The Annals of Statistics
1988, Vol. 16, No. 2, 808–841

SMALL SAMPLE EFFECTS IN TIME SERIES ANALYSIS: A NEW ASYMPTOTIC THEORY AND A NEW ESTIMATE[1]

By Rainer Dahlhaus

Universität Essen

To estimate the parameters of a stationary process, Whittle (1953) introduced an approximation to the Gaussian likelihood function. Although the Whittle estimate is asymptotically efficient, the small sample behavior may be poor if the spectrum of the process contains peaks. We introduce a mathematical model that covers such small sample effects. We prove that the exact maximum likelihood estimate is still optimal in this model, whereas the Whittle estimate and the conditional likelihood estimate are not. Furthermore, we introduce tapered Whittle estimates and prove that these estimates have the same optimality properties as exact maximum likelihood estimates.

1. Introduction. Consider a real-valued stationary process $X(t)$, $t \in \mathbb{Z}$ with mean 0 and spectral density $f_\theta(\lambda)$, $\lambda \in \Pi := (-\pi, \pi]$, depending on a vector of unknown parameters $\theta \in \Theta \subset \mathbb{R}^p$. If θ has to be estimated from a sample $\mathbf{X}_T' = (X_0, \ldots, X_{T-1})$ a natural approach is to maximize the likelihood function, or to minimize $-1/T$ log likelihood function, which, in the Gaussian case, takes the form

$$(1.1)\qquad \mathscr{L}_T(\theta) := \frac{1}{2}\log(2\pi) + \frac{1}{2T}\log\det B_T(f_\theta) + \frac{1}{2T}\mathbf{X}_T' B_T(f_\theta)^{-1}\mathbf{X}_T,$$

where $B_T(f)$ is the Toeplitz matrix of f,

$$(1.2)\qquad B_T(f) := \left\{\int_\Pi f(\alpha)\exp(i\alpha(r-s))\,d\alpha\right\}_{r,s=0,\ldots,T-1}.$$

Unfortunately, the preceding function, especially $B_T(f_\theta)^{-1}$, is difficult to calculate. One way to handle these difficulties is to approximate $B_T(f_\theta)^{-1}$ by $B_T(\{4\pi^2 f_\theta\}^{-1})$ [cf. Shaman (1975, 1976)]. Together with the Szegö identity [see Grenander and Szegö (1958), Section 5.2],

$$\lim_{T\to\infty} \det B_T(f)^{1/T} = 2\pi\exp\left(\frac{1}{2\pi}\int_\Pi \log f(\lambda)\,d\lambda\right),$$

this leads to the likelihood approximation suggested by Whittle (1953, 1954). We will call this approximation of $\mathscr{L}_T(\theta)$ the Whittle function. It is

$$(1.3)\qquad \mathscr{L}_T^W(\theta) = \frac{1}{4\pi}\int_\Pi\left\{\log 4\pi^2 f_\theta(\lambda) + \frac{I_T(\lambda)}{f_\theta(\lambda)}\right\}d\lambda, \qquad \Pi = (-\pi, \pi],$$

Received April 1986; revised August 1987.

[1]Work partially supported by the Deutsche Forschungsgemeinschaft.

AMS 1980 *subject classifications.* Primary 62M15; secondary 62F10.

Key words and phrases. Small sample effects, time series, data tapers, Whittle estimates, maximum likelihood estimates.

where $I_T(\lambda)$ is the periodogram,

$$I_T(\lambda) = \frac{1}{2\pi T}\left|\sum_{t=0}^{T-1} X_t \exp(-i\lambda t)\right|^2.$$

A treatise on Whittle estimates may be found in Dzhaparidze and Yaglom (1983), and the approximation of $\mathscr{L}_T(\theta)$ by $\mathscr{L}_T^W(\theta)$ is studied in Coursol and Dacunha-Castelle (1982). Let us denote by $\tilde{\theta}_T$ the maximum likelihood estimate and by $\hat{\theta}_T$ the Whittle estimate, i.e., the estimates obtained by minimizing $\mathscr{L}_T(\theta)$ and $\mathscr{L}_T^W(\theta)$, respectively.

Asymptotically, both estimates are Fisher efficient and therefore regarded as equivalent [cf. Dzhaparidze and Yaglom (1983), Theorem 5.5]. However, in the small sample situation the Whittle estimate may be bad. For example, for autoregressive processes the Whittle estimate is identical to the Yule–Walker estimate [cf. Dzhaparidze and Yaglom (1983), Example 5.1], and it is known [cf. Priestley (1981), page 351] that Yule–Walker estimates are rather bad for short time series or if a root of the corresponding characteristic equation is close to the unit circle. Although this small sample behavior of the Whittle estimate is known from Monte Carlo studies, it has never been described theoretically. Furthermore, the question arises whether the Whittle estimate can be improved. Both problems are discussed in this paper.

Minimizing the Whittle function $\mathscr{L}_T^W(\theta)$ may also be interpreted as a minimization of the information divergence between $f_\theta(\lambda)$ and $I_T(\lambda)$ [cf. Parzen (1983), Section 3], i.e., as the search for that function $f_\theta(\lambda)$ that approximates the nonparametric estimate $I_T(\lambda)$ best. In order to improve the estimate, we therefore could try to replace $I_T(\lambda)$ in (1.3) by a better nonparametric estimate. Apart from the nonconsistency, which is of no importance for the Whittle estimate, the periodogram has another important disadvantage, which is called the leakage effect.

Consider the bias of the periodogram. We obtain

$$EI_T(\lambda) = \int_\Pi f(\lambda - \alpha) K_T(\alpha)\, d\alpha, \tag{1.4}$$

where

$$K_T(\alpha) = (2\pi T)^{-1} \sin^2(T\alpha/2)/\sin^2(\alpha/2)$$

is the Fejér kernel. The Fejér kernel has side peaks of magnitude $O(T^{-1})$ at frequencies $\alpha = 2\pi(s + 1/2)/T$ with $0 < c \leq \alpha \leq \pi$. If in the preceding convolution such a side peak is multiplied with a strong peak of f this may result in an expectation that is too large, i.e., the spectrum is overestimated. The result is that other lower peaks are superimposed and possibly not discovered. A nice example of the leakage effect may be found in Bloomfield (1976), Sections 5.2 and 5.3, who analyzed the variable star data. Due to the fact that $\mathscr{L}_T^W(\theta)$ is the information divergence between $I_T(\lambda)$ and $f_\theta(\lambda)$ the leakage effect seems to transfer to the parameter estimate $\hat{\theta}_T$.

To avoid the leakage effect, we suggest the following estimation procedure adapted from nonparametric spectral density estimation. Instead of the ordinary periodogram we use a tapered version of the periodogram, i.e., we multiply the process X_t with a data taper $h_{t,T}$, $t = 0, \dots, T-1$. We therefore define

$$I_T(\lambda) = \{2\pi H_{2,T}\}^{-1} \left| \sum_{t=0}^{T-1} h_{t,T} X_t^{(T)} \exp(-i\lambda t) \right|^2,$$

where

$$H_{k,T} = \sum_{t=0}^{T-1} h_{t,T}^k, \qquad k \in \mathbb{N}_0,$$

and replace the ordinary periodogram in (1.3) by this tapered version. The data taper normally has a maximum at $t = [T/2]$ and decreases as t tends to 0 or $T-1$. As in (1.4), the expectation of the tapered periodogram is the convolution of the true spectral density with a kernel, but the tapered kernel has a faster rate of convergence to 0 at frequencies $\alpha \not\equiv 0 \bmod 2\pi$ (cp. Lemma 5.4). The use of data tapers in nonparametric time series was suggested by Tukey (1967).

To demonstrate the advantages of data tapers and to motivate the theoretical results of this paper, we now present briefly a simulation example for autoregressive processes where the Whittle estimate is identical to the Yule–Walker estimate [a similar study was presented in Dahlhaus (1984)].

$T = 256$ Gaussian observations were generated of an AR(14) process with characteristic roots $z_j = q_j e^{i\lambda_j}$ and $\bar{z}_j$, where

$$\begin{aligned}
q_1 &= 0.95, & \lambda_1 &= 0.5,\\
q_2 &= 0.95, & \lambda_2 &= 1.0,\\
q_3 &= 0.99, & \lambda_3 &= 1.5,\\
q_4 &= 0.99, & \lambda_4 &= 1.5,\\
q_5 &= 0.95, & \lambda_5 &= 2.0,\\
q_6 &= 0.95, & \lambda_6 &= 2.5,\\
q_7 &= 0.95, & \lambda_7 &= 2.5.
\end{aligned}$$

In Figure 1 we see the theoretical spectral density (dark line) the classical Yule–Walker spectral estimate without data taper (dashed line) and the Yule–Walker spectral estimate with the Tukey–Hanning taper with $\rho = 0.8$ (Example 5.2). In Table 1 we have listed the values of the estimation for different data tapers (see Examples 5.2 and 5.3) and in the last column for the popular Burg algorithm [the complete algorithm is, e.g., described in Haykin and Kesler (1979); a computer program may be found in Ulrich and Bishop (1975)]. The last line contains the values of the weighted sum of squares for the parameter estimates $T(\hat{a} - a)\Sigma(\hat{a} - a)$, where Σ is the covariance matrix of the process. We see that there is considerable leakage without taper, and marked improvement when a taper is applied. Obviously, the use of data tapers in the Yule–Walker equations solves the classical problem of roots close to the unit circle. The tapered estimate turned out to be as good as the Burg algorithm (cf. the more detailed discussion in Section 8).

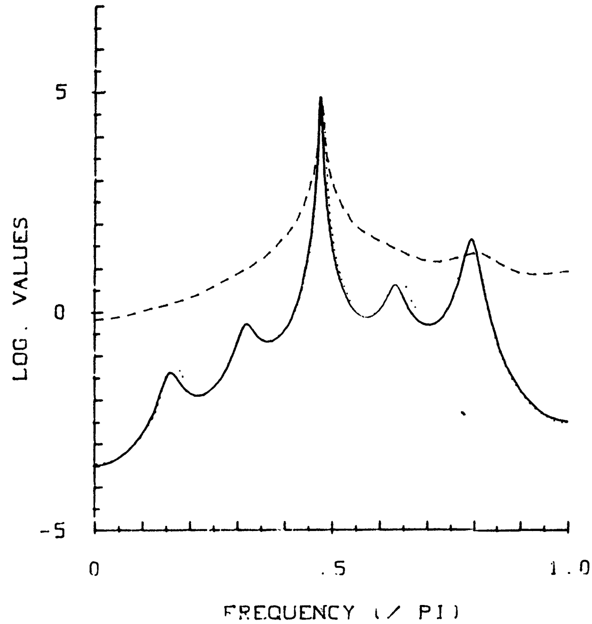

FIG. 1. *Yule–Walker estimates for an* AR(14) *process.*

TABLE 1
Exact and estimated coefficients of an AR(14) *process*

	True coefficient	Nontapered	Polynomial taper $k = 2$ $\rho = 0.3$	Polynomial taper $k = 2$ $\rho = 0.8$	Polynomial taper $k = 2$ $\rho = 1.0$	Tukey–Hanning $\rho = 1.0$	Burg algorithm
a_0	1.00			Not estimated			
a_1	0.86	0.49	0.83	0.83	0.83	0.83	0.96
a_2	2.28	1.22	2.52	2.35	2.35	2.35	2.35
a_3	1.60	0.72	1.78	1.56	1.56	1.56	1.79
a_4	2.68	0.41	3.11	2.87	2.89	2.90	2.82
a_5	1.20	0.14	1.47	1.26	1.28	1.29	1.51
a_6	2.56	0.17	2.91	2.86	2.91	2.92	2.81
a_7	1.01	0.01	1.18	1.16	1.19	1.19	1.43
a_8	2.31	0.04	2.68	2.71	2.75	2.76	2.67
a_9	0.97	0.05	1.28	1.15	1.15	1.14	1.40
a_{10}	1.98	0.02	2.29	2.40	2.42	2.43	2.39
a_{11}	1.04	0.08	1.17	1.25	1.26	1.26	1.45
a_{12}	1.41	0.00	1.41	1.70	1.73	1.74	1.73
a_{13}	0.56	0.04	0.50	0.68	0.69	0.69	0.77
a_{14}	0.58	0.01	0.46	0.67	0.69	0.69	0.68
σ^2	1.00	80.97	2.00	1.17	1.08	1.06	1.01
$T\|a - \alpha\|_\Sigma^2$		145.39	26.76	3.59	4.55	4.83	7.89

The question arises how the previously demonstrated advantage can be described theoretically. In Dahlhaus (1984) we have proved that tapered Whittle estimates are asymptotically normal with the same increase of the asymptotic variance as in the nonparametric case [cf. Brillinger (1981), (5.6.21)]. Thus, asymptotic results do not describe the effects observed in the small sample situation. On the other hand, due to extreme technical complexity, exact calculations for finite T are only possible in very special situations, e.g., for AR(1) processes.

To handle these difficulties, we introduce a special asymptotic model that covers in some sense the small sample situation. We require that the estimates behave well for an increasing number of processes if the sample size increases. Consider again (1.4). Since the magnitude of the side peaks of $K_T(\alpha)$ for $0 < c \leq \alpha \leq p$ is of order T^{-1}, the leakage effect disappears asymptotically. To preserve this small sample effect, we therefore let the magnitude of the main peak of the spectrum increase with T and study the properties of the estimates concerning the lower peaks of the spectrum. The spectral density of an AR(p) process is of the form

$$f(\lambda) = \frac{\sigma^2}{2\pi} \prod_{j=1}^{p} |1 - z_j e^{i\lambda}|^{-2},$$

where z_j are the roots of the process. The simulation indicates that it is quite easy to estimate the strongest peak, i.e., the roots closest to the unit circle. Therefore, it is justified to assume that these roots are known, whereas the other roots have to be estimated. Because of the preceding product form of f, we therefore assume that the spectral density is of the form $f_{\theta_0} = g_T f^*_{\theta_0}$ with a "strong peak" part g_T (whose peaks may increase with T) and a part $f^*_{\theta_0}$, which is independent of T and whose parameter θ_0 has to be estimated. This estimation model is introduced in Section 2.

In Section 3 we study the behavior of the exact (Gaussian) maximum likelihood estimate for double indexed sequences of processes $X_t^{(T)}$ with spectral density $f_{\theta_0} = g_T f^*_{\theta_0}$ and prove that the estimate is still asymptotically efficient in the introduced model (Theorem 3.3).

In Section 4 we consider the conditional likelihood estimate and prove that this estimate may even be inconsistent in the introduced model.

In Section 5 we introduce data tapers, classify them and prove a fundamental inequality for the Fourier transform of data tapers.

In Section 6 we study the behavior of tapered Whittle estimates for the same sequences of processes $X_t^{(T)}$ and prove that a suitable choice of the data taper also leads to an efficient estimate of the parameter θ_0 (Theorem 6.3).

In Section 7 we consider the classical nontapered Whittle estimate. We prove that the estimate may even be inconsistent in the assumed model (Theorem 7.1). Furthermore, we prove that the same holds if strong troughs (values close to 0—caused, e.g., by MA roots close to the unit circle) are present in the spectrum (Theorem 7.2). This effect is called the trough effect. It is of special importance because of the widespread practice of taking differences to remove trends in the data. Furthermore, we prove that data tapers may reduce not only the bias but also the variance and moreover the cumulants of all order of the estimate (Theorem 7.3).

In Appendix 1 we prove a central limit theorem for time-series statistics and in Appendix 2 we state some results on Toeplitz matrices.

The approach of this paper to handle small sample effects is new. Data tapers for Whittle estimates and Yule–Walker estimates were suggested in Dahlhaus (1984). Their importance goes far beyond the scope of this paper. Their use improves, for example, the classical order estimation criteria and the properties of Whittle estimates as starting values in a Newton–Raphson iteration to determine exact maximum likelihood estimates. Furthermore, the use of data tapers removes the so-called edge effect of parameter estimates for spatial processes [Dahlhaus and Künsch (1987)]. Further applications will be considered in forthcoming papers. Pukkila (1979) considered data tapers for similar estimates and made some numerical calculations for an AR(2) process.

A key role in our calculations 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

$$L_T^*(\alpha) := T, \qquad |\alpha| \le 1/T,$$
$$:= 1/|\alpha|, \qquad 1/T < |\alpha| \le \pi.$$

The function $L_T(\alpha)$ is used to describe the properties of data tapers, to describe spectral densities with strong peaks or troughs and as a tool for handling the cumulants of time-series statistics. For the latter purpose it was introduced and discussed in Dahlhaus (1983). It allows the treatment of the cumulants of time-series statistics under extremely weak conditions on the cumulant spectra (e.g., integrability conditions), and is therefore an important tool in difficult situations (e.g., if processes are long-range dependent). The properties of L_T needed in this paper are summarized in Lemma A1.1.

We further use cumulants and cumulant spectra of stationary processes. For the definitions and the basic properties we refer to Brillinger (1981), Section 2.3, especially Theorem 2.3.2.

2. The estimation model. In this section we introduce the estimation model. As mentioned previously we have T observations $X_0^{(T)}, \dots, X_{T-1}^{(T)}$ of a stationary process $X_t^{(T)}$ with second-order spectrum $f_{2,\theta_0,T}(\gamma) = g_{2,T}(\gamma) f_{2,\theta_0}^*(\gamma)$, where f_{2,θ_0}^* is bounded and independent of T, whereas $g_{2,T}$ is a known function with strong peaks or troughs whose magnitude may increase with T. This model may also be interpreted in the following way: Suppose we are interested in the process Y_{t,θ_0} with spectral density f_{2,θ_0}^*, but we only observe a filtered version $X_{t,\theta_0}^{(T)}$, where the squared transfer function of the linear filter is $g_{2,T}(\lambda)$. With a view to the discussion in the Introduction, we include all $g_{2,T}$, which are the spectral density of an ARMA process whose roots are up to $1/T$ close to the unit circle (cf. Theorem 2.4).

We now set down the assumptions on the parameter set Θ and on $f_{2,\theta}^*$. The assumptions on $g_{2,T}$ are stated in Definition 2.2.

ASSUMPTION 2.1.

(a) Let Θ be a compact subset of $\mathbb{R}^P$ such that $Y_{t,\theta}$, $\theta \in \Theta$, is a family of strictly stationary processes with spectral densities $f_{2,\theta}^*$ (uniformly bounded from above and below) and fourth-order spectra $f_{4,\theta}^*$. Suppose that $f_{2,\theta}^*(\lambda)$ and $\partial/\partial\theta_i\, f_{2,\theta}^*(\lambda)$ are continuous on $\Theta \times \Pi$ and $\theta_1 \neq \theta_2$ implies $f_{2,\theta_1}^* \neq f_{2,\theta_2}^*$ on a set of positive Lebesgue measure. (Note that Y_{t,θ_0} is *not* the observed process.)
(b) Suppose, in addition to (a), that $\partial^2/\partial\theta_i\,\partial\theta_j\, f_{2,\theta}^*(\lambda)$ is continuous on $\Theta \times \Pi$.
(c) Let $\theta_0 \in \operatorname{Int}\Theta$, $0 < c_1 \le f_{2,\theta_0}^*(\lambda)$ and $f_{2,\theta_0}^* \in \operatorname{Lip}_{\kappa_0}$ with $\kappa_0 \in (0,1]$.
(d) Let $f_{4,\theta_0}^* \in \operatorname{Lip}_{\kappa_0}$, i.e., $|f_{4,\theta}^*(\alpha_1,\alpha_2,\alpha_3) - f_{4,\theta}^*(\beta_1,\beta_2,\beta_3)| \le K\sum_{j=1}^4 |\alpha_i - \beta_i|^{\kappa_0}$, where $\alpha_4 = -\alpha_1 - \alpha_2 - \alpha_3$, $\beta_4 = -\beta_1 - \beta_2 - \beta_3$.
(e) Let $s_1, s_2 \in \mathbb{N}_0$, $C > 1$ and $X_t^{(T)} \in \mathscr{X}(f_{2,\theta_0}^*, f_{4,\theta_0}^*, C, T, s_1, s_2)$ (defined in Definition 2.2) be a sequence of stationary processes (observed).
(f) Suppose that $\tilde{\theta}_T$ is the maximum likelihood estimate [obtained by minimizing $\mathscr{L}_T(\theta)$], and that $\tilde{\theta}_T$ exists uniquely and lies in $\operatorname{Int}\Theta$.
(g) Suppose that $\hat{\theta}_T$ is the Whittle estimate [obtained by minimizing $\mathscr{L}_T^w(\theta)$] and that $\hat{\theta}_T$ exists uniquely and lies in $\operatorname{Int}\Theta$.

We now define the set $\mathscr{X}(f_{2,\theta_0}^*, f_{4,\theta_0}^*, C, T, s_1, s_2)$ from part (e) of the preceding assumption, i.e., we set down the conditions on $g_{2,T}$. For $r_1, r_2 \in \mathbb{N}_0$, $r_1 + r_2 \ge 1$, $\lambda \in \Pi^{r_1}$, $\mu \in \Pi^{r_2}$ and $s_1, s_2 \in \mathbb{N}_0$, we define

$$b(T,\alpha,\beta,\lambda,\mu,s_1,s_2) := \sum_{\substack{j_i=1\\ i=1,2}}^{r_i} \sum_{\substack{k_i=0\\ k_1+k_2\ge 1}}^{s_i} |\beta|^{k_1+k_2} L_T(\lambda_{j_1} + \alpha + \beta)^{k_1} L_T(\mu_{j_2} + \alpha)^{k_2}, \quad \text{if } s_1 + s_2 \ge 1,$$

and

$$b(T,\alpha,\beta,\lambda,\mu,s_1,s_2) := 0, \quad \text{if } s_1 + s_2 = 0.$$

DEFINITION 2.2. Let $C > 1$, $s_1, s_2 \in \mathbb{N}_0$ and $f_2^*(\gamma)$ with $0 < C^{-2} \le f_2^* \le C^2$ and $f_4^*(\gamma_1,\gamma_2,\gamma_3) \le C^4$ be second- and fourth-order spectra. We define by $\mathscr{X}(f_2^*, f_4^*, C, T, s_1, s_2)$ the set of all processes

$$X_t = \sum_{s=-\infty}^{\infty} a_s Y_{t-s},$$

where Y_t is a stationary process with mean 0 whose moments of all order exist and whose kth-order cumulant spectrum f_k^* fulfills $f_k^* \le C^k$ for all $\gamma \in \mathbb{R}^{k-1}$ and all $k > 2$, and the transfer function $A(\lambda) := \sum_{s=-\infty}^{\infty} a_s \exp(i\lambda s)$ fulfills the following condition:

There exist $r_1, r_2 \in \mathbb{N}_0$, $r_1 + r_2 \ge 1$ and $\lambda = (\lambda_1, \dots, \lambda_{r_1}) \in \Pi^{r_1}$, $\mu = (\mu_1, \dots, \mu_{r_{12}}) \in \Pi^{r_2}$ such that

$$\left| \frac{A(\alpha+\beta)}{A(\alpha)} - 1 \right| \le Cb(T,\alpha,\beta,\lambda,\mu,s_1,s_2), \tag{2.1}$$

where r_1, r_2, λ and μ may depend on T. In this case we assume that $r_1(T)$ and $r_2(T)$ are bounded in T.

Our $g_{2,T}$ from the preceding now is equal to $|A_T(\alpha)|^2$, where $A_T(\alpha)$ fulfills (2.1) for all T. $g_{2,T}$ thus may have peaks at all λ_i of multiplicity less equal s_1 and troughs at all μ_i of multiplicity less equal s_2. This is made precise in Theorem 2.4, where $g_{2,T}$ is an ARMA spectral density.

REMARK 2.3. When dealing with parametric estimation the correct point of view clearly is to regard the parametric model only as an approximation to a (possibly nonparametric) spectral density. In this case the true spectral density is not in the class of models considered. All results of this paper remain valid in this case if one assumes that the true spectral density of the process $X_t^{(T)}$ is of the form $g_{2,T} f_2^*$ with a sufficiently smooth f_2^*. The limit theorems 3.3 and 6.2 then take a similar form as described in Findley (1985), (2.11), where also further references can be found. In order not to complicate the already technical presentation of this paper, we omit this generalization.

THEOREM 2.4. *Let $C^* > 1$, $\delta > 0$, $s_1, s_2 \in \mathbb{N}_0$ and Y_t be a stationary process with mean 0 and kth-order cumulant spectrum f_k^*, where $0 < C^{*-2} \le f_2^* \le C^{*2}$ and $f_k^* \le C^{*k}$ for all $k > 2$. Let further $X_t^{(T)}$ for all $T \in \mathbb{N}$ be defined by*

$$\sum_{j=0}^{p} a_{jT} X_{t-j}^{(T)} = \sum_{j=0}^{q} b_{jT} Y_{t-j},$$

where $\sum_{j=0}^{p} a_{jT} e^{i\alpha j} = \prod_{j=1}^{p}(1 - q_{jT} e^{i\lambda_{jT}} e^{i\alpha})$ with $0 < q_j \le 1 - 1/T$ and $\#\{\lambda_{jT}: |\lambda_{jT} - \nu| \le \delta\} \le s_1$ for all $\nu \in \Pi$, and $\sum_{j=0}^{q} b_{jT} e^{i\alpha j} = \prod_{j=1}^{q}(1 - q'_{jT} e^{i\mu_{jT}} e^{i\alpha})$ with $0 < q'_{jT} \le 1 - 1/T$ and $\#\{\mu_{jT}: |\mu_{jT} - \nu| \le \delta\} \le s_2$ for all $\nu \in \Pi$. Then there exists a C independent of T such that

$$X_t^{(T)} \in \mathscr{X}(f_2^*, f_4^*, C, T, s_1, s_2),$$

for all $T \in \mathbb{N}$.

PROOF. $X_t^{(T)}$ can be represented as $X_t^{(T)} = \sum_{s=0}^{\infty} c_{sT} Y_{t-s}$ with

$$A_T(\alpha) = \sum_{s=0}^{\infty} c_{sT} e^{i\alpha s} = \frac{\prod_{k=1}^{q}\left(1 - q'_{kT} e^{i(\mu_{kT}+\alpha)}\right)}{\prod_{k=1}^{p}\left(1 - q_{kT} e^{i(\lambda_{kT}+\alpha)}\right)}.$$

Using the relations

$$\left\{\prod_{j=1}^{n} x_j\right\} - 1 = \sum_{\substack{M \subset \{1,\dots,n\} \\ M \ne \varnothing}} \prod_{j \in M} (x_j - 1), \quad \text{for } x_j \in \mathbb{R}, \tag{2.2}$$

and

$$\left|\frac{1 - pe^{i(\alpha+\beta)}}{1 - pe^{i\alpha}} - 1\right| \le 4|\beta| L_{1/(1-p)}(\alpha), \tag{2.3}$$

$$\text{for all } \alpha, \beta \in \mathbb{R} \text{ and } 0 < p < 1,$$

we obtain

$$\left|\frac{A_T(\alpha+\beta)}{A_T(\alpha)} - 1\right| \le B_1 B_2 + B_1 + B_2,$$

with

$$B_1 = \sum_{\substack{M \subset \{1,\dots,p\} \\ M \neq \varnothing}} \prod_{j \in M} \left\{ 4|\beta| L_T(\lambda_{jT} + \alpha + \beta) \right\}$$

and

$$B_2 = \sum_{\substack{N \subset \{1,\dots,q\} \\ N \neq \varnothing}} \prod_{j \in N} \left\{ 4|\beta| L_T(\mu_{jT} + \alpha) \right\}.$$

By using Lemma A1.1(f) and the relation $x^k y^l \leq x^{k+l} + y^{k+l}$, $x, y > 0$, we obtain

$$|B_1| \leq K \sum_{j_1=1}^{r_1} \sum_{k_1=1}^{s_1} |\beta|^{k_1} L_T\left(\lambda_{j_1 T} + \alpha + \beta\right)^{k_1}$$

and

$$|B_2| \leq K \sum_{j_2=1}^{r_2} \sum_{k_2=1}^{s_2} |\beta|^{k_2} L_T(\mu_{j_2 T} + \alpha)^{k_2},$$

which implies the result. □

In the following sections we consider different parameter estimates θ_T of θ_0 based on observations $X_0^{(T)}, \dots, X_{T-1}^{(T)}$ with spectral density $g_{2,T} f_{2,\theta_0}^*$, and investigate the validity of a central limit theorem for θ_T. If $g_{2,T}$ does not depend on T this is the ordinary CLT. If the ordinary CLT holds for two estimates $\theta_T^{(1)}$ and $\theta_T^{(2)}$ and the extended CLT (with $g_{2,T}$ dependent on T) only holds for say $\theta_T^{(1)}$ this difference between the two estimates may formally be termed "small sample effect." The reason why this difference really reflects what happens in small samples has been explained in the Introduction.

This is the main purpose of the preceding model. Furthermore, we mentioned at the beginning of this section the situation that a process is observed through a linear filter with transfer function that has sharp peaks and troughs. A special example of such a linear filter is the situation where we form differences of the observed process to remove polynomial trends. This procedure causes a trough at $\lambda = 0$. The same holds if we remove a seasonal component by the s-step difference filter [cf. Priestley (1981), Section 7.7]. Furthermore, the model is of importance for the study of long-range dependence. In the context of long-range dependence, it is normally assumed that the spectrum has a singularity of the form $|\lambda|^{-1+\varepsilon}$ in the neighborhood of 0 [cf. Rosenblatt (1985), Section III.5]. Our model includes spectral densities of the form $|1 - (1 - 1/T)e^{i\lambda}|^{-2s} f_\theta^*(\lambda)$. Although this spectrum is bounded by KT^{2s}, its shape for large T is even more dramatic since it is approximately of the form $|\lambda/2|^{-2s} f_\theta^*(\lambda)$ for small λ. This admits a stronger long-range dependence than the usually considered model, and the following results should also be seen with a view to this application.

3. Maximum likelihood estimates. In this section we prove that the maximum likelihood estimate $\tilde{\theta}_T$ is efficient under Assumption 2.1 if the process $X_t^{(T)}$ [and the process Y_t from Assumption 2.1(a) and Definition 2.2] are Gaussian. In the proofs we make strong use of Lemmas A2.2 and A2.3.

We set $\mathbf{X}_T = (X_0^{(T)}, \dots, X_{T-1}^{(T)})'$, $f_\theta = f_{2,\theta,T}$ and $B_\theta = B_T(f_{2,\theta,T})$.

THEOREM 3.1. *Suppose that Assumptions* 2.1(a), (c), (e) *and* (f) *hold. Then*

$$\tilde{\theta}_T \to_P \theta_0.$$

PROOF. The method of proof is taken from Walker (1964), Section 2. We start by proving that for all $\theta_1 \in \Theta$ there exists a constant $c(\theta_1) > 0$ with

$$\lim_{T\to\infty} E_{\theta_0}\{\mathscr{L}_T(\theta_1) - \mathscr{L}_T(\theta_0)\} \geq c(\theta_1). \tag{3.1}$$

We obtain

$$E_{\theta_0}\{\mathscr{L}_T(\theta_1) - \mathscr{L}_T(\theta_0)\} = \frac{1}{2T}\log\det\left(B_{\theta_1}B_{\theta_0}^{-1}\right) + \frac{1}{2T}\operatorname{tr}\left\{B_{\theta_0}B_{\theta_1}^{-1} - I\right\}. \tag{3.2}$$

Let $\lambda_{1T}, \dots, \lambda_{TT}$ be the eigenvalues of $B_{\theta_0}B_{\theta_1}^{-1}$. Since these are also the eigenvalues of $B_{\theta_0}^{1/2}B_{\theta_1}^{-1}B_{\theta_0}^{1/2}$ we get from Lemma A2.2 that $|\lambda_{jT}| \leq \|B_{\theta_0}^{1/2}B_{\theta_1}^{-1}B_{\theta_0}^{1/2}\| \leq K$ uniformly in j and T. By a Taylor expansion of $\log\det(I + tA)$ around $t = 0$ we now obtain that there exists a $\tau \in [0,1]$ such that (3.2) is equal to

$$\frac{1}{4T}\sum_{j=1}^{T}\left(\frac{\lambda_{jT} - 1}{1 + \tau(\lambda_{jT} - 1)}\right)^2 \geq K\frac{1}{T}\sum_{j=1}^{T}(\lambda_{jT} - 1)^2 = K\frac{1}{T}\operatorname{tr}\left\{\left(B_{\theta_0}B_{\theta_1}^{-1} - I\right)^2\right\}.$$

Theorem A2.3 implies that this tends to

$$K\frac{1}{2\pi}\int_\Pi\left(\frac{f_{2,\theta_0}^*(\lambda)}{f_{2,\theta_1}^*(\lambda)} - 1\right)^2 d\lambda \geq c(\theta_1).$$

Furthermore, we have

$$\begin{aligned}
&\lim_{T\to\infty} \operatorname{var}_{\theta_0}\{\mathscr{L}_T(\theta_1) - \mathscr{L}_T(\theta_0)\}\\
&= \lim_{T\to\infty} \frac{1}{2T^2}\left|B_{\theta_0}^{1/2}\left(B_{\theta_1}^{-1} - B_{\theta_0}^{-1}\right)B_{\theta_0}^{1/2}\right|^2\\
&\leq \lim_{T\to\infty} \frac{\delta}{2T^2}\sum_{i=1}^{p}\left|B_{\theta_0}^{1/2}B_{\theta^{(i)}}^{-1}B_T\left(\left|\frac{\partial}{\partial\theta_i}f_{\theta^{(i)}}\right|\right)B_{\theta^{(i)}}^{-1}B_{\theta_0}^{1/2}\right|^2 \quad (\text{with } \theta^{(i)} \in \Theta)\\
&\leq \lim_{T\to\infty} \frac{K\delta}{T} = 0,
\end{aligned}$$

by using Lemmas A2.1(d) and (l) and A2.2. Using Lemma A2.1(j) and (k), we obtain with a mean value $\theta \in \Theta$,

$$\begin{aligned}
&\mathscr{L}_T(\theta_2) - \mathscr{L}_T(\theta_1)\\
&= \frac{1}{2T}\sum_{i=1}^{p}(\theta_2 - \theta_1)_i\left[\operatorname{tr}\left\{B_\theta^{-1}B_T\left(\frac{\partial}{\partial\theta_i}f_\theta\right)\right\} - \mathbf{X}'B_\theta^{-1}B_T\left(\frac{\partial}{\partial\theta_i}f_\theta\right)B_\theta^{-1}\mathbf{X}\right].
\end{aligned}$$

Lemmas A2.1(d), (g) and (h) and A2.2 now imply, with $U_\delta(\theta_1) := \{\theta_2 \in \Theta: |\theta_2 - \theta_1| < \delta\}$,

$$\sup_{\theta_2 \in U_\delta(\theta_1)} |\mathscr{L}_T(\theta_2) - \mathscr{L}_T(\theta_1)| \le K\delta\left\{1 + \frac{1}{T}\mathbf{X}'B_{\theta_0}^{-1}\mathbf{X}\right\}.$$

Furthermore, $E_{\theta_0}(1/T)\mathbf{X}'B_{\theta_0}^{-1}\mathbf{X} = 1$ and $\operatorname{var}_{\theta_0}(1/T)\mathbf{X}'B_{\theta_0}^{-1}\mathbf{X} = 2/T$. It follows with (3.1) that there exists for all $\theta_1 \neq \theta_0$ a $c(\theta_1) > 0$ with

$$\begin{aligned}
&\lim_{T\to\infty} P_{\theta_0}\left(\inf_{\theta_2 \in U_\delta(\theta_1)} \mathscr{L}_T(\theta_2) - \mathscr{L}_T(\theta_0) \ge c(\theta_1)/4\right)\\
&\quad \ge 1 - \lim_{T\to\infty} P_{\theta_0}\left(\mathscr{L}_T(\theta_1) - \mathscr{L}_T(\theta_0) < c(\theta_1)/2\right)\\
&\qquad - \lim_{T\to\infty} P_{\theta_0}\left(\sup_{\theta_2 \in U_\delta(\theta_1)} |\mathscr{L}_T(\theta_2) - \mathscr{L}_T(\theta_1)| \ge c(\theta_1)/4\right)\\
&\quad = 1,
\end{aligned}$$

for sufficiently small δ. As in Walker (1964) this implies the result. □

In the following lemma we calculate the limit of the Fisher information matrix. Let

$$\Gamma(\theta)_{ij} := \frac{1}{4\pi}\int_\Pi \left(\frac{\partial}{\partial\theta_i}\ln f_{2,\theta}^*(\alpha)\right)\left(\frac{\partial}{\partial\theta_j}\ln f_{2,\theta}^*(\alpha)\right) d\alpha. \tag{3.3}$$

Lemma 3.2. *Suppose that Assumptions* 2.1(a) *and* (e) *hold and* $X_t^{(T)}$ *is Gaussian. Then* $\Gamma(\theta) = \lim_{T\to\infty}(1/T)\Gamma_T(\theta)$ *where* Γ_T *is the Fisher information matrix.*

Proof. Direct calculation gives [cf. Dzhaparidze and Yaglom (1983)]

$$\begin{aligned}
\frac{1}{T}\Gamma_T(\theta) &= \frac{1}{2T}\operatorname{tr}\left\{B_\theta^{-1}B_T(\nabla f_\theta)B_\theta^{-1}B_T(\nabla f_\theta)\right\}\\
&= \frac{1}{2T}\operatorname{tr}\left\{B_\theta^{-1}B_T(f_\theta \nabla \ln f_{2,\theta}^*)B_\theta^{-1}B_T(f_\theta \nabla \ln f_{2,\theta}^*)\right\},
\end{aligned} \tag{3.4}$$

which, by using Theorem A2.3, tends to $\Gamma(\theta)$. □

Theorem 3.3. *Suppose that Assumptions* 2.1(a)–(c), (e) *and* (f) *hold and* $X_t^{(T)}$ *is Gaussian. Then*

$$\sqrt{T}\left(\tilde{\theta}_T - \theta_0\right) \to_{\mathscr{D}} \mathscr{N}\left(0, \Gamma(\theta_0)^{-1}\right).$$

Proof. We use the usual proof technique with the mean-value theorem. Define

$$\nabla g_\theta := \left(\frac{\partial}{\partial\theta_i} g_\theta\right)_{i=1,\dots,p} \quad \text{and} \quad \nabla^2 g_\theta := \left(\frac{\partial^2}{\partial\theta_i\,\partial\theta_j} g_\theta\right)_{i,j=1,\dots,p}.$$

Then

$$\nabla \mathscr{L}_T(\tilde{\theta}_T)_i - \nabla \mathscr{L}_T(\theta_0)_i = \sum_{j=1}^{p} \nabla^2 \mathscr{L}_T(\bar{\theta}_T^{(i)})_{ij}(\tilde{\theta}_T - \theta_0)_j,$$

with $|\bar{\theta}_T^{(i)} - \theta_0| \le |\tilde{\theta}_T - \theta_0|$. Since $\nabla \mathscr{L}_T(\tilde{\theta}_T) = 0$ and $\tilde{\theta}_T \to_P \theta_0$ it is sufficient to prove

(i) $$\theta_T \to_P \theta_0 \Rightarrow \nabla^2 \mathscr{L}_T(\theta_T) - \nabla^2 \mathscr{L}_T(\theta_0) \to_P 0,$$

(ii) $$\nabla^2 \mathscr{L}_T(\theta_0) \to_P \Gamma(\theta_0),$$

(iii) $$T^{1/2} \nabla \mathscr{L}_T(\theta_0) \to_{\mathscr{D}} \mathscr{N}(0, \Gamma(\theta_0)).$$

We only sketch the proofs.

(i) We obtain with Lemma A2.1(j) and (k),

$$\begin{aligned} \nabla^2 \mathscr{L}_T(\theta) = & -\frac{1}{2T}\operatorname{tr}\{B_\theta^{-1} B_T(\nabla f_\theta) B_\theta^{-1} B_T(\nabla f_\theta)\} \\ & + \frac{1}{2T}\operatorname{tr}\{B_\theta^{-1} B_T(\nabla^2 f_\theta)\} \\ & + \frac{1}{T}\mathbf{X}_T B_\theta^{-1} B_T(\nabla f_\theta) B_\theta^{-1} B_T(\nabla f_\theta) B_\theta^{-1}\mathbf{X}_T \\ & - \frac{1}{2T}\mathbf{X}_T B_\theta^{-1} B_T(\nabla^2 f_\theta) B_\theta^{-1}\mathbf{X}_T. \end{aligned} \tag{3.5}$$

Let $\varepsilon > 0$. Choose $\delta > 0$ such that $|\theta_1 - \theta_0| \le \delta$ implies $\sup_{\lambda \in \Pi}|\nabla^2 f_{2,\theta_1}^*(\lambda) - \nabla^2 f_{2,\theta_0}^*(\lambda)| \le \varepsilon$. To estimate the difference $\nabla^2 \mathscr{L}_T(\theta_1) - \nabla^2 \mathscr{L}_T(\theta_0)$, we have to consider the preceding four terms separately. The difference of the two second terms is

$$\begin{aligned} & \frac{1}{2T}\operatorname{tr}\Big\{B_T(\nabla^2 f_{\theta_1})^{1/2}\big(B_{\theta_1}^{-1} - B_{\theta_0}^{-1}\big)B_T(\nabla^2 f_{\theta_1})^{1/2}\Big\} \\ & \quad + \frac{1}{2T}\operatorname{tr}\Big\{B_{\theta_0}^{-1/2} B_T\big(\nabla^2 f_{\theta_1} - \nabla^2 f_{\theta_0}\big)B_{\theta_0}^{-1/2}\Big\}, \end{aligned}$$

which, by Lemmas A2.1(d) and (g) and A2.2, is less than

$$K\big\|B_{\theta_0}^{1/2}\big(B_{\theta_1}^{-1} - B_{\theta_0}^{-1}\big)B_{\theta_0}^{1/2}\big\| + K\varepsilon.$$

The first expression is less than $K\delta$ by using Lemma A2.1(m). The difference of the two last terms in (3.5) is by Lemma A2.1(h) less than

$$K\frac{1}{T}\mathbf{X}_T' B_{\theta_0}^{-1}\mathbf{X}_T \Big\|B_{\theta_0}^{1/2}\Big\{B_{\theta_1}^{-1} B_T\big(\nabla^2 f_{\theta_1}\big)B_{\theta_1}^{-1} - B_{\theta_0}^{-1} B_T\big(\nabla^2 f_{\theta_0}\big)B_{\theta_0}^{-1}\Big\}B_{\theta_0}^{1/2}\Big\|.$$

The matrix norm is small by the same methods as used previously. Similar estimations also hold for the first and third terms in (3.5). Since

$$E\frac{1}{T}\mathbf{X}_T' B_{\theta_0}^{-1}\mathbf{X}_T = 1,$$

this implies (i).

(ii) By using $\operatorname{var}(\mathbf{X}_T' A \mathbf{X}_T) = 2|\Sigma^{1/2} A \Sigma^{1/2}|^2$, where $\Sigma = E\mathbf{X}_T\mathbf{X}_T'$, $E\mathbf{X}_T = 0$, we obtain with Lemma A2.2 $\operatorname{var} \nabla^2 \mathscr{L}_T(\theta_0)_{ij} \le KT^{-1}$. Furthermore,

$$E\,\nabla^2 \mathscr{L}_T(\theta_0) = \frac{1}{2T}\operatorname{tr}\left\{B_{\theta_0}^{-1} B_T(\nabla f_{\theta_0}) B_{\theta_0}^{-1} B_T(\nabla f_{\theta_0})\right\},$$

which (cp. Lemma 3.2) tends to $\Gamma(\theta_0)$.

(iii) We obtain $T^{1/2} E\, \nabla \mathscr{L}_T(\theta_0) = 0$ and

$$T\operatorname{cov}\left(\nabla \mathscr{L}_T(\theta_0)_i, \nabla \mathscr{L}_T(\theta_0)_j\right) = \frac{1}{2T}\operatorname{tr}\left\{B_{\theta_0}^{-1} B_T(\nabla f_{\theta_0}) B_{\theta_0}^{-1} B_T(\nabla f_{\theta_0})\right\} \to \Gamma(\theta_0).$$

Since the process is Gaussian the product theorem for cumulants [Brillinger (1981), Theorem 2.3.2] implies

$$\begin{aligned} &T^{l/2}\operatorname{cum}\left(\nabla \mathscr{L}_T(\theta_0)_{i_1}, \dots, \nabla \mathscr{L}_T(\theta_0)_{i_l}\right) \\ &\quad = \frac{1}{2} T^{-l/2}(-1)^l \frac{1}{l} \sum_{\substack{(j_1,\dots,j_l) \\ \text{permutation of} \\ (i_1,\dots,i_l)}} \operatorname{tr}\left[\prod_{k=1}^{l} \left\{B_\theta^{-1} B_T\left(\frac{\partial}{\partial \theta_{i_k}} f_{\theta_0}\right)\right\}\right], \end{aligned}$$

which, by using Lemmas A2.1(g) and (d) and A2.2 is bounded by $KT^{-l/2+1}$. This implies (iii). □

Thus, the maximum likelihood estimate is still optimal in the model of Assumption 2.1. In the following sections we study several other estimates.

4. Conditional likelihood estimates. One way to overcome numerical difficulties in calculating the exact maximum likelihood estimate is to choose adequate approximations to the inverse $B_T(f_\theta)^{-1}$. Suppose the observed process admits an autoregressive representation

$$\sum_{s=0}^{\infty} a_\theta(s) X_{t-s} = \varepsilon_t, \quad \text{with } \varepsilon_t \sim \mathscr{N}\left(0, \sigma^2(\theta)\right).$$

Then one may choose the approximation

$$B_T(f_\theta)^{-1} \approx \sigma^2(\theta)^{-1} D_T(\theta) D_T(\theta)',$$

where $D_T(\theta)$ is the upper triangular matrix of order T whose (m, n)-entry, $m \le n$, is $a_\theta(n-m)$. Thus, $\tilde{\theta}_T^c$ is the value that minimizes

$$\mathscr{L}_T^c(\theta) = \frac{1}{2}\log\{2\pi\sigma^2(\theta)\} + \frac{1}{2T\sigma^2(\theta)} \mathbf{X}_T' D_T(\theta) D_T(\theta)' \mathbf{X}_T.$$

Ljung and Caines (1979) have proved that $\tilde{\theta}_T^c$ is Fisher efficient. However, with the model of Section 2 we now prove that $\tilde{\theta}_T^c$ is worse than $\tilde{\theta}_T$.

THEOREM 4.1. *There exists a parametric model such that Assumptions 2.1(a)–(e) are fulfilled (with arbitrary s_1 and s_2) and $\tilde{\theta}_T^c$ is not a consistent estimate of θ_0.*

PROOF. We define the sequence $X_t^{(T)}$ as the Gaussian AR(1) process with characteristic root $p_T = 1 - 1/T$ and innovation variance θ, i.e., with the spectral density

$$f_{2,\theta,T}(\lambda) = \frac{\theta}{2\pi}|1 - p_T e^{i\lambda}|^{-2}.$$

From Theorem 2.4 we get

$$X_t^{(T)} \in \mathscr{X}\left(\frac{\theta}{2\pi}, 0, C, T, s_1, s_2\right),$$

for some $C > 0$ and all $s_1, s_2 \in \mathbb{N}$. We assume that p_T is known and θ has to be estimated. Straightforward calculation gives

$$\tilde{\theta}_T^c = \left\{1 + \left(1 - \frac{1}{T}\right)p_T^2\right\}\frac{1}{T}\sum_{t=0}^{T-1} X_t^{(T)2} - 2\left(1 - \frac{1}{T}\right)p_T\frac{1}{T}\sum_{t=0}^{T-2} X_t^{(T)}X_{t+1}^{(T)}$$

and therefore

$$E\tilde{\theta}_T^c = \theta_0 + \frac{1}{T}\frac{p_T^2}{1 - p_T^2} = \theta_0 + \frac{1}{2} + O(T^{-1}). \qquad \square$$

Thus, $\tilde{\theta}_T^c$ is worse than $\tilde{\theta}_T$ (cf. Theorem 3.3). The same holds for the classical Whittle estimate $\hat{\theta}_T$, which is studied in more detail in Section 7.

5. Data tapers. In order to investigate tapered Whittle estimates, we need some properties of data tapers. For a sequence of data tapers $h_{t,T}$, $t = 0, \dots, T-1$; $T \in \mathbb{N}$, we define

$$H_k^{(T)}(\alpha) := \sum_{t=0}^{T-1} h_{t,T}^k \exp(-i\alpha t), \quad H_T(\alpha) := H_1^{(T)}(\alpha) \quad \text{and} \quad H_{k,T} := H_k^{(T)}(0).$$

DEFINITION 5.1. Let $k \in \mathbb{N}_0$ and $\kappa \in [0, 1/2)$. Suppose $h_{t,T} = h_T(t/T)$ is a sequence of data tapers with $h_T(x) = 0$ for all $x \notin [0,1)$ that fulfills the following conditions.

(i) h_T is $(k-1)$-times continuously differentiable (in the case $k = 1$ we assume continuity and in the case $k = 0$ we make no assumption).

(ii) There exists a finite set $P_T = \{p_{1T}, \dots, p_{rT})$ such that h_T is $(k+1)$-times differentiable in all $x \notin P_T$.

(iii) Let

$$s_{jT} := \lim_{y \downarrow p_{jT}} h_T^{(k)}(y) - \lim_{y \uparrow p_{jT}} h_T^{(k)}(y).$$

There exists a $c > 0$ such that $\sum_{j=1}^{r} s_{jT}^2 \geq c$ for all $T \in \mathbb{N}$.

(iv) $H_{2,T} \sim T$ and

$$D_T^{(k)} := \sup_{x \notin P_T} \left| h_T^{(k)}(x) \right| + \sup_{x \notin P_T} \left| h_T^{(k+1)}(x) \right| \leq K T^{\kappa}, \quad \text{with } \kappa \in [0, 1/2).$$

Then we say that the taper (the sequence of tapers) is of degree (k, κ).

In the nontapered case $h_{t,T} = \chi_{[0,1)}(t/T)$ the degree therefore is $(0,0)$.

EXAMPLE 5.2 (Tukey–Hanning taper). A certain proportion of the data is tapered with a cosine bell, i.e.,

$$\begin{aligned} h_\rho(x) &= (1/2)[1 - \cos(2\pi x/\rho)], & x &\in [0, \rho/2), \\ &= 1, & x &\in [\rho/2, 1/2], \\ &= h_\rho(1-x), & x &\in (1/2, 1]. \end{aligned}$$

Thus, h_ρ is continuously differentiable and 3-times differentiable in $x \notin P = \{0, \rho/2, 1-\rho/2, 1\}$. We obtain $|h_\rho^{(3)}(x)| \leq 4\pi^3/\rho^3$ for $x \notin P$, which means that the taper $h_{t,T} = h_\rho(t/T)$ with ρ fixed has degree $(2,0)$. If we choose, e.g., $\rho = \rho_T = T^{-\kappa/3}$, $h_{t,T} = h_{\rho_T}(t/T)$ is of degree $(2, \kappa)$ with $h_{\rho_T}(x) \to \chi_{(0,1)}(x)$, which will lead to efficient Whittle estimates (see Theorem 6.3).

EXAMPLE 5.3 (Polynomial taper). The function

$$\begin{aligned} h_\rho(x) &= 4^k (x/\rho)^k (1 - x/\rho)^k, & x &\in [0, \rho/2), \\ &= 1, & x &\in [\rho/2, 1/2], \\ &= h_\rho(1-x), & x &\in (1/2, 1], \end{aligned}$$

is $(k-1)$-times continuously differentiable and $(k+1)$-times differentiable in $x \notin P = \{0, \rho/2, 1-\rho/2, 1\}$. Thus, the taper $h_{t,T} = h_\rho(t/T)$, where ρ is fixed, has degree $(k, 0)$. Furthermore, we have $\sup_{x \notin P} |h_\rho^{(l)}(x)| \leq K\rho^{-l}$, $0 \leq l \leq 2k$, with K independent of ρ. Thus, if, e.g., $\rho = \rho_T = T^{-\kappa/(k+1)}$, the taper $h_{t,T} = h_{\rho_T}(t/T)$ has degree (k, κ) and $h_{\rho_T}(x) \to \chi_{(0,1)}(x)$.

We now prove a fundamental inequality for data tapers.

LEMMA 5.4. *Let $k \in \mathbb{N}_0$, $\kappa \in [0, 1/2)$ and $(h_{t,T})_{T \in \mathbb{N}}$ be a sequence of data tapers of degree (k, κ). Then there exists a constant $K \in \mathbb{R}$ such that for all $\alpha \in \mathbb{R}$ and $T \in \mathbb{N}$,*

$$|H_T(\alpha)| / H_{2,T}^{1/2} \leq K T^{-k-1/2+\kappa} L_T(\alpha)^{k+1}. \tag{5.1}$$

PROOF. We consider the case $k \geq 1$ (in the case $k = 0$ the result follows easily by summation by parts). Let $D_{l,t}(h_{t,T})$ be the lth difference sequence of $h_{t,T}$, i.e., $D_{0,t}(h_{t,T}) = h_{t,T}$, $D_{l,t}(h_{t,T}) = D_{l-1,t}(h_{t+1,T}) - D_{l-1,t}(h_{t,T})$ for $l \geq 1$. We have $D_{l,t}(t^m) = 0$ if $m < l$ and $D_{l,t}(t^l) = l!$, $l, m \in \mathbb{N}$. Since Definition 5.1(i) implies $h_T(0) = h_T(1) = 0$ we obtain by repeated summation by parts

$$(5.2) \qquad H_T(\alpha) = [\exp(i\alpha) - 1]^{-(k+1)} \sum_{t=-k-1}^{T-1} D_{k+1,t}(h_{t,T})\exp(-i\alpha t).$$

There are at most $(k+2)r$ intervals $[t/T, (t+k+1)/T]$, $t = -k-1, \dots, T-1$, that contain a point $x \in P_T$. For those t we obtain with a Taylor expansion

$$D_{k-1,t}(h_{t+j,T}) = T^{-(k-1)} h_T^{(k-1)}\left(\frac{t+j}{T}\right) + O\left(T^{-k} \sup_{x \notin P_T} |h_T^{(k)}(x)|\right),$$

for $j = 0, 1, 2,$

which implies

$$D_{k+1,t}(h_{t,T}) = O\left(T^{-k} \sup_{x \notin P_T} |h_T^{(k)}(x)|\right).$$

For the other t we also obtain with a Taylor expansion

$$D_{k+1,t}(h_{t,T}) = O\left(T^{-(k+1)} \sup_{x \notin P_T} |h_T^{(k+1)}(x)|\right)$$

and therefore

$$|H_T(\alpha)| \le K \frac{L_T(\alpha)^{k+1}}{T^k} D_T^{(k)}, \quad \text{if } |\alpha - 2\pi l| \ge \frac{1}{T}, \text{ for all } l \in \mathbb{Z}.$$

If there exists a l with $|\alpha - 2\pi l| \le 1/T$ we obtain

$$|H_T(\alpha)| \le T \sup_{x \in [0,1)} |h_T(x)| \le K \frac{L_T(\alpha)^{k+1}}{T^k} D_T^{(k)}$$

and therefore relation (5.1). □

The lemma implies $|H_T(\alpha)|/H_{2,T}^{1/2} = O(T^{-k-1/2+\kappa})$ for all $\alpha \not\equiv 0 \mod 2\pi$. Therefore, the convergence rate to 0 at these α increases with k.

For all further calculations we only need property (5.1), which we therefore could alternatively take as the definition of the degree (k, κ). This would also include other tapers [e.g., the Kolmogorov–Zhurbenko taper—cf. Zhurbenko (1980), (2.10)].

In practical situations we do not want to drop the first observation X_0 completely, which would happen by using the taper $h_{t,T} = h_T(t/T)$ with $h_T(0) = 0$. One therefore chooses in practice the taper $h_{t,T} = h_T((t + 1/2)/T)$, which also fulfills (5.1).

6. Tapered Whittle estimates. In this section we prove that a suitable choice of the data taper leads to efficient Whittle estimates. The technical details of the calculations are given in Appendix 1.

If $\phi\colon \Pi \to \mathbb{C}$ is an integrable function and $X_t^{(T)} \in \mathscr{X}(f_2^*, f_4^*, C, T, s_1, s_2)$ is a sequence of stationary processes with spectral density $f_{2,T}$ we define

$$J_T(\phi) = \int_\Pi \phi(\lambda) \frac{I_T(\lambda)}{f_{2,T}(\lambda)}\, d\lambda$$

and

$$J(\phi) = \int_\Pi \phi(\lambda)\, d\lambda.$$

THEOREM 6.1. *Suppose that Assumptions* 2.1(a), (c), (e) *and* (g) *hold. If the data taper has degree* (k, κ) *with* $k \geq \max\{s_1 + s_2 - 1, s_1, s_2\}$ *and* $\kappa < \kappa_0/2$, *where* $\kappa_0 \in (0, 1]$, *then*

$$\hat{\theta}_T \to_P \theta_0.$$

PROOF. By using Lemmas A1.3(a) and A1.4(a) the result follows analogously to the proof of Theorem 3.1. We omit the details. □

We now investigate the Fisher efficiency of $\hat{\theta}_T$. In addition to $\Gamma(\theta)$ [cp. (3.3)] we define $B(\theta)$ by

$$B(\theta)_{ij} := \frac{1}{8\pi} \int_{\Pi^2} \left(\frac{\partial}{\partial \theta_i} \ln f_{2,\theta}^*(\alpha_1) \right) \times \left(\frac{\partial}{\partial \theta_j} \ln f_{2,\theta}^*(\alpha_2) \right) \frac{f_{4,\theta}^*(\alpha_1, -\alpha_1, \alpha_2)}{f_{2,\theta}^*(\alpha_1) f_{2,\theta}^*(\alpha_2)}\, d\alpha. \tag{6.1}$$

THEOREM 6.2. *Suppose that Assumptions* 2.1(a)–(e) *and* (g) *hold. If the data taper has degree* (k, κ) *with* $k \geq \max\{s_1 + s_2 - 1, s_1, s_2\}$ *and* $\kappa < \kappa_0/4$, *where* $\kappa_0 \in (0, 1]$, *then*

$$\sqrt{T}\left(\hat{\theta}_T - \theta_0\right) \to_{\mathscr{D}} \mathscr{N}\left(0, \left\{ \lim_{T \to \infty} \frac{T H_{4,T}}{H_{2,T}^2} \right\} \Gamma(\theta_0)^{-1} (\Gamma(\theta_0) + B(\theta_0)) \Gamma(\theta_0)^{-1} \right).$$

PROOF. As in the proof of Theorem 3.3 we obtain

$$\nabla \mathscr{L}_T^W(\hat{\theta}_T)_i - \nabla \mathscr{L}_T^W(\theta_0)_i = \sum_{j=1}^p \nabla^2 \mathscr{L}_T^W(\bar{\theta}_T^{(i)})_{ij} (\bar{\theta}_T - \theta_0)_j,$$

with $|\bar{\theta}_T^{(i)} - \theta_0)| \leq |\hat{\theta}_T - \theta_0|$. Using Lemmas A1.3(a) and A1.4(a) and the continuity properties of $f_{2,\theta}^*$, it follows that

$$\nabla^2 \mathscr{L}_T^W(\theta_0) = J_T\left(\frac{1}{4\pi} f_{2,\theta_0}^* \nabla^2 f_{2,\theta_0}^{*-1} \right) - J\left(\frac{1}{4\pi} f_{2,\theta_0}^* \nabla^2 f_{2,\theta 0}^{*-1} \right) + \Gamma(\theta_0) \to_P \Gamma(\theta_0)$$

and

$$\nabla^2 \mathscr{L}_T^W(\bar{\theta}_T^{(i)}) - \nabla^2 \mathscr{L}_T^W(\theta_0) \to_P 0.$$

Furthermore, we have $\nabla \mathscr{L}_T^W(\theta_T) = 0$ and

$$\sqrt{T}\,\nabla \mathscr{L}_T^W(\theta_0) = \sqrt{T}\left\{ J_T\left(\frac{1}{4\pi} f_{2,\theta_0}^* \,\nabla f_{2,\theta_0}^{*-1}\right) - J\left(\frac{1}{4\pi} f_{2,\theta_0}^* \,\nabla f_{2,\theta_0}^{*-1}\right)\right\},$$

which, by Theorem A1.2, tends weakly to

$$\mathscr{N}\left(0, \left\{\lim_{T\to\infty} \frac{TH_{4,T}}{H_{2,T}^2}\right\}(\Gamma(\theta_0) + B(\theta_0))\right).$$

Thus, the theorem is proved. □

If we choose a taper $h_{t,T}$ with $\lim_{T\to\infty}(TH_{4,T}/H_{2,T}^2) = 1$ (e.g., the taper from Example 5.3), we obtain for the limit covariance $\Gamma(\theta_0)^{-1}(\Gamma(\theta_0) + B(\theta_0))\Gamma(\theta_0)^{-1}$, or, in the Gaussian case, $\Gamma(\theta_0)^{-1}$, i.e., the estimate is efficient.

THEOREM 6.3. *Suppose that Assumptions* 2.1(a)–(e) *and* (g) *hold. Let* $X_t^{(T)} \in \mathscr{X}(f_{2,\theta_0}^*, 0, C, T, s_1, s_2)$ *be a sequence of stationary Gaussian processes and* $h_{t,T}$ *be a data taper of degree* (k, κ) *with* $k \geq \max\{s_1 + s_2 - 1, s_1, s_2\}$, $\kappa < \kappa_0/4$ *and* $\lim_{T\to\infty}(TH_{4,T}/H_{2,T}^2) = 1$. *Then* $\hat{\theta}_T$ *is Fisher efficient.*

REMARK 6.4. All results of this section remain valid if the integral in the Whittle likelihood $\mathscr{L}_T^W$ is replaced by the corresponding sum over the Fourier frequencies $2\pi s/T$, $s = 0, \dots, T-1$, which can be calculated efficiently by using the fast-Fourier algorithm. The proofs (especially the proof of Theorem A1.2) are nearly the same as for the preceding results. The main property needed for the modification is Lemma A1.1.(i).

Thus, tapered Whittle estimates, with a suitable choice of the taper, have the same optimality properties as maximum likelihood estimates.

7. Nontapered Whittle estimates. In this section we consider the classical nontapered Whittle estimate (or, more generally, the situation where the degree of the applied taper is too low). A treatment of the classical Whittle estimate may be found, for example, in Dzhaparidze and Yaglom (1983). Consistency and efficiency in the "normal" case (where the spectral density is bounded from above and below) also follow from our Theorems 6.1 and 6.2 with $s_1 = s_2 = 0$.

We now show in this section that the nontapered estimate also may even be inconsistent if there are sharp peaks or troughs in the spectrum. We first consider the case where peaks are present.

THEOREM 7.1. *Let* $s_1 > 0$ *and* $s_2 \leq 1$. *Suppose that the applied data taper is of degree* (k, κ) *with* $k < s_1$. *Then there exists a parametric model such that Assumptions* 2.1(a)–(e) *and* (g) *are fulfilled and the Whittle estimate* $\hat{\theta}_T$ *is not a consistent estimate of* θ_0.

PROOF. We define the sequence $X_t^{(T)}$ as the Gaussian AR$(k+1)$ process with $(k+1)$-times the characteristic root $p_T = 1 - 1/T$ and innovation variance θ, i.e., with the spectral density

$$f_{2,T}(\lambda) = \frac{\theta}{2\pi}|1 - p_T e^{i\lambda}|^{-2(k+1)}.$$

Since $k + 1 \le s_1$ we obtain from Theorem 2.4 $X_t^{(T)} \in \mathscr{X}(\theta/2\pi, 0, C, T, s_1, s_2)$ for a $C > 1$. We assume that p_T is known and the innovation variance has to be estimated. Minimizing the Whittle function leads to the estimate

$$\hat{\theta}_T = J_T\left(\frac{\theta_0}{2\pi}\right), \quad \text{whereas } \theta_0 = J\left(\frac{\theta_0}{2\pi}\right).$$

We first note that

$$L_{1/(1-p)}(\lambda)^2 \le |1 - pe^{i\lambda}|^{-2} \le 2\pi^2 L_{1/(1-p)}(\lambda)^2, \tag{7.1}$$

for all $0 < p < 1$ and all $\lambda \in \Pi$. Elementary calculation now gives

$$\begin{aligned}\frac{f_{2,T}(\lambda+\alpha)}{f_{2,T}(\lambda)} &= \left\{1 + 2p_T\frac{2\cos(\lambda+\alpha)\sin^2(\alpha/2) - \sin(\lambda+\alpha)\sin\alpha}{|1 - p_T e^{i(\lambda+\alpha)}|^2}\right\}^{k+1} \\ &= 1 + \left\{4p_T\cos(\lambda+\alpha)\sin^2\left(\frac{\alpha}{2}\right)\right\}^{k+1}\frac{2\pi}{\theta_0}f_{2,T}(\lambda+\alpha) \\ &\quad + O\left(\sum_{j=1}^{2k+1}|\alpha|^j L_T(\lambda+\alpha)^j\right).\end{aligned}$$

Using the same arguments as in Lemma A1.3, we therefore obtain

$$\begin{aligned}E\hat{\theta}_T - \theta_0 = {}& \{4p_T\}^{k+1}\{2\pi H_{2,T}\}^{-1} \\ & \times \int_{\Pi^2}\sin^{2k+2}(\alpha/2)\cos^{k+1}(\lambda+\alpha)f_{2,T}(\lambda+\alpha)|H_T(\alpha)|^2\,d\alpha\,d\lambda \\ & + O(T^{2\kappa-1}\log^2 T).\end{aligned}$$

Using (5.2), the first term is equal to

$$p_T^{k+1}H_{2,T}^{-1}\int_\Pi f_{2,T}(\lambda)\cos^{k+1}\lambda\,d\lambda \sum_{t=-k-1}^{T-1} D_{k+1,t}(h_{t,T})^2. \tag{7.2}$$

(7.1) implies

$$\begin{aligned}\int_\Pi f_{2,T}(\lambda)\cos^{k+1}\lambda\,d\lambda &\ge \frac{\theta_0}{\pi}\left(\frac{1}{2}\right)^{k+1}\int_0^{\pi/4}L_T(\lambda)^{2k+2}\,d\lambda - \frac{\theta_0}{\pi}\int_{\pi/2}^{\pi}L_T(\lambda)^{2k+2}\,d\lambda \\ &\ge KT^{2k+1}.\end{aligned}$$

Let s_{jT}, p_{jT} and $D_T^{(k)}$ be as in Definition 5.1 and $h_T^0(x)$ be the "jump function" of $h_T^{(k)}(x)$, i.e., $h_T^0(x) = \sum_{j=1}^r s_{jT}\chi_{[p_{jT},\infty)}(x)$ and $h_T^*(x) = h_T^{(k)}(x) - h_T^0(x)$ on $(-1,2)\setminus P_T$. Using the Taylor expansion of degree $k-2$ and integration by parts, we obtain for y, $y_0 \in (-1, 2)$,

$$h_T(y) = \sum_{j=0}^{k-1} \frac{h_T^{(j)}(y_0)}{j!}(y - y_0)^j + \frac{h_T^*(y_0)}{k!}(y - y_0)^k + O(|y - y_0|^{k+1} D_T^{(k)})$$
$$+ \sum_{p_{jT} < y_0} \frac{s_{jT}}{k!}(y - y_0)^k + \sum_{y_0 \le p_{jT} < y} \frac{s_{jT}}{k!}(y - p_{jT})^k.$$

Taking this expansion at $y_0 = t/T$ and $y = (s + t)/T$, $s = 0, \dots, k + 1$, gives with $D_{l,t}(t^m) = 0$ for $m < l$,

$$D_{k+1,t}(h_{t,T}) = O(T^{-(k+1)} D_T^{(k)})$$
$$+ D_{k+1,s}\left(\sum_{t/T \le p_{jT} \le s/T} \frac{s_{jT}}{k!} \left(\frac{s}{T} - p_{jT} \right)^k \right) \Bigg|_{s=t}.$$

We now assume that $T \ge T_0$, where $p_{jT} - p_{j-1,T} > (k + 2)/T_0$, $j = 2, \dots, r$. Furthermore, let $t_j = [p_{jT} T]$, $j = 1, \dots, r$. Then

$$\sum_{t=t_{j-1}+1}^{t_j} D_{k+1,t}(h_{t,T})^2 = O\left(T^{-2k-1}\left\{D_T^{(k)2} + |s_{jT}| D_T^{(k)}\right\}\right)$$
$$+ \frac{s_{jT}^2}{(k!)^2} \sum_{t=t_j-k}^{t_j} D_{k+1,s}\left(\left(\frac{s}{T} - p_{jT} \right)^k \chi_{\{s > t_j\}} \right)^2 \Bigg|_{s=t}.$$

We obtain

$$D_{k+1,s}\left(\left(\frac{s}{T} - p_{jT} \right)^k \chi_{\{s > t_j\}} \right)^2 \Bigg|_{s=t_j-k} = \left(\frac{t_j + 1}{T} - p_{jT} \right)^{2k}$$

and

$$D_{k+1,s}\left(\left(\frac{s}{T} - p_{jT} \right)^k \chi_{\{s > t_j\}} \right)^2 \Bigg|_{s=t_j}$$
$$= \left[D_{k+1,s}\left(\left(\frac{s}{T} - p_{jT} \right)^k \right) \Bigg|_{s=t_j} + (-1)^k \left(\frac{t_j}{T} - p_{jT} \right)^k \right]^2$$
$$= \left(\frac{t_j}{T} - p_{jT} \right)^{2k}.$$

By considering the cases $p_{jT} \le (t_j + 1/2)/T$ and $p_{jT} \ge (t_j + 1/2)/T$, we therefore obtain that (7.2) is bounded from below. Therefore, $\hat{\theta}_T$ is not consistent. □

The theorem establishes theoretically a parametric leakage effect for the classical Whittle estimate (for $s_1 = 1$, $s_2 = 0$, $k = 0$). It explains the small sample behavior of the estimate if peaks are present in the spectrum. We now prove that the same result holds if the spectrum has troughs.

THEOREM 7.2. *Let $s_2 > 0$ and $s_1 \le 1$. Suppose that the applied data taper is of degree (k, κ) with $k < s_2$. Then there exists a parametric model such that Assumptions* 2.1(a)–(e) *and* (g) *are fulfilled and the Whittle estimate $\hat{\theta}_T$ is not a consistent estimate of θ_0.*

PROOF. We consider the MA$(k+1)$ process with the root $p_T = 1 - 1/T$ and the innovation variance θ that has to be estimated. Since

$$\int_{\Pi^2}\left[\frac{f_{2,T}(\lambda+\alpha)}{f_{2,T}(\lambda)} - 1\right]\frac{|H_T(\alpha)|^2}{2\pi H_{2,T}}\,d\alpha\,d\lambda$$

$$= \int_{\Pi^2}\left[\frac{f_{2,T}(\lambda)}{f_{2,T}(\lambda+\alpha)} - 1\right]\frac{|H_T(\alpha)|^2}{2\pi H_{2,T}}\,d\alpha\,d\lambda,$$

we obtain the same result as before. □

Thus, we have also established theoretically another effect for the classical Whittle estimate, which we call the "trough effect" (if we set $s_1 = 0$, $s_2 = 1$, $k = 0$). If the process contains troughs the Whittle estimate is inconsistent (in the situation considered in the preceding proof the innovation variance was *over*estimated). This effect is especially important with a view to the usual taking of differences to remove trends in time-series data. This procedure causes a value of 0 at $\lambda = 0$. For our knowledge, this effect has never been described before. Simulation studies (not presented in this paper) confirmed the problems arising from taking differences as well as the advantages of data tapers in this situation.

The preceding results imply that the degree of the data taper (k, κ) with $k \ge \max\{s_1 + s_2 - 1, s_1, s_2\}$ is also necessary for the CLT Theorem 6.2. As far as we know, it is the first time that the advantage of data tapers for the previous type of estimates has been established.

However, the preceding theorems do not answer all questions of interest. For example, the situation of general s_1 and s_2 is not easy to handle. It is possible to prove for certain $n = s_1 + s_2$ that the polynomial taper from Example 5.3 with $k = n - 2$ is not sufficient for the estimate to be asymptotically unbiased. Since these calculations are very technical and the result is not very general we omit the proof.

We now consider the nontapered case in more detail. In Theorem 6.2 we have proved a central limit theorem where the additional factor $\{\lim TH_{4,T}/H_{2,T}^2\}$, which is ≥ 1, occurred in the asymptotic variance (this factor is missing in the CLT for the classical Whittle estimate). For nonparametric estimates we have the same factor [cf. Brillinger (1981), Section 5.6—note, that the nonparametric results have always been proved under the assumption that the spectra are bounded]. It is therefore a common opinion [see, e.g., Brillinger (1981), page 151, Hannan (1970), page 272, and Priestley (1981), page 562] that tapering may reduce in many situations the bias while it increases the variance of nonparametric estimates. However, these considerations presume implicitly that the variance of the estimate converges, in the situation where a strong peak in the spectrum is present and no data taper is applied, to the same limit as in Theorem 6.2 with $TH_{4,T}/H_{2,T}^2 = 1$. In the next theorem we show that this is not true. More generally, we will consider not only the variance but the cumulants of all order. We consider for an AR(1) process the behavior of the statistic $J_T(1)$. The following result therefore is essentially of a nonparametric nature. However, in Theorem 7.1 the Whittle estimate $\hat{\theta}_T$ of the innovation variance was equal to $(\theta_0/2\pi)J_T(1)$. Therefore, the following result also describes the property of the Whittle estimate.

THEOREM 7.3. *Let $X_t^{(T)}$ be the Gaussian* AR(1) *process with root p_T, where $|1 - p_T| \geq 1/T$ [thus $X_t^{(T)} \in \mathscr{X}(\sigma^2/2\pi, 0, C, T, 1, 0)$ for a $C \geq 1$ and all $\delta > 0$]. Suppose that no data taper is applied. Then we obtain for all $l \in \mathbb{N}$,*

$$\begin{aligned}\operatorname{cum}_l(J_T(1) - J(1)) &\\ = \left(\frac{4\pi}{T}\right)^l (l-1)!\,(1 - p_T^2)^{-l} p_t^{2l} \sum_{k=0}^{[l/2]} \binom{l}{2k} p_T^{2kT} + O(T^{-1}\log^2 T).\end{aligned}$$

PROOF. We consider the case $l \geq 2$ (the case $l = 1$ is obtained by similar arguments). Analogously to the proof of Lemma A1.5, we obtain for $\operatorname{cum}_l(J_T(1))$ an expression similar to (A.10), where we have only to sum over all indecomposable partitions with $|P_i| = 2$ (because $X_t^{(T)}$ is Gaussian). Since $\phi = 1$ and $f_{2,T}(\alpha) = f_{2,T}(-\alpha)$, all $(l-1)!2^{l-1}$ partitions lead to the same value. Therefore, we obtain with $\alpha_{l+1} = \alpha_1$,

$$\begin{aligned}\operatorname{cum}_l(J_T(1)) = (2\pi T)^{-l}(l-1)!2^{l-1} &\\ \times \int_{\Pi^l}\int_{\Pi^l} \left\{ \prod_{j=1}^{l} \frac{f_{2,T}(\gamma_j)}{f_{2,T}(\alpha_j)} \right\} &\\ \times \left\{ \prod_{j=1}^{l} H_T(\alpha_j - \gamma_j) H_T(\gamma_j - \alpha_{j+1}) \right\} d\gamma\, d\alpha.\end{aligned}$$

Using

$$\frac{1 - pe^{i\alpha}}{1 - pe^{i\gamma}} = 1 + pe^{i\alpha}\frac{e^{i(\gamma-\alpha)} - 1}{1 - pe^{i\gamma}},$$

this is equal to

$$\begin{aligned}(2\pi T)^{-l}(l-1)!2^{l-1} \int_{\Pi^l}\int_{\Pi^l} \left\{ \prod_{j=1}^{l} \left(1 + p_T e^{-i\alpha_j} \frac{e^{-i(\gamma_j - \alpha_j)} - 1}{1 - p_T e^{-i\gamma_j}}\right)\right. &\\ \left.\times \left(1 + p_T e^{i\alpha_{j+1}} \frac{e^{i(\gamma_j - \alpha_{j+1})} - 1}{1 - p_T e^{i\gamma_j}}\right)\right\} &\\ \times \left\{ \prod_{j=1}^{l} H_T(\alpha_j - \gamma_j) H_T(\gamma_j - \alpha_{j+1}) \right\} d\gamma\, d\alpha.\end{aligned}$$

Using

$$\left| p_T e^{i\alpha} \frac{e^{i(\gamma-\alpha)} - 1}{1 - p_T e^{i\gamma}} \right| \le K|\gamma - \alpha| L_T(\gamma)$$

[see (7.1)], $|H_T(\alpha)| \le K L_T(\alpha)$ and $|\alpha| L_T(\alpha) \le K$, this is equal to

$$(2\pi T)^{-l} p_T^{2l}(l-1)! 2^{l-1} \int_{\Pi^l} \int_{\Pi^l} \left\{ \prod_{j=1}^{l} \frac{e^{-i(\gamma_j - \alpha_j)} - 1}{1 - p_T e^{-i\gamma_j}} H_T(\alpha_j - \gamma_j) \right.$$

$$\left. \times \frac{e^{i(\gamma_j - \alpha_{j+1})} - 1}{1 - p_T e^{i\gamma_j}} H_T(\gamma_j - \alpha_{j+1}) \right\} d\gamma \, d\alpha + O(T^{-1}\log^2 T).$$

Using (5.2), this is equal to $(\gamma_{l+1} = \gamma_1)$

$$T^{-l} p_T^{2l}(l-1)! 2^{l-1} \left(\frac{\sigma^2}{2\pi} \right)^{-l}$$

$$\times \sum_{t_1, \ldots, t_l = -1}^{T-1} \int_{\Pi^l} \left\{ \prod_{j=1}^{l} f_{2,T}(\gamma_j) D_{1,t_j}\left(h_{t_j,T}\right)^2 e^{-i(\gamma_j - \gamma_{j+1})t_j} \right\} d\gamma$$

$$+ O(T^{-1}\log^2 T),$$

where $h_{t,T} = \chi_{[0,1)}(t/T)$. Since $D_{1,T}(h_{t,T})^2 = 1$ if $t = -1, \; T-1$ and $D_{1,t}(h_{t,T})^2 = 0$ otherwise, and

$$\int_{\Pi} f_{2,T}(\alpha) \exp(i\alpha u) \, d\alpha = \sigma^2 \frac{p_T^{|u|}}{1 - p_T^2}$$

this is equal to $(t_{l+1} = t_1)$

$$T^{-l} p_T^{2l}(l-1)! 2^{l-1} (2\pi)^l (1 - p_T^2)^{-l} \sum_{\substack{t_j = 0,1 \\ j = 1, \ldots, l}} p_T^{T \Sigma_{j=1}^{l} |t_j - t_{j+1}|} + O(T^{-1}\log^2 T).$$

Elementary calculation now gives the result. □

If we set $p_T = 1 - 1/T$ we obtain

$$\operatorname{cum}_l(J_T(1)) = (2\pi)^l (l-1)! \sum_{k=0}^{[l/2]} \binom{l}{2k} e^{-2k} + O(T^{-1}\log^2 T).$$

Thus, neither bias nor variance of the estimate tend to 0, and with a view to Lemmas A1.3 and A1.4 tapering reduces both. Furthermore, we see that even the higher-order cumulants ($l \ge 3$) do not tend to 0, which implies that the limit distribution of $J_T(1)$ is not Gaussian. With a view to Lemma A1.5 tapering also reduces the order of the higher-order cumulants of the statistic.

Theorem 7.3 also shows clearly the relation between the distance of p_T from the unit circle and the rate of convergence.

8. Concluding remarks. In this paper we have presented a mathematical model for the investigation of parameter estimates for stationary time series. Although the model is still an asymptotic one, it reflects up to a certain extent small sample effects by the requirement that estimates should behave well with increasing sample size for an increasing number of stationary processes.

By using this model we were able to explain theoretically several small sample effects: We could distinguish several Fisher efficient estimates (maximum likelihood estimates, conditional likelihood estimates and tapered and nontapered Whittle estimates), we explained theoretically a parametric leakage effect for the Whittle estimate and showed that this effect can be removed by applying a data taper, we found the so far unknown trough effect and we proved contrary to widespread conjectures that data tapers do not only reduce the bias, but may also reduce the variance of the estimates.

The theoretical results are in accordance with simulation studies. The advantage is extreme if roots of ARMA processes are close to the unit circle. From our considerations it is clear that a short sample length with "moderate" roots has the same effect on the estimates as a larger sample length with roots extremely close to the unit circle. This is confirmed by simulation studies (not presented in this paper) where similar results on the advantages of data tapers were also obtained for very short time series (e.g., with $T = 32$).

From the practical point of view we found that using data tapers for Whittle estimates (which are identical to Yule–Walker estimates in the autoregressive case), leads to a new powerful estimate. In 100 replications of the simulation presented in the Introduction the tapered Yule–Walker estimate (with the Tukey–Hanning taper with $\rho = 1.0$) led in 62 (73) cases to a better value of σ^2 $(T(\hat{a} - a)'\Sigma(\hat{a} - a))$ than the Burg algorithm. In 100 replications with a triple peak instead of a double peak at $\lambda = 1.5$ the tapered estimate was better in 28 (51) cases. The best results were obtained in a one-step iterative procedure where the series was extended to double length with back- and forecasting with the estimated values and the estimation was repeated afterward. This iterative tapered estimate was better than Burg's algorithm in 100 (85) cases out of 100 for the double peak model and in 98 (86) cases for the triple peak model, whereas back- and forecasting led to no, or only to minor, improvements both for the nontapered Yule–Walker estimate and for the Burg algorithm.

Tapered Whittle estimates can be used for arbitrary parametric models, whereas, e.g., the Burg algorithm is a special algorithm for autoregressive processes. The Whittle estimate can be calculated, e.g., by a Newton–Raphson procedure. In the special case of autoregressive processes we obtain the estimate from the Yule–Walker equations where the covariances are calculated with tapered data. In this case we have the important Durbin–Levinson algorithm [Durbin (1960)] for these estimates that remains the same as in the nontapered case since we only have to modify the data.

We finally remark that the methods of this paper can also be applied to nonparametric estimates. In fact, most of the results are essentially of a nonparametric nature (e.g., Theorems 7.3 and A1.2).

APPENDIX 1

A central limit theorem. This appendix contains the technical details used in Section 6. We first list the properties of the function $L_T(\alpha)$.

LEMMA A1.1. *Let $k \in \mathbb{N}$ and $\alpha, \beta, \gamma, \nu, \mu \in \mathbb{R}$. 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) *$\int_{\Pi} L_T(\alpha)^k \, d\alpha \le KT^{k-1}$ for all $k > 1$.*
(c) *$\int_{\Pi} L_T(\alpha) \, d\alpha \le K \log T$.*
(d) *$\pi^{-1} \le L_T(\alpha)$.*
(e) *$|\alpha| L_T(\alpha) \le K$.*
(f) *$L_T(\nu)^k L_T(\mu)^l \le L_T((\nu - \mu)/2)^k L_T(\mu)^l + L_T(\nu)^k L_T((\nu - \mu)/2)^l$.*
(g) *$L_T(c\alpha) \le K_c L_T(\alpha)$.*
(h) *$\int_{\Pi} L_T(\gamma + \alpha) L_T(\beta - \alpha) \, d\alpha \le K L_T(\beta + \gamma) \log T$.*
(i) *$L_T(\alpha) \le K L_T(\beta)$ for all α, β with $|\alpha - \beta| \le 2\pi/T$.*

PROOF. The proofs are straightforward. Some of them may be found in Dahlhaus (1983). □

We now consider sequences of processes $X_t^{(T)} \in \mathscr{X}(f_2^*, f_4^*, C, T, s_1, s_2)$ and prove a central limit theorem for the statistic $J_T(\phi)$.

There exist numerous central limit theorems for $J_T(\phi)$ in the literature, proved under different conditions on ϕ, f_2 and the process X_t [see, e.g., Ibragimov (1963), Brillinger (1969), Dunsmuir (1979), Hosoya and Taniguchi (1982) and Dahlhaus (1983)]. However, no results exist if the process $X_t^{(T)}$ depends on T. As we will show, the degree of the data taper plays the central role for the validity of such a central limit theorem.

THEOREM A1.2. *Let $X_t^{(T)} \in \mathscr{X}(f_2^*, f_4^*, C, T, s_1, s_2)$ be a sequence of stationary processes with $f_2^* \in \mathrm{Lip}_{\kappa_0}$ and*

$$\text{(A.1)} \quad |f_4^*(\alpha_1, \alpha_2, \alpha_3) - f_4^*(\beta_1, \beta_2, \beta_3)| \le K \sum_{j=1}^{4} |\alpha_j - \beta_j|^{\kappa_0}, \qquad \kappa_0 \in (0,1],$$

where $\alpha_4 = -\alpha_1 - \alpha_2 - \alpha_3$ and $\beta_4 = -\beta_1 - \beta_2 - \beta_3$. Suppose that the data taper used is of degree (k, κ) with $k \ge \max\{s_1 + s_2 - 1, s_1, s_2\}$ and $\kappa < \min\{1/12, \kappa_0/2 - 1/4\}$. Let further $\phi_1, \dots, \phi_l$ be bounded functions. If $\lim_{T \to \infty} T H_{4,T}/H_{2,T}^2$ exists, then

$$\left\{\sqrt{T}\left(J_T(\phi_j) - J(\phi_j)\right)\right\}_{j=1,\dots,l}$$

tends weakly to a Gaussian random vector ξ *with mean* 0 *and covariance*

$$\operatorname{cov}(\xi_i, \xi_j) = 2\pi\left\{\lim_{T\to\infty} \frac{TH_{4,T}}{H_{2,T}^2}\right\}$$

$$\times\left\{\int_{\Pi}\phi_i(\alpha)\left(\overline{\phi_j(\alpha)} + \overline{\phi_j(-\alpha)}\right)d\alpha \right. \tag{A.2}$$

$$\left. + \int_{\Pi^2}\phi_i(\alpha_1)\overline{\phi_j(-\alpha_2)}\frac{f_4^*(\alpha_1, -\alpha_1, \alpha_2)}{f_2^*(\alpha_1)f_2^*(\alpha_2)}\,d\alpha\right\}.$$

Proof. The theorem is proved by a cumulant method [cf. Brillinger (1981), Lemma P4.5]. The assertion follows from Lemmas A1.3(b), A1.4(b) and A1.5. □

Lemma A1.3. *Let* $X_t^{(T)} \in \mathscr{X}(f_2^*, f_4^*, C, T, s_1, s_2)$ *be a sequence of processes with* $f_2^* \in \mathrm{Lip}_{\kappa_0}$, $\kappa_0 \in (0,1]$. *Suppose that the data taper is of degree* (k, κ).

(a) *If* $k \geq \max\{s_1 + s_2 - 1, s_1, s_2\}$ *and* $\kappa < \kappa_0/2$, *then*

$$\lim_{T\to\infty} EJ_T(\phi) - J(\phi) = 0.$$

(b) *If* $k \geq \max\{s_1 + s_2 - 1, s_1, s_2\}$ *and* $\kappa < \kappa_0/2 - 1/4$, *then*

$$\lim_{T\to\infty} \sqrt{T}\left(EJ_T(\phi) - J(\phi)\right) = 0.$$

Proof. (a) We have

$$EJ_T(\phi) - J(\phi)$$

$$= \{2\pi H_{2,T}\}^{-1}\int_{\Pi^2}\left\{\frac{f_{2,T}(\alpha+\beta)}{f_{2,T}(\alpha)} - 1\right\}\phi(\alpha)|H_T(\beta)|^2\,d\alpha\,d\beta. \tag{A.3}$$

Since

$$|H_T(\beta)|/H_{2,T}^{1/2} \leq KL_T(\beta)^{k+1}T^{-(k+1/2-\kappa)}$$

(Lemma 4.4), we obtain

$$|EJ_T(\phi) - J(\phi)| \leq K\int_{\Pi^2}\left|\frac{f_{2,T}(\alpha+\beta) - f_{2,T}(\alpha)}{f_{2,T}(\alpha)}\right|\frac{L_T(\beta)^{2k+2}}{T^{2k+1-2\kappa}}\,d\alpha\,d\beta. \tag{A.4}$$

Furthermore, if $f_{2,T}(\alpha) = g_{2,T}(\alpha)f_2^*(\alpha)$, (2.2) implies

$$\frac{f_{2,T}(\alpha+\beta) - f_{2,T}(\alpha)}{f_{2,T}(\alpha)} = \left(\frac{f_2^*(\alpha+\beta)}{f_2^*(\alpha)} - 1\right) + \left(\frac{g_{2,T}(\alpha+\beta)}{g_{2,T}(\alpha)} - 1\right)$$

$$+ \left(\frac{f_2^*(\alpha+\beta)}{f_2^*(\alpha)} - 1\right)\left(\frac{g_{2,T}(\alpha+\beta)}{g_{2,T}(\alpha)} - 1\right).$$

By using $f_2^* \in \mathrm{Lip}_{\kappa_0}$ and (2.1), we obtain that (A.4) is less than

$$K\int_{\Pi^2}\{b(T,\alpha,\beta,\lambda,\mu,2s_1,2s_2) + |\beta|^{\kappa_0}\}\frac{L_T(\beta)^{2k+2}}{T^{2k+1-2\kappa}}\,d\alpha\,d\beta$$

$$\leq K\sum_{\substack{j_i=1\\ i=1,2}}^{r_i}\sum_{\substack{k_i=0\\ k_1+k_2\geq 1}}^{2s_i}\int_{\Pi^2}T^{-(2k+1-2\kappa)}L_T(\lambda_{j_1}+\alpha+\beta)^{k_1}L_T(\mu_{j_2}+\alpha)^{k_2}$$

$$\times L_T(\beta)^{2k+2-k_1-k_2}\,d\alpha\,d\beta + KT^{2\kappa-\kappa_0}.$$

Using $\pi^{-1} \leq L_T(\beta) \leq T$, the first summand is less than

$$KT^{2\kappa-1}\sum_{j_i}\int_{\Pi^2}\{\left[L_T(\lambda_{j_1}+\alpha+\beta)+L_T(\mu_{j_2}+\alpha)\right]L_T(\beta)$$

$$+L_T(\lambda_{j_1}+\alpha+\beta)L_T(\mu_{j_2}+\alpha)\}\,d\alpha\,d\beta,$$

which, by using Lemma A1.1(c) is less than $KT^{2\kappa-1}\log^2 T$.

(b) This follows analogously. □

LEMMA A1.4. *Suppose that the assumptions of Theorem* A1.2 *hold.*

(a) *If* $\kappa < (\kappa_0+1)/4$, *then* $\lim_{T\to\infty}\mathrm{var}\{J_T(\phi)\} = 0$.
(b) *If* $\kappa < \kappa_0/4$, *then* $\lim_{T\to\infty}T\,\mathrm{cov}\{J_T(\phi_i), J_T(\phi_j)\} = \mathrm{cov}\{\xi_i, \xi_j\}$.

PROOF. (b) Cumulant calculations give [cf., e.g., Dahlhaus (1983), (3)]

$$T\,\mathrm{cov}\{J_T(\phi_1), J_T(\phi_2)\}$$

$$\begin{aligned}
&= T\{2\pi H_{2,T}\}^{-2}\Bigg\{\int_{\Pi^2}\phi_1(\alpha_1)\left[\overline{\phi_2(\alpha_2)}+\overline{\phi_2(-\alpha_2)}\right]\int_{\Pi^2}\frac{f_{2,T}(\gamma_1)f_{2,T}(\gamma_2)}{f_{2,T}(\alpha_1)f_{2,T}(\alpha_2)}\\
&\qquad\times H_T(\alpha_1-\gamma_1)H_T(\gamma_1-\alpha_2)H_T(\alpha_2-\gamma_2)\\
&\qquad\times H_T(\gamma_2-\alpha_1)\,d\alpha\,d\gamma\\
&\qquad\times\int_{\Pi^2}\phi_1(\alpha_1)\overline{\phi_2(\alpha_2)}\int_{\Pi^3}\frac{f_{4,T}(\gamma_1,\gamma_2,\gamma_3)}{f_{2,T}(\alpha_1)f_{2,T}(\alpha_2)}\\
&\qquad\times H_T(\alpha_1-\gamma_1)H_T(-\alpha_1-\gamma_2)H_T(\alpha_2-\gamma_3)\\
&\qquad\times H_T(-\alpha_2+\gamma_1+\gamma_2+\gamma_3)\,d\gamma\,d\alpha\Bigg\}.
\end{aligned}\tag{A.5}$$

With $\psi(\alpha_1,\alpha_2) = \phi_1(\alpha_1)[\overline{\phi_2(\alpha_2)}+\overline{\phi_2(-\alpha_2)}]$ and (5.1) the first term is equal to

$$\begin{aligned}
&T\{2\pi H_{2,T}\}^{-2}\int_{\Pi^2}\psi(\alpha_1,\alpha_2)\int_{\Pi^2}H_T(\alpha_1-\gamma_1)H_T(\gamma_1-\alpha_2)H_T(\alpha_2-\gamma_2)\\
&\qquad\times H_T(\gamma_2-\alpha_1)\,d\alpha\,d\gamma + R,
\end{aligned}\tag{A.6}$$

with

$$
(\text{A.7})\qquad
\begin{aligned}
|R| \le KT^{-4k-1+4\kappa}\int_{\Pi^4}\left|\frac{f_{2,T}(\gamma_1)f_{2,T}(\gamma_2)}{f_{2,T}(\alpha_1)f_{2,T}(\alpha_2)}-1\right| \\
\times\{L_T(\alpha_1-\gamma_1)L_T(\gamma_1-\alpha_2)L_T(\alpha_2-\gamma_2) \\
\times L_T(\gamma_2-\alpha_1)\}^{k+1}\,d\alpha\,d\gamma.
\end{aligned}
$$

Using (2.1) and (2.2), we obtain

$$
(\text{A.8})\qquad
\begin{aligned}
&\left|\frac{f_{2,T}(\gamma_1)f_{2,T}(\gamma_2)}{f_{2,T}(\alpha_1)f_{2,T}(\alpha_2)}-1\right| \\
&\le \underbrace{C^8\sum_{\substack{M\subset\{1,\dots,4\}\\ M\neq\varnothing}}\prod_{j\in M}\{Cb_j\}}_{\text{(i)}} + \underbrace{\left|\frac{f_2^*(\gamma_1)f_2^*(\gamma_2)}{f_2^*(\alpha_1)f_2^*(\alpha_2)}-1\right|}_{\text{(ii)}},
\end{aligned}
$$

with $b_1 = b(T, \alpha_1, \gamma_1 - \alpha_1, \lambda, \mu, s_1, s_2)$, $b_2 = b(T, \alpha_2, \gamma_1 - \alpha_2, \lambda, \mu, s_1, s_2)$, $b_3 = b(T, \alpha_2, \gamma_2 - \alpha_2, \lambda, \mu, s_1, s_2)$ and $b_4 = b(T, \alpha_1, \gamma_2 - \alpha_1, \lambda, \mu, s_1, s_2)$. We now consider an M of the preceding sum and assume without loss of generality $1 \in M$. We have

$$
\begin{aligned}
&b_1 L_T(\alpha_1-\gamma_1)^{k+1} \\
&\quad\le K\sum_{\substack{j_i=1\\ i=1,2}}^{r_i}\sum_{\substack{k_i=0\\ k_1+k_2\ge 1}}^{s_i} L_T(\lambda_{j_1}+\gamma_1)^{k_1}L_T(\mu_{j_2}+\alpha_1)^{k_2}L_T(\alpha_1-\gamma_1)^{k+1-k_1-k_2}
\end{aligned}
$$

and since $L_T(\alpha) \le T$ and $k \ge \max\{s_1 + s_2 - 1, s_1, s_2\}$,

$$
\le KT^k\sum_{j_i}\{L_T(\lambda_{j_1}+\gamma_1)+L_T(\mu_{j_2}+\alpha_1)\}.
$$

Analogously, we get

$$
\begin{aligned}
&L_T(\gamma_1-\alpha_2)^{k+1}\max\{b_2,1\} \\
&\quad\le KT^k\sum_{j_i}\min\{L_T(\mu_{j_2}+\alpha_2)+L_T(\gamma_1-\alpha_2), L_T(\lambda_{j_1}+\gamma_1)+L_T(\gamma_1-\alpha_2)\}, \\
&L_T(\alpha_2-\gamma_2)^{k+1}\max\{b_3,1\} \\
&\quad\le KT^k\sum_{j_i}\min\{L_T(\lambda_{j_1}+\gamma_2)+L_T(\alpha_2-\gamma_2), L_T(\mu_{j_2}+\alpha_2)+L_T(\alpha_2-\gamma_2)\}
\end{aligned}
$$

and

$$
\begin{aligned}
&L_T(\gamma_2-\alpha_1)^{k+1}\max\{b_4,1\} \\
&\quad\le KT^k\sum_{j_i}\min\{L_T(\mu_{j_2}+\alpha_1)+L_T(\gamma_2-\alpha_1), L_T(\lambda_{j_1}+\gamma_2)+L_T(\gamma_2-\alpha_1)\}.
\end{aligned}
$$

We now obtain with $\int_\Pi L_T(\alpha)\,d\alpha \le K \log T$,

$$\int_{\Pi^4} L_T(\lambda_{j_{11}} + \gamma_1)\{L_T(\mu_{j_{22}} + \alpha_2) + L_T(\gamma_1 - \alpha_2)\}\{L_T(\lambda_{j_{13}} + \gamma_2) + L_T(\alpha_2 - \gamma_2)\}$$
$$\times \{L_T(\mu_{j_{24}} + \alpha_1) + L_T(\gamma_2 - \alpha_1)\}\,d\alpha\,d\gamma \le K \log^4 T$$

and

$$\int_{\Pi^4} L_T(\mu_{j_{21}} + \alpha_1)\{L_T(\lambda_{j_{12}} + \gamma_1) + L_T(\gamma_1 - \alpha_2)\}\{L_T(\mu_{j_{23}} + \alpha_2) + L_T(\alpha_2 - \gamma_2)\}$$
$$\times \{L_T(\lambda_{j_{14}} + \gamma_2) + L_T(\gamma_2 - \alpha_1)\}\,d\alpha\,d\gamma \le K \log^4 T.$$

Thus, we get the upper bound $KT^{4\kappa-1}\log^4 T$ for the term of (A.7) that corresponds to (i). Since (ii) is less than $K|\alpha_1 - \gamma_1|^{\kappa_0} + K|\alpha_2 - \gamma_2|^{\kappa_0}$, the corresponding term of (A.7) is less than $KT^{4\kappa-\kappa_0}\log^4 T$. Thus, (A.6) is equal to

$$TH_{2,T}^{-2}\int_{\Pi^2}\psi(\alpha_1, \alpha_2)\left|H_2^{(T)}(\alpha_1 - \alpha_2)\right|^2 d\alpha_1\,d\alpha_2 + o(1),$$

which tends to

$$2\pi\frac{TH_{4,T}}{H_{2,T}^2}\int_\Pi \psi_T(\alpha, \alpha)\,d\alpha + R,$$

by usual kernel calculations. The second term of (A.5) is treated similarly, which gives the result. Part (a) of the lemma follows analogously. □

LEMMA A1.5. *Let $X_t^{(T)} \in \mathscr{X}(f_2^*, f_4^*, C, T, s_1, s_2)$ be a sequence of stationary processes with $0 < c_1 \le f_2^* \le c_2$ and $\phi_1, \dots, \phi_l$ be bounded functions ($j = 1, \dots, l$). If the data taper is of degree (k, κ) with $k \ge \max\{s_1 + s_2 - 1, s_1, s_2\}$, then we have for $l \ge 3$,*

$$\left|\operatorname{cum}(J_T(\phi_1), \dots, J_T(\phi_l))\right| \le KT^{-l+1+2l\kappa}\log^{2l-1} T.$$

PROOF. We have with $d_T(\alpha) = \sum_{t=0}^{T-1} h_{t,T}X_t^{(T)}\exp(-i\alpha t)$,

$$\left|\operatorname{cum}(J_T(\phi_1), \dots, J_T(\phi_l))\right|$$
$$\text{(A.9)}\qquad \le KH_{2,T}^{-l}\left|\int_{\Pi^l}\left\{\prod_{j=1}^{l}\frac{\phi_j(\alpha_j)}{f_{2,T}(\alpha_j)}\right\}\right.$$
$$\left.\times \operatorname{cum}\{d_T(\alpha_1)d_T(-\alpha_1), \dots, d_T(\alpha_l)d_T(-\alpha_l)\}\,d\alpha\right|.$$

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

$$\begin{matrix} c_1 & d_1 \\ \vdots & \vdots \\ c_l & d_l \end{matrix},$$

where c_i and d_i stand for the position of $d_T(\alpha_i)$ and $d_T(-\alpha_i)$, respectively. This sum will be denoted by $\Sigma_{\text{i.p.}}$. The elements of a set P_i from such a partition shall be assumed to be in a fixed order. If $P_i = \{a_1, \dots, a_n\}$ we set $\bar{P}_i := \{a_1, \dots, a_{n-1}),$ $\beta_{\bar{P}_i} := (\beta_{a_1}, \dots, \beta_{a_{n-1}})$, $\beta_{a_n} = -\Sigma_{j=1}^{n-1}\beta_{a_j}$ and $f_{n,T}(\beta_{\bar{P}_i}) = f_{n,T}(\beta_{a_1}, \dots, \beta_{a_{n-1}})$. Let further m be the size of the corresponding partition and $\beta = (\beta_{\bar{P}_1}, \dots, \beta_{\bar{P}_m})$. Using the product theorem for cumulants and these notation, we get as an upper bound of (A.9)

$$KH_{2,T}^{-l} \sum_{\text{i.p.}} \left| \int_{\Pi^l} \left\{ \prod_{j=1}^{l} \frac{\phi_j(\alpha_j)}{f_{2,T}(\alpha_j)} \right\} \int_{\Pi^{2l-m}} \left\{ \prod_{k=1}^{m} f_{|P_k|,T}\left(\beta_{\bar{P}_k}\right) \right\} \right.$$
$$\text{(A.10)} \qquad \left. \times \left\{ \prod_{j=1}^{l} H_T\left(\alpha_j - \beta_{c_j}\right) H_T\left(-\alpha_j - \beta_{d_j}\right) \right\} d\beta\, d\alpha \right|.$$

If $P_k = \{a_1, \dots, a_n\}$ we obtain from Definition 2.2

$$\frac{f_{|P_k|,T}\left(\beta_{a_1}, \dots, \beta_{a_{n-1}}\right)}{\prod_{j=1}^{n} f_{2,T}(\gamma_j)^{1/2}}$$
$$\leq K \left| \prod_{j=1}^{n} \frac{A_T\left(\beta_{a_j}\right)}{A_T(\gamma_j)} \right| \leq K \sum_{M \subset \{1, \dots, n\}} \prod_{j \in M} b\left(T, \gamma_j, \beta_{a_j} - \gamma_j, \lambda, \mu, s_1, s_2\right),$$

where $\beta_{a_n} = -\Sigma_{j=1}^{n-1}\beta_{a_j}$. Therefore, we obtain by a suitable arrangement of the $f_{2,T}(\alpha_j)$,

$$\frac{\prod_{k=1}^{m} \left| f_{|P_k|,T}\left(\beta_{\bar{P}_k}\right) \right|}{\prod_{j=1}^{l} f_{2,T}(\alpha_j)} \leq K \sum_{M \subset \{1, \dots, 2l\}} \prod_{j \in M} b_j,$$

with $b_j = b(T, \alpha_j, \beta_{c_j} - \alpha_j, \lambda, \mu, s_1, s_2)$ and $b_{j+l} = b(T, -\alpha_j, \beta_{d_j} + \alpha_j, \lambda, \mu, s_1, s_2)$, $j = 1, \dots, l$. Using this and the generalized Hölder inequality, we obtain as an upper bound of (A.10),

$$KT^{-l-2l(k-\kappa)} \sum_{\text{i.p.}} \sum_{M \subset \{1, \dots, 2l\}} \int_{\Pi^l} \int_{\Pi^{2l-m}} \left\{ \prod_{j \in M} b_j \right\}$$
$$\text{(A.11)} \qquad \times \left\{ \prod_{j=1}^{l} L_T\left(\alpha_j - \beta_{c_j}\right)^{k+1} L_T\left(-\alpha_j - \beta_{d_j}\right)^{k+1} \right\} d\beta\, d\alpha.$$

The special structure of each partition is expressed in the special structure of the corresponding β. Every α_j and every $\beta_a (a \in \cup_{k=1}^{m} P_k)$ is contained in $\prod_{j=1}^{l} L_T(\alpha_j - \beta_{c_j})^{k+1} L_T(-\alpha_j - \beta_{d_j})^{k+1}$ exactly twice as an argument, once with positive sign and once with negative sign. Consider a partial product $\prod_{i \in A} L_T(\alpha_i - \beta_{c_i})^{k+1} \prod_{j \in B} L_T(-\alpha_j - \beta_{d_j})^{k+1}$. Because of the indecomposability of the corresponding partition, we see that there exists at least one argument $\gamma \in \{\alpha_j | j \in \{1, \dots, l\}\} \cup \{\beta_a | a \in \cup_{k=1}^{m} \bar{P}_k\}$ that is contained exactly once in the partial sum. As in the proof of Lemma A1.4 we now have

$$
\begin{aligned}
&L_T(\alpha_j - \beta_{c_j})^{k+1} \max\{b_j, 1\} \\
\text{(A.12)} \qquad &\leq KT^k \sum_{\substack{j_i=1 \\ i=1,2}}^{r_i} \min\{L_T(\lambda_{j_1} + \beta_{c_j}) + L_T(\alpha_j - \beta_{c_j}), L_T(\mu_{j_2} + \alpha_j) \\
&\qquad\qquad + L_T(\alpha_j - \beta_{c_j})\}
\end{aligned}
$$

and

$$
\begin{aligned}
&L_T(-\alpha_j - \beta_{d_j})^{k+1} \max\{b_{j+l}, 1\} \\
&\leq KT^k \sum_{\substack{j_i=1 \\ i=1,2}}^{r_i} \min\{L_T(\lambda_{j_1} + \beta_{d_j}) + L_T(-\alpha_j - \beta_{d_j}), L_T(\mu_{j_2} - \alpha_j) \\
&\qquad\qquad + L_T(-\alpha_j - \beta_{d_j})\}.
\end{aligned}
$$

We now calculate an upper bound of (A.11) for a fixed partition and a fixed M. We have $L_T(\alpha_1 - \beta_{c_1})^{k+1}\max\{b_1, 1\} \leq KT^{k+1}$. We know that there exists an argument that is contained in $\prod_{i\neq 1} L_T(\alpha_i - \beta_{c_i})^{k+1} \prod_{j=1}^{l} L_T(-\alpha_j - \beta_{d_j})^{k+1}$ exactly once as an argument (in this case β_{c_1} or, if β_{c_1} is of the form $\beta_{c_1} = -\sum_{j=1}^{n-1}\beta_{a_j}$, one of the β_{a_j} involved). Because of Lemma A.11 integration in (A.11) over this argument gives an additional factor $KT^k \log T$. Proceeding like this, we obtain as an upper bound of (A.9)

$$KT^{-l-2l(k-\kappa)}T^{2kl+1}\log^{2l-1} T,$$

which is the result. □

In Theorem A1.2 we have proved that a data taper of degree (k, κ) with $k \geq \max\{s_1 + s_2 - 1, s_1, s_2\}$ and suitable κ is sufficient for a central limit theorem of $J_T(\phi)$. Our results of Section 7 (cf. the proof of Theorem 7.1) prove for $(s_1 > 0,\ s_2 \leq 1)$ and $(s_1 \leq 1,\ s_2 > 0)$ the necessity of the preceding degree.

APPENDIX 2

Matrix norms and Toeplitz matrices. In this appendix we briefly summarize some results on matrix norms and Toeplitz matrices needed in Section 3.

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

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

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

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

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

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

LEMMA A2.1. *Let A,B be $n \times n$ matrices. Then*

(a) $|\operatorname{tr}(AB)| \leq |A|\,|B|$,
(b) $|AB| \leq \|A\|\,|B|$,
(c) $|AB| \leq |A|\,\|B\|$,
(d) $\|A\| \leq |A| < \sqrt{n}\|A\|$,
(e) $\|AB\| \leq \|A\|\,\|B\|$,
(f) $\|A\| = \|A^*\|$,
(g) $|\operatorname{tr}(A)| \leq \sqrt{n}\,|A|$,
(h) $|x^*Ax| \leq x^*x\|A\|$, $x \in \mathbb{C}^n$,
(i) $\log\det A \leq \operatorname{tr}\{A - I\}$, $A \geq 0$.

Suppose now, that the elements of A are continuously differentiable functions of θ and $\theta_1 < \theta_2$. Then

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

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

(l)
$$|A(\theta_1) - A(\theta_2)| \leq |\theta_1 - \theta_2|\left|\frac{\partial}{\partial\theta}A(\theta)\right|, \quad \textit{with } \theta \in [\theta_1, \theta_2],$$

(m)
$$\|A(\theta_1) - A(\theta_2)\| \leq |\theta_1 - \theta_2|\left\|\frac{\partial}{\partial\theta}A(\theta)\right\|, \quad \textit{with } \theta \in [\theta_1, \theta_2].$$

The following results are proved in Dahlhaus (1986), Lemma 4.1 and Theorem 4.2.

LEMMA A2.2. *Suppose that f and h are real nonnegative, symmetric functions. Suppose further that f is positive on an interval of positive Lebesgue measure. Then*

$$\left\|B_T(f)^{-1/2}B_T(fh)^{1/2}\right\| = \left\|B_T(fh)^{1/2}B_T(f)^{-1/2}\right\| \leq \sup_{\lambda\in\Pi}|h(\lambda)|^{1/2}.$$

LEMMA A2.3. *Let $l \in \mathbb{N}$, $s_1, s_2 \in \mathbb{N}_0$, $f^* \in \operatorname{Lip}_{\kappa_0}$ with a $\kappa_0 \in (0,1]$ and $0 < C^{-2} \leq f^* \leq C^2$ and f_T be a sequence of spectral densities of processes $X_t^{(T)} \in \mathscr{X}(f^*, 0, C, T, s_1, s_2)$. Let further $g_1, \dots, g_l$ be bounded integrable functions, and $h_1, \dots, h_l$ be positive integrable functions that are bounded from* *above and below. Then*

$$\lim_{T\to\infty}\frac{1}{T}\operatorname{tr}\left\{\prod_{j=1}^{l}B_T(f_Tg_j)B_T(f_Th_j)^{-1}\right\}=\frac{1}{2\pi}\int_\Pi\left\{\prod_{j=1}^{l}\frac{g_j(\lambda)}{h_j(\lambda)}\right\}d\lambda\Bigg|.$$

Acknowledgments. This work is part of the author's Habilitationsschrift. It was partially done while the author was a visitor at the Department of Statistics, University of California, Berkeley. The author is grateful to Professor D. R. Brillinger, Berkeley, and Professor P. L. Davies, Essen, for their support.

REFERENCES

BLOOMFIELD, P. (1976). *Fourier Analysis of Time Series: An Introduction.* Wiley, New York.

BRILLINGER, D. R. (1969). Asymptotic properties of spectral estimates of second-order. *Biometrika* **56** 375–390.

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

COURSOL, J. and DACUNHA-CASTELLE, D. (1982). Remarks on the approximation of the likelihood function of a stationary Gaussian process. *Theory Probab. Appl.* **27** 162–167.

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

DAHLHAUS, R. (1984). Parameter estimation of stationary processes with spectra containing strong peaks. *Robust and Nonlinear Time Series Analysis. Lecture Notes in Statist.* **26** 50–86. Springer, New York.

DAHLHAUS, R. (1986). Approximations for the inverse of covariance matrices of stationary processes. Preprint, Universität Essen.

DAHLHAUS, R. and KÜNSCH, H. (1987). Edge effects and efficient parameter estimation for stationary random fields. *Biometrika* **74** 877–882.

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.

DURBIN, J. (1960). The fitting of time series models. *Rev. Inst. Internat. Statist.* **28** 233–244.

DZHAPARIDZE, K. O. and YAGLOM, A. M. (1983). Spectrum parameter estimation in time series analysis. In *Developments in Statistics* (P. R. Krishnaiah, ed.) **4** 1–181. Academic, New York.

FINDLEY, D. F. (1985). On the unbiasedness property of AIC for exact or approximating linear stochastic time series models. *J. Time Ser. Anal.* **6** 229–252.

GRAYBILL, F. A. (1983). *Matrices with Applications in Statistics*, 2nd ed. Wadsworth, Belmont, Calif.

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

HANNAN, E. J. (1970). *Multiple Time Series.* Wiley, New York.

HAYKIN, S. and KESLER, S. (1979). Prediction-error filtering and maximum-entropy spectral estimation. In *Nonlinear Methods of Spectral Analysis. Topics in Applied Physics* (S. Haykin, ed.) **34**. Springer, Berlin.

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.

IBRAGIMOV, I. A. (1963). On estimation of the spectral function of a stationary Gaussian process. *Theory Probab. Appl.* **8** 366–401.

LJUNG, L. and CAINES, P. (1979). Asymptotic normality of prediction error estimates or approximate system models. *Stochastics* **3** 29–45.

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. (1981). *Spectral Analysis and Time Series* **1**. Academic, London.

PUKKILA, T. (1979). The bias in periodogram ordinates and the estimation of ARMA-models in the frequency domain. *Austral. J. Statist.* **21** 121–128.

ROSENBLATT, M. (1985). *Stationary Sequences and Random Fields*. Birkhäuser, Boston.

SHAMAN, P. (1975). An approximate inverse for the covariance matrix of moving average and autoregressive processes. *Ann. Statist.* **3** 532–538.

SHAMAN, P. (1976). Approximations for stationary covariance matrices and their inverses with application to ARMA models. *Ann. Statist.* **4** 292–301.

TUKEY, J. W. (1967). An introduction to the calculations of numerical spectrum analysis. In *Advanced Seminar on Spectral Analysis of Time Series* (B. Harris, ed.) 25–46. Wiley, New York.

ULRICH, T. J. and BISHOP, T. N. (1975). Maximum entropy spectral analysis and autoregressive decomposition. *Rev. Geophys.* **13** 183–200.

WALKER, A. M. (1964). Asymptotic properties of least-squares estimates of parameters of the spectrum of a stationary non-deterministic time-series. *J. Austral. Math. Soc.* **4** 363–384.

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. Appendix to *A Study in the Analysis of Stationary Time Series*, by H. Wold, 2nd ed. 196–228. Almqvist and Wiksell, Uppsala.

ZHURBENKO, I. G. (1980). On the efficiency of spectral density estimates of a stationary process. *Theory Probab. Appl.* **25** 466–480.

FACHBEREICH 6, MATHEMATIK
UNIVERSITÄT GESAMTHOCHSCHULE ESSEN
POSTFACH 103 764
4300 ESSEN 1
WEST GERMANY