Fitting Time Series Models to Nonstationary Processes

R. Dahlhaus, Universität Heidelberg

Abstract. A general minimum distance estimation procedure is presented for nonstationary time series models that have an evolutionary spectral representation. The asymptotic properties of the estimate is derived under the assumption of possible model misspecification. For autoregressive processes with time varying coefficients the estimate is compared to the least squares estimate. Furthermore, the behaviour of estimates is explained when a stationary model is fitted to a nonstationary process.

1. Introduction.

Stationarity has always played a major role in the theoretical treatment of time series procedures. For example, the spectral density is defined for stationary processes and the important ARMA-model is a stationary time series model. Furthermore, the assumption of stationarity is the basis for a general asymptotic theory: it guarantees that the increase of the sample size leads to more and more information of the same kind which is basic for an asymptotic theory to make sense.

On the other hand many series show a nonstationary behaviour (e.g. in economics or sound analysis). Special techniques (such as taking differences or the consideration of the data on small time intervalls) have been applied to make an analysis with stationary techniques possible.

If one resigns from stationarity the number of possible models for time series data explodes. For example, one may consider ARMA models with time varying coefficients. In that case the time behaviour of the coefficients may again be modeled in different ways. Therefore, we try to consider in this paper a general class of nonstationary processes together with a general estimation method which is a generalisation of Whittle's method for stationary processes (Whittle, 1953).

Whittle's method (cf. Dzhaparidze, 1986; Azencott and Dacunha-Castelle, 1986) is based on minimization of the function

$$L_T(\theta) = \frac{1}{4\pi}\int_{-\pi}^{\pi} \left\{ \log f_\theta(\lambda) + \frac{I_T(\lambda)}{f_\theta(\lambda)} \right\} d\lambda$$

Key words and phrases. Nonstationary processes, time series, evolutionary spectra, minimum distance estimates, model selection.

AMS 1990 subject classifications. Primary 62M15; secondary 62F10.

where $f_\theta(\lambda)$ is the model spectral density and $I_N(\lambda)$ is the periodogram. The Whittle estimate is asymptotically efficient and $L_T(\theta)$ is (up to a constant) an approximation to the Gaussian likelihood function. Since $L_T(\theta)$ may be interpreted as a distance between the parametric spectral density $f_\theta(\lambda)$ and the nonparametric estimate $I_N(\lambda)$, the Whittle-estimate is a minimum distance estimate. In the case where the model is misspecified minimization of $L_T(\theta)$ therefore leads to an estimate of the parameter with the best approximating parametric spectral density. This best approximating paramter also minimizes the asymptotic Kullback-Leibler information divergence. For autoregressive processes the Whittle estimate is identical to the Yule-Walker estimate. If a data taper is applied in the calculation of the periodogram then the estimate also has good small sample properties (cf. Dahlhaus, 1988). Asymptotic normality of the Whittle estimate also holds for non-Gaussian processes. However, this requires identifiability of the model which basically only holds for linear processes.

In this paper we generalise the method of Whittle to processes that only show locally a stationary behaviour (cp. Definition 2.1). We replace the periodogram $I_N(\lambda)$ in $L_N(\theta)$ by a local version and integrate over time (cp. Section 3.1). The resulting estimate again is efficient.

If the model is misspecified the estimate again may be regarded as an estimate for the best approximating model ('best' in the sense of distances between spectral densities or in the sense of the Kullback-Leibler information divergence - cp. Section 3). We prove asymptotic normality also in the misspecified case. In particular we can describe the behaviour of the estimate if a stationary model is fitted and the true process is nonstationary (Section 5).

Although we use a spectral density approach our goal in this paper is not the estimation of the spectral density. We mainly are interested in parametric inference for nonstationary time series models that may be defined purely in the time domain. An example are autoregressive processes with time varying coefficients. Such models are studied in detail in section 4. In particular, we give the estimation equations for such models and study the relation of our estimate to the least squares estimate.

Section 6 contains some practical considerations and a simulation example and Section 7 concluding remarks.

2. Asymptotic theory and locally stationary processes

One of the difficult problems to solve when dealing with nonstationary processes is how to set up an adequate asymptotic theory. Asymptotic considerations are needed in time series analysis to simplify the situation since it is usually hopeless to make calculations for a finite sample size.

However, if $X_1,...,X_T$ are observations from an arbitrary nonstationary process, then letting T tend to infinity, i.e. extending the process into the future will not give any information on the behaviour of the process at the beginning of the time intervall. We therefore need a different asymptotic concept.

Suppose for example that we observe

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

for t = 1, … , T. Inference in this case means inference for the unknown function a(t) on the intervall [1,T]. We have informations on a(t) on the grid {1,2,3, … , T}. Analogously to nonparametric regression it seems natural to set down the asymptotic theory in a way that we "observe" a(t) on a finer grid (but on the same intervall), i.e. that we observe the process

$$X_{t,T} = a\left(\frac{t}{T}\right) X_{t-1,T} + \varepsilon_t \qquad \text{for } t = 1, \dots, T \tag{2.1}$$

(where a is now rescaled to the intervall [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

$$X_{t,T} = \mu\left(\frac{t}{T}\right) + \int_{-\pi}^{\pi} \exp(i\lambda t)\, A\left(\frac{t}{T}, \lambda\right) d\xi(\lambda) \,. \tag{2.2}$$

(similar to the analogous representation for stationary processes). However, it turns out that the equation (2.1) has not exactly but only approximately a solution of the form (2.2). We therefore only require that (2.2) holds approximately which leads to the following definition.

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

$$X_{t,T} = \mu\left(\frac{t}{T}\right) + \int_{-\pi}^{\pi} \exp(i\lambda t)\, A^{\circ}_{t,T}(\lambda)\, d\xi(\lambda) \tag{2.3}$$

where

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

$$\operatorname{cum}\{d\xi(\lambda_1), \dots, d\xi(\lambda_k)\} = \eta\Big(\sum_{j=1}^{k} \lambda_j\Big)\, g_k(\lambda_1, \dots, \lambda_{k-1})\, d\lambda_1 \dots d\lambda_k$$

where cum{ … } denotes the cumulant of k-th order, $g_1 = 0$, $g_2(\lambda) = 1$, $|g_k(\lambda_1, \ldots, \lambda_{k-1})| \leq \text{const}_k$ 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

$$\sup_{t,\lambda} \left| A^{\circ}_{t,T}(\lambda) - A\left(\frac{t}{T}, \lambda\right) \right| \leq K T^{-1} \tag{2.4}$$

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

(2.2) Examples. (i) Suppose Y_t is a stationary process and $\mu, \sigma : [0,1] \to \mathbb{R}$ are continuous. Then

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

is locally stationary with $A^{\circ}_{t,T}(\lambda) = A(\frac{t}{T}, \lambda)$. 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.

(ii) Suppose ε_t is an iid sequence and

$$X_{t,T} = \sum_{j=0}^{\infty} a_j\left(\frac{t}{T}\right) \varepsilon_{t-j}.$$

Then $X_{t,T}$ is locally stationary with $A^{\circ}_{t,T}(\lambda) = A(\frac{t}{T}, \lambda) = \sum_{j=0}^{\infty} a_j\left(\frac{t}{T}\right) \exp(-i\lambda j)$.

(iii) Autoregressive processes with time varying coefficients (cp. Section 4) are locally stationary. This was proved in Dahlhaus (1994, Theorem 2.3). However, in this case we only have (2.4) instead of $A^{\circ}_{t,T}(\lambda) = A(\frac{t}{T}, \lambda)$.

The above definition does not mean that a fixed continuous time process is discretized on a finer grid as T tends to infinity. Instead it means heuristically that with increasing T more and more data of each local structure are observed. If μ and A^o do not depend on t and T then X does not depend on T as well and we obtain the spectral representation of an ordinary stationary process. Thus, the classical theory for stationary processes is a special case of our approach.

Letting T tend to infinity no longer means looking into the future. Nevertheless, a prediction theory within this framework is still possible. One may e.g. assume that $X_{t,T}$ is observed for $t \leq T/2$ (i.e. on the time interval (0,1/2)) and one tries to predict the next observations. A result on the local prediction error similar to Kolmogorov's formula for stationary processes has been proved in Dahlhaus (1994, Theorem 3.2).

By $f(u,\lambda) := |A(u,\lambda)|^2$ we denote the spectral density of our process. In Dahlhaus (1994, Theorem 2.2) we show under smoothness conditions on A that

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

where $X_{s,T}$ is defined by (2.3) (with $A^{\circ}_{t,T}(\lambda) = A(0,\lambda)$ for $t < 1$ and $A^{\circ}_{t,T}(\lambda) = A(1,\lambda)$ for $t > T$ - with respect to λ the above convergence is in quadratic mean). This means that if there exists a spectral representation of the form (2.3) with a <u>smooth</u> $A(u,\lambda)$ then $|A(u,\lambda)|^2$ is uniquely determined (there may exist several other non-smooth representations).

There are similarities of our definition to Priestley's definition of an oscillatory process (cf. Priestley, 1981, chapter 11). However, there is the major difference that we consider double indexed processes and make asymptotic considerations.

3. Fitting parametric models to locally stationary processes.

In this section we discuss the fitting of a locally stationary model with time varying spectral density f_θ, $\theta \in \Theta \subset \mathbb{R}^P$ to observations $X_{1,T}, \ldots, X_{T,T}$. As motivated in the introduction we obtain the parameter estimate by minimization of a generalisation of the Whittle function where the usual periodogram is replaced by local periodograms over (possibly overlapping) data segments.

Let $h: \mathbb{R} \to \mathbb{R}$ be a data taper with $h(x) = 0$ for $x \notin [0,1)$ and (for N even)

$$d_N(u,\lambda) \quad = \quad d_N^X(u,\lambda) = \sum_{s=0}^{N-1} h\left(\frac{s}{N}\right) X_{[uT]-N/2+s+1,T} \exp(-i\lambda s),$$

$$H_{k,N}(\lambda) = \sum_{s=0}^{N-1} h\left(\frac{s}{N}\right)^k \exp(-i\lambda s),$$

$$I_N(u,\lambda) = \frac{1}{2\pi\, H_{2,N}(0)} \; |d_N(u,\lambda)|^2.$$

Thus, $I_N(u,\lambda)$ is the periodogram over a segment of length N with midpoint [uT]. The shift from segment to segment is denoted by S, i.e we calculate I_N over segments with midpoints $t_j := S(j-1) + N/2 \quad (j = 1, \ldots, M)$ where $T = S(M-1) + N$, or, written in rescaled time, at time points $u_j := t_j / T$. We now set

$$\mathcal{L}_T(\theta) = \frac{1}{4\pi}\,\frac{1}{M} \sum_{j=1}^{M} \int_{-\pi}^{\pi} \left\{\log f_\theta(u_j,\lambda) + \frac{I_N(u_j,\lambda)}{f_\theta(u_j,\lambda)}\right\} d\lambda$$

and

$$\hat{\theta}_T = \arg\min_{\theta\in\Theta} \mathcal{L}_T(\theta).$$

The use of a data taper which tends smoothly to zero at the boundaries has two benefits: First it reduces leakage (as in the stationary case). Second it reduces the bias due to nonstationarity by downweighting the observations at the boundaries of the segment. It is interesting to see that the taper does not lead to an increase of the asymptotic variance for overlapping segments (Theorem 3.3). Furthermore, some estimates are even approximately independent of the taper (cp. Theorem 4.2 and the discussion after that theorem).

The above motivation of the function $\mathcal{L}_T(\theta)$ is heuristic. We now give a stronger justification for the particular form of $\mathcal{L}_T(\theta)$. Suppose $\bar{f}$ is the true probability-density of the observations $X_{1,T},\ldots,X_{T,T}$ and f the true spectral-density. Analogously, let $\bar{f}_\theta$ and f_θ be the corresponding densities of our model. If $\bar{f}$ and $\bar{f}_\theta$ are Gaussian distributions with mean zero then we have shown in Dahlhaus (1994, Theorem 3.4) that the asymptotic Kullback-Leibler information divergence is

$$\lim_{T\to\infty} \frac{1}{T} E_{\bar{f}} \log(\bar{f} / \bar{f}_\theta)$$

$$= \frac{1}{4\pi}\int_0^1\int_{-\pi}^{\pi}\left\{\log\frac{f_\theta(u,\lambda)}{f(u,\lambda)} + \frac{f(u,\lambda)}{f_\theta(u,\lambda)} - 1\right\} d\lambda\, du = \frac{1}{4\pi}\int_0^1\int_{-\pi}^{\pi}\left\{\log f_\theta(u,\lambda) + \frac{f(u,\lambda)}{f_\theta(u,\lambda)}\right\} d\lambda\, du + \text{const}$$

where the constant is independent of the model spectral density. Therefore, we may regard

$$\mathcal{L}(\theta) := \frac{1}{4\pi}\int_0^1 \int_{-\pi}^{\pi} \left\{ \log f_\theta(u,\lambda) + \frac{f(u,\lambda)}{f_\theta(u,\lambda)} \right\} d\lambda \, du$$

as a distance between the true process with spectral density $f(u,\lambda)$ and the model with spectral density $f_\theta(u,\lambda)$. The best approximating parameter value from our model class then is

$$\theta_0 := \underset{\theta\in\Theta}{\operatorname{argmin}} \, \mathcal{L}(\theta) .$$

If the model is correct, i.e. $f = f_{\theta^*}$, then it is easy to show that $\theta_o = \theta^*$.

The function $\mathcal{L}_T(\theta)$ is now obtained from $\mathcal{L}(\theta)$ by replacing the unknown true spectral density f by the nonparametric estimate I_N. We conjecture that $\mathcal{L}_T(\theta)$ is an approximation to the exact Gaussian likelihood function (as in the stationary case - cf. Azencott and Dacunha-Castelle, 1986, Chapter XIII). This means that $\hat{\theta}_T$ is an approximate Gaussian MLE (the benefits of $\hat{\theta}_T$ over the exact MLE are discussed at the end of Section 4).

We now prove convergence of $\widehat{\theta}_T$ to θ_0 in the case where the mean is known (i.e. we assume $\mu(u) \equiv 0$). The situation of an unkonown mean is treated in Theorem 3.6 and Remark 3.7. A key step in the proof is the use of the more general central limit theorem A.2 which is of independent interest.

(3.1) Assumption.

(i) We observe the realisation $X_{1,T}, \dots, X_{T,T}$ of a locally stationary process with true transfer function A^o and mean $\mu(u)$. The true spectral density is $f(u,\lambda) = |A(u,\lambda)|^2$ with A as in Definition 2.1. $A(u,\lambda)$ is differentiable in u and λ with uniformly bounded derivative $\frac{\partial}{\partial u}\frac{\partial}{\partial \lambda} A$. g_4 is continuous.

(ii) As a model we fit a class of locally stationary processes with spectral density $f_\theta(u,\lambda)$, $\theta\in\Theta\subset\mathbb{R}^p$, Θ compact. The $f_\theta(u,\lambda)$ are uniformly bounded from above and below. The components of $f_\theta(u,\lambda)$, $\nabla f_\theta(u,\lambda)$ and $\nabla^2 f_\theta(u,\lambda)$ are continuous on $\Theta\times[0,1]\times[-\pi,\pi]$ (∇ denotes the gradient with respect to θ). $\nabla f_{\theta_0}^{-1}$ and $\nabla^2 f_{\theta_0}^{-1}$ are differentiable in u and λ with uniformly bounded derivative $\frac{\partial}{\partial u}\frac{\partial}{\partial \lambda} g$ where $g = \frac{\partial}{\partial\theta_i} f_{\theta_0}^{-1}$ or $g = \frac{\partial}{\partial\theta_i}\cdot\frac{\partial}{\partial\theta_j} f_{\theta_0}^{-1}$.

(iii) θ_0 exists uniquely and lies in the interior of Θ.

(iv) N,S and T fulfill the relations $T^{1/4} << N << T^{1/2} / \ln T$ and $S = N$ or $S / N \to 0$.

(v) The data taper $h: \mathbb{R}\to\mathbb{R}$ with $h(x) = 0$ for all $x\notin[0,1]$ is continuous on $\mathbb{R}$ and twice differentiable at all $x \notin P$ where P is a finite set and $\sup_{x\notin P} |h''(x)| < \infty$.

The assumptions on N,S and h are discussed below Theorem 4.2, in Section 6 and in Remark A.3.

(3.2) Theorem. Suppose that Assumption 3.1 holds with $\mu(u) \equiv 0$. Then

$$\hat{\theta}_T \to \theta_0$$

in probability.

Proof. Below we prove that

$$\sup_{\theta} | \mathcal{L}_T(\theta) - \mathcal{L}(\theta) | \to 0 \tag{3.1}$$

in probability. Since $\mathcal{L}(\theta)$ is minimized by θ_0 we have $\mathcal{L}_T(\hat{\theta}_T) \le \mathcal{L}_T(\theta_0)$ and $\mathcal{L}(\theta_0) \le \mathcal{L}(\hat{\theta}_T)$ which implies $\mathcal{L}(\hat{\theta}_T) \to \mathcal{L}(\theta_0)$ and therefore also $\hat{\theta}_T \to \theta_0$ in probability. To prove (3.1) we follow the idea of Hannan (1973, Lemma 1) and approximate the function $g_\theta(u,\lambda) = f_\theta(u,\lambda)^{-1}$ by the Cesaro sum of its Fourier series

$$g_\theta^{(L)}(u,\lambda) := \frac{1}{(2\pi)^2} \sum_{\ell,m=-L}^{L} \Big(1 - \frac{|\ell|}{L}\Big) \Big(1 - \frac{|m|}{L}\Big) \hat{g}_\theta(\ell,m) \exp(-i\, 2\pi\, u\, \ell - i\, \lambda\, m)$$

with L such that $\sup_\theta | g_\theta(u,\lambda) - g_\theta^{(L)}(u,\lambda)| \le \varepsilon$. We obtain

$$\sup_\theta | \mathcal{L}_T(\theta) - \mathcal{L}(\theta) | \le O(M^{-1}) + \varepsilon \frac{1}{4\pi} \frac{1}{M} \sum_{j=1}^{M} \int_{-\pi}^{\pi} \{I_N(u_j,\lambda) + f(u_j,\lambda)\}\, d\lambda\ .$$

$$+ \frac{1}{16\pi^3} \sum_{\ell,m=-L}^{L} \Big(1 - \frac{|\ell|}{L}\Big) \Big(1 - \frac{|m|}{L}\Big) \sup_\theta |\hat{g}_\theta(\ell,m)|$$

$$\cdot \Big| \frac{1}{M} \sum_{j=1}^{M} \int_{-\pi}^{\pi} \exp(-i\, 2\pi\, u_j\, \ell - i\lambda m) \{I_N(u_j,\lambda) - f(u_j,\lambda)\}\, d\lambda \Big|\ .$$

By using Lemma A.8 and Lemma A.9 the $| \dots |$ -term converges for all ℓ and m to zero in probability, while $\frac{1}{M} \sum \int I_N(u_j,\lambda)\, d\lambda$ converges to $\iint f(u,\lambda)\, d\lambda\, du$. This proves the result.

(3.3) Theorem. Suppose that Assumption 3.1 holds with $\mu(u) \equiv 0$. Then we have

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \xrightarrow{D} \mathcal{N}(0, c_h\, \Gamma^{-1}(V + W)\, \Gamma^{-1})$$

with

$$\Gamma = \frac{1}{4\pi} \int_0^1 \int_{-\pi}^{\pi} (f - f_{\theta_0})\ \nabla^2 f_{\theta_0}^{-1}\, d\lambda\, du + \frac{1}{4\pi} \int_0^1 \int_{-\pi}^{\pi} (\nabla \log f_{\theta_0})\, (\nabla \log f_{\theta_0})'\, d\lambda\ du,$$

$$V = \frac{1}{4\pi} \int_0^1 \int_{-\pi}^{\pi} f^2 \nabla f_{\theta_0}^{-1}\, \nabla f_{\theta_0}^{-1}\,'\, d\lambda\ du,$$

$$W = \frac{1}{8\pi} \int_0^1 \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} f(u,\lambda)\, f(u,\mu)\, \nabla f_{\theta_0}^{-1}(u,\lambda)\, \nabla f_{\theta_0}^{-1}(u,\mu)'\, h(\lambda, -\lambda, \mu)\, d\lambda\, d\mu\, du,$$

and $c_h = H_4 / H_2^2$ if $S = N$ and $c_h = 1$ if $S/N \to 0$.

Proof. We obtain with the mean value theorem

$$\nabla L_T(\hat{\theta}_T)_i - \nabla L_T(\theta_0)_i = \{\nabla^2 L_T(\theta_T^{(i)})\, (\hat{\theta}_T - \theta_0)\}_i$$

with $|\theta_T^{(i)} - \theta_0| \le |\hat{\theta}_T - \theta_0|$ $(i = 1, \dots, p)$. 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 $\delta > 0$, i.e., we obtain $P(\sqrt{N}\, |\nabla L_T(\hat{\theta}_T)| \ge \varepsilon) \le P(|\hat{\theta}_T - \theta_0| \ge \delta) \to 0$ for all $\varepsilon > 0$. Thus, the result follows if we prove

(i) $$\nabla^2 L_T(\theta_T^{(i)}) - \nabla^2 L_T(\theta_0) \xrightarrow{p} 0\,;$$

(ii) $$\nabla^2 L_T(\theta_0) \xrightarrow{p} \Gamma\,;$$

(iii) $$\sqrt{T}\, \nabla L_T(\theta_0) \xrightarrow{D} \mathcal{N}(0, c_h(V + W)).$$

We have

$$\nabla L_T(\theta) = \frac{1}{4\pi}\frac{1}{M}\sum_j \int_{-\pi}^{\pi} \{I_N(u_j,\lambda) - f_\theta(u_j,\lambda)\}\, \nabla f_\theta^{-1}(u_j,\lambda)\, d\lambda$$

and

$$0 = \nabla L(\theta_0) = \frac{1}{4\pi}\int_0^1 \int_{-\pi}^{\pi} \{f(u,\lambda) - f_{\theta_0}(u,\lambda)\}\; \nabla f_{\theta_0}^{-1}(u,\lambda)\, d\lambda\, du.$$

Therefore

$$\sqrt{T}\, \nabla L_T(\theta_0) = \frac{\sqrt{T}}{4\pi}\frac{1}{M}\sum_j \int_{-\pi}^{\pi} \{I_N(u_j,\lambda) - f(u_j,\lambda)\}\, \nabla f_{\theta_0}^{-1}(u_j,\lambda)\, d\lambda + O\Big(\frac{\sqrt{T}}{M}\Big)$$

which, by using Theorem A.2 implies (iii). Furthermore

$$\nabla^2 L_T(\theta) = \frac{1}{4\pi}\frac{1}{M}\sum_j \int_{-\pi}^{\pi} \{(I_N - f_\theta)\, \nabla^2 f_\theta^{-1} - \nabla f_\theta \nabla f_\theta^{-1}{}'\}\, d\lambda\, .$$

The smoothness conditions and Lemma A.8 and Lemma A.9 imply (i) and (ii).

(3.4) Corollaries and Remarks.

(i) If the model class contains the true model, then we have $f_{\theta_0} = f$. In this situation Γ, V and W simplify . In particular, we have $V = \Gamma$.

(ii) If $g_4(\lambda,-\lambda,\mu) = 0$ (for example if the process is Gaussian) then $W = 0$. If in addition $f = f_{\theta_0}$ and $c_h = 1$, then

$$\sqrt{T}\,(\hat\theta_T - \theta_0) \xrightarrow{D} N(0,\Gamma^{-1}).$$

In Dahlhaus (1994, Theorem 3.6) we prove that Γ is the limit of the Fisher information matrix. Thus, $\hat\theta_T$ is (Fisher-) efficient in this situation.

(iii) If the model is stationary (all f_θ do not depend on u) then the above theorem gives the asymptotic distribution also in the case where the true underlying process is nonstationary (cp. section 5).

(iv) Alternatively, we get the asymptotic distribution if a nonstationary model is fitted to a stationary process.

(v) If both the model and the true process are stationary, then the above distribution becomes the same as for the classical MLE and the Whittle estimate (cf. Hosoya and Taniguchi, 1982). We therefore have proved efficiency also for a new estimate (minimum distance fit to segment spectral estimates) in the classical stationary situation.

(3.5) Remark (model selection). In a practical application the problem of model selection arises. For example we might wish to compare an AR(2)-model where the coefficients are polynomials in time with a stationary AR(p) model of higher order. We will not solve this problem satisfactorily in this paper. However, we now give a heuristic derivation of the AIC-criterion (Akaike, 1974) in this situation. The criterion is used in the example of Section 6.

As a criterion of the quality of our fit we take $\mathbf{E}\mathcal{L}(\hat{\theta}_T)$, i.e. we estimate the expected Kullback-Leibler information divergence between the model and the true process (up to a constant). A quadratic expansion of $\mathcal{L}(\theta)$ around θ_0 and $\mathcal{L}_T(\theta)$ around $\hat{\theta}_T$ gives

$$\mathcal{L}(\hat{\theta}_T) \approx \mathcal{L}(\theta_0) + \frac{1}{2}\,(\hat{\theta}_T - \theta_0)'\,\nabla^2\mathcal{L}(\theta_0)\,(\hat{\theta}_T - \theta_0) \tag{3.2}$$

and

$$\mathcal{L}_T(\theta_0) \approx \mathcal{L}_T(\hat{\theta}_T) + \frac{1}{2}\,(\hat{\theta}_T - \theta_0)'\,\nabla^2\mathcal{L}_T(\hat{\theta}_T)\,(\hat{\theta}_T - \theta_0).$$

Since $\mathbf{E}\mathcal{L}_T(\theta_0) \approx \mathcal{L}(\theta_0)$, $\nabla^2\mathcal{L}(\theta_0) = \Gamma$ and $\nabla^2\mathcal{L}_T(\hat{\theta}_T) \xrightarrow{P} \Gamma$ with Γ as in Theorem 3.3 we may now estimate $\mathbf{E}\mathcal{L}(\hat{\theta}_T)$ by

$$\mathcal{L}_T(\hat{\theta}_T) + \mathbf{E}(\hat{\theta}_T - \theta_0)\,\Gamma\,(\hat{\theta}_T - \theta_0) \approx \mathcal{L}_T(\hat{\theta}_T) + \frac{1}{T}\,\mathrm{tr}\,\{\Gamma^{-1}(V + W)\} \qquad (\text{if } S/N \to 0)$$

with V,W and Γ as in Theorem 3.3. If the model is Gaussian and correctly specified ($f = f_{\theta_0}$), 0then W = 0 and V = Γ, leading to

$$\approx \mathcal{L}_T(\hat{\theta}_T) + \frac{p}{T}$$

which is the AIC (the AIC is usually $2\mathcal{L}_T(\hat{\theta}_T) + \frac{2p}{T} + \text{const.}$).

Apart from the crucial assumption $f = f_{\theta_0}$ there is another problem: Inspection of the proof of

Lemma A.8 shows that

$$\mathbf{E}\, \mathcal{L}_T(\theta_0) - \mathcal{L}(\theta_0) = 0\left(\frac{1}{M} + \frac{1}{N^2} + \frac{N}{T} \ln N\right)$$

which is of a higher order than p/T. To get rid of this problem it may be helpful to look only at the difference of $\mathcal{L}_T(\widehat{\theta}_T)$ for different models as in Findley (1985).

If a stationary model is fitted the above considerations still hold. However, a stationary model usually is fitted with a different empirical likelihood (e.g. the "exact" stationary Gaussian likelihood function or with the stationary Whittle function). Those likelihoods will in general <u>not</u> converge to $\mathcal{L}(\theta)$ if the true distribution of the process is nonstationary. However, for Yule-Walker estimates it follows from the proof of Theorem 5.1 that

$$\frac{1}{4\pi} \int \{\log f_\theta(\lambda) + \frac{I_T(\lambda)}{f_\theta(\lambda)}\}\, d\lambda$$

converges to $\mathcal{L}(\theta)$ also for nonstationary processes (where $I_T(\lambda)$ is the ordinary periodogramm). Thus, for AR(k)-processes and Yule-Walker estimates we may take the usual

$$\frac{1}{2} \log \frac{\hat{\sigma}_k^2}{2\pi} + \frac{1}{2} + \frac{k+1}{T}$$

and compare it to the above $\mathcal{L}_T(\widehat{\theta}_T) + p/T$ for a nonstationary fit.

The first term in (3.2) ($\mathcal{L}(\theta_0)$) may be regarded as a bias term (between the true f and the fitted $f_{\hat{\theta}_T}$) while the second is the variability of the estimate. Thus, minimizing the criterion $\mathcal{L}_T(\hat{\theta}_T) + p/T$ means balancing these two terms (for example for a higher model order the first term usually becomes smaller while the second gets larger).

A careful investigation of the problems arising in model selection go beyond the scope of this paper. In particular such an investigation would require a different asymptotics where the model order is allowed to increase with the sample size.

We now discuss the situation where the mean function $\mu(u)$ is unknown and estimated by $\hat{\mu}\left(\frac{t}{T}\right)$ at points u = t/T. Let

$$I_N^{\mu}(u,\lambda) := \frac{1}{2\pi H_{2,N}(0)} |d_N^{X-\mu}(u,\lambda)|^2,$$

$$\mathcal{L}_T(\theta,\mu) = \frac{1}{4\pi}\,\frac{1}{M}\sum_{j=1}^{M}\int_{-\pi}^{\pi}\left\{\log f_\theta(u_j,\lambda) + \frac{I_N^\mu(u_j,\lambda)}{f_\theta(u_j,\lambda)}\right\}d\lambda,$$

$$\hat\theta_T := \underset{\theta\in\Theta}{\operatorname{argmin}}\, \mathcal{L}_T(\theta,\mu) \quad\text{and}\quad \tilde\theta_T := \underset{\theta\in\Theta}{\operatorname{argmin}}\, \mathcal{L}_T(\theta,\hat\mu).$$

The asymptotic properties of $\hat\theta_T$ follow from Theorem 3.2 and Theorem 3.3.

(3.6) Theorem. Suppose that Assumption 3.1 holds and in addition that

(3.3) $$\hat\mu\left(\frac{t}{T}\right) - \mu\left(\frac{t}{T}\right) = o_p\left(\left(\frac{N}{T}\right)^{1/2}\right)$$

and

(3.4) $$\left\{\hat\mu\left(\frac{t}{T}\right) - \mu\left(\frac{t}{T}\right)\right\} - \left\{\hat\mu\left(\frac{t-1}{T}\right) - \mu\left(\frac{t-1}{T}\right)\right\} = o_p\left((NT)^{-1/2}\right)$$

uniformly in t. Then

$$\sqrt{T}(\tilde\theta_T - \hat\theta_T) \xrightarrow{P} 0,$$

i.e $\tilde\theta_T$ is consistent and has the same asymptotic distribution as $\hat\theta_T$.

Proof. The result is proved in the appendix.

(3.7) Remark. If the trend function is parametric with parameter τ then conditions (3.3) and (3.4) are e.g. fulfilled for $\hat\mu(u) = \mu_{\hat\tau}(u)$ where $\hat\tau$ is the least squares estimate. For a kernel estimate $\hat\mu$ with bandwidth b_T we need a bandwidth $b_T >> T^{-1/2}$. This means that the segment length of the local periodogram is not long enough for the mean estimate.

4. Fitting autoregressive models with time varying coefficients.

In this section we discuss autoregressive models with time varying coefficients. Such models have e.g. been studied before by Subba Rao (1970), Grenier (1983), Hallin (1978), Kitagawa and Gersch (1985) and Melard and Herteleer-de Schutter (1989). For simplicity we assume throughout this chaper that the mean of the process is zero. Let $X_{t,T}$ be a solution of the system of difference equations

$$\sum_{j=0}^{p} a_j\left(\frac{t}{T}\right) X_{t-j,T} = \sigma\left(\frac{t}{T}\right) \varepsilon_t \qquad \text{for } t \in \mathbb{Z} \tag{4.1}$$

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.The existence of such a process $X_{t,T}$ is discussed in Miller (1968). In Dahlhaus (1994, Theorem 2.3) we have proved that $X_{t,T}$ is locally stationary with spectral density

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

<u>The estimation equations</u>.

Suppose now that $a_\theta(u) = (a_1^\theta(u), \ldots , a_p^\theta(u))$ and $\sigma_\theta^2(u)$ depend on a finite dimensional parameter (they may be e.g. polynomials in time). With the above form of the spectrum $f_\theta(u,\lambda)$ and Kolmogorov's formula (c.f. Brockwell and Davis, 1987, Theorem 5.8.1) we obtain after some straightforward calculations

$$\mathcal{L}_T(\theta) = \frac{1}{2}\,\frac{1}{M} \sum_{j=1}^{M} \Big\{\log \sigma_\theta^2(u_j) + \frac{1}{\sigma_\theta^2(u_j)} \cdot$$

$$\big[(\Sigma_N(u_j)a_\theta(u_j) + C_N(u_j))' \,\Sigma_N(u_j)^{-1}(\Sigma_N(u_j)a_\theta(u_j) + C_N(u_j)) + c_N(u_j,0) - C_N(u_j)' \,\Sigma_N(u_j)^{-1}C_N(u_j)\big]\Big\}$$

with

$$c_N(u,j) = \int_{-\pi}^{\pi} I_N(u,\lambda) \exp(i\lambda j)\, d\lambda$$

$$= H_{2,N}(0)^{-1} \sum_{\substack{s,t=0 \\ s-t=j}}^{N-1} h\left(\frac{s}{N}\right) h\left(\frac{t}{N}\right) X_{[Tu]-N/2+s+1,T}\; X_{[Tu]-N/2+t+1,T} \;,$$

$$C_N(u) = (c_N(u,1), \ldots , c_N(u,p))' \quad \text{and} \quad \Sigma_N(u) = \{c_N(u,i-j)\}_{i,j=1, \ldots ,p}$$

(the analogous relation holds for $\mathcal{L}(\theta)$ with $\frac{1}{M}\sum_j$ replaced by the integral over time and $I_N(u,\lambda)$

replaced by the true spectrum $f(u,\lambda)$).

A nice explanation of the nature of the estimate $\hat{\theta}_T$ can be obtained from the following heuristics. The Yule-Walker estimate of a(u) in the segment of length N with midpoint u is

$$\hat{a}_N(u) = -\hat{\Sigma}_N(u)^{-1} C_N(u)$$

with asymptotic variance proportional to $\sigma^2(u)\Sigma(u)^{-1}$, and

$$\hat{\sigma}_N^2(u) = c_N(u,0) - C_N(u)' \, \Sigma_N(u)^{-1} \, C_N(u)$$

with asymptotic variance $2\sigma^4(u)$. If the model is reasonably close to the true process we can expect $\hat{\sigma}^2_{\hat{\theta}_T}(u) \approx \hat{\sigma}^2_N(u)$. Since $\log x = (x-1) - \frac{1}{2}(x-1)^2 + o\big((x-1)^2\big)$ we therefore obtain for $\mathcal{L}_T(\theta)$ in a neighbourhood of the minimum

$$\mathcal{L}_T(\theta) \approx \frac{1}{2}\,\frac{1}{M}\sum_{j=1}^{M}\left\{2\hat{\sigma}_N^4(u_j)\right\}^{-1}\left(\sigma_\theta^2(u_j) - \hat{\sigma}_N^2(u_j)\right)^2$$

$$(4.2) \qquad + \frac{1}{2}\frac{1}{M}\sum_{j=1}^{M}\left(a_\theta(u_j) - \hat{a}_N(u_j)\right)' \hat{\sigma}_N^2(u_j)^{-1}\,\hat{\Sigma}_N(u_j)\left(a_\theta(u_j) - \hat{a}_N(u_j)\right) + \frac{1}{2}\frac{1}{M}\sum_{j=1}^{M}\log \hat{\sigma}_N^2(u_j) + \frac{1}{2}.$$

Therefore, $\hat{\theta}_T$ is (approximately) obtained by a weighted least squares fit of $a_\theta(u)$ and $\sigma_\theta^2(u)$ to the Yule-Walker estimates on the segments (note that the Yule-Walker estimate with data-taper has good small sample properties - cf. Dahlhaus, 1988). If the parameters separate, i.e. $\theta = (\tau,\nu)$ with $a_\theta(u) = a_\tau(u)$ and $\sigma_\theta^2(u) = \sigma_\nu^2(u)$, we can estimate τ and ν separately.

The above representation justifies the use of graphical tools for model selection and diagnostics on a plot of the Yule-Walker estimate over time.

A weighted least squares fit to a nonparametric estimate of the AR-coefficients weighted by the asymptotic inverse of the variance has been suggested for time varying AR(1) processes by Young (1994). He used the estimate as a tool for fitting non linear time series models.

We now give an explicit formula for $\hat{\theta}_T$ if the $a_\theta(u)$ are linear in θ and $\sigma^2(u)$ is constant over time. Suppose, that some functions $f_1(u), \ldots, f_K(u)$ are given (e.g. the polynomials $f_k(u) = u^{k-1}$) and we fit the model $a_j(u) = \sum_{k=1}^{K} b_{jk}\, f_k(u)$ with σ^2 constant. Let
$b = (b_{11}, \ldots, b_{1K}, \ldots, b_{p1}, \ldots, b_{pK})'$ i.e. $\theta = (b', \sigma^2)'$. Let further F(u) be the matrix
$F(u) = \{f_i(u)f_j(u)\}_{i,j=1,\ldots,K}$ and $f(u) = (f_1(u), \ldots, f_K(u))'$. If $A \otimes B$ denotes the left direct product of the matrices A and B then direct calculations show that the parameters that minimize $\mathcal{L}_T(\theta)$ are given by

$$\hat{b}_T = -\left(\frac{1}{M}\sum_{j=1}^{M} F(u_j)\otimes\Sigma_N(u_j)\right)^{-1}\left(\frac{1}{M}\sum_{j=1}^{M} f(u_j)\otimes C_N(u_j)\right) \tag{4.3}$$

and

$$\hat{\sigma}_T^2 = \frac{1}{M}\sum_{j=1}^{M} c_N(u_j,0) + \hat{b}_T'\,\frac{1}{M}\sum_{j=1}^{M} f(u_j)\otimes C_N(u_j) \tag{4.4}$$

i.e. we obtain a linear equation system similar to the Yule-Walker equations. In case that the model is incorrect we obtain the same equations for the parameter $\theta_0 = (b_0', \sigma_0^2)$ where $\frac{1}{M}\sum_j$ is replaced by the integral over time and Σ_N and C_N are replaced by the corresponding theoretical values. In particular the minimizing values θ_0 and $\hat{\theta}_N$ exist and are unique. If σ^2 is not modelled to be constant then the estimation equations are not linear.

If different submodels (e.g. polynomials of different orders) are fitted to the $a_j(u)$ for different j, the estimate is obtained as in (4.3) and (4.4) after deleting the corresponding columns and rows in

$$\frac{1}{M}\sum_{j=1}^{M} F(u_j)\otimes\Sigma_N(u_j)$$

and

$$\frac{1}{M}\sum_{j=1}^{M} f(u_j)\otimes C_N(u_j)\,.$$

Alternatively, one may use a Levinson-Durbin type algorithm as in Grenier (1983).

Least Squares Estimates

We now prove that a <u>weighted</u> least squares estimate is an equivalent estimate for autoregressive models. Let $f_\theta(u,\lambda) = \frac{\sigma_\theta^2(u)}{2\pi} k_\theta(u,\lambda)$ where

$$k_\theta(u,\lambda) = \left|\sum_{j=0}^{p} a_j^\theta(u)\exp(i\lambda j)\right|^{-2}$$

where $a_0^\theta(u) \equiv 1$,

$$\tilde{\mathcal{L}}_T(\theta) = \frac{1}{2}\,\frac{1}{T}\sum_{t=p+1}^{T}\left\{\log\frac{\sigma_\theta^2(t/T)}{2\pi} + \frac{1}{\sigma_\theta^2(t/T)}\left|\sum_{j=0}^{p} a_j^\theta\left(\frac{t}{T}\right)X_{t-j,T}\right|^2\right\}$$

and

$$\tilde{\theta}_T = \underset{\theta \in \Theta}{\operatorname{argmin}}\, \tilde{\mathcal{L}}_T(\theta).$$

To derive the asymptotic properties of $\tilde{\theta}_T$ we need the following lemma.

<u>(4.1) Lemma.</u> Suppose $X_{t,T}$ is a locally stationary process with mean $\mu(u) = 0$ and uniformly bounded spectral density and $\phi : [0,1] \to \mathbb{R}$ is differentiable with bounded derivative. Suppose $S / N \to 0$. Then we have for all fixed i, k, t_o and $t_1 \in \mathbb{N}_o$

$$\frac{1}{M}\sum_{j=1}^{M} \phi(u_j)\, c_N(u_j,k) - \frac{1}{T}\sum_{t=t_o}^{T-t_1} \phi\left(\frac{t}{T}\right) X_{t-i,T}\, X_{t+k-i,T} = O_p\left(\frac{N}{T}\right) + O_p\left(\frac{S^2}{N^2}\right).$$

If $\phi = \phi_\theta$ and ϕ_θ and $\frac{\partial}{\partial u}\phi_\theta$ are uniformly bounded in θ, then the supremum over θ of the above difference is also of order $O_p\left(\frac{N}{T}\right) + O_p\left(\frac{S^2}{N^2}\right)$.

Proof. We have with $Y_j := X_{j,T}\, X_{j+|k|,T}$ and $\bar{h}_s = h\left(\frac{s}{N}\right)h\left(\frac{s+|k|}{N}\right)$

$$\frac{1}{M}\sum_{j=1}^{M} \phi(u_j)\, c_N(u_j,k) = \frac{1}{M}\sum_{j=1}^{M} \phi(u_j)\, \frac{1}{H_{2,N}(0)}\sum_{s=0}^{N-1-|k|} \bar{h}_s\, Y_{S(j-1)+s+1}$$

$$= \frac{1}{M}\sum_{j=1}^{M} \frac{1}{H_{2,N}(0)}\sum_{s=0}^{N-1-|k|} \phi\left(\frac{S(j-1)+s+1}{T}\right)\bar{h}_s Y_{S(j-1)+s+1} + O_p\left(\frac{N}{T}\right)$$

$$= \frac{1}{MS}\sum_{t=1}^{T-|k|} \phi\left(\frac{t}{T}\right) Y_t\, c_t + O_p\left(\frac{N}{T}\right)$$

where

$$c_t = \frac{S}{H_{2,N}(0)}\sum_{s \in S_t} \bar{h}_s \quad \text{with } S_t = \{t - S(j-1) - 1 | j = 1,\ldots,M\} \cap \{0,\ldots,N-1-|k|\}.$$

The smoothness properties of h together with $h(0) = h(1) = 0$ imply

$$c_t = 1 + O\left(\frac{S^2}{N^2}\right) \qquad \text{uniformly in t.}$$

Therefore, the above expression is equal to

$$\frac{1}{T}\sum_{t=1}^{T-|k|}\phi\left(\frac{t}{T}\right)Y_t + O_p\left(\frac{N}{T}\right) + O_p\left(\frac{S^2}{N^2}\right) = \frac{1}{T}\sum_{t=t_o}^{T-|t_1|}\phi\left(\frac{t}{T}\right)X_{t-i,T}\,X_{t+k-i,T} + O_p\left(\frac{N}{T}\right) + O_p\left(\frac{S^2}{N^2}\right).$$

(4.2) Theorem. Suppose that Assumption 3.1 holds with $\mu(u) \equiv 0$ and S fulfills $TS^4 / N^4 \to 0$. Then

$$\sqrt{T}\left(\tilde{\theta}_T - \hat{\theta}_T\right) \xrightarrow{P} 0$$

(also in the misspecified case), i.e. $\tilde{\theta}_T$ has the same asymptotic distribution as $\hat{\theta}_T$.

Proof. We only give a sketch. We have in the AR-case

$$\mathcal{L}_T(\theta) = \frac{1}{2}\,\frac{1}{M}\sum_{j=1}^{M}\left\{\log\frac{\sigma_\theta^2(u_j)}{2\pi} + \frac{1}{\sigma_\theta^2(u_j)}\sum_{\ell,m=0}^{p} a_\ell^\theta(u_j)a_m^\theta(u_j)c_N(u_j,\ell-m)\right\}.$$

Lemma 4.1 therefore gives

$$\sup_\theta |\mathcal{L}_T(\theta) - \tilde{\mathcal{L}}_T(\theta)| = o_p(1)$$

which implies as in Theorem 3.2 that

$$\tilde{\theta}_T \xrightarrow{P} \theta_o$$

In the same way we get

$$\sqrt{T}\left(\nabla\mathcal{L}_T(\theta_o) - \nabla\tilde{\mathcal{L}}_T(\theta_o)\right) = o_p(1)$$

and

$$\sup_\theta|\nabla^2\mathcal{L}_T(\theta) - \nabla^2\tilde{\mathcal{L}}_T(\theta)| = o_p(1).$$

By using the same Taylor expansion for $\tilde{\theta}_T$ and $\tilde{\mathcal{L}}_T$ as in the proof of Theorem 3.3 we now obtain the result..

It is remarkable that Theorem 4.2 holds regardless of the choice of the data taper and for most of the S and N. The effect of the choice of these parameters can probably only be seen in a higher order asymptotics. This shows the low sensitivity of $\hat{\theta}_T$ with respect to the choice of S, N and h.

In the general case it is difficult to calculate $\tilde{\theta}_T$. However, in the homoscedastic case $\sigma^2_\theta(t/T) \equiv \sigma^2$, i.e. $\theta = (\sigma^2,\tau)$ we obtain

(4.5) $$\tilde{\tau}_T = \operatorname{argmin} \frac{1}{T} \sum_{t=p+1}^{T} \left| \sum_{j=0}^{p} a_j^\tau\left(\frac{t}{T}\right) X_{t-j,T} \right|^2$$

and

$$\tilde{\sigma}_T^2 = \frac{1}{T} \sum_{t=p+1}^{T} \left| \sum_{j=0}^{p} a_j^{\tilde{\tau}_T}\left(\frac{t}{T}\right) X_{t-j,T} \right|^2 .$$

If the a_j^τ are linear in τ (as in the polynomial case) we therefore have a linear least squares problem.

We now compare the minimum distance estimate $\hat{\theta}_T$ to the least squares approach in the heteroscedastic case. Suppose that the parameters separate, i.e. $\theta = (\tau,\kappa)$ where $a_j^\theta(u) = a_j^\tau(u)$ and $\sigma^2_\theta(u) = \sigma^2_\kappa(u)$. Thus, we have

$$f_\theta(u,\lambda) = \frac{\sigma^2_\kappa(u)}{2\pi}\, k_\tau(u,\lambda).$$

Kolmogorov's formula gives

$$\int_{-\pi}^{\pi} \log f_\theta(u,\lambda)\, d\lambda = 2\pi \log \frac{\sigma^2_\kappa(u)}{2\pi} .$$

Therefore,

$$\int_{-\pi}^{\pi} f_\theta\, \nabla_\tau f_\theta^{-1} d\lambda = 0$$

and

$$\int_{-\pi}^{\pi} f_\theta\, \nabla^2_\tau f_\theta^{-1} d\lambda = \int_{-\pi}^{\pi} (\nabla_\tau \log f_\theta)\,(\nabla_\tau \log f_\theta)'\, d\lambda.$$

Similarly,

$$\int_{-\pi}^{\pi} (\nabla_\tau \log f_\theta)\,(\nabla_\kappa \log f_\theta)'\, d\lambda = 0.$$

If the model is correctly specified ($f = f_{\theta_o}$ where $\theta_o = (\tau_o,\kappa_o)$) we therefore obtain for the minimum distance estimate $\hat{\theta}_T = \left(\hat{\tau}_T,\hat{\kappa}_T\right)$ from Theorem 3.3 that

$$\sqrt{T}\left(\hat{\tau}_T - \tau_o\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(O, V_{\tau_o}^{-1}\right)$$

where

$$V_{\tau_o} = \int_0^1 \bar{V}(u)du$$

and

$$\bar{V}(u) = \frac{1}{4\pi}\int_{-\pi}^{\pi}\left(\nabla_\tau \log f_{\theta_o}(\lambda,u)\right)\left(\nabla_\tau \log f_{\theta_o}(\lambda,u)\right)' d\lambda.$$

We now study the behaviour of the least sqares estimate $\tilde{\tau}_T$ as defined in (4.5) (κ may be estimated afterwards e.g. by some fit of the estimated residuals at time point t/T to σ_κ^2 (t / T)). The following theorem implies that the LSE is less efficient in the heteroscedastic case. For simplicity we restrict ourselves to the case where the model is correct.

(4.3) Theorem. Suppose Assumption 3.1 (i) - (iii) holds with $\mu(u) \equiv 0$ and $f = f_{\theta_o}$. Then we have

$$\sqrt{T}(\tilde{\tau}_T - \tau_o) \xrightarrow{\mathcal{D}} \mathcal{N}\ (0,U)$$

where

$$U = \left\{\int_0^1 \sigma_{\kappa_o}^2(u)\ \tilde{V}(u)du\right\}^{-1}\left\{\int_0^1 \sigma_{\kappa_o}^4(u)\ \tilde{V}(u)du\right\}\left\{\int_0^1 \sigma_{\kappa_o}^2(u)\ \tilde{V}(u)du\right\}^{-1}.$$

We have $U \geq V_{\tau_o}^{-1}$ with $U = V_{\tau_o}^{-1}$ if and only if $\sigma_{\kappa_o}^2(u)$ is constant.

Proof. We only give a sketch. As in Theorem 4.2 we can show by using Lemma 4.1 that $\sqrt{T}\left(\tilde{\tau}_T - \tilde{\tilde{\tau}}_T\right) \xrightarrow{P} 0$ where $\tilde{\tilde{\tau}}_T$ minimizes

$$\tilde{\tilde{\mathcal{L}}}_T(\tau) := \frac{1}{M}\sum_{j=1}^{M}\int_{-\pi}^{\pi}\frac{I_N(u_j,\lambda)}{k_\tau(u_j,\lambda)}\,d\lambda$$

where S = 1 and N and h fulfill Assumption 3.1 (iv) + (v). It is easy to show that τ_o minimizes

$$\tilde{\tilde{L}}(\tau) := \int_0^1 \int_{-\pi}^{\pi} \frac{f_{\theta_o}(u,\lambda)}{k_\tau(u,\lambda)} \, d\lambda \, du.$$

It now follows in exactly the same way as in the proofs of Theorem 3.2 and 3.3 that

$$\tilde{\tilde{\tau}}_T \xrightarrow{P} \tau_o$$

and

$$\sqrt{T}\left(\tilde{\tilde{\tau}}_T - \tau_o\right) \xrightarrow{D} N\left(0, \tilde{\Gamma}^{-1} \tilde{V} \tilde{\Gamma}^{-1}\right)$$

where

$$\tilde{\Gamma} = \frac{1}{4\pi} \int_0^1 \int_{-\pi}^{\pi} f_{\theta_o} \nabla_\tau^2 k_{\tau_o}^{-1} \, d\lambda \, du = \frac{1}{2\pi} \int_0^1 \sigma_{\kappa_o}^2(u) \, \overline{V}(u) \, du$$

and

$$\tilde{V} = \frac{1}{4\pi} \int_0^1 \int_{-\pi}^{\pi} f_{\theta_o}^2 \left(\nabla_\tau k_{\tau_o}\right)^2 d\lambda \, du = \frac{1}{4\pi^2} \int_0^1 \sigma_{\kappa_o}^4(u) \, \overline{V}(u) \, du$$

which proves the first part. The matrix

$$\begin{pmatrix} \int_0^1 \sigma_{\kappa_o}^4(u)\overline{V}(u)du & \int_0^1 \sigma_{\kappa_o}^2(u)\overline{V}(u)du \\ \int_0^1 \sigma_{\kappa_o}^2(u)\overline{V}(u)du & \int_0^1 \overline{V}(u)du \end{pmatrix}$$

is non-negative definite which leads with Theorem 12.2.21(5) of Graybill (1983) to $U \geq V_{\tau_o}^{-1}$. If $\sigma_{\kappa_o}^2(u)$ is constant we have $U = V_{\tau_o}^{-1}$. Conversely let $U = V_{\tau_o}^{-1}$. Theorem 8.2.1(1) of Graybill implies that the above matrix is singular, i.e. there exists a vector $(x',y') \neq 0$ with

$$\int_0^1 \left(\sigma_{\kappa_o}^2(u)\, x + y\right)' \overline{V}(u) \left(\sigma_{\kappa_o}^2(u)\, x + y\right) du = 0$$

Since $\overline{V}(u)$ is positive definite we have $\sigma_{\kappa_o}^2(u) = -y_i/x_i$ which implies the result.

Thus, the least squares estimate is less efficient than the minimum distance estimate $\hat{\theta}_T$ in the heteroscedastic case. It is heuristically clear that a weighted least squares estimate will be fully efficient. However, such an estimate has no computational advantages since the weights depend on the unknown parameters and the estimation equations therefore are nonlinear.

A third candidate for estimation is the exact (Gaussian) maximum likelihood estimate from which we conjecture that it is also efficient. Since a time varying AR-model can be written in state space form the MLE can be calculated by using the prediction error decomposition together with a numerical optimization procedure. However, the system matrices in the state space form are time varying, which leads to an extremely large computation time. Therefore, the MLE is not a suitable candidate - in particular if different models are fitted to the data in a model selection process.

The following procedure seems to be reasonable for autoregressive models in a practical situation: For homoscedastic models one uses the linear equation system (4.3) and (4.4) together with the AIC as in Remark 3.5 for model selection and a graphical investigation of the nonparametric estimate $\hat{a}(u)$ for diagnostic checking. An example is given in Section 6. For heteroscedastic errors one may minimize the modified likelihood (4.2) which also leads to linear estimation equations (for models linear in the parameters). The final estimate may be improved by a one-step MLE. Of course a detailed simulation study is necessary to verify these suggestions.

We finally remark that the minimum distance estimate $\hat{\theta}_T$ can be computed for arbitrary locally stationary models while for the LSE and the state space representation of the MLE a special form of the model is necessary.

5. Fitting stationary models to nonstationary processes.

We now discuss the situation where the fitted model is stationary, i.e. $f_\theta(\lambda) = f_\theta(u, \lambda)$ does not depend on u. In this situation we obtain

$$\mathcal{L}(\theta) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \left\{ \log f_\theta(\lambda) + \frac{\int_0^1 f(u,\lambda)\, du}{f_\theta(\lambda)} \right\} d\lambda$$

and therefore, for $\theta_0 = \arg\min_\theta \mathcal{L}(\theta)$ the equations

$$\int_{-\pi}^{\pi} \left(\int_0^1 f(u,\lambda) du \right) \nabla f_{\theta_0}^{-1}(\lambda) d\lambda = \int_{-\pi}^{\pi} f_{\theta_0}(\lambda) \nabla f_{\theta_0}^{-1}(\lambda)\, d\lambda \,.$$

Thus θ_0 is that parameter for which $f_\theta(\lambda)$ approximates the time-integrated true spectrum $\int_0^1 f(u,\lambda) du$ best.

In the case of a stationary AR(p)-model the above equations are the (theoretical) Yule-Walker equations, i.e. we obtain for $\theta_0 = (a_0', \sigma_0^2)'$ with $a_0 = (a_{01}, \ldots, a_{0p})'$

$$a_0 = -\Sigma^{-1} C \quad \text{and} \quad \sigma_0^2 = c(0) + a_0' C$$

with

$$c(k) = \int_{-\pi}^{\pi} \{ \int_0^1 f(u,\lambda) du \} \exp(i\lambda k)\, d\lambda,$$

$$C = (c(1), \ldots, c(p))' \quad \text{and} \quad \Sigma = \{c(i-j)\}_{i,j=1,\ldots,p} .$$

For $\hat{\theta}_T = (\hat{a}_T', \hat{\sigma}_T^2)'$ we obtain the corresponding equations

$$\hat{a}_T = -\hat{\Sigma}_T^{-1} \hat{C}_T \qquad \text{and} \qquad \hat{\sigma}_T^2 = c_T(0) + \hat{a}_T' \hat{C}_T$$

with

$$\hat{c}_T(k) = \int_{-\pi}^{\pi} \Big\{ \frac{1}{M} \sum_{j=1}^{M} I_N(u_j,\lambda) \Big\} \exp(i\lambda k) d\lambda = \frac{1}{M} \sum_{j=1}^{M} c_N(u_j,k),$$

$$\hat{C}_T = (\hat{c}_T(1), \ldots, \hat{c}_T(p))' \quad \text{and} \quad \hat{\Sigma}_T = \{\hat{c}_T(i-j)\}_{i,j=1,\ldots,p} .$$

The asymptotic distribution of $\sqrt{T}(\hat{\theta}_T - \theta_0)$ is given in Theorem 3.3. Straightforward calculations give in this case

$$\Gamma = \begin{pmatrix} \frac{1}{\sigma^2} c_0(i-j)_{i,j=1,\ldots,p} & 0 \\ 0 & \frac{1}{2\sigma_0^4} \end{pmatrix} .$$

The matrices V and W simplify only minor. (Note, that if the true process is also stationary with $f(\lambda) \neq f_{\theta_0}(\lambda)$ and $g_4(\lambda, -\lambda, \mu)$ is constant, then W disappears - however, this does not hold in the nonstationary case).

However, $\hat{\theta}_T$ is not the estimate one would usually use for stationary models. For example, for AR-processes one would use e.g. (tapered) Yule-Walker estimates, the Burg algorithm or (Gaussian) maximum likelihood estimates. In the following theorem we prove that Yule-Walker estimates have the same asymptotic behaviour as $\hat{\theta}_T$ if the true process is (possibly) nonstationary.

(5.1) Theorem. Suppose the true process is of the form (2.3) with $\mu(u) \equiv 0$. Let $\tilde{\theta}_T = (\tilde{a}_T, \tilde{\sigma}^2)$ be the Yule-Walker estimate for a stationary AR(p)-model, i.e.

$$\tilde{a}_T = -\tilde{\Sigma}_T^{-1} \tilde{C}_T, \quad \tilde{\sigma}_T^2 = \tilde{c}_T(0) + \tilde{a}_T' \tilde{C}_T$$

with $\tilde{c}_T(k) = \frac{1}{T} \sum_{j=1}^{T-|k|} X_j X_{j+|k|}$, $\tilde{C}_T = (\tilde{c}_T(1), \ldots, \tilde{c}_T(p))'$ and $\tilde{\Sigma}_T = \{\tilde{c}_T(i-j)\}_{i,j=1,\ldots,p}$. If $\hat{\theta}_T$ is as in section 3 with S = 1 and N and a taper as in Assumption 3.1, then $\sqrt{T}(\tilde{\theta}_T - \hat{\theta}_T)$ converges to zero in probability and

$$\sqrt{T}(\tilde{\theta}_T - \theta_0) \xrightarrow{D} N(0, \Gamma^{-1}(V+W)\,\Gamma^{-1})$$

with Γ as above and V,W as in Theorem 3.3 .

Proof. With θ_0 as above we have

$$-(\tilde{\Sigma}_T a_0 + \tilde{C}_T) = \tilde{\Sigma}_T (\tilde{a}_T - a_0)$$

and

$$-(\hat{\Sigma}_T a_0 + \hat{C}_T) = \hat{\Sigma}_T (\hat{a}_T - a_0) .$$

Thus, it is sufficient to prove that $\sqrt{T}\,(\tilde{c}_T(k) - \hat{c}_T(k))$ tends to zero in probability. Since $\hat{c}_T(k) = \frac{1}{M} \sum_{j=1}^{M} c_N(u_j,k)$ this follows from Lemma 4.1.Therefore, the first assertion is proved if we choose $T^{1/4} << N << T^{1/2}$. The asymptotic normality then follows from Theorem 3.3 .

For tapered Yule-Walker estimates, i.e. the corresponding estimate with $\tilde{c}_T(k) = \frac{1}{H_{2,T}^0(0)} \sum_{j=1}^{T-|k|} h_0(\frac{j}{T})\, h_0(\frac{j+|k|}{T})\, X_j X_{j+|k|}$ (with a taper h_0 that may be different from the taper h used in $\hat{\theta}_T$), we expect the following result: $\tilde{\theta}_T$ will no longer converge to θ_0 but to

$$\theta_0' = \arg\min \frac{1}{4\pi} \int_{-\pi}^{\pi} \left\{ \log f_\theta(\lambda) + \frac{\int_0^1 h_*(u)\, f(u,\lambda)\, du}{f_\theta(\lambda)} \right\} d\lambda$$

with $h_*(u) = \left\{ \int_0^1 h_0^2(v)\ dv \right\}^{-1} h_0^2(u)$. We conjecture that $\sqrt{T}(\tilde{\theta}_T - \theta_0')$ is asymptotically normal with Γ,V,W in Theorem 3.3 where $\int_0^1 \ldots$ du is always replaced by $\int_0^1 h_*(u) \ldots du$.

A few remarks on the use of data tapers seem to be necessary. For stationary time series tapered estimates are less efficient than nontapered estimates or equally efficient if the taper disappears asymptotically (c.f. Dahlhaus, 1988). On the other hand their small sample behaviour is very often much better, in particular the resolution problems of the nontapered estimate are cured. In the situation of this paper Theorem 5.1 says that the asymptotic behaviour of the nontapered Yule-Walker estimate is the same as of the (tapered) estimate $\hat{\theta}_T$. However, for small samples we conjecture that $\hat{\theta}_T$ will be much better.

6. A simulation example

We now briefly present a simulation example for the estimate $\hat{\theta}_T$ in a misspecified situation. If we have a locally stationary process with smoothly varying characteristics then it is likely that $\hat{\theta}_T$ leads to reasonable results for a large sample size, since then the data within each segment are close to a realisation of a stationary process. The interesting question now is how the estimate behaves for moderate or small sample sizes, i.e. whether the asymptotics together with the model of local stationarity yields to a reasonable description also for small data sets.

We have generated T = 128 observations of a time varying AR(2)-process (4.1) with parameters as described below. Several models were fitted by using the equations (4.3) and (4.4).

The choice of the data taper is different from stationary time series. Theorem 3.3 says that there is no efficiency loss for overlapping segments. Theorem 4.2 even means that all estimates are stochastically equivalent to the least squares estimate, regardless of the taper. We have used the 100 % - Tukey Hanning taper $h(x) = \frac{1}{2}\ [1 - \cos(2\pi x)]$. This taper has in addition to good bias properties with respect to leakage also the advantage that the observations at the edge of each segment are weighted down which makes the estimate heuristically less sensitive against the instationarity within the segments.

The shift should in general be as small as possible - the theoretical results hold even for S = 1. However, this choice is very computer intensive. In the simulation we chose S = 2. For the segment length we chose N = 16 (i.e. M = 57). We also tried other parameters. The results turned out to be very insensitive to the choice of N, S and h which is in accordance with Theorem 4.2.

As the parameters of the true AR(2)-process we chose $\sigma(u) \equiv 1$,

$$a_1(u) = -1.8 \cos(1.5 - \cos 4\pi u)$$

$$a_2(u) = + \ 0.81$$

together with Gaussian innovations ε_t, i.e. for u fixed the roots of the characteristic polynomial are

$$\frac{1}{0.9} \ \exp\left[\pm\, i(1.5 - \cos 4\pi u)\right] .$$

Figure 1 below shows the observations. As it could be expected from the above parameters they show a periodic behaviour with time varying period-length. The left picture of Figure 2 shows the true time varying spectrum of the process. We have fitted a time varying AR-model of order p to the data where the coefficients were modeled as polynomials with different orders. Thus, we have fitted the model

$$a_j(u) = \sum_{k=0}^{K_j} b_{jk}\, u^{k-1} \qquad (j = 1, \dots , p)$$

$$\sigma^2 = c$$

to the data. The model orders p, $K_1, \dots , K_p$ were chosen by minimizing the AIC-criterion

$$\text{AIC}(p, K_1, \dots , K_p) = \log \hat{\sigma}^2(p, K_1, \dots , K_p) + 2\,(p + 1 + \sum_{j=1}^{p} K_j) / T.$$

Table 1 shows these values for p = 2 and different K_1 and K_2 . The values for other p turned out to be larger. Thus, a model with p = 2, $K_1 = 6$, $K_2 = 0$ was fitted.

The corresponding spectrum is the right picture of Figure 2. The difference to the true spectrum is plotted in Figure 3. The function $a_1(u)$ and its estimate are plotted in Figure 4. For $\hat{a}_2(u)$ we obtained 0.71 (a constant was fitted because of $K_2 = 0$) while the true $a_2(u)$ was 0.81. Furthermore, $\hat{\sigma}^2 = 1.71$ while $\sigma^2 = 1.0$.

The quality of the fit is remarkable. However, two negative effects can be observed. The fit of $a_1(u)$ becomes rather bad outside $u_1 = 0.063$ and $u_M = 0.938$. This is not surprising, due to the behaviour of a polynomial and the fact that the use of $\mathcal{L}_T(\theta)$ as a distance only punishes bad fits inside the interval $[u_1,u_M]$. This end effect vanishes if one chooses $K_1 = 8$ instead of $K_1 = 6$. A better way seems to be to modify $\mathcal{L}_T(\theta)$ and to include periodograms of shorter lengths at the end points (e.g. $I_{N/2}\,(N/(4T),\lambda)$). The second effect is that in the frequency representation the peak is underestimated. This is due to the non-stationarity of the process on the intervals $(u_j - N/(2T),\, u_j + N/(2T)]$ where $I_N(u_j,\lambda)$ and $c_N(u_j,k)$ are calculated.

We finally remark that this example is typical. The same properties can be observed for other realisations. Even for T = 64 the results turned out to be quite good.

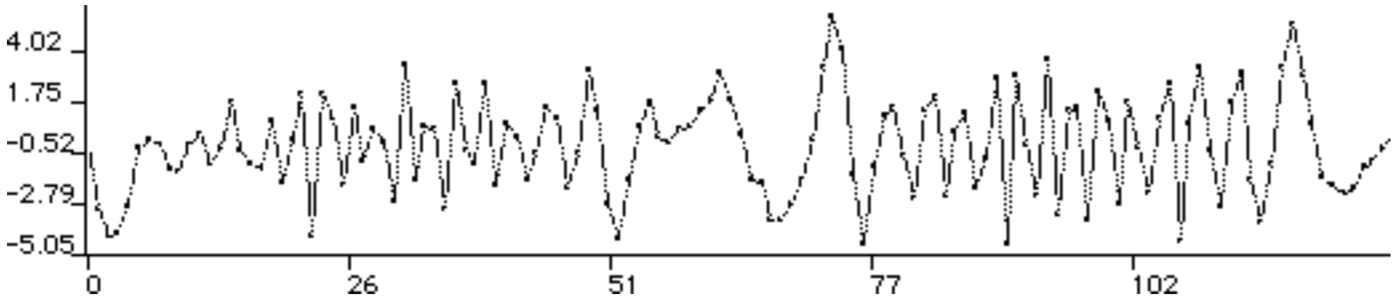

Figure 1. T=128 realisations of a time varying AR-model

K_2 \ K_1	4	5	6	7	8	9
0	0.929	0.888	0.669	0.685	0.673	0.689
1	0.929	0.901	0.678	0.694	0.682	0.698
2	0.916	0.888	0.694	0.709	0.697	0.712

Table 1. Values of AIC for p = 2 and different polynomial orders

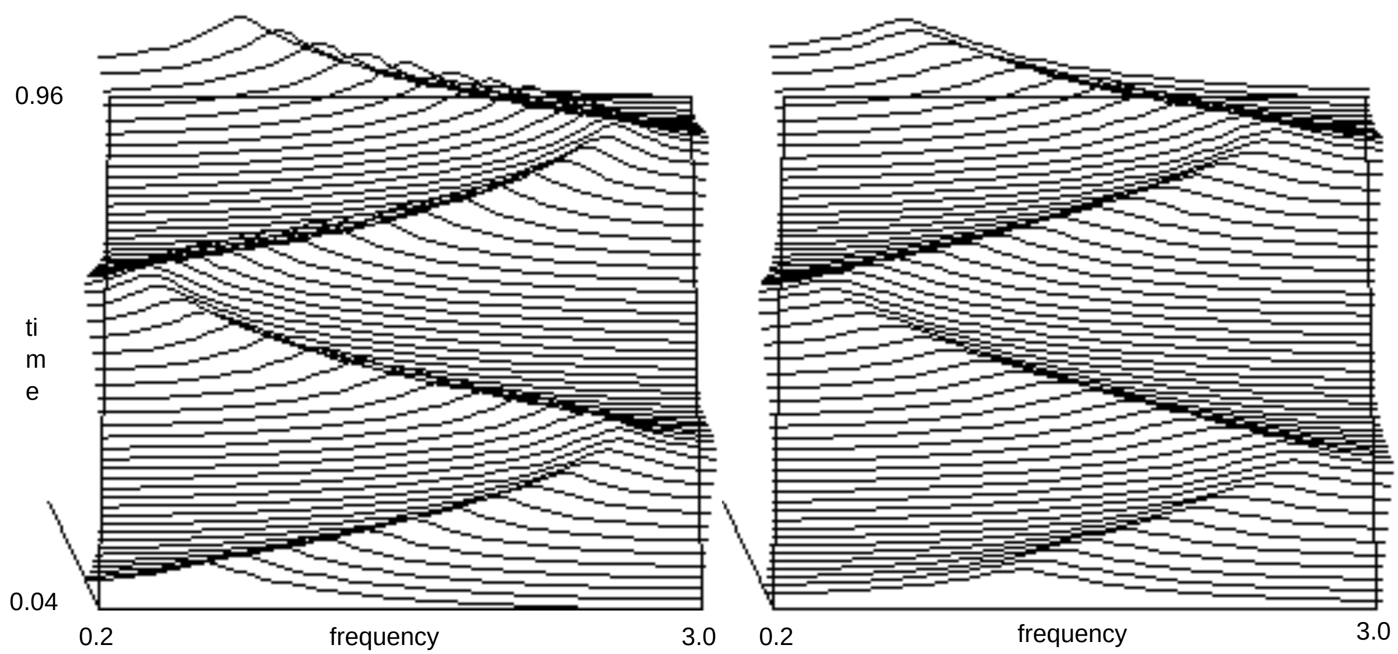

Figure 2. True and estimated spectrum of a time varying AR - process

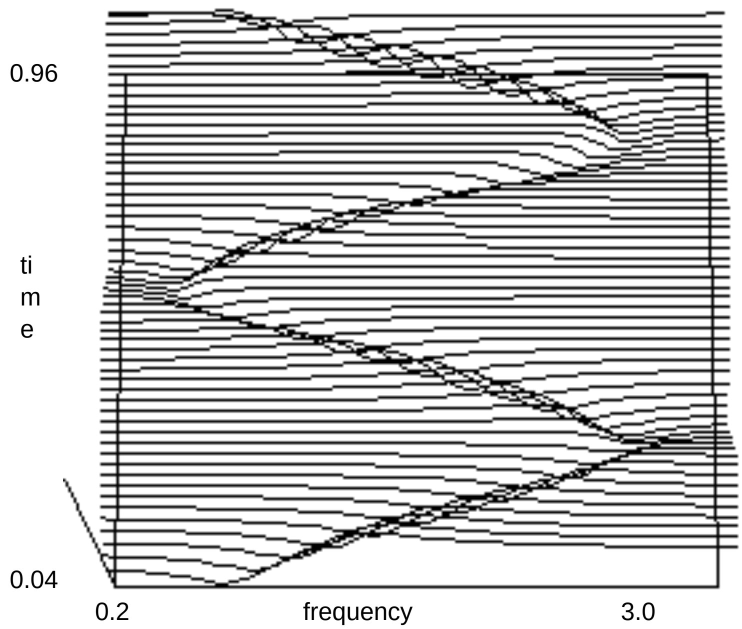

Figure 3. Difference of estimated and true spectrum

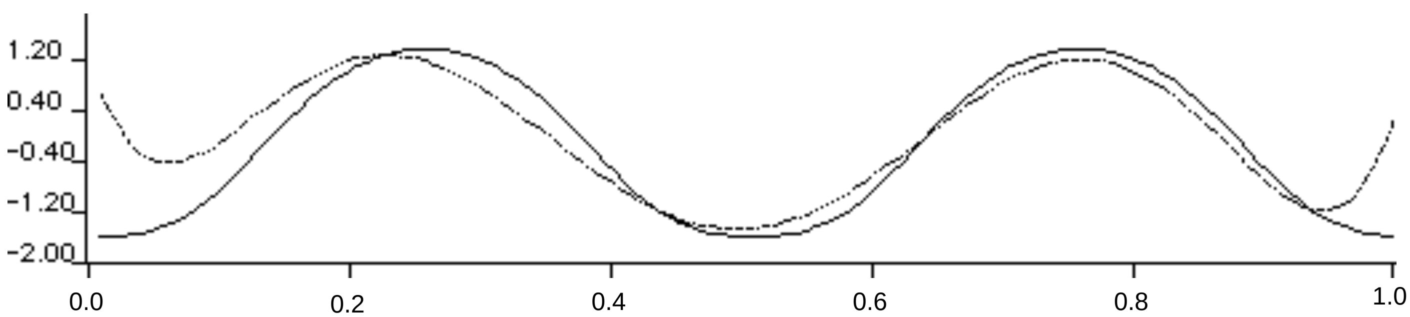

Figure 4. True and estimated time varying coefficient a (u)

7. Concluding remarks.

In this paper we have presented an asymptotic theory for processes that have an evolutionary spectral representation. We have derived the asymptotic behaviour of minimum distance estimates in the spectral domain and of least squares estimates for time varying autoregressive processes. The results also hold when the model is incorrect, i.e. when it does not contain the true process.

The theory leads to a new estimate for various nonstationary models. Simulations show that this estimate works quite well in practice. It is attractive that the classical stationary ARMA model can be included as a special case (as for AR-models in the simulation example). Furthermore, the AIC criterion seems to work reasonably well in this situation (although a strict theoretical justification is still missing). In particular, the AIC can be used to decide between stationary and nonstationary models (as in the example where the stationary model corresponds to $K_1 = K_2 = 0$).

The parameter estimates are minimum distance estimates in the spectral domain. Since our distance function is an approximate Gaussian likelihood the results can in principle only apply to models whose parameters can be identified from this distance function, i.e. to time varying linear models. Here are the limitations of the approach-although it may be possible to derive similar results with other distance functions for nonlinear models.

As any asymptotic theory our approach simplifies the situation (for example, time varying AR-processes have locally the spectral density of a stationary AR-process). The benefit of this simplification is a framework for such processes which makes theoretical results for parameter estimates possible. It is obvious that it is (in principle) possible to study the behaviour of other estimates (e.g. exact MLE'S or local Burg estimates) within this framework. Furthermore, one may look for modifications of the suggested procedures, e.g. with better bias properties (cp. Remark A.3) and better edge properties. For stationary models our asymptotic theory is the same as the classical asymptotic theory.

On the other hand one could argue that with the simplification important features of a nonstationary process are lost, e.g. the special form of $A^{\circ}_{t,T}$ for a time varying AR-process (cf. Mélard and Herteller-de Schutter, 1989). However, one may use this theory also to study some of these effects. For example, one could study the asymptotic properties of the modified estimator for AR-models with $|A^{\circ}_{t,T}(\lambda)|^2$ instead of $|A(u,\lambda)|^2$ in $L(\theta)$ and $L_T(\theta)$.

Appendix: A central limit theorem.

This appendix contains the technical details of the proof of Theorem 3.2 and Theorem 3.3. It basically consists of the proof of the following Theorem A.2. This theorem is of independent interest. It has applications that go beyond the scope of this paper.

Suppose S, M, N, t_j, u_j and $I_n(u,\lambda)$ are defined as in section 3. For ϕ: $[0,1] \times [-\pi,\pi] \to \mathbb{C}$ we set

$$J_T(\phi) := \frac{1}{M} \sum_{j=1}^{M} \int_{-\pi}^{\pi} \phi(u_j,\lambda)\, I_N(u_j,\lambda)\, d\lambda$$

and

$$J(\phi) := \int_0^1 \int_{-\pi}^{\pi} \phi(u,\lambda)\, f(u,\lambda)\, d\lambda\, du\ .$$

To prove asymptotic normality for $\sqrt{T}(J_T(\phi) - J(\phi))$ we need the following assumptions.

(A.1) Assumption.

(i) Let $X_{t,T}$ be a locally stationary process with mean $\mu(u) = 0$ as in Definition 2.1. Suppose that the functions $A(u,\lambda)$ (from Definition 2.1) and $\phi_j(u,\lambda)$ ($j = 1, \dots, k$) are 2π-periodic in λ and the periodic extensions are differentiable in u and λ with uniformly bounded derivative $\frac{\partial}{\partial u} \frac{\partial}{\partial \lambda} A$ (ϕ_j respectively). g_4 is continuous.

(ii) The parameters N,S and T fulfill the relations $T^{1/4} << N << T^{1/2} / \ln T$ and $S = N$ or $S / N \to 0$.

(iii) The data taper $h: \mathbb{R} \to \mathbb{R}$ with $h(x) = 0$ for all $x \notin [0,1]$ is continuous on $\mathbb{R}$ and twice differentiable at all $x \notin P$ where P is a finite set and $\sup_{x \notin P} | h''(x) | < \infty$.

(A.2) Theorem. Suppose $X_{1,T}, \dots, X_{T,T}$ are realisations of a locally stationary process and Assumption A.1 is fulfilled. Then

$$\sqrt{T}(J_T(\phi_j) - J(\phi_j))_{j=1,\dots,k} \xrightarrow{D} (\xi_j)_{j=1,\dots,k}$$

where ξ is a Gausssian random vector with mean zero and

$$\begin{aligned} \mathrm{cov}(\xi_i, \xi_j) = 2\pi\ c_h \int_0^1 \Big[\int_{-\pi}^{\pi} \phi_i(u,\lambda) \{ \overline{\phi_j(u,\lambda)} + \overline{\phi_j(u,-\lambda)} \}\ f(u,\lambda)^2\, d\lambda \\ + \iint_{-\pi}^{\pi} \phi_i(u,\lambda) \overline{\phi_j(u,-\mu)}\ f(u,\lambda)\, f(u,\mu)\, h_4(\lambda, -\lambda, \mu)\, d\lambda\, d\mu \Big] du \end{aligned}$$

with $c_h = (\int_0^1 h(u)^4 du) / (\int_0^1 h(u)^2\, du)^2$ if $S = N$ and $c_h = 1$ if $S/N \to 0$.

(A.3) Remarks. The conditions on M, S and, in particular, on N seem to be restrictive. However, we regard it as remarkable that $\sqrt{T}$ consistency holds at all. Most of the restrictions on N result from the $\sqrt{T}$-unbiasedness (Lemma A.8). Inspection of the proof leads to the conjecture that it is not possible to relax these conditions (apart from some log-terms). This can be made clear by some heuristics: with the periodogram over the first segment we estimate f at time $\frac{N}{2T}$. To conclude from this to f at zero $\sqrt{T}$ - consistently we need $\frac{N}{\sqrt{T}} \to 0$. On the other hand the bias of the periodogram (with a data taper) is $O(N^{-2})$ which leads to the condition $\frac{\sqrt{T}}{N^2} \to 0$. We conjecture that the rate $O(N^{-2})$ cannot be improved with a periodogram type estimator. A periodogram without taper would lead to a bias of $O(N^{-1})$ and therefore to $\frac{\sqrt{T}}{N} \to 0$ which contradicts $\frac{N}{\sqrt{T}} \to 0$. Thus, without taper it is not possible to achieve $\sqrt{T}$-consistency at all. It is noteworthy that the use of a data taper does not lead to an increase of the variance if $S / N \to 0$. However, this is heuristically clear since in this case all observations are used "equally often" (as $T \to \infty$). Note the similarity of the covariance structure to an analogous result in the stationary case (cf. Brillinger, 1981, Theorem 7.6.1).

Theorem A.2 is proved by proving the convergence of the cumulants of all orders (Lemma A.8, Lemma A.9 and Lemma A.10). A key role in the proofs is played by the following function. Let $L_T: \mathbb{R} \to \mathbb{R}$, $T \in \mathbb{R}^+$, be the periodic extension (with period 2π) of

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

<u>(A.4) Lemma.</u> Let k, l, $S, M, N, T \in \mathbf{N}$, $\alpha, \beta, \nu, \mu, x \in \mathbb{R}$ and $\Pi := (-\pi,\pi]$. We obtain with a constant K independent of T:

(a) $L_T(\alpha)$ is monotone increasing in T and decreasing in $\alpha \in [0,\pi]$.

(b) $\int_\Pi L_T(\alpha)^k \, d\alpha \le K T^{k-1}$ for all $k > 1$.

(c) $\int_\Pi L_T(\alpha) \, d\alpha \le K \ln T$ for $T > 1$.

(d) $|\alpha| \, L_T(\alpha) \le K$.

(e) $\int_\Pi L_T(\beta - \alpha) \, L_T(\alpha + \gamma) \, d\alpha \le K \, L_T(\beta + \gamma) \ln T$.

(f) $L_T(\nu)^k \, L_T(\mu)^l \le L_T(\frac{\nu - \mu}{2})^k \, L_T(\mu)^l + L_T(\nu)^k \, L_T(\frac{\nu - \mu}{2})^l$.

(g) $L_T(c\alpha) \le K_C \, L_T(\alpha)$ for $|c\alpha| \le \pi$.

(h) $\int_\Pi L_N(\alpha)^l \, L_M(S(\alpha - \beta))^k \, d\alpha \le K \, \frac{N^l M^{k-1}}{S} \ln M \, \{k = 1\} \ln S \, \{l = 1\}$.

(i) $\int_{\Pi} L_N(\lambda - x) L_N(x - \mu) L_M(S(\alpha - x))\, L_M(S(x - \beta))dx \le K\frac{N}{S}\ln M \ln S\; L_N(\lambda - \mu)\; L_M(S(\alpha - \beta)).$

(j) $\int_{\Pi} L_N(\lambda - x)\; L_N(x - \mu)\; L_M(S(\alpha - x))\; dx \le K\, \frac{N}{S}\; \ln M \ln S\; L_N(\lambda - \mu)$.

Proof. The proofs are technical but straightforward. Some of them may be found in Dahlhaus (1983) or Dahlhaus (1985). (f) is proved by considering the cases $|\nu| \ge \frac{|\nu - \mu|}{2}$ and $|\mu| \ge \frac{|\nu - \mu|}{2}$. (e) is a consequence of (f) and (g) . (h) is proved by splitting the integral into $\int_{|\alpha| \le 1/S} \dots$ and $\int_{|\alpha| \ge 1/S} \dots = \sum_j \int_{[j/S,\ (j+1)/S]} \dots$. (i) and (j) then follow from (f) and (h).

For a complex-valued function f we define

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

and, for the data taper h(x)

$$H_{k,N}(\lambda) := H_N\left(h^k\left(\frac{\cdot}{N}\right), \lambda\right)$$

and

$$H_N(\lambda) = H_{1,N}(\lambda) .$$

Direct calculation gives

$$\int_{-\pi}^{\pi} H_{k,N}(\beta - \alpha)\, H_{l,N}(\alpha - \gamma)\, d\alpha = 2\pi\, H_{k+l,N}(\beta - \gamma) .$$

(A.5) Lemma: Let $N,T \in \mathbb{N}$. Suppose h fulfills Assumption A.1(iii) and $\psi\colon [0,1] \to \mathbb{R}$ is differentiable with bounded derivative. Then we have for $0 \le t \le N$

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

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

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

Proof. Summation by parts gives

$$H_N(\psi\left(\frac{\cdot}{T}\right) h\left(\frac{\cdot}{N}\right), \lambda) - \psi\left(\frac{t}{T}\right) H_N(\lambda) = \sum_{s=0}^{N-1} \{\psi\left(\frac{s}{T}\right) - \psi\left(\frac{t}{T}\right)\} \; h\left(\frac{s}{N}\right) \exp(-i\lambda s)$$

$$= -\sum_{s=0}^{N-1} \{\psi\left(\frac{s}{T}\right) - \psi\left(\frac{s-1}{T}\right)\} H_s(h\left(\frac{\cdot}{N}\right), \lambda) + \{\psi\left(\frac{N-1}{T}\right) - \psi\left(\frac{t}{T}\right)\} \; H_N(h\left(\frac{\cdot}{N}\right), \lambda) \; .$$

We now have (again with summation by parts, c.f. Dahlhaus, 1988, Lemma 5.4)

$$| H_s(h\left(\frac{\cdot}{N}\right), \lambda)| \leq K \; L_s(\lambda) \leq K \; L_N(\lambda)$$

uniformly in $s \leq N$ which gives the result with the mean value theorem.

We remark that Lemma A.5 also holds under weaker assumptions on the data taper (e.g. if h is of bounded variation).

<u>**(A.6) Lemma.**</u> Let ψ be differentiable with bounded derivative and $t_j = S(j-1) + N/2$, $u_j = t_j / T$ with N,M,S and T as in Assumption A.1(ii). Then

$$\left|\sum_{j=1}^{M} \psi(u_j) \;\; \exp(i\lambda Sj)\right| \leq K\left(\sup_u |\psi(u)| + \sup_u |\psi'(u)| \right) L_M(S\lambda) \; .$$

Proof. Similar to the above proof.

<u>**(A.7) Lemma.**</u> Suppose h fullfills Assumption A.1(iii). Then

$$| H_N(\lambda) | \leq K \, N^{-1} \, L_N(\lambda)^2 \; .$$

Proof. Repeated summation by parts (cf. Lemma 5.4 in Dahlhaus, 1988).

(A.8) Lemma. Suppose Assumption A.1 holds. Then

$$\mathbf{E}\, J_T(\phi) = J(\phi) + o(T^{-1/2}) .$$

Proof. We have

$$\mathbf{E}\, J_T(\phi) = \frac{1}{M} \sum_{j=1}^{M} \int_{-\pi}^{\pi} \phi(u_j,\lambda)\, \frac{1}{2\pi H_{2,N}(0)} \operatorname{cum}\left(d_N(u_j,\lambda), d_N(u_j,-\lambda)\right) d\lambda \; .$$

Since

$$\operatorname{cum}\left(X_{s,T}, X_{t,T}\right) = \int_{-\pi}^{\pi} \exp\left(i\gamma\,(s-t)\right) A^{o}_{s,T}(\gamma)\, \overline{A^{o}_{t,T}(\gamma)}\, d\gamma$$

the above expression is equal to

$$\frac{1}{M} \sum_{j=1}^{M} \int\!\!\int_{-\pi}^{\pi} \phi(u_j,\lambda)\, \frac{1}{2\pi H_{2,N}(0)}\;\; H_N\, (A^{\circ}_{t_j-N/2+1+\,\bullet,T}\;(\gamma)\, h\left(\frac{\cdot}{N}\right), \lambda-\gamma)$$
$$H_N\, \overline{(A^{\circ}_{t_j-N/2+1+\,\cdot,T}\,(\gamma)}\, h\left(\frac{\bullet}{N}\right)\;, \gamma-\lambda)\; d\gamma\, d\lambda.$$

Application of Lemma A.5 and A.6 shows that this is equal to

$$\frac{1}{M} \sum_{j=1}^{M} \int\!\!\int_{-\pi}^{\pi} \phi(u_j,\lambda)\, f(u_j\,,\gamma)\;\; \frac{|H_N(\lambda-\gamma)|^2}{2\pi\, H_{2,N}(0)}\; d\gamma\, d\lambda + O\left(\frac{1}{T}\int_{-\pi}^{\pi} L_N(\lambda)^2\, d\lambda\right).$$

Let $g(u,\lambda) = \int_{-\pi}^{\pi} \phi(u,\lambda+\gamma)\, f(u,\gamma)\, d\gamma$. Since ϕ and f are both differentiable g is twice differentiable in λ with bounded second derivative (partial integration). Thus the above expression is with Lemma A.4(b) and Lemma A.7 equal to

$$\text{(A.1)} \qquad \frac{1}{M} \sum_{j=1}^{M} \int_{-\pi}^{\pi} g(u_j,\lambda) \frac{|H_N(\lambda)|^2}{2\pi\, H_{2,N}(0)}\; d\lambda \; + \; O(\frac{N}{T} \ln N)$$

$$= \frac{1}{M} \sum_{j=1}^{M} g(u_j,0) \; + \; O(\int_{-\pi}^{\pi} |\lambda|^2\, \frac{|L_N(\lambda)|^4}{N^3}\, d\lambda\,) \; + \; O(\frac{N}{T} \ln N)$$

$$= J(\phi) + O(M^{-1}) + O(N^{-2}) + O(\tfrac{N}{T} \ln N).$$

(A. 9) Lemma. Suppose Assumption A.1 holds. Then

$$T \operatorname{cov} (J_T(\phi_1), J_T(\phi_i)) = \operatorname{cov}(\xi_i,\xi_j) + o(1)$$

with ξ_i as in Theorem A.2.

Proof. We set i = 1 and j = 2.

(A.2) $$T \operatorname{cov} (J_T(\phi_1), J_T(\phi_2)) = \frac{T}{(2\pi\, M\, H_{2,N}(0))^2} \sum_{j,k=1}^{M} \iint_{-\pi}^{\pi} \phi_1 (u_j, \lambda)\, \overline{\phi_2(u_k,\mu)}$$

$$\cdot [\operatorname{cum} \big(d_N (u_j, \lambda) , d_N(u_k,-\mu)\big) \operatorname{cum} \big(d_N (u_j, -\lambda) , d_N(u_k, \mu) \big)$$

$$+ \operatorname{cum} \big(d_N (u_j, \lambda) , d_N(u_k, \mu)\big) \operatorname{cum} \big(d_N (u_j, -\lambda) , d_N(u_k , -\mu)\big)$$

$$+ \operatorname{cum} (d_N (u_j, \lambda) , d_N (u_j, -\lambda) , d_N (u_k , \mu) , d_N (u_k ,-\mu))]\, d\lambda\, d\mu .$$

We study the behaviour of the three terms separately. The first term is with similar arguments as in the proof of Lemma A.8

$$\iint_{-\pi}^{\pi} H_N \Big(A^{\circ}_{t_j - N/2 + 1 + \cdot,\, T}(\gamma_1)\, h\Big(\frac{\cdot}{N}\Big),\ \lambda - \gamma_1\Big)\, H_N \Big(\overline{A^{\circ}_{t_k - N/2 + 1 + \cdot, T}(\gamma_1)}\ h\Big(\frac{\cdot}{N}\Big), -\mu + \gamma_1\Big)$$

$$H_N \Big(A^{\circ}_{t_j - N/2 + 1 + \cdot,\, T}(\gamma_2)\, h\Big(\frac{\cdot}{N}\Big), -\lambda - \gamma_2\Big)\, H_N \Big(\overline{A^{\circ}_{t_k - N/2 + 1 + \cdot, T}(\gamma_1)}\ h\Big(\frac{\cdot}{N}\Big), \mu + \gamma_2 \Big)$$

$$\cdot \exp \{i(\gamma_1 + \gamma_2) (t_j - t_k)\}\, d\gamma_2\, d\gamma_1$$

which, by using Lemma A.5 , is equal to

(A.3) $$\iint_{-\pi}^{\pi} A\ (u_j,\gamma_1)\, A\big(u_k, -\gamma_1\big)\, A\big(u_j,\gamma_2\big)\, A\big(u_k, -\gamma_2\big) \cdot$$

$$\cdot H_N(\lambda - \gamma_1)\, H_N(\gamma_1 - \mu)\, H_N(\mu + \gamma_2)\, H_N(- \gamma_2 - \lambda) \exp\{i(\gamma_1 + \gamma_2) (t_j - t_k)\}\, d\gamma_2\, d\gamma_1$$

plus a remainder term $R_{j,k}$ with

$$\text{(A.4)}\ \left|\sum_{j,k=1}^{M} \phi_1(u_j,\lambda)\ \overline{\phi_2(u_k,\mu)}\ R_{j,k}\right|$$

$$\leq K\ M\,\frac{N}{T}\int\!\!\int_{-\pi}^{\pi} L_N\ (\lambda-\gamma_1)\, L_N(\gamma_1-\mu)\, L_N(\mu+\gamma_2)\, L_N(-\gamma_2-\lambda)\, L_M\ (S(\gamma_1+\gamma_2))\, d\gamma_2\, d\gamma_1$$

since, by Lemma A.6

$$\sum_{j=1}^{M} \phi_1(u_j,\lambda)\ \ A(u_j,\gamma_1)\ \ A(u_j,\gamma_2\ )\ \exp\{i\ S(\gamma_1+\gamma_2)\ j\} = O(L_M(S(\gamma_1+\gamma_2)).$$

From Lemma A.4 (j) follows that (A.4) is bounded by

$$K\,M\,\frac{N}{T}\ \cdot\frac{N}{S}\ \ (\ln M)\ln S \ln N\, L_N(\lambda-\mu)^2\ .$$

Integration over λ and μ gives with the constants the upper bound $K\,\frac{N}{T}$ (ln M) (ln S) (ln N) which tends to zero. We now replace $\phi_1(u_j\ ,\lambda)$ by $\phi_1(u_j\ ,\gamma_1)$ and then $\phi_2(u_k\ ,\mu)$ by $\phi_2(\ u_k\ ,\gamma_1)$. Lemma A.6 gives

$$\left|\sum_{j=1}^{M}\ (\phi_1(u_j\ ,\lambda)-\ \phi_1(u_j\ ,\gamma_1))\ A(u_j\ ,\gamma_1)\ A(u_j\ ,\gamma_2)\ \exp\big(i(\ \gamma_1+\ \gamma_2)\ t_j\big)\right|$$

$$\leq K\ |\lambda-\ \gamma_1|\ L_M\big(S(\gamma_1+\gamma_2)\big)$$

and therefore we obtain for the corresponding difference term the upper bound

$$K\ \frac{T}{M^2\,N^2}\ \int\!\!\int\!\!\int\!\!\int_{-\pi}^{\pi}\ L_N(\gamma_1-\mu)\, L_N(\mu+\gamma_2)\, L_N(-\gamma_2-\lambda)\, L_M\big(S(\gamma_1+\gamma_2)\big)^2\, d\gamma_2\, d\gamma_1\, d\lambda\, d\mu$$

$$\leq\ K\ \frac{T}{M^2\,N^2}\ \ \ln^2 N\,\frac{N\,M}{S}\ \ \ln\ S\leq K\ \frac{\ln^2 N}{N}\ \ \ln\ S\to\ 0$$

where the integration is done in the order λ, γ_2, μ . Thus, the first term of (A.2) is equal to

$$\frac{T}{\{M\,H_{2,N}(0)\}^2}\sum_{j,k=1}^{M}\iint_{-\pi}^{\pi}\phi_1(u_j,\gamma_1)\,\overline{\phi_2(u_k,\gamma_1)}\,A(u_j,\gamma_1)\,A(u_k,-\gamma_1)\,A(u_j,\gamma_2)\,A(u_k,-\gamma_2)$$

$$|H_{2,N}(\gamma_1+\gamma_2)|^2\exp\{i(\gamma_1+\gamma_2)(t_j-t_k)\}\,d\gamma_1\,d\gamma_2+o(1)\,.$$

Similarly, we now replace $A(u_j,\gamma_2)$ by $A(u_j,-\gamma_1)$ and $A(u_k,-\gamma_2)$ by $A(u_k,\gamma_1)$. Afterwards we substitute $\alpha=\gamma_1+\gamma_2$, $\gamma=\gamma_1$ and obtain with $h_i(u,\gamma)=\phi_i(u,\gamma)\,f(u,\gamma)$ for the above expression

$$\frac{T}{\{MH_{2,N}(0)\}^2}\int_{-\pi}^{\pi}\sum_{r,s=0}^{N-1}\sum_{j,k=1}^{M}h^2\left(\frac{r}{N}\right)h^2\left(\frac{s}{N}\right)h_1(u_j,\gamma)\,\overline{h_2(u_k,\gamma)}\int_{-\pi}^{\pi}\exp\{i\alpha(r-s)+i\alpha S(j-k)\}\,d\alpha d\gamma+o(1).$$

If $S=N$ this is equal to

$$\frac{2\pi\,TH_{4,N}(0)}{\{M\,H_{2,N}(0)\}^2}\int_{-\pi}^{\pi}\sum_{j=1}^{M}h_1(u_j,\gamma)\,\overline{h_2(u_j,\gamma)}\,d\gamma+o(1)=\frac{2\pi H_4}{H_2^2}\int_0^1\int_{-\pi}^{\pi}\phi_1(u,\gamma)\,\overline{\phi_2(u,\gamma)}\,f(u,\gamma)^2d\gamma du+o(M^{-1})$$

where $H_k=\int_0^1 h(u)^k\,du$. If $S\le N$ the above expression is equal to

$$\frac{2\pi T}{\{M\,H_{2,N}(0)\}^2}\int_{-\pi}^{\pi}\sum_{\substack{j,k=1\\|j-k|<\frac{N}{S}}}^{M}h_1(u_j,\gamma)\,\overline{h_2(u_k,\gamma)}\sum_{\substack{r,s=0\\r-s=S(k-j)}}^{N-1}h^2\left(\frac{r}{N}\right)h^2\left(\frac{s}{N}\right)d\gamma+o(1)\,.$$

Straightforward calculations show that this is equal to

$$2\pi\int_0^1\int_{-\pi}^{\pi}\phi_1(u,\gamma)\,\overline{\phi_2(u,\gamma)}\,f(u,\gamma)^2\,d\gamma\,du+o(1)\,.$$

With the substitution $\mu\to-\mu$ we see that the second term of (A.2) converges to the same expression with $\overline{\phi_2(u,-\gamma)}$ instead of $\overline{\phi_2(u,\gamma)}$. An analogous derivation for the third term of (A.2) leads to the result .

(A.10) Lemma. Suppose Assumption A.1 holds. Then

$$T^{\ell/2}\,\mathrm{cum}\,(J_T(\phi_1),\ \dots\ ,J_T(\phi_\ell)) = o(1)\ .$$

Proof. Let $\Pi = (-\pi,\pi]$, $\lambda = (\lambda_1,\ \dots\ ,\lambda_\ell)$

$$T^{\ell/2}\,\mathrm{cum}(J_T(\phi_1),\ \dots\ ,J_T(\phi_\ell)) = T^{\ell/2}\,\{2\pi\, M\, H_{2,N}(0)\}^{-\ell} \sum_{j_1,\ \dots\ ,j_{\ell-1}=1}^{M}$$

$$\int_{\Pi^\ell} \Big\{\prod_{\nu=1}^{\ell} \phi_\nu(u_{j_\nu},\lambda_\nu)\Big\}\ \mathrm{cum}(d_N(u_{j_1},\lambda_1)\, d_N(u_{j_1},-\lambda_1),\ \dots\ ,d_N(u_{j_\ell},\lambda_\ell)\, d_N(u_{j_\ell},-\lambda_\ell)\ \lambda^\ell(d\lambda)\ .$$

Using the product theorem for cumulants (cf. Brillinger, 1981, Theorem 2.3.2) we have to sum over all indecomposable partitions $\{P_1,\ \dots\ ,P_m\}$ with $|P_i| \geq 2$ of the scheme

$$\begin{array}{cc} a_1 & b_1 \\ \vdots & \vdots \\ a_l & b_l \end{array}$$

where a_i and b_i stand for the position of $d_N(u_{j_i},\lambda_i)$ and $d_N(u_{j_i},-\lambda_i)$ respectively. This sum will be denoted by $\sum_{ip}$. The elements of a set P_i from such a partition are assumed to be in a fixed order, so that the following definitions are reasonable. If $P_i = \{c_1,\ \dots\ ,c_k\}$ we set $\overline{P}_i := \{c_1,\ \dots\ ,c_{k-1}\}$, $\beta_{\overline{P}_i} := (\beta_{c_1},\ \dots\ ,\beta_{c_{k-1}})$ and $\beta_{c_k} = -\sum_{j=1}^{k-1}\beta_{c_j}$. Furthermore, let m be the size of the corresponding partition and $\beta := (\beta_{\overline{P}_1},\ \dots\ ,\beta_{\overline{P}_m})$. Using this notation we obtain as in the proof of Lemma A.8 (i) for the above expression

$$= T^{\ell/2}\{2\pi\, M\, H_{2,N}(0)\}^{-\ell}\ \sum_{ip} \sum_{j_1,\dots,j_\ell=1}^{M} \int_{\Pi^\ell} \Big\{\prod_{\nu=1}^{\ell} \phi_\nu(u_{j_\nu},\lambda_\nu)\Big\}$$

$$\int_{\Pi^{2\ell-m}} \Big\{\prod_{\nu=1}^{\ell} H_N\,(A^\circ_{t_{j_\nu}-N/2+1+\,\cdot\,,T}\,(\beta_{a_\nu})h\left(\frac{\cdot}{N}\right),\lambda_\nu-\beta_{a_\nu})\ H_N(A^\circ_{t_{j_\nu}-N/2+1+\,\cdot\,,T}\,(\beta_{b_\nu})h\left(\frac{\cdot}{N}\right),-\lambda_\nu-\beta_{b_\nu})\Big\}$$

$$\Big\{ \prod_{\nu=1}^{m} g_{|P_\nu|}(\beta_{\overline{P}_\nu}) \Big\} \exp(i \sum_{\nu=1}^{\ell} t_{j_\nu} (\beta_{a_\nu} + \beta_{b_\nu})) \, \lambda^{2\ell - m}(d\beta) .$$

As in Lemma A.9 we now replace successively all $H_N(A^{\circ}_{t_{j_\nu}}(\beta)\, h\left(\frac{\cdot}{N}\right), \lambda - \beta)$ by the corresponding $A(u_{j_\nu},\beta)\, H_N(\lambda - \beta)$ terms. We get for example as an upper bound for the error with Lemma A.5

$$K \frac{T^{\ell/2}}{M^\ell N^\ell} \sum_{ip} \int_{\Pi^\ell} \int_{\Pi^{2\ell-m}} M \frac{N}{T} \Big\{ \prod_{\nu=1}^{\ell} L_N(\lambda_\nu - \beta_{a_\nu})\, L_N(-\lambda_\nu - \beta_{b_\nu}) \Big\} \Big\{ \prod_{\nu=2}^{\ell} L_M(S(\beta_{a_\nu} + \beta_{b_\nu})) \Big\}$$
$$\lambda^{2\ell - m}(d\beta)\, \lambda^{\ell}(d\lambda) .$$

The special structure of a partition is expressed in the structure of the corresponding β. Every β_c, $c \in \bigcup_{k=1}^{m} \overline{P}_k$ is contained in $\prod_{\nu=1}^{\ell} L_N(\lambda_\nu - \beta_{a_\nu})\, L_N(-\lambda_\nu - \beta_{b_\nu})$ exactly twice as an argument, once with positive and once with negative sign. We therefore have $\sum_{\nu=1}^{\ell} (-\beta_{a_\nu} - \beta_{b_\nu}) = 0$ while every partial sum is different from 0 by the indecomposability of the partition.

Integration over all λ_ν and afterwards over all β (starting with β_{a_1}) gives as an upper bound

$$K \frac{T^{\ell/2}}{M^\ell N^\ell} M \frac{N}{T} (\ln N)^\ell \frac{N^\ell}{S^{\ell-1}} (\ln M)^{\ell-1} (\ln S)^{\ell-1} \le K \frac{T^{\ell/2}}{T^{\ell-1}} \frac{N}{T} (\ln N \ln M \ln S)^\ell \to 0.$$

Similarly, the resulting main term is bounded by

$$K \frac{T^{\ell/2}}{M^\ell N^\ell} \sum_{ip} \int_{\Pi^\ell} \int_{\Pi^{2\ell-m}} \Big\{ \prod_{\nu=1}^{\ell} L_N(\lambda_\nu - \beta_{a_\nu})\, L_N(-\lambda_\nu - \beta_{b_\nu})\, L_M(S(\beta_{a_\nu} + \beta_{b_\nu})) \Big\} \lambda^{2\ell-m}(d\beta)\, \lambda^{\ell}$$
$(d\lambda)$

$$\le K \frac{T^{\ell/2}}{M^\ell N^\ell} \frac{N^\ell}{S^{\ell-1}} M (\ln M \ln S \ln N)^\ell \le K \frac{T^{\ell/2}}{T^{\ell-1}} (\ln M \ln S \ln N)^\ell \to 0$$

which proves the result.

Proof of Theorem 3.6. Consistency of $\tilde{\theta}_T$ follows with the proof of Theorem 3.2 if we show that

$$\sup_\theta |L_T(\theta,\hat\mu) - L_T(\theta,\mu)| \xrightarrow{p} 0$$

i.e. if we show

$$\sup_\theta \left| \frac{1}{M} \sum_{j=1}^{M} \int_{-\pi}^{\pi} \left\{ I_N^{\hat\mu}(u_j,\lambda) - I_N^{\mu}(u_j,\lambda) \right\} \phi_\theta(u_j,\lambda)\, d\lambda \right| \xrightarrow{p} 0$$

where $\phi_\theta(u_j,\lambda) = f_\theta(u_j,\lambda)^{-1}$. This will be proved below. A Taylor expansion then gives

$$\sqrt{T}\left\{\nabla L_T(\tilde\theta_T,\hat\mu) - \nabla L_T(\theta_0,\hat\mu)\right\} = \nabla^2 L_T(\bar\theta,\hat\mu)\sqrt{T}\,(\tilde\theta_T - \theta_0)$$

with $|\bar\theta - \theta_0| \le |\tilde\theta_T - \theta_0|$. As in the proof of Theorem 3.3 we obtain $\sqrt{T}\,\nabla L_T(\tilde\theta_T,\hat\mu) \xrightarrow{p} 0$. In the proof of Theorem 3.3 we showed that

$$\sqrt{T}\,\nabla L_T(\theta_0,\mu) + \Gamma\sqrt{T}\,(\hat\theta_T - \theta_0) \xrightarrow{p} 0$$

i.e. the result follows if we prove that

$$\sqrt{T}\,\nabla L_T(\theta_0,\hat\mu) - \sqrt{T}\,\nabla L_T(\theta_0,\mu) \xrightarrow{p} 0$$

and

$$\nabla^2 L_T(\bar\theta,\hat\mu) \xrightarrow{p} \Gamma .$$

Together with the proof of Theorem 3.3 the result therefore follows if we show that

$$\text{(A.5)} \qquad \sqrt{T}\frac{1}{M} \sum_{j=1}^{M} \int_{-\pi}^{\pi} \left\{ I_N^{\hat\mu}(u_j,\lambda) - I_N^{\mu}(u_j,\lambda) \right\} \phi_{\theta_0}(u_j,\lambda)\, d\lambda \xrightarrow{p} 0$$

for $\phi_\theta(u,\lambda) = \nabla f_\theta(u,\lambda)^{-1}$ and

$$\text{(A.6)} \qquad \sup_\theta \left|\frac{1}{M} \sum_{j=1}^{M} \int_{-\pi}^{\pi} \left\{ I_N^{\hat\mu}(u_j,\lambda) - I_N^{\mu}(u_j,\lambda) \right\} \phi_\theta(u_j,\lambda)\, d\lambda \right| \xrightarrow{p} 0$$

for $\phi_\theta(u,\lambda) = f_\theta(u,\lambda)^{-1}$ and $\phi_\theta(u,\lambda) = \nabla^2 f_\theta(u,\lambda)^{-1}$. The last expression is equal to

$$\sup_\theta \left|\frac{1}{M} \sum_{j=1}^{M} \int_{-\pi}^{\pi} \phi_\theta(u_j,\lambda) \left\{2\pi H_{2,N}(0)\right\}^{-1}\right.$$

$$\text{(A.7)} \qquad \left\{ d_N^{X-\mu}(u_j,\lambda)\, d_N^{\mu-\hat\mu}(u_j,-\lambda) + d_N^{\mu-\hat\mu}(u_j,\lambda)\, d_N^{X-\mu}(u_j,-\lambda) + d_N^{\mu-\hat\mu}(u_j,\lambda)\, d_N^{\mu-\hat\mu}(u_j,-\lambda) \right\} \, d\lambda|$$

which by means of the Cauchy-Schwarz inequality is with

$$\delta_T := \frac{1}{M} \sum_{j=1}^{M} \int_{-\pi}^{\pi} \{2\pi H_{2,N}(0)\}^{-1} \left| d_N^{\mu-\hat\mu}(u_j,\lambda) \right|^2 d\lambda$$

bounded by

$$\sup_{\theta,u,\lambda} |\phi_\theta(u,\lambda)| \left\{ 2 \left(\frac{1}{M} \sum_{j=1}^{M} \int_{-\pi}^{\pi} I_N^{\mu}(u_j,\lambda)\, d\lambda \right)^{1/2} \cdot \delta_T^{1/2} + \delta_T \right\}.$$

Since $\frac{1}{M} \sum_{j=1}^{M} \int_{-\pi}^{\pi} I_N^{\mu}(u_j,\lambda)\, d\lambda$ is bounded in probability (Theorem A.2) and

$$\delta_T = \frac{1}{M} \sum_{j=1}^{M} H_{2,N}(0)^{-1} \sum_{s=1}^{N} \left\{ \mu\left(\frac{t_j - N/2 + s}{T} \right) - \hat\mu\left(\frac{t_j - N/2 + s}{T} \right) \right\}^2 = o_P\left(\frac{N}{T} \right)$$

(A.6) is proved. To prove (A.5) we note that $\sqrt{T}\delta_T \to 0$. Since $\sqrt{T}\delta_T^{1/2} \not\to 0$ we need a better estimate for the first and second term of (A.7). Summation by parts gives with $c_T := \sqrt{T}\{2\pi\, M\, H_{2,N}(0)\}^{-1}$, $\bar H_{t,N}(\lambda) := \sum_{s=0}^{t-1} h\left(\frac{S}{N}\right) \exp(-i\lambda s)$ and $\bar t_j = t_j - N/2$

$$\sqrt{T}\, \frac{1}{M} \sum_{j=1}^{M} \int_{-\pi}^{\pi} \phi_{\theta_o}(u_j,\lambda)\, \{2\pi H_{2,N}(0)\}^{-1} d_N^{X-\mu}(u_j,\lambda)\, d_N^{\mu-\hat\mu}(u_j,-\lambda)\, d\lambda$$

$$= c_T \sum_{j=1}^{M} \sum_{t=0}^{N-1} \left\{ \mu\left(\frac{\bar t_j + t + 1}{T} \right) - \hat\mu\left(\frac{\bar t_j + t + 1}{T} \right) \right\} \cdot \int_{-\pi}^{\pi} \phi_{\theta_o}(u_j,\lambda)\, d_N^{X-\mu}(u_j,\lambda) \;\; \{\bar H_{t+1,N}(-\lambda) - \bar H_{t,N}(-\lambda)\} \;\; d\lambda$$

$$= -\, c_T \sum_{j=1}^{M} \sum_{t=0}^{N-1} \left[\left\{ \mu\left(\frac{\bar t_j + t + 1}{T} \right) - \hat\mu\left(\frac{\bar t_j + t + 1}{T} \right) \right\} - \left\{ \mu\left(\frac{\bar t_j + t}{T} \right) - \hat\mu\left(\frac{\bar t_j + t}{T} \right) \right\} \right] \cdot$$

$$\cdot \int_{-\pi}^{\pi} \phi_{\theta_o}(u_j,\lambda)\, d_N^{X-\mu}(u_j\lambda) \bar H_{t,N}(-\lambda)\, d\lambda$$

$$+\, c_T \sum_{j=1}^{M} \left\{ \mu\left(\frac{\bar t_j + N}{T} \right) - \hat\mu\left(\frac{\bar t_j + N}{T} \right) \right\} \cdot \int_{-\pi}^{\pi} \phi_{\theta_o}(u_j,\lambda)\, d_N^{X-\mu}(u_j,\lambda) \bar H_{N,N}(-\lambda)\, d\lambda$$

Summation by parts implies $\bar H_{t,N}(-\lambda) \le K L_N(\lambda)$ uniformly in t. We now can prove by similar methods as in the proof of Lemma A.9 that

$$\mathrm{var} \int_{-\pi}^{\pi} \phi_{\theta_o}(u_j,\lambda)\, d_N^{X-\mu}(u_j,\lambda) \overline{H}_{t,N}(-\lambda)\, d\lambda = O(N)$$

uniformly in u_j and t. Since $E\, d_N^{X-\mu}(u_j,\lambda) = 0$ the whole expression tends to zero in probability. The second term of (A.7) is treated in the same way which proves the result.

Acknowledgement. I am grateful to L. Giraitis, R. von Sachs and two anonymous referees whose comments helped to improve the paper. The computations were done together with M. Diller and E. Ioannidis on the basis of the framework "Random & Template" by G. Sawitzki. I am grateful to them for their excellent work.

References.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Automatic control, AC*-19, 716 - 722.

Azencott, R. and Dacunha-Castelle, D. (1986). Series of Irregular Observations. Springer-Verlag, New York.

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

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

Dahlhaus, R. (1983). Spectral analysis with tapered data. *J. Time Ser. Anal.*, 4, 163 - 175.

Dahlhaus, R. (1985). On a spectral density estimate obtained by averaging periodograms. *J. Appl. Probab.*, 22, 598 - 610.

Dahlhaus, R. (1988). Small sample effects in time series analysis: A new asymptotic theory and a new estimate. *Ann. Statist.*, vol. 16, No. 2, 808 - 841.

Dahlhaus, R. (1993). On the Kullback Leibler information divergence for locally stationary processes. In preparation.

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

Findley, D. (1985). On the unbiasedness property of AIC for exact or approximating linear stochastic time series models. *J. Time Ser. Anal.*, vol. 6, No. 4, 229 - 252.

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

Grenier, Y. (1983). Time-dependent ARMA modeling of nonstationary signals. *IEEE Trans. Acoust., Speech, Signal Process.*, ASSP-31, 899-911.

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

Hannan, E. J. (1973). Multiple Time Series. 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.

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

Mélard, G. and Herteleer-de Schutter, A. (1989) Contributions to evolutionary spectral theory. *J. Time Ser. Anal.*, vol. 10, No. 1, 41 - 63.

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

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

Subba Rao, T. (1970). The fitting of non-stationary time series models with time-dependent parameters. *Journ. Roy. Statist. Soc.B* 32, 312 - 322.

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

Young, P.C. (1994). Data-based mechanistic modelling and the rainfall-flow nonlinearity. Environmetrics, to appear.