \to \infty$, Then, as $n \to \infty$, we have*

$$\log \frac{\Gamma_p\left(\frac{n}{2}+t\right)}{\Gamma_p\left(\frac{n}{2}\right)} = pt(\log n - 1 - \log 2) + \gamma_n^2\left[t^2 - \left(p - n + \frac{1}{2}\right)t\right] + o(1)$$

Proof of Theorem 2.1: By the result of Lemma 4.1, it is only need to prove that there exists $\delta > 0$ such that

$$E\exp\left(\frac{\log W_n - \mu_n}{\sigma_n}u\right) \to e^{\frac{u^2}{2}}$$

as $n \to \infty$ for all $|u| < \delta$.

If $y_1 \in (0,1)$, by using the fact that $n_1 < n_2$, $\gamma_{n1}^2 > \gamma_{n2}^2$, when $n \to \infty$, we have $\gamma_{n1}^2 \to -\log(1-y_1)$, $\gamma_{n2}^2 \to -\log(1-y_2)$ and $\sigma_n^2 \to -2[\log(1-y_1) - \log(1-y_2)] > 0$.

If $y_1 = 1$, then $\gamma_{n1}^2 \to \infty$. We can define

$$\delta_0 = \inf\{\gamma_{n1} : n \geq 3\}$$

Fix $|s| < \delta_0$, set $t = t_n = \frac{s}{\gamma_{n1}}$, then we have $|t_n| < 1$ for all $n \geq 3$, and $t^2\gamma_{n1}^2 = s^2$, which means $t = O(\frac{1}{\gamma_{n1}})$. It follows by Lemma 4.2 that

$$\log \frac{\Gamma_p\left(\frac{1}{2}n_1+t\right)}{\Gamma_p\left(\frac{1}{2}n_1\right)} = pt\left(\log n_1 - 1 - \log 2\right) + \gamma_{n_1}^2\left[t^2 - \left(p - n_1 + \frac{1}{2}\right)t\right] + o(1) \tag{4.3}$$

At the same time, we also know that $t^2\gamma_{n2}^2 = s^2\frac{\gamma_{n2}^2}{\gamma_{n1}^2} \leq s^2$, then $t = O(\frac{1}{\gamma_{n2}})$. Using Lemma 4.2 again, then

$$\log \frac{\Gamma_p\left(\frac{1}{2}n_2\right)}{\Gamma_p\left(\frac{1}{2}n_2+t\right)} = -pt(\log n_2 - 1 - \log 2) - \gamma_{n_2}^2\left[t^2 - \left(p - n_2 + \frac{1}{2}\right)t\right] + o(1) \tag{4.4}$$

By the fact (1.4), we have

$$\log E[e^{t\log W_n}] = \log E[W_n^t] = \log \frac{\Gamma_p\left(\frac{1}{2}n_1+t\right)}{\Gamma_p\left(\frac{1}{2}n_1\right)} + \log \frac{\Gamma_p\left(\frac{1}{2}n_2\right)}{\Gamma_p(\frac{1}{2}n_2+t)}$$

This together with (4.3) and (4.4) can lead to

$$\begin{aligned}
&\log E[e^{t\log W_n}] \\
&= \left(\gamma_{n1}^2 - \gamma_{n2}^2\right)t^2 + \left[p\log\frac{n_1}{n_2} - \left(p - n_1 + \frac{1}{2}\right)\gamma_{n_1}^2 + \left(p - n_2 + \frac{1}{2}\right)\gamma_{n_2}^2\right]t + o(1) \\
&= \frac{t^2}{2}\sigma_n^2 + \mu_n t + o(1)
\end{aligned}$$

Let $u = t\sigma_n$, as the fact that $t = \frac{s}{\gamma_{n1}}$, $\sigma_n = \sqrt{2(\gamma_{n1}^2 - \gamma_{n2}^2)}$ and $\gamma_{n1}^2 > \gamma_{n2}^2$, then $|u| < \sqrt{2}\delta_0 := \delta$. We can deduce that

$$E\exp\left(\frac{\log W_n - \mu_n}{\sigma_n}u\right) = e^{\frac{u^2}{2}} + o(1)$$

Downloaded by [Australian National University] at 00:21 02 October 2017

Downloaded by [Australian National University] at 00:21 02 October 2017

which implies that

$$\frac{\log W_n - \mu_n}{\sigma_n} \xrightarrow{d} N(0,1)$$

and the proof of Theorem 2.1 is completed. □

To prove the result of 2.2, we will first present an important lemma, which can be seen in Lemma 5.1 of Jiang and Qi (2015).

Lemma 4.3. *As $x \to \infty$,*

$$\log \frac{\Gamma(x+b)}{\Gamma(x)} = (x+b)\log(x+b) - x\log x - b - \frac{b}{2x} + O\left(\frac{b^2+1}{x^2}\right)$$

holds uniformly on $b \in [-\delta x, \delta x]$ for any given $\delta \in (0,1)$.

Proof of Theorem 2.2: By the result of Lemma 4.1, it suffices to show that

$$E\exp\left(\frac{\log W_n - \mu_n}{\sigma_n}s\right) \to e^{\frac{s^2}{2}}$$

as $n \to \infty$ for all $|s| \le 1$ and the corresponding μ_n and σ_n, or equivalently

$$\log EW_n^t = \mu_n t + \frac{\sigma_n^2 t^2}{2} + o(1)$$

as $n \to \infty$ with $|\sigma_n t| \le 1$.

By the fact (1.4), we have

$$\begin{aligned}
\log EW_n^t &= \log \frac{\Gamma_p[\frac{n_1+t}{2}]}{\Gamma_p[\frac{n_1}{2}]}\frac{\Gamma_p[\frac{n_2}{2}]}{\Gamma_p[\frac{n_2}{2}+t]} = \log \frac{\prod_{i=1}^p \Gamma[\frac{n_1+1-i}{2}+t]}{\prod_{i=1}^p \Gamma[\frac{n_1+1-i}{2}]}\frac{\prod_{i=1}^p \Gamma[\frac{n_2+1-i}{2}]}{\prod_{i=1}^p \Gamma[\frac{n_2+1-i}{2}+t]} \\
&= \sum_{i=1}^p \log \frac{\Gamma[\frac{n_1+1-i}{2}+t]}{\Gamma[\frac{n_1+1-i}{2}]}\frac{\Gamma[\frac{n_2+1-i}{2}]}{\Gamma[\frac{n_2+1-i}{2}+t]} \\
&= \sum_{i=1}^p \left[\log \frac{\Gamma(\frac{n_1+1-i}{2}+t)}{\Gamma(\frac{n_1+1-i}{2})} - \log \frac{\Gamma(\frac{n_2+1-i}{2}+t)}{\Gamma(\frac{n_2+1-i}{2})}\right]
\end{aligned}$$

By applying Lemma 4.3 to the summands, we have

$$\begin{aligned}
\log EW_n^t = \sum_{i=1}^p \Bigg[&\left(\frac{n_1+1-i}{2}+t\right)\log\left(\frac{n_1+1-i}{2}+t\right) - \frac{n_1+1-i}{2}\log\frac{n_1+1-i}{2} - t \\
&- \frac{t}{n_1+1-i} + O\left(\frac{t^2+1}{(\frac{n_1+1-i}{2})^2}\right) - \left(\frac{n_2+1-i}{2}+t\right)\log\left(\frac{n_2+1-i}{2}+t\right) \\
&+ \frac{n_2+1-i}{2}\log\frac{n_2+1-i}{2} + t + \frac{t}{n_2+1-i} + O\left(\frac{t^2+1}{(\frac{n_2+1-i}{2})^2}\right)\Bigg] \\
= \sum_{i=1}^p \Bigg[&\left(\frac{n_1+1-i}{2}+t\right)\log\left(1+\frac{t}{\frac{n_1+1-i}{2}}\right) + t\log\frac{n_1+1-i}{2} - \frac{t}{n_1+1-i} \\
&+ O\left(\frac{t^2+1}{(n_1+1-i)^2}\right) - \left(\frac{n_2+1-i}{2}+t\right)\log\left(1+\frac{t}{\frac{n_2+1-i}{2}}\right) \\
&- t\log\frac{n_2+1-i}{2} + \frac{t}{n_2+1-i} + O\left(\frac{t^2+1}{(n_2+1-i)^2}\right)\Bigg]
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=1}^{p}\left[\left(\frac{n_1+1-i}{2}+t\right)\left(\frac{t}{\frac{n_1+1-i}{2}}-\frac{t^2}{2\left(\frac{n_1+1-i}{2}\right)^2}+O\left(\frac{t^3}{(n_1+1-i)^3}\right)\right)\right.\\
&\quad + t\log\frac{n_1+1-i}{2}-\frac{t}{n_1+1-i}+O\left(\frac{t^2+1}{(n_1+1-i)^2}\right)-\left(\frac{n_2+1-i}{2}+t\right)\\
&\quad \times\left(\frac{t}{\frac{n_2+1-i}{2}}-\frac{t^2}{2\left(\frac{n_2+1-i}{2}\right)^2}+O\left(\frac{t^3}{(n_2+1-i)^3}\right)\right)-t\log\frac{n_2+1-i}{2}\\
&\quad \left.+\ \frac{t}{n_2+1-i}+O\left(\frac{t^2+1}{(n_2+1-i)^2}\right)\right]\\
&= \sum_{i=1}^{p}\left[\left(\frac{1}{n_2+1-i}-\frac{1}{n_1+1-i}\right)-\log\frac{n_2+1-i}{n_1+1-i}\right]t\\
&\quad + \sum_{i=1}^{p}\left(\frac{2}{n_1+1-i}-\frac{2}{n_2+1-i}\right)\frac{t^2}{2}+\sum_{i=1}^{p}O\left(\frac{t^3+t^2+1}{(n_1+1-i)^2}\right)\\
&\quad + \sum_{i=1}^{p}O\left(\frac{t^4}{(n_1+1-i)^3}\right)+\sum_{i=1}^{p}O\left(\frac{t^3+t^2+1}{(n_2+1-i)^2}\right)+\sum_{i=1}^{p}O\left(\frac{t^4}{(n_2+1-i)^3}\right)\\
&= \widetilde{\mu}_n t+\widetilde{\sigma}_n^2\frac{t^2}{2}+\sum_{i=1}^{p}O\left(\frac{t^3+t^2+1}{(n_1+1-i)^2}\right)+\sum_{i=1}^{p}O\left(\frac{t^4}{(n_1+1-i)^3}\right)\\
&\quad + \sum_{i=1}^{p}O\left(\frac{t^3+t^2+1}{(n_2+1-i)^2}\right)+\sum_{i=1}^{p}O\left(\frac{t^4}{(n_2+1-i)^3}\right)
\end{aligned}
$$

Now it is sufficient to prove as $n\to\infty$,

$$
\begin{aligned}
&\sum_{i=1}^{p}\left[O\left(\frac{t^3+t^2+1}{(n_1+1-i)^2}\right)+O\left(\frac{t^4}{(n_1+1-i)^3}\right)+O\left(\frac{t^3+t^2+1}{(n_2+1-i)^2}\right)+O\left(\frac{t^4}{(n_2+1-i)^3}\right)\right]\\
&\quad = o(1)
\end{aligned}\tag{4.5}
$$

for all $|\sigma_n t|\le 1$.

Note that $|\sigma_n t|\le 1$, we have

$$
\begin{aligned}
t &= O\left(\frac{1}{\sigma_n}\right)=O\left(\frac{1}{\sqrt{\sum_{i=1}^{p}\left(\frac{2}{n_1+1-i}-\frac{2}{n_2+1-i}\right)}}\right)=O\left(\frac{1}{\sqrt{\sum_{i=1}^{p}\frac{2(n_2-n_1)}{(n_1+1-i)(n_2+1-i)}}}\right)\\
&= O\left(\frac{1}{\sqrt{\frac{2(n_2-n_1)}{(n_1+1-p)(n_2+1-p)}}}\right)=O\left(\sqrt{\frac{(n_1+1-p)(n_2+1-p)}{n_2-n_1}}\right)
\end{aligned}
$$

Recall the constraint that $O(n_1+1-p)\asymp O(n_2+1-p)\asymp O(n_2-n_1)$. We can verify that

$$
\begin{aligned}
\sum_{i=1}^{p}O\left(\frac{t^3+t^2+1}{(n_1+1-i)^2}\right) &= O\left(\frac{t^3+t^2+1}{(n_1+1-p)^2}\right)=O\left(\frac{t^3}{(n_1+1-p)^2}\right)+O\left(\frac{t^2}{(n_1+1-p)^2}\right)\\
&\quad +O\left(\frac{1}{(n_1+1-p)^2}\right)=O\left(\frac{(n_1+1-p)^{\frac{3}{2}}(n_2+1-p)^{\frac{3}{2}}}{(n_2-n_1)^{\frac{3}{2}}(n_1+1-p)^2}\right)\\
&\quad +O\left(\frac{(n_1+1-p)(n_2+1-p)}{(n_2-n_1)(n_1+1-p)^2}\right)+O\left(\frac{1}{(n_1+1-p)^2}\right)=o(1)
\end{aligned}\tag{4.6}
$$

Downloaded by [Australian National University] at 00:21 02 October 2017

and

$$\sum_{i=1}^{p} O\left(\frac{t^4}{(n_1+1-i)^3}\right) = O\left(\frac{t^4}{(n_1+1-p)^3}\right) = O\left(\frac{(n_1+1-p)^2(n_2+1-p)^2}{(n_2-n_1)^2(n_1+1-p)^3}\right) = o(1) \tag{4.7}$$

Similar to (4.6) and (4.7), respectively, we have

$$\sum_{i=1}^{p} O\left(\frac{t^3+t^2+1}{(n_2+1-i)^2}\right) = O\left(\frac{t^3+t^2+1}{(n_2+1-p)^2}\right) = o(1) \tag{4.8}$$

and

$$\sum_{i=1}^{p} O\left(\frac{t^4}{(n_2+1-i)^3}\right) = O\left(\frac{t^4}{(n_2+1-p)^3}\right) = o(1) \tag{4.9}$$

Combining relations (4.5)–(4.9), we can complete the proof of Theorem 2.2. □

Proof of Corollary 2.1: By the partial sum of harmonic series, $\sum_{i=1}^{p}\frac{1}{i} = \log k + \gamma + \frac{1}{2k} + \tau(k)$, where γ is the Euler–Mascheroni constant and $\tau(k) = o(\frac{1}{k^2})$ as $k \to \infty$, we can deduce that when $n \to \infty$,

$$\begin{aligned}
\widetilde{\sigma}_n^2 &= \sum_{i=1}^{p}\left(\frac{2}{n_1+1-i} - \frac{2}{n_2+1-i}\right) = 2\left(\sum_{i=1}^{n_1}\frac{1}{i} - \sum_{i=1}^{n_1-p}\frac{1}{i}\right) - 2\left(\sum_{i=1}^{n_2}\frac{1}{i} - \sum_{i=1}^{n_2-p}\frac{1}{i}\right)\\
&= 2\left[\log n_1 + \gamma + \frac{1}{2n_1} + o\left(\frac{1}{n_1^2}\right) - \log(n_1-p) - \gamma - \frac{1}{2(n_1-p)} - o\left(\frac{1}{(n_1-p)^2}\right)\right]\\
&\quad -2\left[\log n_2 + \gamma + \frac{1}{2n_2} + o\left(\frac{1}{n_2^2}\right) - \log(n_2-p) - \gamma - \frac{1}{2(n_2-p)} - o\left(\frac{1}{(n_2-p)^2}\right)\right]\\
&= 2\log\left(1-\frac{p}{n_2}\right) - 2\log\left(1-\frac{p}{n_1}\right) + O\left(\frac{1}{n_1}\right) + O\left(\frac{1}{n_1-p}\right) + O\left(\frac{1}{n_2}\right)\\
&\quad + O\left(\frac{1}{n_2-p}\right) = 2\left(\gamma_{n_1}^2 - \gamma_{n_2}^2\right) + o(1) = \sigma_n^2 + o(1)
\end{aligned}$$

Furthermore, we can see that

$$\begin{aligned}
\widetilde{\mu}_n &= \sum_{i=1}^{p}\left[\left(\frac{1}{n_2+1-i} - \frac{1}{n_1+1-i}\right) - \log\frac{n_2+1-i}{n_1+1-i}\right]\\
&= \sum_{i=1}^{p}\left[\log(n_1+1-i) - \log(n_2+1-i)\right] - \frac{1}{2}\left[\sum_{i=1}^{p}\left(\frac{2}{n_1+1-i} - \frac{2}{n_2+1-i}\right)\right]\\
&= \log\frac{n_1!}{(n_1-p)!} - \log\frac{n_2!}{(n_2-p)!} - \frac{1}{2}\widetilde{\sigma}_n^2\\
&= \log\frac{(n_1-1)!}{(n_1-p-1)!} + \log\frac{n_1}{n_1-p} - \log\frac{(n_2-1)!}{(n_2-p-1)!} - \log\frac{n_2}{n_2-p} - \frac{1}{2}\widetilde{\sigma}_n^2\\
&= \log\left(1-\frac{p}{n_2}\right) - \log\left(1-\frac{p}{n_1}\right) + \log\frac{\Gamma(n_1)}{\Gamma(n_1-p)} - \log\frac{\Gamma(n_2)}{\Gamma(n_2-p)} - \frac{1}{2}\widetilde{\sigma}_n^2
\end{aligned}$$

Downloaded by [Australian National University] at 00:21 02 October 2017

$$= [\log \Gamma(n_1) - \log \Gamma(n_1 - p)] - [\log \Gamma(n_2) - \log \Gamma(n_2 - p)] + O\left(\frac{1}{n_1}\right)$$

$$+ O\left(\frac{1}{n_1 - p}\right) + O\left(\frac{1}{n_2}\right) + O\left(\frac{1}{n_2 - p}\right) \tag{4.10}$$

By the Stirling formula in Ahlfors (1979), we have a fact that

$$\log \Gamma(x) = \left(x - \frac{1}{2}\right) \log x - x + \log \sqrt{2\pi} + \frac{1}{12x} + O\left(\frac{1}{x^3}\right)$$

as $x \to \infty$. Then we have

$$\begin{aligned}\log \Gamma(n_1) - \log \Gamma(n_1 - p) &= \left(n_1 - \frac{1}{2}\right) \log n_1 - n_1 + \frac{1}{12n_1} + O\left(\frac{1}{n_1^3}\right) - \left(n_1 - p - \frac{1}{2}\right) \\ &\quad \times \log(n_1 - p) + (n_1 - p) - \frac{1}{12(n_1 - p)} + O\left(\frac{1}{(n_1 - p)^3}\right) \\ &= \left(n_1 - \frac{1}{2}\right) \log n_1 - \left(n_1 - p - \frac{1}{2}\right) \log(n_1 - p) - p + O\left(\frac{1}{n_1^3}\right) \\ &\quad + O\left(\frac{1}{(n_1 - p)^3}\right) + O\left(\frac{p}{n_1(n_1 - p)}\right) \end{aligned} \tag{4.11}$$

Similarly,

$$\begin{aligned}\log \Gamma(n_2) - \log \Gamma(n_2 - p) &= \left(n_2 - \frac{1}{2}\right) \log n_2 - \left(n_2 - p - \frac{1}{2}\right) \log(n_2 - p) - p \\ &\quad + O\left(\frac{1}{n_2^3}\right) + O\left(\frac{1}{(n_2 - p)^3}\right) + O\left(\frac{p}{n_2(n_2 - p)}\right) \end{aligned} \tag{4.12}$$

Putting (4.11) and (4.12) into (4.10), after some elementary calculus, we can deduce that under the assumptions of 2.2,

$$\begin{aligned}\widetilde{\mu}_n &= p \log \frac{n_1}{n_2} - \left(p - n_1 + \frac{1}{2}\right)\left[-\log\left(1 - \frac{p}{n_1}\right)\right] + \left(p - n_2 + \frac{1}{2}\right)\left[-\log\left(1 - \frac{p}{n_2}\right)\right] \\ &\quad + O\left(\frac{1}{n_1}\right) + O\left(\frac{1}{n_1 - p}\right) + O\left(\frac{1}{n_2}\right) + O\left(\frac{1}{n_2 - p}\right) + O\left(\frac{1}{n_1^3}\right) + O\left(\frac{1}{(n_1 - p)^3}\right) \\ &\quad + O\left(\frac{p}{12n_1(n_1 - p)}\right) + O\left(\frac{1}{n_2^3}\right) + O\left(\frac{1}{(n_2 - p)^3}\right) + O\left(\frac{p}{12n_2(n_2 - p)}\right) \\ &= p \log \frac{n_1}{n_2} - (p - n_1 + 0.5)\gamma_{n1}^2 + (p - n_2 + 0.5)\gamma_{n2}^2 + o(1) = \mu_n + o(1)\end{aligned}$$

This completes the proof of the corollary. □

Acknowledgments

The author would like to thank the referee for many valuable comments and suggestions.

Funding

This work is supported by National Natural Science Foundation of China [grant number 11401169], [grant number 11461032], [grant number 11401267], [grant number 11471101] and Teaching Project of Henan University in China [grant number HDXJJG2014-117].

Downloaded by [Australian National University] at 00:21 02 October 2017

References

Ahlfors, L.V. (1979). *ComplexAnalysis*. 3rd Ed. New York: McGraw-Hill, Inc.

Anderson, T.W. (1984). *An Introduction to Multivariate Statistical Analysis*. New York: John Wiley & Sons Inc.

Bai, Z.D., Jiang, D.D., Yao, J.F., Zheng, S.R. (2013). Testing linear hypothesis in high-dimensional regressions. *Statistics* 47(6):1207–1223.

Bai, Z.D., Saranadasa, H. (1996). Effect of high dimension: By an example of a two sample problem. *Stat. Sin.* 6(2):311–329.

Box, G.E.P. (1949). A general distribution theory for a class of likelihood criteria. *Biometrika* 36: 317–346.

Cai, T., Shen, X.T. (2010). *High-Dimensional Data Analysis*. Beijing: Higher Education Press & World Scientific.

Donoho, D. L., et al. (2000). High-dimensional data analysis: The curses and blessings of dimensionality. In: American Mathematical Society Conf. Math. Challenges of the 21st Century.

Fan, J.Q., Li, R.Z. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Am. Stat. Assoc.* 96:1348–1360.

Fan, J.Q., Lv, J.C. (2008). Sure independence screening for ultra-high dimensional feature space. *J. Royal Stat. Soc. Ser. B* 70:849–911.

Feng, L., Zou, C.L., Wang, Z.J., Chen, B. (2013). Rank-based score tests for high-dimensional regression coefficients. *Electr. J. Stat.* 7:2131–2149.

Fujikoshi, Y., Ulyanov, V.V., Shimizu, R. (2010). *Multivariate Statistics: High Dimensional and Large Sample Approximations*. New York: John Wiley & Sons Inc.

Goeman, J., Finos, L., van Houwelingen, J.C. (2011). Testing against a high dimensional alternative in the generalized linear model: Asymptotic type I error control. *Biometrika* 98:381–390.

Gut, A. (2005). *Probability: A Graduate Course*. New York: Springer-Verlag.

Jiang, T.F., Qi, Y.C. (2015). Likelihood ratio tests for high-dimensional normal distributions. *Scand. J. Stat.* 42(4):988–1009.

Jiang, T.F., Yang, F. (2013). Central limit theorems for classical likelihood ratio tests for high-dimensional normal distributions. *Ann. Stat.* 41(4):2029–2074.

Muirhead, R.J. (1982). *Aspects of Multivariate Statistical Theory*. New York: John Wiley & Sons Inc.

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *J. Royal Stat. Soc. Ser. B* 58: 267–288.

Wang, H.S. (2009). Forward regression for ultra-high dimensional variable screening. *J. Am. Stat. Assoc.* 104:1512–1524.

Wang, S.Y., Cui, H.J. (2013). Generalized F test for high dimensional linear regression coefficients. *J. Multivariate Anal.* 117:134–149.

Wang, S.Y., Cui, H.J. (2015). A new test for part of high dimensional regression coefficients. *J. Multivariate Anal.* 137:187–203.

Wen, L., Wang, H.S., Tsai, C.L. (2014). Testing covariatesin high dimensional regression. *Ann. Inst. Stat. Math.* 66(2):279–301.

Zang, Y.G., Zhang, S.G., Li, Q.Z., Zhang, Q.Z. (2016). Jackknife empirical likelihood test for high-dimensional regression coefficients. *Comput. Stat. Data. Anal.* 94:302–316.

Zhong, P.S., Chen, S.X. (2011). Test for high-dimensional regression coefficients with factorial designs. *J. Am. Stat. Assoc.* 106(493):260–274.

Downloaded by [Australian National University] at 00:21 02 October 2017