The Annals of Statistics
2000, Vol. 28, No. 6, 1762–1794

A LIKELIHOOD APPROXIMATION FOR LOCALLY STATIONARY PROCESSES

BY RAINER DAHLHAUS

Universität Heidelberg

A new approximation to the Gaussian likelihood of a multivariate locally stationary process is introduced. It is based on an approximation of the inverse of the covariance matrix of such processes. The new quasi likelihood is a generalization of the classical Whittle likelihood for stationary processes. Several approximation results are proved for the likelihood function. For parametric models, asymptotic normality and efficiency of the resulting estimator are derived for Gaussian locally stationary processes.

1. Introduction. Suppose we observe data $X_1, \dots, X_T$ from some nonstationary process and we want to fit a parametric model to the data. An example is an autoregressive process with time varying coefficients where we model the coefficient functions by polynomials in time. To estimate the parameters of such a model we introduce in this paper a generalization of the Whittle likelihood to nonstationary processes. For a lot of models the maximization of the new likelihood has computational advantages over the maximization of the exact Gaussian likelihood.

For univariate stationary processes with mean zero, Whittle (1953, 1954) introduced

$$\frac{1}{4\pi}\int_{-\pi}^{\pi}\left\{\log 4\pi^2 f_\theta(\lambda) + \frac{I_T(\lambda)}{f_\theta(\lambda)}\right\}d\lambda, \tag{1.1}$$

where

$$I_T(\lambda) = \frac{1}{2\pi T}\left|\sum_{t=1}^{T} X_t \exp(-i\lambda t)\right|^2$$

is the periodogram as an approximation of the negative Gaussian likelihood. This likelihood has been used over the years in many different situations. Among the large number of papers on Whittle estimates we mention the results of Dzhaparidze (1971) and Hannan (1973) for univariate time series, Dunsmuir (1979) for multivariate time series and Hosoya and Taniguchi (1982) for misspecified multivariate time series. A general overview over Whittle estimates for stationary models may be found in the monograph of Dzhaparidze (1986). We also mention the results of Mikosch, Gadrich, Klüppelberg and Adler (1995) on Whittle estimates for linear processes where the innovations have heavy tailed distributions, of Fox and Taqqu (1986) on Whittle estimates

Received January 1999; revised August 2000.
AMS 2000 *subject classifications*. Primary 62M10; secondary 62F10.
Key words and phrases. Locally stationary, Whittle likelihood, local likelihood, preperiodogram, generalized Toeplitz matrices.

for long-range dependent processes and of Robinson (1995) on semiparametric Whittle estimates for long-range dependent processes.

For nonstationary processes Dahlhaus (1997) used a modification of the above likelihood where the periodogram was calculated over a time segment and the resulting likelihood was averaged over the different segments. However, the use of a classical periodogram always contains some implicit smoothing over time [see (1.2) below] resulting in an additional bias in the nonstationary case.

For this reason we use in this paper a likelihood based on the preperiodogram which in the univariate case with mean zero takes the form

$$\widetilde{I}_T(u,\lambda) := \frac{1}{2\pi} \sum_{k:1\le[uT+1/2\pm k/2]\le T} X_{[uT+0.5+k/2]} X_{[uT+0.5-k/2]} \exp(-i\lambda k),$$

where $u \in [0,1]$ is the rescaled time. $\widetilde{I}_T(t/T,\lambda)$ may be regarded as a local version of the periodogram at time t. It was introduced by Neumann and von Sachs (1997) as a starting point for a wavelet estimate of the time-varying spectral density. The above form contains a small modification (time shift).

There exists a nice relation between the preperiodogram and the ordinary periodogram:

$$\begin{aligned}
I_T(\lambda) &= \frac{1}{2\pi T}\left|\sum_{r=1}^{T} X_r \exp(-i\lambda r)\right|^2 \\
&= \frac{1}{2\pi}\sum_{k=-(T-1)}^{T-1}\left(\frac{1}{T}\sum_{t=1}^{T-|k|} X_t X_{t+|k|}\right)\exp(-i\lambda k) \\
&= \frac{1}{T}\sum_{t=1}^{T}\frac{1}{2\pi}\sum_{k:1\le[uT+1/2\pm k/2]\le T} X_{[t+0.5+k/2],T}X_{[t+0.5-k/2],T}\exp(-i\lambda k) \\
&= \frac{1}{T}\sum_{t=1}^{T}\widetilde{I}_T\left(\frac{t}{T},\lambda\right),
\end{aligned} \tag{1.2}$$

that is, the periodogram is the average of the preperiodogram over time. Equation (1.2) means that the periodogram $I_T(\lambda)$ is the Fourier transform of the covariance estimator of lag k over the whole segment while the preperiodogram $\widetilde{I}_T(t/T,\lambda)$ just uses the pair $X_{[t+0.5+k/2]}X_{[t+0.5-k/2]}$ as a kind of "local estimator" of the covariance of lag k at time t (note that $[t+0.5+k/2]-[t+0.5-k/2]=k$). For this reason Neumann and von Sachs also called $\widetilde{I}_T(t/T,\lambda)$ the localized periodogram.

A classical kernel estimator of the spectral density of a stationary process at some frequency λ_0 therefore can be regarded as an average of the preperiodogram over all time points and over the frequencies in the neighbourhood of λ_0. It is therefore plausible that averaging the preperiodogram about some frequency λ_0 and about some time point t_0 gives an estimate of the time-varying spectrum $f(t_o/T, \lambda)$ (e.g., in the framework of a locally stationary process).

If we replace $I_T(\lambda)$ in (1.1) by the above average of the preperiodogram and afterwards replace the model spectral density $f_\theta(\lambda)$ by the time-varying spectral density $f_\theta(u, \lambda)$ of a nonstationary model, we obtain the expression

$$\frac{1}{4\pi}\frac{1}{T}\sum_{t=1}^{T}\int_{-\pi}^{\pi}\left\{\log 4\pi^2 f_\theta\left(\frac{t}{T}, \lambda\right)d\lambda + \frac{\widetilde{I}_T(t/T, \lambda)}{f_\theta(t/T, \lambda)}\right\},$$

which is the univariate form of the likelihood (2.6) investigated in this paper. If the model is stationary, that is, $f_\theta(u, \lambda) = f_\theta(\lambda)$ then the above likelihood is identical to the Whittle likelihood and we have a true generalization to nonstationary processes.

First we investigate in Section 2 the properties of the above quasi likelihood. It is shown that this likelihood can be derived from the Gaussian likelihood by using a certain approximation of the inverse of the covariance matrix appearing in the Gaussian likelihood and by using an extension of the Szegö formula [cf. Grenander and Szegö (1958), Section 5.2] to the nonstationary case.

In Section 3 we prove consistency and asymptotic normality of the resulting estimator and investigate some modifications. The Appendix contains some technical results on norms and matrix products of generalized Toeplitz matrices.

2. The local likelihood approximation. We start with the definition of a multivariate Gaussian locally stationary process. It is given in the form of a time-varying spectral representation. The equivalent form of a time-varying MA(∞)-representation is discussed below.

DEFINITION 2.1. A sequence of Gaussian multivariate stochastic processes $X_{t,T} = (X_{t,T}^{(1)}, \dots, X_{t,T}^{(d)})'$ $(t = 1, \dots, T)$ is called locally stationary with transfer function matrix A^o and mean function vector μ if there exists a representation

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

with the following properties:

(i) $\xi(\lambda)$ is a complex valued Gaussian vector process on $[-\pi, \pi]$ with $\overline{\xi_a(\lambda)} = \xi_a(-\lambda)$, $E\xi_a(\lambda) = 0$ and

$$E\{d\xi_a(\lambda) d\xi_b(\mu)\} = \delta_{ab}\eta(\lambda + \mu)\, d\lambda\, d\mu,$$

where $\eta(\lambda) = \sum_{j=-\infty}^{\infty} \delta(\lambda + 2\pi j)$ is the period 2π extension of the Dirac delta function.

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

$$\sup_{t,\lambda} \left| A^{o}_{t,T}(\lambda)_{ab} - A\left(\frac{t}{T},\lambda\right)_{ab} \right| \leq KT^{-1} \tag{2.2}$$

for all $a, b = 1, \dots, d$ and $T \in \mathbb{N}$. $A(u,\lambda)$ and $\mu(u)$ are assumed to be continuous in u.

$f(u,\lambda) := A(u,\lambda)\overline{A(u,\lambda)}'$ is the time varying spectral density matrix of the process.

REMARK 2.2 (Time-varying MA(∞)-representations). There exists a close connection between the above spectral representation and time-varying MA-representations. Let

$$a_{t,T,k} := \int_{-\pi}^{\pi} A^{o}_{t,T}(\lambda) \exp(i\lambda k)\, d\lambda,$$

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

and

$$\varepsilon_t := \int_{-\pi}^{\pi} \exp(i\lambda t)\, d\xi(\lambda)$$

[note that $a_{t,T,k}$ and $a(u,k)$ are matrices; the Fourier transform is calculated componentwise]. Then $E\varepsilon_t = 0$ and $E\varepsilon_s\varepsilon_t' = 2\pi\delta_{st}I_d$, that is the ε_t are uncorrelated (independent in the present Gaussian case). Since

$$A^{o}_{t,T}(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} a_{t,T,k} \exp(-i\lambda k)$$

and

$$A(u,\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} a_k(u) \exp(-i\lambda k)$$

we obtain

$$X_{t,T} = \mu\left(\frac{t}{T}\right) + \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} a_{t,T,k}\varepsilon_{t-k}. \tag{2.3}$$

Condition (2.2) implies

$$\sup_{t,k} \left| \left\{ a_{t,T,k} - a_k\left(\frac{t}{T}\right) \right\}_{bc} \right| = O(T^{-1})$$

for all $b, c = 1, \dots, d$. If we start conversely with an infinite MA-representation (2.3) where the coefficients fulfill

$$\sup_{t} \sum_{k=-\infty}^{\infty} \left| \left\{ a_{t,T,k} - a_k\left(\frac{t}{T}\right) \right\}_{bc} \right| = O(T^{-1}) \tag{2.4}$$

for all $b, c = 1, \dots, d$, then it can be shown in the same way that a representation (2.1) exists and (2.2) is fulfilled. Note that heteroscedastic ε_t and ε_t with dependent components can be included by choosing other $a_{t,T,k}$ in (2.3). The complicated construction with different functions $A^o_{t,T}(\lambda)$ and $A(t/T, \lambda)$ [$a_{t,T,k}$ and $a_k(t/T)$, respectively] is necessary since we need on the one hand a certain smoothness in time direction [guaranteed by the functions $A(u, \lambda)$ and $a_k(u)$] and on the other hand a class which is rich enough to cover interesting applications. For example, the time varying AR(1)-process $X_{t,T} = \phi(t/T)X_{t-1,T} + \varepsilon_t$ does not have a solution of the form $X_{t,T} = \sum_{k=0}^{\infty} a_k(t/T)\varepsilon_{t-k}$ but only of the form $X_{t,T} = \sum_{k=0}^{\infty} a_{t,T,k}\varepsilon_{t-k}$ with (2.4) where $a_k(u) = \phi(u)^k$.

Processes with an evolutionary spectral representation were introduced and investigated by Priestley (1965, 1981) and Granger and Hatanaka (1964). The above definition is the multivariate generalization of the definition of univariate local stationarity given in Dahlhaus (1997). As in nonparametric regression the time parameter $u = t/T$ in $\mu(u)$ and $A(u, \lambda)$ is rescaled for a meaningful asymptotic theory leading to the above triangular array $X_{t,T}$. The classical asymptotics for stationary sequences is contained as a special case (if μ and A do not depend on t). A detailed discussion of this definition and a comparison with Priestley's approach can be found in Dahlhaus (1996c). Another definition of local stationarity has recently been given by Mallat, Papanicolaou and Zhang (1998). We remark that the methods presented in this paper do not depend on the special definition of local stationarity.

Examples of locally stationary processes in the univariate case can be found in Dahlhaus (1996a). For the multivariate case we give the following examples.

EXAMPLE 2.3. (i) Suppose Y_t is a multivariate stationary process, $\mu(\cdot)$ is a vector function and $\Sigma(\cdot)$ is a matrix function. Then

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

is locally stationary. If Y_t is an iid sequence we have the situation of multivariate nonparametric regression.

(ii) Suppose $X_{t,T}$ is a time-varying multivariate ARMA-model, that is, $X_{t,T}$ is defined by the difference equations

$$\sum_{j=0}^{p} \Phi_j\left(\frac{t}{T}\right)\left[X_{t-j,T} - \mu\left(\frac{t-j}{T}\right)\right] = \sum_{j=0}^{q} \Psi_j\left(\frac{t}{T}\right)\Sigma^{\varepsilon}\left(\frac{t-j}{T}\right)\varepsilon_{t-j},$$

where ε_t are iid with mean zero and variance–covariance matrix I_d and $\Phi_o(u) \equiv \Psi_o(u) \equiv I_d$. For $z \in \mathbb{C}$ let $\Phi(u, z) = \sum_{j=0}^{p} \Phi_j(u)z^j$ and $\Psi(u, z) = \sum_{j=0}^{q} \Psi_j(u)z^j$. If $\det \Phi(u, z) \neq 0$ for all $|z| \leq 1 + c$ with $c > 0$ uniformly in u and all entries of $\Phi_j(u)$ and $\Psi_j(u)$ are continuous is u then it can be shown similarly to the univariate case [Dahlhaus (1996a), Theorem 2.3] that the solution of these difference equations has an infinite time-varying MA-representation, that is, the solution is locally stationary of the form (2.1). The time-varying spectral density of the process is

$$f(u,\lambda) = \frac{1}{2\pi}\Phi(u, e^{i\lambda})^{-1}\Psi(u, e^{i\lambda})\Sigma^{\varepsilon}(u)\Psi(u, e^{-i\lambda})'\Phi(u, e^{-i\lambda})'^{-1}.$$

We omit details of the derivation. However, we remark that in this case the functions $A^{o}_{t,T}(\lambda)$ and $A(t/T,\lambda)$ again do not coincide. They only fulfill (2.2).

In the following we look at parametric locally stationary models. An example is the case where the curves in the above examples are parametrized in time, for example by polynomials [for an example, see Dahlhaus (1997), Section 6]. As an estimator for the parameters we consider

$$\hat{\theta}_T := \operatorname*{argmin}_{\theta\in\Theta} \mathscr{L}_T(\theta), \tag{2.5}$$

where

$$\begin{aligned}\mathscr{L}_T(\theta) = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{4\pi}\int_{-\pi}^{\pi}\Big\{&\log\Big[(2\pi)^{2d}\det f_\theta\Big(\frac{t}{T},\lambda\Big)\Big] \\ &+ \operatorname{tr}\big[f_\theta(t/T,\lambda)^{-1}\tilde{I}_T^{\mu_\theta}(t/T,\lambda)\big]\Big\}\,d\lambda\end{aligned} \tag{2.6}$$

and

$$\begin{aligned}\tilde{I}_T^{\mu}(u,\lambda)_{ab} := \frac{1}{2\pi}\sum_{k:1\le[uT+1/2\pm k/2]\le T}&\Big[X^{(a)}_{[uT+0.5+k/2],T} - \mu^{(a)}\Big(\frac{[uT+0.5+k/2]}{T}\Big)\Big] \\ &\times\Big[X^{(b)}_{[uT+0.5-k/2],T} - \mu^{(b)}\Big(\frac{[uT+0.5-k/2]}{T}\Big)\Big]\exp(-i\lambda k)\end{aligned} \tag{2.7}$$

is the multivariate version of the preperiodogram. Here $[x]$ denotes the largest integer less or equal to x; Θ is assumed to be compact.

In the univariate case and for $\mu_\theta = 0$ this is the likelihood we have already discussed in the introduction. If the mean is not zero and one is not interested in modelling the mean, one may use $\tilde{I}_T^{\hat{\mu}}(u,\lambda)$ instead of $\tilde{I}_T^{\mu_\theta}(u,\lambda)$ where $\hat{\mu}$ is the arithmetic mean or some kernel estimate (if the mean is not believed to be constant over time).

In certain situations the first term reduces to a simpler form and the frequency integral may be replaced by a sum over the Fourier frequencies; see Remark 3.5 for details.

In the main theorem of this paper (Theorem 3.1) we prove $\hat{\theta}_T \to^{\mathscr{P}} \theta_0$ and a central limit theorem for $\sqrt{T}(\hat{\theta}_T - \theta_0)$ where θ_0 is the true parameter. If the model is misspecified the same holds with

$$\theta_0 := \operatorname*{argmin}_{\theta\in\Theta} \mathscr{L}(\theta), \tag{2.8}$$

where

$$\begin{aligned}\mathscr{L}(\theta) := \frac{1}{4\pi}\int_0^1\int_{-\pi}^{\pi}\Big\{\log[(2\pi)^{2d}\det f_\theta(u,\lambda)] + \operatorname{tr}[f_\theta(u,\lambda)^{-1} f(u,\lambda)]\Big\}\,d\lambda\,du \\ + \frac{1}{4\pi}\int_0^1 (\mu_\theta(u) - \mu(u))' f_\theta^{-1}(u,0)(\mu_\theta(u) - \mu(u))\,du\end{aligned} \tag{2.9}$$

is the limit of $\mathscr{L}_T(\theta)$ (see Theorem 2.8 below). In the case where the model is correctly specified, that is, $A(u,\lambda) = A_{\theta^*}(u,\lambda)$ and $\mu(u) = \mu_{\theta^*(u)}$ with some $\theta^* \in \Theta$, one can show that $\theta_0 = \theta^*$.

A deeper investigation of $\mathscr{L}_T(\theta)$ shows that it can be derived from the Gaussian log-likelihood by using a certain approximation of the inverse of the covariance matrix. Since this approximation also plays a key role in our proofs, we have to discuss it in some detail.

Let $\underline{X} = (X'_{1,T}, \dots, X'_{T,T})'$, $\mu = (\mu(1/T)', \dots, \mu(T/T)')'$, and $\Sigma_T(A,B)$ and $U_T(\phi)$ be $T \times T$ block matrices whose (r,s) block is

$$\Sigma_T(A,B)_{r,s} = \int_{-\pi}^{\pi} \exp(i\lambda(r-s)) A^o_{r,T}(\lambda) B^o_{s,T}(-\lambda)'\,d\lambda \tag{2.10}$$

and

$$U_T(\phi)_{r,s} = \int_{-\pi}^{\pi} \exp(i\lambda(r-s))\phi\left(\frac{1}{T}\left[\frac{r+s}{2}\right]^*, \lambda\right) d\lambda \tag{2.11}$$

$(r,s = 1,\dots,T)$ where $A^o_{r,T}(\lambda)$, $B^o_{r,T}(\lambda)$ and $\phi(u,\lambda)$ are $d \times d$-matrices and $[x]^* = [x]$ denotes the largest integer less or equal to x (we have added here the * to discriminate the notation from brackets). Direct calculation shows that

$$\begin{aligned}\mathscr{L}_T(\theta) = \frac{1}{4\pi}\frac{1}{T}\sum_{t=1}^{T}\int_{-\pi}^{\pi} \log\left[(2\pi)^{2d}\det f_\theta\left(\frac{t}{T},\lambda\right)\right] d\lambda \\ + \frac{1}{8\pi^2 T}(\underline{X} - \mu_\theta)' U_T(f_\theta^{-1})(\underline{X} - \mu_\theta).\end{aligned} \tag{2.12}$$

Furthermore, the logarithm of the exact Gaussian likelihood is

$$\widetilde{\mathscr{L}_T}(\theta) := \frac{d}{2}\log(2\pi) + \frac{1}{2T}\log\det\Sigma_\theta + \frac{1}{2T}(\underline{X} - \underline{\mu}_\theta)'\Sigma_\theta^{-1}(\underline{X} - \underline{\mu}_\theta), \tag{2.13}$$

where $\Sigma_\theta = \Sigma_T(A_\theta, A_\theta)$. We set

$$\tilde{\theta}_T := \operatorname*{argmin}_{\theta\in\Theta} \widetilde{\mathscr{L}_T}(\theta). \tag{2.14}$$

In Proposition 2.4 below we now prove that $U_T((1/4\pi^2) f_\theta^{-1})$ is an approximation of Σ_θ^{-1}. Together with the generalization of the multivariate Szegö identity (Proposition 2.5) this implies that $\mathscr{L}_T(\theta)$ is an approximation of $\widetilde{\mathscr{L}_T}(\theta)$ (see Theorem 2.8). If the model is stationary, then A_θ is constant in time and $\Sigma_\theta = \Sigma_T(A_\theta, A_\theta)$ is the Toeplitz matrix of the spectral density $f_\theta(\lambda) = |A_\theta|^2$ while $U_T((1/4\pi^2)f_\theta^{-1})$ is the Toeplitz matrix of $(1/4\pi^2)f_\theta^{-1}$. This is the classical matrix approximation leading to the Whittle likelihood [cf. Dzhaparidze (1986)]. $U_T((1/4\pi^2)f_\theta^{-1})$ is a better approximation of Σ_θ^{-1} than the approximation used in Dahlhaus (1996a) which consisted of overlapping block Toeplitz matrices. This can be seen from the rate of approximation in Proposition 2.4 below and the corresponding rate in Lemma 4.7 of Dahlhaus (1996a). As a consequence this also leads to a better rate of the likelihood approximation.

The technical parts of the following proofs consist of the derivation of properties of products of matrices $\Sigma_T(A, B), \Sigma_T(A, A)^{-1}$ and $U_T(\phi)$. These properties are derived in the Appendix. In particular, Lemmas A.1, A.5, A.7 and A.8 are of relevance for the following proofs.

For convenience we refer in the following proposition to Assumption A.3 in the Appendix concerning the smoothness of the transfer function and the mean. These conditions are fulfilled under Assumption 2.6 below. By $\|A\|$ and $|A|$ we denote the spectral norm and the Euclidean norm of a matrix A [see (A.1) and (A.2)].

PROPOSITION 2.4. *Suppose the matrices A and ϕ fulfill the smoothness conditions of Assumption* A.3(i)–(iii) (*Appendix*) *with existing and bounded derivatives $(\partial^2/\partial u^2)(\partial/\partial\lambda)A(u,\lambda)_{ab}$ and eigenvalues of $\phi(u,\lambda)$ which are bounded from below uniformly in u and λ. Then we have for each $\varepsilon > 0$,*

$$\frac{1}{T}\left|\Sigma_T(A, A)^{-1} - U_T(\{4\pi^2 A\overline{A}'\}^{-1})\right|^2 = O(T^{-1+\varepsilon}) \tag{2.15}$$

and

$$\frac{1}{T}\left|U_T(\phi)^{-1} - U_T(\{4\pi^2\phi\}^{-1})\right|^2 = O(T^{-1+\varepsilon}).$$

PROOF. Let $\Sigma_T = \Sigma_T(A, A)$ and $U_T = U_T(\{4\pi^2 A\overline{A}'\}^{-1})$. We obtain with Lemma A.1(b) and Lemma A.5,

$$\begin{aligned}\frac{1}{T}\left|\Sigma_T^{-1} - U_T\right|^2 &\le \frac{1}{T}\left|I - \Sigma_T^{1/2}U_T\Sigma_T^{1/2}\right|^2\|\Sigma_T^{-1}\|^2\\ &\le K\left(d - \frac{2}{T}\mathrm{tr}\{U_T\Sigma_T\} + \frac{1}{T}\mathrm{tr}\{U_T\Sigma_T U_T\Sigma_T\}\right).\end{aligned}$$

Lemma A.7(i) now implies the result. The second result is obtained in the same way with Lemma A.8. □

By using the above approximation it is possible to prove the following generalization of the Szegö identity [cf. Grenander and Szegö (1958), Section 5.2] to multivariate locally stationary process.

PROPOSITION 2.5. *Suppose A fulfills Assumptions* A.3(i), (ii), *with bounded derivatives* $(\partial^2/\partial u^2)(\partial/\partial\lambda)A(u,\lambda)_{ab}$. *Then we have with* $f(u,\lambda) = A(u,\lambda) A(u,-\lambda)'$ *for each* $\varepsilon > 0$,

$$\frac{1}{T}\log\det\Sigma_T(A,A) = \frac{1}{2\pi}\int_0^1\int_{-\pi}^{\pi}\log[(2\pi)^d\det f(u,\lambda)]\,d\lambda\,du + O(T^{-1+\varepsilon}).$$

If $A = A_\theta$ *depends on a parameter* θ *and fulfills the smoothness conditions of Assumption* 2.6(iii), (iv), *then the* $O(T^{-1+\varepsilon})$ *term is uniform in* θ.

The proof can be found in the Appendix.

In certain situations the right-hand side can be written in terms of the one step prediction error at each time point which often leads to a simpler form of the likelihood; see Remark 3.5(i) for details.

We now discuss the properties of the likelihood $\mathscr{L}_T(\theta)$. We set $\nabla_i = \partial/\partial\theta_i$ and $\nabla_{ij}^2 = \partial^2/\partial\theta_i\partial\theta_j$. The results are proved under the following assumptions.

ASSUMPTION 2.6. (i) We observe a realization $X_{1,T},\dots,X_{T,T}$ of a d-dimensional locally stationary Gaussian process with true mean function vector μ, transfer function matrix A^o and covariance matrix $\Sigma = \Sigma_T(A,A)$. We fit a class of locally stationary Gaussian processes with mean function vector μ_θ, transfer function matrix A_θ^o and covariance matrix $\Sigma_\theta = \Sigma_T(A_\theta, A_\theta)$, $\theta \in \Theta \subset \mathbb{R}^p$, Θ compact.

(ii) $\theta_0 = \operatorname{argmin} \mathscr{L}(\theta)$ exists uniquely and lies in the interior of Θ.

(iii) The components of $A_\theta(u,\lambda)$ are differentiable in θ and u and λ with uniformly continuous derivatives $\nabla_{ij}^2(\partial^2/\partial u^2)(\partial/\partial\lambda)A_\theta(u,\lambda)_{ab}$.

(iv) All eigenvalues of $f_\theta(u,\lambda) = A_\theta(u,\lambda)A_\theta(u,-\lambda)'$ are bounded from below by some constant $C > 0$ uniformly in θ, u and λ.

(v) The components of $A(u,\lambda)$ are differentiable in u and λ with uniformly bounded derivatives $(\partial/\partial u)(\partial/\partial\lambda)A(u,\lambda)_{ab}$.

(vi) The components of $\mu_\theta(u)$ and $\mu(u)$ are differentiable in θ and u with uniformly continuous derivatives $\nabla_{ij}^2(\partial/\partial u)\mu(\theta)_a$ and $(\partial/\partial u)\mu(u)_a$.

First we need equicontinuity of the likelihoods and their derivatives. We call a sequence of random variables $Z_T(\theta)$, $\theta \in \Theta$ equicontinuous in probability, if for each $\eta > 0$ and $\varepsilon > 0$ there exists a $\delta > 0$ with

$$\limsup_{T\to\infty} P\Big(\sup_{|\theta_1-\theta_2|\le\delta}|Z_T(\theta_1) - Z_T(\theta_2)| > \eta\Big) < \varepsilon.$$

LEMMA 2.7. *Suppose Assumption* 2.6 *holds. Then the components of* $\mathscr{L}_T(\theta)$, $\nabla\mathscr{L}_T(\theta)$, $\nabla^2\mathscr{L}_T(\theta)$, $\widetilde{\mathscr{L}_T}(\theta)$, $\nabla\widetilde{\mathscr{L}_T}(\theta)$, $\nabla^2\mathscr{L}_T(\theta)$ *are equicontinuous in probability.*

PROOF. We prove equicontinuity of $\mathscr{L}_T(\theta)$ and $\nabla^2_{ij}\mathscr{L}_T(\theta)$ (which is needed for the asymptotic properties of $\hat{\theta}_T$ in the proof of Theorem 3.1). Equicontinuity of the other quantities follows in the same way. We have with a mean value $\bar{\theta}$,

$$\mathscr{L}_T(\theta_2) - \mathscr{L}_T(\theta_1) = (\theta_2 - \theta_1)'\nabla\mathscr{L}_T(\bar{\theta}),$$

where

$$\begin{aligned}\nabla_i\mathscr{L}_T(\theta) &= \frac{1}{4\pi}\frac{1}{T}\sum_{t=1}^{T}\int_{-\pi}^{\pi}\mathrm{tr}\Big\{f_\theta\Big(\frac{t}{T},\lambda\Big)\nabla_i f_\theta\Big(\frac{t}{T},\lambda\Big)^{-1}\Big\}\,d\lambda\\ &\quad+\frac{1}{8\pi^2 T}(\underline{X}-\underline{\mu}_\theta)'U_T(\nabla_i f_\theta^{-1})(\underline{X}-\underline{\mu}_\theta) \qquad (2.16)\\ &\quad-\frac{1}{4\pi^2 T}(\nabla_i\underline{\mu}_\theta)'U_T(f_\theta^{-1})(\underline{X}-\underline{\mu}_\theta)\\ &=\frac{1}{8\pi^2 T}(\underline{X}-\underline{\mu})'U_T(\nabla_i f_\theta^{-1})(\underline{X}-\underline{\mu})\\ &\quad+\frac{1}{4\pi^2 T}\nabla_i\Big\{(\underline{\mu}-\underline{\mu}_\theta)'U_T(f_\theta^{-1})\Big\}(\underline{X}-\underline{\mu})+\text{const.} \qquad (2.17)\end{aligned}$$

with a constant independent of $\underline{X}$ (but dependent on θ and T). With the Cauchy–Schwarz inequality and Lemma A.1(b) we get

$$\begin{aligned}&\frac{1}{T}(\nabla_{i\underline{\mu}_\theta})'U_T(f_\theta^{-1})(\underline{X}-\underline{\mu}_\theta)\\ &\quad\le\frac{1}{T}\Big\{(\nabla_{i\underline{\mu}_\theta})'U_T(f_\theta^{-1})(\nabla_{i\underline{\mu}_\theta})\cdot(\underline{X}-\underline{\mu}_\theta)'U_T(f_\theta^{-1})(\underline{X}-\underline{\mu}_\theta)\Big\}^{1/2}\\ &\quad\le\Big\{\frac{1}{T}\big|\nabla_{i\mu_\theta}\big|^2\Big\}^{1/2}\Big\{\frac{2}{T}\big|\underline{X}\big|^2+\frac{2}{T}\big|\underline{\mu}_\theta\big|^2\Big\}^{1/2}\|U_T(f_\theta^{-1})\|,\end{aligned}$$

which by Assumption 2.6 and Lemma A.5 is uniformly bounded by $K+K(1/T)|\underline{X}|^2$. Similarly, we can estimate the other terms in (2.16) leading to

$$\sup_{|\theta_1-\theta_2|\le\delta}|\mathscr{L}_T(\theta_2)-\mathscr{L}_T(\theta_1)|\le K\delta\Big(1+\frac{1}{T}\underline{X}'\underline{X}\Big)$$

with some constant K. Since $E(1/T)\underline{X}'\underline{X}=(1/T)\mathrm{tr}\{\Sigma\}+(1/T)|\underline{\mu}|^2$ converges to $\sum_{a=1}^{d}\int_0^1\{\int_{-\pi}^{\pi}f_{aa}(u,\lambda)\,d\lambda+\mu_a(u)^2\,du\}$ and $\mathrm{Var}(1/T)\underline{X}'\underline{X}=(2/T^2)\mathrm{tr}\{\Sigma^2\}\le(2/T)\|\Sigma\|^2\le K/T$ (Lemmas A.1, A.5 and A.7) $T^{-1}\underline{X}'\underline{X}$ is bounded in probability. This implies equicontinuity. The proof of equicontinuity of $\nabla^2_{ij}\mathscr{L}_T(\theta)$ is a bit more involved since we do not want to assume third-order differentiability of $A_\theta(u,\lambda)$ with respect to θ in Assumption 2.6(iii).

We obtain from (2.16),

$$
\begin{aligned}
\nabla^2_{ij}\mathscr{L}_T(\theta) = &-\frac{1}{4\pi}\frac{1}{T}\sum_{t=1}^{T}\int_{-\pi}^{\pi}\mathrm{tr}\left\{f_\theta\left(\frac{t}{T},\lambda\right)\nabla^2_{ij}f_\theta\left(\frac{t}{T},\lambda\right)^{-1}\right\}d\lambda\\
&-\frac{1}{4\pi}\frac{1}{T}\sum_{t=1}^{T}\int_{-\pi}^{\pi}\mathrm{tr}\left\{\nabla_i f_\theta\left(\frac{t}{T},\lambda\right)\nabla_j f_\theta\left(\frac{t}{T},\lambda\right)^{-1}\right\}d\lambda\\
&+\frac{1}{8\pi^2T}(\underline{X}-\underline{\mu}_\theta)'U_T(\nabla^2_{ij}f_\theta^{-1})(\underline{X}-\underline{\mu}_\theta)\\
&+\frac{1}{4\pi^2T}(\nabla_i\underline{\mu}_\theta)'U_T(f_\theta^{-1})(\nabla_j\underline{\mu}_\theta)\\
&-\frac{1}{4\pi^2T}(\nabla_i\underline{\mu}_\theta)'U_T(\nabla_j f_\theta^{-1})(\underline{X}-\underline{\mu}_\theta)\\
&-\frac{1}{4\pi^2T}(\nabla_j\underline{\mu}_\theta)'U_T(\nabla_i f_\theta^{-1})(\underline{X}-\underline{\mu}_\theta)\\
&-\frac{1}{4\pi^2T}(\nabla^2_{ij}\underline{\mu}_\theta)'U_T(f_\theta^{-1})(\underline{X}-\underline{\mu}_\theta).
\end{aligned}
\tag{2.18}
$$

Equicontinuity of $\nabla^2_{ij}\mathscr{L}_{\mathscr{T}}(\theta)$ follows if we prove equicontinuity for all terms separately. The first and second term are deterministic and uniformly continuous in θ and therefore also equicontinuous in probability. The remaining terms of (2.18) can all be written as sums of expressions of the form

$$
\frac{1}{T}\underline{X}'U_\theta\underline{X},\quad \frac{1}{T}\underline{\nu}'_\theta U_\theta\underline{X}\quad\text{or}\quad \frac{1}{T}\underline{\nu}'_{1\theta}U_\theta\underline{\nu}_{2\theta}
\tag{2.19}
$$

with U_θ being equal to $U_T(f_\theta^{-1})$, $U_T(\nabla_i f_\theta^{-1})$ or $U_T(\nabla^2_{ij}f_\theta^{-1})$ and $\underline{\nu}_\theta$ being equal to $\underline{\mu}_\theta$, $\nabla_i\underline{\mu}_\theta$ or $\nabla^2_{ij}\underline{\mu}_\theta$. The last expression is also deterministic and uniformly continuous. For the second term we obtain, with the Cauchy–Schwarz inequality,

$$
\begin{aligned}
&\left|\frac{1}{T}\underline{\nu}'_{\theta_1}U_{\theta_1}\underline{X}-\frac{1}{T}\underline{\nu}'_{\theta_2}U_{\theta_2}\underline{X}\right|\\
&\quad\le\frac{1}{T}\left|(\underline{\nu}_{\theta_1}-\underline{\nu}_{\theta_2})'U_{\theta_1}\underline{X}\right|+\frac{1}{T}\left|\underline{\nu}'_{\theta_2}(U_{\theta_1}-U_{\theta_2})\underline{X}\right|\\
&\quad\le\frac{1}{T}\left\{|\underline{\nu}_{\theta_1}-\underline{\nu}_{\theta_2}|^2|\underline{X}|^2\right\}^{1/2}\|U_{\theta_1}\|+\frac{1}{T}\left\{|\underline{\nu}_{\theta_2}|^2|\underline{X}|^2\right\}^{1/2}\|U_{\theta_1}-U_{\theta_2}\|.
\end{aligned}
$$

Lemma A.5(iii) implies that for all $\varepsilon>0$ there exists a T_0 and a δ such that

$$
\sup_{|\theta_1-\theta_2|\le\delta}\|U_{\theta_1}-U_{\theta_2}\|\le\varepsilon\quad\text{for all } T\ge T_0
$$

for all choices of U_θ. Furthermore, δ can be chosen such that $\sup_{|\theta_1-\theta_2|\le\delta}(1/T)|\underline{\nu}_{\theta_1}-\underline{\nu}_{\theta_2}|^2\le\varepsilon$. Since $\|U_\theta\|$ is uniformly bounded and $(1/T)|\underline{X}|^2$ is bounded in probability this implies equicontinuity of $(1/T)\underline{\nu}'_\theta U_\theta\underline{X}$. Equicontinuity of

$(1/T)\underline{X}'U_\theta\underline{X}$ follows in the same way and we therefore obtain equicontinuity in probability of $\nabla^2_{ij}\mathscr{L}_T(\theta)$. □

THEOREM 2.8. *Suppose Assumption* 2.6 *holds. Then we have for* $k = 0, 1, 2$,

(i)
$$\sup_{\theta\in\Theta} |\nabla^k\{\mathscr{L}_T(\theta) - \widetilde{\mathscr{L}_T}(\theta)\}| \xrightarrow{\mathscr{P}} 0,$$

(ii)
$$\sup_{\theta\in\Theta} |\nabla^k\{\mathscr{L}_T(\theta) - \mathscr{L}(\theta)\}| \xrightarrow{\mathscr{P}} 0,$$

(iii)
$$\sup_{\theta\in\Theta} |\nabla^k\{\widetilde{\mathscr{L}_T}(\theta) - \mathscr{L}(\theta)\}| \xrightarrow{\mathscr{P}} 0.$$

PROOF. Of course (i) and (ii) imply (iii). To prove (i) and (ii) we will show that $\nabla^k\mathscr{L}_T(\theta)$ and $\nabla^k\tilde{\mathscr{L}}_T(\theta)$ consist of sums of quadratic forms whose expectations and variances can be calculated by using Lemma A.7. This leads to pointwise consistency and then with the equicontinuity result of Lemma 2.7 to uniform consistency.

(i) We obtain for $k = 0$ with Proposition 2.5 and $B_T := \Sigma_T(A_\theta, A_\theta)^{-1} - U_T(\{4\pi^2 A_\theta \overline{A}'_\theta\}^{-1})$ for each $\varepsilon > 0$,

$$\mathscr{L}_T(\theta) - \tilde{\mathscr{L}}_T(\theta) = \frac{1}{2T}(\underline{X} - \underline{\mu}_\theta)' B_T(\underline{X} - \underline{\mu}_\theta) + O(T^{-1+\varepsilon}).$$

Since

$$\begin{aligned} &\frac{1}{T}(\underline{X} - \underline{\mu}_\theta)' B_T(\underline{X} - \underline{\mu}_\theta) \\ &\quad = \frac{1}{T}(\underline{X} - \underline{\mu})' B_T(\underline{X} - \underline{\mu}) + \frac{2}{T}(\underline{X} - \underline{\mu})' B_T(\underline{\mu} - \underline{\mu}_\theta) \\ &\qquad + \frac{1}{T}(\underline{\mu} - \underline{\mu}_\theta)' B_T(\underline{\mu} - \underline{\mu}_\theta), \end{aligned} \tag{2.20}$$

we obtain with $\Sigma = \Sigma_T(A, A)$ and Lemma A.7 (note the remark below Lemma A.7),

$$\begin{aligned} E\{\mathscr{L}_T(\theta) - \tilde{\mathscr{L}}_T(\theta)\} &= \frac{1}{2T}\mathrm{tr}\{B_T\Sigma\} + \frac{1}{2T}(\underline{\mu} - \underline{\mu}_\theta)' B_T(\underline{\mu} - \underline{\mu}_\theta) + O(T^{-1+\varepsilon}) \\ &= O(T^{-1+\varepsilon}) \end{aligned}$$

and due to Gaussianity,

$$\begin{aligned} \mathrm{Var}\{\mathscr{L}_T(\theta) - \tilde{\mathscr{L}}_T(\theta)\} &= \frac{1}{2T^2}\mathrm{tr}\{B_T\Sigma B_T\Sigma\} + \frac{1}{T^2}(\underline{\mu} - \underline{\mu}_\theta)' B_T\Sigma B_T(\underline{\mu} - \underline{\mu}_\theta) \\ &= O(T^{-2+\varepsilon}). \end{aligned}$$

This implies for each $\varepsilon > 0$,

$$\mathscr{L}_T(\theta) - \tilde{\mathscr{L}}_T(\theta) = O_p(T^{1-\varepsilon}). \tag{2.21}$$

Together with the equicontinuity of $\mathscr{L}_T(\theta)$ and $\tilde{\mathscr{L}}_T(\theta)$ (Lemma 2.7) this implies uniform convergence.

For $k = 1$ we obtain with Lemma A.8, B_T as above and

$$C_T := -\Sigma_T(A_\theta, A_\theta)^{-1}\{\Sigma_T(\nabla_j A_\theta, A_\theta) + \Sigma_T(A_\theta, \nabla_j A_\theta)\}\Sigma_T(A_\theta, A_\theta)^{-1} - U_T(\nabla_j\{4\pi^2 A_\theta \bar{A}_\theta'\}^{-1}),$$

$$\nabla_j \mathscr{L}_T(\theta) - \nabla_j \tilde{\mathscr{L}}_T(\theta) = \frac{1}{2T}(\underline{X} - \underline{\mu}_\theta)' C_T(\underline{X} - \underline{\mu}_\theta) - \frac{1}{T}(\nabla_j \underline{\mu}_\theta)' B_T(\underline{X} - \underline{\mu}_\theta) + O(T^{-1+\varepsilon}).$$

Analogously to the above we obtain with Lemma A.8,

$$E(\nabla_j \mathscr{L}_T(\theta) - \nabla_j \tilde{\mathscr{L}}_T(\theta)) = O(T^{-1+\varepsilon})$$

and

$$\operatorname{Var}(\nabla_j \mathscr{L}_T(\theta) - \nabla_j \tilde{\mathscr{L}}_T(\theta)) = O(T^{-2+\varepsilon}),$$

which gives

$$\nabla \mathscr{L}_T(\theta) - \nabla \tilde{\mathscr{L}}_T(\theta) = O_p(T^{-1-\varepsilon}) \tag{2.22}$$

and, with Lemma 2.7, uniform convergence. For $k = 2$ the result follows in the same way.

(ii) follows similarily. For $k = 0$ use $B_T = U_T(f_\theta^{-1})$ in the above derivation and apply Lemma A.7 and Proposition 2.5 to get

$$E\mathscr{L}_T(\theta) = \mathscr{L}_T(\theta) + O(T^{-1+\varepsilon})$$

and

$$\operatorname{Var} \mathscr{L}_T(\theta) = O(T^{-1}),$$

which implies with the equicontinuity of $\mathscr{L}_T(\theta)$ and the uniform continuity of $\mathscr{L}(\theta)$ the result. For $k = 1$ we use (2.17) and the Gaussianity of $\underline{X}$ to obtain

$$\begin{aligned}\operatorname{Var}(\nabla_i \mathscr{L}_T(\theta)) &= \frac{1}{32\pi^4 T^2} \operatorname{tr}\{U_T(\nabla_i f_{\theta_0}^{-1})\Sigma U_T(\nabla_i f_{\theta_0}^{-1})\Sigma\} \\ &\quad + \frac{1}{16\pi^4 T^2}[\nabla_i\{(\underline{\mu} - \underline{\mu}_\theta)' U_T(f_\theta^{-1})\}]\Sigma[\nabla_i\{U_T(f_\theta^{-1})(\underline{\mu} - \underline{\mu}_\theta)\}] \\ &= O(T^{-1}) \quad \text{by Lemma A.7.}\end{aligned}$$

Furthermore, we obtain with (2.16), (2.20) and Lemma A.7,

$$
\begin{aligned}
E\nabla_i \mathscr{L}_T(\theta) &= \frac{1}{4\pi}\frac{1}{T}\sum_{t=1}^{T}\int_{-\pi}^{\pi} \operatorname{tr}\left\{ f_\theta\left(\frac{t}{T},\lambda\right)\nabla_i f_\theta\left(\frac{t}{T},\lambda\right)^{-1}\right\} d\lambda \\
&\quad + \frac{1}{8\pi^2 T}\operatorname{tr}\{U_T(\nabla_i f_{\theta_0}^{-1})\Sigma\} \\
&\quad + \frac{1}{8\pi^2 T}(\underline{\mu}-\underline{\mu}_\theta)' U_T(\nabla f_\theta^{-1})(\underline{\mu}-\underline{\mu}_\theta) \qquad (2.23)\\
&\quad - \frac{1}{4\pi^2 T}(\nabla_i \underline{\mu}_\theta)' U_T(f_\theta^{-1})(\underline{\mu}-\underline{\mu}_\theta) \\
&= \nabla_i \mathscr{L}(\theta) + O(T^{-1+\varepsilon}),
\end{aligned}
$$

which implies, with the equicontinuity, the result. For $k = 2$ it follows from (2.18) that

$$
\begin{aligned}
\nabla_{ij}^2 \mathscr{L}_T(\theta) &= \frac{1}{8\pi^2 T}(\underline{X}-\underline{\mu})' U_T(\nabla_{ij}^2 f_\theta^{-1})(\underline{X}-\underline{\mu}) \\
&\quad + \frac{1}{4\pi^2 T}\nabla_{ij}^2\{(\underline{\mu}-\underline{\mu}_\theta)' U_T(f_\theta^{-1})(\underline{X}-\underline{\mu}) + \text{const.}
\end{aligned}
$$

and therefore

$$
\begin{aligned}
\operatorname{Var}(\nabla_{ij}^2 \mathscr{L}_T(\theta)) &= \frac{1}{32\pi^4 T^2}\operatorname{tr}\{U_T(\nabla_{ij}^2 f_\theta^{-1})\Sigma U_T(\nabla_{ij}^2 f_\theta^{-1})\Sigma\} \\
&\quad + \frac{1}{16\pi^4 T^2}[\nabla_{ij}^2\{(\underline{\mu}-\underline{\mu}_\theta)' U_T(f_\theta^{-1})\}]\Sigma[\nabla_{ij}^2\{U_T(f_\theta^{-1})(\mu-\mu_\theta)\}].
\end{aligned}
$$

Lemma A.7 shows that this is of order $O(T^{-1})$. To calculate $E\nabla_{ij}^2 \mathscr{L}_T(\theta)$ we consider all terms separately and prove with Lemma A.7 convergence to the corresponding terms of $\nabla_{ij}^2 \mathscr{L}(\theta)$. As above this implies the result. □

REMARK 2.9. Theorem 2.8(iii) gives the asymptotic Kullback–Leibler information divergence of two Gaussian multivariate locally stationary processes: If $X_{t,T}(\tilde{X}_{t,T})$ are multivariate locally stationary with spectral densities $f = A\overline{A}'(\tilde{f} = \tilde{A}\overline{A})$, mean functions $\mu(\tilde{\mu})$ and Gaussian densities $g(\tilde{g})$, then we obtain for the information divergence

$$
\begin{aligned}
D(\tilde{f},\tilde{\mu},f,\mu) &= \lim_{T\to\infty}\frac{1}{T}E_g \log\frac{g}{\tilde{g}} \\
&= \frac{1}{4\pi}\int_0^1\int_{-\pi}^{\pi}\{\log\det[\tilde{f}(u,\lambda)f(u,\lambda)^{-1}] \\
&\quad + \operatorname{tr}[\tilde{f}(u,\lambda)^{-1}f(u,\lambda) - I]\}\,d\lambda\,du \\
&\quad + \frac{1}{4\pi}\int_0^1(\tilde{\mu}(u)-\mu(u))'\tilde{f}(u,0)^{-1}(\tilde{\mu}(u)-\mu(u))\,du.
\end{aligned}
$$

This is the time average of the Kullback–Leibler divergence in the stationary case [cf. Parzen (1983) for the univariate stationary case with mean zero].

REMARK 2.10. There is another important aspect of the above likelihood approximation: the likelihood is of the form

$$\mathscr{L}_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} \ell_T\left(\theta, \frac{t}{T}\right)$$

with

$$\ell_T\left(\theta, \frac{t}{T}\right) = \frac{1}{4\pi}\int_{-\pi}^{\pi}\left\{\log\left[(2\pi)^{2d} \det f_\theta\left(\frac{t}{T}, \lambda\right)\right] + \operatorname{tr}\left[f_\theta\left(\frac{t}{T}, \lambda\right)^{-1} \tilde{I}_T^{\mu\theta}\left(\frac{t}{T}, \lambda\right)\right]\right\} d\lambda,$$

that is, $\mathscr{L}_T(\theta)$ has a similar form to the negative log-likelihood function of iid observations where $\ell_T(\theta, t/T)$ is the negative log-likelihood at time point t. Heuristically $\ell_T(\theta, t/T)$ may still be regarded as the negative log-likelihood at time point t which now in addition contains the full information on the dependence (correlation) structure of $X_{t,T}$ with all the other variables. This was the reason for choosing the notation "local likelihood." $\ell_T(\theta, t/T)$ may be used for nonparametric estimation of $\theta(u)$ by considering $\ell_T(\theta, t/T)$ in some neighbourhood $|t/T - u| \le \delta$. Suppose we have a locally stationary model which is parametrized by one or several curves $\theta(u)$ in time. By using the local likelihood we may define:

1. A kernel estimate by

$$\hat{\theta}(u) = \operatorname*{argmin}_{\theta} \frac{1}{b_T T}\sum_{t=1}^{T} K\left(\frac{u - t/T}{b_T}\right)\ell_T\left(\theta, \frac{t}{T}\right).$$

2. A local polynomial fit of $\theta(u)$ by

$$\hat{c}(u) = \operatorname*{argmin}_{c} \frac{1}{b_T T}\sum_{t=1}^{T} K\left(\frac{u - t/T}{b_T}\right)\ell_T\left(\sum_{j=0}^{d} c_j\left(\frac{t}{T} - u\right)^j, \frac{t}{T}\right).$$

3. An orthogonal series estimator (e.g., wavelets) by

$$\bar{\alpha} = \operatorname*{argmin}_{\alpha} \frac{1}{T}\sum_{t=1}^{T} \ell_T\left(\sum_{j=i}^{J} \alpha_j \psi_j\left(\frac{t}{T}\right), \frac{t}{T}\right),$$

together with some shrinkage of $\bar{\alpha}$.

In the case of several parameter curves (a vector of curves) θ, the c_j and the α_j are also vectors. It is obvious that the properties of these estimators have to be investigated in detail. In Dahlhaus and Neumann (2000) this has been done for the wavelet estimator from (3). It has been shown that the usual rates of convergence in Besov smoothness classes are attained up to a logarithmic factor by the estimator.

3. Asymptotic properties of local likelihood estimates. In this section we discuss the asymptotic properties of the estimator $\hat{\theta}_T$. As a by-product we also obtain the asymptotic properties of the MLE $\tilde{\theta}_T$. In the univariate case with $\mu = \mu_\theta = 0$, the latter were already proved in Dahlhaus (1996b). We now state our main result.

THEOREM 3.1. *Suppose that Assumption* 2.6 *holds. Then we have*

$$\sqrt{T}(\hat{\theta}_T - \theta_o) \xrightarrow{\mathscr{D}} \mathscr{N}(0, \Gamma^{-1} V \Gamma^{-1}) \text{ and } \sqrt{T}(\tilde{\theta}_T - \theta_0) \xrightarrow{\mathscr{D}} \mathscr{N}(0, \Gamma^{-1} V \Gamma^{-1})$$

with

$$\begin{aligned}\Gamma_{ij} &= \frac{1}{4\pi}\int_0^1 \int_{-\pi}^{\pi} \operatorname{tr}\{(f - f_{\theta_0})\nabla_{ij} f_{\theta_0}^{-1}\}\, d\lambda\, du \\ &\quad - \frac{1}{4\pi}\int_0^1 \int_{-\pi}^{\pi} \operatorname{tr}\{(\nabla_i f_{\theta_0})(\nabla_j f_{\theta_0}^{-1})\}\, d\lambda\, du \\ &\quad + \frac{1}{4\pi}\int_0^1 \nabla_{ij}^2 \big\{(\mu(u) - \mu_{\theta_0}(u))' f_{\theta_0}^{-1}(u,0)(\mu(u) - \mu_{\theta_0}(u))\big\} du\end{aligned}$$

and

$$\begin{aligned}V_{ij} &= \frac{1}{4\pi}\int_0^1 \int_{-\pi}^{\pi} \operatorname{tr}\{f(\nabla_i f_\theta^{-1}) f(\nabla_j f_\theta^{-1})\}\, d\lambda\, du \\ &\quad + \frac{1}{2\pi}\int_0^1 \int_{-\pi}^{\pi} [\nabla_i\{(\mu(u) - \mu_{\theta_0}(u))' f_{\theta_0}^{-1}(u,0)\}] f(u,0) \\ &\qquad \times [\nabla_j\{f_{\theta_0}^{-1}(u,0)(\mu(u) - \mu_{\theta_0}(u))\}]\, du.\end{aligned}$$

PROOF. We start by proving consistency of $\hat{\theta}_T$. We have

$$\mathscr{L}_T(\hat{\theta}_T) \le \mathscr{L}_T(\theta_0) \xrightarrow{\mathscr{P}} \mathscr{L}(\theta_0) \le \mathscr{L}(\hat{\theta}_T).$$

Theorem 2.8 implies $\mathscr{L}_T(\hat{\theta}_T) - \mathscr{L}(\hat{\theta}_T) \to^{\mathscr{P}} 0$ and therefore also $\mathscr{L}(\hat{\theta}_T) - \mathscr{L}(\theta_0) \to^{\mathscr{P}} 0$. Compactness of Θ and the uniqueness of θ_0 then imply $\hat{\theta}_T \to^{\mathscr{P}} \theta_0$. For $\tilde{\theta}_T$ the proof is the same. Furthermore, we obtain with the mean value theorem,

$$\nabla_i \mathscr{L}_T(\hat{\theta}_T) - \nabla_i \mathscr{L}_T(\theta_0) = \{\nabla^2 \mathscr{L}_T(\theta_T^{(i)})(\hat{\theta}_T - \theta_0)\}_i$$

with $|\theta_T^{(i)} - \theta_0| \le |\hat{\theta}_T - \theta_0| (i = 1, \ldots, p)$. If $\hat{\theta}_T$ lies in the interior of Θ we have $\nabla L_T(\hat{\theta}_T) = 0$. If $\hat{\theta}_T$ lies on the boundary of Θ, then the assumption that θ_0 is in the interior implies $|\hat{\theta}_T - \theta_0| \ge \delta$ for some $\delta > 0$; that is, we obtain $P(\sqrt{T}|\nabla \mathscr{L}_T(\hat{\theta}_T)| \ge \varepsilon) \le P(|\hat{\theta}_T - \theta_0| \ge \delta) \to 0$ for all $\varepsilon > 0$. Thus, the result follows if we prove:

(i) $\nabla^2 \mathscr{L}_T(\theta_T^{(i)}) - \nabla^2 \mathscr{L}_T(\theta_0) \xrightarrow{\mathscr{P}} 0$,
(ii) $\nabla^2 \mathscr{L}_T(\theta_0) \xrightarrow{\mathscr{P}} \nabla^2 \mathscr{L}(\theta_0)$ and $\Gamma = \nabla^2 \mathscr{L}_T(\theta_0)$,
(iii) $\sqrt{T} \nabla \mathscr{L}_T(\theta_0) \xrightarrow{\mathscr{D}} \mathscr{N}(0, V)$.

(i) is a simple consequence of $\theta_T^{(i)} \to^{\mathscr{P}} \theta_0$ and the equicontinuity of $\nabla^2 \mathscr{L}_T(\theta)$ proved in Lemma 2.7. The first part of (ii) follows from Theorem 2.8. $\Gamma = \nabla^2 \mathscr{L}(\theta_0)$ follows from elementary calculus. To prove (iii) we use the method of cumulants. We obtain from (2.23) with $\nabla \mathscr{L}(\theta_0) = 0$,

$$\sqrt{T} E \nabla \mathscr{L}_T(\theta_0) = o(1).$$

Furthermore, we get from (2.17),

$$\begin{aligned} &T \operatorname{Cov}(\nabla_i \mathscr{L}(\theta_0), \nabla_j \mathscr{L}_T(\theta_0)) \\ &\quad = \frac{1}{32\pi^4 T} \operatorname{tr}\, \{U_T(\nabla_i f_{\theta_0}^{-1}) \Sigma U_T(\nabla_j f_{\theta_0}^{-1}) \Sigma\} \\ &\qquad + \frac{1}{16\pi^4 T} \Big[\nabla_i \big\{(\mu - \mu_{\theta_0})' U_T(f_{\theta_0}^{-1})\big\}\Big] \Sigma \Big[\nabla_j \big\{U_T(f_{\theta_0}^{-1})(\mu - \mu_{\theta_0})\big\}\Big]. \end{aligned}$$

Lemma A.7 implies that this tends to V_{ij}.

To study the higher-order cumulants we see from (2.17) that $\nabla_i \mathscr{L}_T(\theta_0)$ can be written as

$$\nabla_i \mathscr{L}_T(\theta_0) = \frac{1}{8\pi^2 T} \underline{Y}' A_i \underline{Y} + \frac{1}{4\pi^2 T} \underline{\nu}_i' B \underline{Y} + \text{const.},$$

where $E\underline{Y} = 0$. The cumulants of order more than or equal to 3 of the $\underline{\nu}_i B \underline{Y}$ terms are zero, while the mixed cumulants of the $\underline{Y}' A_i \underline{Y}$ and $\underline{\nu}_i' B \underline{Y}$ terms are nonzero if and only if there are exactly two $\underline{\nu}_i' B \underline{Y}$ terms involved [this follows from the product theorem for cumulants; cf. Brillinger (1981), Theorem 2.3.2, $E\underline{Y} = 0$, and the normality of $\underline{Y}$].

Therefore, we obtain with the product theorem for cumulants,

$$\begin{aligned} &T^{\ell/2} \operatorname{cum}(\nabla_{i_1} \mathscr{L}_T(\theta_0), \dots, \nabla_{i_\ell} \mathscr{L}_T(\theta_0)) \\ &\quad = C_1 T^{-\ell/2} \sum_{\substack{(j_1,\dots,j_\ell) \\ \text{permutation of} \\ (i_1,\dots,i_\ell)}} \operatorname{tr}\Big\{ \prod_{k=1}^{\ell} \Sigma U_T(\nabla_{j_k} f_\theta^{-1}) \Big\} \\ &\qquad + C_2 T^{-\ell/2} \sum_{\substack{(j_1,\dots,j_\ell) \\ \text{permutation of} \\ (i_1,\dots,i_\ell)}} \underline{\nu}_{j_1}' B \Big\{ \prod_{k=2}^{\ell-1} \Sigma U_T(\nabla_{j_k} f_\theta^{-1}) \Big\} \Sigma B \underline{\nu}_{j_\ell}'. \end{aligned}$$

Lemma A.7 implies that all terms are of order $O(T^{-\ell/2+1})$. Therefore, the theorem is proved. Asymptotic normality of $\tilde{\theta}_T$ follows in exactly the same way by using Lemma A.8 instead of Lemma A.7. □

REMARK 3.2 (Special cases). (i) Theorem 3.1 contains the asymptotic distribution of the Whittle estimate and the MLE in the stationary case as a special case (if f, f_{θ_o}, μ and μ_{θ_o} do not depend on u). The result for the classical Whittle estimator is obtained if in addition $\mu = \mu_\theta = 0$ and $f = f_{\theta_o}$.

Theorem 3.1 also gives the asymptotic distribution in the case where a stationary model is used with the classical Whittle likelihood but the process is only locally stationary.

(ii) The matrices Γ and V from Theorem 3.1 simplify in several situations, in particular when the model is correctly specified ($f = f_{\theta_o}, \mu = \mu_{\theta_o}$; cf. Remark 3.3 below), when a stationary model is fitted (f_θ and μ_θ do not depend on u), and when the parameters separate. For univariate processes this has been discussed in Dahlhaus [(1996b), Remark 2.6 and 2.7] in the context of univariate maximum likelihood estimation.

REMARK 3.3 (Correctly specified case/efficiency). In the correctly specified case ($f = f_{\theta_o}, \mu = \mu_{\theta_o}$) it is easy to see that $V = \Gamma$ with

$$\Gamma_{ij} = \frac{1}{4\pi}\int_0^1 \int_{-\pi}^{\pi} \operatorname{tr}\{f_{\theta_o}(\nabla_i f_{\theta_o}^{-1}) f_{\theta_o}(\nabla_i f_{\theta_o}^{-1})\}\, d\lambda\, du + \frac{1}{2\pi}\int_0^1 (\nabla_i \mu_{\theta_o}(u))' f_{\theta_o}^{-1}(u,0)(\nabla_j \mu_{\theta_o}(u))\, du.$$

In that case both estimates are asymptotically efficient. One way to see this is to prove an LAN-expansion and to show that $\sqrt{T}(\hat{\theta}_T - \theta_o)$ and $\sqrt{T}(\tilde{\theta}_T - \theta_o)$ are equivalent to the central sequence. For univariate processes and the MLE $\tilde{\theta}_T$ this has been done in Dahlhaus [(1996b), Theorem 4.1 and 4.2]. By using the technical lemmata of this Appendix the LAN-property and the efficiency of both estimates can be derived in the same way as in that paper. We omit details [for the efficiency concept, LAN-expansions and the importance of the central sequence; cf. Strasser (1985), in particular Remark 83.12].

REMARK 3.4 (Approximation of the MLE). Under the stronger assumption that $A_\theta(u, \lambda)$ is differentiable in θ, u and λ with uniformly continuous derivatives $\nabla^3_{ijk}(\partial^2/\partial u^2)(\partial/\partial\lambda)A_\theta(u,\lambda)_{ab}$ [see Assumption 2.6(iii)] we can prove that

$$\hat{\theta}_T - \tilde{\theta}_T = O_P(T^{-1+\varepsilon}) \tag{3.1}$$

for each $\varepsilon > 0$. We now briefly sketch the proof. Under this stronger assumption we obtain as in Lemma 2.7 the equicontinuity of $\nabla^3 \mathscr{L}_T(\theta)$ and $\nabla^3 \tilde{\mathscr{L}}_T(\theta)$ and as in Theorem 2.8,

$$\sup_{\theta\in\Theta} |\nabla^3\{\mathscr{L}_T(\theta) - \mathscr{L}(\theta)\}| \stackrel{\mathscr{P}}{\to} 0$$

and

$$\sup_{\theta\in\Theta} |\nabla^3\{\tilde{\mathscr{L}}_T(\theta) - \mathscr{L}(\theta)\}| \stackrel{\mathscr{P}}{\to} 0.$$

Taylor expansions of the second order of $\nabla\mathscr{L}_T(\hat{\theta}_T)$ and $\nabla\tilde{\mathscr{L}}_T(\tilde{\theta}_T)$ around θ_0 now yield

$$-\nabla\mathscr{L}_T(\theta_0) = \nabla^2\mathscr{L}_T(\theta_0)(\hat{\theta}_T - \theta_0) + O_p(T^{-1})$$

and

$$-\nabla \tilde{\mathscr{L}}_T(\theta_0) = \nabla^2 \tilde{\mathscr{L}}_T(\theta_0)(\tilde{\theta}_T - \theta_0) + O_p(T^{-1}).$$

From the proof of Theorem 2.8 we have

$$\nabla \mathscr{L}_T(\theta_0) - \nabla \tilde{\mathscr{L}}_T(\theta_0) = O_p(T^{-1+\varepsilon})$$

[see (2.22)] and similarly,

$$\nabla^2 \mathscr{L}_T(\theta_0) - \nabla^2 \tilde{\mathscr{L}}_T(\theta_0) = O_p(T^{-1-\varepsilon}).$$

This implies (3.1)

REMARK 3.5 (Related estimates). There are a number of estimates with similar properties which are based on a simplified or modified form of the likelihood:

(i) The first simplification results from the observation that the first summand of the likelihood is often free from the parameters describing autocorrelations. Suppose, for example, that the process has a one-sided MA(∞)-representation

$$X_{t,T} = \mu\left(\frac{t}{T}\right) + \frac{1}{2\pi} \sum_{k=0}^{\infty} a_{t,T,k} \varepsilon_{t-k}$$

with $a_{t,T,k}$ as in Remark 2.2 and $E\varepsilon_t = 0, E\varepsilon_s\varepsilon_t' = 2\pi\delta_{st}I_d$. Let $\Sigma^{\varepsilon}(u) = (1/2\pi)a_0(u)a_0(u)'$ where $a_0(u) = \int_{-\pi}^{\pi} A(u, \lambda)\, d\lambda$. $\Sigma^{\varepsilon}(t/T)$ is up to an $O(T^{-1})$ error [due to the approximation (2.2) or (2.4)] the covariance matrix of the one-step prediction error at time t and the covariance matrix of the innovations in a standardized MA(∞)-representation. If the time varying spectral density $f(u, \lambda)$ is for all u and λ of full rank [cf. Assumption 2.6(iv)] then it can be shown that

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \log[(2\pi)^d \det f(u, \lambda)]\, d\lambda = \log \det \Sigma^{\varepsilon}(u),$$

leading to a simplified version of $\mathscr{L}_T(\theta)$.

This follows, for example, from Theorem 1.3.2 of Hannan and Deistler (1988) since the *stationary* MA(∞)-process

$$X_t^{(u_0)} = \mu(u_0) + \frac{1}{2\pi} \sum_{k=0}^{\infty} a_k(u_0)\varepsilon_{t-k}$$

with $u_0 \in [0, 1]$ fixed and $a_k(u)$ as in Remark 2.2 fulfills the assumptions of that theorem (this is the stationary approximation of the process $X_{t,T}$ at time $u_0 = t_0/T$).

For time varying ARMA-models as in Example 2.3 the above conditions are fulfilled if $\det \Phi(u, z) \neq 0$ for all $|z| \leq 1 + c$ with $c > 0$ and $\det \Psi(u, z) \neq 0$ for all $|z| < 1$. In this case $\Sigma^{\varepsilon}(t/T)$ is the covariance matrix of the innovations ε_t. The likelihood then becomes

$$\mathscr{L}_T(\theta) = \frac{1}{2T}\sum_{t=1}^{T}\left\{\log\det\Sigma_\theta^{\varepsilon}\left(\frac{t}{T}\right) + \frac{1}{2\pi}\int_{-\pi}^{\pi}\operatorname{tr}\left[f_\theta\left(\frac{t}{T},\lambda\right)^{-1}\tilde{I}_T^{\mu_\theta}\left(\frac{t}{T},\lambda\right)\right]d\lambda\right\}$$

with the spectral density $f_\theta(u,\lambda)$ as in Example 2.3.

(ii) Consider the discrete frequency form of the quasi likelihood

$$\mathscr{L}_T^{(1)}(\theta) = \frac{1}{2T^2}\sum_{t=1}^{T}\sum_{j=0}^{T-1}\left\{\log\left[(2\pi)^{2d}\det f_\theta\left(\frac{t}{T},\lambda_j\right)\right]\right.$$
$$\left.+\operatorname{tr}\left[f_\theta\left(\frac{t}{T},\lambda_j\right)^{-1}\tilde{I}_T^{\mu_\theta}\left(\frac{t}{T},\lambda_j\right)\right]\right\}$$

with the minimizer $\hat{\theta}_T^{(1)}$ [or as in (i) with the simpler form of the first term]. It follows easily that the second term is obtained by using the matrix $U_T^{(1)}((1/4\pi^2)f_\theta^{-1})$ with

$$U_T^{(1)}(\phi)_{r,s} = \frac{2\pi}{T}\sum_{j=0}^{T-1}\exp(i\lambda_j(r-s))\phi\left(\frac{1}{T}\left[\frac{r+s}{2}\right]^*,\lambda_j\right)$$

instead of $U_T((1/4\pi^2)f_\theta^{-1})$ as an approximation of Σ_θ^{-1}. It can be shown that all results of this section continue to hold for $\mathscr{L}_T^{(1)}(\theta)$ and $\hat{\theta}_T^{(1)}$. However, the proofs are *not* a straightforward generalization of the proofs for $\mathscr{L}_T(\theta)$ and $\hat{\theta}_T$. They are sketched at the end of the Appendix. If the true mean function μ is constant over time we can mean correct the likelihood by dropping zero frequency and using the preperiodogram without mean correction (as in the stationary case). Unfortunately, this is no longer possible in the case of a time-varying mean.

(iii) Due to the continuous definition of $\tilde{I}_T^{\mu}(u,\lambda)$ in u it is also possible to define a continuous time version of the likelihood. The likelihood

$$\mathscr{L}_T^{(2)}(\theta) = \frac{1}{4\pi}\int_0^1\int_{-\pi}^{\pi}\left\{\log\left[(2\pi)^{2d}\det f_\theta(u,\lambda)\right]\right.$$
$$\left.+\operatorname{tr}\left[f_\theta(u,\lambda)^{-1}\tilde{I}_T^{\mu_\theta}(u,\lambda)\right]\right\}d\lambda\,du$$

is obtained by using the approximation

$$U_T^{(2)}(\phi)_{r,s} = \int_{-\pi}^{\pi}\exp(i\lambda(r-s))T\int_{(r+s-1)/2T}^{(r+s+1)/2T}\phi(u,\lambda)\,du\,d\lambda$$

[with $\phi(u,\lambda) = 0$ for $u > 1$; note that $\tilde{I}(u,\lambda) = 0$ for $u < 1/(2T)$]. Again all results of this section continue to hold for $\mathscr{L}_T^{(2)}(\theta)$ and its minimizer $\hat{\theta}_T^{(2)}$. In this case the proof is a straightforward generalization of the proofs for $\mathscr{L}_T(\theta)$ and $\hat{\theta}_T$ (it can be shown that all results of the Appendix remain valid with $U_T^{(2)}$ instead of U_T).

REMARK 3.6 (Non-Gaussian case). It is possible to extend Theorem 3.1 to non-Gaussian processes. For two reasons we have not included the non-Gaussian case:

(i) In the present situation where μ_θ may depend on θ non-Gaussianity would not only introduce a fourth-order cumulant term into the asymptotic variance but also a complicated third-order term [as can easily be seen from (2.17)].
(ii) The calculation of the extra terms cannot be handled with the present results from the Appendix but requires a different technical approach.

APPENDIX

Norms and matrix products of generalized Toeplitz matrices. In this section we study the behavior of the matrix $U_T(\phi)$ in some detail. In particular, we prove that $U_T(\{4\pi^2 f\}^{-1})$ with $f(u, \lambda) = A(u, \lambda)A(u, -\lambda)'$ is a reasonable approximation of the inverse of $\Sigma_T(A, A)$. The results of this section are frequently used in Section 3. There are a few similarities to Section 4 of Dahlhaus (1996a) where we have constructed a different (less precise) approximation of the inverse of $\Sigma_T(A, A)$.

For an $n \times n$ matrix A, we denote the spectral norm by

$$\|A\| = \sup_{x\in\mathbb{C}^n} \frac{|Ax|}{|x|} \sup_{x\in\mathbb{C}^n} \left(\frac{x^*A^*Ax}{x^*x}\right)^{1/2} \tag{A.1}$$

$$= [\text{maximum characteristic root of } A^*A]^{1/2},$$

where A^* denotes the conjugate transpose of A, and the Euclidean norm of A by

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

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

LEMMA A.1. *Let A, B be $n \times n$ matrices. Then*:

(a) $|\mathrm{tr}(AB)| \le |A||B|$,
(b) $\|AB\| \le \|A\||B|$, $|AB| \le |A|\|B\|$,
(c) $\|AB\| \le \|A\|\|B\|$,
(d) $\|A\|^2 \le (\sup_i \sum_{j=1}^n |a_{ij}|)(\sup_j \sum_{i=1}^n |a_{ij}|)$,
(e) $\|A\| = \sup_{x\in\mathbb{C}^n} |x^*Ax/x^*x|$ *for A hermite*,
(f) $|x^*Ax| \le x^*x\|A\|$, $x \in \mathbb{C}^n$,
(g) $\log\det A \le \mathrm{tr}\{A - I\}$ *for A positive definite*.

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

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

Properties of $L_T(\alpha)$ are listed in Dahlhaus [(1997). Lemma A.4]. We remark that we have with a generic constant K,

$$\int_{-\pi}^{\pi} L_T(\alpha)d\alpha \text{ is monotone increasing in } T, \tag{A.3}$$

$$L_T(\alpha) \leq 2L_T(2\alpha), \tag{A.4}$$

$$\int_{-\pi}^{\pi} L_T(\beta-\alpha)L_T(\alpha+\gamma)\,d\alpha \leq KL_T(\beta+\gamma)\log T, \tag{A.5}$$

$$\int_{-\pi}^{\pi} L_T(\alpha)d\alpha \leq K\log T, \qquad \int_{-\pi}^{\pi} L_T(\alpha)^k\,d\alpha \leq KT^{k-1} \quad \text{for } k>1. \tag{A.6}$$

Let

$$\Delta_T(\lambda) := \sum_{r=1}^{T} \exp(-i\lambda r).$$

Direct verification shows

$$|\Delta_T(\lambda)| \leq \pi L_T(\lambda). \tag{A.7}$$

LEMMA A.2. (i) *Let $\psi : [0,1] \to \mathbb{C}$ be differentiable with bounded derivative. Then*

$$\sum_{r=1}^{T} \psi\left(\frac{r}{T}\right)\exp(-i\lambda r) = \psi(1)\Delta_T(\lambda) + O\Big(\sup_u |\psi'(u)|L_T(\lambda)\Big)$$
$$= O(L_T(\lambda)).$$

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

(ii) *Suppose ψ: $[0,1]^k \to \mathbb{C}$ has bounded derivative $\partial^k\psi/\partial u_1 \cdots \partial u_k$. Then*

$$\left|\sum_{r_1,\ldots,r_k=1}^{T} \psi\left(\frac{r_1}{T},\ldots,\frac{r_k}{T}\right)\exp\left(-i\sum_{j=1}^{k}\lambda_j r_j\right)\right|$$
$$\leq K \sup_{\ell\leq k}\ \sup_{\{i_1,\ldots,i_\ell\}\subset\{1,\ldots,k\}}\ \sup_u \left|\frac{\partial^\ell}{\partial u_{i_\ell}\cdots\partial u_{i_\ell}}\psi(u)\right| \prod_{j=1}^{k} L_T(\lambda_j) = O\left(\prod_{j=1}^{k} L_T(\lambda_j)\right).$$

PROOF. (i) Summation by parts gives

$$\sum_{r=1}^{T} \psi\left(\frac{r}{T}\right)\exp(-i\lambda r) = -\sum_{r=1}^{T-1}\left\{\psi\left(\frac{r+1}{T}\right) - \psi\left(\frac{r}{T}\right)\right\}\Delta_r(\lambda) + \psi(1)\Delta_T(\lambda),$$

which implies with (A.7) the result. (ii) Let D_j be the difference operator with respect to the jth component, that is,

$$D_j\psi(r_1/T,\ldots,r_k/T) := \psi(r_1/T,\ldots,r_{j-1}/T,(r_j+1)/T,r_{j+1}/T,\ldots,r_k/T)$$
$$-\ \psi(r_1/T,\ldots,r_k/T).$$

Then we obtain with repeated partial summation and the convention $\psi(u) = 0$ for $u \notin [0,1]^k$,

$$\sum_{r_1,\dots,r_k=1}^{T} \psi\left(\frac{r_1}{T},\dots,\frac{r_k}{T}\right) \exp\left(-i \sum_{j=1}^{k} \lambda_j r_j\right)$$
$$= (-1)^k \sum_{r_1,\dots,r_k=1}^{T} \left(D_1 \cdots D_k \psi\left(\frac{r_1}{T},\dots,\frac{r_k}{T}\right)\right) \prod_{j=1}^{k} \Delta_{r_j}(\lambda_j).$$

We have

$$\left| D_1 \cdots D_k \psi\left(\frac{r_1}{T},\dots,\frac{r_k}{T}\right)\right| \le 2^{k-\ell} T^{-\ell} \sup_u \left| \frac{\partial^\ell}{\partial u_{i_1} \cdots \partial u_{i_\ell}} \psi(u)\right|,$$

where $\{i_1,\dots,i_\ell\} = \{i | r_i \neq T\}$, leading to the result. □

ASSUMPTION A.3. (i) Suppose A: $[0,1] \times \mathbb{R} \to \mathbb{C}^{d\times d}$ is a 2π-periodic matrix function with $A(u,\lambda) = \overline{A(u,-\lambda)}$ whose components are differentiable in u and λ with uniformly bounded derivatives $(\partial/\partial u)(\partial/\partial\lambda)A_{ab}$. $A^o_{t,T}$: $\mathbb{R} \to \mathbb{C}^{d\times d}$ are 2π-periodic matrix functions with

$$\sup_{t,\lambda} \left| A^o_{t,T}(\lambda)_{ab} - A\left(\frac{t}{T},\lambda\right)_{ab}\right| \le KT^{-1} \quad \text{for all } a,b \in \{1,\dots,d\}.$$

(ii) Suppose in addition to (i) that all eigenvalues of $A(u,\lambda)\overline{A(u,-\lambda)}'$ are bounded from below by some $C > 0$ uniformly in u and λ.

(iii) Suppose ϕ: $[0,1] \times \mathbb{R} \to \mathbb{C}^{d\times d}$ is a 2π-periodic matrix function whose components are twice differentiable in u and differentiable in λ with uniformly bounded derivative $(\partial^2/\partial u^2)(\partial/\partial\lambda)\phi$.

(iv) Suppose the components of μ: $[0,1] \to \mathbb{R}^d$ are differentiable with uniformly bounded derivatives.

REMARK A.4. All results stated in this Appendix are uniform in the sense that the upper bounds depend only on the bounds of the involved functions A, ϕ and μ and their derivatives and not on the particular values.

LEMMA A.5. (i) *Suppose A and B fulfill Assumption* A.3(i) *and the components of ϕ are differentiable with uniformly bounded derivative $(\partial/\partial u)(\partial/\partial\lambda)\phi_{ab}$. Then we have*

$$\|\Sigma_T(A,B)\| \le C_1$$

and

$$\|U_T(\phi)\| \le C_2$$

with some constants C_1, C_2.

(ii) *More precisely, we have under Assumption* A.3(i)

$$\|\Sigma_T(A, A)\| \leq 2\pi \sup_{u,\lambda} \|A(u, \lambda)\overline{A(u, \lambda)}'\| + C_A o(1)$$

and

$$\begin{aligned}&\|\Sigma_T(A, A) - \Sigma_T(B, B)\| \\ &\quad \leq 2\pi \sup_{u,\lambda} \|A(u, \lambda)\overline{A(u, \lambda)}' - B(u, \lambda)\overline{B(u, \lambda)}'\| + (C_A + C_B)o(1),\end{aligned}$$

where C_A *is a constant depending on the upper bounds of* A *and its derivatives. If in addition* A *fulfills Assumption* A.3(ii) *we have*

$$\|\Sigma_T(A, A)^{-1}\| \leq (2\pi \inf_{u,\lambda} \lambda_{\min}^{|A|^2}(u, \lambda) + C_A o(1))^{-1},$$

where $\lambda_{\min}^{|A|^2}(u, \lambda)$ *is the smallest eigenvalue of* $A(u, \lambda)\overline{A(u, \lambda)}'$.

(iii) *If* ϕ *is Hermite and fulfills Assumption* A.3(iii) *we have*

$$\|U_T(\phi)\| \leq 2\pi \sup_{u,\lambda} \|\phi(u, \lambda)\| + C_\phi o(1),$$

where C_ϕ *is a constant depending on the upper bounds of* ϕ *and its derivatives. If in addition the smallest eigenvalue* $\lambda_{\min}^{\phi}(u, \lambda)$ *of* $\phi(u, \lambda)$ *is uniformly bounded from below, then*

$$\|U_T(\phi)^{-1}\| \leq \Big(2\pi \inf_{u,\lambda} \lambda_{\min}^{\phi}(u, \lambda) + C_\phi o(1)\Big)^{-1}.$$

PROOF. (i) Lemma A.1(g) implies

$$\|U_T(\phi)\| \leq d \sum_{r\in Z} \sup_{\substack{u\in[0,1]\\ a,b\in\{1,\ldots,d\}}} \left| \int \phi(u, \lambda)_{ab} \exp(i\lambda r)\, d\lambda \right| + K.$$

The smoothness conditions then imply the result [cf. Dahlhaus (1996a), page 156]. The upper bound for $\|\Sigma_T(A, B)\|$ is obtained in the same way.

(ii) follows for $d = 1$ from Lemma 4.4 of Dahlhaus (1996a). In the multivariate case the proof is completely analogous to that lemma. We omit the details.

(iii) The bounds for $\|U_T(\phi)\|$ and $\|U_T(\phi)^{-1}\|$ can be established in exactly the same way as the bounds for $\|\Sigma_T(A, A)\|$ and $\|\Sigma_T(A, A)^{-1}\|$ by a straightforward generalization of Lemma 4.4 of Dahlhaus (1996a). We omit the details. □

In the proof of Lemma A.7 we frequently make use of the following result.

LEMMA A.6. *Suppose A and B fulfill Assumption* A.3(i) *and ϕ fulfills Assumption* A.3(iii) *with $d = 1$. Then we have*

$$\begin{aligned}
&\sum_{r,s=1}^{T} \phi\left(\frac{1}{T}\left[\frac{r+s}{2}\right]^{*}, \lambda\right) A^{o}_{s,T}(\gamma_1) B^{o}_{r,T}(-\gamma_2) \exp\{-i(\lambda-\gamma_1)s - i(\gamma_2-\lambda)r\} \\
&\quad = \sum_{r,s=1}^{T} \phi\left(\frac{r+s}{2T}, \lambda\right) A\left(\frac{s}{T}, \lambda\right) B\left(\frac{r}{T}, -\lambda\right) \exp\{-i(\lambda-\gamma_1)s - i(\gamma_2-\lambda)r\} \\
&\qquad + O(L_T(2\lambda - 2\gamma_1)) + O(L_T(2\gamma_2 - 2\lambda)) \\
&\quad = O(L_T(2\lambda - 2\gamma_1) L_T(2\gamma_2 - 2\lambda)).
\end{aligned}$$

PROOF. We start by replacing $A^{o}_{s,T}(\gamma_1)$ by $A(s/T, \gamma_1)$. Lemma A.2(i) and (A.4) imply

$$\left| \sum_{r=1}^{T} \phi\left(\frac{1}{T}\left[\frac{r+s}{2}\right]^{*}, \lambda\right) B^{o}_{r,T}(-\gamma_2) \exp\{-i(\gamma_2-\lambda)r\} \right| \le K L_T(2\gamma_2 - 2\lambda),$$

which gives a replacement error of $K L_T(2\gamma_2 - 2\lambda)$. In the same way we replace $B^{o}_{r,T}(-\gamma_2)$ by $B(r/T, -\gamma_2)$. We then replace $\phi([(r+s)/2]^{*}/T, \lambda)$ by $\phi((r+s)/(2T), \lambda)$. For $r+s$ even those two are the same. The replacement error therefore is ($r = 2k$, $s = 2\ell - 1$)

$$\begin{aligned}
&\sum_{k,\ell=1}^{[T/2]} \left[\phi\left(\frac{2(k+\ell)-2}{2T}, \lambda\right) - \phi\left(\frac{2(k+\ell)-1}{2T}, \lambda\right)\right] A\left(\frac{2\ell-1}{T}, \gamma_1\right) B\left(\frac{2k}{T}, -\gamma_2\right) \\
&\qquad \times \exp\{-i(\lambda-\gamma_1)(2\ell-1) - i(\gamma_2-\lambda)2k\} + \text{a similar term.}
\end{aligned}$$

Since

$$\begin{aligned}
&\phi\left(\frac{2(k+\ell)-2}{2T}, \lambda\right) - \phi\left(\frac{2(k+\ell)-1}{2T}, \lambda\right) \\
&\qquad = \frac{1}{T}\left[\frac{\partial}{\partial u}\phi\left(\frac{k+\ell-1}{T}, \lambda\right) + O(T^{-1})\right]
\end{aligned}$$

we get with Lemma A.2(i) that this expression is bounded by $K L_T(2\lambda - 2\gamma_1)$. Finally, we replace $A(s/T, \gamma_1)$ by $A(s/T, \lambda)$ with a replacement error of $K|\lambda - \gamma_1| L_T(\lambda - \gamma_1) L_T(\gamma_2 - \lambda) \le K L_T(2\gamma_2 - 2\lambda)$ [by using Lemma A.2(ii)]. Similarly, we obtain $K L_T(2\lambda - 2\gamma_1)$ as the replacement error for replacing $B(r/T, -\gamma_2)$ by $B(r/T, -\lambda)$ which leads to the first equation. The second equation then follows from Lemma A.2(ii) and (A.4). □

LEMMA A.7. *Let* $k \in \mathbb{N}$, A_ℓ, B_ℓ *fulfill Assumption* A.3(ii), ϕ_ℓ *fulfill Assumption* A.3(iii) *and* μ_1, μ_2 *fulfill Assumption* A.3(iv). *Then we have*:

(i)

$$\frac{1}{T} \operatorname{tr} \left\{ \prod_{\ell=1}^{k} U_T(\phi_\ell) \Sigma_T(A_\ell, B_\ell) \right\}$$
$$= (2\pi)^{2k-1} \int_0^1 \int_{-\pi}^{\pi} \operatorname{tr} \left\{ \prod_{\ell=1}^{k} \phi_\ell(u, \lambda) A_\ell(u, \lambda) B_\ell(u, -\lambda)' \right\} d\lambda\, du$$
$$+ O(T^{-1} \log^{2k-1} T),$$

(ii)

$$\frac{1}{T} \mu_{1T}' \left\{ \prod_{\ell=1}^{k-1} U_T(\phi_\ell) \Sigma_T(A_\ell, B_\ell) \right\} U_T(\phi_k) \mu_{2T}$$
$$= (2\pi)^{2k-1} \int_0^1 \mu_1(u)' \left\{ \prod_{\ell=1}^{k-1} \phi_\ell(u, 0) A_\ell(u, 0) B_\ell(u, 0)' \right\} \phi_k(u, 0) \mu_2(u)\, du$$
$$+ O(T^{-1} \log^{2k-1} T).$$

REMARK. If $\widetilde{I}$ is the $d \times d$ identity matrix then $(1/2\pi)\Sigma_T(\widetilde{I}, \widetilde{I}) = (1/2\pi)U_T(\widetilde{I})$ is the $dT \times dT$ identity matrix. Therefore Lemma A.7 also gives the asymptotic expressions for

$$\frac{1}{T} \operatorname{tr} \left\{ \prod_{\ell=1}^{k} \Sigma_T(A_\ell, B_\ell) \right\} \quad \text{and} \quad \frac{1}{T} \operatorname{tr} \left\{ \prod_{\ell=1}^{k} U_T(\phi_\ell) \right\}$$

and more generally for the trace of an arbitrary product of Σ_T's and U_T's.

PROOF OF LEMMA A.7. (i) We give the proof for $k = 1$ and afterwards for general $k \geq 2$. We have

$$\frac{1}{T} \operatorname{tr}\{U_T(\phi) \Sigma_T(A, B)\}$$
$$= \frac{1}{T} \sum_{a,b,c=1}^{d} \sum_{r,s=1}^{T} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \phi\left(\frac{1}{T} \left[\frac{r+s}{2} \right]^*, \lambda \right)_{ab} A_{s,T}^o(\gamma)_{bc} B_{r,T}^o(-\gamma)_{ac}$$
$$\times \exp\{i(\lambda - \gamma)(r - s)\}\, d\lambda\, d\gamma,$$

which by using Lemma A.6 and (A.6) is equal to

$$\frac{1}{T} \sum_{a,b,c=1}^{d} \sum_{r,s=1}^{T} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \phi\left(\frac{r+s}{2T}, \lambda \right)_{ab} A\left(\frac{s}{T}, \lambda \right)_{bc} B\left(\frac{r}{T}, -\lambda \right)_{ac}$$
$$\times \exp\{i(\lambda - \gamma)(r - s)\}\, d\lambda\, d\gamma + O(T^{-1} \log T).$$

Integration over γ now gives the result.

To simplify notation we use in the rest of the proof the "trace" notation, keeping in mind that in the calculation of remainders usually the individual components have to be considered. For $k \geq 2$ we then have

$$\begin{aligned}
&\frac{1}{T}\,\mathrm{tr}\Big\{\prod_{j=1}^{k} U_T(\phi_j)\Sigma_T(A_j, B_j)\Big\}\\
&\quad= \frac{1}{T}\sum_{\substack{r_1,\dots,r_k,\\ s_1,\dots,s_k=1}}^{T} \int_{-\pi}^{\pi}\cdots\int \mathrm{tr}\Big\{\prod_{j=1}^{k}\phi_j\Big(\frac{1}{T}\Big[\frac{r_j+s_j}{2}\Big]^*, \lambda_j\Big)\\
&\qquad\qquad\qquad\qquad \times A^o_{j,s_j,T}(\gamma_j)B^o_{j,r_{j+1},T}(-\gamma_j)\Big\}\\
&\qquad\quad \times \exp\Big\{-i\sum_{j=1}^{k}\big[(\lambda_j-\gamma_j)s_j + (\gamma_j-\lambda_{j+1})r_{j+1}\big]\Big\}\,d\lambda\,d\gamma,
\end{aligned}$$

where $r_{k+1} = r_1$ and $\lambda_{k+1} = \lambda_1$. Application of Lemma A.6 together with (A.5) and (A.6) shows that this is equal to

$$\begin{aligned}
&\frac{1}{T}\sum_{\substack{r_1,\dots,r_k,\\ s_1,\dots,s_k=1}}^{T}\int_{-\pi}^{\pi}\cdots\int_{-\pi}^{\pi}\mathrm{tr}\Big\{\prod_{j=1}^{k}\phi_j\Big(\frac{r_j+s_j}{2T},\lambda_j\Big)A_j\Big(\frac{s_j}{T},\lambda_j\Big)B_j\Big(\frac{r_{j+1}}{T},-\lambda_{j+1}\Big)\Big\}\\
&\quad\times\exp\Big\{-i\sum_{j=1}^{k}\big[(\lambda_j-\gamma_j)s_j+(\gamma_j-\lambda_{j+1})r_{j+1}\big]\Big\}\,d\lambda\,d\gamma + O(T^{-1}\log^{2k-1}T).
\end{aligned}$$

Integration over all γ_j shows that this is equal to

$$\begin{aligned}
&\frac{(2\pi)^k}{T}\sum_{r_1,\dots,r_k=1}^{T}\int_{-\pi}^{\pi}\cdots\int_{-\pi}^{\pi}\mathrm{tr}\Big\{\prod_{j=1}^{k}\phi_j\Big(\frac{r_j+r_{j+1}}{2T},\lambda_j\Big)A_j\Big(\frac{r_{j+1}}{T},\lambda_j\Big)\\
&\quad\times B_j\Big(\frac{r_{j+1}}{T},-\lambda_{j+1}\Big)\Big\}\exp\Big\{-i\sum_{j=1}^{k}(\lambda_j-\lambda_{j+1})r_{j+1}\Big\}\,d\lambda + O(T^{-1}\log^{2k-1}T).
\end{aligned}$$

We now replace the argument λ_k in ϕ_k, A_k and B_{k-1} by λ_{k-1}. The replacement error is of the form

$$\frac{1}{T}\sum_{r_1,\dots,r_k=1}^{T}\int_{-\pi}^{\pi}\cdots\int_{-\pi}^{\pi}\psi_\lambda\Big(\frac{r_1}{T},\dots,\frac{r_k}{T}\Big)\exp\Big\{-i\sum_{j=1}^{k}(\lambda_j-\lambda_{j+1})r_{j+1}\Big\}\,d\lambda,$$

where $\sup_u |\partial^\ell/(\partial u_{1_1}\dots\partial u_{i_\ell})\psi(u)| \leq K|\lambda_k - \lambda_{k-1}|$ for all $\{i_1,\dots,i_\ell\} \subset \{1,\dots,k\}$; that is, we obtain for the replacement error with Lemma A.2(ii) and (A.6) as an upper bound,

$$K\frac{1}{T}\int_{-\pi}^{\pi}\cdots\int_{-\pi}^{\pi}|\lambda_k-\lambda_{k-1}|\prod_{j=1}^{k}L_T(\lambda_j-\lambda_{j-1})\,d\lambda \leq KT^{-1}\log^{k-1}T.$$

In the same way we successively replace all λ_j by λ_1 and integrate finally over $\lambda_2, \dots, \lambda_k$ which proves the assertion.

(ii) The proof of (ii) is completely analogous to (i). We therefore only give a brief sketch for the case $k \geq 2$. We have

$$\begin{aligned}
&\frac{1}{T}\mu_{1T}'\left\{\prod_{j=1}^{k-1} U_T(\phi_j)\Sigma_T(A_j, B_j)\right\}U_T(\phi_k)\mu_{2T} \\
&\quad = T^{-1}\sum_{\substack{r_1,\dots,r_k,\\ s_1,\dots,s_k=1}}^{T}\int_{-\pi}^{\pi}\cdots\int_{-\pi}^{\pi}\mu_1\left(\frac{r_1}{T}\right)'\left\{\prod_{j=1}^{k-1}\phi_j\left(\frac{1}{T}\left[\frac{r_j+s_j}{2}\right]^*, \lambda_j\right)\right. \\
&\qquad\qquad\qquad\qquad \left.\times A^o_{j,s_j,T}(\gamma_j)B^o_{j,r_{j+1},T}(-\gamma_j)\right\} \\
&\qquad \times \phi_k\left(\frac{1}{T}\left[\frac{r_k+s_k}{2}\right]^*, \lambda_k\right)\mu_2\left(\frac{s_k}{T}\right) \\
&\qquad \times \exp\left\{-i\sum_{j=1}^{k-1}\left[(\lambda_j-\gamma_j)s_j+(\gamma_j-\lambda_{j+1})r_{j+1}\right]+i\lambda_1 r_1 - i\lambda_k s_k\right\}d\lambda\, d\gamma.
\end{aligned}$$

We now use similar replacement steps as in (i) [note that Lemma A.6 also holds if, e.g., $B^o_{r,T}(-\gamma_2) = \mu_1(r/T)$ and γ_2 is set equal to zero] which leads with $s_k = r_{k+1}$ to

$$\begin{aligned}
&\frac{(2\pi)^{k-1}}{T}\sum_{r_1,\dots,r_{k+1}=1}^{T}\int_{-\pi}^{\pi}\cdots\int_{-\pi}^{\pi}\exp\left\{i\lambda_1 r_1 - i\sum_{j=1}^{k-1}(\lambda_j-\lambda_{j+1})r_{j+1} - i\lambda_k r_{k+1}\right\}d\lambda \\
&\quad \times \mu_1\left(\frac{r_1}{T}\right)'\left\{\prod_{j=1}^{k-1}\phi_j\left(\frac{r_j+r_{j+1}}{2T}, \lambda_j\right)A_j\left(\frac{r_{j+1}}{T}, \lambda_j\right)B_j\left(\frac{r_{j+1}}{T}, -\lambda_{j+1}\right)\right\} \\
&\quad \times \phi_k\left(\frac{r_k+r_{k+1}}{T}, \lambda_k\right)\mu_2\left(\frac{r_{k+1}}{T}\right) + O(T^{-1}\log^{2k-1}T).
\end{aligned}$$

As before we now replace all λ_j by λ_1 and finally λ_1 by 0 leading to the result. □

LEMMA A.8. *Let $k \in \mathbb{N}$ and $\{I_1, \dots, I_4\}$ be a partition of $\{1, \dots, k\}$. Let the matrices A_ℓ, B_ℓ (for $\ell \in I_1$) fulfill Assumption* A.3(ii), *C_ℓ (for $\ell \in I_2$) fulfill Assumption* A.3(i), (ii) *with bounded derivatives $(\partial^2/\partial u^2)(\partial/\partial\lambda)C_\ell(u,\lambda)_{ab}$, ϕ_ℓ (for $\ell \in I_3 \cup I_4$) fulfill Assumption* A.3(iii) *with eigenvalues (for $\ell \in I_4$) that are bounded from below uniformly in u and λ, and μ_1, μ_2 fulfill Assumption* A.3(iv).

Let further

$$V_\ell = \Sigma_T(A_\ell, B_\ell), \qquad \psi_\ell(u,\lambda) = 2\pi A_\ell(u,\lambda) B_\ell(u,-\lambda)', \qquad \ell \in I_1,$$
$$V_\ell = \Sigma_T(C_\ell, C_\ell)^{-1}, \qquad \psi_\ell(u,\lambda) = (1/2\pi) C_\ell(u,-\lambda)'^{-1} C_\ell(u,\lambda)^{-1}, \ \ell \in I_2,$$
$$V_\ell = U_T(\phi_\ell), \qquad \psi_\ell(u,\lambda) = 2\pi\phi_\ell(u,\lambda), \qquad \ell \in I_3,$$
$$V_\ell = U_T(\phi_\ell)^{-1} \qquad \psi_\ell(u,\lambda) = (1/2\pi)\phi_\ell(u,\lambda)^{-1}, \qquad \ell \in I_4.$$

Then we have:

(i)

$$\frac{1}{T}\mathrm{tr}\left\{\prod_{\ell=1}^{k} V_\ell\right\} = \frac{1}{2\pi}\int_0^1\int_{-\pi}^{\pi} \mathrm{tr}\left\{\prod_{\ell=1}^{k}\psi_\ell(u,\lambda)\right\} d\lambda du + O\big(T^{-1}\log^{6k-1} T\big),$$

(ii)

$$\frac{1}{T}\mu_{1T}'\left\{\prod_{\ell=1}^{k} V_\ell\right\}\mu_{2T} = \frac{1}{2\pi}\int_0^1 \mu_1(u)'\left\{\prod_{\ell=1}^{k}\psi_\ell(u,0)\right\}\mu_2(u)du + O\big(T^{-1}\log^{6k-1} T\big),$$

PROOF. (i) Let $j = |I_2| + |I_4|$. More precisely, we prove the result with the rate $O(T^{-1}\log^{2k+4j-1} T)$. For $j = 0$ the assertion follows for all k from Lemma A.7. Suppose now the assertion holds for all k and some fixed j. Consider the case $j+1$. By renumbering the V_ℓ we can assume that $k \in I_2 \cup I_4$. Suppose $k \in I_2$. We approximate $V_k = \Sigma^{-1} := \Sigma_T(C_k, C_k)^{-1}$ by $\widetilde{U} := U_T(\{4\pi^2 C_k \overline{C}_k'\}^{-1})$. We have with Lemma A.1, Lemma A.5 and Proposition 2.4,

$$\left|\frac{1}{T}\mathrm{tr}\left\{\prod_{\ell=1}^{k} V_\ell\right\} - \frac{2}{T}\mathrm{tr}\left\{\Big(\prod_{\ell=1}^{k-1} V_\ell\Big)\widetilde{U}\right\} + \frac{1}{T}\mathrm{tr}\left\{\Big(\prod_{\ell=1}^{k-1} V_\ell\Big)\widetilde{U}\Sigma\widetilde{U}\right\}\right|$$
$$= \left|\frac{1}{T}\mathrm{tr}\left\{\Big(\prod_{\ell=1}^{k-1} V_\ell\Big)(\Sigma^{-1} - \widetilde{U})\Sigma(\Sigma^{-1} - \widetilde{U})\right\}\right|$$
$$\le \Big(\prod_{\ell=1}^{k-1}\|V_\ell\|\Big)\|\Sigma\|\frac{1}{T}|\Sigma^{-1} - \widetilde{U}|^2 = O(T^{-1}\log^3 T).$$

This implies the convergence with rate

$$O(T^{-1}\log^{2(k+2)+4j-1} T) = O(T^{-1}\log^{2k+4(j+1)-1} T)$$

which gives the result. If $k \in I_4$ the result is obtained in the same way by using the second equation of Proposition 2.4. (ii) follows similarly. □

Technically, Lemma A.7 and Lemma A.8 are the key results for proving the asymptotic properties of the local likelihood estimator and of the exact MLE as done in Section 3. For $I_1 = \{\ell | \ell \text{ even}\}$, $I_2 = \{\ell | \ell \text{ odd}\}$, $I_3 = I_4 = \varnothing$. Lemma A.8 is a generalization of a central result for Gaussian stationary processes to the locally stationary case [cf. Taniguchi (1983), Theorem 1].

PROOF OF PROPOSITION 2.5. We replace $\Sigma_T := \Sigma_T(A, A)$ by $U_T := U_T(A\overline{A}')$. We obtain, with Lemma A.1(g),

$$\left|\frac{1}{T}\log\det\Sigma_T - \frac{1}{T}\log\det U_T\right| = \left|\frac{1}{T}\log\det\Sigma_T^{-1/2}U_T\Sigma_T^{-1/2}\right|$$
$$\leq \max\left\{\frac{1}{T}\operatorname{tr}(\Sigma_T^{-1}U_T - I), \frac{1}{T}\operatorname{tr}(U_T^{-1}\Sigma_T - I)\right\}.$$

Lemma A.8 yields that both terms are of $O(T^{-1}\log^{11} T)$. Since $f(u,\lambda) = A(u,\lambda)\overline{A(u,\lambda)}'$ is symmetric and positive definite there exist an orthonormal matrix $B(u,\lambda)$ and a diagonal matrix $D(u,\lambda) = \operatorname{diag}\{d_1(u,\lambda),\dots,d_d(u,\lambda)\}$ with positive $d_j(u,\lambda)$ such that

$$f(u,\lambda) = B(u,\lambda)D(u,\lambda)B(u,\lambda)'.$$

Now let $x \in [0, 1]$ and

$$f^{(x)}(u,\lambda) := B(u,\lambda)D^{(x)}(u,\lambda)B(u,\lambda)'$$

with

$$D^{(x)}(u,\lambda) := \operatorname{diag}\{d_1(u,\lambda)^x,\dots,d_d(u,\lambda)^x\}.$$

We have $U_T(f^{(1)}) = U_T$ and $U_T(f^{(0)}) = 2\pi I$ where I is the $dT \times dT$ identity matrix. We therefore obtain with $U_T^{(x)} := U_T(f^{(x)})$

$$\begin{aligned}\frac{1}{T}\log\det\Sigma_T &= \frac{1}{T}\log\det U_T + O(T^{-1}\log^{11}T)\\ &= \frac{1}{T}\int_0^1 \frac{\partial}{\partial x}\log\det U_T^{(x)}\,dx + \log(2\pi)^d + O(T^{-1}\log^{11}T)\\ &= \frac{1}{T}\int_0^1 \operatorname{tr}\Big[U_T^{(x)-1}\frac{\partial}{\partial x}U_T^{(x)}\Big]\,dx + \log(2\pi)^d + O(T^{-1}\log^{11}T).\end{aligned}$$

Furthermore,

$$\frac{\partial}{\partial x}U_T^{(x)} = \int_{-\pi}^{\pi}\exp(i\lambda(r-s))\frac{\partial}{\partial x}f^{(x)}\left(\frac{1}{T}\Big[\frac{r+s}{2}\Big]^*, \lambda\right)d\lambda$$

with

$$\begin{aligned}\frac{\partial}{\partial x}f^{(x)}(u,\lambda) = B(u,\lambda)\operatorname{diag}\{&d_1(u,\lambda)^x\log d_1(u,\lambda),\dots,d_d(u,\lambda)^x\\ &\times\log d_d(u,\lambda)\}B(u,\lambda)'.\end{aligned}$$

Since $f^{(x)}(u,\lambda)$ and $(\partial/\partial x)f^{(x)}(u,\lambda)$ have the same smoothness properties as $\phi(u,\lambda)$ uniformly in x, we obtain from Lemma A.8, with straightward calculations,

$$\frac{1}{T}\operatorname{tr}\Big[U_T^{(x)-1}\frac{\partial}{\partial x}U_T^{(x)}\Big] = \frac{1}{2\pi}\int_0^1\int_{-\pi}^{\pi}\left\{\sum_{j=1}^{d}\log d_j(u,\lambda)\right\}d\lambda\,du + O(T^{-1}\log^{11}T)$$

$$= \frac{1}{2\pi}\int_0^1\int_{-\pi}^{\pi}\log\det f(u,\lambda)d\lambda\,du + O(T^{-1}\log^{11}T)$$

uniformly in x which implies the result. □

The discrete frequency form of the likelihood. We now briefly indicate how the results are proved for the discrete frequency form of the quasi-likelihood $\mathscr{L}_T^{(1)}(\theta)$ and its minimizer $\hat{\theta}_T^{(1)}$ as defined in Remark 3.5. The main problem is to show that the spectral norm of $U_T^{(1)}(\phi)$ is bounded [Lemma A.5 with $U_T^{(1)}(\phi)$]. Heuristically the reason for the problems is that the elements of $U_T^{(1)}(\phi)$ cannot be approximated uniformly by the elements of $U_T(\phi)$ since the elements in the upper right and the lower left of $U_T^{(1)}(\phi)$ are large while they are small for $U_T(\phi)$. Note that for T even, $r \le T/2$, $s > T/2$ we have

$$U_T^{(1)}(\phi)_{r+T/2,\,s-T/2} = U_T^{(1)}(\phi)_{r,s}, \tag{A.8}$$

that is, $U_T^{(1)}(\phi)_{1+T/2,T/2} = U_T^{(1)}(\phi)_{1,T}$. For this reason we define

$$V_T(\phi)_{r,s} = \begin{cases} U_T^{(1)}(\phi)_{r,s}, & \text{if } |r-s| < T/2,\\ 0, & \text{if } |r-s| \ge T/2.\end{cases}$$

Heuristically, the corner elements i.e., the elements of $U_T^{(1)}(\phi)_{r,s} - V_T(\phi)_{r,s}$ are then "shifted by (A.8) to the center." To be precise let $x = (x_1,\dots,x_T)' \in \mathbb{C}^T$ and $y = (y_1,\dots,y_T)'$ be defined by

$$y_T = \begin{cases} x_{t+T/2}, & \text{for } t = 1,\dots,T/2,\\ x_{t-T/2} & \text{for } t = T/2+1,\dots,T.\end{cases}$$

Since the matrix

$$W_T(\phi)_{r,s} = \begin{cases} U_T^{(1)}(\phi)_{r,s}, & \text{if } r \le T/2 \text{ and } s \le T/2,\\ U_T^{(1)}(\phi)_{r,s}, & \text{if } r > T/2 \text{ and } s > T/2,\\ 0, & \text{elsewhere,}\end{cases}$$

is Hermite and nonnegative definite we obtain

$$\begin{aligned} x^*U_T^{(1)}(\phi)x &\le x^*U_T^{(1)}(\phi)x + y^*W_T(\phi)y\\ &= x^*V_T(\phi)x + x^*[U_T^{(1)}(\phi) - V_T(\phi)]x + y^*W_T(\phi)y\\ &= x^*V_T(\phi)x + y^*V_T(\phi)y\end{aligned}$$

and therefore $||U_T^{(1)}|| \leq 2||V_T(\phi)||$. We can now show analogously to the proof of Lemma 4.4 in Dahlhaus (1996a) that

$$||U_T^{(1)}|| \leq 2||V_T(\phi)|| \leq 4\pi \sup_{u,\lambda} ||\phi(u,\lambda)|| + C_\phi o(1), \tag{A.9}$$

where C_ϕ is a constant dependent on the upper bounds of ϕ and its derivatives.

By using (A.9) we now obtain the result of Lemma A.7 with $U_T^{(1)}(\phi)$ instead of $U_T(\phi)$ [for the proof of this statement we need (A.5) and (A.6) with the integral replaced by the sum over the Fourier frequencies. However, this can easily be derived].

This implies the assertions of (2.15), Lemma 2.7, Theorem 2.8 and Theorem 3.1 for $\mathscr{L}_T^{(1)}(\theta)$ and $\hat{\theta}_T^{(1)}$ with exactly the same proofs as in Sections 2 and 3. □

Acknowledgments. The author is grateful to M. Sahm for correcting a mistake in the proof of Proposition 2.5 and to an Associate Editor and two referees for their helpful remarks.

REFERENCES

BRILLINGER, D. R. (1981). *Time Series: Data Analysis and Theory.* Holden-Day, San Francisco.

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

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

DAHLHAUS, R. (1996c). Asymptotic statistical inference for nonstationary processes with evolutionary spectra. *Athens Conference on Applied Probability and Time Series 2. Lecture Notes in Statist.* **115** 145–159. Springer, New York.

DAHLHAUS, R. (1997). Fitting time series models to nonstationary processes. *Ann. Statist.* **25** 1–37.

DAHLHAUS, R. and NEUMANN, M. H. (2000). Locally adaptive fitting of semiparametric models to nonstationary time series. *Stochastic Process. Appl.* To appear.

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

DUNSMUIR, W. (1979). A central limit theorem for parameter estimation in stationary vector time series and its application to models for a signal observed with noise. *Ann. Statist.* **7** 490–506.

DZHAPARIDZE, K. (1971). On methods for obtaining asymptotically efficient spectral parameter estimates for a stationary Gaussian process with rational spectral density. *Theory Probab. Appl.* **16** 550–554.

DZHAPARIDZE, K. (1986). *Parameter Estimation and Hypothesis Testing in Spectral Analysis of Stationary Time Series.* Springer, New York.

FOX, R. and TAQQU, M. S. (1986). Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Ann. Statist.* **14** 517–532.

GRANGER, C. W. J. and HATTANAKA, M. (1964). *Spectral Analysis of Econometric Time Series.* Princeton Univ. Press.

GRAYBILL, F. A. (1983). *Matrices with Applications in Statistics.* Wadsworth, Belmont, CA.

GRENANDER, U. and SZEGÖ, G. (1958). *Toeplitz Forms and Their Applications.* Univ. California Press, Berkeley.

HANNAN, E. J. (1973). The asymptotic theory of linear time series models. *J. Appl. Probab.* **10** 130–145.

HANNAN, E. J. and DEISTLER, M. (1988). *The Statistical Theory of Linear Systems.* Wiley, New York.

HOSOYA, Y. and TANIGUCHI, M. (1982). A central limit theorem for stationary processes and the parameter estimation of linear processes. *Ann. Statist.* **10** 132–153.

MALLAT, S., PAPANICOLAOU, G. and ZHANG, Z. (1998). Adaptive covariance estimation of locally stationary processes. *Ann. Statist.* **26** 1–47.

MIKOSCH, T., GADRICH, T., KLÜPPELBERG, C. and ADLER, R. J. (1995). Parameter estimation for ARMA models with infinite variance innovations. *Ann. Statist.* **23** 305–326.

NEUMANN, M. H. and VON SACHS, R. (1997). Wavelet thresholding in anisotropic function classes and application to adaptive estimation of evolutionary spectra. *Ann. Statist.* **25** 38–76.

PARZEN, E. (1983). Autoregressive spectral estimation. In *Handbook of Statistics* (D. R. Brillinger and P. R. Krishnaiah, eds.) **3** 221–247. North-Holland, Amsterdam.

PRIESTLEY, M. B. (1965). Evolutionary spectra and nonstationary processes. *J. Roy. Statist. Soc. Ser. B* **27** 204–237.

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

ROBINSON, P. M. (1995). Gaussian semiparametric estimation of long range dependence. *Ann. Statist.* **23** 1630–1661.

STRASSER, H. (1985). *Mathematical Theory of Statistics*. de Gruyter, Berlin.

TANIGUCHI, M. (1983). On the second order asymptotic efficiency of estimators of Gaussian ARMA processes. *Ann. Statist.* **11** 157–169.

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

WHITTLE, P. (1954). Some recent contributions to the theory of stationary processes. In *A Study in the Analysis of Stationary Time Series*, 2nd ed. (H. Wold, ed.) 196–228. Almqvist and Wiksell, Uppsala.

INSTITUT FÜR ANGEWANDTE MATHEMATIK
IM NEUENHEIMER FELD 294
D-69120 HEIDELBERG
GERMANY
E-MAIL: dahlhaus@statlab.uni-heidelberg.de