Reihe Mathematik

Evangelos Ioannidis

On the Asymptotic Behaviour of the Capon Estimator

Verlag Shaker

ON THE ASYMPTOTIC BEHAVIOUR OF THE CAPON ESTIMATOR

Inaugural - Dissertation

zur

Erlangung der Doktorwürde

der

Naturwissenschaftlich - Mathematischen Gesamtfakultät

der

Ruprecht - Karls - Universität

Heidelberg

vorgelegt von

Diplom-Mathematiker Evangelos Ioannidis

aus Athen

1993

Gutachter: Prof. Dr. R. Dahlhaus

PD. Dr. W. Ehm

HEIDELBERG 1993

Reihe Mathematik

Evangelos Ioannidis

On the Asymptotic Behaviour of the Capon Estimator

Verlag Shaker
Aachen 1993

Die Deutsche Bibliothek - CIP-Einheitsaufnahme

Ioannidis, Evangelos:
On the Asymptotic Behaviour of the Capon Estimator / Evangelos Ioannidis. -
Aachen : Shaker, 1993
(Reihe Mathematik)
Zugl.: Heidelberg, Univ., Diss., 1993
ISBN 3-86111-617-0

Printed in Germany

ISBN 3-86111-617-0

Verlag Dr. Chaled Shaker, Hubertusstr. 40, 52064 Aachen
Telefon: 0241 / 406351 - Telefax: 0241 / 406354

Acknowledgement: *We would like to thank Prof. R. Dahlhaus, Dr. W. Ehm and Dr. P. Doukhan whose comments substantially contributed to the contents and the appearance of this thesis.*

CONTENTS

0. INTRODUCTION

0.1 Preliminaries

The subject of this thesis is to study the 'Capon estimator' which is a non-parametric estimator of the spectral density (spectrum) of a stationary stochastic process.

Non-parametric estimators of the spectrum which are quadratic in the observations have been widely used and studied. On the other hand, in the applied sciences, types of non-parametric spectral estimators which are non-quadratic in the observations are also used; these are considered to have better small sample properties than classical periodogram based (quadratic) estimators. Such a non-quadratic estimator was introduced in 1969 by Capon for estimating the wavenumber spectrum of a homogeneous random field. It became known to engineers as a 'High Resolution Estimator'. In 1972 Pisarenko considered a generalization of Capon's method in the context of estimating the continuous spectrum of a univariate time series.

Some existing approaches to analyse the stochastic behaviour of the Capon estimator are based on unrealistic simplifying assumptions. In the present thesis we study this estimator under more realistic assumptions (given in Section 0.6): we prove a Central Limit Theorem, we propose and discuss an automatic selection criterion for its smoothness parameter. A summary of the results can be found in Section 0.7.

Before we introduce the Capon estimator we present some preliminaries concerning a 'classical' non-parametric estimation of the spectrum.

The spectrum f of $\{X_i\}_{i\in Z}$ is a 2π-periodic, symmetric around 0 function, defined as the Fourier transformation of the covariance function $c_u := E(X_t X_{t+u})$, $u \in \mathbf{Z}$. Thus:

$$f(\lambda) := \frac{1}{2\pi}\sum_{u=-\infty}^{\infty} c_u \exp(-i\lambda u)\ ,\ \lambda\in \mathbf{R} \tag{0.1}$$

supposing this limit exists. The spectrum describes the periodicity structure of the time series. Roughly speaking, the time series may be decomposed into a sum of sinusoidal waves of different frequencies: the amplitude of the wave of frequency λ is a stochastic variable with expectation 0 and variance $f(\lambda)$ (for an exact formulation see e.g. Brockwell, Davis (1987) § 4.8: 'The spectral decomposition of a time series').

We suppose that a realization $X_1,\ldots,X_T$ of a zero mean stationary stochastic process $\{X_i\}_{i\in Z}$, $X_i \in \mathbf{R}$ has been observed. Classical non-parametric estimation of the spectrum is based on the periodogram I_T:

$$(0.2) \qquad I_T(\lambda) := (2\pi T)^{-1} \left| \sum_{t=1}^{T} X_t e^{i\lambda t} \right|^2.$$

This is motivated by the fact that the periodogram is the finite Fourier transformation of the empirical covariances $\hat{c}_T(u)$, $u = -(T-1),\ldots,T-1$:

$$(0.3) \qquad \hat{c}_T(u) := T^{-1} \sum_{t=1}^{T-|u|} X_t X_{t+|u|}.$$

The expectation of the periodogram tends to f but its variance does not tend to 0. Consistent spectral estimators are obtained by taking the convolution of the periodogram with a kernel whose bandwidth tends to 0. An alternative smoothing procedure is the Kolmogorov statistic which consists of taking the mean of periodograms based on different segments of the original data (see (0.6) below). Estimators of these two types, which are quadratic in the observations, have been widely used and studied. (e.g. Brillinger (1975), Rosenblatt (1985) and Brockwell, Davis (1987)).

0.2 Introducing the Capon estimator

The Capon estimator is defined as follows:

$$(0.4) \qquad \hat{f}_T(\lambda) \equiv \hat{f}_{d,T}(\lambda) := \frac{d}{2\pi} \left(\bar{b}_\lambda^t \, \hat{\Gamma}_{d,T}^{-1} \, b_\lambda \right)^{-1}, \ \lambda \in [-\pi,\pi],$$

where $\hat{\Gamma}_{d,T}$ (defined in (0.5) below) is an estimator of the d×d covariance matrix $\Gamma = \Gamma_d$ of $(X_1,\ldots,X_d)^t$, $b_\lambda := \{\exp(i\lambda t)\}_{t=0,\ldots,d-1} \in \mathbb{C}^d$ and $d < T/2$ is a smoothing parameter.
(For $b \in \mathbb{C}^d$ we denote by b^t the transposed and by $\bar{b}$ the component-wise conjugation.)

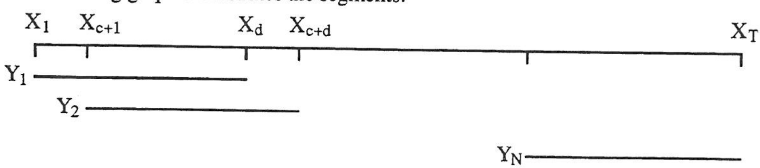
As covariance matrix estimator we use the following 'segment' covariance matrix estimator:

$$(0.5) \qquad \hat{\Gamma}_T \equiv \hat{\Gamma}_{d,T} \equiv \hat{\Gamma}_{d,c,T} := \frac{1}{N} \sum_{i=1}^{N} Y_i Y_i^t, \qquad \text{where}$$

$$N \equiv N_d \equiv N_{d,c,T} := \left[\frac{T-d}{c}\right]+1 \ \text{ and } Y_i \equiv Y_i^{d,c} := (X_{(i-1)c+1},\ldots,X_{(i-1)c+d})^t, \ i=1,\ldots,N$$

(c-displaced segments of length d of the data).

The following graphic illustrates the segments:



Observe that for c=1 we obtain almost fully overlapping segments and for c=d we obtain disjoint segments. This special case, namely the case of $\widehat{\Gamma}_{d,d,T}$ (disjoint segments), was considered in previous works on the Capon estimator (e.g. Pisarenko (1972) and Subba Rao and Gabr (1989)). Note further that $\widehat{\Gamma}_{d,c,T}$ is unbiased and not 'Toeplitz'. This is its main advantage compared to some other estimators which are 'Toeplitz'. (However in Chapter IV we study the Capon estimator also in the case when, in (0.4), instead of $\widehat{\Gamma}_{d,c,T}$ a 'Toeplitz' covariance matrix estimator is used).

Subsequently we also write $\hat{f}_{d,c,T}$ for the Capon estimator when using $\widehat{\Gamma}_{d,c,T}$ as covariance matrix estimator. But we often suppress in our notation the dependence of several quantities on d and c (e.g. $\hat{f}_{d,c,T} \equiv \hat{f}_T$), and of d and c on T, where this does not lead to confusion.

The form of the Capon estimator is surprising since more classical estimators, e.g. the periodogram and the Kolmogorov statistic, may be written as a quadratic form in an estimator of the covariance matrix; on the contrary the Capon estimator is the <u>reciprocal</u> of a quadratic form in the <u>inverse</u> of the estimated covariance matrix.

Let for example the segments over which periodograms are taken in the Kolmogorov statistic to be exactly the Y_i defined in (0.5). Then the Kolmogorov statistic (see also Zhurbenko (1980) and Dahlhaus (1985)) may be written as:

$$I_{d,c,T}(\lambda) := (2\pi d)^{-1} \; \bar{b}_\lambda^t \, \widehat{\Gamma}_{d,c,T} \, b_\lambda \tag{0.6}$$

The periodogram may also be written in this form if one regards $\mathbf{X}^t\mathbf{X}$, where $\mathbf{X} = (X_1,\ldots,X_T)^t$, as an 'estimator' of the covariance matrix Γ_T:

$$I_T(\lambda) := (2\pi T)^{-1} \; \bar{\mathbf{b}}_\lambda^t \;\; \mathbf{X}\,\mathbf{X}^t \;\; \mathbf{b}_\lambda, \quad \mathbf{b}_\lambda = \left\{e^{i\lambda t}\right\}_{t=0,\ldots,T-1} \in \mathbf{C}^T .$$

In the next section a motivation for the form of the Capon estimator will be given.

0.3 Motivating the Capon estimator

An advantage of the Capon estimator when compared to periodogram based estimators is that it is very effective in coping with the 'leakage effect' by using 'adaptive windowing', when the covariance matrix estimator $\widehat{\Gamma}_{d,T}$ involved is unbiased (as the one defined in (0.5)). When using a periodogram type estimator, mass from strong peaks may leak to other frequencies and hide weaker structures of the spectrum. This effect is called 'leakage effect' (see e.g. Bloomfield (1976) Ch. 5.1-5.3). Leakage may be reduced to a certain extend by tapering the data. Tapering consists of multiplying the data X_t by some weights h_t. Usually these are taken as $h_t := h_p(t\,T^{-1})$, where $h_p : [0,1] \to \mathbf{R}$ is a continuously differentiable function which equals 1 on [p/2, 1- p/2] and decreases smoothly to 0 at the edges, e.g. the p % Tukey data taper (see Bloomfield (1976), Ch. 5.2). Then the tapered periodogram is defined as:

$$(0.7) \qquad I_T^{(h)}(\lambda) := (2\pi H_T)^{-1} \left| \sum_{t=1}^{T} h_t X_t e^{i\lambda t} \right|^2, \text{ where } H_T := \sum_{t=1}^{T} h_t^2.$$

Before explaining 'adaptive windowing' more explicitly let us describe very briefly what causes leakage. The expectation of the (non-tapered) periodogram is the convolution of f with the Fejer kernel $\Delta^T(\lambda) := T^{-1} \left| \sum_{t=1}^{T} e^{i\lambda t} \right|^2$. Leakage is caused by the fact that the latter places a lot of mass in the sidelobes. Tapering has the effect of concentrating the mass of the corresponding (modified) kernel $H_T^{-1} \left| \sum_{t=1}^{T} h_t e^{i\lambda t} \right|^2$ closer to 0 and reducing the mass in the sidelobes. The Capon estimator uses implicitly automatic adaptive windowing: in estimating $f(\lambda)$, a kernel is chosen for each λ which minimizes -adaptively to f- the mass contribution coming outside a neighbourhood of λ and which, thus, may differ for different λs not only in location but also in form. Therefore the problem of choosing an optimal data taper is avoided.

Let us now make the property of adaptive windowing more precise by giving the motivation of McDonough (1979). Let $\hat{\Gamma}^{\circ}_{d,T}$ be any unbiased, positive definite estimator of Γ_d (e.g. $\hat{\Gamma}^{\circ}_{d,T} := \hat{\Gamma}_{d,c,T}$). Fix λ and consider the following class of estimators $\hat{f}_{T,w_\lambda}(\lambda)$:

$$\hat{f}_{T,w_\lambda}(\lambda) := \overline{w}_\lambda^t \, \hat{\Gamma}^{\circ}_{d,T} \, w_\lambda, \quad w_\lambda \in \mathbb{C}^d$$

Note that e.g. the Kolmogorov statistic (as specified in (0.6)) belongs to this class: put $\hat{\Gamma}^{\circ}_{d,T} := \hat{\Gamma}_{d,c,T}$ and $w_\lambda = (2\pi\, d)^{-1/2} b_\lambda$. More generally the average over h_t-tapered periodograms belongs to this class: put $w_\lambda = (2\pi\, H_d)^{-1/2} \{ h_t \exp[i\lambda t] \}_{t=1,\dots,d}$ and $\hat{\Gamma}^{\circ}_{d,T} := \hat{\Gamma}_{d,c,T}$ (see also Zhurbenko (1980) and Dahlhaus (1985)).

Then the expectation of $\hat{f}_{T,w_\lambda}(\lambda)$ is given by

$$E\left[\hat{f}_{T,w_\lambda}(\lambda)\right] = \overline{w}_\lambda^t \, \Gamma_d \, w_\lambda = \int_{-\pi}^{\pi} f(\mu) \left| \overline{w}_\lambda^t b_\mu \right|^2 d\mu$$

The aim is to choose a 'good' kernel $\left| \overline{w}_\lambda^t b_\mu \right|^2$, that is to choose a 'good' w_λ. A possibility to formalize this is the following: in order to choose a direction for w_λ (a form for the kernel) a logical demand is that the corresponding kernel has a peak in λ. At the same time the mass contribution coming from other frequencies should be minimized. This can be obtained by minimizing $\int f(\mu) \left| \overline{w}_\lambda^t b_\mu \right|^2 d\mu$ under the constraint that $\left| \overline{w}_\lambda^t b_\lambda \right| = C$ for some $C \in \mathbf{R}$. (The proper normalization C, on which the total mass of the kernel depends, can be chosen afterwards). This yields:

$$\inf\left\{ \overline{w}_\lambda^t \Gamma_d w_\lambda \,\middle|\, \left|\overline{w}_\lambda^t b_\lambda\right| = C \right\} = C^2 \left[\overline{b}_\lambda^t \Gamma_d^{-1} b_\lambda \right]^{-1},$$

the infimum being attained at $w_\lambda := C\left[\overline{b}_\lambda^t \Gamma_d^{-1} b_\lambda \right]^{-1} \Gamma_d^{-1} b_\lambda$.

The proper normalization can be obtained by setting $f = 1/(2\pi)$. In this case Γ_d is the identity matrix; in order that $C^2 \left[\overline{b}_\lambda^t b_\lambda \right]^{-1} = (2\pi)^{-1}$ one has to take $C^2 = d\,(2\pi)^{-1}$. This normalization seems natural for general f as well as can be seen from Relations (A.2.2) and (A.2.3). We remark that the mass of the thus obtained kernel $\int_{-\pi}^{\pi} \left|\overline{w}_\lambda^t b_\mu\right|^2 d\mu$ is not constant in λ but that it will converge to 1 uniformly in λ as d tends to infinity.

With this choice of w_λ,the quantity $\overline{w}_\lambda^t \Gamma_d w_\lambda$ becomes $d\,(2\pi)^{-1} \left[\overline{b}_\lambda^t \Gamma_d^{-1} b_\lambda \right]^{-1}$,which can be naturally estimated by the Capon estimator. Let us underline that the above motivation of the Capon estimator is only heuristics which does not even prove the convergence of $(2\pi)^{-1} d \left[\overline{b}_\lambda^t \Gamma_d^{-1} b_\lambda \right]^{-1}$ to $f(\lambda)$.

0.4 An illustration example

To illustrate that the Capon estimator copes with the 'leakage effect' we present a simulation example. T=512 observations were generated from an autoregressive AR(18) process with innovation variance 1 and characteristic roots (for definitions see Section V.1.2) z_j^{-1} and $\overline{z}_j^{-1}$, with:

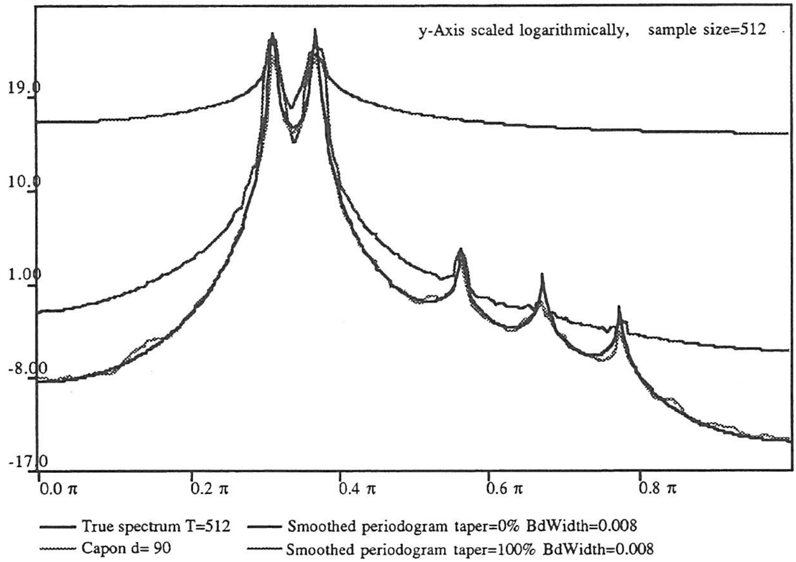
		(radius , angle/π , order)
AR(18)	(z_j)	(0.99, 0.3672, 3), (0.99, 0.3086, 3),
		(0.995, 0.5625, 1), (0.9954, 0.6719, 1)
		(0.9958, 0.7734, 1).

This process was chosen in a way that leakage is extreme in order to make visible the differences between the estimators. For a more moderate example see Fig V.1.

In the next picture we present the Capon estimator and the smoothed periodograms $I_T^{(h)} * K_b$, non-tapered and 100%-tapered. We use the Tukey data taper. As smoothing kernel, we use the Barlett-Priestley kernel with bandwidth b: $K_b(x) := b^{-1}\left[1 - (x / b\pi)^2\right]$ on $(-b\pi, b\pi)$ and elsewhere it equals 0. The parameters (bandwidth for the kernel and d for the Capon estimator) were chosen empirically. All estimators are computed at the points $\lambda_k := 2\pi / T$, k=0,...,T/2. The estimators are plotted on a logarithmic scale (natural logarithm).

From Figure 0.1 it is clear that the untapered periodogram suffers from leakage and fails to discover weaker structures of the spectrum beyond the two main peaks. Leakage is suppressed to some extend by tapering. On the contrary, one has the impression that the Capon estimator is not affected by leakage although no taper was used for it.

Fig. 0.1. A realization of some spectral estimators



0.5 Burg's relation

Several interpretations of the Capon estimator were given by Burg (1972), Marzetta (1983) and Byrne (1984). The most known is the first; it states that the Capon estimator with parameter d is the harmonic mean over autoregressive (Burg) estimators of orders 0 to d-1. Let, for example:

- $\hat{f}^{(h)}_{d,T}$ be the Capon estimator as defined in (0.4) but with a Toeplitz covariance matrix estimator given by some empirical covariances, e.g. (0.3) or a tapered analogon, instead of (0.5). Let also:
- $\hat{f}^{(h,AR)}_{p,T}$ be the autoregressive spectral estimator of order p, defined in a classical manner by estimating the coefficients from the Yule-Walker equations (see V.1.3) using the same empirical covariances as for the Capon estimator.

Then Burg's relation states that:

$$(0.8) \qquad \hat{f}^{(h)}_{d,T}(\lambda) := \left\{ d^{-1} \sum_{p=0}^{d-1} \left[\hat{f}^{(h,AR)}_{p,T}(\lambda) \right]^{-1} \right\}^{-1},$$

Burg's relation has to be slightly modified when using a covariance matrix estimator which is not 'Toeplitz' as (0.5) is not. We give this modification in II.1.4.

From Burg's relation it is clear that the Capon estimator will in general have a larger bias than the corresponding autoregressive estimator -since lower order autoregressive estimators are also involved in the mean- but will, on the other hand, be smoother (see also Baggerroer (1976)). Therefore the comparison of the two estimators is an interesting point which will be partially dealt with in this work.

0.6 The model, the assumptions and some notations

In this thesis we assume that the spectral density f of the time series exists and is bounded above and below by some constants m, M >0: $m < f < M$. This assures that the $d \times d$ covariance matrix $\Gamma = \Gamma_d$ of $\{X_i\}_{i \in Z}$, with (i,j) element $c_{|i-j|}$, is positive definite for each d. The latter also holds for $\widehat{\Gamma}_{d,c,T}$, defined in (0.2), with probability 1 if $d < N_{d,c,T}$.

We make the following assumptions:
(A) f is continuous and there exist m,M >0 with $m < f < M$. Further it fulfills

$$\exists\, C : |f(x) - f(y)| \leq C |\log|x-y||^{-1} \quad \forall\, x,y \in [-\pi, \pi].$$

Assumption (A) is the minimal assumption under which the convergence of orthogonal polynomials to f^{-1} is obtained (see Relation (A.2.3)). The (stronger) assumption (B) allows to study rates for this convergence (see Lemma I.3):

$$(B) \quad [f^{-1}]^{(r)} \in Lip^{\alpha} := \left\{ g \mid \exists\, M < \infty : |g(x) - g(y)| \leq M\, |x-y|^{\alpha} \quad \forall\, x,y \in [-\pi,\pi] \right\},$$

where $g^{(r)}$ is the r-th derivative of g.

(C) $\{X_i\}_{i \in Z}$ has higher order spectral densities $f^{(k)}$ bounded above for all orders k:

$$\|f^{(k)}\|_{\infty} < \infty \quad \forall\, k > 0.$$

A spectral density of higher order is defined as a function $f^{(k)} : [-\pi, \pi]^{k-1} \to C$, such that

$$cum(X_{t_1},\dots,X_{t_k}) = \int f^{(k)}(\alpha_1,\dots,\alpha_{k-1}) \exp\left(i \sum_{j=1}^{k-1} t_j \alpha_j - t_q\, i \sum_{j=1}^{k-1} \alpha_j \right) d(\alpha_1 \dots \alpha_{k-1}).$$

(Integrals are always taken over $[-\pi, \pi]^k$, for some k depending on the integration measure.)

We remark that if $\{X_i\}_{i\in Z}$ is Gaussian then $f^{(k)} \equiv 0$ for $k \geq 3$. Further if $\{X_i\}_{i\in Z}$ is a linear process, that is if it admits a representation as an infinite moving average over some sequence of iid random variables $\{\varepsilon_t\}_{t\in Z}$, then condition (C) is equivalent to existence of all higher order cumulants of ε_t.

We then have the relations:

(0.9) $$c_u = \int f(\lambda)\ \exp(i\lambda u)\ d\lambda\ , \qquad \Gamma_d = \int f(\lambda)\ b_\lambda \overline{b}_\lambda^t\ d\lambda\ ,\ b_\lambda \text{ as in (0.4)}$$

Finally we make the following two general notational remarks:

a) bounds of the form $O[a_T \ln(b_T)]$ with $a_T, b_T \in R$ are to be read as $O[a_T]$ when $b_T \equiv 1$.

b) for any matrix A we denote by $||A||$ the operator norm of the matrix: $||A|| := \sup\{\ ||Ax||_2 \ |\ ||x||_2 = 1\}$.

0.7 Summary of results

In this section we give a very brief summary of the results of the thesis. A more extensive presentation of the results can be found in the sections 'Introduction, results' of each chapter. The proofs can be found in the sections 'Detailed results, proofs' of each chapter.

We first give some references concerning previous work on the Capon estimator.

Contrary to classical nonparametric estimators the Capon estimator is non-linear / non-quadratic in the observations, being a functional of the inverse of the estimated covariance matrix. There are therefore many difficulties in analyzing its statistical behaviour. Until now very little theoretical work has been done in this direction. Its asymptotic distribution has been studied by Capon and Goodman (1971), Pisarenko (1972) and Subba Rao and Gabr (1989). Their approaches are based on the unrealistic, simplifying assumption that independent, Gaussian segments Y_i (of length d) of the time series are available (i.e. that $\widehat{\Gamma}_T$ is exactly Wishart distributed).

In Chapter I of the present work we show the asymptotic normality of the Capon estimator, when the latter is defined by using the segment covariance matrix estimator, as in (0.5), and when $d, T \to \infty$. We drop the assumption of the Gaussianity of the original process, substituting it by the boundness of the higher order spectral densities (see above assumption (C)). We study the bias and variance of the Capon estimator and,consequently, we recommend the use of an 'almost-fully-overlapping-segments covariance matrix estimator' (c=1).

A further result of more general interest is the improvement of the condition sufficient for consistency in the operator norm of the covariance matrix estimator: whereas other authors require $d^2/T \to 0$, we only need $d^{1+\varepsilon}/T \to 0$ for arbitrary $\varepsilon > 0$.

In Chapter II we develop an automatic data adaptive dimension selection criterion for the 'smoothness' parameter d -the dimension of the covariance matrix estimator used. The necessary expansion is carried through up to including second order terms thus yielding as by-product an improvement of the AIC (an order selection criterion for autoregressive type spectral estimators e.g. in Brockwell, Davis (1987) § 9.3).

Finally, based on these considerations, we propose a (quasi) bias correction for the Capon estimator (which can also be carried over to autoregressive estimators); we also show that the corrected estimator has asymptotically, in mean, a smaller error.

In Chapter III we show that the dimension selection criterion developed in Chapter II is asymptotically efficient in the sense of Shibata (1980), using the Whittle distance as a discrepancy: the error in the estimation of f when using this dimension selection criterion is asymptotically equivalent to the minimum possible error (for d smaller than some $d_{max} \to \infty$).
In Chapters II and III we assume almost fully overlapping segments (c=1).

In Chapter IV we show the asymptotic normality of the Capon estimator, when using other covariance matrix estimators instead of (0.5). Namely, we first discuss the use of a symmetrized-segment covariance matrix estimator, and secondly a Toeplitz covariance matrix estimator based on tapered data. In the latter case the necessity of tapering, which was supposed to be avoided by the Capon estimator, is reintroduced: from Burg's relation follows that the estimator will be affected by leakage, since the (Toeplitz) autoregressive one also is. However we study this case for reasons of completeness.

In Chapter V we study, with a simulation, the performance of the Capon estimator as compared to other non-parametric estimators of the spectral density. We also study the performance of the dimension selection criterion and the (quasi) bias correction proposed in Chapter II.

In the Appendix we state and prove some lemmata concerning the properties of 'orthogonal polynomials' associated with f and some derived quantities, as well as two other technical lemmata.

I. THE CENTRAL LIMIT THEOREM

I.1 INTRODUCTION, RESULTS

The aim of this chapter is to prove the asymptotic normality of the Capon estimator $\hat{f}_{d,c,T}$, as defined in (0.4) and (0.5) and to study its bias and variance assuming $d,T\to\infty$. In our proof we use a refinement of Pisarenko's expansion argument, the theory of Orthogonal Polynomials of Szegö (1959) and for the cumulant calculations the concept of 'L^T functions' of Dahlhaus (1983).

A technical result of a more general interest is Lemma I.2 in which we prove $\left\|\hat{\Gamma}_{d,c,T} - \Gamma_d\right\| \to 0$ under the assumption $c\, d^{1+\varepsilon}\, T^{-1} \to 0$. This convergence is also used in the context of non-parametric spectral density estimation or prediction of time series via autoregressive approximation (e.g. Berk (1974), Shibata (1980), Lewis et al. (1985)). In these papers it is proven under the stronger assumption $d^2\, T^{-1} \to 0$ and $c=1$.

We now proceed to the statement of the main theorem of this chapter.

Besides the assumptions on the regularity of the spectrum f ((A) and (B)) and the boundness of spectral densities of higher order (C), we also need some assumptions on the parameters c and d of the estimator which are allowed to converge to ∞ together with T. Assumptions i) and ii) of Theorem I.1 are 'consistency conditions'; they will be discussed in Remarks I.1 and I.2 below. Assumption iii) enforces $d\to\infty$. It is only needed to assure the convergence of the covariances of the estimator to the values given in the theorem (see also Lemma I.9).

The asymptotic normality of the Capon estimator will be stated in terms of $\hat{f}_{d,c,T}$ divided by some proper normalization so that the variance does not depend asymptotically on the spectrum f. As such normalization we first take the theoretical quantity corresponding to the Capon estimator (statement 'a' of the theorem) which is defined as follows:

$$\text{(I.1)} \qquad \tilde{f}_d(\lambda) := \frac{d}{2\pi} \left(\bar{b}_\lambda^t\, \Gamma_d^{-1}\, b_\lambda\right)^{-1}, \ \lambda \in \mathbf{R}.$$

Secondly we take the spectrum f as normalization (statement 'b' of the theorem), which is the quantity one wants to estimate. In doing this we introduce a bias term due to the approximation of f via $\tilde{f}_d$. In statement 'c' of the theorem, which could be used e.g. for constructing confidence intervals for f, we impose an additional condition on the speed of convergence of d to infinity forcing the above mentioned bias term to converge to 0 (when blown up by $\sqrt{T/d}$).

The asymptotic variance of the (standardized) estimator is given by $\lim_{T\to\infty} \theta(d_T / c_T)$, where

$$\theta(x) := x^{-1} \sum_{|u|<x} [1 - |u| x^{-1}]^2, \; x \in \mathbf{R}^+.$$

The appearance of the function θ in the variance is better understood by observing that if $x \equiv k \in \mathbf{Z}^+$ then $\theta(k) = (2\pi k)^{-1} \int [\Delta^k(\lambda)]^2 d\lambda$, where $\Delta^k(\lambda) := k^{-1} |\sum_{t=1}^{k} e^{i\lambda t}|^2$ which is the Fejer kernel. From Lemma A.5 we have that $\theta(x)$ will tend to its minimum value 2/3 e.g. when $c_T \equiv 1$ and $d_T \to \infty$ (almost fully overlapping segments) and it will equal 1 when $d_T \equiv c_T$ (disjoint segments) (see also Remark I.2 below).

For fixed $v \in \mathbf{R}^+$ and $\lambda_k \in [-\pi, \pi]$, k=1,...,K let $\zeta^v = (\zeta_1^v, \ldots, \zeta_K^v) \in \mathbf{R}^K$ be a Gaussian random variable with expectation 0 and covariance $\mathrm{cov}(\zeta_j^v, \zeta_k^v) = v [\delta_{\lambda_j + \lambda_k} + \delta_{\lambda_j - \lambda_k}]$, where δ denotes the Dirac function, extended to be 2π-periodic.

Theorem I.1 Suppose that (A), (B) and (C) hold and the sequences c_T, d_T fulfill

i) $c\, d^{1+\varepsilon} / T \to 0$ for some $\varepsilon > 0$,

ii) $c / d \leq C$ for some constant $C < \infty$.

iii) $d^{-\beta} \ln(T) (\ln(d))^2 \ln(c) \to 0$ (where $\beta := \frac{r+\alpha}{1+r+\alpha}$, r, α as in (B)) as T tends to infinity.

Then setting $v := \lim_{T\to\infty} \theta(d_T/c_T)$ (assuming it exists) we have:

a) $$\sqrt{T d^{-1}} \left\{ \frac{\hat{f}_{d,c,T}}{\tilde{f}_d} (\lambda_k) - 1 \right\}_{k=1,\ldots,K} \Rightarrow_D \zeta^v$$

b) $$\sqrt{T d^{-1}} \left\{ \frac{\hat{f}_{d,c,T}}{f} (\lambda_k) - 1 - B_d(\lambda_k) \right\}_{k=1,\ldots,K} \Rightarrow_D \zeta^v, \text{ where } B_d := \tilde{f}_d f^{-1} - 1.$$

c) If in addition $T d^{-(1+2\gamma)} \ln^4(d) \to 0$, where $\gamma := (r+\alpha) \wedge 1$, then

$$\sqrt{T d^{-1}} \left\{ \frac{\hat{f}_{d,c,T}}{f} (\lambda_k) - 1 \right\}_{k=1,\ldots,K} \Rightarrow_D \zeta^v.$$

•

Remark I.1: Bias considerations. The bias of the Capon estimator $E\, \hat{f}_{d,c,T} - f$ consists of two parts. The first results from the approximation of f via $\tilde{f}_d$. It is the part represented by B_d in the above theorem. Its order of magnitude is studied in Lemma I.3 below. The second part of the bias $E\, \hat{f}_{d,c,T} - \tilde{f}_d$ results from the non-linearity of $\hat{f}_{d,c,T}$. It is shown below, in Lemma I.4, to be of order $O(c\, d^{1+\varepsilon} \ln(T) \ln(c) \ln(d) / T)$, $\forall\, \varepsilon > 0$.

The (quasi) bias correction of $\hat{f}_{d,c,T}$ proposed in Section II.1.3 concerns this second part of the bias.

Remark I.2: Choice of c. From the properties of θ (Lemma A.5) it is clear that the variance tends to the minimum possible value 2/3 when $c / d \rightarrow 0$, while it equals 1 when c=d (disjoint segments). Moreover for fixed d the coefficient of the first order expansion term of the variance and of the bound for $E\,\hat{f}_{d,c,T} - \tilde{f}_d$ are minimized by c=1 (see also Remark I.1). Note that for c=d the consistency condition ii) in Theorem I.1 is $d^{2+\varepsilon} / T \rightarrow 0$ whereas for c=1 it is $d^{1+\varepsilon} / T \rightarrow 0$.

Remark I.3: Interpretation of d. The role of d in the Capon estimator and in the Kolmogorov statistic $I_{d,c,T}$, defined in (0.6), is analogous to the inverse b^{-1} of the bandwidth b in the kernel smoothed periodogram $I_T * K_b$, where $I_T(\lambda)$ is defined in (0.2), $K_b(x) := b^{-1} K(x\, b^{-1})$ and K is a kernel. This may be seen by a comparison of bias and variance; the variance of the limiting distribution of $\hat{f}_{d,1,T}\, f^{-1}$ (and of $f^{-1} I_{d,1,T}(\lambda)$) is approximately $(2\pi T)^{-1} \int \left(\Delta^d\right)^2 \approx \frac{2}{3}\, d / T$, where Δ^d is the Fejer Kernel (see Section 0.3) and that of $(I_T * K_b)\, f^{-1}$ is approximately $(Tb)^{-1} \int K^2$. In the same way, if f fulfills (B) with r=0, the (first part of the) bias of the Capon estimator is of order $O\left(d^{-\alpha} \ln(d)\right)$ (see Lemma I.3), that of $I_{d,1,T}(\lambda)$ is of order $O\left(d^{-\alpha}\right)$, while the bias of the kernel smoothed periodogram is, as can easily be seen, $O\left(b^{\alpha}\right)$ if K is Lipschitz continuous.
On the other hand, if ones imposes stronger regularity conditions on f, e.g. that the k-th derivative of f is bounded from above, then the bias of the Capon estimator and of $I_{d,1,T}(\lambda)$ remains in general of order not smaller than $O\left(d^{-1}\right)$, whereas the bias of the kernel smoothed periodogram becomes, as can easily be seen, $O\left(b^{k}\right)$ if the first k-1 moments of K vanish.

Remark I.4: Choice of d. If c equals 1 (following Remark I.1) the problem of choosing d arises. This problem is dealt with in Chapters II and III, where an AIC-type criterion is developed and is shown to be asymptotically efficient in the sense of Shibata (1980).

Remark I.5: Efficiency considerations. In order to compare the performance of the Capon estimator to that of other estimators, we consider bounds of the convergence rate to 0 of any error criterion, which has the form squared bias + variance. Assuming $f \in \mathrm{Lip}^{\alpha}$ the error of the Capon estimator, the Kolmogorov statistic, the kernel smoothed periodogram and the autoregressive estimator have almost the same rate (up to ln(T) terms): we get by a minimization of $d^{-2\alpha} \ln^2(T) + dT^{-1}$ (see Lemma I.3) that the error of the Capon estimator is $O\left(T^{-(2\alpha)/(2\alpha+1)} \ln^2(T)\right)$ (the power of the ln-term has to be doubled when $\alpha=1$). This is almost the same (up to the ln-terms) as the rate for the kernel smoothed periodogram (Parzen (1957)), for the Kolmogorov statistic and the

autoregressive estimator.
On the other hand, if one imposes stronger regularity conditions on f, e.g. that the k-th derivative of f is bounded from above (it follows $f^{(k-1)} \in \mathrm{Lip}^1$), the error of the Capon estimator and the Kolmogorov statistic has a slower rate than the one for the kernel smoothed periodogram and the autoregressive estimator. The error of the Capon estimator and of $I_{d,1,T}(\lambda)$ remains of order not smaller than $O(T^{-2/3})$. The error of the kernel smoothed periodogram becomes, as can easily be seen, $O\left(T^{-(2k)/(2k+1)}\right)$ if the first k-1 moments of K vanish; already for k=2 and K symmetric this is $O(T^{-4/5})$. The error of the autoregressive estimator is of order $O\left(T^{-(2k)/(2k+1)} \ln^2(T)\right)$ (which follows easily from Lemma I.3). The difference between the first two estimators, on the one hand, and the two latter, on the other, is due to the different order of the bias (see Remark I.3 above). Note that for the two latter estimators (not for the Capon) the error rate approaches the 'parametric rate' $O(T^{-1})$ as the smoothness of the function increases. Note also that $O\left(T^{-(2k)/(2k+1)}\right)$ is the optimal rate under the conditions used here (see e.g. Rosenblatt (1985), Ch. V.6).

The proof of Theorem I.1 is obtained from several technical lemmata in Section I.2. We now sketch the basic idea of the proof and present two results that are of a more general interest.

To obtain the asymptotic distribution of the Capon estimator we need an expansion of it. It turns out that it is technically more convenient to expand the standardized quantity

$$\hat{f}_{d,c,T} / \tilde{f}_d = \left(\bar{b}_\lambda^{*t} \left(\hat{\Gamma}^*_{d,T}\right)^{-1} b^*_\lambda\right)^{-1},$$

where

$$\text{(I.2)} \qquad \hat{\Gamma}^*_{d,T} = \hat{\Gamma}^*_{d,c,T} = U_d^{-1}\, \hat{\Gamma}_{d,c,T} \left(U_d^t\right)^{-1} \text{ and } b^*_\lambda := U_d^{-1} b_\lambda \Big/ \left\|U_d^{-1} b_\lambda\right\|_2.$$

Here $\Gamma_d = U_d\, U_d^t$ is the Cholesky decomposition of Γ_d ($U_d = U_{d,\Gamma}$ is a lower triangular d×d matrix). Quantities involving a standardization with U_d^{-1} will be denoted by '*'. Thus $\hat{\Gamma}^*_{d,T}$ is a 'standardization' of $\hat{\Gamma}_{d,T}$.

By a Neumann expansion of $\left(\hat{\Gamma}^*_{d,T}\right)^{-1}$ we show that $\hat{f}_{d,c,T} / \tilde{f}_d$ is asymptotically equivalent to

$$\text{(I.3)} \qquad I^*_T(\lambda) = I^*_{d,c,T}(\lambda) := \bar{b}_\lambda^{*t}\, \hat{\Gamma}^*_{d,T}\, b^*_\lambda.$$

$I^*_{d,c,T}$ may be interpreted as a 'standardization' of the Kolmogorov statistic $I_{d,c,T}$, defined in (0.6). $I^*_{d,c,T}$ is a random variable which involves the (unknown) true covariance matrix but is quadratic in the observations and can therefore be studied by standard cumulant methods. For the expansion to be valid we need $\left\| \hat{\Gamma}^*_{d,T} - I \right\| \to_P 0$, where I is the d×d identity matrix. This is assured by the following:

Lemma I.2 Assume that (A) and (C) hold and that the sequences c_T, d_T fulfill $d\ c\ \ln(T) \ln(c) / T \to 0$ and $c / d < C$ for some C as T tends to infinity. We then have:

$$E \left\| \hat{\Gamma}^*_{d,T} - I \right\|^K = O\left(c\, d^{1+\varepsilon} \ln(T) \ln(c) \ln(d) / T\right)^{K/2}, \ \varepsilon>0,\ K \in \mathbf{Z}^+ \text{ arbitrary.} \quad \bullet$$

Another result of general interest concerns the (first part of the) bias of the Capon estimator (see Remark I.1). It gives upper bounds for the convergence rate of $\tilde{f}_d - f$ to zero. By $\{\phi_k(\lambda)\}_{k\in N}$ we denote the system of orthogonal polynomials associated with f (see Appendix: A.2). The quantity $\tilde{f}_d^{-1}$ may be written as the mean over the squared modulus of orthogonal polynomials (see (A.2.2)):

$$\tilde{f}_d^{-1} = d^{-1} \sum_{i=0}^{d-1} \left| \phi_i(\lambda) \right|^2$$

Since lower order polynomials are also involved in this mean, it is clear that in general $\tilde{f}_d - f$ will be of order not smaller than d^{-1}.

Lemma I.3 Assume that (A) and (B) hold and let r and α as in (B). Then the following hold:

$$\text{a)} \left\| \left| \phi_d \right|^2 - f^{-1} \right\|_\infty = O\left(d^{-(r+\alpha)} \ln(d)\right)$$

$$\text{b)} \left\| \tilde{f}_d^{-1} - f^{-1} \right\|_\infty = \begin{cases} O\left(d^{-[(r+\alpha)\wedge 1]} \ln(d)\right) & \text{if } r+\alpha \neq 1 \\ O\left(d^{-1} \ln^2(d)\right) & \text{if } r+\alpha = 1 \end{cases} \quad \bullet$$

I.2 DETAILED RESULTS, PROOFS

Here we prove the results stated in I.1. In Section I.2.1 we prove Lemma I.3, in Section I.2.2 the asymptotic equivalence of $\hat{f}_{d,c,T} / \tilde{f}_d$ and $I^*_{d,c,T}$, in Section I.2.3 we bound the moments of $\left\| \hat{\Gamma}^*_{d,T} - I \right\|$ (Lemma I.2), in Section I.2.4 we study the cumulants of $I^*_{d,c,T}$ and in Section I.2.5 we prove Theorem I.1. Some intermediate steps of the proofs will be stated as separate propositions since they will be also used in subsequent chapters.

I.2.1 Bias considerations (part 1)

In this chapter we prove Lemma I.3 by utilizing properties of orthogonal polynomials which are proven in the Appendix. This lemma concerns only the component of the bias of the Capon estimator resulting from the approximation of f via $\tilde{f}_d$ (see also Remark I.1).

Proof of Lemma I.3

a) The result follows directly from Lemma A.3 a), taking $p_n \equiv n \equiv d$. Observe that by (A.2.4) we have $|\psi_{d,d}|^2 \equiv |t_d|^2$.

b) Follows directly from a), since $\left\| \tilde{f}_d^{-1} - f^{-1} \right\|_\infty \leq d^{-1} \sum_{p=0}^{d-1} \left\| |\phi_p|^2 - f^{-1} \right\|_\infty$. □

I.2.2 Asymptotic expansion and bias considerations (part 2)

In this section we prove the asymptotic equivalence between $\hat{f}_{d,c,T} / \tilde{f}_d$ and $I^*_{d,c,T}$. As a by-product we also obtain a bound on the bias $E\,\hat{f}_{d,c,T}\,\tilde{f}_d^{-1} - 1$. We state the result in a somehow more general context, allowing also for other covariance matrix estimators, because it will be needed in this form in Chapter IV. In this order let $\hat{\hat{\Gamma}}_{d,T}$ be any d×d covariance matrix estimator and define $\hat{\hat{f}}_{d,T}$, $\hat{\hat{\Gamma}}^*_{d,T}$ and $\hat{\hat{I}}^*_{d,T}$ analogously to $\hat{f}_{d,c,T}$, $\hat{\Gamma}^*_{d,c,T}$ and $I^*_{d,c,T}$ respectively (see (0.4), (I.2), (I.3)). Then the following holds:

Lemma I.4 The following two implications hold:

i) If $\sqrt{T/d}\,\left\| \hat{\hat{\Gamma}}^*_{d,T} - I \right\|^2 \to_P 0$ as $T \to \infty$ <u>then</u> $\sqrt{T/d}\,\left[\hat{\hat{f}}_{d,T} / \tilde{f}_d - \hat{\hat{I}}^*_{d,T} \right] \to_P 0$.

ii) If $E\left\| \hat{\hat{\Gamma}}^*_{d,T} - I \right\|^2 = O(\alpha_T)$ for some sequence α_T <u>then</u>

$$E\left[\hat{\hat{f}}_{d,T} / \tilde{f}_d - \hat{\hat{I}}^*_{d,T} \right] = O(\alpha_T)$$ •

Proof. First note that :

$$\frac{\hat{\hat{f}}_{d,T}}{\tilde{f}_d} = \frac{\left(\overline{b}_\lambda^t (U_d^t)^{-1} U_d^t \left(\hat{\hat{\Gamma}}_{d,T}\right)^{-1} U_d U_d^{-1} b_\lambda\right)^{-1}}{\left(\overline{b}_\lambda^t (U_d^t)^{-1} U_d^{-1} b_\lambda\right)^{-1}} = \left(\overline{b}_\lambda^{*t} \left(\hat{\hat{\Gamma}}^*_{d,T}\right)^{-1} b_\lambda^*\right)^{-1}.$$

Now let the following event be denoted by $A_T = \left\{ \left\| \hat{\hat{\Gamma}}^*_{d,T} - I \right\| < 1/2 \right\}$. Then on A_T we may expand $\left[\hat{\hat{\Gamma}}^*_{d,T}\right]^{-1}$:

$$\left[\hat{\hat{\Gamma}}^*_{d,T}\right]^{-1} = \left[I - \left(I - \hat{\hat{\Gamma}}^*_{d,T}\right)\right]^{-1} = \sum_{j=0}^{\infty} \left(I - \hat{\hat{\Gamma}}^*_{d,T}\right)^j$$

This implies

$$\left(\overline{b}^{*t}_\lambda \left(\hat{\hat{\Gamma}}^*_{d,T}\right)^{-1} b^*_\lambda\right) - 1 = \overline{b}^{*t}_\lambda \left(I - \hat{\hat{\Gamma}}^*_{d,T}\right) b^*_\lambda + O\left(\left\| \hat{\hat{\Gamma}}^*_{d,T} - I \right\|^2 \right) \text{ on } A_T.$$

Since we have:

$$(1) \qquad \left| \left(\overline{b}^{*t}_\lambda \left(\hat{\hat{\Gamma}}^*_{d,T}\right)^{-1} b^*_\lambda\right)^{-1} - 1 \right| \le \left\| \hat{\hat{\Gamma}}^*_{d,T} - I \right\|$$

it follows from the preceding that

$$(2) \qquad \left(\overline{b}^{*t}_\lambda \left(\hat{\hat{\Gamma}}^*_{d,T}\right)^{-1} b^*_\lambda\right)^{-1} - 1 = \overline{b}^{*t}_\lambda \left(\hat{\hat{\Gamma}}^*_{d,T} - I\right) b^*_\lambda + O\left(\left\| \hat{\hat{\Gamma}}^*_{d,T} - I \right\|^2 \right) \quad \text{on } A_T$$

Since $\overline{b}^{*t}_\lambda \left(\hat{\hat{\Gamma}}^*_{d,T} - I\right) b^*_\lambda = \hat{\tilde{I}}^*_{d,T} - 1$, $\sqrt{T/d}\left\| \hat{\hat{\Gamma}}^*_{d,T} - I \right\|^2 \to_P 0$ and $P(A^c_T) \to 0$ i) is proven.

We now turn to the proof of ii). First note that (1) above holds everywhere, not only on A_T. From this and (2) we have that:

$$E\left[\hat{\hat{f}}_{d,T}/\tilde{f}_d - \hat{\tilde{I}}^*_{d,T} \right] \le O(1)\int_{A_T} \left\| \hat{\hat{\Gamma}}^*_{d,T} - I \right\|^2 dP + O(1)\int_{A^C_T} \left\| \hat{\hat{\Gamma}}^*_{d,T} - I \right\| dP$$

$$\le O(1)\, E\left\| \hat{\hat{\Gamma}}^*_{d,T} - I \right\|^2 + O(1)\left[E\left\| \hat{\hat{\Gamma}}^*_{d,T} - I \right\|^2 P(A^c_T) \right]^{1/2}$$

$$= O(1)\, E\left\| \hat{\hat{\Gamma}}^*_{d,T} - I \right\|^2.$$

This concludes the proof. □

Remark I.6 We consider the case $\hat{\hat{\Gamma}}_{d,T} = \hat{\Gamma}_{d,c,T}$. In this case we obtain $E\, I^*_{d,c,T}=1$. This together with Lemma I.2 yields a bound on the second part of the bias of the Capon estimator:

$$E\, \hat{f}_{d,c,T}\, \tilde{f}_d^{-1} - 1 = O\left(c\, d^{1+\varepsilon} \ln(T) \ln(c) \ln(d) / T\right), \ \forall\ \varepsilon>0.$$

I.2.3 Consistency of the covariance matrix estimator

In this section we prove Lemma I.2 as a direct consequence of the following Lemma I.5. Let $N = N_{d,c,T}$ as in (0.5).

Lemma I.5 If (A), (C) hold, $K \in \mathbb{N}$ is even, $d\, N^{-1} \ln(N) \ln(c) \to 0$ and $c/d < C$ (for some constant C) then we have

$$E\,\mathrm{tr}\left[\hat{\Gamma}^{*}_{d,T} - I\right]^{K} = O\left(N^{-K/2}\, d^{K/2+1}\, [\ln(N) \ln(c) \ln(d)]^{K/2}\right). \qquad \bullet$$

The proof of Lemma I.5 is rather complicated. The first step is given in Proposition I.7. The actual argument is given in Proposition I.8. A technical tool used here and in subsequent chapters is Proposition I.6. First we introduce the following notation:

Table I.1 We consider a $2 \times k$ table of variables of the form:

$$\begin{matrix} \alpha_1 & \beta_1 \\ \vdots & \vdots \\ \alpha_k & \beta_k \end{matrix}$$

and partitions $\mathcal{P}^{(k)} = \{P_1, \ldots, P_S\}$ of the $2 \times k$ table. Let $s_i := |P_i|$. For a partition subset $P_i := (\kappa_1, \ldots, \kappa_{s_i})$ (κ stands for some α or β) denote by $\tilde{\kappa}_i := (\kappa_1, \ldots, \kappa_{s_i - 1})$ and set $\kappa_{s_i} := -\sum_{j=1}^{s_i - 1} \kappa_j$. Let $\sum_{ip,(k)}$ denote summation over the indecomposable partitions of the $2 \times k$ table, $\sum_{ap,(k)}$ denote summation over **all** partitions of the $2 \times k$ table and $\sum_{ap\backslash *,(k)}$ denote summation over **all** partitions of the $2 \times k$ table **excluding** those which contain a partition subset consisting of exactly one row of the table.

For a fixed partition $\mathcal{P}^{(k)}$ we will call <u>indecomposable row-subtable</u> a union of rows of the original $2 \times k$-table, if and only if it can be written as a union of some partition subsets and cannot be split up any more in this sense. Similarly we will call <u>indecomposable diagonal-subtable</u> a union of diagonals $\{\alpha_i, \beta_{i+1}\}$ of the original $2 \times k$-table, if and only if it can be written as a union of some partition subsets and cannot be split up any more in this sense.

Subsequently let $\mathbf{b}_\lambda = \{e^{i\lambda t}\}_{t=0,\ldots,T-1} \in \mathbb{C}^T$ and $\mathbf{X} = (X_1, \ldots, X_T)^t$.

Proposition I.6 Assume (C). Let A_j, j=1,...,k be arbitrary T×T matrices. Using the notation in Table I.1 (e.g. S is the number of partition subsets of, $s_i := |P_i|$) we have:

a) $$\operatorname{cum}\left(\prod_{j=1}^{k} X^t A_j X\right) = \sum_{ap,(k)} \int \prod_{i=1}^{S} f^{(s_i)}(\tilde{\kappa}_i) \prod_{j=1}^{k} \mathbf{b}_{\alpha_j}^t A_j \mathbf{b}_{\beta_j} \prod_{i=1}^{S} d\tilde{\kappa}_i$$

b) $$\operatorname{cum}(X^t A_1 X, \ldots, X^t A_k X) = \sum_{ip,(k)} \int \prod_{i=1}^{S} f^{(s_i)}(\tilde{\kappa}_i) \prod_{j=1}^{k} \mathbf{b}_{\alpha_j}^t A_j \mathbf{b}_{\beta_j} \prod_{i=1}^{S} d\tilde{\kappa}_i.$$

c) $$\operatorname{cum}\left(\prod_{j=1}^{k} [X^t A_j X - E X^t A_j X]\right) = \sum_{ap\backslash *,(k)} \int \prod_{i=1}^{S} f^{(s_i)}(\tilde{\kappa}_i) \prod_{j=1}^{k} \mathbf{b}_{\alpha_j}^t A_j \mathbf{b}_{\beta_j} \prod_{i=1}^{S} d\tilde{\kappa}_i \bullet$$

Proof. The results follow from the product theorem for cumulants by straight forward calculations and by using the spectral representation

$$\operatorname{cum}(X_{t_1}, \ldots, X_{t_q}) = \int f^{(q)}(\alpha_1, \ldots, \alpha_{q-1}) \exp\left(\sum_{j=1}^{q-1} t_j \alpha_j - t_q \sum_{j=1}^{q-1} \alpha_j\right) d(\alpha_1 \ldots \alpha_{q-1}). \qquad \square$$

Let further $E_j = E_j^{d,c,T} := [0_{d\times(j-1)c} \ \ I_d \ \ 0_{d\times(T-d-(j-1)c)}] \in \mathbf{R}^{d\times T}$. Then $Y_j = Y_j^{d,c} = E_j X$.

Denote by $\Delta^N(\lambda) := N^{-1} |\Theta^N(\lambda)|^2$ the Fejer Kernel, where $\Theta^N(\lambda) := \sum_{t=1}^{N} e^{i\lambda t}$ and also by

$$K_d(\alpha, \beta) := \bar{b}_\alpha^t \Gamma_d^{-1} b_\beta.$$

For a given partition $\mathcal{P}^{(k)}$ of the 2 × k table I.1 we denote by

$$V(\mathcal{P}^{(k)}) := N^{-k} \int \prod_{i=1}^{S} f^{(s_i)}(\tilde{\kappa}_i) \prod_{i=1}^{k} \Theta^N(c\alpha_i + c\beta_i) \ K_d(-\alpha_i, \beta_{i+1}) \prod_{i=1}^{S} d\tilde{\kappa}_i$$

and by

$$\tilde{V}(\mathcal{P}^{(k)}) := N^{-k} \int \prod_{i=1}^{k} L^N(c\alpha_i + c\beta_i) \ L^d(\alpha_i + \beta_{i+1}) \prod_{i=1}^{S} d\tilde{\kappa}_i$$

where the L^N functions are defined in the Appendix (A1). In these expressions indices are always taken mod(k): e.g. $\beta_{k+1} \equiv \beta_1$.

The proof of Lemma I.5 will be obtained directly from the following two propositions:

Proposition I.7 Assume (C). Then we have (using the notation of table I.1)

$$E\,\mathrm{tr}\left[\hat{\Gamma}^*_{d,T} - I\right]^K = \sum_{ap\backslash *, (K)} V\left(\mathcal{P}^{(K)}\right) \qquad \bullet$$

Proof. From $I = (2\pi)^{-1}\int b_\lambda \overline{b}^t_\lambda \, d\lambda$ (I is the d×d identity matrix) we obtain:

$$\mathrm{tr}\left[\hat{\Gamma}^*_{d,T} - I\right]^K = (2\pi)^{-K}\int \prod_{i=1}^{K} \overline{b}^t_{\lambda_i}\left[\hat{\Gamma}^*_{d,T} - I\right] b_{\lambda_{i+1}} \prod_{i=1}^{K} d\lambda_i$$

Observe that:

$$\overline{b}^t_{\lambda_i}\left[\hat{\Gamma}^*_{d,T} - I\right] b_{\lambda_{i+1}} = X^t A_i X - \overline{b}^t_{\lambda_i} b_{\lambda_{i+1}} \text{ where } A_i := N^{-1}\sum_{j=1}^{N} E^t_j \left(U^t_d\right)^{-1} b_{\lambda_{i+1}} \overline{b}^t_{\lambda_i} U^{-1}_d E_j.$$

We have that $E_j\, \mathbf{b}_\beta = \exp\left[-i\,(j-1)\,c\beta\right] b_\beta$. This yields:

$$\mathbf{b}^t_\alpha\ A_j\, \mathbf{b}_\beta = N^{-1}\,\Theta^N\left(c\alpha + c\beta\right)\ b^t_\alpha \left(U^t_d\right)^{-1} b_{\lambda_{j+1}}\ \overline{b}^t_{\lambda_j} U^{-1}_d\, b_\beta\ \exp\left[-i\left(c\alpha + c\beta\right)\right].$$

With these we get directly from Proposition I.6 c) that the expectation of $\mathrm{tr}\left[\hat{\Gamma}^*_{d,T} - I\right]^K$ equals (using the notation of table I.1)

$$(2\pi N)^{-K} \sum_{ap\backslash *, (K)} \int \prod_{i=1}^{S} f^{(s_i)}\left(\tilde{\kappa}_i\right) \prod_{i=1}^{K}\left[\Theta^N\left(c\alpha_i + c\beta_i\right) b^t_{\alpha_i}\left(U^t_d\right)^{-1} b_{\lambda_{i+1}}\ \overline{b}^t_{\lambda_i} U^{-1}_d b_{\beta_i}\right] \prod_{i=1}^{K} d\lambda_i \prod_{i=1}^{S} d\tilde{\kappa}_i$$

The desired result follows after integration with respect to λ_i, i=1,...,K noting that:

$$(2\pi)^{-1}\int b^t_{\alpha_i}\left(U^t_d\right)^{-1} b_{\lambda_{i+1}}\ \overline{b}^t_{\lambda_{i+1}} U^{-1}_d\, b_{\beta_{i+1}}\, d\lambda_{i+1} = b^t_{\alpha_i}\Gamma^{-1}_d\, b_{\beta_{i+1}} = K_d\left(-\alpha_i, \beta_{i+1}\right) \qquad \square$$

Proposition I.8 Assume that (A), (C) hold and c / d < C (for some constant C). Further let for K arbitrary $\mathcal{P}^{(K)}$ be a partition of the 2 × K table I.1 which does not contain any partition subset consisting of exactly one row of the table. Assume that $\mathcal{P}^{(K)}$ consists of M row-subtables and L diagonal-subtables. Then we have

$$\tilde{V}\left(\mathcal{P}^{(K)}\right) = O\left(N^{M-K}\, d^{1+K-M}\, c^{M-K+L-1}\, \ln(N)^{K-M}\, \ln(c)^{K-M-L+1}\, \ln(d)^{M}\right). \qquad \bullet$$

Proof. First enumerate the elements of each partition subset P in such a way that its last element appears in the first row of those containing elements of P.

To prove our assertion, we will integrate with respect to (and successively eliminate) the variables under the integral in $\widetilde{V}\left(\mathcal{I}^{(K)}\right)$ using Lemmata A.1 i) and A.1 ii). Note that this is possible since we have the structure needed to use them: each variable appears once with positive and once with negative sign in the arguments of the L^N-factors as well as in the arguments of the L^d-factors (if it appears at all). We will use the following integration-elimination scheme: first integrate with respect to / eliminate the variables (which are not the last of their partition subsets) in the first row of the 2×K table, then in the second, and so on. Use Lemma A.1 i) where it is possible.

Observe that the factor $L^N(c\alpha_j+c\beta_j)$ appears in the integrand in $\widetilde{V}$ as long as neither α_j nor β_j have been eliminated. In this case we call the row j 'unconnected'. We also call 'connecting' an unconnected row when we integrate with respect to / eliminating α_j or β_j.

In this integration / elimination process $L^N(0)=N$ will appear exactly when all rows of a row-subtable have been connected; thus we obtain the factor N^M for the final bound. In the same way we also obtain the factor d^L. Finally Lemma A.1 ii) will only -but not necessarily- be used when connecting a row, that is at most K-M times. Moreover we claim:

(1): the number of times Lemma A.1 ii) will have to be used instead of Lemma A.1 i) does not exceed K-M-L+1.

Accordingly in the final bound the factor

- d ln(N) ln(c) /c will appear whenever Lemma A.1 ii) is used thus K-M-L+1 times;
- ln(N) will additionally appear whenever Lemma A.1 i) is used (on two L^N functions), thus K-M times in total;
- ln(d) will appear whenever Lemma A.1 i) is used (on two L^d functions), thus K-L+1- (number of times Lemma A.1 ii) is used).

These yield:

$$\widetilde{V}\left(\mathcal{I}^{(K)}\right) < O(1)\ N^{M-K}\, d^L \left(\frac{d}{c}\right)^{K-M-L+1} \ln(N)^{K-M}\ \ln(c)^{K-M-L+1}\ \ln(d)^{(K-L+1)-(K-M-L+1)}$$

This proves our assertion. It remains to prove (1). In view of our remark preceding (1), it is sufficient to prove (2).

(2): for each diagonal-subtable up to the one containing α_1 and β_k one may use Lemma A.1 i) instead of Lemma A.1 ii) for connecting its last row.

Proof of (2):

Let β_{j+1} be the element of the diagonal-subtable considered in (2) with maximal index. Due to our enumeration we may assume -without loss of generality- that

i) β_{j+1} is not the last element of any partition subset;

ii) the row j+1 has not yet been connected;

iii) the row j has been connected and α_j, β_j have been eliminated, if they are not the last elements of a partition subset.

There are two cases:

i) α_j is the last element of a partition subset P. Then the P must contain β_{j+1}. The reason is that since β_{j+1} has maximal index in its diagonal-subtable, our enumeration allows only β_j, β_{j+1} as probable candidates for elements of P. The assumption that there is no one-row partition subset excludes the case $\beta_{j+1} \notin P$.

ii) (i) is not fulfilled, that is α_j is not the last element of any partition subset P.

In both cases β_{j+1} does not occur in the argument of an L^d-factor and thus the j+1 row may be connected by using Lemma A.1 i). This proves (2). □

Proof of Lemma I.5 According to Proposition I.7 it is sufficient to show that for any partition $\mathcal{P}^{(K)}$ of the table I.1 $\left| V\left(\mathcal{P}^{(K)}\right) \right|$ has the desired order. Now, because of Lemma A.2, we have $\left| V\left(\mathcal{P}^{(K)}\right) \right| = O\left(\widetilde{V}\left(\mathcal{P}^{(K)}\right) \right)$ and Proposition I.8 yields the result with the help of the following three observations:

a) $M \leq K/2$, as each row-subtable consists of at least two rows.

b) $d\,N^{-1} \ln(N)\, \ln(c) \to 0$ by assumption and

c) $M-K+L-1 \leq 0$, since each m-row-subtable contains maximally m-1 diagonal-subtables, except the m-row-subtable which includes the first and last row; it may contain maximally m diagonal-subtables.

Concerning c) let us remark that there are partitions for which $M-K+L-1=0$; for example the partition $\{(\alpha_1, \beta_2), \ldots, (\alpha_K, \beta_1)\}$. Thus we do not really loose anything by this inequality. □

Finally we prove Lemma I.2 as a simple consequence of Lemma I.5.

Proof of Lemma I.2 For any $p \in \mathbf{Z}_+$, p even, Hölder' s inequality and Lemma I.5 yield with some constant $C_{K,p}$:

$$E\left\|\widehat{\Gamma}_T^* - I\right\|^K \leq \left(E\left\|\widehat{\Gamma}_T^* - I\right\|^{pK}\right)^{1/p} \leq \left(E \operatorname{tr}\left[\widehat{\Gamma}_T^* - I\right]^{pK}\right)^{1/p}$$

$$\leq C_{K,p}\left(\frac{c\, d^{1+2/Kp}\ \ln(T)\ln(c)\ln(d)}{T}\right)^{K/2} = O\left(c\, d^{1+\varepsilon}\ \ln(T)\ln(c)\ln(d)/T\right)^{K/2}$$

if p is chosen in a way that $\frac{2}{pK} < \varepsilon$ □

I.2.4 Cumulants of $I^*_{d,c,T}(\lambda)$.

In this section we study the second and higher order cumulants of the random variable $I^*_{d,c,T}(\lambda)$. Let θ and δ be defined as in Chapter I.1 and $N = N_{d,c,T}$. We prove the following results:

Lemma I.9 Under (A), (B) and (C), assuming $T\to\infty$ and $c/d < C$ for some C we have:

$E\left[I^*_{d,c,T}(\lambda) - 1\right] = 0$, and

$$\mathrm{cov}\left[\sqrt{\frac{Nc}{d}}\, I^*_{d,c,T}(\lambda),\ \sqrt{\frac{Nc}{d}}\, I^*_{d,c,T}(\mu)\right] = \begin{cases} O(R_2), \text{ if } \lambda\neq\pm\mu \bmod(2\pi) \\ \theta(d/c)\left[\delta_{\lambda+\mu}+\delta_{\lambda-\mu}\right] + O(R_1), \text{ else} \end{cases}$$

with $R_2 := d^{-2}\left[\left|\lambda-\mu\right|^{-2}+\left|\lambda+\mu\right|^{-2}\right]\ \ln(N)\ln(d)\ln(c) + R_1$ and

$R_1 := d^{-\beta}\ln(N)\,(\ln(d))^2\ln(c) + d\,(c\,N)^{-1}$, where $\beta := \frac{r+\alpha}{1+r+\alpha}$, r,α as in (B). •

For the next lemma we use the notation of table I.1 (with r instead of k). Further we regard the partitions of table I.1 as naturally and simultaneously translated to partitions of the following table:

$$\begin{matrix} \lambda_1 & -\lambda_1 \\ \vdots & \vdots \\ \lambda_r & -\lambda_r \end{matrix}$$

For a partition subset P of this table we write $\sum_{\pm\lambda_i\in P}\pm\lambda_i$ to indicate the sum of its elements.

The next result is broken into two parts: only the second bound is needed for the proof of Theorem I.1; the first is needed in Chapter III.

Lemma I.10 Under (A), (B) and (C), assuming $T\to\infty$ and $c / d < C$ (for some C) the following holds:

$$\operatorname{cum}\left[\sqrt{\frac{Nc}{d}}\, I^*_{d,c,T}(\lambda_1),\ldots,\sqrt{\frac{Nc}{d}}\, I^*_{d,c,T}(\lambda_r)\right] =$$

$$=O\left[d^{-(r/2+1)}(Nc)^{-(r/2-1)}(\ln(c)\ln(N))^{r-1}\right] \sup_{\mathcal{I}(r)} \ln(d)^{r-S+1}\prod_{j=1}^{S} L^d\left(\sum_{\pm\lambda_i\in P_j}\pm\lambda_i\right)$$

$$=O\left(\left[dc^{-1}N^{-1}\right]^{r/2-1}\left[\ln(N)\ln(c)\right]^{r-1}\ln(d)\right) \qquad \bullet$$

In order to prove these two lemmata we need the following technical tool (using the same notation as in Table I.1):

Proposition I.11 Assume (C). Then the following holds:

$$\operatorname{cum}\left[\sqrt{\frac{Nc}{d}}\, I^*_{d,c,T}(\lambda_1),\ldots,\sqrt{\frac{Nc}{d}}\, I^*_{d,c,T}(\lambda_r)\right] = \left(\frac{c}{dN}\right)^{r/2} d^{-r}(2\pi)^r\left[\prod_{j=1}^{r}\tilde{f}_d(\lambda_j)\right]\cdot$$

$$\sum_{ip,\,(k)}\int\prod_{i=1}^{S} f^{(s_i)}(\tilde{\kappa}_i)\prod_{j=1}^{r}\left[K_d(\lambda_j,\alpha_j)\,K_d(-\lambda_j,\beta_j)\,\Theta^N(c\alpha_j+c\beta_j)\right]\prod_{i=1}^{S} d\tilde{\kappa}_i. \qquad \bullet$$

Proof. The proof is a direct consequence of Proposition I.6 b), by observing that

$$I^*_T(\lambda) = 2\pi\, d^{-1}\,\tilde{f}_d(\lambda)\quad X^t A_\lambda X \text{ with } A_\lambda := N^{-1}\sum_{j=1}^{N} E^t_j\,\Gamma_d^{-1}\,\bar{b}_\lambda\, b^t_\lambda\,\Gamma_d^{-1}\, E_j \text{ and that}$$

$$b^t_\alpha A_\lambda b_\beta = K_d(\lambda,\alpha)\,K_d(-\lambda,\beta)\,N^{-1}\,\Theta^N(c\alpha+c\beta)\,\exp\left(-i\,(c\alpha+c\beta)\right). \qquad \square$$

Using this proposition one may prove the lemma concerning the first and second order cumulants of $I^*_T(\lambda)$.

Proof of Lemma I.9 The first statement concerning the expectation of $I^*_T(\lambda)$ follows easily by observing that $E\,\hat{\Gamma}^*_T = I_d$.

Next we study the covariance structure of $I^*_T(\lambda)$. The indecomposable partitions of the 2x2 table I.1 are

$$\text{i) } \{\alpha_1,\beta_2\},\{\beta_1,\alpha_2\} \quad \text{ii) } \{\alpha_1,\alpha_2\},\{\beta_1,\beta_2\} \quad \text{iii) } \{\alpha_1,\alpha_2,\beta_1,\beta_2\}.$$

Thus by Proposition I.11 and setting

$$P_{d,c,T}(f,\lambda,\mu) := \int f(\alpha)\, f(\beta)\, \Delta^N(c\beta - c\alpha)\, K_d(\lambda,\alpha)\, K_d(-\mu,-\alpha)\, K_d(\mu,\beta)\, K_d(-\lambda,-\beta)\, \mathbf{d\alpha\, d\beta},$$

we obtain:

$$\operatorname{cov}\left[\sqrt{\tfrac{Nc}{d}}\, I_T^*(\lambda),\ \sqrt{\tfrac{Nc}{d}}\, I_T^*(\mu)\right] =$$

$$d^{-3}c\ (2\pi)^2\, \tilde f_d(\lambda)\, \tilde f_d(\mu) \cdot \Big[P_{d,c,T}(f,\lambda,\mu) + P_{d,c,T}(f,\lambda,-\mu) +$$

$$\int f^{(4)}(\alpha,\beta,\gamma)\ \Delta^N(c\alpha + c\beta)\ K_d(\lambda,\alpha)\ K_d(-\lambda,\beta)\ K_d(\mu,\gamma)\, K_d(-\mu,-(\alpha+\beta+\gamma))\ \mathbf{d(\alpha\beta\gamma)}\Big]$$

Now the last integral in the above expression multiplied by $d^{-3}c$ may, by Lemmata A.2 and A.1, be shown to be $O(d^{-1}\ \ln(N)\ln^2(d)\ln(c))$ under (C).
With the same method we obtain

$$d^{-3}\, c\, P_{d,c,T}(f,\lambda,\mu) = O(\ln(N)\ln(d)\ln(c))\, d^{-2}\, (L^d)^2(\mu-\lambda)$$

Since for $\mu \neq \pm\lambda \bmod(2\pi)$ and for d sufficiently large we have $(L^d)^2(\mu-\lambda) = |\mu-\lambda|^2$, it remains to prove that:

$$d^{-3}\, c\ \ (2\pi)^2\, \tilde f_d^2(\lambda)\ \ P_{d,c,T}(f,\lambda,\lambda) = \theta(d\, c^{-1}) + O(R_1).$$

To prove this, let $g_d(\lambda)$ be a sequence of $AR(p_d)$ spectral densities, $p_d := d^{1/(1+r+\alpha)}$, r, α as in (B) with $\| g_d - f \|_\infty = O(p_d^{-(r+\alpha)})$ (Lemma A.3). Then it follows that:

$$d^{-3}\, c\ \ \left| P_{d,c,T}(f,\lambda,\lambda) - P_{d,c,T}(g_d,\lambda,\lambda) \right| = O\left(d^{-\beta}\ln(N)\ln^2(d)\ln(c)\right),\ \beta := \frac{r+\alpha}{1+r+\alpha}.$$

This is proven by substituting in $P_{d,c,T}(f,\lambda,\lambda)$ successively all quantities depending on f, that is f and $K_d(f,\lambda,\mu)$, by the quantities corresponding to g_d. Observe that the error for each substitution is $O(\ p_d^{-(r+\alpha)}\ln(d)\ )\left| d^{-3}c\ \ P_{d,c,T}(f,\lambda,\lambda)\right| = O(\ d^{-\beta}\ln(N)\ln^2(d)\ln(c))$.

Now let $K^*_{d,p}(f,\lambda,\mu) := (2\pi)^{-1} \sum_{j=p}^{d-1} \overline{\phi_j(\lambda)}\, \phi_j(\mu)$, $\{\phi_k(\lambda)\}_{k\in N}$ as in A.2 and define $P^*_{d,p,c,T}(f,\lambda,\mu)$ as $P_{d,c,T}(f,\lambda,\mu)$ by substituting $K^*_{d,p}(.,.)$ instead of $K_d(.,.)$.

Again we obtain, with $p := p_d$,

$$d^{-3} c \; \left| P^*_{d,p_d,c,T}\left(g_d,\lambda,\lambda\right) - P_{d,c,T}\left(g_d,\lambda,\lambda\right)\right| = O\left(d^{-\beta} \ln(N) \ln^2(d) \ln(c)\right).$$

This follows from

$$\left| K_d\left(\lambda,\mu\right) - K^*_{d,p}\left(\lambda,\mu\right)\right| \le O(1)\, L^p\!\left(\mu-\lambda\right) \le O\left(\frac{p}{d}\right) L^d\!\left(\mu-\lambda\right) = O\left(d^{-\beta}\right) L^d\!\left(\mu-\lambda\right)$$

using the same argument as above. Thus remains to prove that:

$$d^{-3} c \;\left(2\pi\right)^2 g_d^2\!\left(\lambda\right) \; P^*_{d,p_d,c,T}\left(g_d,\lambda,\lambda\right) = \theta\left(d\, c^{-1}\right) + O\!\left(R_1\right)$$

But in $P^*_{d,p_d,c,T}\left(g_d,\lambda,\lambda\right)$ enter only the polynomials orthogonal to g_d of degree $k \ge p_d$, which are exactly known (care of (A.2.4)). By substituting them, all the terms depending on g_d are canceled. We now substitute the remaining $\sum_{j=p}^{d-1} e^{i\,(d-1)(\mu-\lambda)}$ by $\sum_{j=0}^{d-1} e^{i\,(d-1)(\mu-\lambda)}$ having again a total error of $O(R_1)$. As elementary calculations show the remaining quantity equals

$$c d^{-1} (2\pi)^{-2} \int \Delta^N\!\left(c\alpha-c\beta\right) \Delta^d\!\left(\alpha-\lambda\right) \Delta^d\!\left(\beta-\lambda\right) \; \mathbf{d}(\alpha\beta) \; = \theta\left(d\, c^{-1}\right) + O\!\left(d\, c^{-1} N^{-1}\right) \qquad \square$$

Finally we prove the lemma concerning the higher order cumulants of $I^*_T(\lambda)$:

Proof of Lemma I.10 The proof follows partly the lines of the proof of Lemma 4.5 of Dahlhaus (1985). Fix a partition $\mathcal{P}^{(r)} = \{ P_1, \ldots, P_S\}$ of the $2 \times r$ Table I.1 and let $s_i := | P_i |$. According to Proposition I.11 and Lemma A.2 it is sufficient to show:

$$\int \prod_{j=1}^{r} \left[L^d\!\left(\lambda_j+\alpha_j\right) L^d\!\left(-\lambda_j+\beta_j\right) L^N\!\left(c\alpha_j+c\beta_j\right)\right] \prod_{i=1}^{S} \mathbf{d}\tilde{\kappa}_i$$

$$= O\!\left(N \left(d/c \;\ln(N) \ln(c)\right)^{r-1} \ln(d)^{\,r-S+1}\right) \prod_{j=1}^{S} L^d\!\left(\sum_{\pm\lambda_i \in P_j} \pm\lambda_i \right)$$

since $S \le r$ and $L^d \le d$.

To prove this we assume, without loss of generality, that the P_j, $j=1,\ldots,S$ are enumerated in such a way that for each P_j exists a P_k (for some $k < j$) and a row of the table, such that P_j and P_k contain at least one element of this row. This is possible by the indecomposibility of the partition. Now let:

$$U_t := \left\{ \kappa \in P_q \,,\; q > t \text{ such that: } \kappa \text{ has a row-neighbour} \in P_{q'} \,,\; q' \le t \right\} \text{ and by}$$

$$V_t := \left\{ \kappa \in P_t \text{ such that: } \kappa \text{ has a row-neighbour} \in P_t \right\}$$

(where κ stands for some α or β of the Table I.1) and let

$$n_t := \frac{|V_t|}{2}\ ,\quad m_t := \left| U_{t-1} \cap P_t \right|,\quad m_1 := 0.$$

We claim that:

(1) for $t \leq S$ the integral over $\widetilde{\kappa}_i$, $i=1,\ldots,t$ of terms in the expression to be bounded, involving these variables, is less than or equal to some constant times

$$L^N\Big(-\sum_{\kappa\in U_t} c\kappa\Big)\prod_{i:=1}^{t} L^d\Big(\sum\nolimits_{\pm\lambda_j\in P_i}\pm\lambda_j\Big)\left[\frac{d\ \ln(N)\ \ln(c)}{c}\right]^{\sum_{i=1}^{t}(s_i-1)-(m_i-1)^+-n_i}\Big[\ln(d)\Big]^{\sum_{i=1}^{t}(m_i-1)^++n_i}$$

To see that the desired result follows from (1), observe that:

a) $(m_1-1)^+=0$, $m_t\geq 1$ for $t\geq 2$, since $U_{t-1}\cap P_t \neq \emptyset$, which follows from the enumeration we chose. This implies that $\sum_{i=1}^{t}(m_i-1)^+ = \sum_{i=2}^{t}(m_i-1)$.

b) $\sum_{t=1}^{S} n_t + \sum_{t=2}^{S} m_t = r$, since the first sum equals the number of rows occupied by elements of the same partition subset and the second that of rows occupied by elements of different partition subsets .

Thus:
$$\sum_{t=1}^{S} n_t + \sum_{t=1}^{S}(m_t-1)^+ \;=\; r-S+1.$$

These in turn yield a final bound of:

$$N\left[\prod_{j=1}^{S} L^d\Big(\sum\nolimits_{\pm\lambda_i\in P_j}\pm\lambda_i\Big)\right]\Big[d/c\ \ln(N)\ \ln(c)\Big]^{\sum_{i:=1}^{S}[(s_i-1)-(m_i-1)^+-n_i]}\Big[\ln(d)\Big]^{\sum_{i:=1}^{S}[(m_i-1)^++n_i]}$$

which is exactly of the desired order.

To check the validity of (1) it is sufficient to prove it a) for t=1 and b) for t+1, assuming it holds for t. Set $P_t := \{\kappa_1,\ldots,\kappa_{s_t}\}$. We indicate the proof of b) assuming further that $\kappa_1 \in U_{t-1}$, $\kappa_{s_t} \notin U_{t-1}$ (note that the chosen enumeration of the partition subsets guaranties $U_{t-1}\cap P_t \neq \emptyset$). The other case (which may be treated by similar arguments) is that $P_t \subseteq U_{t-1}$. Assume $\kappa_1 = a_p$, $\kappa_{s_t} = b_q$. Lemma A.1 ii) yields:

$$\int L^d(\lambda_p+\alpha_p)\ L^d\Big(-\lambda_q-\alpha_p-\sum_{j=2}^{s_t-1}\kappa_j\Big)\ L^N\Big(-\sum_{\kappa\in U_{t-1}} c\kappa\Big)\ L^N\Big(c\alpha_q - c\alpha_p - \sum_{j=2}^{s_t-1} c\kappa_j\Big)\ d\alpha_p \leq$$

$$O(d/c\ \ln(N)\ln(c))\ L^d\left(\lambda_p-\lambda_q-\sum_{j=2}^{s_t-1} c\kappa_j\right) L^N\left(c\alpha_q-\sum_{\kappa\in U_{t-1}\setminus P_t} c\kappa+\sum_{\kappa\in P_t\setminus U_{t-1}} c\kappa\right).$$

Notice that the $\kappa\in U_{t-1}\cap P_t$ do not appear any more in the argument of an L^N-factor (we call this (*)). Therefore they can be eliminated by using Lemma A.1.i). After having integrated with respect to / eliminated a $\kappa\in V_t$, (*) holds also for its row-neighbour. So the number of times (*) occurs while integrating with respect to elements of P_t is $(m_t-1)^++n_t$. The proof is completed by integrating with respect to / eliminating successively all $\kappa\in P_t$, bearing in mind the above remark. It yields that the power of the factor ln(d) in the bound will be $(m_t-1)^++n_t$, and the power of the factor $d/c\ \ln(N)\ln(c)$ will be $s_t-1-(m_t-1)^+-n_t$.
What remains is exactly $L^d\left(\sum_{\pm\lambda_j\in P_{t+1}}\pm\lambda_j\right)$. This finishes the proof of b) □

I.2.5 Proof of Theorem I.1

Under our assumptions Lemmata I.2 and I.4 yield that $\sqrt{T/d}\ \hat{f}_T\tilde{f}_d^{-1}(\lambda)$ and $\sqrt{T/d}\ I_T^*(\lambda)$ are asymptotically equivalent. On the other hand, using the cumulant method together with Lemmata I.9 and I.10, one gets that $\sqrt{T/d}\ [I_T^*(\lambda)-1]$ is asymptotically normal and has the desired covariance structure. This proves a). Our second assertion now follows from the fact that

$$\hat{f}_T f^{-1}-1-B_d=\tilde{f}_d f^{-1}\left[\hat{f}_T\tilde{f}_d^{-1}-1\right]=\left[\hat{f}_T\tilde{f}_d^{-1}-1\right]\left(1+O\left(d^{-[(r+\alpha)\wedge 1]}\ln^2(d)\right)\right)$$

by Lemma I.3. Finally c) is a direct consequence of b) and Lemma I.3. □

II. CONSTRUCTING A DIMENSION SELECTION CRITERION

II.1 INTRODUCTION, RESULTS

Our aim in this chapter is to construct a 'dimension selection criterion' for the parameter d of the Capon estimator, that is a procedure allowing the automatic, data adaptive, choice of this parameter.

In previous chapters we discussed the role of d as a smoothing parameter. In our examination of the asymptotic distribution of the Capon estimator we let $d \to \infty$, assuming this convergence is not too fast. This approach does not offer any solutions on the choice of d in a finite sample situation. Thus the need for a dimension selection criterion is obvious: a too big value of d will inflate the variance of the Capon estimator; a too small will lead to a larger bias. Therefore we try to estimate the value of d which minimizes a certain error criterion or discrepancy of the form squared bias + variance.

From Remark I.2 it is clear that the parameter c of the Capon estimator should be chosen as equal to 1, the value which minimizes the first order expansion term of the variance and the bias. Therefore in this and the next chapter we make the assumption that **c=1** and suppress c in our notation.

Obviously a procedure aiming to the choice of d will depend on the distance measure (discrepancy) between f and $\hat{f}_{d,T}$. For this purpose we use the Whittle discrepancy $\Delta(f, \hat{f}_{d,T})$, defined in Section II.1.1 below, and obtain an AIC-like criterion for the choice of d (Akaike (1970)). Its development is based on a stochastic expansion of this discrepancy whose leading terms do not depend on f (Section II.1.2). The theoretically interesting contribution of this chapter, proven in the main Lemmata II.1 and II.2, is that this expansion is carried through up to second order terms included. These second order terms play a considerable role in finite sample situations (see Figure II.1). Our criterion will be the following:

$$\hat{d} := \underset{d \le d_{max}}{\operatorname{argmin}} (2\pi)^{-1} \int \ln \hat{f}_{d,T}(\lambda)\, d\lambda + \mu_d, \text{ where } \mu_d = \mu_{d,T} := 1 + \frac{d}{N} + \left(\frac{d}{N}\right)^2,$$

where $N \equiv N_d$ as in (0.5).

As in the AIC, a quantity of the form 'estimated innovations variance' + 'penalty' is minimized over a parameter region $d \le d_{max}$. The typical condition on the speed of $d_{max} \to \infty$ is $d_{max}^2 T^{-1} \to 0$ (e.g. Shibata (1980)). We relax this to $d_{max}^{1+\varepsilon} T^{-1} \to 0$ for some $\varepsilon > 0$. This is important in finite sample situations where one is interested in allowing 'large enough' values of d.

Further, based on these considerations, we propose and discuss a (quasi) bias correction of the Capon estimator (Section II.1.3). Our results carry over to the AIC for autoregressive least squares spectral estimators yielding an improvement on it (Section II.1.4).

Questions concerning the stochastic properties of $\Delta(f,\hat{f}_{d,T})$ and $\Delta(f,\hat{f}_{\hat{d},T})$ (where $\hat{d}$ is the dimension selected according to the criterion) will be dealt with in Chapter III.

II.1.1 The discrepancy

For measuring the distance between the Capon estimator and the spectrum f we use the Whittle discrepancy which is defined as follows:

$$\Delta(f,\hat{f}_{d,T}) := \frac{1}{2\pi} \int \ln\left(\frac{\hat{f}_{d,T}}{f}\right)(\lambda) + \frac{f}{\hat{f}_{d,T}}(\lambda) - 1 \, \mathbf{d}\lambda$$

The Whittle discrepancy (between a parametric estimator, say $\hat{f}_T(\theta)$, and the spectrum f) has been often used for defining minimum distance estimators which are also called quasi-maximum likelihood (e.g. Hosoya and Taniguchi (1982)).

The main <u>motivation</u> for its use is that in the Gaussian case the Whittle discrepancy $\Delta(f,g)$ of a spectral density g with respect to a spectral density f may be considered to be an approximation of the Kullback-Leibler distance of the distributions induced by these spectral densities. Let us make this more precise.

Suppose that observations $Y_1,\ldots,Y_T$ are drawn from a (zero mean) Gaussian process whose spectral density is f. Then, writing $L_{(y)}(f)$ for their likelihood, when the true spectrum is f and $B_T(f)$ for the T×T Toeplitz matrix associated with f, the Kullback-Leibler discrepancy (multiplied by 2/T) of $L_{(.)}(g)$ relative to the true distribution of the process is defined as (Brockwell and Davis (1987) § 9.3):

$$d_T(f,g) := -\frac{2}{T} E_f \ln L_{(Y)}(g) + \frac{2}{T} E_f \ln L_{(Y)}(f).$$

We have

$$-\frac{2}{T} E_f \ln L_{(.)}(g) = \ln(2\pi) + T^{-1} \ln\left[\det B_T(g)\right] + T^{-1} \operatorname{tr}\left[B_T(f)\, B_T^{-1}(g)\right].$$

Thus

$$d_T(f,g) = T^{-1} \ln\left[\det B_T^{-1}(f) B_T(g)\right] + T^{-1} \operatorname{tr}\left[B_T(f)\, B_T^{-1}(g)\right] - 1 .$$

Now using the Szegö approximation (Grenander and Szegö (1958) § 6.2) one obtains that, when $T \to \infty$

$$d_T(f, g) - \frac{1}{2\pi} \int \ln\left[\frac{g}{f}(\lambda)\right] + \frac{f}{g}(\lambda) - 1 \, d\lambda \quad \to 0.$$

The error of this approximation has been studied in the case where f and g are fixed functions (e.g. Coursol and Dacuhna-Castelle (1982)). But for the approximation of $d_T(f,\hat{f}_{d,T})$ by $\Delta(f,\hat{f}_{d,T})$ things are more complicated: first $g = \hat{f}_{d,T}$ is random and secondly $\hat{f}_{d,T}$ approximates f (yielding that both discrepancies tend stochastically to 0). Therefore it would be interesting to have a result of the type:

$$\frac{d_T(f,\hat{f}_{d,T})}{\Delta(f,\hat{f}_{d,T})} \to_P 1 \text{ uniformly in } d \le d_{max}$$

It is an open question if this holds. However it does not affect our reasoning, since it plays a role only in the motivation of the Whittle discrepancy. In the Appendix (Lemma A.6) we show that a weaker result holds, namely that:

$$\left| d_T(f,\hat{f}_{d,T}) - \Delta(f,\hat{f}_{d,T}) \right| = o_p(d/T) \text{ when } d,T \to \infty.$$

II.1.2 Constructing a penalty term (from the discrepancy to the criterion)

In order to construct a dimension selection criterion, once a discrepancy is given, one could try to select the value of d minimizing $\Delta(f,\hat{f}_{d,T})$. But this quantity is unknown, since it involves the unknown spectrum f. A possible strategy, proposed e.g. by Linhart-Zucchini (1986), is to aim to the selection of d which would in mean do best, minimizing $E\,\Delta(f,\hat{f}_{d,T})$. In order to construct a dimension selection criterion based on this approach, one needs an asymptotic expansion of $E\,(2\pi)^{-1} \int \left[f\,\hat{f}_{d,T}^{-1}\right](\lambda)\, d\lambda$ which does not depend on f. In many papers dealing with similar problems, e.g. the derivation of the AIC, such an approximation is obtained by arguments of the following type: '$(2\pi)^{-1} \int \left[f\,\hat{f}_{d,T}^{-1}\right](\lambda)\, d\lambda$ is asymptotically chi-square (or F) distributed, thus its expectation can be approximated by the expectation of this chi-square (or F) distribution' (see e.g. Hurvich and Tsai (1989) and Brockwell and Davis (1987) § 9.3). Since however we want to avoid this incorrect type of argument -even in the motivation- and on the other hand we could not obtain an asymptotic expansion of $E\,(2\pi)^{-1} \int \left[f\,\hat{f}_{d,T}^{-1}\right](\lambda)\, d\lambda$, we adopt an indirect argument.

First observe that $(2\pi)^{-1} \int \left[f\,\hat{f}_{d,T}^{-1}(\lambda)\right] d\lambda = d^{-1}\, \mathrm{tr}\left(\hat{\Gamma}^*_{d,T}\right)^{-1}$. Thus the minimization of $\Delta(f,\hat{f}_{d,T})$ is equivalent to the minimization of $(2\pi)^{-1} \int \ln \hat{f}_{d,T}(\lambda)\, d\lambda + d^{-1}\, \mathrm{tr}\left(\hat{\Gamma}^*_{d,T}\right)^{-1}$. Now instead of aiming to a minimization of the latter, we try to minimize a stochastic approximation of it namely $(2\pi)^{-1}\int \ln \hat{f}_{d,T}(\lambda)\, d\lambda + D_{d,T}$, where $D_{d,T} := d^{-1}\sum_{k:=0}^{5} \mathrm{tr}\left[I-\hat{\Gamma}^{*}_{d,T}\right]^{k}$. This approximation is justified by the Lemma II.1 below, obtained by an expansion of $\left(\hat{\Gamma}^{*}_{d,T}\right)^{-1}$ to higher order terms.

Next we estimate $D_{d,T}$, which still involves the unknown matrix Γ_d, by its expectation and obtain a dimension selection criterion by minimizing $(2\pi)^{-1}\int \ln \hat{f}_{d,T}(\lambda)\, d\lambda + E\, D_{d,T}$. This is possible because $E\, D_{d,T} \approx 1 + d\, N^{-1} + d^{2} N^{-2}$, N as in (0.5), does not depend asymptotically on f, as shown in Lemma II.2 below.

Accordingly, the dimension selection criterion $\hat{d}$ we propose consists of taking:

$$\hat{d} := \underset{d\le d_{max}}{\mathrm{argmin}}\, (2\pi)^{-1}\int \ln \hat{f}_{d,T}(\lambda)\, d\lambda + \mu_d, \text{ where } \mu_d = \mu_{d,T} := 1 + \frac{d}{N} + \left(\frac{d}{N}\right)^{2}.$$

Our development is justified by Lemmata II.1 and II.2 below. For their statement we make the following assumptions:

(D) Assume that (A), (B) and (C) hold, that c=1 and let $\beta := \frac{r+\alpha}{1+r+\alpha}$, r, α as in (B). Finally assume that $\frac{d_{max}^{1+\varepsilon}}{T} \to 0$ for some $\varepsilon>0$ and for a sequence $d_{max} \to \infty$.

(D1) Assume that (A), (B) and (C) hold, that c=1 and let $\beta := \frac{r+\alpha}{1+r+\alpha}$, r, α as in (B). Finally assume that $\frac{d \ln(T) \ln(d)}{T} \to 0$ for some sequence $d=d_T$.

Lemma II.1 Assume (D). Then we have

$$\sup_{d\le d_{max}} \left| d^{-1}\, \mathrm{tr}\left[\hat{\Gamma}^{*\,-1}_{d,T}\right] - D_{d,T} \right| = O_P\left(\left[\frac{d_{max}^{1+\varepsilon}}{T} \ln(T)\right]^{3}\right). \qquad \bullet$$

Lemma II.2 Assume that (D1) holds with $d\to\infty$, β as in (D1) and $N=N_{d,T}$ as in (0.5). Then:

$$E\, D_{d,T} = 1 + \left[\frac{d}{N} + \left(\frac{d}{N}\right)^{2}\right]\left[1+O\left(d^{-\beta}\ln(N)\right)\right]. \qquad \bullet$$

Remark II.1 The above considerations, which motivated the dimension selection criterion, do not give any information about its properties. Thus they do not, for example, give any answer to the question of how close $\Delta(f,\hat{f}_{\hat{d},T})$ is to $\inf_{d\le d_{max}} \Delta(f,\hat{f}_{d,T})$.

This question will be dealt with in the next chapter, where we will show that these two quantities are asymptotically equivalent (asymptotic efficiency of the dimension selection criterion). This will be the actual justification for the criterion (see Theorems III.1 and III.2 as well as Lemma III.3 and the remark following it).

Remark II.2 The penalty term μ_d may be regarded as a multiplicative (quasi) bias correction for $\hat{f}_{d,T}$, since $-(\mu_d+1)$ is an additive bias correction for $D_{d,T}$ which is a stochastic approximation of $(2\pi)^{-1} \int \left[f \hat{f}_{d,T}^{-1} \right](\lambda)\, d\lambda$ (see II.1.3).

Let us discuss some of the consequences of the last lemma. First we observe that the expansion $E\, D_{d,T} \cong \mu_d$ is valid only for $d \to \infty$. It is obvious that for the dimension selection criterion to be 'good' -whatever this may mean- it is necessary that the expansion is a 'good' approximation at least in a neighbourhood of the 'truly optimal parameter', say d^* (we make this more precise in Chapter III.). Thus $d^* \to \infty$ is necessary for the dimension selection criterion to be 'good'. In Chapter III we show that $d^* \to \infty$ is almost sufficient (see Theorem III.2) for the dimension selection criterion to be efficient, in the sense already mentioned in Remark II.1.

In order to get a better impression of the quality of the approximation $E\, D_{d,T} = \mu_d + R_d$, where R_d is the error, let us regard the order of magnitude of R_d. It is easy to see that this order depends on the rapidity of $d \to \infty$. To make this clear, note that Lemma II.2 yields an error term of order $d^{1-\beta} N^{-1}$. Thus when $T\, d^{-1-\beta} \to 0$ we have $R_d = o\left[d^2 N^{-2} \right]$. But how big is R_d when d is close to d^*? The next lemma helps us answer this question.

Lemma II.3 Assume that (D1) holds. Then we have

$$E\, D_{d,T} = 1 + \frac{d}{N} + \frac{d}{N} \tilde{B}_d' + O\left[\left(\frac{d}{N}\right)^2 \ln(T) \right] + O\left[d T^{-\gamma-1} \right] + O\left[T^{-1} \ln(T) \ln(d) \right],$$

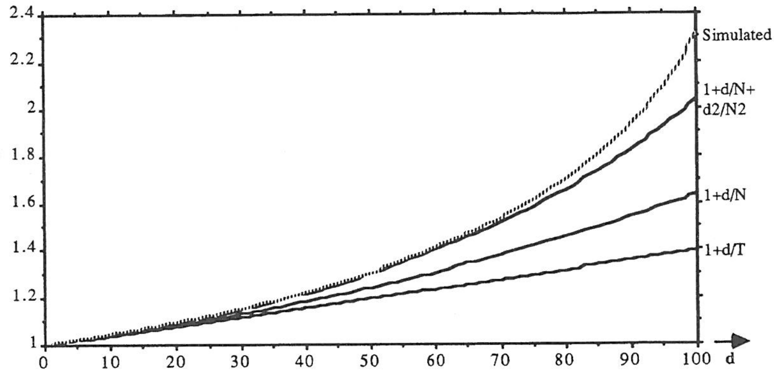
where
$$\tilde{B}_d' := \frac{1}{2\pi} \int \left[\tilde{f}_d^{-1} f - 1 \right]^2 (\lambda)\, d\lambda$$

and $\gamma := \alpha$ if $r=0$ and $\gamma := 1$ else (r, α as in (D1)). •

Unfortunately for d close to d^* we have $\tilde{B}_d' \approx dN^{-1}$. Thus for d close to d^* Lemma II.3 indicates that $R_d = O\left[d^2 N^{-2} \right]$. So, why do we consider it an improvement to include quadratic terms in μ_d? First, because the coefficient 1 for $d^2 N^{-2}$ in μ_d goes in the right direction: $\frac{d}{N} \tilde{B}_d'$ is positive. Secondly, because the inclusion of the quadratic term in μ_d yields $R_d = o\left[d^2 N^{-2} \right]$ for the large values of d, e.g. d close to d_{max} when d_{max} tends rapidly to infinity: $T\, d_{max}^{-1-\beta} \to 0$. In this case this prevents us from choosing a too large value of d. A look at Figure II.1 convinces us that this is a real danger when we use only the linear term in μ_d.

In Figure II.1 we show the performance of different penalty terms for the selection criterion of the Capon estimator compared with the corresponding simulated quantity under white noise (500 samples, T=256). It is evident that the linear penalty terms 1+d/N and 1+d/T are good approximations only for small values of d whereas μ_d is a much better approximation also for larger values of d.

Fig. II.1 Simulated and 'estimated' ∑f/f^ for the Capon estimator when f= white noise and T=256.



II.1.3 (Quasi) bias correction of the estimator

Based on Lemmata II.1 and II.2, one may propose a modification of the Capon estimator. As discussed above $(2\pi)^{-1} \int f\, \hat{f}_{d,T}^{-1}(\lambda)\, d\lambda \;\cong \mu_{d,T}$. Thus by setting

$$\hat{\hat{f}}_{d,T} := \mu_{d,T}\, \hat{f}_{d,T}$$

we obtain $(2\pi)^{-1} \int f\, \hat{\hat{f}}_{d,T}^{-1}(\lambda)\, d\lambda \cong 1$. Following Remark II.2 we call this modification a (quasi) bias correction. In the next lemma we show that, in mean, the estimator $\hat{\hat{f}}_{d,T}$ has a discrepancy to f smaller than the one $\hat{f}_{d,T}$ has.

Lemma II.4 Assume that (A), (B) and (C) hold. Further assume that $\frac{d^{1+\varepsilon}}{T} \to 0$ for some $\varepsilon>0$ and a sequence $d=d_T \to \infty$. Then

$$E\Delta\left(f,\hat{\hat{f}}_{d,T}\right) \leq E\Delta\left(f,\hat{f}_{d,T}\right) - \frac{1}{2}\left(\frac{d}{N}\right)^2 + o\left(\frac{d}{N}\right)^2. \qquad \bullet$$

Note that the dimension selection criterion for the corrected estimator would be

$$\hat{\hat{d}} := \underset{d \leq d_{max}}{\text{argmin}} (2\pi)^{-1} \int \ln \hat{\hat{f}}_{d,T}(\lambda)\, d\lambda \; = \underset{d \leq d_{max}}{\text{argmin}} (2\pi)^{-1} \int \ln \hat{f}_{d,T}(\lambda)\, d\lambda + \ln(\mu_d)$$

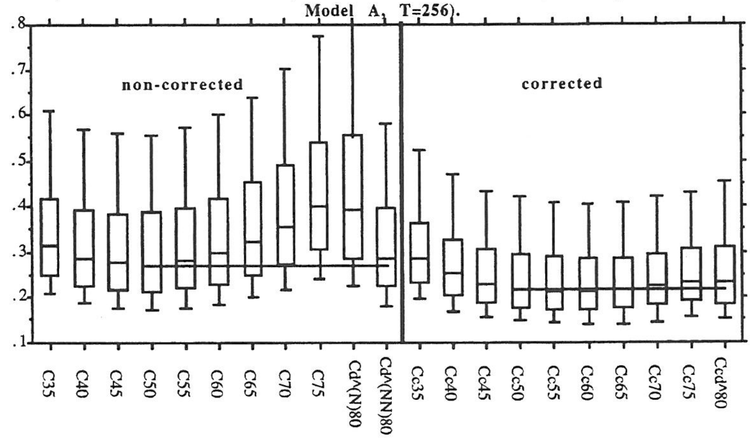
and compare this with the dimension selection criterion for the non-corrected estimator

$$\hat{d} := \underset{d \leq d_{max}}{\text{argmin}} (2\pi)^{-1} \int \ln \hat{f}_{d,T}(\lambda)\, d\lambda + \mu_d.$$

In the next Figure II.2 we show the box plots for the discrepancies of the Capon estimator, non-corrected and corrected for fixed and estimated orders. The simulation consists of 500 samples, each of size T=256 drawn from 'Model A' which is given in Chapter V. Note that:

- the corrected Capon estimator performs better than the non-corrected.
- dimension selection criteria, including quadratic terms, yield discrepancies very close to the optimal d (lines in Figure II.2). (For the corrected estimator, the 'quadratic term' is implicitly present in the correction).
- the dimension selection criterion including only the linear term performs significantly worse.

Fig. II.2 Box-plots of discrepancies for (segment-) Capon-estimators (corrected and non-corrected) for d fixed and estimated (500 samples under Model A, T=256).



Cd: (Segment-) Capon spectral estimator for fixed dimension d.
Cd^(N)80 : Cd, d estimated (p≤80) with penalty term d/N.
Cd^(NN)80 : the same with penalty term d/N + (d/N)2.
Ccd: Cd, multiplied with 1+ d/N + (d/N)2. Ccd^80: Ccd, d estimated (p≤80) (with penalty term 0).

II.1.4 Carrying over to autoregressive estimators and the AIC

Our results carry over to autoregressive estimators yielding an improvement of the AIC as an order selection criterion and a (quasi) bias correction for autoregressive least squares estimators. Let $\hat{f}^{(AR)}_{p,T}$ be the autoregressive least squares estimator of order p and $\hat{\sigma}^2_{p,T}$ the estimated 'innovations variance' defined as follows (Kay and Marple (1981)):

Let $\mathbf{a}^t_p := (1, a_{p,1}, \ldots, a_{p,p})$ and define the 'residuals' $e_{p,t}(\mathbf{a}) := X_t + a_{p,1}X_{t-1} + \ldots + a_{p,p}X_{t-p}$ $t = p+1, \ldots, T$. Further define $\hat{\sigma}^2_{p,T} := \inf_{\mathbf{a}} \sum_{t=p+1}^{T} e^2_{p,t}(\mathbf{a})$ the infimum being attained at $\hat{\mathbf{a}}_{p,T}$.

Then

$$\hat{f}^{(AR)}_{p,T}(\lambda) := \hat{\sigma}^2_{p,T} / 2\pi \; \left| \hat{\mathbf{a}}^t_{p,T}\, b_\lambda \right|^{-2}.$$

The link for carrying over our results concerning the Capon estimator to the autoregressive least squares estimators is Burg's relation (see Section 0.5), which, in this case, has to be slightly modified, as elementary calculations show, into the following:

$$\hat{f}_{d,T}(\lambda) := \left\{ d^{-1} \sum_{p=0}^{d-1} \left[\hat{f}^{(AR)}_{p,T+p-d+1}(\lambda) \right]^{-1} \right\}^{-1}$$

It follows:

$$\hat{f}^{(AR)\,-1}_{p,T} = (p+1)\, \hat{f}^{-1}_{(p+1),T} - p\, \hat{f}^{-1}_{p,T-1}$$

For constructing an order selection criterion one could, as above, try to minimize

$$\frac{1}{2\pi} \int \ln\left(\hat{f}^{(AR)}_{p,T}\right)(\lambda)\, d\lambda + \frac{1}{2\pi} \int \frac{f}{\hat{f}^{(AR)}_{p,T}}(\lambda)\, d\lambda$$

But the second term in the sum above can be approximated (with the same magnitude of error as in Lemma II.2) by

$$(p+1)\, \mu_{(p+1),T} - p\, \mu_{p,T-1} = 1 + 2\frac{p}{N} + 4\left(\frac{p}{N}\right)^2 + O\left[\left(\frac{p}{N}\right)^3 \vee N^{-1}\right],$$

which follows from $(2\pi)^{-1} \int f\, \hat{f}^{-1}_{p,T}(\lambda)\, d\lambda \cong \mu_{p,T}$ together with Burg's relation. For the first term we have $\ln\left(\hat{\sigma}^2_{p,T}\right) = (2\pi)^{-1} \int \ln\left(\hat{f}^{(AR)}_{p,T}\right)(\lambda)\, d\lambda$ (follows e.g. from Brockwell, Davis (1987) § 5.8). The resulting order selection criterion $\hat{p}$ is:

$$\hat{p} := \underset{p \le p_{max}}{\operatorname{argmin}}\; \ln\left(\hat{\sigma}^2_{p,T}\right) + 2\frac{p}{N} + 4\left(\frac{p}{N}\right)^2$$

Observe that the first order expansion term $2\frac{p}{N}$ resembles the penalty term of the AIC: $2\frac{p}{T}$. The use of N instead of T is due to the fact that our expansion was done for the autoregressive least squares estimator, that is for the conditional maximum likelihood estimator.

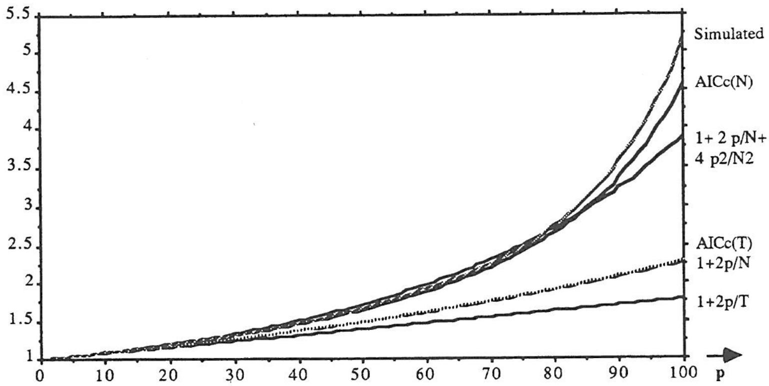
In 1989 Hurvich and Tsai proposed AICc as improvement of the AIC, which is known to be biased in finite sample situations. In the AICc the penalty term equals

$$\frac{1+p/T}{1-(p+2)/T} = 1+2\frac{p}{T}+4\left(\frac{p}{T}\right)^2+O\left[\left(\frac{p}{T}\right)^3 \vee T^{-1}\right].$$

Observe that, if one substitutes N instead of T in the AICc, the resulting penalty term equals the one above developed up to terms of order $O\left[p^3T^{-3} \vee T^{-1}\right]$.

In Figure II.3 we show the performance of different penalty terms for the order selection criterion of the autoregressive LS estimator compared with the corresponding simulated quantity under white noise (500 samples, T=256). With AICc(T) we denote the AICc given above and with AICc(N) the same quantity with N substituted instead of T. Note that AICc(N) as well as $1+2\,p/N+4\,(p/N)^2$ are good approximations of the simulated value even for large values of p whereas the others (including AICc(T)) are not.

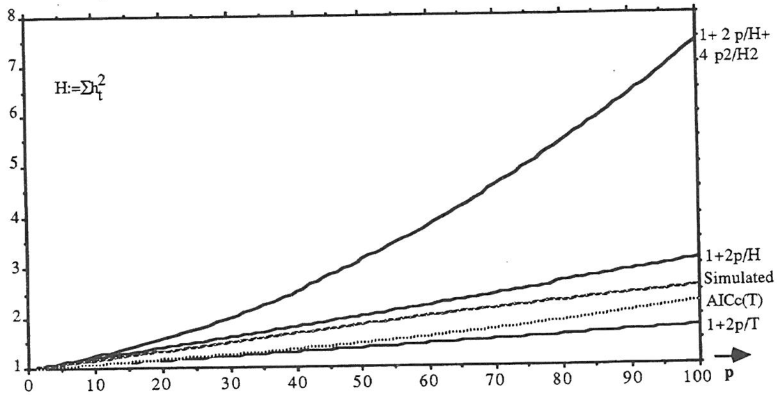
Fig. II.3. Simulated and 'estimated' $\Sigma f/f^\wedge$ for the AR-LS estimator when f=white noise and T=256.



In the literature dealing with order selection criteria enough attention has not been paid to the following matter: 'T' (in the penalty term) should be substituted by other quantities, depending on the type of the autoregressive estimator used. The reason for this negligence is obvious: the effect of the above substitution is reflected only in terms of order $O(p^2T^{-2})$, which until now were regarded as negligible and whose effect is not reflected in known efficiency properties (see Lemma III.3 below and the remark following it). The only justification for such a substitution are expansions of the type used in Section II.1.2, which make it 'seem appropriate'. Its role in finite sample situations is obvious: set T=256 and p=80 then p/T=0.31 and p/N=0.45 making a difference of about 50%.

Our conjecture is that for the Yule-Walker autoregressive estimator based on tapered data (defined in Section V.1.3) one should use the sum of the squared taper instead of N or T, whereas one should use T (no substitution) for the maximum likelihood autoregressive estimator.
However the next Figure II.4 indicates that this is not the only problem. A comparison with Figure II.3 shows that the simulated values for the Toeplitz autoregressive estimator have really a different form than for the LS: it seems that the higher order terms in the expansion are different for different estimators. Note that the linear expansion approximates the simulated values for the Toeplitz autoregressive estimator much better than for the LS estimator. This explains why the AIC works better with Toeplitz than LS estimators.

Fig II.4 Simulated and 'estimated' $\Sigma f/f^{\wedge}$ for the 100% tapered Toeplitz AR estimator when f=white noise (500 samples, T=256).



The (quasi) bias correction of the Capon estimator and the modification of the dimension selection criterion (see Section II.1.3) can also be applied to the autoregressive estimator:

$$\hat{\hat{f}}_{p,T}^{(AR)} := \left[1 + 2\frac{p}{N} + 4\left(\frac{p}{N}\right)^2 \right] \hat{f}_{p,T}^{(AR)}$$

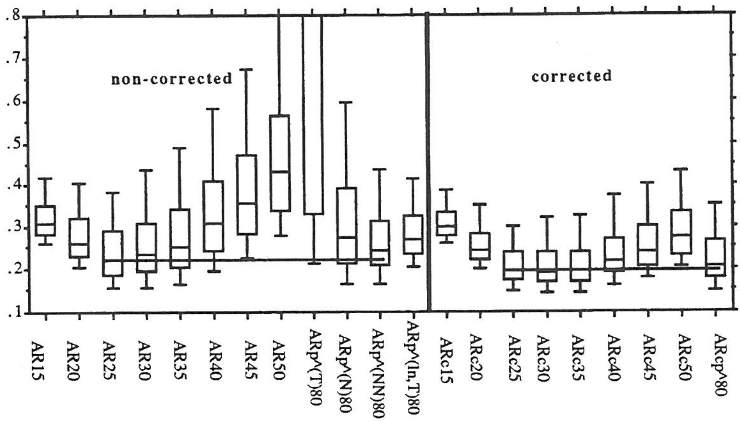
The order selection criterion for the corrected estimator would be

$$\hat{\hat{p}} := \underset{d \le p_{max}}{\text{argmin}} \, (2\pi)^{-1} \int \ln \hat{\hat{f}}_{p,T}^{(AR)}(\lambda)\, d\lambda \;=\; \underset{p \le p_{max}}{\text{argmin}} \ln\left(\hat{\sigma}_{p,T}^2\right) + \ln\left[1 + 2\frac{p}{N} + 4\left(\frac{p}{N}\right)^2\right]$$

In Figure II.5 we show the box plots for the discrepancies of the autoregressive LS estimator, non-corrected and corrected for fixed and estimated orders. The simulation consists of 500 samples, each of size T=256 drawn from 'Model A' which is given in Chapter V. Note that:

- The corrected estimators perform better than the non-corrected.
- Dimension selection criteria including quadratic terms yield discrepancies very close to the optimal d (lines in Fig II.5). (For the corrected estimator, the 'quadratic term' is implicitly present in the correction).
- The dimension selection criterion including only the linear term 2 d/N performs considerably worse. The simple AIC with penalty term 2 p/T performs very badly. The performance of the BIC is surprisingly good.

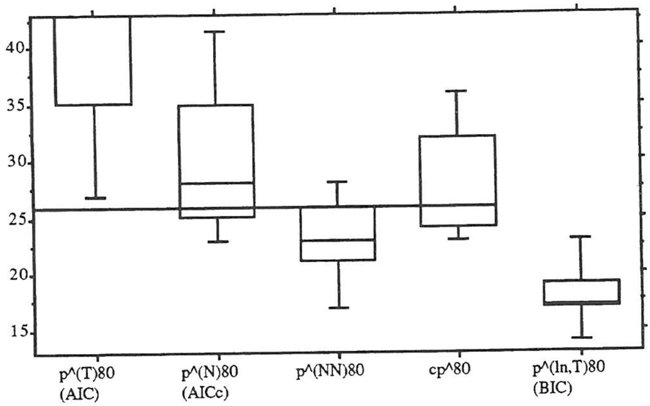
Fig II.5 Box-plots of discrepancies for Autoregressive-LS-estimators (corrected and non-corrected) for d fixed and estimated (500 samples under Model A, T=256).



ARp: Autoregresive Least-squares spectral estimator of fixed order p.
ARp^(T)80 : ARp, p estim., penalty = 2p/T (AIC).
ARp^(N)80 : ARp, p estim., penalty = 2p/N (almost AICc(T)).
ARp^(NN)80 : ARp, p estim., penalty = 2p/N + 4 (p/N)2.
ARp^(ln, T)80 : ARp, p estim., penalty = ln(T) p/T (BIC).
ARcp: ARp, multiplied with 1+ 2p/N +4 (p/N)2.
ARcp^80: ARcp, p estim., penalty= 0.

In Figure II.6 we show the box plots <u>for the selected orders</u> of the autoregressive LS estimator for different dimension selection criteria. The simulation consists of 500 samples, each of size T=256 drawn from 'Model A' which is given in Chapter V.

Fig II.6 **Box-Plots of the estimated order for the AR-LS estimator for different order-selection-criteria (500 samples under model A , T=256).**



$\hat{p}(T)80 := \text{arcmin}\ (p\leq 80) \int \ln(\hat{f}_p) + 2p/T$, $\hat{p}(N)80 := \text{arcmin}\ (p\leq 80) \int \ln(\hat{f}_p) + 2\,p/N$,
$\hat{p}(NN)80 := \text{arcmin}\ (p\leq 80) \int \ln(\hat{f}_p) + 2p/N + 4\,(p/N)^2$,
$c\hat{p}80 := \text{arcmin}\ (p\leq 80) \int \ln(\hat{f}_p) + \ln(1 + 2p/N + 4\,(p/N)^2)$
$\hat{p}(\ln,T)80 := \text{arcmin}\ (p\leq 80) \int \ln(\hat{f}_p) + \ln(T)\,p/T$.

Note that:

- the only criterion whose median is 26, the truly optimal order, is the one of the corrected estimator. It estimates the optimal order better than BIC.
- the dimension selection criterion including only the linear term 2 d/N performs worse. The simple AIC with penalty term 2 p/T performs very badly.

II.2 DETAILED RESULTS, PROOFS

In this chapter we prove the results stated in Section II.1. In the first Section II.2.1 we give the proof of Lemma II.1. In the second II.2.2, the proof of Lemma II.2, which is the most tedious, and we leave the proof of the remaining lemmata for the final Section II.2.3.

II.2.1 Asymptotic expansion (proof of Lemma II.1)

In this section we prove Lemma II.1, the stochastic approximation of $d^{-1}\mathrm{tr}\left[\hat{\Gamma}^{*}_{d,T}{}^{-1}\right]$ by $D_{d,T}$.

Proof of Lemma II.1 Let the following event be denoted by $A_{d,T} := \left\{ \left\| \hat{\Gamma}^{*}_{d,T} - I \right\| < 1/2 \right\}$. Then on $A_{d,T}$ we may expand

$$\hat{\Gamma}^{*\,-1}_{d,T} = \left[I - \left(I - \hat{\Gamma}^{*\,-1}_{d,T}\right)\right]^{-1} = \sum_{j=0}^{\infty} \left(I - \hat{\Gamma}^{*}_{d,T}\right)^{j}$$

This implies $$\left| d^{-1}\, tr\left[\hat{\Gamma}^{*\,-1}_{d,T}\right] - D_{d,T}\right| \le 2 \left|\left| \hat{\Gamma}^{*}_{d,T} - I \right|\right|^{6} \text{ on } A_{d,T}.$$

Recall (Lemma I.2) for any $k \in \mathbf{Z}^{+}$, $\eta>0$ and $\delta>0$ we have

$$P\left(A^{C}_{d,T}\right) \le 2^{2k}\, E \left|\left| \hat{\Gamma}^{*}_{d,T} - I \right|\right|^{2k} = O\left(d^{1+\delta} \ln(T) \ln(d)/ T\right)^{k}, \text{ yielding:}$$

$$P\left\{ \left[T\, d^{-1-\varepsilon}_{max} \ln^{-1}(T)\right]^{3} \sup_{d \le d_{max}} \left| d^{-1}\, tr\left[\hat{\Gamma}^{*\,-1}_{d,T}\right] - D_{d,T}\right| \ge \eta \right\} \le$$

$$\sum_{d=1}^{d_{max}} P\left\{ A^{C}_{d,T}\right\} + \sum_{d=1}^{d_{max}} P\left\{ \left(T\, d^{-1-\varepsilon}_{max} \ln^{-1}(T) \left|\left| \hat{\Gamma}^{*}_{d,T} - I \right|\right|^{2} \right)^{3k} \ge \eta^{k} \right\} \le$$

$$O\left(d^{1/k}_{max}\, T^{-1}\, d^{1+\delta}_{max} \ln^{2}(T)\right)^{k} + O\left(d^{1/(3k)}_{max}\; T\, d^{-1-\varepsilon}_{max} \ln^{-1}(T)\;\; T^{-1}\, d^{1+\delta}_{max} \ln(T) \ln(d_{max})\right)^{3k} \rightarrow 0 \,,$$

if δ and k are chosen such that $\delta + k^{-1} < \varepsilon/4$. □

II.2.2 Approximating the expectation of $\mathbf{D}_{d,T}$ *(Proof of Lemma II.2)*

In this section we prove Lemma II.2. In order to calculate the expectation of $D_{d,T}$ one has to calculate the expectation of $tr\left[\hat{\Gamma}^{*}_{d,T} - I\right]^{K}$ for k=2,...,5. As already shown in Proposition I.7

$$E\, tr\left[\hat{\Gamma}^{*}_{d,T} - I\right]^{K} = \sum_{ap\backslash *,\,(K)} V\left(\mathcal{P}^{(K)}\right),$$

where we adopt the same notation as in Table I.1. The proof will consist of the following steps:

i) first we show that for each K the sum is dominated by some partitions. These are the following:

K=2: $\mathcal{P}^{(2)}_{1} = \left\{ (\alpha_1,\beta_2), (\alpha_2,\beta_1) \right\}$

K=3: $\mathcal{P}^{(3)}_{1} = \left\{ (\alpha_1,\beta_2), (\alpha_2,\beta_3), (\alpha_3,\beta_1) \right\}$ and $\mathcal{P}^{(3)}_{2} = \left\{ (\alpha_1,\beta_3), (\alpha_2,\beta_1), (\alpha_3,\beta_2) \right\}$

K=4: $\mathcal{P}^{(4)}_{1} = \left\{ (\alpha_1,\beta_2), (\alpha_2,\beta_1), (\alpha_3,\beta_4), (\alpha_4,\beta_3) \right\}$

$\mathcal{P}^{(4)}_{2} = \left\{ (\alpha_1,\beta_4), (\alpha_4,\beta_1), (\alpha_2,\beta_3), (\alpha_3,\beta_2) \right\}$

$\mathcal{P}^{(4)}_{3} = \left\{ (\alpha_1,\beta_3), (\alpha_3,\beta_1), (\alpha_2,\beta_4), (\alpha_4,\beta_2) \right\}$

These can be represented graphically as follows:

$$\mathcal{P}_1^{(2)}: \begin{array}{cc}\alpha & \beta \\ \beta & \alpha\end{array} \qquad \mathcal{P}_1^{(3)}: \begin{array}{cc}\alpha & \gamma \\ \beta & \alpha \\ \gamma & \beta\end{array} \qquad \mathcal{P}_2^{(3)}: \begin{array}{cc}\gamma & \alpha \\ \alpha & \beta \\ \beta & \gamma\end{array} \qquad \mathcal{P}_1^{(4)}: \begin{array}{cc}\alpha & \beta \\ \beta & \alpha \\ \gamma & \delta \\ \delta & \gamma\end{array} \qquad \mathcal{P}_2^{(4)}: \begin{array}{cc}\alpha & \beta \\ \gamma & \delta \\ \delta & \gamma \\ \beta & \alpha\end{array} \qquad \mathcal{P}_3^{(4)}: \begin{array}{cc}\alpha & \beta \\ \gamma & \delta \\ \beta & \alpha \\ \delta & \gamma\end{array}$$

All other partitions contributing to the sum will be shown to be of lower order.

ii) In a second step we show that $\mathbf{V}\left(\mathcal{P}_j^{(K)}\right)$, K=2,...,4, j=1,...,K-1 do not asymptotically depend on f -up to a multiplicative error of $O\left(d^{-\beta}\right)$- and we evaluate them by setting $f := \frac{1}{2\pi}$.

Thus Lemma II.2 will be a direct consequence of the following two lemmata:

Lemma II.5 Assume that (D1) holds. Then

i) If $\mathcal{P}^{(2)} \neq \mathcal{P}_1^{(2)}$ then $d^{-1}\,\widetilde{\mathbf{V}}\left(\mathcal{P}^{(2)}\right) = O\left[\,N^{-1}\ln(N)\ln(d)\right]$.

ii) If $\mathcal{P}^{(3)} \notin \left\{\mathcal{P}_1^{(3)}, \mathcal{P}_2^{(3)}\right\}$ then $d^{-1}\,\widetilde{\mathbf{V}}\left(\mathcal{P}^{(3)}\right) = O\left[\,d\,N^{-2}\ln^2(N)\ln(d)\,\right]$

iii) $\mathcal{P}^{(4)} \notin \left\{\mathcal{P}_1^{(4)}, \mathcal{P}_2^{(4)}, \mathcal{P}_3^{(4)}\right\}$ then $d^{-1}\,\widetilde{\mathbf{V}}\left(\mathcal{P}^{(4)}\right) = O\left[\,d\,N^{-2}\ln^2(N)\ln^2(d)\,\right]$

iv) $d^{-1}\,\widetilde{\mathbf{V}}\left(\mathcal{P}^{(5)}\right) = O\left[\,d^3 N^{-3}\ln^3(N)\ln^2(d)\,\right]$ for any $\mathcal{P}^{(5)}$. •

and

Lemma II.6 Assume that (D1) holds. Then

i) $d^{-1}\,\mathbf{V}\left(\mathcal{P}_2^{(2)}\right) = \frac{d}{N}\left[\,1 + O\left(d^{-\beta}\ln(d)\right)\right]$

ii) $d^{-1}\,\mathbf{V}\left(\mathcal{P}_j^{(K)}\right) = \frac{d^2}{N^2}\left[\,1 + O\left(\frac{d}{N}\right)\right]\left[\,1 + O\left(d^{-\beta}\ln(d)\right)\right]$, for K=3,...,4, j=1,...,K-1 •

We postpone the proof of Lemma II.2 until the end of Section II.2.2, after the proofs of Lemmata II.5 and II.6. For proving these two, a further number of propositions and lemmata will be needed. We turn first to the proof of Lemma II.5. For each K we will prove the claim by a 'clever' exhaustive enumeration of possible partitions. We now state some helpful propositions.

Proposition II.7 Assume that a partition $\mathcal{P}^{(3)}$ of the 2×3 table I.1 contains the subset $P:=\{\alpha_j , \beta_{j+1}\}$. Let the partition $\mathcal{P}^{(2)}$ of the 2×2 table be defined by removing P from $\mathcal{P}^{(3)}$. Then we have:

$$\widetilde{V}\left(\mathcal{P}^{(3)}\right) \leq O\left(\frac{d \ln(N)}{N}\right) \widetilde{V}\left(\mathcal{P}^{(2)}\right). \qquad \bullet$$

Proof. Bound the factor $L^d(\alpha_j - \alpha_j) = L^d(0)$ in $\widetilde{V}\left(\mathcal{P}^{(3)}\right)$ by d and use

$$\int L^N(\alpha_j+\beta_j)\ L^N(\alpha_{j+1}-\alpha_j)\, d\alpha_j \ \leq O(\ln(N))\, L^N(\alpha_{j+1}+\beta_j).\text{(Lemma A.1 i).}$$

What remains is exactly $N^{-1}\, \widetilde{V}\left(\mathcal{P}^{(2)}\right)$. □

Proposition II.8 For arbitrary distinct variables $\alpha_i \in \mathbf{R}$, $i=1,\ldots,n$ let $s_1 := \sum_{i=1}^{n} \alpha_i$ and $s_2 := \sum_{i=1}^{n} (-1)^{t_i} \alpha_i$, for arbitrary $t_i \in \{0,1\}$. Assume $s_1 \neq \pm s_2$. Then:

$$\int L^N(s_1)\, L^N(-s_1)\, L^d(s_2)\, L^d(-s_2)\, d(\alpha_1 \ldots \alpha_n) = O[d\, N \ln(N) \ln(d)] . \qquad \bullet$$

Proof. First observe that the quantity to be bounded is smaller than $d\,N \int L^N(s_1)\ L^d(s_2)\ d(\alpha_1 \ldots \alpha_n)$. Without loss of generality, we may assume that $t_1 := 0$. For some j we have $t_j = -1$, since $s_1 \neq \pm s_2$. Set $\alpha := s_1$ and $\beta := \sum_{i \neq j} \alpha_i - \alpha_j$. Then by a variable transformation we obtain

$$\int L^N(s_1)\ L^d(s_2)\ d(\alpha_1 \ldots \alpha_n) \leq O(1) \int L^N(\alpha)\ L^d(\beta)\ d(\alpha\beta) = O(\ln(N) \ln(d)). \square$$

Proof of Lemma II.5

i) If $\mathcal{P}^{(2)} \neq \mathcal{P}_1^{(2)}$ and $\mathcal{P}^{(2)}$ has no 'one-row' partition subsets, there are two possibilities left: $\mathcal{P}^{(2)} = \{(\alpha_1,\alpha_2), (\beta_2,\beta_1)\}$ and $\mathcal{P}^{(2)} = \{(\alpha_1,\alpha_2,\beta_2,\beta_1)\}$. In both cases one may easily check the assertion by using Lemma A.1 and Proposition II.8. We do this in the first case:

$$d^{-1}\, \widetilde{V}\left(\mathcal{P}^{(2)}\right) = d^{-1} N^{-2} \int \left(L^N\right)^2(\alpha-\beta) \left(L^d\right)^2(\alpha+\beta)\, d(\alpha\beta)$$

and the result follows from Proposition II.8.

ii) Assume that $\mathcal{P}^{(3)} \notin \left\{\mathcal{P}_1^{(3)}, \mathcal{P}_2^{(3)}\right\}$ and $\mathcal{P}^{(3)}$ has no 'one-row' partition subsets. Further, because of Proposition II.7, we may assume that it has no subset of the form $P:=\{\alpha_j, \beta_{j+1}\}$. If the contrary was the case, it would follow that $\widetilde{V}(\mathcal{T}^{(3)}) \leq d\, N^{-1} \ln(N)\, \widetilde{V}(\mathcal{T}^{(2)})$ with $\mathcal{T}^{(2)} \neq \mathcal{T}_1^{(2)}$ and our assertion would follow from i).

We observe that: a) for j,k arbitrary $\alpha_j+\beta_j \not\equiv \alpha_k+\beta_{k+1}$ and b) if K=3 and at the same time for some j,k arbitrary $\alpha_j+\beta_j \equiv -(\alpha_k+\beta_{k+1})$ then k=j+1 (if not, then two of the four variables would be identical -remember indices are taken modulo(K)-, yielding an equation in the variables vanishing identically with one coefficient equal 2).

After these remarks we proceed to the proof of ii).
Using Lemma A.1 ii) we integrate with respect to one of the variables of the table, say κ, for which κ and $-\kappa$ do not belong to the same row. We get

$$d^{-1}\, \widetilde{V}(\mathcal{T}^{(3)}) \leq \ln(N)\, N^{-3} \int (L^N)^2(\alpha_j+\beta_j)(L^d)^2(\alpha_k+\beta_{k+1}) \prod_{i=1}^{S} d\widetilde{\kappa}_i\,,$$

where j is the row and k the diagonal to which neither κ nor $-\kappa$ belong. Of course j and k depend on which κ is chosen. Observe that our assumptions on $\mathcal{T}^{(3)}$ guarantee that $\alpha_j+\beta_j$ and $\alpha_k+\beta_{k+1}$ do not vanish identically. From a) above it follows that $\alpha_j+\beta_j \not\equiv \alpha_k+\beta_{k+1}$. There are now two cases: if $\alpha_j+\beta_j \not\equiv -(\alpha_k+\beta_{k+1})$ the desired result follows directly from Proposition II.8. If this is not the case, that is if we have $\alpha_j+\beta_j \equiv -(\alpha_k+\beta_{k+1})$, then from b) above it follows that k=j+1. It follows further that $\mathcal{T}^{(3)}$ must contain the subset $\{\alpha_{j+3}, \beta_{j+1}\}$. Let us assume, for notational convenience (without restricting generality), that j=1, k=2. Then $\mathcal{T}^{(3)}$ must have the form:

$$\mathcal{T}^{(3)}: \begin{matrix} \alpha_1 & \beta_1 \\ \alpha_2 & \kappa \\ -\kappa & \beta_3 \end{matrix}$$

Now either the second or the third row have at least one further variable, say κ', in common with the first row. Repeating our argument from the beginning with κ' instead of κ we obtain the same bound as above with some $(j', k') \neq (j, k)$. By arguing as before if $\alpha_{j'}+\beta_{j'} \not\equiv -(\alpha_{k'}+\beta_{k'+1})$ the proof is finished. If this is not the case, it follows again that either $\{\alpha_2, \beta_1\}$ or $\{\alpha_1, \beta_3\}$ belong to $\mathcal{T}^{(3)}$. But this can only happen if $\mathcal{T}^{(3)}=\mathcal{T}_2^{(3)}$, contradicting our original assumption on $\mathcal{T}^{(3)}$.

iii) Assume that $\mathcal{T}^{(4)} \notin \{\mathcal{T}_1^{(4)}, \mathcal{T}_2^{(4)}, \mathcal{T}_3^{(4)}\}$ and that $\mathcal{T}^{(4)}$ has no 'one-row' partition subsets. Further we may assume that $\mathcal{T}^{(4)}$ has exactly two row-subtables, since it cannot have more than 2 (see the second assumption), and if it had less than 2, already Proposition I.8 would yield the assertion. Finally we may assume that it has no subset of the form

$P:=\{\alpha_j, \beta_{j+1}\}$. If it had two such subsets, then it would be identical to $\mathcal{P}_1^{(4)}$ or $\mathcal{P}_2^{(4)}$ (since it has exactly two row-subtables). If, on the other hand, it had one such subset, say in the first row-subtable, then by setting $L^d(0) = d$, integrating with respect to the variables contained in this row-subtable and using Lemma A.1 i), it would follow that $\widetilde{\mathbf{V}}\left(\mathcal{P}^{(4)}\right) \leq d\, N^{-1} \ln(N)\, \widetilde{\mathbf{V}}\left(\mathcal{P}^{(2)}\right)$ with $\mathcal{P}^{(2)} \neq \mathcal{P}_1^{(2)}$ and our assertion would follow from i).

We distinguish two cases: **a)** each row-subtable consists of two neighbour rows, say the first and the second row and **b)** they consist of two non-neighbour rows, say the first and the third row.

Let us first study **a)**. Integrating with respect to the variables in the first two rows, using once Lemma A.1 ii) and once Lemma A.1 i) we get:

$$d^{-1}\, \widetilde{\mathbf{V}}\left(\mathcal{P}^{(4)}\right) \leq d^{-1} N^{-3} d \ln(N) \ln(d) \int \left(L^N\right)^2\left(\alpha_3+\beta_3\right)\left(L^d\right)^2\left(\alpha_3+\beta_4\right) \prod_{i=1}^{S'} d\widetilde{\kappa}_i\,,$$

with some $S'<S$. Since $\left(\alpha_3+\beta_3\right) \not\equiv 0 \not\equiv \left(\alpha_3+\beta_4\right)$ (because of our assumptions on $\mathcal{P}^{(4)}$) and $\left(\alpha_3+\beta_3\right) \not\equiv \pm \left(\alpha_3+\beta_4\right)$ we may apply Proposition II.8 and our assertion follows.

We now turn to **b)**. Again we distinguish two subcases: **b1)** There exists a partition subset with 4 elements, say $\alpha_1, \beta_1, \alpha_3, \beta_3 := -\alpha_1 - \beta_1 - \alpha_3$ and **b2)** b1) is not fulfilled.

For **b1)** we get, using Lemma A.1, after integrating with respect to α_1 (for some $S', S''<S$) that $d^{-1}\, \widetilde{\mathbf{V}}\left(\mathcal{P}^{(4)}\right)$ is less than or equal to:

$$O(1)\, N^{-3} \ln(N) \int \left(L^N\right)^2\left(\alpha_2+\beta_2\right) L^d\left(\beta_2-\beta_1+\alpha_2-\alpha_3\right) L^d\left(\alpha_3+\beta_4\right)\, L^d\left(\alpha_4+\beta_1\right) \prod_{i=1}^{S'} d\widetilde{\kappa}_i$$

$$\leq O(1)\, N^{-3} \ln(N) \ln^2(d) \int \left(L^N\right)^2\left(\alpha_2+\beta_2\right) L^d\left(\beta_2+\beta_4+\alpha_4+\alpha_2\right) \prod_{i=1}^{S''} d\widetilde{\kappa}_i$$

$$=O\left(d\, N^{-2} \ln^2(N) \ln^2(d)\right), \text{ since } \beta_2+\beta_4+\alpha_4+\alpha_2 \equiv 0.$$

If **b2)** holds there are exactly 3 partitions left up to $\mathcal{P}_3^{(4)}$, namely:

$\left\{\left(\alpha_1,\beta_3\right), \left(\alpha_3,\beta_1\right), \left(\alpha_2,\alpha_4\right), \left(\beta_2,\beta_4\right)\right\}, \left\{\left(\alpha_1,\alpha_3\right), \left(\beta_1,\beta_3\right), \left(\alpha_2,\beta_4\right), \left(\alpha_4,\beta_2\right)\right\}$

$\left\{\left(\alpha_1,\alpha_3\right), \left(\beta_1,\beta_3\right), \left(\alpha_2,\alpha_4\right), \left(\beta_2,\beta_4\right)\right\}$.

The first two cases are straight forward as above. We treat separately only the last one:

$$d^{-1}\,\widetilde{\mathbf{V}}\left(\mathcal{P}^{(4)}\right) \leq O(1)\, d\, N^{-2} \int L^N(\alpha_1+\beta_1)\, L^N(\alpha_1+\beta_2)\, L^d(\alpha_1+\beta_2)\, L^d(\alpha_2-\beta_1) \prod_{i=1}^{S'} d\widetilde{\kappa}_i$$

By a variable transformation, this quantity is less than or equal to:

$$O(1) d\, N^{-2} \int L^N(\alpha)\, L^N(\beta)\, L^d(\gamma)\, L^d(d)\;\; d(\alpha\beta\gamma\delta) = O\left(d\, N^{-2} \ln^2(d) \ln^2(N)\right).$$

iv) $d^{-1}\,\mathbf{V}\left(\mathcal{P}^{(5)}\right) = O\left[d^3 N^{-3} \ln^3(N) \ln^2(d)\right]$ by Proposition I.8, since the number of row-subtables has to be less or equal to two (because no 'one-row' partition subsets are allowed). □

We now turn to the proof of Lemma II.6. The key will be the following lemma which shows that $\mathbf{V}\left(\mathcal{P}^{(K)}\right)$ does not depend asymptotically on f.

> **Lemma II.9** Assume that (D1) holds. Then for partitions $\mathcal{P}^{(K)}$ containing only 2-element partition subsets we have:
>
> $$\mathbf{V}_f\left(\mathcal{P}^{(K)}\right) = \mathbf{V}_{(2\pi)^{-1}}\left(\mathcal{P}^{(K)}\right) + O\left(d^{-\beta}\ln(d)\right)\widetilde{\mathbf{V}}\left(\mathcal{P}^{(K)}\right). \qquad \bullet$$

Proof. To prove this, let $g_d(\lambda)$ be a sequence of AR(p_d) spectral densities, $p=p_d := d^{1/(1+r+\alpha)}$ with $\| g_d - f \|_\infty = O(d^{-\beta})$. Then we get (Lemma A.3):

$$\left| K_d\left(f,\lambda,\mu\right) - K_d\left(g_d,\lambda,\mu\right)\right| = O\left(d^{-\beta}\ln(d)\right) L^d(\mu-\lambda)\,.$$

First we show $\mathbf{V}_f\left(\mathcal{P}^{(K)}\right) = \mathbf{V}_{g_d}\left(\mathcal{P}^{(K)}\right) + O\left(d^{-\beta}\ln(d)\right)\widetilde{\mathbf{V}}\left(\mathcal{P}^{(K)}\right)$. This may be proven by substituting in $\mathbf{V}_f\left(\mathcal{P}^{(K)}\right)$ successively all quantities depending on f, that is f and $K_d\left(f,\lambda,\mu\right)$, by the quantities corresponding to g_d. Observe that the error we have each time is $O\left(d^{-\beta}\ln(d)\right)\widetilde{\mathbf{V}}\left(\mathcal{P}^{(K)}\right)$.

Next we substitute successively in $\mathbf{V}_{g_d}\left(\mathcal{P}^{(K)}\right)$ the quantities $K_d\left(g_d,\lambda,\mu\right) := (2\pi)^{-1} \sum_{j=0}^{d-1} \phi_j(\lambda)\, \overline{\phi_j(\mu)}$, where $\left\{\phi_k(\lambda)\right\}_{k\in N}$ denote the orthogonal polynomials with respect to g_d, by the truncated

$$K^*_{d,p}\left(g_d,\lambda,\mu\right) := (2\pi)^{-1} \sum_{j=p}^{d-1} \phi_j(\lambda)\, \overline{\phi_j(\mu)} \;\; (p:=p_d).$$

Observe that we have: $\left|K_d(\lambda,\mu) - K^*_{d,p}(\lambda,\mu)\right| \le L^p(\mu-\lambda) \le \frac{p}{d} L^d(\mu-\lambda) = d^{-\beta} L^d(\mu-\lambda)$. We conclude that in each substitution the error is again $O(d^{-\beta}\ln(d))\ \widetilde{\mathbf{V}}\left(\mathcal{I}^{(K)}\right)$.

Finally in the resulting quantity enter only the polynomials orthogonal to g_d of degree $k \ge p_d$, which are exactly known (care of (A.2.4)). By substituting all of them the g_d's are canceled. We now substitute in the expression the remaining $\sum_{j=p}^{d-1} e^{i(d-1)(\mu-\lambda)}$ by $\sum_{j=0}^{d-1} e^{i(d-1)(\mu-\lambda)}$ having again a total error of $O(d^{-\beta}\ln(d))\ \widetilde{\mathbf{V}}\left(\mathcal{I}^{(K)}\right)$. The remaining quantity is exactly $\mathbf{V}_{(2\pi)^{-1}}\left(\mathcal{I}^{(K)}\right)$. □

Proof of Lemma II.6 In view of Lemma II.9 it is sufficient to set $f = \frac{1}{2\pi}$ and to show that

i) $d^{-1}\, \mathbf{V}\left(\mathcal{I}_2^{(2)}\right) = \frac{d}{N}$

ii) $d^{-1}\, \mathbf{V}\left(\mathcal{I}_j^{(K)}\right) = \frac{d^2}{N^2}\left[1 + O\left(\frac{d}{N}\right)\right]$, for K=3, 4, j=1,...,K-1.

The calculations are straight forward. We will only show as examples the cases i) and ii) K=3, j=2.

i) $d^{-1}\, \mathbf{V}\left(\mathcal{I}_2^{(2)}\right) =$

$$\frac{1}{dN}(2\pi)^{-2}\int \Delta^N(\lambda-\mu)\, d(\lambda\mu)\Theta^d(\lambda-\lambda)\, \Theta^d(\mu-\mu) = \frac{d}{N}(2\pi)^{-2}\int \Delta^N(\lambda-\mu)\, d(\lambda\mu) = \frac{d}{N}.$$

ii) K=3, j=2. Denote $\alpha_1{:=}\lambda$, $\alpha_2{:=}\mu$, $\alpha_3{:=}\kappa$. Then

$$d^{-1}\, \mathbf{V}\left(\mathcal{I}_2^{(3)}\right) = \frac{1}{dN^{-3}}(2\pi)^{-3}\int \Theta^N(\lambda-\mu)\, \Theta^N(\mu-\kappa)\, \Theta^N(\kappa-\lambda)\, \Theta^d(\lambda-\kappa)\, \Theta^d(\mu-\lambda)\, \Theta^d(\kappa-\mu)\, d(\lambda\mu\kappa)$$

$$= (2\pi)^{-3}\frac{1}{dN^{-3}}\sum_{s,t,r=1}^{N}\ \sum_{j,m,k=1}^{d-1}\int e^{i\lambda[(s-t)+(k-m)]}\, d\lambda \int e^{i\mu[(r-s)+(m-j)]}\, d\mu \int e^{i\kappa[(t-r)+(j-k)]}\, d\kappa$$

$$= \frac{1}{dN^{-3}}\sum_{s,t,r=1}^{N}\ \sum_{j,m,k=1}^{d-1} \delta_{t=r+k-j}\, \delta_{s=r-j+m} = \frac{1}{dN^{-3}}\sum_{r=1}^{N}\ \sum_{j,m,k=1}^{d-1} \delta_{1\le r+k-j\le N\ \wedge\ 1\le r-j+m\le N} =$$

$$= \frac{1}{dN^{-3}} \sum_{j,m,k=1}^{d-1} N+j-m\wedge k = \frac{d^2}{N^2}\left[1 + O\left(\frac{d}{N}\right)\right] .$$ □

Proof of Lemma II.2 From Lemma II.5 we have

$$E\, D_{d,T} = 1 + \sum_{K=2}^{4} \sum_{j=1}^{K-1} (-1)^K d^{-1} V\left(\mathcal{I}_j^{(K)}\right) + O\left[N^{-1} \ln(N) \ln(d)\right] ,$$

since $V\left(\mathcal{I}^{(K)}\right) = O\left[\tilde{V}\left(\mathcal{I}^{(K)}\right)\right]$ and the 'error' terms in Lemma II.5 are also $O\left[N^{-1} \ln(N) \ln(d)\right]$. The desired result follows from the above together with Lemma II.6, observing that $N^{-1} \ln(N) \ln(d) = O\left(\frac{d}{N} d^{-\beta} \ln(N) \ln(d)\right)$, since $\beta<1$. □

II.2.3 Proof of the remaining Lemmata in Chapter II

In order to prove Lemma II.3 we need two more propositions.

Proposition II.10 Assume that (A) and (B) hold. Then we have:

i) $\left|\tilde{f}_d^{-1}(\lambda) - \tilde{f}_d^{-1}(\mu)\right| = O\left(d\,|\lambda-\mu| \wedge 1\right)$ and

ii) $d^{-1}\left| K_d(\mu,\mu-\lambda) - K_d(\mu,\mu)\right| = O\left(d\,|\lambda-\mu| \wedge 1\right)$ •

Proof. Denote by $\{\phi_k(\lambda)\}_{k\in N}$ the orthogonal polynomials with respect to f. Then from the Theorem of S. Bernstein (Grenander and Szegö (1958) § 1.16) we have

$$\left|\phi_p(\lambda) - \phi_p(\mu)\right| \leq C\, p\, |\lambda-\mu|,$$

where the constant C depends only on the maximum and minimum of f. Now from (A.2.2) we have :

$$\left|\tilde{f}_d^{-1}(\lambda) - \tilde{f}_d^{-1}(\mu)\right| = \left|\frac{1}{d}\sum_{p=0}^{d-1}\left(|\phi_p(\lambda)|^2 - |\phi_p(\mu)|^2\right)\right| \leq C' \frac{1}{d}\sum_{p=0}^{d-1} p\, |\lambda-\mu| \leq C'd\, |\lambda-\mu|,$$

for some C'.The result i) follows from the additional observation that $\tilde{f}_d^{-1}(\lambda) \leq C''$ for some C''.

ii) follows similarly by observing:

$$d^{-1}\left| K_d(\mu,\mu-\lambda) - K_d(\mu,\mu)\right| = \left|\frac{1}{d}\sum_{p=0}^{d-1} \bar{\phi}_p(\mu)\left(\phi_p(\mu-\lambda) - \phi_p(\mu)\right)\right| \leq O(1)\frac{1}{d}\sum_{p=0}^{d-1} p\, |\lambda-\mu| .$$ □

Proposition II.11 We have: $\int \Delta^N(\alpha)\left[\, d\,\alpha \wedge 1\right] d\alpha = O\!\left(\frac{d}{N}\ln(N)\right)$. •

Proof: The proof is straight forward by bounding $\Delta^N(\alpha) \le N^{-1}\left(L^N(\alpha)\right)^2$ and braking the integration region in $\left\{|\alpha| \le N^{-1}\right\} \cup \left\{N^{-1} < |\alpha| \le d^{-1}\right\} \cup \left\{|\alpha| > d^{-1}\right\}$. In the first two we have $d\,\alpha \wedge 1 = d\,\alpha$ and in the last it equals 1. □

Proof of Lemma II.3 From Lemmata II.5 and II.6 we have that:

$$E\, D_{d,T} = 1 + d^{-1}\, V\left(\mathcal{I}_1^{(2)}\right) + O\!\left[\left(\frac{d}{N}\right)^2\right] + O\!\left[\, N^{-1}\ln(N)\ln(d)\right].$$

Now we have that $d^{-1}\, V\left(\mathcal{I}_1^{(2)}\right) = \frac{d}{N}\int \Delta^N(\lambda-\mu)\frac{f}{\tilde{f}_d}(\lambda)\frac{f}{\tilde{f}_d}(\mu)\, d(\lambda\mu)$.

Further using $(2\pi)^{-1}\int \frac{f}{\tilde{f}_d}(\lambda)\, d\lambda = 1$:

$$\int \Delta^N(\lambda-\mu)\frac{f}{\tilde{f}_d}(\lambda)\frac{f}{\tilde{f}_d}(\mu)\, d(\lambda\mu) = \int \frac{f}{\tilde{f}_d}(\mu)\int \Delta^N(\lambda-\mu)\frac{f}{\tilde{f}_d}(\lambda)\, d\lambda\, d\mu$$

$$= \int \left(\frac{f}{\tilde{f}_d}(\mu)\, d\mu\right)^2 + O\!\left(\frac{d}{N}\ln(T)\right) + O\!\left(T^{-\gamma}\right) = 1 + \tilde{B}_d' + O\!\left(\frac{d}{N}\ln(N)\right) + O\!\left(T^{-\gamma}\right).$$

The second to last equality following from

i) $\tilde{f}_d^{-1} * \Delta^N - \tilde{f}_d^{-1} = O\!\left(\frac{d}{N}\ln(N)\right)$, which in turn follows from Propositions II.10 and II.11, and

ii) $f * \Delta^N - f = O\!\left(N^{-\gamma}\right)$. (e.g. Dahlhaus (1985)) □

Proof of Lemma II.4 We write $Z_{d,T} := \frac{1}{2\pi}\int \left(f\,\hat{f}_{d,T}^{-1}\right)(\lambda)\, d\lambda = d^{-1}\,\mathrm{tr}\left[\hat{\Gamma}_{d,T}^{*\,-1}\right]$. We define further the events $A_T := \left\{\left\|\hat{\Gamma}_{d,T}^{*} - I\right\| < \frac{1}{2}\right\}$ and $B_T := \{\, Z_{d,T} < 2\,\}$.

Now observe that:

i) $B_T^c \subseteq A_T^c$, since $d^{-1}\,\mathrm{tr}\left[\hat{\Gamma}_{d,T}^{*\,-1}\right] \ge 2 \Rightarrow \lambda_{\min}\left(\hat{\Gamma}_{d,T}^{*}\right) \le \frac{1}{2} \Rightarrow \left\|\hat{\Gamma}_{d,T}^{*} - I\right\| \ge \frac{1}{2}$

ii) $\Delta(f,\hat{\hat{f}}_{p,T}) - \Delta(f,\hat{f}_{d,T}) = \ln(\mu_d) + (\mu_d^{-1} - 1)\, Z_{d,T}$

iii) on A_T we have $|Z_{d,T} - D_{d,T}| \leq 2\left|\left|\hat{\Gamma}^*_{d,T} - I\right|\right|^6$ (obtained by an expansion of $\hat{\Gamma}^{*\,-1}_{d,T}$ on A_T).

iv) $P(B^c_T) \leq P(A^c_T) = O\left(\left[\frac{d^{1+\varepsilon}}{N} \ln(N) \ln(d)\right]^k\right)$ for any integer $k>0$ and any $\varepsilon>0$ when $d\rightarrow\infty$, which follows from i) and Lemma I.2.

v) $\left|\int_{B_T} Z_{d,T}\, dP - ED_{d,T}\right| = O\left(\left[\frac{d^{1+\varepsilon}}{N} \ln(N) \ln(d)\right]^3\right)$.

Proof of v): The quantity on the left in v) is less than or equal to

$$\int_{B^c_T} |D_{d,T}|\, dP + \int_{B_T\cap A_T} |Z_{d,T} - D_{d,T}|\, dP + \int_{B_T\cap A^c_T} |Z_{d,T}| + |D_{d,T}|\, dP$$

$$\leq 2 \int_{A^c_T} |D_{d,T}|\, dP + 2\, E\left|\left|\hat{\Gamma}^*_{d,T} - I\right|\right|^6 + \int_{A^c_T} 2\, dP \text{ because of i), iii)}$$

$$= O\left(\left[\frac{d^{1+\varepsilon}}{N} \ln(N) \ln(d)\right]^3\right) \text{ because of iv), Lemma I.2 and } ED^2_{d,T} = O(1).$$

From these the assertion of the lemma follows , since from ii) we have:

$$E\Delta(f,\hat{\hat{f}}_{p,T}) \leq E\Delta(f,\hat{f}_{d,T}) + \ln(\mu_d) + (\mu_d^{-1} - 1)\int_{B_T} Z_{d,T}\, dP + (\mu_d^{-1} - 1)\int_{B^c_T} Z_{d,T}\, dP$$

$$\leq E\Delta(f,\hat{f}_{d,T}) + \ln(\mu_d) + (\mu_d^{-1} - 1)\int_{B_T} Z_{d,T}\, dP \text{ , since } \mu_d^{-1} - 1 < 0$$

$$= E\Delta(f,\hat{f}_{d,T}) + \ln(\mu_d) + (\mu_d^{-1} - 1)\, ED_{d,T} + o\left[\left(\frac{d}{N}\right)^2\right] \text{ because of v)}$$

$$= E\Delta(f,\hat{f}_{d,T}) + \ln(\mu_d) + 1 - \mu_d + o\left[\left(\frac{d}{N}\right)^2\right] \text{ (because of Lemma II.2)}$$

$$= E\Delta(f,\hat{f}_{d,T}) - \frac{1}{2}\left(\frac{d}{N}\right)^2 + o\left(\frac{d}{N}\right)^2. \qquad \square$$

III. ASYMPTOTIC EFFICIENCY OF THE DIMENSION SELECTION CRITERION

III.1 INTRODUCTION, RESULTS

In the previous chapter we introduced the discrepancy $\Delta(f,\hat{f}_{d,T})$ and developed a dimension selection criterion for the parameter d of the Capon estimator. Our argumentation aimed to the minimization of $\Delta(f,\hat{f}_{d,T})$, but it did not clarify if this minimum is really approximated by $\Delta(f,\hat{f}_{\hat{d},T})$, where $\hat{d} := \operatorname{argmin}_{d \le d_{max}} (2\pi)^{-1} \int \ln \hat{f}_{d,T}(\lambda)\, d\lambda + \frac{d}{N} + \left(\frac{d}{N}\right)^2$.

In this chapter we prove that there exists a stochastic lower bound of the loss $\Delta(f,\hat{f}_{d,T})$ which is asymptotically attained by the proposed dimension selection criterion (Theorems III.1, III.2 and Lemma III.3). A corollary is that the loss $\Delta(f,\hat{f}_{\hat{d},T})$ (when using the dimension selection criterion) is asymptotically equivalent -in probability- to the minimum possible loss $\inf_{d \le d_{max}} \Delta(f,\hat{f}_{d,T})$. In this sense our dimension selection criterion is asymptotically efficient.

In this chapter we also discuss the 'estimation' of $\Delta(f,\hat{f}_{d,T})$ and $\Delta(f,\hat{f}_{\hat{d},T})$ (see Remark III.2).

For the definition of the notion of the asymptotic efficiency of a selection criterion we have adopted the approach of Shibata (1980). He showed the asymptotic efficiency of the AIC as an order selection criterion when autoregressive LS spectral estimators are used and when the one-step prediction error is taken as a discrepancy. Taniguchi (1980) extended the results of Shibata to the case where i) the underlying process is not necessarily Gaussian and ii) the estimators (approximating models) are not necessarily autoregressive. Tanigutchi, however, uses a stochastic approximation of the one-step prediction error as a discrepancy for defining asymptotic efficiency. This is not very problematic since, in both papers, the definition of asymptotic efficiency itself is a statement in probability. Both papers assume that the true spectral density does not belong to the approximating model class: the 'truly optimal' order, say p^*, which minimizes the expectation of the discrepancy, tends to infinity with the number of observations. For the validity of our results we assume a similar condition.

Before we proceed to the statement of the results we introduce some notation:
We write $\Delta_T(d) := \Delta(f,\hat{f}_{d,T})$. We may decompose $\Delta_T(d)$ as follows:

$$\Delta_T(d) = \tilde{B}_d + V_{d,T} + W_{d,T},$$

where

$$\tilde{B}_d := \frac{1}{2\pi}\int \ln\left(\frac{\tilde{f}_d}{f}\right)(\lambda) + \frac{f}{\tilde{f}_d}(\lambda) - 1 \, d\lambda \ , \quad V_{d,T} := \frac{1}{2\pi}\int \ln\left(\frac{\hat{f}_{d,T}}{\tilde{f}_d}\right)(\lambda) + \frac{\tilde{f}_d}{\hat{f}_{d,T}}(\lambda) - 1 \, d\lambda$$

$$\text{and } W_{d,T} := \frac{1}{2\pi}\int \left[\frac{\tilde{f}_d}{\hat{f}_{d,T}}(\lambda) - 1\right] \left[\frac{f}{\tilde{f}_d}(\lambda) - 1\right] d\lambda$$

(since $(2\pi)^{-1} \int \frac{f}{\tilde{f}_d}(\lambda)\, d\lambda = d^{-1}\,\mathrm{tr}\left(\Gamma_d^{-1} \int f(\lambda)\ b_\lambda \bar{b}_\lambda^t\, d\lambda\right) = 1$).

We further define the approximations $V'_{d,T}$, $W'_{d,T}$ (see Lemma III.8 below) of $V_{d,T}$, $W_{d,T}$ respectively:

$$V'_{d,T} := \frac{1}{4\pi}\int \left[I^*_{d,T}(\lambda) - 1\right]^2 d\lambda \text{ and } W'_{d,T} := \frac{1}{2\pi}\int \left[1 - I^*_{d,T}(\lambda)\right] \left[\frac{f}{\tilde{f}_d}(\lambda) - 1\right] d\lambda,$$

where $I^*_{d,T}(\lambda) := \bar{b}_\lambda^{*t}\, \hat{\Gamma}^*_{d,T}\, b^*_\lambda$ as in (I.3).

Finally let $L_T(d) := \tilde{B}_d + EV'_{d,T}$ and $d^* = d^*_T := \mathrm{argmin}_{d \le d_{max}} L_T(d)$ for some sequence $d_{max} \to \infty$. Observe that from Lemma I.9 we obtain $L_T(d) = \tilde{B}_d + \frac{\theta(d)}{2}\frac{d}{N_d} + o\left(d\, N_d^{-1}\right)$.

Remark III.1 As will be shown below (Lemmata III.5, III.8 and III.11), the decomposition of $\Delta_T(d)$ reads:

$$\Delta_T(d) = \tilde{B}_d + V'_{d,T} + o_P\left(L_T(d)\right) = L_T(d) + o_P\left(L_T(d)\right),$$

uniformly in $d \le d_{max}$ which may be viewed as a stochastic decomposition into a 'squared-bias' and a 'variance' term.

Remark III.2 In the praxis one is interested in an approximation or at least an upper bound of $\Delta\left(f, \hat{f}_{d,T}\right)$ and $\Delta\left(f, \hat{f}_{\hat{d},T}\right)$. From the preceding remark it follows that these quantities may be approximated by $L_T(d)$ and $L_T(\hat{d})$ respectively. The problem is that these two quantities involve the unknown $\tilde{B}_d$. One possibility to overcome this difficulty is to observe that if $d\tilde{B}_d$ is falling at d^* (which is the case for sufficiently regular f and d^* large enough, see Proposition A.4), then $\tilde{B}_d \le \frac{\theta(d)}{2}\frac{d}{N_d}$ for $d \ge d^*$. Thus in this case $\frac{\theta(d)}{2}\frac{d}{N_d} + o\left(d\, N_d^{-1}\right) \le L_T(d) \le \theta(d)\frac{d}{N_d} + o\left(d\, N_d^{-1}\right)$.

In this chapter we make the following assumption:

(D2) Assume that (A), (B) and (C) hold and that c=1. Further assume that $\frac{d_{max}^{1+\varepsilon}}{T} \rightarrow 0$ for some sequence $d_{max} \rightarrow \infty$ and some $\varepsilon > 0$.

Our first theorem ensures that $L_T(d^*)$ is a lower bound for the loss.

Theorem III.1 Assume that (D2) and $d^{*-1} \ln^4(T) \ln^6(d^*) \rightarrow 0$ hold. Let $\tilde{d} = \tilde{d}_T \leq d_{max}$ be any sequence of random variables, measurable with respect to $X_1, \ldots, X_T$. Then for any $\varepsilon > 0$ we have:

$$P\left\{\frac{\Delta_T(\tilde{d})}{L_T(d^*)} \geq 1 - \varepsilon\right\} \rightarrow 1. \qquad \bullet$$

This justifies the definition of an asymptotically efficient dimension selection criterion $\tilde{d}$ as one for which

$$\frac{\Delta_T(\tilde{d})}{L_T(d^*)} \rightarrow_P 1.$$

The next theorem states that the selection criterion $\hat{d} := \operatorname{argmin}_{d \leq d_{max}} S_T(d)$ is asymptotically efficient, where $S_T(d) := (2\pi)^{-1} \int \ln \hat{f}_{d,T}(\lambda)\, d\lambda + \frac{d}{N_d}$ and $N \equiv N_d$ as in (0.5). (Note the change in the meaning of $\hat{d}$ compared to the previous chapter.)

Theorem III.2 Assume that (D2) and $d^{*-1} \ln^4(T) \ln^6(d^*) \rightarrow 0$ hold. Then we have:

$$\text{i)}\ \frac{\Delta_T(\hat{d})}{L_T(d^*)} \rightarrow_P 1 \qquad \text{ii)}\ \frac{\Delta_T(\hat{d})}{\inf_{d \leq d_{max}} \Delta_T(d)} \rightarrow_P 1 \qquad \bullet$$

As a corollary we obtain from the next lemma that $\hat{d}^\circ := \operatorname{argmin}_{d \leq d_{max}} \mathring{S}_T(d)$ is also asymptotically efficient, where $\mathring{S}_T(d) := (2\pi)^{-1} \int \ln \hat{f}_{d,T}(\lambda)\, d\lambda + \frac{d}{N_d} + \left(\frac{d}{N_d}\right)^2$.

Lemma III.3 Assume that (D2) and $d^{*-1} \ln^4(T) \ln^6(d^*) \rightarrow 0$ hold. Further assume that the dimension selection criterion $\hat{d} := \operatorname{argmin}_{d \leq d_{max}} S_T(d)$ is asymptotically efficient and that a second dimension selection criterion consists of $\hat{d}^\circ := \operatorname{argmin}_{d \leq d_{max}} \mathring{S}_T(d)$. Assume that $\mathring{S}_T(d)$ fulfills:

$$\sup_{d \le d_{max}} L_T^{-1}(d)\left[S_T(d)-S_T(d^*)-\overset{\circ}{S}_T(d)+\overset{\circ}{S}_T(d^*)\right] \to 0 .$$

Then $\hat{d}^P$ is also asymptotically efficient. •

We remark that, as follows from Lemma III.3, the notion of asymptotic efficiency -here as well as in the case of examining order selection criteria when using an autoregressive spectral estimator- 'looks' only at the coefficient of d/ N in the penalty term. The property of asymptotic efficiency of a dimension selection criterion does not change if one changes the coefficient of $(d/N)^2$. Thus this concept does not give any answer to the questions of i) whether one should include $(d/N)^2$ in the penalty term, ii) whether one should use the AICc instead of the AIC (see Section II.1.4), and iii) whether one should use N instead of T in the AIC and the AICc (see Section II.1.4).

Finally let us remark that from the above considerations it follows that if $\hat{\hat{f}}_{p,T} := \mu_{p,T}\,\hat{f}_{p,T}$, as in Section II.1.3, and if a dimension selection criterion is defined by $\hat{\hat{d}} := \underset{d \le d_{max}}{\operatorname{argmin}}\,(2\pi)^{-1}\int \ln \hat{\hat{f}}_{d,T}(\lambda)\, d\lambda$, then this dimension selection criterion is also asymptotically efficient for $\hat{f}_{d,T}$ and for $\hat{\hat{f}}_{d,T}$:

Lemma III.4 Assume that (D2) and $d^{*-1}\ln^4(T)\ln^6(d^*) \to 0$ hold. Then we have:

$$\text{i) } \frac{\Delta\left(f,\hat{f}_{\hat{\hat{d}},T}\right)}{L_T(d^*)} \to_P 1 \qquad \text{ii) } \frac{\Delta\left(f,\hat{\hat{f}}_{\hat{\hat{d}},T}\right)}{L_T(d^*)} \to_P 1$$ •

III.2 DETAILED RESULTS, PROOFS

In this section we prove the main results stated in Section III.1. A first explanation of the reasons why these results hold, will be given in Section III.2.1, where we also state the basic Lemmata III.5, III.6 and III.7 leading to the proofs of the main results. In Section III.2.2 we prove Lemmata III.5 and III.6, in Section III.2.3 Lemmata III.7 and finally, in Section III.2.4, we prove the main results stated in Section III.1.

We remark that the basic idea for the proofs of the main results, namely the steps represented by Lemmata III.5, III.6 and III.7, is adopted from Shibata (1980) where analogous results are proven. On the other hand we use rather different methods for our proofs.

III.2.1 Some key results and a first explanation

Before we proceed to the proofs of the above theorems and lemmata, we state and explain the main points which lead to these results.

The first critical property is that $\Delta_T(d)$ concentrates stochastically on $L_T(d)$, uniformly in $d \le d_{max}$.

III.5 Lemma. Assume that (D2) and $d^{*-1} \ln^4(T) \ln^6(d^*) \to 0$ hold. Then we have

$$\sup_{d \le d_{max}} \left| \frac{\Delta_T(d)}{L_T(d)} - 1 \right| \to_P 0. \qquad \bullet$$

This helps us understand the behavior of $S_T(d)$. We can write:

$$S_T(d) - (2\pi)^{-1} \int \ln(f(\lambda))\, d\lambda = \Delta_T(d) + \frac{d}{N_d} - d^{-1} \operatorname{tr}\left[\hat{\Gamma}^{*\,-1}_{d,T}\right] + 1 =$$

$$= \Delta_T(d) + d^{-1} \operatorname{tr}\left[\hat{\Gamma}^{*}_{d,T} - I\right] - \left(d^{-1} \operatorname{tr}\left[\hat{\Gamma}^{*}_{d,T} - I\right]^2 - \frac{d}{N_d}\right) + o_P\left(\frac{d}{N_d}\right),$$

the last equality following from Lemmata II.1 and I.2.

Since $\Delta_T(d)$ concentrates stochastically on $L_T(d)$ (Lemma III.5), the minimization of $\Delta_T(d)$ -our actual aim- could be approximated by the minimization of $S_T(d)$, if $S_T(d)$ also concentrated on $L_T(d)$. But it does not; while the last and second-to-last term on the right hand of the last equality converge stochastically to zero when divided by $L_T(d)$ (Lemma III.6 below), the second term does not. On the other hand, this second term divided by $L_T(d)$ asymptotically does not depend on d (Lemma III.7 below), not playing any role in the minimization of $S_T(d)$. It follows that the minimization of $S_T(d)$ is equivalent to the minimization of

$$\Delta_T(d) - \left(d^{-1} \operatorname{tr}\left[\hat{\Gamma}^{*}_{d,T} - I\right]^2 - \frac{d}{N_d}\right) + o_P\left(\frac{d}{N_d}\right),$$

which -as mentioned above- will be shown to converge uniformly to 1 when divided by $L_T(d)$. This is exactly what we need.

Accordingly the results we need for the proof of Theorem III.2 are given in the two following lemmata:

Lemma III.6 Assume that (D2) and $d^{*-1} \ln^4(T) \ln^2(d^*) \to 0$ hold. Then we have

$$\sup_{d \le d_{max}} L_T^{-1}(d) \left| d^{-1} \operatorname{tr}\left[\hat{\Gamma}^{*}_{d,T} - I\right]^2 - \frac{d}{N_d} \right| \to_P 0. \qquad \bullet$$

and

Lemma III.7 Assume that (D2) and $d^{*-1} \ln^2(T) \to 0$ hold. Then we have

$$\sup_{d \le d_{max}} L_T^{-1}(d) \left| d^{-1} \operatorname{tr} \hat{\Gamma}^*_{d,T} - d^{*-1} \operatorname{tr} \hat{\Gamma}^*_{d^*,T} \right| \to_P 0. \qquad \bullet$$

III.2.2 The quadratic terms

In this section we prove Lemmata III.5 and III.6. In both, a random variable depending on d is shown to converge uniformly in the stochastic sense. It turns out that it is sufficient to bound the fourth order moments of these variables in order to prove the convergence. For Lemma III.6 this can be done directly. For the proof of Lemma III.5, which is more complicated, we use the decomposition of $\Delta_T(d)$: first we show that $V_{d,T}$ and $W_{d,T}$ may be approximated stochastically by $V'_{d,T}$ and $W'_{d,T}$ respectively. Secondly, we examine the behaviour of the two latter random variables by considering their higher order centered moments.

We first prove Lemma III.5. For this we will state and prove a proposition and three other lemmata.

Lemma III.8 Assume that (D2) holds. Then we have:

$$\text{i) } \sup_{d \le d_{max}} \left| V_{d,T} - V'_{d,T} \right| = o_P(L_T(d))$$

$$\text{ii) } \sup_{d \le d_{max}} \left| W_{d,T} - W'_{d,T} \right| = o_P(L_T(d)). \qquad \bullet$$

Proof. First we fix a sequence $d = d_T \le d_{max}$. Let the following event be denoted by $A_{d,T} := \left\{ \left\| \hat{\Gamma}^*_{d,T} - I \right\| < 1/2 \right\}$. Further denote by $J^*_{d,T}(\lambda) := \left[\bar{b}_\lambda^{*t} \left(\hat{\Gamma}^*_{d,T} \right)^{-1} b^*_\lambda \right]^{-1}$ and let $I^*_{d,T}$ as in (I.3). Note that we have:

(1) $$\left| J^*_{d,T}(\lambda) - 1 \right|, \left| I^*_{d,T}(\lambda) - 1 \right| \le \left\| \hat{\Gamma}^*_{d,T} - I \right\|$$

(2) $$\left| \left(\left[J^*_{d,T}(\lambda) \right]^{-1} - 1 \right) - \left(1 - I^*_{d,T}(\lambda) \right) \right| \le 2 \left\| \hat{\Gamma}^*_{d,T} - I \right\|^2 \text{ on } A_{d,T}$$

(3) $$\left| \left(J^*_{d,T}(\lambda) - 1 \right) - \left(I^*_{d,T}(\lambda) - 1 \right) \right| \le 4 \left\| \hat{\Gamma}^*_{d,T} - I \right\|^2 \text{ on } A_{d,T}.$$

((2) is obtained by an expansion of $\left[\widehat{\Gamma}^*_{d,T}\right]^{-1}$; (3) follows from (2) and (1))

We first prove i). Expanding the functions ln(.) and 1/(.) by using (1), we obtain on $A_{d,T}$:

$$V_{d,T} = \frac{1}{2\pi}\int \ln\left(J^*_{d,T}(\lambda)\right) + \left(J^*_{d,T}(\lambda)\right)^{-1} - 1 \, d\lambda = \frac{1}{4\pi}\int \left(J^*_{d,T}(\lambda) - 1\right)^2 d\lambda \; + \; O\left(\left\|\widehat{\Gamma}^*_{d,T} - I\right\|^3\right).$$

From this and using (1) and (3) we obtain :

$$(4) \qquad \left|V_{d,T} - V'_{d,T}\right| = O\left(\left\|\widehat{\Gamma}^*_{d,T} - I\right\|^3\right) \text{ on } A_{d,T}.$$

Finally for any even k and any $\delta>0$ and $\varepsilon<1/2$ we obtain from (4) and Lemma I.2:

$$P\left\{\sup_{d\leq d_{max}} T\, d^{-1}\left|V_{d,T} - V'_{d,T}\right| > \varepsilon\right\} \leq$$

$$\sum_{d=1}^{d_{max}} P\left\{A^C_{d,T}\right\} + \sum_{d=1}^{d_{max}} P\left\{T^k d^{-k}\left\|\widehat{\Gamma}^*_{d,T} - I\right\|^{3k} \geq \varepsilon^k\right\}$$

$$= O(1)\, \varepsilon^{-k}\, T^{-k/2} \ln(T)^{3k} \ln(d)^{3k} \;\; d_{max}^{k/2+\delta k+1} \to 0$$

The convergence to 0 follows from (D2) after choosing k large enough and δ small enough.

Assertion ii) is similarly proven: from (2) and Hölder's inequality, observing that $\int \left[f\, \tilde{f}_d^{-1}(\mu) - 1\right]^2 d\mu = O(\tilde{B}_d)$, we obtain

$$L_T^{-1}(d)\left|W_{d,T} - W'_{d,T}\right| \;\leq\; 2\, L_T^{-1}(d)\left\|\widehat{\Gamma}^*_{d,T} - I\right\|^2 \sqrt{\tilde{B}_d}$$

$$= O\left[T^{1/2}\, d^{-1/2}\left\|\widehat{\Gamma}^*_{d,T} - I\right\|^2\right] \text{ on } A_{d,T}. \qquad \square$$

The next step in the proof of Lemma III.5 is to bound higher order centered moments of $V'_{d,T}$ and $W'_{d,T}$ (Lemmata III.10 and III.11). For this we need the following technical proposition:

Proposition III.9 Assume that (D2) and $d = d_T \leq d_{max}$ hold. The setting $N := N_d$ we have for any $r \geq 2$:

$$\int \left(\mathrm{cum}\left[I^*_{d,T}(\lambda_1), \ldots, I^*_{d,T}(\lambda_r)\right]\right)^2 d(\lambda_1 \ldots \lambda_r) = O\left(d^{r-1} N^{-2(r-1)} \ln(d)^{r+1} \ln(N)^{2(r-1)}\right). \quad \bullet$$

Proof. Consider indecomposable partitions $\mathcal{P}^{(r)} = \{P_1, \ldots, P_S\}$ of the $2 \times r$ table:

$$\begin{array}{cc} \lambda_1 & -\lambda_1 \\ \vdots & \vdots \\ \lambda_r & -\lambda_r \end{array}$$

with $|P_i| \geq 2$ (it follows $S \leq r$). Then from Lemma I.10 we know that:

$$\mathrm{cum}\left[I^*_{d,T}(\lambda_1), \ldots, I^*_{d,T}(\lambda_r)\right] = O\left(d^{-1} N^{-(r-1)} \ln(N)^{r-1}\right) \sup_{\mathcal{P}^{(r)}} \ln(d)^{r-S+1} \prod_{j=1}^{S} L^d\left(\sum_{\pm\lambda_i \in P_j} \pm\lambda_i\right)$$

And from this we obtain that the integral over the square of the cumulant is less than or equal to:

$$\sup_{\mathcal{P}^{(r)}} O\left(d^{-2} N^{-2(r-1)} \ln(N)^{2(r-1)} \ln(d)^{2(r-S+1)}\right) \int \prod_{j=1}^{S} (L^d)^2\left(\sum_{\pm\lambda_i \in P_j} \pm\lambda_i\right) d(\lambda_1 \ldots \lambda_r).$$

Now we first bound $(L^d)^2(.) \leq d\, L^d(.)$ and then integrate with respect to $\lambda_1, \ldots, \lambda_r$ by using Lemma A.1 i) maximally r-1 times. The factor $L^d(0) = d$ will appear exactly once, since $\mathcal{P}^{(r)}$ is indecomposable. We obtain a final bound of:

$$\sup_{\mathcal{P}^{(r)}} O\left(d^{-2} N^{-2(r-1)} \ln(N)^{2(r-1)} \ln(d)^{2(r-S+1)}\right) d^{S+1} \ln(d)^{r-1}$$

which proves (1), since $S \leq r$. □

We are now able to prove the lemmata dealing with the asymptotic behaviour of $V'_{d,T}$ and $W'_{d,T}$.

Lemma III.10 Assume that (D2) and $d^{*-1} \ln^4(T) \ln^6(d^*) \to 0$ hold. Then we have

$$\sup_{d \leq d_{max}} L_T^{-1}(d) \left| V'_{d,T} - EV'_{d,T} \right| \to_P 0. \quad \bullet$$

Proof. First we show, for a fixed sequence $d \leq d_{max}$ and setting $N := N_d$, that $\mathrm{cum}_k\left(V'_{d,T}, \ldots, V'_{d,T}\right)$, the k-th order cumulant of $V'_{d,T}$, equals:

$$(1) \quad (4\pi)^{-k} \int \mathrm{cum}\left[\left(I^*_{d,T}-1\right)^2(\lambda_1),\ldots,\left(I^*_{d,T}-1\right)^2(\lambda_k)\right] d(\lambda_1\ldots\lambda_k) = O\left(d^{k/2} N^{-k} \ln(d)^{3k/2} \ln(N)^k\right).$$

To see that this holds consider again indecomposable partitions $\mathcal{P}^{(k)} = \{P_1, \ldots, P_S\}$ of the $2 \times k$ table:

$$\begin{matrix} \left(I^*_{d,T}-1\right)(\lambda_1) & \left(I^*_{d,T}-1\right)(-\lambda_1) \\ \vdots & \vdots \\ \left(I^*_{d,T}-1\right)(\lambda_k) & \left(I^*_{d,T}-1\right)(-\lambda_k) \end{matrix}$$

with $|P_i| \geq 2$ (it follows $S \leq k$). Then from the Product Theorem for Cumulants we have:

$$\int \mathrm{cum}\left[\left(I^*_{d,T}-1\right)^2(\lambda_1),\ldots,\left(I^*_{d,T}-1\right)^2(\lambda_k)\right] d(\lambda_1\ldots\lambda_k) =$$

$$= O(1) \sup_{\mathcal{P}(k)} \int \prod_{j=1}^{S} \mathrm{cum}\left[\left\{I^*_{d,T}(\pm\lambda_i), \pm\lambda_i \in P_j\right\}\right] d(\lambda_1\ldots\lambda_k)$$

$$\leq O(1) \sup_{\mathcal{P}(k)} \sqrt{\prod_{j=1}^{S} \int \mathrm{cum}^2\left[\left\{I^*_{d,T}(\pm\lambda_i), \pm\lambda_i \in P_j\right\}\right] d(\lambda_1\ldots\lambda_k)}$$

$$\leq O(1) \sup_{\mathcal{P}(k)} \sqrt{d^{2k-S} N^{-2(2k-S)} \ln(N)^{2(2k-S)} \ln(d)^{2k+S}},$$

by using Proposition III.9, which proves (1) since $S \leq k$.

From (1) and the Product Theorem for Cumulants we obtain directly:

$$(2) \qquad E\left(V'_{d,T} - EV'_{d,T}\right)^4 = O\left(d^2 N^{-4} \ln(N)^4 \ln(d)^6\right).$$

From (2) we obtain our assertion by using $L_T(d) \geq L_T(d^*)$ for $d \leq d^*$:

$$P\left\{ \sup_{d \leq d_{max}} L_T^{-1}(d) \left|V'_{d,T} - EV'_{d,T}\right| > \varepsilon \right\} \leq O\left(\varepsilon^{-4}\right) \sum_{d=1}^{d_{max}} d^2 \ln(T)^4 \ln(d)^6 \left(d \vee d^*\right)^{-4}$$

$= O\left(\ln(T)^4 \ln(d^*)^6\right)\left(d^*\right)^{-1} \to 0$, because of our assumption. □

Lemma III.11 Assume that (D2) and $d^{*-1} \ln^4(T) \ln^6(d^*) \to 0$ hold. Then we have

$$\sup_{d \leq d_{max}} L_T^{-1}(d) \left|W'_{d,T}\right| \to_P 0. \qquad \bullet$$

Proof. First we show, for a fixed sequence $d \le d_{max}$ and setting $N := N_d$, that :

(1) $$E\left(W'_{d,T}\right)^8 = O\left(\tilde{B}_d^4\, d^2\, N^{-4} \ln(N)^4 \ln(d)^6\right).$$

To see this observe that for even k we have:

$$\left[E\left(W'_{d,T}\right)^k\right]^2 = (2\pi)^{-2k} \left\{\int E\left[\left(I^*_{d,T}(\lambda_1) - 1\right)\cdots\left(I^*_{d,T}(\lambda_k) - 1\right)\right] \prod_{j=1}^{k} \left[f\, \tilde{f}_d^{-1}(\lambda_j) - 1\right] d(\lambda_1 \ldots \lambda_k)\right\}^2$$

$$= O(1)\, \tilde{B}_d^k \int \left\{E\left[\left(I^*_{d,T}(\lambda_1) - 1\right)\cdots\left(I^*_{d,T}(\lambda_k) - 1\right)\right]\right\}^2 d(\lambda_1 \ldots \lambda_k),$$

since $\int \left[f\, \tilde{f}_d^{-1}(\mu) - 1\right]^2 d\mu = O(\tilde{B}_d)$, which follows from an expansion of the ln-term in $\tilde{B}_d$ together with the facts that f is bounded from above and away from 0 (assumption (A)) and that $\tilde{f}_d$ converges uniformly to f (Lemma I.3).

Let $\mathcal{P} = \{P_1, \ldots, P_S\}$ be partitions of $I^*_{d,T}(\lambda_1), \ldots, I^*_{d,T}(\lambda_k)$ with $|P_i| \ge 2$ (it follows $S \le k/2$). Let us write cum(P) for $\mathrm{cum}\left(I^*_{d,T}(\lambda_1), \ldots, I^*_{d,T}(\lambda_r)\right)$ if $P = \left\{I^*_{d,T}(\lambda_1), \ldots, I^*_{d,T}(\lambda_r)\right\}$. Then from the Product Theorem for Cumulants, Proposition III.9 and since $EI^*_{d,T} = 1$, we obtain:

$$\left[E\left(W'_{d,T}\right)^k\right]^2 = O(1)\, \tilde{B}_d^k\, \sup_{\mathcal{P}} \prod_{j=1}^{S} \int \{\mathrm{cum}(P_j)\}^2\, d(\lambda \in P_j)$$

$$= O(1)\, \tilde{B}_d^k\, d^{k-S}\, N^{-2(k-S)} (\ln(d))^{k+S} [\ln(N)]^{2(k-S)}.$$

This proves (1) (setting k=4) because $S \le k/2$.

From (1) we obtain our assertion by using $L_T^{-1}(d)\, \tilde{B}_d \le 1$ and $L_T(d) \ge L_T(d^*)$ for $d \le d^*$:

$$P\left\{\sup_{d \le d_{max}} L_T^{-1}(d) \left|W'_{d,T}\right| > \varepsilon\right\} \le O\left(\varepsilon^{-8}\right) \sum_{d=1}^{d_{max}} d^2 \ln(T)^4 \ln(d)^6 \left(d \vee d^*\right)^{-4}$$

$$= O\left(\ln(T)^4 \ln(d^*)^6\right)\left(d^*\right)^{-1} \to 0,$$

because of our assumption. □

The proof of Lemma III.5 will now follow easily.

Proof of Lemma III.5 We have:

$$\left|\frac{\Delta_T(d)}{L_T(d)} - 1\right| = \left|\frac{V_{d,T} - EV'_{d,T} + W_{d,T}}{L_T(d)}\right| = \left|\frac{V'_{d,T} - EV'_{d,T} + W'_{d,T}}{L_T(d)}\right| + o_P(1) = o_P(1),$$

the $o_P(1)$ uniformly in $d \le d_{max}$. The second equality follows from Lemma III.8, the last equality from Lemmata III.10 and III.11. □

We now proceed to the proof of Lemma III.6. We first introduce some notation and prove a lemma which gives a bound for the fourth moment of $d^{-1}\, tr\left[\hat{\Gamma}^*_{d,T} - I\right]^2$. Let:

$$X_{d,T}(-\lambda,\mu) = \overline{b}^t_\lambda\left[\hat{\Gamma}^*_{d,T} - I\right]b_\mu \text{ and } Z_{d,T} := d^{-1}(2\pi)^{-2}\int \left|X_{d,T}(-\lambda,\mu)\right|^2 - E\left|X_{d,T}(-\lambda,\mu)\right|^2 d(\lambda\mu).$$

Observe that $d^{-1}\, tr\left[\hat{\Gamma}^*_{d,T} - I\right]^2 = d^{-1}(2\pi)^{-2}\int \left|X_{d,T}(-\lambda,\mu)\right|^2 d(\lambda\mu)$.

Lemma III.12 Assume that (D2) and $dN_d^{-1}\ln(T)\to 0$ hold. Then we have

$$EZ^4_{d,T} = O\left(\frac{d^2\ln^4(N_d)\ln^2(d)}{N_d^4}\right).$$ •

Proof. For a fixed sequence $d \le d_{max}$ and setting $N := N_d$ we have from the Product Theorem for Cumulants that $EZ^4_{d,T} = 3\,cum_2^2(Z_{d,T}) + cum_4(Z_{d,T})$, where $cum_k(.)$ denotes the k-th order cumulant. Therefore we first calculate $cum_r(Z_{d,T})$.

Consider indecomposable partitions $\mathcal{O}^{(r)} = \{O_1,\ldots,O_m\}$ of the $2\times r$ table

$$\begin{matrix} X_{d,T}(-\lambda_1,\mu_1) & X_{d,T}(\lambda_1,-\mu_1) \\ \vdots & \vdots \\ X_{d,T}(-\lambda_r,\mu_r) & X_{d,T}(\lambda_r,-\mu_r) \end{matrix},$$

which do not contain any one-element partition subset. Observe that, because of the indecomposibility of the partitions, they also do not contain any one-row partition subset.

We obtain from the Product Theorem for Cumulants that

$$cum_r(Z_{d,T}) := d^{-r}(2\pi)^{-2r}\sum_{ip,\mathcal{O}^{(r)}}\int \prod_{i=1}^{m} cum(O_i)\prod_{j=1}^{r} d(\lambda_j\mu_j),$$

where $\sum_{ip,\mathcal{O}^{(r)}}$ denotes summation over indecomposable partitions $\mathcal{O}^{(r)}$ and cum(O) denotes the cumulant of the random variables contained in the partition subset O, e.g. $\text{cum}(O) := \text{cum}\big(X_{d,T}(\lambda_1,\mu_1),\dots, X_{d,T}(\lambda_k,\mu_k) \big)$ if $O = \{ X_{d,T}(\lambda_1,\mu_1),\dots, X_{d,T}(\lambda_k,\mu_k)\}$.

Let us try to calculate cum(O), O as above.We consider partitions $\mathcal{P}^{(k)} = \{ P_1, \dots, P_S\}$ of the $2\times k$ table I.1 and adopt the notation used there. Observing that $X_{d,T}(\lambda,\mu) = \mathbf{X}^t A(\lambda,\mu)\,\mathbf{X} - b_\lambda^t b_\mu$, with $A(\lambda,\mu) := N^{-1}\sum_{i=1}^{N} E_i^t \left(U_d^t\right)^{-1} b_\mu b_\lambda^t U_d^{-1} E_i$, and

$$\mathbf{b}_\alpha^t\, A(\lambda,\mu)\, \mathbf{b}_\beta = N^{-1}\, \Theta^N(\alpha+\beta)\;\; b_\alpha^t\left(U_d^t\right)^{-1} b_\mu\;\; b_\lambda^t U_d^{-1} b_\beta\;\; \exp\left[-i\,(\alpha+\beta)\right]$$

we obtain from Proposition I.6 b):

$$\text{cum}(O) = N^{-k} \sum_{ip,\,\mathcal{P}^{(k)}} \int \prod_{j=1}^{S} f^{(s_j)}(\tilde{\kappa}_j) \prod_{i=1}^{r} \left[\Theta^N(\alpha_i+\beta_i)\, b_{\alpha_i}^t\left(U_d^t\right)^{-1} b_{\mu_i}\;\; b_{\lambda_i}^t U_d^{-1} b_{\beta_i}\right] \prod_{j=1}^{S} \mathbf{d}\tilde{\kappa}_j$$

Now we are ready to give a closed expression for $\text{cum}_r(Z_{d,T})$: regard the above mentioned partitions $\mathcal{O}^{(r)}$ as partitions of the $4 \times r$ table

$$\begin{matrix} \alpha_1 & \beta_1 & \gamma_1 & \delta_1 \\ \vdots & \vdots & \vdots & \vdots \\ \alpha_r & \beta_r & \gamma_r & \delta_r \end{matrix}$$

by identifying (α_i,β_i) with $X_{d,T}(-\lambda_i,\mu_i)$ and (γ_i,δ_i) with $X_{d,T}(\lambda_i,-\mu_i)$. Further consider partitions $\mathcal{P}(\mathcal{O})$ of this table which are defined by the above mentioned two steps: first take the indecomposable partitions $\mathcal{O}^{(r)}$ of the $2\times r$ table which is formed when regarding (α_i,β_i) and (γ_i,δ_i) as one element respectively; then for each partition-subset O with $|O|=k$ (remember one element of O is a pair) form a $2\times k$ table and take indecomposable partitions of it. Note that this two-step procedure gives a partition $\mathcal{P}(\mathcal{O})$ of the $4\times r$ table, which does not contain subsets of the form $\{\alpha_i\}$ or $\{\beta_i\}$ or $\{\gamma_i\}$ or $\{\delta_i\}$ or $\{\alpha_i,\beta_i\}$ or $\{\gamma_i,\delta_i\}$ or $\{\alpha_i,\beta_i,\gamma_i,\delta_i\}$.

Now using $(2\pi)^{-1} \int b_\alpha^t\left(U_d^t\right)^{-1} b_\lambda\;\; \overline{b}_\lambda^t U_d^{-1} b_\beta\; \mathbf{d}\lambda = K_d(-\alpha,\beta)$ and Lemma A.2 we get:

(1)

$$|\text{cum}_r(Z_{d,T})| = O(1)\, d^{-r} N^{-2r} \sup_{\mathcal{P}(\mathcal{O})} \int \prod_{i=1}^{r} L^N(\alpha_i+\beta_i)\, L^N(\gamma_i+\delta_i)\, L^d(\alpha_i+\gamma_i)\, L^d(\beta_i+\delta_i)\; \mathbf{d}\tilde{\tilde{\kappa}}$$

By $\mathbf{d}\tilde{\tilde{\kappa}}$, and in order to avoid more complicated notation, from this point on we mean integration with respect to all variables $\widetilde{\kappa}_i$ involved in the current preceding expression under the integral sign.

It remains to show that the quantity in (1) has the desired order. The following remark will be helpful (to avoid confusion let us remind that the elements of the partition-subsets of $\mathcal{d}^{(r)}$ are 'pairs'):

(2) If $\mathcal{d}^{(r)}$ contains a 2-element partition subset, say $\{(\alpha_1,\beta_1),(\gamma_2,\delta_2)\}$, then the integral over the L^N, L^d factors depending on these variables is $O(d^3 N \ln(N) \ln(d))$:

$$\int L^N(\alpha_1+\beta_1)\, L^N(\gamma_2+\delta_2)\, L^d(\alpha_1+\gamma_1)\, L^d(\beta_1+\delta_1)\, L^d(\alpha_2+\gamma_2)\, L^d(\beta_2+\delta_2)\, d\widetilde{\widetilde{\kappa}}$$

$$= O(d^3 N \ln(N) \ln(d))$$

This follows from Lemma A.1 for each indecomposable partition of

$$\begin{matrix} \alpha_1 & \beta_1 \\ \gamma_2 & \delta_2 \end{matrix} .$$

We will prove that the quantity in (1) has the desired order by a clever exhaustive enumeration using, on the one hand, remark (2) and, on the other,

(3) we will bound all remaining L^d factors by d. The integral of the remaining L^N factors will, by Lemma A.1 i), give an additional factor $N \ln^s(N)$ for each partition subset (PSS) O of $\mathcal{d}^{(r)}$ with $|O| = s+1$.

We have the following cases:

I) r=2.

i) $\mathcal{d}^{(2)}$ consists of 2 PSS each with 2 elements. Using (2) for one of the PSS and (3) for the rest we get a bound of $O(d^3 N^2 \ln^2(N) \ln(d))$.

ii) $\mathcal{d}^{(2)}$ consists of one PSS with 4 elements. Using directly (3) we obtain $O(d^4 N \ln^3(N))$ as bound.

II) r=4.

i) $\mathcal{d}^{(4)}$ consists of 4 PSS each with 2 elements. Then at least two of the PSS occupy different rows. Using (2) for these two PSS and (3) for the rest we get $O(d^6 N^4 \ln^4(N) \ln^2(d))$ as bound.

ii) i) is not fulfilled but $\mathcal{A}^{(4)}$ contains a PSS with 2 elements. After applying (2) for this PSS there are 4 L^d factors left and the number of remaining PSS is maximally 2. By (3) we get a final bound of $O\left(d^7 N^3 \ln^5(N) \ln(d)\right)$.

iii) If there is non PSS with 2 elements, then the total number of PSS has to be less than or equal to 2. Accordingly (3) yields a bound of $O\left(d^8 N^2 \ln^6(N)\right)$.

In case I, the bounds obtained are $O\left(d^3 N^2 \ln^2(N) \ln(d)\right)$ and in case II $O\left(d^6 N^4 \ln^4(N) \ln^2(d)\right)$ (because $dN^{-1}\ln(T)\to 0$) which, with (1), yields the assertion of the lemma. □

Proof of Lemma III.6 From Lemmata II.3 and I.2 it follows that

$$E\, d^{-1} \operatorname{tr}\left[\hat{\Gamma}^*_{d,T} - I\right]^2 = dN^{-1} + o(L_T(d)),$$

the o-term being uniformly in $d \le d_{max}$.

Thus it is sufficient to prove $\quad \sup_{d\le d_{max}} L_T^{-1}(d)\, Z_{d,T} \to_P 0.$

From Lemma III.12 we have

$$P\left\{\sup_{d\le d_{max}} L_T^{-1}(d)\, Z_{d,T} \ge \varepsilon\right\} \le \varepsilon^{-4} \sum_{d=1}^{d^*} L_T^{-4}(d^*)\, EZ^4_{d,T} + \varepsilon^{-4} \sum_{d=d^*}^{d_{max}} L_T^{-4}(d)\, EZ^4_{d,T}.$$

$$= O\left(d^{*-1} \ln^4(T) \ln^2(d^*)\right) \to 0,$$

which proves our assertion. □

III.2.3 The linear term

In this section we prove Lemma III.7, whose assertion is that the term $d^{-1} \operatorname{tr} \hat{\Gamma}^*_{d,T}$ is smooth as a function in d when compared to $L_T(d)$. We will prove this by bounding the fourth order cumulants of the increment. This will be done in the lemma which we state and prove immediately.

First a notational convention: assume that $d_1 \le d_2 \le d_{max}$ and that quantities indicated by d_j will subsequently be indicated by 'j', e.g. we write $\hat{\Gamma}^*_{1,T}$ instead of $\hat{\Gamma}^*_{d_1,T}$ and $\tilde{B}_j$ instead of $\tilde{B}_{d_j}$.

Lemma III.13 Assume that (D2) holds. Then

$$E\left[d_1^{-1} \operatorname{tr} \hat{\Gamma}^*_{1,T} - d_2^{-1} \operatorname{tr} \hat{\Gamma}^*_{2,T} \right]^4 = O(1) \max_{j=1,2}\left[N_j^{-1}\left(\tilde{B}_j + \frac{d_j \ln(T)}{N_j}\right)\right]^2 \vee \left(\frac{d_2-d_1}{N_2N_1}\right)^2. \bullet$$

Proof. For any random variable Δ with $E\Delta=0$ we have: $E\Delta^4 = 3\,\text{cum}_2^2(\Delta) + \text{cum}_4(\Delta)$. Our first task in the proof will be to calculate the k-th order cumulant of $\Delta := d_1^{-1} \operatorname{tr} \hat{\Gamma}^*_{1,T} - d_2^{-1} \operatorname{tr} \hat{\Gamma}^*_{2,T}$. We have:

$$\Delta := X^t[A_1 - A_2]\,X, \text{ with } A_j := (d_jN_j)^{-1}\sum_{i=1}^{N_j} E_i^t\Gamma_j^{-1}E_i,$$

where E_1 is also considered to be a $d_2{\times}T$ matrix (filling up with 0's) and we consider Γ_1 embedded in a $d_2{\times}d_2$ matrix (again filling up with 0's).

Setting $W_{j,T}(\alpha,\beta) := b^t_\alpha A_j b_\beta = (d_jN_j)^{-1}\,\Theta^{N_j}(\alpha+\beta)\;K_j(-\alpha,\beta)\;\exp[-i(\alpha+\beta)]$ Proposition I.6 b) yields (we adopt the same notation as in table I.1):

$$\text{cum}_k(\Delta) = \sum_{ip\,\mathcal{P}^{(k)}} \int \prod_{i=1}^{S} f^{(s_i)}(\tilde{\kappa}_i) \prod_{i=1}^{k}\left[W_{1,T}(\alpha_i,\beta_i) - W_{2,T}(\alpha_i,\beta_i)\right]\prod_{i=1}^{S} d\tilde{\kappa}_i.$$

By the Hölder inequality and a variable transformation we see that:

$$|\text{cum}_k(\Delta)| = O(1)\left[\int\left[N_1^{-1}\Theta^{N_1}(\lambda)\,d_1^{-1}K_1(-\mu,\lambda-\mu) - N_2^{-1}\Theta^{N_2}(\lambda)\,d_2^{-1}K_2(-\mu,\lambda-\mu)\right]^2 d\lambda\mu\right]^{k/2}$$

The triangular inequality for the L_2-norm yields that $E\Delta^4$ is less than or equal to some constant times:

$$\left\{\sum_{j=1}^{2}\left\|N_j^{-1}\Theta^{N_j}(\lambda)\left[d_j^{-1}2\pi\,K_j(-\mu,\lambda-\mu) - f^{-1}(\mu)\right]\right\|_2 + \left\|f^{-1}(\mu)\left[N_1^{-1}\Theta^{N_1}(\lambda) - N_2^{-1}\Theta^{N_2}(\lambda)\right]\right\|_2\right\}^4$$

Now we have:

i) $\int \left|N_1^{-1}\Theta^{N_1}(\lambda) - N_2^{-1}\Theta^{N_2}(\lambda)\right|^2 d\lambda = O\left(\frac{d_2 - d_1}{N_2\,N_1}\right)$, as seen by an elementary calculation.

ii) $\int \left|N^{-1}\Theta^N(\lambda)\left[d^{-1}2\pi\,K_d(-\mu,\lambda-\mu) - f^{-1}(\mu)\right]\right|^2 d(\lambda\mu) \le O(N^{-1})\left[\tilde{B}_d + dN^{-1}\ln(N)\right]$.

This follows directly, observing that Proposition II.10 ii) yields that N times the left hand side in ii) equals

$$\int \Delta^N(\lambda)\left(\left[d^{-1}2\pi\,K_d(\mu,\mu) - f^{-1}(\mu)\right]^2 + \left[\left(O(d)\lambda\right)\wedge 1\right]\right)d(\lambda\mu).$$

The proof of ii) is obtained from this together with Proposition II.11 and from $\int \left[\tilde{f}_d^{-1}(\mu) - f^{-1}(\mu)\right]^2 d\mu = O(\tilde{B}_d)$, which follows from an expansion of the ln-term in $\tilde{B}_d$ together with the facts that f is bounded from above and away from 0 (assumption (A)) and that $\tilde{f}_d$ converges uniformly to f (Lemma I.3).

The assertion of the lemma follows directly from i) and ii). □

Proof of Lemma III.7 Setting $\Delta_{d,T} := d^{-1} \operatorname{tr} \hat{\Gamma}^*_{d,T} - d^{*-1} \operatorname{tr} \hat{\Gamma}^*_{d^*,T}$ we obtain :

(1) $$E\Delta^4_{d,T} L_T^{-4}(d) = o\left[\ln^2(T) \left(d \vee d^*\right)^{-2}\right]$$

by using Lemma III.13 and the following points (2), (3) and (4):

(2) $$\tilde{B}_{d_o} + N_{d_o}^{-1} d_o \ln(T) = O(L_T(d) \ln(T)) \text{ for } d_o = d, d^*.$$

(3) $$N_{d_o}^{-1} L_T^{-1}(d) = o\left[\left(d \vee d^*\right)^{-1}\right] \text{ for } d_o = d, d^*$$

(4) $$N_d^{-1} N_{d^*}^{-1} \left| d - d^* \right| L_T^{-2}(d) = o\left[\left(d \vee d^*\right)^{-1}\right].$$

Let us prove (2). For $d_o = d$ we have:

$$\frac{\tilde{B}_d + N_d^{-1} d \ln(T)}{L_T(d)} \le \frac{\tilde{B}_d}{\tilde{B}_d} + \frac{N_d^{-1} d \ln(T)}{N_d^{-1} d} = O(\ln(T)).$$

For $d_o = d^*$ (2) follows from the preceding together with $L_T(d^*) \le L_T(d)$. By similar arguments, (3) and (4) can be proven by using $L_T(d^*) \le L_T(d)$ for $d \le d^*$.

From (1) it follows that:

$$P\left\{\sup_{d \le d_{max}} \Delta_{d,T} L_T^{-1}(d) \ge \varepsilon\right\} = O(\ln^2(T)) \sum_{d=1}^{d_{max}} \left(d \vee d^*\right)^{-2} = o\left(d^{*-1} \ln^2(T)\right) \to 0,$$

which proves the assertion of the lemma. □

III.2.4 Proofs of the main results

We now proceed to the proof of Theorems III.1 and III.2 as well as Lemma III.3.

Proof of Theorem III.1 Is a direct consequence of Lemma III.5 and the definition of d^*. □

Proof of Theorem III.2 We will first prove

(1) $$\frac{L_T(\hat{d})}{L_T(d^*)} \rightarrow_P 1.$$

Let $S_T'(d) := S_T(d) - (2\pi)^{-1} \int \ln(f(\lambda))\, d\lambda \; - d^{-1} \operatorname{tr}\left[\hat{\Gamma}^*_{d,T} - I\right]$. From Lemma III.6 we have:
$S_T'(d) = \Delta_T(d) + o_P(L_T(d))$ uniformly in $d \le d_{max}$ and thus from Lemma III.5:

(2) $$L_T(d) = S_T'(d) + o_P(L_T(d)) = S_T'(d^*) + \left(S_T'(d) - S_T'(d^*) \right) + o_P(L_T(d))$$

$$= L_T(d^*) + \left(S_T'(d) - S_T'(d^*) \right) + o_P(L_T(d)) \text{ uniformly in } d \le d_{max}.$$

From Lemma III.7 we have:

(3) $$S_T'(d) - S_T'(d^*) = S_T(d) - S_T(d^*) + o_P(L_T(d)) \text{ uniformly in } d \le d_{max}.$$

From (2) and (3), setting $d := \hat{d}$ and observing $S_T(\hat{d}) - S_T(d^*) \le 0$ we obtain:

(4) $$1 \le \frac{L_T(d^*)}{L_T(\hat{d})} + o_P(1),$$

which together with the definition of d^* proves (1).

The assertion i) of the theorem follows directly from (1) and Lemma III.5.
To prove the second assertion observe that (1) yields that it is sufficient to show:

$$L_T^{-1}(d^*) \inf_{d \le d_{max}} \Delta_T(d) \rightarrow_P 1.$$

This is true, since, on the one hand, $L_T^{-1}(d^*) \inf_{d \le d_{max}} \Delta_T(d) \le L_T^{-1}(d^*)\, \Delta_T(d^*) \rightarrow_P 1$ (because of Lemma III.5) and, on the other, $L_T^{-1}(d^*) \inf_{d \le d_{max}} \Delta_T(d) \ge \inf_{d \le d_{max}} L_T^{-1}(d)\, \Delta_T(d) \rightarrow_P 1$ (again because of Lemma III.5). This completes the proof of ii). □

Proof of Lemma III.3 Follows exactly the lines of the proof of Theorem III.2, by substituting $\hat{d}$ by $\hat{d}^P$ and modifying (3) into

(3) $$S_T'(d) - S_T'(d^*) = S_T(d) - S_T(d^*) + o_P(L_T(d)) = \mathring{S}_T(d) - \mathring{S}_T(d^*) + o_P(L_T(d))$$

uniformly in $d \le d_{max}$, the last equality following from the assumption of the lemma. □

Proof of Lemma III.4 Statement i) follows from Lemma III.3 by observing that $(2\pi)^{-1}\int \ln \hat{\hat{f}}_{d,T}(\lambda)\, d\lambda = (2\pi)^{-1}\int \ln \hat{f}_{d,T}(\lambda)\, d\lambda + \ln(\mu_d)$ and $\ln(\mu_d) = d/N + o(d/N)$.

In order to prove ii) we first show that:

$$\frac{\Delta(f,\hat{\hat{f}}_{d,T})}{L_T(d)} \to_P 1 \tag{1}$$

uniformly in $d \le d_{max}$. To show this note that we have:

$$\Delta(f,\hat{\hat{f}}_{d,T}) = \Delta(f,\hat{f}_{d,T}) + \ln(\mu_d) + \left[\mu_d^{-1}-1\right] d^{-1}\,\mathrm{tr}\left[\hat{\Gamma}^{*\,-1}\right]$$

$$= \Delta(f,\hat{f}_{d,T}) + \ln(\mu_d) + \mu_d^{-1}-1 + o_p(L_T(d)),$$

since from Lemmata II.1 and I.2 follows that $d^{-1}\,\mathrm{tr}\left[\hat{\Gamma}^{*\,-1} - I\right] = o_P(1)$ uniformly in $d \le d_{max}$. From this, together with Lemma III.5, we obtain (1).

Now ii) follows easily:

$$\frac{\Delta(f,\hat{\hat{f}}_{\hat{\hat{d}},T})}{L_T(d^*)} = \frac{\Delta(f,\hat{\hat{f}}_{\hat{\hat{d}},T})}{L_T(\hat{\hat{d}})}\,\frac{L_T(\hat{\hat{d}})}{\Delta(f,\hat{f}_{\hat{\hat{d}},T})}\,\frac{\Delta(f,\hat{f}_{\hat{\hat{d}},T})}{L_T(d^*)} \to_P 1,$$

since the first factor on the right converges to 1 because of (1), the second because of Lemma III.5 and the third because of i). □

IV. USING OTHER COVARIANCE MATRIX ESTIMATORS

IV.1 INTRODUCTION, RESULTS

In all previous chapters we were concerned with the Capon estimator as defined in (0.4) using the segment covariance matrix estimator defined in (0.5). In this chapter we will deal with the behaviour of the Capon estimator when in its definition we use other covariance matrix estimators. Two possibilities are:

a) The symmetrized segment covariance matrix estimator, which consists of a symmetrization of the segment covariance matrix estimator (0.5) (defined in (IV.1) below). In terms of Burg's relation ((0.8), a modification similar to (II.1), is here also necessary) this corresponds to a harmonic mean of 'forward-backward' LS autoregressive estimators (see Section V.1.3).

b) The (tapered) Toeplitz covariance matrix estimator, which is the Toeplitz matrix corresponding to a tapered periodogram (defined in (IV.2) below). In terms of Burg's relation (0.8) this corresponds to an harmonic mean of autoregressive estimators based on the Yule-Walker equations when using the empirical covariances of a tapered periodogram (see Section V.1.3).

In this chapter we will prove that the Central Limit Theorem (eventually properly modified) holds, when using these covariance matrix estimators for the definition of the Capon estimator (Theorems IV.1 and IV.2).

For the estimator in 'a' this will be a direct consequence from what is already known from Chapter I: it turns out that the first order expansion term of the thus defined Capon estimator is identical to $I^*_{d,c,T}$ (see relation (I.3)). For the estimator in 'b' the methods and some of the results of Chapter I will be helpful, but the proofs are still rather different. For example, in order to bound $E\ \mathrm{tr}\left[\hat{\Gamma}^*_{d,T} - I\right]^k$ or the higher order cumulants of the estimator, we again have to bound integral expressions as in the proofs of Lemmata I.5 and I.10. Their structure however is totally different; the arguments used in these lemmata to bound the expressions do not work. This is the reason why we could not treat all cases by a unique argument and we refered cases 'a' and 'b' to this chapter. One meets similar difficulties when trying, for cases 'a' and 'b', an expansion of the type obtained in Chapter II (in order to develop a dimension selection criterion) or when trying to prove asymptotic efficiency of a dimension selection criterion.

IV.1.1 The symmetrized covariance matrix estimator

We define the symmetrized covariance matrix estimator as

$$\text{(IV.1)} \qquad \hat{\Gamma}^{(sym)}_{d,c,T} := \frac{1}{2}\left[\hat{\Gamma}_{d,c,T} + \hat{\Gamma}^{(rev)}_{d,c,T}\right],$$

with $\hat{\Gamma}^{(rev)}_{d,c,T} := \frac{1}{N}\sum_{i:=1}^{N} Y_i^{(rev)} \left(Y_i^{(rev)}\right)^t$, where $N = N_{d,c,T}$ as in (0.5) and

$Y_i^{(rev)} = Y_i^{d,c\,(rev)} := (X_{(i-1)c+d},\ldots,X_{(i-1)c+1})^t$, that is the $Y_i^{d,c}$ (see (0.5)) in reversed order.

Further we define the Capon estimator as

$$\hat{f}_T^{(sym)}(\lambda) = \hat{f}^{(sym)}_{d,c,T}(\lambda) := \frac{d}{2\pi}\left(\bar{b}^t_\lambda \left[\hat{\Gamma}^{(sym)}_{d,c,T}\right]^{-1} b_\lambda\right)^{-1},\ \lambda \in [-\pi,\pi].$$

For fixed $v \in \mathbf{R}^+$ and $\lambda_k \in [-\pi, \pi]$, $k=1,\ldots,K$ let ζ^v and θ be defined as in Chapter I. Then:

Theorem IV.1 Suppose that (A), (B) and (C) hold and that the sequences c_T, d_T fulfill

i) $c\, d^{1+\varepsilon} / T \to 0$ for some $\varepsilon > 0$

ii) $c / d \le C$ for some constant $C < \infty$.

iii) $d^{-\beta} \ln(T) (\ln(d))^2 \ln(c) \to 0$ (where $\beta := \frac{r+\alpha}{1+r+\alpha}$, r, α as in (B)) as T tends to infinity. Then setting $v := \lim_{T\to\infty} \theta(d_T / c_T)$ (assuming it exists) <u>we have</u>:

a) $\sqrt{T d^{-1}} \left\{\frac{\hat{f}^{(sym)}_{d,c,T}}{\tilde{f}_d}(\lambda_k) - 1\right\}_{k=1,\ldots,K} \Rightarrow_D \zeta^v$

b) $\sqrt{T d^{-1}} \left\{\frac{\hat{f}^{(sym)}_{d,c,T}}{f}(\lambda_k) - 1 - B_d(\lambda_k)\right\}_{k=1,\ldots,K} \Rightarrow_D \zeta^v$, where $B_d := \tilde{f}_d f^{-1} - 1$.

c) If in addition $T d^{-(1+2\gamma)} \ln^4(d) \to 0$, where $\gamma := (r+\alpha)\wedge 1$, then

$$\sqrt{T d^{-1}} \left\{\frac{\hat{f}^{(sym)}_{d,c,T}}{f}(\lambda_k) - 1\right\}_{k=1,\ldots,K} \Rightarrow_D \zeta^v. \qquad \bullet$$

IV.1.2 The Toeplitz covariance matrix estimator

We define the Toeplitz covariance matrix estimator $\hat{\Gamma}^{(h)}_{d,T}$ as follows:

$$\left[\hat{\Gamma}^{(h)}_{d,T}\right]_{i,j} := \hat{c}^{(h)}_T(|i-j|), \tag{IV.2}$$

where the empirical covariances are given by $\hat{c}^{(h)}_T(u) := H_T^{-1} \sum_{t=1}^{T-u} h_t X_t h_{t+u} X_{t+u}$, $u=0,\ldots,d-1$. Here $h_t := h(t\,T^{-1})$, $t=1,\ldots,T$, is a 'data taper' with $h:[0,1] \to \mathbf{R}$ a continuously differentiable function of bounded variation fulfilling $\|h\|_2 > 0$ and $H_T := \sum_{t=1}^{T} h_t^2$.

We may write $\hat{\Gamma}^{(h)}_{d,T} = \int \mathbf{I}^{(h)}_T(\alpha)\, b_\alpha \bar{b}^t_\alpha \, d\alpha$, where $\mathbf{I}^{(h)}_T(\alpha)$ is the tapered periodogram: $\mathbf{I}^{(h)}_T(\alpha) := (2\pi H_T)^{-1} \left| \mathbf{X}^t D^{(h)}_T \mathbf{b}_\alpha \right|^2$, $D^{(h)}_T$ being the diagonal matrix $\mathrm{diag}(h_1,\ldots,h_T)$.

Further we define the Capon estimator as

$$\hat{f}^{(h)}_T(\lambda) = \hat{f}^{(h)}_{d,T}(\lambda) := \frac{d}{2\pi} \left(\bar{b}^t_\lambda \left[\hat{\Gamma}^{(h)}_{d,T}\right]^{-1} b_\lambda \right)^{-1}, \quad \lambda \in [-\pi,\pi].$$

Finally we introduce the kernels $\Theta^{(T,h)}(\lambda) := \sum_{t=1}^{T} h_t e^{i\lambda t}$ and $\Delta^{(T,h)}(\lambda) := H_T^{-1} \left|\Theta^{(T,h)}(\lambda)\right|^2$ and the function $f^{(T,h)} := f * \Delta^{(T,h)}$ as well as the $d \times d$ Toeplitz matrix $G^{(T,h)}_d$ corresponding to $f^{(T,h)}$.

Concerning the data taper we have the following properties:

(T1) $H_T T^{-1} \to \|h\|_2^2 > 0$

(T2) $\left|\Theta^{(T,h)}\right| \le C\, L^T$ for some C (Dahlhaus (1983)).

(T3) If f fulfills (B) then $\|f^{(T,h)} - f\|_\infty = O(T^{-\gamma})$ with $\gamma := (r+\alpha) \wedge 1$ (Dahlhaus (1985)).

We now may state the Central Limit Theorem for the thus defined Capon estimator. The difference with the previous analogous theorems is that, beyond the bias which comes from the approximation of f via $\tilde{f}_d$, we now obtain a second bias term resulting from the approximation of f via $f^{(T,h)}$. It is well known that in small sample situations this approximation may be spoiled by leakage which can be reduced only by the choice of a good data taper (see e.g. Bloomfield (1976) Ch. 5.1-5.3).

Let $B_d^{(T,h)}(\lambda) := \left(\vec{b}_\lambda^t \Gamma_d^{-1} b_\lambda\right)^{-1} \vec{b}_\lambda^t \Gamma_d^{-1}\left[G_d^{(T,h)} - \Gamma_d\right] \Gamma_d^{-1} b_\lambda$. Then $\left\|B_d^{(T,h)}\right\|_\infty \leq \left\| f^{(T,h)} f^{-1} - 1\right\|_\infty$

holds, since $B_d^{(T,h)}(\lambda) = \int \left[f^{(T,h)} f^{-1} - 1\right](\mu)\ f(\mu) \left|\vec{b}_\lambda^t \Gamma_d^{-1} b_\mu\right|^2 \left(\vec{b}_\lambda^t \Gamma_d^{-1} b_\lambda\right)^{-1} d\mu.$

For fixed $v \in \mathbf{R}^+$ and $\lambda_k \in [-\pi, \pi]$, k=1,...,K let ζ^v and θ be defined as in Chapter I. Then:

Theorem IV.2 Suppose that (A), (B) and (C) hold and that the sequence d_T fulfills

i) $d^{1+\varepsilon} / T \to 0$ for some $\varepsilon > 0$

ii) $d^{-1/2} T^{1/2-\gamma} \to 0$, γ as in (T3) and

iii) $d^{-\beta} \ln(T) \ln(d) \to 0$ (where $\beta := \frac{r+\alpha}{1+r+\alpha}$, r, α as in (B)) as T tends to infinity.

Then with $v := 2/3\ \|h\|_2^{-2}$ we have:

a) $\sqrt{T d^{-1}} \left\{ \frac{\hat{f}_{d,T}^{(h)}}{\tilde{f}_d} (\lambda_k) - B_d^{(h)}(\lambda_k) - 1 \right\}_{k=1,\ldots,K} \Rightarrow_D \zeta^v$

b) $\sqrt{T d^{-1}} \left\{ \frac{\hat{f}_{d,T}^{(h)}}{f} (\lambda_k) - 1 - B_d(\lambda_k) - B_d^{(T,h)}(\lambda_k) \right\}_{k=1,\ldots,K} \Rightarrow_D \zeta^v$, where

$B_d := \tilde{f}_d f^{-1} - 1$ and $B_d^{(T,h)}$ as above.

c) If in addition $T d^{-(1+2\gamma)} \ln^4(d) \to 0$, where $\gamma := (r+\alpha) \wedge 1$, then

$\sqrt{T d^{-1}} \left\{ \frac{\hat{f}_{d,T}^{(h)}}{f} (\lambda_k) - 1 \right\}_{k=1,\ldots,K} \Rightarrow_D \zeta^v$. •

IV.2 DETAILED RESULTS, PROOFS

IV.2.1 The symmetrized covariance matrix estimator

Proof of Theorem IV.1 Let us define $\hat{\Gamma}_{d,c,T}^{(sym)*}$ and $I_{d,c,T}^{(sym)*}(\lambda)$ analogously to $\hat{\Gamma}_{d,c,T}^{*}$ and $I_{d,c,T}^{*}(\lambda)$ (relations (I.2) and (I.3)). We claim that:

(1) $$\left\|\hat{\Gamma}_{d,c,T}^{(sym)*} - I\right\| \leq \left\|\hat{\Gamma}_{d,c,T}^{*} - I\right\|.$$

For (1) it is sufficient to show

$$(2) \qquad \left\| \widehat{\Gamma}^{(rev)*}_{d,c,T} \right\| = \left\| \widehat{\Gamma}^{*}_{d,c,T} \right\|.$$

To see this, observe that if F_d is the d×d matrix with 1's in the diagonal from right-top to left-bottom, then we have for any $v \in \mathbf{R}^d$: $v^t \widehat{\Gamma}^{(rev)}_{d,c,T} v = (F_d v)^t \widehat{\Gamma}^{*}_{d,c,T} (F_d v)$. Accordingly: $v^t \widehat{\Gamma}^{(rev)*}_{d,c,T} v = w^t \widehat{\Gamma}^{*}_{d,c,T} w$ with $w := U_d^t F_d \left(U_d^t\right)^{-1} v$. From this, (2) follows directly by observing $\| v \|_2 = \| w \|_2$ (use $F_d \Gamma_d F_d = \Gamma_d$).

From (1) together with Lemmata I.2 and I.4, we obtain that under the conditions of the theorem

$$\sqrt{T d^{-1}} \left| \widehat{f}^{(sym)}_{d,c,T} \tilde{f}_d^{-1} (\lambda) - I^{(sym)*}_{d,c,T}(\lambda) \right| \to_P 0.$$

We now claim that $I^{(sym)*}_{d,c,T}(\lambda) \equiv I^{*}_{d,c,T}(\lambda)$. To see this observe that $Y_i^{(rev)} = F_d Y_i$, which yields:

$$N^{-1} \sum_{i=1}^{N} \left| \left(Y_i^{(rev)}\right)^t \Gamma_d^{-1} b_\lambda \right|^2 = N^{-1} \sum_{i=1}^{N} \left| Y_i^t F_d \Gamma_d^{-1} F_d b_\lambda \right|^2 = N^{-1} \sum_{i=1}^{N} \left| Y_i^t \Gamma_d^{-1} b_\lambda \right|^2.$$

Here we used $F_d \Gamma_d^{-1} F_d = \Gamma_d^{-1}$, which holds because the inverse of a Toeplitz matrix is persymmetric. The rest of the theorem follows immediately from Lemmata I.9 and I.10.

□

IV.2.2 The Toeplitz covariance matrix estimator

For proving Theorem IV.2 we follow a concept similar to the one we used in Chapter I. Let $\widehat{\Gamma}^{(h)*}_{d,T}$ and $I^{(h)*}_{d,T}(\lambda)$ (not to be confused with $I_T^{(h)}(\alpha)$) be defined analogously to $\widehat{\Gamma}^{*}_{d,c,T}$ and $I^{*}_{d,c,T}(\lambda)$ respectively (see relations (I.2) and (I.3)). First we examine the cumulants of $I^{(h)*}_{d,T}(\lambda)$. Secondly we show that, under the conditions of the theorem, $\widehat{f}^{(h)}_{d,T} \tilde{f}_d^{-1}$ and $I^{(h)*}_{d,T}(\lambda)$ are asymptotically equivalent.

Lemma IV.3 Assume that (A), (B) and (C) hold. Then we have:

i) $E I^{(h)*}_{d,T}(\lambda) - 1 = B_d^{(T,h)}(\lambda)$

ii)

$$\operatorname{cov}\left[\sqrt{\frac{T}{d}} I^{(h)*}_{d,T}(\mu_1), \sqrt{\frac{T}{d}} I^{(h)*}_{d,T}(\mu_2) \right] = \begin{cases} O(R_2), \text{ if } \mu_1 \neq \pm \mu_2 \bmod (2\pi) \\ \frac{2}{3} \| h \|_2^{-2} \left[\delta_{\mu_1+\mu_2} + \delta_{\mu_1-\mu_2} \right] + O(R_1), \text{ else} \end{cases}$$

with $\quad R_2 := d^{-2}\left[|\mu_1-\mu_2|^{-2} + |\mu_1+\mu_2|^{-2}\right] \ln(T)\ln(d) + R_1$ and

$$R_1 := d^{-\beta}\ln(T)\ln^2(d) + d\,T^{-1}, \text{ where } \beta := \frac{r+\alpha}{1+r+\alpha}, \; r,\alpha \text{ as in (B).} \quad \bullet$$

Proof. i) Is straight forward using $I_{d,T}^{(h)*}(\mu) = \tilde{f}_d(\mu)\, d^{-1}\, 2\pi \int I_T^{(h)}(\gamma)\,|K_d(\gamma,\mu)|^2 d\gamma$ and $EI_T^{(h)} = f^{(T,h)}$.

ii) Using $I_{d,T}^{(h)*}(\mu_j) = \tilde{f}_d(\mu_j)\, d^{-1}\, 2\pi \int I_T^{(h)}(\gamma_j)\,|K_d(\gamma_j,\mu_j)|^2 d\gamma_j$ and Proposition I.6 b) together with $I_T^{(h)}(\gamma_j) := (2\pi H_T)^{-1}\; X^t D_T^{(h)} b_{\gamma_j} \bar{b}_{\gamma_j}^t D_T^{(h)}\; X$ we obtain that the cumulant of interest equals (using the notation of table I.1):

$$d^{-3} T H_T^{-2} \prod_{j=1}^{2} \tilde{f}_d(\mu_j)\; \bullet$$

$$\sum_{ip,\,(2)} \int \prod_{i=1}^{S} f^{(s_i)}(\tilde{\kappa}_i) \prod_{j=1}^{2} \Theta^{(T,h)}(\alpha_j+\gamma_j)\, \Theta^{(T,h)}(\beta_j-\gamma_j) \prod_{j=1}^{2} |K_d(\gamma_j,\mu_j)|^2 \prod_{i=1}^{S} d\tilde{\kappa}_i \prod_{j=1}^{2} d\gamma_j$$

Now using Lemmata A.1 and A.2 and (T2), one may verify that each of the three terms in this sum corresponding to different partitions of the table is bounded. Moreover it is $O(R_1)$ for the partition $\mathcal{P} = \{(\alpha_1, \beta_1, \alpha_2, \beta_2)\}$. For the two other partitions the corresponding terms in the sum are bounded. This together with Lemma A.3 and (A.2.4) allows (as in the proof of Lemma I.9) to substitute all quantities depending on f in these terms by the ones corresponding to $f = (2\pi)^{-1}$ having a total error of $O(R_1)$.

Next for the partition $\mathcal{P} = \{\alpha_1, \beta_2\}, \{\alpha_2, \beta_1\}$ the corresponding term in the sum may be shown to be $O(d^{-2})(L^d)^2(\mu_1-\mu_2)$ if $\mu_1 \neq \mu_2 \bmod(2\pi)$ and to converge to $2/3\,\|h\|_2^2$ if $\mu_1 = \mu_2 \bmod(2\pi)$. The same statements hold with μ_2 substituted by $-\mu_2$ for the partition $\mathcal{P} = \{\alpha_1, \beta_1\}, \{\alpha_2, \beta_2\}$. □

Lemma IV.4 Assume that (A), (B) and (C) hold and $d\,T^{-1}\ln(T) \to 0$. Then we have:

$$\operatorname{cum}\left[\sqrt{\frac{T}{d}}\, I_{d,T}^{(h)*}(\lambda_1), \ldots, \sqrt{\frac{T}{d}}\, I_{d,T}^{(h)*}(\lambda_r)\right] = O\left(\left[d T^{-1}\right]^{r/2-1} \ln(T)^{2r-1} \ln(d)\right). \quad \bullet$$

Proof. The cumulant in which we are interested equals

$$\left(\frac{T}{d}\right)^{r/2}\left(\frac{2\pi}{d}\right)^{r}\left(\prod_{j=1}^{r}\tilde{f}_d(\lambda_j)\right)\int \mathrm{cum}\left(I_T^{(h)}(\gamma_1),\ldots,I_T^{(h)}(\gamma_r)\right)\prod_{j=1}^{r}\left|K_d(\gamma_j,\lambda_j)\right|^2\prod_{j=1}^{r}d\gamma_j$$

Considering partitions of the table I.1 and adopting the notation used there, we have from Proposition I.6 b) and (T2) that:

$$\mathrm{cum}\left(I_T^{(h)}(\gamma_1),\ldots,I_T^{(h)}(\gamma_r)\right)=O(H_T^{-r})\sum_{ip,\,(r)}\int\prod_{j=1}^{r}L^T(\alpha_j+\gamma_j)\,L^T(\beta_j-\gamma_j)\prod_{i=1}^{S}d\tilde{\kappa}_i.$$

Now one may, in a natural way, consider the partitions $\mathcal{P}=\{P_1,\ldots,P_S\}$ of the table I.1 as partitions of the table:

$$\begin{array}{cc}\gamma_1 & -\gamma_1\\ \vdots & \vdots\\ \gamma_r & -\gamma_r\end{array}$$

Then writing $\sum_{P_k}\pm\gamma_j$ for the sum of all elements of P_k we obtain by Lemma A.1 that:

$$\mathrm{cum}\left(I_T^{(h)}(\gamma_1),\ldots,I_T^{(h)}(\gamma_r)\right)=O(H_T^{-r})\ln(T)^{2r-S}\sup_{\mathcal{P}}\prod_{i=1}^{S}L^T\left(\sum_{P_i}\pm\gamma_j\right).$$

Thus we obtain that the cumulant in which we are interested is bounded by:

$$O\left(H_T^{-r/2}\,d^{-3r/2}\ln(T)^{2r-S}\right)\sup_{\mathcal{P}}\int\prod_{i=1}^{S}L^T\left(\sum_{P_i}\pm\gamma_j\right)\prod_{j=1}^{r}L^d(\mu_j-\gamma_j)\,L^d(-\mu_j+\gamma_j)\prod_{j=1}^{r}d\gamma_j$$

Now we may proceed to integrate with respect to γ_j, $j=1,\ldots,r$ successively using Lemma A.1. In this process $L^T(0)=T$ will occur exactly once, since $\mathcal{P}$ is indecomposable. Further Lemma A.1 i) will be applied maximally S-1 times and accordingly Lemma A.1 ii) maximally 2r-2S+1 times. What will remain is a product of maximally S L^d factors, which may be bounded by d^S. Thus we obtain a total bound of

$$O\left(T^{-r/2}\,d^{-3r/2}\ln(T)^{2r-S}\,(d\ln(T))^{S-1}(\ln(d))^{2r-2S+1}\,d^S\,T\right)$$

$$=O\left(T^{1-r/2}\,d^{2S-1-3r/2}\ln(T)^{2r-1}(\ln(d))^{2r-2S+1}\right)$$

which yields the result since $S\leq r$. □

Lemma IV.5 Assume that (A), (B) and (C) hold, that $d^{1+\varepsilon}T^{-1}\to 0$ for some $\varepsilon>0$ and that $d^{-1/2}T^{1/2-\gamma}\to 0$, γ as in (T3). Then we have:

$$\sqrt{\frac{T}{d}}\left\|\hat{\Gamma}_{d,T}^{(h)*}-I\right\|\to 0.$$ •

Proof. Let $G_d^{(T,h)} = V_d^{(T,h)} \left(V_d^{(T,h)}\right)^t$ be the Cholesky decomposition of the matrix $G_d^{(T,h)}$, which was defined above. Further let $G_d^{(T,h)*} := U_d^{-1}\, G_d^{(T,h)} \left(U_d^t\right)^{-1}$. Then we have

$$\left\| G_d^{(T,h)*} - I \right\| \leq \left\| f^{(T,h)} f^{-1} - 1 \right\|_\infty = O(T^{-\gamma}),\ \gamma \text{ as in (T3)}$$

Set $A_{d,T}^{(h)} := \left(V_d^{(T,h)}\right)^{-1} \hat{\Gamma}_{d,T}^{(h)} \left[\left(V_d^{(T,h)}\right)^t\right]^{-1}$. We will show:

(1) $$\sqrt{\frac{T}{d}}\ \left\| A_{d,T}^{(h)} - I \right\| \to 0.$$

This is sufficient since:

(2) $$\left\| \hat{\Gamma}_{d,T}^{(h)*} - I \right\| \leq O(1) \left\| A_{d,T}^{(h)} - I \right\| + O(1) \left\| G_d^{(T,h)*} - I \right\|.$$

To see that (2) holds, note that

$$\hat{\Gamma}_{d,T}^{(h)*} - G_d^{(T,h)*} = U_d^{-1}\ V_d^{(T,h)} \left[A_{d,T}^{(h)} - I \right] \left(V_d^{(T,h)}\right)^t \left(U_d^t\right)^{-1}.$$

As in Chapter I, (1) follows from:

(3) $$E\, \mathrm{tr}\left[A_{d,T}^{(h)} - I \right]^k = O\left(\frac{d^{k/2+1}}{T^{k/2}} \ln(T)^{2k} \right) \quad \text{for } k \text{ even.}$$

In order to prove (3) observe that we have (indices always taken mod(k)):

$$\mathrm{tr}\left[A_{d,T}^{(h)} - I \right]^k = \int \prod_{i=1}^{k} \left[I_T^{(h)} - f^{(T,h)}\right](\mu_i)\ \bar{b}_{\mu_{i-1}}^t \left[G_d^{(T,h)}\right]^{-1} b_{\mu_i} \prod_{i=1}^{k} d\mu_i.$$

Next we calculate $E\left(\prod_{i=1}^{k} \left[I_T^{(h)} - f^{(T,h)}\right](\mu_i)\right)$. For this purpose consider all partitions $\mathcal{P}^{(k)} = \{P_1,\ldots,P_S\}$ (not only indecomposable), which do not contain any one-row or one-element partition subset of the $2\times k$ table:

$$\begin{matrix} \gamma_1 & \delta_1 \\ \vdots & \vdots \\ \gamma_k & \delta_k \end{matrix}\ .$$

Then we have, from Proposition I.6 c) and (T2), and adopting the notation used there:

$$E\left(\prod_{i=1}^{k} \left[I_T^{(h)} - f^{(T,h)}\right](\mu_i)\right) = O\left(H_T^{-k}\right) \sup_{\mathcal{P}^{(k)}} \int \prod_{i=1}^{k} L^T(\mu_i+\gamma_i)\, L^T(-\mu_i+\delta_i) \prod_{i=1}^{S} d\tilde{\kappa}_i,$$

where the supremum is taken over partitions $\mathcal{P}^{(k)}$, which do not contain subsets consisting of one element or one row of the table ($E\left[I_T^{(h)} - f^{(T,h)}\right] = 0$, $EX_t = 0$).

Now one may, in a natural way, consider the partitions $\mathcal{P}^{(k)}$ as partitions of the table:

$$\begin{matrix} \mu_1 & -\mu_1 \\ \vdots & \vdots \\ \mu_k & -\mu_k \end{matrix}$$

Then writing $\sum_{P_j} \pm\mu_i$ for the sum of all elements of P_j we obtain by applying Lemma A.1 i) repeatedly maximally 2k-S times:

$$E\left(\prod_{i=1}^{k} \left[I_T^{(h)}-f^{(T,h)}\right](\mu_i)\right) = O\left(H_T^{-k}\right) \ln(T)^{2k-S} \sup_{\mathcal{P}^{(k)}} \prod_{i=1}^{S} L^T\left(\sum_{P_i} \pm\mu_j\right).$$

Thus in total we get:

$$E\, tr\left[A_{d,T}^{(h)} - I \right]^k = O\left(H_T^{-k}\right) \ln(T)^{2k-S} \sup_{\mathcal{P}^{(k)}} \int \prod_{i=1}^{S} L^T\left(\sum_{P_i} \pm\mu_j\right) \prod_{j=1}^{k} L^d(\mu_j-\mu_{j-1}) \prod_{i=1}^{k} d\mu_i$$

Now suppose $\mathcal{P}^{(k)}$ consists of M row-subtables (a subtable is a union of rows which can be expressed as a union of partition subsets). Since $\mathcal{P}^{(k)}$ does not contain any one-row-partition subsets, it follows that $M \leq k/2$. For bounding the above expression Lemma A.1 ii) will be used maximally m-1 times for a row-subtable consisting of m rows, thus maximally k-M times in total. Further $L^T(0) = T$ will appear exactly M times and $L^d(0)=d$ exactly once. From these arguments we obtain:

$$E\, tr\left[A_{d,T}^{(h)} - I \right]^k = O\left(H_T^{-k} \ln(T)^{2k-S} \left(d \ln(T)\right)^{k-M} d\, T^M\right) = O\left(T^{-k/2}\, d^{k/2+1} \ln(T)^{2k}\right),$$

which proves (3) □

Proof of Theorem IV.2 From Lemmata I.4 and IV.5 we obtain that under the conditions of the theorem $\sqrt{T d^{-1}} \left| \hat{f}_{d,T}^{(h)} \tilde{f}_d^{-1}(\lambda) - I_{d,T}^{(h)*}(\lambda) \right| \to_P 0$. Further from Lemmata IV.3 and IV.4 we obtain that $\sqrt{T d^{-1}} \left\{ I_{d,T}^{(h)*} - B_d^{(T,h)} - 1 \right\}$ is asymptotically normal and has the desired covariance structure. This proves a). Assertion b) follows from a) with

$$\hat{f}_{d,T}^{(h)} f^{-1} - 1 - B_d - B_d^{(T,h)} = \tilde{f}_d f^{-1} \left[\hat{f}_{d,T}^{(h)} \tilde{f}_d^{-1} - 1 \right] - B_d^{(T,h)}$$

$$= \left[\hat{f}_{d,T}^{(h)} \tilde{f}_d^{-1} - 1 - B_d^{(T,h)} \right] \left(1 + O\left(d^{-[(r+\alpha)\wedge 1]} \ln^2(d) \right)\right)$$

Finally c) is an immediate consequence of b) with Lemma I.3 and (T3). □

V. A SIMULATION STUDY

V.1 INTRODUCTION

In previous chapters we studied the asymptotic properties of the Capon estimator and those of an automatic selection criterion for the parameter d.

But first it is not clear in advance that these asymptotics reflect well enough a finite sample situation. In time series analysis there are cases where the contrary is true. Thus, for example, (classical) asymptotic theory fails to describe the leakage effect (which is a finite sample effect) and to demonstrate the superiority of the tapered periodogram over the non-tapered: the bias of both estimators tends to zero while the variance of the tapered periodogram is larger than in the non-tapered case (see also Dahlhaus (1990)).

Moreover in the Introduction of this thesis we claimed that the Capon estimator copes with the leakage effect without using tapering. The confirmation of such a claim cannot be obtained by the (classical) asymptotic theory.

However a simulation experiment offers some indications to these and other similar problems. We are interested in studying the following points:

- How does the Capon estimator perform compared to other non-parametric estimators? Does it perform as well as an 'optimally' tapered smoothed periodogram? How does it perform as compared to its natural competitor, the autoregressive estimator?
- With what covariance matrix estimator does the Capon estimator perform better, with the 'segment', the 'symmetrized segment' or the 'Toeplitz'?
- Does the (quasi) bias correction proposed in Section II.1.3 bring significant benefits?
- How do the several order dimension selection criteria perform? Is the inclusion of quadratic terms really an improvement? Does the property of asymptotic efficiency really reflect the finite sample behaviour?

We will present the results of our simulation concerning these questions in separate sections of this chapter. As far as the two last questions are concerned, we have already presented some results in previous chapters (see Sections II.1.3 and II.1.4). A very brief summary of the simulation results can be found in the Section ***'Conclusions of the simulation study'*** of this chapter. Before describing the model we used for the simulation in the Section ***'The model (Model A)'*** and the different estimators we compared in the Section ***'The estimators'***, we will present the method we followed to measure the performance of an estimator.

V.1.1 Method of comparison

To measure the performance of an estimator $\hat{f}_T$ we used the Whittle discrepancy defined in Section II.1.1. More precisely we approximated this discrepancy by a Riemann sum over the Fourier frequencies. Thus we used:

$$\Delta'(f,\hat{f}_T) := T^{-1} \sum_{j=1}^{T} \left[\ln\left(f^{-1}\hat{f}_T\right)(\lambda_j) + f\,\hat{f}_T^{-1}(\lambda_j) - 1\right], \quad \text{where } \lambda_j := \frac{2\pi\,(j-1)}{T}$$

and T is the sample size, which was chosen T=256.

In order to summarize the information obtained, we present for each estimator the box plot of $\Delta'(f,\hat{f}_T)$ over the 500 simulated samples. Obviously the more the box plot concentrates close to 0 the better the estimator is. One can look at the median and at the inter-quartile range to obtain an impression of this concentration.

In order to help judge the difference between two estimators in this visual comparison, we also show the mean and the 95% and 99.9% confidence intervals for the mean in each box plot. This suggests implicitly a test to compare two estimators: an estimator is 'significantly better' than another if the confidence interval for the mean of the error of the first lies at a smaller location than the one of the second without intersecting it. We underline that this 'implicit test' is very conservative, since it does not take into account the strong dependence between the two samples: for example the 'pair-wise t-test', that is the t-test on the sample $X_i - Y_i$, $i=1,\ldots,n$, would often recognize as significant differences which are not recognized as such by the above test.

There is a second important point concerning the methodology of the comparison: all estimators depend on additional parameters such as the dimension d for the Capon estimator, the order of an autoregressive estimator, the bandwidth for a smoothed periodogram and the percentage of tapering for periodograms and Toeplitz covariance matrix estimator. The problem is how to choose the parameter value for the presentation of the results.

For some of these parameters exist automatic selection procedures in the performance of which we are interested (the first two examples above). In this case it is natural to use them for selecting the parameter value of the estimator. In other cases exist automatic criteria but we are not particularly interested in their performance in this context (e.g. 'Plug in' for the bandwidth). Moreover there are cases where automatic criteria do not exist (or they are not known to us), as for example for the percentage of tapering. In these latter cases, we simulated the estimator over a grid of values for the parameters and we present the estimator with the parameter value which empirically performed best. Thus, when we present smoothed periodograms 30% tapered and with bandwidth 0.025, we implicitly mean that these values for the taper and the bandwidth empirically minimized the mean of the error in our simulation.

Of course this is equivalent to giving 'prior information' to these estimators. One should keep that in mind when comparing their performance with estimators for which no 'prior information' is used. For this reason we also present errors of estimators with an 'optimally chosen parameter' even when automatic selection criteria for this parameter exist and are studied here. This also helps to judge, for example, if a bad performance of an estimator is due to the estimator itself or to the method used to select its parameter values.

V.1.2 The model (Model A)

For the simulation we used a Gaussian ARMA(p,q) model. A stationary process $\{X_t\}_{t\in Z}$ is said to be a Gaussian ARMA(p,q) process if there exist coefficients α_j, j=1,...,p and β_j, j=1,...,q and a sequence of iid normal variables $\{\varepsilon_t\}_{t\in Z} \sim N(0, \sigma)$ such that the following equation is fulfilled:

$$X_t - \alpha_1 X_{t-1} - \ldots - \alpha_p X_{t-p} = \varepsilon_t + \beta_1 \varepsilon_{t-1} + \ldots + \beta_q \varepsilon_{t-q},\ t\in \mathbf{Z}.$$

It is well known that if the polynomials $\phi(z) := z^p - \alpha_1 z^{p-1} - \ldots - \alpha_p$ and $\theta(z) := z^p + \beta_1 z^{p-1} + \ldots + \beta_q$ have roots outside the unit circle and no roots in common, then the above equation has stationary solutions which are causal and invertible (Brockwell, Davis (1987)). The spectral density of this process is given by :

$$f(\lambda) := \frac{\sigma^2}{2\pi} \frac{|\theta(z)|^2}{|\phi(z)|^2},\quad \lambda\in[-\pi, \pi],\ \text{where } z=e^{i\lambda}$$

Writing z_j^{-1}, j= 1,...,p for the roots of the 'autoregressive polynomial' $|\phi(z)|^2$ and w_j^{-1}, j= 1,...,q for the roots of the 'moving average' polynomial $|\theta(z)|^2$ f can be written as:

$$f(\lambda) := \frac{\sigma^2}{2\pi} \left\{\prod_{j=1}^{q} [1 - z\, w_j]\right\} \left\{\prod_{j=1}^{p} [1 - z\, z_j]\right\}^{-1},\quad \lambda\in[-\pi, \pi],\ \text{where } z=e^{i\lambda}$$

The simulation consists of 500 samples, with a sample size of T=256 drawn from an ARMA(12,4) model with Gaussian innovations of variance 1. The spectral density was chosen to contain strong peaks and gaps, which are typical structures of ARMA models, but also to contain a flat peak, which is more a-typical for ARMA models. We made this choice since our estimators are to be studied as nonparametric estimators, so the underlying model should not grand special favors to any type of estimator used in the simulation. Clearly autoregressive estimators would perform better than any others, if the true model was purely autoregressive.

Below we define the model in terms of the z_j, j= 1,...,p (AR) and w_j, j= 1,...,q (MA) (omitting complex conjugate roots) and show a plot of the true spectrum, together with a realisation of a tapered and a non-tapered periodogram, the Capon estimator and the autoregressive estimator in Fig V.1. The roots were chosen by mouse-clicks, which explains the number of decimal places.

(radius , angle/π , order)

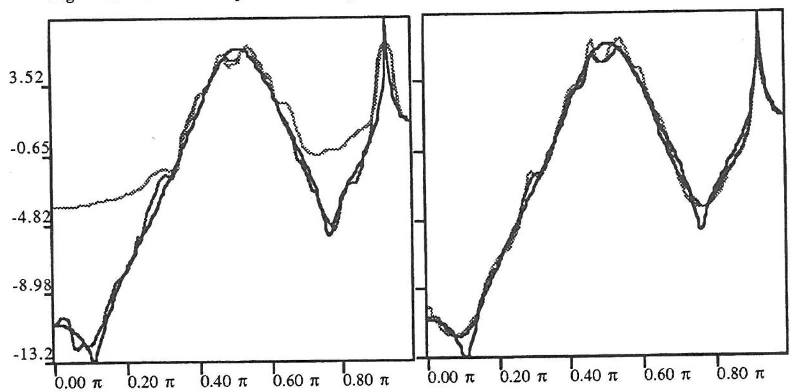
<u>Model A</u> : MA (w_j) (0.9579,0.1094,1), (0.9639,0.7656,1).

AR (z_j) (0.995, 0.9297, 1), (0.7279, 0.5391, 1),

(0.7242, 0.5234, 1), (0.7224, 0.5078, 1),

(0.7166, 0.4922, 1), (0.7203, 0.4688, 1).

Fig. V.1 The true spectral density and a realization of some specctral etimators.



y-Axis scaled logarithmically, sample size=256

—— True spectrum
········ Smoothed Periodogram 0% tapered Bw=0.02
—— Smoothed Periodogram 30% tapered Bw=0.025
—— Capon (Segm.) Dim= 49
········ Autoreg.- LS order= 26

In this figure one sees clearly how the non-tapered periodogram is affected by leakage. A second interesting point is the difference in the behaviour of the estimators close to the main peak: the periodograms seem to 'come from above' whereas the other two estimators to 'come from below' and estimate it much better. One can observe an 'inverse' behavior close to the gap at 0.76 π.

V.1.3 The estimators

Most of the estimators presented in the simulation such as the Capon estimator, the autoregressive least squares estimator and the smoothed periodograms have already been defined in previous chapters. We define here two further estimators: the 'autoregressive least squares forward-backward estimator' and the 'autoregressive (tapered) Toeplitz estimator'.

The autoregressive least squares forward-backward estimator $\hat{f}_{p,T}^{(ARfb)}$ of order p is defined as follows (Kay and Marple (1981)): let $\mathbf{a}_p^t := (1, a_{p,1}, \ldots, a_{p,p})$ and define

- the 'forward residuals' $e_{p,t}(\mathbf{a}) := X_t + a_{p,1}X_{t-1} + \ldots + a_{p,p}X_{t-p}$ $t = p+1,\ldots,T$ and
- the 'backward residuals' $b_{p,t}(\mathbf{a}) := X_t + a_{p,1}X_{t+1} + \ldots + a_{p,p}X_{t+p}$ $t = 1,\ldots,T-p$.

Further define $\hat{\sigma}_{p,T}^2 := \inf_{\mathbf{a}} \sum_{t=p+1}^{T} e_{p,t}^2(\mathbf{a}) + \sum_{t=1}^{T-p} b_{p,t}^2(\mathbf{a})$ the infimum being attained at $\hat{\mathbf{a}}_{p,T}$.

Then

$$\hat{f}_{p,T}^{(ARfb)}(\lambda) := \hat{\sigma}_{p,T}^2 / 2\pi \left| \hat{\mathbf{a}}_{p,T}^t b_\lambda \right|^{-2}.$$

The autoregressive Toeplitz estimator $\hat{f}_{p,T,\alpha}^{(ARt)}$ of order p and α-% tapered is defined as follows: let the Toeplitz covariance matrix estimator $\hat{\Gamma}_{d,T}^{(h)}$ be defined as in Section IV.1.2, where $h \equiv h_\alpha$ is chosen to be the α-% Tukey data taper (see 0.3). Let $\mathbf{a}_p^t := (1, a_{p,1}, \ldots, a_{p,p})$ and define $\hat{\mathbf{a}}_{p,T}$, $\hat{\sigma}_{p,T}^2$ as the solution of the Yule-Walker equations: $\hat{\Gamma}_{p+1,T}^{(h)} \; \hat{\mathbf{a}}_{p,T} = \left(\hat{\sigma}_{p,T}^2, 0, \ldots, 0 \right)^t$. Then $\hat{f}_{p,T,\alpha}^{(ARt)}(\lambda) := \hat{\sigma}_{p,T}^2 / 2\pi \left| \hat{\mathbf{a}}_{p,T}^t b_\lambda \right|^{-2}$.

In the simulation study we included the following estimators.

a) Capon estimators.

- with 'segment', 'symmetrized segment' and 'Toeplitz' covariance matrix estimators, defined in Sections 0.2, IV.1.1, IV.1.2 respectively.
- corrected and non-corrected. By this we refer to the (quasi) bias correction discussed in Section II.1.3. There it was only considered in the case of a 'segment' covariance matrix estimator. It consists of the multiplication of the estimator with $1 + d/N + (d/N)^2$, N as in (0.5). As simulations similar to those presented in Figure II.1 suggest, we use in the 'symmetrized segment ' case the same correction factor and in the 'Toeplitz ' case the factor $1 + d/H_T$ (H_T as in Section 0.3) (see also Figure II.4).
- with orders fixed and estimated. By 'estimated orders' we mean that they were selected according to the dimension selection criterion proposed in Section II.1.2 (and Section II.1.3 for the corrected estimator). For the non-corrected estimator we considered both a penalty term including and not including the quadratic term $(d/N)^2$ in order to observe the difference in their performance. For the 'symmetrized segment' estimator we used the same penalties as above and for the 'Toeplitz' estimator we took d/H_T and d/T, for the same reasons as for the correction factors.

b) Autoregressive estimators.

We underline that autoregressive estimators may have been introduced as parametric estimators for an autoregressive model, but they may also be regarded as non-parametric estimators.

• defined over 'least squares' (Section II.1.4), 'least squares forward-backward' and 'Toeplitz Yule-Walker' equations (both defined in this section).
• corrected and non-corrected. By this we mean again the (quasi) bias correction discussed in Section II.1.4. There, it was only considered in the case of the least squares estimator. It consists of the multiplication of the estimator with $1 + 2\ p/N + 4\ (p/N)^2$, N as in (0.5). For the same reasons as for the Capon estimator we use the same correction factor for the 'forward-backward' estimator; for the 'Toeplitz ' case we use $1 + 2\ d/H_T$ (H_T as in Section 0.3) (see also Figure II.4).
• with orders fixed and estimated. For estimating the orders we used the 'order selection criterion' proposed in Section II.1.4. For the non-corrected estimator we considered various forms of the penalty in order to observe the difference in their performance: 2p/T (**AIC**), 2p/N (almost equal to **AICc**), $2p/N + 4\ (p/N)^2$ (the criterion proposed in Section II.1.4) and also ln(T)p/T (**BIC**). For the 'symmetrized segment ' estimator we took 2p/N and $2p/N+4(p/N)^2$ and for the 'Toeplitz' estimator we took $2\ d/H_T$ and 2 d/T (see Figure II.4).

c) Kernel smoothed tapered periodograms: $I_T^{(h)} * K_b$ where $I_T^{(h)}$ is defined in (0.7) and K_b the Barlett-Priestley kernel (see 0.4) with bandwidth b.

Generally we used the following rules in the abbreviations for the cases a) and b).

C : Capon (segment Cov.Mat.Est.).	**AR**: autoregressive (least squares)
-s- : symmetrized segment Cov.Mat.Est.	**-fb-**: least squares forward-backward
-t- : Toeplitz Cov.Mat.Est. (20% taper)	**-t-**: Toeplitz Yule-Walker (20% taper)

-c-: (quasi) bias correction by multiplication with an appropriate factor.
-d^(N)80- estimated order with penalty including only linear terms, maximum=80.
-d^(T)80- the same but with 'T' instead of 'N', maximum=80.
-d^(T; ln)80- the same (only for AR) but with penalty ln(T) p/T (BIC), maximum=80.
-d^(NN)80- estimated order with penalty including also quadratic terms, maximum=80.
-cd^80- estimated order for (quasi) bias corrected estimator with penalty = 0, maximum=80.

The estimators and their abbreviations

(For Capon or autoregressive estimators $\hat{f}_{d,T}$ orders are estimated by $\text{argmin}_{d \leq 80}\ (2\pi)^{-1} \int \ln(\hat{f}_{d,T})$ + penalty (d) with different penalties, given below. We use N as in (0.5) and H= H_T as in 0.3).

P/ 30%/ Bw=0.02	$I_T^{(h)} * K_b$ where $I_T^{(h)}$ the 30%-tapered periodogram, defined in Section 0.3, relation (0.7) and K_b the Barlett-Priestley kernel (see Section 0.4) withbandwidth b=0.02.
C49	'segment Capon', defined in (0.4) and (0.5), with d=49.
Cd^(N)80	as above but with estimated d: penalty = d/N.
Cd^(NN)80	as above but with penalty= $d/N + (d/N)^2$.
Cc58	'segment Capon' ((quasi) bias corrected), defined in (0.4) and (0.5),with d=58, multiplied by $1+ d/N+ (d/N)^2$.
Ccd^80	as above but with estimated d: penalty= 0.
Cs51	'symmetrized segment Capon', defined in IV.1.1, with d=51.
Csd^(N)80	as above but with estimated d: penalty= d/N.
Csd^(NN)80	as above but with penalty= $d/N + (d/N)^2$.
Csc61	'symmetrized segment Capon' ((quasi) bias corrected), defined in IV.1.1, d=61, multiplied by $1+ d/N+ (d/N)^2$.
Cscd^80	as above but with estimated d: penalty= 0.
Ct67	'Toeplitz Capon', defined in Section IV.1.2, with d=67 and an 20% Tukey data taper (see Section 0.3)
Ctd^(H)80	as above but with estimated d: penalty= d/H.
Ctd^(T)80	as above but with penalty= d/T .
Ctc81	'Toeplitz Capon' ((quasi) bias corrected), defined in IV.1.2, with d=81 and 20% taper, multiplied by 1+ d/H.
Ctcd^80	as above but with estimated d: penalty= 0.
AR26	autoregressive least squares, defined in II.1.4, with p=26.
ARp^(T)80	as above but with estimated p: penalty = 2 p/T (AIC).
ARp^(N)80	as above but with penalty = 2 p/N.
ARp^(NN)80	as above but with penalty= $2 p/N + 4 (p/N)^2$.
ARp^(ln,T)80	as above but with penalty= ln(T) p/T **(BIC)**.
ARc26	autoregressive least squares ((quasi) bias corrected), defined in Section II.1.4, p=26, multiplied by $1+ 2p/N+ 4(p/N)^2$.
ARcp^80	as above but with estimated p: penalty= 0.
ARfb26	autoregressive least squares forward-backward, defined in Section V.1.3, with p=26.
ARfbp^(T)80	as above but with estimated p: penalty= 2 p/T (AIC).
ARfbp^(N)80	as above but with penalty= 2 p/N.
ARfbp^(NN)80	as above but with penalty= $2 p/N + 4 (p/N)^2$.
ARfbc26	autoregressive least squares forward-backward ((quasi) bias corrected), defined in Section V.1.3, p=26, multiplied by

	$1+2p/N+4(p/N)^2$.
ARfbcp^80	as above but with estimated p: penalty= 0.
ARt26	Toeplitz autoregressive, defined in Section V.1.2, with p=26, 20% tapered.
ARtp^(H)80	as above but with estimated p: penalty= 2 p/H.
ARtp^(T)80	as above but with penalty= 2 p/T (AIC).
ARtc26	Toeplitz autoregressive ((quasi) bias corrected), defined in Section V.1.2, with p=26, 20% taper, multiplied by 1+ 2 d/H.
ARtcp^80	as above but with estimated p: penalty= 0.

V.2 SIMULATION RESULTS

We will now present the simulation results following the methodology discussed in Section V.1.1. In each of the following sections we present a set of box plots concerning some of the questions posed at the beginning of this chapter.

V.2.1 Comparison of 'segment' Capon and LS AR with tapered estimators

In the following image we present three groups of estimators:

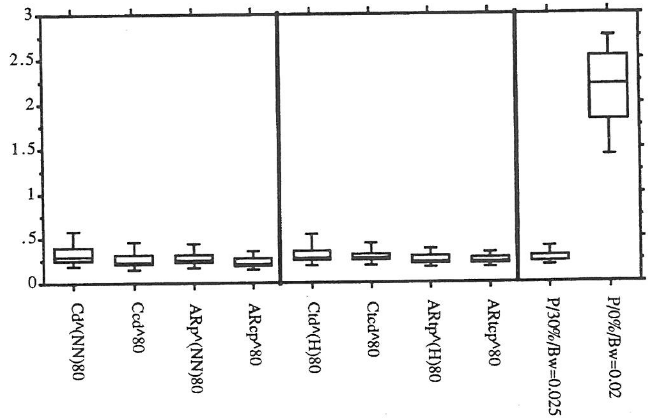
- 'Segment' Capon and autoregressive least squares, corrected and non-corrected.
- optimally tapered 'Toeplitz' Capon and autoregressive estimators.
- Smoothed periodograms with optimal bandwidth, one optimally tapered and one non- tapered.

In the first group we used estimated orders in order to stress the fact that these estimators do not have any prior information at all. In the second group we did the same so that it is comparable with the first group.

What can be observed is: first, that the non-tapered periodogram is much worse than the others; non-tapered 'Toeplitz' Capon and autoregressive estimators would exhibit a similar behaviour. Secondly, the 'segment' Capon and autoregressive estimators have discrepancy distributions very close to that of optimally tapered estimators. The finer differences which can already be observed among them will be further analysed in subsequent sections.

Fig. V.2 Box-plots of discrepancies for different types of estimators (500 samples under Model A).

Capon and AR- estimators with no prior information. | Toeplitz Capon and AR- estimators optimally tapered (20%). | Smoothed periodograms.



C: Segment Capon. Ct: Toeplitz (20% tapered).
'-c-: corrected (multiplication with (1+ d/H) for Ct and with (1+ d/N + (d/N)2) for C.
'-d^(NN)-', '-d^(H)-' : estimated d, penalty- term= d/N + (d/N)2, d/H. For Ccd^, Ctcd^: no penalty.
AR: LS Autoreg. ARt: Toeplitz (20% tapered).
'-c-': corrected (multiplication with (1+ 2p/H) for ARtp and with (1+ 2p/N + 4(p/N)2) for ARp.
'-p^(H)- ', '-p^(NN)-' : estimated p, penalty- term= 2p/H, 2p/N + 4(p/N)2. For ARcp^, ARtcp^ : no penalty.
P/t%/Bw=b: Smoothed periodogram, t% tapered, kernel bandwidth=b.

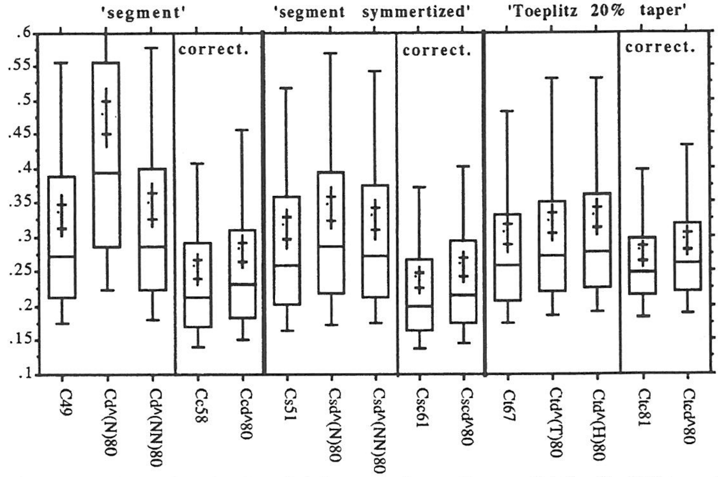
V.2.2 Comparison of different types of Capon estimators

In the following image we present three groups of estimators:

- 'Segment' Capon.
- 'Symmetrized segment' Capon.
- optimally tapered (20%) 'Toeplitz' Capon.

In each group there are corrected and non-corrected estimators and also estimators with fixed (empirically optimal) and estimated orders (by different selection criteria).

Fig. V.3 Box-plots and estimated means with 95% and 99.9% confidence intervalls of discrepancies for different types of Capon-estimators (500 samples under Model A).



Cd: Segment Capon (dimension d). Csd: Symmetrized-segm. Capon. Ctd: Toeplitz (20% tapered).
'-c-: corrected (multiplication with (1+d/H) for Ctd and with (1+ d/N + (d/N)2) for Cd and Csd.
'-d^(N)-, -d^(T)-, -d^(H)- : estimated d, penalty- term= d/N, d/T, d/H.
'-d^(NN)- : estimated d, penalty- term= d/N + (d/N)2. For Ccd^ ,Cscd^, Ctcd^ : no penalty.

The following points can be observed:

- Symmetrization does not seem to bring significant benefits to the 'segment' estimator: the discrepancy distributions for Cs51 and C49 hardly differ.
- The optimally tapered Toeplitz Capon seems to perform slightly better than the 'segment' Capon: only 62% of the mass of C49 lies below the 75% quantile of Ct67. But the corrected 'segment' Capon Cc58 seems to be better than both the corrected and the non-corrected optimally tapered 'Toeplitz' estimators. All its quartiles lie strongly below those of Ct67 and Ctc81.
- As mentioned above, correction brings significant benefits to the 'segment' and 'symmetrized segment' Capon. The benefits are less important for the Toeplitz Capon but one has still the impression that the inter-quartile range of Ctc81 is narrower than the one of Ct67.
- For the first two groups dimension selection criteria including quadratic terms (for corrected estimators these are implicitly present in the correction) work well. Dimension selection criteria including only linear terms perform worse.
- This is not the case for the 'Toeplitz' estimators where the linear approximation is good enough (see also Figure II.4).

V.2.3 Comparison of different types of non-parametric estimators

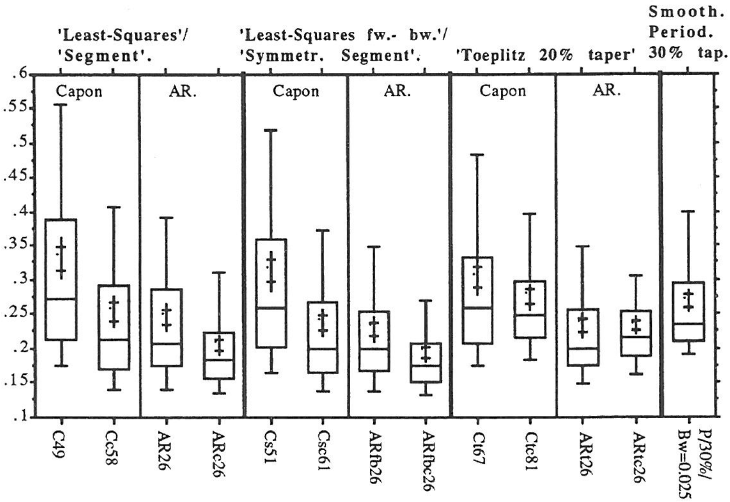
In the following image we present four groups of estimators:

- 'Segment' Capon and 'least squares' autoregressive.
- 'Symmetrized segment' Capon and 'least squares forward-backward' autoregressive.
- optimally tapered (20%) 'Toeplitz' Capon and 'Toeplitz' autoregressive
- an optimally tapered and smoothed periodogram.

In each group there are corrected and non-corrected estimators, but only estimators with fixed (empirically optimal) orders.

Fig V.4.

Box-plots and estimated means with 95% and 99.9% confidence intervalls of discrepancies for different types of Capon- and Autoregressive estimators with 'optimal parameters' (500 samples under Model A).



Cd: Segment Capon (dimension d). Csd: Symmetrized-segm. Capon. Ctd: Toeplitz (20% tapered).
'-c-': corrected (multiplication with (1+d/H) for Ctd and with (1+ d/N + (d/N)2) for Cd and Csd.
ARp: LS Autoreg. (order p). ARfbp: LS-forw-backw. Autor. ARtp: Toeplitz (20% tapered).
'-c-': corrected (multiplication with (1+ 2p/H) for ARtp and (1+ 2p/N + 4(p/N)2) for ARp and ARfbp.
P/30%/Bw...: Smoothed periodogram, 30% tapered, kernel bandwidth= 0.025.

Beyond the points already described in the previous section and which can be seen to hold here also for autoregressive estimators, one can observe in Figure V.4 the following:

- Capon estimators are **always worse** than the corresponding autoregressive ones:

for all types of estimation of the underlying covariance matrix, for both corrected and non-corrected estimators, **all** quartiles of the discrepancy distribution of the Capon estimators **always** lie above those of the corresponding autoregressive estimators.

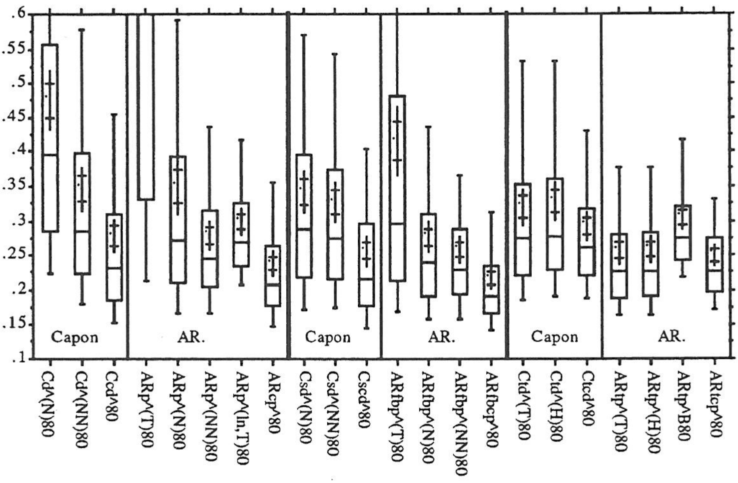
- The optimally tapered periodogram seems to perform better than the 'segment' Capon: only 58% of the mass of C49 lies below the 75% quartile of P/30%/Bw=0.025. But the corrected 'segment' Capon Cc58 seems to be better than the (optimally tapered and smoothed) periodogram. All its quartiles lie below those of P/30%/Bw=0.025.
- 'Least squares' autoregressive estimators corrected and non-corrected also seem to perform better than the (optimally tapered and smoothed) periodogram.

In the image which follows we present the same as above but with estimated orders. Beyond the optimal taper for the third group there is no 'prior information' entering. This is the reason we present it separately.

Fig V.5.

Box-plots and estimated means with 95% and 99.9% confidence intervalls of discrepancies for different types of Capon- and Autoregressive estimators with 'estimated orders' (500 samples under Model A).

'Least-Squares' 'Segment'.   **'Least-Squares fw.- bw.'/ 'Symmetr. Segment'.**   **'Toeplitz 20% taper'**



C: Segment Capon.   Cs: Symmetrized-segm. Capon.   Ct: Toeplitz (20% tapered).
'-c-: corrected (multiplication with (1+ d/H) for Ct and with (1+ d/N + (d/N)2) for C and Cs.
'-d^(N)-', '-d^(NN)-', '-d^(T)-', '-d^(H)-' : estimated d, penalty- term= d/N, d/N + (d/N)2, d/T, d/H.
AR: LS Autoreg.   ARfb: LS-forw-backw. Autor.   ARt: Toeplitz (20% tapered).
'-c-': corrected (multiplication with (1+ 2p/H) for ARtp and with (1+ 2p/N + 4(p/N)2) for ARp and ARfbp.
'-p^(N)-', ' -p^(T)-', '-p^(ln,T)-', '-p^(H)- ', '-p^(NN)-' : estimated p, penalty- term=
2p/N (AICc), 2p/T (AIC), ln(T)p/T (BIC), 2p/H, 2p/N + 4(p/N)2.
For Ccd^ ,Cscd^, Ctcd^, ARcp^ ,ARfbcp^, ARtcp^ : no penalty.

One can verify the observations already mentioned for the first image (with fixed orders). This is natural since 'estimated orders' approximate the behavior of fixed 'optimal' orders. One observes again rather clearly the superiority of corrected estimators as well as the superiority of order selection criteria including quadratic terms when 'segment Cov.Mat.Est.' are used. What is also very interesting is the difference of the performance of the simple AIC:

• the AIC (penalty 2p/T) for autoregressive least squares estimators performs **very badly**; not even the median of ARp^(T)80 lies in the image. When symmetrization (forward-backward) is used, AIC performs a bit better than previously, but still much worse than the other selection criteria presented. On the contrary AIC performs as **well** as other selection criteria in connection with a 'Toeplitz' autoregressive estimator. This difference is explained in Section II.1.4. It is due to the fact that the approximation $(2\pi)^{-1} \int f \hat{f}_{p,T}^{-1}(\lambda)\, d\lambda \cong 2p/T$ is 'good enough' when $\hat{f}_{p,T}$ is the 'Toeplitz' autoregressive estimator but very bad when $\hat{f}_{p,T}$ is the autoregressive least squares estimator.

V.2.4 Conclusions of the simulation study

In this section we summarize the points observed in the simulation study.

- Capon estimators with a 'segment' or a 'symmetrized segment' Cov.Mat.est. and autoregressive LS or LS forward-backward estimators perform as well as optimally tapered smoothed periodograms or optimally tapered 'Toeplitz' Capon and autoregressive estimators.
- (Quasi) bias correction brings important benefits to the 'segment', 'symmetrized segment' Capon and autoregressive LS or LS forward-backward estimators. The benefits are less important for the Toeplitz Capon and autoregressive estimators.
- For the 'segment', 'symmetrized segment' Capon and for the autoregressive LS or LS forward-backward estimators dimension selection criteria including quadratic terms (for corrected estimators these are implicitly present in the correction) work well: the resulting discrepancy distributions can be hardly distinguished from those with an optimal parameter choice (asymptotic efficiency). Dimension selection criteria including only linear terms perform less well. This is not the case for the 'Toeplitz' estimators where the criteria including only linear terms are good enough.

A. APPENDIX

A.1 The L^N functions

We introduce the $L^N(\lambda)$ functions (Dahlhaus (1983)) and state some of their properties which we use for the cumulant calculations.

Let $L^N(\lambda)$ be a 2π periodic and symmetric around 0 function, defined on $[-\pi,\pi]$ as follows:

$$L^N(\lambda) = \begin{cases} N \text{ if } |\lambda| \leq N^{-1} \\ |\lambda|^{-1}, \text{ else} \end{cases}$$

Then the following holds:

Lemma A.1 There exists $K \in \mathbf{R}$, such that for $N,d,c \in \mathbf{N}^+$, $\alpha,\beta,\gamma,\lambda,\mu,x \in \mathbf{R}$:

i) $\int L^N(\gamma+\alpha)\, L^N(\beta-\alpha)\, d\alpha \leq K \ln(N)\, L^N(\gamma+\beta)$.

ii) If $c / d \leq C$ for some C holds, then there exists $K' \in \mathbf{R}$, such that

$$\int L^d(\gamma+x)\, L^d(\beta-x)\, L^N(c\lambda+cx)\, L^N(c\mu-cx)\, dx \leq K'\, d\, c^{-1} \ln(N) \ln(c)\, L^d(\gamma+\beta)\, L^N(c\lambda+c\mu) \bullet$$

Proof. i) Dahlhaus (1983)
ii) Follows easily from the proof of Dahlhaus (1985), Lemma 4.1.

A.2 Orthogonal polynomials

In this chapter we study some properties of quantities related to the system $\{\phi_k(\lambda)\}_{k\in N}$ of polynomials orthogonal with respect to a spectral density f, where f fulfills (A) and (B) (see also Szegö (1959)). We study especially bounds and approximations of kernels of the type $K_d(\lambda,\mu) = K_d(f,\lambda,\mu)$ defined in Section I.2.3. Let also Γ_d be defined as in Chapter 0.

Let $\{\phi_k(\lambda)\}_{k\in N_0^+}$ denote the system of polynomials orthogonal with respect to f, that is ϕ_k is a polynomial of degree k in $e^{i\lambda}$ and $(2\pi)^{-1}\int f(\lambda)\phi_i(\lambda)\overline{\phi_j(\lambda)} = \delta_{ij}$. Then it may be easily seen that

(A.2.1) $$\phi_{d-1}(\lambda) = \sqrt{2\pi}\ \ e_d^t U_d^{-1} b_\lambda, \quad e_d^t := (0,\dots,0,1) \in \mathbf{R}^d$$

where $\Gamma_d = U_d U_d^t$ is the Cholesky decomposition of Γ_d (U_d is a lower triangular d×d matrix).

(A.2.2) $$K_d(\lambda, \mu) := \overline{b_\lambda^t}\, \Gamma_d^{-1}\, b_\mu = (2\pi)^{-1} \sum_{i=0}^{d-1} \overline{\phi_i(\lambda)}\phi_i(\mu)$$

It is further well known that under (A) (Szegö (1959), Theorem 12.1.3.)

(A.2.3) $$|\phi_d(\lambda)|^2 \to f^{-1}(\lambda) \text{ uniformly in } \lambda \text{ for } d \to \infty.$$

Moreover if $f = |h_p|^{-2}$, where h_p is a polynomial of degree p in $e^{i\lambda}$ (f autoregressive) then

(A.2.4) $$\text{for } j \geq p \quad \phi_j(\lambda) = \overline{h_p(\lambda)}\, e^{i\lambda j}$$

Finally the Christoffel-Darboux (Szegö (1959), Theorem 11.4.2.) formula holds:

(A.2.5) $$K_d(\lambda, \mu) = (2\pi)^{-1} \frac{e^{i(d)(\mu-\lambda)}\, \overline{\phi_d(\lambda)}\, \phi_d(\mu) - \phi_d(\lambda)\, \overline{\phi_d(\mu)}}{1 - e^{i(\mu-\lambda)}}$$

Proof of (A.2.1) and (A.2.2). Let $\underline{\phi}_d(\lambda) \in \mathbf{C}^d$, $\underline{\phi}_d(\lambda) := (\phi_0(\lambda),\dots,\phi_{d-1}(\lambda))^t$. We first show

(1) $$\underline{\phi}_d(\lambda) = \sqrt{2\pi}\ U_d^{-1} b_\lambda.$$

If (1) is true, (A.2.1) follows directly. Furthermore (A.2.2) also follows from (1):

$$\overline{b_\lambda^t}\, \Gamma_d^{-1}\, b_\mu = \left(U_d^{-1}\, \overline{b_\lambda}\right)^t U_d^{-1} b_\mu = (2\pi)^{-1}\, \overline{\underline{\phi}_d^t(\lambda)}\, \underline{\phi}_d(\mu) = (2\pi)^{-1} \sum_{i=0}^{d-1} \overline{\phi_i(\lambda)}\phi_i(\mu).$$

In order to check the validity of (1) it is sufficient to prove that:

i) the j-th component of $\sqrt{2\pi}\ U_d^{-1} b_\lambda$ is a polynomial of degree j-1 in $e^{i\lambda}$, which is the case.

ii) the polynomials given by the components of $\sqrt{2\pi}\ U_d^{-1} b_\lambda$ fulfill the orthogonality property, which follows from

$$\int f(\lambda)\, (U_d^{-1} b_\lambda)_j\, \overline{(U_d^{-1} b_\lambda)}_k\, d\lambda = \left[U_d^{-1} \int f(\lambda)\, b_\lambda\, \overline{b_\lambda^t}\, d\lambda \left(U_d^{-1}\right)^t\right]_{j,k} = \left[U_d^{-1}\, \Gamma_d \left(U_d^{-1}\right)^t\right]_{j,k} = \delta_{j,k}$$

iii) the first d components of $U_{d+1}^{-1} b_\lambda$ are identical to $U_d^{-1} b_\lambda$ which is equivalent to

(2) $$\exists\ \underline{v} \in \mathbf{R}^d, b \in \mathbf{R}: \quad U_{d+1}^{-1} = \begin{pmatrix} U_d^{-1} & 0 \\ \underline{v}^t & b \end{pmatrix}.$$

In order to prove (2) we first prove:

(3) $$\exists\ \underline{u}\in \mathbf{R}^d, a\in \mathbf{R}:\quad U_{d+1} = \begin{matrix} U_d & 0 \\ \underline{u}^t & a \end{matrix}.$$

Define the vector $\underline{c} := (c_d,\ldots,c_1)^t$ and the matrix U° as on the right hand of (3) by setting $\underline{u} := (U_d)^{-1}\,\underline{c}$ and $a := \sqrt{c_0 - \underline{u}^t\underline{u}}$. It is easily verified that the thus defined matrix U° has the property that $U^\circ(U^\circ)^t = \Gamma_{d+1}$ thus $U_{d+1} = U^\circ$.

It still remains to prove $c_0 - \underline{u}^t\underline{u} \geq 0$. This follows from the next argument:
Set $\underline{x}^t := \left(\underline{c}^t\Gamma_d^{-1}, -1\right)$. Then, since Γ_{d+1} is positive definite we have $\underline{x}^t\,\Gamma_{d+1}\underline{x} = c_0 - \underline{u}^t\underline{u} \geq 0$.

We now turn to the proof of (2). Define the matrix $\overset{\circ}{V}_{d+1}$ as the right hand side of (2) with $b := a^{-1}$ and $\underline{v} := -b\left(U_d^t\right)^{-1}\underline{u}$, with a and $\underline{u}$ as above. Then it is easily verified that $\overset{\circ}{V}_{d+1} = U_{d+1}^{-1}$. This proves (2). □

In the following lemma we bound the $K_d\left(f,\lambda,\mu\right)$-kernels by an L_d function:

> **Lemma A.2** Let f fulfill (A). Then there exists a constant M', which depends only on m,M (from (A)), such that
>
> $$K_d\left(f,\lambda,\mu\right) \leq M'\, L^d(\lambda-\mu). \qquad \bullet$$

Proof. It is easy to see that for any $\underline{x}\in\mathbf{R}^d$ with $||\underline{x}||_2 = 1$ we have $\underline{x}^t\,\Gamma_d\,\underline{x} \geq 2\pi$ m.
It follows $\left|\left|\Gamma_d^{-1}\right|\right| \leq (2\,\pi\,m)^{-1}$. This together with Cauchy's inequality yield:

$$\left|\overline{b}_\lambda^t\,\Gamma_d^{-1}\,b_\mu\right|^2 \leq \overline{b}_\lambda^t\,\Gamma_d^{-1}\,b_\lambda\quad \overline{b}_\mu^t\,\Gamma_d^{-1}\,b_\mu \leq d^2\,m^{-2}\,(2\pi)^{-2}.$$

On the other hand, from (A.2.5), we have:

$$\left|\overline{b}_\lambda^t\,\Gamma_d^{-1}\,b_\mu\right|^2 \leq (2\pi)^{-2}\left(\left|e^{i(d)(\mu-\lambda)}\,\phi_d(\lambda)\,\overline{\phi_d(\mu)}\right|^2 + \left|\overline{\phi_d(\lambda)}\,\phi_d(\mu)\right|^2\right)\left|1 - e^{i(\mu-\lambda)}\right|^{-2}$$

$$\leq 2\,(2\pi)^{-2}\left|\phi_d(\lambda)\right|^2\left|\phi_d(\mu)\right|^2\,(1/2)\left(1 - \cos(\mu-\lambda)\right)^{-1}.$$

Thus:

$$\left|\overline{b}_\lambda^t\,\Gamma_d^{-1}\,b_\mu\right| \leq (2\pi)^{-1}\left|\phi_d(\lambda)\right|\left|\phi_d(\mu)\right|\left|\sqrt{2}\,\sin\frac{(\mu-\lambda)}{2}\right|^{-1}$$

Because of (A.2.3), the result follows from $\left|x\,\sin^{-1}(x/2)\right| \leq M'$, $x\in[-\pi,\pi]$. □

We now turn to the study of approximations of f by trigonometric polynomials. From

Butzer and Nessel (1971) Theorem 2.2.3, it is known that if f fulfills (A) and (B) then (for n big enough) there exists a sequence $|t_n|^2$ of positive trigonometric polynomials of degree p_n with $\| |t_n|^2 - f^{-1} \|_\infty = O(p_n^{-\gamma})$, where $\gamma := r+\alpha$, r,α as in (B). We show that the orthogonal polynomials $|\phi_n|^2$ corresponding to f approximate f^{-1} at almost the same rate (up to a ln(n) factor). This will be a consequence of statement 'a' of our next lemma (see also Lemma I.3 a).

Hannan (1989) uses the above mentioned result (Butzer and Nessel) to prove that

$$\sup_{\lambda,\mu \in \mathbf{R}} \left| K_n(f,\lambda,\mu) - K_n(|t_n|^{-2},\lambda,\mu) \right| = O(n\, p_n^{-\gamma}) .$$

In statement 'b' of the next lemma we prove a refinement of this result.

Lemma A.3 Let f fulfill the conditions (A) and (B). Then for a given sequence of integers $p_n \to \infty$ with $p_n \leq n$ there exists a sequence $|t_n|^2$ of positive trigonometric polynomials of degree p_n, such that:

a) $\| |t_n|^2 - f^{-1} \|_\infty = O(p_n^{-\gamma})$ and $\| \phi_n - \psi_{n,n} \|_\infty = O(p_n^{-\gamma} \ln(n))$, where $\{\psi_{n,k}(\lambda)\}_{k \in N}$ denotes the system of orthogonal polynomials associated with $|t_n|^{-2}$ and $\gamma := r+\alpha$, r,α as in (B).

b) $\sup_{\lambda,\mu \in \mathbf{R}} \left| K_n(f,\lambda,\mu) - K_n(|t_n|^{-2},\lambda,\mu) \right| = O(p_n^{-\gamma} \ln(n))\ L^n(\mu-\lambda)$ •

Proof. From Butzer and Nessel (1971) Theorem 2.2.3, follows that (for n big enough) there exists a sequence $|t_n|^2$ of positive trigonometric polynomials of degree p_n with $\| |t_n|^2 - f^{-1} \|_\infty = O(p_n^{-\gamma})$. We will show that this sequence also fulfills a) and b).

Denote by G_n be the n×n Toeplitz matrix associated with $|t_n|^{-2}$ and let $G_n = V_n V_n^t$ be its Cholesky decomposition. First observe that $\| \Gamma_n - G_n \| \leq 2\pi \| |t_n|^{-2} - f \|_\infty$ and further that

$$\| \Gamma_n^{-1} - G_n^{-1} \| = O(p_n^{-\gamma}).$$

To see this observe that $\Gamma_n^{-1} - G_n^{-1} = (U_n^t)^{-1} \left[I - (U_n^{-1} G_n (U_n^t)^{-1})^{-1} \right] U_n^{-1}$ and further that a Neuman expansion yields:

$$\| I - (U_n^{-1} G_n (U_n^t)^{-1})^{-1} \| = \| I - (U_n^{-1} G_n (U_n^t)^{-1}) \| = O\left(\| \Gamma_n - G_n \| \right) = O(p_n^{-\gamma})$$

To prove a) let us expand ϕ_n with respect to $\{\psi_{n,k}(\lambda)\}_{k \in N}$. We obtain:

$$\phi_n(\lambda) = \alpha_n \psi_{n,n}(\lambda) + \int \left[|t_n|^{-2} - f \right](\mu) \sum_{v=0}^{n-1} \psi_{n,v}(\lambda) \overline{\psi_{n,v}(\mu)}\ \phi_n(\mu)\ d\mu$$
$$+ \int f(\mu) \sum_{v=0}^{n-1} \psi_{n,v}(\lambda) \overline{\psi_{n,v}(\mu)}\ \phi_n(\mu)\ d\mu$$

with $\alpha_n = \int |t_n|^{-2}\ \overline{\psi_{n,n}(\mu)}\ \phi_n(\mu)\ d\mu$.

The last term vanishes because of the orthogonality property of ϕ_n with respect to f. The second term on the right side is less than or equal to (note (A.2.3) and Lemma A.2)

$$\left\| |t_n|^{-2} - f \right\|_\infty \ \sup_{\lambda,n} |\phi_n(\lambda)| \int L^n(\lambda)\ d\lambda = O\left(p_n^{-\gamma} \ln(n)\right)$$

Thus for a) it is sufficient to show $|\alpha_n - 1| = O\left(p_n^{-\gamma}\right)$. Now from (A.2.1):

$$\alpha_n = \int |t_n|^{-2}\ \overline{\psi_{n,n}(\mu)}\ \phi_n(\mu)\ d\mu = e_n^t\, U_n^{-1}\, G_n \left(V_n^t\right)^{-1} e_n \ = \ e_n^t\, U_n^{-1}\, V_n\, e_n = \left(U_n^{-1}\right)_{nn} \left(V_n\right)_{nn},$$

where we denote by $(A)_{nn}$ the n,n-th element of the matrix A.

Since $\left(V_n^{-1}\right)_{nn} = \left(V_n\right)_{nn}^{-1}$ and $\left(V_n^{-1}\right)_{nn}^2 = \left(G_n^{-1}\right)_{nn} \geq (\pi m)^{-1}$ it is sufficient to show:

$$\left(U_n^{-1}\right)_{nn} - \left(V_n^{-1}\right)_{nn} = O\left(p_n^{-\gamma}\right).$$

But this follows from

$$\left(U_n^{-1}\right)_{nn}^2 = \left(\Gamma_n^{-1}\right)_{nn},\ \left(V_n^{-1}\right)_{nn}^2 = \left(G_n^{-1}\right)_{nn},\ \left\|\Gamma_n^{-1} - G_n^{-1}\right\| = O\left(p_n^{-\gamma}\right) \text{ and } \left\|\Gamma_n^{-1}\right\| \geq 2\pi\, m^{-1}.$$

This finishes the proof of a).

Let us now prove b). First we obtain from Cauchy's inequality:

$$\left| \overline{b}_\lambda^t\, \Gamma_n^{-1}\, b_\mu - \overline{b}_\lambda^t\, G_n^{-1}\, b_\mu \right| \leq n \left\|\Gamma_n^{-1} - G_n^{-1}\right\| = n\, O\left(p_n^{-\gamma}\right).$$

On the other hand from (A.2.5)

$$\left| K_n\left(f,\lambda,\mu\right) - K_n\left(|t_n|^{-2},\lambda,\mu\right) \right| \leq O(1)\ \left\| \phi_n - \psi_{n,n} \right\|_\infty \left| 1 - \cos\,(\mu-\lambda) \right|^{-1/2}$$

and the result follows from a) and from $x \sin^{-1}(x/2) \leq M'$, $x \in [-\pi,\pi]$. □

Let $\widetilde{B}_d := \frac{1}{2\pi} \int \ln\left(f^{-1}\tilde{f}_d\right)(\lambda) + f\,\tilde{f}_d^{-1}(\lambda) - 1\ d\lambda$ be defined as in Chapter III. In the next proposition we show that $d\,\widetilde{B}_d$ is a sequence falling in d, at least for d large enough.

Proposition A.4 Let f fulfill the conditions (A) and (B) with $r \geq 2$ and $\alpha > 0$. Then we have:

$$d\,\widetilde{B}_d - (d+1)\,\widetilde{B}_{d+1} = \frac{1}{4\pi} \int \left(f^{-1}\tilde{f}_d - 1\right)^2(\lambda)\ d\lambda + o\left(d^{-2}\right)$$

Proof. From $(2\pi)^{-1} \int f \tilde{f}_d^{-1}(\lambda)\, d\lambda = d^{-1} \operatorname{tr}\left(\Gamma_d^{-1} \int f(\lambda)\; b_\lambda \bar{b}_\lambda^t\, d\lambda\right) = 1$ we have:

$$d\, \tilde{B}_d - (d+1)\, \tilde{B}_{d+1} = \frac{1}{2\pi} \int d\left[\ln(\tilde{f}_d) - \ln(f)\right](\lambda) - (d+1)\left[\ln(\tilde{f}_{d+1}) - \ln(f)\right](\lambda)\, d\lambda$$

$$= \frac{1}{2\pi} \int \left[\ln(f) - \ln(\tilde{f}_d)\right](\lambda) + (d+1) \ln\left[\tilde{f}_{d+1}^{-1} \tilde{f}_d\right](\lambda)\, d\lambda$$

Now since $\frac{\tilde{f}_{d+1}^{-1}}{\tilde{f}_d^{-1}} = 1 + (d+1)^{-1}\left[\frac{|\phi_d|^2}{\tilde{f}_d^{-1}} - 1\right]$, where ϕ_d is the orthogonal polynomial of order d with respect to f, we obtain that the quantity above equals:

$$= \frac{1}{2\pi} \int - \ln\left[\tilde{f}_d f^{-1}\right](\lambda) + (d+1) \ln\left[1 + (d+1)^{-1}\left(|\phi_d|^2 \tilde{f}_d - 1\right)\right](\lambda)\, d\lambda$$

Expanding the ln-terms we obtain with Lemma I.3 that the quantity above equals:

$$= \frac{1}{2\pi} \int -\left(f^{-1}\tilde{f}_d - 1\right)(\lambda) + \frac{1}{2}\left(f^{-1}\tilde{f}_d - 1\right)^2(\lambda) + \left(|\phi_d|^2 \tilde{f}_d - 1\right)(\lambda)\, d\lambda + o\left(d^{-2}\right)$$

which with $|\phi_d|^2 \tilde{f}_d - 1 = f^{-1}\tilde{f}_d - 1 + o\left(d^{-2}\right)$ (Lemma I.3) yields our assertion. □

A.3 The variance function θ

Let θ be defined as in Chapter I. The following lemma gives its relevant properties:

Lemma A.5 θ fulfills the following:

i) $\theta(x) \to \frac{2}{3}$, $x \to \infty$

ii) $\theta(x) = x^{-1}$, $0 < x \leq 1$.

iii) $\theta(x) \geq \frac{2}{3}$, $x \geq 0$

iv) $\arg\inf\left\{\theta(dc^{-1}),\, c \in \mathbf{N}^+\right\} = 1$, $\forall\, d \in \mathbf{N}^+$

Proof. For $x \in \mathbf{R}^+$ let $\alpha_x := x - [x] \in [0,1)$. Then θ may be written as:

(1) $$\theta(x) = \frac{2}{3} + \frac{x^{-2}}{3} - \left(\frac{2}{3}\alpha_x^3 - \alpha_x^2 + \frac{\alpha_x}{3}\right) x^{-3}.$$

If $x \in \mathbf{N}$ then $\alpha_x = 0$ and (2) $\theta(x) = \frac{2}{3} + \frac{x^{-2}}{3}$.

If $x \leq 1$ then $\alpha_x = x$ and $\theta(x) = x^{-1}$. This proves ii).

(3) For $k \leq x < k+1$ for some k it is easily seen that θ is falling.

Let us now prove iv). We have for $c \in \mathbb{N}^+$ and $x := dc^{-1}$:

$$3d^3 \left[\theta(d) - \theta(dc^{-1})\right] = c^3 \left(2\, \alpha_x^3 - 3\alpha_x^2 + \alpha_x\right) - c^2 d + d \leq A\, c^3 - c^2 d + d, \text{ with}$$

$$A := \sup_{0 \leq \alpha \leq 1} \left(2\, \alpha^3 - 3\alpha^2 + \alpha\right) = \sqrt{3} / 18.$$

Now for fixed d and $c \leq d$ the function $A\, c^3 - c^2 d + d$ is falling in c and is negative for c=2, thus it is negative for $2 \leq c \leq d$. This together with ii) proves iv). Together with (2) it proves iii) and taking (3) into account i) follows too. □

A.4 Approximation of the Kullback-Leibler by the Whittle discrepancy

In this section we prove the result announced in Section II.1.1 concerning the approximation of the Kullback-Leibler distance $d_T\left(f, \hat{f}_{d,T}\right)$ between the Capon estimator $\hat{f}_{d,T}$ and the spectrum f by the Whittle discrepancy $\Delta\left(f, \hat{f}_{d,T}\right)$. Adopting the same notation as in Section II.1.1 we have:

Lemma A.6 Assume that (A) and (C) hold and that $\|f'\|_\infty < \infty$. Further assume that $d = d_T$ fulfills

i) $\dfrac{d^{1+\varepsilon}}{T} \to 0$ for some $\varepsilon > 0$ and

ii) $d^{-4}\, T\, \ln^3(T) \to 0$.

Then as $T \to \infty$ <u>we have</u>: $\left| d_T\left(f, \hat{f}_{d,T}\right) - \Delta\left(f, \hat{f}_{d,T}\right)\right| = o_p(d/T)$ •

Proof. Let $A_T := \left\{ \left\| \hat{f}_{d,T}\, f^{-1} - 1 \right\|_\infty \leq 1/2 \right\}$ and

$$\alpha_T := \left(d^{\,1+\delta}\, T^{-1} \ln(d) \ln(T) \right)^{1/2} \vee d^{-1} \ln(T)$$

with some $\delta \leq \varepsilon\,(1-\varepsilon)/6$. Observe that for d big enough from Lemmata I.2 and I.3 we have:

$$\left\| \hat{f}_{d,T}\, f^{-1} - 1 \right\|_\infty \leq O(1) \left\| \hat{f}_{d,T}\, \tilde{f}_d^{-1} - 1 \right\|_\infty + \left\| \tilde{f}_d\, f^{-1} - 1 \right\|_\infty$$

$$\leq O(1) \left\| \hat{\Gamma}^*_{d,T} - I \right\| + \left\| \tilde{f}_d\, f^{-1} - 1 \right\|_\infty = O_P(\alpha_T).$$

Let us denote by $B_T^* := U_T^{-1}\; B_T\left(\hat{f}_{d,T}\right) \left(U_T^{-1}\right)^t$ and its eigenvalues by $\mu_1, \ldots, \mu_T$. Then it follows from the previous that $\left\| B_T^* - I \right\| \leq \left\| \hat{f}_{d,T}\, f^{-1} - 1 \right\|_\infty = O_P(\alpha_T)$.

Thus on A_T we may expand:

$$d_T\left(f,\hat{f}_{d,T}\right) = T^{-1}\sum_{i=1}^{T}\left[\ \ln(\mu_i) + \mu_i^{-1} - 1\right] = (2T)^{-1}\sum_{i=1}^{T}\left[\ \mu_i - 1\right]^2 + O\left(\alpha_T^3\right)$$

$$= (2T)^{-1}\ \mathrm{tr}\left[B_T^* - I\right]^2 + O\left(\alpha_T^3\right)$$

$$= 1/2\int\left[\hat{f}_{d,T} - f\right](\lambda)\left[\hat{f}_{d,T} - f\right](\mu)\ T^{-1}\left|K_T(\lambda,\mu)\right|^2 d\lambda\ d\mu\ \ + O\left(\alpha_T^3\right)$$

On the other hand on A_T we may also expand

$$\Delta\left(f,\hat{f}_{d,T}\right) = (4\pi)^{-1}\int\left[\hat{f}_{d,T}f^{-1} - 1\right]^2(\lambda)\ d\lambda\ \ + O\left(\alpha_T^3\right)$$

Thus, since $\int f(\mu)\ T^{-1}\left|K_T(\lambda,\mu)\right|^2\ d\mu\ = (2\pi)^{-1}\ \tilde{f}_T^{-1}(\lambda)$, we have (uniformly) on A_T:

$$\left|d_T\left(f,\hat{f}_{d,T}\right)\ -\ \Delta\left(f,\hat{f}_{d,T}\right)\right| \leq$$

$$= 1/2\int\left|\hat{f}_{d,T} - f\right|(\lambda)\left|\left[\hat{f}_{d,T} - f\right](\mu) - \left[\hat{f}_{d,T} - f\right](\lambda)\ \frac{f(\mu)\ \tilde{f}_T(\lambda)}{f^2(\lambda)}\right| T^{-1}\left[L^T(\mu-\lambda)\right]^2 d\lambda\ d\mu + O\left(\alpha_T^3\right)$$

$$= O_p\left(\alpha_T\right)\int\left|\left[\hat{f}_{d,T} - f\right](\mu) - \left[\hat{f}_{d,T} - f\right](\lambda)\right| T^{-1}\left[L^T(\mu-\lambda)\right]^2 d\lambda\ d\mu + O_P\left(\alpha_T^2\ \frac{\ln(T)}{T}\right) + O\left(\alpha_T^3\right)$$

It is easy to verify that:

i) $\left|\hat{f}_{d,T}(\mu) - \hat{f}_{d,T}(\lambda)\right| = O_P(1)\left[\ d^{-1}\|b_\lambda\|_2\ \|b_\lambda - b_\mu\|_2\ \wedge\ 1\right] = O_P(1)\left[\ d\left|\lambda - \mu\right|\ \wedge\ 1\right]$

ii) $\int\left[\ d|x|\ \wedge\ 1\right] T^{-1}\left[L^T(x)\right]^2 dx = O\left(\frac{d\ \ln(T)}{T}\right)$

It follows:

$$\frac{T}{d}\left|d_T\left(f,\hat{f}_{d,T}\right)\ -\ \Delta\left(f,\hat{f}_{d,T}\right)\right|\ =\ O_P\left(\alpha_T \ln(T)\right)\ +\ O\left(\frac{\alpha_T^3\ T}{d}\right).$$

This quantity converges to 0 under the conditions of the lemma. Since also $P\left(A_T^c\right) \rightarrow 0$ the proof is finished. □

REFERENCES

Akaike, H. (1970). Statistical Predictor Identification. *Ann. Inst. Statist. Math.* **22** 203-217.

Baggerroer, A.B. (1976). Confidence Intervals for Regression (MEM) Spectral Estimates. *IEEE Trans.Inform.Theory* **22** 534-545.

Berk, K.N. (1974). Consistent Autoregressive Spectral Estimates. *Annals Stat.* **2** 489-502.

Bloomfield, P. (1976). Fourier Analysis of Time Series: An Introduction. John Wiley and Sons, New York, London, Sydney, Toronto.

Brillinger, R. (1975). Time Series: Data Analysis and Theory. Holt, Rinehart and Winston Inc.

Brockwell, P. and Davis, R. (1987). Time Series: Theory and Methods. Springer-Verlag, New York, Berlin, Heidelberg, London, Paris, Tokyo.

Burg, J.P. (1972). The Relationship between Maximum Entropy and Maximum Likelihood Spectra. *Geophys.***37** 375-376.

Butzer, P. and Nessel, R. (1971).Fourier Analysis and Approximation, Vol 1: One Dimensional Theory. Birkhaeuser Verlag, Basel und Stuttgart.

Byrne, C. L. and Fitzgerald, R. M. (1984). Spectral Estimators that Extend the Maximum Entropy and the Maximum Likelihood Methods. *SIAM J.Appl.Math.***44** 425 -442.

Capon, J. (1969). High Resolution Frequency Wavenumber Spectrum Analysis. *Proc IEEE* **57** 1408 - 1418.

Capon, J. and Goodman, N.R. (1970). Probability Distributions for Estimates of the Frequency Wavenumber Spectrum. *Proc IEEE* **58** 1785 - 1786.

Coursol, J. and Dacuhna-Castelle. (1982). Remarks on the Approximation of the Likelihood Function of a Stationary Gaussian Process. *Theory of Probability and its Applications* **27** 162-167.

Dahlhaus, R (1983). Spectral Analysis with Tapered Data. *Journal of Time Series Analysis* **4** 163-175.

Dahlhaus,R (1985). On a Spectral Density Estimate Obtained by Averaging Periodograms. *J.Appl. Prob.* **22** 598-610.

Dahlhaus,R (1990). Nonparametric High Resolution Spectral Estimation. *Probability Theory Rel.Fields* **85** 147-180.

Grenander, Ul. and Szegö, G. (1958). Toeplitz Forms and their Applications, University of California Press, Berkeley and Los Angeles.

Hannan, E.J. and Wahlberg, B.(1989). Convergence Rates for Inverse Toeplitz Matrix Forms. *Journal of Multivariate Analysis* **31** 127-135.

Hurvich, C.M. and Tsai, C.L. (1989). Regression and Time Series Model Selection in Small Samples. *Biometrica* **76** 297-307.

Hosoya, Y. and Taniguchi, M. (1982). A Central Limit Theorem for Stationary Processes and the Parameter Estimation of Linear Processes. *Annals Stat.* **10** 132-153.

Kay, S.M. and Marple, S.L. (1981). Spectrum Analysis - A Modern Perspective. *Proc IEEE* **69** 1380 - 1419.

Lewis, R. and Reinsel, G.C. (1985). Prediction of Multivariate Time Series by Autoregressive Model Fitting. *Journal Mult. Anal.* **16** 393-411.

Linhart and Zucchini (1986). Model selection. Wiley Series in Probability and Mathematical Statistics.

McDonough, R.N. (1979). Application of the Maximum Likelihood Method and the Maximum Entropy Method to Array Processing. In: Nonlinear Methods of Spectral analysis, Edit : Haykin, 181 - 245, Springer Verlag Heidelberg, New York 1979

Marzetta, T.L. (1983). A New Interpretation for Capon's Maximum Likelihood Method of Frequency Wavenumber Spectral Estimation. *IEEE Trans. Acoust. Speech. Sign.Proc.***31** 445.

Parzen, E. (1957). On Consistent Estimates of the Spectrum of a Stationary Time Series. *Ann. Math. Statist.* **28** 329-348.

Pisarenko, V.F. (1972). On the Estimation of Spectra by Means of Non-linear Functions of the Covariance Matrix. *Geophys. J. Royal Astr. Soc.* **28** 511 - 531.

Rosenblatt, M. (1985). Stationary Sequences and Random Fields. Birkhäuser, Boston, Basel, Stuttgart.

Shibata, R. (1980). Asymptotically Efficient Selection of the Order of the Model for Estimating Parameters of a Linear Process. *Annals Stat.* **8** 147-164.

Subba Rao, T. and Gabr M.M.(1989). The Estimation of Spectrum, Inverse Spectrum and Inverse Autocovariances of a Stationary Time Series. *Journal of Time Series Analysis* **10** 183-202.

Szegö, G. (1959). Orthogonal Polynomials, American Mathematical Society Colloquium Publications vol XXIII, New York 1959.

Taniguchi, M. (1980). On Selection of the Order of the Spectral Density Model for a Stationary Process. *Ann. Inst. Statist. Math.* **32** 401-419.

Zhurbenko, I.G. (1980). On the Efficiency of Spectral Density Estimates of a Stationary Process. *Theory Prob. Appl.* **25** 466-480.

ISBN 3-86111-617-0