

# Empirical Likelihood Approach for Non-Gaussian Vector Stationary Processes and Its Application to Minimum Contrast Estimation

HIROAKI OGATA

*Waseda University, Tokyo, Japan*  
hiroakiogata@suou.waseda.jp

AND

MASANOBU TANIGUCHI\*

*Waseda University, Tokyo, Japan*  
taniguchi@waseda.jp

**ABSTRACT.** For a class of vector-valued non-Gaussian stationary processes with unknown parameters, we develop the empirical likelihood approach. In time series analysis it is known that Whittle likelihood is one of the most fundamental tools to get a good estimator of unknown parameters, and that the score functions are asymptotically normal. Motivated by the Whittle likelihood, we apply the empirical likelihood approach to its derivative with respect to unknown parameters. We also consider the empirical likelihood approach to a minimum contrast estimation based on a spectral disparity measure, and apply the approach to the derivative of the spectral disparity.

This paper provides rigorous proofs on convergence of our two empirical likelihoods to sums of Gamma distributions. Since the fitted spectral model may be different from the true spectral structure, the results enable us to construct confidence regions for various important time series parameters without Gaussianity. Numerical studies are given, and illuminate some interesting features of the empirical approach.

## 1. Introduction.

Empirical likelihood method is used when the distribution of an appropriate pivotal quantity is unknown. It is shown that empirical likelihood ratio is asymptotically chi-square distributed (e.g. Owen (2001)). However, most of studies on this topic are aimed to independent observations.

For dependent observations, Monti (1997) applied the empirical likelihood approach to the derivative of the Whittle likelihood, and showed that the empirical likelihood ratio is asymptotically chi-square distributed. The results were applied to the problem of testing and construction of confidence region. In this paper we develop a rigorous asymptotic theory for the empirical Whittle likelihood approach when

- (i) the process concerned is a vector-valued non-Gaussian linear process, and

---

*AMS 2000 subject classifications.* Primary 62G30; 62F25; 62M15.

*Key words and phrases.* Empirical likelihood, spectral density matrix, Whittle likelihood, Non-Gaussian vector stationary process, Estimating function, Minimum contrast estimation.

\*corresponding author

(ii) a fitted spectral model may be different from the true one.

Concretely we derive the asymptotic distribution of the empirical likelihood  $\mathcal{R}(\theta)$  based on the derivative of the Whittle likelihood with respect to unknown parameter  $\theta$ . Then it is shown that  $-2 \log \mathcal{R}(\theta)$  converges to a sum of Gamma distribution. In view of (i) and (ii), using the result we can construct confidence regions for various important time series indices without assuming specified spectral structures and Gaussianity of the process.

As an example, we deal with the estimation of autocorrelations. When the fitted spectral model function is chosen appropriately, we can estimate the autocorrelations with the empirical likelihood method. Moreover, when the autocovariance satisfies a certain condition we can show that the estimation with the empirical likelihood method is superior to that with the sample autocorrelations in the sense of asymptotic covariances of both estimators. Especially in the case of a scalar-valued linear process, the asymptotic distribution of the empirical likelihood ratio is chi-square while that of the sample autocorrelation is normal whose variance includes unknown true autocorrelation (e.g. Brockwell and Davis (1990)). This point is also an advantage of empirical likelihood method. In addition, we deal with a prediction problem. We fit AR(1) model to the stock data and make prediction intervals by both methods of the empirical likelihood and the ordinary MLE. The empirical likelihood method gives a narrower prediction interval. Numerical studies for the above results are provided.

We also consider the empirical likelihood approach to a minimum contrast estimation based on a disparity measure between a fitted spectral model and the true spectral density. Then it is shown that  $-2 \log$  (the empirical likelihood for the derivative of the disparity measure) converges to a sum of Gamma distribution. Similarly we can construct confidence regions for unknown parameters by using the result. In this method, choosing the disparity measure appropriately we can give non-iterative efficient estimators of  $\theta$  in explicit forms, whereas the (quasi) maximum likelihood estimators procedure requires iterative methods except for autoregressive models.

This paper is organized as follows. Section 2 describes our setting. In Section 3, we explain the empirical likelihood approach for the Whittle likelihood. The asymptotic distribution of the empirical likelihood ratio is derived. Section 4 deals with the estimation of autocorrelations and comparison between the estimation with empirical likelihood method and that with sample autocorrelations. We also provides some numerical studies on confidence regions. In addition, a prediction problem is dealt. Two prediction intervals obtained by the empirical likelihood method and the MLE method are compared. In Section 5, we consider the empirical likelihood approach for the minimum contrast estimation, and give the asymptotic distribution of the empirical likelihood ratio. The proof of theorem is relegated to Section 6.

As for notations used in this paper, we denote the  $\alpha$ th component of vector  $\mathbf{a}$  by  $a_\alpha$  and denote the  $(\alpha, \beta)$  component of matrix  $\mathbf{A}$  and  $\mathbf{A}^{-1}$  by  $A_{\alpha\beta}$  and  $A^{\alpha\beta}$ , respectively, and we denote the set of all integers by  $\mathbf{Z}$ , and denote Kronecker's delta by  $\delta(t, l)$ .

## 2. Setting.

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

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

where the  $\mathbf{X}(t)$ 's have  $s$  components and the  $\mathbf{e}(t)$ 's are  $s$  dimensional vectors satisfying  $E[\mathbf{e}(t)] = \mathbf{0}$  and  $E[\mathbf{e}(t)\mathbf{e}(l)'] = \delta(t, l)\mathbf{K}$ , with  $\mathbf{K}$  a nonsingular  $s$  by  $s$  matrix;  $\mathbf{G}(j)$ 's are constant  $s$  by  $s$  matrices; and the components of  $\mathbf{X}$ ,  $\mathbf{e}$  and  $\mathbf{G}$  are all real. If  $\sum_{j=0}^{\infty} \text{tr}\{\mathbf{G}(j)\mathbf{K}\mathbf{G}(j)'\} < \infty$  (this condition is assumed throughout), the process  $\{\mathbf{X}(t)\}$  is a second-order stationary process and has the spectral

density matrix which is representable as

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

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

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

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

ASSUMPTION 1. (i)  $\{\mathbf{X}(t)\}$  is strictly stationary with all of whose moments exist.

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

$$\sum_{u_1, \dots, u_{k-1} = -\infty}^{\infty} (1 + |u_j|) |c_{X^k}(u_1, \dots, u_{k-1})_{\beta_1 \dots \beta_k}| < \infty \quad (3)$$

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

ASSUMPTION 2. For the sequence  $\{C_k\}$  defined by

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

it holds that

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

for  $z$  in a neighborhood of 0.

Assumption 1 (ii) means that the own dependence of the stretch  $\mathbf{X}(t)$  becomes weak when time lag becomes large. If there exists some independent pair in the sequence  $X(t)_{\beta_1}, X(t+u_1)_{\beta_2}, \dots, X(t+u_{k-1})_{\beta_k}$ , then  $c_{X^k}(u_1, \dots, u_{k-1})_{\beta_1 \beta_2 \dots \beta_k}$  becomes 0. Therefore (ii) seems a natural assumption for the dependence of  $\{\mathbf{X}(t)\}$ . Assumption 2 is needed to control the maximum of important terms which will appear in the proofs.

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

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

### 3. Empirical likelihood approach.

Empirical likelihood is a nonparametric method of statistical inference. It 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  $R(F) = \prod_{i=1}^n n w_i$  where  $F$  is an arbitrary distribution which has probability  $w_i$  on the data  $X_i$ . We use this  $R(F)$  as a basis for hypothesis testing and confidence intervals.

When we are interested in parameter  $\boldsymbol{\theta} \in \mathbf{R}^q$  which satisfies  $E[\mathbf{m}(X; \boldsymbol{\theta})] = \mathbf{0}$ , where  $\mathbf{m}(X; \boldsymbol{\theta}) \in \mathbf{R}^q$  is a vector-valued function, called estimating function, we consider the empirical likelihood ratio function  $\mathcal{R}(\boldsymbol{\theta})$  (defined in (11) below). As a test statistic, it is shown that  $-2 \log \mathcal{R}(\boldsymbol{\theta})$  converges in distribution to the chi-square distribution with degree of freedom  $q$ , when  $X'_i$ 's are independent and identically distributed, (e.g. Owen (2001)).

For a scalar-valued linear process whose true spectral density is written parametrically as  $g(\omega; \boldsymbol{\theta})$ , Monti (1997) applied the empirical likelihood approach to the Whittle likelihood, that is,

$$\int_{-\pi}^{\pi} \left\{ \log g(\omega; \boldsymbol{\theta}) + \frac{I_T(\omega)}{g(\omega; \boldsymbol{\theta})} \right\} d\omega, \quad (4)$$

and used its discretized derivative  $(\partial/\partial\boldsymbol{\theta}) \{\log g(\lambda_t; \boldsymbol{\theta}) + I_T(\lambda_t)/g(\lambda_t; \boldsymbol{\theta})\}$ , where  $\lambda_t = 2\pi t/T$  (throughout this paper), as a counterpart of Owen's estimating function. If  $\boldsymbol{\theta}_0$  is the true value of  $\boldsymbol{\theta}$ , then she showed that  $-2 \log \mathcal{R}(\boldsymbol{\theta}_0)$  tends to chi-square distribution with degree of freedom  $q$ .

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

$$\int_{-\pi}^{\pi} \left[ \log \det \mathbf{f}(\omega; \boldsymbol{\theta}) + \text{tr}\{\mathbf{f}(\omega; \boldsymbol{\theta})^{-1} \mathbf{I}_T(\omega)\} \right] d\omega.$$

Here, we impose the following assumption on the parametric spectral model  $\mathbf{f}(\omega; \boldsymbol{\theta})$ .

ASSUMPTION 3. (i)  $\Theta$  is a compact subset of  $\mathbf{R}^q$ .

(ii)  $\mathbf{f}(\omega; \boldsymbol{\theta})$  is continuously twice differentiable with respect to  $\boldsymbol{\theta} \in \Theta$ .

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

$$\mathbf{f}(\omega; \boldsymbol{\theta}) = \left( \sum_{j=0}^{\infty} \mathbf{B}_j(\boldsymbol{\theta}) e^{ij\omega} \right) \boldsymbol{\Sigma} \left( \sum_{j=0}^{\infty} \mathbf{B}_j(\boldsymbol{\theta}) e^{ij\omega} \right)^* \quad (5)$$

where  $\mathbf{B}_j(\boldsymbol{\theta})$  is  $s \times s$  matrices,  $\mathbf{B}_0(\boldsymbol{\theta})$  is  $s \times s$  unit matrix and  $\boldsymbol{\Sigma}$  is an  $s \times s$  symmetric matrix.

The above model (5) is the spectral form of the general linear process so this setting is quite natural. Note that the parameter  $\boldsymbol{\theta}$  does not depend on  $\boldsymbol{\Sigma}$ , which corresponds to the covariance matrix of the innovation. Like this, when  $\boldsymbol{\theta}$  depends on only the coefficient parts  $\mathbf{B}_j$  and does not depend on the innovation part  $\boldsymbol{\Sigma}$ , we call  $\boldsymbol{\theta}$  "innovation-free". Let  $\boldsymbol{\theta}_0$  be the value defined by

$$\frac{\partial}{\partial \boldsymbol{\theta}} \int_{-\pi}^{\pi} \left[ \log \det \mathbf{f}(\omega; \boldsymbol{\theta}) + \text{tr}\{\mathbf{f}(\omega; \boldsymbol{\theta})^{-1} \mathbf{g}(\omega)\} \right] d\omega \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \mathbf{0}, \quad (6)$$

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

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

as a disparity measure between  $\mathbf{f}(\omega; \boldsymbol{\theta})$  and  $\mathbf{g}(\omega)$ , so  $\boldsymbol{\theta}_0$  means the point minimizing the  $D(\mathbf{f}_{\boldsymbol{\theta}}, \mathbf{g})$ . If  $\boldsymbol{\theta}$  is innovation-free, then  $\int_{-\pi}^{\pi} \log \det \mathbf{f}(\omega; \boldsymbol{\theta}) d\omega$  is independent of  $\boldsymbol{\theta}$  (Brockwell-Davis (1991, p.191) or Priestley (1981, p.760)). Therefore (6) becomes

$$\frac{\partial}{\partial \boldsymbol{\theta}} \int_{-\pi}^{\pi} \text{tr}\{\mathbf{f}(\omega; \boldsymbol{\theta})^{-1} \mathbf{g}(\omega)\} d\omega \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \mathbf{0}. \quad (7)$$

Furthermore, this setting is unexpectedly useful for many other situations, such as, prediction, interpolation and smoothing problems. In what follows, we give explanations of these three examples in a scalar case.

First, consider the  $\nu$ -step prediction problem as in Hannan (1970 Chapter III Section 2). We predict  $X_t$  using a linear combination of  $X_{t-j}, j \geq \nu$ , that is, we use  $\tilde{X}_t = \sum_{j \geq \nu} a_j(\boldsymbol{\theta}) X_{t-j}$  as a predictor. We measure the error of the predictor by  $E[|X_t - \tilde{X}_t|^2]$  and seek the best linear predictor which minimizes this error. If we specify the spectral model  $\mathbf{g}$ , the functions  $a_j(\boldsymbol{\theta})$ 's are specified (see Grenander-Rosenblatt (1957, p.261)). It is seen that the spectral representations of  $X_t$  and  $\tilde{X}_t$  are

$$X_t = \int_{-\pi}^{\pi} e^{-it\omega} z(d\omega), \quad \tilde{X}_t = \int_{-\pi}^{\pi} e^{-it\omega} \left( \sum_{j \geq \nu} a_j(\boldsymbol{\theta}) e^{ij\omega} \right) z(d\omega)$$

where  $E(|z(d\omega)|^2) = g(\omega) d\omega$ ,  $E(z(d\omega) \overline{z(d\eta)}) = 0$ ,  $\omega \neq \eta$ . Then, the prediction error is expressed as

$$\int_{-\pi}^{\pi} \left| 1 - \sum_{j \geq \nu} a_j(\boldsymbol{\theta}) e^{ij\omega} \right|^2 g(\omega) d\omega.$$

We seek the  $\boldsymbol{\theta}$  which minimizes this error, that is, seek  $\boldsymbol{\theta}_0$  satisfying

$$\left. \frac{\partial}{\partial \boldsymbol{\theta}} \int_{-\pi}^{\pi} \left| 1 - \sum_{j \geq \nu} a_j(\boldsymbol{\theta}) e^{ij\omega} \right|^2 g(\omega) d\omega \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \mathbf{0}. \quad (8)$$

Compare (7) and (8). This is exactly our problem if we set

$$f(\omega; \boldsymbol{\theta}) = \left| 1 - \sum_{j \geq \nu} a_j(\boldsymbol{\theta}) e^{ij\omega} \right|^{-2}.$$

Second, we consider the interpolation problem. Let  $\{X_t\}$  be a stationary process with spectral density  $g$ . Assume that the entire time series has been observed except for the time point  $t = 0$ . Thus  $X_t, t \neq 0$ , have been observed. We would like to estimate  $X_0$  by a linear combination of the observed stochastic variables, that is,  $\tilde{X}_0 = \sum_{j \neq 0} a_j(\boldsymbol{\theta}) X_j$ . The function form of  $a_j(\boldsymbol{\theta})$  is given in Hannan(1970). Similar to the prediction problem, the error of interpolation becomes

$$\int_{-\pi}^{\pi} \left| 1 - \sum_{j \neq 0} a_j(\boldsymbol{\theta}) e^{ij\omega} \right|^2 g(\omega) d\omega.$$

We seek the  $\boldsymbol{\theta}$  minimizing this error, that is, seek  $\boldsymbol{\theta}_0$  satisfying

$$\left. \frac{\partial}{\partial \boldsymbol{\theta}} \int_{-\pi}^{\pi} \left| 1 - \sum_{j \neq 0} a_j(\boldsymbol{\theta}) e^{ij\omega} \right|^2 g(\omega) d\omega \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \mathbf{0}. \quad (9)$$

Compare (7) and (9). This is exactly our problem if we set

$$f(\omega; \boldsymbol{\theta}) = \left| 1 - \sum_{j \neq 0} a_j(\boldsymbol{\theta}) e^{ij\omega} \right|^{-2}.$$

Third, we consider the following smoothing problem. We smooth the trajectory of  $X_t$  by

$$X_t \longrightarrow \sum_{j=-N}^N \theta_j X_{t+j}.$$

Then, similar to the previous problem, the error of this smoothing is expressed as

$$\int_{-\pi}^{\pi} \left| 1 - \sum_{j=-N}^N \theta_j e^{ij\omega} \right|^2 g(\omega) d\omega.$$

We seek the  $\boldsymbol{\theta} = (\theta_{-N}, \dots, \theta_N)'$  minimizing this error, that is, seek  $\boldsymbol{\theta}_0$  satisfying

$$\frac{\partial}{\partial \boldsymbol{\theta}} \int_{-\pi}^{\pi} \left| 1 - \sum_{j=-N}^N \theta_j e^{ij\omega} \right|^2 g(\omega) d\omega \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \mathbf{0}. \quad (10)$$

Compare (7) and (10). This is exactly our problem if we set

$$f(\omega; \boldsymbol{\theta}) = \left| 1 - \sum_{j=-N}^N \theta_j e^{ij\omega} \right|^{-2}.$$

Like these, our approach based on fitting model  $f(\omega; \boldsymbol{\theta})$  which may be different from the true spectral density  $g$ , has a very wide applications and potentials for time series analysis.

From (7), we naturally set

$$\mathbf{m}(\lambda_t; \boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} \text{tr}\{f(\lambda_t; \boldsymbol{\theta})^{-1} \mathbf{I}_T(\lambda_t)\}$$

as an estimating function. We use the following empirical likelihood ratio function  $\mathcal{R}(\boldsymbol{\theta})$  defined by

$$\mathcal{R}(\boldsymbol{\theta}) = \max \left\{ \prod_{t=1}^T T w_t \mid \sum_{t=1}^T w_t \mathbf{m}(\lambda_t; \boldsymbol{\theta}) = \mathbf{0}, w_t \geq 0, \sum_{t=1}^T w_t = 1 \right\}. \quad (11)$$

Then we get the following theorem.

**THEOREM 1.** *Let  $\{X(t)\}$  be the linear process defined in (1) satisfying Assumptions 1 - 3. Then*

$$-2 \log \mathcal{R}(\boldsymbol{\theta}_0) \xrightarrow{d} (\mathbf{AN})'(\mathbf{AN}) \quad (12)$$

as  $T \rightarrow \infty$ , where  $\mathbf{N}$  have a  $q$ -dimensional normal random vector with zero mean vector and covariance matrix  $\mathbf{I}$  (identity matrix) and  $\mathbf{A} = \boldsymbol{\Sigma}_2^{-\frac{1}{2}} \boldsymbol{\Sigma}_1^{\frac{1}{2}}$ . Here  $\boldsymbol{\Sigma}_1$  is  $q$  by  $q$  matrix whose  $(\gamma_1, \gamma_2)$  element is

$$\begin{aligned} (\boldsymbol{\Sigma}_1)_{\gamma_1 \gamma_2} &= \frac{1}{\pi} \int_{-\pi}^{\pi} \text{tr} \left[ \mathbf{g}(\omega) \frac{\partial f(\omega; \boldsymbol{\theta})^{-1}}{\partial \theta_{\gamma_1}} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \mathbf{g}(\omega) \frac{\partial f(\omega; \boldsymbol{\theta})^{-1}}{\partial \theta_{\gamma_2}} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right] d\omega \\ &\quad + \frac{1}{2\pi} \sum_{\beta_1, \dots, \beta_4=1}^s \iint_{-\pi}^{\pi} \frac{\partial f(\omega_1; \boldsymbol{\theta})^{\beta_1 \beta_2}}{\partial \theta_{\gamma_1}} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \frac{\partial f(\omega_2; \boldsymbol{\theta})^{\beta_3 \beta_4}}{\partial \theta_{\gamma_2}} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\ &\quad \times g_4(-\omega_1, \omega_2, -\omega_2)_{\beta_1 \dots \beta_4} d\omega_1 d\omega_2, \end{aligned}$$

and  $\boldsymbol{\Sigma}_2$  is  $q$  by  $q$  matrix whose  $(\gamma_1, \gamma_2)$  element is

$$\begin{aligned} (\boldsymbol{\Sigma}_2)_{\gamma_1 \gamma_2} &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \text{tr} \left[ \mathbf{g}(\omega) \frac{\partial f(\omega; \boldsymbol{\theta})^{-1}}{\partial \theta_{\gamma_1}} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \mathbf{g}(\omega) \frac{\partial f(\omega; \boldsymbol{\theta})^{-1}}{\partial \theta_{\gamma_2}} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right] d\omega \\ &\quad + \frac{1}{2\pi} \int_{-\pi}^{\pi} \text{tr} \left[ \mathbf{g}(\omega) \frac{\partial f(\omega; \boldsymbol{\theta})^{-1}}{\partial \theta_{\gamma_1}} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right] \text{tr} \left[ \mathbf{g}(\omega) \frac{\partial f(\omega; \boldsymbol{\theta})^{-1}}{\partial \theta_{\gamma_2}} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right] d\omega. \end{aligned}$$

REMARK 1. Denote the eigenvalues of  $A'A$  by  $a_1, \dots, a_q$ , then we can write

$$(\mathbf{AN})'(\mathbf{AN}) = \sum_{\gamma=1}^q V_{\gamma} \quad (13)$$

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

$\Sigma_1$  and  $\Sigma_2$  contain the unknown spectral density matrix  $\mathbf{g}(\omega)$  and the fourth-order spectral density  $g_4(-\omega_1, \omega_2, -\omega_2)_{\beta_1 \dots \beta_4}$ . In practice, we can make appropriate consistent estimators  $\hat{\Sigma}_1$  and  $\hat{\Sigma}_2$  of  $\Sigma_1$  and  $\Sigma_2$  respectively as follows. We can use non-parametric spectral estimator  $\hat{\mathbf{g}}_T(\omega)$  (defined in Section 5) and substitute it into  $\mathbf{g}(\omega)$  in  $\Sigma_1$  and  $\Sigma_2$ , then we get the consistent estimator for the integral of function of  $\mathbf{g}(\omega)$ . It is complicated to give the explicit form of consistent estimator for the general integrals of fourth-order spectral density  $g_4(-\omega_1, \omega_2, -\omega_2)_{\beta_1 \dots \beta_4}$  in  $\Sigma_1$ . 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  $\hat{\Sigma}_1$  and  $\hat{\Sigma}_2$ . Then, from Slutsky's theorem it follows that

$$(\hat{\mathbf{A}}\mathbf{N})'(\hat{\mathbf{A}}\mathbf{N}) \xrightarrow{d} (\mathbf{AN})'(\mathbf{AN}) = \sum_{\gamma=1}^q V_{\gamma}, \quad (14)$$

where  $\hat{\mathbf{A}} = \hat{\Sigma}_2^{-\frac{1}{2}} \hat{\Sigma}_1^{\frac{1}{2}}$ . Using this theorem, we can construct confidence regions for  $\theta$ . First, we choose a proper threshold value  $z_{\alpha}$ , which is  $\alpha$  percentail of estimated distribution of (13) based on the relation (14). Then we calculate  $-2 \log \mathcal{R}(\theta)$  at numerous points over the range and construct the region

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

REMARK 2. In the scalar case, we can easily see  $\Sigma_1 = \Sigma_2$ . Then the asymptotic distribution of  $-2 \log \mathcal{R}(\theta_0)$  becomes  $\chi_q^2$ , which is independent of unknown parameter.

#### 4. Numerical simulation

In this section, we provide applications of Section 3.

First we discuss the estimation of autocorrelations and provide a numerical simulation. Denote  $\Gamma(h) = \text{Cov}[\mathbf{X}(t), \mathbf{X}(t+h)]$  as the autocovariance matrix of  $\mathbf{X}$  with lag  $h$ . Let us consider the linear process defined in (1). If we set

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

and

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

then the condition

$$\int_{-\pi}^{\pi} \frac{\partial}{\partial \theta} \text{tr}\{\mathbf{f}(\omega; \theta)^{-1} \mathbf{g}(\omega)\} \Big|_{\theta=\theta_0} d\omega = \mathbf{0} \quad (15)$$

shows

$$\sum_{j=1}^s [\theta_0]_{\beta_1, j} \int_{-\pi}^{\pi} g(\omega)_{j\beta_2} d\omega = \int_{-\pi}^{\pi} e^{i h \omega} g(\omega)_{\beta_2 \beta_1} d\omega \quad (\beta_1, \beta_2 = 1, \dots, s). \quad (16)$$

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

$$\Gamma(h) = \int_{-\pi}^{\pi} e^{i h \omega} \mathbf{g}(\omega) d\omega. \quad (17)$$

From (16) and (17), we obtain

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

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

REMARK 3. *In this example, we do not assume specified spectral structures and Gaussianity for the process  $\mathbf{X}(t)$ .*

The usual estimator for  $\Gamma(h)$  is

$$\hat{\Gamma}(h) = \begin{cases} T^{-1} \sum_{t=1}^{T-h} (\mathbf{X}(t) - \bar{\mathbf{X}}_T)(\mathbf{X}(t+h) - \bar{\mathbf{X}}_T)' & \text{for } 0 \leq h \leq n-1, \\ T^{-1} \sum_{t=-h+1}^T (\mathbf{X}(t) - \bar{\mathbf{X}}_T)(\mathbf{X}(t+h) - \bar{\mathbf{X}}_T)' & \text{for } -n+1 \leq h < 0, \end{cases}$$

where  $\bar{\mathbf{X}}_T = \sum_{t=1}^T \mathbf{X}(t)$ . Therefore we can estimate the quantity  $\Gamma(h)\Gamma(0)^{-1}$  by  $\hat{\Gamma}(h)\hat{\Gamma}(0)^{-1}$ . By Hosoya and Taniguchi (1982 Theorem 2.2) and Slutsky's lemma, we can see that  $\sqrt{T}(\hat{\Gamma}(h)\hat{\Gamma}(0)^{-1} - \Gamma(h)\Gamma(0)^{-1})$  have a joint asymptotic normal distribution whose mean is zero and the asymptotic covariance between  $[\sqrt{T}(\hat{\Gamma}(h)\hat{\Gamma}(0)^{-1} - \Gamma(h)\Gamma(0)^{-1})]_{\beta_1\beta_2}$  and  $[\sqrt{T}(\hat{\Gamma}(h)\hat{\Gamma}(0)^{-1} - \Gamma(h)\Gamma(0)^{-1})]_{\beta_3\beta_4}$  is given as

$$\begin{aligned} & \sum_{\beta'_1, \dots, \beta'_6=1}^s \left[ \frac{\Gamma(0)^{\beta'_3\beta'_2}\Gamma(0)^{\beta'_6\beta'_4}}{s^4} \sigma(h, h)_{\beta_1\beta'_3, \beta_3\beta'_6} \right. \\ & \quad - \frac{\Gamma(0)^{\beta'_3\beta'_2}\Gamma(h)_{\beta_3\beta'_4}\Gamma(0)^{\beta'_4\beta'_5}\Gamma(0)^{\beta'_6\beta'_4}}{s^2} \sigma(h, 0)_{\beta_1\beta'_3, \beta'_5\beta'_6} \\ & \quad - \frac{\Gamma(0)^{\beta'_6\beta'_4}\Gamma(h)_{\beta_1\beta'_1}\Gamma(0)^{\beta'_1\beta'_2}\Gamma(0)^{\beta'_5\beta'_2}}{s^2} \sigma(0, h)_{\beta'_2\beta'_3, \beta_3\beta'_6} \\ & \quad \left. + \Gamma(h)_{\beta_1\beta'_1}\Gamma(h)_{\beta_3\beta'_4}\Gamma(0)^{\beta'_1\beta'_2}\Gamma(0)^{\beta'_4\beta'_5}\Gamma(0)^{\beta'_3\beta'_2}\Gamma(0)^{\beta'_6\beta'_4} \sigma(0, 0)_{\beta'_2\beta'_3, \beta'_5\beta'_6} \right], \end{aligned}$$

where

$$\begin{aligned} & \sigma(h_1, h_2)_{\beta_1\beta_2, \beta_3\beta_4} \\ = & 2\pi \int_{-\pi}^{\pi} \left[ g(\omega)_{\beta_1\beta_3} \overline{g(\omega)_{\beta_2\beta_4}} \exp\{-i(h_2 - h_1)\omega\} + g(\omega)_{\beta_1\beta_4} \overline{g(\omega)_{\beta_2\beta_3}} \exp\{i(h_2 + h_1)\omega\} \right] d\omega \\ & + 2\pi \sum_{\alpha_1, \dots, \alpha_4=1}^p \iint_{-\pi}^{\pi} \exp\{ih_1\omega_1 + ih_2\omega_2\} k(\omega_1)_{\beta_1\alpha_1} k(-\omega_1)_{\beta_2\alpha_2} k(\omega_2)_{\beta_3\alpha_3} k(-\omega_2)_{\beta_4\alpha_4} \\ & \quad \cdot g_4^e(\omega_1, -\omega_2, \omega_2)_{\alpha_1 \dots \alpha_4} d\omega_1 d\omega_2. \end{aligned}$$



and  $g_4^e(\cdot, \cdot, \cdot, \cdot) \dots$  is the fourth order spectral density of the process  $\{e(t)\}$ .

Let us consider the scalar case of (1), that is,  $s = p = 1$ . Denote the estimator  $\hat{\Gamma}(h)/\hat{\Gamma}(0)$  by  $\hat{\rho}(h)$ . Then, the asymptotic variance of  $\sqrt{T}(\hat{\rho}(h) - \rho(h))$  becomes

$$V = \frac{1}{\Gamma(0)^2} \left[ \rho(h)^2 \sigma(0, 0) - \rho(h) \{ \sigma(0, h) + \sigma(h, 0) \} + \sigma(h, h) \right] \quad (18)$$

where

$$\sigma(h_1, h_2) = 2\pi \int_{-\pi}^{\pi} g(\omega)^2 (e^{-i(h_2-h_1)\omega} + e^{i(h_1+h_2)\omega}) d\omega.$$

On the other hand, from (??) in Section 6 we can see the corresponding asymptotic variance of the empirical likelihood method is  $\Sigma_1$ , and easily show that

$$\Sigma_1 = \left( \frac{\Gamma(0)}{\pi} \right)^2 V.$$

Therefore if  $\Gamma(0) < \pi$ , the empirical likelihood method is better than that based on the asymptotics of  $\sqrt{T}(\hat{\rho}(h) - \rho(h))$ . Furthermore, (18) shows that the asymptotic distribution of sample autocorrelation depends on the unknown autocorrelation  $\rho(h)$  while that of empirical likelihood ratio is independent of it in scalar case (see Remark 2). This point can be also an advantage of empirical likelihood method.

Next, we give a numerical simulation. Let us consider the following two dimensional AR(1) model

$$\begin{pmatrix} X_{t,1} \\ X_{t,2} \end{pmatrix} = \begin{pmatrix} 0.3 & 0 \\ 0 & 0.5 \end{pmatrix} \begin{pmatrix} X_{t-1,1} \\ X_{t-1,2} \end{pmatrix} + \begin{pmatrix} e_{t,1} \\ e_{t,2} \end{pmatrix}, \quad (19)$$

where  $(e_{t,1}, e_{t,2})'$ 's are independent and identically distributed as two dimensional  $t$ -distribution with mean zero and covariance matrix  $\mathbf{K}$ . Here we assume that innovation part has correlation, i.e.,  $\mathbf{K}$  is not a diagonal matrix so this setting is not a trivial extension from scalar case even if the coefficient matrix is diagonal. Theoretically we can estimate the  $\Gamma(h)\Gamma(0)^{-1}$  but this matrix is  $2 \times 2$  and has 4 elements so it is difficult to express its confidence region. Therefore, we especially set

$$\mathbf{A}(\boldsymbol{\theta}) = \begin{pmatrix} \theta_1 & 0 \\ 0 & \theta_2 \end{pmatrix}$$

then (15) shows

$$\mathbf{A}(\boldsymbol{\theta}_0) = \begin{pmatrix} \Gamma_{11}(h)/\Gamma_{11}(0) & 0 \\ 0 & \Gamma_{22}(h)/\Gamma_{22}(0) \end{pmatrix}.$$

In this situation the true value is 2 dimensional and we can display the confidence region. Denote  $\boldsymbol{\theta}_0 = (\theta_{0,1}, \theta_{0,2})' = (\Gamma_{11}(h)/\Gamma_{11}(0), \Gamma_{22}(h)/\Gamma_{22}(0))' = (\rho_1(h), \rho_2(h))' = \boldsymbol{\rho}(h)$ . We estimate the autocorrelation with lag  $h = 1$ , that is,  $\boldsymbol{\theta}_0 = \boldsymbol{\rho}(1)$ . In AR(1) model (19), it becomes  $(0.3, 0.5)'$ . Figure 1 shows that 90% confidence region of  $\boldsymbol{\theta}_0$  by use of the empirical likelihood method and sample autocorrelation method with covariance matrix of innovation  $\mathbf{K}_1 = \begin{pmatrix} 5.125 & 3.5 \\ 3.5 & 2.5 \end{pmatrix}$ , sample size  $T = 300$ . To construct the empirical region, we calculate  $-2 \log \mathcal{R}(\boldsymbol{\theta})$  at numerous points over the range and gather the points which satisfy  $|\boldsymbol{\theta}| - 2 \log \mathcal{R}(\boldsymbol{\theta}) < z_{0.90}$ , where  $z_{0.90}$  is the 90 % percentail of the asymptotic distribution which is expressed in (12). Both regions include the true value  $(0.3, 0.5)'$ , but the region using the empirical likelihood method is much narrower than that

using the sample autocorrelation method. Figure 2 shows that the case of covariance matrix of innovation  $\mathbf{K}_2 = \begin{pmatrix} 4.27 & 3.45 \\ 3.45 & 5.18 \end{pmatrix}$ . In this case, both regions are roughly same size but the region using the sample autocorrelation method does not include the true value  $(0.3, 0.5)'$ . Simulation result shows the empirical likelihood method is superior to autocorrelation method.

As the final of this section, we consider a prediction problem. Here is a stock price data of KDDI (Japanese communication company) from June 13, 2005 to January 30, 2006 and denote their log returns by  $X_1, \dots, X_{155}$  ( $X_i$ 's are modified so that their mean is zero). Assume that we have observed  $X_1, \dots, X_{154}$  and fitting AR(1) model, consider to predict  $X_{155}$ . Its predictor  $\hat{X}_{155}$  is expressed as  $\hat{X}_{155} = \hat{b} X_{154}$  where  $\hat{b}$  is a certain estimator of coefficient of AR(1) model. We plug in lower and upper bound of  $\hat{b}$  in both empirical likelihood case and usual MLE case. In Figure 3, the solid line shows  $X_1, \dots, X_{155}$  and the dashed lines show 99% confidence intervals of  $X_{155}$  (from the top, MLE's upper bound, empirical upper bound, empirical lower bound and MLE's lower bound). The concrete values of them are provided in Table 2. We obtain the result that the confidence interval via empirical likelihood approach is narrower than that via MLE's approach.

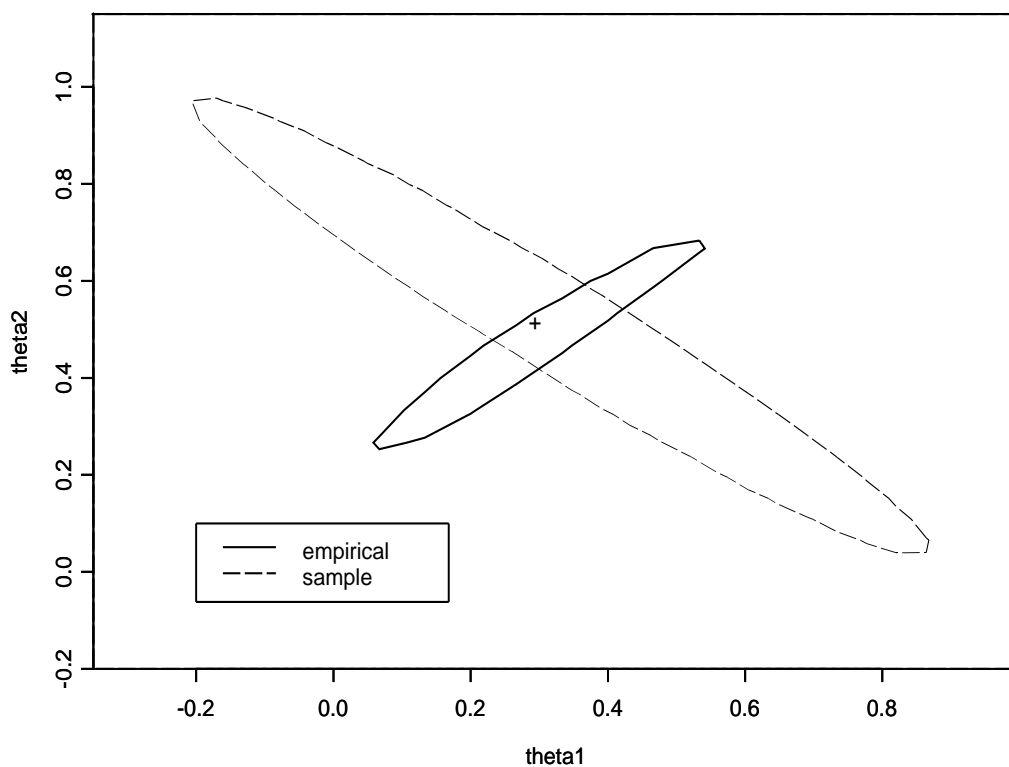


Figure 1: 90% confidence region of  $(\theta_1, \theta_2)$  with  $\mathbf{K}_1$ .

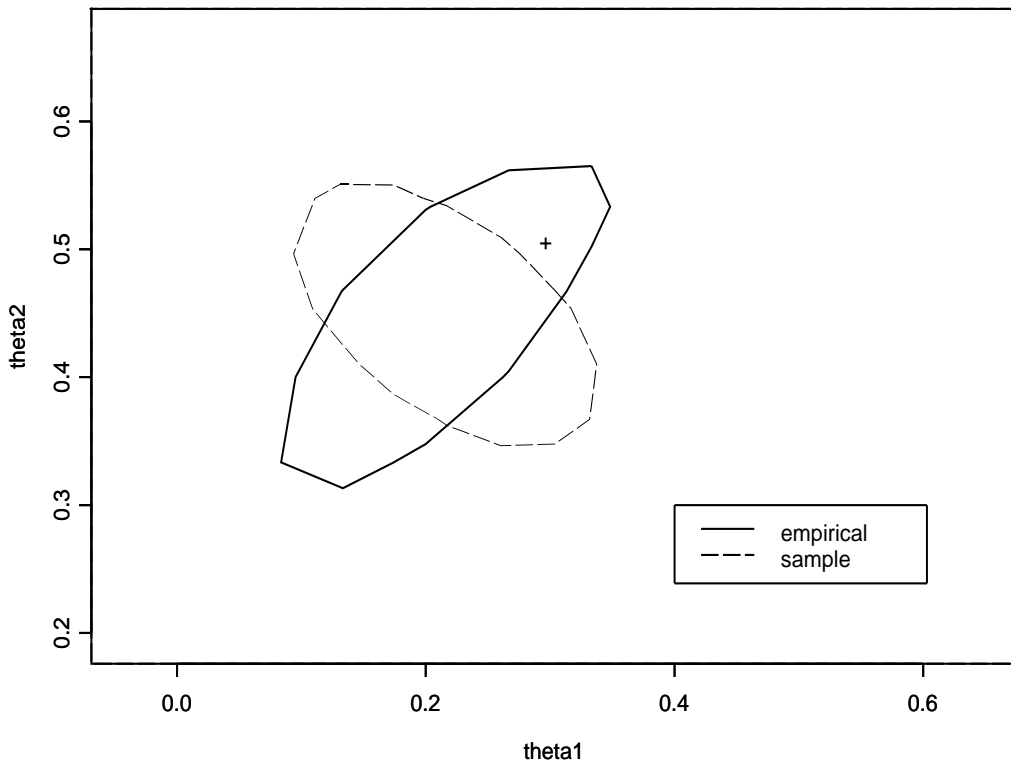


Figure 2: 90% confidence region of  $(\theta_1, \theta_2)$  with  $\mathbf{K}_2$

Table 1: The values of prediction

	lower bound	upper bound	range
MLE	$-2.12 \times 10^{-3}$	$5.97 \times 10^{-3}$	$8.08 \times 10^{-3}$
empirical	$-1.68 \times 10^{-3}$	$5.64 \times 10^{-3}$	$7.32 \times 10^{-3}$

$$X_{155} = 1.75 \times 10^{-4}$$

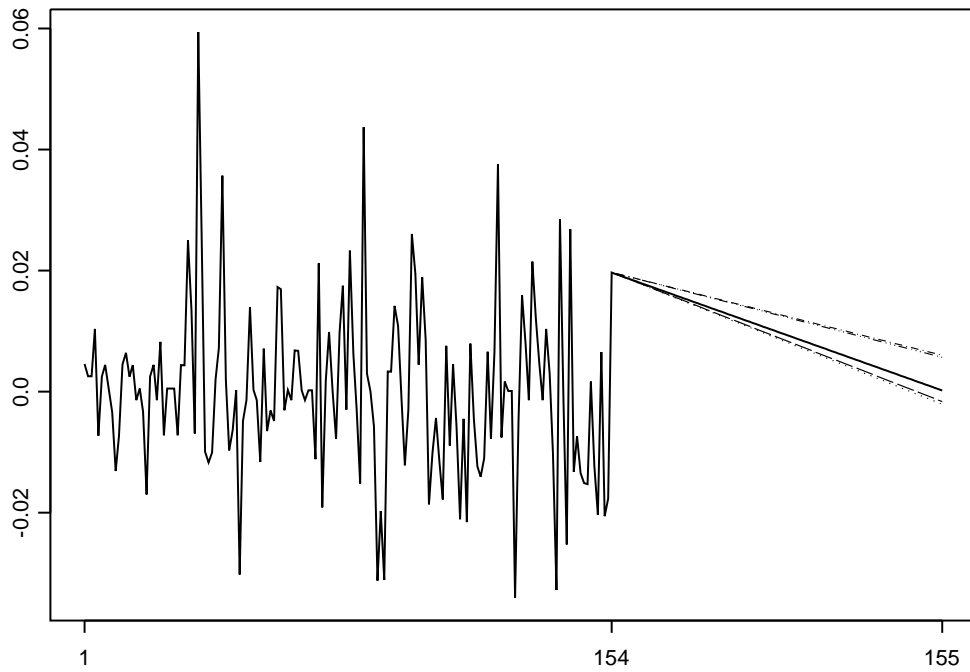


Figure 3: 99% confidence interval of  $X_{155}$

## 5. Minimum contrast estimation

In this section, we apply the empirical likelihood method to minimum contrast estimation. The minimum contrast estimation is based on a disparity measure between the true spectral density  $\mathbf{g}(\omega) \in \mathcal{F}$  ( $\mathcal{F}$  is a space of spectral density matrices) and a certain parametric family  $\mathcal{P} = \{\mathbf{f}(\omega; \boldsymbol{\theta}); \mathbf{f}(\omega; \boldsymbol{\theta}) \in \mathcal{F}, \boldsymbol{\theta} \in \Theta \subset \mathbf{R}^q\}$  of spectral densities. The disparity measure is defined as  $\mathcal{D}(\mathbf{f}_\theta, \mathbf{g}) = \int_{-\pi}^{\pi} K(\mathbf{f}_\theta, \mathbf{g}, \omega) d\omega$ , where  $K(\cdot, \cdot, \cdot)$  is an appropriate holomorphic function which measures a nearness between  $\mathbf{f}(\omega; \boldsymbol{\theta})$  and  $\mathbf{g}(\omega)$ . (see, Taniguchi and Kakizawa, 2000, Section 6.2). Since the true  $\mathbf{g}(\omega)$  is actually unknown, we introduce a non-parametric window type estimator  $\hat{\mathbf{g}}_T(\omega)$  for  $\mathbf{g}(\omega)$ .

We set down the following assumptions.

ASSUMPTION 4.  $W(x)$  is real, bounded, even, non-negative and satisfies

$$\int_{-\infty}^{\infty} W(x) dx = 1.$$

ASSUMPTION 5. For  $M_T = O(T^\alpha)$ , ( $1/4 < \alpha < 1/2$ ), the function  $W_T(\omega) = M_T W(M_T \omega)$  can be expanded as

$$W_T(\omega) = \frac{1}{2\pi} \sum_l w\left(\frac{l}{M_T}\right) e^{-il\omega},$$

where  $w(x)$  is a continuous, even function with  $w(0) = 1$ ,  $|w(x)| \leq 1$  and  $\int_{-\infty}^{\infty} w(x)^2 dx < \infty$ , and satisfies

$$\lim_{|x| \rightarrow 0} \frac{1 - w(x)}{|x|^2} = \kappa_2 < \infty.$$

Henceforth we use the following non-parametric spectral estimator

$$\hat{\mathbf{g}}_T(\omega) = \int_{-\pi}^{\pi} W_T(\omega - \mu) \mathbf{I}_T(\mu) d\mu.$$

REMARK 4. We can also define the discretized non-parametric estimator  $\hat{\mathbf{g}}_T^{dis}(\omega)$  by

$$\hat{\mathbf{g}}_T^{dis}(\omega) = \frac{2\pi}{T} \sum_{s=1}^T W_T(\omega - \lambda_s) \mathbf{I}_T(\lambda_s).$$

Due to Brillinger (2001, Theorem 5.9.1), the two estimator  $\hat{\mathbf{g}}_T(\omega)$  and  $\hat{\mathbf{g}}_T^{dis}(\omega)$  are differ by the magnitude  $O_p(M_T T^{-1}(M_T + \log T))$ .

A functional  $\Pi$  defined on  $\mathcal{F}$  is determined by the requirement that for the parametric family of spectral densities  $\mathcal{P}$ ,

$$\mathcal{D}(\mathbf{f}_{\Pi(\mathbf{g})}, \mathbf{g}) = \min_{\boldsymbol{\theta} \in \Theta} \mathcal{D}(\mathbf{f}_\theta, \mathbf{g}) \quad \text{for every } \mathbf{g} \in \mathcal{F}.$$

Taniguchi and Kakizawa(2000, Theorem 6.2.3.) showed that the functional  $\Pi$  is expanded as

$$\Pi(\hat{\mathbf{g}}_T) = \Pi(\mathbf{g}) - \mathbf{D}_g^{-1} \left( \int_{-\pi}^{\pi} \text{tr}[\boldsymbol{\psi}_j(\omega) (\hat{\mathbf{g}}_T(\omega) - \mathbf{g}(\omega))] d\omega \right)_{j=1, \dots, q}' + O(\|\hat{\mathbf{g}}_T(\omega) - \mathbf{g}(\omega)\|^2)$$

where

$$\begin{aligned} \mathbf{D}_g &= \int_{-\pi}^{\pi} \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} K(\mathbf{f}_{\boldsymbol{\theta}}, \mathbf{g}, \omega) \Big|_{\boldsymbol{\theta}=\Pi(\mathbf{g})} d\omega \quad (q \times q \text{ matrix}), \\ (\boldsymbol{\psi}_j(\omega))_{\beta_1 \beta_2} &= \frac{\partial}{\partial Z_{\beta_1 \beta_2}} \left\{ \frac{\partial}{\partial \theta_j} K(\mathbf{f}_{\boldsymbol{\theta}}, \mathbf{Z}, \omega) \Big|_{\boldsymbol{\theta}=\Pi(\mathbf{g})} \right\} \Big|_{\mathbf{Z}=\mathbf{g}(\omega)} \quad (s \times s \text{ matrix}). \end{aligned}$$

Therefore,  $-\mathbf{D}_g^{-1} \left( \int_{-\pi}^{\pi} \text{tr}[\boldsymbol{\psi}_j(\omega) (\hat{\mathbf{g}}_T(\omega) - \mathbf{g}(\omega))] d\omega \right)'_{j=1, \dots, q}$  can be regarded as the derivative of functional  $\Pi$ , so we set

$$\mathbf{m}(\lambda_t; \boldsymbol{\theta}) = -\mathbf{D}_g^{-1} \left( \text{tr}[\boldsymbol{\psi}_j(\lambda_t) \hat{\mathbf{g}}_T(\lambda_t)] \right)'_{j=1, \dots, q} \quad (20)$$

as an estimating function.

For  $\sum_{t=1}^T E[\mathbf{m}(\lambda_t, \boldsymbol{\theta})] \rightarrow \mathbf{0}$ , we assume the following.

ASSUMPTION 6. For any  $\epsilon > 0$ , we can take  $\boldsymbol{\psi}_j(\omega)$  such that

$$\int_{-\pi}^{\pi} \text{tr}[\boldsymbol{\psi}_j(\omega) \mathbf{g}(\omega)] d\omega = O(\epsilon). \quad (j = 1, \dots, q)$$

Then, we get the following result.

THEOREM 2. Let  $\{\mathbf{X}(t)\}$  be the linear process defined in (1) satisfying Assumptions 1 - 6. Then

$$-2 \log \mathcal{R}(\Pi(\mathbf{g})) \xrightarrow{d} (\mathbf{BN})'(\mathbf{BN})$$

as  $T \rightarrow \infty$ , where  $\mathbf{N}$  have a  $q$ -dimensional normal random vector with zero mean vector and covariance matrix  $\mathbf{I}$  (identity matrix) and  $\mathbf{B} = \boldsymbol{\Sigma}_4^{-\frac{1}{2}} \boldsymbol{\Sigma}_3^{\frac{1}{2}}$  where  $\boldsymbol{\Sigma}_3 = \mathbf{D}_g^{-1} \mathbf{U}_g \mathbf{D}_g^{-1}$  and  $\boldsymbol{\Sigma}_4 = \mathbf{D}_g^{-1} \mathbf{V}_g \mathbf{D}_g^{-1}$ . Here  $\mathbf{U}_g$  is  $q$  by  $q$  matrix whose  $(\gamma_1, \gamma_2)$  element is

$$\begin{aligned} (\mathbf{U}_g)_{\gamma_1 \gamma_2} &= \frac{1}{\pi} \int_{-\pi}^{\pi} \text{tr}[\mathbf{g}(\omega) \boldsymbol{\psi}_{\gamma_1}(\omega) \mathbf{g}(\omega) \boldsymbol{\psi}_{\gamma_2}(\omega)] d\omega \\ &\quad + \frac{1}{2\pi} \sum_{\beta_1, \dots, \beta_4=1}^s \iint_{-\pi}^{\pi} \{\boldsymbol{\psi}_{\gamma_1}(\omega)\}_{\beta_1 \beta_2} \{\boldsymbol{\psi}_{\gamma_2}(\omega)\}_{\beta_3 \beta_4} \\ &\quad \times g_4(-\omega_1, \omega_2, -\omega_2)_{\beta_1 \dots \beta_4} d\omega_1 d\omega_2, \end{aligned}$$

and  $\mathbf{V}_g$  is  $q$  by  $q$  matrix whose  $(\gamma_1, \gamma_2)$  element is

$$(\mathbf{V}_g)_{\gamma_1 \gamma_2} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \text{tr}[\mathbf{g}(\omega) \boldsymbol{\psi}_{\gamma_1}(\omega)] \text{tr}[\mathbf{g}(\omega) \boldsymbol{\psi}_{\gamma_2}(\omega)] d\omega$$

REMARK 5. Assuming that  $q/s^2$  becomes a positive integer, let  $J = q/s^2$  and if we set

$$\begin{aligned} K(\boldsymbol{\theta}, \mathbf{Z}, \omega) &= \left\{ \log \det(\mathbf{Z} \mathbf{f}(\omega; \boldsymbol{\theta})^{-1}) \right\}^2 \\ \mathbf{f}(\omega; \boldsymbol{\theta}) &= \exp \left\{ \sum_{j=1}^J \mathbf{H}(j; \boldsymbol{\theta}) \cos(j\omega) + \mathbf{S}(\omega) \right\} \end{aligned}$$

where  $\exp\{\cdot\}$  is the matrix exponential,  $\mathbf{H}(j; \boldsymbol{\theta})$  is an  $s \times s$  matrices which satisfies

$$\text{vec}\{\mathbf{H}(j; \boldsymbol{\theta})\} = (\theta_{s^2(j-1)+1}, \dots, \theta_{s^2 j})'$$

and  $S(\omega)$  is an  $s \times s$  matrix which is independent of  $\theta$ , then it is not difficult to show

$$\int_{-\pi}^{\pi} \text{tr}[\psi_j(\omega)\mathbf{g}(\omega)] d\omega = \mathbf{0} \quad (j = 1, \dots, q).$$

Hence we can construct a model satisfying Assumption 6.

This method has the following desirable property. For various spectra  $f(\omega; \theta)$ , choosing the function  $K(\cdot, \cdot, \cdot)$  in  $D(f_\theta, \mathbf{g})$  appropriately we can give non-iterative efficient estimators of  $\theta$  in explicit forms, whereas the (quasi) maximum likelihood estimators procedure requires iterative methods except for autoregressive models. Moreover, for the following density

$$\mathbf{g}(\omega) = (1 - \delta)f_\theta(\omega) + \delta h_\eta(\omega)$$

where  $h_\eta$  is a perturbation which is the density with pointmass one at  $\eta$ , a robustness of functional  $\Pi$  is reported in numerical simulation. In detail, see Chandra and Taniguchi (2006 Theorem 5 and p.31).

**Acknowledgements** The authors are grateful to three referees for their helpful comments, which improved the original version greatly.

## REFERENCES

- [1] BRILLINGER, D. R. (2001). *Time Series: Data Analysis and Theory, expanded ed.* Holden-Day, San Francisco.
- [2] BROCKWELL P. J. AND DAVIS. R. A (1991). *Time Series: Theory and Methods, second ed.* Springer-Verlag.
- [3] HOSOYA, Y. AND 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.
- [4] KEENAN, D. M. (1987). Limiting Behavior of Functionals of Higher-Order Sample Cumulant Spectra. *Ann. Statist.* **15**, 134-151.
- [5] MONTI, A. C. (1997). Empirical likelihood confidence regions in time series models. *Biometrika.* **84**, 395-405.
- [6] OWEN, A. B. (2001). *Empirical Likelihood.* Chapman-Hall/CRC, Boca Raton, London, New York, Washinton, D.C.
- [7] TANIGUCHI, M. AND KAKIZAWA, Y. (2000). *Asymptotic Theory of Statistical Inference for Time Series.* Springer-Verlag, New York.