ELSEVIER Stochastic Processes and their Applications 62 (1996) 139–168

stochastic processes and their applications

On the Kullback–Leibler information divergence of locally stationary processes[1]

R. Dahlhaus

Universität Heidelberg, Inst. für Angewandte Mathematik, Im Neuenheimer Feld 294, 69120 Heidelberg, Germany

Received March 1995; revised November 1995

Abstract

A class of processes with a time varying spectral representation is established. As an example we study time varying autoregressions. Several results on the asymptotic norm behaviour and trace behaviour of covariance matrices of such processes are derived. As a consequence we prove a Kolmogorov formula for the local prediction error and calculate the asymptotic Kullback–Leibler information divergence.

Keywords: Locally stationary processes; Evolutionary spectra; Kullback–Leibler divergence; time varying autoregressions

1. Introduction

There exists a large variety of statistical techniques for stationary processes (e.g. methods based on the spectrum or methods based on parametric models, such as ARMA models). These techniques are usually well investigated and therefore very often used in applications. Even in situations where it is obvious that a nonstationary model is more adequate, stationary models and techniques are used frequently (e.g. after removing trends or by looking at segments of the data). An example is the LPC (linear predictive coding) approach to signal processing where autoregressive models are fitted locally to the data (Thomson and de Souza, 1985). However, in the theoretical treatment of the estimates it is usually assumed that the data are coming from a stationary sequence.

In this paper we set up a more realistic framework for such considerations by assuming that the observed process has a time varying spectral representation similar to the one for stationary sequences. Such an approach was first suggested by Priestley (1965) (also Priestley (1981)). However, the approach of Priestley does not allow for rigorous local asymptotic considerations. This is important for handling in a satisfactory way the difficult expressions arising in the statistics for such processes. To

[1] This work was supported by the Deutsche Forschungsgemeinschaft.

SSDI 0304-4149(95)00090-9

overcome these problems we suggest in this paper to consider a triangular array of data. To clarify our view we give an example.

Suppose that we observe

$$X_t = g(t) X_{t-1} + \varepsilon_t \quad \text{with } \varepsilon_t \text{ iid N}(0, \sigma^2)$$

for $t = 1, \ldots, T$. Inference in this case means inference for the unknown function g on the grid $\{1, \ldots, T\}$. It is obvious that an asymptotic approach where $T \to \infty$ is not suitable for describing a statistical method since future "observations" of $g(t)$ do not necessarily contain any information on $g(t)$ on $\{1, \ldots, T\}$. On the other hand we need some kind of asymptotics which simplifies the situation (in order to compare e.g. least squares estimates with maximum likelihood estimates in a parametric model $g(t) = g_\theta(t)$).

Analogous to nonparametric regression it seems natural to set down the asymptotic theory in a way that we "observe" $g(t)$ on a finer grid (but on the same interval), i.e. that we observe the process

$$X_{t,T} = g\left(\frac{t}{T}\right) X_{t-1,T} + \varepsilon_t \quad \text{for } t = 1, \ldots, T \tag{1.1}$$

(where g is now rescaled to the interval $[0, 1)$].

To define a general class of nonstationary processes which includes the above example we may try to take the time varying spectral representation

$$X_{t,T} = \mu\left(\frac{t}{T}\right) + \int_{-\pi}^{\pi} \exp(\mathrm{i}\lambda t)\, A\left(\frac{t}{T}, \lambda\right) \mathrm{d}\xi(\lambda). \tag{1.2}$$

(similar to the analogous representation for stationary processes). However, it turns out that the Eq. (1.1) has not exactly but only approximately a solution of the form (1.2). We therefore only require that (1.2) holds approximately which leads to the definition of a locally stationary process given in Section 2.

Furthermore, we prove in Section 2 a uniqueness property of the spectral representation and define the time varying spectral density of the process. We also show that time varying ARMA models are locally stationary.

In Section 3 we discuss some properties of Gaussian locally stationary processes. In particular, we calculate the local prediction error and show that a local Kolmogorov formula holds (Theorem 3.2). Furthermore, we calculate the asymptotic Kullback–Leibler information divergence of two locally stationary sequences (Theorem 3.4) and the limit of the Fisher information matrix (Theorem 3.6).

To establish these results we need several properties on matrix norms and traces of covariance matrices of locally stationary processes. These properties are proved in Section 4. The significance of these results goes beyond this paper since they form e.g. the basis for a comprehensive treatment of the maximum likelihood estimator for such processes (Dahlhaus, 1996a).

2. Locally stationary processes

Definition 2.1. A sequence of stochastic processes $X_{t,T}$ $(t = 1, \ldots, T)$ is called locally stationary with transfer function A^0 and trend μ if there exists a representation

$$X_{t,T} = \mu\left(\frac{t}{T}\right) + \int_{-\pi}^{\pi} \exp(\mathrm{i}\lambda t)\, A_{t,T}^0(\lambda)\, \mathrm{d}\xi(\lambda) \tag{2.1}$$

where

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

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

where $\mathrm{cum}\{\cdots\}$ denotes the cumulant of kth order, $h_1 = 0$, $h_2(\lambda) = 1$, $|h_k(\lambda_1, \ldots, \lambda_{k-1})| \leqslant \mathrm{const}_k$ for all k and $\eta(\lambda) = \sum_{j=-\infty}^{\infty} \delta(\lambda + 2\pi j)$ is the period 2π extension of the Dirac delta function.

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

$$\sup_{t,\lambda} | A_{t,T}^0(\lambda) - A(t/T, \lambda)| \leqslant K T^{-1} \tag{2.2}$$

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

The smoothness of A in u guarantees that the process has locally a "stationary behaviour". Below we will require additional smoothness properties for A, namely differentiability in both components.

In the following we denote by s and t always time points in the interval $[1, T]$ while u and v are time points in the rescaled interval $[0, 1]$, i.e. $u = t/T$.

To give a simple example of a locally stationary process let Y_t be a stationary sequence with spectral representation

$$Y_t = \int_{-\pi}^{\pi} \exp(\mathrm{i}\lambda t)\, A(\lambda)\, \mathrm{d}\xi(\lambda)$$

and $\mu, \sigma: [0, 1] \to \mathbb{R}$ be continuous. Then

$$X_{t,T} := \mu\left(\frac{t}{T}\right) + \sigma\left(\frac{t}{T}\right) Y_t$$

is locally stationary with $A_{t,T}^0(\lambda) = A(t/T, \lambda) = \sigma(t/T) A(\lambda)$. If Y_t is an AR(2) process with (complex) roots close to the unit circle then Y_t shows a periodic behaviour and σ may be regarded as a time varying amplitude function of the process $X_{t,T}$. If T tends to infinity more and more cycles of the process with $u = t/T \in [u_0 - \varepsilon, u_0 + \varepsilon]$, i.e. with amplitude close to $\sigma(u_0)$, are observed.

Asymptotics of this kind have frequently been used e.g. in nonparametric regression where Y_t is iid and σ usually is assumed to be constant. For a similar example in a time series context see Robinson (1989).

In particular, the above definition does not mean that a fixed continuous time process is discretized on a finer grid as T tends to infinity. If μ and A^0 do not depend on t and T then X does not depend on T as well and we obtain the spectral representation of a stationary process. Thus, the classical asymptotic theory for stationary processes is a special case of our approach.

There are similarities of our definition to Priestley's definition of an oscillatory process (see Priestley (1981), Chapter 11), for other approaches see Cohen (1989) and Tjøstheim (1976). However, there is a major difference and it is that we consider double indexed processes and we also make use of asymptotic considerations. While Priestley's concern was a stochastic representation of the process itself our concern is mainly a representation which allows for a rigorous asymptotic treatment of statistical inference problems. A deeper justification of our approach and a comparison with the approach of Priestley can be found in Dahlhaus (1996b, Section 3). One important consequence of this asymptotic approach is a uniqueness property of our spectral representation (proved below).

The Wigner–Ville spectrum for fixed T (Martin and Flandrin, 1985) is

$$f_T(u,\lambda) := \frac{1}{2\pi}\sum_{s=-\infty}^{\infty} \operatorname{cov}(X_{[uT-s/2],T}, X_{[uT+s/2],T})\exp(-\mathrm{i}\lambda s),$$

where $X_{s,T}$ is defined by (2.1) (with $A^0_{t,T}(\lambda) = A(0,\lambda)$ for $t<1$ and $A^0_{tT}(\lambda) = A(1,\lambda)$ for $t>T$). Below we prove that $f_T(u,\lambda)$ tends in squared mean to

$$f(u,\lambda) := |A(u,\lambda)|^2,$$

the spectrum which corresponds to the spectral representation. Therefore we call $f(u,\lambda)$ the (time varying) spectral density of the process.

Theorem 2.2. *If $X_{t,T}$ is locally stationary and $A(u,\lambda)$ is uniformly Lipschitz continuous in both components with index $\alpha > \frac{1}{2}$ then we have for all $u \in (0,1)$*

$$\int_{-\pi}^{\pi} |f_T(u,\lambda) - f(u,\lambda)|^2 \,\mathrm{d}\lambda = \mathrm{o}(1). \tag{2.3}$$

Proof. We have

$$f_T(u,\lambda) = \frac{1}{2\pi}\sum_{s=-\infty}^{\infty} \exp(-\mathrm{i}\lambda s)\int_{-\pi}^{\pi} \exp(\mathrm{i}\mu s)\, A^0_{[uT-s/2],T}(\mu)\,\overline{A^0_{[uT+s/2],T}(\mu)}\,\mathrm{d}\mu$$

and

$$f(u,\lambda) = \frac{1}{2\pi}\sum_{s=-\infty}^{\infty} \exp(-\mathrm{i}\lambda s)\int_{-\pi}^{\pi} \exp(\mathrm{i}\mu s)\, A(u,\mu)\,\overline{A(u,\mu)}\,\mathrm{d}\mu.$$

After replacing A^0 by A we therefore have

$$\int_{-\pi}^{\pi} |f_T(u,\lambda) - f(u,\lambda)|^2 \,\mathrm{d}\lambda = \frac{1}{2\pi}\sum_{s=-\infty}^{\infty} |c_s|^2 + \mathrm{o}(1)$$

with $c_s = \int_{-\pi}^{\pi} \exp(\mathrm{i}\mu s) g(s/2T, \mu) \mathrm{d}\mu$ and

$$g\left(\frac{s}{2T}, \mu\right) = A\left(u + \frac{s}{2T}, \mu\right) A\left(u - \frac{s}{2T}, -\mu\right) - A(u, \mu) A(u, -\mu)$$

where $A(u, \mu) = A(0, \mu)$ for $u < 0$ and $A(u, \mu) = A(1, \mu)$ for $u > 1$. By a standard argument for Fourier coefficients (Bary, 1964, Chapter 2.3) we obtain $|c_s| \leqslant C s^{-\alpha}$ and therefore

$$\sum_{s=n}^{\infty} |c_s|^2 = \mathrm{O}(n^{-2\alpha+1}).$$

Let $\Delta_s(\lambda) = \sum_{r=0}^{s-1} \exp(-\mathrm{i}\lambda r)$. Summation by parts gives

$$\begin{aligned}\sum_{s=0}^{n-1} |c_s|^2 &= \int_{-\pi}^{\pi} \int \sum_{s=0}^{n-1} \exp(\mathrm{i}(\lambda - \mu)s) g\left(\frac{s}{2T}, \lambda\right) \overline{g\left(\frac{s}{2T}, \mu\right)} \mathrm{d}\lambda \, \mathrm{d}\mu \\ &\leqslant \int_{-\pi}^{\pi} \int \left| - \sum_{s=0}^{n-1} \left\{ g\left(\frac{s}{2T}, \lambda\right) \overline{g\left(\frac{s}{2T}, \mu\right)} \right.\right. \\ &\qquad \left. - g\left(\frac{s-1}{2T}, \lambda\right) \overline{g\left(\frac{s-1}{2T}, \mu\right)} \right\} \Delta_s(\mu - \lambda) \\ &\qquad \left. + g\left(\frac{n-1}{2T}, \lambda\right) \overline{g\left(\frac{n-1}{2T}, \mu\right)} \Delta_n(\mu - \lambda) \right| \mathrm{d}\lambda \, \mathrm{d}\mu = \mathrm{O}\left(\frac{n \ln n}{T^\alpha}\right).\end{aligned}$$

The same holds for $\sum_{s=n}^{\infty} |c_{-s}|^2$. Choosing a suitable n gives the result. □

Usually, $f_T(u, \lambda)$ does not converge pointwise to $f(u, \lambda)$. This can be seen from the example $A(u, \lambda) = 1$ and $A_{t,T}^0(\lambda) = 1 + (1/T)\mathrm{e}^{-\mathrm{i}\lambda(2t-T)}$. In this case $f_T(\frac{1}{2}, 0) \to 2$.

Theorem 2.2 has an important consequence for the uniqueness of the spectral representation (2.1). It is well known (Priestley, 1981, Chapter 11.1) that the spectral representation (2.1) is not unique. However, Theorem 2.2 says that if there exists a spectral representation of the form (2.3) with a smooth $A(u, \lambda)$ then $|A(u, \lambda)|^2$ is uniquely determined from the whole triangular array (there may exist other non-smooth representations). Furthermore, it is the limit of the Wigner–Ville spectrum (with the asymptotics of this paper). Since $\mu(u)$ is the mean of the process it is also uniquely determined. If in addition the process $\xi(\lambda)$ is non-Gaussian then even $A(u, \lambda)$ is uniquely determined which may be proved similarly by considering higher-order spectra.

Inspection of the above proof shows that only the values of $X_{t,T}$ in the time interval

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

contribute to $f(u, \lambda)$. 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. The requirement $n \ln n/T^{\alpha} \to 0$ defines in some sense the interval on which the observations can be considered as stationary.

We feel that our approach describes mathematically very well what people mean when they speak of the spectrum at a timepoint t_0 of a nonstationary process $X_1, \dots, X_T$. Since the process is nonstationary only a few points around t_0 may have the same spectral structure. It is clear that the probability structure of these few points does not specify a spectral density uniquely. This is only guaranteed by an infinite number of observations. Our approach says that $f(u, \lambda) = |A(u, \lambda)|^2$ is the spectral density if one had infinitely many observations of the same kind at a fixed time point.

We now prove that time varying AR processes are locally stationary in the sense of the above definition. Consider the following system of difference equations:

$$\sum_{j=0}^{p} a_j\left(\frac{t}{T}\right)\left(X_{t-j,T} - \mu\left(\frac{t-j}{T}\right)\right) = \sigma\left(\frac{t}{T}\right)\varepsilon_t, \quad t \in \mathbb{Z} \tag{2.4}$$

where $a_0(u) \equiv 1$ and the ε_t are independent random variables with mean zero and variance 1. We assume that $\sigma(u)$ and the $a_j(u)$ are continuous on $\mathbb{R}$ with $\sigma(u) = \sigma(0)$, $a_j(u) = a_j(0)$ for $u < 0$; $\sigma(u) = \sigma(1)$, $a_j(u) = a_j(1)$ for $u > 1$, and differentiable for $u \in (0, 1)$ with bounded derivatives.

Since ε_t is an iid sequence we have a representation

$$\varepsilon_t = \int_{-\pi}^{\pi} \exp(\mathrm{i}\lambda t)(2\pi)^{-1/2}\,\mathrm{d}\xi(\lambda) \quad \text{for all } t$$

with $\xi(\lambda)$ as in Definition 2.1. Our goal is to prove that (2.1) holds with

$$A(u, \lambda) := \frac{\sigma(u)}{\sqrt{2\pi}}\left(1 + \sum_{j=1}^{p} a_j(u)\exp(-\mathrm{i}j\lambda)\right)^{-1}. \tag{2.5}$$

We now demonstrate the situation for the case $p = 1$. Direct verification shows that

$$X_{t,T} := \mu\left(\frac{t}{T}\right) + \int_{-\pi}^{\pi} \exp(\mathrm{i}\lambda t)\,A_{t,T}^{0}(\lambda)\,\mathrm{d}\xi(\lambda)$$

with

$$A_{t,T}^{0}(\lambda) = 1/\sqrt{2\pi}\sum_{\ell=0}^{\infty}(-1)^{\ell}\left\{\prod_{j=0}^{\ell-1} a_1\left(\frac{t-j}{T}\right)\right\}\sigma\left(\frac{t-\ell}{T}\right)\exp(-\mathrm{i}\lambda\ell)$$

is a solution of Eqs. (2.4). Straightforward calculation gives

$$\sup_{t,\lambda}|A_{t,T}^{0}(\lambda) - A\left(\frac{t}{T}, \lambda\right)| \leqslant K\frac{\log^3 T}{T},$$

i.e. $X_{t,T}$ is locally stationary in the sense of Definition 2.1 (the $\log^3 T$ can be dropped – see Theorem 2.3 below). The situation for general p is much more difficult. The existence of a purely nondeterministic solution of (2.4) for general p is usually answered under conditions on the Green's function of the autoregressive operator (Miller, 1968; Hallin, 1978, 1986; Melard, 1985). Künsch (1995) has proved that

Eqs. (2.4) have a solution of the form

$$X_{t,T} = \mu\left(\frac{t}{T}\right) + \sum_{\ell=0}^{\infty} \psi_{t,T,\ell}\, \varepsilon_{t-\ell} \tag{2.6}$$

with $\sum_{\ell=0}^{\infty} |\psi_{t,T,\ell}| < \infty$ uniformly in t and T if $\sum_{j=0}^{p} a_j(u) z^j \neq 0$ for all $|z| \leqslant 1 + c$ with $c > 0$ uniformly in u and the $a_j(u)$ are continuous in u. Replacing $\varepsilon_{t-\ell}$ in (2.6) by the left-hand side of (2.4) divided by $\sigma(t/T)$ leads to

$$\sum_{j=0}^{p} \psi_{t,T,\ell-j}\, a_j\left(\frac{t-\ell+j}{T}\right) \sigma\left(\frac{t-\ell+j}{T}\right)^{-1} = \begin{cases} 1, & \text{if } \ell = 0, \\ 0. & \text{if } \ell \neq 0 \end{cases} \tag{2.7}$$

with $\psi_{t,T,\ell-j} = 0$ if $\ell < j$. As above we obtain the spectral representation

$$X_{t,T} = \mu\left(\frac{t}{T}\right) + \int_{-\pi}^{\pi} \exp(\mathrm{i}\lambda t)\, A_{t,T}^{0}(\lambda)\, \mathrm{d}\xi(\lambda)$$

with

$$A_{t,T}^{0}(\lambda) = \frac{1}{\sqrt{2\pi}} \sum_{\ell=0}^{\infty} \psi_{t,T,\ell} \exp(-\mathrm{i}\lambda\ell).$$

Part (i) of the following theorem now gives the local stationarity of $X_{t,T}$. In part (ii) we prove an additional property needed in Dahlhaus (1996a) for the treatment of maximum likelihood estimates (∇ denotes the gradient with respect to θ).

Theorem 2.3. (i) *Suppose that* $\sum_{j=0}^{p} a_j(u) z^j \neq 0$ *for all* $|z| \leqslant 1 + c$ *with* $c > 0$ *uniformly in u and the coefficient functions* $a_j(u)$ *are continuous on* $\mathbb{R}$. *Then the difference equations* (2.4) *have a solution of the form* (2.1) *with* $A(u, \lambda)$ *as in* (2.5) *and time varying spectral density*

$$f(u, \lambda) = \frac{\sigma^2(u)}{2\pi} \left| \sum_{j=0}^{p} a_j(u) \exp(\mathrm{i}\lambda j) \right|^{-2}.$$

(ii) *If in addition the* $a_j(u)$ *depend on a parameter* $\theta \in \Theta \subset \mathbb{R}^p$ *and the components of* $a_j^\theta, \nabla a_j^\theta, \nabla^2 a_j^\theta$ *are differentiable in* $u \in (0, 1)$ *with uniformly bounded derivatives, then*

$$\sup_{t,\lambda} \left| \nabla^k \left\{ A_{\theta,t,T}^{0}(\lambda) - A_\theta\left(\frac{t}{T}, \lambda\right) \right\} \right| \leqslant K T^{-1}$$

for $k = 1, 2$.

Proof. The existence of a solution of the form (2.6) follows from Künsch (1995). It is straightforward to show that the components of A_θ, ∇A_θ, and $\nabla^2 A_\theta$ are differentiable in u and λ with uniformly bounded derivatives. Since $X_{t,T}$ fulfills difference equations (2.4) we obtain

$$\frac{\sigma\left(\frac{t}{T}\right)}{\sqrt{2\pi}} = \sum_{j=0}^{p} a_j\left(\frac{t}{T}\right) \exp(-\mathrm{i}\lambda j)\, A_{t-j,T}^{0}(\lambda) \tag{2.8}$$

for all $t \in \mathbb{Z}$, $T \in \mathbb{N}$, $\lambda \in (-\pi, \pi]$. Since

$$\frac{\sigma\left(\frac{t}{T}\right)}{\sqrt{2\pi}} = \sum_{j=0}^{p} a_j\left(\frac{t}{T}\right)\exp(-\mathrm{i}\lambda j)A\left(\frac{t}{T},\lambda\right)$$

$$= \sum_{j=0}^{p} a_j\left(\frac{t}{T}\right)\exp(-\mathrm{i}\lambda j)A\left(\frac{t-j}{T},\lambda\right)$$

$$+ \sum_{j=0}^{p} a_j\left(\frac{t}{T}\right)\exp(-\mathrm{i}\lambda j)\left\{A\left(\frac{t}{T},\lambda\right) - A\left(\frac{t-j}{T},\lambda\right)\right\}$$

for all t, T and λ, we get

$$\sum_{j=0}^{p} a_j\left(\frac{t}{T}\right)\exp\{\mathrm{i}\lambda(t-j)\}\left\{A_{t-j,T}^{0}(\lambda) - A\left(\frac{t-j}{T},\lambda\right)\right\}$$

$$= \sum_{j=0}^{p} a_j\left(\frac{t}{T}\right)\exp\{\mathrm{i}\lambda(t-j)\}\left\{A\left(\frac{t}{T},\lambda\right) - A\left(\frac{t-j}{T},\lambda\right)\right\}$$

$$=: \sigma\left(\frac{t}{T}\right)\delta_{t,T}(\lambda) \quad \text{with } \delta_{t,T}(\lambda) = 0 \quad \text{for } t \leqslant 0.$$

We set $\Delta_{t,T}(\lambda) := \{A_{t,T}^{0}(\lambda) - A(t/T,\lambda)\}\exp(\mathrm{i}\lambda t)$. It follows with (2.7) that

$$\Delta_{t,T}(\lambda) = \sum_{\ell=0}^{t} \psi_{t,T,\ell}\,\delta_{t-\ell,T}(\lambda).$$

Since $\delta_{t,T}(\lambda) = \mathrm{O}(T^{-1})$ uniformly in t and λ this gives

$$\sup_{t,\lambda}\left|A_{t,T}^{0}(\lambda) - A\left(\frac{t}{T},\lambda\right)\right| = \mathrm{O}(T^{-1}),$$

i.e. we have proved (i). To prove (ii) we proceed similarly. Let $A(t/T,\lambda)'$ denote the derivative with respect to θ_i. Eq. (2.8) implies

$$\sum_{j=0}^{p} a_j\left(\frac{t}{T}\right)\exp\{\mathrm{i}\lambda(t-j)\}\left\{A_{t-j,T}^{0}(\lambda)' - A\left(\frac{t-j}{T},\lambda\right)'\right\}$$

$$= \sum_{j=0}^{p} a_j\left(\frac{t}{T}\right)\exp\{\mathrm{i}\lambda(t-j)\}\left\{A\left(\frac{t}{T},\lambda\right)' - A\left(\frac{t-j}{T},\lambda\right)'\right\}$$

$$+ \sum_{j=0}^{p} a_j\left(\frac{t}{T}\right)'\exp(-\mathrm{i}\lambda j)\left\{A\left(\frac{t}{T},\lambda\right) - A_{t-j,T}^{0}(\lambda)\right\}$$

$$=: \sigma\left(\frac{t}{T}\right)\delta_{t,T}^{(1)}(\lambda) \quad \text{with } \delta_{t,T}^{(1)} = 0 \quad \text{for } t \leqslant 0.$$

We set $\Delta_{t,T}^{(1)}(\lambda) := \{A_{t,T}^{0}(\lambda)' - A(t/T,\lambda)'\}\exp(\mathrm{i}\lambda t)$. It follows that

$$\Delta_{t,T}^{(1)}(\lambda) = \sum_{\ell=0}^{t} \psi_{t,T,\ell}\,\delta_{t-\ell,T}^{(1)}(\lambda).$$

Since $\delta^{(1)}_{t,T}(\lambda) = \mathrm{O}(T^{-1})$ uniformly in t and λ this leads to

$$\sup_{t,\lambda}\left| A^0_{t,T}(\lambda)' - A\left(\frac{t}{T},\lambda\right)'\right| = \mathrm{O}(T^{-1}),$$

which implies (ii) for $k = 1$. For $k = 2$ the result is proved analogously. □

The representation for a MA process can be obtained easily from the above representation for ε_t. It is

$$X_{t,T} = \mu\left(\frac{t}{T}\right) + \int_{-\pi}^{\pi} \exp(\mathrm{i}\lambda t)\, A^0_{t,T}(\lambda)\,\mathrm{d}\xi(\lambda)$$

with

$$A^0_{t,T}(\lambda) = A\left(\frac{t}{T},\lambda\right) := \frac{\sigma\left(\frac{t}{T}\right)}{\sqrt{2\pi}} \sum_{j=0}^{q} b_j\left(\frac{t}{T}\right)\exp(-\mathrm{i}\lambda j).$$

where $b_j(u)$ are the coefficients of the MA part. In the mixed ARMA case we can combine the above arguments which leads to the above representation with

$$\sup_{t,\lambda}\left| A^0_{t,T}(\lambda) - A\left(\frac{t}{T},\lambda\right)\right| = \mathrm{O}(T^{-1}),$$

where

$$A(u,\lambda) = \frac{\sigma(u)}{\sqrt{2\pi}} \cdot \frac{\sum_{j=0}^{q} b_j(u)\exp(-\mathrm{i}\lambda j)}{\sum_{j=0}^{p} a_j(u)\exp(-\mathrm{i}\lambda j)}.$$

The above results are surprising and interesting since without our framework $f(u,\lambda)$ as in Theorem 2.3 cannot be interpreted as the spectral density of a time varying AR process (Melard and Herteleer-de Schutter, 1989). $f(u,\lambda)$ from Theorem 2.3 is usually called instantaneous spectrum of a time varying AR process. The definition was motivated by the relationship between a stationary AR process and the theoretical spectrum of the process (Kitagawa and Gersch, 1985). Theorem 2.3 gives a theoretical justification for this definition.

3. The Kullback–Leibler information divergence

Suppose we observe data $X_{1,T}, \ldots, X_{T,T}$ and fit a locally stationary model to the data e.g. a time varying AR model where the parameters $a^\theta_j(u)$, $\sigma_\theta(u)$ and $\mu_\theta(u)$ depend on a finite dimensional parameter $\theta \in \Theta \subset \mathbb{R}^p$ (all functions may be e.g. polynomials in time with the parameters being the coefficients). Suppose we estimate the parameters by maximizing the likelihood. Then there are several important questions related to the Kullback–Leibler information divergence e.g. the problem what happens if the model is not correct or the aspect of model selection. To be precise let

$$\Sigma_T(A,B) = \left\{\int_{-\pi}^{\pi} \exp\{\mathrm{i}\lambda(r-s)\}\, A^0_{r,T}(\lambda)\,\overline{B^0_{s,T}(\lambda)}\,\mathrm{d}\lambda\right\}_{r,s=1,\ldots,T}.$$

If the true process is locally stationary with transfer function A^0 and trend function μ, then $\Sigma = \Sigma_T(A, A)$ is the true covariance matrix of the process. Suppose the model is Gaussian and has transfer function A_θ^0 and trend function μ_θ. Then $\Sigma_\theta = \Sigma_T(A_\theta, A_\theta)$ is the model covariance matrix. The maximum likelihood estimate is

$$\hat{\theta}_T := \arg\min_{\theta \in \Theta} \mathscr{L}_T(\theta)$$

where

$$\mathscr{L}_T(\theta) := -\frac{1}{T} \text{ Gaussian log likelihood}$$

$$= \frac{1}{2}\log(2\pi) + \frac{1}{2T}\log\det\Sigma_\theta + \frac{1}{2T}(\boldsymbol{X} - \boldsymbol{\mu}_\theta)'\,\Sigma_\theta^{-1}\,(\boldsymbol{X} - \boldsymbol{\mu}_\theta) \tag{3.1}$$

with $\boldsymbol{X} = (X_{1,T}, \dots, X_{T,T})'$ and $\boldsymbol{\mu}_\theta = (\mu_\theta(1/T), \dots, \mu_\theta(T/T))'$. Under certain regularity conditions $\hat{\theta}_T$ *will converge to*

$$\theta_0 := \arg\min_{\theta \in \Theta} \mathscr{L}(\theta)$$

where

$$\mathscr{L}(\theta) := \lim_{T \to \infty} E\ \mathscr{L}_T(\theta)$$

(Dahlhaus, 1996a, Theorem 2.3). If the model is correct, i.e. $A^0 = A_{\theta^*}^0$ and $\mu = \mu_\theta^*$ then typically $\theta_0 = \theta^*$. It is therefore important to calculate $\mathscr{L}(\theta)$ which is equivalent to the calculation of the Kullback–Leibler information divergence. This is done in Theorem 3.4. For this calculation we need a result on the local prediction error derived in Theorem 3.2. This result is a generalization of Kolmogorov's formula for stationary processes (Brockwell and Davis, 1987, Theorem 5.8.1). As an application the best approximating parameter θ_0 is calculated in the situation where a stationary model is fitted to a nonstationary process (Example 3.5). The results are also important with respect to model selection since an estimate of $\mathscr{L}(\hat{\theta}_T)$ usually serves as a model selection criterion.

If the model is correct then the estimator $\hat{\theta}_T$ *is called Fisher efficient if its asymptotic covariance matrix is equal to the limit of the Fisher information matrix*

$$\Gamma := \lim_{T \to \infty} T\, E_{\theta_0}(\nabla \mathscr{L}_T(\theta_0))(\nabla \mathscr{L}_T(\theta_0))'$$

Γ is calculated in Theorem 3.6. More generally, a LAN property is proved in Dahlhaus (1996a). Suppose now that the true process fulfills

Assumption 3.1. $X_{1,T}, \dots, X_{T,T}$ are realizations of a locally stationary process with transfer function A^0 where the corresponding A is bounded from below and has uniformly bounded derivative

$$\frac{\partial}{\partial u}\frac{\partial}{\partial \lambda} A$$

and mean function μ which has uniformly bounded derivative. $f(u,\lambda) = |A(u,\lambda)|^2$ denotes the time varying spectral density of $X_{t,T}$.

Theorem 3.2. *Suppose Assumption* 3.1 *holds with* $\mu(u) = 0$.

(i) *Let* $\hat{X}_{t,T}$ *be the best linear predictor of* $X_{t,T}$ *given* $X_{1,T}, \ldots, X_{t-1,T}$ *and* $v_{t,T}$ *be the prediction error, i.e.* $v_{t,T} = E(X_{t,T} - \hat{X}_{t,T})^2$. *Then*

$$v_{t,T} = \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \log 2\pi f\left(\frac{t}{T},\lambda\right)\mathrm{d}\lambda\right\} + \mathrm{o}_t(1) + \mathrm{o}_T(1)$$

where the $\mathrm{o}_t(1)$ *term is uniform in* T *and the* $\mathrm{o}_T(1)$ *term is uniform in* t.

(ii) *Furthermore, we have*

$$\lim_{T\to\infty} \frac{1}{T} \log\det \Sigma_T(A,A) = \frac{1}{2\pi}\int_0^1\int_{-\pi}^{\pi} \log 2\pi f(u,\lambda)\,\mathrm{d}\lambda\,\mathrm{d}u.$$

Theorem 3.2(i) is a nonstationary version of Kolmogorov's formula (Brockwell and Davis, 1987, Theorem 5.8.1). It is proved in the appendix. The Kolmogorov formula for processes with an evolutionary spectral representation in the sense of Priestley has been stated by Subba Rao (1973). He has looked at the prediction error given the whole (infinite) past of the series.

Lemma 3.3. *Suppose Assumption* 3.1 *holds and* B^0 *(together with the corresponding* B*) and* v *fulfill the same smoothness assumptions as* A^0 *and* μ*, respectively. Then we have with*

$$Y_T = \frac{1}{T}(\boldsymbol{X} - \boldsymbol{v})' \,\Sigma_T(B,B)^{-1}\,(\boldsymbol{X} - \boldsymbol{v}),$$

$$EY_T = \frac{1}{2\pi}\int_0^1\int_{-\pi}^{\pi} \frac{f(u,\lambda)}{|B(u,\lambda)|^2}\,\mathrm{d}\lambda\,\mathrm{d}u + \frac{1}{2\pi}\int_0^1 \frac{(\mu(u) - v(u))^2}{|B(u,0)|^2}\,\mathrm{d}u + \mathrm{O}(T^{-1/2}\ln^4 T)$$

and

$$\operatorname{var} Y_T = \mathrm{O}(T^{-1}).$$

Proof. We have with $\Sigma_A = \Sigma_T(A,A)$ and $\Sigma_B = \Sigma_T(B,B)$

$$Y_T = \frac{1}{T}(\boldsymbol{X} - \boldsymbol{\mu})'\,\Sigma_B^{-1}\,(\boldsymbol{X} - \boldsymbol{\mu}) + \frac{2}{T}(\boldsymbol{X} - \boldsymbol{\mu})'\,\Sigma_B^{-1}\,(\boldsymbol{\mu} - \boldsymbol{v}) + \frac{1}{T}(\boldsymbol{\mu} - \boldsymbol{v})'\Sigma_B^{-1}(\boldsymbol{\mu} - \boldsymbol{v})$$

leading with Lemma 4.8 to the assertion for EY_T. In the Gaussian case the first two terms are independent with

$$\operatorname{var}\frac{1}{T}(\boldsymbol{X} - \boldsymbol{\mu})'\,\Sigma_B^{-1}\,(\boldsymbol{X} - \boldsymbol{\mu}) = \frac{2}{T^2}\operatorname{tr}\{(\Sigma_A\,\Sigma_B^{-1})^2\}$$

and

$$\operatorname{var} \frac{1}{T} (X - \mu)' \Sigma_B^{-1} (\mu - \nu) = \frac{1}{T^2} (\mu - \nu)' \Sigma_B^{-1} \Sigma_A \Sigma_B^{-1} (\mu - \nu)$$

leading again with Lemma 4.8 to the assertion for var Y_T. The non-Gaussian case needs more technical considerations. Since the behaviour of var Y_T is less important in this paper we omit the proof. □.

Theorem 3.4. *Suppose that Assumption* 3.1 *holds and the model consists of locally stationary processes with transfer function* A_θ^0 *and trend function* μ_θ *that also fulfills Assumption* 3.1. *Then we have with* $f_\theta(u, \lambda) = |A_\theta(u, \lambda)|^2$

$$\mathscr{L}(\theta) = \lim_{T \to \infty} E \mathscr{L}_T(\theta) = \frac{1}{4\pi} \int_0^1 \int_{-\pi}^{\pi} \left\{ \log 4\pi^2 f_\theta(u, \lambda) + \frac{f(u, \lambda)}{f_\theta(u, \lambda)} \right\} \mathrm{d}\lambda \, \mathrm{d}u + \frac{1}{4\pi} \int_0^1 \frac{(\mu_\theta(u) - \mu(u))^2}{f_\theta(u, 0)} \mathrm{d}u$$

and

$$\mathscr{L}_T(\theta) \to \mathscr{L}(\theta)$$

in probability.

Proof. The assertions follow immediately from Theorem 3.2 and Lemma 3.3. □

In the same way Theorem 3.2 and Lemma 3.3 lead to the asymptotic Kullback–Leibler information divergence of two locally stationary Gaussian processes (for a discussion in the stationary case see Parzen (1983)). If $X_{1,T}, \ldots, X_{T,T}$ $(\tilde{X}_{1,T}, \ldots, \tilde{X}_{T,T})$ are locally stationary Gaussian processes with densities g $(\tilde{g})$, spectral densities $f = |A|^2$ $(\tilde{f} = |\tilde{A}|^2)$ and trend functions μ $(\tilde{\mu})$ then we obtain for the asymptotic Kullback–Leibler information divergence in the same way

$$\begin{aligned} D(\tilde{f}, \tilde{\mu}, f, \mu) &= \lim_{T \to \infty} \frac{1}{T} E_g \log \frac{g}{\tilde{g}} \\ &= \frac{1}{4\pi} \int_0^1 \int_{-\pi}^{\pi} \left\{ \log \frac{\tilde{f}(u, \lambda)}{f(u, \lambda)} + \frac{f(u, \lambda)}{\tilde{f}(u, \lambda)} - 1 \right\} \mathrm{d}\lambda \, \mathrm{d}u \\ &\quad + \frac{1}{4\pi} \int_0^1 \frac{(\tilde{\mu}(u) - \mu(u))^2}{\tilde{f}(u, 0)} \mathrm{d}u \end{aligned}$$

i.e. the information divergence between two locally stationary Gaussian processes is a distance between the spectral densities and the trend functions. Of course θ_0 also minimizes $D(f_\theta, \mu_\theta, f, \mu)$, i.e. θ_0 is the value such that f_{θ_0} and μ_{θ_0} are the best approximations of the true f and μ in the sense of the above distance. This is the value to which the maximum likelihood estimate converges if the true process is not in the fitted model. This is proved in Dahlhaus (1996a). The above distance is the time average of the Kullback–Leibler divergence in the stationary case (Parzen, 1983).

Example 3.5. Suppose that the model is stationary, i.e. $f_\theta(\lambda) := f_\theta(u, \lambda)$ and $m := \mu_\theta(u)$ do not depend on u. Then

$$\mathscr{L}(\theta) = \frac{1}{4\pi}\int_{-\pi}^{\pi}\left\{\log 4\pi^2 f_\theta(\lambda) + \frac{\int_0^1 f(u,\lambda)\,\mathrm{d}u}{f_\theta(\lambda)}\right\}\mathrm{d}\lambda + \frac{1}{4\pi} f_\theta(0)^{-1}\int_0^1 (m - \mu(u))^2\,\mathrm{d}u,$$

i.e. $m_0 = \int_0^1 \mu(u)\,\mathrm{d}u$, and $f_{\theta_0}(\lambda)$ is the best approximation to the time integrated true spectrum.

The technical results used in the proof of Theorem 3.4 also enable us to calculate the limit of the Fisher information matrix.

Theorem 3.6. *Suppose that Assumption* 3.1 *holds with* $A = A_{\theta_0}$, $\mu = \mu_{\theta_0}$ *and that also*

$$\frac{\partial}{\partial\theta} A_{\theta_0} \text{ and } \frac{\partial}{\partial\theta}\mu_{\theta_0}$$

fulfill Assumption 3.1. *Then we have with* $f_\theta(u,\lambda) = |A_\theta(u,\lambda)|^2$

$$\Gamma = \frac{1}{4\pi}\int_0^1\int_{-\pi}^{\pi} (\nabla \log f_{\theta_0})(\nabla \log f_{\theta_0})'\,\mathrm{d}\lambda\,\mathrm{d}u + \frac{1}{2\pi}\int_0^1 (\nabla \mu_{\theta_0}(u))(\nabla u_{\theta_0}(u))' f_{\theta_0}^{-1}(u,0)\,\mathrm{d}u.$$

Proof. Let

$$C_\theta^{(i)} = \frac{\partial}{\partial\theta_i}\Sigma_\theta = \Sigma_T\left(\frac{\partial}{\partial\theta_i}A_\theta, A_\theta\right) + \Sigma_T\left(A_\theta, \frac{\partial}{\partial\theta_i}A_\theta\right).$$

We have with Lemma 4.1

$$\frac{\partial}{\partial\theta_i}\mathscr{L}_T(\theta) = \frac{1}{2T}\operatorname{tr}\{\Sigma_\theta^{-1}C_\theta^{(i)}\} - \frac{1}{2T}(\boldsymbol{X} - \boldsymbol{\mu}_\theta)'\Sigma_\theta^{-1}C_\theta^{(j)}\Sigma_\theta^{-1}(\boldsymbol{X} - \boldsymbol{\mu}_\theta)$$
$$- \frac{1}{T}(\nabla_i \mu_\theta)'\Sigma_\theta^{-1}(\boldsymbol{X} - \boldsymbol{\mu}_\theta).$$

Therefore

$$E_{\theta_0}\frac{\partial}{\partial\theta_i}\mathscr{L}_T(\theta) = 0$$

and

$$TE_{\theta_0}(\nabla_i\mathscr{L}_T(\theta_0)\nabla_j\mathscr{L}_T(\theta_0)) = \frac{1}{4T}\operatorname{tr}\{\Sigma_{\theta_0}^{-1}C_{\theta_0}^{(i)}\Sigma_{\theta_0}^{-1}C_{\theta_0}^{(j)}\} + \frac{1}{T}(\nabla_i\mu_{\theta_0})'\Sigma_{\theta_0}^{-1}(\nabla_j\mu_{\theta_0}).$$

Lemma 4.8 now implies the result. □

The special form of $D(f_\theta, f)$ or of $\mathscr{L}(\theta)$ suggests an alternative (minimum distance) estimate of θ_0 which is obtained by replacing the unknown f in $\mathscr{L}(\theta)$ by a non-parametric estimate of f and minimizing the resulting function with respect to θ. An estimate of this type has been investigated in Dahlhaus (1993).

4. The behaviour of covariance matrices of locally stationary processes

In this section we establish several results on matrix norms and the trace behaviour of $\Sigma_T(A, A)$ and products of matrices of this type. Apart from this paper the results are also essential for the investigation of the maximum likelihood estimate.

Suppose A is an $n \times n$ matrix. We denote

$$\|A\| = \sup_{x \in C^n} \frac{|Ax|}{|x|} = \sup_{x \in C^n} \left(\frac{x^* A^* Ax}{x^* x}\right)^{1/2}$$
$$= [\text{maximum characteristic root of } A^* A]^{1/2},$$

where A^* denotes the conjugate transpose of A, and

$$|A| = [\operatorname{tr}(AA^*)]^{1/2}.$$

If A is a real nonnegative symmetric matrix, i.e. $A = P'DP$ with $PP' = P'P = I$ and $D = \operatorname{diag}\{\lambda_1, \ldots, \lambda_n\}$, where $\lambda_i \geq 0$, then we define $A^{1/2} = P'D^{1/2}P$, where $D^{1/2} = \operatorname{diag}\{\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n}\}$. Thus, $A^{1/2}$ is also nonnegative definite and symmetric with $A^{1/2}A^{1/2} = A$. Furthermore, $A^{-1/2} = (A^{1/2})^{-1}$ if A is positive definite.

The following results are well known [see e.g. Davies (1973, Appendix II) or Graybill (1983, Section 5.6)].

Lemma 4.1. *Let A, B be $n \times n$ matrices. Then*

(a) $|\operatorname{tr}(AB)| \leq |A|\,|B|$,

(b) $|AB| \leq \|A\|\,|B|$,

(c) $|AB| \leq |A|\,\|B\|$,

(d) $\|AB\| \leq \|A\|\,\|B\|$,

(e) $|x^* Ax| \leq x^* x \|A\| \quad x \in \mathbb{C}^n$.

Suppose now, that the elements of A are continuously differentiable functions of θ. Then

(f) $$\frac{\partial}{\partial\theta} A^{-1} = -A^{-1}\left(\frac{\partial}{\partial\theta} A\right) A^{-1},$$

(g) $$\frac{\partial}{\partial\theta} \log\det A = \operatorname{tr}\left\{A^{-1}\frac{\partial}{\partial\theta} A\right\}.$$

Furthermore, let $L_T: \mathbb{R} \to \mathbb{R}$, $T \in \mathbb{R}^+$, be the periodic extension (with period 2π) of

$$L_T^*(\alpha) := \begin{cases} T, & |\alpha| \leq 1/T, \\ 1/|\alpha|, & 1/T \leq |\alpha| \leq \pi. \end{cases}$$

Properties of $L_T(\alpha)$ are listed in Dahlhaus (1993, Lemma A.4). We remark that we have with a generic constant K

$$\int_{-\pi}^{\pi} L_T(\alpha)\,d\alpha \text{ is monotone increasing in } T, \tag{4.1}$$

$$\int_{-\pi}^{\pi} L_T(\beta - \alpha) L_T(\alpha + \gamma)\,d\alpha \leq K L_T(\beta + \gamma) \ln T, \tag{4.2}$$

$$\int_{-\pi}^{\pi} L_T(\alpha)^k \, d\alpha \leqslant K\, T^{k-1} \ln T^{\{k=1\}}. \tag{4.3}$$

$$\int_{-\pi}^{\pi} L_N(\alpha)^l L_M(S(\alpha-\beta))^k \, d\alpha \leqslant K \frac{N^l M^{k-1}}{S} \ln M^{\{k=1\}} \ln S^{\{l=1\}}. \tag{4.4}$$

$$\int_{-\pi}^{\pi} L_N(\lambda - x) L_N(x-\mu) L_M(S(\alpha - x))\, dx \leqslant K \frac{N}{S} \ln M \ln S\, L_N(\lambda-\mu). \tag{4.5}$$

Furthermore, let

$$H_N(f(\cdot),\lambda) := \sum_{s=0}^{N-1} f(s) \exp(-i\lambda s)$$

and

$$H_N(\lambda) := H_N(1,\lambda).$$

Lemma 4.2. *Let $N, T \in \mathbb{N}$ with $N \leqslant T$. Suppose $\psi : [0,1] \to \mathbb{R}$ is differentiable with bounded derivative. Then we have for $0 \leqslant t \leqslant N$*

$$H_N\left(\psi\left(\frac{\cdot}{T}\right),\lambda\right) = \psi\left(\frac{t}{T}\right) H_N(\lambda) + O\left(\sup_u |\psi'(u)| \frac{N}{T} L_N(\lambda)\right)$$

$$= O\left(\sup_{u \leqslant N/T} |\psi(u)|\, L_N(\lambda) + \sup_u |\psi'(u)| L_N(\lambda)\right).$$

The same holds, if $\psi(\cdot/T)$ is replaced on the left side by numbers $\psi_{s,T}$ with $\sup_s |\psi_{s,T} - \psi(s/T)| = O(T^{-1})$.

Proof. The proof is similar to the proof of Lemma 4.5 in Dahlhaus (1993). □

Essentially, we need upper bounds for the norms $\| \Sigma_T(A,A) \|$ and $\| \Sigma_T(A,A)^{-1} \|$ and the results of Lemmas 4.5 and 4.8. If A is constant over time (stationary case) these results are well known. However, the time dependence causes serious technical problems. To establish our results we approximate $\Sigma_T(A,A)$ by overlapping block Toeplitz matrices where the blocks are along the diagonal (Lemma 4.4). In Lemmas 4.7 and 4.8 we use the same technique for the approximation of $\Sigma_T(A,A)^{-1}$.

We divide the observation domain $\{1, \ldots, T\}$ into overlapping blocks of length N with a shift S. At the edges we keep the shift and use a smaller block length (such that each observation is contained exactly in N/S blocks).

To be specific let S be a natural number and N a multiple of S. We start by assuming that T can be divided by S. We use blocks of length

$$L_j = \begin{cases} jS & j = 1, \ldots, N/S, \\ N & j = N/S, \ldots, T/S, \\ N + T - jS & j = T/S, \ldots, (T+N)/S - 1 \end{cases}$$

with midpoints

$$t_j = \begin{cases} jS/2 & j = 1, \dots, N/S, \\ jS - N/2 & j = N/S, \dots, T/S, \\ \dfrac{T-N}{2} + jS/2 & j = T/S, \dots, (T+N)/S - 1. \end{cases}$$

Let $M = (T + N)/S - 1$ be the number of blocks and $u_j = t_j/T$ be the midpoints in rescaled time. Note, that each point is contained exactly in N/S blocks. If T cannot be divided by S the last N/S blocks are chosen smaller which does not affect our argumentation. Furthermore, let $v_j \in [0, 1]$ (usually we take $v_j = u_j = t_j/T$),

$$W_T^{(j)}(\phi) = \left\{ \int_{-\pi}^{\pi} \phi(v_j, \lambda) \exp(\mathrm{i}\lambda(k - \ell))\, \mathrm{d}\lambda \right\}_{k, \ell = 1, \dots, L_j}$$

and

$$K_T^{(j)} = (0_{j1}\, \mathrm{I}_{L_j}\, 0_{j2})$$

where I_{L_j} is the $L_j \times L_j$ identity matrix and 0_{j1} is the $L_j \times (t_j - L_j/2)$ matrix with zero entries and 0_{j2} is the $L_j \times (T - t_j + L_j/2)$ matrix with zero entries (i.e. $K_T^{(j)}$ contains an $L_j \times L_j$ identity matrix "centred" around t_j). $K_T^{(j)} X$ then gives the jth block of observations. We define

$$W_T(\phi) = \frac{S}{N} \sum_{j=1}^{M} K_T^{(j)\prime} W_T^{(j)}(\phi) K_T^{(j)}$$

and set

$$\Sigma_T^{(j,k)}(A, B) = K_T^{(j)} \Sigma_T(A, B) K_T^{(k)\prime}.$$

We now approximate $\Sigma_T(A, B)$ by $W_T(A, \overline{B})$. First, we summarize the assumptions used in this chapter.

Assumption 4.3. (i) Suppose $A: [0, 1] \times \mathbb{R} \to \mathbb{C}$ is a 2π-periodic function with $A(u, \lambda) = \overline{A(u, -\lambda)}$ and $|A(u, \lambda)| \geqslant C > 0$ which is differentiable in u and λ with uniformly bounded derivative

$$\frac{\partial}{\partial u} \frac{\partial}{\partial \lambda} A.\ A_{t,T}^0 : \mathbb{R} \to \mathbb{C}$$

are 2π-periodic functions with

$$\sup_{t, \lambda} \left| A_{t,T}^0(\lambda) - A\left(\frac{t}{T}, \lambda\right) \right| \leqslant K T^{-1}.$$

(ii) Suppose $\phi: [0, 1] \times \mathbb{R} \to \mathbb{C}$ is a 2π-periodic function which is differentiable in u and λ with uniformly bounded derivative

$$\frac{\partial}{\partial u} \frac{\partial}{\partial \lambda} \phi.$$

(iii) Suppose $\mu:[0,1]\to\mathbb{R}$ is differentiable with uniformly bounded derivative.
(iv) Suppose $S/N \to 0$ as T tends to infinity.

Remark. All results proved in the rest of this section are uniform in the sense that the constants depend only on the bounds of the involved functions A, ϕ and μ and their derivatives and not on the particular values.

Lemma. 4.4. *Suppose A and B fulfill Assumption* 4.3(i) *and $S/N\to 0$, $N\ln^2 N/T\to 0$. Then we have for each $x=(x_1,\ldots,x_T)'\in\mathbb{C}^T$ as $T\to\infty$*

$$x^*(\Sigma_T(A,B)-W_T(A\bar{B}))x = x^*x\,\mathrm{o}(1)$$

where the v_j in the definition of W_T may be arbitrary time points in the jth block, i.e. with $|v_j-u_j|\leqslant L_j/(2T)$. As a consequence we obtain with $C_1=\sup_{u,\lambda}|A(u,\lambda)\,B(u,\lambda)|$ and $C_2=\inf_{u,\lambda}|A(u,\lambda)|^2$

$$\sup_{|x|=1}|x^*\,\Sigma_T(A,B)\,x|\leqslant 2\pi\,C_1+\mathrm{o}(1),$$

$$\inf_{|x|=1}|x^*\,\Sigma_T(A,A)\,x|\geqslant 2\pi\,C_2+\mathrm{o}(1)$$

and

$$\|\Sigma_T(A,A)\|\leqslant 2\pi\,C_1+\mathrm{o}(1),\ \|\Sigma_T(A,A)^{-1}\|\leqslant(2\pi\,C_2+\mathrm{o}(1))^{-1}.$$

Proof. Let $c_{r,t}$ be the components of $\Sigma_T(A,B)$. Straightforward considerations yield

$$\begin{aligned}
&x^*(\Sigma_T(A,B)-W_T(A\bar{B}))x\\
&=x^*\left\{\frac{S}{N}\sum_{j=1}^{M}K_T^{(j)\prime}(\Sigma_T^{(j,j)}(A,B)-W_T^{(j)}(A\bar{B}))K_T^{(j)}\right\}x\\
&\quad+\sum_{\substack{j,k=0\\ j\neq k}}^{T/S-1}\min\left\{|j-k|\frac{S}{N},1\right\}\sum_{r,t=1}^{S}\bar{x}_{jS+r}\,c_{jS+r,kS+t}\,x_{kS+t}
\end{aligned}$$

(note that in the first expression j denotes the jth block while in the second it denotes the jth shift). We now show that the second term tends to zero. To show this we replace c_{rt} by

$$\tilde{c}\left(\frac{r+t}{2T},r-t\right)$$

where

$$\tilde{c}(u,k)=\int_{-\pi}^{\pi}\exp(\mathrm{i}\lambda k)\,A\left(u+\frac{k}{2T},\lambda\right)B\left(u-\frac{k}{2T},-\lambda\right)\mathrm{d}\lambda.$$

Then the second term is bounded by

$$\sum_{d=1}^{T/S-1}\min\left\{d\frac{S}{N},1\right\}\sum_{\substack{r,t=1\\ (d-1)S<|r-t|\leqslant dS}}^{T}\left|\bar{x}_r\,\tilde{c}\left(\frac{r+1}{2T},r-t\right)x_t\right|+R$$

with a remainder R estimated below. With $c^*(k) := \sup_{u\in(0,1)}|\tilde{c}(u,k)|$ this is bounded by

$$2x^* x \sum_{d=1}^{T/S-1} \min\left\{d\frac{S}{N}, 1\right\} \sum_{k=(d-1)S+1}^{dS} c^*(k) + R$$

$$\leqslant 2x^* x\left(\frac{S+\sqrt{N}}{N}\sum_{k=1}^{\infty} c^*(k) + \sum_{k>\sqrt{N}} c^*(k)\right) + R.$$

The first term tends to zero if $\sum_{k=1}^{\infty} c^*(k) < \infty$. To show this we first note that we get for the Fourier coefficients by a similar argument as in Bary (1964, Section 2.3) $|\tilde{c}(u,k) - \tilde{c}(u',k)| \leqslant K|u-u'|(1/k)$ uniformly in u, u' with some constant K. Let $u_j^{(k)} = j/\ln^2 k$ ($j = 0, \ldots, \ln^2 k$). Then

$$\sum_{k=1}^{\infty} c^*(k) \leqslant K \sum_{k=1}^{\infty} \frac{1}{\ln^2 k}\frac{1}{k} + K \sum_{k=1}^{\infty}\sum_{j=0}^{\ln^2 k} |\tilde{c}(u_j^{(k)}, k)|$$

$$\leqslant \text{const.} + K \sum_{j=0}^{\infty} \sup_u \sum_{k=c\sqrt{j}}^{\infty} \tilde{c}(u,k)$$

The smoothness properties of A and B imply (Bary, 1964, Section 2.3)

$$\sum_{k=n}^{\infty} |\tilde{c}(u,k)| = O(n^{-1/2}).$$

Therefore, the above expression is bounded. To estimate the remainder we note that

$$c_{r,t} - \tilde{c}\left(\frac{r+t}{2T}, r-t\right)$$

$$= \int_{-\pi}^{\pi} \exp(i\lambda(r-t))\left[A\left(\frac{r}{T},\lambda\right)\overline{\left\{B_{t,T}^0(\lambda) - B\left(\frac{t}{T},\lambda\right)\right\}}\right.$$

$$+ \left\{A_{r,T}^0(\lambda) - A\left(\frac{r}{T},\lambda\right)\right\}\overline{B\left(\frac{t}{T},\lambda\right)}$$

$$\left. + \left\{A_{r,T}^0(\lambda) - A\left(\frac{r}{T},\lambda\right)\right\}\overline{\left\{B_{t,T}^0(\lambda) - B\left(\frac{t}{T},\lambda\right)\right\}}\right] d\lambda$$

$$=: \delta_{r,t}^{(1)} + \delta_{r,t}^{(2)} + \delta_{r,t}^{(3)}.$$

Let

$$a(u,k) = \int_{-\pi}^{\pi} A(u,\lambda)\exp(i\lambda k)\, d\lambda$$

and

$$\Delta_{t,T}(k) = \int_{-\pi}^{\pi} \left\{B_{t,T}^0(\lambda) - B\left(\frac{t}{T},\lambda\right)\right\}\exp(i\lambda k)\, d\lambda.$$

As above we have $\sum_{k=n}^{\infty} \sup_u |a(u,k)| = O(n^{-1/2})$ and with the Parseval identity

$$\delta_{r,t}^{(1)} = \frac{1}{2\pi} \sum_{|\ell| \leqslant n} a\left(\frac{r}{t}, \ell\right) \Delta_{t,T}(\ell + t - r) + O(T^{-1} n^{-1/2}).$$

Therefore, we obtain for the corresponding term of the remainder as an upper bound

$$\sum_{r,t=1}^{T} |x_r x_t \delta_{r,t}^{(1)}| \leqslant \frac{K}{2\pi} \sum_{|\ell| \leqslant n} \sum_{t=1}^{T} |x_t| \sum_{r=1}^{T} |x_r| \Delta_{t,T}(\ell + t - r) + x^* x\, O(n^{-1/2})$$

$$\leqslant K n T^{1/2} x^* x \left\{ \sup_t \int_{-\pi}^{\pi} |B_{t,T}^o(\lambda) - B\left(\frac{t}{T}, \lambda\right)|^2 \, d\lambda \right\}^{1/2} + x^* x\, O(n^{-1/2}).$$

$$\leqslant K x^* x \{ n T^{-1/2} + n^{-1/2} \}.$$

Choosing e.g. $n = T^{1/4}$ gives convergence to zero. The term with $\delta_{r,t}^{(2)}$ is handled analogously and the term with $\delta_{r,t}^{(3)}$ can be estimated directly by $x^* x O(T^{-1})$.
To get an upper bound for the first term of (4.6) we consider the different j separately. We have with $r_j = t_j - L_j/2 + 1$

$$x^* K_T^{(j)\prime} (\Sigma_T^{(j,j)}(A,B) - W_T^{(j)}(A, \overline{B})) K_T^{(j)} x$$

$$= \sum_{r,t=0}^{L_j - 1} \bar{x}_{r_j + r}\, x_{r_j + t} \int_{-\pi}^{\pi} \left\{ A_{r_j + r, T}^0(\lambda)\, B_{r_j + t, T}^0(-\lambda) \right.$$

$$\left. - A\left(\frac{v_j}{T}, \lambda\right) B\left(\frac{v_j}{T}, -\lambda\right) \right\} \exp(i\lambda(r-t))\, d\lambda. \tag{4.7}$$

Let

$$z_k = \sum_{r=0}^{L_j - 1} x_{r_j + r} \exp\left(-i \frac{2\pi}{L_j} kr\right).$$

We have

$$x_{r_j + r} = \frac{1}{L_j} \sum_{k=0}^{L_j - 1} z_k \exp\left(i \frac{2\pi}{L_j} kr\right)$$

and

$$\frac{1}{L_j} \sum_{k=0}^{L_j - 1} |z_k|^2 = \sum_{r=0}^{L_j - 1} |x_{r_j + r}|^2.$$

With this notation (4.7) is equal to

$$\frac{1}{L_j^2} \sum_{k,\ell=0}^{L_j - 1} \bar{z}_k z_\ell \int_{-\pi}^{\pi} \left\{ H_{L_j}\left(A_{r_j + \cdot, T}^0(\lambda), \frac{2\pi k}{L_j} - \lambda\right) H_{L_j}\left(B_{r_j + \cdot, T}^0(-\lambda), \lambda - \frac{2\pi \ell}{L_j}\right) \right.$$

$$\left. - A\left(\frac{v_j}{T}, \lambda\right) B\left(\frac{v_j}{T}, -\lambda\right) H_{L_j}\left(\frac{2\pi k}{L_j} - \lambda\right) H_{L_j}\left(\lambda - \frac{2\pi \ell}{L_j}\right) \right\} d\lambda.$$

Lemma 4.2 implies that this is bounded by

$$K\frac{L_j}{T}\frac{1}{L_j^2}\sum_{k,\ell=0}^{L_j-1}|\bar{z}_k z_\ell|\int_{-\pi}^{\pi}L_{L_j}\left(\frac{2\pi k}{L_j}-\lambda\right)L_{L_j}\left(\lambda-\frac{2\pi\ell}{L_j}\right)\mathrm{d}\lambda$$

$$\leqslant K\frac{\ln L_j}{TL_j}\sum_{k,\ell=0}^{L_j-1}|\bar{z}_k z_\ell|\,L_{L_j}\left(\frac{2\pi(k-\ell)}{L_j}\right)\text{(cf. (4.2))}$$

$$\leqslant K\frac{\ln L_j}{T}\sum_{k=0}^{L_j-1}|z_k|^2+K\frac{\ln L_j}{T}\sum_{k\neq\ell}|\bar{z}_k z_\ell|\,|k-\ell|^{-1}$$

$$\leqslant K\frac{L_j\ln^2 L_j}{T}\sum_{r=0}^{L_j-1}|x_{r_j+r}|^2.$$

This implies that the first term of (4.6) is bounded by $K\,(N\ln^2 N/T)\,x^*x$ which proves the first result.

As a consequence we obtain with $v_j=u_j$ and $r_j=t_j-L_j/2+1$

$$\sup_{|x|=1}|x^*\Sigma_T(A,B)x|\leqslant\sup_{|x|=1}\frac{S}{N}\sum_{j=1}^{M}|x^*K_T^{(j)\prime}\,W_T^{(j)}(A\bar{B})K_T^{(j)}x|+\mathrm{o}(1)$$

$$\leqslant C_1\sup_{|x|=1}\frac{S}{N}\sum_{j=1}^{M}\int_{-\pi}^{\pi}\left|\sum_{r=0}^{L_j-1}x_{r_j+r}\exp(-\mathrm{i}\lambda r)\right|^2\mathrm{d}\lambda+\mathrm{o}(1)$$

which leads to the result since each x_r is contained in exactly N/S blocks. The lower bound is obtained in the same way. Since $\Sigma=\Sigma_T(A,A)$ is symmetric and positive definite we get

$$\|\Sigma\|\leqslant\|\Sigma^{1/2}\|^2=\left(\sup_{|x|=1}x^*\Sigma x\right)\leqslant C_1$$

and

$$\|\Sigma^-\|\leqslant\|\Sigma^{-1/2}\|^2=\sup_x\frac{x^*\Sigma^{-1}x}{x^*x}=\sup_x\frac{x^*x}{x^*\Sigma x}=\left(\inf_{|x|=1}x^*\Sigma x\right)^{-1}\leqslant C_2^{-1}.$$

Lemma 4.5. *Suppose A fulfills Assumption* 4.3(*i*) *and there exists a* t^* *with* $x_j=0$ *for all* $j\notin\{t^*,\dots,t^*+L\}$. *Then we have for each* $t_0\in\{t^*,\dots,t^*+L\}$

$$x^*\Sigma_T(A,A)x=\int_{-\pi}^{\pi}\left|\sum_{j=t^*}^{t^*+L}x_j\exp(\mathrm{i}\lambda j)\right|^2\left|A\left(\frac{t_0}{T},\lambda\right)\right|^2\mathrm{d}\lambda+x^*x\,\mathrm{O}\left(\frac{L}{T}\ln^2 L\right).$$

Proof. The proof is completely analogous to the proof of Lemma 4.4. We only have to estimate the first term in (4.6) which leads to the result. □

By using Lemma 4.5 and the approximation of $\Sigma_T(A,A)$ by $W_T(|A|^2)$ we are now able to prove Theorem 3.2.

Proof of Theorem 3.2. We start by giving a lower bound for $v_{t+1,T}$. Let $L \in \mathbb{N}$, $b_0, \ldots, b_t$ be arbitrary real numbers with $b_0 = 1$, and

$$x_j = \begin{cases} b_{t+1-j} & j = 1, \ldots, t+1, \\ 0 & j = t+2, \ldots, T. \end{cases}$$

Lemma 4.4 gives with $r_j = t_j - L_j/2 + 1$ and

$$E\left(\sum_{j=0}^{t} b_j X_{t+1-j,T}\right)^2 = x^* \Sigma_T(A, A) x$$

$$= \frac{S}{N} \sum_{j=1}^{M} \int_{-\pi}^{\pi} \left| \sum_{r=0}^{L_j-1} x_{r_j+r} \exp(\mathrm{i}\lambda r) \right|^2 f(v_j, \lambda)\, \mathrm{d}\lambda + x^* x\, \mathrm{o}(1)$$

uniformly in t. We know that the index $t+1$ is contained exactly in N/S blocks. Selecting those blocks and choosing $v_j = t/T$ for those blocks gives as a lower bound for this expression

$$\frac{S}{N} \sum_{j=1}^{N/S} \int_{-\pi}^{\pi} \left| \sum_{r=0}^{\ell_j} b_r \exp(\mathrm{i}\lambda r) \right|^2 f\left(\frac{t}{T}, \lambda\right) \mathrm{d}\lambda + x^* x\, \mathrm{o}(1)$$

with some numbers $\ell_j \geqslant 1$. Each integral represents the prediction error of a predictor for a *stationary* time series with spectral density $f(t/T, \lambda)$, which leads to a lower bound

$$\exp\left\{\frac{1}{2\pi} \int_{-\pi}^{\pi} \log 2\pi f\left(\frac{t}{T}, \lambda\right) \mathrm{d}\lambda\right\} + x^* x\, \mathrm{o}(1)$$

(Kolmogorov's formula, see Brockwell and Davis (1987, Theorem 5.8.1)). Since $x^* x \leqslant x^* \Sigma_T(A, A) x \, \| \Sigma_T(A, A)^{-1} \|$ we have

$$x^* \Sigma_T(A, A) x (1 + \mathrm{o}(1)) \geqslant \exp\left\{\frac{1}{2\pi} \int_{-\pi}^{\pi} \log 2\pi f\left(\frac{t}{T}, \lambda\right) \mathrm{d}\lambda\right\}$$

and therefore,

$$\log v_{t+1,T} \geqslant \frac{1}{2\pi} \int_{-\pi}^{\pi} \log 2\pi f\left(\frac{t}{T}, \lambda\right) \mathrm{d}\lambda + \mathrm{o}_T(1) \tag{4.8}$$

uniformly in t.

To get an upper bound for $v_{t+1,T}$ we set $\bar{t} = \min(t, L)$ with $(L/T) \ln^2 L \to 0$, $L \to \infty$ and take those $b_0^*(\bar{t}), \ldots, b_{\bar{t}}^*(\bar{t})$ $(b_0^*(\bar{t}) = 1)$ that lead to an optimal one-step predictor based on $\bar{t}$ observations for a *stationary* time series with spectral density $h(\lambda) = f(t/T, \lambda)$. Let

$$x_j = \begin{cases} b_{t+1-j} & j = t+1-\bar{t}, \ldots, t+1, \\ 0 & \text{otherwise}. \end{cases}$$

We obtain with Lemma 4.5

$$v_{t+1,T} \leqslant E\left(\sum_{j=0}^{\bar{t}} b_j^*(\bar{t}) X_{t+1-j,T}\right)^2 = x^* \Sigma_T(A, A) x$$

$$= \int_{-\pi}^{\pi} \left| \sum_{j=0}^{\bar{t}} b_j^*(\bar{t}) \exp(i\lambda j) \right|^2 f\left(\frac{t}{T}, \lambda\right) d\lambda + x^* x \, O\left(\frac{L}{T} \ln^2 L\right).$$

Let $\Sigma_h = \{\int_{-\pi}^{\pi} h(\lambda) \exp(i\lambda(j-k)) \, d\lambda\}_{j,k=1,\dots,T}$. We have

$$x^* x \leqslant x^* \Sigma_h x \| \Sigma_h^{-1} \| \leqslant K \int_{-\pi}^{\pi} \left| \sum_{j=0}^{\bar{t}} b_j^*(\bar{t}) \exp(i\lambda j) \right|^2 f\left(\frac{t}{T}, \lambda\right) d\lambda$$

which leads to

$$v_{t+1,T} \leqslant \left\{ \int_{-\pi}^{\pi} \left| \sum_{j=0}^{\bar{t}} b_j^*(\bar{t}) \exp(i\lambda j) \right|^2 f\left(\frac{t}{T}, \lambda\right) d\lambda \right\} \left(1 + O\left(\frac{L}{T} \ln^2 L\right)\right\}.$$

Since the { }-term is the prediction error of a stationary time series we have

$$v_{t+1,T} \leqslant \exp\left\{\frac{1}{2\pi} \int_{-\pi}^{\pi} \log 2\pi f\left(\frac{t}{T}, \lambda\right) d\lambda + o_t(1)\right\} (1 + o_T(1))$$

which implies (i). To prove (ii) we note that for Gaussian processes

$$\det \Sigma_T(A, A) = \prod_{t=1}^{T} v_{t,T}.$$

Cesaro summability now gives

$$\frac{1}{T} \log \det \Sigma_T(A, A) = \frac{1}{T} \sum_{t=1}^{T} \log v_{t,T}$$

$$= \frac{1}{2\pi} \int_0^1 \int_{-\pi}^{\pi} \log 2\pi f(u, \lambda) \, d\lambda \, du + o_T(1).$$

Lemma 4.6. *Let* $k \in \mathbb{N}$, A_ℓ, B_ℓ *fulfill Assumption* 4.3(*i*), ϕ_ℓ *fulfill Assumption* 4.3(ii), μ_1, μ_2 *fulfill Assumption* 4.3(iii) *and* N, S *fulfill Assumption* 4.3(iv). *Then we have* (*with* $v_j = u_j = t_j/T$ *in the definition of* W_T)

(i) $$\frac{1}{T} \operatorname{tr}\left\{ \prod_{\ell=1}^{k} W_T(\phi_\ell) \Sigma_T(A_\ell, B_\ell) \right\}$$

$$= (2\pi)^{2k-1} \int_0^1 \int_{-\pi}^{\pi} \left\{ \prod_{\ell=1}^{k} \phi_\ell(u, \lambda) A_\ell(u, \lambda) B_\ell(u, -\lambda) \right\} d\lambda \, du$$

$$+ O(N^{-1} \ln^{2k} T + N/T)$$

(ii) $$\frac{1}{T}\mu'_{1T}\left\{\prod_{\ell=1}^{k-1} W_T(\phi_\ell)\,\Sigma_T(A_\ell,B_\ell)\right\} W_T(\phi_k)\,\mu_{2T}$$

$$= (2\pi)^{2k-1}\int_0^1 \left\{\sum_{\ell=1}^{k-1} \phi_\ell(u,0)\,A_\ell(u,0)\,B_\ell(u,0)\right\}\phi_k(u,0)\,\mu_1(u)\,\mu_2(u)\,\mathrm{d}u$$

$$+\,\mathrm{O}(N^{-1}\ln^{2k}T + N/T)$$

where $\mu_{iT} = (\mu_i(1/T), \ldots, \mu_i(T/T))'$.

In particular, we have the rate $\mathrm{O}(T^{-1/2}\ln^{2k}T)$ *with* $N = T^{1/2}$. *If in addition the* ϕ_ℓ *are twice differentiable in u with uniformly bounded derivative we obtain for the remainder terms the rate* $\mathrm{O}(N^{-1}\ln^{2k}T + N^2/T^2 + S/T)$ *leading with* $N = T^{2/3}$ *and* $S \leqslant T^{1/3}$ *to the rate* $\mathrm{O}(T^{-2/3}\ln^{2k}T)$.

Proof. (i) We give the proof for $k = 1$ and $k = 2$ in detail. Since the general case is similar to the case $k = 2$ we afterwards only give a sketch for general k. We have with $r_j = t_j - L_j/2 + 1$

$$\frac{1}{T}\operatorname{tr}\{W_T(\phi)\,\Sigma_T(A,B)\} = \frac{S}{NT}\sum_{j=1}^{M}\operatorname{tr}\{W_T^{(j)}(\phi)\,\Sigma_T^{(j,j)}(A,B)\}$$

$$= \frac{S}{NT}\sum_{j=1}^{M}\iint_{-\pi}^{\pi}\phi(u_j,\lambda)\,H_{L_j}(A^0_{r_j+\cdot,T}(\gamma),\lambda-\gamma)\,H_{L_j}(B^0_{r_j+\cdot,T}(-\gamma),\gamma-\lambda)\,\mathrm{d}\lambda\,\mathrm{d}\gamma. \tag{4.9}$$

We now replace $\phi(u_j,\lambda)$ by $\phi(u_j,\gamma)$ and integrate afterwards over λ. The replacement error is with Lemma 4.2 and (4.3) bounded by $K\,N^{-1}\ln N$, i.e. (4.9) is equal to

$$2\pi\frac{S}{NT}\sum_{j=1}^{M}\sum_{r=0}^{L_j-1}\int_{-\pi}^{\pi}\phi(u_j,\gamma)\,A^0_{r_j+r,T}(\gamma)\,B^0_{r_j+r,T}(-\gamma)\,\mathrm{d}\gamma + \mathrm{O}(N^{-1}\ln N)$$

$$= 2\pi\frac{S}{NT}\sum_{j=1}^{M}\sum_{r=0}^{L_j-1}\int_{-\pi}^{\pi}\phi(u_j,\gamma)\,A\left(\frac{r_j+r}{T},\gamma\right)B\left(\frac{r_j+r}{T},-\gamma\right)\mathrm{d}\gamma + \mathrm{O}(N^{-1}\ln N).$$

Since each point of $\{1,\ldots,T\}$ is contained in exactly N/S segments, this is equal to

$$2\pi\int_0^1\int_{-\pi}^{\pi}\phi(u,\gamma)\,A(u,\gamma)\,B(u,-\gamma)\,\mathrm{d}\gamma\,\mathrm{d}u + \mathrm{O}(N^{-1}\ln N) + \mathrm{O}\left(\frac{N}{T}\right).$$

If ϕ is twice differentiable in u we can obtain $\mathrm{O}(N^2/T^2)$ instead of $\mathrm{O}(N/T)$.

For $k = 2$ we get

$$\frac{1}{T}\operatorname{tr}\{W_T(\phi_1)\,\Sigma_T(A_1,B_1)\,W_T(\phi_2)\,\Sigma_T(A_2,B_2)\}$$

$$= \frac{S^2}{TN^2}\sum_{j,k=1}^{M}\int_{-\pi}^{\pi}\cdots\int_{-\pi}^{\pi}\phi_1(u_j,\lambda_1)\,\phi_2(u_k,\lambda_2)\exp\{\mathrm{i}(\gamma_1-\gamma_2)(r_j-r_k)\}$$

$$\times H_{L_j}(A^0_{1,r_j+\cdot,T}(\gamma_1),\lambda_1-\gamma_1)\,H_{L_k}(B^0_{1,r_k+\cdot,T}(-\gamma_1),\gamma_1-\lambda_2)$$

$$\times H_{L_k}(A^0_{2,r_k+\cdot,T}(\gamma_2),\lambda_2-\gamma_2)\,H_{L_j}(B^0_{2,r_j+\cdot,T}+(-\gamma_2),\gamma_2-\lambda_1)\,\mathrm{d}\gamma_1\,\mathrm{d}\gamma_2\,\mathrm{d}\lambda_1\,\mathrm{d}\lambda_2. \tag{4.10}$$

We now replace $\phi_1(u_j, \lambda_1)$ by $\phi_1(u_j, \gamma_1)$. We obtain with Lemma 4.2 and Lemma A.6 of Dahlhaus (1993)

$$\left| \sum_{k=1}^{M} \phi_2(u_k, \lambda_2)\, H_{L_k}(B^0_{1,r_k+\cdot,T}(-\gamma_1), \gamma_1 - \lambda_2) \right.$$

$$\left. H_{L_k}(A^0_{2,r_k+\cdot,T}(\gamma_2), \lambda_2 - \gamma_2) \exp\{-\mathrm{i}(\gamma_1 - \gamma_2) r_k\} \right|$$

$$\leqslant K\, L_N(\gamma_1 - \lambda_2)\, L_N(\lambda_2 - \gamma_2) \left\{ \frac{N}{S} + L_{T/S}(S(\gamma_1 - \gamma_2)) \right\}$$

which leads to the following upper bound for the replacement error

$$K \frac{S^2}{TN^2} \frac{T}{S} \int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} L_N(\gamma_2 - \lambda_1) L_N(\gamma_1 - \lambda_2) L_N(\lambda_2 - \gamma_2)$$

$$\times \left\{ \frac{N}{S} + L_{T/S}(S(\gamma_1 - \gamma_2)) \right\} \mathrm{d}\gamma_1\, \mathrm{d}\gamma_2\, \mathrm{d}\lambda_1\, \mathrm{d}\lambda_2 \leqslant K N^{-1} \ln^4 T$$

by using (4.3) and (4.4). Analogously, we replace $\phi_2(u_k, \lambda_2)$ by $\phi_2(u_k, \gamma_1)$ and integrate over λ_1 and λ_2. Thus, (4.10) is equal to

$$\frac{(2\pi)^2 S^2}{TN^2} \sum_{j,k=1}^{M} \iint_{-\pi}^{\pi} \phi_1(u_j, \gamma_1)\, \phi_2(u_k, \gamma_1) \exp\{\mathrm{i}(\gamma_1 - \gamma_2)(r_j - r_k)\}$$

$$\times \sum_{r=0}^{L_j - 1} A^0_{1,r_j+r,T}(\gamma_1)\, B^0_{2,r_j+r,T}(-\gamma_2) \exp\{\mathrm{i}(\gamma_1 - \gamma_2) r\}$$

$$\times \sum_{t=0}^{L_k - 1} A^0_{2,r_k+t,T}(\gamma_2)\, B^0_{1,r_k+t,T}(-\gamma_1) \exp\{-\mathrm{i}(\gamma_1 - \gamma_2) t\}\, \mathrm{d}\gamma_1\, \mathrm{d}\gamma_2$$

$$+ \mathrm{O}(N^{-1} \ln^4 T).$$

We now replace $A^0_{1,r_j+r,T}(\gamma_1)$ by $A_1((r_j + r)/T, \gamma_1)$. Similar to the above replacement the error is bounded by

$$K \frac{S^2}{TN^2} \frac{T}{S} \frac{N}{T} \iint_{-\pi}^{\pi} L_N(\gamma_1 - \gamma_2) \left\{ \frac{N}{S} + L_{T/S}(S(\gamma_1 - \gamma_2)) \right\} \mathrm{d}\gamma_1\, \mathrm{d}\gamma_2 \leqslant K T^{-1} \ln^2 T.$$

Analogously, we replace A^0_2, B^0_2 and B^0_1. We then replace $B_2((r_j + r)/T, -\gamma_2)$ by $B_2((r_j + r)/T, -\gamma_1)$. We obtain with Lemma 4.2 and Lemma A.6 of Dahlhaus (1993)

$$\left| \sum_{j=1}^{M} \phi_1(u_j, \gamma_1) \sum_{r=0}^{L_j - 1} A_1\left(\frac{r_j + r}{T}, \gamma_1\right) \left\{ B_2\left(\frac{r_j + r}{T}, -\gamma_2\right) \right. \right.$$

$$\left. \left. - B_2\left(\frac{r_j + r}{T}, -\gamma_1\right) \right\} \exp\{\mathrm{i}(\gamma_1 - \gamma_2)(r_j + r)\} \right|$$

$$\leqslant K |\gamma_1 - \gamma_2| L_N(\gamma_1 - \gamma_2) \left\{ \frac{N}{S} + L_{T/S}(S(\gamma_1 - \gamma_2)) \right\}$$

$$\leqslant K \left\{ \frac{N}{S} + L_{T/S}(S(\gamma_1 - \gamma_2)) \right\},$$

i.e. the replacement error is with (A.4) bounded by

$$K\frac{S^2}{TN^2}\iint_{-\pi}^{\pi} L_N(\gamma_1-\gamma_2)\left\{\frac{N}{S}+L_{T/S}(S(\gamma_1-\gamma_2))\right\}^2 \mathrm{d}\gamma_1\,\mathrm{d}\gamma_2 \leqslant KN^{-1}\ln T.$$

Similarly, we replace $A_2((r_k+t)/T,\gamma_2)$ by $A_1((r_k+t)/T,\gamma_1)$. We now set $C_1(u,\gamma_1)=A_1(u,\gamma_1)B_2(u,-\gamma_1)$ and $C_2(u,\gamma_1)=A_2(u,\gamma_1)B_1(u,-\gamma_1)$. Thus, (4.10) is equal to

$$\frac{(2\pi)^2S^2}{TN^2}\sum_{j,k=1}^{M}\sum_{r,t=0}^{L_j-1,L_k-1}\int_{-\pi}^{\pi}\phi_1(u_j,\gamma)\phi_2(u_k,\gamma)C_1((r_j+r)/T,\gamma)\,C_2((r_k+t)/T,\gamma)\,\mathrm{d}\gamma$$

$$\times\int_{-\pi}^{\pi}\exp\{\mathrm{i}\alpha(r_j+r-r_k-t)\}\,\mathrm{d}\alpha+\mathrm{O}(N^{-1}\ln^4 T)$$

$$=\frac{(2\pi)^3S}{TN}\sum_{j=1}^{M}\sum_{r=0}^{L_j-1}\int_{-\pi}^{\pi}\phi_1(u_j,\gamma)C_1((r_j+r)/T),\gamma)C_2((r_j+r)/T),\gamma)$$

$$\times\frac{S}{N}\sum_{k\in K_{j,r}}\phi_2(u_k,\gamma)\,\mathrm{d}\gamma$$

where $K_{j,r}$ is the set of all k such that $t:=r_j+r-r_k\in\{0,\ldots,L_k-1\}$, i.e. such that $r_k\leqslant r_j+r\leqslant r_k+L_k-1$. Thus, the sum is over all k where r_j+r lies in the kth segment. Due to the construction of the segments there are exactly N/S segments with this property. We now replace $\phi_2(u_k,\gamma)$ by $\phi_2((r_j+r)/T,\gamma)$. Since $|(r_j+r)/T-u_k|\leqslant N/T$ the replacement error is of order $\mathrm{O}(N/T)$. If ϕ_2 is twice differentiable in u we use a second order expansion of $\phi_2(u_k,\gamma)$ around $(r_j+r)/T$ leading to a replacement error of $\mathrm{O}(N^2/T^2)$ for the second order term and a replacement error of

$$K\frac{S}{TN}\sum_{j=1}^{M}\sum_{r=0}^{L_j-1}\frac{S}{N}\left|\sum_{k\in K_{j,r}}\frac{t_k-r_j-r}{T}\right|$$

for the first order term. For $2N/S\leqslant j\leqslant (T-N)/S$ we have

$$\sum_{k\in K_{j,r}}\frac{t_k-r_j-r}{T}=\sum_{\substack{t\\ -N/2\leqslant \ell S+N/2-r-1\leqslant N/2-1}}\frac{\ell S+N/2-r-1}{T}=\mathrm{O}\left(\frac{N}{T}\right)$$

(since all summands up to at most one can be grouped in groups of two with different signs whose sum is bounded by S/T). For the other j this sum is of order $\mathrm{O}(N^2/(ST))$. Therefore, we obtain for the replacement error of the first order term

$$\mathrm{O}\left(\frac{N^2}{T^2}+\frac{S}{T}\right).$$

Finally we replace the sum over j and r by the integral over u and obtain the assertion of the lemma with remainder $\mathrm{O}(N^{-1}\ln^4 T+N/T)$ leading with $N=T^{1/2}$ to the result. Under the stronger conditions on ϕ we obtain as the remainder $\mathrm{O}(N^{-1}\ln^4 T+N^2/T^2+S/T)$ leading with $N=T^{2/3}$ and $S=T^{1/3}$ to the better rate.

For general k the proof is analogous. We only indicate the main steps. We have with $\Pi = (-\pi, \pi]$

$$\frac{1}{T}\operatorname{tr}\left\{\prod_{l=1}^{k} W_T(\phi_l)\Sigma_T(A_l, B_l)\right\}$$

$$= \frac{1}{T}\frac{S^k}{N^k}\sum_{j_1,\ldots,j_k=1}^{M}\int_{\Pi^{2k}}\left\{\prod_{v=1}^{k}\phi_v(u_{jv}, \lambda_v)\right\}\exp\left\{i\sum_{v=1}^{k}\gamma_v(j_v - j_{v+1})\right\}$$

$$\times\left\{\prod_{v=1}^{k} H_{L_{j_v}}(A^0_{v,r_{j_v}+\cdot}(\gamma_v), \lambda_v - \gamma_v)H_{L_{j_{v+1}}}(B^0_{v,r_{j_{v+1}}+\cdot}(-\gamma_v), \gamma_v - \lambda_{v+1})\right\}\mathrm{d}\lambda\,\mathrm{d}\gamma$$

where $\lambda_{k+1} = \lambda_1$ and $j_{k+1} = j_1$. As in the case $k = 2$ we now replace $\phi_v(u_{j_v}, \lambda_v)$ by $\phi_v(u_{j_v}, \gamma_v)$ $(v = 1, \ldots, k)$ with replacement error $O(N^{-1}\ln^{2k} T)$, integrate over $\lambda_1, \ldots, \lambda_k$, and replace $A^0_{v,r_{j_v}+r}(\gamma_v)$ by $A_v((r_{j_v} + r)/T), \gamma_v)$ $(v = 1, \ldots, k)$ with replacement error $O(T^{-1}\ln^{2k} T)$ (the same for B^0_v), leading to

$$\frac{(2\pi)^k S^k}{TN^k}\sum_{j_v}\int_{\Pi^k}\left\{\prod_{v=1}^{k}\phi_v(u_{j_v}, \gamma_v)\right\}\exp\left\{i\sum_{v=1}^{k}\gamma_v(j_v - j_{v+1})\right\}$$

$$\prod_{v=1}^{k}\left[\sum_{t_v=0}^{L_{j_v}-1} A_v\left(\frac{r_{j_v} + t_v}{T}, \gamma_v\right)B_{v-1}\left(\frac{r_{j_{v-1}} + t_v}{T}, -\gamma_{v-1}\right)\exp\{i(\gamma_v - \gamma_{v-1})t\}\right]\mathrm{d}\gamma$$

where $j_0 = j_k, \gamma_0 = \gamma_k$. We now replace successively, e.g. $\phi_{k-1}(u, \gamma_{k-1})$ by $\phi_{k-1}(u, \gamma_{k-2})$, then the same by $\phi_{k-1}(u, \gamma_{k-3})$ etc. and finally by $\phi_{k-1}(u, \gamma_1)$ with replacement error $O(N^{-1}\ln^{2k} T)$; similarly the arguments in all A and B by γ_1 and $-\gamma_1$ with replacement error $O(N^{-1}\ln^{2k} T)$. Integration over $\gamma_2, \ldots, \gamma_k$ now leads to $(2\pi)^{k-1}$ if $r_{j_1} + t_1 = \cdots = r_{j_k} + t_k$ and 0 otherwise. The same arguments as in the case $k = 2$ now give the result.

(ii) Apart from a few changes the proof is analogous. We therefore only show the differences in the case $k = 2$. We have

$$\frac{1}{T}\mu_1' W_T(\phi_1)\Sigma_T(A, B)W_T(\phi_2)\mu_2$$

$$= \frac{S^2}{TN^2}\sum_{j,k=1}^{M}\sum_{t_1,\ldots,t_4=0}^{L_j-1,L_k-1}\mu_1\left(\frac{r_j + t_1}{T}\right)W_T^{(j)}(\phi_1)_{t_1t_2}\Sigma_T(A, B)^{(j,k)}_{t_2t_3}W_T^{(k)}(\phi_2)_{t_3t_4}\mu_2\left(\frac{r_k + t_4}{T}\right)$$

$$= \frac{S^2}{TN^2}\sum_{j,k=1}^{M}\iiint_{-\pi}^{\pi}\phi_1(u_j, \lambda_1)\phi_2(u_k, \lambda_2)\exp\{i\gamma(r_j - r_k)\}.$$

$$\times H_{L_j}\left(\mu_1\left(\frac{r_j + \cdot}{T}\right), -\lambda_1\right)H_{L_j}(A^0_{r_j+\cdot}(\gamma), \lambda_1 - \gamma)H_{L_k}(B^0_{r_k+\cdot}(-\gamma), \gamma - \lambda_2)$$

$$\times H_{L_k}\left(\mu_2\left(\frac{r_k + \cdot}{T}\right), \lambda_2\right)\mathrm{d}\lambda_1\,\mathrm{d}\lambda_2\,\mathrm{d}\gamma.$$

We now replace $\phi_1(u_j, \lambda_1)$ by $\phi_1(u_j, \gamma)$ and $\phi_2(u_j, \lambda_2)$ by $\phi_2(u_j, \gamma)$. The replacement error is $O(N^{-1}\ln^4 T)$. We then integrate over λ_1 and λ_2 which leads to

$$(2\pi)^2 \frac{S^2}{TN^2} \sum_{j,k=1}^{M} \int_{-\pi}^{\pi} \phi_1(u_j, \gamma)\phi_2(u_k, r)\exp\{i\gamma(r_j - r_k)\}.$$
$$\times H_{L_j}\left(\mu_1\left(\frac{r_j + \cdot}{T}\right)A^0_{r_j+\cdot}(\gamma), -\gamma\right)H_{L_k}\left(\mu_2\left(\frac{r_k + \cdot}{T}\right)B^0_{r_k+\cdot}(-\gamma), \gamma\right)d\gamma.$$

We then replace $A^0_{r_j+\cdot}(\gamma)$ by $A((r_j + \cdot)/T, \gamma)$ and $B^0_{r_k+\cdot}(-\gamma)$ by $B((r_k + \cdot)/T, -\gamma)$ with replacement error $O(T^{-1}\ln^2 T)$, i.e. we obtain

$$\frac{(2\pi)^2 S^2}{TN^2} \sum_{j,k=1}^{M} \sum_{r,t=0}^{L_j-1, L_k-1} \int_{-\pi}^{\pi} \phi_1(u_j, \gamma)\phi_2(u_k, \gamma)\mu_1\left(\frac{r_j + r}{T}\right)A\left(\frac{r_j + r}{T}, \gamma\right)$$
$$\times \mu_2\left(\frac{r_k + t}{T}\right)B\left(\frac{r_k + t}{T}, -\gamma\right)\exp\{i\gamma(r_j + r - r_k - t)\}\, d\gamma.$$

We now successively replace all arguments γ by 0 leading to an replacement error $O(N^{-1}\ln^2 T)$. After integration over γ the main term becomes

$$\frac{(2\pi)^3 S}{TN} \sum_{j=1}^{M} \sum_{r=0}^{L_j-1} \phi_1(u_j, 0)\,\mu_1\left(\frac{r_j + r}{T}\right)\mu_2\left(\frac{r_j + r}{T}\right)A\left(\frac{r_j + r}{T}, 0\right)$$
$$\times B\left(\frac{r_j + r}{T}, 0\right)\cdot \frac{S}{N} \sum_{k \in K_{j,r}} \phi_2(u_k, 0)$$

with the same $K_{j,r}$ as above. The result now follows analogously. □

We now prove that $W_T(\{4\pi^2 |A|^2\}^{-1})$ is an approximate inverse of $\Sigma_T(A, A)$.

Lemma 4.7. *Let A fullfill Assumption* 4.3(i) *and* $N = T^{1/2}$, $S \ll N$. *We then have with the* $T \times T$ *identity matrix* I_T

$$\frac{1}{\sqrt{T}} |I_T - \Sigma_T(A, A)^{1/2} W_T(\{4\pi^2 |A|^2\}^{-1})\Sigma_T(A, A)^{1/2}| = O(T^{-1/4}\ln^2 T). \quad (4.11)$$

If $N = T^{2/3}$, $S \leqslant T^{1/3}$ *and* $|A|^2$ *is twice differentiable in u, then the remainder is of order* $O(T^{-1/3}\ln^2 T)$.

Proof. The squared expression of (4.11) is equal to

$$1 - 2\frac{1}{T}\operatorname{tr}\{W_T \Sigma_T\} + \frac{1}{T}\operatorname{tr}\{W_T \Sigma_T W_T \Sigma_T\}.$$

Thus, Lemma 4.6(i) implies the result. □

Below we use this approximation result to establish a lemma on the trace behaviour of the matrices $\Sigma_T(A_\ell, B_\ell)$ and their inverses. Apart from the present paper this result is needed for the asymptotic treatment of the maximum likelihood estimator (Dahlhaus, 1996a).

The above approximation of $\Sigma_\theta = \Sigma_T(A_\theta, A_\theta)$ may also be used (together with Theorem 3.2(ii)) to construct an approximation of the Gaussian likelihood function (3.1), namely

$$\tilde{\mathscr{L}}_T(\theta) = \frac{1}{4\pi}\int_0^1\int_{-\pi}^{\pi} \log 4\pi^2 f_\theta(u,\lambda)\,\mathrm{d}\lambda\,\mathrm{d}u + \frac{1}{2T}(\boldsymbol{X}-\boldsymbol{\mu}_\theta)' W_T(\{4\pi^2|A_\theta|^2\}^{-1})(\boldsymbol{X}-\boldsymbol{\mu}_\theta)$$

$$= \frac{1}{4\pi}\int_0^1\int_{-\pi}^{\pi} \log 4\pi^2 f_\theta(u,\lambda)\,\mathrm{d}\lambda\,\mathrm{d}u + \frac{1}{4\pi}\frac{S}{N}\sum_{j=1}^{M}\int_{-\pi}^{\pi}\frac{I_{L_j}^{\mu_\theta}(u_j,\lambda)}{f_\theta(u_j,\lambda)}\mathrm{d}\lambda$$

where

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

i.e. $I_{L_j}^{\mu_\theta}(u_j,\lambda)$ is the periodogram on the jth segment.

This approximation is a generalization of the classical Whittle approximation for stationary processes (Whittle, 1953). The asymptotic properties of the resulting estimate $\tilde{\theta}_T = \arg\min \mathscr{L}_T(\theta)$ may be studied by using Lemma 4.6.

A similar estimate has been investigated in Dahlhaus (1993). The estimate studied in that paper does not use the first and the last N/S blocks but a data taper with the remaining blocks. Furthermore, we have replaced μ_θ by the empirical mean.

Lemma 4.8. *Let $k \in \mathbb{N}$; A_ℓ, B_ℓ, C_ℓ fulfill Assumption* 4.3(i) *and μ_1, μ_2 fulfill Assumption* 4.3(iii). *Let $\Sigma_\ell = \Sigma_T(A_\ell, B_\ell)$, $\Gamma_\ell = \Sigma_T(C_\ell, C_\ell)$, and $W_\ell = W_T(\{4\pi^2|C_\ell|^2\}^{-1})$ (with $v_j = u_j = t_j/T$ in the definition of W_T). Furthermore, let $V_\ell = \Gamma_\ell^{-1}$* or *$V_\ell = W_\ell$ ($\ell = 1, \ldots, k$). Then we have*

(i) $$\frac{1}{T}\operatorname{tr}\left\{\prod_{\ell=1}^{k} V_\ell \Sigma_\ell\right\} = \frac{1}{2\pi}\int_0^1\int_{-\pi}^{\pi}\left\{\prod_{\ell=1}^{k}\frac{A_\ell(u,\lambda)B_\ell(u,-\lambda)}{|C_\ell(u,\lambda)|^2}\right\}\mathrm{d}\lambda\,\mathrm{d}u$$
$$+\,\mathrm{O}(T^{-1/2}\ln^{2k+2}T).$$

(ii) $$\frac{1}{T}\mu_{1T}'\left\{\prod_{\ell=1}^{k-1} V_\ell\Sigma_\ell\right\}V_k\mu_{2T}$$
$$= \frac{1}{2\pi}\int_0^1\left\{\prod_{\ell=1}^{k-1}\frac{A_\ell(u,0)B_\ell(u,0)}{|C_\ell(u,0)|^2}\right\}|C_k(u,0)|^{-2}\mu_1(u)\mu_2(u)\,\mathrm{d}u$$
$$+\,\mathrm{O}(T^{-1/2}\ln^{2k+2}T).$$

If the C_ℓ are in addition twice differentiable in u, then the remainder terms are of order $\mathrm{O}(T^{2/3}\ln^{2k+2}T)$.

Proof. We start with a preliminary consideration. Let $A_i = \prod_{j=1}^{k_i}\Sigma_{ij}W_{ij}$ where W_{ij} and Σ_{ij} are matrices as in the assumption and $k_i \in \mathbb{N}_0$ (if $k_i = 0$ then A_i is the identity matrix). Furthermore, let D be a $T \times T$ matrix with $\|D\| \leqslant K$ and Σ is as in the assumption. Then we have

$$\left| \frac{1}{T} \operatorname{tr} \{ A_1' (\Gamma_1^{-1} - W_1) D (\Gamma_2^{-1} - W_2) A_2 \Sigma \} \right|$$

$$\leqslant \frac{1}{T} | \Sigma^{1/2} A_1' (\Gamma_1^{-1} - W_1) | \, |(\Gamma_2^{-1} - W_2) A_2 \Sigma^{1/2} | \, \| D \| . \tag{4.12}$$

Since $x^* W x / x^* x$ is bounded from below W^{-1} exists and $\| W^{-1/2} \|$ is bounded. Therefore,

$$\frac{1}{T} | \Sigma^{1/2} A_1' (\Gamma_1^{-1} - W_1) |^2 \leqslant \frac{1}{T} | \Sigma^{1/2} A_1' (\Gamma_1^{-1} - W_1) \Gamma_1 W_1^{1/2} |^2 \| \Gamma_1^{-1/2} \|^4 \| W_1^{-1/2} \|^2$$

$$\leqslant \frac{K}{T} \operatorname{tr} \{ A_1' (\Gamma_1^{-1} - W_1) \Gamma_1 W_1 \Gamma_1 (\Gamma_1^{-1} - W_1) A_1 \Sigma \}$$

$$= \frac{K}{T} (\operatorname{tr} \{ A_1' W_1 A_1 \Sigma \} - 2 \operatorname{tr} \{ A_1' W_1 \Gamma_1 W_1 A_1 \Sigma \} + \operatorname{tr} \{ A_1 W_1 \Gamma_1 W_1 \Gamma_1 W_1 A_1 \Sigma \})$$

$$= \mathrm{O}(T^{-1/2} (\ln T)^{4k_1 + 6})$$

by Lemma 4.6 which means that (4.12) is of magnitude $\mathrm{O}(T^{-1/2} (\ln T)^{2(k_1 + k_2) + 6})$.

Suppose now that the assertion holds for all $k \in \mathbb{N}$ and fixed $j := \# \{ \ell \mid V_\ell = \Gamma_\ell^{-1} \}$. For $j = 0$ this was proved in Lemma 4.6. For $j = 1$ we obtain from (4.12) with $D = \Gamma = \Gamma_1 = \Gamma_2$

$$\frac{1}{T} \operatorname{tr} \{ A_1' \Gamma^{-1} A_2 \Sigma \} = \frac{2}{T} \operatorname{tr} \{ A_1' W A_2 \Sigma \} - \frac{1}{T} \operatorname{tr} \{ A_1' W \Gamma W A_2 \Sigma \}$$

$$+ \mathrm{O}(T^{-1/2} (\ln T)^{2(k_1 + k_2) + 4})$$

leading to the result. For $j + 1$ let

$$\left(\prod_{\ell=1}^{k} V_\ell \Sigma_\ell \right) = \left(\prod_{\ell=1}^{k_1 - 1} W_\ell \Sigma_\ell \right) \Gamma_{k_1}^{-1} \Sigma_{k_1} \left(\prod_{\ell = k_1 + 1}^{k_2 - 1} V_\ell \Sigma_\ell \right) \Gamma_{k_2}^{-1} \left(\prod_{\ell = k_2 + 1}^{k} \Sigma_{\ell - 1} W_\ell \right) \Sigma_k$$

$$=: A_1' \Gamma_{k_1}^{-1} D \Gamma_{k_2}^{-1} A_2 \Sigma_k .$$

Eq. (4.12) again leads to the required upper bound. Under additional assumptions on the C_ℓ we get with the same arguments the stronger rate of convergence. To prove (ii) we show by similar arguments that

$$\left| \frac{1}{T} \mu_{1T}' A_1' (\Gamma_1^{-1} - W_1) D (\Gamma_2^{-1} - W_2) A_2 \mu_{2T} \right| = \mathrm{O}(T^{-1/2} (\ln T)^{2(k_1 + k_2) + 6}). \tag{4.13}$$

The rest of the proof is analogous.

Acknowledgements

I am grateful to L. Giraitis, A. Rijkeboer and R. von Sachs for several helpful discussions.

References

N.K. Bary, A Treatise on Trigonometric Series, Vol. I (Pergamon, Oxford, 1964).

P. Brockwell and R.A. Davis, Time Series: Theory and Methods (Springer, New York, 1987).

L. Cohen, Time-frequency distributions – a review. Proc. IEEE 77 (1989) 941–981.

R. Dahlhaus, Fitting time series models to nonstationary processes, Beiträge zur Statistik No. 4, (Universität Heidelberg, 1993).

R. Dahlhaus, Maximum likelihood estimation and model selection for nonstationary processes, J. Nonparametric Statist. 6 (1996a) to appear.

R. Dahlhaus, Asymptotic statistical inference for nonstationary processes with evolutionary spectra. Preprint, Universität Heidelberg (1996b).

R.B. Davies, Asymptotic inference in stationary Gaussian time-series, Adv. Appl. Probab. 5 (1973) 469–497.

F.A. Graybill, Matrices with Applications in Statistics (Wadsworth. Belmont, California, 2nd ed., 1983).

M. Hallin, Mixed autoregressive moving-average multivariate processes with time-dependent coefficients, J Multivariate Anal. 8 (1978) 567–572.

M. Hallin, Nonstationary q-dependent processes and time-varying moving average models: invertibility properties and the forecasting problem, Adv. Appl. Probab. 18 (1986) 170–210.

G. Kitagawa and W. Gersch, A smoothness priors time-varying AR coefficient modeling of nonstationary covariance time series, IEEE Trans. Automat. Control, AC 30 (1985) 48–56.

H.R. Künsch, A note on causal solutions for locally stationary AR processes, preprint (ETH Zürich, 1995).

W. Martin and P. Flandrin, Wigner-Ville spectral analysis of nonstationary processes, IEEE Trans. Acoust. Speech Signal Process. 33 (1985) 1461–1470.

G. Mélard, An example of the evolutionary spectrum theory, J. Time Ser. Anal. 6 (1985) 81–90.

G. Mélard and A. Harteleer-de Schutter, Contributions to evolutionary spectral theory, J. Time Ser. Anal. 10 (1989) 41–63.

K.S. Miller, Linear Difference Equations (Benjamin, New York, 1968).

E. Parzen, in: D.R. Brillinger, P.R. Krishnaiah, eds., Handbook of Statistics, Vol. 3, (Elsevier, Amsterdam 1983) pp. 221-247.

M.B. Priestley, Evolutionary spectra and non-stationary processes, J. Roy. Statist. Soc. Ser. B 27 (1965) 204–237.

M.B. Priestley, Spectral Analysis and Time Series, Vol. 2, (Academic Press, London, 1981).

P.M. Robinson, Nonparametric estimation of time-varying parameters, in: P. Hackl, ed., Statistical Analysis and Forecasting of Economic Structural Change (Springer, Berlin, 1989) pp. 253–264.

T. Subba Rao, Discussion on the paper by Priestley and Tong "On the analysis of bivariate non-stationary processes (153-166)", J.S. Roy. Statist. Soc. Ser. B 35 (1973) 182–183.

P.J. Thomson and P. de Souza, P., in: E.J. Hannan, P.R. Krishnaiah, M.M. Rao, eds., Handbook of Statistics, Vol. 5 (Elsevier, Amsterdam, 1985) pp. 389–412.

D. Tjøstheim, Spectral generating operators for non-stationary processes, Adv. Appl. Probab 8 (1976) 831–846.

P. Whittle, Estimation and information in stationary time series, Ark. Mat. 2 (1953) 423–434.