数理解析研究所講究録
第1621巻 2009年 88-103

Application of empirical likelihood method to time series model

早稲田大学・国際教養学部　小方　浩明 (Hiroaki Ogata)
School of International Liberal Studies,
Waseda University

1 Introduction

Empirical likelihood method is one of the nonparametric methods for statistical inference proposed by Owen (1988, 1990). It is shown that empirical likelihood ratio is asymptotically chi-square distributed and used for constructing confidence regions for the sample mean, for a class of M-estimates that includes quantile, and for differentiable statistical functionals. Empirical likelihood has been studied extensively in the literature because of its generality and effectiveness. We can name many applications, such as general estimating equations (Qin and Lawless (1994)), regression models (Owen (1991), Chen (1993, 1994)), biased sample models (Qin (1993)), etc. Although empirical likelihood method has been studied by many authors, it seems to have been investigated mainly under i.i.d. setting. For dependent observations, Kitamura (1997) developed blockwise empirical likelihood for estimating equations and for smooth functions of means. Monti (1997) applied the empirical likelihood method to dependent observations, essentially under circular Gaussian assumption, using a spectral method.

In this resume, we introduce some parts of our previous works on the extension of the empirical likelihood method to non-Gaussian stationary processes by use of spectral approach. In Section 2, we deal with non-Gaussian scalar stationary processes. Motivated by the Whittle likelihood, we introduce estimating functions for dependent observations and derive the asymptotic distribution of the empirical likelihood ratio. In Section 3, we extend the setting to non-Gaussian vector stationary processes. The method of fitting parametric model is also considered and by choosing this parametric function properly, we can consider the estimation problem of the autocorrelation, which is one of the most important indices for time series analysis. In Section 4, we study an application of the method with Cressie-Read power-divergence statistic (CR statistic) to non-Gaussian vector stationary processes. CR method is more general than empirical likelihood method. In this setting, we consider the problem of testing, too. Various numerical studies are also given in Section 5 and illuminate interesting features.

2 Empirical likelihood method for non-Gaussian scalar stationary processes

We consider a scalar-valued linear process $\{X(t); t \in \mathbf{R}\}$, generated as

$$X(t) = \sum_{j=0}^{\infty} G(j)\, e(t-j), \qquad t \in \mathbf{R}, \tag{1}$$

where $\{e(t)\}$ is a sequence of random variables satisfying $E\{e(t)\} = 0$ and $E\{e(t)e(s)\} = \delta(t, s)\sigma^2$, with $\sigma^2 > 0$, $G(j)$'s are constants, and the X, e and G are all real. If $\sum_{j=0}^{\infty} G(j)^2 < \infty$ (this condition is assumed throughout in this section), the process $\{X(t)\}$ is a second-order stationary process, and has the spectral density function

$$g(\omega) = \frac{\sigma^2}{2\pi}\Big|\sum_{j=0}^{\infty} G(j)\mathrm{e}^{-i\omega j}\Big|^2, \qquad -\pi < \omega \le \pi. \tag{2}$$

For the stretch $X(t)$, $t = 1, \ldots, T$, we denote by $I_T(\omega)$, the periodogram; namely

$$I_T(\omega) = \frac{1}{2\pi T}|d_T(\omega)|^2, \qquad \text{where} \qquad d_T(\omega) = \sum_{t=1}^{T} X(t)\exp\{-i\omega t\} \quad -\pi < \omega \le \pi.$$

We set down the following assumptions.

Assumption 2.1. *(i) $\{X(t)\}$ is strictly stationary with all of whose moments exist.*

(ii) The joint k-th order cumulant $c_{X^k}(u_1, \ldots, u_{k-1})$ of $X(t), X(t+u_1), \ldots, X(t+u_{k-1})$ satisfies

$$\sum_{u_1,\ldots,u_{k-1}=-\infty}^{\infty} (1 + |u_j|)\, |c_{X^k}(u_1, \ldots, u_{k-1})| < \infty$$

for $j = 1, \ldots, k-1$ and any k, $k = 2, 3, \ldots$.

Assumption 2.2. *For the sequence $\{C_k\}$ defined by*

$$C_k = \sum_{u_1,\ldots,u_k=-\infty}^{\infty} |c_{X^k}(u_1, \ldots, u_{k-1})|,$$

it holds that

$$\sum_{k=1}^{\infty} C_k z^k/k! < \infty$$

for z in a neighborhood of 0.

Denote by $g_k(\omega_1, \ldots, \omega_{k-1})$, the k-th order spectral density of the process $\{X(t)\}$; namely

$$g_k(\omega_1, \ldots, \omega_{k-1}) = (2\pi)^{-k+1} \sum_{u_1, \ldots, u_k = -\infty}^{\infty} c_{X^k}(u_1, \ldots, u_{k-1}) \exp\Big(-i \sum_{j=1}^{k-1} u_j \omega_j\Big).$$

Henceforth we assume that spectral density depends on an unknown parameter θ in this section: thus $g(\omega) = g(\omega, \theta)$, $g_k(\omega_1, \ldots, \omega_{k-1}) = g_k(\omega_1, \ldots, \omega_{k-1}; \theta)$. In what follows, we state the fundamental results on periodogram.

Lemma 2.1. *Let $\{X(t)\}$ satisfy Assumption 2.1. Let $\boldsymbol{A}(\omega), -\pi < \omega \leq \pi$ be a q-dimensional vector valued continuous function, satisfying $\boldsymbol{A}(\omega) = \boldsymbol{A}(-\omega)$. Then*

$$T^{-1/2} \sum_{t=1}^{T} \boldsymbol{A}(\lambda_t)\{I_T(\lambda_t) - EI_T(\lambda_t)\} \xrightarrow{d} N(0, \Sigma_1) \qquad (T \to \infty),$$

where $\lambda_t = 2\pi t/T$ and

$$\begin{aligned} \Sigma_1 &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \boldsymbol{A}(\omega_1)\boldsymbol{A}(\omega_2)' g_4(\omega_1, -\omega_1, \omega_2)\, d\omega_1 d\omega_2 \\ &\quad + \frac{1}{\pi} \int_{-\pi}^{\pi} \boldsymbol{A}(\omega)\boldsymbol{A}(\omega)' g(\omega)^2\, d\omega. \end{aligned}$$

Lemma 2.2. *Under the same assumption as in Lemma 2.1, it holds that*

$$T^{-1} \sum_{t=1}^{T} \{\boldsymbol{A}(\lambda_t) I_T(\lambda_t)\}\{\boldsymbol{A}(\lambda_t) I_T(\lambda_t)\}' \xrightarrow{p} \Sigma_2 \qquad (T \to \infty),$$

where

$$\Sigma_2 = \frac{1}{\pi} \int_{-\pi}^{\pi} \boldsymbol{A}(\omega)\boldsymbol{A}(\omega)' g(\omega)^2\, d\omega.$$

Empirical likelihood allows us to use likelihood methods, without assuming that the data come from a known family of distribution. Empirical likelihood method is based on the nonparametric likelihood ratio $R(F) = \prod_{i=1}^{n} np_i$ where F is an arbitrary distribution which has probability p_i on the obtained data $\boldsymbol{X}_i$. We use this ratio $R(F)$ as a basis for hypothesis testing and confidence intervals.

When we are interested in parameter $\theta \in \mathbf{R}^q$ which satisfies $E[\boldsymbol{m}(\boldsymbol{X}, \theta)] = 0$, where $\boldsymbol{m}(\boldsymbol{X}, \theta) \in \mathbf{R}^q$ is a vector-valued function, called estimating function, we consider the empirical likelihood ratio function $\mathcal{R}(\theta)$ (defined in (5) below). As a test statistic, it is shown that $-2 \log \mathcal{R}(\theta)$ tends to chi-square with degree of freedom q, when $\boldsymbol{X}_i$'s are i.i.d. (e.g. Owen (1988, 1990)).

Here, we consider the case of dependent sample. When $\{X(t)\}$ is a Gaussian circular ARMA process, Anderson (1977) showed that the log likelihood $LL_c(\theta)$ for $\boldsymbol{X} = (X(1), \ldots, X(T))'$ becomes, disregarding a constant term,

$$LL_c(\theta) = -\sum_{t=1}^{T} \left\{ \log g(\lambda_t, \theta) + \frac{I_T(\lambda_t)}{g(\lambda_t, \theta)} \right\},$$

and that $2I_T(\lambda_t)/g(\lambda_t, \theta), t = 1, \ldots, (T/2) - 1$ or $(T-1)/2$, are independently distributed, each with χ_2^2-distribution, where $I_T(\lambda_t)$ is the periodogram of $\boldsymbol{X}$ and $g(\lambda_t, \theta)$ is the spectral density which depends on an unknown parameter θ. Without the assumption of circular Gaussian ARMA process, it is known that Anderson's results hold asymptotically (e.g. Taniguchi and Kakizawa (2000, Section 7.2.2)). That is, if $\{X(t)\}$ is an appropriate stationary process, $2I_T(\lambda_t)/g(\lambda_t, \theta)$, $t = 1, \ldots, (T/2) - 1$ or $(T-1)/2$ are asymptotically independent and asymptotically χ_2^2-distributed.

Monti (1997) applied the spectral approach of this type to the empirical likelihood, and considered an integral version of $LL_c(\theta)$, which is called the Whittle likelihood, that is,

$$WL(\theta) \equiv \int_{-\pi}^{\pi} \left\{ \log g(\omega, \theta) + \frac{I_T(\omega)}{g(\omega, \theta)} \right\} d\omega, \tag{3}$$

and used $\psi_t(\theta) = (\partial/\partial\theta) \{\log g(\lambda_t, \theta) + I_T(\lambda_t)/g(\lambda_t, \theta)\}$ as a counterpart of Owen's $\boldsymbol{m}(\boldsymbol{X}, \theta)$. Then, Monti (1997) showed that $-2 \log \mathcal{R}(\theta)$ tends to chi-square with degree of freedom q. However, her proof of the above result essentially relies on Anderson's results.

In this section, assuming that $\{X(t)\}$ is a non-Gaussian scalar stationary process, we give the rigorous proof of it. First, we impose the following assumptions.

Assumption 2.3. *$g(\omega, \theta)$ is continuously twice differentiable with respect to θ.*

Assumption 2.4. *(i) θ_0 is the true value of a parameter of interest θ.*

(ii) θ_0 is innovation free, that is,

$$\int_{-\pi}^{\pi} \frac{\partial}{\partial \theta} \{g(\omega, \theta)\}^{-1} g(\omega, \theta)\, d\omega \Big|_{\theta=\theta_0} = 0. \tag{4}$$

If θ is innovation-free, $(\partial/\partial\theta) WL(\theta)|_{\theta=\theta_0} = \boldsymbol{0}$ becomes

$$\int_{-\pi}^{\pi} \frac{\partial}{\partial \theta} \left\{ \frac{I_T(\omega)}{g(\omega, \theta)} \right\} d\omega \Big|_{\theta=\theta_0} = \boldsymbol{0}$$

and its discriterized version of the left hand side is

$$\frac{2\pi}{T} \sum_{t=1}^{T} \frac{\partial}{\partial \theta} \left\{ \frac{I_T(\lambda_t)}{g(\lambda_t, \theta)} \right\} \Big|_{\theta=\theta_0}.$$

Because it is known that $E\{I_T(\lambda_t)\}$ converges to $g(\lambda_t, \theta_0)$, we can see that

$$\frac{1}{T}\sum_{t=1}^{T} E\left[\frac{\partial}{\partial \theta}\left\{\frac{I_T(\lambda_t)}{g(\lambda_t,\theta)}\right\}\Big|_{\theta=\theta_0} \right] \to \mathbf{0}$$

which motivates our empirical likelihood ratio function $\mathcal{R}(\theta)$ defined by

$$\mathcal{R}(\theta) = \max_{w}\left\{\prod_{t=1}^{T} T w_t \mid \sum_{t=1}^{T} w_t \boldsymbol{m}(\lambda_t,\theta) = \mathbf{0},\ w_t \geq 0,\ \sum_{t=1}^{T} w_t = 1\right\}, \tag{5}$$

where $\boldsymbol{w} = (w_1, \ldots, w_T)'$ and

$$\boldsymbol{m}(\lambda_t,\theta) = \frac{\partial}{\partial \theta}\left\{\frac{I_T(\lambda_t)}{g(\lambda_t,\theta)}\right\}.$$

We set down the following further assumption.

Assumption 2.5. *The process* $\{e(t)\}$ *satisfies*

$$cum\{e(t_1), e(t_2), e(t_3), e(t_4)\} = \begin{cases} \kappa^4 & (t_1 = t_2 = t_3 = t_4) \\ 0 & (otherwise) \end{cases}.$$

Then we get the following theorem.

Theorem 2.1. *Let* $\{X(t)\}$ *be a scalar-valued linear process defined in (1), and satisfy Assumptions 2.1 ~ 2.5. Then* $-2\log \mathcal{R}(\theta_0) \xrightarrow{d} \chi_q^2$ *as* $T \to \infty$.

Using this theorem, we can construct a confidence regions of θ. First, we choose a proper threshold value z_α, which is α percentile of χ_q^2. Then we calculate $-2\log \mathcal{R}(\theta)$ at division points over the range and construct the region

$$C_{\alpha,T} = \{\theta \mid -2\log \mathcal{R}(\theta) < z_\alpha\}.$$

3 Empirical likelihood method for non-Gaussian vector stationary processes with fitting parametric spectral model

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

$$\boldsymbol{X}(t) = \sum_{j=0}^{\infty} \boldsymbol{G}(j)\, \boldsymbol{e}(t-j), \qquad t \in \mathbf{Z}, \tag{6}$$

where $\boldsymbol{X}(t)$'s have s components and $\boldsymbol{e}(t)$'s are s dimensional vectors satisfying $E[\boldsymbol{e}(t)] = \mathbf{0}$ and $E[\boldsymbol{e}(t)\boldsymbol{e}(l)'] = \delta(t,l)\boldsymbol{K}$, with $\boldsymbol{K}$ a nonsingular s by s matrix, $\boldsymbol{G}(j)$'s are constant s by

s matrices, and the components of $\boldsymbol{X}, \boldsymbol{e}$ and $\boldsymbol{G}$ are all real. If $\sum_{j=0}^{\infty} \mathrm{tr}\{\boldsymbol{G}(j)\boldsymbol{K}\boldsymbol{G}(j)'\} < \infty$ (this condition is assumed throughout in this section), the process $\{\boldsymbol{X}(t)\}$ is a second-order stationary process and has the spectral density matrix which is expressed as

$$\boldsymbol{g}(\omega) = \frac{1}{2\pi}\boldsymbol{k}(\omega)\boldsymbol{K}\boldsymbol{k}(\omega)^*, \qquad -\pi < \omega \leq \pi, \tag{7}$$

where $\boldsymbol{k}(\omega) = \sum_{j=0}^{\infty} \boldsymbol{G}(j)\mathrm{e}^{i\omega j}$. For the stretch $\boldsymbol{X}(t)$, $t = 1, \ldots, T$, we denote by $\boldsymbol{I}_T(\omega)$, the periodogram; namely

$$\boldsymbol{I}_T(\omega) = \frac{1}{(2\pi T)}\boldsymbol{d}_T(\omega)\boldsymbol{d}_T(\omega)^*, \qquad -\pi < \omega \leq \pi. \tag{8}$$

where $\boldsymbol{d}_T(\omega) = \sum_{t=1}^{T} \boldsymbol{X}(t)\mathrm{e}^{-i\omega t}$. We set down the following assumptions.

Assumption 3.1. *(i) $\{\boldsymbol{X}(t)\}$ is strictly stationary with all of whose moments exist.*

(ii) The joint k-th order cumulant $c_{X^k}(u_1, \ldots, u_{k-1})_{\beta_1\beta_2\ldots\beta_k}$ of $X(t)_{\beta_1}, X(t+u_1)_{\beta_2}, \ldots, X(t+u_{k-1})_{\beta_k}$ satisfies

$$\sum_{u_1,\ldots,u_{k-1}=-\infty}^{\infty} (1 + |u_j|)\,|c_{X^k}(u_1, \ldots, u_{k-1})_{\beta_1\ldots\beta_k}| < \infty \tag{9}$$

for $j = 1, \ldots, k-1, \beta_1, \ldots, \beta_k = 1, \ldots, s$ and any k, $k = 2, 3, \ldots$.

Assumption 3.2. *For the sequence $\{C_k\}$ defined by*

$$C_k = \sup_{\beta_1,\ldots,\beta_k} \sum_{u_1,\ldots,u_k=-\infty}^{\infty} |c_{X^k}(u_1, \ldots, u_{k-1})_{\beta_1\ldots\beta_k}|,$$

it holds that

$$\sum_{k=1}^{\infty} C_k z^k/k! < \infty$$

for z in a neighborhood of 0.

We denote by $g_k(\omega_1, \ldots, \omega_{k-1})_{\beta_1\ldots\beta_k}$, the k-th order spectral density of the process $\{\boldsymbol{X}(t)\}$; namely

$$g_k(\omega_1, \ldots, \omega_{k-1})_{\beta_1\ldots\beta_k} = (2\pi)^{-k+1} \sum_{u_1,\ldots,u_k=-\infty}^{\infty} c_{X^k}(u_1, \ldots, u_{k-1})_{\beta_1\ldots\beta_k} \exp\Big(-i\sum_{j=1}^{k-1} u_j\omega_j\Big).$$

In Section 2, we extended the empirical likelihood method to non-Gaussian scalar stationary processes based on the Whittle likelihood. In this section, we consider to

apply the method to non-Gaussian vector stationary processes. The difference from Section 2 is not only that we deal with vector processes but also that we use a fitting parametric spectral model instead of parametrized true spectral density.

For the vector-valued non-Gaussian linear process (6) with the true spectral density matrix $g(\omega)$, we fit a parametric spectral model $f(\omega, \theta)$ with $\theta \in \Theta \subset \mathbf{R}^q$, to $g(\omega)$. Here $f(\omega, \theta)$ may be different from $g(\omega)$. Consider the multivariate Whittle likelihood

$$\int_{-\pi}^{\pi} \left[\log \det f(\omega, \theta) + \mathrm{tr}\{f(\omega, \theta)^{-1} I_T(\omega)\}\right] d\omega.$$

Here, we impose the following assumption on the parametric spectral model $f(\omega, \theta)$.

Assumption 3.3. *(i)* Θ *is a compact subset of* $\boldsymbol{R}^q$.

(ii) $f(\omega, \theta)$ *is continuously twice differentiable with respect to* $\theta \in \Theta$.

(iii) $f(\omega, \theta) \in \mathcal{F}$. *Here* $\mathcal{F}$ *is the parametric spectral family whose element is expressed as*

$$f(\omega, \theta) = \left(\sum_{j=0}^{\infty} C_j(\theta) e^{ij\omega}\right) \Sigma \left(\sum_{j=0}^{\infty} C_j(\theta) e^{ij\omega}\right)^* \tag{10}$$

where $C_j(\theta)$ *is* $s \times s$ *matrices,* $C_0(\theta)$ *is* $s \times s$ *unit matrix and* Σ *is an* $s \times s$ *positive definite matrix which is independent of* θ.

The above model (10) is the spectral form of the general linear process so this assumption is quite natural. Note that the parameter θ does not depend on Σ, which corresponds to the covariance matrix of the innovation. Like this, when θ depends on only the coefficient parts C_j and does not depend on the innovation part Σ, we call θ "innovation-free". Let θ_0 be the value defined by

$$\frac{\partial}{\partial \theta} \int_{-\pi}^{\pi} \left[\log \det f(\omega, \theta) + \mathrm{tr}\{f(\omega, \theta)^{-1} g(\omega)\}\right] d\omega \Big|_{\theta=\theta_0} = \mathbf{0}, \tag{11}$$

which is called the pseudo-true vale of θ. We use

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

as a disparity measure between $f(\omega, \theta)$ and $g(\omega)$, so θ_0 means the point minimizing the $D(f_\theta, g)$. If θ is innovation-free, then $\int_{-\pi}^{\pi} \log \det f(\omega, \theta)\, d\omega$ is independent of θ (Brockwell-Davis (1991, p.191). Therefore (11) becomes

$$\frac{\partial}{\partial \theta} \int_{-\pi}^{\pi} \mathrm{tr}\{f(\omega, \theta)^{-1} g(\omega)\}\, d\omega \Big|_{\theta=\theta_0} = \mathbf{0}. \tag{12}$$

Furthermore, by choosing $f(\omega, \theta)$ appropriately, θ_0 can show various important indices of time series model. One of such examples is the autocorrelation, which is introduced in the following.

Example 3.1. *Denote* $\Gamma(\delta) = cov\{X(t), X(t+\delta)\}$ *as the autocovariance matrix of* X *with lag* δ. *Let us consider the linear process defined in (6). If we set*

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

and

$$f(\omega, \theta) = (I - A(\theta)e^{i\delta\omega})^{-1}(I - A(\theta)e^{i\delta\omega})^{-1^*},$$

then condition (12) shows

$$\sum_{j=1}^{s} A(\theta_0)_{\beta_1 j} \int_{-\pi}^{\pi} g(\omega)_{j\beta_2}\, d\omega = \int_{-\pi}^{\pi} e^{i\delta\omega} g(\omega)_{\beta_2\beta_1}\, d\omega \quad (\beta_1, \beta_2 = 1, \ldots, s). \tag{13}$$

It is well known that the autocovariance and the spectral density have the following relation

$$\Gamma(\delta) = \int_{-\pi}^{\pi} e^{i\delta\omega} g(\omega)\, d\omega. \tag{14}$$

From (13) and (14), we obtain

$$A(\theta_0)\Gamma(0) = \Gamma(\delta)' \quad \Leftrightarrow \quad A(\theta_0) = \Gamma(\delta)\Gamma(0)^{-1}.$$

Hence, we can estimate the quantity $\Gamma(\delta)\Gamma(0)^{-1}$, *which is a generalized quantity of the usual autocorrelation* $\rho(\delta) = \Gamma(\delta)/\Gamma(0)$ *in scalar case.*

As a natural extension from the scalar case in Section 2, we define an estimating function $m(\lambda_t, \theta)$ as

$$m(\lambda_t, \theta) = \frac{\partial}{\partial\theta}\mathrm{tr}\{f(\lambda_t, \theta)^{-1} I_T(\lambda_t)\} \quad \text{where} \quad \lambda_t = \frac{2\pi t}{T},\ t = 1, \ldots, T \tag{15}$$

according to (12). In addition, we use the empirical likelihood ratio function $\mathcal{R}(\theta)$ defined in (5). Then we obtain the following theorem.

Theorem 3.1. *Let* $\{X(t)\}$ *be the linear process defined in (6) satisfying Assumptions 3.1 - 3.3. Then*

$$-2\log\mathcal{R}(\theta_0) \xrightarrow{d} (AN)'(AN) \tag{16}$$

as $T \to \infty$, *where* N *has a* q*-dimensional stamdard normal distribution and* $A = \Sigma_4^{-1/2}\Sigma_3^{1/2}$. *Here* Σ_3 *is* q *by* q *matrix whose* (γ_1, γ_2) *element is*

$$(\Sigma_3)_{\gamma_1\gamma_2} = \frac{1}{\pi}\int_{-\pi}^{\pi} tr\left[g(\omega)\left.\frac{\partial f(\omega,\theta)^{-1}}{\partial\theta_{\gamma_1}}\right|_{\theta=\theta_0} g(\omega)\left.\frac{\partial f(\omega,\theta)^{-1}}{\partial\theta_{\gamma_2}}\right|_{\theta=\theta_0}\right] d\omega$$
$$+\frac{1}{2\pi}\sum_{\beta_1,\ldots,\beta_4=1}^{s}\iint_{-\pi}^{\pi}\left.\frac{\partial f(\omega_1;\theta)^{\beta_1\beta_2}}{\partial\theta_{\gamma_1}}\right|_{\theta=\theta_0}\left.\frac{\partial f(\omega_2;\theta)^{\beta_3\beta_4}}{\partial\theta_{\gamma_2}}\right|_{\theta=\theta_0}$$
$$\times g_4(-\omega_1,\omega_2,-\omega_2)_{\beta_1\ldots\beta_4}\; d\omega_1 d\omega_2,$$

and Σ_4 *is* q *by* q *matrix whose* (γ_1, γ_2) *element is*

$$(\Sigma_4)_{\gamma_1\gamma_2} = \frac{1}{2\pi}\int_{-\pi}^{\pi} tr\left[g(\omega)\left.\frac{\partial f(\omega,\theta)^{-1}}{\partial\theta_{\gamma_1}}\right|_{\theta=\theta_0} g(\omega)\left.\frac{\partial f(\omega,\theta)^{-1}}{\partial\theta_{\gamma_2}}\right|_{\theta=\theta_0}\right] d\omega$$
$$+\frac{1}{2\pi}\int_{-\pi}^{\pi} tr\left[g(\omega)\left.\frac{\partial f(\omega,\theta)^{-1}}{\partial\theta_{\gamma_1}}\right|_{\theta=\theta_0}\right] tr\left[g(\omega)\left.\frac{\partial f(\omega,\theta)^{-1}}{\partial\theta_{\gamma_2}}\right|_{\theta=\theta_0}\right] d\omega.$$

Remark 3.1. *Denote the eigenvalues of* $A'A$ *by* $a_1, \ldots, a_q$, *then we can write*

$$(AN)'(AN) = \sum_{\gamma=1}^{q} Z_\gamma \tag{17}$$

where Z_γ *is distributed as Gamma distribution* $\Gamma(2^{-1}, (2a_\gamma)^{-1})$.

Σ_3 and Σ_4 contain the unknown spectral density matrix $g(\omega)$ and the fourth-order spectral density $g_4(-\omega_1, \omega_2, -\omega_2)_{\beta_1\ldots\beta_4}$. In practice, we can make appropriate consistent estimators $\hat{\Sigma}_3$ and $\hat{\Sigma}_4$ of Σ_3 and Σ_4, respectively as follows. We can use nonparametric spectral estimator $\hat{g}_T(\omega)$ (see Brillinger (2001) for example) and substitute it into $g(\omega)$ in Σ_3 and Σ_4, then we get the consistent estimator for the integral of function of $g(\omega)$. It is complicated to give the explicit form of consistent estimator for the general integrals of fourth-order spectral density $g_4(-\omega_1, \omega_2, -\omega_2)_{\beta_1\ldots\beta_4}$ in Σ_3. Basically we substitute the fourth-order weighted periodograms into the fourth-order spectral. The consistent estimators can be found in Keenan (1987 Section 2). Thus we can obtain consistent estimators $\hat{\Sigma}_3$ and $\hat{\Sigma}_4$. Then, from Slutsky's theorem, it follows that

$$(\hat{A}N)'(\hat{A}N) \xrightarrow{d} (AN)'(AN) = \sum_{\gamma=1}^{q} Z_\gamma, \tag{18}$$

where $\hat{A} = \hat{\Sigma}_4^{-1/2}\hat{\Sigma}_3^{1/2}$. Using this theorem, we can construct confidence regions for θ. First, we choose a proper threshold value z_α, which is α percentail of estimated distribution of (17) based on the relation (18). Then we calculate $-2\log\mathcal{R}(\theta)$ at division points over the range and construct the region

$$C_{\alpha,T} = \{\theta \mid -2\log\mathcal{R}(\theta) < z_\alpha\}. \tag{19}$$

Remark 3.2. *In the scalar case, we can easily see* $\Sigma_3 = \Sigma_4$. *Then the asymptotic distribution of* $-2\log\mathcal{R}(\theta_0)$ *becomes* χ_q^2, *which is independent of unknown parameter.*

4 Application of Cressie-Read power-divergence statistics to non-Gaussian vector stationary processes with fitting parametric spectral model

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

$$\boldsymbol{X}(t) = \sum_{j=0}^{\infty} \boldsymbol{G}(j)\,\boldsymbol{U}(t-j), \qquad t \in \mathbf{Z}, \tag{20}$$

where the $\boldsymbol{U}(t)$'s are i.i.d. s-vector random variables with probability density $p(\boldsymbol{u}) > 0$ on $\mathbf{R}^s$ and $\boldsymbol{G}(j)$'s are s by s matrices. The components of $\boldsymbol{X}, \boldsymbol{U}$ and $\boldsymbol{G}$ are all real. We make the following assumptions.

Assumption 4.1.

(i) The coefficient matrices $\boldsymbol{G}(j)$'s satisfy

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

where $\|\boldsymbol{G}(j)\|$ denotes the sum of all the absolute values of the entries of $\boldsymbol{G}(j)$.

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

$$\lim_{\|\boldsymbol{u}\| \to \infty} p(\boldsymbol{u}) = 0, \quad \int \boldsymbol{u} p(\boldsymbol{u}) d\boldsymbol{u} = \boldsymbol{0}, \quad \text{and} \int \boldsymbol{u}\boldsymbol{u}' p(\boldsymbol{u}) d\boldsymbol{u} = \boldsymbol{I}_s,$$

where $\|\boldsymbol{u}\| = \sqrt{\boldsymbol{u}'\boldsymbol{u}}$ and $\boldsymbol{I}_s$ denotes the s by s identity matrix.

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

The spectral density of the process $\{\boldsymbol{X}(t)\}$ and the periodogram are expressed as (7) and (8), respectively. (We set $\boldsymbol{K} = \mathbf{I}_s$ in this section.)

Let $\boldsymbol{\theta} \in \Theta$ be a quantity of interest, and be characterized by an s by s nonnegative definite matrix-valued function $\boldsymbol{f}(\omega, \boldsymbol{\theta})$ as is seen in Section 3. Further we impose Assumption 3.3, and assume that true value $\boldsymbol{\theta}_0$ satisfies (12).

In Section 3, we considered the derivative of an extended Whittle likelihood, i.e.,

$$\boldsymbol{m}(\omega, \boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} \mathrm{tr}\{\boldsymbol{f}(\omega, \boldsymbol{\theta})^{-1} \boldsymbol{I}_T(\omega)\} \in \mathbf{R}^q$$

as an estimating function. Then, for the empirical likelihood ratio $\mathcal{R}(\boldsymbol{\theta})$, we showed that $-2 \log \mathcal{R}(\boldsymbol{\theta}_0)$ converges to a sum of Gamma distributions.

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

$$CR_\nu(\boldsymbol{\theta}) = \min_{\boldsymbol{w}} \left\{ \frac{2}{\nu(\nu+1)} \sum_{t=1}^{T} \{(Tw_t)^{-\nu} - 1\} \,\middle|\, \sum_{t=1}^{T} w_t \boldsymbol{m}(\lambda_t, \boldsymbol{\theta}) = \mathbf{0},\ \sum_{t=1}^{T} w_t = 1,\ w_t \geq 0 \right\}, \tag{21}$$

where $\nu \in (-\infty, \infty)$. CR statistic contains user-specified parameter $\nu \in (-\infty, \infty)$ and encompasses several commonly-used tests, i.e., Neyman-modified χ^2 statistic ($\nu = -2$), the maximum entropy, minimum information or Kullback-Leibler statistic ($\nu = -1$), the Freeman-Tukey statistic ($\nu = -1/2$), the empirical likelihood statistic ($\nu = 0$), and Pearson's χ^2 statistic ($\nu = 1$). Hence, Cressie-Read power-divergence statistic is much broader criterion than the empirical likelihood ratio and its asymptotic theory covers the results of Section 3.

The asymptotic results of the Cressie-Read power-divergence statistic are given as follows.

Theorem 4.1. *For any given $\nu \in (-\infty, \infty)$, as $T \to \infty$,*

$$CR_\nu(\boldsymbol{\theta}_0) \xrightarrow{d} (\boldsymbol{AN})'(\boldsymbol{AN}), \tag{22}$$

where the asymptotic distribution $(\boldsymbol{AN})'(\boldsymbol{AN})$ is same one in Theorem 3.1.

In addition, we consider a power property of the test based on Theorem 4.1. From now on, let the coefficient matrices $\boldsymbol{G}(j)$'s of (20) be parametrized by $\boldsymbol{\theta} \in \Theta$, $\Theta \subset \mathbf{R}^q$. Write

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

We make the following assumptions.

Assumption 4.2.

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

$$|(\partial/\partial\theta_{u_1}) \ldots (\partial/\partial\theta_{u_k}) G_{\boldsymbol{\theta}, l_1 l_2}(j)| = O\{j^{-1+D}(\log j)^k\}, \quad k = 0, 1, 2$$

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

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

$$\boldsymbol{G}_{\boldsymbol{\theta}}^{-1}(z) = \boldsymbol{I}_s + \boldsymbol{B}_{\boldsymbol{\theta}}(1)z + \boldsymbol{B}_{\boldsymbol{\theta}}(2)z^2 + \cdots,$$

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

$$|(\partial/\partial\theta_{u_1})\ldots(\partial/\partial\theta_{u_k})B_{\theta,l_1l_2}(j)| = O\{j^{-1-D}(\log j)^k\}, \quad k = 0, 1, 2$$

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

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

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

Consider the problem of testing

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

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

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

Then we get the following theorem.

Theorem 4.2. *Let* $\boldsymbol{A}$, $\boldsymbol{\Sigma}_3$, $\boldsymbol{\Sigma}_4$ *and* $\boldsymbol{N}$ *be the same matrices and q-dimensional standard normal vector as defined in Theorem 4.1. Under the sequence of local alternatives* A_T, *for any given* $\nu \in (-\infty, \infty)$,

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

where $\boldsymbol{\mu} = 2\boldsymbol{\Sigma}_4^{-1/2}\boldsymbol{\tau}$. *Here* $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_q)'$ *with*

$$\tau_i = \int_{-\pi}^{\pi} tr\left[\boldsymbol{g}(\omega)\left.\frac{\partial \boldsymbol{f}(\omega,\boldsymbol{\theta})^{-1}}{\partial\theta_i}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\boldsymbol{g}^{\kappa X}(\omega)\Big\{\sum_{j=1}^{\infty}\boldsymbol{B}_{\boldsymbol{h}'\partial\boldsymbol{\theta}_0}(j)e^{i\omega j}\Big\}\right] d\omega.$$

where

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

The difference with Theorem 4.1 is that we are considering the asymptotic distribution of the test under a sequence of "contiguous alternatives A_T", and that its normal factorization $\boldsymbol{AN} + \boldsymbol{\mu}$ has mean $\boldsymbol{\mu}$. This difference $\boldsymbol{\mu}$ means the distance from the asymptotic distribution under the null hypothesis, so the magnitude $|\boldsymbol{\mu}|$ shows the magnitude of the power of the test.

5 Numerical simulations

In this section, we introduce the results of numerical simulations for Theorems 4.1 and 4.2. Let us consider the following scalar-valued AR(1) model

$$X(t) = bX(t-1) + U(t) \tag{23}$$

where $|b| < 1$, and $U(t)'s$ are independent and identically distributed, and the distribution of $U(t)$ satisfies (ii) and (iii) of Assumption 4.1.

As an application of Theorem 4.1, we can discuss the estimation of the autocorrelation with lag δ, which is denoted by $\rho(\delta)$. As is seen in Example 3.1, we set $f(\omega, \theta) = |1 - \theta e^{i\delta\omega}|^{-2}$ and calculate $CR_\nu(\theta)$ at division points over $(-1, 1)$. Since the process (23) is scalar, the asymptotic distribution of $CR_\nu(\theta_0)$ is chi-square with degree of freedom 1, χ_1^2 (see Remark 3.2). Then we construct the interval $C_{\alpha,T}(\theta)$ in (19) where z_α is the α percentail of χ_1^2 and get the α percent confidence interval of $\theta_0 = \rho(\delta)$.

Let the innovation $U(t)$ have t-distribution with degree of freedom 5 and generate $X(1), \ldots, X(200)$ from (23), *i.e.* $T = 200$. Then we estimate the autocorrelation with lag $\delta = 2$. In AR(1) model (23), the autocorrelation $\rho(\delta)$ is $b^{|\delta|}$, hence $\theta_0 = b^2$. Table 1 shows that 90% confidence interval of θ_0 by use of the Cressie-Read power-divergence method ($\nu = -2, -1, -1/2, 0, 1, 2$) and the usual sample autocorrelation (SAC) method for $b = 0.1, 0.5$, and 0.9. The upper side in each cell shows the 90% confidence interval and the lower side shows the length of the interval. Except for a few cases, the length of interval by use of the Cressie-Read power-divergence method is shorter than that by use of the sample autocorrelation.

Next, as an application of Theorem 4.2, we discuss the power property of the test

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

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 4.2, we can see that the mean difference $|\mu|$ shows a magnitude of the power. When we consider the AR(1) model (23), the magnitude $|\mu|$ is expressed as

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

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

If the innovation $U(t)$ is distributed as a standard normal we can easily check $|M_p| = 1$. To see the effect of non-Gaussianity we consider the generalized exponential distributions $GE(\eta)$, 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 standard normal distribution and $GE(\eta)$, $\eta < 2$ is heavier-tailed distribution than normal. Therefore we see the behavior of $|M_p|$ when $\eta < 2$ to check the effect of non-Gaussianity. Figure 1 shows the relation of η and $|M_p|$. 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. Finally we see the magnitude of $K(b, \delta)$. Figures 2 shows that the relation of $K(b, \delta)$ and b with $\delta = 2, 3$ and 4. In every case the magnitude of $K(b, \delta)$ becomes larger when the value of b tends to 1. Therefore the test based on Cressie-Read power-divergence method works well for the near unit root process.

References

[1] Anderson, T. W. (1977). Estimation for autoregressive moving average models in the time and frequency domains. *Ann. Statist.* **5**, 842-865.

[2] Baggerly, K. A. (1998). Empirical likelihood as a goodness-of-fit measure. *Biometrika.* **85**, 535-547.

[3] BRILLINGER, D. R. (2001). *Time Series: Data Analysis and Theory, expanded ed.* Holden-Day, San Francisco.

[4] Brockwell P. J. and Davis. R. A (1991). *Time Series: Theory and Methods, second ed.* Springer-Verlag.

[5] Chen, S. X. (1993). On the accuracy of empirical likelihood confidence regions for linear regression model. *Ann. Inst. Statist. Math.* **45**, 621-637.

[6] Chen, S. X. (1994). Empirical likelihood confidence intervals for linear regression coefficients. *J. multivariate Anal.* **49**, 24-40.

[7] Keenan, D. M. (1987). Limiting Behavior of Functionals of Higher-Order Sample Cumulant Spectra. *Ann. Statist.* **15**, 134-151.

[8] Kitamura, Y. (1997). Empirical likelihood methods with weekly dependent processes. *Ann. Statist.* **25**, 2084-2102.

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

[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] Owen, A. B. (1991). Empirical likelihood for linear models. *Ann. Statist.* **19**, 1725-1747.

[13] Qin, J. (1993). Empirical likelihood in biased sample problems. *Ann. Statist.* **21**, 1182-1196.

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

[15] Taniguchi, M. and Kakizawa, Y. (2000). *Asymptotic Theory of Statistical Inference for Time Series*. Springer-Verlag, New York.

表 1: 90 % confidence intervals of autocorrelation $\rho(2)$

	$b = 0.1, \theta_0 = 0.01$	$b = 0.5, \theta_0 = 0.25$	$b = 0.9, \theta_0 = 0.81$
$\nu = -2$	(-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

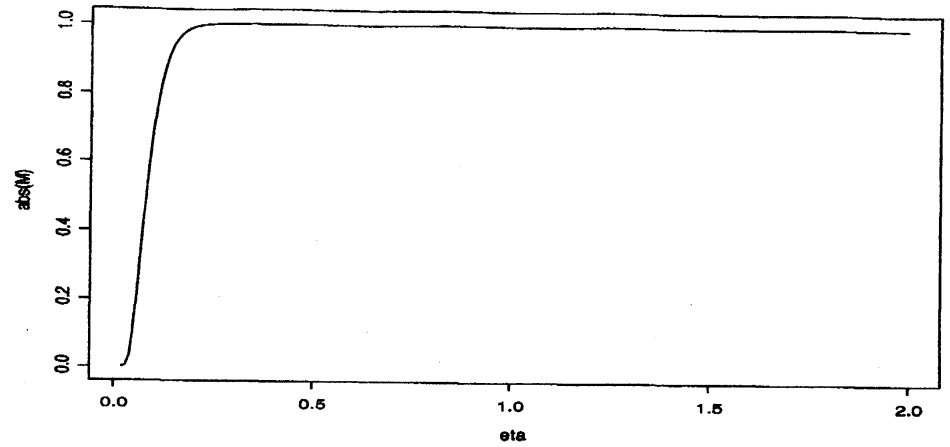

図 1: The relation of $|M_p|$ and η

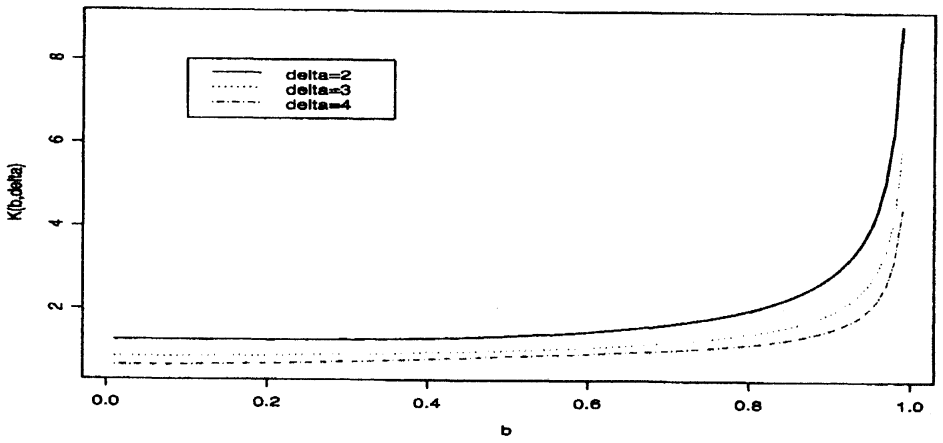

図 2: The relation of $K(b, 1)$ and b