

Consistency for non-linear functions of the periodogram of tapered data

By Daniel Janas and Rainer von Sachs

Institut für Angewandte Mathematik, Universität Heidelberg and

Fachbereich Mathematik, Universität Kaiserslautern

Abstract

In this paper we investigate the merits of using a data taper in non-linear functionals of the periodogram of a stationary time series. To this end, we show consistency for a general class of statistics of the form $\int_{-\pi}^{\pi} \Lambda(\omega) \zeta(I_T(\omega)) d\omega$, where $\Lambda(\omega)$ is a function of bounded variation and where ζ is allowed to be a non-linear function of the periodogram $I_T(\omega)$ of the *tapered* data. The key step in deriving our asymptotic results is an Edgeworth expansion for the finite Fourier transform of the tapered data, which do not have to follow a particular distribution (i.e., we allow for non-Gaussianity). Important applications are estimation of $\int_{-\pi}^{\pi} \Lambda(\omega) g(f(\omega)) d\omega$, choosing ζ to be a suitable transform of a given function g (see Taniguchi, 1980), the peak-insensitive spectrum estimator of von Sachs (1992), where ζ is chosen to be a bounded (robustifying) Ψ -function, and the parametric approach of Chiu (1990) on robust estimation of the parameters of the continuous spectrum of the time series.

Keywords: EDGEWORTH EXPANSION; PERIODOGRAM; DATA-TAPER; NON-LINEAR FUNCTIONALS; PEAK-INSENSITIVE SPECTRAL ESTIMATOR.

MSC Subject Classification 1991: Primary 62M10; Secondary 62E20.

1. Introduction

There is quite a number of important problems which demand for replacing a *linear* functional of the periodogram of a stationary time series by a *non-linear* one (in estimating the spectrum, autocovariance or related functionals): Either this is motivated by the goal of estimating itself a non-linear function of the spectral density (cf. Taniguchi, 1980). Or it is rather the situation of the data sampled from an underlying model which demands for some modification of the pure periodogram (as in the peak-insensitive approaches of Chiu, 1990, and von Sachs, 1992, i.e. in the situation of outliers in the frequency domain).

However, one will hardly find the introduction of the use of *data tapers* into the mentioned problems, though using a taper in periodogram-based estimation problems is a well-known remedy to reduce leakage effects (see, e.g., Tukey, 1967, Bloomfield, 1976, Dahlhaus, 1983). The difficulty in deriving both asymptotic and finite-sample properties of these non-linear functionals *with* the use of data tapers lies in the changing correlation structure of the tapered periodogram, which, up to now, was successfully solved only for linear functionals (see Brillinger, 1981; Dahlhaus, 1985, 1990; Janas, 1993).

In this paper, we provide both the theoretical and practical background for using data tapers. To this end, we show consistency of a general class of these taper-modified non-linear functionals, without assuming a particular underlying distribution of the considered time series (as, e.g., the Gaussian, which facilitates proofs immediately; see the remark to Theorem 3.1). This asymptotic result enables us to investigate how using a taper considerably improves the performance of the mentioned estimating procedures. In particular, this has an important application in the context of detecting periodic components in a stationary time series (see Chen, 1988, and von Sachs, 1993): We show

how replacing classical (non-consistent and biased) *linear* spectrum estimators by (consistent and asymptotically unbiased) peak-insensitive (i.e. *non-linear*) estimates gains testing power, especially in situations where the detecting procedure would fail without use of a data taper.

In the following preliminaries we give the global assumptions needed throughout the whole work. The main results are collected in section 3, i.e. we give our main theorem on consistency and all of the applications yielded. Section 4 deals with the key step in deriving this theorem, i.e. Edgeworth expansion of finite Fourier transforms of tapered data. In Section 5 we present some simulations, and, as a particular striking example, we demonstrate the improvements of the taper-modified peak-insensitive estimator of von Sachs (1992) in the problem of detecting periodic components. The last section, finally, gives the proofs.

2. Preliminaries

We gather the assumptions needed in this paper:

(A1) $\{X_t\}_{t \in \mathbf{Z}}$ is a real-valued linear process such that $X_t = \sum_{u \in \mathbf{Z}} a_u \varepsilon_{t-u}$, where ε_t are i.i.d. random variables satisfying $\mathbf{E} \varepsilon_1 = 0$, $\mathbf{E} \varepsilon_1^2 = 1$, $\mathbf{E} \varepsilon_1^{s+1} < \infty$ for some fixed $s \geq 3$.

(A2) ε_1 fulfills Cramér's condition, i.e.

$$\exists \delta > 0, d > 0 \quad \forall |t| > d \quad |\mathbf{E} \exp(it\varepsilon_1)| \leq 1 - \delta.$$

(A3) The filter coefficients a_u decrease exponentially, i.e.

$$\exists 0 < \rho < 1 \quad \forall \text{ large } u \quad |a_u| < \rho^{|u|}.$$

(A4) $h: \mathbf{R} \rightarrow [0,1]$ denotes a *data taper* with bounded variation,

$$h(x) = 0 \text{ for } x \notin (0,1) \text{ and } H_2 := \int_0^1 h^2(x) dx > 0 .$$

Given a sample X_1, \dots, X_T of size T , let $(h_t \cdot X_t)_{t=1, \dots, T}$ denote the sample based on the tapered data where $h_t = h(t/T)$, $t = 1, \dots, T$.

We consider the periodogram of $(h_t \cdot X_t)_{t=1, \dots, T}$, i.e.

$$I_T(\omega) = (H_{2,T})^{-1} \cdot d_T(\omega) d_T(-\omega), \quad \omega \in \Pi := (-\pi, \pi), \quad (2.1)$$

where $d_T(\omega) = \sum_{t=1}^T h_t \cdot X_t e^{-i\omega t}$ is the finite Fourier transform of the tapered data and where $H_{2,T} = \sum_{t=1}^T h_t^2$ denotes the appropriate norming factor (with $H_{2,T} \sim T H_2$).

Let $f(\omega)$, $\omega \in \Pi$, denote the spectral density of X .

Finally, we assume

(A5) $\sum = \lim_{T \rightarrow \infty} D((H_{2,T})^{-1/2} (d_T(\omega_1), \dots, d_T(\omega_d))')$ is positive definite for fixed $d \in \mathbf{N}$,

where D denotes the dispersion matrix.

3. Main Results

3.1 The main theorem

Given a sample X_1, \dots, X_T of size T , consider the following kind of statistics

$$H_T := \int_{-\pi}^{\pi} \Lambda(\omega) \zeta(I_T(\omega)) d\omega. \quad (3.1)$$

In (3.1) assume that

(B1) $\Lambda(\omega)$, $\omega \in \Pi$, is a real-valued function of bounded variation.

Let $M_r(m) := \sup_x (1 + \|x\|)^{-r} |m(x)|$ ($r \in \mathbf{N}$) for any measurable function $m: \mathbf{R} \rightarrow \mathbf{R}$. (3.2)

Assume further that

(B2) ζ is a measurable function with $M_{s_0}(\zeta^2) < \infty$, where s_0 is s or $(s - 1)$ according to s is even or odd.

Note that unlike in most of the classical problems ζ may be a non-linear function!

Of course, the choice of ζ depends on the kind of problem considered (cf. Introduction): If one wants to estimate the theoretical counterpart H of the quantity H_T , e.g., a certain (smooth) function g of the spectral density $f(\omega)$, then ζ has to be chosen dependent on g (as it is the case in our first application below, in Theorem 3.2).

If, however, the choice of ζ is motivated by the sampling situation, i.e. the underlying model (an outlier model, e.g.), then the form of H as asymptotic limit (in the mean) of $H_T, T \rightarrow \infty$, determines the form of ζ (as it is in Theorem 3.6 below). Hence, we merely define H to be the limit in the mean of H_T , i.e.

$$H := \lim_{T \rightarrow \infty} \mathbf{E} H_T, \quad (3.3)$$

and do not specify it further, at this place.

With this definition (3.3) the main theorem of this work states as follows:

Theorem 3.1: *Assume (A1) – (A5), (B1) and (B2). Then, for all $\varepsilon > 0$,*

$$P \{ |H_T - H| > \varepsilon \} \rightarrow 0 \quad \text{as } T \rightarrow \infty,$$

i.e. H_T converges weakly to H , as $T \rightarrow \infty$.

The proof is mainly a consequence of an Edgeworth expansion of the finite Fourier transform of the tapered data X_t . As this is a result of its own interest, it is derived in section 4. How this is used to show Theorem 3.1 can be found in the appendix.

Remark: Assuming Gaussianity of the time series X_t facilitates proofs considerably, because, in this case, the p.d.f of $(H_{2,T})^{-1/2} \cdot d_T(\omega)$ is the same as its asymptotic limit, i.e. Gaussian (all of the asymptotics is in the elements of the covariance matrix of $(H_{2,T})^{-1/2} \cdot d_T(\omega)$, which still is not straightforward to handle, cf. the comments introducing Theorem 4.3). In this case, under slightly more stringent moment conditions, it is possible to derive asymptotic normality of the statistic H_T (see von Sachs, 1992, Theorem 3.2).

Let us replace the integral H_T by the respective sum over the appropriate grid frequencies of the interval Π , the so-called Fourier frequencies $\omega_k = \frac{2\pi k}{T}$, $k = -N, \dots, N$, $N = [T/2]$. This makes no difference for our result as the error of approximation is always of smaller order than the convergence considered in Theorem 3.1.

We now give a list of important consequences of this main result, where the following two theorems and their corollaries deal with a nonparametric set up:

3.2. Application: Estimating a function of the spectral density

First we consider the following situation: Let $\zeta(x) = L^{-1}\{g(1/t) 1/t\}_{\{x\}}$ for an arbitrary function g , such that $\zeta(x)$ fulfills Assumption (B2), where $L^{-1}\{G(u)\}_{\{x\}}$ denotes the Laplace inverse transform of $G(u)$ at argument x (see Taniguchi, 1980). Note that

$$L\{F(x)\}_{\{u\}} = \int_0^{\infty} F(x) \exp\{-u x\} dx$$

denotes the Laplace transform of $F(x)$ at argument u , whereas the inverse writes as

$$L^{-1}\{G(u)\}_{\{x\}} = (2\pi i)^{-1} \int_{\sigma-i\infty}^{\sigma+i\infty} G(u) \exp\{u x\} du ,$$

where σ is greater than the abscissa of absolute convergence.

With this, $H = \int \Lambda(\omega) g(f(\omega)) d\omega$, and we have the following

Theorem 3.2: *Under the assumptions of Theorem 3.1*

$$H_T = \int_{-\pi}^{\pi} \Lambda(\omega) L^{-1}\{g(1/t) 1/t\}\{I_T(\omega)\} d\omega \text{ converges weakly to}$$

$$H = \int_{-\pi}^{\pi} \Lambda(\omega) g(f(\omega)) d\omega, \text{ as } T \rightarrow \infty.$$

Note that (A1) and (A3) imply that the spectral density $f(\omega)$ is bounded away from above and below. Furthermore, we observe, for comparison only, that with (A1), (A3) and (B2) the assumptions (1 – 3) of Taniguchi (1980) are either fulfilled or weakened as far as we are concerned with proving consistency results (by a completely different technique, not restricted to the case of a Gaussian time series)! Note also that unlike Taniguchi we do not restrict to $\Lambda(\omega)$ being continuous; in particular we consider $\Lambda(\omega) = \chi_{(0,\lambda)}(\omega)$, for fixed $0 < \lambda < \pi$, as a choice of interest.

Theorem 3.2 has a couple of corollaries for particular choices of g :

Corollary 3.3: *Let $g(x) = x^r$, $0 < r < \infty$. If the number s_0 in (B2) obeys $s_0 \geq 4r$, then Theorem 3.2 provides a consistent estimate of a functional of the r -th power of the spectral density with $L^{-1}\{g(1/t) 1/t\}\{u\} = u^r / \Gamma(r+1)$.*

The next corollary provides an estimate of the prediction error variance. The following additional assumption on the characteristic function of ε_1 is needed:

(B3) For some integer $p > 0$, $\int |\mathbf{E} \exp(it\varepsilon_1)|^p dt < \infty$.

Corollary 3.4: *Let $g(x) = \log x$. Then, under the additional assumption (B3),*

$\int \Lambda(\omega) \log\{\alpha I_T(\omega)\} d\omega$ *is a consistent estimate of* $\int \Lambda(\omega) \log\{f(\omega)\} d\omega$, *where* $\alpha := \exp \gamma$ *(with Euler's constant* $\gamma = 0.57721\dots$ *).*

This has important consequences: In situations which are governed by strong leakage, the estimation of the prediction error variance, which is proportional to $\int \log\{f(\omega)\} d\omega$, is heavily biased, if one does without using a data taper.

Corollary 3.5: *Let $g(x) = 1/(x+\varepsilon)$, for some $\varepsilon > 0$. Then, with $L^{-1}\{g(1/t)1/t\}\{u\} = \varepsilon^{-1} \exp(-u/\varepsilon)$ a consistent estimate of the "almost-inverse" of the spectral density is provided (which might be useful for deriving MA-type spectral estimates).*

3.3. Application: Nonparametric peak-insensitive spectrum estimation

Secondly, we turn to the situation of peak-insensitive kernel spectral estimation, as introduced in von Sachs (1992):

There we are dealing with a nonparametric spectral estimator $f_T(\alpha)$, which is defined as the root, pointwise in $\alpha \in \mathbb{I}$, of the following equation in $s > 0$

$$H_T(\alpha, s) := T^{-1} \sum_{k=-N}^N K_b(\alpha - \omega_k) \Psi\left(\frac{I_T(\omega_k)}{s} - 1\right) = 0, \quad (3.4)$$

where $K_b(\vartheta) := b^{-1} K(\vartheta/b)$ with a smooth kernel function K with compact support, with smoothing parameter (bandwidth) $b = b_T \rightarrow 0$ and $Tb \rightarrow \infty$, as $T \rightarrow \infty$, and where Ψ is some smooth bounded function (motivated by robust M -estimation in the frequency domain), with $\int_0^\infty \Psi(x-1) e^{-x} dx = 0$ to ensure asymptotic unbiasedness of the resulting estimator $f_T(\alpha)$. Note that (3.4) derives as a modification of the classical kernel spectral estimator (smoothed periodogram) of the form,

$$\hat{f}_T(\alpha) = T^{-1} \sum_{k=-N}^N K_b(\alpha - \omega_k) I_T(\omega_k) \quad (3.5)$$

(choose $\Psi(x) = x$ to end up with an estimator being asymptotically equivalent to (3.5)).

For the following we need an additional set of assumptions (which correspond to the respective ones in Theorem 3.1 of von Sachs, 1992, i.e., (A1) and (A5). – Note that (A4) of that theorem implies our Assumption (A5)):

(B4) Ψ is a bounded and Lipschitz-continuous real function with $\Psi(0) = 0$, having bounded derivative Ψ' (except at a finite number of points) with $\Psi'(0) > 0$. (If Ψ is not differentiable at 0, we assume continuity of Ψ' from the right and from the left with $\Psi'(0^+) > 0$ and $\Psi'(0^-) > 0$.) Further let $\int_0^\infty \Psi(x-1) e^{-x} dx = 0$.

(B5) For the bandwidth b_T in (3.4) assume that

$$Tb_T^{2+\varepsilon} \rightarrow \infty \text{ as } T \rightarrow \infty, \text{ for some } \varepsilon > 0.$$

The convergence of $f_T(\alpha)$ to $f(\alpha)$, for a fixed $\alpha \in \Pi$, is studied by the convergence of $H_T(\alpha, s)$ for $s = f(\alpha)$:

To match with our notation, $\Lambda(\omega) = (2\pi)^{-1} K((\alpha-\omega)/b_T)/b_T$ and $\zeta(x) = \Psi(x/f(\alpha) - 1)$

such that, by the above assumption on Ψ , $\lim_{T \rightarrow \infty} \mathbf{E} \zeta(I_T(\alpha)) = 0$. Then, consider the

respective integral version of (3.4) with $s = f(\alpha)$, i.e.

$$H_T = H_T(\alpha) := (2\pi)^{-1} \int_{-\pi}^{\pi} b_T^{-1} K((\alpha-\omega)/b_T) \Psi(I_T(\omega)/f(\alpha) - 1) d\omega$$

with $H = \lim_{T \rightarrow \infty} \mathbf{E} H_T = (2\pi)^{-1} \int K(\beta) d\beta \cdot \int \Psi(x-1) e^{-x} dx = 0$

(note that $K_b(\vartheta)$ is an approximate convolution identity). For details, see von Sachs (1992), Theorem 3.3 for the convergence of $H_T(\alpha, f(\alpha))$ and Theorem 3.1 on the consistency of $f_T(\alpha)$, where, in the proofs, the tapered situation for non-Gaussian data was not covered. This gap is now closed by the following

Theorem 3.6: *Under the assumptions of Theorem 3.1 and (B4) and (B5) the peak-insensitive estimator $f_T(\alpha)$, implicitly defined by (3.4), is a consistent estimator of the spectral density $f(\alpha)$, for all $\alpha \in \Pi$.*

Note that the consistency of $f_T(\alpha)$ is yielded by the weak convergence of H_T to H (see the proof in the Appendix), where again the error between sums and integrals is negligible.

Moreover, as important application, Theorem 3.6 holds true for time series data Y_t which are modelled as superposition of X_t with some periodic components S_t , consisting of P periodic components ($P \geq 0$ unknown) with unknown constants A_p and $\lambda_p \neq 0$, and with phases Φ_p uniformly distributed in $(-\pi, \pi)$, i.e., $Y_t = S_t + X_t$, where

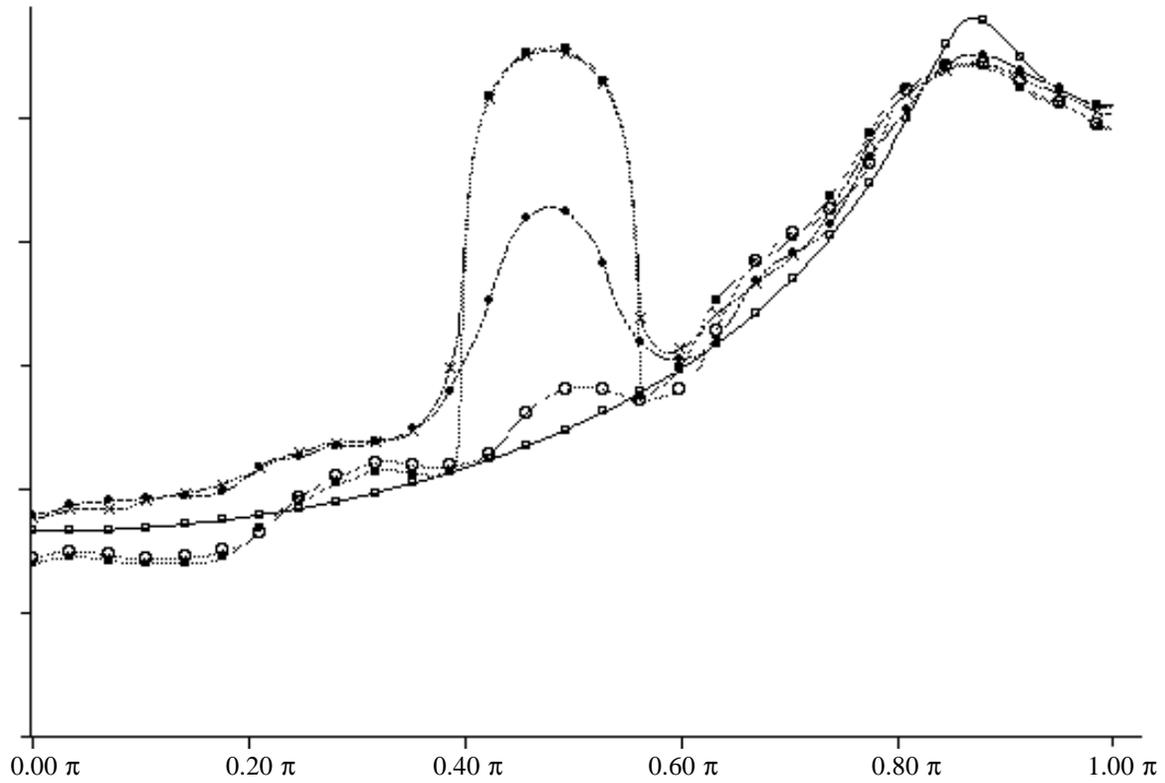
$$S_t = \sum_{p=1}^P A_p \cos\{\lambda_p t + \Phi_p\} . \quad (3.6)$$

I.e., the spectral density f of X can be estimated correctly even at the location of the deterministic frequencies λ_p , i.e. without an asymptotic bias due to the peaks caused by S_t . Note that Assumption (B5) implies the assumption (A5) of Theorem 3.1 of von Sachs (1992) for a model of the form (3.6).

While it is well-known how, for $S = 0$, the performance of classical nonparametric spectral estimators are improved by the use of a taper (see, e.g., Dahlhaus, 1990, Fig. 1), we now want to give an example which is typical for the situation in the presence of periodic components:

Let X_t be an AR[2] – process with a root of radius 0.88 at frequency 0.864π . In the following figures the spectral density $f(\alpha)$ of X_t is shown as 'true spectrum'. As periodic signal we add $S_t = A \cdot \cos(0.48\pi \cdot t)$ with $A = 4.0$ and compare the performance of the classical kernel estimator $\hat{f}_T(\alpha)$ defined by (3.5) with the peak-insensitive $f_T(\alpha)$, where we use a monotone Ψ -function with cut-off point $c = 1.0$ (see (5.1)). Both estimators are with and without tapered data (using a cosine taper, see (5.2)). While a detailed simulation study is postponed to section 5, we want to present two realizations of the simulation runs, where the sample size is chosen to be $T = 512$ and the bandwidth $b = 0.08$ for all estimators:

Figure 3.1: AR[2] – process with one periodic component at 0.48π ($A = 4$; $T = 512$):

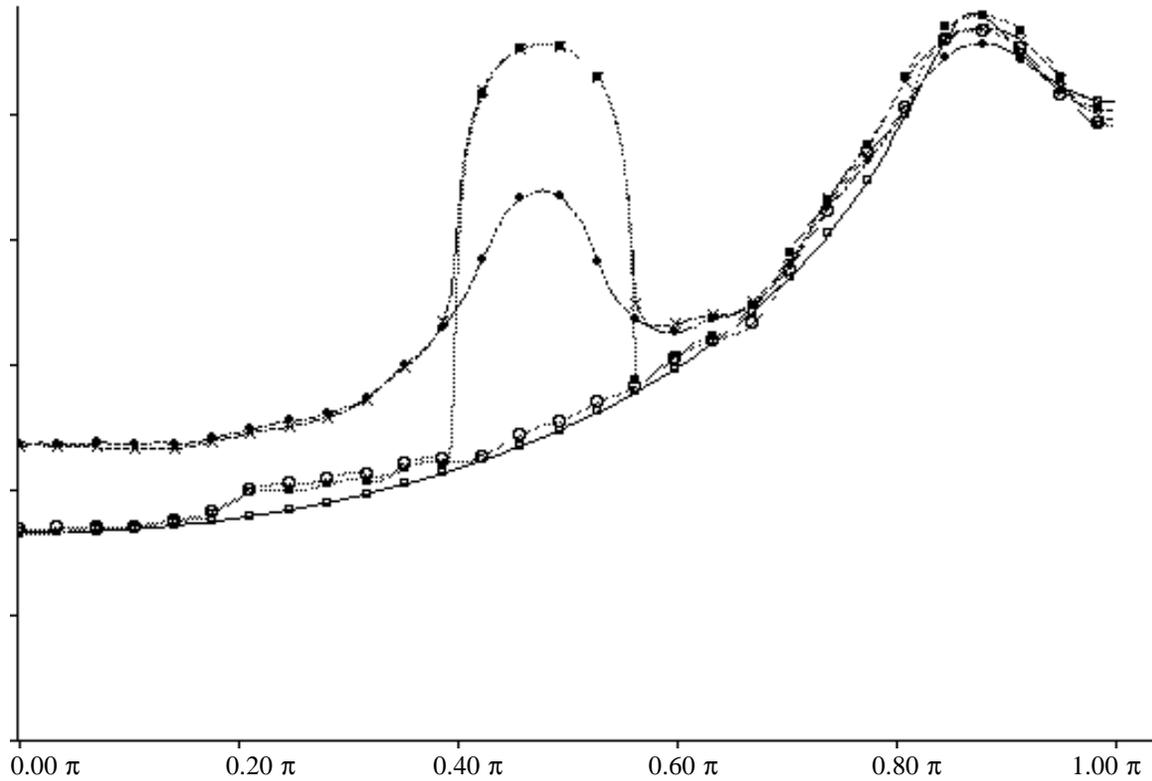


y – axis scaled logarithmically

- true spectrum
- ×— non – robust kernel estimator, non – tapered, $b = 0.08$
- non – robust kernel estimator, 100% – tapered, $b = 0.08$
- robust kernel estimator, non – tapered, monotone Ψ_m , $c = 1.0$, $b = 0.08$
- robust kernel estimator, 100% – tapered, monotone Ψ_m , $c = 1.0$, $b = 0.08$

In the next figure we present an even better performance in a different sample of the same simulation run:

Figure 3.2: same situation as in Figure 3.1, but a different sample of the same simulation run:



y – axis scaled logarithmically

- true spectrum
- ×- non – robust kernel estimator, non – tapered, $b = 0.08$
- non – robust kernel estimator, 100% – tapered, $b = 0.08$
- robust kernel estimator, non – tapered, monotone Ψ_m , $c = 1.0$, $b = 0.08$
- robust kernel estimator, 100% – tapered, monotone Ψ_m , $c = 1.0$, $b = 0.08$

In this second sample of the same simulation run as in Figure 3.1 the 100% –tapered robust estimator nearly performs perfectly: it is completely insensitive to the comparatively strong contamination at frequency 0.48π , whereas in the true spectral autoregressive mode at frequency 0.864π it estimates close to the true spectrum!

It might be interesting also to ask whether using a taper has some influence on the choice of the smoothing parameter b_T : From this example one immediately observes that it can be crucial to be able to choose b_T smaller (without losing robustness!) in order not to underestimate spectral modes of the noise X_t (i.e. not to treat them as outliers in the spectral domain). Heuristically, the leakage of the peaks is reduced by using a data-taper. Indeed, it can be shown theoretically, by some algebra, that, with a taper of degree $k \geq 0$ (see Dahlhaus, 1988, Def. 5.1, where $k = 0$ denotes the nontapered case), we end up with a consistent estimator $f_T(\alpha)$ in the presence of periodic components of type (3.6) if the bandwidth b_T fulfills the following condition :

$$Tb_T^{1+[(1+\varepsilon)/(2k+1)]} \rightarrow \infty \quad \text{as } T \rightarrow \infty, \text{ for some } \varepsilon > 0. \quad (3.7)$$

A proof of (3.7) is delivered by Lemma 3.11 of von Sachs, 1992.

With (3.7) we have the justification for imposing Assumption (B5) in Theorem 3.6.

For non-tapered data b_T should be at least of order $T^{-1/2+\varepsilon}$, where (3.7) allows for a smaller bandwidth if using a taper (i.e., $k \geq 1$), and, in principle, for a smaller bandwidth the higher k is.

So this is another important aspect why one should use data tapers, in particular in the context of peak-insensitive, i.e. *non-linear*, spectral estimators.

The considerations for the peak-insensitive estimator $f_T(\alpha)$ have an important impact on the use of nonparametric spectral estimators in combination with a detecting procedure for hidden frequencies (like that of Chen, 1988). Due to the danger of overestimating the spectrum of X_t at the location λ_p of the peaks, there might be a substantial loss of power of the resulting test procedure which can be overcome by using a peak-insensitive estimator (for details, see von Sachs, 1993). However, in the situation of spectral leakage (e.g., if one spectral line is masked by a nearby second one of stronger signal-to-noise ratio), it is often necessary to use a taper to detect all periodicities! We will examine this very situation in the section 5 on simulations.

3.4. Application: Parametric peak-insensitive spectral estimation

A third field of applications, which we only want to mention rather than study in detail is a *parametric* one, namely peak-insensitively estimating a parametric spectral density $f(\alpha, \theta_0)$, $\theta_0 \in \Theta$ (or its related functionals) as considered in Chiu (1990): In a set up somewhat similar to the one of von Sachs (1992), i.e. with model (3.6), Chiu studies three robustified modifications of traditional estimators. All of them are based on replacing the pure periodogram by a modified one of the form $\tilde{I}_T(\alpha) := \rho \cdot g\{I_T(\alpha) / f(\alpha, \tilde{\theta})\}$, where $\tilde{\theta}$ is an estimate for θ_0 , and where ρ is an unbiased factor to make the means of the modified periodograms approximately equal to $f(\alpha)$. First, Chiu (in his Theorem 2) considers a modified sample autocovariance function from which one can obtain an estimate of θ_0 by the method of moments (such as Yule–Walker equations). Another modification (Theorem 3) is the one for the 'approximate' maximum-likelihood estimate (also known as Whittle estimate), which, in its tapered version, is covered by Janas (1993). A third one, finally, deals with an estimate which minimizes a weighted sum of squares of deviations of the (modified) periodogram and the spectrum (see Chiu, Theorem 4). Note that in contrast to our second application in Theorem 3.6 (i.e. the work of von Sachs, 1992) Chiu's estimators are iterative ones. So they heavily depend on the robustness properties of the respective initial estimates.

Chiu's Theorems 1 – 4 deal more generally with asymptotic normality in case of a Gaussian non-tapered time series. Our Theorem 3.1 shows that the *consistency* assertions of these theorems continue to hold for the tapered, not necessarily Gaussian, case!

Finally we would like to mention that, in general for *non-linear* functionals, it is still an open problem how to prove asymptotic normality for *non-Gaussian* time series data:

Regardless to the use of a taper, the proofs (as in von Sachs, 1992) only work with Gaussian data as they heavily depend on the use of the normal distribution function.

4. Edgeworth expansions

In this section we establish Edgeworth expansions for sums of dependent random vectors using the results of Götze and Hipp (1983) (henceforth referred to as GH). The theory of Edgeworth expansions is our predominant mathematical tool and therefore shall be treated in detail. After giving the general framework we derive expansions for finite Fourier transforms of tapered data.

Let $\{Z_{T,t}\}_{t=1,\dots,T}$ be a triangular array of d -dimensional, real-valued random vectors on an abstract measure space (Ω, \mathbf{A}, P) with $\mathbf{E} Z_{T,t} = 0 \quad \forall t$ and

$$S_T = c_T^{-1/2} \sum_{t=1}^T Z_{T,t} , \quad (4.1)$$

where c_T is a norming constant of order T to be specified. The function $\Psi_{T,s}$ represents the first $(s-1)$ terms of the Edgeworth expansion of the distribution of S_T whenever such an expansion is valid. For any random vector Z , $D(Z)$ denotes the dispersion matrix of Z . Let φ_Σ be the normal density with mean zero and dispersion matrix Σ , and Φ_Σ the corresponding distribution function. c stands for a generic constant. For fixed $r \in \mathbf{N}$ let $f: \mathbf{R}^d \rightarrow \mathbf{R}$ be a measurable function with $M_r(f) = \sup_x (1 + \|x\|)^{-r} |f(x)| < \infty$ (cf. (3.2)).

Define the average modulus of oscillation of f with respect to a finite measure P by

$$\bar{\omega}(f, \varepsilon, P) := \int \sup_{\|y-x\| \leq \varepsilon} |f(y) - f(x)| dP(x). \text{ Now we give the result of GH:}$$

Let \mathbf{D}_j be σ -fields on (Ω, \mathbf{A}, P) (write $\sigma(\bigcup_{j=a}^b \mathbf{D}_j) =: \mathbf{D}_a^b$) and $0 < \rho < 1$ such that

$$(C1) \quad \mathbf{E} Z_{T,t} = 0 \quad \forall t.$$

$$(C2) \quad \mathbf{E} \|Z_{T,t}\|^{s+1} \leq \beta_{s+1} < \infty \quad \forall t \text{ for some } s \geq 3.$$

$$(C3) \quad \exists Y_{T,t,m} \in \mathbf{D}_{t-m}^{t+m} \quad \text{with} \quad \mathbf{E} \|Z_{T,t} - Y_{T,t,m}\| \leq \rho^m.$$

$$(C4) \quad \forall A \in \mathbf{D}_{-\infty}^t, B \in \mathbf{D}_{t+m}^{\infty} \quad |P(A \cap B) - P(A)P(B)| \leq \rho^m.$$

$$(C5) \quad \exists \varepsilon, \eta, \rho > 0 \quad \forall \|\theta\| \geq \varepsilon \quad \forall \rho^{-1} < m < T \\ \# \{t \in \{1, \dots, T\}: \mathbf{E} | \mathbf{E} \exp(i\theta'(Z_{T,t-m} + \dots + Z_{T,t+m})) | \mathbf{D}_j; j \neq t) | \leq 1 - \eta\} \geq \rho T.$$

$$(C6) \quad \forall A \in \mathbf{D}_{t-p}^{t+p} \quad \forall t,p,m \quad \mathbf{E} | P(A | \mathbf{D}_j; j \neq t) - P(A | \mathbf{D}_j; 0 < |j-t| \leq m+p) | \leq \rho^m.$$

$$(C7) \quad \lim_{T \rightarrow \infty} D(S_T) = \Sigma \quad \text{exists and is positive definite.}$$

Remark: The Cramér type condition (C5) is a weaker assumption than the condition (2.5) in GH. Nevertheless, it suffices for the results of GH to hold as is pointed out by remark (3.44) in GH. The weaker condition (C5) means that Cramér's condition is fulfilled for a sufficiently large number of t 's. Whereas condition (2.5) cannot be fulfilled in the situations we will discuss, by some effort it is possible to verify (C5).

Let s_0 be s or $(s-1)$ according to s is even or odd.

Theorem 4.1: *Assume that (C1) – (C7) hold. Then there exists a positive constant δ not depending on f and $M_{s_0}(f)$, and for arbitrary $\kappa > 0$ there exists a positive constant c depending on $M_{s_0}(f)$ but not on f such that*

$$| \mathbf{E} f(S_T) - \int f d\Psi_{T,s} | \leq c \bar{\omega}(f, T^{-\kappa}, \Phi_{\Sigma}) + o(T^{-(s-2+\delta)/2}).$$

The term $o(\cdot)$ depends on f through $M_{s_0}(f)$ only.

Corollary 4.2: *Assume (C1) – (C7). Then the following approximation holds uniformly over convex measurable $C \subseteq \mathbf{R}^d$:*

$$P(S_T \in C) = \Psi_{T,s}(C) + o(T^{-(s-2)/2}).$$

4.1 Expansions for finite Fourier transforms of tapered data

For fixed d and integers $j(1), \dots, j(d)$ in $(0, T/2)$ let

$$d_T(\omega_1, \dots, \omega_d) := \left(\sum_{t=1}^T X_t \exp(-i \omega_k t) \right), \quad k = 1, \dots, d \quad (4.2)$$

be the finite Fourier transform of X_1, \dots, X_T at Fourier frequencies $\omega_k := 2\pi j(k) / T$ (see Brillinger 1981).

Under suitable conditions on the sequence $\{X_t\}$ of random variables the distribution of d_T admits a higher order approximation. Chen and Hannan (1980) have shown the validity of such an expansion when $\{X_t\}$ are i.i.d. (and fulfill Cramér's condition as well as certain moment conditions). GH have generalized their result for strictly stationary Markov-dependent sequences satisfying certain regularity conditions. Here, a generalization of the result of Chen and Hannan (1980) is given for linear processes $\{X_t\}_{t \in \mathbf{Z}}$ with i.i.d. innovations $\{\varepsilon_t\}_{t \in \mathbf{Z}}$. Moreover, we allow the data to be tapered. Whereas the generalization for linear processes is relatively easy to handle by the mathematical tools of GH, tapering destroys the orthogonality relations of sine and cosine functions which causes more trouble.

Theorem 4.3: *Assume (A1) – (A5). Then Theorem 4.1 and Corollary 4.2 hold for*

$$S_T := (H_{2,T})^{-1/2} \sum_{t=1}^T h_t X_t \xi_{T,t},$$

where $\xi_{T,t} := (\cos(\omega_1 t), \dots, \cos(\omega_d t), \sin(\omega_1 t), \dots, \sin(\omega_d t))'$.

The proof is given in the appendix.

Remark. The expansion above allows to compute cumulants for non-linear functions of the periodogram, as, for $d = 1$, $I_T = S_T S_T'$. Thus the consistency results for these statistics are byproducts of Theorem 4.3 (cf. proof of Lemma 6.1).

5. Simulations and applications to detecting periodicities

5.1 A simulation study for the peak-insensitive spectral estimator:

First we want to add a more detailed simulations study of the illustrating example of section 3 on the improvement of the peak-insensitive estimator $f_T(\alpha)$ by using a data taper (see Figures 3.1 and 3.2). To X_t being an AR[2] – process with a root of radius 0.88 at frequency 0.864π , we add $S_t = A \cdot \cos(0.48\pi \cdot t)$ with $A = 4.0$ and compare the performance of the classical kernel estimator $\hat{f}_T(\alpha)$ defined by (3.5) with the peak-insensitive $f_T(\alpha)$, where we use a monotone Ψ -function Ψ_m with cut-off point $c = 1.0$:

$$\Psi_m(x) = \begin{cases} Q_H^{-1} \cdot \max\{-c, x\} & \text{for } x \leq 0 \\ \min\{x, c\} & \text{for } x > 0 \end{cases}, \quad (5.1)$$

where $Q_H := \left| \int_0^1 \Psi_H(x-1) e^{-x} dx / \int_1^\infty \Psi_H(x-1) e^{-x} dx \right|$

and $\Psi_H(x) = \max\{-c, \min\{x, c\}\}$ (see Huber, 1975).

Note that Q_H is used to cope with the asymmetric periodogram distribution, in order to ensure asymptotic unbiasedness of the resulting estimator $f_T(\alpha)$.

We want to compare non-tapered and 100% – tapered versions of both non-robust and robustified kernel estimator, using a so-called 'cosine taper'

$$h_\rho(x) = \begin{cases} (1/2) [1 - \cos(2\pi x/\rho)], & x \in [0, \rho/2], \\ 1, & x \in [\rho/2, 1/2], \\ h_\rho(1-x), & x \in (1/2, 1]. \end{cases} \quad (5.2)$$

where $\rho = 1$ in this case. We do this by using a familiar error criterion, the Mean Integrated Relative Squared Error (MIRSE)

$$\mathbf{E} \left[\int_{-\pi}^{\pi} (\hat{f}_T(\omega; b)/f(\omega) - 1)^2 d\omega \right] \quad (5.3)$$

over the number n of Monte–Carlo replications (simulation runs) for fixed sample size T . (We also give confidence sets belonging to the normal approximation, to roughly indicate significant differences.) The smoothing kernel is chosen to be a Bartlett–Priestley Kernel $K(\vartheta) = 3/(4\pi) \cdot \{1 - (\vartheta/\pi)^2\}$ with compact support $[-\pi, \pi]$.

In the following $n = 50$ simulation runs (which are sufficient to demonstrate the different performance) the sample size is $T = 512$ and the bandwidth $b = 0.08$ for all estimators:

estimator	bandwidth b	MIRSE	confidence sets
non–robust, non–tapered	0.08	4472	[4426; 4517]
non–robust, tapered	0.08	4579	[4520; 4638]
robust Ψ_m , non–tapered	0.08	52.52	[50.23; 54.81]
robust Ψ_m , tapered	0.08	0.147	[0.127; 0.168]

For this comparison see also Figures 3.1 and 3.2 in section 3, which show two different samples out of this simulation run. With both presentations it can be seen clearly that tapering improves the peak–insensitive procedure drastically, where for the usual kernel estimator it has no significant effect apart from reducing leakage outside the peak–contaminated region of the spectral domain (for the price of an increased variance leading to an even higher MIRSE).

5.2 Detecting periodic components:

Now we choose both a simulated example and an interesting data set to demonstrate the ability of our estimator to detect periodic components in situations where it is necessary to use a data taper (as mentioned in the introduction): In the procedure of Chen

(1988) we use the peak-insensitive estimator $f_T(\omega)$ for the noise spectrum $f(\omega)$ as a normalization of the periodogram $I_T^Y(\omega)$. Roughly speaking (for details, see the given reference), this procedure is based on the statistic

$$\hat{z}_j := I_T^Y(\omega_j) / \{\hat{f}_T(\omega_j) \log T\} \quad , \quad j = 1, \dots, T, \quad (5.4)$$

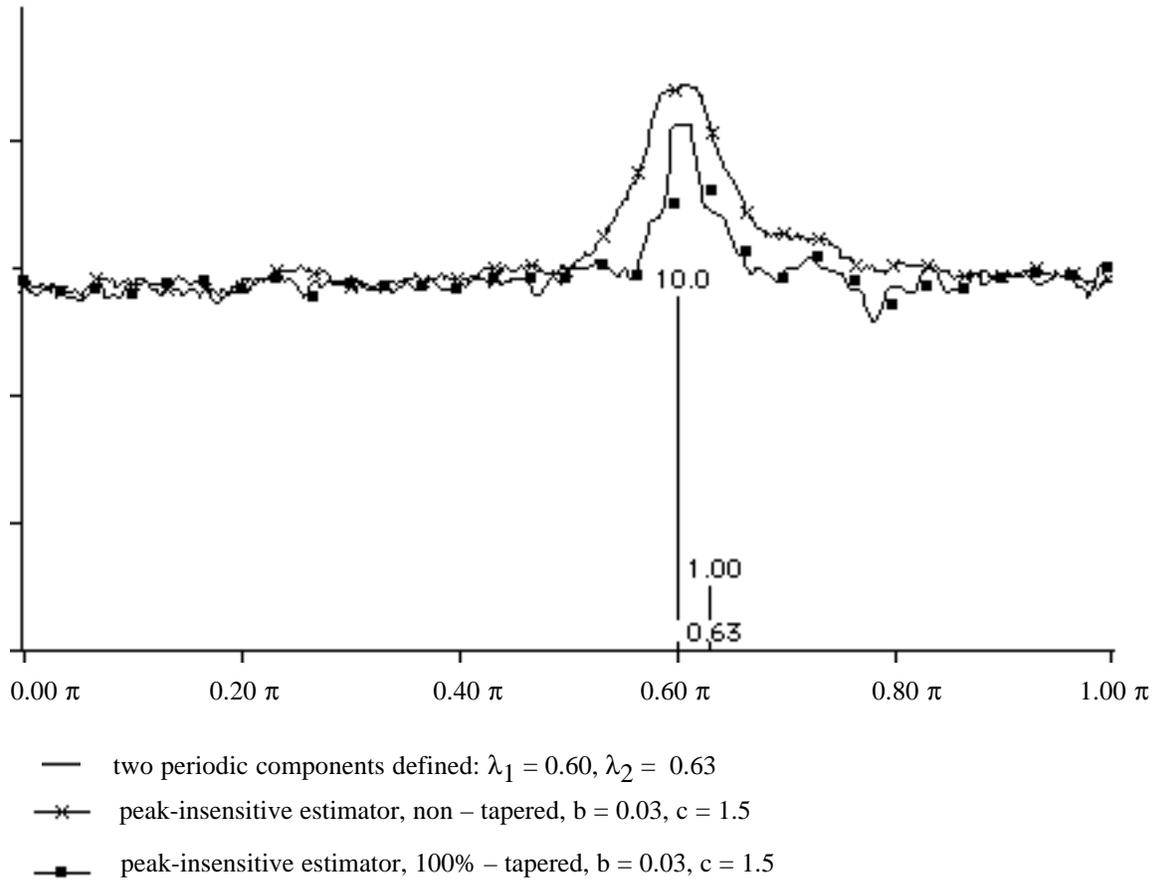
where $\hat{z}_j > 1 + \varepsilon$ should indicate the presence of a periodic component at or close to frequency ω_j (for details on the appropriate choice of $\varepsilon > 0$ see also von Sachs, 1993).

As stated in section 3, choosing the peak-insensitive estimator $f_T(\omega_j)$ will face the danger of losing power of \hat{z}_j at the location of the peaks. Note that in Chen (1988) it was shown that, in contrast to $f_T(\alpha)$, his estimator for the noise spectrum (a kernel estimator of type (3.5), modified by leaving-out a fixed number of periodogram values) is not consistent at the locations of the occurring periodicities. We now want to show that it is not only of theoretical interest to deliver a (non-linear) estimator which is consistent even with the use of a taper.

Our simulated example is a superposition of two periodic components of different signal-to-noise ratio in Gaussian white noise X_t with $T = 512$: one with $A_1 = 10$ and $\lambda_1 = 0.60\pi$, the other closely located at $\lambda_2 = 0.63\pi$ with $A_2 = 1$. Simulations show that the procedure based on the classical estimator $\hat{f}_T(\alpha)$ is not able to detect the small periodicity at λ_2 , regardless to the use of a data taper. However with $f_T(\alpha)$ it is successful ending up with estimates $\hat{\lambda}_1 = 0.598$, $\hat{\lambda}_2 = 0.631$, but only *with* 100% of the data *tapered*. I.e., in any case without using a taper it would not be possible to cope the leakage and unmask the small periodicity!

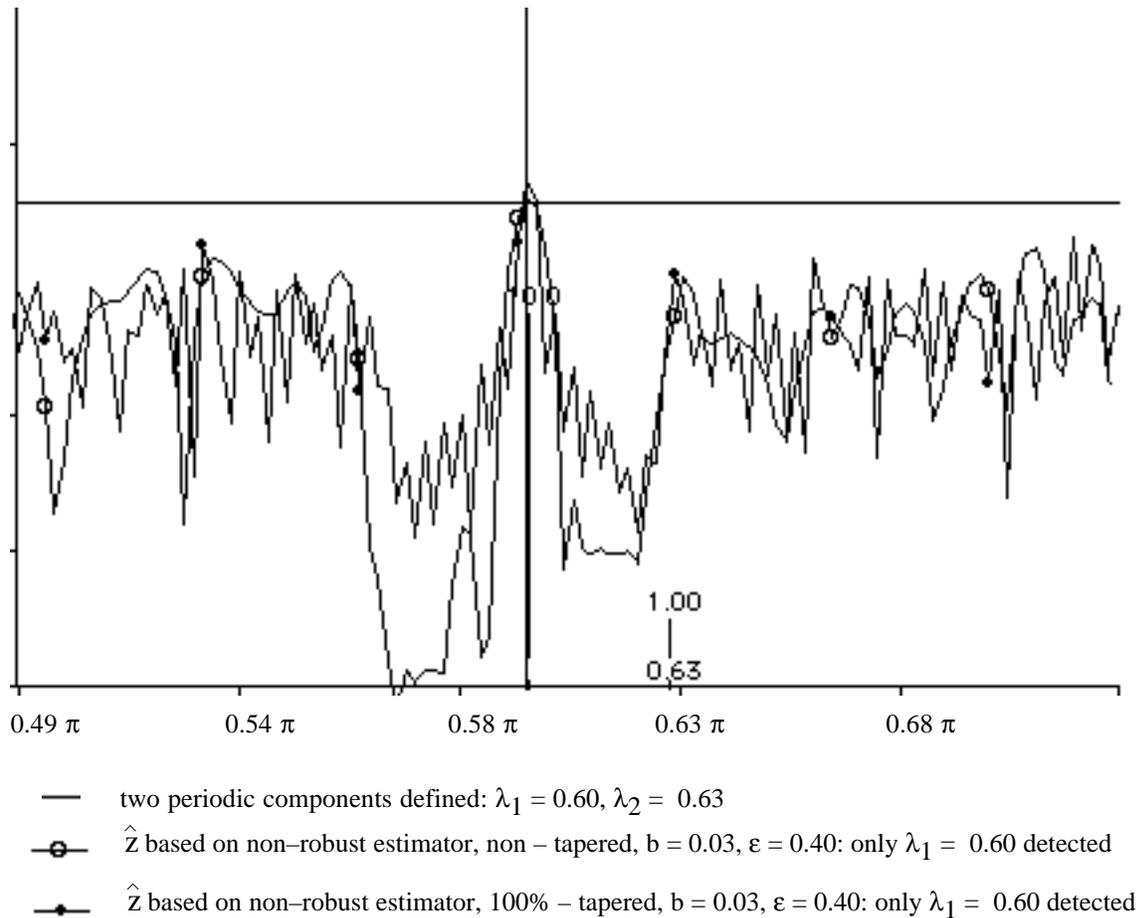
In Figure 5.1 we get an impression of the improvement of $f_T(\alpha)$ by using a taper. Note that the amplitude of the strong periodicity is very high, such that complete insensitivity cannot be expected for finite sample size ($T = 512$).

Figure 5.1: White noise with two periodic components: $A_1 = 10$ at $\lambda_1 = 0.60\pi$,
 $A_2 = 1$ at $\lambda_2 = 0.63\pi$ ($T = 512$):



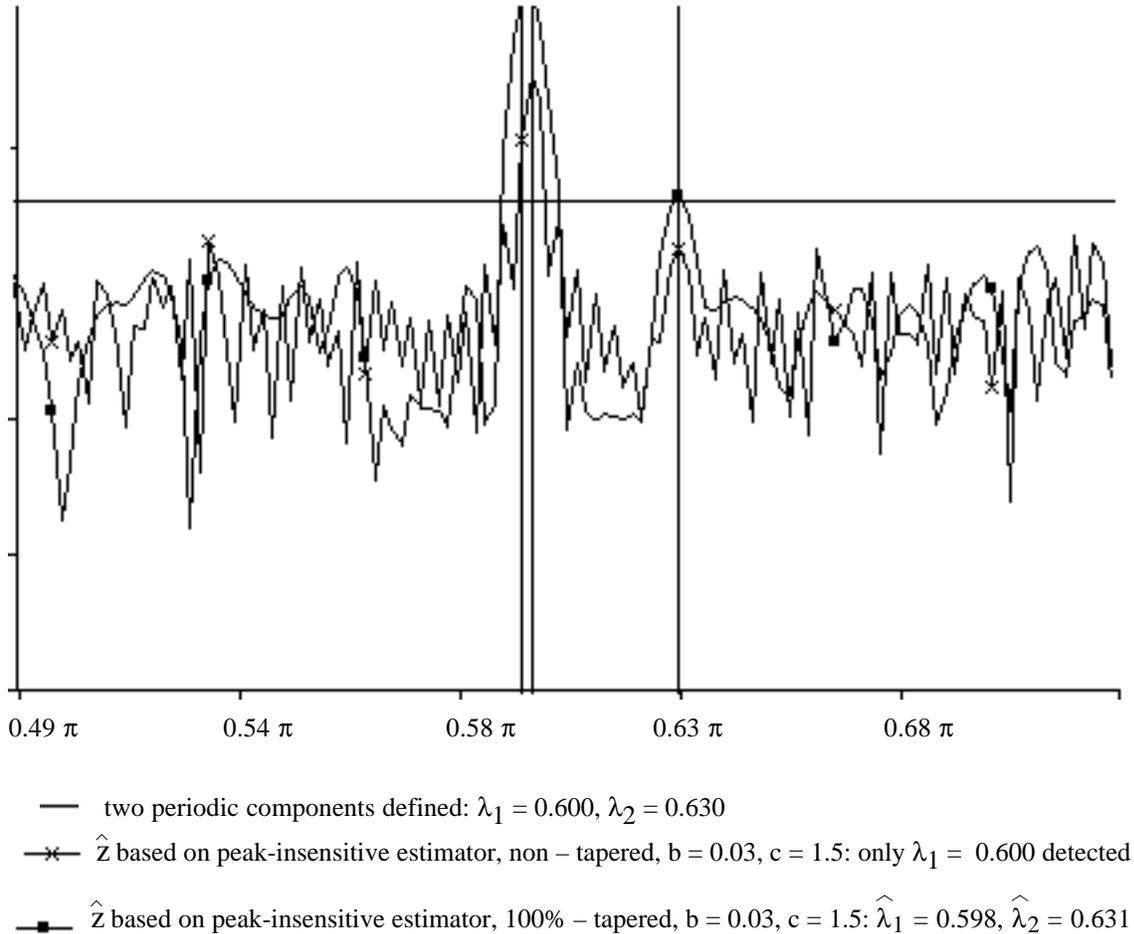
Looking on the window in Figure 5.2 one observes that with the non-robust estimator $\hat{f}_T(\alpha)$ only the strong periodicity can be detected; using a taper does not help at all. Note that the presence of a periodic component is indicated by a vertical line crossing the critical bound $1 + \varepsilon$ at any place where the test statistic \hat{z}_j exceeds this bound, here with $\varepsilon = 0.40$:

Figure 5.2: Detecting procedure based on \hat{z} in a window around 0.6π of Figure 5.1:



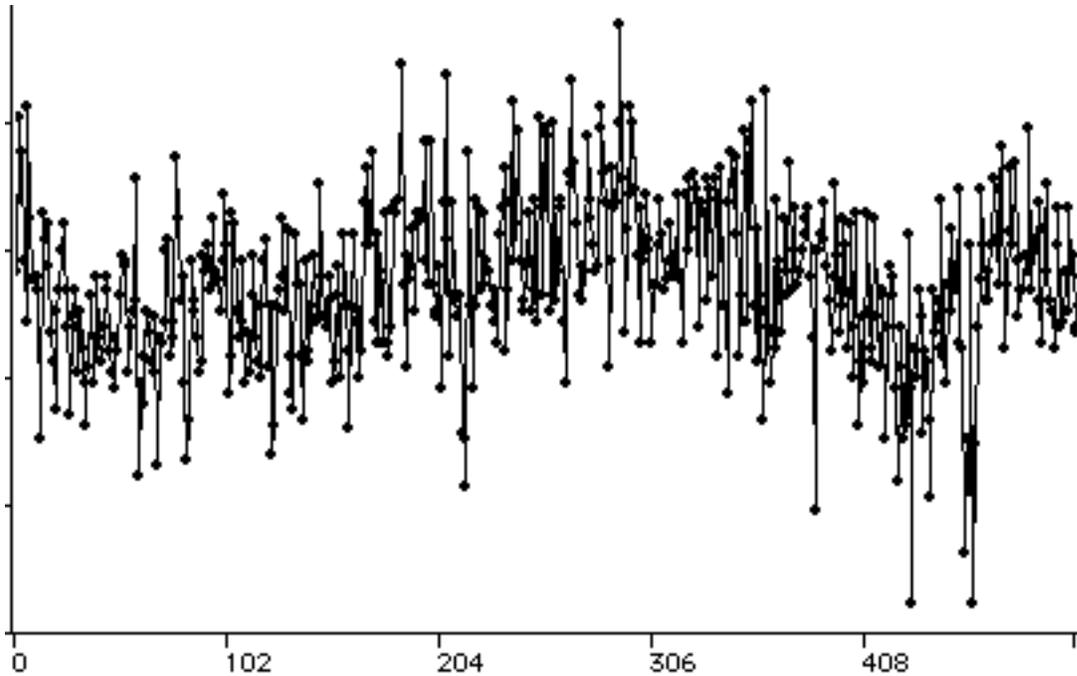
The same window is shown in Figure 5.3 for the detecting procedure based on the peak-insensitive $f_T(\alpha)$: Only *with* the use of a *taper* the test statistic \hat{z} exceeds the critical bound, again chosen to be 1.40, at the location 0.63π of the weak periodicity, too!

Figure 5.3: same situation as in Figure 5.2,
but with \hat{z} based on peak-insensitive estimator:



For the application to a real data-set we choose the following data (which are from Tamar Breus at the Space Research Institute in Moscow and were kindly provided by G. Cornélissen and F. Halberg, University of Minnesota, Minneapolis): It is a segment of length $T = 512$ of a time series of daily collected data, describing the rate of medical infarctions, based on daily ambulance calls in Moscow from Jan 1st, 1979 through Dec 31st, 1981 (i.e., 1096 data points: Note that the procedure is successful with this chosen smaller segment of size 512; there is no difference to the analysis with full data size!):

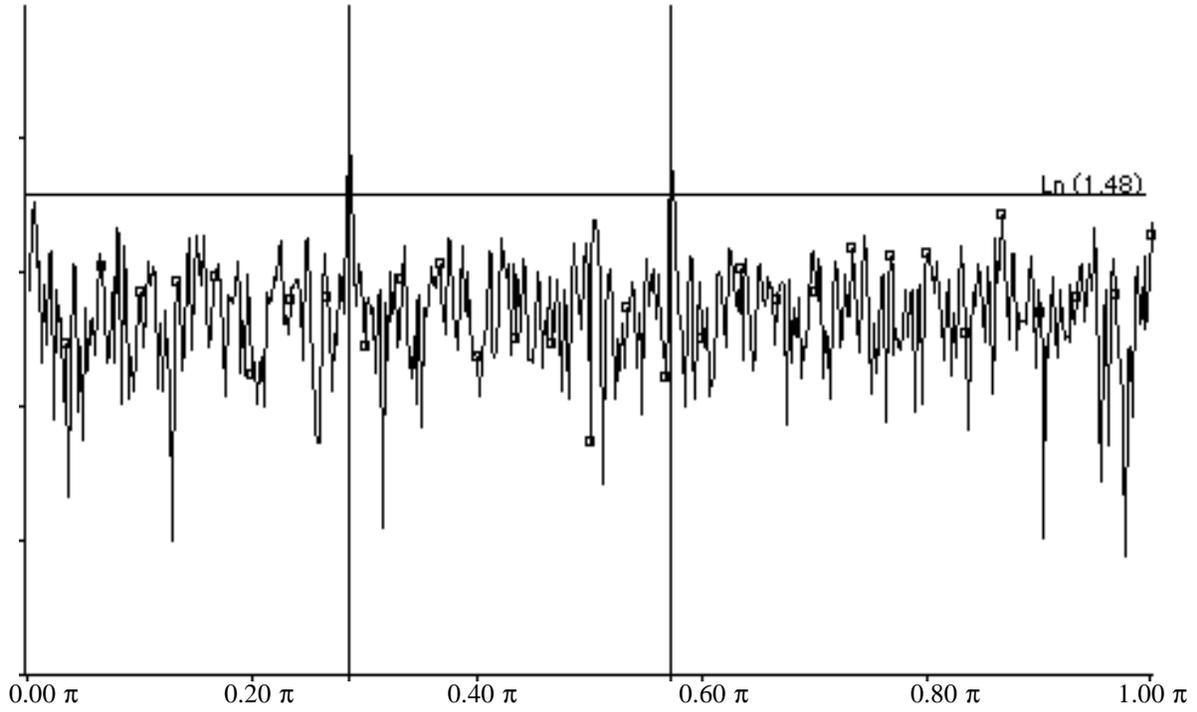
Figure 5.4: Moscow daily infarctions data May 28, 1980 – Oct 21, 1981 (512 data):



From the time series plot one might expect a long-term cycle in addition to the apparent short-term (weekly) periodicity: Obviously, any of the following detecting procedures shows the peak at $0.285\pi = 2\pi/7$, which is indeed due to a weekly periodicity (and also the first harmonic at $0.571\pi = 4\pi/7$). Less obvious, but a rather difficult task, is to detect the yearly cycle at $0.006\pi = 3\pi/512$, which, for this kind of procedure, is only possible with the use of data taper – we compare Figure 5.5 with Figure 5.6, both with $\varepsilon = 0.48$:

Figure 5.5: detection of periodicities in the infarctions data of Figure 5.4:

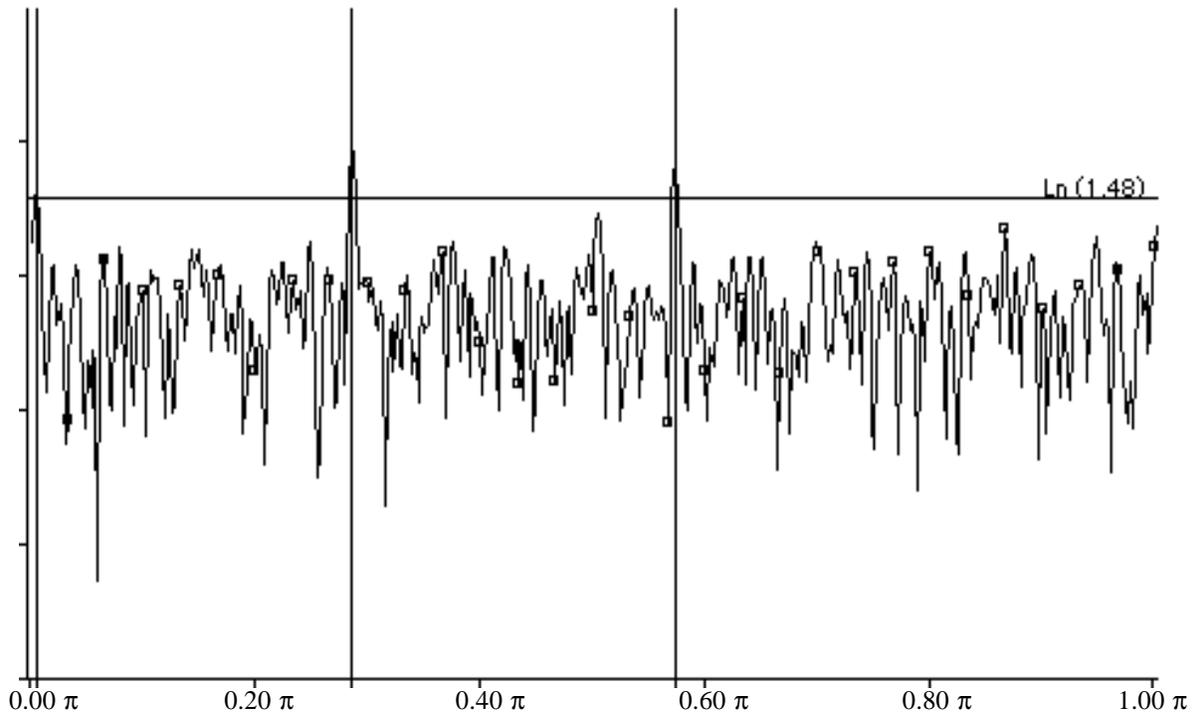
test statistic \hat{z} , based on peak-insensitive estimator, *no taper*, $b = 0.05$, $c = 1.5$:



2 periodicities detected: $0.285\pi = 2\pi/7$ (weekly cycle) and $0.571\pi = 4\pi/7$ (first harmonic)

Figure 5.6: same procedure as in Figure 5.5 but now *with* the use of a data *taper*:

test statistic \hat{z} , based on peak-insensitive estimator, 100% taper, $b = 0.05$, $c = 1.5$



3 periodicities detected: $0.006\pi = 3\pi/512$ (yearly cycle), $0.285\pi = 2\pi/7$ (weekly cycle) and $0.571\pi = 4\pi/7$ (first harmonic).

Note that this analysis is completely in accordance with the results of different detecting procedures (which are not based on Fourier transform analysis).

6. Appendix: Proofs

Proof of Theorem 3.1: With Theorem 4.3, i.e. the Edgeworth expansion of the finite Fourier transform of X_t for the tapered case, this proof runs quite analogously to the proof of Theorem 3.1 of von Sachs (1992), which was for the particular statistic $H_T(\alpha, s)$ as in (3.4), but excluding the tapered case. Note that we replace integrals by sums of Fourier frequencies as the error of approximation is of smaller order than the convergence considered in the theorem. For the reader's convenience we summarize the steps. First, and this is the key step, we use a result similar to Lemma 5.1 of von Sachs (1992):

Lemma 6.1: Let I_T be the periodogram of $\{h_t X_t\}$ as in (2.1), and let Z denote a standard exponentially distributed rv (i.e. with parameter 1). Then

- (i) $\mathbf{E} \zeta(I_T(\omega_j)) = \mathbf{E} \zeta(f(\omega_j) Z) + o(T^{-1/2})$ uniformly in ω_j ,
- (ii) $\text{Var}\{\zeta(I_T(\omega_j))\} = \text{Var}\{\zeta(f(\omega_j) Z)\} + o(T^{-1/2})$ uniformly in ω_j ,
- (iii) $\text{Cov}\{\zeta(I_T(\omega_{j_1})), \zeta(I_T(\omega_{j_2}))\} = o(T^{-1/2})$ uniformly in $\omega_{j_1} \neq \pm \omega_{j_2}$.

Proof:

We only prove (i). By the expansion given in Theorem 4.3 we get, with $s = 3$ and $d = 1$,

as $I_T(\omega_j) = S_T(\omega_j)S_T(\omega_j)'$, where $S_T(\omega_j) = (H_{2,T})^{-1/2} \sum_{t=1}^T h_t X_t (\cos(\omega_j t), \sin(\omega_j t))'$:

$$\mathbf{E} \zeta(I_T(\omega_j)) = \int_{\mathbb{R}^2} \zeta(y_1^2 + y_2^2) d\Psi_{T,3}(y_1, y_2) = \int_0^\infty \zeta(f(\omega_j) x) e^{-x} dx + o(T^{-1/2}).$$

Note that the second order term of the mentioned expansion yields the remainder whereas the first order term cancels due to symmetry arguments: Both ζ and ϕ_{2k} are even functions of y_1 and y_2 , whereas the polynomials occurring in the derivatives of ϕ_{2k} are odd in y_1 or y_2 . This is completely analogous to the proof of Lemma 5.1 of von Sachs (1992).

(ii) and (iii) can be shown quite analogously with $s = 3$ and $d = 2$. □

With the definition (3.3) of $H = \lim_{T \rightarrow \infty} \mathbf{E} H_T$ it is sufficient for proving Theorem 3.1 to

$$\text{show that} \quad H_T - \mathbf{E} H_T \xrightarrow{P} 0 \quad \text{as } T \rightarrow \infty, \quad (6.1)$$

$$\text{which will be implied by} \quad \text{var} \{ H_T \} \rightarrow 0 \quad \text{as } T \rightarrow \infty. \quad (6.2)$$

By Assumption (B2) $\text{Var}\{\zeta(Z)\}$ is finite such that (6.2) follows by Lemma 6.1 (ii), (iii). \square

Proof of Theorem 3.2: With assumption (B2) on $\zeta(x) = L^{-1}\{g(1/t) 1/t\}_{x}$, with $s_0 = 2$, this is an immediate consequence of Theorem 3.1 noting that

$$\mathbf{E} L^{-1}\{g(1/t) 1/t\}_{I_T(\omega_j)} = \mathbf{E} L^{-1}\{g(1/t) 1/t\}_{\{f(\omega_j) Z\}} + o(T^{-1/2}), \quad \text{uniformly in } \omega_j,$$

due to Lemma 6.1 (i), and that $\mathbf{E} L^{-1}\{g(1/t) 1/t\}_{\{f(\omega_j) Z\}} = g(f(\omega_j))$, by the definition of the inverse Laplace transform.

Proof of Corollary 3.4: Like Theorem 3.1 and Theorem 3.2 this result is a consequence of the Edgeworth expansion for the finite Fourier transform. However the singular behaviour of $\log x$ at $x = 0$ causes some trouble. Therefore, Theorem 4.3 is not sufficient. The distribution of the finite Fourier transform has to be absolutely continuous. Under the additional assumption (B3) on the characteristic function of ε_1 this can be proved in the white noise case for tapered data by applying the techniques developed in the proof of Theorem 3.1. The assertion for the general case now follows analogously to the proof of Theorem 1 in Chen and Hannan (1980).

Proof of Theorem 3.6: The proof is analogous to the proof of Theorem 3.1 in von Sachs (1992); we want to give the main ideas: With $H = 0$, Theorem 3.1 yields $H_T = H_T(\alpha, f(\alpha)) \xrightarrow{P} 0$ as $T \rightarrow \infty$. Then, using standard arguments of local monotonicity, as, e.g. in Huber (1964, Lemmata 2 and 3), the weak consistency of $f_T(\alpha)$ can be deduced by the one of H_T .

In extension of Theorem 3.6, consistency in the situation with periodic components $S_t \neq$

0, see (3.6), is completed by Theorem 4.1 of von Sachs (1992), which holds regardless to the use of a data taper. Note that for periodicities as in (3.6) the conditions for this Theorem 4.1 are fulfilled by our Assumption (B5). \square

Proof of Theorem 4.3: We have to check (C1) – (C7) to hold for the process $\{Z_{T,t}\}$. Let $\mathbf{D}_j = \sigma(\varepsilon_j)$. Conditions (C1), (C4) and (C6) hold trivially. (C7) is assumption (A5). By (A1), (C2) follows from

$$\begin{aligned} \mathbf{E} \|Z_{T,t}\|^{s+1} &= \mathbf{E} |h_t X_t|^{s+1} \|\xi_{T,t}\|^{s+1} \\ &\leq \mathbf{E} |X_t|^{s+1} \left(\sum_{j=1}^d (\cos^2(t\omega_j) + \sin^2(t\omega_j)) \right)^{(s+1)/2} \\ &\leq d^{(s+1)/2} \left(\sum_{u \in \mathbf{Z}} |a_u| \right)^{s+1} \mathbf{E} |\varepsilon_1|^{s+1} < \infty. \end{aligned}$$

(C3) can be verified by taking $Y_{T,t,m} \in \mathbf{D}_{t-m}^{t+m}$ defined by

$$Y_{T,t,m} := h_t \left(\sum_{u=-m}^m a_u \varepsilon_{t-u} \right) \xi_{T,t}.$$

$$\mathbf{E} \|Z_{T,t} - Y_{T,t,m}\| = \mathbf{E} |h_t \sum_{|u|>m} a_u \varepsilon_{t-u}| \|\xi_{T,t}\| \leq \sum_{|u|>m} |a_u| \mathbf{E} |\varepsilon_1| d^{1/2} \leq \rho^m$$

by the exponential decay of the coefficients $|a_u|$ and the moment conditions on ε_t .

It remains to check a Cramér type condition. The usual condition (2.5) in GH is not fulfilled in our set up, but we will show the weaker condition (C5).

$$\begin{aligned} \sum_{|t-j|\leq m} Z_{T,j} &= \sum_{|t-j|\leq m} h_j \sum_{u \in \mathbf{Z}} a_{j-u} \varepsilon_u \xi_{T,j} \\ &= \sum_{u \in \mathbf{Z}} \varepsilon_u \sum_{j=-m}^m h_{t+j} a_{t+j-u} \xi_{T,t+j} \\ &= \varepsilon_t A_{T,t,m} + \zeta, \end{aligned}$$

where $A_{T,t,m} := \sum_{j=-m}^m h_{t+j} a_j \xi_{T,t+j}$ and ζ denotes a random vector stochastically independent of ε_t . Thus, with $\theta' := (\theta_1, \tilde{\theta}_1, \theta_2, \tilde{\theta}_2, \dots, \theta_d, \tilde{\theta}_d) \in \mathbf{R}^{2d}$

$$\begin{aligned}
& \mathbf{E} \left| \mathbf{E} \exp(i \theta' \sum_{j=t-m}^{t+m} Z_{T,j} | \mathbf{D}_j; j \neq t) \right| \\
&= \left| \mathbf{E} \exp(i \varepsilon_t \theta' A_{T,t,m} | \cdot \mathbf{E} | \exp(i \theta' \zeta) \right| \\
&\leq 1 - \eta, \quad \eta > 0
\end{aligned}$$

by Cramér's condition on ε_t (i.e. (A2)), if $|\theta' A_{T,t,m}|$ is bounded away from zero. This is verified in the next technical lemma from which the assertion follows. \square

Lemma 6.2: *Assume the conditions of Theorem 4.3. Then*

$$\exists \varepsilon, \rho > 0 \quad \forall \|\theta\| = 1 \quad \forall \rho^{-1} < m < T : \# \{ t \in \{1, \dots, T\} : |\theta' A_{T,t,m}|^2 \geq \varepsilon \} \geq \rho T.$$

Proof: By the Schwarz inequality we obtain the following upper bound

$$\begin{aligned}
|\theta' A_{T,t,m}|^2 &\leq \|\theta\|^2 \left\| \sum_{j=-m}^m h_{t+j} a_j \xi_{T,t+j} \right\|^2 \\
&\leq \left(\sum_{j \in \mathbf{Z}} |a_j| \right)^2 d \leq a, \tag{6.3}
\end{aligned}$$

where a is a positive constant only depending on $\{a_j\}$ and d . Further, assume for a moment that the following lower bound holds for all $\|\theta\| = 1$ and $m < T$ large enough

$$\frac{1}{T} \sum_{t=1}^T |\theta' A_{T,t,m}|^2 \geq b \tag{6.4}$$

with b being a positive constant independent of T and m . Assume w.l.o.g. $a \geq \max(b, 1)$.

Let $c = \frac{b}{a} \leq 1$ and $\rho = \frac{c - \varepsilon}{1 - \varepsilon}$. If less than $\rho \cdot T$ terms had the property $|\theta' A_{T,t,m}|^2 \geq \varepsilon$, we could bound the left-hand side of (6.3) by

$$\frac{1}{T} \sum_{t=1}^T |\theta' A_{T,t,m}|^2 < (1 - \rho) \varepsilon + \rho a \leq a ((1 - \rho) \varepsilon + \rho) = b,$$

which is a contradiction to (6.4). It remains to show the lower bound (6.4).

Since $c_T^{-1} \sum_{t=1}^T |\theta' A_{T,t,m}|^2 = \theta' c_T^{-1} \sum_{t=1}^T A_{T,t,m} A'_{T,t,m} \theta$, it is enough to show that

$$c_T^{-1} \sum_{t=1}^T A_{T,t,m} A'_{T,t,m} \rightarrow \Sigma \quad \text{for } m < T \text{ and } m \rightarrow \infty, \tag{6.5}$$

because Σ is assumed to be positive definite (assumption (A5)). Note that here $c_T = H_{2,T}$.

The left-hand side of (6.5) can be written as a dispersion matrix, i.e.

$$D(c_T^{-1/2} \sum_{t=1}^T \varepsilon_t \sum_{|j| \leq m} h_{t+j} a_j \xi_{T,t+j}) . \quad (6.6)$$

The right-hand side is
$$\lim_{T \rightarrow \infty} D(c_T^{-1/2} \sum_{t=1}^T h_t X_t \xi_{T,t}) . \quad (6.7)$$

We show the convergence of (6.6) to (6.7) for the corresponding complex-valued counterparts, that is we replace the vectors

$$\xi_{T,s} = (\cos(\omega_1 s), \dots, \cos(\omega_d s), \sin(\omega_1 s), \dots, \sin(\omega_d s))' \quad (6.8)$$

by
$$\xi_{T,s}^c = (\exp(-i \omega_1 s), \dots, \exp(-i \omega_d s))' . \quad (6.9)$$

Then the assertion follows, since the real-valued versions (6.6) and (6.7) can be reconstructed, e.g.

$$\Sigma = \frac{1}{2} \begin{pmatrix} \operatorname{Re} \Sigma^c & -\operatorname{Im} \Sigma^c \\ \operatorname{Im} \Sigma^c & \operatorname{Re} \Sigma^c \end{pmatrix}, \quad (6.10)$$

where Σ^c denotes the dispersion matrix corresponding to $\xi_{T,s}^c$ (cf Brillinger, 1981, p. 89).

$$\begin{aligned} & D(c_T^{-1/2} \sum_{t=1}^T \varepsilon_t \sum_{|u| \leq m} h_{t+u} a_u \xi_{T,t+u}^c)_{r,s} \\ &= c_T^{-1} \sum_{t=1}^T \sum_{|u|, |v| \leq m} a_u a_v h_{t+u} h_{t+v} \exp(-i \omega_r (t+u)) \exp(i \omega_s (t+v)) \\ &= c_T^{-1} \sum_{|u|, |v| \leq m} a_u a_v \exp(-i(\omega_r u - \omega_s v)) \sum_{t=1}^T h_{t+u} h_{t+v} \exp(-i(\omega_r - \omega_s) t) \end{aligned} \quad (6.11)$$

By lemma P4.1 in Brillinger (1981) we can substitute the term

$$\sum_{t=1}^T h_{t+u} h_{t+v} \exp(-i(\omega_r - \omega_s) t) \quad (6.12)$$

by $H_{2,T}(\omega_r - \omega_s) = \sum_{t=1}^T h_t^2 \exp(-i(\omega_r - \omega_s) t)$ within the error bound $O(|u| + |v|)$,

which does not cause trouble because of $\sum_{u \in \mathbf{Z}} |u| |a_u| < \infty$.

Thus, continuing in (6.11),

$$\begin{aligned}
& c_T^{-1} H_{2,T}(\omega_r - \omega_s) \sum_{|u|,|v| \leq m} a_u a_v \exp(-i(\omega_r u - \omega_s v)) + O(c_T^{-1}) \\
&= c_T^{-1} H_{2,T}(\omega_r - \omega_s) \sum_{u \in \mathbf{Z}} a_u \exp(-i \omega_r u) \sum_{v \in \mathbf{Z}} a_v \exp(i \omega_s v) + O(c_T^{-1}) \\
&= H_{2,T}(\omega_r - \omega_s) / H_{2,T} \cdot f(\omega_r) + O(c_T^{-1}), \tag{6.13}
\end{aligned}$$

and for $r \neq s$ using the differentiability of the transfer function and the inequality

$$|H_{2,T}(\omega_r - \omega_s)| \leq K |\omega_r - \omega_s|^{-1}, \quad K \text{ being a constant.} \tag{6.14}$$

(see Dahlhaus, 1988, p. 822).

The last expression in (6.11) is the element (r,s) of the matrix in Σ^c (see Brillinger, 1981, lemma 4.3.2.). □

Acknowledgements

We would like to thank Tamar Breus at the Space Research Institute in Moscow for sampling the data example of section 5.2. Moreover we are grateful to G. Cornélissen and F. Halberg for kindly providing these data and also to C. Bingham (all at the University of Minnesota, Minneapolis) for his support analyzing this data example.

References

- BLOOMFIELD, P. (1976). *Fourier Analysis of Time Series*. Wiley, New York.
- BRILLINGER, D.R. (1981). *Time Series: Data Analysis and Theory*. Holden-Day, San Francisco.
- CHEN, Z.-G. and HANNAN, E.J. (1980). The distribution of periodogram ordinates. *J. Time Series Analysis* **1**, 73 – 82.

- CHEN, Z.–G. (1988). An alternative consistent procedure for detecting hidden frequencies. *J. Time Series Analysis* **9**, 301 – 317.
- CHIU, S.–T. (1990). Peak–insensitive parametric spectrum estimation. *Stoch. Proc. Appl.* **35**, 121 – 140.
- DAHLHAUS, R. (1983). Spectral analysis with tapered data. *J. Time Ser. Anal.* **4**, 163 – 175.
- (1985). Asymptotic normality of spectral estimates. *J. Mult. Anal.* **16**, 412 – 431.
- (1988). Small sample effects in time series analysis: A new asymptotic theory and a new estimate. *Ann. Statist.* **16**, 808 – 841.
- (1990). Nonparametric high resolution spectral estimates. *Probab. Th. Rel. Fields* **85**, 147 – 180.
- GÖTZE, F. and HIPPEL, C (1983). Asymptotic expansions for sums of weakly dependent random vectors. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **64**, 211 – 239.
- HUBER, P.J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35**, 73 – 101.
- JANAS, D (1993). Edgeworth expansions for spectral mean estimates with applications to Whittle estimates. Preprint, Universität Heidelberg.
- VON SACHS, R. (1992). Peak–insensitive nonparametric spectrum estimation. To appear in *J. Time Ser. Anal.*
- (1993). Detecting periodic components in a stationary time series by an improved nonparametric procedure. *Proc. of the Int. Conf. Time Series and Appl. Astronomy and Meteorology 1993, Padova, Italy.*
- TANIGUCHI, M. (1980). On estimation of the integrals of certain functions of spectral density. *J. Appl. Prob.* **17**, 73 – 83.
- TUKEY, J.W. (1967). An introduction to the calculations of numerical spectrum analysis. In: *Advanced Seminar on Spectral Analysis of Time Series* (ed. B. Harris), 25 – 46. Wiley, New York.