The Annals of Statistics
2012, Vol. 40, No. 5, 2511–2540
DOI: 10.1214/12-AOS1040

A PENALIZED EMPIRICAL LIKELIHOOD METHOD IN HIGH DIMENSIONS[1]

BY SOUMENDRA N. LAHIRI AND SUBHADEEP MUKHOPADHYAY

Texas A&M University

This paper formulates a penalized empirical likelihood (PEL) method for inference on the population mean when the dimension of the observations may grow faster than the sample size. Asymptotic distributions of the PEL ratio statistic is derived under different component-wise dependence structures of the observations, namely, (i) non-Ergodic, (ii) long-range dependence and (iii) short-range dependence. It follows that the limit distribution of the proposed PEL ratio statistic can vary widely depending on the correlation structure, and it is typically different from the usual chi-squared limit of the empirical likelihood ratio statistic in the fixed and finite dimensional case. A unified subsampling based calibration is proposed, and its validity is established in all three cases, (i)–(iii). Finite sample properties of the method are investigated through a simulation study.

1. Introduction. In a seminal paper, Owen (1988) introduced the empirical likelihood (EL) method for statistical inference on population parameters in a nonparametric framework, and showed that it enjoyed properties similar to the likelihood-based inference methods in a more traditional parametric framework. Following Owen (1988), the EL method has been extended to various complex inference problems; see, for example, Diccicio, Hall and Romano (1991), Hall and Chen (1993), Qin and Lawless (1994), Owen (2001), Bertail (2006), Hjort, McKeague and Van Keilegom (2009), Chen, Peng and Qin (2009) and the references therein. An extension of the EL method in the high-dimensional context, where the dimension p of the observations increases with the sample size n, is given by Hjort, McKeague and Van Keilegom (2009). Hjort, McKeague and Van Keilegom (2009) derives the limit distribution of the EL ratio statistic based on p-dimensional estimating equations when $p \to \infty$ with n at the rate $p = o(n^{1/3})$. Chen, Peng and Qin (2009) improved upon the rate restriction in Hjort, McKeague and Van Keilegom (2009) and established a nondegenerate limit distribution of the EL ratio statistic, allowing $p = o(n^{1/2})$ under suitable regularity conditions.

For applications to high-dimensional problems, such as those involving gene expression data, one encounters a p that is typically much larger than the sample

Received September 2011; revised August 2012.

[1]Supported in part by NSF Grants DMS-07-07139 and DMS-10-07703.

MSC2010 subject classifications. Primary 62G09; secondary 62G20, 62E20.

Key words and phrases. Asymptotic distribution, long-range dependence, Rosenblatt process, Wiener–Itô integral, regularization, simultaneous tests, subsampling.

size n. However, extension of the EL to such high-dimensional problems is itself a daunting task because the (standard) EL method is known to fail in such situations. An important result of Tsao (2004) shows that the definition of *the EL for a p-dimensional population mean based on a sample size n breaks down on a set of positive probability whenever $p > n/2$*; further, this probability is asymptotically nonnegligible. The main reason for this surprising behavior of the EL is that for $p > n/2$, the convex hull of n random vectors in $\mathbb{R}^p$ is too small a set to contain the true mean with high probability. As a result, the standard EL approach cannot be applied to the "large p small n" problems with $p > n/2$. An alternative formulation of the EL in such situations (called the *adjusted EL*) is given by Chen, Variyath and Abraham (2008), which is further refined and studied by Emerson and Owen (2009). The adjusted EL method adds additional pseudo-observations [a single one in Chen, Variyath and Abraham (2008) and two in Emerson and Owen (2009)] so as to cover a hypothesized value of the mean parameter within the convex hull of the augmented data set, thereby making the adjusted EL well-defined. A second approach, due to Bartolucci (2007), is to drop the convex hull constraint in the formulation of the EL altogether and redefine the likelihood of a hypothesized value of the parameter by penalizing the unconstrained EL using the Mahalanobis distance. The penalized EL (PEL) of Bartolucci (2007) is well defined for all values of $p \le n$, as long as the sample covariance matrix is nonsingular. However, due to the use of the inverse of the sample covariance matrix in its formulation, the PEL of Bartolucci (2007) is also not well defined for $p > n$. Bartolucci (2007) establishes a chi-squared limit of the PEL for the population mean in the case where the dimension p is fixed and finite for all n. 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). Both these papers consider the high dimensional set up and establish validity of their methods still requiring p to grow at most as a fractional power of the sample size n. *In this paper, we introduce a modified version of the PEL method of* Bartolucci (2007) *that is computationally simpler and that is applicable to a large class of "large p small n" problems, allowing p to grow faster than n*. This is an important step in generalizing the EL in high dimensions beyond the $p \le n$ threshold where the standard EL and its existing variants fail.

To briefly describe the proposed methodology and the main results of the paper, suppose that $X_1, \ldots, X_n$ are independent and identically distributed (i.i.d.) $\mathbb{R}^p$-valued random vectors with mean $\mu \in \mathbb{R}^p$, $1 < p < \infty$. Denote the jth component of a p-vector x by x_j, $j = 1, \ldots, p$. The proposed PEL employs a multiplicative penalty term to penalize the likelihood of a hypothesized value μ of the population mean as a quadratic function of the distance between the sample mean and μ. However, unlike Bartolucci's (2007) method, the use of the inverse sample covariance matrix is completely avoided, as consistency of the sample covariance matrix in the high dimensional case for all the dependence structures that we consider in this paper is not guaranteed. The proposed PEL instead uses a *component-wise* *scaling* to bring up the varying degrees of variability (variances) along different components to a common level, and then it applies an overall penalty on the sum of the squared rescaled differences; see (2.1) in Section 2 below. As a result, *the proposed PEL is well-defined for all values of* $n, p \geq 1$. Further, this approach has the added advantage that it does not require inversion of a high-dimensional matrix, and therefore, it is computationally much simpler.

For investigations into the theoretical properties of the proposed PEL method, we allow the components of X_1 to be dependent. The range of dependence that we consider covers the cases of:

(i) *short-range dependence* (SRD), where roughly speaking, the average of the components of X_1 satisfies a central limit theorem (CLT) under suitable moment conditions; cf. Ibragimov and Linnik (1971);

(ii) *long-range dependence* (LRD), where under suitable regularity conditions, the average of the components satisfies noncentral limit theorems [Taqqu (1975, 1977), Dobrushin and Major (1979)];

(iii) *nonergodicity* (NE), where the dependence is so strong that the average of the components even fails to satisfy a (strong) law of large numbers.

We refer to the LRD and SRD cases collectively as the ergodic (E)-case, as the negative logarithm of the PEL ratio statistic K_n (say) here satisfies a law of large numbers without further centering and scaling, for *any* rate of growth of p; cf. Remark 4.2 below. However, such degenerate limits laws are not always the most useful in practice as these only lead to conservative large sample inference procedures. By using suitable centering and scaling, we are able to further refine these results and establish convergence in distribution to nondegenerate limits. Specifically, we show that under SRD, K_n with centering at 1 [for $c_* = 1$ in condition (C.2)(ii) below] and scaling by square-root of the dimension p of the observations converges to a Normal limit, very much like the results of Hjort, McKeague and Van Keilegom (2009) and Chen, Peng and Qin (2009), but allowing a much faster rate of growth of p and allowing a more general dependence framework. In the long range dependent (also abbreviated as LRD) case, K_n with a suitable normalization can have both Normal and non-Normal limits. For the Normal limit, the centering and the scaling sequences are the same as those used in the SRD case, except at the boundary layer of dependence where the Normal limit switches over to the non-Normal limit. For the non-Normal limit under LRD, the centering term is the same as that in the SRD case, but the scaling depends on the rate of decay of the auto-correlation coefficient of the components of X_1 (up to a possibly unknown permutation). Finally, in comparison to the E-case, K_n in the NE-case is shown to converge in distribution to a stochastic integral, and it does NOT require any further centering and scaling.

The growth rate of p, for which a *nondegenerate limit law* holds for a suitably transformed K_n, primarily depends on the strength of dependence among the components of the observations; cf. Figure 1. In the NE-case, p can grow *arbitrarily fast* (e.g., polynomial, exponential, super-exponential, etc.) as a function of n. In

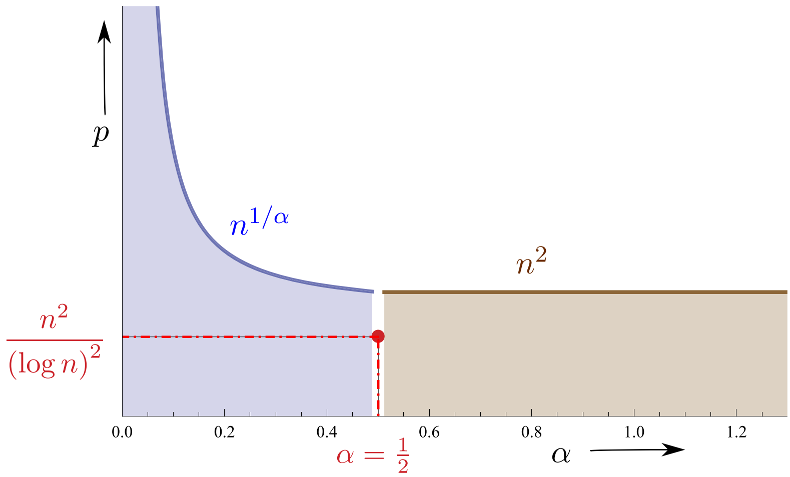

FIG. 1. *Envelope for the growth rates of p for a nondegenerate limit law of the PEL ratio statistic. The smaller the α, the stronger is the dependence.*

the E-case, although a degenerate limit law holds for an arbitrary growth rate of p, for a nondegenerate limit, p must admit a suitable upper bound. In particular, for the Normal limit in the E-case (excluding the boundary case), the growth rate of p as a function of n is $p = o(n^2)$. For the non-Normal limit under the E-case, $p = o(n^{1/\alpha})$ for $0 < \alpha < 1/2$ where, roughly speaking, α denotes the exponent of the rate of decay of the autocorrelation among the components of X_1, up to a permutation; cf. condition $(\text{C.4})_\alpha$, Section 3. The boundary case is given by $\alpha = 1/2$, where the growth rate is slightly smaller and is given by $p = o(n/[\log n]^2)$. From Figure 1, it follows that the stronger the dependence among the components of the observations, the higher is the allowable growth rate of p as a function of n for a nondegenerate limit. The limiting case $\alpha \to 0+$ is the NE-case. Here the nondegenerate limit for the the negative logarithm of the PEL ratio statistic holds for an arbitrary growth rate of p as a function of n.

It is worth pointing out that in most cases, the limit distribution of the PEL ratio statistic is *not* distribution free in the sense that the asymptotic approximation to the distribution of the PEL ratio statistic requires the knowledge of one or more unknown population parameters. As a result, the limit laws are not directly usable in practice. To address this issue, we propose a calibration procedure based on the subsampling method. We show that under mild conditions, the subsampling based calibration method is consistent under *all three* types of dependence structures.

The key step in the proofs is to derive a quadratic asymptotic approximation to K_n under all three cases of dependence. This is presented in Lemma 6.2 for the NE-case and in the proof of Theorem 3.2 for the E-case. The derivation of the limit law in the NE-case uses some weak convergence and operator convergence results on Hilbert spaces. On the other hand, in the E-case, refined approximations to K_n are required to go beyond their degenerate limits. See Section 6 for more details.

The rest of the paper is organized as follows. In Section 2, we describe the PEL methodology. In Section 3, we introduce the asymptotic framework and establish the limit distributions of the logarithm of the PEL ratio statistic under the three dependence scenarios. In Section 4, we describe the subsampling method and prove its validity for all three cases. We report the results from a moderately large simulation study in Section 5. Proofs of the results are given in Section 6.

2. Formulation of the PEL. Let $X_1, \dots, X_n$ be i.i.d. random vectors with mean $\mu \in \mathbb{R}^p$. Let X_{ij} denote the jth component of X_i, $1 \le i \le n$, $1 \le j \le p$. Also, let A' denote the transpose of a matrix A. We define the *penalized empirical likelihood (PEL)* of a plausible value $\mu = (\mu_1, \dots, \mu_p)'$ of the population mean as

$$L_n(\mu) = \sup_{(\pi_1,\dots,\pi_n)' \in \Pi_n} \left\{ \left(\prod_{i=1}^n \pi_i \right) \exp\left(-\lambda \sum_{j=1}^p \delta_j \left[\sum_{i=1}^n \pi_i (X_{ij} - \mu_j) \right]^2 \right) \right\}, \tag{2.1}$$

where $\Pi_n = \{(\pi_1, \dots, \pi_n)' \in [0,1]^n : \sum_{i=1}^n \pi_i = 1\}$, δ_j's are component specific weights (which may be random), and $\lambda = \lambda_n \in [0, \infty)$ is an overall penalty factor. Here we use

$$\delta_j \equiv \delta_{nj} = s_{nj}^{-2} \mathbb{1}(s_{nj} \neq 0), \tag{2.2}$$

where $s_{nj}^2 = n^{-1} \sum_{i=1}^n (X_{ij} - \bar{X}_{nj})^2$ is the sample variance of the jth components of $X_1, \dots, X_n$, $\bar{X}_{nj} = n^{-1} \sum_{i=1}^n X_{ij}$ and where $\mathbb{1}(\cdot)$ denotes the indicator function. This choice of the component-wise scaling allows us to adjust for the heteroscedasticity along different co-ordinates of X_1 and therefore, the overall penalization parameter λ gives *comparable* weights to *all* components. In addition, the choice of the penalty function makes the proposed PEL *invariant* with respect to component-wise scaling, which is an inherently desirable property, particularly while dealing with high-dimensional variables, where the assumption of homoscedasticity among a large number of components is unrealistic. The maximizer of the product $\prod_{i=1}^n \pi_i$ in (2.1) without the penalty term is given by $\pi_i = 1/n$ for $i = 1, \dots, n$. Hence, the *PEL ratio statistic* at a plausible value μ of the mean vector is defined as

$$R_n(\mu) = n^n L_n(\mu).$$

We now compare our formulation with the PEL of Bartolucci (2007), which is defined as

$$L_n^{\mathrm{B}}(\mu) = \sup_{(\pi_1,\dots,\pi_n)' \in \Pi_n} \left\{ \left(\prod_{i=1}^n \pi_i \right) \exp\left(-\frac{n(v - \mu)' V_n^{\dagger -1} (v - \mu)}{2h^2} \right) \right\}, \tag{2.3}$$

where $v = \sum_{i=1}^n \pi_i X_i$, $V_n^\dagger = n^{-1} \sum_{i=1}^n (X_i - \bar{X}_n)(\bar{X}_i - \bar{X}_n)'$ is the sample covariance matrix, and h is a penalty parameter. Note that for a large $p \in (1, n]$, the sample variance matrix V_n is ill-conditioned, if the smallest eigen-value of the underlying covariance matrix Σ (say) of X_1 is not bounded away from zero, and it is always singular when $p > n$, requiring further modifications to make (2.3) well defined. Under either of the two scenarios, the PEL based on (2.3) can be computationally demanding and unstable. In comparison, the component-wise scaling in (2.1) only involves 1-dimensional operations which is computationally much simpler and feasible even for a large p. A second limitation of (2.3) is the lack of attractive theoretical properties of V_n^{-1} (or of its variants) in high dimensions. Indeed, consistency of the sample covariance matrix (and its banded or tapered versions) in the high-dimensional setting is questionable in presence of strong correlations among the components of X_1 that we consider here. Existing work on consistency of the sample covariance matrix is known only under suitable conditions of sparsity or weak dependence; cf. Bickel and Levina (2008), El Karoui (2008), Cai, Zhang and Zhou (2010). Our formulation also avoids this problem altogether by using component-wise scaling.

For the sake of completeness, we also briefly describe the penalized EL approach of Otsu (2007) and Tang and Leng (2010), specialized to the case of the mean parameter μ for simplicity of exposition. Let $L_n^{\mathrm{ST}}(\mu) = \sup\{\prod_{i=1}^n \pi_i : (\pi_1, \dots, \pi_n) \in \Pi_n, \sum_{i=1}^n \pi_i(X_i - \mu) = 0\}$ denote the standard EL for μ. Also, let $p_\lambda(\cdot)$ be a penalty function, such as the smoothly clipped absolute deviation (SCAD) penalty function of Fan and Li (2001). Then, the penalized EL considered by Otsu (2007) and Tang and Leng (2010) is of the form

$$L_n^{\mathrm{OTL}}(\mu) = L_n^{\mathrm{ST}}(\mu) \exp\left(-n \sum_{j=1}^{p} p_\lambda(\mu_j)\right), \tag{2.4}$$

where $\mu = (\mu_1, \dots, \mu_p)'$. For the case of a more general parameter θ defined through a set of estimating equations, the formulation of Otsu (2007) and Tang and Leng (2010) replaces $L_n^{\mathrm{ST}}(\mu)$ in (2.4) by the corresponding version of the standard EL for θ; cf. Qin and Lawless (1994). As a result, irrespective of the target parameter, since (2.4) is directly based on the standard EL, this formulation of the penalized EL also suffers from the same limitations as the standard EL. In particular, this approach also fails in high dimensions whenever $p > n/2$.

In the next section, we investigate theoretical properties of the proposed PEL method (2.1) under the dependence structures described in Section 1.

3. Limit distributions.

3.1. *General framework.* We establish the limit distribution theory for the PEL ratio statistic in a triangular array set up, with n denoting the variable driving the asymptotics. Thus, the vectors $X_1, \dots, X_n$ depend on n as are their distributions and the dimension p. However, we often suppress the dependence on n for simplicity of notation. The limit distribution of the PEL ratio statistic depends on the degree of dependence among the components of $X_1, \ldots, X_n$. As stated in Section 1, we can broadly classify the dependence structure into two categories: (i) Non-Ergodic (NE) and (ii) Ergodic (E). In the NE-case, the dependence among the components of each X_i is so strong [cf. condition (C.3) below] that even the law of large numbers fails. In this case, we show that under appropriate conditions, the PEL ratio statistic has a nondegenerate limit. In contrast, in the E-case, the corresponding limit is degenerate, and further centering and scaling are needed for nondegenerate limit laws which, in turn, depend the type of dependence (SRD or LRD). We begin with the NE-case.

3.2. *Limit distribution for the nonergodic case.* We need to introduce some notation at this stage. Let $\rho_n(j,l)$ denote the correlation between X_{1j} and X_{1l}, $1 \le j, l \le p$. Let $\sigma_{nj}^2 = \mathrm{Var}(X_{1j})$ and $D_{nj}^X = \{x \in \mathbb{R} : P(X_{1j} = x) > 0\}$, $1 \le j \le p$, $n \ge 1$. Write C to denote a generic constant in $(0, \infty)$. Also, for any two sequences $\{a_n\}_{n\ge1}$ and $\{b_n\}_{n\ge1} \in (0,\infty)$, write $a_n \sim b_n$ if $a_n/b_n \to 1$ as $n \to \infty$. Let $L^2[0,1]$ denote the set of all square integrable functions on $[0,1]$ (with respect to the Lebesgue measure on $[0,1]$), equipped with the inner product $\langle f, g\rangle = \int_0^1 fg$, $f, g \in L^2[0,1]$. Let $\{\phi_k : k \in \mathbb{N}\}$ denote a complete orthonormal basis of $L^2[0,1]$, where $\mathbb{N} = \{1, 2, \ldots\}$ denotes the set of all positive integers. For any (bounded) jointly measurable function $h : [0,1]^2 \to \mathbb{R}$, define the operator Υ_h on $L^2[0,1]$ by $\Upsilon_h f = \int_{[0,1]} h(\cdot; u) f(u)\,du$ for $f \in L^2[0,1]$. For $x, y \in \mathbb{R}$, let $x \wedge y = \min\{x, y\}$.

We shall make use of the following conditions for deriving the limit distribution of $-\log R_n(\mu)$. The values of the integers r and s below will be specified later in the statements of the theorems.

CONDITIONS.

(C.1) (i) $\max\{E[\sigma_{nj}^2 \delta_j]^s : 1 \le j \le p\} = O(1)$ for a given $s \in \mathbb{N}$.
(ii) $\limsup_{n\to\infty} \max\{P(X_{nj} = x) : x \in D_{nj}^X, 1 \le j \le p\} < 1$.

(C.2) (i) For a given $r \in \mathbb{N}$, $\max\{E|X_{1j}|^r : 1 \le j \le p\} < C$.
(ii) $\lambda_n = c_* n/p$ for some $c_* \in (0, \infty)$.

(C.3) There exists a correlation function $\rho_0(\cdot,\cdot)$ of a mean-square continuous process on $[0,1]$, and for each $n \ge 1$, there exists a permutation ι_n of $\{1, \ldots, p_n\}$ such that $\rho_n(j,l) = \rho_0(\frac{\iota_n(j)}{p}, \frac{\iota_n(l)}{p})$. Further, with c_* as in (C.2), $\rho_0(\cdot,\cdot)$ satisfies the following:
(i) $4c_*^2 \int_0^1 \int_0^1 \rho_0^2(u,v)\,du\,dv < 1$;
(ii) $\sup\{|\rho_0(u+h_1, v+h_2) - \rho_0(u,v)| : |h_1| \le \delta, |h_2| \le \delta\} \le g(\delta) H(u,v)$ for all $u, v, u+h_1, v+h_2 \in [0,1]$ for some function $g(\cdot)$ satisfying $g(\delta) \to 0$ as $\delta \downarrow 0$ and for some function $H(\cdot,\cdot)$ satisfying $\sum_{k\ge1} \langle |\phi_k|, |\Upsilon_H \phi_k|\rangle < \infty$;
(iii) $\sum_{k\ge1} \langle \phi_k, \Upsilon_0 \phi_k\rangle$ converges, where $\Upsilon_0 = \Upsilon_h$ with $h = \rho_0$.

We now briefly comment on the conditions. Condition (C.1)(i) is a moment condition on the scaled component-wise weights δ_j's and requires finiteness of the sth negative moment of the sample variance s_{nj}^2's (scaled by the respective expected values σ_{nj}^2's). This condition holds in the case of Gaussian X_{ij}'s whenever the the sample size $n > 2s$. Condition (C.1)(ii) is a mild condition—it says that none of the X_{ij}'s take a single value with probability approaching one. This condition trivially holds if the components of X_1 are continuous and also in the discrete case, if the supports of X_{1j}'s contain at least two values with asymptotically nonvanishing probabilities. Condition (C.2)(i) is a moment condition that will be used with different values of r in the main theorems of this section, while (C.2)(ii) specifies the growth rate of the penalty parameter for a nondegenerate limit of $-\log R_n(\mu_0)$. However, unlike the standard usage of the penalty parameter in the context of variable selection, where different choices of the parameter lead to different sets of variables being chosen, here the key role of the penalty parameter is to stabilize the contribution from the sum of component-wise squared differences to the overall "likelihood" in (2.1). Finally, consider condition (C.3) that specifies the nonergodic structure of the X_i's. Note that, up to a (possibly unknown) permutation of the co-ordinates, the components of X_1 are essentially correlated as strongly as the variables $W(t)$'s coming from a constant mean, mean-square continuous process $\{W(t) : t \in [0, 1]\}$ (say) with covariance function $\rho_0(\cdot)$. In this case, the dependence among the variables $W(i/p)$ and $W(j/p)$, $1 \le i < j \le p$ is so strong that the average $p^{-1}\sum_{j=1}^{p} W(j/p)$ may not converge to a constant as $p \to \infty$, as one would expect from the well-known ergodic theorems.

Under conditions (C.1)–(C.3), the limit distribution of the log-PEL ratio statistic is given by a stochastic integral, as shown by the following result.

THEOREM 3.1. *Let conditions* (C.1), (C.2) *and* (C.3) *hold with* $s = 4$ *and* $r = 8$, *let* μ_0 *denote the true value of* μ *and let* $p \to \infty$ *as* $n \to \infty$. *Then*

$$-\log R_n(\mu_0) \to^d c_* \int_0^1 \int_0^1 \Lambda_0(u, v) Z(u) Z(v)\, du\, dv, \tag{3.1}$$

where $Z(\cdot)$ *is a zero mean Gaussian process on* $[0, 1]$ *with covariance function* $\rho_0(\cdot)$ *and where the function* $\Lambda_0(\cdot, \cdot)$ *is defined as*

$$\Lambda_0(u, v) = \sum_{k=0}^{\infty} (-2)^k c_*^k \rho_0^{*(k)}(u, v), \qquad 0 < u, v < 1, \tag{3.2}$$

with $\rho_0^{*(0)}(u, v) = 1$, $\rho_0^{*(1)}(u, v) = \rho_0(u, v)$ *and for* $k \ge 1$,

$$\rho_0^{*(k+1)}(u, v) = \int_0^1 \cdots \int_0^1 \left\{ \prod_{j=1}^{k-1} \rho_0(u_j, u_{j+1}) \right\} \rho_0(u_1, u) \rho_0(u_k, v)\, du_1 \cdots du_k.$$

For a general definition of a stochastic integral of the form (3.1), see Cramer and Leadbetter (1967). Note that under condition (C.3), by repeated application of the Cauchy–Schwarz inequality, for $k \geq 1$,

$$\sup_{u,v\in[0,1]} \left|\rho_0^{*(k+1)}(u,v)\right| \leq (1-\delta)^{k-1}[2c_*]^{-(k+1)} \qquad \text{for some } \delta \in (0,1),$$

and hence, the limiting stochastic integral is well defined.

Theorem 3.1 shows that under a suitable choice of the penalty parameter, namely, $\lambda = c_* n/p$, the negative log PEL ratio statistic has a nondegenerate limit distribution. Note that, unlike the standard version of the EL, we do not use the multiple 2 before $-\log R_n(\mu_0)$. This is a direct artifact of the additional penalty term that we use in the formulation of PEL. Also, unlike most high-dimensional problems where the validity of a large sample inference procedure breaks down beyond a certain (often exponential) rate of growth of p, the PEL and the associated limit distribution of $-\log R_n(\mu_0)$ in the NE-case remains valid for *arbitrary* rate of growth of p as a function of the sample size. Thus, in the NE-case, it is possible to carry out simultaneous hypothesis testing for a very large number of parameters even with a moderately large sample.

3.3. *Limit distribution for the ergodic case.* In the E-case, we shall make use of the following conditions, for $\alpha \in (0,\infty)$:

$(\mathrm{C.4})_\alpha$ There exists a covariance function $\rho_\alpha(\cdot)$ on $\mathbb{Z} \equiv \{0, \pm 1, \pm 2, \ldots\}$, and for each $n \geq 1$, there exists a (possibly unknown) permutation ι_n of $\{1,\ldots,p_n\}$ such that $\rho_\alpha(k) \sim C|k|^{-\alpha}$ as $k \to \infty$ and

$$\sup_{1\leq j,l\leq p} \left| \frac{\rho_n(j,l)}{\rho_\alpha(\iota_n(j)-\iota_n(l))} - 1 \right| = o(1) \qquad \text{as } n \to \infty.$$

$(\mathrm{C.5})_\alpha$ There exists a constant $C > 0$ such that

$$\check{\varrho}_n(k) \leq Ck^{-\alpha} \qquad \text{for all } k \geq 1, n > C,$$

where $\check{\varrho}_n(\cdot)$ denotes the ϱ-mixing coefficient of the variables $\{\tilde{X}_{1j} : 1 \leq j \leq p\}$, defined by $\check{\varrho}_n(k) = \sup\{|P(A\cap B) - P(A)P(B)|/\sqrt{P(A)P(B)} : A \in \mathcal{F}_1^m, B \in \mathcal{F}_{m+k}^p, 1 \leq m \leq p-k\}$. Here, $\mathcal{F}_a^b$ denotes the σ-field generated by $\{\tilde{X}_{1j} : a \leq j \leq b\}$, $1 \leq a \leq b \leq p$, $\tilde{X}_{1j} = X_{1\tau(j)}$, and $\tau = \tau_n$ is the inverse of the permutation ι_n in $(\mathrm{C.4})_\alpha$.

Condition $(\mathrm{C.4})_\alpha$ says that up to a (possibly unknown) permutation of the coordinates, the components of the X_i-vectors have a dependence structure that is asymptotically similar to the one given by $\rho_\alpha(\cdot)$. Note that the sum $\sum_{k=0}^{\infty} |\rho_\alpha(k)|$ diverges if and only if $\alpha \leq 1$, and therefore, we classify the dependence structure of the X_{1j}'s as LRD or SRD according to $\alpha \leq 1$ or $\alpha > 1$, respectively; cf. Beran (1994). Condition $(\mathrm{C.5})_\alpha$ is a decay condition on the ϱ-mixing coefficient of the reordered variables $\{\tilde{X}_{1j} : 1 \leq j \leq p\}$. Note that by $(\mathrm{C}.4)_\alpha$, the correlation coefficient between $\tilde{X}_{1j}$ and $\tilde{X}_{1k}$ is $\rho_\alpha(j-k)(1+o(1))$, and therefore, the reordered sequence $\{\tilde{X}_{1j} : 1 \leq j \leq p\}$ behaves approximately like a stationary time series with the natural time-index j. Thus, condition $(\mathrm{C}.5)_\alpha$ specifies the degree of dependence of the X_{ij}'s, up to a permutation that need not be known to the user.

3.3.1. *Results under short-range dependence*. The following result shows that the log-PEL ratio statistic in the SRD case converges to a Normal limit after a suitable centering and after scaling by the "standard factor" $p^{1/2}$.

THEOREM 3.2. *Let conditions* (C.1), (C.2) *and* $(\mathrm{C}.4)_\alpha$, $(\mathrm{C}.5)_\alpha$ *hold for some* $\alpha > 1$, $s \geq 6$, $r \geq 12$. *Let* $\kappa^2 = 2c_*^2 \sum_{k=0}^{\infty} \rho_0(k)^2$. *Then, for* $p = o(n^2)$,

$$p^{1/2}[-\log R_n(\mu_0) - c_*] \to^d N(0, \kappa^2). \tag{3.3}$$

We now comment on Theorem 3.2. From the proof, it follows that the distribution of the log-PEL ratio statistic, for a given sample size, is close to the sum of p weakly dependent chi-squared random variables with one degree of freedom. As a result, centering at 1 and scaling by $p^{1/2}$ yields a nondegenerate Normal limit. The effect of the weak dependence shows up in the variance of the limiting Normal distribution, which depends on the correlation structure of the components of X_1. It is worth noting that in the SRD case, one can use Normal critical points with an estimated variance to calibrate simultaneous tests of p hypotheses using the EL.

Theorem 3.2 extends existing results on the EL in more than one direction. Hjort, McKeague and Van Keilegom (2009) and Chen et al. (2009) proved a version of the result (i.e., a Normal limit) for the standard log-EL ratio statistic in increasing dimensions with centering at 1 and scaling by $p^{1/2}$. In comparison, Theorem 3.2 relaxes the restriction on the dimension p of the parameter μ, by allowing it to grow faster than the sample size. This should be compared with the best available rate of $p = o(n^{1/2})$, obtained by Chen et al. (2009). Further, Theorem 3.2 covers a wide range of dependence structures of the components of X_{1j}'s which are not covered by the earlier results (e.g., here the minimum eigen-value of the covariance matrix of X_1 need not be bounded away from zero). However, the most important implication of Theorem 3.2 is that under SRD, the penalization step circumvents the limitation of the standard EL which is known to break down beyond the threshold $p \leq n/2$, as shown by Tsao (2004).

3.3.2. *Results under long-range dependence*. For $\alpha \in (0, 1]$, the sum $\sum_{k=1}^{\infty} \rho_0(k)$ fails to converge absolutely, and we refer to this as the LRD case. Sums of LRD random variables are known to have either a Normal or a non-Normal limit, depending on the value of α. The next result deals with the case where α can be very small, and the limit law is non-Normal. Further, the scaling also depends on the correlation decay parameter α, as shown by Theorem 3.3. Let $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t}\, dt$ and $\iota = \sqrt{-1}$.

THEOREM 3.3. *Let conditions* (C.1), (C.2) *and* $(C.4)_\alpha$, $(C.5)_\alpha$ *hold for some* $\alpha \in (0, \frac{1}{2})$, $s \geq 6$ *and* $r \geq 12$. *If* $p = o(n^{1/\alpha})$, *then*

$$p_n^\alpha[-\log R_n(\mu_0) - c_*] \to^d W, \tag{3.4}$$

where W *is defined in terms of a bivariate Wiener–Itô integral with respect to the random spectral measure* ζ *of the Gaussian white noise process as*

$$W = c_*[2\Gamma(\alpha)]^{-1} \int \frac{\exp(\iota[x_1 + x_2]) - 1}{\iota[x_1 + x_2]} |x_1 x_2|^{(\alpha-1)/2} \, d\zeta(x_1) \, d\zeta(x_2).$$

Theorem 3.3 shows that under very strong dependence (i.e., for small values of α) in the E-case, the log-PEL ratio statistic, with the same centering but a different scaling factor, has a nondegenerate limit distribution and the limit law is *non-Normal*. Further, the range p for which the result holds is $p = o(n^{1/\alpha})$, which is a decreasing function of α. Thus, the stronger the dependence among the coordinates of X_{1j}'s, the larger is the allowable growth rate of p as a function of n for the validity of the limit distribution.

Theorems 3.2 and 3.3 exhaust the types of limit laws for the log-PEL ratio statistic in the E-case. However, in terms of the rate of decay of the correlation function, these leave out the case where $\alpha \in [1/2, 1]$. Although $\alpha \in [1/2, 1]$ corresponds to LRD in the traditional sense, the centered and scaled versions of the log-PEL ratio statistic continue to have a Normal limit as shown by the following result. Curiously, the scaling sequence as well as the growth rate of p depend on whether $\alpha = 1/2$ or $\alpha \in (1/2, 1]$.

THEOREM 3.4. *Let conditions* (C.1), (C.2) *and* $(C.4)_\alpha$, $(C.5)_\alpha$ *hold for some* $1/2 \leq \alpha \leq 1$, $s \geq 6$, $r \geq 12$.

(i) *If* $1/2 < \alpha \leq 1$ *and* $p = o(n^2)$, *then* (3.3) *holds.*
(ii) *If* $\alpha = 1/2$ *and* $p = o([n/\log n]^2)$, *then*

$$[p \log p]^{1/2}[-\log R_n(\mu_0) - c_*] \to^d N(0, c_*^2). \tag{3.5}$$

Thus, it follows from Theorem 3.4 that the log-PEL ratio statistic is asymptotically Normal for all $\alpha \geq 1/2$, although the components of X_i's have LRD when $\alpha \in [1/2, 1]$. The peculiar behavior of the scaling sequence at the boundary value $\alpha = 1/2$ is essentially determined by the growth rate of the series $\sum_{j=1}^p \rho_\alpha^2(j)$ as $p \to \infty$, which is asymptotically equivalent to $\log p$ for $\alpha = 1/2$ but it is bounded for $\alpha > 1/2$.

REMARK 3.1. Proofs of Theorems 3.2–3.4 show that for any $p \to \infty$, $-\log R_n(\mu_0) \to_p c_*$ as $n \to \infty$, that is, the log-PEL ratio statistic has a degenerate limit under *all* sub-cases of the E-case, for *arbitrarily large* p as a function of n. However, for nondegenerate limits, refined approximations to the difference

$[-\log R_n(\mu_0) - c_*]$ are needed. Here, we are able to show that an approximation of the form

$$-\log R_n(\mu_0) - c_* = T_n + E_n \tag{3.6}$$

holds for *all* sub-cases of the E-case, where T_n is a centered sum and where E_n is an error term, roughly of the order of $O_p(n^{-1})$. Further, T_n has a nondegenerate limit up to a suitable scaling, as a function of p, depending on the dependence structure of the X_{1j}'s. The bounds on the growth rate of p in the different sub-cases of the E-case are then determined by the requirement that the scaled error term be asymptotically negligible. For example, in the SRD case, $p^{1/2}T_n \to^d N(0, \kappa^2)$ and hence, $p^{1/2}E_n \to_p 0$ if and only if $p^{1/2}/n \to 0$, which is equivalent to the bound $p = o(n^2)$. Similar considerations lead to the respective upper bounds in the other sub-cases of the E-case.

REMARK 3.2. It is worth pointing out that the PEL can be used for constructing "conservative" large sample simultaneous tests of the p hypotheses $H_0 : \mu = \mu_0$ for *arbitrarily* large p in the E-case. Indeed, for p growing faster than the upper bounds given in Theorems 3.2–3.4, a conservative large sample simultaneous test of $H_0 : \mu = \mu_0$ rejects H_0 if $|c_* + \log R(\mu_0)| > n^{-1}\log n$. Note that by (3.6), this test attains the *ideal* level 0 asymptotically.

4. A subsampling based calibration. In this section, we describe a nonparametric calibration method based on subsampling to approximate the quantiles of the nondegenerate limit laws in both E- and NE-cases, which typically involve unknown population parameters and hence, cannot be used directly in practice. Let $\mathcal{X}_n(I) = \{X_i : i \in I\}$ be a subset of $\{X_1, \ldots, X_n\}$ where $I \subset \{1, \ldots, n\}$ is of size m and where $1 < m < n$ (specific conditions on m are given below). On each $\mathcal{X}_n(I)$, we employ the PEL method and obtain a version of the PEL ratio statistic $R^*_m(\mu; I)$, by replacing n with m and $X_1, \ldots, X_n$ by $\mathcal{X}_n(I)$ in the definitions $R_n(\mu)$. First consider the NE case. Here, the subsampling estimator of the distribution function $G_n^{\mathrm{NE}}(\cdot) \equiv P(-\log R_n(\mu_0) \le \cdot)$ under the null hypothesis $H_0 : \mu = \mu_0$ is given by

$$\hat{G}_n^{\mathrm{NE}}(x) = |\mathcal{I}_n|^{-1} \sum_{I \in \mathcal{I}_n} \mathbb{1}\bigl(-\log R^*_m(\mu_0; I) \le x\bigr), \qquad x \in \mathbb{R},$$

where $\mathcal{I}_n$ is a collection of subsets of $\{1, \ldots, n\}$ of size m and where $|A|$ denotes the size of a set A. All possible subsets of size m cannot be used mainly due to the sheer number of such sets, and hence, only a small fraction of these subsets are used to compute $\hat{G}_n^{\mathrm{NE}}(\cdot)$ in practice. In view of the block resampling methods for time series data, here we shall take $\mathcal{I}_n$ to be the collection of all overlapping blocks (subsets) of size m contained in $\{1, \ldots, n\}$. Then, we have the following result in the NE case.

THEOREM 4.1. *Suppose that the conditions of Theorem 3.1 hold and that*

$$m^{-1} + m/n = o(1) \qquad \text{as } n \to \infty. \tag{4.1}$$

Then, $\sup_{x\in\mathbb{R}} |\hat{G}_n^{\mathrm{NE}}(x) - G_n^{\mathrm{NE}}(x)| \to_p 0$ *as* $n \to \infty$.

Next consider the E-case. Note that for $\alpha = 1/2$, the limit of the log-PEL ratio statistic is $N(0, 1)$, which is distribution free. One can carry out a simple test [cf. Beran (1994)] to ascertain if "$H : \alpha = 1/2$" is true and then use the limit distribution directly to conduct the PEL test of the simultaneous p hypotheses $H_0 : \mu = \mu_0$ using the $N(0, 1)$ critical points, without the need for an alternative calibration. As a result, we concentrate on the values of $\alpha \neq 1/2$ in the E-case. Let $\hat{\alpha}_n$ be an estimator of the correlation parameter α; cf. Remark 4.1 below. Let $R_m^*(\mu_0; I)$ denote the PEL ratio statistic based on the subsample $\mathcal{X}(I)$ under μ_0, and define

$$V_m^*(I) = \hat{b}_n[-\log R_m^*(\mu_0; I) - c_*], \qquad I \in \mathcal{I}_n,$$

where $\hat{b}_n = p^{\hat{\alpha}_n \wedge 1/2}$. Then, a subsampling estimator of the distribution of $V_n \equiv \hat{b}_n[-\log R_n(\mu_0) - c_*]$ is given by $\hat{G}_{n,\alpha}(x) = |\mathcal{I}_n|^{-1} \sum_{I\in\mathcal{I}_n} \mathbb{1}(-\log V_m^*(I) \leq x)$, $x \in \mathbb{R}$ and we have the following results.

THEOREM 4.2. *Suppose that there exists a* $c_0 \in \mathbb{R}$ *such that*

$$(\log p)[\hat{\alpha} - \alpha] \to_p c_0 \qquad \text{as } n \to \infty. \tag{4.2}$$

(i) *For* $\alpha \in (1, \infty)$, *let the conditions of Theorem 3.2 and for* $\alpha \in (1/2, 1]$, *let those of Theorem 3.4*(i) *hold. If* $p/m^2 + m/n = o(1)$, *then*

$$\sup_{x\in\mathbb{R}} \left|\hat{G}_{n,\alpha}(x) - P(V_n \leq x)\right| \to_p 0 \qquad \text{as } n \to \infty. \tag{4.3}$$

(ii) *If the conditions of Theorem 3.3 hold for some* $\alpha \in (0, 1/2)$ *and* $p^\alpha/m + m/n = o(1)$, *then* (4.3) *holds.*

Theorem 4.2 shows that for both Normal and non-Normal limit laws under the E-case, the subsampling method provides a valid approximation to the distribution of the log-PEL ratio statistic. Hence, one can use the quantiles of the subsampling estimators to calibrate simultaneous tests on μ in a unified manner. This is specially important in the case of non-Gaussian limit laws for which the quantiles are difficult to derive. However, for $\alpha > 1/2$, the limit distribution is Gaussian, and an alternative approximation can be generated by using a Normal distribution with an *estimated* variance. Indeed, the latter may be preferable to subsampling from the computational point of view.

REMARK 4.1. In practice, the value of α is not known and must be estimated. First consider the case where the permutation $\iota_n(\cdot)$ in $(\mathrm{C}.4)_\alpha$ is known. Then, we are essentially dealing with n i.i.d. copies of a time series of length p as observations. By using the n replicates of the time series, it is easy to modify standard estimators of α based on a single time series [cf. Beran (1994)] to construct an estimator $\hat{\alpha}$ of α satisfying $\hat{\alpha} - \alpha = o_p(n^{-1/2}[\log p]^{-1})$ as $n \to \infty$, which clearly satisfies (4.2) with $c_0 = 0$.

Next consider the case where the permutations $\iota_n(\cdot)$ are unknown. In this case, it is *not* possible to identify pairs (j, l), $1 \le j, l \le p$ that correspond to the lag-k correlation $\rho_\alpha(k)$. However, it is still possible to construct estimators of α that satisfy (4.2). Define

$$\hat{\alpha} = -(\log p)^{-1} \log\left(e_n + n^{-1} \sum_{i=1}^{n} \left\{ p^{-1} \sum_{j=1}^{p} [X_{ij} - \bar{X}_{nj}] \delta_j^{1/2} \right\}^2 \right), \tag{4.4}$$

where $e_n = \prod_{j=1}^{p} \mathbb{1}(s_{nj} = 0)$ and δ_j (and $\bar{X}_{nj}$ and s_{nj}) are as in (2.2). Note that $\hat{\alpha}$ is invariant under permutations of the components of X_i's and also under componentwise location and scale transformations. In Section 6, we show that $\hat{\alpha}$ satisfies (4.2) under the conditions of Theorem 4.2, even when $\iota_n(\cdot)$ is unknown. In the same spirit, we may use the following estimator of the limiting variance κ^2 in Theorem 3.2 in the case where $\iota_n(\cdot)$ is unknown:

$$\hat{\kappa}^2 = 2c_*^2 \sum_{j=2}^{p-1} \hat{c}(1, j)^2 \mathbb{1}(|\hat{c}(1, j)| > 2n^{-1/2} \log n), \tag{4.5}$$

where $\hat{c}(j, k) = n^{-1} \sum_{i=1}^{n} (X_{ij} - \bar{X}_{nj})(X_{kn} - \bar{X}_{nk}) \delta_j \delta_k$, $1 \le j, k \le p$. Consistency of $\hat{\kappa}^2$ holds under mild moment conditions; cf. Section 6.

REMARK 4.2. An important factor that impacts the accuracy of the subsampling method is the choice of the subsample size m. Note that

$$m = C[np]^{\alpha_0/(1+\alpha_0)} \tag{4.6}$$

satisfies the requirements of Theorem 4.2, where $\alpha_0 = \min\{\alpha, 1/2\}$. At this point, we do not know the order of the optimal m for the different cases considered here. In the next section, we address this through a numerical study and explore the effects of different choices of m on the performance of the PEL method.

5. Numerical study.

We assess finite sample performance of the PEL method by simulation in a variety of settings. We considered different combinations of the sample size n and the dimension p, with $p \approx 2n^{1/3}, n/2, 2n$ and $n = 40$ and 200. The testing problem we considered is $H_0 : \mu = 0$, although any other value of μ may be used in H_0, as the PEL criterion is location invariant. We generated i.i.d. random p-vectors $X_1, \dots, X_n$ where the p coordinates of X_i's had one of the three different types of dependence structures, namely: (i) non-Ergodic, (ii) LRD and (iii) SRD, as follows.

5.1. *Algorithms for generating the data.*

5.1.1. *The nonergodic case.*

(1) Consider the basis functions in $L^2[0,1]$ given by $\phi_j(t) = \sin(2\pi jt)/\sqrt{2}$ for $j = 1, 2, \dots, 15$ and $\phi_j(t) = \cos(2\pi(j-15)t)/\sqrt{2}$, for $j = 16, 2, \dots, 30$ and $\phi_0(t) \equiv 1$.

(2) Generate $Z_j \sim \mathrm{N}(0,1)$ i.i.d. and let $\lambda_j = (\exp\{1\} + 1)$ for all j.

(3) Define $X_{1j} = \sigma_j \cdot W(j/p)$, $j = 1, 2, \dots, p$, where $W(\cdot) = \sum_{j=0}^{30} Z_j \phi_j(\cdot)\lambda_j$ and where $\sigma_j \equiv \sigma_{jn}$ are scalars in $(0, \infty)$.

Then, $X_1 = \{X_{11}, X_{12}, \dots, X_{1p}\}$ is a nonergodic series. Replicates of X_1 yield $X_1, \dots, X_n$ in the NE-case.

5.1.2. *Long-range dependence.* For the LRD case, we follow a setup similar to that used in Hall, Jing and Lahiri (1998). We generate stationary increments of a self-similar process with self-similarity parameter (or Hurst constant) $H = \frac{1}{2}(2-\alpha) \in (1/2, 1)$ for $\alpha \in (0,1)$. The algorithm is as follows:

(1) Generate a random sample $\mathbf{Z}_{p0} = \{Z_{10}, \dots, Z_{p0}\}$ from $N(0,1)$.

(2) Define $\mathbf{Z}_p \equiv U^T \mathbf{Z}_{p0}$, where U is obtained by Cholesky factorization of R into $R = U^T U$ and where $R = ((r_{ij}))$ with $r_{ij} = \rho_\alpha(|i-j|)$, and

$$\rho_\alpha(k) = \tfrac{1}{2}\{(k+1)^{2H} + (k-1)^{2H} - 2k^{2H}\}, \qquad k \geq 1, \tag{5.1}$$

and $\rho_\alpha(0) = 1$. Note that $\rho_\alpha(k) \sim Ck^{-\alpha}$ask $\to \infty$.

Replicates of $\mathbf{Z}_p$ give the variables $X_1, \dots, X_n$ in the LRD case.

For the simulation study here, we considered the NE-case ($\alpha = 0$) where the data were generated by the algorithm in Section 5.1.1 and the LRD cases $\alpha = 0.1$ and 0.8 based on the algorithm of Section 5.1.2. For the SRD case ($\alpha = \infty$), X_1 was generated by an ARMA(2, 3) process with $N(0,1)$ error variables and parameter vector $(-0.4, 0.1; 0.3, 0.5, 0.1)$.

5.2. *Choice of the subsample size.* We also considered different choices of the subsample size m in order to get some insight into its effects on the accuracy of the subsampling calibration. Note that the feasible choices of the subsample size depend on the relative growth rates of both n and p as well as on the strength of dependence, here quantified by α. For each pair (p, n), we considered three choices of the subsample size m (denoted by the generic symbols m_1, m_2, m_3), depending on the dependence structure. Specifically, for the SRD case ($\alpha = \infty$), we set

$$m_i = c_i^0 \cdot [np]^{1/3}, \qquad i = 1, 2, 3,$$

where $c_1^0 = 0.5$, $c_2^0 = 1$ and $c_3^0 = 2$. Note that in this case, the random variables in $X_1, \dots, X_n$ form a series of length np and are weakly dependent. Further, the target parameter for the subsampling method in the SRD case (and also in the LRD case with $\alpha > 1/2$) is the variance of the limiting Normal distribution. Hence, in view of the well-known results on optimal block length (for variance estimation) [cf. Hall, Horowitz and Jing (1995), Lahiri (2003)], the above choices of the m_i's are reasonable.

Next consider the case $\alpha = 0.1$ under LRD, where the limit distribution is non-Normal. From the proofs of Theorems 3.2 and 4.2, it follows that the prescription for m in (4.6) attempts to balance the bias of the subsampling approximation to the limit distribution and its variance. However, for a very small value of α, a direct application of (4.6) leads to a very small fractional exponent of np, which may be too small in practice. In such situations, particularly where p is not very large and the LRD exponent α is small, we use the threshold $n^{1/3}$ and set

$$m_i \equiv m_i(\alpha) = c_i^0 \cdot \max\{[np]^{\alpha/(1+\alpha)}, n^{1/3}\}, \qquad i = 1, 2, 3,$$

where c_i^0's are as before. The rationale behind this modification is that for p small, we simply treat $X_1, \ldots, X_n$ as a weakly dependent multivariate time series and again employ the known results on the optimal block size.

Finally, consider the NE-case, $\alpha = 0$. Note that for $\alpha = 0$, p can grow at an arbitrary rate with the sample size n for the validity of Theorems 3.1 and 4.1. Hence, in this case, our choice of m depends only on the sample size. We consider the "canonical" choice $m_1 = n^{1/3}$ as well as the larger values $m_2 = n^{1/2}$ and $m_3 = 2n^{1/2}$ to explore the effects of a larger subsample size on the accuracy of the subsampling calibration method.

5.3. *Results.*

5.3.1. *Levels of significance in simultaneous tests.* Here we consider finite sample accuracy of the proposed PEL method for simultaneous testing of the p hypotheses

$$H_0 : \mu = 0 \quad \text{vs.} \quad \mu \neq 0 \tag{5.2}$$

at the levels of significance $a = 0.1, 0.05$. The correlation parameter α for the subsampling based calibration was estimated by averaging the Taqqu, Teverovsky and Willinger (1995) estimator of the Hurst parameter (H) from each of the p-time series and by setting $\hat{\alpha} = (2 - 2\hat{H})$. Further, we have used the interior-point method [cf. Wright (1997)] as a fast optimization tool for computing the PEL ratio statistic, which can handle high-dimensional optimization problems efficiently. Tables 1 and 2 report the attained levels of significance based on 500 simulation runs and $n = 200$ for the target significance levels of 0.05 and 0.10, respectively, for different values of p, m, and α.

From the tables, it follows that the PEL does a reasonable job of simultaneous testing of p hypotheses for all 4 cases of dependence, for appropriately chosen subsample size. Comparing the attained level of significance, it is clear that the best

TABLE 1
Empirical levels of significance $\hat{a}$ for the subsampling based PEL, with sample size $n = 200$ at 0.05 significance level and $c_* = 1$. Here we have reported $|0.05 - \hat{a}|$

	$p_1 = 20$			$p_2 = 100$			$p_3 = 400$		
α	m_1	m_2	m_3	m_1	m_2	m_3	m_1	m_2	m_3
0.0	0.0092	0.0063	0.0190	0.0090	0.0114	0.0162	0.0089	0.0103	0.0091
0.1	0.0099	0.0081	0.0171	0.0045	0.0069	0.0129	0.0149	0.0134	0.0094
0.8	0.0079	0.0031	0.0061	0.0086	0.0042	0.0081	0.0101	0.0099	0.0190
∞	0.0059	0.0091	0.0010	0.0020	0.0104	0.0039	0.0091	0.0159	0.0078

choice of the subsample size critically depends on the relative sizes of n and p, and more importantly, on the type of dependence among the components of X_1. Further, rather surprisingly, the PEL tests at the level of significance 0.05 turned out to be more accurate (on an absolute scale) than at the level 0.1, for the subsample sizes considered here.

We also considered the effect of the penalty parameter $\lambda_n = c_* n/p$ on the performance of the PEL test. In the supplementary material Lahiri and Mukhopadhyay (2012) (hereafter referred to as [LM]), we report the empirical levels of significance of the PEL test for $n = 200$ and the target level 0.1 for two other choices of the constant c_*, namely, $c_* = 0.5$ and $c_* = 2.0$. The results for the choice $c_* = 2$ are qualitatively similar to those reported in Table 2 (with $c_* = 1$); in comparison, the accuracy for the case $c_* = 0.5$ appears to be slightly better than the $c_* = 1$ case. A similar pattern was observed for the 0.05 level of significance. We also considered the accuracy of the empirical significance levels of the PEL tests at a relatively smaller sample size $n = 40$, for $c_* = 1$ and $\alpha = 0.1$; cf. [LM]. The PEL has a reasonable performance even at this low sample size; see [LM] for details.

5.3.2. *Finite sample power properties.* To get some idea about the power properties of the PEL tests, we computed the probability of Type II error for a

TABLE 2
Empirical levels of significance $\hat{a}$ for the subsampling based PEL, with sample size $n = 200$ at 0.1 significance level and $c_* = 1$. Here we have reported $|0.1 - \hat{a}|$

	$p_1 = 20$			$p_2 = 100$			$p_3 = 400$		
α	m_1	m_2	m_3	m_1	m_2	m_3	m_1	m_2	m_3
0.0	0.170	0.020	0.075	0.221	0.142	0.090	0.075	0.152	0.227
0.1	0.030	0.033	0.033	0.012	0.005	0.011	0.112	0.133	0.066
0.8	0.011	0.045	0.087	0.123	0.082	0.018	0.108	0.027	0.069
∞	0.138	0.135	0.065	0.050	0.011	0.003	0.048	0.054	0.026

TABLE 3
Power of the proposed PEL, with sample size $n = 200$ at 0.1 significance level and $c_* = 1$

	$p_1 = 20$			$p_2 = 100$			$p_3 = 400$		
α	m_1	m_2	m_3	m_1	m_2	m_3	m_1	m_2	m_3
0.0	0.569	0.681	0.929	0.515	0.643	0.791	0.87	0.834	0.766
0.1	0.66	0.71	0.85	0.569	0.903	0.676	0.794	0.8	0.868
0.8	0.515	0.883	0.688	0.75	0.997	0.488	0.70	0.87	0.90
∞	0.510	0.622	0.870	0.739	0.778	0.996	0.802	0.790	0.939

level 0.1 PEL test with $n = 200$ and $c_* = 1$ under the alternative $\mu = \mu_1$ where the first $p/2$ components of μ_1 were equal to 1 and the rest were 0. Table 3 gives the power of the PEL test at level 0.1 under $\mu = \mu_1$ for different combinations of p, α and m. From Table 3, it appears that the power can be reasonably high for a suitable choice of the subsample size, although the maximum value critically depends on the dimension p of the parameters and the strength of dependence α. In particular, the PEL attains a higher (maximum) power under weaker dependence ($\alpha = 0.8, \infty$) than under strong dependence ($\alpha = 0, 0.1$).

5.3.3. *Comparison with Normal calibration.* Note that for $\alpha > 1/2$, the limit distribution of the logarithm of the PEL ratio statistic is Normal and therefore, one can use the limiting Normal distribution with an estimated variance to conduct the PEL test. In this section, we compare the performance of the subsampling-based calibration with the Normal distribution-based calibration. To estimate the asymptotic variance $\kappa^2 = 2c_*^2 \sum_{k=0}^{\infty} \rho_0(k)^2$, we first estimate $\rho_0(k)$ using the sample auto-covariance at lag-k based on the components of individual X_i's and then average them to get an estimate $\hat{\rho}_n(k)$ of $\rho_0(k)$ for $k = 1, \dots, K$ where $K = \min\{p/2, p^{1/2}\}$. Since κ^2 involves the squares of $\rho_0(k)$, the plug-in estimator is positive (with probability 1). Tables 4 and 5 compare the best performance of the subsampling based PEL with the Normal, calibration-based PEL for $n = 40$ and $n = 200$, respectively.

Tables 4 and 5 show that, except for the small values of p, the subsampling-based PEL method has a better accuracy (marked as bold) than the Normal,

TABLE 4
Comparison of the subsampling (SS) and Normal (G) calibrations for $n = 40$

	$p_1 = 7$		$p_2 = 20$		$p_3 = 80$	
α	G	SS	G	SS	G	SS
0.8	**0.122**	0.151	0.132	**0.080**	0.136	**0.081**
∞	**0.098**	0.092	0.030	**0.111**	0.076	**0.093**

TABLE 5
Comparison of the subsampling (SS) and Normal (G) calibrations for $n = 200$

	$p_1 = 7$		$p_2 = 20$		$p_3 = 80$	
α	**G**	**SS**	**G**	**SS**	**G**	**SS**
0.8	0.150	**0.113**	0.141	**0.082**	0.174	**0.127**
∞	**0.111**	0.165	0.048	**0.103**	0.238	**0.126**

calibration-based PEL. However, the computational burden associated with the subsampling method is typically larger than the Normal-based PEL.

6. Proofs. Note that for each $n \ge 1$, the PEL likelihood function in (2.1) is invariant with respect to (i) component-wise scaling and (ii) permutation of the p components. Hence, all through this section, without loss of generality (w.l.g.), we set the component-wise variance $\sigma_{nj}^2 = 1$ and set the permutation $\iota_n(j) = j$ for all $1 \le j \le p$ and $n \ge 1$. 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 $\rho_0(\cdot)$, mixing coefficients, etc.] are dropped to simplify notation, and limits in all order symbols are taken by letting $n \to \infty$. For $x \in \mathbb{R}$, let $\lfloor x \rfloor$ denote the largest integer not exceeding x and let $x_+ = \max\{x, 0\}$.

6.1. *Limit distribution in the nonergodic case.*

LEMMA 6.1. *For each $n \ge 1$, let $\{Y_{ij} = \mathrm{Y}_{ijn}, 1 \le j \le p\}$, $i = 1, 2, \dots, n$ be a collection of $p = p_n$-dimensional random vectors with $\mathrm{E}Y_{1j} = 0$ and $\mathrm{E}Y_{1j}^2 = \sigma_{nj}^2 \in (0, \infty)$. Let $\delta_j \equiv \delta_{nj} = s_{nj}^{-2}\mathbb{1}(s_{nj} \ne 0)$ where $s_{nj}^2 = n^{-1}\sum_{i=1}^n (Y_{ij} - \bar{Y}_{jn})^2$ and $\bar{Y}_{nj} = n^{-1}\sum_{i=1}^n Y_{ij}$. Also, let $Z_{ij} = |Y_{ij}| - \mathrm{E}|Y_{ij}|$, $\bar{Z}_{jn} = n^{-1}\sum_{i=1}^n Z_{ij}$, $W_i(j,l) = Y_{ij}Y_{il} - EY_{ij}Y_{il}$, $\bar{W}_n(j,l) = n^{-1}\sum_{i=1}^n W_i(j,l)$ and $D_{jn} = \{y \in \mathbb{R} : P(Y_{1j} = y) > 0\}$. Suppose that* (L.1) $\max\{\mathrm{E}(\sigma_{nj}^2\delta_j)^4 : 1 \le j \le p\} = O(1)$; (L.2) $\max\{\mathrm{E}|Y_{ij}|^8 : 1 \le j \le p\} < \mathrm{C}$, *and* (L.3) $\limsup_{n\to\infty} \max\{P(Y_{1j} = y) : y \in D_{jn}, j = 1, \dots, p\} < 1$. *Then*:

(a) *For $k = 1, 2, 3$, $\sum_{j=1}^p \delta_j \bar{Y}_{jn}^{2k} = O_p(n^{-k}p)$.*
(b) $\sum_{j=1}^p \delta_j[\bar{Z}_{jn}^2 + \bar{W}_{jn}^2] = O_p(n^{-1}p)$.
(c) $\sum_{j=1}^p \sum_{l=1}^p (\delta_j + 1)\delta_l \bar{W}_n^2(j,l) = O_p(n^{-1}p^2)$.
(d) *For $r = 1, 2$, $\max_{1\le j\le n} \sum_{j=1}^p \delta_j |\mathrm{Y}_{1j}|^r = O([pn]^{1/4})$.*
(e) $\sum_{j=1}^p (\delta_j - 1)^2 = O_p(n^{-1}p)$.

PROOF. By replacing Y_{ij}'s with Y_{ij}/σ_{nj} for all i, j, w.l.g., we can assume that $\sigma_{nj} = 1$ for all j, n. First consider part (a), $k = 3$; the proofs of $k = 1, 2$ are similar.

By repeated use of Hölder's inequality,

$$\mathrm{E}\sum_{j=1}^{p}\delta_j|\bar{Y}_{nj}|^6 \le \mathrm{E}\left(\sum_{j=1}^{p}\delta_j^4\right)^{1/4}\left(\sum_{j=1}^{p}\bar{Y}_{nj}^8\right)^{3/4} \le \left(\mathrm{E}\sum_{j=1}^{p}\delta_j^4\right)^{1/4}\left(\mathrm{E}\sum_{j=1}^{p}\bar{Y}_{nj}^8\right)^{3/4},$$

which is $O(pn^{-3})$, by (L.1), (L.2). This proves (a). Parts (b) and (c) follow by similar arguments. As for part (d), note that (for $r=2$)

$$\begin{aligned}\mathrm{E}\left(\max_{1\le i\le n}\sum_{j=1}^{p}\delta_j Y_{ij}^2\right) &\le \left[n\mathrm{E}\left(\sum_{j=1}^{p}\delta_j Y_{1j}^2\right)^4\right]^{1/4}\\ &\le n^{1/4}\left[\mathrm{E}\left(\sum_{j=1}^{p}\delta_j^4\right)^{1/4}\left(\mathrm{E}\sum_{j=1}^{p}|Y_{1j}|^{8/3}\right)^{3/4}\right]^{1/4} \le Cn^{1/4}p^{1/4}.\end{aligned}$$

Finally consider part (e). Note that for $j=1,2,\ldots,p$

$$\begin{aligned}\delta_j - 1 &= \frac{1-s_{nj}^2}{s_{nj}^2}\mathbb{1}(s_{nj}\ne 0) - \mathbb{1}(s_{nj}=0)\\ &= \delta_j\left[\bar{Y}_{nj}^2 - n^{-1}\sum_{i=1}^{n}(Y_{ij}^2-1)\right] - \mathbb{1}(s_{nj}=0)\\ &= \delta_j\bar{Y}_{nj}^2 - \delta_j\overline{\mathrm{W}}_n(j,j) - \mathbb{1}(s_{nj}=0),\end{aligned}$$

so that

$$\delta_j = \mathbb{1}(s_{nj}\ne 0) + \delta_j\bar{Y}_{nj}^2 - \delta_j\overline{\mathrm{W}}_n(j,j). \tag{6.1}$$

Part (e) can now be proved using (L.1)–(L.3) and the Cauchy–Schwarz inequality. We omit the details to save space. □

LEMMA 6.2. *Under the conditions of Theorem* 3.1,

$$-\log R_n(\mu_0) = n\gamma_n\left(\frac{\bar{Y}_{n1}}{\sigma_{n1}},\ldots,\frac{\bar{Y}_{np}}{\sigma_{np}}\right)(\mathrm{I}_p+2\gamma_n\mathrm{A}_n)^{-1}\left(\frac{\bar{Y}_{n1}}{\sigma_{n1}},\ldots,\frac{\bar{Y}_{np}}{\sigma_{np}}\right)' + o_p(1),$$

where $\bar{Y}_{nj} = n^{-1}\sum_{i=1}^{n}Y_{ij}$, $Y_{ij} = X_{ij} - \mu_j$, $1\le j\le p, 1\le i\le n$ *and* $\mathrm{A}_n = ((\rho_n(i-j)))_{p\times p}$.

PROOF. W.l.g., let $\sigma_{nj}=1$ for all $1\le j\le p, 1\le i\le n$. Note that

$$-\log R_n(\mu_0) = \min\{f(\pi_1,\ldots,\pi_n) : (\pi_1,\ldots,\pi_n)' \in \Pi_n\}, \tag{6.2}$$

where $f(\pi_1,\ldots,\pi_n) = -\sum_{i=1}^{n}\log(n\pi_i) + \lambda_n\sum_{j=1}^{p}\delta_j(\sum_{i=1}^{n}\pi_i Y_{ij})^2$. Since $f(\cdot)$ is strictly convex in $\pi_1,\ldots,\pi_n$ over a closed convex set $\Pi_n\subset\mathbb{R}^n$, it has a unique minimizer in Π_n. [The maximum of $f(\cdot)$ over Π_n is $+\infty$.] To find the minimizer, we use a Lagrange multiplier η and solve the set of equations

$$\frac{\partial}{\partial \pi_k} g(\pi_1, \ldots, \pi_n; \eta) = 0, \qquad 1 \le k \le n,$$
$$\frac{\partial}{\partial \eta} g(\pi_1, \ldots, \pi_n; \eta) = 0,$$

where $g(\pi_1, \ldots, \pi_n; \eta) = \sum_{i=1}^n \log(n\pi_i) - \lambda_n \sum_{j=1}^p \delta_j (\sum_{i=1}^n \pi_i Y_{ij})^2 + \eta(\sum_{i=1}^n \pi_i - 1)$. This leads to the equations

$$0 = \pi_k^{-1} - 2\lambda_n \sum_{j=1}^p \delta_j \left(\sum_{i=1}^n \pi_i Y_{ij}\right) \mathrm{Y}_{kj} + \eta, \qquad 1 \le k \le n \quad \text{and} \quad 1 = \sum_{i=1}^n \pi_i,$$

which, in turn, yield the implicit solution

$$\begin{aligned} &\eta = 2\lambda_n \sum_{j=1}^p \delta_j M_{nj}^2 - n \quad \text{and} \\ &\pi_k^{-1} = n\left\{1 + 2\gamma_n \sum_{j=1}^p \delta_j M_{nj} \mathrm{Y}_{kj} - 2\gamma_n \sum_{j=1}^p M_{nj}^2 \delta_j\right\}, \qquad 1 \le k \le n, \end{aligned} \tag{6.3}$$

where $\gamma_n = \lambda_n/n$ and $M_{nj} = \sum_{i=1}^n \pi_i Y_{ij}$. To obtain a more explicit approximation, we show that π_k's are of the form $\pi_k = n^{-1}(1 + o_p(1))$ uniformly in k. In view of Brouwer's fixed point theorem [cf. Milnor (1965)], it is enough to show that, with $a_n^{-1} = n^{-1/2} \log(n)$,

$$\begin{aligned} &\max_{1 \le k \le n} \left| \frac{1}{n} \left\{1 + 2\gamma_n \sum_{j=1}^p \delta_j M_{nj} \mathrm{Y}_{kj} - 2\gamma_n \sum_{j=1}^p M_{nj}^2 \delta_j\right\}^{-1} - \frac{1}{n}\right| \\ &\quad = O_p(n^{-1} a_n^{-1}) \end{aligned} \tag{6.4}$$

whenever $\max\{|\pi_k - n^{-1}| : 1 \le k \le n\} = O(n^{-1} a_n^{-1})$. To prove (6.4), we first show that

$$\sup\left\{\gamma_n \sum_{j=1}^p \delta_j \mathrm{M}_{nj}^2 : (\pi_1, \ldots, \pi_n)' \in \Pi_n^0\right\} = O_p(a_n^{-2}), \tag{6.5}$$

where $\Pi_n^0 = \{(\pi_1, \ldots, \pi_n)' \in \Pi_n : |\pi_k - n^{-1}| \le C a_n^{-1} n^{-1}$ for all $1 \le k \le n\}$. Note that for any $(\pi_1, \ldots, \pi_n)' \in \Pi_n^0$, $\sum_{i=1}^n \pi_i = 1 = \sum_{i=1}^n (1/n) \Longrightarrow \sum_{i=1}^n (1 - n\pi_i) = 0 \Longrightarrow \sum_i (1 - n\pi_i)_+ = \sum_i (n\pi_i - 1)_+$. Also, $(n\pi_i - 1)_+ > 0$ if and only if (iff) $n\pi_i > 1$ and similarly, $(1 - n\pi_i)_+ > 0$ iff $n\pi_i < 1$. Hence, using the bound "$|n\pi_i -$

$1| \le \mathrm{C}a_n^{-1}$ for all $i = 1, \dots, n$," we have

$$\begin{aligned}\left|\sum_{i=1}^{n}(n\pi_i)^{-1} - n\right| &= \left|\sum_{i=1}^{n}\frac{(1-n\pi_i)_+}{n\pi_i} - \sum_{i=1}^{n}\frac{(n\pi_i - 1)_+}{n\pi_i}\right| \\ &= \left|\sum_{i=1}^{n}\frac{(1-n\pi_i)_+}{1-(1-n\pi_i)_+} - \sum_{i=1}^{n}\frac{(n\pi_i - 1)_+}{1+(n\pi_i - 1)_+}\right| \\ &= \left|\sum_{i=1}^{n}(1-n\pi_i)_+^2[1+O(a_n)] + \sum_{i=1}^{n}(n\pi_i - 1)_+^2[1+O(a_n)]\right| \\ &\le 2\mathrm{C}^2 n a_n^{-2} \qquad \text{for large } n.\end{aligned}$$

By (6.3), $n^{-1}\sum_{k=1}^{n}(n\pi_k)^{-1} - 1 = 2\gamma_n \sum_{j=1}^{p}\delta_j \mathrm{M}_{nj}[\bar{Y}_{nj} - \mathrm{M}_{nj}]$. Hence by Lemma 6.1 and the Cauchy–Schwarz inequality,

$$\begin{aligned}\frac{2\mathrm{C}^2}{a_n^2} &\ge 2\gamma_n\left|\sum_{j=1}^{p}\delta_j \mathrm{M}_{nj}[\bar{Y}_{nj} - \mathrm{M}_{nj}]\right| \\ &\ge 2\left(\gamma_n \sum_{j=1}^{p}\delta_j \mathrm{M}_{nj}^2\right)^{1/2}\left[\left(\gamma_n \sum_{j=1}^{p}\delta_j \mathrm{M}_{nj}^2\right)^{1/2} - O_p(n^{-1/2})\right]\end{aligned}$$

uniformly in $(\pi_1, \dots, \pi_n) \in \Pi_n^0$, for n large. Consequently, (6.5) holds. Now using (6.5), (6.3), the Cauchy–Schwarz inequality and Lemma 6.1, (6.4) follows. Hence, by Brouwer's fixed point theorem [cf. Milnor (1965)], there exists a solution $(\pi_1^0, \dots, \pi_n^0)$ of (6.3) satisfying the bound

$$\max\{|\pi_k^0 - n^{-1}| : 1 \le k \le n\} = O_p(n^{-1}a_n^{-1}). \tag{6.6}$$

Using the second derivative condition, it is easy to verify that $(\pi_1^0, \dots, \pi_n^0)$ is a local minimizer of $f(\cdot)$. In view of the strict convexity of $f(\cdot)$ on Π_n, it also follows that $(\pi_1^0, \dots, \pi_n^0)$ is the unique minimizer of $f(\cdot)$ over Π_n.

Next let $\Gamma_{1n} = 2n\gamma_n^2 \sum_{j=1}^{p}\sum_{l=1}^{p}\rho_n(j,l)\delta_j\delta_l \mathrm{M}_{nj}^0 \mathrm{M}_{nl}^0$, where $\mathrm{M}_{nj}^0 = \sum_{i=1}^{n}\pi_i^0 Y_{ij}$, $1 \le j \le p$. Then, from (6.3), we have

$$\begin{aligned}-\log R_n(\mu_0) &\equiv f(\pi_1^0, \dots, \pi_n^0) \\ &= \sum_{i=1}^{n}\log\left\{1 + 2\gamma_n \sum_{j=1}^{p}\delta_j \mathrm{M}_{nj}^0 Y_{ij} - 2\gamma_n \sum_{j=1}^{p}\delta_j (\mathrm{M}_{nj}^0)^2\right\} \\ &\quad + n\gamma_n \sum_{j=1}^{p}\delta_j (\mathrm{M}_{nj}^0)^2 \\ &\equiv \sum_{i=1}^{n}\left\{2\gamma_n \sum_{j=1}^{p}\delta_j \mathrm{M}_{nj}^0 Y_{ij} - 2\gamma_n \sum_{j=1}^{p}\delta_j (\mathrm{M}_{nj}^0)^2\right\} - \Gamma_{1n}\end{aligned}$$

$$
\begin{aligned}
&+ \mathrm{R}_{1n} + n\gamma_n \sum_{j=1}^{p} \delta_j (\mathrm{M}_{nj}^0)^2 \\
&= 2n\gamma_n \sum_{j=1}^{p} \delta_j \mathrm{M}_{nj}^0 \bar{Y}_{ij} - n\gamma_n \sum_{j=1}^{p} \delta_j (\mathrm{M}_{nj}^0)^2 - \Gamma_{1n} + \mathrm{R}_{1n} \\
&\equiv 2n\gamma_n \sum_{j=1}^{p} \mathrm{M}_{nj}^0 \bar{Y}_{ij} - n\gamma_n \sum_{j=1}^{p} (\mathrm{M}_{nj}^0)^2 \\
&\quad - 2n\gamma_n^2 \sum_{j=1}^{p} \sum_{l=1}^{p} \rho_n(j,l) \mathrm{M}_{nj}^0 \mathrm{M}_{nl}^0 + \mathrm{R}_{2n} \\
&\equiv n\gamma_n (\bar{Y}_{n1}, \ldots, \bar{Y}_{np})(\mathrm{I}_p + 2\gamma_n \mathrm{A}_n)^{-1} (\bar{Y}_{n1}, \ldots, \bar{Y}_{np})' + \mathrm{R}_{3n},
\end{aligned}
$$

where the remainder terms R_{kn}'s are defined by subtraction. By the next lemma, $\mathrm{R}_{kn} = o_p(1)$ for $k = 1, 2, 3$. Hence, Lemma 6.2 is proved. □

LEMMA 6.3. *Under the conditions of Theorem* 3.1, $\sum_{k=1}^{3} |\mathrm{R}_{kn}| = o_p(1)$.

PROOF. See [LM] for details. □

PROOF OF THEOREM 3.1. Recall that $\sigma_{nj} = 1, \forall j, n$. We carry out the proof in 2 steps.

Step (I): Let $\mathrm{Z}_n(t) = \sum_{j=1}^{p} (\sqrt{n}\bar{Y}_{nj}) \mathbb{1}_{((j-1)/p, j/p]}(t)$, $t \in (0, 1]$ and let $\mathrm{Z}_n(0) \equiv \mathrm{Z}_n(0+)$. Then, $P(\mathrm{Z}_n \in L^2[0, 1]) = 1$. The first step is to prove that $\mathrm{Z}_n(\cdot) \to^d \mathrm{Z}(\cdot)$ as elements of $L^2[0, 1]$. Recall that $\{\phi_j : j \in \mathbb{N}\}$ is a complete orthonormal basis for $L^2[0, 1]$. By Theorem 1.84 of Van der Vaart and Wellner (1996), it enough to show that:

(i) For any $0 < t_1 < \cdots < t_r \le 1$, $1 \le r < \infty$,

$$
\big(\mathrm{Z}_n(t_1), \ldots, \mathrm{Z}_n(t_r)\big) \to^d \big(\mathrm{Z}(t_1), \ldots, \mathrm{Z}(t_r)\big), \tag{6.7}
$$

(ii) For any $\varepsilon > 0, \delta > 0$, there exists $N = N(\varepsilon, \delta) \in \mathbb{N}$ such that

$$
\limsup_{n \to \infty} P\left(\sum_{k=N}^{\infty} \left| \langle \mathrm{Z}_n, \phi_k \rangle \right|^2 > \delta \right) < \varepsilon. \tag{6.8}
$$

Part (i) can be proved using Theorem 11.1.6 of Athreya and Lahiri (2006); we omit the routine details. For part (ii), it is enough to show that

$$
\lim_{N \to \infty} \limsup_{n \to \infty} \mathrm{E}\left(\sum_{k=N}^{\infty} \left| \langle Z_n, \phi_k \rangle \right|^2 \right) = 0. \tag{6.9}
$$

Let $I_j = (\frac{j-1}{p}, \frac{j}{p}]$, $j = 1, \dots, p$. Then, by Fubini's theorem and (C.3),

$$\begin{aligned}\mathrm{E}\sum_{k=1}^{\infty}|\langle \mathrm{Z}_n, \phi_k\rangle|^2 &= \sum_{k=1}^{\infty}\sum_{j=1}^{p}\sum_{l=1}^{p}\int_{I_j}\int_{I_l}\phi_k(t)\phi_k(s)\rho_0(j/p, l/p)\,ds\,dt \\ &= \sum_{k=1}^{\infty}\int_0^1\int_0^1 \phi_k(t)\phi_k(s)\rho_0(s,t)\,ds\,dt + o(1),\end{aligned} \tag{6.10}$$

which equals $\sum_{k=1}^{\infty}\langle \phi_k, \Lambda_0\phi_k\rangle + o(1)$. Next, using $|\rho_n(\cdot,\cdot)| \le 1$ and $\int |\phi_k(t)|\,dt \le (\int \phi_k^2(t))^{1/2} = 1$, one can show that for each fixed $k \in \mathbb{N}$,

$$\begin{aligned}\mathrm{E}|\langle \mathrm{Z}_n, \phi_k\rangle|^2 &= \sum_{j=1}^{p}\sum_{l=1}^{p}\int_{I_j}\int_{I_l}\phi_k(t)\phi_k(s)\rho_0(j/p, l/p)\,ds\,dt \\ &\to \int_0^1\int_0^1 \phi_k(t)\phi_k(s)\rho_0(s,t)\,ds\,dt = \langle \phi_k, \Upsilon_0\phi_k\rangle.\end{aligned} \tag{6.11}$$

By (6.10), (6.11) and (C.3), (6.8) follows. Thus, $\mathrm{Z}_n \to^d \mathrm{Z}$ on $L^2[0,1]$.

Step (II): Next we establish weak convergence of the quadratic form:

$$np^{-1}\bar{\mathbf{Y}}_n'(\mathrm{I}_p + 2\gamma_n \mathrm{A}_n)^{-1}\bar{\mathbf{Y}}_n = \sum_{k=0}^{\infty} np^{-1}\bar{\mathbf{Y}}_n'(-2\gamma_n\mathrm{A}_n)^k\bar{\mathbf{Y}}_n. \tag{6.12}$$

Note that by condition (C.3), $\|\gamma_n\mathrm{A}_n\|^2 \le \gamma_n^2\sum_{j=1}^{p}\sum_{l=1}^{p}\rho_0^2(j/p, l/p) \to c_*^2\int_0^1\int_0^1 \rho_0^2(x,y)\,dx\,dy \in (0, 1/4)$. Hence there exists a nonrandom $\varepsilon_o \in (0,1)$ such that for any fixed $m \in \mathbb{N}$ and for n sufficiently large (not depending on m),

$$\begin{aligned}&\sum_{k=m}^{\infty}|np^{-1}(\bar{Y}_{n1},\dots,\bar{Y}_{np})(-2\gamma_n\mathrm{A}_n)^k(\bar{Y}_{n1},\dots,\bar{Y}_{np})'| \\ &\quad \le \sum_{k=m}^{\infty} p^{-1}\sum_{j=1}^{p}(\sqrt{n}\bar{Y}_{nj})^2\|2\gamma_n\mathrm{A}_n\|^k \le \|\mathrm{Z}_n(\cdot)\|^2\sum_{k=m}^{\infty}\varepsilon_o^k.\end{aligned} \tag{6.13}$$

Next, let $a_{n.k}(j,l)$ denote the (j,l)th element of $(\gamma_n\mathrm{A}_n)^k$. Note that $\rho_0(\cdot,\cdot)$ is uniformly continuous on $[0,1]^2$. Hence, using induction and condition (C.3), it can be shown that for any $k \in \mathbb{N}$,

$$\sum\{|a_{n,k}(j,l) - p^{-1}c_*^k\rho_0^{*(k)}(j/p, l/p)| : 1 \le j, l \le p\} = o(1).$$

By the continuous mapping theorem, it now follows that for any fixed m,

$$\begin{aligned}\sum_{k=0}^{m}\frac{n}{p}\bar{\mathbf{Y}}_n'(-2\gamma_n\mathrm{A}_n)^k\bar{\mathbf{Y}}_n &= \sum_{k=0}^{m}\frac{(-2)^k}{p}\sum_{j=1}^{p}\sum_{k=1}^{p}\mathrm{Z}_n(j/p)\mathrm{Z}_n(l/p)a_{n.k}(j,l) \\ &\to^d \sum_{k=0}^{m}(-2)^k\int_0^1\int_0^1 \mathrm{Z}(u)\mathrm{Z}(v)c_*^k\rho_0^{*(k)}(u,v)\,du\,dv.\end{aligned} \tag{6.14}$$

Thus, by (6.14), partial sums of the infinite series in (6.12) converge to the partial sums of the limit series for any fixed m. By (6.13), the tail of the infinite series in (6.12) is negligible. It can be shown (cf. [LM]) that the tail of the limit series is also negligible. Since, $\|\mathrm{Z}_n(\cdot)\|_2 \to^d \|\mathrm{Z}(\cdot)\|_2$, by (6.12)–(6.14) and Lemmas 6.1, 6.3, the theorem is proved. $\square$

6.2. *Limit distribution for the ergodic case.* We prove Theorem 3.2 and 3.3, using different arguments than the proof of Theorem 3.1. This is necessitated by the fact that *we need more accurate bounds on the remainder terms that must become negligible after the scaling* (*e.g., by* p^{α_0}). We will also use the notation $O_p^u(\cdot)$ to denote a bound that holds uniformly over $i \in \{1, \dots, n\}$ as $n \to \infty$. For example, $\Delta_{in} = O_p^u(a_n^{-1})$ means $\max\{|\Delta_{in}| : 1 \le i \le n\} = O_p(a_n^{-1})$ an $n \to \infty$. Similarly, define $o_p^u(\cdot)$.

PROOF OF THEOREM 3.2. For $1 \le i \le n$, let $\Delta_{in} = 2\gamma_n \sum_{j=1}^p \delta_j \mathrm{M}_{nj}^0 Y_{ij}$, $\mathrm{D}_n = 2\gamma_n \sum_{j=1}^p \delta_j (\mathrm{M}_{nj}^0)^2$ and $\Delta_{in}^0 = \Delta_{in} - \mathrm{D}_n$. Then, $\pi_i^0 = 1/[n(1 + \Delta_{in}^0)]$. Now using $|\pi_i^0 - 1/n| \le a_n^{-1} n^{-1}$ for $1 \le i \le n$, one can show (cf. [LM]) that

$$|\Delta_{in}^0| = O_p^u(a_n^{-1}) \quad \text{and} \quad |\Delta_{in}| = O_p^u(a_n^{-1}). \tag{6.15}$$

Hence, by Taylor's expansion of $\log(1 + x)$ around $x = 0$,

$$\begin{aligned}
-\log R_n(\mu_0) &= \sum_{i=1}^n \log(1 + \Delta_{in}^0) + \lambda \sum_{j=1}^p \delta_j (M_{nj}^0)^2 \\
&= \left[2n\gamma_n \sum_{j=1}^p \delta_j M_{nj}^0 \bar{Y}_{nj} - \frac{n}{2} D_n\right] - 2^{-1}\left[\sum_{i=1}^n \Delta_{in}^2 + 2D_n \sum_{j=1}^n \Delta_{jn}\right] \\
&\quad + 3^{-1} \sum_{i=1}^n \Delta_{in}^3 + O_p(na_n^{-4}).
\end{aligned} \tag{6.16}$$

There exist $E_{nj} = n^{-1} \sum_{i=1}^n [Y_{ij} \cdot O_p^u(a_n^{-4})]$, $1 \le j \le p$, such that (cf. [LM])

$$\begin{aligned}
M_{nj}^0 &= \bar{Y}_{nj} - n^{-1} \sum_{i=1}^n Y_{ij} \Delta_{in} + D_n \bar{Y}_{nj} + n^{-1} \sum_{i=1}^n Y_{ij} [\Delta_{in}^2 + 2\Delta_{in} D_n] \\
&\quad - n^{-1} \sum_{i=1}^n Y_{ij} (\Delta_{in})^3 + E_{nj};
\end{aligned} \tag{6.17}$$

$$L_{1n} \equiv 2\gamma_n n \sum_{j=1}^p \delta_j M_{nj}^0 \bar{Y}_{nj} - \frac{n}{2} D_n \qquad \text{the lead term of } -\log R_n(\mu)$$

$$= \gamma_n n \sum_{j=1}^{p} \delta_j M_{nj}^0 \bar{Y}_{nj} + \sum_{i=1}^{n} \Delta_{in}^2/2 + \sum_{i=1}^{n} \Delta_{in}^4/2 - \frac{1}{2}\left(\sum_{i=1}^{n} \Delta_{in}\right) D_n - \frac{1}{2}\sum_{i=1}^{n} \Delta_{in}^3 - \sum_{i=1}^{n} \Delta_{in}^2 D_n + \sum_{i=1}^{n} \Delta_{in} \cdot O_p^u(a_n^{-4}).$$

Hence, from (6.16),

$$-\log R_n(\mu) = n\gamma_n \sum_{j=1}^{p} \delta_j M_{nj}^0 \bar{Y}_{nj} - \frac{3}{2}\left(\sum_{i=1}^{n} \Delta_{in}\right) D_n - \frac{1}{6}\sum_{i=1}^{n} \Delta_{in}^3 + O_p(na_n^{-4}). \tag{6.18}$$

Next, define $\beta_n^2 = p\sum_{k=1}^{p} k^{-\alpha}$. Then,

$$\beta_n^2 \sim \begin{cases} Cp, & \text{if } \alpha > 1, \\ Cp\log p, & \text{if } \alpha = 1, \\ Cp^{2-\alpha}, & \text{if } 0 < \alpha < 1. \end{cases} \tag{6.19}$$

Note that for any $1 \le i \le n$, $\mathrm{E}\bar{Y}_{nj} Y_{ij} = n^{-1}\sum_{k=1}^{n} \mathrm{E}Y_{kj}Y_{ij} = 1/n$. Let $V_{in} \equiv \gamma_n \sum_{j=1}^{p}(\delta_j \bar{Y}_{nj} Y_{ij} - 1/n)$ and $D_{nj}^Y \equiv \{y : P(Y_{1j} = y) > 0\}$. Then there exists a $\delta \in (0, 1)$ such that by using (6.1) and the conditions of Theorem 3.2, one gets (cf. [LM])

$$\begin{aligned} \mathrm{E}\sum_{i=1}^{n} V_{in}^2 &\le C\Bigg[n\mathrm{E}\left\{\gamma_n \sum_{j=1}^{p}(\bar{Y}_{nj}Y_{1j} - 1/n)\right\}^2 + n\mathrm{E}\left\{\gamma_n \sum_{j=1}^{p} \delta_j \bar{Y}_{nj}^3 Y_{1j}\right\}^2 \\ &\quad + n\mathrm{E}\left\{\gamma_n \sum_{j=1}^{p} \delta_j \bar{W}_n(j, j)\bar{Y}_{nj}Y_{1j}\right\}^2 \\ &\quad + n\mathrm{E}\left\{\gamma_n \sum_{j=1}^{p} Y_{ij}^2 \mathbb{1}(s_{nj} = 0)\right\}^2\Bigg] \\ &\le Cp^{-2}\beta_n^2 + O(n^{-1}) + n\delta^{n-2} \sum_{y \in D_{nj}^Y} y^2 P(Y_{1j} = y) \\ &\le Cp^{-2}\beta_n^2 + O(n^{-1}). \end{aligned} \tag{6.20}$$

Using (6.17), (6.20) and the Cauchy–Schwartz inequality, one can show that

$$n\gamma_n \sum_{j=1}^{p} \delta_j M_{nj}^0 \bar{Y}_{nj} = n\gamma_n \sum_{j=1}^{p} \delta_j \bar{Y}_{nj}^2 - \sum_{i=1}^{n} \Delta_{in}[V_{in} + 1/n] + O_p(a_n^{-2}) + O_p(a_n^{-1} p^{-1}\beta_n). \tag{6.21}$$

Also, from (6.17), for any $1 \le k \le n$, we have

$$\begin{aligned}\Delta_{kn} &= 2\gamma_n \sum_{j=1}^{p} \delta_j M_{nj}^0 Y_{kj}\\ &\equiv 2\gamma_n \sum_{j=1}^{p} \delta_j \bar{Y}_{nj} Y_{kj} - \frac{2}{n}\Delta_{kn} + R_{1n}(k)\\ &\equiv \left(1+\frac{2}{n}\right)^{-1} 2\gamma_n \left(\sum_{j=1}^{p} \bar{Y}_{nj} Y_{kj}\right) + R_{2n}(k),\end{aligned}$$

where $R_{1n}(k)$ and $R_{2n}(k)$ are remainder terms satisfying (cf. [LM])

$$\sum_{k=1}^{n} R_{ln}^2(k) = O_p\bigl(na_n^{-2} p^{-2}\beta_n^2 + na_n^{-4}\bigr), \qquad l = 1, 2.$$

Next, using similar arguments and noting that $\mathrm{E}\bar{W}_n(j, j)\bar{Y}_{nj}^2 = O(n^{-2})$ and $\mathrm{Var}(\bar{W}_n(j, j)\bar{Y}_{nj}^2) \le Cn^{-3}$ for all j, one can show (cf. [LM]) that

$$\sum_{i=1}^{n} \Delta_{in}(V_{in} + 1/n) = \left(\frac{n}{n+2}\right) 2\gamma_n^2 \sum_{i=1}^{n} \left(\sum_{j=1}^{p} \bar{Y}_{nj} Y_{ij}\right)^2 (1 + o_p(1)), \tag{6.22}$$

$$n\gamma_n \sum_{j=1}^{p} \delta_j \bar{Y}_{nj}^2 = n\gamma_n \sum_{j=1}^{p} \bar{Y}_{nj}^2 + O_p\bigl(n^{-1} + n^{-1/2} p^{-1}\beta_n\bigr), \tag{6.23}$$

$$\begin{aligned}\sum_{i=1}^{n} \Delta_{in}^3 &= \left(\frac{n}{n+2}\right) 8\gamma_n^2 \left[\sum_{i=1}^{n} \left(\sum_{j=1}^{p} \bar{Y}_{nj} Y_{ij}\right)^3\right] (1 + o_p(1))\\ &= O_p\bigl(n^{-1} + n^{-1/2} p^{-1}\beta_n\bigr).\end{aligned} \tag{6.24}$$

Using (6.18), (6.21) and (6.22)–(6.24) and the fact that $\mathrm{E}\bar{Y}_{nj} Y_{ij} = n^{-1}$ and $\mathrm{Var}(\bar{Y}_{nj} Y_{ij}) \le Cn^{-1}$ for all i, j, one can conclude (cf. [LM])

$$-\log R_n(\mu_0) = n\gamma_n \sum_{j=1}^{p} \bar{Y}_{nj}^2 - \frac{2}{n} + O_p\bigl(na_n^{-4}\bigr) + O_p\bigl(a_n^{-1} p^{-1}\beta_n\bigr). \tag{6.25}$$

Set $a_n = n^{1/2}(p/n^2)^{1/10}$. Then $a_n \to \infty$, $a_n = o(n^{1/2})$ and $\sqrt{p}.n.a_n^{-4} = o(1)$, there by making the last two terms in (6.25) $o(p^{-1/2})$ whenever $p = o(n^2)$. Now Theorem 3.2 follows by adapting the proof of the CLT for a stationary sequence of ϱ-mixing random variables to triangular arrays. □

PROOF OF THEOREM 3.3. Arguments in the proof of Theorem 3.2 yield the asymptotic approximation for $-\log R_n(\mu_0)$ in (6.25) with $\beta_n^2 \sim cp^{2-\alpha}$ for all $0 <$

$\alpha < 1/2$. Now choose $a_n \to \infty$ to satisfy $p^{\alpha}[n^{-1} + na_n^{-4} + a_n^{-1}p^{-1}\beta_n] \to 0$, for example, $a_n = n^{1/2}[p^{\alpha}/n]^{1/5}$. Then it follows that

$$p^{\alpha}(-\log R_n(\mu_0) - c_*) = c_* p^{\alpha}\left(\frac{1}{p}\sum_{j=1}^{p}\{(\sqrt{n}\bar{Y}_{nj})^2 - 1\}\right) + o_p(1).$$

In the case where X_{nj}'s are Gaussian with $\varrho_n(i, j) = \varrho_\alpha(i - j)$, the leading term has the same distribution as $W_n \equiv (p^{-2+2\alpha})^{1/2}\sum_{j=1}^{p}(Z_j^2 - 1)$, where $\{Z_j\}$ is a stationary Gaussian process with correlation function $\varrho_\alpha(\cdot)$. Then the result of Taqqu (1975) implies the $W_n \to^d W$, and the theorem follows. In the general case when X_{nj}'s are not Gaussian, the theorem follows by using convergence of moments of $\sqrt{n}\bar{Y}_n$ to the moments of $N(0, 1)$ and a variant of the diagram formula; cf. Arcones (1994). □

PROOF OF THEOREM 3.4. Similar to the proof of Theorem 3.3; see [LM]. □

PROOF OF THEOREM 4.1. Note that $\mathcal{I}_n = \{I_i : 1 \le i \le n - m + 1\}$ with $I_i = \{i, i + 1, \dots, i + m - 1\}$. Let

$$U_j(x) \equiv \sum_{i=(j-1)m+1}^{jm \wedge n-m+1} \mathbb{1}(-\log R_m^*(\mu_0, I_i) \le x), \qquad x \in \mathbb{R},$$

for $1 \le j \le M$ where $M = \lceil (n - m + 1)/m \rceil$ is the smallest integer not less than $(n - m + 1)/m$. By the independence of $U_j(x)$ and $U_{j+k}(x)$ for $k \ge 2$, one can show (cf. [LM]) that for each $x \in \mathbb{R}$,

$$\begin{aligned} &\mathrm{E}(\hat{G}_n^{\mathrm{NE}}(x) - P(-\log R_m^*(\mu_0, I_i) \le x))^2 \\ &\quad \le Cn^{-2}Mm^2 = o(1). \end{aligned} \tag{6.26}$$

The next arguments are similar to the proof of the Glivenko–Cantelli theorem [cf. Theorem 13.3 of Billingsley (1999)] and the continuity of the limit distribution of $-\log R_n(\mu_0)$; one can complete the proof; see [LM]. □

PROOF OF THEOREM 4.2. Similar to the proof of Theorem 4.1. □

PROOF OF REMARK 4.1. Here we outline a proof of consistency of the permutation invariant estimators $\hat{\alpha}$ and $\hat{\kappa}$ of Remark 4.1. W.l.g, suppose that $\mu = 0$ and $\sigma_{nj} = 0$ for all j, n. First consider the estimator $\hat{\alpha}$ in (4.4). Let $A_n = \{s_{nj} \ne 0$ for all $j = 1, \dots, p\}$, $n \ge 1$. Then, by condition (C.1), $P(A_n^c) = O(p\delta^{n-1})$ for some $\delta \in (0, 1)$. On the set A_n, $e_n = 0$, and using arguments similar to those in the proof of Lemma 6.2, one can show that

$$e_n + n^{-1}\sum_{i=1}^{n}\left\{p^{-1}\sum_{j=1}^{p}[X_{ij} - \bar{X}_{nj}]\delta_j^{1/2}\right\}^2 = n^{-1}\sum_{i=1}^{n}\bar{X}_{ip}^2 + R_n,$$

where $R_n = o_p(p^{-\alpha})$. Note that $EI_{11} \sim \frac{2Cp^{-\alpha}}{(1-\alpha)(2-\alpha)}$ for $0 < \alpha < 1$ and $\mathrm{Var}(I_{11}) = O(n^{-1}p^{-2\alpha})$. Now it is easy to verify that $\hat{\alpha}$ satisfies (4.2).

To prove the consistency of $\hat{\kappa}$ of (4.4), using moderate deviation inequalities [cf. Götze and Hipp (1978)], one can conclude that

$$P\left(\left|\hat{c}(j,k) - c(j,k)\right| > n^{-1/2}\log n \text{ for some } 1 \le j,k \le p\right) = o(1). \tag{6.27}$$

Next note that $Ck^{-\alpha} \le n^{-1/2}\log n$ for all $k > n^{1/(2\alpha)}$. This implies that only $O(n^{1/(2\alpha)})$-many correlation terms contribute to $\hat{\kappa}$, with probability tending to one. Now using (6.27) for the nonvanishing terms and the fact that $\alpha > 1/2$, one can prove consistency of $\hat{\kappa}$. $\square$

Acknowledgments. We thank three anonymous reviewers and Professors Peter Bühlmann and Art Owen for a number of constructive comments that improved an earlier version of the paper.

SUPPLEMENTARY MATERIAL

Numerical results and proofs (DOI: 10.1214/12-AOS1040SUPP; .pdf). Additional simulation results and some details of proofs.

REFERENCES

ARCONES, M. A. (1994). Limit theorems for nonlinear functionals of a stationary Gaussian sequence of vectors. *Ann. Probab.* **22** 2242–2274. MR1331224

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

BARTOLUCCI, F. (2007). A penalized version of the empirical likelihood ratio for the population mean. *Statist. Probab. Lett.* **77** 104–110. MR2339024

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

BERTAIL, P. (2006). Empirical likelihood in some semiparametric models. *Bernoulli* **12** 299–331. MR2218557

BICKEL, P. J. and LEVINA, E. (2008). Covariance regularization by thresholding. *Ann. Statist.* **36** 2577–2604. MR2485008

BILLINGSLEY, P. (1999). *Convergence of Probability Measures*, 2nd ed. Wiley, New York. MR1700749

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

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

CHEN, J., VARIYATH, A. M. and ABRAHAM, B. (2008). Adjusted empirical likelihood and its properties. *J. Comput. Graph. Statist.* **17** 426–443. MR2439967

CRAMÉR, H. and LEADBETTER, M. R. (1967). *Stationary and Related Stochastic Processes*. Wiley, New York.

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

DOBRUSHIN, R. L. and MAJOR, P. (1979). Non-central limit theorems for nonlinear functionals of Gaussian fields. *Z. Wahrsch. Verw. Gebiete* **50** 27–52. MR0550122

EL KAROUI, N. (2008). Operator norm consistent estimation of large-dimensional sparse covariance matrices. *Ann. Statist.* **36** 2717–2756. MR2485011

EMERSON, S. C. and OWEN, A. B. (2009). Calibration of the empirical likelihood method for a vector mean. *Electron. J. Stat.* **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

GÖTZE, F. and HIPP, C. (1978). Asymptotic expansions in the central limit theorem under moment conditions. *Z. Wahrsch. Verw. Gebiete* **42** 67–87. MR0467882

HALL, P., HOROWITZ, J. L. and JING, B.-Y. (1995). On blocking rules for the bootstrap with dependent data. *Biometrika* **82** 561–574. MR1366282

HALL, P., JING, B.-Y. and LAHIRI, S. N. (1998). On the sampling window method under long range dependence. *Statist. Sinica* **8** 1189–1204.

HJORT, N. L., MCKEAGUE, I. W. 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. MR0322926

LAHIRI, S. N. (2003). *Resampling Methods for Dependent Data*. Springer, New York. MR2001447

LAHIRI, S. N. and MUKHOPADHYAY, S. (2012). Supplement to "A penalized empirical likelihood method in high dimensions." DOI:10.1214/12-AOS1040SUPP.

MILNOR, J. W. (1965). *Topology from the Differentiable Viewpoint*. The Univ. Press of Virginia, Charlottesville, VA. MR0226651

OTSU, T. (2007). Penalized empirical likelihood estimation of semiparametric models. *J. Multivariate Anal.* **98** 1923–1954. MR2396947

OWEN, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75** 237–249. MR0946049

OWEN, A. (2001). *Empirical Likelihood*. Chapman & Hall, Boca Raton, FL.

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** 905–919. MR2746160

TAQQU, M. S. (1975). Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Z. Wahrsch. Verw. Gebiete* **31** 287–302. MR0400329

TAQQU, M. S. (1977). Law of the iterated logarithm for sums of non-linear functions of Gaussian variables that exhibit a long range dependence. *Z. Wahrsch. Verw. Gebiete* **40** 203–238. MR0471045

TAQQU, M., TEVEROVSKY, V. and WILLINGER, W. (1995). Estimators for long-range dependence: An empirical study. *E. Fractals* **3** 785–798.

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

VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York. MR1385671

WRIGHT, S. J. (1997). *Primal-dual Interior-point Methods*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA. MR1422257

DEPARTMENT OF STATISTICS
TEXAS A&M UNIVERSITY
TAMU MS 3143
COLLEGE STATION, TEXAS 77843-3143
USA
E-MAIL: snlahiri@stat.tamu.edu
deep@stat.tamu.edu