

On the Mahalanobis-distance based penalized empirical likelihood method in high dimensions

S. N. LAHIRI^{*,†} AND S. MUKHOPADHYAY

In this paper, we consider the penalized empirical likelihood (PEL) method of Bartolucci (2007) for inference on the population mean which is a modification of the standard empirical likelihood and employs a penalty based on the Mahalanobis-distance. We derive the asymptotic distributions of the PEL ratio statistic when the dimension of the observations increases with the sample size. Finite sample properties of the method are investigated through a small simulation study.

KEYWORDS AND PHRASES: Asymptotic null distribution, Empirical likelihood, High dimension, Regularization, Simultaneous tests.

1. INTRODUCTION

Owen (1988) introduced the empirical likelihood (EL) method for inference on population parameters in a non-parametric framework. Since its introduction, the EL method has been extended to various inference problems by several authors, notably by Diccicio, Hall and Romano (1991), Chen and Hall (1993), Qin and Lawless (1994), Owen (2001), Bertail (2006), Hjort, McKeague and van Keilegom (2009), Chen, Peng and Qin (2009), among others. In a recent work, Hjort et al. (2009) considered the asymptotic distribution of the EL ratio statistic in the high dimensional context, where the dimension p of the observations increases with the sample size n . Hjort et al. (2009)'s results allow the parameter dimension p to increase to ∞ with n at the rate $p = o(n^{1/3})$. Chen, Peng and Qin (2009) improved upon the rate restriction in Hjort et al. (2009) and established a non-degenerate limit distribution of the EL ratio statistic, allowing $p = o(n^{1/2})$, under suitable regularity conditions.

An important result of Tsao (2004) showed that for $p > n/2$, the definition of the EL breaks down on a set of non-negligible positive probability, as the convex hull of n random vectors in \mathbb{R}^p is too small a set to contain the true mean with probability one. Alternative formulations of the EL in such situations are given by Chen, Variyath and Abraham (2008), Emerson and Owen (2009) and Bartolucci (2007). The first two papers modify the EL by augmenting the original data set with additional pseudo-observations (a

single one in Chen et al. (2008) and two in Emerson and Owen (2009)) which cover a hypothesized value of the mean parameter within the convex hull of the augmented data set. In comparison, the last paper takes a penalization approach. Specifically, Bartolucci (2007) drops the convex hull constraint in the formulation of the EL and redefines the likelihood of a hypothesized value of the parameter by penalizing the unconstrained EL using the Mahalanobis distance. Since the convex hull constraint is dropped, the penalized EL (PEL) of Bartolucci (2007) is well defined for all values of $p \leq n$, as long as the sample covariance matrix is nonsingular. Other variants of the PEL where a *penalty function is added to the standard EL*, in the spirit of the penalized likelihood work of Fan and Li (2001) and Fan and Peng (2004), are considered by Otsu (2007) and Tang and Leng (2010). Bartolucci (2007) establishes a chi-squared limit of negative two times the logarithm of the PEL ratio statistic for the population mean in the case where the dimension p is *fixed and finite* for all n . *In this paper, we investigate asymptotic distribution of Bartolucci (2007)'s PEL method allowing p to grow with n .*

To briefly describe the main result of the paper, suppose that $\mathbf{X}_1, \dots, \mathbf{X}_n$ are independent and identically distributed (iid) \mathbb{R}^p -valued random vectors with mean $\boldsymbol{\mu} \in \mathbb{R}^p$, $1 < p < \infty$. The PEL of Bartolucci (2007) employs a multiplicative penalty term that replaces the convex hull constraint in the EL criterion function; It penalizes the likelihood of a hypothesized value of the population parameter $\boldsymbol{\mu}$ by a penalty factor times the Mahalanobis distance between the (weighted) sample mean and the hypothesized value of $\boldsymbol{\mu}$. See Section 2 for the exact formulation. The main result of the paper gives regularity conditions that ensure a nondegenerate limit distribution of the PEL ratio statistic. Specifically, we show that the right growth rate of the penalty factor (say, λ_n) for a non-degenerate limit is n/p . Further, with a suitable centering and with scaling by \sqrt{p} , the limit distribution of the normalized logarithm of the PEL ratio statistic is Gaussian with mean zero and a known variance. It is interesting to note that here the centering and the asymptotic variance both depend on the penalty factor λ_n . This result may be compared with the behavior of the standard EL, where twice the negative log-EL ratio statistic, with centering at 1 and scaling by $p^{1/2}$, converges in distribution to a $N(0, 1)$ variate; See, for example, Chen, Peng and Qin (2009).

We also report the results from a small simulation study to get some idea about the finite sample properties of the

*Research partially supported by NSF grant no. DMS 1007703.

†Corresponding author.

PEL. The simulation results show that large sample tests on the mean $\boldsymbol{\mu}$ based on the PEL are fairly accurate even for samples of size $n = 36$ under a varying degree of dependence among the components of the observations. Thus, the PEL can be used to carry out simultaneous tests of p hypotheses on the components of the mean vector $\boldsymbol{\mu}$ at any desired level of significance. In particular, the PEL gives a viable alternative to the commonly used approach of multiple testing in the high dimensional problems involving the mean.

The rest of the paper is organized as follows. In Section 2, we describe the PEL method of Bartolucci (2007). In Section 3, we describe the regularity conditions and establish the limit distribution of the logarithm of the PEL ratio statistic. In Section 4, we report the results from the simulation study. Finally, a proof of the main result is given in Section 5.

2. FORMULATION OF THE PEL

Let $\mathbf{X}_1, \dots, \mathbf{X}_n$ be iid random vectors with mean $\boldsymbol{\mu} \in \mathbb{R}^p$ and $p \times p$ covariance matrix Σ_p . Let

$$S_n = n^{-1} \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}}_n)(\mathbf{X}_i - \bar{\mathbf{X}}_n)'$$

denote the sample covariance matrix of $\mathbf{X}_1, \dots, \mathbf{X}_n$, where $\bar{\mathbf{X}}_n = n^{-1} \sum_{i=1}^n \mathbf{X}_i$. We shall suppose that S_n and Σ_p are nonsingular. Write A' to denote the transpose of a matrix A . Bartolucci (2007) defines the *penalized empirical likelihood (PEL)* of a plausible value $\boldsymbol{\mu}$ of the population mean as

$$(2.1) \quad L_n(\boldsymbol{\mu}) = \sup \left\{ \left(\prod_{i=1}^n \pi_i \right) \exp(-\lambda(\mathbf{v} - \boldsymbol{\mu})' S_n^{-1}(\mathbf{v} - \boldsymbol{\mu})) : \right. \\ \left. (\pi_1, \dots, \pi_n) \in \Pi_n \right\}$$

where $\Pi_n = \{(\pi_1, \dots, \pi_n) : \pi_i \in [0, 1] \text{ for all } i \text{ and } \sum_{i=1}^n \pi_i = 1\}$, $\mathbf{v} = \sum_{i=1}^n \pi_i \mathbf{X}_i$, and $\lambda = \lambda_n \in [0, \infty)$ is a penalty factor. Thus, the PEL of $\boldsymbol{\mu}$ penalizes the likelihood $\prod_{i=1}^n \pi_i$ by an exponential function of the Mahalanobis distance between $\boldsymbol{\mu}$ and \mathbf{v} . Note that \mathbf{v} is a point in the convex hull $\mathcal{C}(\mathcal{X}_n)$ (say) of the observations $\mathbf{X}_1, \dots, \mathbf{X}_n$ when the i th observation \mathbf{X}_i is assigned the probability π_i . For a value of $\boldsymbol{\mu}$ that lies in $\mathcal{C}(\mathcal{X}_n)$, the penalization term attains the value 1 and drops out. In this case, one may also use the standard EL for $\boldsymbol{\mu}$:

$$L_n^{\text{STD}}(\boldsymbol{\mu}) = \sup \left\{ \left(\prod_{i=1}^n \pi_i \right) : (\pi_1, \dots, \pi_n) \in \Pi_n, \right. \\ \left. \sum_{i=1}^n \pi_i \mathbf{X}_i = \boldsymbol{\mu} \right\}.$$

However, for a $\boldsymbol{\mu}$ that lies outside $\mathcal{C}(\mathcal{X}_n)$, the standard EL for $\boldsymbol{\mu}$ becomes undefined and therefore, can not be used in

practice. As pointed by Tsao (2004), this undesirable phenomenon occurs with a non-trivial probability in high dimensions, which delimits the use of the standard EL in high dimensional problems. In contrast, the PEL of Bartolucci (2007) remains well-defined for all values of $\boldsymbol{\mu} \in \mathbb{R}^p$. Since Bartolucci (2007)'s PEL is based on the Mahalanobis distance, it is well defined for all $p \leq n$, provided S_n^{-1} exists. In principle, the same basic approach can be utilized for $p > n$ if we replace S_n^{-1} by a suitable location invariant matrix that remains invertible for $p > n$ (cf. Lahiri and Mukhopadhyay (2012)). However, in this paper, we shall restrict our attention to the original formulation of PEL by Bartolucci (2007) and assume that S_n^{-1} exists.

The PEL $L_n(\cdot)$ can be used to conduct simultaneous tests of hypotheses of the form

$$(2.2) \quad H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0 \quad \text{against} \quad H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$$

where $\boldsymbol{\mu}_0$ is a specified vector in \mathbb{R}^p . A PEL likelihood ratio test rejects the null hypotheses for small values of

$$R_n(\boldsymbol{\mu}_0) \equiv \frac{L_n(\boldsymbol{\mu}_0)}{\sup_{\boldsymbol{\mu}} L_n(\boldsymbol{\mu})}.$$

It is easy to check that the denominator equals the maximum of the product $\prod_{i=1}^n \pi_i$ over Π_n and is attained when $\pi_1 = \dots = \pi_n = n^{-1}$. Hence, the PEL ratio statistic for testing (2.2) is given by

$$(2.3) \quad R_n(\boldsymbol{\mu}_0) = n^n L_n(\boldsymbol{\mu}_0).$$

For a given $\alpha \in (0, 1)$, the PEL ratio-test for testing (2.2) with nominal level α has a rejection region, given by

$$(2.4) \quad \{-2 \log R_n(\boldsymbol{\mu}_0) > \rho_n(\alpha)\}$$

where $\rho_n(\alpha)$ is an upper α -critical point of the distribution of $-2 \log R_n(\boldsymbol{\mu}_0)$ under H_0 . Exact determination of $\rho_n(\alpha)$ is a difficult task, as it heavily depends on the sampling distribution of the observations $\mathbf{X}_1, \dots, \mathbf{X}_n$. In the next Section, we derive the asymptotic distribution of $-2 \log R_n(\boldsymbol{\mu}_0)$ under H_0 that can be used to construct an approximate large sample level α test for (2.2).

It is worth noting that we can similarly define a version of the PEL for testing about more general parameters $\boldsymbol{\theta}$ that are specified by a set of p estimating equations

$$E\psi_{jn}(\mathbf{X}_1; \boldsymbol{\theta}) = 0 \quad \text{for } j = 1, \dots, p.$$

In this case, the PEL for a given value of $\boldsymbol{\theta}$ is defined by

$$(2.5) \quad L_n(\boldsymbol{\theta}) = \sup \left\{ \left(\prod_{i=1}^n \pi_i \right) \exp(-\lambda \mathbf{u}' \tilde{S}_n^{-1} \mathbf{u}) : \right. \\ \left. (\pi_1, \dots, \pi_n) \in \Pi_n \right\}$$

where $\mathbf{u} = \sum_{i=1}^n \pi_i \psi_{jn}(\mathbf{X}_i; \theta)$ and where \tilde{S}_n is an estimator of the covariance of the random vector $(\psi_{1n}(\mathbf{X}_1; \theta), \dots, \psi_{pn}(\mathbf{X}_1; \theta))$. The case of the population mean $\theta = \mu = E\mathbf{X}_1$ corresponds to $\psi_{jn}(\mathbf{x}; \theta) = \mathbf{e}'_j(\mathbf{x} - \theta)$ where $\mathbf{e}_1, \dots, \mathbf{e}_p$ denotes the standard basis of \mathbb{R}^p . Distributional results similar to those presented in Section 3 can also be derived for the more general case, by retracing the arguments given in Section 5 below.

3. MAIN RESULTS

3.1 Conditions

In this section, we derive the asymptotic null distribution of the log-PEL ratio statistic (cf. (2.3)) in the high dimensional case where the dimension $p = p_n \rightarrow \infty$ with the sample size n . Since p changes with n , here we adopt a triangular array framework to study the asymptotic behavior of $-2 \log R_n(\mu_0)$. More precisely, we suppose that $\{\mathbf{X}_{in} : 1 \leq i \leq n\}_{n \geq 1}$ is a triangular array of iid random vectors with $E\mathbf{X}_{1n} = \boldsymbol{\mu}_n \equiv \boldsymbol{\mu} \in \mathbb{R}^p$ and nonsingular covariance matrix $\text{Cov}(\mathbf{X}_{1n}) = \Sigma_{pn} \equiv \Sigma_p$. Here and in the following, we shall drop the subscript n to simplify notation, provided it does not create any confusion. Next define $\mathbf{Y}_{in} \equiv \mathbf{Y}_i = \Sigma_p^{-1/2}(\mathbf{X}_{in} - \boldsymbol{\mu}_0)$, $1 \leq i \leq n$ where $\boldsymbol{\mu}_0$ is the value of $\boldsymbol{\mu}$ specified under the null hypothesis $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ and where $\Sigma_p^{-1/2} = [\Sigma_p^{1/2}]^{-1}$ and $\Sigma_p^{1/2}$ is a $p \times p$ matrix satisfying $\Sigma_p^{1/2}(\Sigma_p^{1/2})' = \Sigma_p$. Denote the j th component of \mathbf{Y}_i by Y_{ij} , $1 \leq i \leq n$, $1 \leq j \leq p$. Also, let

$$\tilde{S}_n = n^{-1} \sum_{i=1}^n (\mathbf{Y}_i - \bar{\mathbf{Y}}_n)(\mathbf{Y}_i - \bar{\mathbf{Y}}_n)'$$

denote the sample covariance matrix of $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ where $\bar{\mathbf{Y}}_n = n^{-1} \sum_{i=1}^n \mathbf{Y}_i$ is the mean of the \mathbf{Y}_i 's. Note that \tilde{S}_n is an estimator of \mathbb{I}_p , the identity matrix of order p . Let $\mathbb{N} = \{1, 2, \dots\}$ be the set of all positive integers. For any $k \in \mathbb{N}$ and $\mathbf{x} = (x_1, \dots, x_k)' \in \mathbb{R}^k$, let $\|\mathbf{x}\| = (x_1^2 + \dots + x_k^2)^{1/2}$ and for a $k \times k$ matrix A , let $\|A\| = \sup\{\|A\mathbf{x}\| : \mathbf{x} \in \mathbb{R}^k, \|\mathbf{x}\| = 1\}$.

We shall use the following conditions to derive the limit distribution of the PEL ratio statistic:

Conditions:

(C.1) There exist $\delta \in (0, \infty)$ and $C \in (0, \infty)$ such that

$$\max\{E|Y_{1j}|^{4+\delta} : 1 \leq j \leq p\} < C$$

for all $n \geq 1$.

(C.2) There exists a sequence of real numbers $\{\alpha(n)\}_{n \geq 1} \subset [0, 1]$ such that

(i) $\sum_{n=1}^{\infty} \alpha(n)^{\frac{\delta}{4+\delta}} < \infty$, where δ is as in Condition (C.1);

(ii) For any $m \in \mathbb{N}$,

$$\sup \left\{ |P(A \cap B) - P(A)P(B)| : A \in \mathcal{F}_{n,1}^j, \right. \\ \left. B \in \mathcal{F}_{n,j+m}^p, n, j \in \mathbb{N} \right\} \leq \alpha(m)$$

where for any integers $a, b, n \in \mathbb{N}$, $\mathcal{F}_{n,a}^b = \sigma(Y_{1jn} : j \in \{1, \dots, p\} \cap [a, b])$ (with the convention that the σ -field generated by an empty collection of random variables is the trivial σ -field).

$$(C.3) \quad \|\tilde{S}_n - \mathbb{I}_p\| = o_p(p^{-1/2}).$$

Now we comment on the conditions. Condition (C.1) is a moment condition that requires slightly more than the fourth moment of the components of \mathbf{Y}_1 to be finite. This condition can be reduced to a suitable uniform integrability condition on the fourth power of the Y_{1j} 's (see the statement of the Proposition in Section 5 which is applied with the choice $\delta = 2$ therein) at the cost of strengthening the strong mixing condition in (C.2) (to a suitable ρ -mixing condition, say). Condition (C.2) is a strong mixing condition, which together with Condition (C.1) yields the Central Limit Theorem (CLT) for the log PEL ratio statistic. From the proofs given in Section 5 below, it follows that the leading term of the stochastic approximation to the centered and scaled version of $-2 \log R_n(\mu_0)$ is given by (a multiple of) the sum

$$T_n \equiv p^{-1/2} \sum_{j=1}^p ([\sqrt{n}\bar{Y}_{jn}]^2 - 1).$$

Thus, for asymptotic normality of $-2 \log R_n(\mu_0)$, the main task is to establish a CLT for T_n . Although by scaling with $\Sigma_p^{-1/2}$, the components of \mathbf{Y}_1 become *uncorrelated*, in general they are *not independent*. The strong mixing condition specifies the degree of dependence among the components of the scaled variable \mathbf{Y}_1 that ensures validity of the normal limit (cf. Ibragimov and Linnik (1971)). In the special case, where \mathbf{X}_1 has a multivariate normal distribution, the scaling by $\Sigma_p^{-1/2}$ makes the components of \mathbf{Y}_1 completely *independent* and Condition (C.2) is trivially satisfied. As commented by a referee, in general, the components in \mathbf{X}_1 can be arbitrarily ordered, which does not directly follow the serial dependence structure of our time series formulation. However, it is easy to check that the PEL criterion function is invariant under permutation of the co-ordinates of \mathbf{X}_1 and therefore, the distributional result remains valid if the components of \mathbf{X}_1 has a time series ordering up to a permutation, as in Lahiri and Mukhopadhyay (2012). Next consider Condition (C.3) which requires the sample covariance matrix of the centered and scaled variables $\{\mathbf{Y}_i : 1 \leq i \leq n\}$ to converge to the corresponding population covariance matrix \mathbb{I}_p at a rate faster than $p^{-1/2}$. Since the sample covariance matrix of the \mathbf{Y}_i is based on n iid observations, the best possible rate of convergence for the difference $\|\Sigma_p^{-1/2} S_n (\Sigma_p^{-1/2})' - \mathbb{I}_p\|$ is $O_p(n^{-1/2})$. Hence, Condition (C.3) necessarily requires

p to grow at most at the rate $p = o(n)$. Also, it is easy to check that under the moment condition (C.1), the trace norm bound on $\|\tilde{S}_n - \mathbb{I}_p\|$ is $o_p(p^{-1/2})$ (a sufficient condition), provided $p = o(n^{1/3})$. Hence, (C.3) allows a growth rate of p that is at least as fast as $p = o(n^{1/3})$.

3.2 The main result

The following result gives the asymptotic distribution of the PEL ratio statistic. For two sequences $\{a_n\}, \{b_n\} \subset (0, \infty)$, write $a_n \sim b_n$ if $\lim_{n \rightarrow \infty} a_n/b_n = 1$.

Theorem. *Suppose that Conditions (C.1)–(C.4) holds, $p \rightarrow \infty$ and $\lambda \sim c_0 \frac{2\lambda p}{n}$ as $n \rightarrow \infty$, for some $c_0 \in (0, \infty)$. Then, under $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$,*

$$p^{1/2} \left(\log R_n(\boldsymbol{\mu}_0) + \frac{\lambda p}{n} \right) \rightarrow^d N(0, 2c_0^2).$$

Thus, it follows that the log PEL ratio statistic, with centering at $-\frac{2\lambda p}{n}$ and scaling by $p^{1/2}$ converges in distribution to a normal limit. An alternative but equivalent statement that gives the asymptotic distribution of the more standard quantity, namely, twice the negative logarithm of the PEL ratio statistic, is the following:

$$p^{1/2} \left(-2 \log R_n(\boldsymbol{\mu}_0) - \frac{2\lambda p}{n} \right) \rightarrow^d N(0, 8c_0^2).$$

In contrast to the chi-squared p limit law of $-2 \log R_n(\boldsymbol{\mu}_0)$ in the fixed p case, in the high dimensional case, $-2 \log R_n(\boldsymbol{\mu}_0)$ has a somewhat non-standard limit behavior when the penalization is based on the Mahalanobis distance. It needs a non-standard centering to converge to a nondegenerate limit under scaling by $p^{1/2}$. Further, the variance of the limit distribution also depends on the penalization term through the penalty factor λ . A simple consequence of the Theorem is that under the conditions (C.1)–(C.4),

$$-2 \log R_n(\boldsymbol{\mu}_0) \rightarrow_p 2c_0 \quad \text{as } n \rightarrow \infty.$$

In general, it is not possible to replace the centering $\frac{2\lambda p}{n}$ by $2c_0$. However, note that all three quantities λ , p and n are *known* and hence, these may be used directly to center $-2 \log R_n(\boldsymbol{\mu}_0)$.

It is worth noting that due to the effect of the penalization in the variance of the limit distribution, choosing a penalty factor λ corresponding to a small c_0 reduces the variability of $-2 \log R_n(\boldsymbol{\mu}_0)$ from the centering constant $\frac{2\lambda p}{n}$. However, if λ is chosen such that

$$\lambda/[n/p] \rightarrow 0$$

(i.e. $c_0 = 0$), then a scaling sequence different from $p^{1/2}$ is needed to ensure a nondegenerate limit. A similar comment applies when $c_0 = \infty$.

From the proof, it follows that $-2 \log R_n(\boldsymbol{\mu}_0)$ behaves approximately like a $\frac{2c_0}{p} \chi^2(p)$ random variate. In particular,

when $c_0 = p/2$ and p is finite, one gets the chi-squared distribution with p degrees of freedom as the limit as in the case of Wilk's theorem. Bartolucci (2007) proved a version of Wilk's theorem for $-2 \log R_n(\boldsymbol{\mu}_0)$ in the case when p is fixed and finite. Thus, the results of this paper supplements Bartolucci (2007)'s result by extending it to the increasing dimensional case.

4. SIMULATION

In this Section, we report the results from a small simulation study to gain some insight into the finite sample behavior of Bartolucci (2007)'s PEL method. We consider two pairs of values of the sample size and the dimension $(n, p) = (36, 6)$ and $(n, p) = (100, 10)$, where the growth rate of p is given by $p \sim \sqrt{n}$. Also, for the sake of comparison, we have included results on higher values of p where the sample covariance matrix is replaced with a regularized version. Specifically, we replaced S_n by the tapered covariance estimator by $\tilde{\Sigma}_R = S_n * W$, where $*$ denotes the Hadamaard product of matrices and where the weights w_{ij} are given by (cf. Cai, Zhang and Zhou (2010))

$$w_{ij} = \begin{cases} 1 & \text{when } |i - j| \leq l/2 \\ 2(1 - \frac{|i-j|}{l}) & \text{when } l/2 \leq |i - j| \leq l \\ 0 & \text{otherwise.} \end{cases}$$

We generated iid samples from two models, specified by a stretch of a time series of length p . The time series models we considered are (i) a Moving Average (MA) model of order 5 with co-efficients given by $(.5)^k : k = 0, 1, \dots, 5$ with iid $N(0, 1)$ errors and (ii) a stationary Gaussian process with autocorrelation function $r(\cdot)$ where $r(0) = 1$ and

$$r(k) = \frac{1}{2} [(k+1)^{2H} + (k-1)^{2H} - 2k^{2H}], \quad k \geq 1,$$

with $H = .9$. Note that the first model leads to a banded covariance matrix Σ_p while the second model yields a Σ_p that has all its entries non-zero. Further, model (i) corresponds to weak dependence among the components of \mathbf{X}_1 while model (ii) corresponds to long range dependence (LRD) (cf. Beran (1994)). We have set $l = 5$ for the MA model and $l = \lfloor n^{1/(2\alpha+1)} \rfloor$ with $\alpha = 2 - 2H$ for the LRD model.

Tables 1 and 2 below give the attained level of significance in a test of the hypotheses $H_0 : \boldsymbol{\mu} = \mathbf{0}$ against $H_1 : \boldsymbol{\mu} \neq \mathbf{0}$ at a nominal .05 level of significance for $n = 36$ and $n = 100$ respectively, for various choices of p , which reflect the four cases $p \ll n$, $p \sim n$, $p > n$, $p \gg n$. In each case, the results are based on 500 simulation runs and the choice of the penalty parameter is given by $\lambda = n/p$, i.e., $c_0 = 1$.

It follows from Tables 1 and 2 that the PEL ratio tests based on the normal limit of the Theorem of Section 3 gives a reasonable answer even for samples of size $n = 36$, under both forms of dependence among the components of \mathbf{X}_1 when $p = \sqrt{n}$. The regularized version performs well for all

Table 1. Attained level of significance of a PEL test of $H_0 : \boldsymbol{\mu} = \mathbf{0}$ vs. $H_1 : \boldsymbol{\mu} \neq \mathbf{0}$ for $n = 36$ at a nominal .05 level of significance

Methods	$p = 6$		$p = 36$		$p = 50$		$p = 500$	
	MA	LRD	MA	LRD	MA	LRD	MA	LRD
Bartolucci	.046	.045	NA	NA	NA	NA	NA	NA
Regularized – Bartolucci	.043	.069	.054	.083	.065	.116	.061	.165

Table 2. Attained level of significance of a PEL test of $H_0 : \boldsymbol{\mu} = \mathbf{0}$ vs. $H_1 : \boldsymbol{\mu} \neq \mathbf{0}$ for $n = 100$ at a nominal .05 level of significance

Methods	$p = 10$		$p = 100$		$p = 150$		$p = 1000$	
	MA	LRD	MA	LRD	MA	LRD	MA	LRD
Bartolucci	.048	.057	NA	NA	NA	NA	NA	NA
Regularized – Bartolucci	.049	.071	.047	.091	.059	.129	.064	.198

cases of $p \geq n$ under weak dependence. The performance deteriorates for large p 's under LRD.

5. PROOFS

Let $a_n^2 = p/n$. Note that $a_n \rightarrow 0+$ as $n \rightarrow \infty$. Let $\mathbb{1}(\cdot)$ denote the indicator function. Let $C, C(\cdot)$ denote generic constants that depend only on their arguments (if any), but not on n . Unless otherwise specified, dependence on (limiting) population quantities (such as the moments, mixing co-efficients, etc.) are dropped to simplify notation, and limits in all order symbols are taken by letting $n \rightarrow \infty$.

Proposition. Let $\{\mathbf{X}_{1n}, \dots, \mathbf{X}_{nn}\}_{n \geq 1}$ be a triangular array of identically distributed p dimensional random vectors and let $\mathbf{X}_{1n} \equiv \mathbf{X}_1 = (X_{11}, \dots, X_{1p})$. Suppose that $\{p^{-1} \sum_{j=1}^p |X_{1j}|^{2+\delta} : n \geq 1\}$ is uniformly integrable, i.e.,

$$\lim_{t \rightarrow \infty} \sup_{n \geq 1} E \left[\left\{ p^{-1} \sum_{j=1}^p |X_{1j}|^{2+\delta} \right\} \times \mathbb{1} \left(p^{-1} \sum_{j=1}^p |X_{1j}|^{2+\delta} > t \right) \right] = 0.$$

Let $M_n = \max\{\|\mathbf{X}_i\| : 1 \leq i \leq n\}$. Then $M_n = o_P(n^{\frac{1}{2+\delta}} \sqrt{p})$.

Proof. Note that by Hölder's inequality,

$$\begin{aligned} \|\mathbf{X}_1\|^{2+\delta} &= \left(\sum_{j=1}^p X_{1j}^2 \right)^{\frac{2+\delta}{2}} \\ &\leq \left[\left(\sum_{j=1}^p |X_{1j}|^{2+\delta} \right)^{\frac{2}{2+\delta}} p^{1-\frac{2}{2+\delta}} \right]^{\frac{2+\delta}{2}} \\ &= p^{\delta/2} \sum_{j=1}^p |X_{1j}|^{2+\delta}. \end{aligned}$$

Fix $\epsilon \in (0, 1)$. Let $\epsilon_{1n} = \epsilon n^{\frac{1}{2+\delta}} \sqrt{p}$. Using the above relation and the uniform integrability condition, it is easy to check that

$$\begin{aligned} P(M_n > \epsilon_{1n}) &\leq \epsilon_{1n}^{-(2+\delta)} E M_n^{2+\delta} \mathbb{1}(M_n > \epsilon_{1n}) \\ &\leq \epsilon_{1n}^{-(2+\delta)} \sum_{k=1}^n E \|\mathbf{X}_k\|^{2+\delta} \mathbb{1}(\|\mathbf{X}_k\| > \epsilon_{1n}) \mathbb{1}(M_n = \|\mathbf{X}_k\|) \\ &\leq n \epsilon_{1n}^{-(2+\delta)} E \|\mathbf{X}_k\|^{2+\delta} \mathbb{1}(\|\mathbf{X}_k\| > \epsilon_{1n}) \\ &\leq n \epsilon_{1n}^{-(2+\delta)} p^{\delta/2} E \left(\sum_{j=1}^p |X_{1j}|^{2+\delta} \right) \\ &\quad \times \mathbb{1} \left(\sum_{j=1}^p |X_{1j}|^{2+\delta} > p^{-\delta/2} \epsilon_{1n}^{2+\delta} \right) \\ &= \epsilon^{-(2+\delta)} E \left(p^{-1} \sum_{j=1}^p |X_{1j}|^{2+\delta} \right) \\ &\quad \times \mathbb{1} \left(p^{-1} \sum_{j=1}^p |X_{1j}|^{2+\delta} > \epsilon^{2+\delta} n \right) \\ &= o(1). \end{aligned}$$

Hence the proposition follows. \square

Proof of the Theorem. Note that by the definition of

$$L(\boldsymbol{\mu}_0) = \max \left\{ \prod_{i=1}^n \pi_i \exp(-\lambda(\mathbf{v} - \boldsymbol{\mu}_0)' S_n^{-1}(\mathbf{v} - \boldsymbol{\mu}_0)) : (\pi_1, \dots, \pi_n) \in \Pi_n \right\}$$

with $\mathbf{v} = \sum_{i=1}^n \pi_i \mathbf{X}_i$, the PEL ratio statistic $R_n(\boldsymbol{\mu}_0) = n^n L(\boldsymbol{\mu}_0)$ is invariant under translation and matrix scaling. Hence, without loss of generality, for the rest of the proof, we

may therefore assume that $\boldsymbol{\mu}_0 = \mathbf{0}$ and $\Sigma_p = \mathbb{I}_p$, the identity matrix of order p . This, in particular, implies that $\mathbf{Y}_i = \mathbf{X}_i$ for all $1 \leq i \leq n$ and that $\hat{S}_n = S_n$. Therefore, by Condition (C.3), S_n^{-1} exists for n large and $\|S_n^{-1}\| = O_p(1)$.

Next note that the maximizing π_i 's must all be positive (otherwise, we get $L_n(\boldsymbol{\mu}_0) = 0$, which is the minimum) and therefore, $\log R_n(\boldsymbol{\mu}_0)$ is well-defined. To find $R_n(\boldsymbol{\mu}_0)$, we use the standard Lagrangian method and consider the function

$$\begin{aligned} g(\pi_1, \dots, \pi_n; \gamma) &= \log R_n(\boldsymbol{\mu}_0) - n\gamma \left(\sum_{i=1}^n \pi_i - 1 \right) \\ &= \sum_{i=1}^n \log(n\pi_i) - \lambda \mathbf{v}' S_n^{-1} \mathbf{v} - n\gamma \left(\sum_{i=1}^n \pi_i - 1 \right) \end{aligned}$$

for $\pi_i > 0$ for $1 \leq i \leq n$ and $\gamma \in \mathbb{R}$. The maximizing π_i 's is a solution to the system of equations:

$$\frac{\partial g}{\partial \pi_i} = 0 \quad \text{for } 1 \leq i \leq n \quad \text{and} \quad \frac{\partial g}{\partial \gamma} = 0$$

which, in turn, lead to the equations for $1 \leq i \leq n$,

$$(5.1) \quad 0 = \frac{1}{\pi_i} - \lambda \sum_{k=1}^p \sum_{j=1}^p (S_n^{-1})_{jk} \{X_{ij}v_k + X_{ik}v_j\} - n\gamma,$$

$$(5.2) \quad 0 = \sum_{i=1}^n \pi_i - 1,$$

where $(S_n^{-1})_{jk}$ is the (j, k) th element of S_n^{-1} . Multiplying the i th equation in (5.1) by π_i and summing over the resulting n equations, we get

$$\gamma = 1 - \frac{\lambda}{n} \sum_{i=1}^n \sum_{j=1}^p \sum_{k=1}^p (S_n^{-1})_{jk} \{X_{ij}v_k + X_{ik}v_j\} \pi_i.$$

Now substituting this in (5.1) and using (5.2), after some algebra, one can show that

$$\pi_i = \frac{1}{n} \cdot \frac{1}{1 + \lambda_1 [\mathbf{v}' S_n^{-1} \mathbf{X}_i - \mathbf{v}' S_n^{-1} \mathbf{v}]}, \quad 1 \leq i \leq n$$

where $\lambda_1 = 2\lambda/n$. As a consequence, $\mathbf{v} = \sum_{i=1}^n \pi_i \mathbf{X}_i$ must satisfy the equation

$$(5.3) \quad \begin{aligned} \mathbf{v} &= \frac{1}{n} \sum_{i=1}^n \frac{\mathbf{X}_i}{1 + \lambda_1 [\mathbf{v}' S_n^{-1} \mathbf{X}_i - \mathbf{v}' S_n^{-1} \mathbf{v}]} \\ &= \bar{\mathbf{X}}_n + \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i \frac{\lambda_1 [\mathbf{v}' S_n^{-1} \mathbf{v} - \mathbf{v}' S_n^{-1} \mathbf{X}_i]}{1 + \lambda_1 [\mathbf{v}' S_n^{-1} \mathbf{X}_i - \mathbf{v}' S_n^{-1} \mathbf{v}]} \end{aligned}$$

We shall use a fixed point theorem to derive an approximation to \mathbf{v}_0 , the value of \mathbf{v} corresponding to the maximizing π_i 's. To that end, let

$$\gamma_i = \lambda_1 [\mathbf{v}' S_n^{-1} \mathbf{X}_i - \mathbf{v}' S_n^{-1} \mathbf{v}], \quad 1 \leq i \leq n.$$

Also, let $a_n^2 = p/n$ and set $M_n = \max\{\|\mathbf{X}_i\| : 1 \leq i \leq n\}$. Then, by the Proposition (with $\delta = 2$), for any \mathbf{v} with $\|\mathbf{v}\| = O_p(a_n)$,

$$(5.4) \quad \begin{aligned} &\max\{|\gamma_i| : 1 \leq i \leq n\} \\ &\leq \lambda_1 [\|\mathbf{v}\| \|S_n^{-1}\| M_n + \|\mathbf{v}\|^2 \|S_n^{-1}\|] \\ &= O(p^{-1}) [O_p(a_n) \cdot O_p(1) \cdot o_p(\sqrt{pn}^{\frac{1}{4}}) + O_p(a_n^2) \cdot O_p(1)] \\ &= o_p(n^{-1/4}). \end{aligned}$$

Next note that $\sum_{i=1}^n \|\mathbf{X}_i\|^2 = O_p(np)$ (as the expectation of the left side is $O(np)$). Hence, for $\|\mathbf{v}\| = O_p(a_n)$,

$$\begin{aligned} &\left\| n^{-1} \sum_{i=1}^n \mathbf{X}_i \frac{\lambda_1 [\mathbf{v}' S_n^{-1} \mathbf{v} - \mathbf{v}' S_n^{-1} \mathbf{X}_i]}{1 + \lambda_1 [\mathbf{v}' S_n^{-1} \mathbf{X}_i - \mathbf{v}' S_n^{-1} \mathbf{v}]} \right\| \\ &\leq \frac{\lambda_1}{n} \left\{ \left\| \sum_{i=1}^n \mathbf{X}_i \frac{\mathbf{v}' S_n^{-1} \mathbf{X}_i}{1 + \gamma_i} \right\| + \mathbf{v}' S_n^{-1} \mathbf{v} \left\| \sum_{i=1}^n \frac{\mathbf{X}_i}{1 + \gamma_i} \right\| \right\} \\ &\leq \frac{\lambda_1}{n} \frac{[\sum_{i=1}^n \|\mathbf{X}_i\|^2 \|\mathbf{v}\| \|S_n^{-1}\| + \sum_{i=1}^n \|\mathbf{X}_i\| \|\mathbf{v}\|^2 \|S_n^{-1}\|]}{(1 + \max\{|\gamma_i| : 1 \leq i \leq n\})} \\ &= \frac{\lambda_1}{n} \frac{[O_p(np) \cdot O_p(a_n) \cdot O_p(1) + O_p(n\sqrt{p}) \cdot O_p(a_n^2) \cdot O_p(1)]}{[1 + o_p(1)]} \\ &= O_p(a_n). \end{aligned}$$

Since $E\|\bar{\mathbf{X}}_n\|^2 = \sum_{j=1}^p E\bar{X}_{jn}^2 = O_p(a_n^2)$, it now follows that the right side of (5.3) is $O_p(a_n)$, whenever $\|\mathbf{v}\| = O_p(a_n)$. Hence, by Brouwer's Fixed Point theorem (cf. Milnor (1965)), a solution $\mathbf{v} = \mathbf{v}_0$ to (5.3) exists and $\|\mathbf{v}_0\| = O_p(a_n)$.

Next we refine (5.3) to get a more accurate approximation to \mathbf{v}_0 . Write $\nu_i^0 = \lambda_1 \mathbf{v}_0' S_n^{-1} \mathbf{X}_i$, $1 \leq i \leq n$, $D_n = \max\{|\nu_i^0| : 1 \leq i \leq n\}$, and let $\Gamma_n^0 = \lambda_1 \mathbf{v}_0' S_n^{-1} \mathbf{v}_0$. Note that by (5.4),

$$(5.5) \quad \begin{aligned} \mathbf{v}_0 &= \bar{\mathbf{X}}_n + n^{-1} \sum_{i=1}^n \frac{\mathbf{X}_i [-\nu_i^0 + \Gamma_n^0]}{1 + \nu_i^0 - \gamma_n^0} \\ &= \bar{\mathbf{X}}_n + n^{-1} \sum_{i=1}^n \mathbf{X}_i (-\nu_i^0 + \Gamma_n^0) [1 - \nu_i^0 + (\nu_i^0)^2 \\ &\quad + O_p(D_n^3 + n^{-1})] \\ &= \bar{\mathbf{X}}_n (1 + \Gamma_n^0) - n^{-1} \sum_{i=1}^n \mathbf{X}_i \{\nu_i^0 - (\nu_i^0)^2 + (\nu_i^0)^3\} \\ &\quad + \mathbf{R}_n^{[0]}, \quad \text{say,} \end{aligned}$$

where the vector of remainder terms $\mathbf{R}_n^{[0]} = (R_{1n}^{[0]}, \dots, R_{pn}^{[0]})'$ satisfies $|R_{jn}^{[0]}| \leq (n^{-1} \sum_{i=1}^n |X_{ij}|) \cdot O_p(D_n^4 + n^{-1} D_n)$ for all

$1 \leq j \leq p$ and the O_p^u -term in the bound does *not* depend on $j \in \{1, \dots, p\}$. It is easy to check that

$$\begin{aligned} n^{-1} \sum_{i=1}^n \mathbf{X}_i \nu_i^0 &= \frac{\lambda_1}{n} \left[\sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i' \right] S_n^{-1} \mathbf{v}_0 \\ &= \lambda_1 [\mathbf{v}_0 + \bar{\mathbf{X}}_n \bar{\mathbf{X}}_n' S_n^{-1} \mathbf{v}_0]. \end{aligned}$$

Hence, by rearranging the terms, we have

$$(5.6) \quad (1 + \lambda_1) \mathbf{v}_0 = (1 + \Gamma_n^0) \bar{\mathbf{X}}_n - \lambda_1 \bar{\mathbf{X}}_n (\bar{\mathbf{X}}_n' S_n^{-1} \mathbf{v}_0) + \sum_{r=2}^3 (-1)^r \left[n^{-1} \sum_{i=1}^n \mathbf{X}_i (\nu_i^0)^r \right] + \mathbf{R}_n^{[0]}.$$

We shall next derive a bound on the last but one term. Let $W_{ij} = |X_{ij}| \mathbb{1}(|X_{ij}| > \sqrt{n})$, $1 \leq i \leq n, 1 \leq j \leq p$. Note that

$$\begin{aligned} p^{-1} \sum_{j=1}^p E \left(\sum_{i=1}^n \{W_{ij} - EW_{ij}\} \right)^2 &\leq np^{-1} \sum_{j=1}^p EW_{1j}^2 \\ &\leq p^{-1} \sum_{j=1}^p EX_{1j}^4 = O(1), \end{aligned}$$

and similarly, $p^{-1} \sum_{j=1}^p [EW_{ij}]^2 \leq n^{-3} [\max\{EX_{1j}^4 : 1 \leq j \leq p\}]^2 = O(n^{-3})$. Also, in (5.10) below, we show that $\sum_{i=1}^n (\nu_i^0)^2 = O_p(p^{-1} + n^{-1})$. Hence, for $r = 2, 3$, we get

$$(5.7) \quad \begin{aligned} \left\| n^{-1} \sum_{i=1}^n \mathbf{X}_i (\nu_i^0)^r \right\|^2 &= n^{-2} \sum_{j=1}^p \left(\sum_{i=1}^n X_{ij} [\nu_i^0]^r \right)^2 \\ &\leq n^{-2} \sum_{j=1}^p \left(\sqrt{n} \sum_{i=1}^n [\nu_i^0]^r + D_n^r \sum_{i=1}^n |X_{ij}| \mathbb{1}(|X_{ij}| > \sqrt{n}) \right)^2 \\ &\leq 2pn^{-1} \left\{ \sum_{i=1}^n [\nu_i^0]^2 \right\}^2 D_n^{2(r-2)} + 2D_n^{2r} \cdot n^{-2} \\ &\quad \times \sum_{j=1}^p \left(nEW_{1j} + \sum_{i=1}^n [W_{ij} - EW_{ij}] \right)^2 \\ &= D_n^{2(r-2)} \cdot O_p(pn^{-1} \{p^{-2} + n^{-2}\}) + D_n^{2r} \cdot O_p \\ &\quad \times (n^{-2} p [1 + n^2 n^{-3}]) \\ &= O_p(p [n^{-1} p^{-2} + n^{-3}]). \end{aligned}$$

Hence, using (5.4) (with $\mathbf{v} = \mathbf{v}_0$), (5.7) and the bounds on the components of $\mathbf{R}_n^{[0]}$ following (5.5), from (5.6), we get

$$(5.8) \quad \begin{aligned} \|\mathbf{v}_0 - \bar{\mathbf{X}}_n\| &\leq \|\bar{\mathbf{X}}_n\| \left| \frac{1 + \Gamma_n^0}{1 + \lambda_1} - 1 \right| + \lambda_1 \|\bar{\mathbf{X}}_n\|^2 \|S_n^{-1}\| \|\mathbf{v}_0\| \\ &\quad + \sum_{r=2}^3 \left\| n^{-1} \sum_{i=1}^n \mathbf{X}_i (\nu_i^0)^r \right\| + \|\mathbf{R}_n^{[0]}\| \\ &= a_n [O_p(p^{-1}) + o_p(n^{-1/2})]. \end{aligned}$$

Next we derive the limit distribution of the PEL ratio statistic. It is easy to check that

$$(5.9) \quad \begin{aligned} -2 \log R_n(\boldsymbol{\mu}_0) &= 2 \sum_{i=1}^n \log \left(1 + \lambda_1 [\mathbf{v}_0' S_n^{-1} \mathbf{X}_i - \mathbf{v}_0' S_n^{-1} \mathbf{v}_0] \right) + 2\lambda \mathbf{v}_0' S_n^{-1} \mathbf{v}_0 \\ &= 2 \sum_{i=1}^n [\nu_i^0 - \Gamma_n^0] + 2\lambda \mathbf{v}_0' S_n^{-1} \mathbf{v}_0 + R_{1n}, \quad \text{say} \\ &= 2\lambda_1 \sum_{i=1}^n \nu_i^0 - 2n\Gamma_n^0 + 2\lambda \mathbf{v}_0' S_n^{-1} \mathbf{v}_0 + R_{1n} \\ &= 2\lambda_1 n \mathbf{v}_0' S_n^{-1} \bar{\mathbf{X}}_n - n\lambda_1 \mathbf{v}_0' S_n^{-1} \mathbf{v}_0 + R_{1n} \end{aligned}$$

where R_{1n} is a remainder term. Using the bound $|\log(1+x) - x| \leq x^2/[2(1-\eta)^2]$ for all $|x| \leq \eta < 1$, on the set $\{D_n + \Gamma_n^0 < 1/2\}$, we have

$$|R_{1n}| \leq 8 \left[\sum_{i=1}^n (\nu_i^0)^2 + n(\Gamma_n^0)^2 \right].$$

Note that

$$(5.10) \quad \begin{aligned} \sum_{i=1}^n (\nu_i^0)^2 &= \lambda_1^2 \mathbf{v}_0' S_n^{-1} \left(\sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i' \right) S_n^{-1} \mathbf{v}_0 \\ &= n\lambda_1^2 [\mathbf{v}_0' S_n^{-1} \mathbf{v}_0 + (\mathbf{v}_0' S_n^{-1} \bar{\mathbf{X}}_n)^2] \\ &= O_p(np^{-2} [a_n^2 + a_n^4]) \\ &= O_p(n^{-1} + p^{-1}). \end{aligned}$$

Hence, it follows that

$$(5.11) \quad |R_{1n}| = O_p(n^{-1} + p^{-1}).$$

By (5.8), (5.9) and (5.11), it now follows that

$$\begin{aligned} -2 \log R_n(\boldsymbol{\mu}_0) &= 2\lambda_1 n \mathbf{v}_0' S_n^{-1} \bar{\mathbf{X}}_n - n\lambda_1 \mathbf{v}_0' S_n^{-1} \mathbf{v}_0 \\ &\quad + O_p(n^{-1} + p^{-1}) \\ &= n\lambda_1 \bar{\mathbf{X}}_n' S_n^{-1} \bar{\mathbf{X}}_n + O_p(p^{-1}) + o_p(n^{-1/2}) \\ &= n\lambda_1 \sum_{j=1}^p \bar{X}_{jn}^2 + O_p(\|\bar{\mathbf{X}}_n\|^2 \cdot \|S_n^{-1} - \mathbb{I}_p\| p^{-1} n) \\ &\quad + O_p(p^{-1}) + o_p(n^{-1/2}). \end{aligned}$$

Next note that by Condition (C.3),

$$\|S_n^{-1} - \mathbb{I}_p\| \leq \|S_n^{-1}\| \cdot \|S_n - \mathbb{I}\| = o_p(p^{-1/2}).$$

Now using the fact that $E\bar{X}_{jn}^2 = n^{-1}$ for all $1 \leq j \leq p$, it now follows that

$$(5.12) \quad p^{1/2}[-2 \log R_n(\boldsymbol{\mu}_0) - p\lambda_1] = p^{1/2}\lambda_1 \sum_{j=1}^p [n\bar{X}_{jn}^2 - 1] + o_p(1).$$

Next by using the independence of the \mathbf{X}_i 's and that under the assumption that $\Sigma_p = \mathbb{I}_p$, one can show that for any $1 \leq j, k \leq p$,

$$\begin{aligned} & |\text{Cov}([\sqrt{n}\bar{X}_{jn}]^2, [\sqrt{n}\bar{X}_{kn}]^2) - 2[\text{Cov}(\sqrt{n}\bar{X}_{jn}, \sqrt{n}\bar{X}_{kn})]^2| \\ & \leq n^{-1}|\chi_4(X_{1j}, X_{1j}, X_{1k}, X_{1k})| \\ & \leq Cn^{-1} \end{aligned}$$

and

$$\text{Cov}(\sqrt{n}\bar{X}_{jn}, \sqrt{n}\bar{X}_{kn}) = \delta_{j,k}$$

for all $n \geq 1$ where $\delta_{j,k} = 1$ if $j = k$ and $= 0$ otherwise. Now using the strong mixing condition and a 'big-block little block' argument (cf. Athreya and Lahiri (2006), Ch. 16; Ibragimov and Linnik (1971)), one can show that

$$(5.13) \quad p^{-1/2} \sum_{j=1}^p [n\bar{X}_{jn}^2 - 1] \rightarrow^d N(0, \sigma_\infty^2)$$

where

$$\begin{aligned} \sigma_\infty^2 & \equiv \lim_{n \rightarrow \infty} p^{-1} \sum_{k=-(p-1)}^{p-1} (p - |k|) \text{Cov}([\sqrt{n}\bar{X}_{jn}]^2, [\sqrt{n}\bar{X}_{jn}]^2) \\ & = 2. \end{aligned}$$

The theorem now follows from (5.12) and (5.13). \square

ACKNOWLEDGEMENTS

The authors acknowledge gratefully the helpful comments of referees that led to the improvement of the presentation and the results of the paper.

Received 16 January 2012

REFERENCES

ATHREYA, K. B. and LAHIRI, S. N. (2006). *Measure Theory and Probability Theory*. Springer, New York, NY. [MR2247694](#)

BARTOLUCCI, F. (2007). A penalized version of the empirical likelihood ratio for the population mean. *Statistics & Probability Letters* **77** 104–110. [MR2339024](#)

BERAN, J. (1994). *Statistical Methods for Long Memory Processes*. Chapman & Hall, London.

BERTAİL, P. (2006). Empirical likelihood in some semi-parametric models. *Bernoulli* **12** 299–331. [MR2218557](#)

CAI, T., ZHANG, C-H. and ZHOU, H. (2010). Optimal rates of convergence for covariance matrix estimation. *Ann. Statist.* **38** 2118–2144. [MR2676885](#)

CHEN, J., VARIYATH, A. M. and ABRAHAM, B. (2008). Adjusted empirical likelihood and its properties. *Journal of Computational and Graphical Statistics* **17** 426–443. [MR2439967](#)

CHEN, S. X. and HALL, P. (1993). Smoothed empirical likelihood confidence intervals for quantiles. *Ann. Statist.* **21** 1166–1181. [MR1241263](#)

CHEN, S. X., PENG, L. and QIN, Y-L. (2009). Effects of data dimension on empirical likelihood. *Biometrika* **96** 711–722. [MR2538767](#)

DICICCIO, T., HALL, P. and ROMANO, J. (1991). Empirical likelihood is Bartlett-correctable. *Ann. Statist.* **19** 1053–1061. [MR1105861](#)

EMERSON, S. C. and OWEN, A. (2009). Calibration of the empirical likelihood method for a vector mean. *Electronic Journal of Statistics* **3** 1161–1192. [MR2566185](#)

FAN, J. and LI, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* **96** 1348–1360. [MR1946581](#)

FAN, J. and PENG, H. (2004). Nonconcave penalized likelihood with a diverging number of parameters. *Ann. Statist.* **32** 928–961. [MR2065194](#)

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

IBRAGIMOV, I. A. and LINNIK, Y. V. (1971). *Independent and stationary Sequences of Random Variables*. Wolters-Noordhoff Publishing, Groningen, The Netherlands. [MR0322926](#)

LAHIRI, S. N. and MUKHOPADHYAY (2012). A penalized empirical likelihood method in high dimensions. *Ann. Statist.* (To appear).

MILNOR, J. W. (1965). *Topology from the Differentiable Viewpoint*. The University Press of Virginia, Charlottesville, Va. [MR0226651](#)

OTSU, T. (2007). Penalized empirical likelihood estimation of semi-parametric models. *J. Mult. Analysis* **98** 1923–1954. [MR2396947](#)

OWEN, A. (1988). Empirical likelihood ratio confidence intervals for a single function. *Biometrika* **75** 237–249. [MR0946049](#)

OWEN, A. (2001). *Empirical Likelihood*. Chapman and Hall-CRC.

QIN, J. and LAWLESS, J. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.* **22** 300–325. [MR1272085](#)

TANG, C. Y. and LENG, C. (2010). Penalized high-dimensional empirical likelihood. *Biometrika* **97** 905–920. [MR2746160](#)

TSAO, M. (2004). Bounds on coverage probabilities of the empirical likelihood ratio confidence regions. *Ann. Statist.* **32** 1215–1221. [MR2065203](#)

S. N. Lahiri
Department of Statistics, Texas A&M University
College Station, TX 77843-3143
USA
E-mail address: snl@stat.tamu.edu

S. Mukhopadhyay
Department of Statistics, Texas A&M University
College Station, TX 77843-3143
USA
E-mail address: deep@stat.tamu.edu