On the asymptotic inference for locally stationary processes

新潟大学・理学部   蛭川   潤一 (Junichi Hirukawa)
Faculty of Science,
Niigata University

1   Introduction

Time series analysis has been developed under stationarity. However, the assumption of stationarity is insufficient to describe the real time series data. Many empirical studies show that most of time series data such as financial and biological time series exhibit nonstationary behavior.

Random walk process

$$(1) \qquad Y_t = \sum_{j=1}^{t} u_j, \quad \left(\{u_j\} \sim i.i.d. \ (0, \sigma^2) \right),$$

is the most fundamental nonstationary process. This process will be reasonable for economic indices in which we suppose a value at the present time t is represented by the sum of random shocks over $t = 1, 2, \ldots, t$. We can rewrite the equation (1) as

$$(2) \qquad Y_t = Y_{t-1} + u_t, \quad (Y_0 = 0),$$

which corresponds to AR(1) process

$$(3) \qquad Y_t = bY_{t-1} + u_t,$$

with $b = 1$. In this case we say that AR model (3) has a unit root. It is known that the local asymptotic normality (LAN) does not hold for models including the unit root.

We now turn to discuss nonstationary models which are regular in the sense that they satisfy the LAN property. At the sight of actual time series data we often find that they seem locally stationary and contain several changes of the structure in its entirety. To meet this Dahlhaus (1996a, b, c) introduced an important class of nonstationary processes with rigorous asymptotic framework, called locally stationary processes. Locally stationary processes include stationary processes as a special case $g(u, \lambda) \equiv g(\lambda)$. Here we explain the extension of techniques in stationary models to locally stationary models.

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*

$$(4) \qquad X_{t,T} = \int_{-\pi}^{\pi} \exp{(i\lambda t)} A_{t,T}^\circ(\lambda) d\xi(\lambda),$$

where

(i) $\xi(\lambda)$ *is a stochastic spectral measure on* $[-\pi, \pi]$ *of uncorrelated sequence* ε_t, *i.e.,*

$$(5) \qquad \varepsilon_t = \int_{-\pi}^{\pi} \exp(i\lambda t) d\xi(\lambda).$$

(ii) $\exists K$ & $\exists A : [0,1] \times \mathbf{R} \to \mathbf{C}$ *with* $A(u, -\lambda) = \overline{A(u, \lambda)}$ *and*

$$(6) \qquad \sup_{t, \lambda} \left| A_{t,T}^\circ(\lambda) - A\left(\frac{t}{T}, \lambda\right) \right| \le KT^{-1}$$

for all T. $A(u, \lambda)$ *is assumed to be continuous in* u *and*

$$(7) \qquad g(u, \lambda) := |A(u, \lambda)|^2$$

is called the time varying spectral density of the process.

One major difficulty in developing the general nonstationary theory is the problem of asymptotics. The asymptotic theory is needed since investigation of e.g., the MLE for a fixed sample size is too much complicated. The classical asymptotic theory with assumption that more and more observations of future become available does not make sense since future observations of general nonstationary processes do not necessarily contain any information on the structure at present.

The Wigner-Ville spectrum for fixed T is

$$(8) \qquad g_T(u, \lambda) := \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} \text{Cov}\left(X_{[uT-s/2],T}, X_{[uT+s/2],T}\right) e^{(-i\lambda s)},$$

where $X_{s,T}$ is defined by (4) (with $A_{t,T}^\circ(\lambda) = A(0, \lambda)$ for $t < 1$ and $A_{t,T}^\circ(\lambda) = A(1, \lambda)$ for $t > T$). If $X_{t,T}$ is locally stationary and $A(u, \lambda)$ is uniform Lipschitz continuous in both components with index $\alpha > \frac{1}{2}$ then we have for all $u \in (0, 1)$

$$(9) \qquad \int_{-\pi}^{\pi} |g_T(u, \lambda) - g(u, \lambda)|^2 d\lambda = o(1).$$

This result is important because it shows the uniqueness of the evolutionary (time varying) spectral density $g(u, \lambda)$. The uniqueness property of spectral density is a major difference between locally stationary processes and other approaches to modeling nonstationary time series, such as oscillatory processes (Priestley, 1981). In contrast with Priestley's definition, locally stationary time series are doubly indexed and their time varying spectral density is rescaled on the time interval $[0, 1]$. This is the key point that allows us to make use of asymptotic considerations. Only the values of $X_{t,T}$ in the interval

$$(10) \qquad \frac{t}{T} \in \left[u - \frac{n}{T}, u + \frac{n}{T}\right]$$

contribute to $g(u, \lambda)$. Since the length of this interval tends to zero and $A(u, \lambda)$ is smooth, the observations become "asymptotically stationary" on this interval which leads to the above uniqueness. When people speak of the spectrum at a time point t_0 of a nonstationary process $X_1, \ldots, X_T$, since the process is nonstationary, only a few points around t_0 may have the same spectral structure. It is clear that the pobability structure of these few points does not specify a spectral density uniquely. Our approach says that $g(u, \lambda) = |A(u, \lambda)|^2$ is the spectral density if one had infinitely many observations of the same kind at a fixed time point.

2 Time Series Locally Stationary Regression Models

When we consider problems of statistical inference on trend functions linear in an unknown vector coefficient β, the BLUE contains the residual covariance matrix Σ_T which is usually unknown. On the other hand, the LSE is feasible regardless of the knowledge of the residuals part. It is of particular interest to study the efficiency of the LSE relative to the BLUE.

Let $\mathbf{y}^T = (Y_{1,T}, \ldots, Y_{T,T})'$ have mean vector and covariance matrix

$$(11) \qquad E\left(\mathbf{y}^T\right) = \mathbf{Z}\beta,$$

$$(12) \qquad E\left(\mathbf{y}^T - \mathbf{Z}\beta\right)\left(\mathbf{y}^T - \mathbf{Z}\beta\right)^* = \Sigma_T,$$

where $\mathbf{Z} = (z_{jt})'$ is a $T \times q$ known matrix and has (t, j)-th component $(\mathbf{Z})_{t,j} = z_{jt} = z_j(t)$. Let

$$(13) \qquad a_{jk}^T(h) = \sum_{t=1}^{T-h} z_j(t+h)z_k(t) \quad h = 0, 1, \ldots.$$

We make the following assumption (which we shall call Grenander's conditions):

$$(14) \qquad \lim_{T \to \infty} a_{jj}^T(0) = \infty,$$

$$(15) \qquad \lim_{T \to \infty} \frac{z_j(T)^2}{a_{jj}^T(0)} = 0, \quad j = 1, \ldots, q,$$

$$(16) \qquad \lim_{T \to \infty} \frac{a_{jk}^T(h)}{\sqrt{a_{jj}^T(0)a_{kk}^T(0)}} = \rho_{jk}(h), \quad j, k = 1, \ldots, q.$$

Let $\mathbf{R}(h) = [\rho_{jk}(h)]$ and $\mathbf{R}(0)$ be nonsingular. It follows that there is a Hermitian matrix function $\mathbf{M}(\lambda)$ with positive semidefinite increments such that

$$(17) \qquad \mathbf{R}(h) = \int_{-\pi}^{\pi} e^{ih\lambda} d\mathbf{M}(\lambda).$$

Define

$$(18) \quad \mathbf{D}_T = \begin{pmatrix} \left\{a_{11}^{(T)}(0)\right\}^{1/2} & 0 & \cdots & & 0 \\ 0 & \ddots & \left\{a_{22}^{(T)}(0)\right\}^{1/2} & & \vdots \\ \vdots & \ddots & \ddots & & 0 \\ 0 & \cdots & & 0 & \left\{a_{qq}^{(T)}(0)\right\}^{1/2} \end{pmatrix}.$$

Then

$$(19) \quad \lim_{T\to\infty} \mathbf{D}_T^{-1}\mathbf{Z}^*\mathbf{Z}\mathbf{D}_T^{-1} = \mathbf{R}(0).$$

Here β is a q-component vector of coefficients to be estimated from an observation on $\mathbf{y}^T$. The BLUE and LSE of β are given by

$$(20) \quad \mathbf{b}_{\text{BLUE}} = \left(\mathbf{Z}^*\boldsymbol{\Sigma}_T^{-1}\mathbf{Z}\right)^{-1}\mathbf{Z}^*\boldsymbol{\Sigma}_T^{-1}\mathbf{y},$$

$$(21) \quad \mathbf{b}_{\text{LSE}} = \left(\mathbf{Z}^*\mathbf{Z}\right)^{-1}\mathbf{Z}^*\mathbf{y}$$

and the covariance matrices are

$$(22) \quad E\left(\mathbf{b}_{\text{BLUE}} - \beta\right)\left(\mathbf{b}_{\text{BLUE}} - \beta\right)^* = \left(\mathbf{Z}^*\boldsymbol{\Sigma}_T^{-1}\mathbf{Z}\right)^{-1},$$

$$(23) \quad E\left(\mathbf{b}_{\text{LSE}} - \beta\right)\left(\mathbf{b}_{\text{LSE}} - \beta\right)^* = \left(\mathbf{Z}^*\mathbf{Z}\right)^{-1}\mathbf{Z}^*\boldsymbol{\Sigma}_T\mathbf{Z}\left(\mathbf{Z}^*\mathbf{Z}\right)^{-1}.$$

We discuss under what conditions the two covariance matrices are asymptotically equivalent in the sense that

$$(24) \quad \lim_{T\to\infty} \mathbf{D}_T E\left(\mathbf{b}_{\text{LSE}} - \beta\right)\left(\mathbf{b}_{\text{LSE}} - \beta\right)^*\mathbf{D}_T$$
$$= \lim_{T\to\infty} \mathbf{D}_T E\left(\mathbf{b}_{\text{BLUE}} - \beta\right)\left(\mathbf{b}_{\text{BLUE}} - \beta\right)^*\mathbf{D}_T.$$

Lemma 1. *A necessary and sufficient condition that the LSE be asymptotically efficient for all stationary processes with continuous, positive spectral densities is that* $M(\lambda)$ *increases at not more than q values of λ, $0 \le \lambda \le \pi$, and the sum of the ranks of the increases in* $M(\lambda)$ *is q.*

Now we turn to discuss locally stationary disturbance case. To simplify we restrict ourselves to the trend functions of the form

$$(25) \quad (\mathbf{Z})_{tj} = z_j(t) = t^j \sum_{l=1}^{r_j} \alpha_j^{(l)} e^{i\theta_j^{(l)}t},$$

$$\alpha_j^{(l)} \ne 0, \quad -\pi \le \theta_j^{(1)} < \cdots < \theta_j^{(r_j)} < \pi, \quad j = 0.\ldots,p.$$

$\boldsymbol{\Sigma}_T$ is the covariance matrix of locally stationary process $\mathbf{x}^T = (X_{1,T}, \ldots, X_{T,T})'$ given by

$$(26) \quad \boldsymbol{\Sigma}_T = \left\{\int_{-\pi}^{\pi} A_{s,T}^\circ(\lambda)A_{t,T}^\circ(-\lambda)e^{i(s-t)\lambda}d\lambda\right\}_{s,t=1,\ldots,T}.$$

Theorem 1.

(27)
$$\lim_{T\to\infty} \mathbf{D}_T E \left(\mathbf{b}_{\text{LSE}} - \beta\right) \left(\mathbf{b}_{\text{LSE}} - \beta\right)^* \mathbf{D}_T$$
$$= 2\pi \mathbf{R}(0)^{-1} \mathbf{L} \mathbf{R}(0)^{-1},$$

where

(28)
$$\mathbf{L}_{jk} = \left\{ \frac{(2j+1)(2k+1)}{\left(\sum_{l=1}^{r_j} \left|\alpha_j^{(l)}\right|^2\right)\left(\sum_{l=1}^{r_k} \left|\alpha_k^{(l)}\right|^2\right)} \right\}^{1/2}$$
$$\sum_{l_1=1}^{r_j} \sum_{l_2=1}^{r_k} \overline{\alpha_j^{(l_1)}} \alpha_k^{(l_2)} \delta \left(\theta_k^{(l_2)} - \theta_j^{(l_1)}\right) \int_0^1 u^{j+k} f\left(u, \theta_j^{(l_1)}\right) du.$$

Theorem 2.

(29)
$$\lim_{T\to\infty} \mathbf{D}_T E \left(\mathbf{b}_{\text{BLUE}} - \beta\right) \left(\mathbf{b}_{\text{BLUE}} - \beta\right)^* \mathbf{D}_T$$
$$= 2\pi \mathbf{B}^{-1},$$

where

(30)
$$\mathbf{B}_{jk} = \left\{ \frac{(2j+1)(2k+1)}{\left(\sum_{l=1}^{r_j} \left|\alpha_j^{(l)}\right|^2\right)\left(\sum_{l=1}^{r_k} \left|\alpha_k^{(l)}\right|^2\right)} \right\}^{1/2}$$
$$\sum_{l_1=1}^{r_j} \sum_{l_2=1}^{r_k} \overline{\alpha_j^{(l_1)}} \alpha_k^{(l_2)} \delta \left(\theta_k^{(l_2)} - \theta_j^{(l_1)}\right) \int_0^1 u^{j+k} f^{-1}\left(u, \theta_j^{(l_1)}\right) du.$$

Theorem 3.

(31)
$$\lim_{T\to\infty} \mathbf{D}_T E \left(\mathbf{b}_{\text{BLUE}} - \beta\right) \left(\mathbf{b}_{\text{BLUE}} - \beta\right)^* \mathbf{D}_T$$
$$\leq \lim_{T\to\infty} \mathbf{D}_T E \left(\mathbf{b}_{\text{LSE}} - \beta\right) \left(\mathbf{b}_{\text{LSE}} - \beta\right)^* \mathbf{D}_T,$$

where the equality holds if and only if there exists $(p+1) \times (p+1)$ *matrix* $\mathbf{V}$ *which does not depend on* u *and* λ *and satisfies*

(32)
$$\chi(u,\lambda) + \mathbf{V}\phi(u,\lambda) = 0, \quad a.e.,$$

where

(33)
$$\mathbf{L} = \int_0^1 \int_{-\pi}^{\pi} \phi(u,\lambda) \phi(u,\lambda)^* \, d\lambda du$$

and

(34)
$$\mathbf{B} = \int_0^1 \int_{-\pi}^{\pi} \chi(u,\lambda) \chi(u,\lambda)^* \, d\lambda du.$$

Example 1. *We now assume that* $A\left(u, \theta_j^{(l_j)}\right) \equiv A\left(\theta_j^{(l_j)}\right)$, *for all* $l_j = 1, \ldots, r_j$, $j = 0, \ldots p$. *In such case, the results coincide with those of the stationary case, since the asymptotic variances of LSE and BLUE are independent of* u. *Indeed, if we take* $v_{jj} = f^{-1}\left(\theta_j^{(1)}\right) = \cdots = f^{-1}\left(\theta_j^{(r_j)}\right)$, $j = 0, \ldots, p$ *and* $v_{jk} = 0$ *if* $j \neq k$, *then the equation (32) holds. In the case* r_j *is at most 2 and* $-\theta_j^{(1)} = \theta_j^{(2)}$, *we can obtain such a matrix* **V** *for arbitrary* $f(\lambda)$.

Example 2. *Next, we assume that* $A(u, \lambda) \equiv A(\lambda) e^{iu\lambda}$ *with* $A(-\lambda) = \overline{A(\lambda)}$. *If we take* $v_{jj} = \left|A\left(\theta_j^{(1)}\right)\right|^{-2} = \cdots = \left|A\left(\theta_j^{(r_j)}\right)\right|^{-2}$, $j = 0, \ldots, p$ *and* $v_{jk} = 0$ *if* $j \neq k$, *then the equation (32) holds. In the case* r_j *is at most 2 and* $-\theta_j^{(1)} = \theta_j^{(2)}$, *we can obtain such a matrix* **V**.

3 Model selection for locally stationary processes

In the actual statistical analysis the order of proposed parametric models $q = \dim \Theta$ must be inferred. The best known rule for determining the true value of q is probably Akaike's information criterion (AIC). In the i.i.d. case Takeuchi (1976) and Konishi and Kitagawa (1996) gave derivation of a generalized AIC, which includes the original AIC as a special case.

To evaluate models, we usually assume structures of stochastic processes are specified by some functions. As examples of such function we can take the probability distribution function $G(x)$ for i.i.d. case, the trend function $\mu(u)$ for regression model, the spectral density function $g(\lambda)$ for stationary process and the dynamic system function $\mathbf{F}(X_{t-1}, \ldots, X_{t-p})$ for nonlinear models. The structure of locally stationary processes is specified by the smooth function, namely time varying spectral density function $g(u, \lambda)$. Dahlhaus (1996a) and Bellegem and Dahlhaus (2006) proposed model selection criterion for locally stationary processes based on Gaussian Kullback-Leibler information measure. We introduce model selection criteria for locally stationary processes based on functionals of a time varying spectral density.

Many important quantities in time series are often expressed as functionals of spectra. For a linear functional, a natural idea of constructing an estimator is to replace an unknown time varying spectral density by the local periodogram based on the data. The functional of interest is, however, not always linear with respect to time varying spectral density (Dahlhasu and Weffelmeyer (1996) and Taniguchi and Kakizawa (2000)). In these cases we use nonparametric kernel type time varying spectral density estimator instead of local periodogram to avoid the inconsistency. Let $\{X_{t,T}\}$ $(t = 1, \ldots, T; T \geq 1)$ be locally stationary processes with mean zero and time varying spectral density $g(u, \lambda)$. Suppose a stretch $\mathbf{X}_T = \{X_{2-N/2,T}, \ldots, X_{1,T}, \ldots, X_{T,T}, \ldots, X_{T+N/2,T}\}$ is available from this locally stationary processes. We want to fit a class of time varying spectral models $\mathcal{P} = \{f_\theta(u, \lambda) : \theta \in \Theta \subset \mathbf{R}^q\}$ without assuming that the true time

varying spectral density $g(u, \lambda)$ belongs to $\mathcal{P}$.

Here we consider the local distance function at time u of the form

$$(35) \qquad D(\theta, g, u) = \int_{-\pi}^{\pi} K\{\theta, g(u, \lambda), u, \lambda\} d\lambda$$

associated with function $K(\cdot, \cdot, \cdot, \cdot)$ defined below. We set a functional S by the requirement that

$$(36) \qquad D\{S_g(u), g, u\} = \min_{\theta \in \Theta} D(\theta, g, u).$$

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 has a unique minimum zero at $x = 1$, such as

$$(37) \qquad H_1(x) = -\log(x) + x - 1$$

$$(38) \qquad H_2(x) = \frac{1}{\alpha(1 - \alpha)}\{\log(\alpha x + 1 - \alpha) - \alpha \log(x)\},$$

$$0 < \alpha < 1$$

$$(39) \qquad H_3(x) = \frac{1}{2}(x - 1)^2.$$

Then, we define the contrast type function

$$(40) \qquad K(\theta, z(u, \lambda), u, \lambda) = H\{z(u, \lambda)f_\theta(u, \lambda)^{-1}\}$$

or

$$(41) \qquad K(\theta, z(u, \lambda), u, \lambda) = H\{f_\theta(u, \lambda)z(u, \lambda)^{-1}\}$$

or

$$(42) \qquad K(\theta, z(u, \lambda), u, \lambda)$$
$$= \frac{1}{2}[H\{z(u, \lambda)f_\theta(u, \lambda)^{-1}\} + H\{f_\theta(u, \lambda)z(u, \lambda)^{-1}\}].$$

Weighted squared function:

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

$$(43) \qquad K(\theta, z(u, \lambda), u, \lambda) = \frac{1}{2}[\psi(u, \lambda)\{f_\theta(u, \lambda) - z(u, \lambda)\}]^2.$$

The estimator is naturally defined by $S_{\hat{g}_T}(u)$, where $\hat{g}_T$ is a nonparametric kernel type time varying spectral density estimator

$$(44) \qquad \hat{g}_T(u, \lambda) = \int_{-\pi}^{\pi} W_M(\lambda - \mu) I_N(u, \mu) d\mu$$

and

$$(45) \qquad I_N(u, \lambda) = \frac{1}{2\pi H_N} \left| \sum_{s=0}^{N-1} h\left(\frac{s}{N}\right) X_{[uT]-N/2+s+1,T} \exp(i\lambda s) \right|^2$$

is the data tapered local periodogram at time u.

Here $W_M(\omega) = M \sum_{\nu=-\infty}^{\infty} W\{M(\omega + 2\pi\nu)\}$ is the weight function, $h : [0, 1] \to \mathbf{R}$ is a data taper and $H_N = \sum_{s=0}^{N-1} h^2(s/N) \sim N \int_0^1 h^2(x)dx$. The parameters $M = M(T)$ and $N = N(T)$, $M \ll N \ll T$ depend on T in such a way that

$$(46) \qquad \frac{M}{N^{1/2}} + \frac{N^{1/4}}{M} + \frac{N^5}{T^4} \to 0.$$

Furthermore, we define

$$(47) \qquad D_g(u) = \int_{-\pi}^{\pi} \frac{\partial^2}{\partial\theta\partial\theta'} K\{\theta, g(u, \lambda), u, \lambda\}\big|_{\theta=S_g(u)} d\lambda$$

and assume $D_g(u)$ is nonsingular for all u. From Corollary 4.1 of Dahlhaus and Giraitis (1998), we see that

Lemma 2.

$$(48) \qquad \sqrt{N} \left(\int_{-\pi}^{\pi} \phi_j(u, \lambda) \left[I_N(u, \lambda) - E\{I_N(u, \lambda)\} \right] \right)'_{j=1,\ldots,q}$$

$$\xrightarrow{d} \{\nu(h)\}^{1/2} (\xi(\phi_j))'_{j=1,\ldots,q},$$

where

$$(49) \qquad \nu(h) = \frac{\int_0^1 h^4(x)dx}{\cdot\{\int_0^1 h^2(x)dx\}^2},$$

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

$$(50) \qquad \{\Gamma_g(u)\}_{j,k} = E\{\xi(\phi_j)\xi(\phi_k)\}$$

$$= 4\pi \int_{-\pi}^{\pi} \phi_j(u, \lambda)\phi_k(u, \lambda)g^2(u, \lambda)d\lambda$$

$$+ 2\pi \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \phi_j(u, \lambda)\phi_k(u, \mu)g(u, \lambda)g(u, \mu)\kappa_4(\lambda, -\lambda, \mu)d\lambda d\mu.$$

Therefore, we have

Theorem 4.

(51)
$$\sqrt{N}\{S_{\hat{g}_T}(u) - S_g(u)\}$$

$$\xrightarrow{d} N(0, \nu(h)D_g(u)^{-1}\Gamma_g(u)D_g(u)^{-1}).$$

Recall that we fit a class of parametric models $\mathcal{P} = \{f_\theta(u, \lambda) : \theta \in \Theta \subset \mathbf{R}^q\}$ to g by use of measure of local disparity at time u, $D(\theta, g, u)$, and we estimate θ by the value $S_{\hat{g}_T}(u)$ which minimizes $D(\theta, \hat{g}_T, u)$, 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 $E_{X_T}\{D(S_{\hat{g}_T}(u), g, u)\}$. 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 $D(S_{\hat{g}_T}(u), \hat{g}_T, u)$. Ordinary this provides underestimate of $D(S_{\hat{g}_T}(u), g, u)$. Writing bias as

(52)
$$b_g(u) = E_{X_T}\{D(S_{\hat{g}_T}(u), \hat{g}_T, u) - D(S_{\hat{g}_T}(u), g, u)\},$$

we define generalised information criterion as

(53)
$$D(S_{\hat{g}_T}(u), \hat{g}_T, u) - b_{\hat{g}_T}(u).$$

We define the pseudo true value of θ in terms of the functional $S_g(u)$ by the requirement $D(S_g(u), g, u) = \min_{\theta \in \Theta} D(\theta, g, u)$. We see that

(54)
$$-b_g(u) \approx E_{X_T}\{(S_{\hat{g}_T}(u) - S_g(u))'D_g(u)(S_{\hat{g}_T}(u) - S_g(u))\}$$

$$\approx \frac{v(h)}{N}\mathrm{tr}\{D_g(u)^{-1}\Gamma_g(u)\}.$$

Here $D_g(u)$ and $\Gamma_g(u)$ depend on g, we replace g by nonparametric time varying spectral density estimator $\hat{g}_T$. Then we estimate $E_{X_T}\{D(S_{\hat{g}_T}(u), g, u)\}$ by $G_N(q) = D(S_{\hat{g}_T}(u), \hat{g}_T, u) + N^{-1}\nu(h)\mathrm{tr}\{D_{\hat{g}_T}(u)^{-1}\Gamma_{\hat{g}_T}(u)\}$. Multiplying $G_N(q)$ by N we call

(55)
$$\mathrm{GIC}(q) = ND(S_{\hat{g}_T}(u), \hat{g}_T, u) + \nu(h)\mathrm{tr}\{D_{\hat{g}_T}(u)^{-1}\Gamma_{\hat{g}_T}(u)\}$$

a generalized information criterion.

In particular, for the contrast type $K(\cdot, \cdot, \cdot, \cdot)$, if the true time varying spectral density is $f_\theta(u, \lambda) \in \mathcal{P}$, where θ lies in the interior of Θ and the process is Gaussian, or the parameter θ is innovation-free, then the generalized information criterion $\mathrm{GIC}(q)$ becomes Akaike's information criterion

(56)
$$\mathrm{AIC}(q) = ND(S_{\hat{g}_T}(u), \hat{g}_T, u) + 4\pi H^{(2)}(1)\nu(h)q.$$

3.1 Numerical examples

Here we give the concrete examples of the quantity $v(h)\mathrm{tr}\{D_g(u)^{-1}\Gamma_g(u)\}$ in (54), which is interpreted as the penalized term in GIC. Here we take the contrast type function (40) 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 $f_\theta(u, \lambda) = f_\theta(\lambda)$ to locally stationary process which has the true time varying spectral density $g(u, \lambda)$. Here we assume that θ is innovation free (the innovation ε_t's are i.i.d. and θ satisfies $\frac{\partial}{\partial \theta} \int_{-\pi}^{\pi} \log f_\theta(\lambda) d\lambda = 0$). Then, θ satisfies

$$(57) \quad \frac{\partial^2}{\partial \theta \partial \theta'} \int_{-\pi}^{\pi} \log f_\theta(\lambda) d\lambda$$

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

Therefore, we have

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

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

If the true time varying spectral density is of the form $g(u, \lambda) = f_\theta(\lambda) \alpha(u)$, then from (57) the penalized term becomes

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

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

$$D_g(u) = \begin{pmatrix} D_g(u)_{(11)} & 0 \\ 0 & D_g(u)_{(22)} \end{pmatrix}$$

and

$$\Gamma_g(u) = \begin{pmatrix} \Gamma_g(u)_{(11)} & * \\ * & \Gamma_g(u)_{(22)} \end{pmatrix},$$

where

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

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

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

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

$$+ 8\pi^3 \kappa_4 \left[f_{2,\theta_{(q)}}(u)^{-2} \left\{ \frac{\partial\theta_{(q)} f_{2,\theta_{(q)}}(u)}{\partial\theta_{(q)}} \right\}^2 \right]_{\theta=S_g(u)}$$

and

$$\mathrm{tr}\left\{ v(h) D_g(u)^{-1} \Gamma_g(u) \right\}$$
$$= \mathrm{tr}\left\{ v(h) D_g(u)_{(11)}^{-1} \Gamma_g(u)_{(11)} \right\} + \mathrm{tr}\left\{ v(h) D_g(u)_{(22)}^{-1} \Gamma_g(u)_{(22)} \right\}.$$

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

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

3.2 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 $\left\{ X_{1-N/2,T}, \ldots, X_{0,T}, \ldots, X_{T,T}, \ldots, \right.$ $\left. X_{T+N/2,T} \right\}$ of 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 $\left\{ X_{k+1-N/2,T}, \ldots, X_{k+N/2,T} \right\}$ in terms of Yule-Walker equations for each $u_k = k/T$, $k = 0, \ldots, T$. The estimated models are

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

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

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

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

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

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 $\text{GIC}(\hat{q}(u_k))$ are plotted in Figures 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.
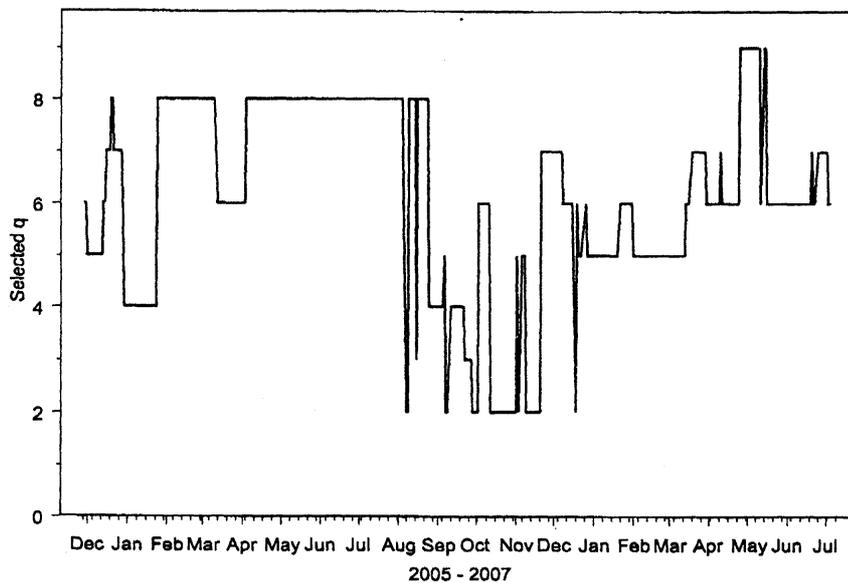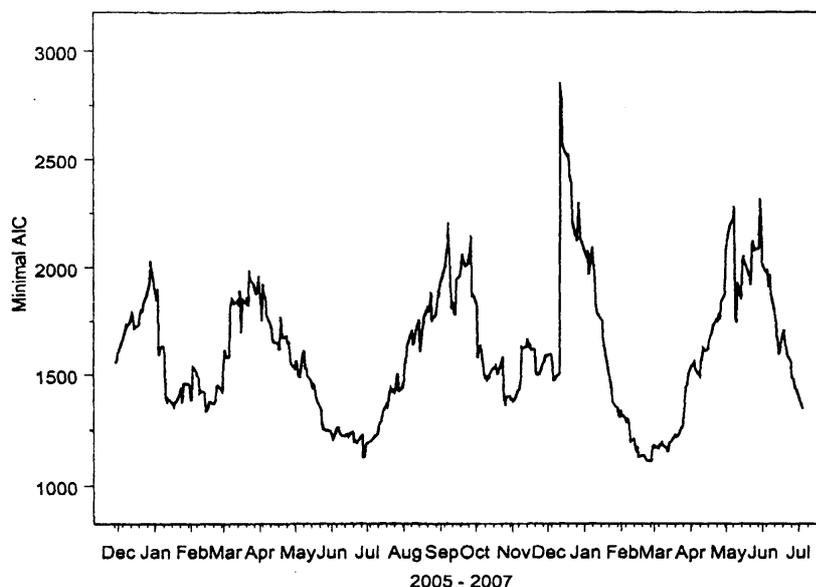

Figure 1: The selected $\hat{q}(u_k)$.

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

References

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

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

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

[4]   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.

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

[6]   Konishi, S. and Kitagawa, G. (1996). Generalised information criteria in model selection. *Biometrika*, **83**, 875–890.

[7]   Priestley, M. B. (1981). *Spectral analysis and time series*, Vol. II, (London: Academic Press).

[8] Takeuchi, K. (1976). Distribution of information statistics and criteria for adequacy of models. *Suri-Kagaku (Math. Sci.)*, **153**, 12–18, (in Japanese).

[9] Taniguchi, M. and Kakizawa, Y. (2000). *Asymptotic theory of statistical inference for time series*, (New York: Springer-Verlag).

[10] 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.