Statistica Sinica **12**(2002), 387-407

EMPIRICAL LIKELIHOOD FOR AUTOREGRESSIVE MODELS, WITH APPLICATIONS TO UNSTABLE TIME SERIES

Chin-Shan Chuang and Ngai Hang Chan

Goldman, Sachs and Co. and Chinese University of Hong Kong

Abstract: Empirical likelihood is developed for autoregressive models with innovations that form a martingale difference sequence. Limiting distributions of the log empirical likelihood ratio statistic for both the stable and unstable cases are established. Behavior of the log empirical likelihood ratio statistic is considered in nearly nonstationary models to assess the local power of unit root tests and to construct confidence intervals. Resampling methods are proposed to improve the finite-sample performance of empirical likelihood statistics. This paper shows that empirical likelihood methodology compares favorably with existing methods and demonstrates its potential for time series with more general innovation structures.

Key words and phrases: Empirical likelihood, dual likelihood, autoregressive models, unit root tests.

1. Introduction

Empirical likelihood was introduced by Owen (1988, 1990) as a way to extend likelihood based inference ideas to certain nonparametric situations. In the simplest situation, one is interested in obtaining a confidence region for the mean μ of some unknown distribution F which gives rise to n independent and identically distributed (i.i.d.) $(p \times 1)$ observations $X_1, \ldots, X_n$. Without specifying a parametric form for F, Owen (1988, 1990) obtains the empirical likelihood ratio $L(\mu_0)$ for testing $\mu = \mu_0$ and constructs the corresponding confidence region for μ with approximate coverage probability $1 - \alpha$ given by

$$\{\mu_0 : l(\mu_0) \leq c_{1-\alpha}\}, \tag{1.1}$$

where $c_{1-\alpha}$ is the $1 - \alpha$ quantile of χ_p^2 and χ_p^2 is a chi-square random variable with p degrees of freedom. The confidence region (1.1) has the advantage that its shape is determined by the observations through $l(\mu)$ and is not necessarily ellipsoidal, as in the case for confidence regions based on normal approximations. Furthermore, empirical likelihood in this situation is Bartlett-correctable so that a simple analytic correction gives (1.1) an actual coverage probability that differs from the nominal coverage probability of $1-\alpha$ by a term of order n^{-2}. These attractive properties have motivated various authors to extend empirical likelihood methodology to other situations; see Owen (1991) for linear models, Kolaczyk (1994) for generalized linear models, and Qin and Lawless (1994) for connections with general estimating equations.

In this paper, empirical likelihood methodology is developed for autoregressive models with innovations that form a martingale difference sequence, and we focus on the use of this methodology for a general unstable autoregressive model. In Section 2, we study how one can define an empirical likelihood ratio for autoregressive models and we review some of the recent literature on empirical likelihood methodology for dependent data. In Section 3, we derive the limiting distribution of the log empirical likelihood ratio statistic introduced in Section 2 for both the stable and unstable cases, and consider a second-order model as an example. We characterize the limiting distribution for the unstable case using Brownian functionals, as in Chan and Wei (1988). A special case of our results appears in Wright (1999). In Section 4, the unstable case is studied further and unit root tests obtained via empirical likelihood in the first-order model are compared with existing tests in the literature by using simulations and local power analyses. We also construct confidence intervals for the largest autoregressive root in nearly nonstationary models. Section 5 concludes and the appendix contains the proofs.

2. Empirical Likelihood for Autoregressive Models

We first review the definition of empirical likelihood for the mean, as discussed by Owen (1988, 1990). Let $X_1, \ldots, X_n$ be i.i.d. $(p \times 1)$ random vectors having a common unknown distribution F, and let $\boldsymbol{X} = (X_1, \ldots, X_n)$. The nonparametric maximum likelihood estimate of F is the empirical distribution $\widehat{F}$ which puts equal mass n^{-1} on $X_1, \ldots, X_n$. Letting $\widehat{F}^{(\mu_0)}$ be the distribution that maximizes the empirical likelihood $L(\boldsymbol{X}; F) = \prod_{i=1}^n F(\{X_i\})$ among all distributions F with $F(\{X_i\}) > 0$ subject to the constraint that the mean of F is μ_0, Owen (1988, 1990) obtains the log empirical likelihood ratio statistic

$$l(\mu_0) = 2\{\log L(\boldsymbol{X}; \widehat{F}) - \log L(\boldsymbol{X}\widehat{F}^{(\mu_0)})\}. \tag{2.1}$$

He shows further that $l(\mu_0)$ has a limiting chi-square distribution with p degrees of freedom provided $E||X_1||^2 < \infty$. Let $p_i = F(\{X_i\})$. To maximize $L(\boldsymbol{X}, F) = \prod_{i=1}^n F(\{X_i\}) = \prod_{i=1}^n p_i$ subject to the constraint F has mean μ_0, one can introduce a Lagrange multiplier λ and arrive at the following equation for λ:

$$\sum_{i=1}^n (X_i - \mu_0)/\{1 + \lambda'(X_i - \mu_0)\} = 0, \tag{2.2}$$

where λ' denotes the transpose of λ, with $p_i^{-1} = n\{1 + \lambda'(X_i - \mu_0)\}$. Since $L(\boldsymbol{X}; \widehat{F}) = n^{-n}$, (2.1) can be written as

$$l(\mu_0) = 2\sum_{i=1}^n \log\{1 + \lambda'(X_i - \mu_0)\}. \tag{2.3}$$

By appealing to the chi-square approximation, $l(\mu_0)$ can be used to provide a confidence region for the mean using (1.1). More generally, if θ is associated with F, information about θ and F is available through r estimating equations $g_1, \ldots, g_r$ so that $E(g_i(X_1, \theta)) = 0$ for $i = 1, \ldots, r$. For simplicity, consider the case where $p = r$. A similar constrained optimization problem leads to the consideration of

$$l(\theta) = 2\sum_{i=1}^{n} \log\{1 + \lambda' g(X_i, \theta)\} \tag{2.4}$$

as the log empirical likelihood ratio statistic, where $g(X_i, \theta) = (g_1(X_i, \theta), \ldots, g_r(X_i, \theta))'$ and the Lagrange multiplier λ satisfies

$$\sum_{i=1}^{n} g(X_i, \theta)/\{1 + \lambda' g(X_i, \theta)\} = 0. \tag{2.5}$$

Note that (2.4) is simply (2.3) with $X_i - \mu_0$ replaced by $g(X_i, \theta)$, and a similar connection can be made between (2.5) and (2.2). This formulation is discussed by Owen (1988, 1990) for M-estimates, and by Qin and Lawless (1994) for general estimating equations, where they permit $r > p$.

Mykland (1995) generalizes the definition of empirical likelihood for i.i.d. data to statistical models with a martingale structure by using the concept of dual likelihood. In these models, which include conditional least squares estimation for autoregressive models, the derivative of the objective function with respect to the unknown parameter given by the "score" function is a martingale under the true parameter. The "score" function is then used to construct the dual likelihood ratio statistic. Consider an AR(p) model given by

$$y_t = \beta_1 y_{t-1} + \cdots + \beta_p y_{t-p} + \epsilon_t, \tag{2.6}$$

where y_t is the observation, ϵ_t is the unobservable random disturbance at time t, p is the order of the model, and $\beta_1, \ldots, \beta_p$ are the parameters of the model. In what follows, we assume that the disturbances $\{\epsilon_t\}$ form a martingale difference sequence with respect to an increasing sequence of σ-fields $\mathcal{F}_t$, i.e., ϵ_t is $\mathcal{F}_t$-measurable and $E(\epsilon_t|\mathcal{F}_{t-1}) = 0$ a.s. for every t. We also assume that $E(\epsilon_t^2|\mathcal{F}_{t-1}) = \sigma^2$ a.s. for every t. The initial values $y_0, \ldots, y_{1-p}$ are assumed to be $\mathcal{F}_1$-measurable so that y_t is $\mathcal{F}_t$-measurable. The unknown parameter $\beta = (\beta_1, \ldots, \beta_p)'$ can be estimated by the conditional least squares estimate $\widehat{\beta}$ which maximizes the conditional least squares criterion

$$-\frac{1}{2}\sum_{t=p+1}^{n} (y_t - E(y_t|\mathcal{F}_{t-1}))^2 = -\frac{1}{2}\sum_{t=p+1}^{n} (y_t - \beta' \mathbf{y}_{t-1})^2, \tag{2.7}$$

where $\mathbf{y}_t = (y_t, \ldots, y_{t-p+1})'$; see Hall and Heyde (1980, Section 6.3). Thus we have

$$\widehat{\beta} = \left(\sum_{t=p+1}^{n} \mathbf{y}_{t-1}\mathbf{y}'_{t-1} \right)^{-1} \sum_{t=p+1}^{n} \mathbf{y}_{t-1} y_t. \tag{2.8}$$

Partial differentiation of (2.7) with respect to β yields the "score" function $\sum_{t=p+1}^{n} (y_t - \beta' \mathbf{y}_{t-1})\mathbf{y}_{t-1} = \sum_{t=p+1}^{n} m_t$, where $m_t = (y_t - \beta' \mathbf{y}_{t-1})\mathbf{y}_{t-1}$. Let β_0 denote the true parameter value for β. When $\beta = \beta_0, m_t = \epsilon_t \mathbf{y}_{t-1}$ forms a martingale difference sequence, and the "score" function then forms a martingale. The estimate $\widehat{\beta}$ is the value of β which equates the "score" function to 0.

For autoregressive models, Mykland's approach leads to the dual/empirical likelihood ratio statisic

$$l(\beta) = 2 \sum_{t=p+1}^{n} \log(1 + \lambda' m_t), \tag{2.9}$$

where λ satisfies

$$\sum_{t=p+1}^{n} m_t/(1 + \lambda' m_t) = 0. \tag{2.10}$$

One could have arrived at the expressions (2.9) and (2.10), in analogy with (2.4) and (2.5), by treating m_t as i.i.d. with $E(m_t) = 0$. Note that (2.3) is well defined only if μ_0 lies in the convex hull of $X_1, \ldots, X_n$. Otherwise $l(\mu_0) = \infty$ since it is then impossible to reweight the data so that the weighted mean is μ_0, as pointed out by Owen (1990, p.106). Similarly, (2.10) is well-defined only if 0 lies in the convex hull of $m_{p+1}, \ldots, m_n$ and $l(\beta) = \infty$ otherwise. With this convention, the value of β which minimizes $l(\beta)$ is also the conditional least squares estimate $\widehat{\beta}$.

One reason that Mykland calls (2.9) a dual likelihood statistic is that it shares properties of parametric likelihood ratio statistics when viewed as a likelihood ratio statistic in the dual parameter λ for each fixed β. Let $l(\psi)$ be the log likelihood based on i.i.d. observations and $\widehat{\psi}$ be the maximum likelihood estimate of ψ. The log likelihood ratio statistic for testing $\psi = \psi_0$ is given by $2(l(\widehat{\psi}) - l(\psi_0))$. Let $l_\psi(\psi_0)$ and $l_{\psi\psi}(\psi_0)$ be the first and second derivatives of the log likelihood function evaluated at ψ_0. It is well known that the log likelihood ratio statistic is asymptotically equal to the quadratic score (Wald) statistic $-l'_\psi(\psi_0) l^{-1}_{\psi\psi}(\psi_0) l_\psi(\psi_0)$, where $l'_\psi(\psi_0)$ denotes the transpose of $l_\psi(\psi_0)$, or the score function at ψ_0, and $l^{-1}_{\psi\psi}(\psi_0)$ denotes the inverse of $l_{\psi\psi}(\psi_0)$, or the Hessian matrix. Now, when the dual parameter λ in (2.9) is regarded as freely varying and β is regarded as fixed, the first and second derivatives of (2.9) with respect to λ evaluated at $\lambda = 0$ are given by $\sum_{t=p+1}^{n} m_t$ and $-\sum_{t=p+1}^{n} m_t m'_t$. We shall see in the next section that familiar Taylor expansion arguments for (2.9) and (2.10), such as those given in Owen (1990) and Mykland (1995), show that $l(\beta)$ and

$$Q(\beta) = \sum_{t=p+1}^{n} m_t'(\sum_{t=p+1}^{n} m_t m_t')^{-1} \sum_{t=p+1}^{n} m_t \tag{2.11}$$

are asymptotically equivalent, and $Q(\beta)$ is the analogue of the parametric quadratic score statistic.

Recently, Kitamura (1997) and Monti (1997) define empirical likelihood for certain types of dependent data using different ideas. Kitamura considers models with weakly dependent observations and uses blocks of observations to define empirical likelihood by extending the approach of Qin and Lawless (1994) based on estimating equations. For autoregressive models, (2.9) and (2.10) seem more natural since we do not need to study the choice of block size. Monti considers autoregressive models with moving average errors and defines empirical likelihood in the frequency domain.

3. Asymptotic Distribution of Empirical Likelihood Ratio and Confidence Regions

In this section, the asymptotic distribution of the dual likelihood ratio (2.9) is obtained for both stable and unstable autoregressive models. Let

$$\phi(z) = 1 - \beta_1 z - \cdots - \beta_p z^p \tag{3.1}$$

denote the characteristic polynomial of the autoregressive model (2.6). Then (2.6) is termed stable when the roots of (3.1) lie outside the unit circle, and unstable when the roots of (3.1) lie either on or outside the unit circle, with at least one root on the unit circle. For the rest of the paper, we write $\sum$ for $\sum_{t=p+1}^{n}$. Also, we assume that the sequence $\{\epsilon_t\}$ satisfies the additional condition $\sup_{t\geq p} E(|\epsilon_t|^{2+\alpha}|\mathcal{F}_{t-1}) < \infty$ for some $\alpha > 0$. Proofs are given in the Appendix and all limits are taken as the sample size n tends to infinity. The following lemma provides the asymptotic distribution for $\widehat{\beta}$ and $l(\beta)$ in the stable case:

Lemma 1. *Assume all roots of* (3.1) *lie outside the unit circle. Then*
(i) $(\sum \mathbf{y}_{t-1}\mathbf{y}_{t-1}')^{1/2}(\widehat{\beta} - \beta)$ *converges in distribution to the normal distribution with mean 0 and dispersion matrix* $\sigma^2 I_p$, *where* I_p *is the* $p \times p$ *identity matrix.*
(ii) $l(\beta)$ *given by* (2.9) *converges in distribution to* χ_p^2.

The Appendix provides arguments that $l(\beta)$ and the quadratic score $Q(\beta)$ given by (2.11) are asymptotically equivalent and that $\sum m_t m_t' = \sum \mathbf{y}_{t-1}\mathbf{y}_{t-1}'\epsilon_t^2$ can be replaced by $\sigma^2 \sum \mathbf{y}_{t-1}\mathbf{y}_{t-1}'$ without affecting the asymptotic distribution of $l(\beta)$, so part (ii) then follows from part (i). Let $\widehat{\sigma}^2 = n^{-1}\sum(y_t - \widehat{\beta}'\mathbf{y}_{t-1})^2$. The following lemma states that $\widehat{\sigma}^2$ is strongly consistent for σ^2.

Lemma 2. *Assume all roots of* (3.1) *lie either on or outside the unit circle. Then* $\widehat{\sigma}^2 \to \sigma^2$ *a.s.*

Consequently, if the model is known to be stable, an approximate $1-\alpha$ confidence region for β can be based on either $l(\beta) = 2\sum_{t=p+1}^{n} \log(1+\lambda' m_t)$ defined in (2.9), or $S(\beta) = \widehat{\sigma}^{-2}(\widehat{\beta}-\beta)'(\sum \mathbf{y}_{t-1}\mathbf{y}'_{t-1})(\widehat{\beta}-\beta)$, and such a confidence region is given by $\{\beta : R(\mathbf{y},\beta) \le c_{1-\alpha}\}$, where $R(\mathbf{y},\beta)$ is either $l(\beta)$ or $S(\beta)$ and $c_{1-\alpha}$ is the $1-\alpha$ quantile of χ_p^2. The choice of $S(\beta)$ as the "root" $R(\mathbf{y},\beta)$ leads to the familiar ellipsoidal confidence region for β.

We now turn to the unstable case. An unstable AR(1) model is given by the random walk model with $\beta = \beta_1 = 1$. When ϵ_t are i.i.d. with common mean 0 and variance σ^2, White (1958) shows that $(\sum_{t=1}^{n-1} y_t^2)^{1/2}(\widehat{\beta}-1)/\sigma$ converges in distribution to

$$\mathcal{L}(0) = \frac{1}{2}(W(1)^2 - 1)/(\int_0^1 W(t)^2 dt)^{1/2}, \tag{3.2}$$

where $W(t)$ is standard Brownian motion. For a general unstable AR(p) model, and when ϵ_t forms a martingale difference sequence satisfying the conditions in Section 2, Chan and Wei (1988) show that $(\sum \mathbf{y}_{t-1}\mathbf{y}'_{t-1})^{1/2}$ is of the same order as a block diagonal non-random matrix $(Q'G'_n)^{-1}$ (Q and G_n are defined on page 379 of Chan and Wei (1988)), and characterize the limiting distribution of $(Q'G'_n)^{-1}(\widehat{\beta}-\beta)$ with stochastic integrals. The analogue of their result for empirical likelihood ratio statistics is given by the following theorem (the proof is given in the Appendix).

Theorem 1. *Assume all roots of* (3.1) *lie either on or outside the unit circle, with at least one root lying on the unit circle. Then* $Q(\beta), S(\beta)$ *and* $l(\beta)$ *all have the same limiting distribution.*

The limit distribution is characterized by Theorem 3.5.1 in Chan and Wei (1988), which provides the limiting distributions for the three terms in braces in $S(\beta) = \widehat{\sigma}^{-2}\{(\widehat{\beta}-\beta)'(G_nQ)^{-1}\}\{G_nQ\sum \mathbf{y}_{t-1}\mathbf{y}'_{t-1}Q'G'_n\}\{(Q'G'_n)^{-1}(\widehat{\beta}-\beta)\}$.

For the remainder of this section, we use an AR(2) model to explicitly illustrate the ideas. Let $\{y_t\}$ be an AR(2) process given by

$$y_t = \beta_1 y_{t-1} + \beta_2 y_{t-1} + \epsilon_t, \tag{3.3}$$

with characteristic roots $|z_1|$ and $|z_2|$ and $|z_1| \ge |z_2|$.

Corollary 1. *Let* $W_1(u)$ *and* $W_2(u)$ *be two independent standard Brownian motions,* $\widetilde{W}_1(u) = \int_0^u W_1(s)ds, \xi = (\int_0^1 \widetilde{W}_1(u)dW_1(u), \int_0^1 W_1(u)dW_1(u))'$, *and*

$$F = \begin{pmatrix} \int_0^1 W_1^2(u)du & \int_0^1 W_1(u)\widetilde{W}_1(u)du \\ \int_0^1 W_1(u)\widetilde{W}_1(u)du & \int_0^1 \widetilde{W}_1^2(u)du \end{pmatrix}.$$

Let Z be a standard normal random variable independent of W_1 and W_2. Consider the AR(2) *model* $\{y_t\}$ *from* (3.3) *with* $|z_2| = 1$. *Then* $Q(\beta), S(\beta)$, *and* $l(\beta)$ *all have the same limiting distribution given by*

(i) $\frac{1}{4}(W_1^2(1)-1)^2/\int_0^1 W_1^2(u)du + Z^2$, *if* z_1 *and* z_2 *are both real and* $|z_1| > 1$,

(ii) $\frac{1}{4}(W_1^2(1)-1)^2/\int_0^1 W_1^2(u)du + \frac{1}{4}(W_2^2(1)-1)^2/\int_0^1 W_2^2(u)du$, *if* $z_1 = -z_2 = 1$,

(iii) $\{(\int_0^1 W_1(u)dW_1(u) + \int_0^1 W_2(u)dW_2(u))^2 + (\int_0^1 W_1(u)dW_2(u) - \int_0^1 W_2(u)dW_1(u))^2\}$
$/\{\int_0^1 W_1^2(u) + W_2^2(u)du\}$, *if* z_1 *and* z_2 *form a pair of complex roots, and*

(iv) $\xi' F^{-1}\xi$, *if* $z_1 = z_2 = \pm 1$.

Corollary 1 can be used for testing the null hypothesis that $\beta = \beta_0$. Let $\triangle$ be the (closed) triangle in the (β_1, β_2)-plane with vertices $(-2,-1), (2,-1)$ and $(0,1)$, and let $\text{int}(\triangle)$ denote its interior, $\text{bd}(\triangle)$ its boundary. It is well known that $\{y_t\}$ is stable when $\beta' = (\beta_1, \beta_2)$ lies in $\text{int}(\triangle)$ and unstable when β' lies on $\text{bd}(\triangle)$; see, for example, Brockwell and Davis (1990). One then rejects the null that $\beta = \beta_0$ at level α if $l(\beta_0)$ (or $S(\beta_0)$) exceeds $u_{1-\alpha}(\beta_0)$, with $u_{1-\alpha}(\beta_0)$ chosen as follows. When β_0 lies in $\text{int}(\triangle)$, $u_{1-\alpha}(\beta_0) = c_{1-\alpha}$. When β_0 lies on $\text{bd}(\triangle)$, then $u_{1-\alpha}(\beta_0)$ is either $u_{1-\alpha}(\text{i}), u_{1-\alpha}(\text{ii}), u_{1-\alpha}(\text{iii})$ or $u_{1-\alpha}(\text{iv})$, the $1-\alpha$ quantiles corresponding to the limiting distributions in cases (i), (ii), (iii) or (iv) of Corollary 1, respectively. The choice $u_{1-\alpha}(\beta_0)$ when β_0 lies on $\text{bd}(\triangle)$ depends on well-known relations between β_0 and the roots z_1 and z_2. Case (i) in Corollary 1 corresponds to β' lying on the edge between $(0,1)$ and $(2,-1)$ (when $\beta_1 + \beta_2 = 1$ and $z_2 = 1$) or to β' lying on the edge between $(0,1)$ and $(-2,-1)$ (when $\beta_1 + \beta_2 = -1$ and $z_2 = -1$). Case (ii) corresponds to $\beta' = (0,1)$. Case (iii) corresponds β' lying on the edge between $(-2,-1)$ and $(2,-1)$, whereas case (iv) corresponds to $\beta' = (-2,-1)$ (when $z_1 = z_2 = -1$) or $\beta' = (2,-1)$ (when $z_1 = z_2 = 1$).

To see how we can compute quantiles, it is useful first to consider the unstable AR(1) model with $y_t = y_{t-1} + \epsilon_t$. Then $S^{1/2}(1) = \widehat{\sigma}^{-1}(\sum_{t=2}^n y_{t-1}^2)^{1/2}(\widehat{\beta} - 1)$ has limiting distribution given by $\mathcal{L}(0)$ at (3.2), and $l(1)$ has limiting distribution given by $\mathcal{L}^2(0)$ by Theorem 1. Let $u_{1-\alpha}$ be the $1-\alpha$ quantile of $\mathcal{L}^2(0)$. Let Z_{1t} be i.i.d. standard normal random variables, $x_{10} = 0$ and $x_{1m} = \sum_{t=1}^m Z_{1t}$. Let $\lfloor x \rfloor$ denote the greatest integer not exceeding x. Since $n^{-1/2}x_{1,\lfloor nu \rfloor}$ converges weakly to $W(u)$, the limiting distribution of $\mathcal{L}_{1n} = \frac{1}{2}(x_{1n}^2 - n)/(\sum_{t=1}^n x_{1,t-1}^2)^{1/2}$ is $\mathcal{L}(0)$. Thus $u_{1-\alpha}$ can be approximated by the $1-\alpha$ quantile of $\mathcal{L}_{1n}^2$, which can be obtained by simulation. The numerical results from Chan (1988) suggest that $n = 5000$ is sufficiently large to ensure that the effects of discretization are small. The second row of Table 1 reports the simulated quantiles $u_{1-\alpha}$ for $1-\alpha = 0.5, 0.8, 0.9, 0.95, 0.975$ and 0.99. Each entry was obtained by averaging the results over 25 simulations based on 99,999 replications (standard errors are given in parentheses). Abadir (1995) has computed these quantiles by numerical methods based on Laplace inversion and his numerical values are summarized in the first row of Table 1 as $u_{1-\alpha}(A)$. A comparison of the first two rows of Table 1 suggests that this simulation method gives sufficiently accurate answers.

The preceding discussion suggests the following simulation methods for obtaining the quantiles for the AR(2) model. Let Z_{2t} be another sequence of i.i.d. standard normal random variables which are independent of Z_{1t}, and let $x_{20} = 0$ and $x_{2m} = \sum_{t=1}^m Z_{2t}$. We then approximate the quantiles $u_{1-\alpha}(\text{i})$ and $u_{1-\alpha}(\text{ii})$ by the $1-\alpha$ quantiles of $\mathcal{L}_{1n}^2 + Z_{21}^2$ and $\mathcal{L}_{1n}^2 + \mathcal{L}_{2n}^2$, respectively, where $\mathcal{L}_{2n}$ has a similar expression as $\mathcal{L}_{1n}$ with x_{1t} replaced by x_{2t}. The quantile $u_{1-\alpha}(\text{iii})$ is approximated by the $1-\alpha$ quantile of $\{\frac{1}{4}(x_{1n}^2 + x_{2n}^2 - 2n)^2 + (\sum_{t=1}^n x_{1,t-1}Z_{2t} - x_{2,t-1}Z_{1t})^2\}/(\sum_{t=1}^n x_{1,t-1}^2 + x_{2,t-1}^2)$. As before the entries in Table 1 are the averages over 25 simulations based on 99,999 replications and $n = 5000$. The entries in Table 1 for $u_{1-\alpha}(\text{iv})$ were computed in the following way. Since case (iv) corresponds to $\beta' = (2, -1)$ or $(-2, -1)$, we simulated $v_{1t} = 2v_{1,t-1} - v_{1,t-2} + Z_{1t}$ and $v_{2t} = -2v_{2,t-1} - v_{2,t-2} + Z_{2t}$ and obtained the $1-\alpha$ quantiles of

$$(\sum_{t=3}^{5000} v_{i,t-1}Z_{it})'(\sum_{t=3}^{5000} v_{i,t-1}v_{i,t-1}')^{-1}\sum_{t=3}^{5000} v_{i,t-1}Z_{it}$$

for $i = 1, 2$ based on 99,999 replications. Table 1 gives averages based on 25 quantiles from v_{1t} and 25 quantiles from v_{2t}.

Table 1. Percentiles of the limiting distribution for the empirical likelihood ratio in unstable AR(1) and AR(2) models (standard errors are given in parentheses).

	Percentiles					
	0.5	0.8	0.9	0.95	0.975	0.99
$u_{1-\alpha}(A)$	0.601	1.885	2.978	4.129	5.321	6.938
$u_{1-\alpha}$	0.601	1.887	2.982	4.133	5.322	6.932
	(0.001)	(0.002)	(0.003)	(0.005)	(0.007)	(0.011)
$u_{1-\alpha}(\text{i})$	1.536	3.424	4.826	6.232	7.633	9.481
	(0.002)	(0.003)	(0.004)	(0.005)	(0.009)	(0.012)
$u_{1-\alpha}(\text{ii})$	1.687	3.623	5.051	6.461	7.875	9.723
	(0.002)	(0.002)	(0.004)	(0.005)	(0.007)	(0.010)
$u_{1-\alpha}(\text{iii})$	1.505	3.408	4.824	6.230	7.640	9.477
	(0.001)	(0.002)	(0.003)	(0.005)	(0.006)	(0.013)
$u_{1-\alpha}(\text{iv})$	1.971	4.021	5.508	6.966	8.404	10.305
	(0.001)	(0.002)	(0.003)	(0.005)	(0.007)	(0.010)

We now illustrate the construction of confidence regions through empirical likelihood.

Example 1. Consider the "deseasonalized" U.S. monthly housing starts data for the period January 1965 through December 1974, given in Example 6.3 of Reinsel (1997). The series y_t has length $n = 120$ and has been identified as AR(2). The fitted least-squares model is $\widehat{y}_t = 0.688y_{t-1} - 0.262y_{t-2}$. Also $\widehat{\sigma} = (\sum_{t=3}^{120}(y_t - \widehat{\beta}\mathbf{y}_{t-1})^2/120)^{1/2} = 5.808, \sum_{t=3}^{120} y_{t-1}^2 = 30424.75, \sum_{t=3}^{120} y_{t-2}^2 = 30283.85$, and $\sum_{t=3}^{120} y_{t-1}y_{t-2} = 28114.77$. Let $S(\beta) = \widehat{\sigma}^{-2}(\widehat{\beta} - \beta)' \sum_{t=3}^{120} \mathbf{y}_{t-1}\mathbf{y}'_{t-1}$ $(\widehat{\beta} - \beta)$. Given these numbers, an approximate 90% ellipsoidal confidence region subject to the restriction that the true model is stable is given by $\{\beta : S(\beta) \leq 4.61\} \cap \text{int}(\triangle)$, since $c_{0.9} = 4.61$ for χ_2^2. This is the region bounded between the dashed and dotted lines in Figure 1. Note that a point $\beta = (\beta_1, \beta_2)$ lies on the dotted segment in Figure 1 if it satisfies $\beta_1 + \beta_2 = 1$. Also shown in Figure 1 is the approximate 90% empirical likelihood confidence region subject to $\{\beta : l(\beta) \leq 4.61\}$, which is bounded by the solid and dotted lines. This region was computed by adapting the Splus function `elm` written by Art Owen for computing empirical likelihood ratios for the mean, available on his web page http://www-stat.stanford.edu/~art.

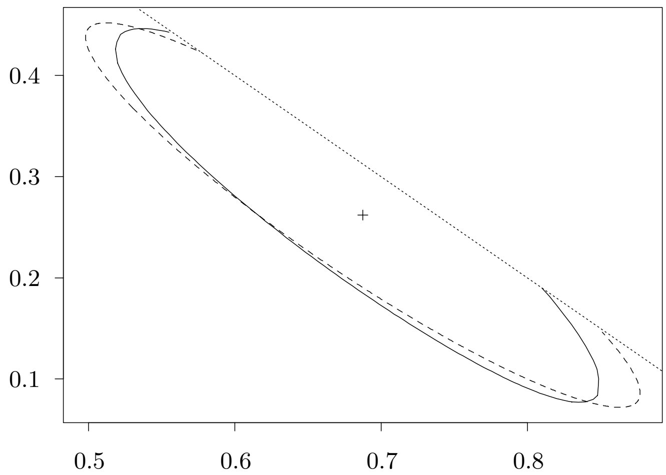

Figure 1. Approximate ellipsoidal (dashed) and empirical likelihood (solid) 90% confidence regions for β for the housing starts data fitted as an AR(2) model. Values of β for which the fitted AR(2) model is stable are shown, and the dotted segment comprises β which satisfy $\beta_1 + \beta_2 = 1$. The least squares estimate given by $(0.688, 0.262)'$ is marked with +.

Example 1 simply serves to illustrate that the shapes of empirical likelihood confidence regions are not constrained to be elliptical. For these data, a casual look at Figure 1 suggests that the data are consistent with $\beta_1 + \beta_2 = 1$ since $\{\beta : S(\beta) \leq 4.61\}$ contains values of β for which $\beta_1 + \beta_2 = 1$ and $\widehat{\beta}_1 + \widehat{\beta}_2 = 0.95$; note, however, that the chi-square approximation is not valid when $\beta_1 + \beta_2 = 1$ since we are then in case (i) of Corollary 1. A formal test of $\beta_1 + \beta_2 = 1$ is considered in the next section. Furthermore, if there is one root of (3.1) that is near 1 while the remaining roots lie outside the unit circle, the methods of Stock (1991) can be used to provide asymptotic confidence intervals for the autoregressive root that is near 1. We return to this point in the next section.

For general AR(2) (or AR(p)) models with possibly multiple roots near or on the unit circle, the chi-square approximation is poor or invalid. Here the hybrid resampling method of Chuang and Lai (2000) can be used to obtain a confidence region for β based on inverting hypothesis tests given by $\{\beta : R(\mathbf{y}, \beta) \leq \widehat{u}_{1-\alpha}(\beta)\}$ when ϵ_t are assumed to be i.i.d. To describe the resampling method, let the centered residuals be given by $\widetilde{\epsilon}_t = \widehat{\epsilon}_t - (n-p)^{-1} \sum_{t=p+1}^{n} \widehat{\epsilon}_t$ and $\widehat{\epsilon}_t = y_t - \widehat{\beta} y_{t-1}$ for $t = p+1, \ldots, n$. For a given value of β, we simulate $y_t(\beta) = \beta_1 y_{t-1}(\beta) + \cdots + \beta_p y_{t-p}(\beta) + \epsilon_t^*$ for $t = p+1, \ldots, n$, where ϵ_t^* are generated independently from the centered residuals $\widetilde{\epsilon}_t$ with the initial values $y_t(\beta)$ set to be y_t for $t = 1, \ldots, p$. Based on each series $\mathbf{y}(\beta) = \{y_t(\beta)\}_{t=1}^{n}$, we compute the root $R(\mathbf{y}(\beta), \beta)$ and then use the $1-\alpha$ quantile based on B replications as the quantile $\widehat{u}_{1-\alpha}(\beta)$. Note that the hybrid resampling method differs from the usual bootstrap method in that the simulated quantiles depend on the parameter β (so that one inverts hypothesis tests), whereas the bootstrap method uses only the least-squares estimate $\widehat{\beta}$ to obtain a single $1-\alpha$ quantile of the bootstrap distribution.

4. Empirical Likelihood Ratio Unit Root Tests and Confidence Intervals

In this section, we apply empirical likelihood methodology to the unit root autoregressive models which have received so much attention in the statistics and econometrics literature since the seminal work of Dickey and Fuller (1979). For a general AR(p) autoregressive model given by (2.6), we are concerned with testing the null hypothesis that exactly one root of $\phi(z)$, as given by (3.1), is 1 with the remaining roots outside the unit circle. Note that

$$\beta_{\boldsymbol{\cdot}} = \beta_1 + \cdots + \beta_p = 1 \tag{4.1}$$

if and only if $\phi(1) = 0$. An AR(p) model with $\beta_{\boldsymbol{\cdot}} = 1$ and all of its remaining roots outside of the unit circle could have arisen from a model whose difference is a stationary AR($p-1$) model. An AR(1) model with $\beta_1 = 1$ corresponds to the random walk model given by $y_t = y_{t-1} + \epsilon_t$.

In parametric problems, testing a constraint such as (4.1) using likelihood entails that the nuisance parameters be profiled out. The same principle applies when empirical likelihood is used. As we have pointed out in Section 2, Mykland's dual likelihood for an AR(p) model is given by

$$L_D(\beta) = \prod_{t=p+1}^{n} \frac{1}{n(1+\lambda' m_t)}, \tag{4.2}$$

where λ satisfies (2.10). The unconstrained likelihood is given by $L_D(\widehat{\beta}) = \prod_{t=p+1}^{n} n^{-1}$. The empirical likelihood ratio test rejects the null hypothesis that $\beta_{\bullet} = 1$ if

$$l_1 = 2\log L_D(\widehat{\beta}) - \max_{\beta\,:\,\beta_{\bullet}=1} 2\log L_D(\beta) = \min_{\beta\,:\,\beta_{\bullet}=1} l(\beta) \tag{4.3}$$

is too large, where $l(\beta)$ is defined in (2.9).

One way to obtain a critical value for such a test is to consider the asymptotic distribution of l_1. As we have pointed out, the empirical likelihood ratio $l(\beta)$ is approximately equal to $S(\beta)$. Let $\widehat{\beta}_{\bullet} = \widehat{\beta}_1 + \cdots + \widehat{\beta}_p$. The minimum of $S(\beta)$ subject to the constraint that $\beta_{\bullet} = 1$ is given by

$$S_1 = \widehat{\sigma}^{-2}(\widehat{\beta}_{\bullet} - 1)^2 / \mathbf{1}' \left(\sum \mathbf{y}_{t-1}\mathbf{y}_{t-1}'\right)^{-1} \mathbf{1}, \tag{4.4}$$

where $\mathbf{1} = (1,\ldots,1)'$, which is the square of the usual t-ratio for testing whether $\beta_{\bullet} = 1$. A common procedure for testing whether $\beta_{\bullet} = 1$ is to compare the coefficient of y_{t-1} with 1 through the t-ratio in a linear regression of y_t on y_{t-1}, $y_{t-1} - y_{t-2}, \ldots, y_{t-p+1} - y_{t-p}$, and the square of this t-ratio is equal to (4.4). Also, when $\beta_{\bullet} = 1$, it is well known that the limiting distribution of S_1 is given by $\mathcal{L}^2(0)$, as defined in (3.2). Taylor expansion arguments similar to those used to establish Theorem 2 in Owen (1990) for smooth functions of means show that the limiting distribution of l_1 is $\mathcal{L}^2(0)$.

To apply these ideas, consider again the housing starts data discussed in the previous section.

Example 2. The observed $S(1)$ for these data is $(-1.486)^2 = 2.208$; note that Reinsel (1997, p.189) obtains a t-ratio of -1.47 because he uses a divisor of $n-2 = 118$ rather than $n = 120$ in the estimate of σ^2. From $u_{1-\alpha}(A)$ given in Table 1, since $1.885 < 2.208 < 2.978$, the p-value for the null hypothesis that $\beta_1 + \beta_2 = 1$ is between 10% and 20%. We used `nlmin` in Splus to minimize $l(\beta)$ subject to the constraint that $\beta_1 + \beta_2 = 1$ and found a minimum of $l_1 = 1.819$ at $(0.688, 0.312)'$. Therefore the p-value for the null hypothesis that $\beta_1 + \beta_2 = 1$, using the empirical likelihood ratio unit root test, is larger than 20% since $1.819 < 1.885$. This analysis shows that the data are consistent with the hypothesis that $\beta_1 + \beta_2 = 1$.

The behavior of l_1 and S_1 under local alternatives provides information on the power of unit root tests. Consider first the AR(1) model

$$y_t = \beta y_{t-1} + \epsilon_t. \tag{4.5}$$

By Theorem 1, rejecting $\beta = 1$ when l_1 (or S_1) $\geq u_{1-\alpha}$ is an asymptotically level α test. We now show that the power of the two tests under local alternatives of the form $\beta = 1 - \gamma/n$ for some fixed γ in (4.5) are asymptotically the same. Chan and Wei (1987) study the distribution of the least squares estimate in this situation and show that $(\sum_{t=2}^{n} y_{t-1}^2)^{1/2}(\widehat{\beta} - \beta)/\sigma$ converges in distribution to $\mathcal{L}(\gamma) = \int_0^1 X(t)dW(t)/(\int_0^1 X^2(t)dt)^{1/2}$, where $X(t)$ is the Ornstein-Uhlenbeck process

$$dX(t) = -\gamma X(t)dt + dW(t), \quad X(0) = 0, \tag{4.6}$$

and $W(t)$ is standard Brownian motion. A direct argument shows that $\widehat{\sigma}$ converges to σ in probability so that $S(\beta)$ is asymptotically $\mathcal{L}^2(\gamma)$. Theorem 2 below states that the limiting distribution of $l(\beta)$ is also $\mathcal{L}^2(\gamma)$. Furthermore, under the same sequence of local alternatives, l_1 and S_1 have the same limiting distribution so that the two unit root tests have the same power against local alternatives, to first order.

Theorem 2. *Let $y_t = \beta y_{t-1} + \epsilon_t$, where $\beta = 1 - \gamma/n$ and γ is a fixed constant. Let $X(t)$ be the Ornstein-Uhlenbeck process given by* (4.6). *Then the limiting distribution of $l(\beta)$ is $\mathcal{L}^2(\gamma)$, and l_1 and S_1 have the same limiting distribution $\{\mathcal{L}(\gamma) - \gamma(\int_0^1 X^2(t)dt)^{1/2}\}^2$.*

Table 2 summarizes the results of a simulation study which assesses the small-sample properties of the two unit root tests based on l_1 and S_1. When n is large, a test with approximate level 5% rejects the null that $\beta = 1$ when l_1 (or S_1) ≥ 4.129. To assess whether this approximation is adequate in small samples, we generated time series according to the model (4.5), starting from $y_0 = 0$ with independent ϵ_t. We used three different distributions for ϵ_t: the standard normal; t with 8 degrees of freedom, denoted t_8 in Table 2; $\chi_4^2 - 4$ (so that $E(\epsilon_t) = 0$), and $n = 50$. When the ϵ_t were chosen to be standard normal, we also considered $n = 25$. Table 2 reports the observed proportions of rejections based on 10,000 replications. We computed empirical likelihood ratios for l_1 by using Brent's method (Press, Teukolsky, Vetterling and Flannery (1992)) to solve for λ, as explained in Owen (1988).

As is evident from Table 2, the actual levels of the two unit root tests can be rather different from the nominal level of 5% when n is small and the distributions of S_1 and l_1 are not well approximated by $\mathcal{L}^2(0)$. DiCiccio, Hall and Romano (1991) show that empirical likelihood for the mean is Bartlett-correctable (as are parametric likelihoods) so that an analytic correction renders empirical likelihood confidence regions that have coverage probabilities that differ from their nominal ones by a term of order $O(n^{-2})$. Mykland (1994, 1995) and Kitamura (1997) study Bartlett corrections for dependent data. However, in most of these situations, the Bartlett factor is already complicated. For unit root tests, Bartlett correction appears to be impossible, and Nielsen (1997) has in fact shown that the usual conditions for Bartlett correction fail even in a simple parametric Gaussian unit root autoregressive model.

Table 2. Power comparisons for l_1 and S_1 of nominal level 0.05 and their modified versions l_{1r} and S_{1r}. The observed levels of the tests are given under the column $\beta = 1$.

			β								
n	ϵ_t	Test	0.70	0.80	0.85	0.90	0.95	0.99	1.00	1.02	1.10
25	$N(0,1)$	S_1	0.59	0.34	0.24	0.15	0.09	0.06	0.06	0.09	0.66
		l_1	0.59	0.36	0.26	0.18	0.11	0.08	0.08	0.11	0.67
		S_{1r}	0.49	0.27	0.18	0.11	0.06	0.05	0.05	0.07	0.64
		l_{1r}	0.44	0.25	0.17	0.11	0.07	0.05	0.05	0.08	0.63
50	$N(0,1)$	S_1	0.97	0.76	0.54	0.31	0.13	0.06	0.05	0.18	0.96
		l_1	0.95	0.73	0.53	0.31	0.14	0.06	0.06	0.19	0.96
		S_{1r}	0.96	0.71	0.48	0.27	0.11	0.05	0.05	0.17	0.96
		l_{1r}	0.92	0.67	0.47	0.26	0.11	0.05	0.05	0.16	0.96
50	t_8	S_1	0.97	0.76	0.54	0.30	0.13	0.06	0.06	0.18	0.96
		l_1	0.93	0.73	0.55	0.33	0.15	0.07	0.07	0.19	0.96
		S_{1r}	0.96	0.71	0.49	0.27	0.11	0.05	0.05	0.17	0.96
		l_{1r}	0.88	0.65	0.46	0.27	0.11	0.05	0.05	0.17	0.95
50	$\chi^2_4 - 4$	S_1	0.97	0.75	0.53	0.30	0.13	0.06	0.06	0.18	0.96
		l_1	0.91	0.73	0.55	0.33	0.16	0.07	0.07	0.20	0.96
		S_{1r}	0.94	0.68	0.46	0.25	0.11	0.05	0.05	0.17	0.96
		l_{1r}	0.87	0.64	0.45	0.25	0.12	0.05	0.06	0.17	0.96

Nevertheless, one can use resampling to modify the two tests so that the levels are closer to the nominal ones. Consider the resampled time series given by $y_t^* = y_{t-1}^* + \epsilon_t^*, t = 2, \ldots, n$ and $y_1^* = y_1$, where ϵ_t^* are independent from the centered residuals $\widetilde{\epsilon}_t = \widehat{\epsilon}_t - (n-1)^{-1} \sum_{t=2}^{n} \widehat{\epsilon}_t$ and $\widehat{\epsilon}_t = y_t - \widehat{\beta} y_{t-1}$. To obtain an alternative critical value for S_1, compute $\widehat{\sigma}^{*-2} \sum_{t=2}^{n} y_{t-1}^{*2} (\widehat{\beta}^* - 1)^2$ for each resampled series, where $\widehat{\sigma}^*$ and $\widehat{\beta}^*$ are the quantities $\widehat{\sigma}$ and $\widehat{\beta}$ computed based on y_t^* instead of y_t. The upper 5th percentile after repeating this procedure B times is then used as the critical value. A similar procedure can be used to obtain an alternative critical value for l_1. Note that this resampling method uses the value of $\beta = 1$ as prescribed by the null hypothesis rather then $\widehat{\beta}$ used in the usual bootstrap. Basawa *et al.* (1991) show that the bootstrap method based on $\widehat{\beta}$ gives inconsistent critical values, and the method used here is based on that of Ferreti and Romo (1996) and Chuang and Lai (2000). In principle the m out of n bootstrap method can be used, as pointed out by Bickel in his discussion to Chuang and Lai (2000). However the above resampling method can be easier to implement since it does not require the user to choose m. The modifications of l_1 and S_1 thus obtained based on $B = 1999$ are denoted by l_{1r} and S_{1r} in Table 2, and the levels are closer to the nominal level of 5%. Again, the observed levels were obtained based on 10,000 replications.

Table 2 also presents the results of a simulation of the power of the two tests based on S_1 and l_1 and their modified versions S_{1r} and l_{1r}, by reporting the observed proportions of rejections based on 10,000 replications when the true value of β ranges from 0.7 to 1.1. Similar power studies are reported by Dickey and Fuller (1979) and Ferreti and Romo (1996). The test based on l_1 has slightly higher power than that based on S_1 for values of β ranging from 0.85 to 1.1 for the cases considered. However, note that the test based on l_1 also has slightly higher levels than that based on S_1. For $\beta = 0.7$ and 0.8, the test based on S_1 has slightly higher power. Table 2 also indicates that the two unit root tests based on S_1 and l_1 have uniformly higher power functions than their modified versions S_{1r} and l_{1r}, but this appears to be due to the slightly inflated levels of the tests based on S_1 and l_1. The power functions of the modified versions S_{1r} and l_{1r} are quite similar, but S_{1r} has slightly higher power than l_{1r} for $\beta = 0.7$ and 0.8.

Finally we generalize the nearly nonstationary AR(1) model (4.5) with $\beta = 1-\gamma/n$ to an AR(p) situation. Consider $y_t = \theta_1 y_{t-1} + x_t$, where $\theta_1 = 1-\gamma/n$ and x_t satisfies an AR($p-1$) model given by $x_t = \theta_2 x_{t-1} + \cdots + \theta_p x_{t-p+1} + \epsilon_t$, with the roots of $1-\theta_2 z - \cdots - \theta_p z^{p-1}$ lying outside of the unit circle. Stock (1991) shows that the process $\{y_t\}$ can be rewritten as $y_t = \beta_1 y_{t-1} + \cdots + \beta_p y_{t-p} + \epsilon_t$, where $\beta_1 = \theta_1 + \theta_2$, $\beta_i = \theta_{i+1} - \theta_1\theta_i$ for $i = 2, \ldots, p-1$, and $\beta_p = -\theta_1\theta_p$. The characteristic polynomial (3.1) factors as $(1-\theta_1 z)(1-\theta_2 z - \cdots - \theta_p z^{p-1})$ so that there is a single root near the unit circle, given by $z = 1/\theta_1 = (1-\gamma/n)^{-1}$. The arguments in Stock (1991) can be used to establish that the limiting distribution of S_1 as defined in (4.4) is $\{\mathcal{L}(\gamma) - \gamma(\int_0^1 X^2(t)dt)\}^2$, which is the same as that given in Theorem 2 for a nearly nonstationary AR(1) model. Taylor expansion arguments then show that l_1 as defined in (4.3) has the same limiting distribution. These results can be used to construct confidence intervals for $\theta_1 = 1-\gamma/n$ and $1/\theta_1$ by using signed square roots of either l_1 or S_1, $\text{sgn}(\widehat{\beta}. - 1)\sqrt{l_1}$ or $\text{sgn}(\widehat{\beta}. - 1)\sqrt{S_1}$. The limiting distribution of the signed roots is $\mathcal{L}(\gamma) - \gamma(\int_0^1 X^2(t)dt)$, or equivalently by

$$\frac{1}{2}(X^2(1)-1)/(\int_0^1 X^2(t)dt)^{1/2}, \tag{4.7}$$

where $X(t)$ is the Ornstein-Uhlenbeck process given in (4.6); see Chan (1988, p.859). To find confidence intervals, we need the quantiles of (4.7), which can be tabulated as follows. We approximate the integral in (4.7) by $\frac{1}{n}\sum_{i=0}^{n-1} X^2(i/n)$, where $X_0 = 0$. To simulate $X(i/n)$, note that its conditional distribution given $X((i-1)/n)$ is normal with mean $\exp(-\gamma/n)X((i-1)/n)$ and variance $\{1 - \exp(-2\gamma/n)\}/(2\gamma)$; see Karatzas and Shreve (1991, p.358). The 0.01, 0.025, 0.05, 0.1, 0.8, 0.9, 0.95 and 0.99 percentiles are provided in Table 3 for selected values of γ. We used $n = 500$ and 99,999 replications to compute the quantiles, each averaged over 25 such simulations.

Table 3. Percentiles of the limiting distribution of $(X^2(1)-1)/\{2(\int_0^1 X^2(t)\,dt)^{1/2}\}$, where $X(t)$ is the Ornstein-Uhlenbeck process $dX(t) = -\gamma X(t) + dW(t)$ with $X(0) = 0$ (standard errors are given in parentheses for $\gamma < 0$, the standard errors for the entries with asterisks are at most 0.002 and the standard errors for the rest of the entries are at most 0.001).

	Percentiles							
	0.01	0.025	0.05	0.1	0.9	0.95	0.975	0.99
−5	−0.975	1.08	4.02	9.06	122.7	146.2	167.3	192.5
	(0.007)	(0.012)	(0.014)	(0.021)	(0.07)	(0.083)	(0.119)	(0.149)
−4	−1.599	−0.86	0.46	2.75	45.14	53.82	61.58	70.82
	(0.004)	(0.006)	(0.007)	(0.008)	(0.028)	(0.031)	(0.04)	(0.053)
−3	−1.929	−1.452	−0.948	0.059	16.67	19.90	22.81	26.23
	(0.003)	(0.003)	(0.002)	(0.004)	(0.010)	(0.012)	(0.017)	(0.023)
−2	−2.179	−1.779	−1.416	−0.951	6.262	7.532	8.66	10.00
	(0.002)	(0.002)	(0.002)	(0.002)	(0.004)	(0.005)	(0.007)	(0.009)
−1	−2.384	−2.026	−1.713	−1.347	2.437	3.037	3.559	4.166
	(0.002)	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)	(0.003)
1	−2.730*	−2.407*	−2.137	−1.835	0.126	0.444	0.721	1.051*
2	−2.878*	−2.566*	−2.309	−2.022	−0.322	−0.051	0.192	0.486*
3	−3.014*	−2.712	−2.463	−2.188	−0.618	−0.388	−0.171	0.098*
4	−3.141*	−2.849	−2.606	−2.338	−0.835	−0.637	−0.443	−0.196*
5	−3.262*	−2.975	−2.739	−2.477	−1.014	−0.834	−0.660	−0.433
6	−3.377*	−3.095	−2.863	−2.606	−1.170	−1.000	−0.840	−0.633
7	−3.486*	−3.209	−2.980	−2.728	−1.311	−1.147	−0.997	−0.805
8	−3.59*	−3.317	−3.092	−2.843	−1.441	−1.280	−1.138	−0.958
9	−3.689	−3.421	−3.199	−2.953	−1.563	−1.404	−1.266	−1.095
10	−3.786*	−3.520	−3.301	−3.058	−1.677	−1.520	−1.384	−1.221
15	−4.222	−3.970	−3.759	−3.526	−2.175	−2.021	−1.891	−1.741
20	−4.604	−4.361	−4.156	−3.929	−2.595	−2.440	−2.311	−2.164
25	−4.950	−4.711	−4.512	−4.289	−2.964	−2.808	−2.679	−2.533
50	−6.344	−6.122	−5.933	−5.721	−4.414	−4.254	−4.120	−3.971

Example 3. We use the housing starts data considered in Examples 1 and 2 to illustrate how approximate 90% confidence intervals can be constructed. The t-statistic $\text{sgn}(\widehat{\beta}_{\cdot} - 1)\sqrt{S_1}$ equals -1.486. Values of γ consistent with the observed data are those for which -1.486 lies between the 5th and 95th percentiles of (4.7). Table 3 indicates that a rough range of γ values consistent with the data is $-1 \leq \gamma \leq 9$. Since $n = 120$, an interval for the autoregressive root near 1 is given by $(1-\gamma/n)^{-1}$ or $[0.99, 1.08]$. Similarly, the signed root of l_1 is $-\sqrt{1.819} = -1.349$ and this leads to the interval $-2 \leq \gamma \leq 8$, which corresponds to $[0.98, 1.07]$ for the largest autoregressive root.

5. Conclusion

This paper shows that empirical likelihood is an effective method for analyzing unstable autoregressive time series. It offers a promising direction to study general unstable autoregressive models such as those generated by the long-memory processs studied in Chan and Terrin (1995), or by the heteroskedastic processes studied in Ling and Li (1998).

Acknowledgement

Research was supported in part by the Hong Kong Research Grant Council Grants CUHK 6082/98P and 2060205. We also thank the Editor and two referees for their helpful comments.

Appendix (Proofs)

Proof of Lemma 1. Part (i) is a special case of Theorem 3 of Lai and Wei (1982) since $n^{-1}\sum \mathbf{y}_{t-1}\mathbf{y}'_{t-1}$ converges a.s. to a nonsingular limit. This can be shown by applying a strong law for martingales due to Chow (1965, Theorem 2) and modifying arguments for Theorem 6.6 in Hall and Heyde (1980).

Part (ii) follows by first showing that $l(\beta)$ is asymptotically equivalent to $\sum m'_t(\sum m_t m'_t)^{-1} m_t$, where $m_t = \mathbf{y}_{t-1}\epsilon_t$. This asymptotic equivalence uses familiar arguments in the literature; see for example, the proof of Theorem 1 below for the unstable case. The assertion then follows from part (i) and $n^{-1}(\sum m_t m'_t - \sigma^2 \sum \mathbf{y}_{t-1}\mathbf{y}'_{t-1}) \to 0$ a.s., which can be established using arguments similar to those used to prove part (iii) of Lemma 5 below.

Proof of Lemma 2. The assertion follows from Lemma 3 of Lai and Wei (1982) and Corollary 1 of Lai and Wei (1985).

We now turn our attention to establishing the limiting distribution of $l(\beta)$ in the unstable case. First, follow Chan and Wei (1988) and factorize (3.1) using its roots, so that $\phi(z) = (1-z)^a(1+z)^b\prod_{k=1}^{l}(1-2\cos\theta_k z+z^2)^{d_k}\psi(z)$, where $a+b+2d_1+\cdots+2d_l = p$. Let $(1-B)^j u_t(j) = \epsilon_t$, $j = 1,\ldots,a$, $(1+B)^j v_t(j) = \epsilon_t$, $j = 1,\ldots,b$, $(1-2\cos\theta_k B+B^2)^{d_k} y(d_k)_t(j) = \epsilon_t$, $j = 1,\ldots,d_k$, and $\psi(B)z_t = \epsilon_t$. Let

$$\mathbf{y}(d_k)_t = (n^{-1}y(d_k)_t(1), n^{-1}y(d_k)_{t-1}(1),\ldots,n^{-d_k}y(d_k)_t(d_k), n^{-d_k}y(d_k)_{t-1}(d_k))'$$

for $k = 1,\ldots,l$. Define the matrices G_n and Q, as in Chan and Wei (1988), so that $(G_nQ)\mathbf{y}_t = Y_t$, where $Y_t = (n^{-a}u_t(a),\ldots,n^{-1}u_t(1), n^{-b}v_t(b),\ldots,n^{-1}v_t(1), \mathbf{y}(d_1)'_t, \ldots, \mathbf{y}(d_l)'_t, n^{-1/2}z_t,\ldots,n^{-1/2}z_{t-q+1})'$. Multiplying $\mathbf{y}_t$ by the matrix G_nQ transforms the AR(p) model into its individual components and simplifies the analysis.

Now, let $\widetilde{\lambda} = (Q'G'_n)^{-1}\lambda$, and rewrite (2.10) as

$$\sum_{t=p+1}^{n} n_t/(1+\widetilde{\lambda}'n_t) = 0, \tag{A.1}$$

where $n_t = (G_nQ)m_t = Y_{t-1}\epsilon_t$. In Theorem 1 below, we modify Owen's (1990) argument to derive the distribution of $l(\beta)$; see also Mykland (1995). This analysis is preceded by three lemmas.

Lemma 3. *Let* $0 < \alpha' < \alpha$.

(i) $\max_{p+1\le t\le n}|\epsilon_t| = o(n^{1/(2+\alpha')})$ *a.s.*

(ii) $n^{-1}\sum\epsilon_t^2 \to \sigma^2$ *a.s.*

(iii) $\sum|\epsilon_t|^3 = o(n^{1+1/(2+\alpha')})$ *a.s.*

Proof. Part (i) is proved in Theorem 1 of Lai and Wei (1982). Part (ii) follows from Chow's strong law for martingales. Part (iii) then follows from parts (i) and (ii), and the fact that $\sum|\epsilon_t|^3 \le \max_{p+1\le t\le n}|\epsilon_t|\sum|\epsilon_t|^2$.

Lemma 4.

(i) *Let* $M_{tn}, t = 1,\ldots,k_n$, *be a* $p\times 1$ *martingale difference array adapted to a sequence of filtrations* $\mathcal{G}_{tn}$ *for each* n. *Let* $U_{nn} = \sum_{t=1}^{k_n} M_{tn}M'_{tn}$ *and* $V_{nn} = \sum_{t=1}^{k_n} E(M_{tn}M'_{tn}|\mathcal{G}_{t-1,n})$. *Suppose that* $\sup_n P(||V_{nn}|| > a) \to 0$ *when* $a\to\infty$ *and for all* $\delta > 0$, $\sum_{t=1}^{k_n} E(||M_{tn}||^2 1(||M_{tn}|| > \delta)|\mathcal{G}_{t-1,n}) \to 0$ *in probability. Then* $V_{nn} - U_{nn} \to 0$ *in probability.*

(ii) *Let* X_t *be random variables and* $\mathcal{F}_t$ *be a filtration. Suppose* $\sup_t E(|X_t|^p|\mathcal{F}_{t-1}) < \infty$ *a.s. for* $p > 1$. *If* $\max_{1\le t\le n} P(A_t|\mathcal{F}_{t-1}) \to 0$ *a.s., then* $\max_{1\le t\le n} E(|X_t|1\{A_t\}|\mathcal{F}_{t-1}) \to 0$ *a.s.*

Proof. Part (i) can be proved by adapting the proof of Theorem 2.23 of Hall and Heyde (1980) for scalar M_{tn} to vector-valued M_{tn}. The proof of (ii) is similar to proofs for uniform integrability of random variables; see Billingsley (1995, pp.217-218).

Lemma 5.

(i) $\max_{p\le t\le n}||Y_t|| = O(n^{-1/2}(\log\log n)^{1/2})$ *a.s.*

(ii) $\max_{p+1\le t\le n}||n_t|| = o(1)$ *a.s.*

(iii) $\sum Y_{t-1}Y'_{t-1}$ *converges in distribution to a nonsingular limit and* $\sum ||Y_{t-1}||^2 = O_p(1)$.

(iv) $\sum n_t n'_t - \sum Y_{t-1}Y'_{t-1}\sigma^2$ *converges in probability to* 0.

(v) *Let* σ_n *be the smallest eigenvalue of* $\sum n_t n'_t$. *Then* σ_n *converges in distribution to a strictly positive random variable.*

Proof. Consider the case when $\psi(z) \equiv 1$ first, that is, when all the roots of (3.1) lie on the unit circle. The first component of Y_t involves $n^{-a}u_t(a)$, where $u_t(a)$ satisfies $(1-B)^a u_t(a) = \epsilon_t$, so that $\max_{p\le t\le n}||u_t(a)|| = O(n^{a-1/2}(\log\log n)^{1/2})$ a.s. by Theorem 1 of Lai and Wei (1982). A similar analysis can be carried out for the other components. Part (ii) holds since $\max_{p+1\le t\le n}||n_t|| \le \max_{p\le t\le n-1}||Y_t|| \max_{p+1\le t\le n}|\epsilon_t|$.

Part (iii) follows from Chan and Wei (1988), who show that $n^{1/2}Y_t$ converges in distribution to a random process and that $\sum Y_{t-1}Y'_{t-1}$ converges in distribution to a nonsingular limit. Thus $n^{-1}\sum ||n^{1/2}Y_{t-1}||^2$ converges in distribution by the Continuous Mapping Theorem.

To prove part (iv), we apply part (i) of Lemma 4, with $M_{tn} = n_t = Y_{t-1}\epsilon_t$. Thus $U_{nn} = \sum M_{tn}M'_{tn} = \sum Y_{t-1}Y'_{t-1}\epsilon_t^2$ and $V_{nn} = \sum E(Y_{t-1}Y'_{t-1}\epsilon_t^2|\mathcal{F}_{t-1}) = \sigma^2\sum Y_{t-1}Y'_{t-1}$. Since V_{nn} converges in distribution by part (iii), $\sup_n P(||V_{nn}|| > a) \to 0$ when $a\to\infty$. Then $\sum E(||M_{tn}||^2 1(||M_{tn}|| > \delta)|\mathcal{F}_{t-1}) \le \sum ||Y_{t-1}||^2 \max_{p+1\le t\le n} E(\epsilon_t^2 1(||M_{tn}|| > \delta))$. Since $\sum ||Y_{t-1}||^2 = O_p(1)$, it suffices to show that $\max_{p+1\le t\le n} E(\epsilon_t^2 1(||M_{tn}|| > \delta)) \to 0$ a.s. This follows from part (ii) of Lemma 4 and $P(\epsilon_t^2 > \delta^2/||Y_{t-1}||^2|\mathcal{F}_{t-1}) \le E(\epsilon_t^2|\mathcal{F}_{t-1})||Y_{t-1}||^2/\delta^2 = \sigma^2||Y_{t-1}||^2/\delta^2$, and then $\max_{p\le t\le n}||Y_t||^2 \to 0$ a.s. by part (i).

Part (iv) follows from parts (iii) and (iv) and the Continuous Mapping Theorem.

If (3.1) has roots outside of the unit circle as well, the results still hold by noting that for $\mathbf{z}_t = (z_t,\ldots,z_{t-q+1})'$, $||\mathbf{z}_t|| = o(n^{1/(2+\alpha')})$ a.s. for $0<\alpha'<\alpha$, as proved in Lai and Wei (1985, Theorem 1), and that $n^{-1}\sum \mathbf{z}_t\mathbf{z}'_t$ converges a.s. to a nonsingular limit, as indicated in the proof of Lemma 1.

Proof of Theorem 1. First we establish that $\widetilde{\lambda} = O_p(1)$. Let $\widetilde{\lambda} = \rho\psi$, where $\rho = ||\widetilde{\lambda}||$ and $\psi = \lambda/||\widetilde{\lambda}||$, so that $||\psi|| = 1$. From (A.1), we obtain that $\psi'(\sum n_t - \rho\sum \frac{n_t\psi' n_t}{1+\rho\psi' n_t} = 0$. Solving for ρ yields $\rho = (\psi'\sum\frac{n_t n'_t}{1+\rho\psi' n_t}\psi)^{-1}\psi'\sum n_t$. Since $||\psi|| = 1$, $\psi'\sum n_t n'_t\psi \ge \sigma_n$, where σ_n is the smallest eigenvalue of $\sum n_t n'_t$. The argument in Owen (1990, p.101) shows that $1+\rho\psi' n_t > 0$, so that $\rho \le (\psi'\sum n_t n'_t\psi)^{-1}(\psi'\sum n_t)(1+\rho\max_{p+1\le t\le n}||n_t||) \le \sigma_n^{-1}(\psi'\sum n_t)(1+\rho\max_{p+1\le t\le n} ||n_t||)$. By part (v) of Lemma 5, σ_n^{-1} converges in distribution to an a.s. finite random variable. Also, $\sum n_t = (\sum Y_{t-1}Y'_{t-1})(Q'G'_n)^{-1}(\widehat{\beta}-\beta) = O_p(1)$; see Chan and Wei (1988, p.379), and $\max_{p+1\le t\le n}||n_t|| = o(1)$ a.s. Thus $\widetilde{\lambda} = O_p(1)$.

By using the identity $(1+x)^{-1} = 1 - x + x^2(1+x)^{-1}$ to expand (A.1), we obtain an expression involving $\widetilde{\lambda}$: $\sum n_t(1-\widetilde{\lambda}'n_t + \frac{1}{2}(\widetilde{\lambda}'n_t)^2(1+\widetilde{\lambda}'n_t)^{-1}) = 0$. This implies that $\widetilde{\lambda} = (\sum n_t n_t')^{-1}\sum n_t + \delta$, where $\delta = (\sum n_t n_t')^{-1}\sum \frac{1}{2}n_t(\lambda' n_t)^2(1+\widetilde{\lambda}n_t)^{-1}$. Now we show that $\delta = o_p(1)$. Note that $\sum ||n_t(\widetilde{\lambda}'n_t)^2(1+\widetilde{\lambda}'n_t)^{-1}|| \le ||\widetilde{\lambda}||^2\sum||n_t||^3\max_{p+1\le t\le n}(1+\widetilde{\lambda}'n_t)^{-1}$. Since $\widetilde{\lambda} = O_p(1), n_t = o(1)$ a.s., $(\sum n_t n_t')^{-1} = O_p(1)$, and $\sum ||n_t||^3 \le \max_{p\le t\le n}||Y_t||^3\sum|\epsilon_t|^3 = o(1)$ a.s. by part (iii) of Lemma 3 and part (i) of Lemma 5, $\delta = o_p(1)$.

Next, noting that $\widetilde{\lambda}'n_t = o_p(1)$, we expand (2.9) as

$$\begin{aligned}2\sum\log(1+\widetilde{\lambda}'n_t) &= 2\sum(\widetilde{\lambda}'n_t - (\widetilde{\lambda}'n_t)^2/2 + r_t)\\ &= \sum n_t'(\sum n_t n_t')^{-1}\sum n_t - \delta'(\sum n_t n_t')^{-1}\delta + 2\sum r_t,\end{aligned}$$

using the relation $\widetilde{\lambda} = (\sum n_t n_t')^{-1}\sum n_t + \delta$. Taylor's Theorem with remainder implies that $6\sum||r_t|| \le ||\widetilde{\lambda}||^3\sum||n_t||^3 = o_p(1)$, as we have shown before. The term $\delta'(\sum n_t n_t')^{-1}\delta$ is of smaller order. Thus, $l(\beta)$ and $\sum n_t'(\sum n_t n_t')^{-1}\sum n_t = \sum m_t'(\sum m_t m_t')^{-1}\sum m_t = Q(\beta)$ have the same limiting distribution. By part (iv) of Lemma 5, $Q(\beta)$ and $S(\beta)$ are asymptotically equivalent.

The proof of Theorem 2 relies on the following Lemma.

Lemma 6.

(i) *Let $a_{t,n}, t = 1,\ldots,n$ and $n = 1,2,\ldots$ be a double array of constants satisfying $a_{n,n}\to a$, finite, and $\sup_n\sum_{t=2}^n|a_{t,n}-a_{t-1,n}| < \infty$. Then for any $\delta > 1/2, n^{-\delta}\sum_{t=1}^n a_{t,n}\epsilon_t \to 0$ a.s.*

(ii) *Let $y_t = \beta y_{t-1}+\epsilon_t$, where $\beta = 1-\gamma/n$ and γ is a fixed constant. Then $\max_{2\le t\le n}|y_t|/n^\delta = o(1)$ a.s. for any $\delta > 1/2$.*

Proof. Let $S_0 = 0$ and $S_k = \sum_{t=1}^k\epsilon_t$. By Chow's strong law for martingales, $S_n/n^\delta \to 0$ a.s. and $\max_{1\le k\le n}|S_k|/n^\delta \to 0$ a.s. Summing by parts yields $\sum_{t=1}^n a_{t,n}\epsilon_t = a_{n,n}S_n - \sum_{t=1}^n S_{t-1}(a_{t,n}-a_{t-1,n})$. Part (i) follows from $\sum_{t=1}^n|S_{t-1}||a_{t,n}-a_{t-1,n}| \le \max_{1\le t\le n}|S_t|\sum_{t=1}^n|a_{t,n}-a_{t-1,n}|$.

To prove part (ii), note that $y_t = \sum_{t=1}^n(1-\gamma/n)^{n-t}\epsilon_t$. Apply part (i) to the double array of constants $a_{t,n} = (1-\gamma/n)^{n-t}$.

Proof of Theorem 2. Let $y_t = \beta y_{t-1}+\epsilon_t$, where $\beta = 1-\gamma/n$. The proof of Theorem 1 shows that as long as $\sum_{t=2}^n n_t^2$ converges in distribution to an a.s. positive random variable and $\max_{2\le t\le n}|n_t| = o(1)$ a.s., $l(\beta) = Q(\beta)+o_p(1)$, where $Q(\beta) = (\sum_{t=2}^n n_t)^2/\sum_{t=2}^n n_t^2$ and $n_t = n^{-1}y_{t-1}\epsilon_t$. Note that $\sum_{t=2}^n n_t^2$ converges to $\sigma^2\int_0^1 X(t)^2$ in distribution by applying part (i) of Lemma 4 and results in Chan and Wei (1987). Also, by part (i) of Lemma 3 and part (ii) of Lemma

6, $\max_{2\le t\le n}|n_t| = o(1)$ a.s., since $\max_{2\le t\le n}|n_t| \le \max_{1\le t\le n}|y_t|\max_{1\le t\le n}|\epsilon_t|$. Thus, $l(\beta)$ and $Q(\beta)$ have limiting distribution given by $\mathcal{L}^2(\gamma)$.

Under the sequence of local alternatives $\beta = 1-\gamma/n$, $S_1^{1/2} = (\sum_{t=2}^n y_{t-1}^2)^{1/2}(\widehat{\beta}-1)/\widehat{\sigma} = (\sum_{t=2}^n y_{t-1}^2)^{1/2}(\widehat{\beta}-\beta)/\widehat{\sigma} - \gamma(\widehat{\sigma}^{-2}n^{-2}\sum_{t=2}^n y_{t-1}^2)^{1/2}$, which converges in distribution to $\mathcal{L}(\gamma) - \gamma\int_0^1 X^2(t)dt$, since $\widehat{\sigma}$ is still consistent for σ. Thus S_1 converges in distribution to $\{\mathcal{L}(\gamma) - \gamma\int_0^1 X^2(t)dt\}^2$.

To derive the distribution of l_1 under the same sequence of local alternatives, note that the argument in Theorem 1 shows that $l_1 = S_1 + o_p(1)$ provided $\sum_{t=2}^n n_t^2$ converges in distribution to a positive random variable and $\max_{2\le t\le n}|n_t| = o(1)$ a.s., where $n_t = n^{-1}y_{t-1}(y_t - y_{t-1}) = n^{-1}y_{t-1}\epsilon_t - n^{-2}\gamma y_{t-1}^2$. By part (ii) of Lemma 6, $\max_{2\le t\le n}|n_t| = o(1)$ a.s. One can also show that $\sum_{t=2}^n n_t^2$ still converges in distribution to $\int_0^1 X^2(t)dt$.

References

Abadir, K. M. (1995). The limiting distribution of the t ratio under a unit root. *Econom. Theory* **11**, 775-793.

Basawa, I. V., Mallik, A. K., McCormick, W. P., Reeves, J. H. and Taylor, R. L. (1991). Bootstrapping unstable first-order autoregressive processes. *Ann. Statist.* **19**, 1098-1101.

Billingsley, P. (1995). *Probability and Measure.* 3rd edition. Wiley, New York.

Brockwell, P. J. and Davis, R. A. (1990). *Time Series: Theory and Methods.* 2nd edition. Springer-Verlag, New York.

Chan, N. H. (1988). Parameter inference for nearly nonstationary time series. *J. Amer. Statist. Assoc.* **83**, 857-862.

Chan, N. H. and Terrin, N. (1995). Inference for unstable long-memory processes with applications to fractional unit root autoregressions. *Ann. Statist.* **23**, 1662-1683.

Chan, N. H. and Wei, C. Z. (1987). Asymptotic inference for nearly nonstationary AR(1) processes. *Ann. Statist.* **15**, 1050-1063.

Chan, N. H. and Wei, C. Z. (1988). Limiting distributions of least squares estimates of unstable autoregressive processes. *Ann. Statist.* **16**, 367-401.

Chow, Y. (1965). Local convergence of martingales and the law of large numbers. *Ann. Math. Statist.* **36**, 552-558.

Chuang, C. and Lai, T. L. (2000). Hybrid resampling methods for confidence intervals with discussion. *Statist. Sinica* **10**, 1-50.

DiCiccio, T. J., Hall, P. and Romano, J. P. (1991). Empirical likelihood is Bartlett-correctable. *Ann. Statist.* **19**, 1053-1061.

Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *J. Amer. Statist. Assoc.* **74**, 427-431.

Ferretti, N. and Romo, R. (1996). Unit root bootstrap tests for AR(1) models. *Biometrika* **83**, 849-860.

Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and its Applications.* Academic Press, New York.

Karatzas, I. and Shreve, S. (1991). *Brownian Motion and Stochastic Calculus.* 2nd edition. Springer, New York.

Kitamura, Y. (1997). Empirical likelihood methods for weakly dependent processes. *Ann. Statist.* **25**, 2084-2102.

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

Lai, T. L. and Wei, C. Z. (1982). Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. *Ann. Statist.* **10**, 154-166.

Lai, T. L. and Wei, C. Z. (1985). Asymptotic properties of multivariate weighted sums with applications to stochastic regression in linear dynamic systems. In *Multivariate Analysis—VI: Proceedings of the Sixth International Symposium on Multivariate Analysis* (Edited by P. R. Krishnaiah), 375-394. North-Holland, Amsterdam.

Ling, S. and Li, W. K. (1998). Limiting distributions of maximum likelihood estimators for unstable autoregressive moving-average time series with general autoregressive heteroskedastic errors. *Ann. Statist.* **26**, 84-125.

Monti, A. C. (1997). Empirical likelihood confidence regions in time series models. *Biometrika* **84**, 395-405.

Mykland, P. A. (1994). Bartlett type identities for martingales. *Ann. Statist.* **22**, 21-38.

Mykland, P. A. (1995). Dual likelihood. *Ann. Statist.* **23**, 396-421.

Nielsen, B. (1997). Bartlett correction of the unit root test in autoregressive models. *Biometrika* **84**, 500-504.

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.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P. (1992). *Numerical Recipes in C: The Art of Scientific Computing.* 2nd edition. Cambridge University Press.

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

Reinsel, G. (1997). *Elements of Multivariate Time Series Analysis.* 2nd edition. Springer-Verlag, New York.

Stock, J. H. (1991). Confidence intervals for the largest autoregressive root in U.S. macroeconomic time series. *J. Monetary Econom.* **29**, 435-459.

White, J. S. (1958). The limiting distribution of the serial correlation coefficient in the explosive case. *Ann. Math. Statist.* **29**, 1188-1197.

Wright, J. (1999). An empirical likelihood ratio test for a unit root, problem 99.2.1. *Econom. Theory* **15**, 257.

Goldman, Sachs and Co., 85 Broad Street, New York, NY 10004, U.S.A.

E-mail: chinshan.chuang@stanfordalumni.org

Department of Statistics, Chinese University of Hong Kong, Shatin, N.T., Hong Kong.

E-mail: nhchan@sta.cuhk.edu.hk

(Received December 1999; accepted August 2001)