The Annals of Statistics
1997, Vol. 25, No. 5, 2084–2102

EMPIRICAL LIKELIHOOD METHODS WITH WEAKLY DEPENDENT PROCESSES[1]

BY YUICHI KITAMURA

University of Minnesota

This paper studies the method of empirical likelihood in models with — 总起
weakly dependent processes. In such cases, if the likelihood function is — 现实问题
formulated as if the data process were independent, obviously empirical
likelihood fails. We propose to use empirical likelihood of blocks of obser- — 解决方案
vations to solve this problem in a nonparametric manner. This method of
"blockwise empirical likelihood" preserves the dependence of data, and the — 新方案优点
resulting likelihood ratios can be used to construct asymptotically valid
confidence intervals. We consider general estimating equations, for which — 工作1
an efficient estimator is derived by maximizing blockwise empirical likeli-
hood. We also introduce "blocks-of-blocks empirical likelihood" to conduct — 工作2
inference for parameters of the infinite dimensional joint distribution of
data; the smooth function model is used for such cases. We show that — 效果
blockwise empirical likelihood of the smooth function model with weakly
dependent processes is Bartlett correctable. A wide variety of problems, — 应用前景
such as time series regressions and spectral densities, can be treated
using our methodology.

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), page 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 [see also Hall and La Scala (1990)]. 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 many applications: smooth function models, regression models [Owen (1991), Chen (1993, 1994a, b)], generalized

Received May 1996; revised February 1997.
[1] Research supported in part by NSF Grant SBR-96-32101 and the University of Minnesota Graduate School.
AMS 1991 *subject classifications*. Primary 62M10; secondary 62G10, 62E20.
Key words and phrases. Bartlett correction, Edgeworth expansion, empirical likelihood, estimating function, generalized method of moments, nonparametric likelihood, spectral density, strong mixing, time series regression, weak dependence.

linear models [Kolaczyk (1994)], 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 (1989, 1991), DiCiccio and Romano (1989, 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-squared 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, that 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 vectors X_t, $t = 1, \dots, N$. Treating X_t *as if they were independent*, the empirical likelihood for θ is the value of the likelihood of the multinomial distribution $\prod_{t=1}^{N} p_t$ maximized under the constraints $\sum_t p_t = 1$ and $\sum_t p_t X_t = \theta$. Let $L(\theta)$ denote the maximum value. Without the second constraint, clearly the likelihood is maximized at the empirical distribution $p_t = 1/N$ for all t, thereby yielding the estimate $\overline{X} = N^{-1}\sum X_t$. Writing this $L(\overline{X})$, the empirical likelihood ratio is

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

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

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

where $\overline{\Sigma} = N^{-1}\sum_t (X_t - \theta_0)(X_t - \theta_0)'$. 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 $\overline{\Sigma}$ provides a "wrong metric" for the score $\sqrt{N}(\overline{X} - \theta_0)$; $\overline{\Sigma}$ converges to $\mathrm{Var}(X_t)$ in probability, instead of the desired term $\sum_{-\infty}^{\infty} \mathrm{Cov}(X_t, X_{t-j})$. 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, an ARMA model) so that the transformed innovations become 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 blockwise resampling [see Hall (1985), Carlstein (1986), Künsch (1989) and Bühlman and Künsch (1993)], which preserves the dependence property of the data nonparametrically by appropriately choosing blocks of data.

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. The sample blocking allows us to treat (weak) dependence of time series in a nonparametric way.

Our methodology is quite general. Its applications include the following.

1. *Time series regression*. In this example the parameter of interest is the coefficient vector in the following regression model:

$$Y_t = X_t \theta + \varepsilon_t, \qquad t = 1, \dots, N,$$

where $\{(X_t, \varepsilon_t)\}$ is weakly dependent and $\mathbf{E}(X_t \varepsilon_t) = 0$.
2. *Spectral density*. Consider a weakly dependent time series $\{X_t\}$, with the jth autocovariance $\gamma_X(j)$. The parameter of interest is the spectral density of $\{X_t\}$ at the frequency $\omega \in [-\pi, \pi]$: $\theta = (2\pi)^{-1}\sum_{j=-\infty}^{\infty}\gamma_X(j)e^{-i\omega j}$.

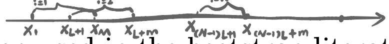
Note that example (1) deals with a parameter of a finite dimensional distribution, and example (2) is concerned with a parameter of an infinite dimensional joint distribution.

The paper is organized as follows. In Section 2 some basic concepts that will be used repeatedly throughout this paper are laid out. Section 3 considers the empirical likelihood for GEE, originally analyzed by Qin and Lawless (1994) in an 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. 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. Then we show that the blockwise empirical likelihood ratio statistic is Bartlett correctable. Section 5 offers some conclusions. Proofs of theorems are included in the Appendix.

2. Weak dependence and data blocking. In this section we state some important concepts that are used throughout the subsequent development. (我早) In this paper we allow for weakly dependent processes; in particular, we consider the following form of dependence. Throughout the paper, $\{X_t\}$ denotes an $\mathbb{R}^d$-valued stationary stochastic process, satisfying the following strong mixing condition:

$$\alpha_X(k) \to 0, \qquad k \to \infty,$$

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



We use blocking methods that have been used in the bootstrap literature; the reader is referred to Politis and Romano (1992) for an example. 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 $(X_{(i-1)L+1}, \ldots, X_{(i-1)L+M})$. Note that M is the "window width," whereas L is the separation between block starting points. Also define $Q = [(N - M)/L] + 1$, where $[\cdot]$ is the integer part of $\cdot$. 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. Define $A_N = QM/N$.

We introduce a more general blocking scheme to deal with inference regarding parameters of the infinite dimensional joint distribution. Define the sth "block of blocks" $\beta_s = (B_{(s-1)h+1}, \ldots, B_{(s-1)h+b})$. Then b and h are dependent on Q or N. Let $q = [(Q - b)/h] + 1$ and $a_N = qb/Q$; they are analogs of Q and A_N above.

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, e.g., 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-squared 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 point estimator 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 $f\colon \mathbb{R}^d \times \Theta \to \mathbb{R}^r$, which is assumed to satisfy the moment condition

约束数

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

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 ith block of observations as defined in Section 2 and the mapping ϕ_M: $\mathbb{R}^{rM} \times \Theta \to \mathbb{R}^r$ has the following particular form:

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

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

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

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 $\gamma = (\gamma_1, \dots, \gamma_r)'$:

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

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

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

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 denote the empirical likelihood without constraints, that is, 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

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

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

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

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

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

$$H_0\colon \Psi(\theta_0) = \psi, \tag{3.7}$$

where $\psi \in \mathbb{R}^q$, $q \le p$ and $\Delta = \partial\Psi/\partial\theta'|_{\theta_0}$ 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

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

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. Similarly, the second statistic LR_1 is a blockwise version of "*ELR*" in Qin and Lawless (1995), who considered estimating equations with $p = r$.

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 $\Gamma(z, \delta)$ be an open sphere with center z and radius δ; $\|\cdot\|$ 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* $\delta > 0$ *and* $\eta > 0$, $\mathbf{E}\sup_{\theta^* \in \Gamma(\theta, \delta)} \|f(X_t, \theta^*)\|^{2(1+\eta)} < \infty$ *for all* $\theta \in \Theta$;

(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 *except perhaps 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) $\operatorname{Var}(N^{-1/2}\sum_{t=1}^{N} f(X_t, \theta_0)) \to S > 0 (N \to \infty)$;

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

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

Then

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

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 $(D'W^{-1}D)^{-1}D'W^{-1}SW^{-1} \times D(D'W^{-1}D)^{-1}$ where $W = \mathbf{E}f(x, \theta_0)f(x, \theta_0)'$. 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

$$\bar{f}(\theta)'\hat{S}^{-1}\bar{f}(\theta), \tag{3.9}$$

where $\bar{f}(\theta) = N^{-1}\sum f(x_t, \theta)$ 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 $\overline{D}\overline{S}\bar{f}(\theta) = o_p(1/\sqrt{N})$, where $\overline{D} = N^{-1}\sum \partial f(x_i, \theta_0)/\partial\theta'$ and $S = Q^{-1}M\sum T_i(\theta_0)T_i(\theta_0)'$. When $L = 1$ (the "fully overlapped" case), the matrix 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 (3.9) in applications of GMM. If $L \to \infty$ as $N \to \infty$, 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), Priestley (1981) and Politis and Romano (1993b).

THEOREM 2. *Suppose all the assumption in Theorem* 1 *hold. Then*:

(i) $LR_0 \to_d \chi^2_{r-p}$;

(ii) *Under* H_0, $LR_1 \to_d \chi^2_q$.

REMARK. As in Qin and Lawless (1995), 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:

$$\begin{aligned}\text{Wald} &= N\big(\Psi(\hat{\theta}) - \psi\big)'\big(\hat{\Delta}\hat{V}_\theta\hat{\Delta}'\big)^{-1}\big(\Psi(\hat{\theta}) - \psi\big),\\ LM &= M^{-2}N\gamma_N(\hat{\theta}^c)'\hat{D}\hat{V}_\theta\hat{D}'\gamma_N(\hat{\theta}^c),\end{aligned}$$

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 [see also Kitamura and Stutzer (1995) and Imbens, Johnson and Spady (1996)]. It can be shown that these statistics have the same chi-squared limiting distribution with q degrees of freedom in the same manner as the results in Theorem 2.

4. Smooth function model.

4.1. *Blockwise empirical likelihood for the smooth function model.* Consider the smooth function model $\theta = H(\mu)$, where $\theta \in \mathbb{R}^p$ and $\mu \in \mathbb{R}^r$ $(p \le r)$. As in Politis and Romano (1992), let $\phi_M\colon \mathbb{R}^{dM} \to \mathbb{R}^r$ denote the mapping of blocks of observations and define the "estimating function" $T_i = \phi_M(B_i)$. Let $\overline{T} = \sum_{i=1}^{Q} T_i/Q$.

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

$$\sup\left\{\prod_{i=1}^{Q} p_i \,|\, p_i > 0,\ \sum_1^Q p_i = 1,\ H\left(\sum_1^Q p_i T_i\right) = \theta\right\}. \tag{4.1}$$

However, in order to cover inference for the infinite dimensional joint distribution, we need to use the "blocks of blocks" technique. Let $\Phi_b\colon \mathbb{R}^{rb} \to \mathbb{R}^r$ be a mapping such that $\Phi_b(\beta_s) = \sum_{i=1}^{b} T_{(s-1)h+i}/b$, 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:

$$L(\theta) = \sup\left\{\prod_{s=1}^{q} p_s \,|\, p_s > 0,\ \sum_1^q p_s = 1,\ H\left(\sum_1^q p_s U_s\right) = \theta\right\}. \tag{4.2}$$

Note that 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 $\overline{U} = \sum_{s=1}^{q} U_s/q$. Hence the blockwise empirical likelihood ratio statistic is

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

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-squared 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) $L = A^{-1}M$ *for some* $A \geq 1$;
(ii) $b \to \infty$, $h = O(b)$ *and* $b = o(Q^{1/2})$;
(iii) $\mathbf{E}\|T_i\|^{2c} < K$ *for* $c > 1$ *defined in Section* 2, *some* $K < \infty$ *and all* M;
(iv) $\mathbf{E}T_i = \mu_0 + o(Q^{-1/2})$;
(v) $\mathrm{Var}(\sqrt{Q}\,\overline{T}) \to \Lambda$ *as* $Q \to \infty (N \to \infty)$;
(vi) H: $\mathbb{R}^r \to \mathbb{R}^p$ *is continuously differentiable and* $\mathrm{rank}(\partial H/\partial\mu'|_{\mu_0}) = p$.

Then

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

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 (1993a) 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.

4.2. *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); see Corcoran, Davison and Spady (1995) for more information. It should be noted that Mykland (1995) showed a general Bartlett-correctability result using the concept "dual likelihood." In Mykland's analysis, continuous time models are allowed, but martingale properties are maintained. In contrast, we confine ourselves to discrete time models, but our data generating process may not be martingale. Note that there is no need for blocking martingale difference sequences. 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 (4.1). In particular, we consider the nonoverlapping blocking method, such 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 $\Sigma_N^{-1/2} T_i = \Xi_i = (\Xi_i^1, \dots, \Xi_i^r)'$. For a sequence of d integers satisfying $0 < k(1) < \cdots < k(d) = k$, $k \geq 3$, define

$$\begin{aligned}
&\tilde{\kappa}^{j_1 \cdots j_{k(1)}, j_{k(1)+1} \cdots j_{k(2)}, j_{k(2)+1} \cdots j_{k(d-1)}, j_{k(d-1)+1} \cdots j_{k(d)}} \\
&\quad = Q^{-1} \sum_{1 \leq i(1), \dots, i(d) \leq Q} \mathbf{E}\Big\{ M^{-1}\Big(M^{k(1)} \Xi_{i(1)}^{j_1} \cdots \Xi_{i(1)}^{j_{k(1)}}\Big) \\
&\qquad\qquad \times \Big(M^{k(2)} \Xi_{i(2)}^{j_{k(1)+1}} \cdots \Xi_{i(2)}^{j_{k(2)}}\Big) \\
&\qquad\qquad \times \cdots \Big(M^{k(d)} \Xi_{i(d)}^{j_{k(d-1)+1}} \cdots \Xi_{i(d)}^{j_{k(d)}}\Big)\Big\} \\
&\qquad\qquad \times I\Big\{ \max_{p, q < d} |i(p) - i(q)| \leq k - 2 \Big\},
\end{aligned} \tag{4.4}$$

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

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

Our Bartlett factor is

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

where

$$\begin{aligned}
\mu &= (\mu^1, \dots, \mu^r)', \quad \bar{\mu}_0 = \Sigma_N^{-1/2} \mu_0, \\
\overline{H}(\mu) &= H\big(\Sigma_N^{1/2} \mu\big), \quad \overline{H}^l_{j_1 \cdots j_k} = \partial^k \overline{H}^l(\mu) / \partial\mu^{j_1} \cdots \partial\mu^{j_k} \big|_{\mu = \bar{\mu}_0}, \\
\nabla \overline{H} &= \big(\overline{H}^i_j\big), \quad G = (\nabla\overline{H}\, \nabla\overline{H}')^{-1}, \quad W = \Delta \overline{H}' G\, \nabla\overline{H}, \quad N = \nabla\overline{H}'\, G, \\
t_{1a} &= (1/3)\, \kappa^{jkl} \tilde{\kappa}^{m,n,o} W^{jo} W^{km} W^{ln}, \\
t_{1b} &= (3/8)\, \tilde{\kappa}^{jk,l} \tilde{\kappa}^{nm,o} W^{jo} W^{km} W^{ln} \\
&\quad - (5/6)\, \kappa^{jkl} \tilde{\kappa}^{mn,o} (W^{jm} W^{kn} W^{lo} + W^{jo} W^{km} W^{ln}) \\
&\quad + (8/9)\, \kappa^{jkl} \kappa^{mno} W^{jm} W^{kn} W^{lo},
\end{aligned}$$

$$
\begin{aligned}
t_{1c} &= (1/4)\,\kappa^{jkl}\tilde{\kappa}^{mn,o}W^{jm}W^{ko}W^{ln}\\
&\quad +\big\{(-2/3)\,\kappa^{jkl}\tilde{\kappa}^{mn,o} + (2/9)\,\kappa^{jkl}\kappa^{mno}\big\}W^{jm}W^{kn}W^{lo},\\
t_{2a} &= (3/8)\,\tilde{\kappa}^{jk,l}\tilde{\kappa}^{mn,o}W^{jl}W^{km}W^{no}\\
&\quad +\big\{(-5/6)\,\kappa^{jkl}\tilde{\kappa}^{mn,o} + (4/9)\,\kappa^{jkl}\kappa^{mno}\big\}W^{jk}W^{lm}W^{no},\\
t_{2b} &= \big\{(1/4)\,\tilde{\kappa}^{jk,l}\tilde{\kappa}^{mn,o} - (1/3)\,\kappa^{jkl}\tilde{\kappa}^{mn,o}\big\}W^{jm}W^{kl}W^{no}\\
&\quad +(1/9)\,\kappa^{jkl}\kappa^{mno}W^{jm}W^{kl}W^{no},\\
t_{3a} &= (-1/2)\,\tilde{\kappa}^{jk,l,m}W^{jm}W^{kl},\\
t_{3b} &= \big\{(3/8)\,\tilde{\kappa}^{jl,km} + \tilde{\kappa}^{jkl,m} - (3/4)\,\kappa^{jklm}\big\}W^{jk}W^{lm},\\
t_{3c} &= (1/4)\,\tilde{\kappa}^{jk,lm}W^{jl}W^{km},\\
t_{4a} &= (1/2)\,\kappa^{jkl}N^{ju}\overline{H}^{u}_{mn}(I-W)^{mk}(I-W)^{nl},\\
t_{4b} &= -\,\tilde{\kappa}^{jk,l}N^{ju}\overline{H}^{u}_{mn}(I-W)^{mk}(I-W)^{nl},\\
t_{5} &= (1/4)G^{uv}\overline{H}^{u}_{jk}\overline{H}^{v}_{lm}\big\{(I-W)^{jk}(I-W)^{lm} + 2(I-W)^{jl}(I-W)^{km}\big\},\\
t_{6a} &= \big\{(-1/4)\,\tilde{\kappa}^{jk,l} + (1/3)\,\kappa^{jkl}\big\}N^{ju}\overline{H}^{u}_{mn}(I-W)^{mn}W^{kl},\\
t_{6b} &= \big\{-(1/2)\,\tilde{\kappa}^{jk,l} + (1/3)\,\kappa^{jkl}\big\}N^{ju}\overline{H}^{u}_{mn}(I-W)^{mn}W^{kl}.
\end{aligned}
$$

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 [Bhattacharya and Ghosh (1978)] 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 (1) the existence of sufficiently many moments, (2) a conditional Cramer condition and (3) 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 [see also Lahiri (1991, 1992)].

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, 2 and 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 the Appendix we shall show that

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

Moreover,

$$\mathbf{P}\{LR(\theta_0)(1 - N^{-1}a) \le z\} = \mathbf{P}\{\chi_p^2 \le z\} + O(N^{-5/6}). \tag{4.5b}$$

That is, the blockwise empirical likelihood ratio statistic is Bartlett correctable. The coverage error of confidence intervals is improved up to the order of $O(N^{-5/6})$. This rate is slower than the rate of N^{-2} obtained for the standard empirical likelihood assuming iid samples [DiCiccio, Hall and Romano (1991)], since our nonparametric treatment of dependence slows it down. [A similar phenomenon is observed for the blockwise bootstrap; see, e.g., Götze and Künsch (1996).] Nevertheless, these results demonstrate that the (Bartlett-corrected) empirical likelihood with blocking is a 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 $\{X_i\}$ 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 estimates using sample moments of T_i, or by the bootstrap. This replacement does not affect the conclusion of the above result.

5. 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. In practical applications, we need to select block length and the length of time shift. The sensitivity of our method to the choice of these parameters needs to be investigated.

APPENDIX

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

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

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, e.g., 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 $\max_Q \|T_i(\theta_0)\| = o(N^{1/2}M^{-1})$ 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. Define $C_N = \{x\colon \|f(x,\theta)\| \le N^{1/(2+2\eta)}$, all $\theta \in \Theta\}$, and $f_N(x,\theta) = f(x,\theta)I\{C_N\}$. Let $q_{\theta,N}(g) = \mathbf{E}[-\log(1+g'f_N(X_t,\theta))]$ for small g. Then $\lim_{N\to\infty}(\partial/\partial g)q_{\theta,N}(g) = \mathbf{E}f(x,\theta)$ uniformly in $g \in \Gamma(0, N^{-1/(2+\eta)})$. Let $\Gamma_N = \{g\colon\ g = N^{-1/(2+\eta)}u,\ \|u\| = 1\}$, $g_N(\theta) = \operatorname{argmin}_{g\in\Gamma_N} \mathbf{E}[-\log(1+g'f_N(X_t,\theta))]$ and $u_N(\theta) = g_N(\theta)/\|g_N(\theta)\|$. Using the mean value theorem, the minorant is approximated by

$$\mathbf{E}\big[-N^{1/(2+\eta)}\log(1+g_N(\theta)'f_N(X_t,\theta))\big] = -\|\mathbf{E}f(x,\theta)\| + o(1), \tag{A.2}$$

with $\lim_{N\to\infty} u_N(\theta) = \mathbf{E}f(x,\theta)/\|\mathbf{E}f(x,\theta)\|$. By assumption (iv),

$$\begin{aligned}\lim_{N\to\infty}\lim_{\delta\downarrow 0} N^{1/(2+\eta)}\mathbf{E}\sup_{\theta^*\in\Gamma(\theta,\delta)} -\log(1+g_N(\theta^*)'f_N(X_t,\theta^*))\\ = -\|\mathbf{E}f(x,\theta)\|.\end{aligned} \tag{A.3}$$

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

$$N^{1/(2+\eta)}\mathbf{E}\sup_{\theta^*\in\Gamma(\theta_j,\delta_j)} -\log(1+g_N(\theta^*)'f_N(X_t,\theta^*)) + o(1) = -2H_j, \qquad j = 1,\dots,h,$$

for positive numbers H_j, $j = 1,\dots,h$. Note assumption (iii) implies that $\max_t \sup_{\theta^*\in\Gamma(\theta_j,\delta)}\|f(X_t,\theta^*)\| = o(N^{1/(2+2\eta)})$ with probability 1 as $N\to\infty$ [see Lemma 3 of Owen (1990)]. Thus there exists a sufficiently large integer N_j such that for small $\varepsilon > 0$,

$$\begin{aligned}&\mathbf{P}\Big\{(1/N)\sum_t \sup_{\theta^*\in\Gamma(\theta_j,\delta_j)} -\log(1+g_N(\theta^*)'f(X_t,\theta^*)) > -N^{-1/(2+\eta)}H_j\Big\}\\ &\quad< \varepsilon/(2h), \qquad j = 1,\dots,h,\end{aligned}$$

for all $N > N_j$ [note $f(\cdot,\cdot)$, not $f_N(\cdot,\cdot)$, is used]. These h inequalities imply

$$\begin{aligned}&\mathbf{P}\Big\{\sup_{\theta^*\in\Theta(\delta)} (1/N)\sum_t -\log(1+g_N(\theta^*)'f(X_t,\theta^*)) > -N^{-1/(2+\eta)}H\Big\}\\ &\quad< \varepsilon/2, \qquad H = \min_j H_j,\end{aligned}$$

for all $N > \max_j N_j$. Now note that the optimality of $\gamma_N(\theta)$ implies

$$\begin{aligned}&(1/Q)\sum_i -\log(1+\gamma_N(\theta)'T_i(\theta))\\ &\quad\le (1/N)\sum_t -\log(1+g_N(\theta)'f(X_t,\theta)) + o_p(M/N).\end{aligned}$$

Therefore there exists a sufficiently large integer N_A such that

$$\begin{aligned}&\mathbf{P}\Big\{\sup_{\theta^*\in\Theta(\delta)} (1/Q)\sum_i -\log(1+\gamma_N(\theta^*)'T_i(\theta^*)) > -N^{-1(2+\eta)}H\Big\}\\ &\quad< \varepsilon/2.\end{aligned} \tag{A.4}$$

for all $N > N_A$. Notice that

$$-\gamma_N(\theta_0)'(1/Q)\sum_i T_i(\theta_0) \leq (1/Q)\sum_i -\log(1+\gamma_N(\theta_0)'T_i(\theta_0)) \leq 0,$$

where the first term is $O_p(M/N^{1/2})O_p(N^{-1/2}) = o_p(N^{-1/2})$. Thus there exists a large integer N_B such that

$$\text{(A.5)} \qquad \mathbf{P}\left\{(1/Q)\sum_i -\log(1+\gamma_N(\theta_0)'T_i(\theta_0)) < -N^{-1/2}H\right\} < \varepsilon/2$$

for all $N > N_B$. By (A.4) and (A.5), for any small δ, $\mathbf{P}\{\hat{\theta} \in \Gamma(\theta_0, \delta)\} \geq 1 - \varepsilon$ 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

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

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

Also define

$$l_{\gamma\gamma}(\theta,\gamma) = (\partial/\partial\gamma)l_\gamma(\theta,\gamma), \qquad l_{\theta\gamma}(\theta,\gamma) = (\partial/\partial\gamma)l_\theta(\theta,\gamma)$$

and

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

As in Qin and Lawless (1994), the consistency and assumption (v) imply the following FOCs:

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

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

$$\begin{aligned} 0 &= Q^{-1}N^{1/2}\Sigma T_i(\theta_0) + Ml_{\gamma\gamma}(\theta^*,\gamma^*)N^{1/2}M^{-1}(\hat{\gamma}-0) \\ &\quad + l_{\gamma\theta}(\theta^*,\gamma^*)N^{1/2}(\hat{\theta}-\theta_0), \\ 0 &= 0 + l_{\theta\gamma}(\theta^*,\gamma^*)N^{1/2}M^{-1}(\hat{\gamma}-0) + M^{-1}l_{\theta\theta}(\theta^*,\gamma^*)N^{1/2}(\hat{\theta}-\theta_0), \end{aligned}$$

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

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

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

PROOF OF THEOREM 2. (i) Let $\overline{T}(\theta) = \Sigma_{i=1}^{Q} T_i(\theta)/Q$. By the asymptotic results in Theorem 1, $M^{-1}N^{1/2}\hat{\gamma} = S^{-1}N^{1/2}\overline{T}(\hat{\theta}) + o_p(1)$ and $\sqrt{N}\,\overline{T}(\hat{\theta}) \to_d N(0, [S - DV_\theta D'])$. Then we obtain

$$LR_0 = 2A_N^{-1}\Sigma \log\bigl(1 + \hat{\gamma}'T_i(\hat{\theta})\bigr) = N\overline{T}(\hat{\theta})'S^{-1}\overline{T}(\hat{\theta}) + o_p(1) \to_d \chi^2_{r-p}.$$

(ii) The Lagrangean for the constrained estimation is $\mathscr{L} = \log(L(\theta)) + \zeta'(\psi - \Psi(\theta))$, 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:

$$\frac{\sqrt{N}}{MQ}\sum_{i=1}^{Q}\frac{\partial T_i(\hat{\theta}^c)/\partial\theta'}{1 + \gamma_N(\hat{\theta}^c)'T_i(\hat{\theta}^c)}\gamma_N(\hat{\theta}^c)$$
$$= \left\{N^{-1}\sum_{t=1}^{N}\bigl(\partial f\bigl(X_t, \hat{\theta}^c\bigr)/\partial\theta'\bigr)\right\}'S^{-1}\left\{N^{-1/2}\sum_{t=1}^{N} f\bigl(X_t, \hat{\theta}^c\bigr)\right\} + o_p(1).$$

Since the approximation term of the right-hand side is the optimally weighted estimating functions, the stated chi-squared limiting distributions of likelihood-ratio type statistics are obtained by the conventional argument for nonlinear dynamic models [see, e.g., 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

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

Also note

$$q^{-1}\sum_{s=1}^{q} U_s U_s' = (bq)^{-1}\sum_{s=1}^{q}\left[b^{-1}\left(\sum_{i=1}^{b} T_{(s-1)h+1}\right)^2\right] = b^{-1}\bigl(\Lambda + o_p(1)\bigr).$$

Using these results, as in Owen's proof we obtain

$$\begin{aligned} LR(\theta_0) &= a_N^{-1}bq\bigl(\overline{U} - \mu_0\bigr)'\Lambda^{-1}\bigl(\overline{U} - \mu_0\bigr) + o_p(1) \\ &\to_d \chi_p^2. \end{aligned}$$ □

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

$$C_i^{j_1 - j_k} = T_i^{j_1} \cdots T_i^{j_k} - M^{-(k-1)} \kappa^{j_1 \cdots j_k}$$

and

$$K^{j_1 \cdots j_k} = Q^{-1} M^{k-1} \sum_{i=1}^{Q} C_i^{j_1 \cdots j_k}.$$

First we consider the empirical likelihood for the mean parameter with $\mu_0 = 0$ and $\Sigma_N = I$. Define R_1, R_2 and R_3 as in DHR (1991) with α, A, θ, Q and M replaced by κ, K, H, G and W. Then moment bounds by Yokoyama (1980) and Kim (1993) imply that $LR(\theta_0)$ is approximated using $R = R_1 + R_2 + R_3$:

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

Next we calculate the third and fourth cumulants of R. For our purpose, it is enough to show that $\operatorname{cum}(R^u R^v R^w) = O(N^{-7/3})$ and $\operatorname{cum}(R^u R^v R^w R^x) = O(N^{-17/6})$. 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 kth order $\tilde{\kappa}$ only involves moments of T_i's within $k-1$ consecutive periods; see definition (4.4). This is a consequence of the mixing condition (iii).] To this end, note

$$\begin{aligned} \alpha_T(k) &\le \alpha_X(kL - M) \le c \exp[-d(kL - M)] \\ &= c \exp[-d(k-1)M] \\ &= c \exp[-d(k-1)N^{1/3}] \end{aligned} \tag{A.6}$$

for $k \ge 2$, by assumptions (i) and (iv). Equation (A.6) implies that T_i is an asymptotically 1-dependent process with small asymptotic approximation errors. The mixing inequality and (A.6) imply formulas as in Step 6 of DHR (1988), though they need to be appropriately modified to take account of serial correlations among T_i's. Moreover, under certain moment conditions,

$$\begin{aligned} \mathbf{E}\left(T_i^j T_{i+1}^k T_{i+2}^l\right) &< C \alpha_X^{1-r}([M/2] + 1) \le Cc \exp(-d(1-r)([M/2] + 1)) \\ &= O\left(\exp(-d(1-r)N^{1/3}/2)\right), \end{aligned}$$

where C is a positive constant and $r < 1$. To see this, notice that third-order moments of X_i, $\mathbf{E} X_\tau X_{\tau+p} X_{\tau+p+q}$, say, that appear in $\mathbf{E}(T_i^j T_{i+1}^k T_{i+2}^l)$ 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, $\mathbf{E}(T_i^j T_{i+1}^k T_{i+2}^l T_{i+3}^m) = O(\exp(-d(1-r')N^{1/3}/2))$ for some $r' < 1$.

Using the above results, it can be shown that $\mathrm{cum}(R^u R^v R^w) = O(N^{-7/3})$. The error term of order $O(N^{-7/3})$ is due to the bias $\mathbf{E}(K^{j_1 j_2}) = O(M^{-1})$ through $\mathbf{E}(R_2^u R_1^v R_1^w)$ [see DHR (1988), equation (3.9)]. Note that $K^{j_1 \cdots j_k}$, $k \geq 3$ has no such bias by definition. The expectation $\mathbf{E}(R_2^u R_1^v R_1^w)$ includes the term $\mathbf{E}(K^{jk} K^l K^m K^n) = \mathbf{E}[(K^{jk} - \mathbf{E}K^{jk}) K^l K^m K^n] + O(M^{-1}N^{-2})$. The last remainder term does not cancel with any other terms and it determines the order of the third-order cumulants. Other error terms are $O(N^{-5/2})$. Similar calculations show that $\mathrm{cum}(R^u R^v R^w R^x) = O(N^{-3})$.

The above results imply that the third- and fourth-order cumulants of $N^{1/2}R$ are $O(N^{-5/6})$ and it can be shown that the sth order cumulant of $NR'R\{\mathbf{E}[NR'R/p]\}^{-1}$ coincides with that of χ_p^2 up to errors of order $O(N^{-5/6})$. Given the validity of Edgeworth expansions, we have

$$\mathbf{P}\{LR(\theta_0)/\mathbf{E}[NR'R/p] \leq z\} = \mathbf{P}\left\{\chi_p^2 \leq z\right\} + O(N^{-5/6}).$$

Finally, a straightforward calculation shows that $\mathbf{E}(NR'R) = p + N^{-1}a + O(N^{-5/6})$, which implies (4.5). (Recall $a = O(M)$.) For the general case $\Sigma_N \neq I$, we replace $H(\lambda)$ with $\overline{H}(\lambda) = H(\Sigma_N^{1/2}\lambda)$, and the desired result follows. □

Acknowledgments. The helpful comments of two referees and an Associate Editor are gratefully acknowledged.

REFERENCES

BHATTACHARYA, R. N. and GHOSH, J. K. (1978). On the validity of the formal Edgeworth expansion. *Ann. Statist.* **6** 434–451.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. 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.

CARLSTEIN, E. (1986). The use of subseries values for estimating the variance of a general statistic from stationary sequence. *Ann. Statist.* **14** 1171–1179.

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. (1993). On the accuracy of empirical likelihood confidence regions for linear regression models. *Ann. Inst. Statist. Math.* **45** 621–637.

CHEN, S. X. (1994a). Comparing empirical likelihood and bootstrap hypothesis tests. *J. Multivariate Anal.* **51** 277–293.

CHEN, S. X. (1994b). Empirical likelihood confidence intervals for linear regression coefficients. *J. Multivariate Anal.* **49** 24–40.

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

CORCORAN, S. A., DAVISON, A. C. and SPADY, R. H. (1995). Reliable inference from empirical likelihoods. Mimeo, Dept. Statistics, Oxford Univ.

DAVISON, A. C. and HALL, P. (1993). On studentizing and blocking methods for implementing the bootstrap with dependent data. *Austral. J. Statist.* **35** 215–224.

DICICCIO, T., HALL, P. and ROMANO, J. (1988). Bartlett adjustment for empirical likelihood. Technical Report 298, Dept. Statistics, Stanford Univ.

DICICCIO, T., HALL, P. and ROMANO, J. (1989). Comparison of parametric and empirical likelihood functions. *Biometrika* **76** 465–476.

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

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. *Internat. Statist. Rev.* **58** 59–76.

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

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

Götze, F. and Hipp, C. (1983). Asymptotic expansions for sums of weakly dependent random vectors. *Z. Wahrsch. Verw. Gebiete* **64** 211–239.

Götze, F. and Künsch, H. R. (1996). Second order correctness of the blockwise bootstrap for stationary observations. *Ann. Statist.* **24** 1914–1933.

Hall, P. (1985). Resampling a coverage process. *Stochastic Process. Appl.* **19** 259–269.

Hall, P. (1990). Pseudo-likelihood theory for empirical likelihood. *Ann. Statist.* **18** 121–140.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. Springer, New York.

Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and Its Applications*. Academic Press, San Diego.

Hall, P. and La Scala, B. (1990). Methodology and algorithms of empirical likelihood. *Internat. Statist. Rev.* **58** 109–127.

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

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 Probab. Appl.* **7** 349–382.

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

Imbens, G., Johnson, P. and Spady, R. (1996). Information theoretic approaches to inference in moment condition models. Mimeo, Dept. Economics, Harvard Univ.

Kim, T. Y. (1993). A note on moment bounds for strong mixing sequences. *Statist. Probab. Lett.* **16** 163–168.

Kitamura, Y. and Stutzer, M. (1995). An information-theoretic alternative to generalized method of moments estimation. Mimeo, Depts. Economics and Finance, Univ. Minnesota.

Kolaczyk, E. D. (1994). Empirical likelihood for generalized linear models. *Statist. Sinica* **4** 199–218.

Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *Ann. Statist.* **17** 1217–1241.

Lahiri, S. N. (1991). Second order optimality of stationary bootstrap. *Statist. Probab. Lett.* **11** 335–341.

Lahiri, S. N. (1992). Edgeworth correction by 'moving block' bootstrap for stationary and nonstationary data. In *Exploring the Limits of Bootstrap* (R. LePage and L. Billard, eds.) 183–214. Wiley, New York.

Mykland, P. A. (1995). Dual likelihood. *Ann. Statist.* **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. *Ann. Statist.* **18** 90–120.

Owen, A. (1991). Empirical likelihood for linear models. *Ann. Statist.* **19** 1725–1747.

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. *Ann. Statist.* **20** 1985–2007.

Politis, D. N. and Romano, J. P. (1993a). Nonparametric resampling for homogeneous strong mixing random fields. *J. Multivariate Anal.* **47** 301–328.

Politis, D. N. and Romano, J. P. (1993b). On the sample variance of linear statistics derived from mixing sequences. *Stochastic Process. Appl.* **45** 155–167.

Politis, D. N. and Romano, J. P. (1994). Large sample confidence regions based on subsamples under minimal assumptions. *Ann. Statist.* **22** 2031–2050.

PRIESTLEY, M. D. (1981). *Spectral Analysis and Time Series*. Academic Press, New York.

QIN, J. (1993). Empirical likelihood in biased sample problems. *Ann. Statist.* **21** 1182–1196.

QIN, J. and LAWLESS, J. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.* **23** 300–325.

QIN, J. and LAWLESS, J. (1995). Estimating equations, empirical likelihood and constraints on parameters. *Canad. J. Statist.* **23** 300–325.

ROSENBLATT, M. (1985). *Stationary Sequences and Random Fields*. Birkhäuser, Boston.

SINGH, K. (1981). On the asymptotic accuracy of Efron's bootstrap. *Ann. Statist.* **9** 1187–1195.

WALD, A. (1949). Note on the consistency of the maximum likelihood estimate. *Ann. Math. Statist.* **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. *Ann. Math. Statist.* **20** 601–603.

YOKOYAMA, R. (1980). Moment bounds for stationary mixing sequences. *Z. Wahrsch. Verw. 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.

DEPARTMENT OF ECONOMICS
UNIVERSITY OF MINNESOTA
MINNEAPOLIS, MINNESOTA 55455
E-MAIL: kitamura@atlas.socsci.umn.edu