EMPIRICAL LIKELIHOOD METHODS WITH WEAKLY DEPENDENT PROCESSES

Yuichi Kitamura[1]
Department of Economics
University of Minnesota

July 25, 1996

comments welcome
Minneapolis, MN 55455
(612) 625-9571
kitamura@atlas.socsci.umn.edu

---

[1]The author thanks Bill Brown, Phoebus Dhrymes, Mark Dwyer, John Geweke, Doug Hodgson, Guido Imbens, Max King, John Knight, Larry Marsh, Akimichi Takemura, Tao Zha and seminar participants at the Conference on Recent Developments in Econometrics at Hitotsubashi University, August 1995, the 1996 North American Meetings of the Econometric Society, the 1996 Annual Meetings of Midwest Economics Association, the 1996 Annual Meetings of Society for Economics Dynamics and Control, the Federal Reserve Bank of Atlanta, Columbia University, the University of Minnesota and the University of Rochester for their comments and suggestions.

ABSTRACT

This paper studies the method of empirical likelihood in models with weakly dependent processes. In such cases, if the likelihood function is formulated as if the data process is independent, obviously empirical likelihood fails. We propose to use empirical likelihood of blocks of observations to solve this problem in a nonparametric manner. This method of "blockwise empirical likelihood" preserves the dependence of data, and the resulting likelihood ratios can be used to construct valid confidence intervals. We consider general estimating equations, for which an efficient estimator is derived by maximizing blockwise empirical likelihood. We also introduce "blocks-of-blocks empirical likelihood" to conduct inference of parameters of the infinite dimensional joint distribution of data, such as spectra; the smooth function model is used for such cases. We show that blockwise empirical likelihood of the smooth function model with weakly dependent processes is Bartlett-correctable.

KEYWORDS: Bartlett correction, Edgeworth expansion, Empirical likelihood, Estimating function, Generalized method of moments, Nonparametric likelihood, Random walk, Smooth function model, Spectral density, Strong mixing, Time series regression, Weak dependence

1. Introduction

The method of empirical likelihood, introduced by Owen (1988), is a technique which has many parallels with the bootstrap. Both are based on nonparametric likelihood; while the bootstrap assigns $1/N$ probability mass to each observation, the empirical likelihood method "chooses" probability mass under linear constraints. The former uses simulations, while the latter uses numerical calculation to obtain confidence intervals. These confidence intervals calculated by the two methods share similar properties. In fact, as Hall (1992, p.161) puts it, empirical likelihood provides confidence regions "that have coverage accuracy properties at least comparable with those of bootstrap confidence regions". Efron and Tibshirani (1993) provide a nice discussion on the two methods. Chen (1994a) compares the power of the two methods in the context of mean parameter tests in terms of higher order asymptotics.

Empirical likelihood has been studied extensively in the literature because of its generality and effectiveness. It has numerous applications: smooth function models, regression models (Owen, 1991, Chen, 1993, 1994a), generalized linear models (Kolaczyk, 1992), quantiles (Chen and Hall, 1993), biased sample models (Qin, 1993) general estimating equations (GEE) (Qin and Lawless, 1994), to name a few. Recent studies suggest desirable properties of empirical likelihood; see DiCiccio, Hall and Romano (1991 a,b), DiCiccio and Romano (1990) and Hall (1990), for example. The empirical likelihood ratio statistic has much in common with its conventional parametric counterpart. In particular, it has a chi-square limiting distribution as in Wilk's theorem. Furthermore, its confidence interval is Bartlett correctable, thus the coverage error can be reduced to the order of $O(N^{-2})$.

It should be noted, however, the existing literature seems to focus on independent data generating processes. If one wishes to use empirical likelihood for general stationary time series, it seems that a new device is called for. To realize the problem of empirical likelihood in a dependent data setting, consider the following simple example. Suppose the researcher's parameter of interest is the mean θ_0 of identically distributed random variables X_t, $t = 1,...,N$. Treating X_t *as if they are independent*, the empirical likelihood for θ is the value of the likelihood of the multinomial distribution $\Pi_{i=1}^{N} p_i$ maximized under the constraints $\Sigma_t p_t = 1$ and

$\Sigma_t p_t X_t = \theta$. Let $L(\theta)$ denote the maximum value. Without the second constraint, clearly the likelihood is maximized at the empirical distribution $p_i = 1/N$ for all i, thereby yielding the estimate $\bar{X} = N^{-1}\Sigma X_t$. Writing this $L(\bar{X})$, the empirical likelihood ratio is

$$R(\theta) = L(\theta)/L(\bar{X}).$$

Under mild regularity conditions, the following approximation result can be shown to hold for possibly dependent processes:

$$-2\log R(\theta_0) = N(\bar{X} - \theta_0)'\bar{\Sigma}^{-1}(\bar{X} - \theta_0) + o_p(1),$$

where $\bar{\Sigma} = N^{-1}\Sigma_t(X_t - \theta_0)(X_t - \theta_0)'$. If in fact X_t is iid, clearly the above likelihood ratio statistic is asymptotically chi-squared distributed. If, however, X_t is a dependent (and stationary) process, the matrix $\bar{\Sigma}$ provides a "wrong metric" for the score $\sqrt{N}(\bar{X} - \theta_0)$; $\bar{\Sigma}$ converges to $\mathrm{Var}(X_t)$ in probability, instead of the desired term $\Sigma_{-\infty}^{\infty}\mathrm{Cov}(X_t, X_{t-j})$. In this case the empirical likelihood method fails.

Obviously this failure occurs because the empirical likelihood was constructed ignoring the dependence structure of the data. In this aspect, again there is a similarity between the empirical likelihood and the bootstrap. Since the remark by Singh (1981), it has been recognized that independent resampling generally leads to results which are not consistent if dependence is present in the data series. One possible remedy is to fit a parametric model (typically ARMA model) so that the transformed innovations becomes iid; such a parameterization is often too restrictive, and the result can be quite sensitive to the specification of the unknown data dependence structure. (In fact, the problem of dependent processes in empirical likelihood can be treated in the same way, leading to the same problem.) Thus the recent studies of the bootstrap in stationary time series mainly utilize the "blockwise resampling" (see, Hall, 1985, Carlstein, 1986 and Künsch, 1989), which preserves the dependence property of the data nonparametrically by appropriately choosing blocks of data. Note that Carlstein's blocking scheme is "non-overlapping", while Künsch proposes "overlapping" blocking.

Observing the close connection between the empirical likelihood and the bootstrap, one

might conjecture that the blocking technique could be effectively adapted to the method of empirical likelihood. In this paper we show that this conjecture is correct. Empirical likelihood of blocks of observations, not observations themselves, is proposed. We shall call it the method of blockwise empirical likelihood. Blocks can be overlapped or non-overlapped, again in the same manner as in the blockwise bootstrap technique. The sample blocking allows us to treat (weak) dependence of time series in a nonparametric way.

The paper is organized as follows. In Section 2 some basic concepts that will be repeatedly used throughout this paper are laid out. Section 3 considers the empirical likelihood for GEE, originally analyzed by Qin and Lawless (1994) in iid framework, and extends their results to blockwise empirical likelihood with weakly dependent processes. This model is chosen because of its extreme generality. The weak consistency and the asymptotic normality of the maximum blockwise empirical likelihood estimator are proved and various likelihood ratios are shown to be asymptotically chi-squared approximable. In Section 4 the smooth function model is discussed. Using this model, we first show that we can conduct empirical likelihood-based inference regarding parameters of the infinite dimensional joint distribution of the data series by using "blocks of blocks" techniques. This demonstrates the applicability of empirical likelihood to the spectral density estimation or tests of random walks (see Section 4.4). Then we show that the blockwise empirical likelihood ratio statistic is Bartlett correctable. The coverage error of the blockwise empirical likelihood-based confidence intervals can be improved up to the order of $O(N^{-4/3})$. Though this rate is not as fast as the rate of N^{-2} obtained by DiCiccio, Hall and Romano (1991a) for iid cases, this result demonstrates that the (Bartlett-corrected) empirical likelihood with blocking is a quite powerful and accurate method. Section 5 considers two alternative measures of distance, and a Euclidean-likelihood interpretation of Hansen's (1982) generalized method of moments (GMM) is suggested. Section 6 offers some conclusions. Proofs of theorems are included in Appendix.

2. Weak Dependence and Data Blocking

In this section we state some important concepts that are used throughout the subsequent development. In this paper we allow for weakly dependent processes; in particular, we consider

the following form of dependence. Throughout the paper, $\{X_i\}$ denotes a $\mathbb{R}^d$-valued stationary stochastic process, satisfying the following strong mixing condition:

$$\alpha_X(k) \longrightarrow 0 \quad (k \longrightarrow \infty),$$

where $\alpha_X(k) = \sup_{A,B} |\boldsymbol{P}(A \cap B) - \boldsymbol{P}(A)\boldsymbol{P}(B)|$, $A \in \mathscr{F}_{-\infty}^{0}$, $B \in \mathscr{F}_{k}^{\infty}$ and $\mathscr{F}_{m}^{n} = \sigma(X_i: m \le i \le n)$. Further, we assume $\sum_{k=1}^{\infty} \alpha_X(k)^{1-1/c} < \infty$ for some constant $c > 1$.

As suggested in the introduction, we deal with weakly dependent processes by choosing blocks of data so that the information regarding the dependence property is captured. In what follows we state our general blocking scheme. Since the idea is close to the data blocking for bootstrapping, in particular the method by Politis and Romano (1992), we follow some of their notation. Let M and L be integers that depend on N, where $M \to \infty$, $M = o(N^{1/2}), L = O(M)$ as $N \to \infty$, and $L \le M$. We let B_i, $i \in \mathbb{N}$ be a vector of M consecutive observations $(X_{(i-1)L+1}, ..., X_{(i-1)L+M})$. Note that M is the "window width", whereas L is the index for the block's time shift. Also define $Q = [(N-M)/L]+1$, where $[\cdot]$ is the integer part of $\cdot$; this is the number of blocks taken from the observations. We further consider mapping each block by a function ϕ_M and define "observations" $T_i = \phi_M(B_i)$; we discuss details regarding the T_i in later sections. Let $A_N = QM/N$, then the number of overlaps for each typical observation is $A_N - 1$. Our blocking scheme is "increasingly-overlapping" if $L = o(M)$ and "constant-overlapping" if $L = M/A$, where the constant $A \ge 1$. In the case where $A = 1$, the blocks are non-overlapping. As we shall see, for the first order asymptotics of (blockwise) empirical likelihood for parameters of finite dimensional distributions (see Section 3), the "increasing-overlap" scheme is allowed. To obtain the asymptotic validity of the empirical likelihood for parameters of infinite-dimensional distributions, however, it seems that the "constant-overlap" scheme is required.

We introduce a more general blocking scheme to deal with inference regarding parameters of the infinite dimensional joint distribution. In such a case, "blocks of blocks" successfully capture the dependence in the triangular array $\{T_i\}$ (recall that T_i also depends on M and L,

which in turn depend on N). This idea of the use of "blocks of blocks" is originally developed and investigated by Politis and Romano (1992, 1993) in the context of resampling methods; here we propose to apply such a technique to empirical likelihood. Our s-th "block of blocks" is defined as $\beta_s = (B_{(s-1)h+1},...,B_{(s-1)h+b})$. Note that B_i, b and h play the same roles as X_t, M and L in the definition of B_i; β_s is the s-th block of "observations" T_i. The block (-of-blocks) width b and the index of time shift h are dependent on Q, or N. Also we define the number of "blocks of blocks" $q = [(Q-b)/h] + 1$. Finally, define $a_N = qb/Q$ so that the number of overlaps for a typical block is $a_N - 1$.

3. General Estimating Equations (GEE)

3.1 Overview

Recently Qin and Lawless (1994) considered the application of empirical likelihood to general estimating equations (GEE) with iid data. They allow for a situation in which the number of estimating equations may exceed the number of parameters; such models, often said to be "overidentified", are typical in econometric applications (see, for example, Hansen and Singleton, 1982). They showed that the maximum empirical likelihood estimator (MELE) is asymptotically efficient assuming iid samples. Qin and Lawless also proposed statistics based on empirical likelihood to test parameter restrictions and "overidentifying restrictions" (i.e., whether the moment condition holds or not). In both cases the statistic converges to a chi-square distribution, where the degrees of freedom are equal to the number of restrictions (in the former case) or the number of overidentifying restrictions (in the latter case). In this section we consider the same model, but allowing for weakly dependent data generating processes.

When the underlying processes are dependent, the original MELE method is indeed consistent under regularity conditions, but not efficient if the model is "overidentified". More importantly, the empirical likelihood ratio statistics as originally defined generally are not asymptotically chi-squared distributed. (This is true for other empirical likelihood ratio statistics; see Section 1). We solve these problems by using blockwise empirical likelihood.

3.2 Blockwise Empirical Likelihood for GEE

Consider the estimating function f: $\mathbb{R}^d \times \mathbb{R}^p \to \mathbb{R}^r$, which is assumed to satisfy the moment condition:

(3.1) $\boldsymbol{E} f(X_t,\theta_0) = 0$

where $\theta_0 \in \Theta \subset \mathbb{R}^p$ is the true parameter and $r \geq p$. Instead of considering the empirical likelihood of the original estimating functions, we use the function of an observation block $T_i(\theta) = \phi_M(B_i,\theta)$, where B_i is the i-th block of observations as defined in Section 2 and the mapping ϕ_M: $\mathbb{R}^{rM} \times \mathbb{R}^p \to \mathbb{R}^r$ has the following particular form:

(3.2) $\phi_M(B_i,\theta) = \sum_{n=1}^{M} f(X_{(i-1)L+n},\theta)/M$.

Though more flexibility would be obtained by considering M-moving averages with various weights, for simplicity we shall focus on the equally weighted sum as defined above. The profile blockwise empirical likelihood function is

(3.3) $L(\theta) = \sup\left\{\prod_{i=1}^{Q} p_i \,|\, p_i > 0,\ \sum_1^Q p_i = 1,\ \sum_1^Q p_i T_i(\theta) = 0\right\}$.

Note that the dependence of p_i on M and L is suppressed for notational convenience. As in Qin and Lawless (1994) this optimization problem is solved by considering a Lagrangean with multipliers λ and $\gamma = (\gamma_1,...,\gamma_r)'$:

(3.4) $\mathcal{L} = \sum_{i=1}^{Q} \log(p_i) + \lambda\left(1 - \sum_1^Q p_i\right) - Q\gamma' \sum_1^Q p_i T_i(\theta)$.

From the first order conditions (see Qin and Lawless, 1994) it is easily seen that the profile blockwise empirical likelihood (3.3) is rewritten as

(3.5) $L(\theta) = \prod_{i=1}^{Q} \{(1/Q)1/\{1 + \gamma_N(\theta)' T_i(\theta)\}\}$,

where $\gamma_N(\theta)$ is a stationary point of the function $q(\gamma) = -\sum_1^Q \log(1 + \gamma' T_i(\theta))$. (If the conditioning set for (3.3) is empty, we simply let $L(\theta) = -\infty$.) If we further assume that $\sum T_i(\theta) T_i(\theta)'$ is of full rank, $q(\cdot)$ is shown to be strictly convex, implying that $\gamma_N(\cdot)$ is a continuously differentiable function. (See Owen, 1990, and Qin and Lawless, 1994, for these and

other basic properties of empirical likelihood.) We define the maximum blockwise empirical likelihood estimator as the maximizer of (3.5), which is henceforth denoted by $\hat{\theta}$.

We can also define various useful statistics using blockwise empirical likelihood. Let L_{F_Q} denote the empirical likelihood without constraints, i.e., empirical likelihood evaluated at the empirical M-dimensional marginal $F_Q = Q^{-1}\sum_1^Q \delta_{B_i}$ where δ_x denotes the point mass at $x \in \mathbb{R}^M$. The blockwise empirical log likelihood ratio is then defined as

$$-2\log(L(\theta)/L_{F_Q}) = 2\sum_1^Q \log(1 + \gamma_N(\theta)' T_i(\theta)),$$

which could be used to construct a likelihood ratio statistic to test the moment condition (3.1):

(3.6) $$LR_0 = 2A_N^{-1}\sum_1^Q \log(1 + \gamma_N(\hat{\theta})' T_i(\hat{\theta})) .$$

Notice the presence of the factor A_N^{-1}, which is necessary to obtain the proper chi-square asymptotics (see the next section). Intuitively, this is an adjustment factor for the effect of overlapped use of observation, and it disappears when there is no overlap.

Next, suppose we are interested in the following parametric hypothesis,

(3.7) $$H_0: \Psi(\theta_0) = \psi,$$

where $\psi \in \mathbb{R}^q$, $q \leq p$ and $\Delta = \partial\Psi/\partial\theta' \big|_{\theta_0}$ is of full row rank. Now we maximize the profile likelihood (3.3) under the above constraint to obtain $\hat{\theta}^c$ and $L(\hat{\theta}^c)$. In exactly the same manner as in the likelihood ratio statistic for a conventional fully parametric model, we let

(3.8) $$LR_1 = -2A_N^{-1}\log(L(\hat{\theta}^c)/L(\hat{\theta})),$$

in order to test the parametric hypothesis (3.7).

Note that LR_0 is a blockwise version of Qin and Lawless' (1994) W_1; as we shall see immediately, LR_0 is asymptotically chi-squared distributed due to the blocking technique. The second statistics LR_1 is also similar to W_2 in Qin and Lawless (1994), though our hypothesis is slightly more general than theirs, which restricts (a subset of) the parameters to be equal to some hypothesized values.

3.3 Asymptotic Results

Qin and Lawless (1994) showed the asymptotic normality of MELE for GEE in the $N^{-1/3}$ neighborhood of θ_0, in an iid setting. We prove the consistency and the asymptotic normality of our blockwise version of MELE with weakly dependent processes. In the proof we show the weak consistency of $\hat{\theta}$, utilizing the classical argument developed by Wald (1949) and Wolfowitz (1949).

Let $g(\theta)$ be the minimizer of $\boldsymbol{E}[-\log(1 + g'f(X_t,\theta))]$, and $\Gamma(\theta,\delta)$ be an open sphere with center θ and radius δ. $g(\cdot)$ is, loosely speaking, the population version of $\gamma(\cdot)$ (with $L = M = 1$). $\|\cdot\|$ denotes the Euclidean norm.

THEOREM 1 *Assume*

(i) *The parameter space* Θ *is compact.*

(ii) θ_0 *is the unique root of* (3.1).

(iii) *For sufficiently small* δ, $\boldsymbol{E} \sup_{\theta^* \in \Gamma(\theta,\delta)} -\log(1 + g(\theta^*)'f(x,\theta^*)) < \infty$.

(iv) *If a sequence* θ_j, $j = 1,2,\ldots$ *converges to some* $\theta \in \Theta$ *as* $j \to \infty$, $f(x,\theta_j)$ *converges to* $f(x,\theta)$ *for all* x *on a null set which may vary with* θ.

(v) θ_0 *is an interior point of* Θ *and* $f(x,\theta)$ *is twice continuously differentiable at* θ_0.

(vi) $\mathrm{Var}(N^{-1/2}\Sigma_{t=1}^{N} f(X_t,\theta_0)) \to S > 0$ $(N \to \infty)$

(vii) $\boldsymbol{E}\|f(x,\theta_0)\|^{2c} < \infty$ *for* $c > 1$ *defined in Section 2,* $\boldsymbol{E} \sup_{\theta^* \in \Gamma(\theta_0,\delta)} \|f(x,\theta^*)\|^{2+\varepsilon} < K$, $M = o(N^{1/2 - 1/(2+\varepsilon)})$ *for some* $\varepsilon > 0$, $\boldsymbol{E} \sup_{\theta^* \in \Gamma(\theta_0,\delta)} \|\partial f(x,\theta^*)/\partial\theta\|^2 < K$ *and* $\boldsymbol{E} \sup_{\theta^* \in \Gamma(\theta_0,\delta)} \|\partial^2 f_j(x,\theta^*)/\partial\theta\partial\theta'\| < K$ *for all* j, *where* $K < \infty$ *and* $f_j(x,\theta)$ *denotes the j-th element of* $f(x,\theta)$.

(viii) $D = \boldsymbol{E}\partial f(x,\theta_0)/\partial\theta'$ *is of full column rank.*

Then

$$\begin{pmatrix} N^{1/2}(\hat{\theta} - \theta_0) \\ M^{-1}N^{1/2}(\gamma_N(\hat{\theta}) - 0) \end{pmatrix} \Rightarrow_d N\left(0, \begin{pmatrix} V_\theta & 0 \\ 0 & V_\gamma \end{pmatrix}\right)$$

where $V_\theta = (D'S^{-1}D)^{-1}$ *and* $V_\gamma = S^{-1}\{I - DV_\theta D'S^{-1}\}$.

Remark: The assumptions made here are by no means the weakest possible. For example, assumption (i) could be replaced with a weaker condition at the expense of greater complexity.

It is easy to check that asymptotic normality holds for Qin and Lawless' maximum empirical likelihood estimator without blocking; however, its asymptotic covariance matrix has the form

$$(D'W^{-1}D)^{-1}D'W^{-1}SW^{-1}D(D'W^{-1}D)^{-1}$$

where $W = \boldsymbol{E}f(x,\theta_0)f(x,\theta_0)'$. Therefore such an estimator is asymptotically less efficient than $\hat{\theta}$. In fact, our V_θ coincides with the asymptotic covariance matrix of the GMM estimator with optimal weighting (see, Hansen, 1982), defined as the minimizer of

(3.5) $\bar{f}(\theta)'\hat{S}^{-1}\bar{f}(\theta)$,

where $\bar{f}(\theta) = N^{-1}\Sigma f(x_t,\theta)$ and $\hat{S}$ is a consistent estimator of S. Note that our blockwise MELE $\hat{\theta}$ is, to the first order approximation, the solution to the first order condition

$$\bar{D}'\bar{S}\bar{f}(\theta) = o_p(1/\sqrt{N}),$$

where $\bar{D} = N^{-1}\Sigma\partial f(x_t,\theta_0)/\partial\theta$ and $\bar{S} = Q^{-1}M\Sigma T_i(\theta_0)T_i(\theta_0)'$. When $L = 1$ (the "fully overlapped" case), the matrix $\bar{S}$ is a nonparametric estimator of S (which is the spectral density matrix of $f(x_t,\theta_0)$ at the origin) with the Bartlett kernel and the truncation parameter M; if we allow for various weighting patterns in the definition of ϕ_M, we would obtain other kernel estimators.

These estimators are frequently used to calculate $\hat{S}$ of (3.5) in applications of GMM. If $L \to \infty$ as $N \to \infty$, $\bar{S}$ corresponds to the spectral density estimator using the time-averaged subsample periodogram, which is extensively studied by Zhurbenko (1979, 1986); see also Welch (1967).

THEOREM 2: *Suppose all the assumptions in theorem 1 hold. Then*

(a) $LR_0 \Longrightarrow_d \chi^2_{r-p}$.

(b) *Under* H_0, $LR_1 \Longrightarrow_d \chi^2_q$.

Remark: Though our main concern is likelihood ratio-type statistics, it is possible to consider other test statistics that are asymptotically chi-squared distributed. Define, for example, the Wald and Lagrange Multiplier (LM)-type statistics:

$$\text{Wald} = N(\Psi(\hat{\theta}) - \psi)'(\hat{\Delta}\hat{V}_\theta\hat{\Delta}')^{-1}(\Psi(\hat{\theta}) - \psi)$$

$$LM = M^{-2}N\gamma_N(\hat{\theta}^c)'\hat{D}\hat{V}_\theta\hat{D}'\gamma_N(\hat{\theta}^c),$$

where $\hat{\Delta}$, $\hat{D}$ and $\hat{V}_\theta$ are consistent estimates of Δ, D, and V_θ; these can be obtained by using the unconstrained estimator $\hat{\theta}$ (Wald statistic) or the constrained $\hat{\theta}^c$ (LM statistic) in place of the unknown θ_0. It can be shown that these statistics have the same chi-square limiting distribution with q-degrees of freedom in the same manner as the results in Theorem 2.

4. Smooth Function Model

4.1 Overview

In many statistical models, the parameter of interest θ can be written as a smooth function of the mean μ of the distribution of observations. Such a model is called the smooth function model and studied by Bhattacharya and Ghosh (1978). They showed the validity of the formal Edgeworth expansions in such a model if observations are iid. The smooth function model in weakly dependent processes is also considered, especially in the blockwise

bootstrapping literature (see, for example, Götze and Künsch, 1993, or Lahiri, 1992). The empirical likelihood for the smooth function model was considered by DiCiccio, Hall and Romano (1991) in iid setups, who derived higher order expansion and gave a formula for the Bartlett correction. This Bartlett correctability of empirical likelihood is rather remarkable, implying that the coverage error of the empirical likelihood-based confidence interval can be improved to the order of $O(N^{-2})$.

As noted by DiCiccio, Hall and Romano, the smooth function models are quite general; they could be used, for example, for the inference about (ratios of) various parameters of finite dimensional distributions (e.g., means, variances, skewness). In this section, we first consider the extension of the existing literature of smooth function models and empirical likelihood to inference regarding parameters of infinite dimensional joint distributions. Note that, in the literature of the bootstrap, such a problem was solved by Politis and Romano (1992) in their seminal work on the blocks-of-blocks bootstrap. Here we propose "blocks-of-blocks empirical likelihood" to deal with parameters of the infinite dimensional joint distribution. The leading example of applications is inference about spectra; some concrete examples will be given later. Second, using the smooth function model, we further show that blocks-of-blocks empirical likelihood for weakly dependent processes is Bartlett correctable if additional conditions are satisfied.

4.2 Blockwise Empirical Likelihood for the Smooth Function Model

Consider the smooth function model $\theta = H(\mu)$, where $\theta \in \mathbb{R}^p$ is the parameter of interest, $\mu \in \mathbb{R}^r$ $(p \leq r)$ is a parameter of the (finite or infinite dimensional joint) distribution of $\{X_t\}_{t\in\mathbb{N}}$. $H\colon \mathbb{R}^r \to \mathbb{R}^p$ is continuously differentiable and $\text{rank}(\partial H/\partial \mu' |_{\mu_o}) = p$. As in Politis and Romano (1992), let $\phi_M\colon \mathbb{R}^{lM} \to \mathbb{R}^r$ denote the mapping of blocks of observations and define the "estimating function" $T_i = \phi_M(B_i)$. Let $\bar{T} = \sum_{i=1}^{Q} T_i/Q$, which is, loosely speaking, the sample counterpart of μ. If the parameter of interest is the spectral density of $\{X_t\}_{t\in\mathbb{N}}$, we would let $\phi_M(B_i)$ be the sample periodogram calculated from the subsample B_i.

In cases where the parameter of interest is a smooth function of parameters of the finite dimensional distributions of the data, it suffices to consider the blockwise empirical likelihood for θ:

(4.1) $\quad \sup\left\{\Pi_{i=1}^{Q} p_i \,|\, p_i > 0,\ \Sigma_1^Q p_i = 1,\ H(\Sigma_1^Q p_i T_i) = \theta\right\}.$

However, in order to cover the inference of the infinite dimensional joint distribution, we need to use the "blocks of blocks" technique. Let Φ_b: $\mathbb{R}^{rb} \to \mathbb{R}^r$ be a mapping such that $\Phi_b(\beta_s) = \Sigma_{i=1}^{b} T_{(s-1)h+i}/b$, where the notation introduced in Section 2 is used. We then use the (double) array of new observations $U_s = \Phi_b(\beta_s)$ to construct the "blocks of blocks" empirical likelihood:

(4.2) $\quad L(\theta) = \sup\left\{\Pi_{s=1}^{q} p_s \,|\, p_s > 0,\ \Sigma_1^q p_s = 1,\ H(\Sigma_1^q p_s U_s) = \theta\right\}.$

Note U_s implicitly depends on N. To deal with a parameter of a finite (m-) dimensional marginal in this framework, simply let $M = m$ and $L = 1$; the rest of our theory remains valid. The maximum value of the empirical likelihood function without restriction is q^{-q} at $\hat{\theta} = H(\overline{U})$, where $\overline{U} = \Sigma_{s=1}^{q} U_s/q$. Hence the blockwise empirical likelihood ratio statistic is

(4.3) $\quad LR(\theta) = -2a_N^{-1}\log(L(\theta)/L(\hat{\theta})) = -2a_N^{-1}\log(q^q L(\theta))$.

The factor $a_N^{-1} = Q/qb$ adjusts the effect of overlaps in blocks; see (3.6) and (3.8). The statistic (4.3) can be used to test the hypothesis H_0: $\theta = \theta_0 = H(\mu_0)$ using asymptotic chi-square criteria, as the next theorem implies. It should be noted that some of the assumptions made in the theorem are essentially the same as assumptions used in Politis and Romano (1992).

THEOREM 3 *Assume*

(i) $\quad L = A^{-1}M$ *for some* $A \geq 1$.

(ii) $\quad b \longrightarrow \infty$, $h = O(b)$ *and* $b = o(Q^{1/2})$.

(iii) $\quad \boldsymbol{E}\|T_i\|^{2c} < K$ *for* $c > 1$ *defined in Section 2, some* $K < \infty$ *and all* M.

(iv) $\boldsymbol{E}T_i = \mu_0 + o(Q^{-1/2})$.

(v) $\mathrm{Var}(\sqrt{Q}\,\bar{T}) \longrightarrow \Lambda$ *as* $Q \longrightarrow \infty$ $(N \longrightarrow \infty)$.

Then

$$LR(\theta_0) \longrightarrow_d \chi_p^2.$$

Remark: We need the "constant-overlapping" scheme (Assumption (i)) to ensure the strong mixing properties of $\{T_i\}$. This condition is automatically satisfied when we construct the empirical likelihood for parameters of finite (m) dimensional marginals, with $M = m, L = 1$ and $A = m$. Assumption (i) is also essentially important to show the Bartlett-correctability in a weakly dependent framework.

The extension of the above to the (homogeneous mixing) random field (see, Rosenblatt, 1985, for example) might be of interest. In such cases, we need to take rectangles of observations, instead of the blocks used in the time series case. Bootstrapping for the random field using blocking techniques has been studied in the literature (see Politis and Romano, 1993, and the papers cited therein) and the consistency of such techniques has been proved. Such bootstrapping methods are basically an extension of the blockwise bootstrapping for weakly dependent time series. The blockwise empirical likelihood for the smooth function models may be generalized in a similar fashion.

4.3 Bartlett Correction

In iid settings, the Bartlett correction of the empirical likelihood ratio statistic in the smooth function models was developed by DiCiccio, Hall and Romano (1991). It should be noted that Mykland (1995) showed a general Bartlett-correctability result using the concept "dual likelihood". As will be shown below, the empirical likelihood for the smooth function model with weakly dependent observations is Bartlett correctable if a particular data blocking scheme is used and additional regularity conditions are satisfied.

This section studies blockwise empirical likelihood for the smooth function models introduced in Section 4.1; see Eq. (4.1). In particular, we consider the non-overlapping blocking

method as Carlstein's (1986) with $M/N^{1/3} \to C$, $0<C<\infty$, as $N \to \infty$. Though more flexibility could be allowed, here we limit ourselves to a discussion of this blocking method, which simplifies our proof.

We define the Bartlett correction factor for the blockwise empirical likelihood ratio, modifying the formula for the Bartlett factor derived by DiCiccio, Hall and Romano (1991). Let $\Sigma_N = \mathrm{Var}(N^{-1/2}\sum_{t=1}^{N} X_t)$ and $T_i = (T_i^1,...,T_i^r)'$. For a sequence of d integers satisfying $0<k(1)<\cdots<k(d)=k$, $k \geq 3$, define

$$\begin{aligned}\tilde{\kappa}^{j_1\cdots j_{k(1)},j_{k(1)+1}\cdots j_{k(2)},j_{k(2)+1}\cdots j_{k(d-1)},j_{k(d-1)+1}\cdots j_{k(d)}} \\ = Q^{-1} \sum_{1\leq i(1),\cdots,i(d)\leq Q} \boldsymbol{E}\Big\{M^{-1}(M^{k(1)}T_{i(1)}^{j_1}\cdots T_{i(1)}^{j_{k(1)}})(M^{k(2)}T_{i(2)}^{j_{k(1)+1}}\cdots T_{i(2)}^{j_{k(2)}})\cdots(M^{k(d)}T_{i(d)}^{j_{k(d-1)+1}}\cdots T_{i(d)}^{j_{k(d)}})\Big\} \\ \cdot I\{\max_{p,q<d}|i(p) - i(q)| \leq k-2\},\end{aligned} \tag{4.4}$$

where $I\{\cdot\}$ denotes the indicator function. For the special case where $d = 1$, we sometimes use the notation

$$\begin{aligned}\kappa^{j_1\cdots j_k} &= Q^{-1}\sum_{i=1}^{Q}\boldsymbol{E}(M^{k-1}T_i^{j_1}\cdots T_i^{j_k}) \\ &= \boldsymbol{E}(M^{k-1}T_i^{j_1}\cdots T_i^{j_k}).\end{aligned}$$

Our Bartlett factor is

$$a = p^{-1}(2t_{1a} + 2t_{1b} + t_{1c} + 2t_{2a} + t_{2b} + 2t_{3a} + 2t_{3b} + t_{3c} + 2t_{4a} + 2t_{4b} + t_5 + 2t_{6a} + t_{6b}),$$

where

$$\mu = (\mu^1,...,\mu^r)',\ \bar{\mu}_0 = \Sigma_N^{-1/2}\mu_0,\ \bar{H}(\mu) = H(\Sigma_N^{1/2}\mu),\ \bar{H}_{j_1\cdots j_k}^{l} = \partial^k \bar{H}^l(\mu)/\partial\mu^{j_1}\cdots\partial\mu^{j_k}\big|_{\mu=\bar{\mu}_0},$$

$$\nabla\bar{H} = (\bar{H}_j^i),\ G = (\nabla\bar{H}\nabla\bar{H}')^{-1},\ W = \nabla\bar{H}'G\nabla\bar{H},\ N = \nabla\bar{H}'G,$$

$$t_{1a} = (1/3)\kappa^{jkl}\tilde{\kappa}^{m,n,o}W^{jo}W^{km}W^{ln},$$

$$\begin{aligned}t_{1b} &= (3/8)\tilde{\kappa}^{jk,l}\tilde{\kappa}^{nm,o}W^{jo}W^{km}W^{ln} - (5/6)\kappa^{jkl}\tilde{\kappa}^{mn,o}(W^{jm}W^{kn}W^{lo} + W^{jo}W^{km}W^{ln}) \\ &\quad + (8/9)\kappa^{jkl}\kappa^{mno}W^{jm}W^{kn}W^{lo},\end{aligned}$$

$$t_{1c} = (1/4)\kappa^{jkl}\tilde{\kappa}^{mn,o}W^{jm}W^{ko}W^{\ln} + \left\{(-2/3)\kappa^{jkl}\tilde{\kappa}^{mn,o} + (2/9)\kappa^{jkl}\kappa^{mno}\right\}W^{jm}W^{kn}W^{lo},$$

$$t_{2a} = (3/8)\tilde{\kappa}^{jk,l}\tilde{\kappa}^{mn,o}W^{jl}W^{km}W^{no} + \left\{(-5/6)\kappa^{jkl}\tilde{\kappa}^{mn,o} + (4/9)\kappa^{jkl}\kappa^{mno}\right\}W^{jk}W^{lm}W^{no},$$

$$t_{2b} = \left\{(1/4)\tilde{\kappa}^{jk,l}\tilde{\kappa}^{mn,o} - (1/3)\kappa^{jkl}\tilde{\kappa}^{mn,o}\right\}W^{jm}W^{kl}W^{no} + (1/9)\kappa^{jkl}\kappa^{mno}W^{jm}W^{kl}W^{no},$$

$$t_{3a} = (-1/2)\tilde{\kappa}^{jk,l,m}W^{jm}W^{kl},\ t_{3b} = \left\{(3/8)\tilde{\kappa}^{jl,km} + \tilde{\kappa}^{jkl,m} - (3/4)\kappa^{jklm}\right\}W^{jk}W^{lm},$$

$$t_{3c} = (1/4)\tilde{\kappa}^{jk,lm}W^{jl}W^{km},\ t_{4a} = (1/2)\kappa^{jkl}N^{ju}\bar{H}^{u}_{mn}(I - W)^{mk}(I - W)^{nl},$$

$$t_{4b} = -\tilde{\kappa}^{jk,l}N^{ju}\bar{H}^{u}_{mn}(I - W)^{mk}(I - W)^{nl},$$

$$t_{5} = (1/4)G^{uv}\bar{H}^{u}_{jk}\bar{H}^{v}_{lm}\left\{(I - W)^{jk}(I - W)^{lm} + 2(I - W)^{jl}(I - W)^{km}\right\},$$

$$t_{6a} = \left\{(-1/4)\tilde{\kappa}^{jk,l} + (1/3)\kappa^{jkl}\right\}N^{ju}\bar{H}^{u}_{mn}(I - W)^{mn}W^{kl},$$

$$t_{6b} = \left\{-(1/2)\tilde{\kappa}^{jk,l} + (1/3)\kappa^{jkl}\right\}N^{ju}\bar{H}^{u}_{mn}(I - W)^{mn}W^{kl}.$$

Repeated subscripts are used to denote summations as the conventional notation. Note that the coefficients defined above depend on M and Q (or N, in general).

In addition to the assumptions made in Theorem 3, in the rest of this section, we assume the validity of the Edgeworth expansions that are required to show the coverage error results stated below. Götze and Hipp (1983) showed the validity of Edgeworth expansion for sums of dependent processes assuming (i) the existence of sufficiently many moments, (ii) a conditional Cramer condition and (iii) the random processes are approximated by other exponentially strong mixing processes that satisfies a Markov type condition. Note that we assume the validity of Edgeworth expansions for sums of (strong mixing) blocks of data. Davison and Hall (1993) used an Edgeworth expansion for sums of data blocks to analyze the bootstrap of Studentized statistics with dependent processes.

In the derivation of our coverage error results, we assume that $\alpha_X(m) \leq ce^{-dm}$ for all m, where c and d are positive constants. This can be relaxed, since Götze and Hipp's (1983)

results only require that the observations are L_1-approximable by some "base" random sequence that has exponentially decaying mixing coefficients. In fact, Theorems 1-3 also can be proved to hold under somewhat weaker conditions, under which observations are approximated by some mixing processes. Such conditions are introduced by Ibragimov (1962) and Billingsley (1968); various laws of large numbers and central limit theorems are available for such processes.

In Appendix we shall show that

(4.5a) $\boldsymbol{P}\{LR(\theta_0) \le z\} = \boldsymbol{P}\{\chi_p^2 \le z\} + O(N^{-2/3})$.

Moreover,

(4.5b) $\boldsymbol{P}\{LR(\theta_0)(1 - N^{-1}a) \le z\} = \boldsymbol{P}\{\chi_p^2 \le z\} + O(N^{-4/3})$.

That is, the blockwise empirical likelihood ratio statistic is Bartlett correctable. The coverage error of confidence intervals are improved up to the order of $O(N^{-4/3})$. This rate is not as fast as the rate of N^{-2} obtained for the standard empirical likelihood assuming iid samples (DiCiccio, Hall and Romano, 1991a), since our nonparametric treatment of dependence slows it down. (A similar phenomenon is observed for the blockwise bootstrap; see, for example, Götze and Künsch, 1993.) Nevertheless, these results demonstrate that the (Bartlett-corrected) empirical likelihood with blocking is a quite powerful and accurate method. It would be possible to extend the above results to blocks-of-blocks empirical likelihood (4.2). In this case we would replace $\{T_i\}$ and $\{X_t\}$ with $\{U_s\}$ and $\{T_i\}$.

In practice the Bartlett factor a needs to be estimated; this could be done by replacing unknown population parameters with their consistent estimates using sample moments of T_i, or by the bootstrap. This replacement does not affect the conclusion of the above theorem. This is a consequence of general parity properties of polynomials in Edgeworth expansions (e.g., Barndorff-Nielsen and Hall, 1988), as pointed out by Hall and La Scala (1990).

4.4 Some Examples

In the preceding sections, we have seen the effectiveness of the blockwise empirical likelihood for the smooth function model. Examples of applications of the technique include

inference regarding the spectral density. There are also some interesting econometric applications, which could be viewed as variants of spectral analysis. In macroeconomics or financial economics, often the researcher is interested in the analysis of the "long-run" or "permanent" behavior of economic data series, as opposed to "short run" or "temporary" fluctuations. Therefore it is natural to consider the spectral density at zero to analyze the long-run components of time series.

In what follows we discuss three examples; the first example deals with a parameter of a finite dimensional distribution, and the second is concerned with parameters of infinite dimensional joint distributions. The third addresses statistics involving parameters of both finite and infinite dimensional distributions.

EXAMPLE 1 (Time Series Regression)

The empirical likelihood-based inference for linear regression models (or more precisely, correlation models) were considered by Owen (1991) in independent settings. We can extend his results to time series regression with weakly dependent processes by using blockwise empirical likelihood for the smooth function model proposed in Section 4. Note that, in the conventional OLS-based inference, we need to estimate $\sum_{j=-\infty}^{\infty}\mathrm{Cov}(X_0X_j\varepsilon_0\varepsilon_j)$ to construct valid confidence interval when the data is dependent (see Eicker, 1967). For the empirical likelihood analysis, the blocking technique is necessary.

This type of regression is important in certain econometric applications where GLS may not be applicable due to the data structure implied by economic theory (see Hansen and Hodrick, 1980, for this point). Suppose, for example, we are interested in the regression $Y_t = X_t\theta + \varepsilon_t, t = 1,...,N$, where $\{(X_t,\varepsilon_t)\}_{t\in\mathbb{N}}$ is weakly dependent (strong mixing) and $\boldsymbol{E}(X_t\varepsilon_t) = 0$. Let $\theta = H(\mu) = \mu_1/\mu_2$, where $\mu_1 = \boldsymbol{E}(X_tY_t)$ and $\mu_2 = \boldsymbol{E}(X_t^2)$. $T_i = (X_iY_i,X_i^2)'$ can be used to construct the confidence interval for the regression coefficient θ. Furthermore, if the assumptions in Section 4.3 are satisfied, the empirical likelihood ratio statistic for θ would be Bartlett-correctable.

EXAMPLE 2 (Spectral Density)

Let $\omega \in [-\pi,\pi]$ and consider the spectral density $f(\omega)$ of a stationary time series $\{X_t\}_{t\in\mathbb{N}}$ with $EX_1 = 0$; here we consider the scalar case $d = 1$ for simplicity, though the generalization to the multivariate case is straightforward. We take the subsample (tapered) periodogram as the ϕ_M-function, thus

$$T_i(\omega) = (2\pi M)^{-1} \left| \sum_{t=L(i-1)+1}^{L(i-1)+M} k_N(t) X_t \exp(-\sqrt{-1}\, t\omega) \right|^2,$$

(see also Politis and Romano, 1992) where the kernel $k_N(\cdot)$ satisfies the usual regularity conditions (see Priestley, 1981, for example). Then $\bar{T}$ defined in section 4.2 is the time averaged periodogram estimator of the spectrum

$$(2\pi M)^{-1}(1/Q) \sum_{i=1}^{Q} \left| \sum_{t=L(i-1)+1}^{L(i-1)+M} k_N(t) X_t \exp(-\sqrt{-1}\, t\omega) \right|^2.$$

If the spectral density is of interest in itself, obviously H is the identity function. Note that the constant kernel can be used to consistently estimate the spectral density at frequency ω; a rigorous proof is given by Zhurbenko (1980). However, if no tapering is used, the block length growth rate has the lower bound of $1/3$ so that assumption (iv) in theorem 3 is satisfied (Politis and Romano, 1992). By using tapering, it is possible to decrease the block length growth rate, or increase the growth rate of the number of blocks-of-blocks q. This may help to achieve higher order coverage accuracy, since the error terms in (4.5a,b) for blocks-of-blocks empirical likelihood would be $O(q^{-1})$ and $O(q^{-2})$. It should be noted that Politis and Romano (1992) demonstrate the consistency of the blocks-of-blocks bootstrap for the spectral density.

EXAMPLE 3 (Variance Ratio Test of Random Walk)

To test the random walk (RW) hypothesis for stock prices, Lo and MacKinlay (1988) considered a simple specification test and rejected RW. Let p_t be the log of the stock price at time t. They calculated

$$\hat{\sigma}_b^2(M)/\hat{\sigma}_a^2 - 1 \text{ or } \hat{\sigma}_c^2(M)/\hat{\sigma}_a^2 - 1,$$

where

$$\hat{\sigma}_a^2(M) = \sum_{t=1}(p_i - p_{i-1} - \hat{m})^2/N,\ \hat{\sigma}_b^2(M) = \sum_{i=1}(p_{Mi} - p_{Mi-M} - M\hat{m})^2/N,$$

$$\hat{\sigma}_c^2(M) = \sum_{i=M}(p_i - p_{i-M} - M\hat{m})^2/NM, \text{ and } \hat{m} = (p_N - p_0)/N.$$

They also investigated the small sample properties of their test (Lo and MacKinlay, 1989). Note that similar ideas have been exploited by other researchers in financial and macro-economics (e.g., Campbell and Mankiw, 1987, Cochrane, 1987, Fama and French, 1987 and Poterba and Summers, 1988). Some authors interpret these variance ratio statistics as the measure of persistence of economic time series.

Clearly the above statistics can be interpreted as the ratio between the spectral density of $\Delta p_t = p_t - p_{t-1}$ at the origin and $\mathrm{Var}(\Delta p_t)$. We can construct an analogous (and asymptotically equivalent) statistic using our blockwise smooth function framework. Let $X_t = \Delta p_t$. Note that we need to deal with parameters of both finite and infinite dimensional distributions. Define $T_i = (T_{1i}, T_{2i}, T_{3i})'$ (in the s-th block of blocks), where

$$T_{1i} = \sum_{t=1}^{M} X_{L(i-1)+t}/M,\ T_{2i} = \sum_{t=1}^{M} X^2_{L(i-1)+t}/M,\ T_{3i} = (\sum_{t=1}^{M} X_{L(i-1)+t} - M\Sigma\hat{m}_s)^2/M,$$

and $\hat{m}_s = \sum_{t=hL(s-1)+1}^{hL(s-1)+L(b-1)+M} X_t/(L(b-1)+M)$. Define the smooth function $\theta = H(\mu_1,\mu_2,\mu_3) = \mu_3/(\mu_2 - \mu_1^2)$. Then $\bar{\theta} = H(\bar{U}_1, \bar{U}_2, \bar{U}_3)$ and we can calculate confidence intervals using the empirical likelihood ratio.

5. Other Distance Measures

As suggested by Owen (1990, 1991), it is possible to conduct statistical inference similar to empirical likelihood methods by considering other distances in place of likelihood. As examples, he suggests the use of KLIC and "Euclidean likelihood". In fact, estimators and test statistics based on these other distance measures have asymptotic properties that are quite similar to empirical likelihood-based counterparts discussed in this paper (see, Owen, 1990, 1991, and DiCiccio and Romano, 1990, Section 2.5). Therefore, it would be natural to consider the blockwise versions of these various nonparametric likelihood methods.

The KLIC from the empirical distribution F_N to $F \ll F_N$ is defined as $D(F,F_N) = \sum_t p_t \log(Np_t)$ (which is justified in the iid case). By maximizing this under the usual constraints as in (3.2) (i.e., requirements for F as a distribution plus moment restrictions), we can calculate various estimators and statistics that are similar to those discussed in this paper. KLIC is known to have many desirable properties and has been extensively investigated in the literature. Its use in density estimation under moment restrictions was considered by Sheehy (1988) and others. Kitamura and Stutzer (1995) considered a KLIC-based estimation method with a blocking scheme which is a special case of the ones considered in this paper.

If we define F and F_N as above, the Euclidean log likelihood ratio between these two distributions is $l_E = -1/2\sum_i (np_i - 1)^2$ (without blocking). It is interesting to apply this distance measure to the analysis of GEE. First suppose the data is iid. Using the same notation as in Section 3, a little calculation shows that the "maximum Euclidean likelihood estimator" for θ is the maximizer of

$$l_E(\theta) = -(N/2)\bar{f}(\theta)' \tilde{S}^{-1} \bar{f}(\theta),$$

where $\tilde{S} = (1/N)\sum_t (f(X_t,\theta) - \bar{f}(\theta))(f(X_t,\theta) - \bar{f}(\theta))'$. Note that this is not an approximation result. Since $\tilde{S}$, evaluated at θ_0, is a consistent estimator of $\mathrm{Var}(f(X_t,\theta_0))$ in the iid case, we can interpret this estimator as a version of Hansen's (1982) optimal GMM estimator, where the optimal weighting matrix and the parameter of interest are estimated simultaneously. Hansen, Heaton and Yaron, (1994) discuss such estimators, which they call the GMM estimators with "continuous updating". If the data process is weakly dependent (and stationary), one would use the blocking technique discussed above and maximize "blockwise Euclidean likelihood". Obviously this is equivalent to the time series optimal GMM with "continuous updating", where the population value of the optimal metric matrix is $S = \lim_{N\to\infty} \mathrm{Var}(1/N \sum_{t=1}^N f(x_t,\theta_0))$.

6. Conclusions

By using blocks to capture the weak dependence of data, we have seen that the method

of empirical likelihood could be applied to models with strong mixing time series. Our approach is nonparametric, and thus is expected to be rather immune from specification errors. It has been also shown that inference regarding the parameters of the infinite joint distribution of the data can be conducted based on the blocks-of-blocks empirical likelihood. The Bartlett-correctability of the blockwise empirical likelihood with strong mixing sequences has been demonstrated.

In practical applications, we need to select block length and the length of time shift for both observations and blocks; one also may wish to use some "tapers" in the construction of blocks. Just as other nonparametric procedures, the efficacy of our procedure could be affected by these choices. As we have seen, the appropriate choice of the block length tends to more be important for the results in Section 4, so care should be taken when implementing these methods. The sensitivity of our method to the choice of these parameters needs to be investigated.

Another issue left out of the paper is the power properties. Though it is a straightforward exercise to calculate the power of the statistics obtained in Section 3 under Pitman drifts, it would not be useful for the comparison of, for example, statistics based on different distance measures. As for the higher order results in Section 4.3, recently Chen (1994a) compared empirical likelihood tests and bootstrap tests by calculating Edgeworth expansions of the power functions in iid sample cases, and suggested some practical rules for choosing from the two methods. The same analysis of our blockwise method with weakly dependent time series would be of interest.

APPENDIX

PROOF OF THEOREM 1: First, it can be shown that

(A.1) $\gamma_N(\theta_0) \to_p 0$,

by following the argument in the proof of Owen (1990); use weak laws of large numbers (WLLN) and a central limit theorem (CLT) for strong mixing processes (see, for example, Ibragimov and Linnik, 1971), which hold under the mixing rate and moment conditions assumed here, in place of the classical WLLN and CLT. In place of Equation (2.5) of Owen's proof, we make use of the fact that $\max_Q T_i(\theta) = o(N^{1/2}M^{-1})$ with probability 1, which follows from Lemma 3.2 of Künsch (1989) and Assumption (vii). By using Owen's argument, it is shown that $\|\gamma_N(\theta_0)\| = O_p(M/N^{1/2})$, which implies the consistency of $\gamma_N(\theta_0)$.

In what follows we show that $\hat{\theta}$, which is the maximizer of $(1/Q)\Sigma_i \log(1 + \gamma_N(\theta)' T_i(\theta))$, is consistent. By assumption (ii)

(A.2) $\boldsymbol{E}[\log(1 + g(\theta)' f(x_t,\theta))] < \boldsymbol{E}[\log(1 + g(\theta_0)' f(x_t,\theta_0))] = 0$

for all $\theta \in \Theta$, $\theta \neq \theta_0$. This is the empirical likelihood version of the conventional information inequality. By assumptions (iii) and (iv),

(A.3) $\lim_{\delta \downarrow 0} \boldsymbol{E} \sup_{\theta^* \in \Gamma(\theta,\delta)} \log(1 + g(\theta^*)' f(x_t,\theta^*)) = \boldsymbol{E}[\log(1 + g(\theta)' f(x_t,\theta))]$.

By (A.2) and (A.3), there exist a finite number of open spheres $\Gamma(\theta_j,\delta_j)$, $j = 1,...,h$, that cover the set $\Theta(\delta) = \Theta/\Gamma(\theta_0,\delta)$, where the small numbers δ_j's are chosen so that

$$\boldsymbol{E} \sup_{\theta^* \in \Gamma(\theta,\delta)} \log(1 + g(\theta^*)' f(x_t,\theta^*)) < \boldsymbol{E} \log(1 + g(\theta_0)' f(x_t,\theta_0)) \ (= 0).$$

By the definition of θ_j, $j = 1,...,h$, we can find positive numbers H_j, $j = 1,...,h$, such that

$$\boldsymbol{E} \sup_{\theta^* \in \Gamma(\theta_j,\delta)} \log(1 + g(\theta^*)' f(x_t,\theta^*)) = -2H_j, \ j = 1,...,h.$$

WLLN ensures the existence of a sufficiently large integer N_j such that for small $\varepsilon > 0$

$$\boldsymbol{P}\left\{(1/N)\Sigma_t \sup_{\theta^* \in \Gamma(\theta_j,\delta)} -\log(1 + g(\theta^*)' f(x_t,\theta^*)) > -H_j\right\} < \varepsilon/(2h),\ j = 1,...,h,$$

for all $N > N_j$.

These h inequalities imply

$$\boldsymbol{P}\left\{\sup_{\theta^* \in \Theta(\delta)} (1/N)\Sigma_t -\log(1 + g(\theta^*)' f(x_t,\theta^*)) > -H\right\} < \varepsilon/2,\ H = \min_j H_j,$$

for all $N > \max_j N_j$.

Now note that

$$(1/Q)\Sigma_i -\log(1 + \gamma_N(\theta)' T_i(\theta))$$

$$\leq (1/Q)\Sigma_i -\log(1 + g(\theta)' \Sigma_{s=1}^{M} f(X_{(i-1)L+s},\theta)/M) \qquad \text{(optimality of } \gamma(\theta))$$

$$\leq (1/QM)\Sigma_i \Sigma_s -\log(1 + g(\theta)' f(X_{(i-1)L+s},\theta)) + o_p(1) \text{ (convexity)}$$

$$= (1/N)\Sigma_t -\log(1 + g(\theta)' f(X_t,\theta)) + o_p(1).$$

Therefore there exists a sufficiently large integer N_A such that,

(A.4) $$\boldsymbol{P}\left\{\sup_{\theta^* \in \Theta(\delta)} (1/Q)\Sigma_i -\log(1 + \gamma_N(\theta^*)' T_i(\theta^*)) > -H\right\} < \varepsilon/2.$$

for all $N > N_A$. Notice that

$$-\log(1 + \gamma_N(\theta_0)'[(1/Q)\Sigma_i T_i(\theta_0)]) \leq (1/Q)\Sigma_i -\log(1 + \gamma_N(\theta_0)' T_i(\theta_0)) \leq 0.$$

The first term converges to 0 in probability since $\gamma_N(\theta_0) \to_p 0$ by (A.1), and the inside of the square bracket in the first logarithm also converges to 0 in probability by WLLN. Thus there exists a large integer N_B such that

(A.5) $$\boldsymbol{P}\left\{(1/Q)\Sigma_i -\log(1 + \gamma_N(\theta_0)' T_i(\theta_0)) < -H/2\right\} < \varepsilon/2$$

for all $N > N_B$. By (A.3) and (A.4), for any small δ,

$$\boldsymbol{P}\{\hat{\theta} \in \Gamma(\theta_0,\delta)\}$$

$$\geq \boldsymbol{P}\left\{(1/Q)\Sigma_i -\log(1 + \gamma_N(\theta_0)' T_i(\theta_0)) > \sup_{\theta^* \in \Theta(\delta)} (1/Q)\Sigma_i -\log(1 + \gamma_N(\theta^*)' T_i(\theta^*))\right\}$$

$$\geq \boldsymbol{P}\left\{(1/Q)\Sigma_i -\log(1 + \gamma_N(\theta_0)' T_i(\theta_0)) \geq -H/2, \sup_{\theta^* \in \Theta(\delta)} (1/Q)\Sigma_i -\log(1 + \gamma_N(\theta^*)' T_i(\theta^*)) \leq -H\right\}$$

$$\geq 1 - \varepsilon$$

for all $N > \max(N_A, N_B)$, thus $\hat{\theta} \to_p \theta_0$.

The asymptotic normality follows by the Taylor expansion of the first order condition just as in Qin and Lawless (1994, Theorem 1) with some modifications. Let

$$l_\gamma(\theta,\gamma) = Q^{-1}\Sigma_{i=1}^{Q} T_i(\theta)/(1 + \gamma' T_i(\theta)),\ l_\theta(\theta,\gamma) = Q^{-1}\Sigma_{i=1}^{Q} (\partial T_i(\theta)/\partial\theta')'\gamma/(1 + \gamma' T_i(\theta)).$$

Also define

$$l_{\gamma\gamma}(\theta,\gamma) = (\partial/\partial\gamma) l_\gamma(\theta,\gamma),\ l_{\theta\gamma}(\theta,\gamma) = (\partial/\partial\gamma) l_\theta(\theta,\gamma) \text{ and } l_{\theta\theta}(\theta,\gamma) = (\partial/\partial\theta) l_\theta(\theta,\gamma).$$

As in Qin and Lawless (1994), the consistency and Assumption (v) implies the following FOCs:

$$l_\gamma(\hat{\theta},\hat{\gamma}) = 0,\ l_\theta(\hat{\theta},\hat{\gamma}) = 0,$$

where we write $\hat{\gamma} = \gamma_N(\hat{\theta})$. Expanding these equations around $(\theta_0, 0)$, we get

$$0 = \frac{\sqrt{N}}{Q}\Sigma T_i(\theta_0) + M l_{\gamma\gamma}(\theta^*,\gamma^*)\frac{\sqrt{N}}{M}(\hat{\gamma} - 0) + l_{\gamma\theta}(\theta^*,\gamma^*)\sqrt{N}(\hat{\theta} - \theta_0),$$

$$0 = 0 + l_{\theta\gamma}(\theta^*,\gamma^*)\frac{\sqrt{N}}{M}(\hat{\gamma} - 0) + \frac{1}{M} l_{\theta\theta}(\theta^*,\gamma^*)\sqrt{N}(\hat{\theta} - \theta_0),$$

where (θ^*,γ^*) is on the line segment joining $(\hat{\theta},\hat{\gamma})$ and $(\theta_0, 0)$, hence $(\theta^*,\gamma^*) \to_p (\theta_0, 0)$, and in particular $\|\gamma^*\| = O_p(M/N^{1/2})$. By using the last result, Assumption (vii) and Künsch's (1989) Lemma 3.2, the argument of the proof of Theorem 1 by Owen (1990) shows that $\max_{1 \leq i \leq Q} \gamma^{*\prime} T_i(\theta^*) = o_p(1)$. Using this result, we obtain

$$M l_{\gamma\gamma}(\theta^*,\gamma^*) = \frac{M}{Q}\Sigma -T_i(\theta^*)T_i(\theta^*)' + o_p(1) \to_p -S,$$

since Assumptions (v)-(vii), ergodicity and stationarity imply that $MQ^{-1}\Sigma T_i(\theta^*)T_i(\theta^*)'$ converges to S in probability. Similarly, $l_{\gamma\theta}(\theta^*,\gamma^*) \Longrightarrow_p D$ and $M^{-1}l_{\theta\theta}(\theta^*,\gamma^*) \Longrightarrow_p 0$. Following Qin and Lawless' argument, the theorem is proved. □

PROOF OF THEOREM 2: (a) Let $\bar{T}(\theta) = \Sigma_{i=1}^{Q} T_i(\theta)/Q$. By the asymptotic normality results in Theorem 1, $M^{-1}N^{1/2}\hat{\gamma} = S^{-1}N^{1/2}\bar{T}(\hat{\theta}) + o_p(1)$ and $\sqrt{N}\bar{T}(\theta_0) \Longrightarrow_d N(0,[S-DV_\theta D'])$. Then the following result follows:

$$
\begin{aligned}
LR_0 &= 2A_N^{-1}\Sigma\log(1 + \hat{\gamma}' T_i(\hat{\theta})) \\
&= 2A_N^{-1}\left(\Sigma\hat{\gamma}' T_i(\hat{\theta}) - (1/2)\Sigma(\hat{\gamma}' T_i(\hat{\theta}))^2\right) + o_p(1) \\
&= 2A_N^{-1}\left(MQ\bar{T}(\hat{\theta})' S^{-1}\bar{T}(\hat{\theta}) - (1/2)M^2\bar{T}(\hat{\theta})S^{-1}M^{-1}Q[\Sigma M' T_i(\hat{\theta})T_i(\hat{\theta})'/Q]S^{-1}\bar{T}(\hat{\theta})\right) + o_p(1) \\
&= N\bar{T}(\hat{\theta})' S^{-1}\bar{T}(\hat{\theta}) + o_p(1) \\
&\Longrightarrow_d \chi^2_{r-p}.
\end{aligned}
$$

(b) The Lagrangean for the constrained estimation is $\mathfrak{L} = \log(L(\theta)) + \zeta'(\psi - \Psi(\theta))$, where $L(\theta)$ is given by (3.5) and ζ is a vector of Lagrange multipliers. Under H_0, the first order condition for the first term of the Lagrangean has the following approximation:

$$
\frac{\sqrt{N}}{MQ}\Sigma_{i=1}^{Q}\frac{\partial T_i(\hat{\theta}^c)/\partial\theta'}{1 + \gamma_N(\hat{\theta}^c)' T_i(\hat{\theta}^c)'}\gamma_N(\hat{\theta}^c) = \left[\frac{1}{N}\Sigma_{t=1}^{N}(\partial f(X_t,\hat{\theta}^c)/\partial\theta')\right]' S^{-1}\left[\frac{1}{\sqrt{N}}\Sigma_{t=1}^{N} f(X_t,\hat{\theta}^c)\right] + o_p(1).
$$

Since the approximation term of the RHS is the optimally weighted estimating functions, the stated chi-square limiting distributions of likelihood-ratio type statistics are obtained by the conventional argument for nonlinear dynamic models (see, for example, Gallant, 1987). □

PROOF OF THEOREM 3: It suffices to show the result for the case in which H is the identity function, since the general case can be treated as in Section 4 of Owen (1990). Then we can

show the theorem following the argument of the proof of Theorem 1 of Owen (1990). First we check that a CLT holds. Let $\alpha_T(k)$ denote the strong mixing measure of T_i. Recall that T_i depends on M and L, which in turn depends on N; therefore α_T implicitly depends on N. As noted by Politis and Romano (1992, Lemma 1(b)), T_i is a strong-mixing sequence and $\alpha_T(k) \le \alpha_X(kL - M)$ for all $k \ge 2$. Then trivially $\Sigma_k \alpha_T^{1-1/c}(k) < \infty$, and assumptions (i)-(iii) ensure the CLT for a strong-mixing triangular array

$$\sqrt{Q}\,(\bar{T} - \mu) \Longrightarrow_d N(0,\Lambda).$$

Also note $\frac{1}{q}\Sigma_{s=1}^{q} U_s U_s' = \frac{1}{bq}\Sigma_{s=1}^{q}\left[\frac{1}{b}\left(\Sigma_{i=1}^{b} T_{(s-1)h+i}\right)^2\right] = \frac{1}{b}(\Lambda + o_p(1))$. Using these results, as in Owen's proof we obtain,

$$LR(\theta_0) = a_N^{-1} bq(\bar{U} - \mu_0)'\Lambda^{-1}(\bar{U} - \mu_0) + o_p(1)$$

$$\Longrightarrow_d \chi_p^2. \quad \square$$

DERIVATION OF (4.5a,b): Our derivation heavily relies upon the argument by DiCiccio, Hall and Romano (hereafter, DHR) (1991); see the working paper version (DHR, 1988), for the details. Throughout our derivation, we assume that appropriate moment conditions are satisfied.

Let $C_i^{j_1\cdots j_k} = T_i^{j_1}\cdots T_i^{j_k} - M^{-(k-1)}\kappa^{j_1\cdots j_k}$ and $K^{j_1\cdots j_k} = Q^{-1}M^{k-1}\Sigma_{i=1}^{Q} C_i^{j_1\cdots j_k}$. By the moment bound for sums of strong mixing random variables (Yokoyama, 1980, or Kim, 1993) and a CLT for strong mixing triangular arrays,

$$\kappa^{j_1\cdots j_k} = O(M^{k/2-1}) = O(N^{(1/3)(k/2-1)}) \text{ and}$$

$$N^{1/2}K^{j_1\cdots j_k} = O_p(N^{1/2}Q^{-1/2}M^{(k/2-1)}) = O_p(M^{(k-1)/2}) = O_p(N^{(k-1)/6}).$$

Therefore $\kappa^{j_1\cdots j_k}$ (and $\tilde{\kappa}^{j_1\cdots j_{k(1)},j_{k(1)+1}\cdots j_{k(2)},j_{k(2)+1}\cdots j_{k(d-1)},j_{k(d-1)+1}\cdots j_{k(d)}}$, $k(d) = k$) and $N^{1/2}K^{j_1\cdots j_{k'}}$ have growing orders of magnitude if $k > 2$ and $k' > 1$, respectively. Call $k - 2$ and $k' - 1$ "excess growth rates" of $\kappa^{j_1\cdots j_k}$ and $K^{j_1\cdots j_{k'}}$ (or $\tilde{\kappa}^{j_1\cdots j_{k(1)},j_{k(1)+1}\cdots j_{k(2)},j_{k(2)+1}\cdots j_{k(d-1)},j_{k(d-1)+1}\cdots j_{k(d)}}$, $k(d) = k'$) for the subsequent developments.

(Recall $\alpha^{j_1 \cdots j_k}$ and $N^{1/2}A^{j_1 \cdots j_{k'}}$ of DHR are of order $O(1)$ and $O_p(1)$, respectively.)

First we consider the empirical likelihood for the mean parameter with $\mu_0 = 0$ and $\Sigma_N = I$. Let γ denote the Lagrangean multipliers as before. Expand the first order condition:

$$\begin{aligned} & Q^{-1}\Sigma(1 + \gamma' T_i)^{-1} T_i \\ & \quad = Q^{-1}\Sigma T_i - M^{-1}\gamma'(Q^{-1}\Sigma M T_i T_i' - I) - M^{-1}\gamma + Q^{-1}\Sigma(\gamma' T_i)^2 T_i - Q^{-1}\Sigma(\gamma' T_i)^3 T_i \cdots \\ & \quad = 0 \end{aligned}$$

(note $Q^{-1}\Sigma M T_i T_i' \Longrightarrow_p I$). Let $\bar{\gamma} = M^{-1}\gamma$, so that $\bar{\gamma}$ is of order $O_p(N^{-1/2})$. The j-th element of the above vector is,

$$\begin{aligned} & K^j - \bar{\gamma}^j - K^{jk}\bar{\gamma}^k + \kappa^{jkl}\bar{\gamma}^k\bar{\gamma}^l + K^{jkl}\bar{\gamma}^k\bar{\gamma}^l - \kappa^{jklm}\bar{\gamma}^k\bar{\gamma}^l\bar{\gamma}^m + O_p(N^{-3/2}) \\ & \quad = 0, \end{aligned}$$

which can be solved for $\bar{\gamma}$. Note

$$\text{(A.6)} \quad N^{-1}LR(\mu_0) = N^{-1}\Sigma_{i=1}^{Q}\left\{(\gamma' T_i)^2 - (4/3)(\gamma' T_i)^3 + (3/2)(\gamma' T_i)^4\right\} + O_p(N^{-2}).$$

By substituting the expression for $\bar{\gamma}$ obtained above into (A.6), we get

$$\begin{aligned} \text{(A.7)} \quad N^{-1}LR(\mu_0) &= K^j K^j - K^{jk}K^j K^k + 2/3\kappa^{jkl}K^j K^k K^l + K^{jl}K^{kl}K^j K^k + 2/3K^{jkl}K^j K^k K^l \\ &\quad - 2\kappa^{jkm}K^{lm}K^j K^k K^l + \kappa^{jkn}\kappa^{lmn}K^j K^k K^l K^m - 1/2\kappa^{jklm}K^j K^k K^l K^m + O_p(N^{-2}). \end{aligned}$$

Define R_1, R_2 and R_3 as in DHR (1991) with A, α, θ, P, Q and M replaced by κ, K, H, J, G and W. Then $LR(\theta_0)$ is approximated using $R = R_1 + R_2 + R_3$:

$$N^{-1}LR(\theta_0) = R'R + O_p(N^{-1}), \text{ or } LR(\theta_0) = NR'R + O(N^{-2}).$$

Next we calculate the third and fourth cumulants of R. For our purpose, it is enough to show that

$$\text{cum}(R^u R^v R^w) = O(N^{-5/2}),$$

$$\text{cum}(R^u R^v R^w R^x) = O(N^{-3}).$$

Define $O = \nabla H'(\nabla H \nabla H')^{-1/2}$ and $J = (\nabla H \nabla H')^{-1/2}$ as in DHR. In what follows we derive

various moments of R_1, R_2 and R_3, which are functions of K's. Moments of K's, which are centered sums of mixing random variables T_i, can be expressed in terms of $\tilde{\kappa}$'s. (Note the k-th order $\tilde{\kappa}$ only involves moments of T_i's within a $k-1$ consecutive periods; see the definition (4.4). This is a consequence of the mixing condition (iii).) To this end, note

$$\begin{aligned} \alpha_T(k) &\le \alpha_X(kL - M) \le c\exp[-d(kL - M)] \\ &= c\exp[-d(k-1)M] \\ &= c\exp[-d(k-1)N^{1/3}] \end{aligned} \tag{A.8}$$

for $k \ge 2$, by Assumptions (i) and (iv). (A.8) implies that T_i is asymptotically 1-dependent process with small asymptotic approximation errors. The mixing inequality and (A.8) imply formulae as in Step 6 of DHR (1988), though they need to be appropriately modified to take account of serial correlations among T_i's. Also note that $M^{k/2}C_i^{j_1\cdots j_k}$ and $M^{m/2}C_{i+1}^{l_1\cdots l_m}$ are asymptotically independent for k, $m \in \mathbb{N}$. Moreover, under certain moment conditions, $\boldsymbol{E}(T_i^{\,j}T_{i+1}^{\,k}T_{i+2}^{\,l}) < C\alpha_X^{1-r}([M/2]+1) \le Cc\exp(-d(1-r)([M/2]+1)) = O(\exp(-d(1-r)N^{1/3}/2))$, where C is a positive constant and $r < 1$. To see this, notice that third order moments of X_t, $\boldsymbol{E}X_\tau X_{\tau+p}X_{\tau+p+q}$, say, that appear in $\boldsymbol{E}(T_i^{\,j}T_{i+1}^{\,k}T_{i+2}^{\,l})$ can be bounded uniformly by $C'\alpha_X^{1-r}([M/2]+1)$ for some $C' > 0$, using the mixing inequality (e.g., Corollary A.2, Hall and Heyde, 1980), since $\max(p,q) > [M/2]$. Similarly, $\boldsymbol{E}(T_i^{\,j}T_{i+1}^{\,k}T_{i+2}^{\,l}T_{i+3}^{\,m}) = O(\exp(-d(1-r')N^{1/3}/2))$ for some $r' < 1$. Using these results, it can be shown that the following equalities hold ignoring terms of order $O(N^{-5/2})$:

$$\begin{aligned} \boldsymbol{E}(R_1^{\,u}R_1^{\,v}R_1^{\,w}) &= N^{-2}\tilde{\kappa}^{j,k,l}O^{\,ju}O^{\,kv}O^{\,lw}, \\ \boldsymbol{E}(R_2^{\,u}) &= N^{-1}\{(-1/2)\tilde{\kappa}^{jk,l} + (1/3)\kappa^{jkl}O^{\,ju}W^{kl} + (1/2)J^{uv}H_{jk}^{\,v}(I - W)^{jk}\}, \\ \boldsymbol{E}(R_2^{\,u}R_1^{\,v}R_1^{\,w}) &= N^{-2}\{((-1/2)\tilde{\kappa}^{jk,l} + (1/3)\kappa^{jkl})\delta^{vw}O^{\,ju}W^{kl} + (-\tilde{\kappa}^{jk,l} + (2/3)\kappa^{jkl})O^{\,ju}O^{\,kv}O^{\,lw} \\ &\quad + (1/2)J^{uv}H_{jk}^{\,v}(I - W)^{jk}\delta^{vw}\}. \end{aligned} \tag{A.9}$$

Also by definition,

$$\boldsymbol{E}(R_1^{\,u}R_1^{\,v}) \;=\; N^{-1}\delta^{uv}.$$

Thus

$$\begin{aligned}\boldsymbol{E}(R_2^{\,u}R_1^{\,v}R_1^{\,w}) \;-\; \boldsymbol{E}(R_2^{\,u})\boldsymbol{E}(R_1^{\,v}R_1^{\,w}) \;+\; \frac{1}{3}\boldsymbol{E}(R_1^{\,u}R_1^{\,v}R_1^{\,w}) \\ = \frac{1}{(3N^2)}(\tilde{\kappa}^{j,k,l} \;-\; 3\tilde{\kappa}^{jk,l} \;+\; 2\kappa^{jkl})O^{\,ju}O^{\,kv}O^{\,lw} \;+\; O(N^{-5/2}) \\ = O(N^{-5/2}),\end{aligned}$$

and

$$\text{(A.10)}\quad \text{cum}(R^{\,u}R^{\,v}R^{\,w}) \;=\; \boldsymbol{E}(R_1^{\,u}R_1^{\,v}R_1^{\,w}) \;+\; \boldsymbol{E}(R_2^{\,u}R_1^{\,v}R_1^{\,w})[3] \;-\; \boldsymbol{E}(R_2^{\,u})\boldsymbol{E}(R_1^{\,v}R_1^{\,w})[3] \;+\; O(N^{-5/2})$$

(cf., DHR, 1988), therefore we obtain the desired result.

Next we evaluate the forth order cumulants. Let

$$S_{1a} \;=\; \tilde{\kappa}^{j,k,l,m}\; O^{\,ju}O^{\,kv}O^{\,lw}O^{\,mx}$$

$$S_{1b} \;=\; -6\tilde{\kappa}^{jk,l,m}\; O^{\,ju}O^{\,kv}O^{\,lw}O^{\,mx}$$

$$S_{1c} \;=\; 3\tilde{\kappa}^{jk,lm}\; O^{\,ju}O^{\,kv}O^{\,lw}O^{\,mx}$$

$$S_{1d} \;=\; (8\tilde{\kappa}^{jkl,m} \;-\; 6\kappa^{jklm})\; O^{\,ju}O^{\,kv}O^{\,lw}O^{\,mx}$$

$$S_2 \;=\; M(\delta^{uv}\delta^{wx} \;+\; \delta^{uw}\delta^{vx} \;+\; \delta^{ux}\delta^{vw})$$

$$\begin{aligned}S_{3a} \;=\; &((-1/2)\tilde{\kappa}^{j,k,l}\tilde{\kappa}^{mn,o}+(1/3)\tilde{\kappa}^{j,k,l}\kappa^{mno})W^{no} \\ &\cdot(O^{\,ju}O^{\,kv}O^{\,lw}O^{\,mx} \;+\; O^{\,ju}O^{\,kv}O^{\,lx}O^{\,mw} \;+\; O^{\,ju}O^{\,kw}O^{\,lx}O^{\,mv} \;+\; O^{\,ju}O^{\,kw}O^{\,lx}O^{\,\mu})\end{aligned}$$

$$\begin{aligned}S_{3b} \;=\; &((3/2)\tilde{\kappa}^{j,kl}\tilde{\kappa}^{mn,o} \;-\; \kappa^{jkl}\tilde{\kappa}^{mn,o} \;-\; \tilde{\kappa}^{j,kl}\kappa^{mno} \;+\; (2/3)\kappa^{jkl}\kappa^{mno})W^{no} \\ &\cdot(O^{\,ju}O^{\,kv}O^{\,lw}O^{\,mx} \;+\; O^{\,ju}O^{\,kv}O^{\,lx}O^{\,mw} \;+\; O^{\,ju}O^{\,kw}O^{\,lx}O^{\,mv} \;+\; O^{\,ju}O^{\,kw}O^{\,lx}O^{\,\mu})\end{aligned}$$

$$S_{4a} \;=\; (-2\tilde{\kappa}^{jk,l}\tilde{\kappa}^{m,n,o} \;+\; (8/3)\kappa^{jkl}\tilde{\kappa}^{m,n,o})W^{km}(O^{\,ju}O^{\,kv}O^{\,lw}O^{\,mx} \;+\; O^{\,ju}O^{\,kw}O^{\,lv}O^{\,mx} \;+\; O^{\,ju}O^{\,kx}O^{\,lv}O^{\,mw})$$

$$\begin{aligned}S_{4b} \;=\; &(\tilde{\kappa}^{jk,l}\tilde{\kappa}^{mn,o} \;+\; 2\tilde{\kappa}^{jk,l}\tilde{\kappa}^{no,m} \;-\; (8/3)\kappa^{jkl}\tilde{\kappa}^{mn,o} \;-\; (8/3)\kappa^{jkl}\tilde{\kappa}^{no,m} \;+\; (16/9)\kappa^{jkl}\kappa^{mno}) \\ &\cdot W^{km}(O^{\,ju}O^{\,kv}O^{\,lw}O^{\,mx} \;+\; O^{\,ju}O^{\,kw}O^{\,lv}O^{\,mx} \;+\; O^{\,ju}O^{\,kx}O^{\,lv}O^{\,mw})\end{aligned}$$

$$S_{4c} = (3\tilde{\kappa}^{jk,l}\tilde{\kappa}^{mn,o} - (20/3)\kappa^{jkl}\tilde{\kappa}^{mn,o} + (32/9)\kappa^{jkl}\kappa^{mno}) \cdot W^{km}(O^{ju}O^{kv}O^{lw}O^{mx} + O^{ju}O^{kw}O^{lv}O^{mx} + O^{ju}O^{kx}O^{lv}O^{mw})$$

$$S_{5a} = (1/2)H_{mn}^{y}(I - W)^{mn}\tilde{\kappa}^{j,k,l} \cdot (J^{yu}O^{jv}O^{kw}O^{lw} + J^{yv}O^{ju}O^{kw}O^{lx} + J^{yw}O^{ju}O^{kv}O^{lx} + J^{yx}O^{ju}O^{kv}O^{lw})$$

$$S_{5b} = H_{mn}^{y}(I - W)^{mn}((-3/2)\tilde{\kappa}^{j,kl} + \kappa^{jkl}) \cdot (J^{yu}O^{jv}O^{kw}O^{lw} + J^{yv}O^{ju}O^{kw}O^{lx} + J^{yw}O^{ju}O^{kv}O^{lx} + J^{yx}O^{ju}O^{kv}O^{lw}).$$

Then ignoring terms that are $O(N^{-3})$ (note S_1's, S_2, S_3's and S_4's are of order $O(M)$ and $S_5 = O(M^{1/2})$),

$$\boldsymbol{E}(R_1^{u}R_1^{v}R_1^{w}R_1^{x}) - \boldsymbol{E}(R_1^{u}R_1^{v})\boldsymbol{E}(R_1^{w}R_1^{x})[3] = N^{-3}(S_{1a} - S_2),$$

$$\boldsymbol{E}(R_2^{u}R_1^{v}R_1^{w}R_1^{x})[4] - \boldsymbol{E}(R_2^{u}R_1^{v})\boldsymbol{E}(R_1^{w}R_1^{x})[12] = N^{-3}(S_{1b} + 2S_2 + S_{3a} + S_{4a} + S_{5a}),$$

$$\boldsymbol{E}(R_2^{u}R_2^{v}R_1^{w}R_1^{x})[6] - \boldsymbol{E}(R_2^{u}R_2^{v})\boldsymbol{E}(R_1^{w}R_1^{x})[6] = N^{-3}(S_{1c} - S_2 + S_{3b} + S_{4b} + S_{5b}),$$

$$\boldsymbol{E}(R_3^{u}R_1^{v}R_1^{w}R_1^{x})[4] - \boldsymbol{E}(R_3^{u}R_1^{v})\boldsymbol{E}(R_1^{w}R_1^{x})[12] = N^{-3}(S_{1d} + S_{4c}).$$

Note

$$S_{1a} + S_{1b} + S_{1c} + S_{1d} = 0$$

$$S_{3a} + S_{3b} + S_{3c} + S_{3d} = \left\{(-1/2)(\tilde{\kappa}^{j,k,l} - 3\tilde{\kappa}^{j,kl} + 2\kappa^{jkl})\tilde{\kappa}^{mn,o} + (1/3)(\tilde{\kappa}^{j,k,l} - 3\tilde{\kappa}^{j,kl} + 2\kappa^{jkl})\kappa^{mno}\right\}W^{no} \cdot (O^{ju}O^{kv}O^{lw}O^{mx} + O^{ju}O^{kv}O^{lx}O^{mw} + O^{ju}O^{kw}O^{lx}O^{mv} + O^{ju}O^{kw}O^{lx}O^{\mu}) = 0$$

$$S_{4a} + S_{4b} + S_{4c} = \{(-2\tilde{\kappa}^{m,n,o} + 4\tilde{\kappa}^{mn,o} - 2\tilde{\kappa}^{no,m})\tilde{\kappa}^{jk,l} - 4\kappa^{mno}\tilde{\kappa}^{jk,l} + ((8/3)\tilde{\kappa}^{m,n,o} - (28/3)\tilde{\kappa}^{mn,o} + (8/3)\tilde{\kappa}^{no,m} + (16/3)\kappa^{mno})\kappa^{jkl} + 4\tilde{\kappa}^{mn,o}\kappa^{jkl}\} \cdot W^{km}(O^{ju}O^{kv}O^{lw}O^{mx} + O^{ju}O^{kw}O^{lv}O^{mx} + O^{ju}O^{kx}O^{lv}O^{mw}) = (\tilde{\kappa}^{m,n,o} - 2\tilde{\kappa}^{mn,o} - \tilde{\kappa}^{no,m} + 2\kappa^{mno})(-2\tilde{\kappa}^{jk,l} + (8/3)\kappa^{jkl}) \cdot W^{km}(O^{ju}O^{kv}O^{lw}O^{mx} + O^{ju}O^{kw}O^{lv}O^{mx} + O^{ju}O^{kx}O^{lv}O^{mw}) = 0$$

and

$$S_{5a} + S_{5b} = (1/2)H^{y}_{mn}(I - W)^{mn}(\tilde{\kappa}^{j,k,l} - 3\tilde{\kappa}^{j,kl} + 2\kappa^{jkl}) \cdot (J^{yu}O^{jv}O^{kw}O^{lw} + J^{yv}O^{ju}O^{kw}O^{lx} + J^{yw}O^{ju}O^{kv}O^{lx} + J^{yx}O^{ju}O^{kv}O^{lw}) = 0 .$$

Finally, by (A.10),

$$\boldsymbol{E}(R_2)\boldsymbol{E}(R_1^{v}R_1^{w}R_1^{x})[4] - \boldsymbol{E}(R_2)\boldsymbol{E}(R_2^{v}R_1^{w}R_1^{x})[12] + 2\boldsymbol{E}(R_2^{u})\boldsymbol{E}(R_2^{v})\boldsymbol{E}(R_1^{w}R_x^{x})[6] = O_p(N^{-5/2})O(N^{-5/6}) = O(N^{-10/3})$$

Using (3.10) of DHR (1989), it is shown that $\text{cum}(R^{u}R^{v}R^{w}R^{x}) = O(N^{-3})$.

The above results imply that the third- and fourth order cumulants of $N^{1/2}R$ are $O(N^{-1}) = O(Q^{-3/2})$, and it can be shown that the s-th order cumulant of $NR'R\{\boldsymbol{E}[NR'R/p]\}^{-1}$ coincides with that of χ^2_p up to errors of order $O(Q^{-3/2})$. Given the validity of Edgeworth expansions, we have

$$\boldsymbol{P}\{LR(\theta_0)/\boldsymbol{E}[NR'R/p] \le z\} = \boldsymbol{P}\{\chi^2_p \le z\} + O_p(Q^{-3/2}).$$

By the same argument as in DHR, the error rate is in fact $O(Q^{-2})$ $(= O(N^{-4/3}))$, due to parity properties of polynomials in Edgeworth expansions (see Götze and Hipp, 1983, for Edgeworth expansions for dependent processes).

The last step is the derivation of the Bartlett factor. Note

$$\boldsymbol{E}(R^{u}R^{u}) = \boldsymbol{E}(R_1^{u}R_1^{u}) + 2\boldsymbol{E}(R_1^{u}R_2^{u}) + 2\boldsymbol{E}(R_1^{u}R_3^{u}) + \boldsymbol{E}(R_2^{u}R_2^{u}) + O(N^{-5/2}),$$

where the order of the remainder term reflects the stochastic order of $E(R_2^{u}R_3^{u})$: without the excess growth rates, this term is $O(N^{-3})$, and the sums of excess growth rates are at most 3, hence we get $\boldsymbol{E}(R_2^{u}R_3^{u}) = O(N^{-3}M^{3/2}) = O(N^{-5/2})$. Obviously

$$\boldsymbol{E}(R_1^{u}R_1^{u}) = N^{-1}p.$$

Also a long but straightforward calculation shows that, ignoring terms of order $O(N^{-7/3})$,

$$\boldsymbol{E}(R_1^{u}R_2^{u}) = N^{-2}\{t_{1a} + t_{3a} + t_{4a} + (1/2)p\},$$

$$\boldsymbol{E}(R_1^{u} R_3^{u}) = N^{-2}\{t_{1b} + t_{2a} + t_{3b} + t_{4b} + t_{6a} - (3/8)p\},$$

$$\boldsymbol{E}(R_2^{u} R_2^{u}) = N^{-2}\{t_{1c} + t_{2b} + t_{3c} + t_5 + t_{6b} - (1/4)p\}.$$

Using the above formulae,

$$\begin{aligned}
&\boldsymbol{E}(NR'R) \\
&\quad = p + N^{-1}(2t_{1a} + 2t_{1b} + t_{1c} + 2t_{2a} + t_{2b} + 2t_{3a} + 2t_{3b} + t_{3c} + 2t_{4a} + 2t_{4b} + t_5 + 2t_{6a} + t_{6b}) \\
&\quad + O(N^{-4/3}) \\
&\quad = p + N^{-1}a + O(Q^{-2}),
\end{aligned}$$

therefore

$$[\boldsymbol{E}(NR'R)/p]^{-1} = 1 - (pN)^{-1}a + O(Q^{-2}).$$

(Recall $a = O(M)$.) For the general case $\Sigma_N \neq I$, we replace $H(\lambda)$ with $\bar{H}(\lambda) = H(\Sigma_M^{1/2}\lambda)$, and the desired result follows. □

REFERENCES

Barndorff-Nielsen, O. E. and Hall, P. (1988). On the level-error after Bartlett adjustment of the likelihood ration statistics. *Biometrika* 75, 374-378.

Bhattacharya, R. N. and Ghosh, J. K. (1978). On the validity of the formal Edgeworth expansion. *Annals of Statistics* 6, 434-451

Billingsley, P. (1968). *Convergence of Probability Measures*. John Wiley: New York.

Bühlmann, P. and Künsch, H. R. (1993). The blockwise bootstrap for general parameters of a stationary time series. Research Report 70, ETH Zürich.

Campbell, J. Y. and Mankiw, N. G. (1987). Are output fluctuations transitory? *Quarterly Journal of Economics* 102, 857-880.

Carlstein, E. (1986). The use of subseries values for estimating the variance of a general statistic from stationary sequence, *Annals of Statistics* 14, 1171-1179.

Chen, S. X. (1993). On the accuracy of empirical likelihood confidence regions for linear regression models. *Annals of the Institute of Statistical Mathematics* 45, 621-637.

Chen, J. and Qin, J. (1993). Empirical likelihood estimation for finite population and the effective usage of auxiliary information. *Biometrika* 80, 107-116.

Chen, S. X. (1994a). Comparing empirical likelihood and bootstrap hypothesis tests. *Journal of Multivariate Analysis* 51, 277-293.

Chen, S. X. (1994b). Empirical likelihood confidence intervals for linear regression coefficients. *Journal of Multivariate Analysis* 49, 24-40.

Chen, S. X. and Hall, P. (1993). Smoothed empirical likelihood confidence intervals for quantiles. *Annals of Statistics* 21, 1166-1181.

Cochrane, J. H. (1988). How big is the random walk in GNP? *Journal of Political Economy* 96, 893-920.

Davison, A. C. and Hall, P. (1993). On studentizing and blocking methods for implementing the bootstrap with dependent data. *Australian Journal of Statistics* 35, 215-224.

Davison, A. C., Hinkley, D. V. and Worton, B. J. (1992). Bootstrap likelihoods. *Biometrika* 79, 113-130.

DiCiccio, T., Hall, P. and Romano, J. (1988). Bartlett adjustment for empirical likelihood.

Technical Report No. 298, Department of Statistics, Stanford University.

DiCiccio, T., Hall, P. and Romano, J. (1991a). Empirical likelihood is Bartlett-correctable. *Annals of Statistics* 19, 1053-1061.

DiCiccio, T., Hall, P. and Romano, J. (1991b). Comparison of parametric and empirical likelihood functions. *Biometrika* 76, 465-476

DiCiccio, T. and Romano, J. (1989). On adjustments to the signed root of the empirical likelihood ratio statistics. *Biometrika* 76 447-456.

DiCiccio, T. and Romano, J. (1990). Nonparametric confidence limits by resampling methods and least favorable families. *International Statistical Review* 58, 59-76.

Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. Chapman and Hall: New York.

Eicker, F. (1967). Limit theorems for regressions with unequal and dependent errors. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Vol 1, 59-82. Berkeley: University of California Press.

Fama, E. and French, K. (1988). Permanent and temporary components of stock prices. *Journal of Political Economy* 96, 246-273.

Gallant, A.R. (1987). *Nonlinear Statistical Models*. John Wiley: New York.

Götze, F. and Hipp, C. (1983). Asymptotic expansions for sums of weakly dependent random vectors. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 64, 211-239.

Götze, F. and Künsch, H. R. (1993). Second order correctness of the blockwise bootstrap for stationary observations. Mimeo.

Haberman, S.J. (1984). Adjustment by minimum discriminant information. *Annals of Statistics* 12, 971-988.

Hall, P. (1985). Resampling a coverage process. *Stochastic Processes and Their Applications* 19, 259-269.

Hall, P. (1987). Edgeworth expansion for student's t statistic under minimal moment conditions. *Annals of Probability* 15, 920-931.

Hall, P. (1990). Pseudo-likelihood theory for empirical likelihood. *Annals of Statistics* 18, 121-140.

Hall. P. (1992). *The Bootstrap and Edgeworth Expansion*. Springer-Verlag: New York.

Hall, P. and Heyde, C. C. *Martingale Limit Theory and Its Applications*. Academic Press: San Diego, California.

Hall, P. and La Scala, B. (1990). Methodology and algorithms of empirical likelihood. *International Statistical Review* 58, 109-127.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica* 50, 1029-1054.

Hansen, L. P., Heaton, J. and Yaron, A. (1995). Finite sample properties of some alternative GMM estimators. University of Chicago, mimeo.

Hansen, L. P. and Hodrick, R. J. (1980). Forward exchange rates as optimal predictors of future spot rates: an econometric analysis. *Journal of Political Economy* 88, 829-853.

Hansen, L. P. and Singleton, K. J. (1982). Generalized instrumental variables estimation of nonlinear rational expectation models. *Econometrica* 50, 1269-1286.

Ibragimov, I.A. (1962). Some limit theorems for stationary processes. *Theory of Probability and Its Applications* 7, 349-382.

Ibragimov, I.A. and Linnik, Y.V. (1971). *Independent and Stationary Sequences of Random Variables*. Wolters-Noordhoff: Groningen.

Janas, D. (1994). Edgeworth expansions for spectral mean estimates with applications to Whittle estimates. *Annals of the Institute of Statistical Mathematics* 46, 667-682.

Kim, T. Y. (1993). A note on moment bounds for strong mixing sequences. *Statistics and Probability Letters* 16, 163-168.

Kitamura, Y. and Stutzer, M. (1995). An information-theoretic alternative to generalized method of moments estimation. University of Minnesota, mimeo.

Kolaczyk, E. D. (1992). Empirical likelihood for generalized linear models. *Statistica Sinica* 4, 199-218.

Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *Annals of Statistics* 17, 1217-1241.

Lahiri, S. N. (1991). Second order optimality of stationary bootstrap. *Statistics and Probability Letters* 11, 335-341.

Lahiri, S. N. (1992). Edgeworth correction by 'moving block' bootstrap for stationary and nonstationary data. *Exploring the Limits of Bootstrap*. LePage, R. and Billard, L., Editors, John Wiley: New York.

Leger, C., Politis, D. N. and Romano, J. P. (1992). Bootstrap technology and applications. *Technometrics* 34, 378-398.

Lo, A. W. and MacKinlay, A. C. (1988). Stock market prices do not follow random walks: Evidence from a simple specification test. *Review of Financial Studies* 1, 41-66.

Lo, A. W. and MacKinlay, A. C. (1989). The size and power of the variance ration test in finite samples: A Monte Carlo investigation. *Journal of Econometrics* 40, 203-238.

Mykland, P. A. (1995) Dual likelihood. *Annals of Statistics* 23, 396-421.

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

Owen, A. (1990). Empirical likelihood ratio confidence regions. *Annals of Statistics* 18, 90-120.

Owen, A. (1991). Empirical likelihood for linear models. *Annals of Statistics* 19, 1725-1747.

Poterba, J. and Summers, L. (1988). Mean reversion in stock returns: Evidence and implications. *Journal of Financial Economics* 22, 27-59.

Politis, D.N. and Romano, J.P. (1992). A general resampling scheme for triangular arrays of α-mixing random variables with application to the problem of spectral density estimation. *Annals of Statistics* 20, 1985-2007.

Politis, D. N. and Romano, J. P. (1993). Nonparametric resampling for homogeneous strong mixing random fields. *Journal of Multivariate Analysis* 47, 301-328.

Politis, D. N. and Romano, J. P. (1993). On the sample variance of linear statistics derived from mixing sequences. *Stochastic Processes and their Applications* 45, 155-167.

Politis, D. N. and Romano, J. P. (1995). Large sample confidence regions based on subsamples under minimal assumptions. *Annals of Statistics*, 22, 2031-2050.

Priestley, M. B. (1981). *Spectral Analysis and Time Series*, Academic Press: New York.

Qin, J. (1993). Empirical likelihood in biased sample problems. *Annals of Statistics* 21, 1182--381196.

Qin, J. (1994). Semi-empirical likelihood ratio confidence intervals for the difference of two

sample means. *Annals of the Institute of Statistical Mathematics* 46, 117-126.

Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *Annals of Statistics* 22, 300-325.

Rosenblatt, M. (1985) *Stationary sequences and random fields*. Birkhäuser: Boston.

Sheehy, A. (1988). Kullback-Leibler constrained estimation of probability measures, Technical Report 137, Department of Statistics, Stanford University.

Singh, K. (1981). On the asymptotic accuracy of Efron's bootstrap. *Annals of Statistics* 9, 1187-1195.

Taniguchi, M. (1991). *Higher Order Asymptotic Theory for Time Series Analysis*. Lecture Notes in Statistics, Vol.68, Springer-Verlag: Heidelberg.

Wald, A. (1949). Note on the consistency of the maximum likelihood estimate. *Annals of Mathematical Statistics* 20, 595-601.

Welch, P. D. (1967). The use of the fast Fourier transform for estimation of spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics* AU-15, 70-73.

Wolfowitz, J. (1949). On the Wald's proof of the consistency of the maximum likelihood estimate. *Annals of Mathematical Statistics* 20, 601-603.

Yokoyama, R. (1980). Moment bounds for stationary mixing sequences. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 52, 45-57.

Zhurbenko, I. G. (1979). On the spectral density statistics obtained by means of time shift. *Soviet Math. Dokl* 20, 547-552.

Zhurbenko, I. G. (1986). *The Spectral Analysis of Time Series*. North-Holland: Amsterdam.