Biometrika (2010), **94**, 2, *pp.* 1–11
Advance Access publication on ? 2008

Printed in Great Britain

Confidence intervals for spectral mean and ratio statistics

By Xiaofeng Shao

Department of Statistics, University of Illinois at Urbana–Champaign, 725 South Wright St, Champaign, Illinois, 61820 USA

xshao@uiuc.edu

Summary

We propose a new method to construct confidence intervals for spectral mean and related ratio statistics of a stationary process, that avoids direct estimation of their asymptotic variances. By introducing a bandwidth, a self-normalization procedure is adopted and the distribution of the new statistic is asymptotically nuisance-parameter free. The bandwidth is chosen using information criteria and a moving average sieve approximation. Through a simulation study, we demonstrate good finite sample performance of our method when the sample size is moderate, while a comparison with empirical likelihood based method for ratio statistics is made confirming wider applicability of our method.

Some key words: Autocorrelation; Cumulant; Ratio statistic; Spectral density; Spectral distribution function.

1. Introduction

Consider a zero-mean stationary process $\{X_t\}_{t\in\mathbb{Z}}$. Denote by

$$I_n(\lambda) = (2\pi n)^{-1}\left|\sum_{t=1}^{n} X_t e^{it\lambda}\right|^2$$

the periodogram and let $\phi : [-\pi, \pi] \to \mathbb{R}$ be a symmetric function with bounded variation. In this paper, we use $\mathbb{R}$, $\mathbb{Z}$, $\mathbb{N}$ to denote the set of real, integer and natural numbers. In time series analysis, a large class of statistics admits the form

$$S(f, \phi) = \int_0^{\pi} \phi(\lambda) f(\lambda) d\lambda,$$

where $f(\cdot)$ is the spectral density function of $\{X_t\}_{t\in\mathbb{Z}}$. As the periodogram $I_n(\lambda)$ is the sample analogue of $f(\lambda)$, a natural estimator of $S(f, \phi)$ is $S(I_n, \phi) = \int_0^{\pi} \phi(\lambda) I_n(\lambda) d\lambda$. In practice, if one is only interested in the patterns of dependence described in terms of autocorrelations, then the ratio statistic $R(f, \phi) = S(f, \phi)/S(f, 1)$, which is estimated by its sample counterpart $R(I_n, \phi) = S(I_n, \phi)/S(I_n, 1)$, is of more practical relevance.

Important examples include $\phi(\lambda) = 2\cos(k\lambda)$, $k \in \mathbb{N}$ and $\phi(\lambda) = 1_{[0,x]}(\lambda)$, $x \in [0, \pi]$. The former corresponds to $S(I_n, \phi) = \hat\gamma_k = n^{-1}\sum_{t=1}^{n-|k|} X_t X_{t+|k|}$, $k \in \mathbb{Z}$, and $R(I_n, \phi) = \hat\rho_k = \hat\gamma_k/\hat\gamma_0$, which consistently estimate the true underlying autocovariance $\gamma_k = \mathrm{cov}(X_t, X_{t-k})$ and the autocorrelation $\rho_k = \gamma_k/\gamma_0$ at the kth lag. The latter corresponds to $S(I_n, \phi) = F_n(x) = \int_0^x I_n(\lambda) d\lambda$ and $R(I_n, \phi) = F_n(x)/F_n(\pi)$, which are $n^{1/2}$-consistent estimators of the spectral distribution function $S(f,\phi) = F(x)$ and its normalized version $R(f,\phi) = F(x)/F(\pi)$. Under some moment assumption and weak dependence conditions on X_t (Brillinger, 1975; Rosenblatt, 1985; Dahlhaus, 1985), we have

$$n^{1/2}\{S(I_n,\phi) - S(f,\phi)\} \to N\{0, \sigma^2(\phi)\} \tag{1}$$

in distribution as $n \to \infty$, where

$$\sigma^2(\phi) = 2\pi \left\{ \int_0^\pi \phi^2(\lambda) f^2(\lambda) d\lambda + \int_0^\pi \int_0^\pi \phi(w_1)\phi(w_2) f_4(w_1, -w_1, -w_2) dw_1 dw_2 \right\},$$

and $f_4(\cdot,\cdot,\cdot)$ is the fourth order cumulant spectral density of X_t. To construct a confidence interval for $S(f,\phi)$, one can replace $\sigma^2(\phi)$ by a consistent estimator. Direct estimation of $\sigma^2(\phi)$ inevitably involves the estimation of the integral of the fourth order cumulant spectra, which has been studied by Taniguchi (1982), Keenan (1987) and Chiu (1988). The mean squared consistency of the proposed estimators requires the existence of the eighth moment and the empirical performance of their methods has not been investigated.

For ratio statistics, Dahlhaus & Janas (1996) proved the validity of the frequency domain bootstrap procedure (Franke & Härdle, 1992) for linear processes with independent and identically distributed innovations. For linear processes with infinite autoregressive representation, Kreiss & Paparoditis (2003) proposed the autoregressive-aided periodogram bootstrap method and showed that the sampling distribution of $n^{1/2}\{S(I_n,\phi) - S(f,\phi)\}$ can be consistently estimated by its bootstrap analogue. Recently, Nordman & Lahiri (2006) developed empirical likelihood based methods to construct confidence intervals for ratio statistics. However, the theory and methods presented in the above-mentioned articles heavily rely on the assumption that X_t is a linear process with independent and identically distributed errors, and seem inapplicable to general stationary process. In particular, the ARMA models with GARCH-type errors are excluded from their framework.

For stationary processes satisfying certain mixing conditions, Romano & Thombs (1996) derived the asymptotic distribution of the sample autocorrelation, which depends on the fourth order cumulants of $\{X_t\}$. The latter authors proposed the subsampling and nonparametric time domain bootstrap methods to approximate the sampling distribution of $n^{1/2}(\hat\rho_k - \rho_k)$, so confidence intervals can be established based on the bootstrap or subsample percentiles. However, the coverage probability is typically sensitive to the choice of the window width or block size. Although there has been some work addressing the optimal choice of the subsampling window width or block size (Politis et al., 1999), these methods typically involve very expensive computations. The main goal of this paper is to propose a new alternative approach to constructing confidence intervals for $S(f,\phi)$ and $R(f,\phi)$, which works for a wide class of stationary processes. The proposed procedures are easy to implement and computationally inexpensive. In particular, no bandwidth selection is involved in the construction of confidence intervals for γ_k and ρ_k.

2. Methodology

Letting $\psi_k = (2\pi)^{-1}\int_0^\pi \phi(\lambda)e^{ik\lambda}d\lambda$, then $S(I_n,\phi) = \sum_{k=1-n}^{n-1}\hat\gamma_k\psi_k$. In the proof of (1), a commonly used strategy (Dahlhaus, 1985) is to prove the central limit theorem for $S_m(I_n,\phi) = \sum_{k=-m}^{m}\hat\gamma_k\psi_k$ for any finite $m \in \mathbb{N}$ and then show that $S(I_n,\phi) - S_m(I_n,\phi)$ is negligible for sufficiently large m. This motivates us to propose $\tilde{S}(I_n,\phi)=\sum_{k=-B_n}^{B_n}\hat{\gamma}_k\psi_k$ as an estimator for $S(f,\phi)$. Here B_n is a sequence of numbers that satisfy $1/B_n+B_n/n\to 0$ as $n\to\infty$. To facilitate the introduction of our idea, we assume that $\{X_{-B_n+1},\ldots,X_n\}$ are observed. Let $g_k=\psi_k+\psi_{-k}$ and $g_0=\psi_0$. Then

$$\tilde{S}(I_n,\phi)=\sum_{k=0}^{B_n}\hat{\gamma}_k g_k\approx\frac{1}{n}\sum_{t=1}^{n}X_t\sum_{k=0}^{B_n}X_{t-k}g_k=\check{S}(I_n,\phi).$$

Now we introduce a partial sum process $K_n(r)=\sum_{t=1}^{\lfloor nr\rfloor}X_t\sum_{k=0}^{B_n}X_{t-k}g_k$, for $r\in[0,1]$; then $K_n(1)=n\check{S}(I_n,\phi)$. Let $\mathcal{D}[0,1]$ be the space of functions on $[0,1]$ which are right continuous and have left limits, endowed with the Skorokhod topology (Billingsley, 1968). In Section 3, the following functional central limit theorem is shown:

$$\frac{1}{n^{1/2}}\{K_n(r)-\lfloor nr\rfloor S(f,\phi)\}\to\sigma(\phi)B(r)$$

in $\mathcal{D}[0,1]$ in a weak sense, where $B(r)$ is Brownian motion and $\lfloor a\rfloor$ stands for the integer part of a. By the continuous mapping theorem, we have

$$\frac{G_{1n}}{G_{2n}}=\frac{n\{K_n(1)-nS(f,\phi)\}^2}{\sum_{t=1}^{n}\{K_n(t/n)-(t/n)K_n(1)\}^2}\to\frac{B^2(1)}{\int_0^1\{B(r)-rB(1)\}^2dr}=U$$

in distribution. Let U_α be the $100(1-\alpha)$th percentile of the distribution for U. Then an asymptotic $100(1-\alpha)\%$ confidence interval for $S(f,\phi)$ is

$$[\{K_n(1)-(U_\alpha G_{2n}/n)^{1/2}\}/n,\{K_n(1)+(U_\alpha G_{2n}/n)^{1/2}\}/n].$$

The upper critical values for the distribution U have been tabulated by Lobato (2001).

Note that $\sigma^2(\phi)$ is canceled out in the above procedure since both the numerator G_{1n} and the denominator G_{2n} are proportional to $\sigma^2(\phi)$. This idea of using random normalization is similar in spirit to Lobato (2001), who applied it to test for non-correlation of a dependent process. Our proposal avoids the thorny issue of estimating $\sigma^2(\phi)$ by introducing a bandwidth parameter B_n, the choice of which is certainly crucial and is described in Section 4. If interest focuses on confidence intervals of γ_{k_0} and ρ_{k_0}, for $k_0\in\mathbb{N}$, then $B_n=k_0$ and thus no bandwidth selection is involved.

In practice, our interest often centers on the ratio statistic $R(f,\phi)$, which is estimated by $R(I_n,\phi)$. Under some suitable conditions, the following approximately holds:

$$\left\{\begin{matrix}\check{S}(I_n,\phi)\\ \check{S}(I_n,1)\end{matrix}\right\}\sim N\left[\left\{\begin{matrix}S(f,\phi)\\ S(f,1)\end{matrix}\right\},n^{-1}\begin{pmatrix}C_{11}\ C_{12}\\ C_{21}\ C_{22}\end{pmatrix}\right].$$

Then we have that approximately,

$$\check{S}(I_n,\phi)-R(f,\phi)\check{S}(I_n,1)\sim N[0,n^{-1}\{C_{11}-2R(f,\phi)C_{12}+R^2(f,\phi)C_{22}\}],$$

where C_{ij}, for $i,j=1,2$ involve the fourth-order cumulants of $\{X_t\}$. Following the idea described above, we define another partial sum process $M_n(r)=\sum_{t=1}^{\lfloor nr\rfloor}X_t^2g_0,\ 0\le r\le 1$, which implies $M_n(1)=n\check{S}(I_n,1)$. By a similar argument to that used in Theorem 1, we have the following functional central limit theorem

$$n^{-1/2}\{K_n(r)-R(f,\phi)M_n(r)\}\to\sigma_R(\phi)B(r)$$

in $\mathcal{D}[0,1]$ for some positive $\sigma_R(\phi)$. Then it follows from the continuous mapping theorem that

$$H_n(\phi) = \frac{n\{K_n(1) - R(f,\phi)M_n(1)\}^2}{\sum_{t=1}^n [K_n(t/n) - (t/n)K_n(1) - R(f,\phi)\{M_n(1) - (t/n)M_n(t/n)\}]^2} \to U$$

in distribution. Let $P_{1n} = \sum_{t=1}^n \{K_n(t/n) - (t/n)K_n(1)\}^2$, $P_{2n} = \sum_{t=1}^n \{M_n(t/n) - (t/n)M_n(1)\}^2$ and $P_{12n} = \sum_{t=1}^n \{K_n(t/n) - (t/n)K_n(1)\}\{M_n(t/n) - (t/n)M_n(1)\}$. Since $\lim_{n\to\infty} \mathrm{pr}(H_n(\phi) \le U_\alpha) = 1 - \alpha$, the approximate $100(1-\alpha)\%$ confidence interval of $R(f,\phi)$ can be obtained by solving $a_n R^2(f,\phi) + b_n R(f,\phi) + c_n \ge 0$, where

$$a_n = P_{2n}U_\alpha - M_n^2(1)n, \qquad b_n = 2nK_n(1)M_n(1) - 2P_{12n}U_\alpha,$$
$$c_n = U_\alpha P_{1n} - nK_n^2(1).$$

When $b_n^2 - 4a_nc_n \ge 0$, let

$$L_n = \min\left\{\frac{-b_n + (b_n^2 - 4a_nc_n)^{1/2}}{2a_n}, \frac{-b_n - (b_n^2 - 4a_nc_n)^{1/2}}{2a_n}\right\},$$
$$U_n = \max\left\{\frac{-b_n + (b_n^2 - 4a_nc_n)^{1/2}}{2a_n}, \frac{-b_n - (b_n^2 - 4a_nc_n)^{1/2}}{2a_n}\right\}.$$

An approximate $100(1-\alpha)\%$ confidence interval for $R(f,\phi)$ is $[L_n, U_n]$ if $a_n < 0$, $(-\infty, L_n] \cup [U_n, \infty)$ if $a_n > 0$. When $b_n^2 - 4a_nc_n < 0$, an empty set is delivered by the procedure. When $a_n = 0$, then it is $[-c_n/b_n, \infty)$ if $b_n > 0$, $(-\infty, -c_n/b_n]$ if $b_n < 0$. When $a_n = b_n = 0$, it yields no meaningful confidence interval.

Remark 1. Following Lobato (2001), one can easily extend the above idea to construct a confidence region for a vector of interest, $\{S(f,\phi^{(1)}), \ldots, S(f,\phi^{(m)})\}'$ or $\{R(f,\phi^{(1)}), \ldots, R(f,\phi^{(m)})\}'$, where $\phi^{(1)}, \ldots, \phi^{(m)}$ are symmetric functions with bounded variation. The asymptotic distribution of the self-normalized statistic would be

$$B_m(1)'\left[\int_0^1 \{B_m(r) - rB_m(1)\}\{B_m(r) - rB_m(1)\}'dr\right]^{-1} B_m(1),$$

where $B_m(r)$ is a m-dimensional vector of independent Brownian motions. The derivation of the confidence region differs only notationally from the displayed univariate case, so we omit the details.

3. THEORETICAL RESULTS

Define the kth $(k = 2, \ldots, 8)$ order cumulant spectral density by

$$f_k(w_1, \ldots, w_{k-1}) = (2\pi)^{1-k} \sum_{j_1,\cdots,j_{k-1}\in\mathbb{Z}} \mathrm{cum}(X_0, X_{j_1}, \cdots, X_{j_{k-1}}) \exp\left(-i\sum_{h=1}^{k-1} w_h j_h\right)$$

Let $Y_{t,m} = (X_t, \ldots, X_{t-m})'$, for $m \in \mathbb{N}$. Define $J_n(r) = n^{-1/2}\sum_{t=1}^{\lfloor nr\rfloor}\{X_tY_{t,m} - \mathbb{E}(X_tY_{t,m})\}$ for $r \in [0,1]$.

THEOREM 1. *Assume that* $\sum_{j=0}^{\infty} g_j^2 < \infty$ *and that* $f_k(\lambda_1, \ldots, \lambda_{k-1})$ *is bounded for* $k = 2, \ldots, 8$. *Further, for any fixed* $m \in \mathbb{N}$, *suppose the finite dimensional convergence of* $J_n(r)$ *holds, i.e.*

$$\text{For any } 0 \le r_1, \ldots, r_s \le 1, \{J_n(r_1), \ldots, J_n(r_s)\} \text{ converges to normal distribution with mean zero and a nonnegative definite covariance matrix } \Sigma_{(m+1)s}. \tag{2}$$

Then

$$\frac{1}{n^{1/2}}[K_n(r) - \mathbb{E}\{K_n(r)\}] \to \sigma(\phi)B(r)$$

in $\mathcal{D}[0,1]$. *Further if* $\sum_{k=B_n+1}^{\infty} |\gamma_k| = o(n^{-1/2})$, *then*

$$n^{-1/2}\{K_n(r) - \lfloor nr \rfloor S(f, \phi)\} \to \sigma(\phi)B(r)$$

in $\mathcal{D}[0,1]$.

Remark 2. The boundedness of the kth order cumulant spectra is implied by

$$\sum_{j_1, \cdots, j_{k-1} \in \mathbb{Z}} |\mathrm{cum}(X_0, X_{j_1}, \ldots, X_{j_{k-1}})| < \infty. \tag{3}$$

Summability conditions on joint cumulants (3) are widely used in spectral analysis. For a linear process $X_t = \sum_{j \in \mathbb{Z}} a_j \varepsilon_{t-j}$ with ε_j being independent and identically distributed, (3) holds if $\sum_{j \in \mathbb{Z}} |a_j| < \infty$ and $\mathbb{E}|\varepsilon_1|^k < \infty$. For a nonlinear process

$$X_t = F(\ldots, \varepsilon_{t-1}, \varepsilon_t), \tag{4}$$

where F is a measurable function for which X_t is a well defined random variable, (3) is satisfied under a geometric moment contraction condition with order k (Wu & Shao, 2004). The process $\{X_t\}$ is geometric moment contracting with order $\alpha > 0$, if there exists a $\rho = \rho(\alpha) \in (0,1)$ such that

$$\mathbb{E}(|X_n^* - X_n|^{\alpha}) \le C\rho^n, \ n \in \mathbb{N}, \tag{5}$$

where $X_n^* = F(\ldots, \varepsilon'_{-1}, \varepsilon'_0, \varepsilon_1, \ldots, \varepsilon_n)$ and $\{\varepsilon'_t\}_{t \in \mathbb{Z}}$ is an independent and identically distributed copy of $\{\varepsilon_t\}_{t \in \mathbb{Z}}$. The property (5) indicates that the process $\{X_n\}$ forgets its past exponentially fast and it can be verified for many nonlinear time series models, including GARCH models of various forms (Wu & Min, 2005; Shao & Wu, 2007).

Remark 3. The assumption on the finite dimensional convergence of $J_n(r)$ is not primitive. To give primitive conditions, we restrict our attention to the nonlinear causal process (4). By the Cr*á*mer-Wold device and a martingale approximation argument (Hannan, 1973; Wu & Min, 2005), it holds provided that $\sum_{t=0}^{\infty} \|\mathcal{P}_0(X_t X_{t-j})\| < \infty$, $j = 0, 1, \ldots, m$. Define the physical dependence measure (Wu, 2005) $\delta_q(t) = \|X_t - X'_t\|_q$, where $X'_t = F(\ldots, \varepsilon_{-1}, \varepsilon'_0, \varepsilon_1, \ldots, \varepsilon_t)$, $t \in \mathbb{N}$, $q \ge 1$. Since

$$\|\mathcal{P}_0(X_t X_{t-j})\| \le \|X_t X_{t-j} - X'_t X'_{t-j}\| \le C\{\delta_4(t) + \delta_4(t-j)\}, \tag{6}$$

(2) holds if $\sum_{t=0}^{\infty} \delta_4(t) < \infty$. Wu (2005) showed that geometric moment contracting of order $\alpha \ge 1$ implies that $\sum_{t=0}^{\infty} \delta_{\alpha}(t) < \infty$. Since the geometric moment contracting property is preserved in ARMA modeling (Shao and Wu, 2007, Theorem 5.2), ARMA-GARCH models satisfy (5) for some $\alpha > 0$. A straightforward calculation shows that the models M_3, M_6 and M_9 used in our simulation work satisfy geometric moment contracting of order 4, consequently they satisfy (2) and (3) with $k = 2, 3, 4$. Further, we

conjecture that the finiteness of the fourth moment and boundedness of f_j, $j = 2, 3, 4$ may suffice for Theorem 1 to hold.

Remark 4. If the interest centers on γ_{k_0} or ρ_{k_0}, $k_0 \in \mathbb{N}$, we only need the functional central limit theorem

$$\frac{1}{n^{1/2}} \sum_{t=1}^{\lfloor nr \rfloor} (X_t X_{t-k_0} - \gamma_{k_0}) \to \{2\pi f_{X_t X_{t-k_0}}(0)\}^{1/2} B(r), \tag{7}$$

where $f_{X_t X_{t-k_0}}(0)$ is the spectral density function of $\{X_t X_{t-k_0}\}_{t\in\mathbb{Z}}$ evaluated at zero frequency. According to Hannan (1973), the assertion (7) holds if $\sum_{t=0}^{\infty} \|\mathcal{P}_0(X_t X_{t-k_0})\| < \infty$, which is implied by $\sum_{t=0}^{\infty} \delta_4(t) < \infty$; compare (6). Other type of conditions that imply (7) can be found in Lobato (2001).

4. Moving average sieve approximation and bandwidth selection

For any purely nondeterministic zero-mean stationary process, the Wold decomposition theorem (Brockwell & Davis, 1991) asserts that

$$X_t = \sum_{j=0}^{\infty} b_j Z_{t-j}, \ b_0 = 1,$$

where $\{Z_t\}_{t\in\mathbb{Z}}$ are mean zero uncorrelated random variables with finite variance σ_Z^2. Given a realization $\{X_t\}_{t=1}^n$, we shall adopt the following moving average sieve approximation to choose the bandwidth B_n:

$$X_t = Z_t + b_1 Z_{t-1} + \cdots + b_q Z_{t-q}.$$

Let $\theta_q = (b_1, \ldots, b_q)'$. In practice, we choose $\hat{q}$ through the information criteria, such as AIC and BIC, and let $B_n = \hat{q}$. The approximating MA($\hat{q}$) model implies $\hat{q}$-dependence, which matches the approximation used in forming our statistic $\check{S}(I_n, \phi)$, where the contribution of $\hat{\gamma}(k)$, $|k| \geq B_n + 1$ is ignored. In the literature, autoregressive sieve approximation has been widely used in various statistical problems, such as estimating the spectral density (Berk, 1974) and bootstrapping the residuals of a time series model to approximate the sampling distribution of linear statistics (Bühlmann, 1997). But in our setting, the moving average sieve approximation seems more natural. Here are the main steps involved in determining $\hat{q}$:

1. For each $q = 0, 1, \ldots, Q$, where Q is pre-specified, find the least squares estimator of θ_q, i.e. $\hat{\theta}_q = \text{argmin } \sigma_q^2(\theta_q)$, where $\sigma_q^2(\theta_q) = n^{-1} \sum_{t=1}^n Z_t^2(\theta_q)$. Here we follow convention and set $Z_0 = Z_{-1} = \cdots = Z_{-q} = 0$ in the calculation of $Z_t(\theta_q)$ based on $\{X_t\}_{t=1}^n$.
2. The optimal q is determined via information criteria, i.e. $\hat{q} = \text{argmin}_q \log\{\sigma_q^2(\hat{\theta}_q)\} + \text{pen}(q)$, where $\text{pen}(q)$ penalizes the model complexity; $\text{pen}(q) = 2(q+1)/n$ if we use AIC, $\text{pen}(q) = 2(q+1)/(n-q-2)$ corresponds to AICC, and $\text{pen}(q) = (q+1)\log n/n$ for BIC.

In the next section we examine the empirical coverage probability of the confidence intervals for both spectral mean and ratio statistics through Monte Carlo simulations.

5. Simulation studies

Let ε_{1t} and ε_{2t} be independent and identically distribution with $N(0,1)$ and $t(5)$ distributions respectively; further let $\varepsilon_{3t} = W_t\{0.5\varepsilon_{3(t-1)}^2 + 0.3\}^{1/2}$, where W_t are independent and identically distributed with the standard normal distribution. Denote by B the backward shift operator. We consider the following models:
M_1: $(1-0.7B)X_t = \varepsilon_{1t}$,
M_2: $(1-0.7B)X_t = 0.6^{1/2}\varepsilon_{2t}$,
M_3: $(1-0.7B)X_t = \varepsilon_{3t}/0.6^{1/2}$,
M_4: $X_t = (1+0.8B)\varepsilon_{1t}$,
M_5: $X_t = (1+0.8B)\{0.6^{1/2}\varepsilon_{2t}\}$,
M_6: $X_t = (1+0.8B)\{\varepsilon_{3t}/0.6^{1/2}\}$,
M_7: $(1-0.7B)X_t = (1+0.8B)\varepsilon_{1t}$,
M_8: $(1-0.7B)X_t = (1+0.8B)\{0.6^{1/2}\varepsilon_{2t}\}$,
M_9: $(1-0.7B)X_t = (1+0.8B)\{\varepsilon_{3t}/0.6^{1/2}\}$.

In the above models, M_1–M_3, M_4–M_6 and M_7–M_9 are AR(1), MA(1) and ARMA(1, 1) models with the three error processes generated from independent and identically distributed standard normal distribution, $t(5)$ distribution and an ARCH(1) process, respectively. The variances for $t(5)$ and ARCH(1) processes are standardized to 1. Note that the processes generated from M_2, M_5 and M_8 do not have finite fourth moment, which is required by our theory. Sample sizes $n = 150$, 600 are investigated.

Table 1 shows the coverages of the confidence intervals for γ_1 based on 10000 replications, and for $F(\pi/4)$ based on 1000 replications. Table 2(a)–(b) gives the coverages for the confidence intervals for corresponding ratio statistics $\rho(1)$ and $F(\pi/4)/F(\pi)$. The number in the round brackets of Table 2 stands for the percentage that produces an empty set. The bandwidth is chosen by AIC criteria. We also tried BIC and AICC criteria. The results are very close to that obtained using AIC, so are not reported here.

[Insert Table 1 about here]

The larger the sample size, the closer the empirical coverage probability is to the nominal level. For the confidence interval of $\gamma(1)$, it appears that the models with normal innovations produce slightly better results than those with $t(5)$ distributions, which outperform those with ARCH(1) innovations, although all of the intervals exhibit under-coverage. In this case the estimated standard error of the true coverage probability is given by $\{\alpha(1-\alpha)/10000\}^{1/2}$, where α is the observed coverage proportion. As seen from Table 2, the coverages for $\rho(1)$ and $F(\pi/4)/F(\pi)$ are fairly close to the nominal level when $n = 600$, and only a small fraction of empty intervals occur when the innovations of the models are from $t(5)$ or ARCH(1). Among non-empty intervals, more than 99.5% of them are of the type $[L_n, U_n]$ when $n = 600$.

A comparison of Tables 1 and 2 suggests that the coverage of the interval for $\rho(1)$ is noticeably closer to the nominal level than that for $\gamma(1)$ for all models and sample sizes. This might be related to the fact that the asymptotic distribution of $\rho(1)$ is less dependent on the variance and the fourth cumulants of the process than that for $\gamma(1)$. The same phenomenon is also observed for $F(\pi/4)$ when compared to its corresponding ratio statistic. Further, it can be seen that MA(1) models slightly outperform corresponding AR(1) and ARMA(1, 1) counterparts almost uniformly in the statistics examined here when $n = 150$, while the advantage is less obvious for $n = 600$. This might be due to very short correlation structure of the MA(1) models.

[**Insert Table 2 about here**]

As suggested by a referee, we compare the finite-sample performance of our method with that offered by the empirical likelihood method (Nordman & Lahiri, 2006). The formulation in the latter paper is limited to ratio statistics and the theoretical validity of empirical likelihood-based confidence interval is only justified for linear processes with independent and identically distributed innovations. Table 2 shows the empirical coverages of the two methods for $\rho(1)$ and $F(\pi/4)/F(\pi)$ based on 10000 replications. For models with independent and identically distributed innovations, $N(0,1)$ or $t(5)$, the performance of the empirical likelihood-based confidence interval is comparable to that delivered by our method, but in the case of ARCH(1) innovation, the empirical likelihood method performs rather poorly.

In terms of computational cost, for the model M_1 at sample size $n = 600$ and 10000 replications, it takes 3.15 seconds of processing time on a Dell PC with Intel Core 2 Duo E 6750 processor to construct a confidence interval for $\rho(1)$. To obtain empirical coverage for the empirical likelihood based method in our simulations, we only need to calculate the empirical likelihood at the true parameter value, which takes 41.69 seconds with the same model, sample size and number of replications. For $F(\pi/4)/F(\pi)$, our method is more expensive computationally due to bandwidth selection involved in our procedure. For $Q = 10$ and 100 replications, it takes 643.04 seconds for our method, while it takes 0.45 seconds for the empirical likelihood method. However, if the goal is to locate the empirical likelihood based confidence interval, one needs to calculate the empirical likelihood at a number of values (Owen, 2001). So constructing the empirical likelihood based confidence interval actually requires more computational time than reported here.

In summary, the finite sample coverage of the proposed confidence intervals seem reasonably good, especially for ratio statistics at a moderate sample size $n = 600$. The ARCH(1) and $t(5)$ errors only slightly affect the sample coverage, when compared to that for the standard normal errors. For ratio statistics, the difference is less apparent, and in some cases, the coverage associated with ARCH(1) and $t(5)$ errors could outperform their counterpart for the standard normal errors. Overall, our method is well supported by the encouraging simulation results. Compared to the empirical likelihood-based method, our approach has wider applicability, and is more appealing to the practitioner, since in general we do not know if the time series at hand is from a linear process with independent and identically distributed innovations.

Acknowledgement

The author thanks Professor A. C. Davison and two anonymous referees for constructive comments that led to improvement of the paper. I also thank Yunwen Yang for helpful discussions on empirical likelihood methods. The work is in part supported by a grant from the U.S. National Science Foundation.

Appendix

In the appendix, $C > 0$ denotes a generic constant that may vary from line to line.

Proof of Theorem 1: Let $Z_{jt} = X_t X_{t-j}$. We first show the finite dimensional convergence. Write

$$n^{-1/2}[K_n(r) - \mathbb{E}\{K_n(r)\}] = n^{-1/2} \sum_{t=1}^{\lfloor nr \rfloor} \sum_{k=0}^{m} \{Z_{kt} - \mathbb{E}(Z_{kt})\} g_k$$
$$+ n^{-1/2} \sum_{t=1}^{\lfloor nr \rfloor} \sum_{k=m+1}^{B_n} \{Z_{kt} - \mathbb{E}(Z_{kt})\} g_k,$$

where the latter term will be shown to be stochastically small (independent of n) when m is sufficiently large. The argument is similar to that used in Dahlhaus (1985); see equation (5) and the discussion therein. Specifically, we shall show that

$$\limsup_{m\to\infty} \sup_{n\in\mathbb{N}} n^{-1} \mathrm{var}\left[\sum_{t=1}^{n} \sum_{j=m+1}^{B_n} \{Z_{jt} - \mathbb{E}(Z_{jt})\} g_j\right] = 0. \quad \text{(A1)}$$

Let $G_m(w) = \sum_{j=m+1}^{B_n} g_j e^{ijw}$. Note that

$$\begin{aligned}
\mathrm{var}\left[\sum_{t=1}^{n} \sum_{j=m+1}^{B_n} \{Z_{jt} - \mathbb{E}(Z_{jt})\} g_j\right] &= \sum_{t,t'=1}^{n} \sum_{j,j'=m+1}^{B_n} \mathrm{cov}(Z_{jt}, Z_{j't'}) g_j g_{j'} \\
&= \sum_{t,t'=1}^{n} \sum_{j,j'=m+1}^{B_n} g_j g_{j'} \{\mathrm{cum}(X_t, X_{t-j}, X_{t'}, X_{t'-j'}) + \mathrm{cov}(X_t, X_{t'}) \\
&\quad \mathrm{cov}(X_{t-j}, X_{t'-j'}) + \mathrm{cov}(X_t, X_{t'-j'}) \mathrm{cov}(X_{t'}, X_{t-j})\} \\
&= I_1 + I_2 + I_3,
\end{aligned}$$

say, where, under the assumption on the boundedness of f_2, f_4,

$$\begin{aligned}
|I_1| &= \left| \int_{[-\pi,\pi]^3} \left| \sum_{t=1}^{n} e^{it(w_2+w_3)} \right|^2 G_m(-w_1) G_m(-w_3) f_4(w_1, w_2, w_3) dw_1 dw_2 dw_3 \right| \\
&\le C \int_{-2\pi}^{2\pi} \left| \sum_{t=1}^{n} e^{it\theta} \right|^2 d\theta \int_{-\pi}^{\pi} |G_m(-w_1)|^2 dw_1 \le Cn \sum_{j=m+1}^{\infty} g_j^2, \\
|I_2| &= \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \left| \sum_{t=1}^{n} e^{it(\lambda+w)} \right|^2 |G_m(w)|^2 d\lambda dw \le Cn \sum_{j=m+1}^{\infty} g_j^2, \\
|I_3| &= \left| \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \left| \sum_{t=1}^{n} e^{it(\lambda+w)} \right|^2 G_m(-\lambda) G_m(w) f(\lambda) f(w) d\lambda dw \right| \\
&\le C \int_{[-\pi,\pi]^2} \left| \sum_{t=1}^{n} e^{it(\lambda+w)} \right|^2 |G_m(-\lambda)|^2 d\lambda dw \le Cn \sum_{j=m+1}^{\infty} g_j^2.
\end{aligned}$$

Thus (A1) follows from the assumption that $\sum_{j=0}^{\infty} g_j^2 < \infty$. It is easy to see that the same argument leads to $\limsup_{m\to\infty} \sup_{n\in\mathbb{N}} n^{-1} \mathrm{var}\left[\sum_{t=1}^{\lfloor nr \rfloor} \sum_{j=m+1}^{B_n} \{Z_{jt} - \mathbb{E}(Z_{jt})\} g_j\right] = 0$ for any $r \in [0,1]$. Therefore, the finite dimensional convergence of $n^{-1/2}[K_n(r) - \mathbb{E}\{K_n(r)\}]$ follows from our assumption on the finite dimensional convergence of $J_n(r)$.

It remains to show the tightness. In view of Theorem 15.6 of Billingsley (1968), it suffices to show that for any $0 \le r_1 < r_2 \le 1$,

$$\mathbb{E} |K_n(r_2) - K_n(r_1) - \mathbb{E}\{K_n(r_2) - K_n(r_1)\}|^4 \le Cn^2 (r_2 - r_1)^2.$$

Write

$$
\begin{aligned}
&\mathbb{E}\,|K_n(r_2) - K_n(r_1) - \mathbb{E}\{K_n(r_2) - K_n(r_1)\}|^4 \\
&= \sum_{t_1,t_2,t_3,t_4=\lfloor nr_1\rfloor+1}^{\lfloor nr_2\rfloor} \sum_{k_1,k_2,k_3,k_4=0}^{B_n} g_{k_1}g_{k_2}g_{k_3}g_{k_4}\mathrm{cum}(Z_{t_1k_1}, Z_{t_2k_2}, Z_{t_3k_3}, Z_{t_4k_4}). \qquad \text{(A2)}
\end{aligned}
$$

Denote by $d\Lambda = d\lambda_1 d\lambda_2 d\lambda_3$ and $dW = dw_1 dw_2 dw_3$. Note that

$$
\mathrm{cum}(Z_{t_1k_1}, Z_{t_2k_2}, Z_{t_3k_3}, Z_{t_4k_4}) = \sum_v \mathrm{cum}(X_{i_j}, i_j \in v_1)\cdots \mathrm{cum}(X_{i_j}, i_j \in v_p),
$$

where the summation is over all indecomposable partitions $v = v_1 \cup \cdots \cup v_p$ of the two-way table

$$
\begin{array}{ll}
t_1 & t_1 - k_1 \\
t_2 & t_2 - k_2 \\
t_3 & t_3 - k_3 \\
t_4 & t_4 - k_4 \quad .
\end{array}
$$

For example, one such term is $\mathrm{cum}(X_{t_1}, X_{t_2-k_2}, X_{t_3}, X_{t_4})\mathrm{cum}(X_{t_1-k_1}, X_{t_2}, X_{t_3-k_3}, X_{t_4-k_4})$, which equals

$$
\begin{aligned}
&\int_{[-\pi,\pi]^3} e^{i(t_2-k_2-t_1)\lambda_1 + i(t_3-t_1)\lambda_2 + i(t_4-t_1)\lambda_3} f_4(\lambda_1,\lambda_2,\lambda_3)d\Lambda \\
&\qquad \times \int_{[-\pi,\pi]^3} e^{i(t_2-t_1+k_1)w_1 + i(t_3-k_3-t_1+k_1)w_2 + i(t_4-k_4-t_1+k_1)w_3} f_4(w_1,w_2,w_3)dW.
\end{aligned}
$$

Let $H_n(\lambda) = \sum_{t=\lfloor nr_1\rfloor+1}^{\lfloor nr_2\rfloor} e^{it\lambda}$ and $W_n(\lambda) = \sum_{j=0}^{B_n} g_j e^{ij\lambda}$. Then the corresponding term in (A2) is

$$
\begin{aligned}
&\int_{[-\pi,\pi]^6} H_n(-\lambda_1-\lambda_2-\lambda_3-w_1-w_2-w_3)H_n(\lambda_1+w_1)H_n(\lambda_2+w_2)H_n(\lambda_3+w_3) \\
&\quad W_n(w_1+w_2+w_3)W_n(-\lambda_1)W_n(-w_2)W_n(-w_3)f_4(\lambda_1,\lambda_2,\lambda_3)f_4(w_1,w_2,w_3)d\Lambda dW,
\end{aligned}
$$

which is smaller in magnitude than

$$
\begin{aligned}
&C\left\{\int_{[-\pi,\pi]^6} |H_n(-\lambda_1-\lambda_2-\lambda_3-w_1-w_2-w_3)|^2|W_n(-\lambda_1)|^2|W_n(-w_2)|^2|W_n(-w_3)|^2 d\Lambda dW\right. \\
&\left.\int_{[-\pi,\pi]^6} |H_n(\lambda_1+w_1)|^2|H_n(\lambda_2+w_2)|^2|H_n(\lambda_3+w_3)|^2|W_n(w_1+w_2+w_3)|^2 d\Lambda dW\right\}^{1/2} \\
&\le C\left\{\int_{-\pi}^{\pi}|H_n(\lambda)|^2 d\lambda \int_{-\pi}^{\pi}|W_n(\lambda)|^2 d\lambda\right\}^2 \le Cn^2(r_2-r_1)^2.
\end{aligned}
$$

Other typical terms are

$$
\begin{aligned}
&\mathrm{cum}(X_{t_1}, X_{t_1-k_1}, X_{t_2}, X_{t_2-k_2}, X_{t_3}, X_{t_3-k_3}, X_{t_4}, X_{t_4-k_4}), \\
&\mathrm{cum}(X_{t_1}, X_{t_2}, X_{t_2-k_2}, X_{t_3})\mathrm{cov}(X_{t_1-k_1}, X_{t_4})\mathrm{cov}(X_{t_3-k_3}, X_{t_4-k_4}), \\
&\mathrm{cum}(X_{t_1}, X_{t_2}, X_{t_2-k_2})\mathrm{cum}(X_{t_3}, X_{t_1-k_1}, X_{t_4-k_4})\mathrm{cov}(X_{t_4}, X_{t_3-k_3}), \\
&\mathrm{cum}(X_{t_1}, X_{t_2}, X_{t_4-k_4})\mathrm{cum}(X_{t_1-k_1}, X_{t_2-k_2}, X_{t_3}, X_{t_3-k_3}, X_{t_4}), \\
&\mathrm{cov}(X_{t_1}, X_{t_2-k_2})\mathrm{cov}(X_{t_1-k_1}, X_{t_3})\mathrm{cov}(X_{t_4}, X_{t_3-k_3})\mathrm{cov}(X_{t_2}, X_{t_4-k_4}).
\end{aligned}
$$

The same bound as above can be established for such terms in view of our assumption on the boundedness of the jth ($j = 2, 3, \ldots, 8$) cumulant spectral density. This establishes the conclusion. ♢

References

Berk, K. N. (1974). Consistent autoregressive spectral estimates. *Ann. Statist.* **2**, 489–502.

Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley.

Brillinger, D. R. (1975). *Time Series: Data Analysis and Theory.* Holden-Day, San Francisco.

Brockwell, P. J. & Davis, R. A. (1991). *Time Series: Theory and Methods 2nd ed.* Springer, New York.

Bühlmann, P. (1997). Sieve bootstrap for time series. *Bernoulli* **3**, 123–148.

Chiu, S. T. (1988). Weighted least squares estimators on the frequency domain for the parameters of a time series. *Ann. Statist.* **16**, 1315–1326.

Dahlhaus, R. (1985). Asymptotic normality of spectral estimates. *J. Mult. Anal.* **16**, 412–431.

Dahlhaus, R. & Janas, D. (1996). A frequency domain bootstrap for ratio statistics in time series analysis. *Ann. Statist.* **24**, 1934–1963.

Franke, J. & Härdle, W. (1992). On bootstrapping kernel spectral estimates. *Ann. Statist.* **20**, 121–145.

Hannan, E. J. (1973). Central limit theorem for time series regression. *Z. Wahrsch. Verw. Gebiete.* **26**, 157–170.

Keenan, D. M. (1987). Limiting behavior of functionals of higher-order sample cumulant spectra. *Ann. Statist.* **15**, 134–151.

Kreiss, J. P. & Paparoditis, E. (2003). Autoregressive-aided periodogram bootstrap for time series. *Ann. Statist.* **31**, 1923–1955.

Lobato, I. N. (2001). Testing that a dependent process is uncorrelated. *J. Am. Statist. Assoc.* **96**, 1066–1076.

Nordman, D. J. & Lahiri, S. N. (2006). A frequency domain empirical likelihood for short- and long-range dependence. *Ann. Statist.* **34**, 3019–3050.

Owen, A. B. (2001). *Empirical Likelihood.* Chapman & Hall, London.

Politis, D. N., Romano, J. P. & Wolf, M. (1999). *Subsampling.* Springer-Verlag, New York.

Romano, J. L. & Thombs, L. A. (1996). Inference for autocorrelations under weak assumptions. *J. Am. Statist. Assoc.* **91**, 590–600.

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

Shao, X. & Wu, W. B. (2007). Asymptotic spectral theory for nonlinear time series. *Ann. Statist.* **35**, 1773–1801.

Taniguchi, M. (1982). On estimation of the integrals of the 4th order cumulant spectral density. *Biometrika* **69**, 117–122.

Wu, W. B. (2005). Nonlinear system theory: another look at dependence. *Proc. Nat. Acad. Sci. USA* **102**, 14150–14154.

Wu, W. B. & Min, W. (2005). On linear processes with dependent innovations. *Stochast. Process. Appl.* **115**, 939–958.

Wu, W. B. & Shao, X. (2004). Limit theorems for iterated random functions. *J. Appl. Prob.* **41**, 425–436.

[*Received January* 2008. *Revised June* 2008]

Table 1. *(a). The percentages of coverage (out of 10000 replications) for the confidence interval for* $\gamma(1)$ *under the nine models. The largest standard error is* 0.44%. *(b). The percentages of coverage (out of 1000 replications) for the confidence interval of* $F(\pi/4)$ *under the nine models. The largest standard error is* 1.41%.

	(1a)				(1b)			
	$n = 150$		$n = 600$		$n = 150$		$n = 600$	
$100(1-\alpha)\%$	90%	95%	90%	95%	90%	95%	90%	95%
M_1	80.1	87.9	88.2	92.8	82.0	87.7	86.9	91.8
M_2	78.2	84.7	86.3	91.9	79.4	85.4	87.7	92.6
M_3	73.6	80.2	82.8	88.7	73.1	79.3	82.6	87.7
M_4	86.3	91.9	89.0	93.7	84.5	90.3	87.5	92.5
M_5	84.0	89.6	87.6	92.9	80.9	87.2	85.1	91.2
M_6	77.5	83.9	83.6	89.0	75.6	82.0	81.6	87.1
M_7	80.5	86.5	87.7	93.3	82.9	87.8	86.3	91.2
M_8	78.5	84.7	86.1	91.6	78.8	84.2	87.4	92.3
M_9	73.5	79.9	82.4	88.0	72.5	79.8	82.2	87.3

Table 2. *(a). The percentages of coverage (out of 10000 replications) for the confidence interval of* $\rho(1)$ *under the nine models. The largest standard error is* 0.45%. *(b). The percentages of coverage (out of 1000 replications) for the confidence interval of* $F(\pi/4)/F(\pi)$ *under the nine models. The largest standard error is* 1.13%. *Here the number in the square brackets stands for the coverage percentage (out of 10000 replications) delivered by the empirical likelihood method; the number in the round brackets is the percentage that produces an empty set by our method. The largest standard error for the empirical likelihood method is* 0.43%.

(2a)	$n = 150$		$n = 600$	
$100(1-\alpha)\%$	90%	95%	90%	95%
M_1	84.0 (0.00) [84.3]	90.3 (0.00) [89.8]	87.7 (0.00) [88.1]	92.8 (0.00) [93.3]
M_2	86.4 (0.44) [84.4]	91.5 (1.28) [90.4]	89.0 (0.03) [88.1]	94.1 (0.09) [93.6]
M_3	86.5 (1.23) [72.5]	90.1 (3.47) [80.6]	88.9 (0.08) [72.6]	94.0 (0.29) [80.4]
M_4	89.1 (0.00) [86.9]	94.6 (0.00) [92.6]	89.9 (0.00) [89.4]	94.9 (0.00) [94.5]
M_5	90.1 (0.41) [87.4]	94.0 (1.29) [92.8]	90.0 (0.00) [89.1]	95.1 (0.06) [94.3]
M_6	88.8 (1.14) [74.3]	91.7 (3.20) [82.0]	89.9 (0.13) [71.0]	95.1 (0.34) [79.0]
M_7	87.0 (0.00) [80.3]	93.0 (0.07) [86.7]	89.5 (0.00) [87.0]	94.3 (0.00) [92.6]
M_8	89.1 (0.75) [80.3]	92.6 (2.15) [86.6]	90.1 (0.09) [87.6]	95.0 (0.13) [92.9]
M_9	88.1 (2.56) [73.6]	88.7 (6.73) [81.1]	90.2 (0.12) [77.1]	95.1 (0.47) [84.5]

(2b)	$n = 150$		$n = 600$	
$100(1-\alpha)\%$	90%	95%	90%	95%
M_1	87.5 (0.00) [84.6]	93.2 (0.00) [90.7]	91.8 (0.00) [88.9]	96.5 (0.00) [94.1]
M_2	86.7 (0.70) [85.0]	91.1 (1.20) [91.0]	91.8 (0.00) [89.2]	96.3 (0.00) [94.4]
M_3	85.0 (1.20) [76.3]	87.2 (4.20) [83.8]	89.9 (0.00) [76.5]	95.1 (0.10) [84.4]
M_4	90.0 (0.00) [87.2]	95.3 (0.00) [92.5]	90.9 (0.00) [89.5]	95.1 (0.00) [94.7]
M_5	90.2 (0.10) [87.7]	93.5 (0.70) [93.1]	88.0 (0.00) [89.8]	93.1 (0.10) [94.9]
M_6	87.6 (0.50) [79.4]	91.4 (2.20) [86.5]	89.4 (0.00) [77.9]	94.7 (0.10) [85.3]
M_7	92.4 (0.00) [83.9]	96.0 (0.00) [90.4]	90.9 (0.00) [88.7]	96.0 (0.00) [94.0]
M_8	90.9 (0.70) [84.4]	91.9 (1.70) [90.4]	92.5 (0.00) [89.4]	96.8 (0.00) [94.4]
M_9	88.4 (2.90) [78.2]	89.6 (6.70) [85.3]	91.4 (0.20) [80.2]	94.9 (0.60) [87.2]