Statistica Sinica **25** (2015), 1659-1677
doi:http://dx.doi.org/10.5705/ss.2013.246

LARGE DIMENSIONAL EMPIRICAL LIKELIHOOD

Binbin Chen[1], Guangming Pan[1], Qing Yang[1] and Wang Zhou[2]

[1]*Nanyang Technological University and* [2]*National University of Singapore*

Abstract: The empirical likelihood is a versatile nonparametric approach to testing hypotheses and constructing confidence regions. However it is not clear if Wilks' Theorem still works in high dimensions. In this paper, by adding two pseudo-observations to the original data set, we prove the asymptotic normality of the log empirical likelihood-ratio statistic when the sample size and the data dimension are comparable. In practice, we suggest using the normalized $F(p, n-p)$ distribution to approximate its distribution. Simulation results show excellent performance of this approximation.

Key words and phrases: Empirical likelihood, large dimensional data.

1. Introduction

Empirical likelihood (EL) method, introduced by Owen (1988, 1990), has been shown to perform remarkably well in a wide range of settings as a tool for nonparametric and semiparametric inference. Its advantage is that EL provides nonparametric analogs of parametric likelihood-based tests and confidence regions while keeping two key properties of the conventional likelihood: Wilks' theorem and Bartlett correction. Applications of EL in statistical inference include the mean of a distribution, quantiles of a distribution, (censored) linear models (Qin and Jing (2001)), partial linear models (Wang and Jing (1999)), estimating equations (Qin and Lawless (1994)) and more.

In this paper, we focus on the EL for the population mean, an important application. Suppose $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ is a sample of n independent and identically distributed (i.i.d.) p-dimensional vectors, distributed according to some unknown distribution F. Let μ and Σ be the p-dimensional population mean vector and the $p \times p$ positive definite population covariance matrix, respectively. The EL ratio statistic for any hypothesized value μ_0 is

$$R_n(\mu_0) = \max\left\{\prod_{i=1}^{n} n\omega_i : \omega_i \geq 0, \sum_{i=1}^{n}\omega_i = 1, \sum_{i=1}^{n}\omega_i \mathbf{x}_i = \mu_0\right\}. \tag{1.1}$$

For a fixed dimension p, Owen (1990) proved that the log EL ratio statistic

$$W(\mu) = -2\log R_n(\mu) \xrightarrow{d} \chi^2_p \tag{1.2}$$

as $n \to \infty$, similar to the parametric likelihood (Wilks (1938)). Here, "$\xrightarrow{d}$" means "convergence in distribution". Then a level-α EL confidence interval for the mean μ is formed by the set $\{\mu_0 : W(\mu_0) \le \chi^2_{p,\alpha}\}$, where $\chi^2_{p,\alpha}$ is the $(1-\alpha)$-quantile of the chi-square distribution with degrees of freedom p.

High dimensional data analysis has attracted widespread attention in scientific areas, bringing new challenges to the EL method. One question is how to formulate (1.2) when p tends to infinity. A natural extension of (1.2) is

$$\frac{1}{\sqrt{2p}}(W(\mu) - p) \xrightarrow{d} N(0,1) \tag{1.3}$$

when both n and p tend to infinity, since χ^2_p is asymptotically normal with mean p and variance $2p$. Hjort, Mckeague, and Van Keilegom (2009) argued that (1.3) still holds when $p = o(n^{1/3})$, under some mild conditions. Chen, Peng and Qin (2009) provided a general rate for the dimension p, which is shown to depend on the trace of the population covariance matrix Σ and the largest eigenvalue of Σ but, roughly speaking, $p = o(n^{1/2})$. Then a natural question is whether we can apply the EL approach to a higher order dimensional data compared with the sample size, say, p is proportional to n.

The answer to this is negative if one only considers the usual log EL ratio statistic. It was pointed out by Chen, Peng, and Qin (2009) that $p = o(n^{1/2})$ is likely the best rate for p such that the log EL ratio is asymptotically normal. When p is relatively large compared to n, care is needed with the EL approach, e.g., to check whether there exists a solution $\omega_i, i = 1, \ldots, n$ to (1.1). In fact, a solution to (1.1) exists only if the zero vector is an interior point of the convex hull of $\{\mathbf{x}_i - \mu, i = 1, \ldots, n\}$. It was noticed in Tsao (2004) that for fixed p and n with $p/n > 1/2$, the EL for a p-dimensional population mean breaks down with a positive probability, so the standard EL method is not then reliable. The situation is definitely more serious for high-dimensional data. A number of suggestions have been proposed for improving the behavior of the EL ratio statistic, mainly in the small sample setting. Among them, adding artificial data points to the observed sample is an easy and efficient approach that solves the convex hull problem (see Chen, Variyath, and Abraham (2008), Emerson and Owen (2009)). Interestingly, this simple strategy allows the EL to perform well in a high-dimensional setting.

In this paper, we extend the scope of the EL method to high-dimensional data for $p/n = c_n \to c \in (0,1)$ by adopting the method of Emerson and Owen (2009), adding two points to the data set, to address the convex hull problem. The asymptotic normality of $W(\mu)$ is established. Our result extends results in Chen, Peng, and Qin (2009), where $p = o(n^{1/2})$, and in Hjort, Mckeague, and Van Keilegom (2009), where $p = o(n^{1/3})$.

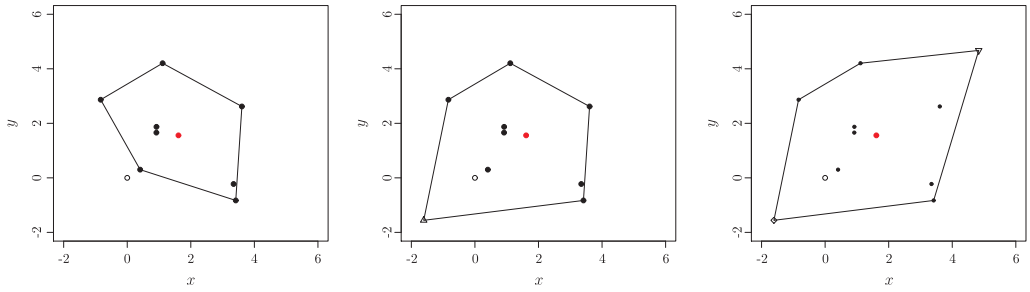

Figure 1. The left figure shows an example in which the zero vector is not contained in the convex hull of the original data set; The middle figure shows the zero vector is contained in the convex hull of the new data set to which one point is added (see Chen, Variyath, and Abraham (2008)); The right figure shows the convex hull of the data set to which two points are added (see Emerson and Owen (2009)).

Another finding of this paper is that using the normalized $F(p, n-p)$ distribution to approximate $W(\mu)$ is more appropriate than using the normal approximation, see our simulation results. The rest of the paper is organized as follows. The main theorems are stated in Section 2. Simulations are reported in Section 3. A case study is given in Section 4. We sketch proofs in the appendix and leave the details to the supplementary document.

2. Balanced Adjusted Empirical Likelihood

Suppose $\{\mathbf{x}_j, j = 1, \ldots, n\} \in \mathbb{R}^p$ are i.i.d. random vectors following the multivariate model

$$\mathbf{x}_j = \Gamma \mathbf{z}_j + \mu, \qquad j = 1, \ldots, n,$$

where $\mathbf{z}_j, j = 1, \ldots, n$ are i.i.d. p-dimensional random vectors. The components of $\mathbf{z}_j$ are also i.i.d. with mean 0, variance 1, and finite fourth moment μ_4. One can generate a rich collection of $\mathbf{x}_j$ from $\mathbf{z}_j$ with the given covariance matrix $\Sigma = \Gamma\Gamma'$, where $\Gamma = \Sigma^{1/2}$. In the rest of this paper, we assume that all the eigenvalues of Σ are between the positive constants c_0 and C_0.

Chen, Variyath, and Abraham (2008) modified the EL function by adding one point, called a pseudo-observation, to make sure that the zero vector was an interior point of the convex hull of the new set $\{\mathbf{x}_i - \mu, i = 1, \ldots, n, n+1\}$, where $\mathbf{x}_{n+1} = \mu - b_n(\bar{\mathbf{x}} - \mu)$, $\bar{\mathbf{x}}$ is the sample mean of the original data set, and b_n is a well chosen constant that may depend on n. Figure 1 provides an example that after adding $\mathbf{x}_{n+1}$ to the data set, the zero vector becomes an interior point of the convex hull of the new data set, even when the zero vector is not contained in the original data set.

It was pointed out in Emerson and Owen (2009) that addressing the convex hull problem by adding one point to the original data set results in two shortcomings: the sample mean of the new data set is changed; the log EL statistic is bounded from above, which leads to poor performance when the dimension is large, say $p = 7$. Simulations in Emerson and Owen (2009) support their arguments. Hence, it is better to add two pseudo-observations to the data set in order to preserve the sample mean. This method is called the Balanced Adjusted Empirical Likelihood (BAEL) method in Emerson and Owen (2009).

2.1. Normal approximation

Let $\bar{\mathbf{x}}$ and $\mathbf{S} = (1/(n-1))\sum_{j=1}^{n}(\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})' = \Gamma\big((1/(n-1))\sum_{j=1}^{n}(\mathbf{z}_j - \bar{\mathbf{z}})(\mathbf{z}_j - \bar{\mathbf{z}})'\big)\Gamma'$ be the sample mean and sample covariance matrix respectively, where $\bar{\mathbf{z}} = (1/n)\sum_{i=1}^{n}\mathbf{z}_j$. Since all the eigenvalues of $\mathbf{S}$ are bounded from above and from below (Jiang (2004) and Xiao and Zhou (2010)), with probability one the inverse of $\mathbf{S}$ exists almost surely. Following Emerson and Owen (2009), the two pseudo-observations are

$$\mathbf{x}_{n+1} = \mu - sc_{\tilde{\mathbf{u}}}\tilde{\mathbf{u}}, \qquad \mathbf{x}_{n+2} = 2\bar{\mathbf{x}} - \mu + sc_{\tilde{\mathbf{u}}}\tilde{\mathbf{u}}, \tag{2.1}$$

where $c_{\tilde{\mathbf{u}}} = \left(\tilde{\mathbf{u}}'\mathbf{S}^{-1}\tilde{\mathbf{u}}\right)^{-1/2}$, $\tilde{\mathbf{v}} = \bar{\mathbf{x}} - \mu$, $\tilde{r} = \|\tilde{\mathbf{v}}\|$ and $\tilde{\mathbf{u}} = \tilde{\mathbf{v}}/\tilde{r}$. As illustrated in Emerson and Owen (2009), the choice of $c_{\tilde{\mathbf{u}}}$ is the inverse Mahalanobis distance of a unit vector from $\bar{\mathbf{x}}$ in the direction of $\tilde{\mathbf{u}}$. By imposing $\mathbf{x}_{n+1}$, the zero vector must be contained in the convex hull of $\{\mathbf{x}_i - \mu, i = 1, \ldots, n, n+1\}$; the second point $\mathbf{x}_{n+2}$ is included to maintain the original sample mean, since $(1/(n+2))\sum_{j=1}^{n+2}\mathbf{x}_j = \bar{\mathbf{x}}$. For the new data set $\{\mathbf{x}_j, j = 1, \ldots, n, n+1, n+2\}$, the BAEL ratio for any hypothesized value μ_0 is

$$R(\mu_0) = \max\left\{\prod_{i=1}^{n+2}(n+2)\omega_i : \omega_i \geq 0, \sum_{i=1}^{n+2}\omega_i = 1, \sum_{i=1}^{n+2}\omega_i(\mathbf{x}_i - \mu_0) = \mathbf{0}\right\},$$

and the log BAEL statistic is

$$W(\mu_0) = -2\log R(\mu_0). \tag{2.2}$$

Our results establish the equivalence of $W(\mu)$ to Hotelling's T^2 statistic and its asymptotic normality under the setting that p is proportional to n.

Theorem 1. *Suppose the following conditions hold*

1. $\{\mathbf{x}_j, j = 1, \ldots, n\} \in \mathbb{R}^p$ *are i.i.d. random vectors satisfying*

$$\mathbf{x}_j = \Gamma\mathbf{z}_j + \mu, \qquad j = 1, \ldots, n, \tag{2.3}$$

where $\mathbf{z}_j, j = 1, \ldots, n$ *are i.i.d. p-dimensional random vectors, and the components of* $\mathbf{z}_j$ *are i.i.d. with mean 0, variance* 1*, and finite fourth moment* μ_4.

2. $\Gamma = \Sigma^{1/2}$, *where* Σ *is a covariance matrix with eigenvalues bounded below and above by positive constants* c_0 *and* C_0, *respectively.*
3. *The* $p's$ *are functions of* n, *satisfying*

$$\frac{p}{n} = c_n \to c \in (0,1), \qquad \frac{n\sqrt{n}}{s} \to 0, \quad as \quad n \to \infty. \tag{2.4}$$

Then,

$$\frac{2s^2 W(\mu)}{(n+2)^2} - \frac{T^2}{n} = o_p(\frac{n}{s}) + o_p(\frac{1}{n}), \quad as \quad n \to \infty, \tag{2.5}$$

where $T^2 = n(\bar{\mathbf{x}} - \mu)'\mathbf{S}^{-1}(\bar{\mathbf{x}} - \mu)$ *is Hotelling's* T^2 *statistic, an* $o_p(\alpha(n))$ *term tends to zero in probability when divided by* $\alpha(n)$, *and* $\xrightarrow{i.p.}$ *signifies convergence in probability.*

Pan and Zhou (2011) proved that Hotelling's T^2 statistic is asymptotically normal by using random matrix theory, see Lemma A.1 in the Appendix. The following is a direct consequence of Theorem 1 and Lemma A.1.

Corollary 1. *Under the assumptions of Theorem* 1,

$$\sqrt{\frac{n}{2c_n(1-c_n)^{-3}}}\Big(\frac{2s^2 W(\mu)}{(n+2)^2} - c_n(1-c_n)^{-1}\Big) \xrightarrow{d} N(0,1), \quad as \quad n \to \infty. \tag{2.6}$$

Remark 1. We require $s \to \infty$ along with n satisfying the condition (2.4). In practice, we may choose $s = n^2$. The condition (2.4) also implies that the dimension p cannot exceed the sample size n.

Remark 2. It is well-known that the log EL ratio statistic is asymptotically equivalent to Hotelling's T^2 statistic when the dimension p is fixed, see Owen (1988, 1990). Theorem 1 now implies the asymptotic equivalence of the log BAEL statistic and Hotelling's T^2 statistic.

2.2. Calibration by normalized $F(p, n-p)$ distribution

A simulation study shows that the normal approximation in Corollary 1 is good, but not perfect, see Table 1 in Section 3.1 and Figure 4 in Section 3.2. To calibrate the normal approximation to $W(\mu)$ for high dimensional data, one approach is to use bootstrap calibration (see Owen (1988), Hjort, Mckeague, and Van Keilegom (2009)) which involves resampling from the original data K times to get the new data sets. For $k = 1, \ldots, K$, the k-th resampled data set is $\{\mathbf{x}_1^{(k)}, \ldots, \mathbf{x}_n^{(k)}\}$. Then compute the EL ratio statistic $W^{(k)}(\bar{\mathbf{x}})$ with this k-th resampled data set. The p-value is given by the distribution of $W^{(k)}(\bar{\mathbf{x}}), k = 1, \ldots, K$. The bootstrap calibration is reasonably good but computationally intensive, especially when the dimension is large.

Theorem 1 indicates the connection between the scaled $W(\mu)$ and Hotelling's T^2 statistic under a high-dimensional setting. Noticing that $(n-p)T^2/(pn)$ follows the F distribution with degrees of freedom p and $n-p$ when the population is $N(\mu, \Sigma)$, we suggest using the normalized $F(p, n-p)$ distribution to approximate the log BAEL statistic.

Theorem 2. *Under the assumptions of Theorem* 1*, we have*

$$\sqrt{\frac{n}{2c_n(1-c_n)^{-3}}}\Big(\frac{2s^2W(\mu)}{(n+2)^2}-c_n(1-c_n)^{-1}\Big)=\frac{F(p,n-p)-1}{\sqrt{2/p+2/(n-p)}}+o_p(1),\ \text{as } n\to\infty.$$

3. Simulation Results

Here we report on simulation studies to investigate the performance of the BAEL statistic. Our p-dimensional data follow the model (2.3). We generated the components of $\mathbf{z}_j$ from three distributions: the standard normal distribution, the standardized Gamma(4, 2), and the standardized student t distribution with 5 degrees of freedom, t(5). Throughout the simulation, we took the covariance of $\mathbf{x}_j$, $\Sigma = \Gamma\Gamma'$, to be a Toeplize matrix with first row $(1, 0.5, 0.5^2, 0.5^3, \ldots, 0.5^{p-1})$, the covariance matrix for the AR(1) model with the parameter $\sigma = 0.5$.

We focussed on the large-dimensional case with $p/n \to c$, using $n = 200$, 400, 800, $c = 0.4, 0.8$. The s satisfying (2.4) was chosen as $s = n^2$. We carried out $M =$5,000 simulations for each (n, c)-combination and for each distribution of $\mathbf{z}_j$.

3.1. Empirical sizes

In the first section, we tabulate the empirical sizes of the BAEL statistic when the standard normal distribution and the normalized $F(p, n-p)$ were used as calibrations. From Table 1, we can see that the empirical sizes according to the normalized $F(p, n-p)$ approximation were closer to the nominal significance level 0.05.

3.2. Q-Q plots

We used Q-Q plots to compare the accuracy of the normalized $F(p, n-p)$ approximation and the $N(0, 1)$ approximation. For each (n, c)-combination and for each distribution of $\mathbf{z}_j$, we simulated the normalized log BAEL statistic $\zeta = \sqrt{n/2c_n(1-c_n)^{-3}}(2s^2W(\mu)/(n+2)^2 - c_n(1-c_n)^{-1})$ for M times and calculated the quantiles of the theoretical $(F(p, n-p)-1)/\sqrt{2/p+2/(n-p)}$ and $N(0, 1)$ distributions at probabilities $i/(M+1)$, $i = 1, \ldots, M$.

Q-Q plots are given in Figures 2 and 3. Figure 2 corresponds to the normalized $F(p, n-p)$ approximation. Here, as the dimension p increases proportionally

Table 1. Empirical sizes for $N(0,1)$ and $F(p,n-p)$ approximations when $\mathbf{z}_j$ is standard normal, standardized Gamma(4,2), or standardized t(5).

	Standard Normal		Standardized Gamma(4,2)		Standardized t(5)	
(n,c)	$N(0,1)$	$F(p,n-p)$	$N(0,1)$	$F(p,n-p)$	$N(0,1)$	$F(p,n-p)$
(200,0.4)	0.0530	0.0478	0.0684	0.0628	0.0548	0.0494
(200,0.8)	0.0746	0.0514	0.0794	0.0554	0.0666	0.0460
(400,0.4)	0.0540	0.0508	0.0584	0.0568	0.0448	0.0436
(400,0.8)	0.0574	0.0456	0.0598	0.0490	0.0590	0.0496
(800,0.4)	0.0532	0.0454	0.0520	0.0500	0.0492	0.0506
(800,0.8)	0.0552	0.0514	0.0548	0.0496	0.0556	0.0510

to the sample size n, the normalized $F(p,n-p)$ distribution approximates the normalized log BAEL with high accuracy. These figures also show that whether the underlying variables were generated from a fatter-tail distribution or not (the fourth moment of three distributions are 3, 4.5, and 6, respectively), the performance of the normalized $F(p,n-p)$ approximation is unaffected.

Figure 3 shows the performance of the $N(0,1)$ approximation under different populations. Compared with the $F(p,n-p)$ approximation, the normal approximation is unsatisfactory and suggest using the normalized $F(p,n-p)$ distribution for calibration.

3.3. Ratio of $||\lambda||$ to $1/s$

As a numerical illustration of the result $||\lambda|| = o_p(s^{-1})$, we report the ratio of $||\lambda||$ to $1/s$ by boxplots (Figure 4) for different (n,c)-combinations and for the three distributions of $\mathbf{z}_j$ mentioned above. These boxplots show that the ratio of $||\lambda||$ to $1/s$ is close to zero regardless of the sample distributions.

3.4. Difference between Hotelling's T^2 and BAEL

Set $r^2 = T^2/n$ and $\varsigma = 2s^2W(\mu)/(n+2)^2$. Since we choose $s = n^2$ in our simulation, (2.5) in Theorem 2.1 indicates $r^2 - \varsigma = o_p(1/n)$. We used simulations to check the difference. Figure 5 reports $(r^2 - \varsigma)$ under different (n,c) pairs and different sample distributions. Overall, this difference is much smaller than $1/n$ and shows a significant decrease as n becomes larger. Another observation is that the difference $(r^2 - \varsigma)$ is always larger than 0, which may imply that BAEL test is less likely to make the Type I error than Hotelling's T^2 test. This phenomenon is also observed in our data case study.

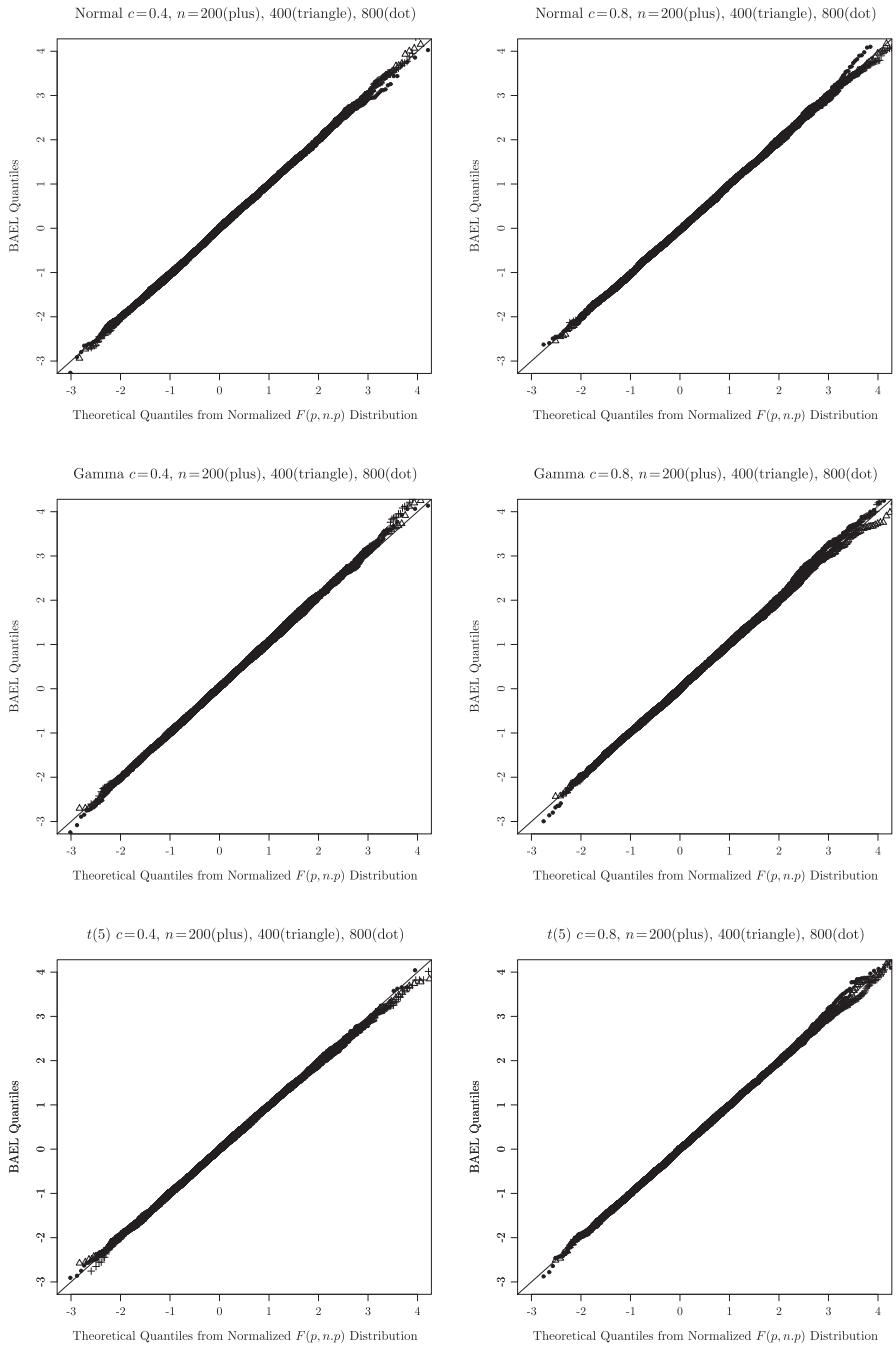

Figure 2. Normalized $F(p, n-p)$ approximation with the sample generating from normal distribution(first row), standardized Gamma(4,2)(second row) and standardized t(5)(third row). $c = p/n$ is 0.4(left) and 0.8(right), and the sample size is 200(plus), 400(triangle) and 800(dot).

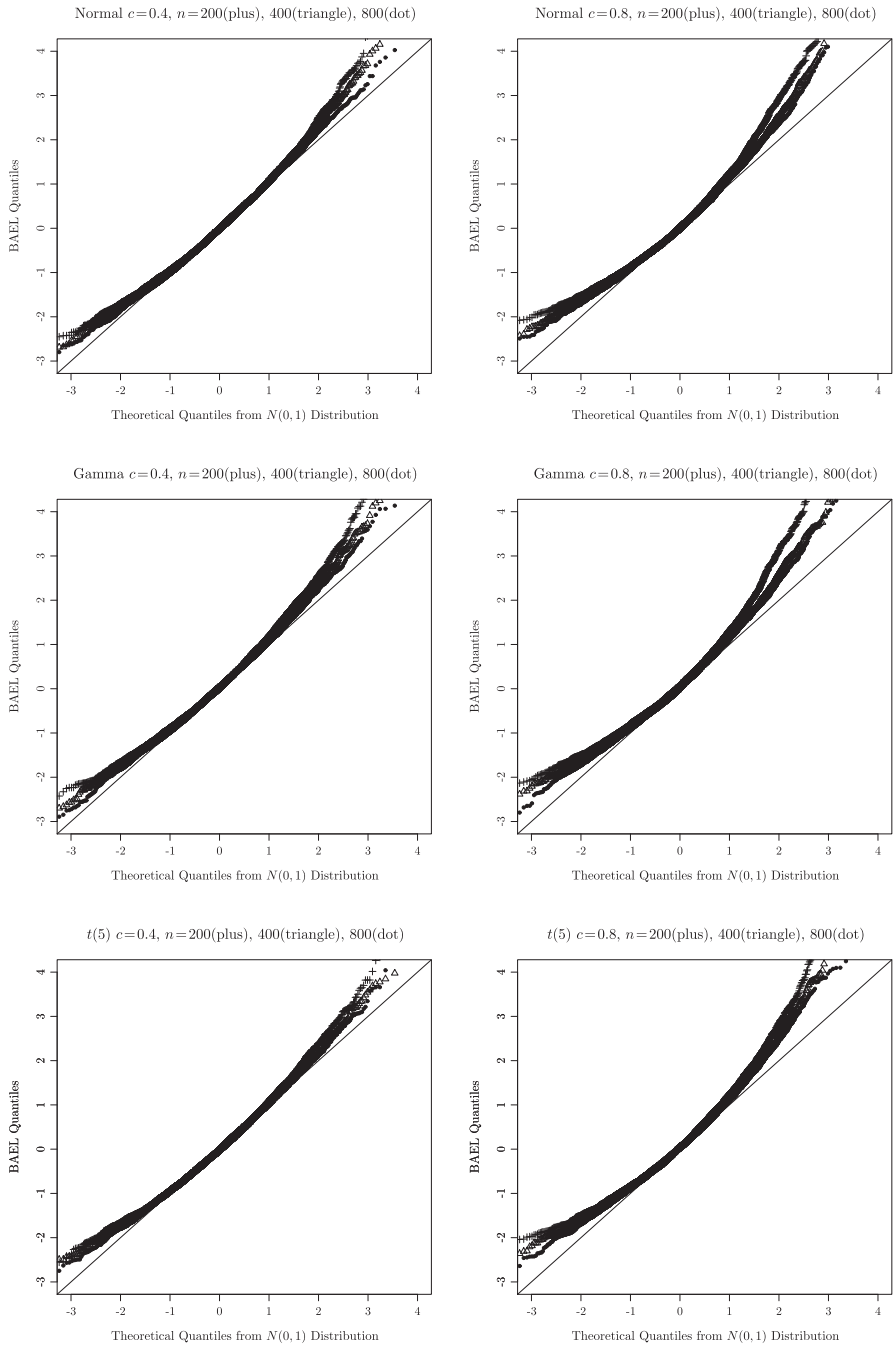

Figure 3. $N(0,1)$ approximation with the sample generating from normal distribution(first row), standardized Gamma(4,2)(second row) and standardized t(5)(third row). $c = p/n$ is 0.4(left) and 0.8(right), and the sample size is 200(plus), 400(triangle) and 800(dot).

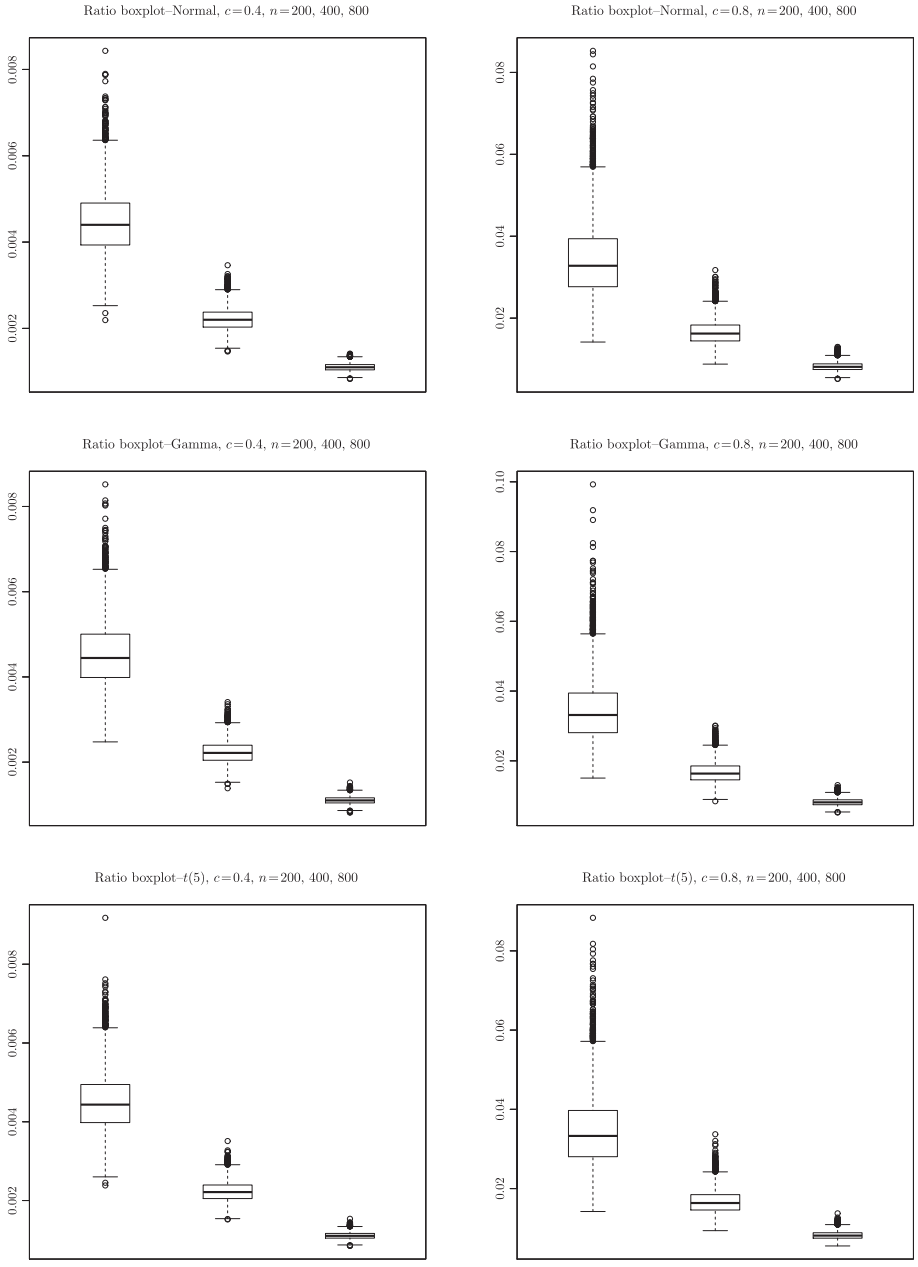

Figure 4. Ratio of $||\lambda||$ to $1/s$ with the sample generating from normal distribution (first row), standardized Gamma(4,2) (second row) and standardized t(5) (third row). $c = p/n$ is 0.4 (left) and 0.8 (right), and the sample size is 200 (left), 400 (middle) and 800 (right).

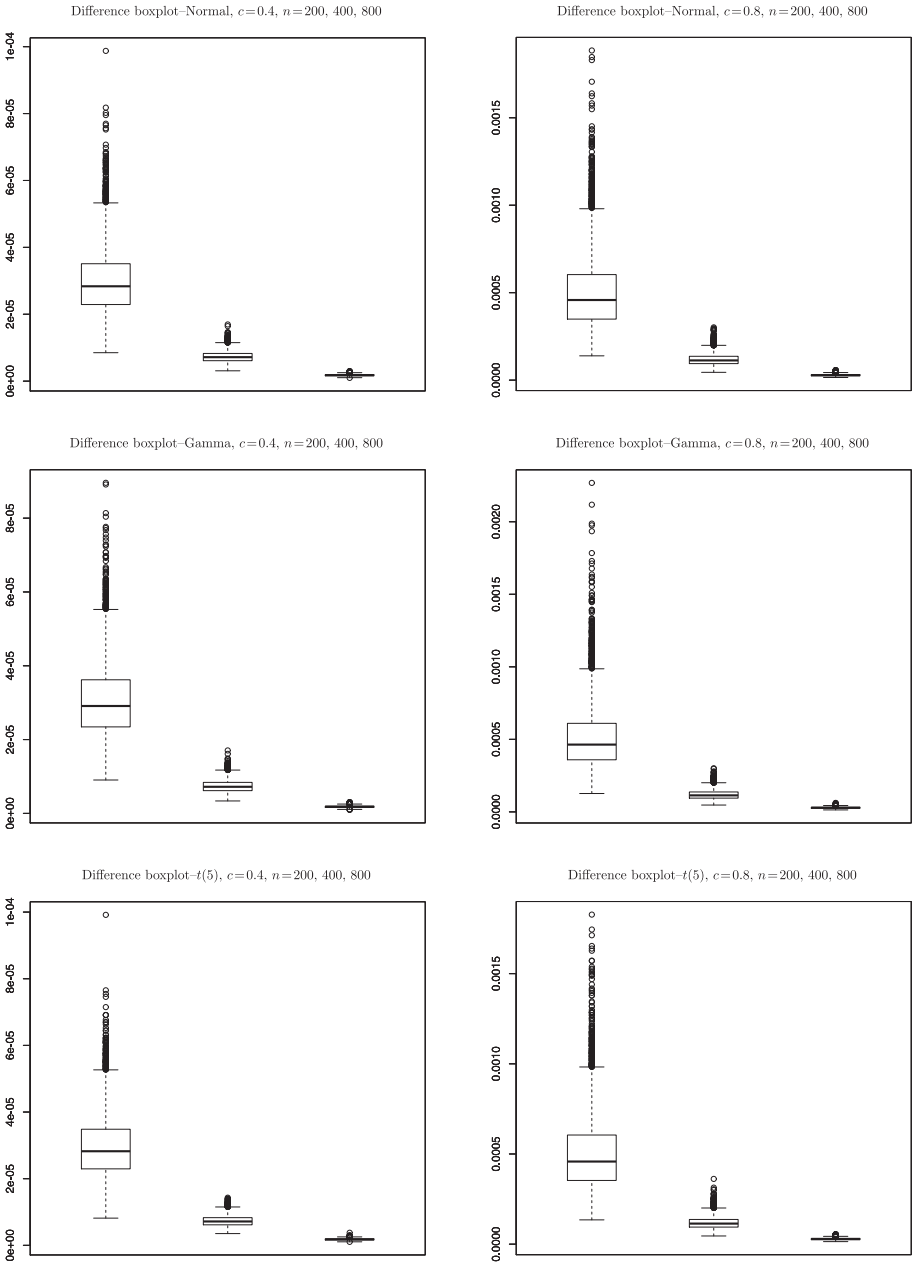

Figure 5. $(r^2 - \varsigma)$ with the sample generating from normal distribution (first row), standardized Gamma(4,2) (second row) and standardized t(5) (third row). $c = p/n$ is 0.4 (left) and 0.8 (right), and the sample size is 200 (left), 400 (middle) and 800 (right).

4. Data Analysis

We considered a financial application of large dimensional BAEL to test the mean of the stock returns of 508 companies from the $S\&P$ 500. The data contain the daily closing bid prices for these stocks from 1 Jan 2010 to 31 Dec 2011 (data can be downloaded from Wharton Research Data Services). We adopted 2-day returns in the hope of making observations independent, leading to 125 observations for each stock. We investigated the returns in each sector of $S\&P$ 500 rather than the overall returns. These sectors are consumer discretionary, consumer staples, energy, financials, health care, industrials, information technology, materials, telecommunications services and utilities. In sector k, $k = 1, \ldots, 10$, let p_k denote the number of stocks contained in the sector. The dimension of sector k is p_k while the number of observations is $n = 125$.

Let $\mathbf{h}_t^{(k)} = (h_{1t}^{(k)}, h_{2t}^{(k)}, \ldots, h_{p_k t}^{(k)})', k = 1, \ldots, 10, t = 1, \ldots, 252$ be the daily closing bid prices for the stocks in sector k at time t. The j-th, $j = 1, \ldots, 125$, log-returns for stocks in sector k is

$$\mathbf{x}_j^{(k)} = \Big(\log \frac{h_{1,2j+1}^{(k)}}{h_{1,2j-1}^{(k)}}, \log \frac{h_{2,2j+1}^{(k)}}{h_{2,2j-1}^{(k)}}, \ldots, \log \frac{h_{p_k,2j+1}^{(k)}}{h_{p_k,2j-1}^{(k)}} \Big)'.$$

Hence the $p_k \times n$ data matrix for sector k is $\mathbf{X}^{(k)} = \Big(\mathbf{x}_1^{(k)}, \mathbf{x}_2^{(k)}, \ldots, \mathbf{x}_n^{(k)} \Big)$. We take $\mathbf{x}_1^{(k)}, \ldots, \mathbf{x}_n^{(k)}$ as $i.i.d.$ random vectors satisfying (2.3). Let $\mu^{(k)} = E\mathbf{x}_1^{(k)}$. We are interested in testing

$$H_0^{(k)}: \quad \mu^{(k)} = \mathbf{0}_{p_k} \quad \text{vs} \quad H_1^{(k)}: \quad \mu^{(k)} \neq \mathbf{0}_{p_k}, \tag{4.1}$$

where $\mathbf{0}_{p_k}$ is a p_k-dimensional vector with all the elements zero. Following (2.1), we can construct two pseudo-observations $\mathbf{x}_{n+1}^{(k)}$ and $\mathbf{x}_{n+2}^{(k)}$ under the null hypothesis where $s = n^2$. Similarly define $W(\mathbf{0}_{p_k})$ as in (2.2). The test statistic is

$$S_n^{(k)} = \sqrt{\frac{n}{c_n^{(k)}(1 - c_n^{(k)})^{-3}}} \Big(\frac{2s^2 W(\mathbf{0}_{p_k})}{n^2} - c_n^{(k)}(1 - c_n^{(k)})^{-1} \Big),$$

where $c_n^{(k)} = p_k/n$. With the normal distribution as the calibration, $S_n^{(k)} \xrightarrow{d} N(0,1)$. If we use the normalized $F(p, n-p)$ as the calibration, Theorem 2 allows us to calculate the critical value.

Following the notation in Section 3.4, we denote

$$r_k^2 = \frac{T^2(\mathbf{0}_{p_k})}{n}, \qquad \varsigma_k = \frac{2s^2 W(\mathbf{0}_{p_k})}{(n+2)^2}, \qquad k = 1, \ldots, 10,$$

satisfying $r_k^2 - \varsigma_k = o_p(1/n)$ according to Theorem 1.

Table 2. BAEL test and Hotelling's T^2 test for large dimensional data in $S\&P$ 500.

Sector	p	S_n	ς	r^2	HT-Pvalue	EL-N-Pvalue	EL-F-Pvalue
CD	86	-0.2175	2.074215	2.074632	0.4142	0.4139	0.3981
CS	42	-1.7398	1.08842	1.088575	0.0410	0.0409	0.0074
EG	40	0.0002	0.4706132	0.4706565	0.5002	0.5001	0.5131
FN	82	-0.2463	1.78189	1.782221	0.4030	0.4027	0.3888
HC	53	-0.7753	0.5900303	0.5900913	0.2192	0.2191	0.1999
IND	62	-0.7610	0.794661	0.7947569	0.2234	0.2233	0.2004
IT	69	0.0207	1.238644	1.238833	0.5085	0.5083	0.5047
MR	30	-0.2434	0.2930243	0.2930453	0.4039	0.4038	0.4214
TS	9	0.0548	0.07966701	0.07966996	0.52188	0.52185	0.5770
UL	35	-0.2893	0.3571932	0.3572217	0.3863	0.3862	0.3985
Overall	508						

CD, CS, EG, FN, HC, IND, IT, MR, TS, UL are abbreviations for the sectors consumer discretionary, consumer staples, energy, financials, health care, industrials, information technology, materials, telecommunications services, and utilities, respectively.

We report the values of $S_n^{(k)}$, ς_k, r_k^2, p-values for Hotelling's T^2 test, and p-values for BAEL test when the $N(0,1)$ and normalized $F(p, n-p)$ are used as approximations. They are denoted as S_n, ς, r^2, HT-Pvalue, EL-N-Pvalue and EL-F-Pvalue, respectively, in Table 2.

The dimension of the sectors can be as large as 80 and comparable to the number of observations. Table 2 has the p-values for consumer staples (CS) all less than 0.05; especially, when the normalized $F(p, n-p)$ calibration is used for the BAEL test, the p-value is 0.0074. Hence we can reject the null hypothesis at the level of significance $\alpha = 0.05$. Indeed, according to a report, shares of consumer staples companies, accounting for a total of about 11.4% of the $S\&P$ 500 Index, were up 5.3%, compared with a 3.3% drop for the $S\&P$ 500. Comparison among the three p-values for CS implies that the normalized $F(p, n-p)$ approximation for the BAEL test performs much better than Hotelling's T^2 test and the $N(0,1)$ approximation since its p-value is significantly smaller than 0.05 when the null hypothesis should be rejected. This again supports our suggestion to use the normalized $F(p, n-p)$ as calibration in the simulation study of Section 3.2.

We are unable to reject the remaining null hypotheses since the p-values for the other sectors are not too low. In particular, the data from the sectors of energy (EG), information technology (IT) and telecommunications services (TS) show little change.

In Table 2, r^2 is always a little bit larger than ς; this phenomenon was also seen in the simulations of Section 3.4.

Acknowledgements

The authors are grateful to an associate editor and a referee for useful comments which helped to improve the paper significantly. Pan's research was supported in part by the Ministry of Education, Singapore, under grant # ARC 14/11, and MOE's Tier 1 grant under RG25/14. Zhou's research was supported in part by the Ministry of Education, Singapore, under grant # ARC 14/11, and a grant R-155-000-131-112 at the National University of Singapore.

Appendix

Lemma A.1 (Theorem 1 in Pan and Zhou (2011))**.** Under the assumptions of Theorem 1, we have

$$\sqrt{\frac{n}{2c_n(1-c_n)^{-3}}}\Big(\frac{T^2}{n}-c_n(1-c_n)^{-1}\Big)\xrightarrow{d}N(0,1),\quad \text{as}\quad n\to\infty,$$

where

$$r^2\triangleq\frac{T^2}{n}=(\bar{\mathbf{x}}-\mu)'\mathbf{S}^{-1}(\bar{\mathbf{x}}-\mu).$$

Lemma A.2 (Lemma 2 in Bai and Yin (1993))**.** Let $\{X_{ij}, i,j=1,\dots,\}$ be a double array of i.i.d. random variables and let $\alpha>1/2$, $\beta\geq 0$ and $M>0$ be constants. Then as $n\to\infty$,

$$\max_{j\leq Mn^{\beta}}\left|n^{-\alpha}\sum_{i=1}^{n}(X_{ij}-c)\right|\to 0\qquad \text{a.s.},$$

if and only if the following hold:

(i) $E|X_{11}|^{(1+\beta)/\alpha}<\infty$;

(ii) $$c=\begin{cases}EX_{11}, & \text{if}\quad \alpha\leq 1,\\ \text{any number}, & \text{if}\quad \alpha>1.\end{cases}$$

Under (2.4), the covariance matrix $\mathbf{S}$ is of full rank with probability one. Hence we have $\mathbf{S}=\mathbf{AA}'$ where $\mathbf{A}$ is a $p\times p$ invertible matrix. To simplify the notation, we work on standardized data. Let $\mathbf{y}_i=\mathbf{A}^{-1}(\mathbf{x}_i-\bar{\mathbf{x}})$, for $i=1,\dots,n$, $\bar{\mathbf{y}}=(1/n)\sum_{i=1}^{n}\mathbf{y}_i=\mathbf{0}$, and $\beta=\mathbf{A}^{-1}(\mu-\bar{\mathbf{x}})$. Take $\mathbf{v}=\bar{\mathbf{y}}-\beta=-\beta$, $r=\|\mathbf{v}\|=\|\beta\|$, $\mathbf{u}=\mathbf{v}/r=-\beta/r$, $\mathbf{y}_{n+1}=\beta-s\mathbf{u}$ and $\mathbf{y}_{n+2}=-\beta+s\mathbf{u}$, where $\|f\|$ denotes the Euclidean norm if f is a vector or the spectral norm if f is a matrix. As the EL method has the invariance property under transformation $\mathbf{x}\longmapsto\tilde{\mathbf{x}}=\mathbf{M}\mathbf{x}$, where $\mathbf{M}$ is any invertible matrix, we have

$$R(\mu)=R(\mu;\mathbf{x}_1,\dots,\mathbf{x}_{n+2})=\breve{R}(\beta;\mathbf{y}_1,\dots,\mathbf{y}_{n+2})=\breve{R}(\beta),$$

where

$$\breve{R}(\beta) = \max\Big\{ \prod_{i=1}^{n+2}(n+2)\omega_i : \omega_i \geq 0, \sum_{i=1}^{n+2}\omega_i = 1, \sum_{i=1}^{n+2}\omega_i(\mathbf{y}_i - \beta) = \mathbf{0}\Big\}. \quad \text{(A.1)}$$

Thus $W(\mu) = \breve{W}(\beta) = -2\log\breve{R}(\beta)$. An explicit expression for $\breve{R}(\beta)$ and $\breve{W}(\beta)$ can be obtained by using Lagrange multipliers. The optimal weights ω_i for $\breve{R}(\beta)$ are

$$\omega_i = \frac{1}{n+2}\frac{1}{1+\lambda'(\mathbf{y}_i - \beta)}, i = 1, \ldots, n+2, \quad \text{(A.2)}$$

where $\lambda \in \mathbb{R}^p$ is the Lagrange multiplier satisfying

$$\sum_{i=1}^{n+2}\frac{\mathbf{y}_i - \beta}{1+\lambda'(\mathbf{y}_i - \beta)} = 0. \quad \text{(A.3)}$$

Hence

$$\breve{W}(\beta) = 2\sum_{i=1}^{n+2}\log\Big(1+\lambda'(\mathbf{y}_i - \beta)\Big). \quad \text{(A.4)}$$

Proof of Theorem 1. The proof of Theorem 1 proceeds in several steps. We only sketch the proofs here. One may refer to the supplementary document for further details.

Step 1 (w.r.t Lemma S1). Step 1 is to establish $\lambda'\mathbf{u} = o_p(s^{-1})$. To this end, using the constraint $\sum_{i=1}^{n+2}\omega_i(\mathbf{y}_i - \beta) = \mathbf{0}$, we can write $s(\omega_{n+1} - \omega_{n+2})$ as a sum of two terms:

$$s(\omega_{n+1} - \omega_{n+2}) = \sum_{i=1}^{n}\omega_i\mathbf{u}'(\mathbf{y}_i - \beta) + 2r\omega_{n+2}.$$

The first term is shown to be $O_p(\sqrt{n})$ by the Hölder inequality; the second term is shown to be bounded according to Hotelling T^2's property. By the expressions of ω_{n+1} and ω_{n+2}, $s(\omega_{n+1} - \omega_{n+2})$ can be re-expressed as

$$\frac{n+2}{s}\cdot s(\omega_{n+1} - \omega_{n+2}) = \frac{1}{1 - s\lambda'\mathbf{u}} - \frac{1}{1+(2r+s)\lambda'\mathbf{u}}.$$

Combing these two gives $1/(1 - s\lambda'\mathbf{u}) - 1/(1 + (2r+s)\lambda'\mathbf{u}) = O_p(n\sqrt{n}/s)$, which leads to $\lambda'\mathbf{u} = o_p(s^{-1})$. Details can be found in Lemma S1 of the supplementary material.

Step 2 (w.r.t Lemma S2). Step 2 is to prove $\|\lambda\| = o_p(s^{-1/2})$ and $\max_{i\leq n}|\lambda'(\mathbf{y}_i - \beta)| = o_p(\sqrt{n/s})$. Let $\lambda = \rho\theta$, $\rho = \|\lambda\|$. We first show that $\max_{i\leq n}|\theta'(\mathbf{y}_i - \beta)|^2 = O_p(n)$ by Lemma 5.2. Then we split the equality $\sum_{i=1}^{n+2}(\lambda'(\mathbf{y}_i$

$-\beta))/(1+\lambda'(\mathbf{y}_i-\beta))=0$ according to the identity $1/(1+x)=1-x/(1+x)$, from which we can get

$$\frac{n+2}{n}r\lambda'\mathbf{u}=\frac{\rho^2}{n}\sum_{i=1}^{n+2}\frac{\theta'(\mathbf{y}_i-\beta)(\mathbf{y}_i-\beta)'\theta}{1+\lambda'(\mathbf{y}_i-\beta)}.$$

Checking both sides here by applying the results $\max_{i\le n}|\theta'(\mathbf{y}_i-\beta)|^2=O_p(n)$ and $\lambda'\mathbf{u}=o_p(s^{-1})$ from Step 1, we can obtain $\|\lambda\|=\rho=o_p(s^{-1/2})$. Finally the bound $\max_{i\le n}|\lambda'(\mathbf{y}_i-\beta)|=o_p(\sqrt{n/s})$ comes from the fact that $\max_{i\le n}|\lambda'(\mathbf{y}_i-\beta)|=\|\lambda\|\cdot\max_{i\le n}|\theta'(\mathbf{y}_i-\beta)|$. Details can be found in Lemma S2 of the supplementary material.

Step 3 (w.r.t Lemma S3). The aim of this step is to improve the bound on the norm of λ to $\|\lambda\|=o_p(s^{-1})$. Based on the conclusion in Step 1, we have $\|\lambda\|\cdot|\mathbf{u}'\theta|=|\lambda'\mathbf{u}|=o_p(s^{-1})$. So once $|\mathbf{u}'\theta|\xrightarrow{i.p.}1$ is proved, $\|\lambda\|=o_p(s^{-1})$ is clear.

To prove this, we construct $\mathbf{y}_i-\beta=k_i\mathbf{u}+\mathbf{r}_i$, where $k_i=(\mathbf{y}_i-\beta)'\mathbf{u}$, $\mathbf{r}_i=(\mathbf{y}_i-\beta)-k_i\mathbf{u}$, $\mathbf{u}'\mathbf{r}_i=\mathbf{0}$, $i=1,\dots,n+2$. Then we show that with probability one there exist $a_1,\dots,a_n$ such that

$$\begin{aligned}\theta&=a_1(\mathbf{y}_1-\beta)+a_2(\mathbf{y}_2-\beta)+\cdots+a_n(\mathbf{y}_n-\beta)\\&=\Big(\sum_{i=1}^n a_ik_i\Big)\mathbf{u}+a_1\mathbf{r}_1+\cdots+a_n\mathbf{r}_n.\end{aligned}$$

Multiplying both sides by $\mathbf{u}'$ gives $\mathbf{u}'\theta=\sum_{i=1}^n a_ik_i$, and multiplying both sides by θ' gives $1=\Big(\sum_{i=1}^n a_ik_i\Big)\theta'\mathbf{u}+\sum_{i=1}^n a_i\theta'\mathbf{r}_i$.

The proof of $|\mathbf{u}'\theta|\xrightarrow{i.p.}1$ reduces to showing $\Big|1-\Big(\sum_{i=1}^n a_ik_i\Big)^2\Big|=\Big|\sum_{i=1}^n a_i\theta'\mathbf{r}_i\Big|$ $\le\Big|\sum_{i=1}^n a_i^2\cdot\sum_{i=1}^n(\theta'\mathbf{r}_i)^2\Big|^{1/2}\xrightarrow{i.p.}0$. To deal with $\sum_{i=1}^n a_i^2$, we observe that $tr\Gamma'\Gamma\sum_{i=1}^n a_i^2$ is the leading term of $\theta'\mathbf{A}'\mathbf{A}\theta$, which is bounded from above in probability. So $\sum_{i=1}^n a_i^2=O_p(1/n)$. To deal with $\sum_{i=1}^n(\theta'\mathbf{r}_i)^2$, as in Step 2, we split $\sum_{i=1}^n\theta'\mathbf{r}_i/(1+\lambda'(\mathbf{y}_i-\beta))$ twice according to the identity $1/(1+x)=1-x/(1+x)$, to get

$$\sum_{i=1}^n\frac{\theta'\mathbf{r}_i}{1+\lambda'(\mathbf{y}_i-\beta)}=\lambda'\mathbf{u}\rho\sum_{i=1}^n\frac{k_i\theta'\mathbf{r}_i\theta'(\mathbf{y}_i-\beta)}{1+\lambda'(\mathbf{y}_i-\beta)}-\rho\sum_{i=1}^n(\theta'\mathbf{r}_i)^2+\rho\sum_{i=1}^n\frac{(\theta'\mathbf{r}_i)^2\lambda'(\mathbf{y}_i-\beta)}{1+\lambda'(\mathbf{y}_i-\beta)}.$$

As $\sum_{i=1}^n\theta'\mathbf{r}_i/(1+\lambda'(\mathbf{y}_i-\beta))$ is shown to be 0 from our construction,

$$\rho\sum_{i=1}^n(\theta'\mathbf{r}_i)^2=\lambda'\mathbf{u}\rho\sum_{i=1}^n\frac{k_i\theta'\mathbf{r}_i\theta'(\mathbf{y}_i-\beta)}{1+\lambda'(\mathbf{y}_i-\beta)}+\rho\sum_{i=1}^n\frac{(\theta'\mathbf{r}_i)^2\lambda'(\mathbf{y}_i-\beta)}{1+\lambda'(\mathbf{y}_i-\beta)}.$$

Using the conclusions $\lambda'\mathbf{u} = o_p(s^{-1})$ in Step 1 and $\|\lambda\| = o_p(s^{-1/2})$, $\max_{i\le n}|\lambda'(\mathbf{y}_i - \beta)| = o_p(\sqrt{n/s})$, $\max_{i\le n}|\theta'(\mathbf{y}_i - \beta)|^2 = O_p(n)$ in Step 2, we obtain $\sum_{i=1}^n(\theta'\mathbf{r}_i)^2 = o_p(n^2/s^2)$ from the above equality.

Then $\Big|\sum_{i=1}^n a_i^2 \cdot \sum_{i=1}^n(\theta'\mathbf{r}_i)^2\Big|^{1/2} = o_p(\sqrt{n}/s) \xrightarrow{i.p.} 0$ given (2.4), by which we can conclude that $|\mathbf{u}'\theta| \xrightarrow{i.p.} 1$ and thus $\|\lambda\| = o_p(s^{-1})$. Details can be found in Lemma S3 of the supplementary material.

Step 4 (w.r.t Lemma S4). λ is further found to satisfy $s^2\lambda'\mathbf{u} = (n+2)r/2 + o_p(n^2/s) + o_p(1)$ and $\|\lambda\| = \rho = o_p(n/s^2)$. Applying the identity $1/(1+x) = 1 - x + x^2/(1+x)$ to expand (A.3), we can get

$$\begin{aligned}
0 &= \sum_{i=1}^{n+2}\frac{\mathbf{u}'(\mathbf{y}_i - \beta)}{1 + \lambda'(\mathbf{y}_i - \beta)}\\
&= \sum_{i=1}^{n+2}\mathbf{u}'(\mathbf{y}_i - \beta) - \sum_{i=1}^{n+2}\mathbf{u}'(\mathbf{y}_i - \beta)(\mathbf{y}_i - \beta)'\lambda + \sum_{i=1}^{n+2}\frac{\mathbf{u}'(\mathbf{y}_i - \beta)\Big((\mathbf{y}_i - \beta)'\lambda\Big)^2}{1 + \lambda'(\mathbf{y}_i - \beta)}\\
&= (n+2)r - \Big(n\mathbf{u}'\mathbf{S}_1\lambda + 2s^2\lambda'\mathbf{u} + (4sr + 4r^2)\lambda'\mathbf{u}\Big)\\
&\quad + \sum_{i=1}^{n}\frac{\mathbf{u}'(\mathbf{y}_i - \beta)\Big((\mathbf{y}_i - \beta)'\lambda\Big)^2}{1 + \lambda'(\mathbf{y}_i - \beta)} - s^3(\lambda'\mathbf{u})^2\Big[(n+2)(\omega_{n+2} - \omega_{n+1})\Big]\\
&\quad + \frac{(6s^2r + 12sr^2 + 8r^3)(\lambda'\mathbf{u})^2}{1 - (2r+s)\lambda'\mathbf{u}}.
\end{aligned}$$

The right hand side here can be reduced to $(n+2)r - 2s^2\lambda'\mathbf{u} + s^2\lambda'\mathbf{u}\cdot o_p(s\lambda'\mathbf{u}) + o_p(1)$ by using the bounds proved in the first three steps. Its equivalence to zero further implies the conclusion. Details can be found in Lemma S4 of the supplementary material.

Step 5. To prove Theorem 1, with a Taylor's expansion, $\breve{W}(\beta)$ in (A.4) can be expanded as

$$\begin{aligned}
\breve{W}(\beta) &= 2\sum_{i=1}^{n+2}\log\Big(1 + \lambda'(\mathbf{y}_i - \beta)\Big)\\
&= 2\Big[\sum_{i=1}^{n+2}\lambda'(\mathbf{y}_i - \beta) - \frac{1}{2}\sum_{i=1}^{n+2}\Big(\lambda'(\mathbf{y}_i - \beta)\Big)^2 + \frac{1}{3}\sum_{i=1}^{n+2}\Big(\lambda'(\mathbf{y}_i - \beta)\Big)^3 - \sum_{i=1}^{n+2}\eta_i\Big].
\end{aligned}$$

Using the bounds in the previous steps, this expansion can be further simplified to

$$\breve{W}(\beta) = 2\Big[(n+2)r\lambda'\mathbf{u} - s^2(\lambda'\mathbf{u})^2 + O_p\Big(\frac{n^4}{s^4}\Big)\Big].$$

Multiplying both sides by $2s^2/(n+2)^2$ and using $s^2\lambda'\mathbf{u} = (n+2)r/2 + o_p(n^2/s) + o_p(1)$ from Step 4, we have

$$\frac{2s^2\breve{W}(\beta)}{(n+2)^2} - r^2 = o_p(\frac{n}{s}) + o_p(\frac{1}{n}).$$

Noting that $r^2 = T^2/n$, the proof of Theorem 1 is completed.

Proof of Corollary 1. Combining Theorem 1 and Lemma A.1, we have

$$\begin{aligned}
&\sqrt{\frac{n}{2c_n(1-c_n)^{-3}}}\Big(\frac{2s^2W(\mu)}{(n+2)^2} - c_n(1-c_n)^{-1}\Big) \\
&= \sqrt{\frac{n}{2c_n(1-c_n)^{-3}}}\Big(\frac{T^2}{n} - c_n(1-c_n)^{-1}\Big) + \sqrt{\frac{n}{2c_n(1-c_n)^{-3}}}\Big(\frac{2s^2\breve{W}(\beta)}{(n+2)^2} - \frac{T^2}{n}\Big) \\
&\xrightarrow{d} N(0,1), \quad \text{as} \quad n \to \infty.
\end{aligned}$$

Proof of Theorem 2. By the Law of Large Numbers and the Central Limit Theorem, we have

$$\frac{Y_n - 1}{\sqrt{2/p + 2/(n-p)}} \xrightarrow{d} N(0,1), \tag{A.5}$$

where Y_n follows the $F(p, n-p)$ distribution. Then Theorem 2 follows directly from Corollary 1 and (A.5).

References

Bai, Z. D. and Yin, Y. Q. (1993). Limit of the smallest eigenvalue of a large dimensional sample covariance matrix. *Ann. Probab.* **21** , 1275-1294.

Chen, S., Peng, L. and Qin Y. L. (2009). Effects of data dimension on empirical likelihood. *Biometrika* **96**, 711-722.

Chen, J., Variyath, A. M. and Abraham, B. (2008). Adjusted empirical likelihood and its properties. *J. Comput. Graph. Statist.* **17**, 426-443.

Emerson, S. and Owen, A. (2009). Calibration of the empirical likelihood method for a vector mean. *Electron. J. Statist.* **3**, 1161-1192.

Hjort, H. L., Mckeague, I. W. and Van Keilegom, I. (2009). Extending the scope of empirical likelihood. *Ann. Statist.* **37** , 1079-1115.

Jiang, T. F. (2004). The limiting distributions of eigenvalues of sample correlation matrices. *Sankhyā* **66**, 35-48.

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

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

Pan, G. M. and Zhou, W. (2011). Central limit theorem for Hotelling's T^2 statistic under large dimension. *Ann. Appl. Probab.* **21**, 1860-1910.

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

Qin, G. S. and Jing, B. Y. (2001). Empirical likelihood for censored linear regression. *Scand. J. Statist.* **28**, 661-673.

Tsao, M. (2004). Bounds on coverage probabilities of the empirical likelihood ratio confidence regions. *Ann. Statist.* **32** , 1215-1221.

Wang, Q. H. and Jing, B. Y. (1999). Empirical likelihood for partial linear models with fixed designs. *Statist. Probab. Lett.* **41**, 425-433.

Xiao, H. and Zhou, W. (2010). On the limit of the smallest eigenvalue of some sample covariance matrix. *J. Theoret. Probab.* **23**, 1-20.

Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore 637371.

E-mail: CHEN0635@e.ntu.edu.sg

Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore 637371.

E-mail: gmpan@ntu.edu.sg

Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore 637371.

E-mail: QYANG1@e.ntu.edu.sg

Department of Statistics and Applied Probability, National University of Singapore, Singapore, 117546.

E-mail: stazw@nus.edu.sg

(Received August 2013; accepted November 2014)