J. Japan Statist. Soc.
Vol. 38 No. 1 2008 157–171

GENERALIZED INFORMATION CRITERIA IN MODEL SELECTION FOR LOCALLY STATIONARY PROCESSES

Junichi Hirukawa*, Hiroko Solvang Kato**, Kenichiro Tamaki*** and Masanobu Taniguchi****

The problem of fitting a parametric model of time series with time varying parameters attracts our attention. We evaluate a goodness of time varying spectral models from an information theoretic point of view. We propose model selection criteria for locally stationary processes based on nonlinear functionals of a time varying spectral density without assuming that the true time varying spectral density belongs to the model. Also, we obtain a sufficient condition such that our information criteria coincide with Akaike's information criterion.

Key words and phrases: Generalized information criterion, locally stationary process, minimum distance estimation, misspecified models, time varying spectral density.

1. Introduction

The problem of evaluating goodness of statistical models has been well investigated from an information theoretic point of view. To evaluate models, we usually assume their structure is specified by some function. As examples of such functions we can take the probability distribution function $p(x)$ for the i.i.d. case, the trend function $\mu(u)$ for regression models, the spectral density function $f(\lambda)$ for stationary processes, and the dynamic system function $\boldsymbol{F}(X_{t-1}, \ldots, X_{t-p})$ for nonlinear models.

Time series analysis has been developed under stationarity. However, the assumption of stationarity is insufficient to describe the real time series data. Recently, an important class of nonstationary processes has been proposed by Dahlhaus (1996a, 1996b, 1996c), called locally stationary processes. We give the precise definition which is due to Dahlhaus (1996a, 1996b).

Definition 1. *A sequence of stochastic processes* $X_{t,T}$ $(t = 1, \ldots, T; T \geq 1)$ *is called locally stationary with transfer function* A° *if there exists a representation*

$$X_{t,T} = \int_{-\pi}^{\pi} \exp(i\lambda t) A^\circ_{t,T}(\lambda) d\xi(\lambda),$$

Accepted February 6, 2008.

*Department of Mathematics, Faculty of Science, Niigata University, 8050 Ikarashi 2-no-cho, Nishi-ku, Niigata City, Niigata 950-2181, Japan.

**Department of Genetics, Institute for Cancer Research, Radiumhospitalet, University of Oslo, Norway.

***School of Political Science and Economics, Waseda University, 1-6-1, Nishiwaseda, Shinjuku-ku, Tokyo 169-8050, Japan.

****Department of Applied Mathematics, School of Fundamental Science and Engineering, Waseda University, 3-4-1, Okubo, Shinjuku-ku, Tokyo 169-8555, Japan.

where

(i) $\xi(\lambda)$ *is a stochastic process on* $[-\pi,\pi]$ *with* $\overline{\xi(\lambda)}=\xi(-\lambda)$ *and*

$$\operatorname{cum}\{d\xi(\lambda_1),\ldots,d\xi(\lambda_k)\} = \eta\left(\sum_{j=1}^{k}\lambda_j\right)\kappa_k(\lambda_1,\ldots\lambda_{k-1})d\lambda_1\cdots d\lambda_{k-1}, \quad (1.1)$$

where $\operatorname{cum}\{\cdots\}$ *denotes the cumulant of k-th order,* $\kappa_1=0$, $\kappa_2(\lambda)=(2\pi)^{-1}$, $|\kappa_k(\lambda_1,\ldots,\lambda_{k-1})|\le \text{const}_k$ *for all* $k\ge 3$ *and* $\eta(\lambda)=\sum_{j=-\infty}^{\infty}\delta(\lambda+2\pi j)$ *is the period* 2π *extension of the Dirac delta function. Write* $\varepsilon_t=\int_{-\pi}^{\pi}\exp(i\lambda t)d\xi(\lambda)$, *then* $\{\varepsilon_t\}$ *becomes a white noise sequence and is called the innovation of the process. If the innovation* ε_t*'s are i.i.d., then* $\kappa_k(\lambda_1,\ldots,\lambda_{k-1})\equiv\kappa_k$ *(constant).*

(ii) *There exists a constant* K *and a* 2π*-periodic function* $A:[0,1]\times\boldsymbol{R}\to\boldsymbol{C}$ *with* $\overline{A(u,\lambda)}=A(u,-\lambda)$ *and*

$$\sup_{t,\lambda}\left|A^{\circ}_{t,T}(\lambda)-A\left(\frac{t}{T},\lambda\right)\right|\le KT^{-1} \quad (1.2)$$

for all T. $A(u,\lambda)$ *is assumed to be continuous in* u *and* $g(u,\lambda):=|A(u,\lambda)|^2$ *is called the time varying spectral density of the process.*

A major difference between the above definition and Priestley's definition of an oscillatory process (see Priestley (1981), Chapter 11) is that we consider double indexed processes. That is, the locally stationary process depends on both t and T, which allows us to use asymptotic considerations. A justification of the locally stationary approach and a comparison with the approach of Priestley can be found in Dahlhaus (1996b, 1996c).

The structure of locally stationary processes is specified by the smooth function, namely the time varying spectral density function $g(u,\lambda)$. We want to fit a class of time varying spectral models $\mathcal{P}=\{f_{\boldsymbol{\theta}}(u,\lambda):\boldsymbol{\theta}\in\boldsymbol{\Theta}\subset\boldsymbol{R}^q\}$ to $g(u,\lambda)$. Dahlhaus (1996a) and Van Bellegem and Dahlhaus (2006) proposed model selection criterion for locally stationary processes based on the Gaussian Kullback-Leibler distance. In this paper, more generally, we suggest model selection criteria for locally stationary processes based on nonlinear functionals of a time varying spectral density without assuming that the true time varying spectral density $g(u,\lambda)$ belongs to the model $\mathcal{P}$.

The distance functions based on nonlinear functionals of time varying spectral densities are defined in Section 2. The asymptotic normality of our minimum distance estimators are proved in Section 3. The generalized information criteria based on nonlinear integral functionals are derived in Section 4. Numerical examples and empirical study are shown in Sections 5 and 6. Some technical lemmas are given in the Appendix.

2. Nonlinear distance functions

Many important quantities in locally stationary processes are often expressed as functionals of time varying spectral densities. For a linear functional, a natural idea of constructing an estimator is to replace an unknown time varying density by the local periodogram based on the data. The functional of interest is, however, not always linear with respect to the time varying spectral density. In these cases we use the nonparametric kernel type time varying spectral density estimator instead of a local periodogram to avoid the inconsistency.

To simplify, we restrict ourselves to univariate locally stationary processes $\{X_{t,T}\}$ $(t = 1, \ldots, T; T \geq 1)$ with mean zero and time varying spectral density $g(u, \lambda)$. Suppose a stretch $\boldsymbol{X}_T = \{X_{2-N/2,T}, \ldots, X_{1,T}, \ldots, X_{T,T}, \ldots, X_{T+N/2,T}\}$ is available from this locally stationary processes. We want to fit a class of time varying spectral models $\mathcal{P} = \{f_{\boldsymbol{\theta}}(u, \lambda) : \boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \boldsymbol{R}^q\}$ without assuming that the true time varying spectral density $g(u, \lambda)$ belongs to $\mathcal{P}$. Dahlhaus (1996a) and Van Bellegem and Dahlhaus (2006) proposed model selection criterion for locally stationary processes based on the Gaussian Kullback-Leibler distance. More generally, we consider in this paper the local distance function at time u of the form

$$D(\boldsymbol{\theta}, g, u) = \int_{-\pi}^{\pi} K\{\boldsymbol{\theta}, g(u, \lambda), u, \lambda\} d\lambda \tag{2.1}$$

associated with function $K(\cdot, \cdot, \cdot, \cdot)$ defined below. Some concrete examples of distance functions of this form for stationary processes are found in Dahlhaus and Wefelmeyer (1996) and Taniguchi and Kakizawa (2000), and they will be naturally extended to locally stationary processes.

We set a functional S by the requirement

$$D\{S_g(u), g, u\} = \min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} D(\boldsymbol{\theta}, g, u), \tag{2.2}$$

and make the following assumption, as did Dahlhaus and Wefelmeyer (1996).

Assumption 1.

(i) The parameter space $\boldsymbol{\Theta} \subset \boldsymbol{R}^q$ is compact and the function $K(\boldsymbol{\theta}, z, u, \lambda) : \boldsymbol{\Theta} \times (0, \infty) \times [0, 1] \times [-\pi, \pi] \to \boldsymbol{R}$ in (2.1) is three time differentiable in $(\boldsymbol{\theta}, z)$ with continuous derivatives in $(\boldsymbol{\theta}, z, u, \lambda)$.

(ii) For the true time varying spectral density g and for all u, $S_g(u)$ exists, is unique and lies in the interior of $\boldsymbol{\Theta}$.

Denote the first and second derivatives of $K(\boldsymbol{\theta}, z, u, \lambda)$ by

$$K_j(z, \cdot, \cdot) = \frac{\partial}{\partial \theta_j} K(\boldsymbol{\theta}, z, \cdot, \cdot) \mid_{\boldsymbol{\theta} = S_g(u)}$$

and

$$K_{j,k}(z, \cdot, \cdot) = \frac{\partial^2}{\partial \theta_j \partial \theta_k} K(\boldsymbol{\theta}, z, \cdot, \cdot) \mid_{\boldsymbol{\theta} = S_g(u)},$$

and the first derivative of $K_j(z, u, \lambda)$ by

$$K_j^{(1)}(z, \cdot, \cdot) = \frac{\partial}{\partial z} K_j(z, \cdot, \cdot).$$

Furthermore, we define

$$D_g(u) = \int_{-\pi}^{\pi} \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} K\{\boldsymbol{\theta}, g(u, \lambda), u, \lambda\} \mid_{\boldsymbol{\theta}=S_g(u)} d\lambda \tag{2.3}$$

and assume the following:

Assumption 2.

(i) The first derivative of $K_j(z, u, \lambda)$ satisfies

$$K_j^{(1)}\{g(u, -\lambda), u, -\lambda\} = K_j^{(1)}\{g(u, \lambda), u, \lambda\}.$$

(ii) The $q \times q$ matrix $D_g(u)$ is nonsingular for all u.

We now give two specific forms of the function $K(\cdot, \cdot, \cdot, \cdot)$.

Contrast type: Let $H(x)$ on $(0, \infty)$ be an appropriate smooth function which is three times continuously differentiable and has a unique minimum zero at $x = 1$, such as

$$\begin{aligned} H_1(x) &= -\log(x) + x - 1, \\ H_2(x) &= \frac{1}{\alpha(1-\alpha)}\{\log(\alpha x + 1 - \alpha) - \alpha \log(x)\}, \qquad 0 < \alpha < 1, \\ H_3(x) &= \frac{1}{2}(x-1)^2. \end{aligned}$$

Then, we define the contrast type function

$$K(\boldsymbol{\theta}, z(u, \lambda), u, \lambda) = H\{f_{\boldsymbol{\theta}}(u, \lambda) z(u, \lambda)^{-1}\} \tag{2.4}$$

or

$$K(\boldsymbol{\theta}, z(u, \lambda), u, \lambda) = H\{z(u, \lambda) f_{\boldsymbol{\theta}}(u, \lambda)^{-1}\} \tag{2.5}$$

or

$$\begin{aligned} &K(\boldsymbol{\theta}, z(u, \lambda), u, \lambda) \\ &\quad = \frac{1}{2}[H\{z(u, \lambda) f_{\boldsymbol{\theta}}(u, \lambda)^{-1}\} + H\{f_{\boldsymbol{\theta}}(u, \lambda) z(u, \lambda)^{-1}\}]. \end{aligned} \tag{2.6}$$

Weighted squared function: Let $\psi(u, \lambda)$ be a given weighted function which satisfies $\psi(u, -\lambda) = \psi(u, \lambda)$. Then, we define

$$K(\boldsymbol{\theta}, z(u, \lambda), u, \lambda) = \frac{1}{2}[\psi(u, \lambda)\{f_{\boldsymbol{\theta}}(u, \lambda) - z(u, \lambda)\}]^2.$$

In these cases, the identifiability condition (S1) below on parametrization of the model $\mathcal{P} = \{f_{\boldsymbol{\theta}}(u, \lambda) : \boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \boldsymbol{R}^q\}$ implies that

$$S_{f_{\boldsymbol{\theta}}}(u) = \boldsymbol{\theta} \text{ uniquely for every } \boldsymbol{\theta} \in \boldsymbol{\Theta},$$

since $K\{\boldsymbol{t}, f_{\boldsymbol{\theta}}(u,\lambda), u, \lambda\} \geq K\{\boldsymbol{\theta}, f_{\boldsymbol{\theta}}(u,\lambda), u, \lambda\}$ and the equality holds if and only if $\boldsymbol{t} = \boldsymbol{\theta}$. On the other hand Assumption 1 (i) is fulfilled, if the model time varying densities fulfill the assumption (S2) below.

We make the following assumption on the model time varying spectral density $f_{\boldsymbol{\theta}}$.

ASSUMPTION 3.
(S1) For fixed u, if $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$, then $f_{\boldsymbol{\theta}_1}(u,\lambda) \neq f_{\boldsymbol{\theta}_2}(u,\lambda)$ on a set of positive Lebesgue measure.
(S2) The model time varying spectral density $f_{\boldsymbol{\theta}}$ is three times continuously differentiable with respect to $\boldsymbol{\theta}$ and these derivatives are continuous in u and λ. Furthermore, $f_{\boldsymbol{\theta}}$ and its derivatives are uniformly bounded and bounded away from 0.

3. Parameter estimation based on functionals

Now, we turn to developing estimation theory for estimators corresponding to $S_g(u)$, where S is the functional defined in (2.2). The estimator is naturally defined by $S_{\hat{g}_T}(u)$, where $\hat{g}_T$ is a suitable time varying spectral density estimator. We will examine the large sample behavior of $S_{\hat{g}_T}(u)$ when $\hat{g}_T$ is a nonparametric kernel type time varying spectral density estimator

$$\hat{g}_T(u,\lambda) = \int_{-\pi}^{\pi} W_M(\lambda-\mu) I_N(u,\mu) d\mu, \tag{3.1}$$

where

$$I_N(u,\lambda) = \frac{1}{2\pi H_N} \left| \sum_{s=0}^{N-1} h\left(\frac{s}{N}\right) X_{[uT]-N/2+s+1,T} \exp(i\lambda s) \right|^2 \tag{3.2}$$

is the data tapered local periodogram at time u, $W_M(\lambda) = M \sum_{\nu=-\infty}^{\infty} W\{M(\lambda + 2\pi\nu)\}$ is the weight function, $h : [0,1] \to \boldsymbol{R}$ is a data taper and $H_N = \sum_{s=0}^{N-1} h^2(s/N) \sim N \int_0^1 h^2(x)dx$. We impose the following assumptions on the parameters M and N.

ASSUMPTION 4. The parameters $M = M(T)$ and $N = N(T)$, $M \ll N \ll T$ depends on T in such a way that

$$\frac{M}{N^{1/2}} + \frac{N^{1/4}}{M} + \frac{N^5}{T^4} \to 0.$$

Furthermore, we assume that the functions $W(\cdot)$ and $h(\cdot)$ satisfy the following:

ASSUMPTION 5.
(i) The weight function $W : \boldsymbol{R} \to [0,\infty]$ satisfies $W(x) = 0$ for $x \notin [-\pi,\pi]$, and is a continuous and even function satisfying $\int_{-\pi}^{\pi} W(x)dx = 1$ and $\int_{-\pi}^{\pi} x^2 W(x)dx < \infty$.

(ii) For $M = O(N^\alpha)$, $(1/4 < \alpha < 1/2)$, the function $W_M(\lambda) = M \sum_{\nu=-\infty}^{\infty} \cdot W\{M(\lambda + 2\pi\nu)\}$ can be expanded as

$$W_M(\lambda) = \frac{1}{2\pi} \sum_{j=-(N-1)}^{N-1} \omega\left(\frac{j}{M}\right) e^{-ij\lambda},$$

where $\omega(x)$ is a continuous, even function with $\omega(0) = 1$, $|\omega(x)| \le 1$ and $\int_{-\infty}^{\infty} \omega(x)dx < \infty$.

Assumption 6. The data taper $h : \boldsymbol{R} \to \boldsymbol{R}$ satisfies (i) $h(x) = 0$ for all $x \notin [0,1]$ and $h(x) = h(1-x)$, (ii) $h(x)$ is continuous on $\boldsymbol{R}$, twice differentiable at all $x \notin U$ where U is a finite set of $\boldsymbol{R}$, and $\sup_{x\notin U} |h''(x)| < \infty$.

Write

$$K_t(x) := \left\{\int_0^1 h(x)^2 dx\right\}^{-1} h(x+1/2)^2, \qquad x \in [-1/2, 1/2],$$

which plays a role of kernel in the time domain.

Using Lemma 3 and (ii), (iii) of Lemma 2, we obtain

$$\begin{aligned}\sqrt{N}&\{S_{\hat{g}_T}(u) - S_g(u)\} \\ &= -D_g(u)^{-1}\sqrt{N}\left(\int_{-\pi}^{\pi} \phi_j(u,\lambda)[I_N(u,\lambda) - E\{I_N(u,\lambda)\}]d\lambda\right)'_{j=1,\dots,q} + o_P(1),\end{aligned}$$

where $\phi_j(u,\lambda) = K_j^{(1)}(g(u,\lambda), u, \lambda)$. The following theorem is a consequence of Lemma 1.

Theorem 1. *Under Assumptions* 1, 2, 4–6, *we have*:

$$\begin{aligned}\sqrt{N}&\{S_{\hat{g}_T}(u) - S_g(u)\} \\ &\xrightarrow{d} N(\mathbf{0}, v(h)D_g(u)^{-1}\Gamma_g(u)D_g(u)^{-1}),\end{aligned}$$

where $\Gamma_g(u)$ *is the covariance matrix of a Gaussian vector* $\xi(\phi_j)$, $j = 1, \dots, q$ *given in* (A.1) *below.*

In particular, if we take the contrast type function (2.4) as $K(\cdot,\cdot,\cdot,\cdot,)$, then we have

$$D_g(u) = \int_{-\pi}^{\pi}\left[\frac{H^{(2)}\{f_{\boldsymbol{\theta}}(u,\lambda)g(u,\lambda)^{-1}\}}{g(u,\lambda)^2}\frac{\partial}{\partial\boldsymbol{\theta}}f_{\boldsymbol{\theta}}(u,\lambda)\frac{\partial}{\partial\boldsymbol{\theta}'}f_{\boldsymbol{\theta}}(u,\lambda) + \frac{H^{(1)}\{f_{\boldsymbol{\theta}}(u,\lambda)g(u,\lambda)^{-1}\}}{g(u,\lambda)}\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}f_{\boldsymbol{\theta}}(u,\lambda)\right]_{\boldsymbol{\theta}=S_g(u)} d\lambda \tag{3.3}$$

and

$$\begin{aligned}\{\Gamma_g(u)\}_{j,k} = 4\pi &\int_{-\pi}^{\pi} \phi_j(u,\lambda)\phi_k(u,\lambda)g^2(u,\lambda)d\lambda \\ &+ 2\pi\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \phi_j(u,\lambda)\phi_k(u,\mu)g(u,\lambda)g(u,\mu) \\ &\qquad \times \kappa_4(\lambda,-\lambda,\mu)d\lambda d\mu\end{aligned} \tag{3.4}$$

with

$$\phi_j(u,\lambda) = -\Big[\Big[\frac{H^{(2)}\{f_{\boldsymbol\theta}(u,\lambda)g(u,\lambda)^{-1}\}f_{\boldsymbol\theta}(u,\lambda)}{g(u,\lambda)^3} + \frac{H^{(1)}\{f_{\boldsymbol\theta}(u,\lambda)g(u,\lambda)^{-1}\}}{g(u,\lambda)^2}\Big]\frac{\partial}{\partial\theta_j}f_{\boldsymbol\theta}(u,\lambda)\Big]_{\boldsymbol\theta=S_g(u)},$$

where $H^{(i)}(\cdot)$, $i=1,2$ are the first and second derivatives of $H(\cdot)$. In addition, we introduce the following assumptions.

Assumption 7. The true time varying spectral density of the process is $f_{\boldsymbol\theta}$, where $\boldsymbol\theta$ lies in the interior of $\boldsymbol\Theta$.

Assumption 8. The process is Gaussian, or $\boldsymbol\theta$ is innovation free (namely, the innovation ε_t's are i.i.d. and $\boldsymbol\theta$ satisfies $\frac{\partial}{\partial\boldsymbol\theta}\int_{-\pi}^{\pi}\log f_{\boldsymbol\theta}(u,\lambda)d\lambda = \mathbf{0}$).

Under Assumptions 1–7, we have

$$D_g(u) = H^{(2)}(1)\int_{-\pi}^{\pi}\frac{1}{f_{\boldsymbol\theta}(u,\lambda)^2}\frac{\partial}{\partial\boldsymbol\theta}f_{\boldsymbol\theta}(u,\lambda)\frac{\partial}{\partial\boldsymbol\theta'}f_{\boldsymbol\theta}(u,\lambda)d\lambda$$

and

$$\Gamma_g(u) = 4\pi H^{(2)}(1)D_g(u) + 2\pi\{H^{(2)}(1)\}^2\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\frac{\kappa_4(\lambda,-\lambda,\mu)}{f_{\boldsymbol\theta}(u,\lambda)f_{\boldsymbol\theta}(u,\mu)}\frac{\partial}{\partial\boldsymbol\theta}f_{\boldsymbol\theta}(u,\lambda)\frac{\partial}{\partial\boldsymbol\theta'}f_{\boldsymbol\theta}(u,\mu)d\lambda d\mu.$$

Similarly we have the following results for general contrast type distance function:

Theorem 2. *For the contrast type $K(\cdot,\cdot,\cdot,\cdot)$ of the form* (2.4) *or* (2.5) *or* (2.6), *denote the functional $S_g(u)$ of* (2.2) *by $S_g^c(u)$. Under Assumptions* 1–7, *we have*:

(i) *Then $\sqrt{N}\{S^c_{\hat g_T}(u)-\boldsymbol\theta\}$ is asymptotically normal with mean vector zero and covariance matrix $v(h)\tilde D_g(u)^{-1}\{4\pi\tilde D_g(u)+\Pi_g(u)\}\tilde D_g(u)^{-1}$. Here $\tilde D_g(u)$ and $\Pi_g(u)$ are given by*

$$\tilde D_g(u) = \int_{-\pi}^{\pi}\frac{1}{f_{\boldsymbol\theta}(u,\lambda)^2}\frac{\partial}{\partial\boldsymbol\theta}f_{\boldsymbol\theta}(u,\lambda)\frac{\partial}{\partial\boldsymbol\theta'}f_{\boldsymbol\theta}(u,\lambda)d\lambda$$

and

$$\Pi_g(u) = 2\pi\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\frac{\kappa_4(\lambda,-\lambda,\mu)}{f_{\boldsymbol\theta}(u,\lambda)f_{\boldsymbol\theta}(u,\mu)}\frac{\partial}{\partial\boldsymbol\theta}f_{\boldsymbol\theta}(u,\lambda)\frac{\partial}{\partial\boldsymbol\theta'}f_{\boldsymbol\theta}(u,\mu)d\lambda d\mu.$$

(ii) *In addition if Assumption* 8 *holds, then the asymptotical covariance matrix of $\sqrt{N}\{S^c_{\hat g_T}(u)-\boldsymbol\theta\}$ becomes $4\pi v(h)\tilde D_g(u)^{-1}$.*

4. Model selection criterion

We now turn to discuss model selection criterion. Recall that we fit a class of parametric models $\mathcal{P} = \{f_{\boldsymbol{\theta}}(u,\lambda) : \boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \boldsymbol{R}^q\}$ to g by use of a measure of local disparity at time u, $D(\boldsymbol{\theta}, g, u)$, and we estimate $\boldsymbol{\theta}$ by the value $S_{\hat{g}_T}(u)$ which minimizes $D(\boldsymbol{\theta}, \hat{g}_T, u)$, where $\hat{g}_T(u,\lambda)$ is a nonparametric kernel type time varying spectral density estimator at time u. Nearness between $f_{S_{\hat{g}_T}(u)}$ and g is measured by $D(S_{\hat{g}_T}(u), g, u)$. A simple estimator of $D(S_{\hat{g}_T}(u), g, u)$ is given by substituting for g the nonparametric time varying spectral density estimator $\hat{g}_T$, yielding $D(S_{\hat{g}_T}(u), \hat{g}_T, u)$. Ordinarily this provides an underestimate of $D(S_{\hat{g}_T}(u), g, u)$. Writing bias as

$$b_g(u) = E_{X_T}\{D(S_{\hat{g}_T}(u), \hat{g}_T, u) - D(S_{\hat{g}_T}(u), g, u)\}, \tag{4.1}$$

we define the generalised information criterion as

$$D(S_{\hat{g}_T}(u), \hat{g}_T, u) - b_{\hat{g}_T}(u).$$

Recall we define the pseudo true value of $\boldsymbol{\theta}$ in terms of the functional $S_g(u)$ by the requirement $D(S_g(u), g, u) = \min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} D(\boldsymbol{\theta}, g, u)$. Expanding $D(S_{\hat{g}_T}(u), g, u)$ at $S_g(u)$ we obtain the approximation

$$\begin{aligned}
&D(S_{\hat{g}_T}(u), g, u) \\
&\quad\approx D(S_g(u), g, u) + (S_{\hat{g}_T}(u) - S_g(u))' \frac{\partial}{\partial\boldsymbol{\theta}} D(\boldsymbol{\theta}, g, u) \mid_{\boldsymbol{\theta}=S_g(u)} \\
&\qquad + \frac{1}{2}(S_{\hat{g}_T}(u) - S_g(u))' \frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} D(\boldsymbol{\theta}, g, u) \mid_{\boldsymbol{\theta}=S_g(u)} (S_{\hat{g}_T}(u) - S_g(u)) \\
&\quad= D(S_g(u), g, u) \\
&\qquad + \frac{1}{2}(S_{\hat{g}_T}(u) - S_g(u))' \frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} D(\boldsymbol{\theta}, g, u) \mid_{\boldsymbol{\theta}=S_g(u)} (S_{\hat{g}_T}(u) - S_g(u)).
\end{aligned}$$

On the other hand, expanding $D(S_g(u), \hat{g}_T, u)$ at $S_{\hat{g}_T}(u)$ we have

$$\begin{aligned}
&D(S_g(u), \hat{g}_T, u) \\
&\quad\approx D(S_{\hat{g}_T}(u), \hat{g}_T, u) + (S_g(u) - S_{\hat{g}_T}(u))' \frac{\partial}{\partial\boldsymbol{\theta}} D(\boldsymbol{\theta}, \hat{g}_T, u) \mid_{\boldsymbol{\theta}=S_{\hat{g}_T}(u)} \\
&\qquad + \frac{1}{2}(S_{\hat{g}_T}(u) - S_g(u))' \frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} D(\boldsymbol{\theta}, \hat{g}_T, u) \mid_{\boldsymbol{\theta}=S_{\hat{g}_T}(u)} (S_{\hat{g}_T}(u) - S_g(u)) \\
&\quad= D(S_{\hat{g}_T}(u), \hat{g}_T, u) \\
&\qquad + \frac{1}{2}(S_{\hat{g}_T}(u) - S_g(u))' \frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} D(\boldsymbol{\theta}, \hat{g}_T, u) \mid_{\boldsymbol{\theta}=S_{\hat{g}_T}(u)} (S_{\hat{g}_T}(u) - S_g(u)).
\end{aligned}$$

From the fact that $\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} D(\boldsymbol{\theta}, \hat{g}_T, u) \mid_{\boldsymbol{\theta}=S_{\hat{g}_T}(u)} \xrightarrow{P} \frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} D(\boldsymbol{\theta}, g, u) \mid_{\boldsymbol{\theta}=S_g(u)} = D_g(u)$ and $E_{\boldsymbol{X}_T}\{D(S_g(u), \hat{g}_T, u)\} \approx D(S_g(u), g, u)$, we see that

$$\begin{aligned}
&D(S_{\hat{g}_T}(u), g, u) - D(S_{\hat{g}_T}(u), \hat{g}_T, u) \\
&\quad\approx D(S_g(u), g, u) - D(S_g(u), \hat{g}_T, u) \\
&\qquad + (S_{\hat{g}_T}(u) - S_g(u))' D_g(u) (S_{\hat{g}_T}(u) - S_g(u))
\end{aligned}$$

and

$$(4.2)\qquad -b_g(u) \approx E_{\boldsymbol{X}_T}\{(S_{\hat{g}_T}(u) - S_g(u))' D_g(u)(S_{\hat{g}_T}(u) - S_g(u))\} \approx \frac{v(h)}{N}\,\mathrm{tr}\{D_g(u)^{-1}\Gamma_g(u)\},$$

where $D_g(u)$ and $\Gamma_g(u)$ are as in (2.3) and (A.1). Since $D_g(u)$ and $\Gamma_g(u)$ depend on g, we replace g by the nonparametric time varying spectral density estimator $\hat{g}_T$. Then (4.1) and (4.2) validate that $G_N(q) = D(S_{\hat{g}_T}(u), \hat{g}_T, u) + N^{-1}v(h)\,\mathrm{tr}\{D_{\hat{g}_T}(u)^{-1}\Gamma_{\hat{g}_T}(u)\}$ is an asymptotically unbiased estimator of $E_{X_T}\{D(S_{\hat{g}_T}(u), g, u)\}$. Multiplying $G_N(q)$ by N we call

$$\mathrm{GIC}(q) = ND(S_{\hat{g}_T}(u), \hat{g}_T, u) + v(h)\,\mathrm{tr}\{D_{\hat{g}_T}(u)^{-1}\Gamma_{\hat{g}_T}(u)\}$$

a generalized information criterion.

In particular, for the contrast type $K(\cdot,\cdot,\cdot,\cdot)$ of the form (2.4) or (2.5) or (2.6) under Assumptions 1–8, the generalized information criterion GIC(q) becomes Akaike's information criterion

$$(4.3)\qquad \mathrm{AIC}(q) = ND(S_{\hat{g}_T}(u), \hat{g}_T, u) + 4\pi H^{(2)}(1)v(h)q.$$

5. Numerical examples

In this section we give concrete examples of the quantity $v(h)\,\mathrm{tr}\{D_g(u)^{-1}\Gamma_g(u)\}$ in (4.2), which is interpreted as the penalized term in the GIC. Here we take the contrast type function (2.4) as $K(\cdot,\cdot,\cdot,\cdot)$ with $H(x) = H_1(x) = -\log(x) + x - 1$.

(i) First, we consider the misspecified case. Namely, we fit the stationary model $f_{\boldsymbol{\theta}}(u,\lambda) = f_{\boldsymbol{\theta}}(\lambda)$ to a locally stationary process which has the true time varying spectral density $g(u,\lambda)$. Here we assume that $\boldsymbol{\theta}$ is innovation free (the innovation ε_t's are i.i.d. and $\boldsymbol{\theta}$ satisfies $\frac{\partial}{\partial\boldsymbol{\theta}}\int_{-\pi}^{\pi}\log f_{\boldsymbol{\theta}}(\lambda)d\lambda = \mathbf{0}$). Then, $\boldsymbol{\theta}$ satisfies

$$(5.1)\qquad \frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\int_{-\pi}^{\pi}\log f_{\boldsymbol{\theta}}(\lambda)d\lambda = \int_{-\pi}^{\pi} f_{\boldsymbol{\theta}}(\lambda)^{-1}\frac{\partial^2 f_{\boldsymbol{\theta}}(\lambda)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} - f_{\boldsymbol{\theta}}(\lambda)^{-2}\left(\frac{\partial f_{\boldsymbol{\theta}}(\lambda)}{\partial\boldsymbol{\theta}}\right)\left(\frac{\partial f_{\boldsymbol{\theta}}(\lambda)}{\partial\boldsymbol{\theta}'}\right)d\lambda = \mathbf{0}.$$

Therefore, we have

$$D_g(u) = \int_{-\pi}^{\pi}\left\{g(u,\lambda)^{-1}\frac{\partial^2 f_{\boldsymbol{\theta}}(\lambda)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right\}_{\boldsymbol{\theta}=S_g(u)} d\lambda,$$

$$\Gamma_g(u) = \int_{-\pi}^{\pi}\left\{g(u,\lambda)^{-2}\left(\frac{\partial f_{\boldsymbol{\theta}}(\lambda)}{\partial\boldsymbol{\theta}}\right)\left(\frac{\partial f_{\boldsymbol{\theta}}(\lambda)}{\partial\boldsymbol{\theta}'}\right)\right\}_{\boldsymbol{\theta}=S_g(u)} d\lambda.$$

If the true time varying spectral density is of the form $g(u,\lambda) = f_{\boldsymbol{\theta}}(\lambda)\alpha(u)$, then from (5.1) the penalized term becomes

$$\mathrm{tr}\{\nu(h)D_g(u)^{-1}\Gamma_g(u)\} = 4\pi\frac{\nu(h)}{\alpha(u)}q,$$

which coincides with the penalized term of AIC when $\alpha(u) \equiv 1$.

(ii) Next, we consider the overfitted case. Namely, we fit the locally stationary model $f_{\boldsymbol{\theta}}(u,\lambda) = f_{1,\boldsymbol{\theta}_{(q-1)}}(\lambda) f_{2,\theta_{(q)}}(u)$, $\boldsymbol{\theta} = (\boldsymbol{\theta}'_{(q-1)}, \theta_{(q)})'$, $\boldsymbol{\theta}_{(q-1)} \in \boldsymbol{R}^{q-1}$, $\theta_{(q)} \in \boldsymbol{R}^1$ to a stationary process which has the true spectral density $g(\lambda)$. Here we assume that $\boldsymbol{\theta}_{(q-1)}$ is innovation free (the innovation ε_t's are i.i.d. and $\boldsymbol{\theta}_{(q-1)}$ satisfies $\frac{\partial}{\partial \boldsymbol{\theta}_{(q-1)}} \int_{-\pi}^{\pi} \log f_{1,\boldsymbol{\theta}_{(q-1)}}(\lambda) d\lambda = \boldsymbol{0}$). Then, we have

$$D_g(u) = \begin{pmatrix} D_g(u)_{(11)} & \boldsymbol{0} \\ \boldsymbol{0} & D_g(u)_{(22)} \end{pmatrix} \quad \text{and}$$
$$\Gamma_g(u) = \begin{pmatrix} \Gamma_g(u)_{(11)} & * \\ * & \Gamma_g(u)_{(22)} \end{pmatrix},$$

where

$$D_g(u)_{(11)} = \int_{-\pi}^{\pi} \left\{ \frac{f_{2,\theta_{(q)}}(u)}{g(\lambda)} \frac{\partial^2 f_{1,\boldsymbol{\theta}_{(q-1)}}(\lambda)}{\partial \boldsymbol{\theta}_{(q-1)} \partial \boldsymbol{\theta}'_{(q-1)}} \right\}_{\boldsymbol{\theta}=S_g(u)} d\lambda,$$

$$D_g(u)_{(22)} = 2\pi \left[f_{2,\theta_{(q)}}(u)^{-2} \left\{ \frac{\partial \theta_{(q)} f_{2,\theta_{(q)}}(u)}{\partial \theta_{(q)}} \right\}^2 \right]_{\boldsymbol{\theta}=S_g(u)},$$

$$\Gamma_g(u)_{(11)} = 4\pi \int_{-\pi}^{\pi} \left\{ \frac{f_{2,\theta_{(q)}}(u)^2}{g(\lambda)^2} \left(\frac{\partial f_{1,\boldsymbol{\theta}_{(q-1)}}(\lambda)}{\partial \boldsymbol{\theta}_{(q-1)}} \right) \left(\frac{\partial f_{1,\boldsymbol{\theta}_{(q-1)}}(\lambda)}{\partial \boldsymbol{\theta}'_{(q-1)}} \right) \right\}_{\boldsymbol{\theta}=S_g(u)} d\lambda,$$

$$\Gamma_g(u)_{(22)} = 4\pi \int_{-\pi}^{\pi} \left[\left\{ \frac{\partial \theta_{(q)} f_{2,\theta_{(q)}}(u)}{\partial \theta_{(q)}} \right\}^2 \frac{f_{1,\boldsymbol{\theta}_{(q-1)}}(\lambda)^2}{g(\lambda)^2} \right]_{\boldsymbol{\theta}=S_g(u)} d\lambda$$
$$+ 8\pi^3 \kappa_4 \left[f_{2,\theta_{(q)}}(u)^{-2} \left\{ \frac{\partial \theta_{(q)} f_{2,\theta_{(q)}}(u)}{\partial \theta_{(q)}} \right\}^2 \right]_{\boldsymbol{\theta}=S_g(u)}$$

and

$$\begin{aligned} &\mathrm{tr}\{v(h) D_g(u)^{-1} \Gamma_g(u)\} \\ &\quad = \mathrm{tr}\{v(h) D_g(u)_{(11)}^{-1} \Gamma_g(u)_{(11)}\} + \mathrm{tr}\{v(h) D_g(u)_{(22)}^{-1} \Gamma_g(u)_{(22)}\}. \end{aligned}$$

If the true spectral density is of the form $g(\lambda) = f_{1,\boldsymbol{\theta}_{(q-1)}}(\lambda)$ and the pseudo true value $\boldsymbol{\theta}^0(u) = (\boldsymbol{\theta}^0_{(q-1)}(u)', \theta^0_{(q)}(u))' = S_{f_{1,\boldsymbol{\theta}_{(q-1)}}}(u)$ satisfies $\boldsymbol{\theta}^0_{(q-1)}(u) = \boldsymbol{\theta}_{(q-1)}$ and $f_{2,\theta^0_{(q)}(u)}(u) \equiv 1$, then the penalized term becomes

$$\mathrm{tr}\{v(h) D_g(u)^{-1} \Gamma_g(u)\} = 4\pi v(h) \{(q-1) + 1 + \pi \kappa_4\}.$$

6. Empirical study

Because we explained the model selection procedures and the parameter estimation methods, we can now identify the statistical models from real data. We apply our methods to the daily log returns $\{X_{1-N/2,T}, \ldots, X_{0,T}, \ldots, X_{T,T}, \ldots,$

$X_{T+N/2,T}\}$ of the S&P 500 index from September 20, 2005 to September 14, 2007 (500 trading days).

First, we fit stationary AR(q) models to the data $\{X_{k+1-N/2,T}, \dots, X_{k+N/2,T}\}$ in terms of Yule-Walker equations for each $u_k = k/T$, $k = 0, \dots, T$. The estimated models are

$$f_{\hat{\boldsymbol{\theta}}}(u_k, \lambda) = \frac{\sigma(u_k)^2}{2\pi} \left| 1 + \sum_{j=1}^{q} a_j(u_k) e^{ij\lambda} \right|^{-2}, \qquad k = 0, \dots, T.$$

Then, we select the order of models in terms of $\hat{q}(u_k)$, $k = 0, \dots, T$ which minimizes

$$\mathrm{GIC}(q(u_k)) = N \int_{-\pi}^{\pi} K \left(\frac{\hat{g}_T(u_k, \lambda)}{f_{\hat{\boldsymbol{\theta}}}(u_k, \lambda)} \right) d\lambda + 4\pi\nu(h)(q+1),$$

where the parameters are $T = 400$, $N = 100$ and $M = 8$, and we employ the symmetric contrast type function $K = \frac{1}{2}\{x + x^{-1} + 2\}$, the Bartlett-Priestley window

$$W(\lambda) = \begin{cases} \dfrac{3}{4\pi}\{1 - (\lambda/\pi)^2\}, & |\lambda| \leq \pi, \\ 0, & |\lambda| \geq \pi, \end{cases}$$

and taper function

$$h^2(x) = \begin{cases} 6x(1-x), & 0 \leq x \leq 1, \\ 0, & \text{otherwise}. \end{cases}$$

The selected $\hat{q}(u_k)$ and minimal GIC values $\mathrm{GIC}(\hat{q}(u_k))$ are plotted in Figs. 1 and 2. From both figures we see that the model is not constant in time. Therefore, we can conclude time varying spectral models are desirable rather than stationary spectral models.

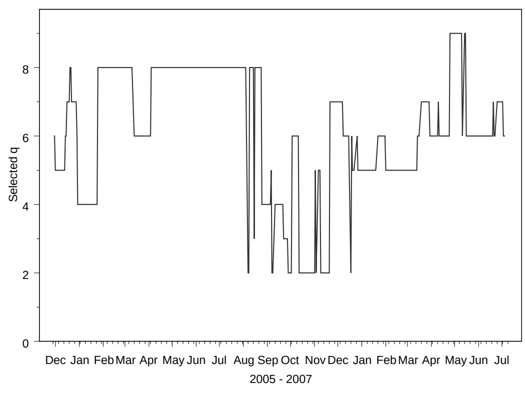

Figure 1. The selected $\hat{q}(u_k)$.

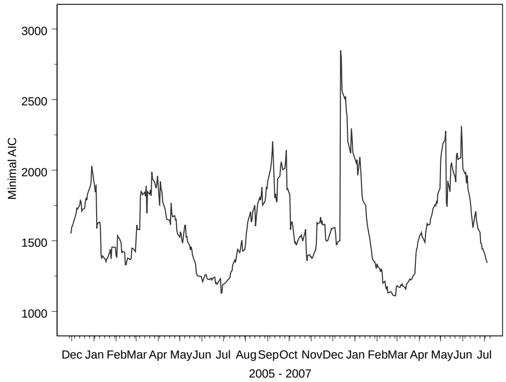

Figure 2. The minimal GIC values GIC($\hat{q}(u_k)$).

Appendix A

In this Appendix we briefly summarize some convergence results for time varying spectral estimates.

Suppose $\{X_{t,T}\}$ is locally stationary with mean function $\mu \equiv 0$ and transfer function A° with corresponding A whose derivatives $\frac{\partial^2}{\partial u^2}A$, $\frac{\partial^2}{\partial \lambda^2}$, $\frac{\partial^2}{\partial\partial u \partial \lambda}$ are continuous and uniformly bounded. Then, the nonparametric kernel type time varying spectral density estimator $\hat{g}_T$ in (3.1) and the data tapered local periodogram I_N in (3.2) possess the following asymptotic properties. First, by Corollary 4.1 of Dahlhaus and Giraitis (1998), we have the following.

Lemma 1. *Let ϕ_j be functions of bounded variation with $\phi_j(u,\lambda) = \phi_j(u,-\lambda)$. Then, under Assumptions* 4 *and* 6, *we have*

$$\sqrt{N}\left(\int_{-\pi}^{\pi} \phi_j(u,\lambda)[I_N(u,\lambda) - E\{I_N(u,\lambda)\}]d\lambda\right)'_{j=1,\ldots,q} \xrightarrow{d} \{v(h)\}^{1/2}(\xi(\phi_j))'_{j=1,\ldots,q},$$

where

$$v(h) = \frac{\int_0^1 h^4(x)dx}{\{\int_0^1 h^2(x)dx\}^2},$$

and $\xi(\phi_j)$, $j = 1,\ldots,q$ is a Gaussian vector with zero mean and covariance matrix

$$\begin{aligned}\{\Gamma_g(u)\}_{j,k} &= E\{\xi(\phi_j)\xi(\phi_k)\} \\ &= 4\pi\int_{-\pi}^{\pi} \phi_j(u,\lambda)\phi_k(u,\lambda)g^2(u,\lambda)d\lambda \\ &\quad + 2\pi\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \phi_j(u,\lambda)\phi_k(u,\mu)g(u,\lambda)g(u,\mu)\kappa_4(\lambda,-\lambda,\mu)d\lambda d\mu,\end{aligned} \tag{A.1}$$

where $\kappa_4(\lambda,-\lambda,\mu)$ is the 4th order cumulant spectrum of innovation $\{\varepsilon_t\}$.

Lemma 2. *Under Assumptions* 4–6, *we have*:

(i) $E|\hat{g}_T(u,\lambda) - g(u,\lambda)|^2 = O(\frac{M}{N})$, *uniformly in* u *and* λ.

(ii) $\sqrt{N}\int_{-\pi}^{\pi}\psi(u,\lambda)\{\hat{g}_T(u,\lambda) - I_N(u,\lambda)\}d\lambda = o_P(1)$, *for continuous* ψ.

(iii) $\sqrt{N}\int_{-\pi}^{\pi}\psi(u,\lambda)[E\{I_N(u,\lambda)\} - g(u,\lambda)]d\lambda = o(1)$, *for bounded* ψ.

(iv) $\max_{u,\lambda\in[0,1]\times[-\pi,\pi]} |\hat{g}_T(u,\lambda) - g(u,\lambda)| \xrightarrow{P} 0$.

PROOF. From Theorem 2.2 of Dahlhaus (1996c), we see that

$$E|\hat{g}_T(u,\lambda) - g(u,\lambda)|^2 = O\left(\frac{M}{N}\right) + O\left(\left(M^{-2} + \frac{N^2}{T^2}\right)^2\right) = O\left(\frac{M}{N}\right)$$

and

$$E\{I_N(u,\lambda)\} - g(u,\lambda) = O\left(\frac{N^2}{T^2} + \frac{\log N}{N}\right),$$

which imply the assertion (i) and (iii). Putting

$$\begin{aligned}\sqrt{N}&\int_{-\pi}^{\pi}\psi(u,\lambda)\{\hat{g}_T(u,\lambda) - I_N(u,\lambda)\}d\lambda\\ &= \sqrt{N}\int_{-\pi}^{\pi}\psi(u,\lambda)\{\hat{g}_T(u,\lambda) - g(u,\lambda)\}d\lambda\\ &\quad - \sqrt{N}\int_{-\pi}^{\pi}\psi(u,\lambda)\{I_N(u,\lambda) - g(u,\lambda)\}d\lambda\\ &= L_N - J_N \qquad \text{(say)},\end{aligned}$$

by the same argument as in the proof of Theorem 1 of Taniguchi *et al.* (1996), we have

$$|L_N - J_N| = o_P(1),$$

which implies the assertion (ii). The assertion (iv) follows from (ii) of Assumption 5, since

$$\begin{aligned}\max_{u,\lambda\in[0,1]\times[-\pi,\pi]} &|\hat{g}_T(u,\lambda) - E\{\hat{g}_T(u,\lambda)\}|\\ &\le \sum_{j=-(N-1)}^{N-1}\left|\omega\left(\frac{j}{M}\right)\right| |\hat{c}_N(u,j) - E\{\hat{c}_N(u,j)\}|,\end{aligned}$$

where

$$\hat{c}_N(u,j) = \frac{1}{2\pi H_N}\sum_{s=0}^{N-1} h\left(\frac{s}{N}\right) h\left(\frac{s+j}{N}\right) X_{[uT]-N/2+s+1,T}X_{[uT]-N/2+s+j+1,T}$$

and from Theorem 4.1 of Dahlhaus and Giraitis (1998), we have

$$E|N^{1/2}[\hat{c}_N(u,j) - E\{\hat{c}_N(u,j)\}]| \le C. \quad \square$$

LEMMA 3. *Under Assumptions* 1, 2, 4–6, *we have*:

(i) $S_{\hat{g}_T}(u) \xrightarrow{P} S_g(u)$.

(ii) *There exists an integrable function $m(u,\lambda)$ with respect to $\lambda \in [-\pi,\pi]$ such that, with*

$$\hat{R}_T = \int_{-\pi}^{\pi} m(u,\lambda)|\hat{g}_T(u,\lambda) - g(u,\lambda)|^2 d\lambda,$$

the relation

$$S_{\hat{g}_T}(u) - S_g(u) + D_g(u)^{-1}\left(\int_{-\pi}^{\pi} \phi_j(u,\lambda)\{\hat{g}_T(u,\lambda) - g(u,\lambda)\}d\lambda\right)'_{j=1,\dots,q}$$
$$= O_P(\hat{R}_T)$$

holds.

(iii) $\hat{R}_T = O_P(\frac{M}{N})$.

Proof. By (i) of Theorem 6.2.3 of Taniguchi and Kakizawa (2000), (iv) of Lemma 2 implies the assertion (i). From the definition of $S_{\hat{g}_T}(u)$ and $S_g(u)$ we obtain

$$\begin{aligned}\mathbf{0} &= \frac{\partial}{\partial\boldsymbol{\theta}} D(\boldsymbol{\theta}, \hat{g}_T, u) \mid_{\boldsymbol{\theta}=S_{\hat{g}_T}(u)} \\ &= \frac{\partial}{\partial\boldsymbol{\theta}} D(\boldsymbol{\theta}, \hat{g}_T, u) \mid_{\boldsymbol{\theta}=S_g(u)} \\ &\quad + \left\{\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} D(\boldsymbol{\theta}, \hat{g}_T, u) \mid_{\boldsymbol{\theta}=S_g(u)}\right\}\{S_{\hat{g}_T}(u) - S_g(u)\} \\ &\quad + O_P(|S_{\hat{g}_T}(u) - S_g(u)|^2),\end{aligned}$$

$$\begin{aligned}\frac{\partial}{\partial\theta_j} D(\boldsymbol{\theta}, \hat{g}_T, u) \mid_{\boldsymbol{\theta}=S_g(u)} = &\int_{-\pi}^{\pi} K_j^{(1)}(g(u,\lambda), u, \lambda)\{\hat{g}_T(u,\lambda) - g(u,\lambda)\}d\lambda \\ &+ O_P\left(\int_{-\pi}^{\pi} m_j(u,\lambda)|\hat{g}_T(u,\lambda) - g(u,\lambda)|^2 d\lambda\right)\end{aligned}$$

and

$$\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} D(\boldsymbol{\theta}, \hat{g}_T, u) \mid_{\boldsymbol{\theta}=S_g(u)} = D_g(u) + O_P\{(\hat{R}'_T)^{1/2}\},$$

where $m_j(u,\lambda)$ is some integrable function with respect to $\lambda \in [-\pi,\pi]$ and

$$\hat{R}'_T = \int_{-\pi}^{\pi} |\hat{g}_T(u,\lambda) - g(u,\lambda)|^2 d\lambda.$$

By the same argument as in the proof of Theorem 6.1.2 of Taniguchi and Kakizawa (2000), the assertions (ii) hold. The assertion (iii) follows from (i) of Lemma 2. □

Acknowledgements

The authors would like to express their sincere thanks to a referee for his/her helpful comments and advices.

References

Dahlhaus, R. (1996a). Maximum likelihood estimation and model selection for locally stationary processes, *J. Nonparametr. Statist.*, **6**, 171–191.

Dahlhaus, R. (1996b). On the Kullback-Leibler information divergence of locally stationary processes, *Stochastic Process. Appl.*, **62**, 139–168.

Dahlhaus, R. (1996c). Asymptotic statistical inference for nonstationary processes with evolutionary spectra, *Athens Conference on Applied Probability and Time Series Analysis*, Vol. II, Lecture Notes in Statist., 115, pp. 145–159, Springer, New York.

Dahlhaus, R. and Giraitis, L. (1998). On the optimal segment length for parameter estimates for locally stationary time series, *J. Time Ser. Anal.*, **19**, 629–655.

Dahlhaus, R. and Wefelmeyer, W. (1996). Asymptotically optimal estimation in misspecified time series models, *Ann. Statist.*, **24**, 952–974.

Priestley, M. B. (1981). *Spectral Analysis and Time Series*, Vol. II, Academic Press, London.

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

Taniguchi, M., Puri, M. L. and Kondo, M. (1996). Nonparametric approach for non-Gaussian vector stationary processes, *J. Multivariate Anal.*, **56**, 259–283.

Van Bellegem, S. and Dahlhaus, R. (2006). Semiparametric estimation by model selection for locally stationary processes, *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **68**, 721–746.