Statistics and Probability Letters 86 (2014) 85–90

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Empirical likelihood test for high dimensional linear models

Liang Peng [a], Yongcheng Qi [b], Ruodu Wang [c,*]

[a] School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332-0160, USA

[b] Department of Mathematics and Statistics, University of Minnesota–Duluth, 1117 University Drive, Duluth, MN 55812, USA

[c] Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, ON N2L3G1, Canada

ARTICLE INFO

Article history:
Received 13 August 2013
Received in revised form 25 December 2013
Accepted 25 December 2013
Available online 3 January 2014

Keywords:
Empirical likelihood
High-dimensional data
Hypothesis test
Linear model

ABSTRACT

方法 + 问题

We propose an empirical likelihood method to test whether the coefficients in a possibly high-dimensional linear model are equal to given values. The asymptotic distribution of the test statistic is independent of the number of covariates in the linear model.

1. Introduction

Regression model is a commonly employed technique to model the relationship between responses and covariates. Consider the following classical and also the simplest linear regression model

$$Y_i = \beta^T X_i + \epsilon_i, \quad i = 1, \dots, n, \tag{1}$$

where $\beta = (\beta_1, \dots, \beta_p)^T$ is the vector of unknown parameters, $X_1 = (X_{1,1}, \dots, X_{1,p})^T, \dots, X_n = (X_{n,1}, \dots, X_{n,p})^T$ are independent and identically distributed (iid) random vectors, and $\epsilon_1, \dots, \epsilon_n$ are iid random variables with zero mean and variance σ^2 with ϵ_i's being independent of X_i's. Statistical inference for β can be based on either least squares estimator or M-estimator when p is fixed. When p depends on the sample size n and goes to infinity as $n \to \infty$, Portnoy (1984, 1985) studied the consistency and asymptotic normality of M-estimators for β, which requires that p cannot be too large in comparison with the sample size.

Statistical inference for the linear model (1) is needed for the case when p is of an exponential order of n, motivated by the studies in bioinformatics. To deal with the case when many of β_i's are zero (sparsity), one first selects variables with nonzero $\beta_i's$ and then makes statistical inference for the selected nonzero $\beta_i's$. It is not surprising that the order of the number of nonzero $\beta_i's$ cannot be larger than the optimal one in Portnoy (1985). We refer to Bradic et al. (2011) for more details and references on the ultrahigh dimensional situation. Sparse estimators like the famous Lasso estimator (Tibshirani, 1996) and its extensions (Zou, 2006; Meinshausen, 2007) are very powerful in the setting of sparse alternative. Meinshausen et al. (2009) studied the variable selection for high-dimensional linear regression models.

On the other hand, when the number of nonzero β's is large, new techniques are needed. In contrast to the sparse model and variable selection techniques, we study a general setting in this paper. We consider the problem of testing $H_0 : \beta = \beta_0$ against $H_a : \beta \neq \beta_0$ for a given value $\beta_0 \in \mathbb{R}^p$ when p is either fixed or goes to infinity as $n \to \infty$. In particular, we are

* Corresponding author.
E-mail addresses: peng@math.gatech.edu (L. Peng), yqi@d.umn.edu (Y. Qi), wang@uwaterloo.ca (R. Wang).

0167-7152/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.spl.2013.12.019

interested in the case when the alternative hypothesis has a dense shift (i.e. small shifts in many dimensions instead of large shifts in a few dimensions). When p is fixed, the traditional test is Hotelling's T^2 test, based on the test statistic

$$HT = \frac{1}{\hat{\sigma}^2}(\hat{\beta} - \beta_0)^T \left(\frac{1}{n}\sum_{i=1}^n X_i X_i^T\right)^{-1} (\hat{\beta} - \beta_0), \tag{2}$$

where $\hat{\beta} = (\frac{1}{n}\sum_{i=1}^n X_i X_i^T)^{-1}\frac{1}{n}\sum_{i=1}^n Y_i X_i$ and $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (Y_i - \hat{\beta}^T X_i)^2$. It is known that $HT \xrightarrow{d} \chi_p^2$ as $n \to \infty$. However, when p is large, finding the inverse matrix in (2) becomes problematic. To overcome such difficulty, we consider the empirical likelihood method.

As a powerful nonparametric likelihood approach, empirical likelihood test is another useful method. More specifically, define the traditional empirical likelihood function for β as

$$L_n^{(T)}(\beta) = \sup\left\{\prod_{i=1}^n (nq_i) : q_1 \ge 0, \dots, q_n \ge 0, \sum_{i=1}^n q_i = 1, \sum_{i=1}^n q_i (Y_i - \beta^T X_i) X_i = 0\right\}.$$

Under some regularity conditions, one can show that the Wilks theorem holds, i.e., $-2\log L_n^{(T)}(\beta_0)$ converges in distribution to a chi-square limit with p degrees of freedom. Therefore, the empirical likelihood test can be constructed by using the test statistic $-2\log L_n^{(T)}(\beta)$. See Owen (2001) for more details on empirical likelihood methods. However, the maximization in computing $L_n^{(T)}(\beta)$ becomes nontrivial and even unavailable when p is large; see Chen et al. (2008) for discussions on this phenomenon. Empirical likelihood method for high dimensional data can be found in Chen et al. (2009), Hjort et al. (2009) and Peng and Schick (2013).

In this paper we propose a new empirical likelihood test for testing $H_0 : \beta = \beta_0$ against $H_a : \beta \ne \beta_0$ regardless of fixed or divergent p. We begin with an estimator of $\theta = (\beta_0 - \beta)^T \Sigma^2 (\beta_0 - \beta)$ where $\Sigma = \mathbb{E}(X_1 X_1^T)$. It is obvious that when Σ is positive definite, testing $H_0 : \beta = \beta_0$ against $H_a : \beta \ne \beta_0$ is equivalent to testing $H_0 : \theta = 0$ against $H_a : \theta \ne 0$. To find such an estimator, we split the data into two parts and introduce an empirical likelihood test based on this estimator. It turns out that the new method works for both fixed and divergent p. The sample splitting method was also used and discussed in Peng et al. (in press) and Wang et al. (2013), where they proposed empirical likelihood tests and jackknife empirical likelihood tests for high-dimensional means. Other methods based on sample splitting techniques for variable selection were discussed in Wasserman and Roeder (2009) and Meinshausen et al. (2009). Note that the purpose of sample splitting in this paper is for testing without variable selection, and hence it is different from their methods.

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

2. Methodology

eg: $n=4,\ m=2$; $X_1\ X_2\ \underset{X_{2+1}}{X_3}\ \underset{X_{2+2}}{X_4}$ $n=5,\ m=[2.5]=2$ $X_1\ X_2\ \underset{X_{2+1}}{X_3}\ \underset{X_{2+2}}{X_4}\ X_5$

We start by splitting the sample into two groups to get a random sample with mean being $\theta = (\beta_0 - \beta)^T \Sigma^2 (\beta_0 - \beta)$, where $\Sigma = \mathbb{E}(X_1 X_1^T)$. Put $\boxed{m = [n/2]}$, the integer part of $n/2$, and define $\tilde{X}_i = X_{m+i}$, $\tilde{Y}_i = Y_{i+m}$, $\tilde{\epsilon}_i = \epsilon_{i+m}$,

$$W_i(\beta) = (\underbrace{Y_i}_{X_i^T\beta_0+\epsilon_i} X_i - X_i X_i^T \beta)^T (\tilde{Y}_i \tilde{X}_i - \tilde{X}_i \tilde{X}_i^T \beta) \quad 1\times 1 \qquad (a_1, a_2)\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = a_1 b_1 + a_2 b_2 \le \sqrt{a_1^2 + a_2^2}\cdot\sqrt{b_1^2 + b_2^2}$$

for $i = 1, \dots, m$. Then

$$\mathbb{E}W_i(\beta_0) = \mathbb{E}[(X_i X_i^T(\beta_0 - \beta) + X_i\epsilon_i)^T (\tilde{X}_i \tilde{X}_i^T(\beta_0 - \beta) + \tilde{X}_i\tilde{\epsilon}_i)] = (\beta_0 - \beta)^T \Sigma^2 (\beta_0 - \beta).$$

When Σ is positive definite, testing $H_0 : \beta = \beta_0$ against $H_a : \beta \ne \beta_0$ is equivalent to testing $H_0 : \mathbb{E}W_1(\beta_0) = 0$ against $H_a : \mathbb{E}W_1(\beta_0) \ne 0$. This motivates us to apply the empirical likelihood method in Qin and Lawless (1994) to the estimating equation $\mathbb{E}W_1(\beta_0) = 0$. However this direct application results in a poor power in general by noting that $\mathbb{E}W_1(\beta_0) = O(\|\beta - \beta_0\|^2)$ instead of $O(\|\beta - \beta_0\|)$ when $\|\beta - \beta_0\|$ is small, where $\|\cdot\|$ denotes the L_2 norm of a vector. The explanation of this weak power is discussed in Peng et al. (in press).

To improve the power, we propose to add one more linear equation $\mathbb{E}W_1^*(\beta_0) = 0$ where $\mathbb{E}W_1^*(\beta_0)$ is close to $O(\|\beta - \beta_0\|_1)$ where $\|\beta - \beta_0\|_1$ is the L_1 norm, and thus it captures the small change of $\beta - \beta_0$. More specifically, define

$$W_i^*(\beta) = (Y_i X_i - X_i X_i^T \beta)^T \mathbf{1}_p + (\tilde{Y}_i \tilde{X}_i - \tilde{X}_i \tilde{X}_i^T \beta)^T \mathbf{1}_p \qquad (a_1, a_2)\begin{pmatrix} 1 \\ 1 \end{pmatrix} + (b_1, b_2)\begin{pmatrix} 1 \\ 1 \end{pmatrix} = a_1 + a_2 + b_1 + b_2$$

for $i = 1, \dots, m$, where $\mathbf{1}_p = (1, 1, \dots, 1)^T \in \mathbb{R}^p$, and then define the empirical likelihood function for β as

$$L_n(\beta) = \sup\left\{\prod_{i=1}^m (mq_i) : q_1 \ge 0, \dots, q_m \ge 0, \sum_{i=1}^m q_i = 1, \sum_{i=1}^m q_i W_i(\beta) = 0, \sum_{i=1}^m q_i W_i^*(\beta) = 0\right\}.$$

The following theorem shows that the Wilks theorem holds for the above empirical likelihood method. We use $\mathbf{tr}(A)$ to denote the trace of a matrix A.

Theorem 1. *Let β_0 be the true value of the parameter β. Assume Σ is positive definite and there exists some $\delta > 0$ such that*

$$\frac{\mathbb{E}|X_1^T \tilde{X}_1|^{2+\delta}}{\{\mathbf{tr}(\Sigma^2)\}^{(2+\delta)/2}} \left(\frac{\mathbb{E}|\epsilon_1|^{2+\delta}}{\sigma^{2+\delta}}\right)^2 = o(\underbrace{m}_{m=[\frac{n}{2}]}{}^{\delta/2}), \tag{3}$$

and

$$\frac{\mathbb{E}|X_1^T \mathbf{1}_p|^{2+\delta}}{\{\mathbb{E}(X_1^T \mathbf{1}_p)^2\}^{(2+\delta)/2}} \left(\frac{\mathbb{E}|\epsilon_1|^{2+\delta}}{\sigma^{2+\delta}} \right) = o(m^{\delta/2}), \tag{4}$$

where $\sigma^2 = \mathrm{Var}(\epsilon_1)$. *Then* $-2 \log L_n(\beta_0)$ *converges in distribution to a chi-square limit with 2 degrees of freedom.*

Remark 1. The distribution of X_1 varies with n when the dimension of X_1 changes with n. In general, the distribution of the error term ϵ_1 may also change with n and thus the moments of ϵ_1 may not be constants.

Remark 2. Recently Tang and Leng (2010) proposed penalized empirical likelihood method for selecting variables and showed the Wilks theorem for the case of finite number of constraints. This is different from the above theorem, where the number of constraints is either fixed or divergent.

Conditions (3) and (4) may impose some restrictions on p implicitly. When the dependence on X_i has some special structures, we can show that indeed little restriction on p is required in Theorem 1; see the following two examples with proofs given in Section 4.

Example 1. Let X_1 be a Gaussian random vector with mean 0 and covariance matrix $\Sigma = (\sigma_{i,j})_{1 \le i,j \le p}$, where Σ is an arbitrary p by p positively definite matrix. Assume $\mathbb{E}(\epsilon_1^4)/\sigma^4 = o(m^{1/2})$, then conditions (3) and (4) hold.

Example 2. Let $0 < \lambda_1 \le \lambda_2 \le \cdots \le \lambda_p$ be the p eigenvalues of Σ. Assume that

A1: $0 < \liminf_{n\to\infty} \lambda_1 \le \limsup_{n\to\infty} \lambda_p < \infty$, $\liminf_{n\to\infty} \sigma^2 > 0$;
A2: For some $\delta > 0$, $\frac{1}{p} \sum_{i=1}^p \mathbb{E}[|X_{1,i}|^{2+\delta}] = O(1)$, $\mathbb{E}|\epsilon_1|^{2+\delta} = O(1)$;
A3: $p = o(m^{\frac{\delta}{2+\delta}})$.

Then conditions (3) and (4) hold.

Note that Example 1 assumes a special dependence structure, which is just a special case of the following model considered by Bai and Saranadasa (1996), Chen et al. (2009), and Chen and Qin (2010):

$$X_i = \Gamma B_i + \mu, \quad 1 \le i \le n,$$

where Γ is $p \times k$ matrix with $\Gamma\Gamma^T = \Sigma$, $k \ge p$, and μ is a non-random vector, and $\{B_i = (B_{i,1}, \ldots, B_{i,k})^T, 1 \le i \le n\}$ are i.i.d. random k-vectors with $\mathbb{E}B_i = 0$, $\mathrm{Var}(B_i) = I_{k\times k}$, $\mathbb{E}B_{i,j}^4 = 3 + \xi < \infty$ and $\mathbb{E}\prod_{l=1}^k B_{i,l}^{\nu_l} = \prod_{l=1}^k \mathbb{E}B_{i,l}^{\nu_l}$ whenever $\nu_1 + \cdots + \nu_k = 4$ for nonnegative integers ν_l's. One can also show that Theorem 1 holds for the above model. Example 2 imposes moment conditions A1 and A2 on X_1 and ϵ_1, and the only restriction on p is A3.

One advantage of the proposed empirical likelihood method is that one can easily add more equations if one has more information on the alternative hypothesis, or replace $W_1^*(\beta)$ by another statistic $\bar{W}_1(\beta)$ satisfying $\mathbb{E}\bar{W}_1(\beta) = O(\|\beta - \beta_0\|_1)$. Although adding more relevant equations may improve the test power, computing the empirical likelihood function becomes more complicated. The simulation study in the next section shows that the test of using $\mathbb{E}W_i(\beta) = 0$ and $\mathbb{E}W_i^*(\beta) = 0$ in Theorem 1 performs well in terms of both size and power.

3. Simulation study

In this section, we examine the finite sample behavior of the proposed empirical likelihood test and compare it with Hotelling's T^2 test and the standard empirical likelihood method in terms of both size and power. Note that it is expected that Hotelling's T^2 test and the standard empirical likelihood test do not work for large p. For the case of dense shift with large p, there are no existing tests in the literature which are proper to compare with.

We draw 10,000 random samples with size $n = 200$ and 1000 from the linear model (1) with $X_i = (X_{i1}, \ldots, X_{ip})^T \sim N(0, \Sigma_0)$, $\Sigma_0 = (0.5^{(|i-j|)})_{1\le i,j\le p}$, $\epsilon_i \sim t_8$ and $\beta = \beta_0 + \Delta/\sqrt{n}$ for some $\Delta \ge 0$, $\beta_0 = \mathbf{1}_p$. Note that when $\Delta > 0$, the alternative hypothesis in this model has a dense shift. Consider testing $H_0: \beta = \beta_0$ against $H_a: \beta \ne \beta_0$. We use TEL, NEL and HT to denote the traditional empirical likelihood test based on $-2\log L_n^{(T)}(\beta)$, the new empirical likelihood test based on $-2\log L_n(\beta)$, and Hotelling's T^2 test in (2).

We compute the sizes ($\Delta = 0$) and powers ($\Delta = 0.3$) of these three tests at level 0.05 in Tables 1 and 2 for $p = 5$, 10, 20, ..., 100.

As expected, the traditional empirical likelihood method and Hotelling's T^2 test do not have a consistent size when p is slightly large. Hence, the power for those tests does not make real sense. In contrast the proposed empirical likelihood test has a very stable size with respect to p and is powerful too. When n becomes large, the size of the proposed empirical likelihood test is more accurate.

In summary, the proposed empirical likelihood test has a very stable size with respect to the number of covariates and is powerful too. Moreover the proposed new test is quite easy to implement by using the R package `emplik`, which does not need to compute the inverse of a high dimensional covariance matrix. Our method does not distinguish between p is fixed or $p \to \infty$, and works particularly well in the dense setting of the alternative hypothesis.

4. Proofs

Throughout we denote

$$u_i := W_i(\beta_0) = (X_i^T \tilde{X}_i)\epsilon_i \tilde{\epsilon}_i, \qquad v_i := W_i^*(\beta_0) = (X_i^T \mathbf{1}_p)\epsilon_i + (\tilde{X}_i^T \mathbf{1}_p)\tilde{\epsilon}_i,$$

$$\sigma_1 = \sqrt{\mathrm{Var}(u_1)} \quad \text{and} \quad \sigma_2 = \sqrt{\mathrm{Var}(v_1)}.$$

Then it is easy to verify that $\mathbb{E}(u_1) = \mathbb{E}(v_1) = \mathbb{E}(u_1 v_1) = 0$. One can also easily show that conditions (3) and (4) imply

$$S := \mathbb{E}\left[\left|\frac{u_1}{\sigma_1}\right|^{2+\delta}\right] + \mathbb{E}\left[\left|\frac{v_1}{\sigma_2}\right|^{2+\delta}\right] = o(m^{\frac{\delta}{2}}). \tag{5}$$

Lemma 1. *Under conditions of Theorem 1, we have*

$$\frac{1}{\sqrt{m}} \sum_{i=1}^m \begin{pmatrix} u_i/\sigma_1 \\ v_i/\sigma_2 \end{pmatrix} \xrightarrow{d} N(0, I_2), \tag{6}$$

$$\frac{\sum_{i=1}^m u_i^2}{m\sigma_1^2} - 1 \xrightarrow{p} 0, \tag{7}$$

$$\frac{\sum_{i=1}^m v_i^2}{m\sigma_2^2} - 1 \xrightarrow{p} 0, \tag{8}$$

$$\frac{\sum_{i=1}^m u_i v_i}{m\sigma_1\sigma_2} \xrightarrow{p} 0, \tag{9}$$

$$\max_{1\le i\le m}\left|\frac{u_i}{\sigma_1}\right| = o_p(m^{1/2}) \quad \text{and} \quad \max_{1\le i\le m}\left|\frac{v_i}{\sigma_2}\right| = o_p(m^{1/2}), \tag{10}$$

where I_2 is a 2×2 identity matrix.

Proof. For any constants a and b with $a^2 + b^2 = 1$, let $Z_i = au_i/\sigma_1 + bv_i/\sigma_2$, $i = 1, \ldots, m$. It is clear that $Z_i,\ i = 1, \ldots, m$ are iid random variables with mean zero and variance one. To show (6)–(10) it suffices to prove that

$$\frac{1}{\sqrt{m}} \sum_{i=1}^m Z_i \xrightarrow{d} N(0, 1), \tag{11}$$

$$\frac{1}{m} \sum_{i=1}^m Z_i^2 \xrightarrow{p} 1, \tag{12}$$

and

$$\max_{1\le i\le m} Z_i = o_p(m^{1/2}) \tag{13}$$

hold for all constants a and b with $a^2 + b^2 = 1$. Indeed, (6) follows from (11), (7)–(9) follow from (12), and (10) follows from (13), by choosing different values of a and b. We check that $Z_i,\ i = 1, \ldots, m$ satisfy the Lindeberg condition for $\eta > 0$:

$$M_m(\eta) := \mathbb{E}[Z_1^2 1(|Z_1| > \eta m^{1/2})] \le \frac{\mathbb{E}[|Z_1|^{2+\delta}]}{\eta^\delta m^{\delta/2}} \le \frac{(|a| + |b|)^{2+\delta} S}{\eta^\delta m^{\delta/2}} \to 0. \tag{14}$$

(11) and (13) follow from (14). To show (12), let

$$T_1 = \frac{1}{m} \sum_{i=1}^m \left(Z_i^2 1(|Z_i| \le \eta m^{1/2}) - \mathbb{E}[Z_i^2 1(|Z_i| \le \eta m^{1/2})]\right),$$

and

$$T_2 = \frac{1}{m} \sum_{i=1}^m \left(Z_i^2 1(|Z_i| > \eta m^{1/2}) - \mathbb{E}[Z_i^2 1(|Z_i| > \eta m^{1/2})]\right).$$

It is straightforward to compute

$$\mathbb{E}[T_1^2] \le \frac{1}{m}\mathbb{E}[Z_1^4 11(|Z_i| \le \eta m^{1/2})] \le \eta^2 \mathbb{E}[Z_1^2] = \eta^2,$$

and $\mathbb{E}[T_2] \le 2M_m(\eta) \to 0$. Therefore, we have that

$$\limsup_{m\to\infty} \mathbb{E}[|T_1 + T_2|] \le \eta,$$

and since η is arbitrary, we have

$$\limsup_{m\to\infty} \mathbb{E}[|T_1 + T_2|] = \limsup_{m\to\infty} \mathbb{E}\left[\left|\frac{1}{m}\sum_{i=1}^{m} Z_i^2 - 1\right|\right] = 0$$

and (12) follows.

Proof of Theorem 1. The theorem follows from Theorem 6.1 of Peng and Schick (2013) and Lemma 1, and so we skip all details.

Proof of Example 1. Let $Z = O\Sigma^{-1/2}X_1$, and $\tilde{Z} = O\Sigma^{-1/2}\tilde{X}$ where O is an orthogonal matrix such that $\Lambda := O\Sigma O^T$ is diagonal. It is obvious that Z and $\tilde{Z}$ are iid standard Gaussian random vectors. It follows that

$$X_1^T\tilde{X}_1 = Z^T O\Sigma O^T\tilde{Z} = Z^T\Lambda\tilde{Z} = \sum_{i=1}^{p} \lambda_i Z_i \tilde{Z}_i,$$

where $\lambda_1, \ldots, \lambda_p$ are eigenvalues of Σ. Then we have that

$$\mathbb{E}[(X_1^T\tilde{X}_1)^4] = \mathbb{E}\left[\left(\sum_{i=1}^{p} \lambda_i Z_i \tilde{Z}_i\right)^4\right] \le 9\sum_{i=1}^{p}\sum_{i=j}^{p} \lambda_i^2\lambda_j^2 = 9(\mathbf{tr}(\Sigma^2))^2.$$

Thus we have that $\mathbb{E}[(X_1^T\tilde{X}_1)^4]/(\mathbf{tr}(\Sigma^2))^2 = O(1)$ is bounded uniformly for p.

Similarly, we can show that the first term on the left-hand side of (4) is also bounded uniformly for p. Therefore, conditions (3) and (4) will be fulfilled with $\delta = 2$ for any p if $\mathbb{E}(\epsilon_1^4)/\sigma^4 = o(m^{1/2})$. □

Proof of Example 2. By A1 and A2, obviously $\mathbb{E}|\epsilon_1|^{2+\delta}/\sigma^{2+\delta} = O(1)$. It follows from the Cauchy–Schwarz inequality that

$$|X_1^T\tilde{X}_1|^2 \le \|X_1\|^2\|X_{m+1}\|^2.$$

Then by using the C_r inequality we conclude that

$$\begin{aligned}
\mathbb{E}|X_1^T\tilde{X}_1|^{2+\delta} &\le \mathbb{E}\left(\sum_{i=1}^{p} X_{1,i}^2\right)^{(2+\delta)/2} \mathbb{E}\left(\sum_{i=1}^{p} X_{(m+1)i}^2\right)^{(2+\delta)/2} \\
&= \left(\mathbb{E}\left(\sum_{i=1}^{p} X_{1,i}^2\right)^{(2+\delta)/2}\right)^2 \\
&= \left(p^{\delta/2}\sum_{i=1}^{p}\mathbb{E}|X_{1,i}|^{2+\delta}\right)^2 \\
&= p^{\delta}\left(\sum_{i=1}^{p}\mathbb{E}|X_{1,i}|^{2+\delta}\right)^2.
\end{aligned}$$

By the above inequality, A1 and A2 we have $\frac{\mathbb{E}|X_1^T\tilde{X}_1|^{2+\delta}}{\{\mathbf{tr}(\Sigma^2)\}^{(2+\delta)/2}} = O(p^{(2+\delta)/2})$, and by A3 and the fact $\mathbb{E}|\epsilon_1|^{2+\delta}/\sigma^{2+\delta} = O(1)$ we have (3). Similarly, from the C_r inequality we have

$$\mathbb{E}|X_1^T\mathbf{1}_p|^{2+\delta} \le \mathbb{E}\left(\sum_{i=1}^{p}|X_{1,i}|\right)^{2+\delta} \le p^{1+\delta}\sum_{i=1}^{p}\mathbb{E}|X_{1,i}|^{2+\delta},$$

and (4) follows from A1–A3. □

Table 1
Sizes of the traditional empirical likelihood test (TEL), the proposed empirical likelihood test (NEL) and Hotelling's T^2 test at level 5% are given for the case of $\Delta = 0$.

p	TEL $n = 200$	NEL $n = 200$	HT $n = 200$	TEL $n = 1000$	NEL $n = 1000$	HT $n = 1000$
5	0.0811	0.0875	0.0620	0.0541	0.0582	0.0534
10	0.1327	0.0816	0.0760	0.0591	0.0554	0.0539
20	0.3104	0.0835	0.1173	0.0840	0.0557	0.0614
30	0.5735	0.0820	0.1992	0.1163	0.0550	0.0659
40	0.8180	0.0802	0.3117	0.1575	0.0521	0.0790
50	0.9510	0.0821	0.4724	0.2237	0.0609	0.0878
60	0.9933	0.0859	0.6343	0.2959	0.0589	0.1123
70	0.9993	0.0828	0.7866	0.3902	0.0577	0.1272
80	0.9997	0.0867	0.9040	0.4897	0.0590	0.1472
90	1	0.0891	0.9655	0.5926	0.0621	0.1830
100	1	0.0817	0.9913	0.6987	0.0555	0.2059

Table 2
Powers of the traditional empirical likelihood test (TEL), the proposed empirical likelihood test (NEL) and Hotelling's T^2 test at level 5% are given for the case of $\Delta = 0.3$.

p	TEL $n = 200$	NEL $n = 200$	HT $n = 200$	TEL $n = 1000$	NEL $n = 1000$	HT $n = 1000$
5	0.1264	0.1550	0.1006	0.0938	0.1266	0.0882
10	0.2250	0.2443	0.1480	0.1255	0.2062	0.1144
20	0.5046	0.4386	0.2725	0.2127	0.4012	0.1723
30	0.7841	0.6006	0.4192	0.3069	0.5680	0.2180
40	0.9454	0.7209	0.5922	0.4193	0.6977	0.2740
50	0.9902	0.8001	0.7565	0.5450	0.8078	0.3445
60	0.9997	0.8665	0.8742	0.6656	0.8785	0.4181
70	1	0.9077	0.9474	0.7618	0.9211	0.4685
80	1	0.9423	0.9859	0.8501	0.9526	0.5389
90	1	0.9624	0.9980	0.9096	0.9693	0.6013
100	1	0.9747	0.9996	0.9550	0.9812	0.6724

Acknowledgments

We thank the reviewer for constructive comments, which weaken the conditions in Theorem 1 and simplify the proofs of Lemma 1, Example 1 and Theorem 1. Peng's research was supported by the NSF Grant DMS-1005336, Qi's research was supported by the NSF Grant DMS-1005345 and Wang's research was supported by the Natural Sciences and Engineering Research Council of Canada.

References

Bai, Z., Saranadasa, H., 1996. Effect of high dimension: by an example of a two sample problem. Statist. Sinica 6, 311–329.
Bradic, J., Fan, J., Wang, W., 2011. Penalized composite quasi-likelihood for ultrahigh dimensional variable selection. J. R. Stat. Soc. Ser. B Stat. Methodol. 73, 325–349.
Chen, S., Peng, L., Qin, Y., 2009. Empirical likelihood methods for high dimension. Biometrika 96, 711–722.
Chen, S., Qin, Y., 2010. A two-sample test for high-dimensional data with applications to gene-set testing. Ann. Statist. 38, 808–835.
Chen, J., Variyath, A.M., Abraham, B., 2008. Adjusted empirical likelihood and its properties. J. Comput. Graph. Statist. 17, 426–443.
Hjort, N.L., McKeague, I.W., Van Keilegom, I., 2009. Extending the scope of empirical likelihood. Ann. Statist. 37, 1079–1111.
Meinshausen, N., 2007. Relaxed lasso. Comput. Statist. Data Anal. 52, 374–393.
Meinshausen, N., Meier, L., Bühlmann, P., 2009. P-values for high-dimensional regression. J. Amer. Statist. Assoc. 104, 1671–1681.
Owen, A., 2001. Empirical Likelihood. Chapman & Hall/CRC.
Peng, L., Qi, Y., Wang, F., 2014. Empirical likelihood method for high dimensional means. Statist. Sci. (in press).
Peng, H., Schick, A., 2013. An empirical likelihood approach to goodness of fit testing. Bernoulli 19 (3), 954–981.
Portnoy, S., 1984. Asymptotic behavior of M-estimators of p regression parameters when p^2/n is large; I, consistency. Ann. Statist. 12, 1298–1309.
Portnoy, S., 1985. Asymptotic behavior of M-estimators of p regression parameters when p^2/n is large; II, normal approximation. Ann. Statist. 13, 1403–1417.
Qin, J., Lawless, J., 1994. Empirical likelihood and general estimating equations. Ann. Statist. 22, 300–325.
Tang, C.Y., Leng, C., 2010. Penalized high dimensional empirical likelihood. Biometrika 97, 905–920.
Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. J. R. Stat. Soc. Ser. B Stat. Methodol. 58, 267–288.
Wang, R., Peng, L., Qi, Y., 2013. Jackknife empirical likelihood test for the equality of two high dimensional means. Statist. Sin. 23, 667–690.
Wasserman, L., Roeder, K., 2009. High dimensional variable selection. Ann. Statist. 37, 2178–2201.
Zou, H., 2006. The adaptive Lasso and its oracle properties. J. Amer. Statist. Assoc. 101, 1418–1429.