

Cressie-Read Power-Divergence Statistics for Non-Gaussian Vector Stationary Processes

HIROAKI OGATA

School of International Liberal Studies, Waseda University

MASANOBU TANIGUCHI

School of Fundamental Science and Engineering, Waseda University

ABSTRACT. For a class of vector-valued non-Gaussian stationary processes, we develop the Cressie-Read power-divergence (CR) statistic approach which has been proposed for the i.i.d. case. The CR statistic includes empirical likelihood as a special case. Therefore, by adopting this CR statistic approach, the theory of estimation and testing based on empirical likelihood is greatly extended. We use an extended Whittle likelihood as score function and derive the asymptotic distribution of the CR statistic. We apply this result to estimation of autocorrelation and the AR coefficient, and get narrower confidence intervals than those obtained by existing methods. We also consider the power properties of the test based on asymptotic theory. Under a sequence of contiguous local alternatives, we give the asymptotic distribution of the CR statistic. The problem of testing autocorrelation is discussed and we introduce some interesting properties of the local power.

Key words: Empirical likelihood; Estimating function; Local asymptotic normality; Spectral density matrix; Whittle likelihood;

SHORT TITLE: *CR statistic for stationary processes*

1. Introduction

The empirical likelihood method was introduced as a nonparametric method of inference based on a data-driven likelihood ratio function (Owen, 1988, 1990) in the i.i.d. case. It is shown that -2 times the logarithm of an empirical likelihood ratio is asymptotically chi-square distributed, and this is used for constructing confidence regions for parameters which are specified by an estimating function. Empirical likelihood approaches have been applied to various statistical models. For example, Zhu & Xue (2006) dealt with an empirical likelihood-based inference for the parameters in a partially linear single index model using the bias-corrected empirical likelihood.

Instead of the empirical likelihood ratio, Baggerly (1998) used the Cressie-Read power-divergence (CR) statistic as a test statistic and showed that it is also asymptotically chi-square distributed in the i.i.d. case. The CR statistic has user-specified parameter $\nu \in (-\infty, \infty)$ and contains the empirical likelihood statistic as a special case ($\nu = 0$). Moreover, it encompasses several commonly-used tests, i.e., the Neyman-modified χ^2 -statistic ($\nu = -2$), the maximum entropy, minimum information or Kullback-Leibler statistic ($\nu = -1$), the Freeman-Tukey statistic ($\nu = -1/2$) and Pearson's χ^2 -statistic ($\nu = 1$). As the asymptotic result of the CR statistic is established for any given value of ν , the theory of empirical likelihood is extended dramatically.

Although the empirical likelihood method has been investigated by many authors, this research has mostly focused on independent observations. For dependent observations, Monti (1997) applied Owen's empirical likelihood approach to scalar linear processes. She used the derivative of the Whittle likelihood, which is an approximated likelihood function in frequency domain, as a score function and showed that the empirical likelihood ratio is asymptotically chi-square distributed. Ogata & Taniguchi (2006) extended Monti's approach to the case of non-Gaussian vector stationary processes and provided a rigorous asymptotic theory for the empirical Whittle likelihood approach. They also applied the empirical likelihood theory to minimum contrast estimation. Details on minimum contrast estimation can be found in Taniguchi (1987).

In this paper, we develop Baggerly's results on the CR statistic for the i.i.d. case and apply this to time series models, especially, non-Gaussian vector stationary processes. Specifically, we derive the asymptotic distribution of the CR statistic $CR_\nu(\boldsymbol{\theta})$ based on the frequency domain estimating function described by an unknown parameter $\boldsymbol{\theta}$. Then it is shown that $CR_\nu(\boldsymbol{\theta})$ converges to a sum of Gamma distributions for any given value of ν . Using this result, we can consider interval estimation and testing problems of various indices in time series analysis.

Specifically, we apply this result to construct confidence intervals of the autocorrelation $\rho(\delta)$, which is a very important and fundamental index in time series analysis. We obtain confidence intervals of autocorrelation using $CR_\nu(\boldsymbol{\theta})$ with various ν . We also get confidence intervals of autocorrelation using the asymptotics of sample autocorrelation. We observe that our CR statistic method gives narrower confidence intervals than the usual sample autocorrelation method when the model is further away from the unit root process. We can also see that the asymptotic distribution of the CR statistic does not include the unknown parameter in the case of scalar processes while that of the usual sample autocorrelation depends on it. Furthermore, we consider estimating the coefficient of the AR model. We compare the length of confidence intervals of the AR coefficient using the CR statistic with the Yule-Walker estimator, which is one of the most important estimators of the AR coefficient. We get the result that the length of confidence intervals with the CR method is shorter than those generated by the Yule-Walker estimator. Besides, for potential practical applications, this CR method can also be applied to prediction problems, interpolation problems and smoothing problems in finance and econometrics, etc. We refer to Ogata & Taniguchi (2006, section 3) for details.

A power property of the test based on the CR statistic is also discussed. The asymptotic distribution under a sequence of contiguous local alternatives is given.

As an example, we consider the problem of testing autocorrelation and investigate its power property. It is seen that the local power is hardly affected by the non-Gaussianity of the process, and that the local power becomes larger as the process tends toward the unit root process.

This paper is organized as follows. In section 2, we summarize existing approaches to asymptotic theory for empirical likelihood ratio in frequency domain. In section 3, we propose a more generalized method of the CR statistic for non-Gaussian vector stationary processes, and give the asymptotic distribution. Section 4 provides power properties of the test based on the CR statistic. The asymptotic distribution under a sequence of contiguous local alternatives is derived. In section 5, we give simulations of confidence intervals of autocorrelation and the AR coefficient. We also give the numerical study on the power of the test of autocorrelation. The advantages of the CR method are summarized in section

6.

As for notations used in this paper, we denote the a th component of a vector \mathbf{X} by X_a and denote the (a, b) -component of the matrices A and A^{-1} by A_{ab} and A^{ab} , respectively. A^* denotes the conjugate transpose of a matrix A , and I_s represents an $s \times s$ unit matrix. We denote the set of all integers by \mathbf{Z} .

2. Setting

Consider a vector-valued linear process $\{\mathbf{X}(t); t \in \mathbf{Z}\}$ generated by

$$\mathbf{X}(t) = \sum_{j=0}^{\infty} A(j) \mathbf{U}(t-j), \quad t \in \mathbf{Z}, \quad (1)$$

where the innovations $\mathbf{U}(t)$ are i.i.d. s -vector random variables with probability density $p(\mathbf{u}) > 0$ on \mathbf{R}^s and the $A(j)$ are s by s matrices. The components of \mathbf{X}, \mathbf{U} and A are all real. We make the following assumptions.

Assumption 1

(i) The coefficient matrices $A(j)$ satisfy

$$\sum_{j=0}^{\infty} j^{1/2} \|A(j)\| < \infty,$$

where $\|A(j)\|$ denotes the sum of all the absolute values of the entries of $A(j)$.

(ii) The probability density $p(\cdot)$ satisfies

$$\lim_{\|\mathbf{u}\| \rightarrow \infty} p(\mathbf{u}) = 0, \quad \int \mathbf{u} p(\mathbf{u}) d\mathbf{u} = \mathbf{0}, \quad \text{and} \quad \int \mathbf{u} \mathbf{u}' p(\mathbf{u}) d\mathbf{u} = I_s,$$

where $\|\mathbf{u}\| = \sqrt{\mathbf{u}'\mathbf{u}}$ and I_s denotes the s by s identity matrix.

(iii) $\int \|\mathbf{u}\|^4 p(\mathbf{u}) d\mathbf{u} < \infty$.

Remark 1 Assumption 1 (i) implies that the dependence between $\mathbf{X}(t)$ and $\mathbf{X}(t+l)$ becomes weaker as the time lag l becomes larger. (ii) is the condition for innovation $\mathbf{U}(t)$, which is commonly used in the literature. We need (iii) to control the maximum of important terms that will appear in the proofs. For example, the non-Gaussian VARMA model with fourth order moments, which is one of the most typical models in time series, satisfies these conditions, so assumption 1 is natural.

The process $\{\mathbf{X}(t)\}$ has spectral density matrix which is expressed as

$$g(\omega) = (2\pi)^{-1} k(\omega) k(\omega)^*, \quad -\pi \leq \omega \leq \pi,$$

where $k(\omega) = \sum_{j=0}^{\infty} A(j) e^{i\omega j}$. For the observed stretch $\mathbf{X}(t)$, $t = 1, \dots, T$, we denote the periodogram by $I_T(\omega)$; namely

$$I_T(\omega) = (2\pi T)^{-1} \mathbf{d}_T(\omega) \mathbf{d}_T(\omega)^*, \quad -\pi \leq \omega \leq \pi.$$

where $\mathbf{d}_T(\omega) = \sum_{t=1}^T \mathbf{X}(t)e^{-i\omega t}$.

First, for the vector-valued non-Gaussian linear process (1) with the true spectral density matrix $g(\omega)$, we consider to fit a parametric spectral model $f_{\boldsymbol{\theta}}(\omega)$ with $\boldsymbol{\theta} \in \Theta \subset \mathbf{R}^q$, to $g(\omega)$. Here $f_{\boldsymbol{\theta}}(\omega)$ may be different from $g(\omega)$. Consider the multivariate Whittle likelihood

$$W(\boldsymbol{\theta}) \equiv \int_{-\pi}^{\pi} [\log \det f_{\boldsymbol{\theta}}(\omega) + \text{tr}\{f_{\boldsymbol{\theta}}(\omega)^{-1}I_T(\omega)\}] d\omega.$$

Here, we impose the following assumption on $f_{\boldsymbol{\theta}}(\omega)$.

Assumption 2

- (i) Θ is a compact subset of \mathbf{R}^q .
- (ii) $f_{\boldsymbol{\theta}}(\omega)$ is continuously twice differentiable with respect to $\boldsymbol{\theta} \in \Theta$.
- (iii) $f_{\boldsymbol{\theta}}(\omega) \in \mathcal{F}$. Here \mathcal{F} is the parametric spectral family whose element is expressed as

$$f_{\boldsymbol{\theta}}(\omega) = \left(\sum_{j=0}^{\infty} B_j(\boldsymbol{\theta})e^{ij\omega} \right) \Sigma \left(\sum_{j=0}^{\infty} B_j(\boldsymbol{\theta})e^{ij\omega} \right)^*, \quad (2)$$

where the $B_j(\boldsymbol{\theta})$ are $s \times s$ matrices, $B_0(\boldsymbol{\theta})$ is an $s \times s$ identity matrix and Σ is an $s \times s$ symmetric matrix.

The model (2) is a spectral form of a general linear process. Suppose that the parameter $\boldsymbol{\theta}$ does not depend on Σ , which corresponds to the covariance matrix of the innovation, and that the B_j depend on $\boldsymbol{\theta}$. Then we call $\boldsymbol{\theta}$ "innovation-free". Let $\boldsymbol{\theta}_0$ be the value defined by

$$\frac{\partial}{\partial \boldsymbol{\theta}} \int_{-\pi}^{\pi} [\log \det f_{\boldsymbol{\theta}}(\omega) + \text{tr}\{f_{\boldsymbol{\theta}}(\omega)^{-1}g(\omega)\}] d\omega \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \mathbf{0}, \quad (3)$$

which is called the pseudo-true value of $\boldsymbol{\theta}$. If we use

$$D(f_{\boldsymbol{\theta}}, g) := \int_{-\pi}^{\pi} [\log \det f_{\boldsymbol{\theta}}(\omega) + \text{tr}\{f_{\boldsymbol{\theta}}(\omega)^{-1}g(\omega)\}] d\omega$$

as a disparity measure between $f_{\boldsymbol{\theta}}(\omega)$ and $g(\omega)$, then $\boldsymbol{\theta}_0$ means the point minimizing the $D(f_{\boldsymbol{\theta}}, g)$ under natural conditions. We know that, if $\boldsymbol{\theta}$ is innovation-free, then $\int_{-\pi}^{\pi} \log \det f_{\boldsymbol{\theta}}(\omega) d\omega$ is independent of $\boldsymbol{\theta}$ (Brockwell & Davis, 1991, p.191; or Priestley, 1981, p.760). Therefore (3) implies

$$\frac{\partial}{\partial \boldsymbol{\theta}} \int_{-\pi}^{\pi} \text{tr}\{f_{\boldsymbol{\theta}}(\omega)^{-1}g(\omega)\} d\omega \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \mathbf{0}. \quad (4)$$

Furthermore, this setting is unexpectedly useful for many other situations. By choosing $f_{\boldsymbol{\theta}}(\omega)$ appropriately, $\boldsymbol{\theta}$ can express various important indices of a time series model. We briefly give the following example of the autocorrelation, which is a very important and fundamental quantity in time series analysis.

Example 1 (Autocorrelation) Denote by $\Gamma(\delta) = Cov[\mathbf{X}(t), \mathbf{X}(t + \delta)]$ the autocovariance matrix of \mathbf{X} with lag δ . Let us consider the linear process defined in (1), and let

$$\boldsymbol{\theta} = (\theta_{11}, \dots, \theta_{1s}, \dots, \theta_{s1}, \dots, \theta_{ss})',$$

$$A(\boldsymbol{\theta}) = \begin{pmatrix} \theta_{11} & \cdots & \theta_{1s} \\ \vdots & \ddots & \vdots \\ \theta_{s1} & \cdots & \theta_{ss} \end{pmatrix}.$$

If we set

$$f_{\boldsymbol{\theta}}(\omega) = (I_s - A(\boldsymbol{\theta})e^{i\delta\omega})^{-1}(I_s - A(\boldsymbol{\theta})e^{i\delta\omega})^{-1*},$$

then condition (4) implies

$$\sum_{j=1}^s [\theta_0]_{\beta_1 j} \int_{-\pi}^{\pi} g(\omega)_{j\beta_2} d\omega = \int_{-\pi}^{\pi} e^{i\delta\omega} g(\omega)_{\beta_2\beta_1} d\omega \quad (\beta_1, \beta_2 = 1, \dots, s). \quad (5)$$

From Helglotz's theorem it is known that

$$\Gamma(\delta) = \int_{-\pi}^{\pi} e^{i\delta\omega} g(\omega) d\omega, \quad (6)$$

(c.f. Brockwell & Davis, 1991). From (5) and (6), we obtain

$$A(\boldsymbol{\theta}_0)\Gamma(0) = \Gamma(\delta), \quad i.e., \quad A(\boldsymbol{\theta}_0) = \Gamma(\delta)\Gamma(0)^{-1}.$$

Hence, using the Whittle likelihood $W(\boldsymbol{\theta})$, we can discuss an estimator for the quantity $\Gamma(\delta)\Gamma(0)^{-1}$, which is a generalized quantity of the usual autocorrelation $\rho(\delta) = \Gamma(\delta)/\Gamma(0)$ in the scalar case.

Besides, $\boldsymbol{\theta}$ can express other important indices in many other situations such as prediction, interpolation and smoothing problems. For detailed information, see Ogata & Taniguchi (2006, sections 3 and 4).

3. Cressie-Read power-divergence statistic for time series

In this section, motivated by Baggerly's (1998) results in the i.i.d. case, we suggest the CR statistic $CR_{\nu}(\boldsymbol{\theta})$ for time series,

$$CR_{\nu}(\boldsymbol{\theta}) = \min_{\boldsymbol{w}} \left\{ \frac{2}{\nu(\nu+1)} \sum_{t=1}^T \{(Tw_t)^{-\nu} - 1\} \left| \sum_{t=1}^T w_t \mathbf{m}_{\boldsymbol{\theta}}(\lambda_t) = \mathbf{0}, \sum_{t=1}^T w_t = 1, w_t \geq 0 \right. \right\}, \quad (7)$$

where

$$\mathbf{m}_{\boldsymbol{\theta}}(\lambda_t) = \frac{\partial}{\partial \boldsymbol{\theta}} \text{tr}\{f_{\boldsymbol{\theta}}(\lambda_t)^{-1} I_T(\lambda_t)\} \in \mathbf{R}^q, \quad \lambda_t = \frac{2\pi t}{T} \quad (t = 1, \dots, T)$$

and $\nu \in (-\infty, \infty)$. The CR statistic contains the user-specified parameter $\nu \in (-\infty, \infty)$ and encompasses several commonly-used tests, i.e., the Neyman-modified χ^2 -statistic ($\nu = -2$), the maximum entropy, minimum information or Kullback-Leibler statistic ($\nu = -1$),

the Freeman-Tukey statistic ($\nu = -1/2$) and Pearson's χ^2 -statistic ($\nu = 1$). Especially, it includes the empirical likelihood statistic ($\nu = 0$), too. Hence, the CR statistic is a much broader criterion than the empirical likelihood ratio, and its asymptotic theory covers the results of Ogata & Taniguchi (2006), which show that -2 times the logarithm of the empirical likelihood ratio converges to a sum of Gamma distributions under time series models. Moreover, as numerical simulations show that the Kullback-Leibler statistic behaves better than the empirical likelihood in some cases (see section 5), this general CR approach is well worth consideration.

The asymptotics of the CR statistic are given as follows.

Theorem 1 Suppose that assumptions 1 and 2 hold. For any given $\nu \in (-\infty, \infty)$, as $T \rightarrow \infty$,

$$CR_\nu(\boldsymbol{\theta}_0) \xrightarrow{d} (G\mathbf{N})'(GN), \quad (8)$$

where \mathbf{N} is a standard q -dimensional normal random vector and $G = W^{-1/2}V^{1/2}$. Here V is a q by q matrix whose (l_1, l_2) element is

$$\begin{aligned} V_{l_1 l_2} &= \frac{1}{\pi} \int_{-\pi}^{\pi} \text{tr} \left[g(\omega) \partial_{l_1} f_{\boldsymbol{\theta}_0}^{-1}(\omega) g(\omega) \partial_{l_2} f_{\boldsymbol{\theta}_0}^{-1}(\omega) \right] d\omega \\ &\quad + \frac{1}{2\pi} \sum_{u_1 u_2 v_1 v_2=1}^s \iint_{-\pi}^{\pi} \partial_{l_1} f_{\boldsymbol{\theta}_0}^{u_1 u_2}(\omega_1) \partial_{l_2} f_{\boldsymbol{\theta}_0}^{v_1 v_2}(\omega_2) g_{u_1 u_2 v_1 v_2}^{X^4}(-\omega_1, \omega_2, -\omega_2) d\omega_1 d\omega_2, \end{aligned}$$

and W is a q by q matrix whose (l_1, l_2) element is

$$\begin{aligned} W_{l_1 l_2} &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \text{tr} \left[g(\omega) \partial_{l_1} f_{\boldsymbol{\theta}_0}^{-1}(\omega) g(\omega) \partial_{l_2} f_{\boldsymbol{\theta}_0}^{-1}(\omega) \right] d\omega \\ &\quad + \frac{1}{2\pi} \int_{-\pi}^{\pi} \text{tr} \left[g(\omega) \partial_{l_1} f_{\boldsymbol{\theta}_0}^{-1}(\omega) \right] \text{tr} \left[g(\omega) \partial_{l_2} f_{\boldsymbol{\theta}_0}^{-1}(\omega) \right] d\omega, \end{aligned}$$

where

$$\partial_{l_1} f_{\boldsymbol{\theta}_0}^{-1}(\omega) = \left. \frac{\partial}{\partial \theta_{l_1}} f_{\boldsymbol{\theta}}^{-1}(\omega) \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$$

and $g_{\dots}^{X^4}(\cdot, \cdot, \cdot)$ is the fourth-order spectral density of the process $\{\mathbf{X}(t)\}$; namely

$$g_{u_1 u_2 u_3 u_4}^{X^4}(\omega_1, \omega_2, \omega_3) = (2\pi)^{-3} \sum_{j_1 j_2 j_3=-\infty}^{\infty} c_{u_1 u_2 u_3 u_4}^{X^4}(j_1, j_2, j_3) \exp\{-i(j_1 \omega_1 + j_2 \omega_2 + j_3 \omega_3)\},$$

where

$$c_{u_1 u_2 u_3 u_4}^{X^4}(j_1, j_2, j_3) = \text{cum}\{X_{u_1}(t), X_{u_2}(t+j_1), X_{u_3}(t+j_2), X_{u_4}(t+j_3)\}.$$

The proofs of the theorems appear in the Appendix.

Remark 2 Hjort *et al.* (2008) provide a general investigation on empirical likelihood ratio estimation, which does not require the assumption of i.i.d. observations. However they give the asymptotics only for the logarithm of the empirical likelihood ratio, whereas we focus on the CR statistic. The CR statistic includes the empirical likelihood ratio as a

special case, so we can compare the various test statistics when constructing confidence intervals. Moreover, we develop a systematic asymptotics in the frequency domain while the main stream of Hjort *et al.* (2008) is for independent samples. By considering the frequency domain approach, we are able to deal with the estimation of various important indices in a time series setting, such as autocorrelation (c.f. example 1), prediction error, interpolation error and so on.

Remark 3 Denote the eigenvalues of $G'G$ by a_1, \dots, a_q . Then we can write the right hand side of (8) as

$$(GN)'(GN) = \sum_{i=1}^q \gamma_i, \quad (9)$$

where γ_i is distributed as $Gamma(2^{-1}, (2a_i)^{-1})$.

V and W contain the unknown spectral density matrix $g(\omega)$ and the fourth-order spectral density $g_{u_1 u_2 u_3 u_4}^{X^4}(-\omega_1, \omega_2, -\omega_2)$. In practice, we can make appropriate consistent estimators \hat{V} and \hat{W} of V and W , respectively by use of methods in Taniguchi (1982) and Keenan (1987). Then, from the Slutsky theorem it follows that

$$(\hat{GN})'(\hat{GN}) \xrightarrow{d} (GN)'(GN) = \sum_{i=1}^q \gamma_i, \quad (10)$$

where $\hat{G} = \hat{W}^{-1/2} \hat{V}^{1/2}$. Using theorem 1, we can construct confidence regions for $\boldsymbol{\theta}$. First, we choose a proper threshold value z_α , which is the α -percentile corresponding to the approximation to (9) based on the relation (10). Then we calculate $CR_\nu(\boldsymbol{\theta})$ at numerous points over Θ and construct the region

$$C_{\alpha,T}(\boldsymbol{\theta}) = \{\boldsymbol{\theta} \mid CR_\nu(\boldsymbol{\theta}) < z_\alpha\}. \quad (11)$$

Remark 4 When the process (1) is scalar, V and W in the theorem are identical. Then we can see that $G = I_q$ and the asymptotic distribution in (8) becomes the chi-square distribution with q degrees of freedom.

4. Power property

In this section, we consider a power property of the test based on theorem 1. From now on, let the coefficient matrices $A(j)$ of (1) be parameterized by $\boldsymbol{\theta} \in \Theta \subset \mathbf{R}^q$. Write

$$A_{\boldsymbol{\theta}}(z) = \sum_{j=0}^{\infty} A_{\boldsymbol{\theta}}(j) z^j, \quad |z| < 1.$$

In this section, we need the local asymptotic normality for a stochastic process, which is seen in Taniguchi & Kakizawa (2000, section 2.2). We thus make the following assumptions (c.f. Taniguchi & Kakizawa, 2000, p.37).

Assumption 3

- (i) (a) Every $A_{\boldsymbol{\theta}}(j)$ is continuously two times differentiable with respect to $\boldsymbol{\theta}$, and the derivatives satisfy

$$|\partial_{u_1} \partial_{u_2} \dots \partial_{u_k} A_{\boldsymbol{\theta}, l_1 l_2}(j)| = O\{j^{-1+\epsilon}(\log j)^k\}, \quad k = 0, 1, 2$$

for $l_1, l_2 = 1, \dots, s$ and for some ϵ ($0 < \epsilon < 1/2$).

- (b) $\det A_{\boldsymbol{\theta}}(z) \neq 0$ for $|z| < 1$ and $A_{\boldsymbol{\theta}}^{-1}(z)$ can be expanded as follows:

$$A_{\boldsymbol{\theta}}^{-1}(z) = I_s + B_{\boldsymbol{\theta}}(1)z + B_{\boldsymbol{\theta}}(2)z^2 + \dots$$

- (c) Every $B_{\boldsymbol{\theta}}(j)$ is continuously two times differentiable with respect to $\boldsymbol{\theta}$, and the derivatives satisfy

$$|\partial_{u_1} \partial_{u_2} \dots \partial_{u_k} B_{\boldsymbol{\theta}, l_1 l_2}(j)| = O\{j^{-1-\epsilon}(\log j)^k\}, \quad k = 0, 1, 2$$

for $l_1, l_2 = 1, \dots, s$.

- (ii) The continuous derivative Dp of $p(\cdot)$ exists on \mathbf{R}^s .

- (iii) $\int \|\boldsymbol{\kappa}(\mathbf{u})\|^4 p(\mathbf{u}) d\mathbf{u} < \infty$, where $\boldsymbol{\kappa}(\mathbf{u}) = p^{-1}(\mathbf{u}) Dp(\mathbf{u})$.

Remark 5 The conditions stated in assumption 3 are not restrictive. The typical models in time series analysis such as the VAR, VMA and VARMA models satisfy these conditions. Besides the more complicated FARIMA model with long range dependence satisfies them. Therefore we can say that assumption 3 is natural.

Consider the problem of testing

$$H : \boldsymbol{\theta} = \boldsymbol{\theta}_0 \quad \text{against} \quad A : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0.$$

To see the goodness of our test we evaluate the local power under the sequence of local alternatives $A_T : \boldsymbol{\theta}_T = \boldsymbol{\theta}_0 + T^{-1/2} \mathbf{h}$, where $\mathbf{h} = (h_1, \dots, h_q)'$. Define $c^{\boldsymbol{\kappa}\mathbf{X}}(j) = \text{cum}\{\boldsymbol{\kappa}(\mathbf{U}(t)), \mathbf{X}(t+j)'\}$. Then the cross-spectral density matrix $g^{\boldsymbol{\kappa}\mathbf{X}}(\omega)$ is given by the following relation,

$$c^{\boldsymbol{\kappa}\mathbf{X}}(j) = \int_{-\pi}^{\pi} e^{ij\omega} g^{\boldsymbol{\kappa}\mathbf{X}}(\omega) d\omega.$$

Then we get the following theorem.

Theorem 2 Let G, V, W and \mathbf{N} be the same matrices and q -dimensional standard normal vector as defined in theorem 1. Under the sequence of local alternatives A_T , for any given $\nu \in (-\infty, \infty)$,

$$CR_{\nu}(\boldsymbol{\theta}_0) \xrightarrow{d} (\mathbf{GN} + \boldsymbol{\mu})'(\mathbf{GN} + \boldsymbol{\mu}),$$

where $\boldsymbol{\mu} = 2W^{-1/2}\boldsymbol{\tau}$. Here $\boldsymbol{\tau} = (\tau_1, \dots, \tau_q)'$ with

$$\tau_i = \int_{-\pi}^{\pi} \text{tr} \left[g(\omega) \partial_i f_{\boldsymbol{\theta}_0}^{-1}(\omega) g^{\boldsymbol{\kappa}\mathbf{X}}(\omega) \left\{ \sum_{j=1}^{\infty} B_{\mathbf{h}'\partial\boldsymbol{\theta}_0}(j) e^{i\omega j} \right\} \right] d\omega$$

and

$$B_{\mathbf{h}'\partial\boldsymbol{\theta}_0}(j) = \sum_{l=1}^q h_l \frac{\partial B_{\boldsymbol{\theta}_0}(j)}{\partial \theta_l}.$$

The difference with theorem 1 is that we are considering the asymptotic distribution of the test statistic under a sequence of "contiguous alternatives A_T ", and that its normed factorization $GN + \mu$ has mean μ . This μ means deviation from the asymptotic distribution under the null hypothesis, so the magnitude of $|\mu|$ indicates the magnitude of the power of the test. For simplicity, let us consider the scalar case. The limiting power of the test with significance level α is expressed as

$$P(\mu) = 1 - Pr\{(GN + \mu)^2 < z_\alpha\} = 1 - Pr\{-\sqrt{z_\alpha} < GN + \mu < \sqrt{z_\alpha}\} = 1 - Q(\mu) \quad (\text{say}),$$

where z_α is the α -percentile of the limiting distribution under the null hypothesis. Obviously, $Q(\mu)$ attains its maximum when $\mu = 0$, and larger $|\mu|$ gives smaller $Q(\mu)$. Therefore, the power $P(\mu)$ becomes larger when the magnitude of $|\mu|$ is larger.

5. Numerical example

In this section, we give numerical simulations for theorems 1 and 2. Let us consider the following scalar-valued AR(1) model,

$$X(t) = bX(t-1) + U(t), \quad (12)$$

where $|b| < 1$, and the $U(t)$ are independent and identically distributed, where the distribution of $U(t)$ satisfies (ii) and (iii) of assumption 1.

As an example of theorem 1, we discuss the estimation of the quantity $\theta_0 = \rho(\delta)$, the autocorrelation with lag δ . As is seen in example 1, we set $f_\theta(\omega) = |1 - \theta e^{i\delta\omega}|^{-2}$ and calculate $CR_\nu(\theta)$ at numerous point over $(-1, 1)$. Since the process (12) is scalar, the asymptotic distribution of $CR_\nu(\theta_0)$ is chi-square with 1 degree of freedom, χ_1^2 (see remark 4). Then we construct the interval $C_{\alpha,T}(\theta)$ in (11) where z_α is the α -percentile of χ_1^2 and get the α percent confidence interval of $\theta_0 = \rho(\delta)$.

The autocorrelation $\rho(\delta)$ can also be estimated by using the sample autocorrelation $\hat{\rho}(\delta) = \hat{\gamma}(\delta)/\hat{\gamma}(0)$, where $\hat{\gamma}(\delta) = T^{-1} \sum_{t=1}^{T-|\delta|} (X(t) - \bar{X}_T)(X(t+\delta) - \bar{X}_T)$ with $\bar{X}_T = T^{-1} \sum_{t=1}^T X(t)$. It is known (e.g., Brockwell & Davis, 1991, theorem 7.2.1) that the quantity $\sqrt{T}(\hat{\rho}(\delta) - \rho(\delta))$ is asymptotically normally distributed with mean zero and variance

$$\sum_{k=-\infty}^{\infty} \{\rho(k+\delta)^2 + \rho(k-\delta)\rho(k+\delta) + 2\rho(\delta)^2\rho(k)^2 - 4\rho(\delta)\rho(k)\rho(k+\delta)\}. \quad (13)$$

Then we can also get the confidence interval of $\theta_0 = \rho(\delta)$ with sample autocorrelation. But the asymptotic variance (13) includes the unknown parameter ρ . Therefore, we recommend the CR method rather than the sample autocorrelation method when the process is scalar.

The results of our simulations are as follows. Let the innovation $U(t)$ have t -distribution with 5 degrees of freedom and generate $X(1), \dots, X(200)$ from (12), *i.e.* $T = 200$. Then we estimate the autocorrelation with lag $\delta = 2$. In the AR(1) model (12), the autocorrelation $\rho(\delta)$ is $b^{|\delta|}$, hence $\theta_0 = b^2$. Table 1 shows 90% confidence intervals of θ_0 by use of the CR method ($\nu = -2, -1, -1/2, 0, 1, 2$) and the sample autocorrelation (SAC) method for $b = 0.1, 0.5$, and 0.9 . The upper line in each cell shows the 90% confidence interval and the lower line shows the length of the interval. Except for a few cases, the length of the interval obtained by using the CR method is shorter than that corresponding interval using sample autocorrelation. Although the case of $\nu = 0$ (empirical likelihood statistic) shows

the best performance when $b = 0.1$ and $b = 0.5$, the case of $\nu = 1$ (Pearson's χ^2 -statistic) is best when $b = 0.9$. Therefore it is also worth considering the general CR approach not only the empirical likelihood method.

Table 1 here

Moreover, we investigate the case $b = 0.9$ with small sample sizes ($T = 50, 100$) for various ν ($\nu = -2, 0, 1, 2, 3, 4$) and SAC. Table 2 shows the resulting 90% confidence intervals of $\rho(2)$ for the model $X(t) = 0.9X(t-1) + U(t)$ where the innovations $U(t)$ have t -distributions with 5 degrees of freedom. The true value of $\rho(2)$ is 0.81. The upper line in each cell shows the 90% confidence interval and the lower line shows the length of the interval. The cases $(\nu, T) = (2, 100), (4, 50), (4, 100)$ give better results than using sample autocorrelation, but we cannot consistently say that the CR method is better than the sample autocorrelation method.

Table 2 here

We then investigate the estimation problem of the AR coefficient. Consider again the AR(1) model (12) whose innovations $U(t)$ are independent and have t -distribution with 5 degrees of freedom. Suppose that we want to estimate the AR(1) coefficient b . If we set $f_\theta(\omega) = (2\pi)^{-1}|1 - \theta e^{i\omega}|^{-2}$, which is the spectral density function of the model (12), then the θ_0 which satisfies condition (4) becomes the coefficient b . Therefore we can construct confidence intervals of the coefficient b with the CR method.

The Yule-Walker estimator $\hat{\theta}_{YW}$ is another existing estimator of the AR coefficient. It is a quasi-MLE and it is known that the quantity $\sqrt{T}(\hat{\theta}_{YW} - \theta_0)$ is asymptotically normally distributed with mean zero and variance

$$4\pi \left\{ \int_{-\pi}^{\pi} \frac{1}{f_\theta(\omega)^2} \left(\frac{\partial f_\theta(\omega)}{\partial \theta} \right)^2 d\omega \right\}^{-1}. \quad (14)$$

From this we can derive confidence intervals of $\theta_0 = b$ by the Yule-Walker estimator.

We compare the length of confidence intervals obtained using the CR method with those obtained from the Yule-Walker estimator. Table 3 shows 90% confidence intervals when $b = 0.4$ using both the CR method ($\nu = -1, -1/2, 0, 1, 2, 3$) and the Yule-Walker method (Y-W), with sample size $T = 200$. In all cases, the CR method is superior to the method by the Yule-Walker estimator. We also compared the 90% confidence intervals when $b = 0.1$, $T = 200$ and $b = 0.1$, $T = 50$ (results not displayed). Similarly, the CR method turned out as superior to the Yule-Walker estimator.

Table 3 here

From the results of various simulations on confidence intervals, we can say that the CR method is better than the existing methods when the process is away from the unit root process.

Next, as an example of theorem 2, we discuss the power property of the test

$$H : \rho(\delta) = \theta_0 \quad \text{versus} \quad A : \rho(\delta) \neq \theta_0.$$

We evaluate the local power under the sequence of local alternatives $A_T : \rho(\delta) = \theta_0 + T^{-1/2}h$, $h \in \mathbf{R}$. From theorem 2, we can see that the mean difference $|\boldsymbol{\mu}|$ indicates the

magnitude of the power. When we consider the AR(1) model (12), the magnitude $|\boldsymbol{\mu}|$ is expressed as

$$|\boldsymbol{\mu}| = (2\pi)^{-1/2} |M_p h| K(b, \delta),$$

where $M_p := \int_{-\infty}^{\infty} u Dp(u) du$ and $K(b, \delta)$ is a positive function of b and δ . Therefore we can see that the larger $|h|$, $|M_p|$ and $K(b, \delta)$, the larger is the power.

If the innovation $U(t)$ has the standard normal distribution, we can easily check that $|M_p| = 1$. To see the effect of non-Gaussianity, we consider the generalized exponential distribution $GE(\eta)$, whose density is expressed as

$$p(u) = c \exp\{-|u|^\eta / 2\zeta\},$$

where $\eta > 0$, $\zeta = 2^{-1/\eta} \Gamma(1/\eta)^{1/2} \Gamma(3/\eta)^{-1/2}$ and $c = \eta \zeta^{-1} 2^{-(1+\eta)/\eta} \Gamma(1/\eta)^{-1}$. $GE(2)$ coincides with the standard normal distribution and $GE(\eta)$, $\eta < 2$, is a more heavy-tailed distribution than normal. Therefore, we study the behavior of $|M_p|$ when $\eta < 2$ to check the effect of non-Gaussianity. Except for the region close to 0, the magnitude of $|M_p|$ is approximately 1 so we can see that the effect of non-Gaussianity is very small.

Figure 1 here

Finally, we consider the magnitude of $K(b, \delta)$. Figure 2 shows the relation between $K(b, \delta)$ and b with $\delta = 2, 3$ and 4, respectively. In every case, the magnitude of $K(b, \delta)$ increases when the value of b tends to 1. Therefore, the test based on the CR method works well for the near unit root process.

Figure 2 here

6. Conclusion

This paper introduces a CR method for time series models, especially vector-valued non-Gaussian stationary processes. The asymptotic distribution of the CR statistic under the null hypothesis is derived. It is shown that we can construct confidence intervals for quantities of interest. As examples, we consider estimation of the autocorrelation and AR(1) coefficient. From the simulation results, we can see that the CR method works better than existing methods such as sample autocorrelation and the Yule-Walker method in the sense of length of confidence intervals. Especially, when the process is distant from the unit root or the sample size is very small, the CR method shows good performance. We also give the asymptotic distribution of the CR statistic under sequences of local alternatives, and we carry out numerical studies about the power property of the test of the autocorrelation. We conclude that the power is hardly affected by the non-Gaussianity of the innovation process, and is larger near the unit root process. It should be noted that if we choose the score function $f_{\boldsymbol{\theta}}$ appropriately, the quantity $\boldsymbol{\theta}_0$ can express various important indices in time series, and hence our approach can be applied to many problems, such as prediction problems, interpolation problems and smoothing problems.

Acknowledgements

The authors are grateful to the editor, an associate editor and two referees for their constructive comments which improved the original manuscript greatly. We would also like to

thank Professor Yu Hayakawa and Research associate James P. Boyd of Waseda University for helpful comments on our English.

References

- [1] Baggerly, K. A. (1998). Empirical likelihood as a goodness-of-fit measure. *Biometrika* **85**, 535-547.
- [2] Brockwell, P. J. & Davis, R. A. (1991). *Time series: Theory and methods*. Springer-Verlag, New York.
- [3] Davis, R. A. & Mikosch, T. (1999). The maximum of the periodogram of a non-Gaussian sequence. *Ann. Probab.* **27**, 522-536.
- [4] Hjort, N. L., McKeague, I.W. & Van Keilegom, I. (2008). Extending the scope of empirical likelihood. To appear in *Ann. Statist.*
- [5] Hosoya, Y. & Taniguchi, M. (1982). A central limit theorem for stationary processes and the parameter estimation of linear processes. *Ann. Statist.* **10**, 132-153. Correction: (1993) **21**, 1115-1117.
- [6] Keenan, D. M. (1987). Limiting behavior of functionals of higher-order sample cumulant spectra. *Ann. Statist.* **15**, 134-151.
- [7] Magnus, J. R. & Neudecker, H. (1999). *Matrix differential calculus with applications in statistics and econometrics*. Wiley, Chichester.
- [8] Monti, A. C. (1997). Empirical likelihood confidence regions in time series models. *Biometrika* **84**, 395-405.
- [9] Ogata, H. & Taniguchi, M. Empirical likelihood approach for non-Gaussian vector stationary processes and its application to minimum contrast estimation. WUTS23, 2006 [Online]. Available: <http://www.math.waseda.ac.jp/~taniguchi/wuts.html>
- [10] Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75**, 237-249.
- [11] Owen, A. B. (1990). Empirical likelihood ratio confidence regions. *Ann. Statist.* **18**, 90-120.
- [12] Priestley, M. B. (1981). *Spectral analysis and time series*. Academic Press.
- [13] Taniguchi, M. (1982). On estimation of the integrals of the fourth order cumulant spectral density. *Biometrika* **69**, 117-122.
- [14] Taniguchi, M. (1987). Minimum contrast estimation for spectral densities of stationary processes. *J. Roy. Statist. Soc. Ser. B* **49**, 315-325.
- [15] Taniguchi, M. & Kakizawa, Y. (2000). *Asymptotic theory of statistical inference for time series*. Springer-Verlag, New York.
- [16] Zhu, L. & Xue, L. (2006). Empirical likelihood confidence regions in a partially linear single-index model. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **68**, 549-570.

Table 1: 90 % confidence intervals (and their length) for the autocorrelation $\theta_0 = \rho(2)$, by use of the CR method (varying ν) and the sample autocorrelation (SAC) method. The model is $X(t) = bX(t-1) + U(t)$, where $U(t)$ is t-distributed with 5 degrees of freedom. Sample size is 200.

	$b = 0.1, \theta_0 = 0.01$	$b = 0.5, \theta_0 = 0.25$	$b = 0.9, \theta_0 = 0.81$
$\nu = -2$	(-0.075, 0.144) 0.219	(0.135, 0.389) 0.254	(0.689, 0.862) 0.173
$\nu = -1$	(-0.074, 0.144) 0.218	(0.143, 0.392) 0.249	(0.710, 0.867) 0.157
$\nu = -1/2$	(-0.074, 0.144) 0.217	(0.146, 0.394) 0.248	(0.715, 0.867) 0.152
$\nu = 0$	(-0.073, 0.144) 0.217	(0.149, 0.396) 0.247	(0.721, 0.869) 0.148
$\nu = 1$	(-0.074, 0.146) 0.220	(0.152, 0.402) 0.250	(0.729, 0.875) 0.146
$\nu = 2$	(-0.076, 0.149) 0.225	(0.153, 0.411) 0.257	(0.735, 0.881) 0.147
SAC	(-0.083, 0.155) 0.237	(0.122, 0.427) 0.305	(0.733, 0.884) 0.152

Table 2: 90 % confidence intervals (and their length) for the autocorrelation $\rho(2) = 0.81$, by use of the CR method (varying ν) and the sample autocorrelation (SAC) method. The model is $X(t) = 0.9X(t-1) + U(t)$, where $U(t)$ is t-distributed with 5 degrees of freedom. The case of small sample sizes: $T = 50, 100$

	$T = 50$	$T = 100$
$\nu = -2$	(-4.437, 0.831) 5.268	(-0.092, 0.843) 0.935
$\nu = 0$	(0.474, 0.845) 0.371	(0.613, 0.853) 0.240
$\nu = 1$	(0.522, 0.855) 0.333	(0.631, 0.861) 0.229
$\nu = 2$	(0.545, 0.864) 0.319	(0.692, 0.822) 0.129
$\nu = 3$	(0.558, 0.872) 0.315	(0.648, 0.876) 0.228
$\nu = 4$	(0.604, 0.844) 0.240	(0.678, 0.853) 0.175
SAC	(0.575, 0.878) 0.303	(0.665, 0.879) 0.214

Table 3: 90% confidence intervals (and their length) for the AR(1) coefficient b , by use of the CR method (varying ν) and the Yule-Walker (Y-W) method. The model is $X(t) = bX(t-1) + U(t)$, where $U(t)$ is t-distributed with 5 degrees of freedom.

	$b = 0.4, T = 200$	$b = 0.1, T = 200$	$b = 0.1, T = 50$
$\nu = -1$	(0.161, 0.553) 0.393	(0.00724, 0.223) 0.216	(-0.493, 0.472) 0.965
$\nu = -1/2$	(0.295, 0.496) 0.202	(0.00777, 0.224) 0.216	(-0.147, 0.278) 0.425
$\nu = 0$	(0.296, 0.498) 0.202	(0.00800, 0.224) 0.216	(-0.147, 0.279) 0.426
$\nu = 1$	(0.298, 0.503) 0.205	(0.00753, 0.227) 0.219	(-0.152, 0.286) 0.438
$\nu = 2$	(0.298, 0.509) 0.212	(0.00551, 0.231) 0.225	(-0.0521, 0.188) 0.240
$\nu = 3$	(0.296, 0.518) 0.222	(0.00151, 0.237) 0.236	(-0.175, 0.309) 0.484
Y-W	(0.093, 0.709) 0.616	(-0.245, 0.479) 0.724	(-0.635, 0.795) 1.45

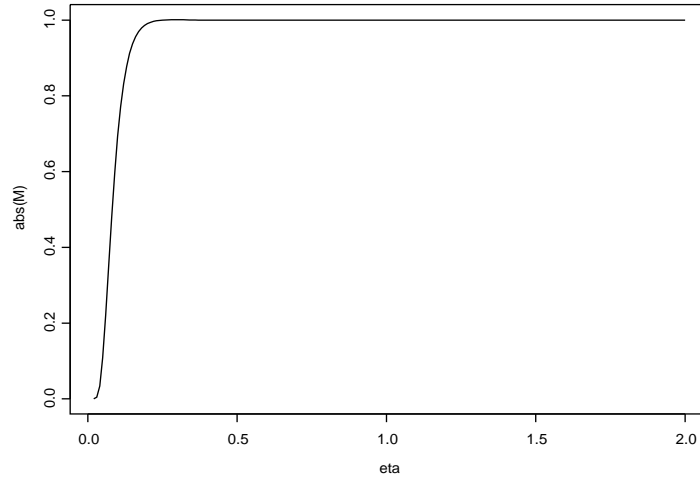


Figure 1: The relation between $|M_p|$ and η is shown. The model is $X(t) = bX(t-1) + U(t)$, where $U(t)$ has a generalized exponential distribution with density $p(u) = c \exp\{-|u|^\eta/2\zeta\}$.

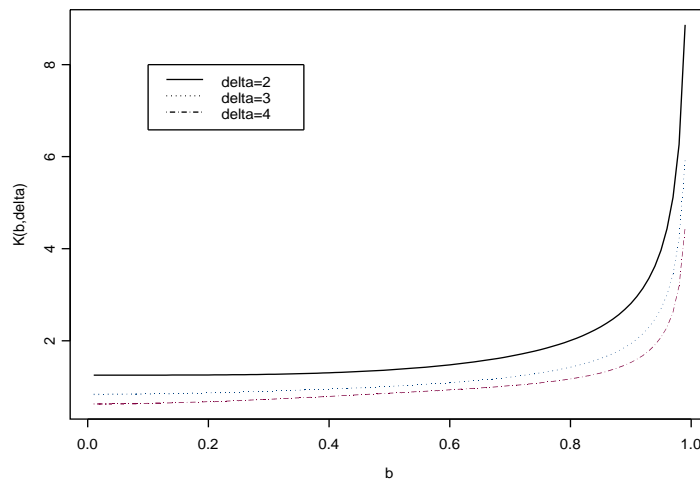


Figure 2: The relation between $K(b, \delta)$ and b is shown for the cases $\delta = 2, 3$ and 4 .