Higher Order Asymptotic Theory for Nonparametric Time Series Analysis and Related Contributions

Carlos Velasco

1996

A Thesis submitted in partial fulfillment of the requirements of the Degree of Doctor of Philosophy in Statistics at the London School of Economics and Political Science (University of London)

Abstract

We investigate higher order asymptotic theory in nonparametric time series analysis. The aim of these techniques is to approximate the finite sample distribution of estimates and test statistics. This is specially relevant for smoothed nonparametric estimates in the presence of autocorrelation, which have slow rates of convergence so that inference rules based on first-order asymptotic approximations may not be very precise.

First we review the literature on autocorrelation-robust inference and higher order asymptotics in time series. We evaluate the effect of the nonparametric estimation of the variance in the studentization of least squares estimates in linear regression models by means of asymptotic expansions. Then, we obtain an Edgeworth expansion for the distribution of nonparametric estimates of the spectral density and studentized sample mean. Only local smoothness conditions on the spectrum of the time series are assumed, so long range dependence behaviour in the series is allowed at remote frequencies, not necessary only at zero frequency but at possible cyclical and seasonal ones.

The nonparametric methods described rely on a bandwidth or smoothing number. We propose a cross-validation algorithm for the choice of the optimal bandwidth, in a mean square sense, at a single point without restrictions on the spectral density at other frequencies.

Then, we focus on the performance of the spectral density estimates around a singularity due to long range dependence and we obtain their asymptotic distribution in the Gaussian case. Semiparametric inference procedures about the long memory parameter based on these nonparametric estimates are justified under mild conditions on the distribution of the observed time series. Using a fixed average of periodogram ordinates, we also prove the consistency of the log-periodogram regression estimate of the memory parameter for linear but non-Gaussian time series.

Acknowledgments

I am deeply indebted to my supervisor Professor Peter M. Robinson. He made possible the beginning of this thesis and provided best guidance towards its completion.

I wish to thank Dr Martin Knott, who has been always ready to help. I am grateful to Dr Luis Borge, Dr Miguel Delgado, Dr Javier Hidalgo and Dr Lola Prada. This thesis is a result of their inestimable advice at different stages. I also wish to thank Professors Antoni Espasa, Alexander Samarov and Masanobu Taniguchi.

Financial support from Fundación Ramón Areces and ESRC (Grant R000235892) is gratefully acknowledged.

I would like to thank Jesús Arteche, Luis Gil Alaña, Ignacio Lobato, Giovanni Marseguerra and Paolo Zaffaroni for constant support.

I wish to express my gratitude to my family for their continuous understanding and to dedicate this thesis to Rosa for her infinite patience and encouragement during these years.

Contents

List of Tables and Figures

Page

Chapter 4. Local Cross Validation for Spectrum Bandwidth Choice

Chapter 1

Introduction

1.1 Autocorrelation-robust inference

The main feature of time series data is their dependence across time. Accounting for serial dependence can considerably complicate statistical inference. The attempt to model the dependence parametrically, or even nonparametrically, can be difficult and computationally expensive.

In some circumstances, the serial dependence is merely a nuisance feature, interest focusing on "static" aspects. Here we can frequently base inference on point or function estimates that are natural ones to use in case of independence, and may well be optimal in that case. Such estimates will generally be less efficient than ones based on a comprehensive model that incorporates the serial dependence. But apart from their relative computational simplicity, they often remain consistent even in the presence of certain forms of dependence, and can be more reliable or "robust" than the "efficient" ones, which can sometimes become inconsistent when the dependence is inappropriately dealt with, leading to statistical inferences that are invalid, even asymptotically.

Exact finite-sample distributions of estimates and test statistics are only available for simple functions of the data and when the likelihood function is completely specified. Then, we have to rely on approximations to the sampling distributions, based typically on the central limit theorem. Asymptotic or large-sample theory is relatively simple to use, and also relies on milder assumptions than finite-sample theory, and can thus be more widely applied, at least given data sets of reasonable size. In the asymptotics for time series, a very important role is played by the stationarity assumption, and indeed by further restrictions on the dependence of a (possibly unobservable) process associated with the data or the statistical model.

Valid inference based on an asymptotically normal statistic requires only a consistent estimate of the variance matrix in the limiting distribution. Then the statistic can be studentized, consistent confidence regions set, and asymptotically valid hypothesis tests carried out. Usually the variance is affected by the dependence, and requires a different, and more complicated, type of estimate than that under independence. This can be based on a parametric model for the autocorrelations. However nonparametric types of estimate are more popular, being consistent under broader conditions.

Alternate methods related with the bootstrap have been proposed recently for dependent observations (Künsch (1989), Politis and Romano (1995), Bertail et al. (1995)). The bootstrap proposed originally by Efron (1979) does not work in this case, but is possible to approximate the distribution of the statistic of interest by the empirical distribution of that statistic calculated from a large number of blocks of consecutive observations. Connected ideas to those and the jackknife apply as well for estimation of the variance of different statistics (see Section 1.3).

We now describe three general frameworks were an adaptation for serial dependence is necessary in the estimation of the asymptotic variance, complicating asymptotic inference. However, there are situations were independence-based rules continue to apply, like nonparametric probability density and regression functions estimation. For more examples see Robinson and Velasco (1996).

1.1.1 Sample mean

Let $\{X_t, t = 1, 2, \ldots\}$ be a real-valued covariance stationary sequence with expectation $\mu = E[X_t]$, lag-j autocovariance

$$\gamma(j) = E\left[(X_t - \mu)(X_{t+j} - \mu)\right],$$

and spectral density $f(\lambda)$ given by

$$\gamma(j) = \int_{-\pi}^{\pi} f(\lambda) \cos j\lambda d\lambda. \tag{1.1}$$

The existence of $f(\lambda)$ requires that the autocovariances $\gamma(j)$ must decay to zero fast enough. A sufficient condition for that is the absolute summability of $\gamma(j)$,

$$\sum_{j=-\infty}^{\infty} |\gamma(j)| < \infty. \tag{1.2}$$

The faster $\gamma(j)$ decay to zero, the smoother $f(\lambda)$ will be. Relation (1.1) gives for $j = 0$

$$\gamma(0) = \int_{-\pi}^{\pi} f(\lambda) d\lambda,$$

a decomposition of the variance of the time series across frequencies. Therefore, the magnitude of $f(\lambda)$ is a measure of the importance of the frequency λ in the variance of the time series.

Inference on the mean μ given observations $X_1, \ldots, X_N$ is typically based on the sample mean

$$\overline{X} = \frac{1}{N} \sum_{t=1}^{N} X_t,$$

which is the ordinary least squares estimate (OLSE) of μ. With dependent observations $\overline{X}$ is no longer a best linear unbiased estimate (BLUE) or maximum likelihood estimate (if X_t is Gaussian) as in the independent case, but it is unbiased and Grenander (1954) showed that if $f(\lambda)$ is continuous and positive at $\lambda = 0$, $\overline{X}$ is asymptotically efficient within the class of linear unbiased estimates, so no differential weighting need to be considered. Also

$$\mathrm{Var}[\overline{X}] = \frac{1}{N} \sum_{t=1-N}^{N-1} \left(1 - \frac{|j|}{N}\right) \gamma(j), \tag{1.3}$$

and under various conditions permitting dependence in X_t, expressed in terms of the α-mixing numbers or linear process conditions (see Ibragimov and Linnik (1971) and Hannan (1979)) we can obtain

$$\frac{\overline{X} - \mu}{\mathrm{Var}[\overline{X}]^{1/2}} \to_d \mathcal{N}(0, 1), \qquad \text{as } N \to \infty. \tag{1.4}$$

These conditions imply that the $\gamma(j)$ decay fast enough to be absolutely summable (c.f. (1.2)), and thus that $f(\lambda)$ is continuous for all λ, so they can be referred to as "weak dependence" conditions. Hence we have from Fejér's theorem (Zygmund (1977), p. 89) and also from (1.2)

$$N \operatorname{Var}[\overline{X}] \to 2\pi f(0), \quad \text{as } N \to \infty.$$

Therefore, to implement large-sample inference it remains to find an estimate $\widehat{f}(0)$ of $f(0)$ such that

$$\widehat{f}(0) \to_p f(0), \quad \text{as } N \to \infty, \tag{1.5}$$

for then

$$\frac{N^{1/2}\left(\overline{X}-\mu\right)}{\left[2\pi \widehat{f}(0)\right]^{1/2}} \tag{1.6}$$

can be approximated by a $\mathcal{N}(0,1)$ variate. Given a parametric model for $\gamma(j)$, equivalently for $f(\lambda)$, like a stationary autoregression (AR), invertible moving average (MA), or stationary and invertible autoregressive moving average (ARMA), we can estimate the unknown parameters, insert them in the formula for $f(0)$, and obtain an $N^{1/2}$-consistent estimate. However, (1.5) requires no rate of convergence, and there has been greater stress on using nonparametric spectrum estimates which are consistent in the absence of assumptions on the functional form, though as even the early papers on studentizing the sample mean of a time series by Jowett (1955) and Hannan (1957) illustrate, the same estimates of $\widehat{f}(0)$ can be interpreted as either parametric or nonparametric. We review the literature about such estimates in Section 1.3.

1.1.2 Linear regression

The simple previous location model extends to the multiple regression

$$Y_t = \beta' Z_t + X_t, \quad t = 1, 2 \ldots \tag{1.7}$$

where X_t is, as before, covariance stationary, but it now has zero mean and is unobservable, the scalar Y_t and q-dimensional column vector Z_t being observed. Again the properties of the OLSE

$$\widehat{\beta} = \left(\sum_{t=1}^{N} Z_t Z_t'\right)^{-1} \left(\sum_{t=1}^{N} Z_t Y_t\right)$$

of the vector β partially extend from independent to dependent situations. In particular, Grenander (1954) gave conditions under which $\widehat{\beta}$ is asymptotically as efficient as the

BLUE of β, in case of nonstochastic Z_t. These conditions are satisfied if, for example Z_t is a vector of polynomials of trigonometric functions in t and $f(\lambda)$ is continuous and positive for all $\lambda \in (-\pi, \pi]$. Even when these conditions are not satisfied $\widehat{\beta}$ is often consistent and asymptotically normal, whereas a generalized least squares estimate of β, obtained under a misspecified parametric model for $f(\lambda)$, might be inconsistent. However, whether or not it is asymptotically efficient, the asymptotic variance of $\widehat{\beta}$ is not the same as it would for uncorrelated disturbances X_t.

Assume that the matrix of sample correlation coefficients of Z_t for both stochastic and nonstochastic regressors achieve a limit as the sample size increases. Denote this limit matrix for lag j as $R(j) = (\rho_{h,i}(j))$ and assume that it is positive definite for $j = 0$ and

$$R(j) = \int_{-\pi}^{\pi} e^{ij\lambda} dG(\lambda), \qquad j = 0, \pm 1, \ldots$$

for some matrix G with Hermitian nonnegative definite increments, continuous from the right and with $G(-\pi) = 0$. G is usually called the spectrum measure of the regressors.

Then, under regularity conditions on the asymptotic behaviour of the regressors, and on the errors X_t similar to those sufficient for (1.4) we can obtain,

$$D(\widehat{\beta} - \beta) \to_d \mathcal{N}(0, R(0)^{-1} S R(0)^{-1}),$$

where $D = \mathrm{diag}\{D_1, \ldots, D_q\}$ with $D_j = (\sum_{t=1}^{N} Z_{jt}^2)^{1/2}$, and

$$S = 2\pi \int_{-\pi}^{\pi} f(\lambda) dG(\lambda). \tag{1.8}$$

Because $R(0)$ is consistently estimated by $D^{-1} \sum_t Z_t Z_t' D^{-1}$, it remains to estimate S. Equivalently, we have the approximation

$$\widehat{\beta} \to_d \mathcal{N}\left(\beta, \left[\sum_{t=1}^{N} Z_t Z_t'\right]^{-1} \mathrm{Var}\left[\sum_{t=1}^{N} Z_t X_t\right] \left[\sum_{t=1}^{N} Z_t Z_t'\right]^{-1}\right),$$

where Var[$\cdot$] refers now to the covariance matrix of its argument. Then

$$\mathrm{Var}\left[\sum_{t=1}^{N} Z_t X_t\right] \sim D^{1/2} S D^{1/2} = T \quad \text{as } N \to \infty,$$

so we require an estimate $\widehat{T}$ of T such that

$$D^{-1/2}(\widehat{T} - T) D^{-1/2} \to_p 0.$$

Several nonparametric estimates of T have been proposed in both the time and frequency domains (see Section 1.3.2). Many of them are also valid for more complicated inference procedures, specially methods used recently in econometrics reviewed in Section 1.1.3.

The OLSE is a special case (with $h \equiv 1$) of

$$\widehat{\beta}(h) = \left(\sum_{j=1}^{N} I_Z(\lambda_j) h(\lambda_j) \right)^{-1} \sum_{j=1}^{N} I_{ZY}(\lambda_j) h(\lambda_j), \tag{1.9}$$

where $I_{ZY}(\lambda) = (2\pi N)(\sum_{t=1}^{N} Z_t e^{it\lambda})(\sum_{t=1}^{N} Y_t e^{it\lambda})$ and $h(\lambda)$ is a real, nonnegative function. For general h,

$$D\left(\widehat{\beta}(h) - \beta\right) \to_d$$
$$\mathcal{N}\left(0, \left[\int_{-\pi}^{\pi} h(\lambda) dG(\lambda)\right]^{-1} 2\pi \int_{-\pi}^{\pi} f(\lambda) h^2(\lambda) dG(\lambda) \left[\int_{-\pi}^{\pi} h(\lambda) dG(\lambda)\right]^{-1}\right), \tag{1.10}$$

under regularity conditions. The user-chosen function h can achieve two distinct goals. One is to reduce bias due to misspecification of (1.7) from errors in the observation of regressors. Then, we take $h(\lambda) = 0$ for frequencies where the signal-to-noise ratio is feared to be low, like can be at high frequencies for white noise errors in the observations or at seasonal ones when the noise is seasonal. The idea of omission of frequencies was proposed by Hannan (1963a), and justified theoretically by Hannan and Robinson (1973), Robinson (1972).

Irrespective of whether or not we eliminate frequencies, there remains the choice of h at the other frequencies. Hannan (1963b) showed, for a modification of (1.9), that the data-dependent choice $h(\lambda) = \widehat{f}^{-1}(\lambda)$, with $\widehat{f}$ representing a nonparametric spectrum estimate, achieves the same asymptotic efficiency as the GLSE using a parametric model for $f(\lambda)$ and is asymptotically normal distributed. Robinson (1991) showed that this remains true in case of a general data dependent bandwidth in $\widehat{f}$, and allowed also for omission of frequencies.

Samarov (1987) investigated if adapting unknown correlation is necessary when there is only a small unspecified autocorrelation from the point of view of loss of efficiency with respect to the OLSE.

One important class of estimates of location and regression models, which generalizes the OLSE, are M-estimates. These estimates are designed to avoid distortions in finite samples from a few "outliers", seemingly contaminated observations that are of much greater magnitude than are the bulk of the data. Here, parallel, but more involved results apply. See Robinson and Velasco (1996), Section 4 for more details.

1.1.3 Econometric models

Due to the occurrence of macroeconomic and financial time series data, many econometric methods are devised with possible serial dependence in mind. In fact relatively early econometric work stressed the efficiency gains due to the GLSE in regression models in the presence of autocorrelated errors, see e.g. Cochrane and Orcutt (1949). This interest has continued, and more recently it has been fashionable to employ point estimates which may well be inefficient, but studentize them to allow for serial dependence. A feature of much econometric work is the relative complexity of modelling, often involving nonlinearity and multivariate data, for example. Some similar models have been used in non-economic applications but we have chosen to categorize as "econometric models" ones which are more complicated than the simple location and linear regression models treated so far.

Many important problems, involving nonlinearities and other complications, are covered by the model

$$Y_t = H_t(\theta) + X_t, \quad t = 1, 2, \dots \tag{1.11}$$

where now Y_t and X_t are $p \times 1$ vectors and $H_t(\theta)$ is a $p \times 1$ vector of possibly nonlinear functions of an unknown $s \times 1$ vector θ, and of observable stochastic or nonstochastic explanatory variables to which explicit reference is suppressed. Again X_t is unobservable.

The function $H_t(\theta)$ could be linear in the observables,

$$H_t(\theta) = \mu(\theta) + \sum_{j=-\infty}^{\infty} A_j(\theta) Z_{t-j}, \tag{1.12}$$

where the Z_t are the $q \times 1$ explanatory variables, the $A_j(\theta)$ are $p \times q$ matrix functions of θ, and $\mu(\theta)$ is a vector. This includes static linear simultaneous equations models of econometrics and distributed lag models. Also nonlinear models are covered by (1.11), like static scalar or multivariate nonlinear regression models.

Suppose first that the processes $\{H_t(\theta)\}$ and $\{X_t\}$ are independent of each other and stationary. A minimum-distance or Gaussian estimate of θ is given by

$$\widehat{\theta} = \arg\min_{\vartheta\in\Theta} \sum_{j=1}^{N} \|\Phi(\lambda_j)^{1/2}\{w_Y(\lambda_j) - w_H(\lambda_j;\vartheta)\}\|^2,$$

where $\|\cdot\|$ means Euclidean norm, $\Phi(\lambda)$ is a $p\times p$ positive definite Hermitian matrix, and

$$w_Y(\lambda) = (2\pi N)^{-1/2}\sum_{t=1}^{N} Y_t e^{it\lambda}, \qquad w_H(\lambda) = (2\pi N)^{-1/2}\sum_{t=1}^{N} H_t(\vartheta)e^{it\lambda},$$

and Θ is a compact subset of $\mathcal{R}^s$. In case (1.12) leads to H including some unobservable Z_t, we can consider instead

$$\widehat{\theta} = \arg\min_{\vartheta\in\Theta} \sum_{j=1}^{N} \|\Phi(\lambda_j)^{1/2}\{w_Y(\lambda_j) - A(\lambda_j;\vartheta)w_Z(\lambda_j)\}\|^2,$$

$$w_Z(\lambda) = (2\pi N)^{-1/2}\sum_{t=1}^{N} Z_t e^{it\lambda}, \qquad A(\lambda;\vartheta) = \sum_{-\infty}^{\infty} A_j(\vartheta)e^{ij\lambda}.$$

Such estimates were considered by Hannan (1971), Robinson (1972), and cover many cases. Under regularity conditions given in these references,

$$N^{1/2}(\widehat{\theta} - \theta) \to_d \mathcal{N}(0,\ \Psi(\Phi)^{-1}\Psi(\Phi f\Phi)\Psi(\Phi)^{-1}), \tag{1.13}$$

where

$$\Psi(\Phi) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\partial}{\partial\vartheta_a\partial\vartheta_b'}\mathrm{Trace}\{\Phi(\lambda)dF(-\lambda;\vartheta_a,\vartheta_b)\}\bigg|_{\vartheta_a,\vartheta_b=\theta}, \tag{1.14}$$

$\Psi(\Phi f\Phi)$ is the same quantity with $\Phi(\lambda)$ replaced by $\Phi(\lambda)f(\lambda)\Phi(\lambda)$, and F satisfies

$$E[H_t(\vartheta_a)H_{t+j}'(\vartheta_b)] = \int_{-\pi}^{\pi} e^{ij\lambda}dF(\lambda;\vartheta_a,\vartheta_b).$$

When $\Phi \equiv I_p$, $\widehat{\theta}$ can be neatly written in time domain form, and when $p=1$,

$$N^{1/2}(\widehat{\theta} - \theta) \to_d \mathcal{N}(0,\ A^{-1}BA^{-1}), \tag{1.15}$$

where $A = E[\partial H_t(\theta)/\partial\theta\,\partial H_t(\theta)/\partial\theta']$, $B = 2\pi f_u(0)$, where $f_u(\lambda)$ is the spectral density matrix of $U_t = X_t\partial H_t(\theta)/\partial\theta$.

One approach much stressed in the subsequent econometric literature is as follows. Let a vector parameter satisfy the equation

$$E[U_t(\theta)] = 0, \qquad t = 1,2,\ldots, \tag{1.16}$$

where the dimension of U_t is at least as great as that of θ and may also depend on other unknown parametric and/or nonparametric functions. In particular (1.16) embodies the econometric model, or the principal part of it. Given a suitable matrix S of the same rank, and proxies $\widehat{U}_t(\theta)$ for $U_t(\theta)$, involving estimates of any nuisance parameters/functions, and possibly each depending on all the data, we estimate θ by

$$\widehat{\theta} = \arg\min_{\vartheta \in \Theta} \sum_{t=1}^{N} \widehat{U}_t(\vartheta)\widehat{S}'\widehat{S} \sum_{t=1}^{N} \widehat{U}_t(\vartheta)$$

for a compact set Θ. Under regularity conditions

$$N^{1/2}(\widehat{\theta} - \theta) \rightarrow_d \mathcal{N}(0, C^{-1}DC^{-1}), \tag{1.17}$$

where

$$C = S\,E\left[\frac{\partial U_t(\theta)}{\partial \theta'}\right], \qquad D = S\,2\pi\,f_u(0)\,S',$$

and $f_u(\lambda)$ is the spectral density matrix of $U_t(\theta)$. The generalized method-of-moments (GMM) estimates of Hansen (1982) fall into this scheme.

1.2 Plan of the Thesis

In this thesis we analyze different aspects of studentization in dependent situations using nonparametric estimates of the variance. Since nonparametric techniques estimate in theory an infinite number of (nuisance) parameters, they have typically slower rates of convergence than parametric ones and poorer small sample properties. Thus, there is a special concern about how this slow convergence will affect the properties of the studentized estimates and, therefore, about the merits of the asymptotic normal approximation, which is not affected by the smoothing number employed in the nonparametric estimation.

We consider the studentization of the sample mean and least squares estimates in a general linear time series regression model with nonstochastic regressors. We start with the regression model in Chapter 2 and then we specialize to the sample mean in Chapter 3 under much milder assumptions.

For the sample mean (or for any other simple trigonometric or polynomial in t regression) only the behaviour of $f(\lambda)$ at a particular frequency is relevant, given the form of the matrix S when $dG(\lambda)$ is a (single) jump function (see equation (1.8)). Therefore, in these situations it is difficult to justify any condition on the dependence structure of the time series that implies a global smoothness restriction on the spectrum, like (1.2) or conditions on the mixing coefficients, and the assumptions should concentrate only at that particular frequency. A minimum requirement across all frequencies is the integrability of the spectral density in $(-\pi, \pi]$, necessary for covariance stationarity.

The analysis of the properties of the nonparametric studentization will be based on higher order asymptotic properties of the distribution of the studentized estimates. We justify asymptotic expansions of the Edgeworth type for those statistics under Gaussian assumptions for X_t. The Gaussianity will allow us to express the restrictions on the dependence of the time series in terms of the autocovariance sequence or in terms of the spectral density, since all the higher order cumulants are zero. Also, this will enable us to concentrate on departures from normality due exclusively to the nonparametric variance estimation.

In the next section of this Introduction we present the topic of nonparametric estimation of the spectral density and of the variance in linear regression and more general models. Then we review in Section 1.4 the main higher order asymptotic methods, focusing in time series applications. The literature survey is not intended to be exhaustive, its aim being just to provide examples of the main developments in each context.

Nonparametric estimation depends heavily on the choice of a smoothing number. This topic is introduced in next section and elaborated further in Section 1.5. In Chapter 4 of this thesis we propose a new cross-validatory technique for spectrum bandwidth choice at a particular frequency, valid assuming only local conditions on the spectral density.

Situations where (1.2) does not hold have also been considered in the literature. In Section 1.6 we focus on long range dependence time series models for which $f(0) = \infty$ and on semiparametric inference for them. Then, in Chapter 5 we discuss some topics about the log-periodogram estimate of the memory parameter for long range dependence time series, extending some previous results available in the literature.

1.3 Nonparametric time series estimation

1.3.1 Spectral density estimation

Nonparametric spectrum estimates have been used for a long time in different sciences. The descriptive characteristics of the spectrum are even more interesting since the estimates for distinct frequencies tend to be nearly statistically independent when the sample size N is large. The estimates of $\gamma(j)$ contain identical information but they do not share that property. Beyond the studentization problem in a variety of frameworks, spectral density estimates can also be employed for detection of hidden periodicities, hypothesis testing, discrimination and classification, model identification, parameter estimation, prediction and smoothing, and seasonal adjustments. Brillinger (1975) and Robinson (1983) contain descriptions of many applications of spectral density estimation.

We consider nonparametric estimates of the spectral density of the quadratic type. These estimates are quadratic forms of the observed stretch of data, and can be written in the time domain in terms of weighted sums of the sample autocovariances,

$$\widehat{\gamma}(j) = \frac{1}{N} \sum_{1 \leq t, t+j \leq N} (X_t - \overline{X})(X_{t+j} - \overline{X}), \qquad j = 0, \pm 1, \ldots$$

and in the frequency domain as averages of the periodogram,

$$I(\lambda) = \frac{1}{2\pi N} \left| \sum_{t=1}^{N} (X_t - \overline{X}) e^{it\lambda} \right|^2 .$$

Consider a real, even, bounded and integrable function $K(\lambda)$, satisfying

$$\int_{-\infty}^{\infty} K(\lambda) d\lambda = 1,$$

and define

$$K_M(\lambda) = M \sum_{j=-\infty}^{\infty} K(M[\lambda + 2\pi j]),$$

where M is called a "lag" or "bandwidth" number. Then the weighted autocovariance estimates of $f(\lambda)$ are given as

$$\widehat{f}_C(\lambda) \;=\; \int_{-\pi}^{\pi} K_M(\lambda - \alpha) I(\alpha) d\alpha = \frac{1}{2\pi} \sum_{j=1-N}^{N-1} \omega\left(\frac{j}{M}\right) \widehat{\gamma}(j) \cos j\lambda,$$

where

$$\omega(x) = \int_{-\infty}^{\infty} K(\lambda) e^{ix\lambda} d\lambda.$$

For the same K and M, $\widehat{f}_C(\lambda)$ is typically closely approximated by a discrete average of periodogram ordinates

$$\widehat{f}_P(\lambda) = \frac{2\pi}{N} \sum_{j=1}^{N-1} K_M(\lambda - \lambda_j) I(\lambda_j) \tag{1.18}$$

where $\lambda_j = 2\pi j/N$ are the Fourier frequencies, and because these $I(\lambda_j)$ are invariant to location shift in X_t no mean correction of X_t is necessary in (1.18).

The choice of M and K has been extensively discussed, mainly from the standpoint of bias and variability. For given K, variances tend to increase with M. On the other hand, when estimating $f(0)$, a small M lead to $K_M(\lambda)$ not being heavily concentrated around the origin, and bias from the influence of frequencies near zero, most likely negative bias if there is a spectral peak there. For given M, a similar dilemma is faced in the choice of K. Distant frequencies can also cause bias. Many K produce side-lobes in K_M, so coincidence of a side-lobe with a large spectral peak in $f(\lambda)$ can give an inflated estimate of $f(0)$.

On computational grounds ω having compact support in $\widehat{f}_C(\lambda)$, $\omega(x) = 0$, $|x| > 1$, is desirable because then only about M $\widehat{\gamma}(j)$ need be computed, whereas the (contradictory) practice of choosing K to have compact support in $\widehat{f}_P(\lambda)$, so $K(\lambda) = 0$, $|\lambda| > \pi$, is desirable because then only about $N/2M$ of the $I(\lambda_j)$ need be computed. On the other hand, all $N-1$ $\widehat{\gamma}(j)$ and $N/2$ $I(\lambda_j)$ can be rapidly computed via the fast Fourier transform, so these considerations are of minor importance. In Chapter 2 we employ compact support ω, mostly for convenience, thus avoiding additional conditions on the tails of ω. We justify compact support kernels K in Chapters 3, 4 and 5, in order to avoid leakage from other frequencies when we assume only local conditions on the spectral density.

The desirable requirement of non-negativity of a variance estimate is implied if

$$K(\lambda) \geq 0, \quad \text{all } \lambda.$$

Several elaborations of $\widehat{f}_C(\lambda)$ and $\widehat{f}_P(\boldsymbol{\lambda})$ have been proposed in the spectrum estimation literature, like tapering and prewhitening. Tapering (Tukey (1967)) multiplies X_t by a sequence which decays smoothly to zero at $t = 1$ and $t = N$ in order to reduce the effect of contamination of $I(\lambda)$ from other frequencies. Prewhitening (Press and Tukey (1956)) entails fitting a preliminary AR to X_t, forming $\widehat{f}_C$ or $\widehat{f}_P$ from the residuals, and then multiplying by the AR transfer function. This recognizes that a quadratic spectrum estimate may not be very good at fitting a sharp peak, such as indeed appears at zero frequency in many empirical series. In fact pure AR spectrum estimation, without the kernel smoothing involved in $\widehat{f}_C$ or $\widehat{f}_P$, became popular, see Burg (1975), Parzen (1969); here the AR order replaces M as the bandwidth and is regarded as increasing slowly to infinity with N in the asymptotics. Mixed ARMA models have also been used, as well as many other nonquadratic estimators, with the idea of obtaining higher resolution techniques (see Robinson (1983), Section 3).

In the operational research literature different quadratic estimators of the variance of the sample mean have been justified under different considerations. For a recent review see Song and Schmeiser (1993).

Brillinger (1973) discussed the problem of inference on the mean of a continuous time process, which is either observed continuously or at possibly unequally-spaced times that are either finite or random. He suggested splitting the observation interval $(0, N)$ into m disjoint subseries of lengths $\ell = N/m$, showed that for fixed m the sample means of each stretch are asymptotically independent and identically distributed as $N \to \infty$, and then used the sample variance of the m sub-series means to estimate the overall sample mean. Carlstein (1986) and Künsch (1989) adapted Brillinger's subsamples method to more general statistics, showing the consistency of a "block" jackknife under conditions on the number and size of the subsamples. These estimates are quadratic functions of the data as well, and in some cases, special versions of $2\pi\widehat{f}_C(0)$.

The asymptotic statistical properties of quadratic spectrum estimates have been studied extensively. Both types of estimates considered, $\widehat{f}_C$ and $\widehat{f}_P$, have similar characteristics. If M is increasing slowly with respect to N, $\widehat{f}(\lambda)$ is a mean square consistent estimate of $f(\lambda)$. Furthermore $\sqrt{N/M}\left(\widehat{f}(\lambda) - f(\lambda)\right)$ converges to a normal distribution under suitable conditions on the dependence structure of the time series and the lag number M (see for example Hannan (1970)).

Given the slow rate of convergence of nonparametric spectrum estimates, $\sqrt{N/M}$, alternative approximations for the distribution of such estimates have been pursued. These approximations can be based on χ^2 distributions (Brillinger (1975)) or on asymptotic expansions (Bentkus and Rudzkis (1982)), but also more nonparametric approaches have been considered. Thus, Franke and Härdle (1992) obtained a bootstrap approximation for the distribution of estimates of discrete periodogram average type, and Politis and Romano (1992) modified the block bootstrap for dependent data of Künsch (1989) to cover smoothed spectral estimates of $\widehat{f}_C$ type.

For the studentization of the sample mean we consider nonparametric estimates of the form $\widehat{f}_C(\lambda)$ at $\lambda = 0$, which can be written down easily as quadratic forms in the vector of observations. Our high order results for this nonparametric estimate are an extension of those of Bentkus and Ruzdkis (1982). The main contribution is that we do not require boundedness of the spectral density for all frequencies, allowing for a much more general class of dependence characterizations.

There are several references about spectral estimation under mild smoothness conditions. Bentkus (1985) and Rudzkis (1985) analyzed the mean squared minimax asymptotic risk and the distribution of the maximum deviation for related (unfeasible) estimates, assuming only $f(\lambda) \in L_2$. Zhurbenko (1984) considered the properties of estimates of the spectral density obtained by a time shift under a perturbation of the spectrum caused by a blip or pulse at a remote frequency. Estimation of the spectrum for long memory time series (with a pole at the origin; see Section 1.6) has received also some attention in recent years: Hidalgo (1994) considered the estimation of $f(\lambda)$, $\lambda \neq 0$, under a linear process condition and Soulier (1993) discussed the same problem for Gaussian fields. See also von Sachs (1994b) for related references about nonparametric spectral estimation in the presence of peaks.

1.3.2 Variance estimation

Eicker (1967) proposed for the estimation of the variance of the vector of OLSE in linear regression the following estimate of T,

$$\widehat{T}_1 = N^2 \sum_{j=1-M}^{M-1} \frac{\widehat{\gamma}(j)\,\widehat{c}(j)}{N-|j|}, \tag{1.19}$$

where

$$\widehat{c}(j) = \frac{1}{N} \sum_{1 \le t, t+j \le N} Z_t Z'_{t+j}$$

and for the calculation of $\widehat{\gamma}(j)$ we use the OLS residuals $\widehat{X}_t = Y_t - \widehat{\beta}' Z_t$, and M has analogous interpretation as before, increasing slowly with N in the asymptotics. $\widehat{T}_1$ is suggested by applying Parseval's equality,

$$S = \sum_{j=-\infty}^{\infty} R(j)\gamma(j),$$

and then inserting sample estimates and truncating the sum. To guarantee a positive semi-definite $\widehat{T}$ it is possible to introduce suitable weights in (1.19), corresponding to weights guaranteeing a nonnegative estimate of $f(0)$. This is best seen in the frequency-domain version

$$\widehat{T}_2 = (2\pi)^2 \sum_{j=1}^{N} I_Z(\lambda_j) \widehat{f}(\lambda_j),$$

considered for example by Hannan and Robinson (1973), where $I_Z(\lambda_j)$ is the periodogram matrix of Z_t and $\widehat{f}(\lambda_j)$ is a smoothed estimate of $f(\lambda)$. This estimate is valid for both stochastic and nonstochastic regressors.

However, the methods used to robustify against autocorrelation many more complicated statistics, specially methods used recently in econometrics, can be seen as extensions of those proposed much earlier for the mean by such authors as Jowett (1955), Hannan (1957) and Brillinger (1979). In the linear regression framework with Z_t stationary and ergodic, we can estimate T by

$$\widehat{T}_3 = 2\pi N \, \widehat{f}_U(0), \tag{1.20}$$

where $\widehat{f}_U(0)$ is an smoothed estimate of the spectral density matrix of $U_t = Z_t X_t$, computed with X_t replaced by the OLS residuals $\widehat{X}_t$. This approach has been stressed in the econometric literature and it has the advantage over $\widehat{T}_1$ and $\widehat{T}_2$ of not requiring independence, at least up to fourth moments, between Z_t and X_t, but only that Z_t and X_t are uncorrelated. On the other hand, it makes less use of the structure of the model if such independence is reasonable, when it might be expected to possess inferior finite-sample properties.

In the more general case of (1.10) we can estimate the covariance matrix of the asymptotic distribution by

$$\left(\frac{1}{N}\sum_{j=1}^{N} h(\lambda_j)I_Z(\lambda_j)\right)^{-1} \frac{2\pi}{N}\sum_{j=1}^{N} \hat{f}(\lambda_j)h^2(\lambda_j)I_Z(\lambda_j) \left(\frac{1}{N}\sum_{j=1}^{N} h(\lambda_j)I_Z(\lambda_j)\right)^{-1}$$

in the same spirit as $\widehat{T}_2$, and when $h(\lambda) = \hat{f}^{-1}(\lambda)$,

$$\left(\sum_{j=1}^{N} \hat{f}(\lambda_j)^{-1}I_Z(\lambda_j)\right)^{1/2} (\hat{\beta}(\hat{f}^{-1}) - \beta)$$

is asymptotically multivariate standard normal under regularity conditions.

Moving to minimum distance estimates and in a similar fashion, we can estimate Ψ in (1.14) by

$$\widehat{\Psi}(\Phi) = \frac{1}{N}\sum_{j=1}^{N} W(-\lambda_j;\hat{\theta})\Phi(\lambda_j)W(\lambda_j;\hat{\theta})',$$

$$W(\lambda) = \frac{\partial}{\partial\theta'}w_H(\lambda;\theta) \qquad \text{or} \qquad W(\lambda) = \frac{\partial}{\partial\theta'}A(\lambda;\theta)w_Z(\lambda).$$

Let $\widehat{\Psi}(\Phi\hat{f}\Phi)$ be defined analogously, with $\hat{f}(\lambda)$ an estimate of $f(\lambda)$ as before, based on residuals. Then $\hat{\theta}$ in (1.13) is approximately

$$\mathcal{N}(\theta,\, N^{-1}\widehat{\Psi}(\Phi)^{-1}\widehat{\Psi}(\Phi\hat{f}\Phi)\widehat{\Psi}(\Phi)^{-1}).$$

This approach was suggested by Hannan and Robinson (1973), Robinson (1976) in special cases where H_t is given by (1.12).

As we have commented before, estimates of the form $\widehat{T}_3$ can alternatively be used as for linear regression, just starting from expression (1.15). This was advocated by White and Domowitz (1984), though they suggested using the truncated version of $\hat{f}_C$. This approach has been stressed in the bulk of subsequent relevant econometric literature, in which special names have been invented for the topic, such as "autocorrelation-consistent variance estimation", "heteroskedasticity and autocorrelation-consistent variance estimation" and "long run variance estimation". Then in (1.17), the $f_u(0)$ in D can be estimated as before, from the $U_t(\hat{\theta})$, while C can be estimated by a sample average. Also optimal matrices S can be approximated in a similar way.

Particular cases of, or modifications of, the various spectrum estimates have been proposed in the econometric literature, concerning the selection of weights, functions K

and prewhitening, among others. The dependence conditions stressed in the econometric literature have been mixing conditions, with rate conditions on the mixing numbers. These are attractive because the $U_t(\theta)$ may be complicated, nonlinear functions of underlying variables, but inherit the mixing properties of these; this is not the case if the underlying variables are linear filters of white noise. Also mixing conditions readily allow a degree of non-trending nonstationarity, indicative of forms of asymptotic stationarity as in Parzen (1962).

Our analysis of the studentization of the OLSE in Chapter 2 will be based in an estimate related to $\widehat{T}_1$, using general weights ω with compact support. We include the normalization $(\sum_t Z_t Z_t')^{-1}$ in $\widehat{c}(j)$, instead of $1/N$, to account for possibly nonstationary behaviour in the nonstochastic regressors.

1.4 Higher order asymptotics

Most of the statistical inference is carried out on the basis of asymptotic or large-sample results as approximations for the distribution and properties of estimates and test statistics. This is almost always the case for dependence and nonparametric situations, since the exact results are here even more intractable. Therefore, it is desirable to describe conditions under which the asymptotic approximations are reasonable and to obtain alternative methods when the asymptotic approximations break down for reasonable finite sample sizes. These approximations, generally of density and distribution functions, but also of moments and other quantities of interest, can be used to improve numerical calculations from the data or to evaluate and compare theoretically alternate statistical procedures.

One interesting question which readily lends itself to higher-order asymptotic study is the cost of correcting for autocorrelation in estimating $\mathrm{Var}[\overline{X}]$ when none exists. This is a special case of a more general problem, that of over-specifying M in $\widehat{f}_C$ or $\widehat{f}_P$ relative to an actual MA order less than M, or over-specifying the AR order. Albers (1978) considered the case when the prescribed MA or AR order is fixed relative to N, and the observations are uncorrelated. He found that while there is no asymptotic loss of power of (1.6) relative to the ordinary t-ratio, the deficiency measure of Hodges and Lehmann

(1970) - the difference between the numbers of observations required to achieve the same power - is non-zero to order $O(N^{-1})$, and he estimated this deficiency.

Many statistical problems can be indexed by the sample size $N = 1, 2, \ldots$, or other number reflecting the quantity of information available. Consider a sequence of estimators $\widehat{\theta}_N$. In many situations, centred and normalized versions of the estimate $\widehat{\theta}_N$ are asymptotically normal. However, it is possible to develop other approximations and to estimate the asymptotic error.

1.4.1 Asymptotic expansions

There are different methods for obtaining higher-order approximations, most based on Fourier inversion of the approximate characteristic function. Let $\psi(t) = E[\exp\{itY_N\}]$ be the characteristic function for the normalized version of $\widehat{\theta}_N$, Y_N, and $\varphi(t) = \log \psi(t)$ its cumulant generating function. Then, if ψ is integrable, the density function ζ_N of Y_N can be written as

$$\zeta_N(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ixt} \psi(t) dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ixt+\varphi(t)} dt. \tag{1.21}$$

The cumulant generating function $\varphi(t)$ can be expanded in a series where the successive terms are increasing powers of $N^{-1/2}$. Then, expanding the exponential and integrating term by term, an approximation can be found for the density ζ_N.

Barndorff-Nielsen and Cox (1989) constitutes a general introduction to asymptotic theory and asymptotic expansions in particular. Bickel (1974), Phillips (1980), Rothenberg (1984a), Bhattacharya (1987) and Reid (1991) are comprehensive reviews, with different emphasis and many interesting references about both the statistics and econometrics literature.

Different approximations can be obtained, depending on the point t where we expand $\varphi(t)$. The Edgeworth approximation is obtained by expanding $\varphi(t)$ around $t = 0$ and the saddlepoint one around the saddlepoint value $t^\star$ that maximizes the integrand. The latter approximation has an alternative interpretation as an Edgeworth expansion around the mean of a transformed random variable. The saddlepoint approximation typically gives more accurate approximations especially in the tails of the distribution, although requires close knowledge of the cumulant function. Barndorff-Nielsen and Cox (1979), Reid (1988) and Jensen (1995) are detailed accounts for the saddlepoint approximation.

We will focus on the Edgeworth approximation, since it is simpler and more common, and can be calculated from the low order cumulants of Y_N. If $Y_N = N^{1/2}(\widehat{\theta}_N - \theta_0)$ is asymptotically normal distributed with zero mean and variance σ, then, integrating the density in (1.21), the distribution function of $N^{1/2}(\widehat{\theta}_N - \theta_0)$ can be expanded as a power series in $N^{-1/2}$,

$$\begin{aligned} &\mathrm{P}\Big\{N^{1/2}(\widehat{\theta}_N - \theta_0)/\sigma \leq x\Big\} \\ &\quad = \quad \Phi(x) + N^{-1/2}p_1(x)\phi(x) + \ldots + N^{-j/2}p_j(x)\phi(x) + \ldots, \end{aligned} \tag{1.22}$$

where $\phi(x) = (2\pi)^{-1/2}e^{-x^2/2}$ is the standard normal density function and

$$\Phi(x) = \int_{-\infty}^{x} \phi(u)du$$

is the standard normal distribution function. Then (1.22) is called an Edgeworth expansion. The functions p_j are polynomials with coefficients depending on the cumulants of $\widehat{\theta}_N - \theta_0$. Usually (1.22) is only available as an asymptotic series, or an asymptotic expansion, meaning that if the series is stopped after a given number of terms then the remainder is of smaller order as $N \to \infty$ than the last term that has been included. Of course, for given sample size N the order of magnitude of the asymptotic error does not tell us anything about the absolute magnitude of the error of the approximation, although it might be a useful indicator of the approximation error for moderate values of N. The terms in $N^{-1/2}$ and in N^{-1} can be described as corrections for skewness and kurtosis, respectively, since for the normal distribution all the cumulants of order bigger than two are zero.

The inverse expansion of (1.22) for the quantiles is denominated Cornish-Fisher expansion. If σ is unknown, it can be substituted by a consistent estimate, obtaining a related expansion. Empirical Edgeworth expansions can be obtained from (1.22), substituting the unknown cumulants (or equivalently, moments) in the functions p_j by sample consistent estimates. Then, it is possible to estimate the magnitude of the stochastic errors for the feasible approximation of the distribution.

This method is a natural extension of traditional large-sample techniques based on the central limit theorem: since the expansions are constructed in terms of the normal and chi-square distributions, the usual asymptotic approximation is the leading term of the Edgeworth expansion. Also, its general availability and simplicity lead to a useful, comprehensive approach to second-order comparisons of alternative procedures (e.g. Akahira and Takeuchi (1981)). This fact has been used extensively in the bootstrap literature for the comparison against first order asymptotics (e.g. Singh (1981), Bhattacharya and Qumsiyeh (1989) and Hall (1992)).

The simplest example to introduce Edgeworth expansions is the sum of independent and identically distributed (i.i.d.) random variables. In the previous references can be found several intuitive expositions on how to arrive to expressions like (1.22) in this case. Proofs are available in general probability texts, like Feller (1971), Chapter XVI, or Petrov (1975), Chapter VI. Bhattacharya and Rao's (1975) monograph presents all the main results for sums of independent random vectors, including the lattice case. The conditions used are mainly stated in terms of regularity of the characteristic function and enough moments. Additionally, some heterogeneity can be allowed.

The theory also applies to smooth functions of such sample sums of i.d.d. variables, as conjectured by Wallace (1958), (e.g. Bhattacharya and Ghosh (1978 and 1989)), including minimum contrast estimators (e.g. Pfanzagl (1973)), estimates of regression models with i.i.d. errors (Qumsiyeh (1990)), discrete Fourier transforms of i.i.d. sequences (Chen and Hannan (1980)) and U-statistics (e.g. Bickel et al. (1986)). In a econometric context the validity of Edgeworth expansions was proved by Sargan (1976) and Phillips (1977b), completing some previous ideas of Chambers (1967).

Other parameters different from the sample size N can index the sequence of problems. For example, in nonparametric estimation of the probability density function for i.d.d. random variables, the index is constructed in terms of Nh, where $h = h_N$ is a sequence of numbers tending to zero as N increases (Hall (1991)). This is also the case in nonparametric estimation of the spectral density and most smoothed estimates.

However, in semiparametric frameworks where we estimate nonparametrically a nuisance function, often the asymptotic distribution and the rate of convergence are not affected by the bandwidth, like in the studentization problem, where only a consistent estimate is required (see Robinson (1995a)). Nevertheless, it appears that the presence of slowly converging nonparametric estimates in $N^{1/2}$-consistent estimates might lead to inferior higher-order asymptotic properties than those of estimates based on a correct parametric model (see, e.g., Linton (1996)).

Varied asymptotic expansions and higher order results of different type are possible. They include Gram-Charlier and Laplace expansions, Berry-Esseen bounds and large deviation approximations. (See for example Phillips (1980) for a neat introduction and more references).

1.4.2 Time series higher order asymptotic theory

Many of the above results can be extended to weak dependence situations. The more general result in this field is due to Götze and Hipp (1983), where earlier related references can be found. They established the validity of an Edgeworth expansion for sums of weakly dependent random vectors (see Lahiri (1993) and the references therein, for recent extensions). The essential regularity conditions required relate with the dependence structure of the time series, expressed in terms of the strong mixing coefficients (Rosenblatt (1956)).

Under equivalent conditions to those of Götze and Hipp, Lahiri (1991, 1994) and Götze and Künsch (1995) compared approximations based on bootstrap methods and Edgeworth expansions for the distribution of time series statistics that can be approximated by sample means. Bentkus et al. (1995) is a recent reference for Berry-Essen bounds in dependence situations. Similar ideas have been applied to approximate the distribution of various statistics. Bose (1988) considered the case of the sample autocovariances and Janas (1993) and Janas and von Sachs (1993) applied his results to functions of the periodogram of stationary time series.

The validity of Edgeworth expansions for a general class of statistics in time series models is analyzed in Götze and Hipp (1994). They extended some results of Taniguchi (1983, 1984 and 1986) concerning Gaussian ARMA models estimators and test statistics. Taniguchi (1991) is a complete survey of higher order asymptotic theory for time series models and contains many references of interest up to 1991. Section 2.5 of Taniguchi

(1991) review the literature for linear time series regression estimation, always assuming a parametric covariance matrix for the observations.

Expansions connected with the saddle-point approximation and suitable for time series statistics are described in Durbin (1980a and 1980b), extending some earlier results of Daniels (1956) for the serial correlation coefficient. Related problems are studied under different sets of conditions in Phillips (1977a) and Kakizawa (1993).

1.5 Bandwidth choice in nonparametric serial dependence estimation

Most of the nonparametric methods that have been used in statistics to allow for serial dependence are appealing because they are justifiable in the absence of precise assumptions about the dependence, which can be very hard to motivate. Nonparametric estimation is involved in one way of another in the previous estimates of the variance we have considered, so discussion on the problem of selecting bandwidth numbers is important. In nonparametric estimation of the spectral density and other variance quantities, the user-chosen parameter M controls the smoothness, and therefore the statistical properties of the estimate. The asymptotics require that this number (proportional to the inverse of the actual bandwidth of K_M) is increasing at a suitable rate with the sample size, but they give not practical guidance when N is finite.

Indeed, the distinction between "nonparametric" and "parametric" estimation resides principally in their interpretation in large samples, and in both cases a particular functional form has to be chosen, the same one may be used in either case, and the outcome of inferences will be dependent, possibly greatly so, on the choice of the functional form. For example, $\widehat{f}_C$ and $\widehat{f}_P$ depend on K and M, AR spectrum estimates depend on AR order, and prewhitened quadratic estimates depend on them all.

Consider the choice of M in $\widehat{f}_C$ or $\widehat{f}_P$. Because bias tends to vary inversely with M, and variance tends to vary directly, minimization of mean squared error (MSE) $E[(\widehat{f}(\lambda)-f(\lambda))^2]$ was proposed by Grenander and Rosenblatt (1957) as a simple criterion for producing a balancing M. Under regularity conditions

$$\lim_{N\to\infty} N^{2v}E[(\widehat{f}(\lambda)-f(\lambda))^2] = \frac{f^2(\lambda)}{c}\int_{-\infty}^{\infty}\omega^2(x)dx\{1+I(\lambda=0)\} + c^{2v}[\kappa^{(v)}f^{(v)}(\lambda)]^2, \tag{1.23}$$

where

$$c = \lim_{N\to\infty}\frac{N^{1/(1+2v)}}{M}, \tag{1.24}$$

$$\kappa^{(v)} = \lim_{x\to 0}\frac{1-\omega(x)}{x^v}, \tag{1.25}$$

$$f^{(v)}(\lambda) = \frac{1}{2\pi}\sum_{j=-\infty}^{\infty}|j|^v\gamma(j)e^{-ij\lambda}, \tag{1.26}$$

and v is the largest real number for which c and $\kappa^{(v)}$ are assumed finite. Often $v = 2$, though for the Barlett window $v = 1$. Then, in view of (1.24) we should choose M proportional to $N^{1/(1+2v)}$, and minimizing (1.23) with respect to c gives the optimal choice of M,

$$M = N^{1/(1+2v)}/c^\star,$$

where

$$c^\star = \left[\frac{f^2(\lambda)\int\omega^2(x)dx\{1+I(\lambda=0)\}}{2v[\kappa^{(v)}f^{(v)}(\lambda)]^2}\right]^{1/(1+2v)}.$$

In practice, though v can be reasonably picked as the largest value such that (1.25) is finite, (i.e. trusting that (1.26) is finite) there is a strong element of circularity in that $c^\star$ depends on $f(\lambda)$ itself, and also on $f^{(v)}(\lambda)$. This problem is standard in the smoothed estimation of nonparametric functions, and a standard proposal to deal with it is to replace f and $f^{(v)}$ by "pilot" estimates based on either a simple parametric model or on an initial choice of bandwidth, in the hope that $M^\star$ will not be too sensitive to the design of the pilot estimates, though it can be very hard to accurately estimate $c^\star$. Andrews (1991) has developed this approach in case of pilot AR spectrum estimates, and also showed that the eventual spectrum estimates are still consistent in the presence of the data-dependent M. Newey and West (1994) justified the optimality of methods using an initial choice of bandwidth. Bülhmann (1995) considered a similar approach and designed an iterative procedure to reduce the high variability of these methods, typical of bandwidth choice algorithms.

For estimates of the variance, say in linear regression, of the form $\widehat{T}_1$ or $\widehat{T}_2$, f has to be estimated across all the Nyquist band $(-\pi,\pi]$. It would be possible to determine suitable

M as in the previous paragraph, in a frequency-dependent way. A simpler approach is to obtain a single M which reflects characteristics of the data across all frequencies. To seek an optimal choice, consider the integrated MSE

$$\int_{-\pi}^{\pi} E[(\widehat{f}(\lambda) - f(\lambda))^2]\chi(\lambda)d\lambda, \tag{1.27}$$

for a weight function $\chi(\lambda)$. Lomnicki and Zaremba (1957) suggested $\chi(\lambda) \equiv 1$, and Jenkins and Watts (1968) suggested $\chi(\lambda) = 1/f^2(\lambda)$ because $\widehat{f}(\lambda)$ has asymptotic variance proportional to $f(\lambda)^2$. We minimize (1.27) asymptotically by

$$M^* = \left[\frac{2vN(\kappa^{(v)})^2 \int_{-\pi}^{\pi} f^{(v)}(\lambda)^2\chi(\lambda)d\lambda}{\int_{-\pi}^{\pi} f^2(\lambda)\chi(\lambda)d\lambda \int_{-\infty}^{\infty} \omega^2(x)dx}\right]^{1/(1+2v)}. \tag{1.28}$$

Again this can be estimated using pilot estimates of f and $f^{(v)}$. Hurvich (1985), Beltrao and Bloomfield (1987) and Robinson (1991) considered a fully automatic cross validation method. One version of this is as follows. Introduce the leave-two-out spectrum estimate

$$\widehat{f}_P^{(j)}(\lambda_j) = \frac{2\pi}{N} \sum_{\ell=1,\neq\{j,N-j\}}^{N-1} K_M(\lambda_j - \lambda_\ell)I(\lambda_\ell),$$

and the pseudo log-likelihood criterion

$$\sum_{j=1}^{N} \left\{\log \widehat{f}_P^{(j)}(\lambda_j) + I(\lambda_j)/\widehat{f}_P^{(j)}(\lambda_j)\right\}.$$

Then, Robinson (1991) showed that M minimizing this is consistent for M^* in (1.28) with $\chi(\lambda) = 1/f^2(\lambda)$ and extended the results about $\widehat{f}$ for a wider class of situations, including statistics related to $\widehat{T}_2$.

Franke and Härdle (1992) proposed a bootstrap approximation for the optimal bandwidth for the discrete periodogram average spectral estimates $\widehat{f}_P$. See the survey by Hannan (1987) for AR order selection in AR spectrum estimation.

The previous cross-validation methods select a global bandwidth for all the range of frequencies $[-\pi, \pi]$ or for a fixed non-null subset of it. In Chapter 4 of this thesis we propose a modified version of cross-validation to justify a local bandwidth choice for a single frequency, as it is relevant, for example, in the case of the sample mean studentization or $\widehat{T}_3$. Concentrating in a single frequency, we only need to use local smoothness properties of the spectral density of the time series around this frequency, allowing for a broader range of dependence models.

1.6 Long memory time series

For the inference using the sample mean we have assumed that $f(0)$ was bounded above (and away from zero), but there has been considerable interest in the possibility that $f(0) = \infty$, though $f(0) = 0$ has also been considered. Both possibilities are covered by the model

$$f(\lambda) \sim G\,\lambda^{-2d}, \quad \text{as } \lambda \to 0^+, \quad -1/2 < d < 1/2, \tag{1.29}$$

for $0 < G < \infty$. The case $d = 0$ covers the bounded $f(0)$ situation. When $0 < d < 1/2$, the case of most interest, we say that there is long range dependence. When $-1/2 < d < 0$ there is antipersistence. Condition (1.29) is closely related to

$$\gamma(j) \sim g\,j^{2d-1}, \quad \text{as } j \to \infty, \tag{1.30}$$

where

$$g = 2G\,\Gamma(1-2d)\cos\pi(d+1/2),$$

so $\gamma(j)$ is not summable when $d > 0$. Examples of (1.29) and (1.30) are given by the fractional ARIMA models, in which

$$f(\lambda) = \frac{\sigma^2}{2\pi}\left|1-e^{i\lambda}\right|^{-2d}\left|\frac{a(e^{i\lambda})}{b(e^{i\lambda})}\right|^2, \quad -\pi < \lambda \leq \pi, \tag{1.31}$$

where $\sigma^2 > 0$ and a and b are polynomials of finite degree having no zeros in or on the unit circle, and the fractional noise model where

$$\gamma(j) = \frac{\gamma_0}{2}\left(|j+1|^{2d+1} - 2|j|^{2d+1} + |j-1|^{2d+1}\right), \quad j = \pm1, \ldots \tag{1.32}$$

Long memory stationary time series models have been described in many sciences for some time now. They characterize strongly dependent stochastic processes where the observations are quite dependent even from large lags, since the autocovariances decay only hyperbolically, and not exponentially, like in ARMA models. The rate of decay of the autocovariances is determined by the parameter d, which determines the long range properties of the time series. Justification, statistical properties of inference methods, and an extensive literature for long memory time series can be found in Robinson (1994a) and Beran (1994).

For a covariance stationary sequence $\{X_t\}$, with mean μ and spectrum and autocovariance satisfying (1.29) and (1.30), $\overline{X}$ is no longer asymptotically a BLUE of μ when $d \neq 0$, see Adenstedt (1979). Samarov and Taqqu (1988) found that for $-1/2 < d < 0$ the efficiency can be poor, though for $0 < d < 1/2$ it is at least 0.98. In any case $\overline{X}$ is still unbiased and consistent for μ and is a computationally simple candidate for use in inference, especially in the Gaussian case.

Under (1.30),

$$\mathrm{Var}[\overline{X}] \sim \frac{g\,N^{2d-1}}{2d(d+1/2)}, \tag{1.33}$$

so $\overline{X}$ is less than $N^{1/2}$-consistent when $0 < d < 1/2$. Hence, when $\overline{X}$ is asymptotically normal, the studentization proposed previously will not produce consistent interval estimates or asymptotically valid hypothesis tests if $d \neq 0$. In practice d is unknown. Given the asymptotic normality of $\overline{X}$, and (1.33), Robinson (1994a) observed that

$$N^{1/2-\widehat{d}}\left[\frac{2\widehat{d}(\widehat{d}+1/2)}{2\,\widehat{G}\,\Gamma(1-2\widehat{d})\cos\pi(\widehat{d}+1/2)}\right]^{1/2}\left(\overline{X}-\mu\right) \to_d \mathcal{N}(0,1)$$

if the estimates $\widehat{G}$, $\widehat{d}$, satisfy

$$\widehat{G} \to_p G, \quad \log N\left(\widehat{d}-d\right) \to_p 0.$$

Robinson (1994b, 1995b, 1995c) has verified these properties for three different types of estimate, all of which are based only on local assumptions for $f(\lambda)$ near $\lambda = 0$, and which do not require parameterization of f across all frequencies. A corrected specified parametric $f(\lambda)$ or $\gamma(j)$ (as in (1.31) and (1.32)) can yield $N^{1/2}$-consistent estimates (e.g. Fox and Taqqu (1986) and Dahlhaus (1989)) and Beran (1989) considered studentization based on such estimates, with an alternative type of approximate distribution to that arising in (1.33). However, if $f(\lambda)$ is incorrectly specified, such estimates will be inconsistent, indicating a cost to modelling high frequency behaviour in a situation in which only low frequency behaviour is of real interest.

Further models and techniques in a long range dependence environment, including OLSE, GLSE and M-estimates of linear regression and nonlinear models, are described in Robinson and Velasco (1996).

Robinson (1995b) shown that a modified version of the log-periodogram estimate of d proposed by Geweke and Porter-Hudak (1983), was consistent and asymptotic normal for Gaussian sequences. Taking logarithms in (1.29),

$$\log f(\lambda) \sim \log G - 2d \log \lambda, \qquad \text{as} \quad \lambda \to 0^+,$$

and introducing the periodogram for Fourier frequencies λ_j close to zero, $j = 1, \ldots, m$, with m increasing in the asymptotics but slower than N, we can obtain

$$\log I(\lambda_j) \approx \log G - 2d \log \lambda_j + \log \frac{I(\lambda_j)}{f(\lambda_j)}, \qquad j = 1, \ldots, m.$$

Believing that the last term in the previous expression is approximately i.i.d. as in the weak dependence case, the estimate of d proposed by Geweke and Porter-Hudak is the ordinary least squares estimate in the linear regression between the logarithm of the periodogram and (a regressor equivalent to) $-2\log\lambda$ for frequencies in a degenerating band around the origin. However the periodogram for frequencies close to the origin does not behave in the same fashion for long range than for weak dependence series, and Robinson (1995b) proposed a trimming of frequencies too close to the origin, following the proposal of Künsch (1986). He also used a pool of contributions from periodogram ordinates at adjacent frequencies.

Two questions arise at this point: the substitution of the periodogram ordinates by consistent, nonparametric smoothed estimates of the spectral density, and the validity of the estimation procedure for non Gaussian time series. In Chapter 5 of this thesis we obtain the consistency of the log-periodogram estimate for linear time series, non necessarily Gaussian, for both situations: with and without smoothing. Again, only local conditions on f are needed, so high frequency behaviour modelling is avoid.

Chapter 2

Edgeworth expansions for time series linear regression

2.1 Introduction

Autocorrelation consistent estimation of covariance matrices of parameter estimates in linear and nonlinear models has been a growing field of research in recent years in econometrics and time series (e.g. Hansen (1982), White and Domowitz (1984), Newey and West (1987 and 1994), Andrews (1991), Robinson (1991). See Robinson and Velasco (1996) for a review). Since statistical inference is typically based on central limit theorems, a consistent estimate of the variance of the distribution is necessary. For time series statistics this variance estimation requires explicit consideration of the dependence, and nonparametric set-ups have been stressed in much of the literature.

Most of the techniques proposed relate with ideas developed in spectral analysis since the seminal works of Jowett (1954) and Hannan (1957). They were concerned with the variance of the sample mean of a stationary process, but their work extends to the estimation of the covariance matrix of the least squares estimator in a linear regression model with serial dependent errors of unknown form. And as we have seen in the Introduction of this thesis, this framework can cover also nonlinear regressions and many more complicated models. An alternate approach is due to Eicker (1967), who proposed an estimate of convolution type for linear regression, using the special structure of the model.

This chapter analyzes higher order properties of Eicker's class of estimators of the variance of least squares estimates in a time series linear regression model. These estimates are defined in terms of a weighting scheme for the sample autocovariances of observed least squares residuals and on the choice of a lag or smoothing parameter. Little is known about how the use of nonparametric estimates affects the distribution of the studentized estimator. Although the asymptotic distribution remains unaffected, small sample properties may deteriorate, since, typically, nonparametric estimates have slower rates of converge than parametric ones. This is reflected in the higher order terms of the Edgeworth expansions for the distributions of the variance and studentized estimates, which depend on the lag-number used in the nonparametric smoothing.

Higher order asymptotic methods have been used in time series analysis and in linear regression for some time now (see for example Taniguchi (1991) and the references there). Among others, Rothenberg (1984b), Toyooka (1986) and Magee (1989) have studied different higher order properties of generalized least squares estimates using parametric estimates of the covariance matrix of the observations. See Taniguchi (1991, Section 2.5) for a general discussion. However, their results rely on parametric assumptions and do not consider the studentization problem. For our purposes most of the previous higher order asymptotic work is not pertinent, as we face two main complications: nonparametric estimation and autocorrelation of unknown form.

A great deal of higher order asymptotic research in nonparametric situations has been based on independent observations. Bentkus and Rudzkis (1982) is a notable exception. They obtained asymptotic expansions and large deviation theorems for the distribution of nonparametric spectral density estimates for Gaussian sequences. We adapt some of their techniques to our situation under Gaussianity assumptions. Next chapter extends some of their results in the context of the studentization of the sample mean of a vector of autocorrelated observations.

Lahiri (1994) obtained an Edgeworth expansion for studentized M-estimates in linear regression, with disturbances satisfying Götze and Hipp's (1983) conditions. He also showed that a block-bootstrap can be second order correct, outperforming the normal approximation. We consider here a similar nonparametric variance estimate for a much more general class of regressors and dependence structures, but our distributional assumptions are significantly stronger. Under Götze and Hipp's (1983) assumptions it has not been possible yet to obtain expansions for smoothed statistics which estimate parameters of the whole distribution of a stationary sequence, like variances of estimates of regression coefficients or spectral estimates.

We first present our main definitions and assumptions, and some auxiliary results, in next section. Then, in Section 2.3 we analyze the asymptotic properties of a nonparametric estimate of the variance of least squares estimates for linear regression. In Section 2.4 we consider the joint distribution of the least squares and variance estimates. Then, we study the properties of the studentized estimate in Section 2.5. First, we concentrate on the studentization problem in a model with only one regressor. Later we cover the situation with two regressors (probably one of them an intercept) and from there, the extension for a general multiple regression seems straightforward, but notationally involved. All the proofs can be found at the end of the chapter in three Appendices, together with the details of the bivariate regression in a final Appendix.

2.2 Assumptions and definitions

We assume Gaussian (unobservable) disturbances and nonstochastic regressors, satisfying Grenander and Rosenblatt (1957) conditions, with some modifications. Thus, we allow for trending and other types of nonstationary regressors. Eicker's (1967) nonparametric convolution estimates that we consider are analogous to the class of weighted autocovariance estimates of spectral density, and therefore can be written as a continuous average of the periodogram of the observed least squares residuals.

Some restrictions are needed on a user-chosen bandwidth number for the nonparametric smoothing as well as on the dependence structure of the time series. We formulate them in terms of summability conditions on the autocovariance sequence of the errors, which imply some smoothness properties of the spectral density.

Let the observable sequence $\{Y_t\}$ be given by the following regression model

$$Y_t = \beta Z_t + X_t, \qquad t = 1, \ldots, N,$$

where the regressor $\{Z_t\}$ is an observable nonstochastic sequence and $\{X_t\}$ is a stationary Gaussian sequence with $E[X_t] = 0$, autocovariance function γ and spectral density $f(\lambda)$,

$$\gamma(j) = E\left[X_t X_{t+j}\right] = \int_{\Pi} f(\lambda) e^{ij\lambda} d\lambda,$$

satisfying $f(\lambda) < \infty$, $\lambda \in \Pi = (-\pi, \pi]$. The analysis of this restricted model will be very helpful for Section 2.6, where we consider in detail the more interesting situation,

$$Y_t = \alpha Z_{1,t} + \beta Z_{2,t} + X_t, \qquad t = 1, \ldots, N.$$

Here, when $Z_{1,t} \equiv 1$, we obtain the typical linear regression with intercept.

Let $\mathbf{X} = (X_1, \ldots, X_N)'$ be a vector of N consecutive observations of X_t. Then $\mathbf{X}$ has a multivariate normal distribution $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_N)$, with covariance matrix

$$[\mathbf{\Sigma}_N]_{r,j} = \gamma(r-j), \qquad r, j = 1, \ldots, N.$$

Denoting $\mathbf{Z} = (Z_1, \ldots, Z_N)'$ and $\mathbf{Y} = (Y_1, \ldots, Y_N)'$ the classical estimate of β is the OLS estimate

$$\widehat{\beta} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y} = \beta + (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X},$$

assuming $\mathbf{Z}'\mathbf{Z} > 0$. Then we have that $(\mathbf{Z}'\mathbf{Z})^{1/2}(\widehat{\beta} - \beta)$ has a normal distribution $\mathcal{N}(0, V_N)$, where

$$V_N = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{\Sigma}_N\mathbf{Z} = \left(\sum_{t=1}^{N} Z_t^2\right)^{-1} \sum_{t=1}^{N}\sum_{t'=1}^{N} Z_t Z_{t'} \gamma(t-t') = \sum_{j=1-N}^{N-1} \gamma(j) R_N(j),$$

and

$$R_N(j) = \left(\sum_{t=1}^{N} Z_t^2\right)^{-1} \sum_{1 \le t, t+j \le N} Z_t Z_{t+j}, \qquad j = 0, \pm 1, \ldots, \pm(N-1).$$

Let us define the centered and standardized variable

$$u_1 = V_N^{-1/2} (\mathbf{Z}'\mathbf{Z})^{1/2} (\widehat{\beta} - \beta),$$

which is distributed as a $\mathcal{N}(0,1)$ variate, and also $\delta_{1,N} = \mathrm{Var}[\widehat{\beta}] = V_N(\mathbf{Z}'\mathbf{Z})^{-1}$.

In a first stage we will consider unfeasible estimates of V_N constructed with the unobserved sequence X_t. We will consider later estimates formulated in terms of the observable residuals given by $\widehat{X}_t = Y_t - \widehat{\beta} Z_t$.

Defining the unfeasible (biased) estimator of the autocovariance function as

$$\widehat{\gamma}(j) = \frac{1}{N} \sum_{1 \leq t, t+j \leq N} X_t X_{t+j}, \qquad j = 0, \pm 1, \ldots, \pm(N-1),$$

we consider the following unfeasible estimate of V_N based on the weighting function ω,

$$\begin{aligned} \widehat{V}_N &= \sum_{\ell=1-N}^{N-1} \omega\left(\frac{\ell}{M}\right) \widehat{\gamma}(\ell) R_N(\ell) \\ &= \frac{1}{N} \sum_{r=1}^{N} \sum_{j=1}^{N} \omega\left(\frac{r-j}{M}\right) X_r X_j R_N(r-j) \\ &= \frac{1}{N} \mathbf{X}' \mathbf{Q}_M \mathbf{X}, \end{aligned}$$

where $\mathbf{Q}_M$ is the $N \times N$ matrix

$$[\mathbf{Q}_M]_{r,j} = Q_M(r-j) \stackrel{def}{=} R_N(r-j) \omega\left(\frac{r-j}{M}\right) = R_N(r-j) \int_{\Pi} K_M(\lambda) e^{i(r-j)\lambda} d\lambda$$

and $K_M(\lambda)$ is a kernel function with smoothing number M, defined as the Fourier transform of $\omega(\frac{j}{M})$ for periodic smoothing in $(-\pi, \pi]$. Here we use the scale-type weights given by the function $\omega(\cdot)$ and by the number M, which in the asymptotics is growing with the sample size N, but slower, to achieve consistency. For the construction of this estimate of V_N we do not require any precise assumption about the asymptotic behaviour of the regressors, and it is consistent under regularity conditions for both stationary and trending or trigonometric regressors.

The estimate $\widehat{V}_N$ is basically the proposal of Eicker (1967) when least squares residuals are used, see $\widehat{T}_1$ in (1.19), here with general weights ω. Given the Gaussianity of X_t, this class of estimates $\widehat{V}_N$ is quite appropriate and easy to handle in our set-up. Estimates of the type $\widehat{T}_3$ (see (1.20)) are not easily tractable with nonstationary regressors.

Define the function $\overline{Q}_M$ as the Fourier transform of Q_M,

$$\overline{Q}_M(\lambda) = \frac{1}{2\pi} \sum_{-\infty}^{\infty} Q_M(j) e^{ij\lambda} = \frac{1}{2\pi} \sum_{-\infty}^{\infty} \omega\left(\frac{j}{M}\right) R_N(j) e^{ij\lambda}.$$

Using the *normalized* periodogram of the regressors $\mathbf{Z}$,

$$\overline{R}_N(\lambda) = \left(2\pi \sum_{t=1}^{N} Z_t^2\right)^{-1} \sum_{t=1}^{N} \sum_{t'=1}^{N} Z_t Z_{t'} e^{i(t-t')\lambda},$$

so that

$$R_N(j) = \int_\Pi \overline{R}_N(\lambda) e^{ij\lambda} d\lambda,$$

it is possible to write $\overline{Q}_M$ as a convolution of K_M and of $\overline{R}_N$,

$$\overline{Q}_M(\lambda) = \int_\Pi K_M(\lambda - \alpha) \overline{R}_N(\alpha) d\alpha.$$

We define the estimates $\tilde{\gamma}(j)$ of the autocovariances and $\tilde{V}_N$ of V_N in the same way as $\hat{\gamma}(j)$ and $\hat{V}_N$ respectively, but using $\hat{X}_t$ instead of X_t. The difference between both estimates of the variance of $\hat{\beta}$ is analyzed in Lemma 2.5.

The function K_M is constructed in the following way. Define, for a sequence of positive integers $M = M_N$ tending to infinity with N but slower, and for an even, integrable function K which integrates to one:

$$K_M(\lambda) = M \sum_{j=-\infty}^{\infty} K(M[\lambda + 2\pi j])$$

so $K_M(\lambda)$ is periodic of period 2π, even and integrates to 1. Then we have

$$\omega(r) = \int_{-\infty}^{\infty} e^{irx} K(x) dx$$

and $\omega(0) = 1$. Positive estimates $\hat{V}_N$ are guaranteed with positive functions K.

We make the following assumptions about the covariance sequence and spectral density of $\{X_t\}$, the asymptotic behaviour of the regressors $\{Z_t\}$ and on the user-chosen functions K and ω. All the limits and error bounds are taken as $N \to \infty$, unless stated otherwise.

Assumption 2.1 *For an integer $d \geq 2$, and $0 < \varrho \leq 1$*

$$\sum_{j=-\infty}^{\infty} |j|^{d+\varrho} |\gamma(j)| < \infty.$$

Assumption 2.2 *The function ω is even, bounded and satisfies $\omega(0) = 1$ and $\omega(x) = 0$ for $|x| > 1$. K is bounded, continuous and integrable.*

Assumption 2.3 *The function ω satisfies for an even integer q, $0 < q \leq d$,*

$$\lim_{|x| \to 0} \frac{1 - \omega(x)}{|x|^q} = \omega_q < \infty.$$

Assumption 2.4 *The regressor sequence* $\{Z_t\}$ *satisfies as* $N \to \infty$

1. $d_N^2 \overset{def}{=} \sum_{t=1}^{N} Z_t^2 \to \infty.$
2. $\max_t \frac{Z_t^2}{d_N^2} = O(N^{-p}), \qquad p \in (0,1].$
3. $R_N(j) = d_N^{-2} \sum_{1 \le t, t+j \le N} Z_t Z_{t+j} \to \rho(j),$

where

$$\rho(j) = \int_{\pi}^{\pi} e^{ij\lambda} dG(\lambda),$$

and $G(\lambda)$ *is a function with nonnegative increments, continuous from the right, satisfying* $G(\pi) - G(-\pi) = 1$. *Further, we assume for the same* p,

$$R_N(j) = \rho(j) + O(N^{-p/2} + j\, N^{-p}), \qquad \textit{uniformly for } j = o(N^p).$$

Assumption 2.5 *We assume*

$$\int_{\Pi} f(\lambda) dG(\lambda) > 0.$$

From Assumption 2.1, $f(\lambda)$ has dth derivative, satisfying a uniform Lipschitz condition of order $\varrho > 0$. We use it in all its strength to evaluate the bias of the nonparametric estimate. In fact, for most of our analysis, continuity of $f(\lambda)$ or weaker conditions on the summability of the autocovariance sequence are enough. For all our results concerning the estimation of the higher order cumulants, $d = 0$ and $\varrho = 1$ are sufficient conditions (see, e.g., the proof of Lemma 2.4 below).

The bounded support for ω in Assumption 2.2 may be relaxed with enough conditions on its tail behaviour, but it simplifies considerably the proofs and many functions used in practice satisfy that condition. For K we require only very mild restrictions on its tails and continuity. Those imply the integrability and continuity of K_M in $(-\pi, \pi]$ for all M.

Assumption 2.3 deals with the smoothness of the weighting function around the origin and it is employed to estimate the bias of the variance estimate in a similar way as with nonparametric spectral density estimates. Assumptions 2.2 and 2.3 are satisfied, for example, with $q = 2$ for the Parzen kernel (Priestley (1981), p. 446).

Assumption 2.5 guarantee that the variance of $\widehat{\beta}$ and of the estimates of V_N are well defined after appropriate normalizations. It holds if we assume $f(\lambda) > 0$ for all λ.

In Assumption 2.4 we follow Grenander and Rosenblatt (1957), except that we impose rates in terms of $O(N^{-p})$ for the limits in conditions 2. and 3., instead of assuming simply $o(1)$. The explicit rate in 2. is used to estimate the error in several approximations. Related conditions have been employed in other higher order asymptotic work for regression models, with dependent and independent errors. Our assumption is milder (even with $p = 1$) than the one of Lahiri (1994), who assumed bounded Z_t and

$$\liminf_{N\to\infty} \frac{d_N^2}{N} > 0, \tag{2.1}$$

so

$$\max_t \frac{Z_t^2}{d_N^2} = O(N^{-1}). \tag{2.2}$$

Qumsiyeh (1990) also assumed (2.1) and $\max_t |Z_t| = O(N^\delta)$ for some $\delta \in [0, \frac{1}{2})$, so the left hand side of (2.2) would be $O(N^{2\delta-1})$, with $-1 \le 2\delta - 1 < 0$. However the condition on the $\max_t |Z_t|$ rule out regressors like the polynomials in t, for which δ can be bigger than 1. In this case, if $Z_t = t^n$ for some $n > 0$, say, then,

$$\max_t \frac{Z_t^2}{d_N^2} \sim \frac{N^{2n}}{\frac{N^{2n+1}}{2n+1}} = O(N^{-1}),$$

where $a \sim b$ means that the ratio a/b is tending to one (as $N \to \infty$).

The asymptotic error rate in Assumption 2.4.3 is needed for the proof of Lemma 2.2. For the intuition on this assumption we first introduce two general conditions on the regressors $\mathbf{Z}$ that include many cases of interest, like 'stationary' and polynomial and trigonometric regressions.

Condition 2.1 *We assume that $G(\lambda)$ is absolutely continuous and*

$$\sum_{j=-\infty}^{\infty} |\rho(j)| < \infty,$$

so $G(\lambda)$ has continuous density $g(\lambda)$.

Condition 2.2 *The spectral measure of the regressors, $G(\lambda)$, has jumps and, possibly, an absolute continuous part. We assume that the discrete part of $G(\lambda)$ has a fixed number* *of jumps, each of size* Δ_j *at frequencies* $\lambda_j \in (-\pi, \pi]$, *and the continuous part has a continuous density* $g^c(\lambda)$, *such that,*

$$\sum_j \Delta_j + \int_\Pi g^c(\lambda) d\lambda = 1.$$

When we use Assumption 2.4 we will always assume that either Condition 2.1 or Condition 2.2 holds. Different and intermediate situations are possible, but we concentrate on these two simple cases.

Then for Assumption 2.4.3, first imagine that under Condition 2.1 the regressors behave roughly like the realization of a stationary and ergodic stochastic process with continuous spectral density (see Anderson (1971), pp. 582 and 596). Then, the bound $O(N^{-p/2})$ would be equivalent to the term for the variance of the sample autocovariances (with $p = 1$ under regularity conditions on the moments and the dependence structure of the series; see for example Robinson (1991, pp. 1333-1334) for sufficient conditions). The error bound $O(j\,N^{-p})$ corresponds to the bias term due to the use of biased autocovariances.

In the case of Condition 2.2, this condition is satisfied with $p = 1$ for the trigonometric and polynomial regression. For example, take as before, $Z_t = t^n$ for some $n > 0$. Then

$$\sum_{t=1}^{N} t^n = \frac{N^{n+1}}{n+1} + O(N^n),$$

and therefore, for $j > 0$,

$$\begin{aligned} R_N(j) &= \frac{\sum_{t=1}^{N-j} t^n (t-j)^n}{\sum_{t=1}^{N} t^{2n}} \\ &= \frac{\sum_{t=1}^{N-j} t^n (t^n - n t^{n-1} j + \cdots)}{\frac{N^{2n+1}}{2n+1}(1 + O(N^{-1}))} \\ &= 1 + O(j\,N^{-1}). \end{aligned}$$

From Assumptions 2.1 and 2.4, since $R_N(j)$ converges boundedly to $\rho(j)$ (see for example Hannan (1963b), p. 25), we have that, as $N \to \infty$,

$$V_N \to \sum_{j=-\infty}^{\infty} \gamma(j)\rho(j) = 2\pi \int_\Pi f(\lambda) dG(\lambda) < \infty.$$

Under *Condition 2.1* V_N converges to

$$2\pi \int_\Pi f(\lambda) g(\lambda) d\lambda$$

and under *Condition 2.2* to

$$2\pi \sum_j \Delta_j f(\lambda_j) + 2\pi \int_\Pi f(\lambda) g^c(\lambda) d\lambda. \tag{2.3}$$

Before finishing this section we give two lemmas that will be useful in the subsequent discussion. First we prove a lemma about the behaviour of the spectral window K_M, both in the origin and in the tails, as $M \to \infty$.

Lemma 2.1 *Under Assumption 2.2, as $M \to \infty$,*

$$K_M(0) = M\, K(0) + o(1),$$

and for all $\lambda \in (0, \pi]$, fixed,

$$K_M(\lambda) = o(1).$$

Proof of Lemma 2.1. From the definition of K_M, because K and ω are even,

$$M^{-1} K_M(0) - K(0) = 2 \sum_{j=1}^{\infty} K(2\pi\, M\, j).$$

Since K is integrable, $|K(x)| = O(|x|^{-1-\epsilon})$, $\epsilon > 0$, as $|x| \to \infty$. Then,

$$\sum_{j=1}^{\infty} |K(2\pi\, M\, j)| = O\left(M^{-1-\epsilon} \sum_{j=1}^{\infty} j^{-1-\epsilon} \right) = o(M^{-1}).$$

The argument for $K_M(\lambda)$, $|\lambda| > 0$, is exactly the same, since now as $M \to \infty$, $K(M\lambda) = O(M^{-1-\epsilon}) = o(M^{-1})$. □

Now it is interesting to study the function $\overline{Q}_M(\lambda)$ in the different situations that may arise as a consequence of the type of regressors. First we will prove that $\overline{Q}_M(\lambda)$ converges to

$$\frac{1}{2\pi} \sum_{j=-M}^{M} \omega\left(\frac{j}{M}\right) \rho(j) = \int_\Pi K_M(\lambda - \alpha) dG(\alpha),$$

under the two conditions we consider in Assumption 2.4. Basically we require that the sequence R_N converges to the sequence ρ faster than the kernel function K_M tends to a Dirac's delta as $N \to \infty$.

Lemma 2.2 *Under Assumptions 2.2, 2.4 and $M^{-1} + M^2 N^{-p} \to 0$,*

$$\overline{Q}_M(\lambda) = \frac{1}{2\pi} \sum_{j=-M}^{M} \omega\left(\frac{j}{M}\right) \rho(j) + o(1).$$

Proof of Lemma 2.2. Under the assumptions of the lemma,

$$\left|\sum_{j=-M}^{M} \omega\left(\frac{j}{M}\right) R_N(j) - \sum_{j=-M}^{M} \omega\left(\frac{j}{M}\right) \rho(j)\right| = O\left(\sum_{j=-M}^{M} \left|\omega\left(\frac{j}{M}\right)\right| \left[N^{-p/2} + |j|\, N^{-p}\right]\right)$$
$$= O\left(MN^{-p/2} + M^2 N^{-p}\right) = o(1),$$

and the lemma follows. □

This result is related to *Helly's Theorems* (see for example Kawata (1972), Theorem 9.1.2.) about the weak convergence of $\overline{R}_N(\lambda)$ to $G(\lambda)$ and the limits

$$\lim_{N\to\infty} \int_{\Pi} h(x)\overline{R}_N(x)dx = \int_{\Pi} h(x)dG(x),$$

for continuous functions $h(x)$. We can not use here these results directly, because K_M is not bounded as $N \to \infty$, and even more, $K_M(x)/M$ is depending on N by means of M. Under Condition 2.2 it could seem that a bound $o(M)$ would suffice in Lemma 2.2, but we will employ later this result to approximate the higher order cumulants of $\widehat{V}_N$ in Lemma 2.6 and in that case the bound $o(M)$ is not enough.

Then from Lemma 2.2, if *Condition 2.1* holds, $\overline{Q}_M(\lambda)$ converges to $g(\lambda) < \infty$, by the continuity of g and using standard arguments. Under *Condition 2.2* $\overline{Q}_M(\lambda)$ tends to

$$\sum\nolimits_j \Delta_j K_M(\lambda - \lambda_j) + g^c(\lambda),$$

since K_M is continuous for all M and because g^c is continuous too. From here, we can see that, asymptotically, in the case of a regressor with mixed, continuous and discrete, spectral measure, only the jumps will matter for the analysis of $\widehat{V}_N$, since the continuous part contribution is bounded and the discrete one diverges with M.

One leading example satisfying *Condition 2.2* is when $Z_t \equiv 1, \forall t$. Then $\widehat{\beta}$ is just then sample mean and $G(\lambda)$ has simply a jump of size 1 in the origin, i.e., $dG(\lambda)$ is a Dirac's delta at zero. Therefore $\overline{Q}_M(\lambda)$ is equal to $K_M(\lambda)$ and we are in the same situation as with the nonparametric estimate of $f(0)$. We will consider this framework in the next chapter of this thesis under different conditions.

2.3 Distribution of the estimate of the variance

In this section we investigate the asymptotic distribution of the estimation of the variance $\widetilde{V}_N$. To that end we concentrate first on the unfeasible estimate $\widehat{V}_N$ and then estimate the difference between both estimates as $N \to \infty$.

We begin with a lemma for the bias of $\widehat{V}_N$, which is closely related with the results for nonparametric estimates of the spectral density. Hannan (1958) studied a bias correction for the spectral density estimate after a general form of trend has been removed. When a trend is present in the regressors (Condition 2.2), to studentize the OLS estimate we only need to estimate the spectral density f at the correspondent frequencies if known (see expression (2.3) with $g^c = 0$). However we do not assume explicitly any structure for the regressors.

We focus now on the smoothing bias, and leave the bias due to the use of regression residuals to Lemma 2.5.

Lemma 2.3 *Under Assumptions 2.1, 2.2, 2.3, 2.4 and* $M^{-1} + MN^{-1} \to 0$,

$$E[\widehat{V}_N] - V_N = a_N N^{-1} + b_N M^{-q} + O(M^{-q-\varrho}),$$

where a_N *is the bias from the use of the biased autocovariances*

$$a_N = \sum_{j=1-N}^{N-1} |j|\, \omega\left(\frac{j}{M}\right) \gamma(j) R_N(j),$$

and, denoting by $f^{(d)}$ *the dth derivative of* f, b_N *is the smoothing bias,*

$$b_N = -2\pi\,(-1)^{q/2}\,\omega_q \int_\Pi f^{(d)}(\lambda) \overline{R}_N(\lambda) d\lambda.$$

Next, we study the cumulants of the distribution of $\widehat{V}_N$, as a first step to approximate its distribution. Since $\mathbf{X}$ is normal distributed, the cumulants of $\widehat{V}_N$ are given by the formula for the cumulants of a quadratic form in a vector of normal variables (see for example Kendall and Stuart (1969), p. 357),

$$\text{Cumulant}_s[\widehat{V}_N] = \frac{(s-1)!\,2^{s-1}}{N^s} \text{Trace}\left[(\mathbf{\Sigma}_N \mathbf{Q}_M)^s\right],$$

so we have to study the trace of the matrix $(\mathbf{\Sigma}_N \mathbf{Q}_M)^s$ as $N \to \infty$.

Lemma 2.4 *Under Assumptions 2.1, 2.2, 2.4 and $M^{-1} + M^2 N^{-p} \to 0$, $s = 2, 3, \ldots$,*

$$\mathrm{Trace}\left[(\mathbf{\Sigma}_N \mathbf{Q}_M)^s\right] = N(2\pi)^{2s-1} \int_\Pi f^s(\lambda) \overline{Q}_M^s(\lambda) d\lambda + O\left(\min\left\{M^s, M^2 \left[\sum\nolimits_j |\rho(j)|\right]^{s-1}\right\}\right),$$

where the bound has to be interpreted as $O(M^s)$ if $\sum |\rho(j)|$ diverges and as $O(M^2)$ otherwise, and the leading term is $O(NM^{s-1})$ in the first case and $O(N)$ in the second.

The properties of the leading term follow easily from Lemma 2.2. Now we are in conditions of estimate the cumulants of $\widehat{V}_N$. Let us analyze first the variance of $\widehat{V}_N$. We have that [with Assumptions 2.1, 2.2, 2.4, 2.5 and $M^{-1} + N^{-p}M^2 \to 0$]

$$\begin{aligned} \delta_{N,2} &\stackrel{def}{=} \mathrm{Var}[\widehat{V}_N] = \frac{2}{N^2} \mathrm{Trace}\left[(\mathbf{\Sigma}_N \mathbf{Q}_M)^2\right] \\ &= \frac{2(2\pi)^3}{N} \int_\Pi f^2(\lambda) \overline{Q}_M^2(\lambda) d\lambda + O\left(N^{-2}M^2\right) \end{aligned}$$

and then the variance $\delta_{N,2}$ can be of order N^{-1} (like in a parametric framework) in the smooth $dG(\lambda)$ situation [*Condition 2.1*], since in this case $\overline{Q}_N(\lambda)$ tends to $g(\lambda)$. However, the variance would be only of the 'nonparametric' rate M/N if the function $G(\lambda)$ has jumps [*Condition 2.2*]. In the later case $\overline{Q}_N(\lambda)$ tends to some linear combination $\sum_j \Delta_j K_M(\lambda - \lambda_j)$, plus a term of smaller order of magnitude due to the continuous contribution to G. Here

$$\int_\Pi f^2(\lambda) \left[\sum\nolimits_j \Delta_j^2 K_M(\lambda - \lambda_j)\right]^2 d\lambda \sim M \int K^2(x) dx \sum\nolimits_j \Delta_j^2 f^2(\lambda_j) = O(M),$$

since the cross-products of two kernels at fixed frequencies apart lead to negligible contributions (cf. Lemma 2.1). Therefore, in this case the estimate has a slower rate of convergence, and the constant term of the variance depends on the magnitude of the noise at the frequencies where the spectral measure of the regressors has jumps.

Normalizing the estimate $\widehat{V}_N$ by $\delta_{N,2}^{-1/2}$, the random variable

$$u_2 \stackrel{def}{=} \delta_{N,2}^{-1/2} \left(\widehat{V}_N - \mathrm{E}[\widehat{V}_N]\right),$$

has zero mean and unit variance. In general, we have that [under the same assumptions that for the variance], $s = 2, 3, \ldots$,

$$\kappa_N[0, s] \stackrel{def}{=} \mathrm{Cumulant}_s[u_2] = \delta_{N,2}^{-s/2} \mathrm{Cumulant}_s[\widehat{V}_N]$$

$$
\begin{aligned}
&= \delta_{N,2}^{-s/2}\frac{(s-1)!\,2^{s-1}}{N^s}\mathrm{Trace}\left[(\mathbf{\Sigma}_N\mathbf{Q}_M)^s\right] \\
&= \delta_{N,2}^{-s/2}\frac{(s-1)!\,2^{s-1}(2\pi)^{2s-1}}{N^{s-1}}\int_{\Pi} f^s(\lambda)\overline{Q}_M^s(\lambda)d\lambda \\
&\quad +O\left(\delta_{N,2}^{-s/2}N^{-s}\min\left\{M^s, M^2\left[\sum\nolimits_j|\rho(j)|\right]^{s-1}\right\}\right).
\end{aligned}
$$

Therefore as $N^{1-s}\int_{\Pi} f^s(\lambda)\overline{Q}_M^s(\lambda)d\lambda = O(\delta_{N,1}^{s-1})$ from Lemma 2.2, u_2 has higher order cumulants of order $O(\delta_{N,1}^{(s-2)/2}) = o(1)$, $s = 3, 4, \ldots$, and tends in distribution to a $\mathcal{N}(0,1)$ variate.

Using the previous results for the cumulants of $\widehat{V}_N$ we investigate higher order asymptotic features of the distribution of this unfeasible estimate for large samples. As $\mathbf{X}$ is $N(\mathbf{0}, \mathbf{\Sigma}_N)$ distributed, the characteristic function of the normalized variable, u_2, is,

$$
\psi(t_2) = \left|\mathbf{I} - \frac{2it_2}{N\delta_{N,2}^{1/2}}\mathbf{\Sigma}_N\mathbf{Q}_M\right|^{-1/2}\exp\left\{-it_2\mathrm{E}[u_2]\right\}.
$$

We now modify and extend a result of Bentkus and Rudzkis (1982), Theorem 1.2, to prove the validity of an Edgeworth expansion for the distribution and density functions of u_2. They studied the case of the estimate of the spectral density at a fixed frequency for a zero mean Gaussian time series.

Bentkus and Rudzkis (1982) use the following result from Feller (1971) and Petrov (1975). Let a random variable ς_Δ, with $\mathrm{E}[\varsigma_\Delta] = 0$, $\mathrm{Var}[\varsigma_\Delta] = 1$ satisfy

$$
|\mathrm{Cumulant}_s[\varsigma_\Delta]| \leq s!\,H\Delta^{2-s}, \qquad s = 3, 4, \ldots, \tag{2.4}
$$

where H and Δ are positive numbers, $\Delta \to \infty$. Then, if for some a, $0 < a < (1+2H)^{-1}$,

$$
\int_{|t|>a\Delta} |\psi(t)|dt = O(\Delta^{-r}), \qquad r \geq 2,
$$

where ψ is the characteristic function of ς_Δ, the density of ς_Δ exists and the Edgeworth expansions up to order $O(\Delta^{1-r})$ for the density and distribution function of ς_Δ are valid.

Concentrating on an expansion of second order, i.e., up to and including terms $O(\delta_{N,2}^{1/2})$ we can obtain the following theorem (further higher order terms are possible from the proof):

Theorem 2.1 *Under Assumptions 2.1, 2.2, 2.4, 2.5 and $M^{-1} + M^2 N^{-p} \to 0$, the distribution function of u_2, $F(x)$, has density $F'(x)$, and they satisfy, uniformly with respect to x,*

$$F(x) = \Phi(x) + \frac{1}{6}\kappa_N[0,3](1-x^2)\phi(x) + O(\delta_{N,2})$$

and

$$F'(x) = \phi(x) + \frac{1}{6}\kappa_N[0,3](x^3 - 3x)\phi(x) + O(\delta_{N,2}),$$

where Φ and ϕ are the distribution function and density, respectively, of a standard Gaussian random variable.

This result simplifies to Bentkus and Rudzkis's Theorem 1.2 with $Z_t = 1$, $\forall t$, and can be used directly if we assume that $E[X_t] = 0$. In the general case we face two problems. First, we can not calculate in practice $\widehat{V}_N$, only $\widetilde{V}_N$ (since β is unknown and we do not observe the disturbances X_t), so we have to evaluate the difference between using the random variables X_t or the OLS residuals $\widehat{X}_t$. The second problem is that, even in the case where the previous Edgeworth expansion refers to a feasible statistic, we would need to estimate the higher order cumulants, the variance and the bias of $\widehat{V}_N$, with negligible asymptotic error, not to make the contribution of the terms of smaller magnitude meaningless with respect to the error term.

Using the previous results for $\widehat{V}_N$ we can approximate to first order the distribution of $\widetilde{V}_N$ after evaluating the effect of the residuals estimation. Later we will consider the studentization of the regression coefficient estimate. For the approximation of $\widetilde{V}_N$ by $\widehat{V}_N$ we use the techniques developed for the evaluation of the cumulants of u_2. This also will allow us to obtain a result about the bias of the estimate of the variance, parallel to that of Hannan (1958) for the bias of the spectral density estimate.

Lemma 2.5 *Under Assumptions 2.1, 2.2, 2.4, 2.5 and $M^{-1} + M^2 N^{-p} \to 0$,*

$$\widetilde{V}_N = \widehat{V}_N + \delta_{N,2}\,\xi_N,$$

where ξ_N is a random variable with bounded moments of all orders, and

$$E\left[\delta_{N,2}\,\xi_N\right] \sim \frac{1}{N}\left(2\pi\, V_N \int_{\Pi} f(\lambda)\overline{Q}_M(\lambda)\overline{R}_N(\lambda)d\lambda - 4\pi \int_{\Pi} f(\lambda)\overline{Q}_M(\lambda)\overline{R}_N(\lambda)d\lambda\right). \quad (2.5)$$

We can now evaluate the bias term (2.5). With Lemma 2.2 and *Condition 2.1* the right hand side of expression (2.5) times N tends to

$$(2\pi)^2\int_\Pi f(\lambda)g(\lambda)d\lambda\int_\Pi f(\lambda)g^2(\lambda)d\lambda - 4\pi\int_\Pi f(\lambda)g^2(\lambda)d\lambda,$$

and under *Condition 2.2*, neglecting the continuous part of dG and the cross products of kernels K_M centered at different frequencies, to

$$(2\pi)^2 MK(0)\sum_i\Delta_i f(\lambda_i)\sum_j\Delta_j^2 f(\lambda_j) - 4\pi MK(0)\sum_j\Delta_j^2 f(\lambda_j).$$

From Theorem 2.1 and Lemmas 2.3 and 2.5 it is immediate to obtain a central limit theorem for the feasible estimate $\widetilde{V}_N$.

Theorem 2.2 *Under Assumptions 2.1, 2.2, 2.4, 2.5 and* $M^{-1} + M^2N^{-p} \to 0$,

$$Var[\widetilde{V}_N]^{1/2}\left\{\widetilde{V}_N - E[\widetilde{V}_N]\right\} \to_{\mathcal{D}} \mathcal{N}(0,1),$$

with $Var[\widetilde{V}_N] \sim \delta_{N,2}$. *We can substitute* $E[\widetilde{V}_N]$ *by* V_N, *if additionally* $N\,M^{-2q} \to 0$ *under Condition 2.1 or if* $N\,M^{-1-2q} \to 0$ *under Condition 2.2.*

Exactly as in the case of many nonparametric estimation problems, we could propose an optimal choice for the bandwidth M in terms of minimizing of the leading term of the asymptotic Mean Square Error of the estimate $\widetilde{V}_N$. The situation under Condition 2.2 is equivalent to the nonparametric estimation of the spectral density f (since then the leading term of the variance V_N is a linear combination of the values of the spectral density at certain frequencies). However, under Condition 2.1, the results are quite different, since the rate of convergence of the estimate is now root-N (not depending on M), so the optimal choice of M should be the one minimizing the bias, given the restriction $N^{-p}M^2 \to 0$.

2.4 Joint distribution of the regression and variance estimates

In this section we give several results about the joint distribution of $\widehat{\beta}$ and $\widehat{V}_N$, that will be used in next section for the analysis of the distribution of the studentized version of u_1.

Since $\mathbf{X}$ is $\mathcal{N}(0, \mathbf{\Sigma}_N)$ distributed, the joint characteristic function of the vector of standardized variables, $\mathbf{u} = (u_1, u_2)$, is given by, $\mathbf{t} = (t_1, t_2)'$,

$$\begin{aligned}\psi(\mathbf{t}) &= \left| \mathbf{I} - \frac{2it_2}{N\delta_{N,2}^{1/2}} \mathbf{\Sigma}_N \mathbf{Q}_M \right|^{-1/2} \\ &\quad \times \exp\left\{ -\frac{1}{2} t_1^2 V_N^{-1} d_N^{-2} \mathbf{Z}' \left(\mathbf{I} - \frac{2it_2}{N\delta_{N,2}^{1/2}} \mathbf{\Sigma}_N \mathbf{Q}_M \right)^{-1} \mathbf{\Sigma}_N \mathbf{Z} - it_2 \mathrm{E}[u_2] \right\}.\end{aligned}$$

To analyze the joint distribution of $\mathbf{u}$ we first study the cross cumulants between the normalized variables u_1 and u_2. Given the Gaussianity of the vector $\mathbf{X}$, the $(2, s)$ cross-cumulants are the only ones different from zero,

$$\kappa_N[2, s] \stackrel{def}{=} \mathrm{Cumulant}_{2,s}[u_1, u_2] = \frac{\delta_{N,2}^{-s/2}}{N^s d_N^2 V_N} 2^s s!\, \mathbf{Z}'(\mathbf{\Sigma}_N \mathbf{Q}_M)^s \mathbf{\Sigma}_N \mathbf{Z}.$$

Let us study the expression $d_N^{-2} \mathbf{Z}'(\mathbf{\Sigma}_N \mathbf{Q}_M)^s \mathbf{\Sigma}_N \mathbf{Z}$ using the techniques we developed for the analysis of $\mathrm{Trace}[(\mathbf{\Sigma}_N \mathbf{Q}_M)^s]$ in Lemma 2.4.

Lemma 2.6 *Under Assumptions 2.1, 2.2, 2.4 and* $M^{-1} + M^2 N^{-p} \to 0$, $s = 2, 3, \ldots$,

$$\begin{aligned}\frac{1}{d_N^2} \mathbf{Z}'(\mathbf{\Sigma}_N \mathbf{Q}_M)^s \mathbf{\Sigma}_N \mathbf{Z} &= (2\pi)^{2s+1} \int_\Pi f^{s+1}(\lambda) \overline{Q}_M^s(\lambda) \overline{R}_N(\lambda) d\lambda \\ &\quad + O\left(N^{-p} \min\left\{ M^{s+1}, M^2 \left[\sum_j |\rho(j)| \right]^{s-1} \right\} \right),\end{aligned}$$

where the bound has to be interpreted as in Lemma 2.4.

Then we can conclude that [with Assumptions 2.1, 2.2, 2.4, 2.5, $M^{-1} + N^{-p} M^2 \to 0$],

$$\begin{aligned}\kappa_N[2, s] &= \frac{(2\pi)^{2s+1} \delta_{N,2}^{-s/2}}{N^s V_N} 2^s s! \int_\Pi f^{s+1}(\lambda) \overline{Q}_M^s(\lambda) \overline{R}_N(\lambda) d\lambda \\ &\quad + O\left(\frac{\delta_{N,2}^{-s/2}}{N^{s+1}} \min\left\{ M^{s+1}, M^2 \left[\sum_j |\rho(j)| \right]^{s-1} \right\} \right), \\ &\sim \mathrm{const.}\, \delta_{N,2}^{s/2} \qquad (2.6)\end{aligned}$$

where under *Condition 2.1*, expression (2.6) is

$$\frac{s!\,(4\pi)^{s/2}}{N^{s/2}} \left[\int_\Pi f(\lambda) g(\lambda) d\lambda \right]^{-1} \left[\int_\Pi f^2(\lambda) g^2(\lambda) d\lambda \right]^{-s/2} \int_\Pi f^{s+1}(\lambda) g^{s+1}(\lambda) d\lambda$$

and under *Condition 2.2*

$$\left(\frac{M}{N}\right)^{s/2} s!(4\pi)^{s/2}\left[\sum_j \Delta_j f(\lambda_j)\right]^{-1}\left[\int K^2(x)dx \sum_j \Delta_j^2 f^2(\lambda_j)\right]^{-s/2} K^s(0)\sum_j \Delta_j^{s+1} f^{s+1}(\lambda_j).$$

We use a more general approach in this section than in the previous one to justify our asymptotic expansions. Chapter 3 will follow the same lines for the sample mean under different assumptions. In order to prove the validity of an Edgeworth expansion for the joint distribution of the vector $\mathbf{u}$ we have to check that the characteristic function of that expansion approximates well the true characteristic function. We do this in two steps. First, using the previous results on the cumulants of $\mathbf{u}$ we can check that the approximation for $\psi(\mathbf{t})$ inside a circle, with radius going to infinity with N, has an error of the desired order of magnitude. And second, we have to see that the contribution of the tails of $\psi(\mathbf{t})$ is negligible outside this circle. After this, we are in conditions to use a smoothing lemma which will measure the difference between the true distribution and the Edgeworth series approximation.

Let us start constructing the approximation for $\psi(\mathbf{t})$. As in Taniguchi (1987, pp. 11-14), using the fact that only the cumulants of the form $\kappa_N[0,s]$ and $\kappa_N[2,s]$ are different from zero, we can write the generating cumulant function as $(\tau = 1, 2, \ldots)$,

$$\begin{aligned}
\log\psi(\mathbf{t}) &= \frac{1}{2}\|i\mathbf{t}\|^2 + \sum_{s=3}^{\tau+1}\frac{1}{s!}\sum_{|\mathbf{r}|=s}\frac{s!}{r_1!\,r_2!}\kappa_N[r_1,r_2](it_1)^{r_1}(it_2)^{r_2} + R_N(\tau)\\
&= \frac{1}{2}\|i\mathbf{t}\|^2 + \sum_{s=3}^{\tau+1}\frac{1}{s!}\left[\kappa_N[0,s](it_2)^s + \frac{s(s-1)}{2}\kappa_N[0,s-2](it_1)^2(it_2)^{s-2}\right]\\
&\quad + R_N(\tau), \qquad (2.7)
\end{aligned}$$

where the vector $\mathbf{r}$ is of the form (r_1, r_2), with $r_j \in \{0,2\}$, and $|\mathbf{r}| = r_1 + r_2$, and the remaining term R_N is of this form, if τ is even:

$$R_N(\tau) = R_N[0,\tau+2](it_2)^{\tau+2} + R_N[2,\tau](it_1)^2(it_2)^{\tau},$$

or of this other form, if τ is odd:

$$\begin{aligned}
R_N(\tau) &= \frac{1}{(\tau+2)!}\left[\kappa_N[0,\tau+2](it_2)^{\tau+2} + \frac{(\tau+2)(\tau+1)}{2}\kappa_N[2,\tau](it_1)^2(it_2)^{\tau}\right]\\
&\quad + R_N[0,\tau+3](it_2)^{\tau+3} + R_N[2,\tau+1](it_1)^2(it_2)^{\tau+1},
\end{aligned}$$

where the $R_N[i,j]$ are $O(\delta_{N,2}^{(i+j-2)/2})$.

We are interested in obtain an approximation for the characteristic function of the vector $\mathbf{u}$ from its cumulant generating function. This approximation, $A(\mathbf{t}, \tau)$ say, should have leading term $\exp\{\frac{1}{2}||i\mathbf{t}||^2\}$, multiplied by a polynomial in $\mathbf{t}$, depending on the higher order cumulants of $\mathbf{u}$. For general τ, this approximation has this general form

$$A(\mathbf{t},\tau)=\exp\left\{\frac{1}{2}||i\mathbf{t}||^2\right\}\left[1+\sum_{j=3}^{\tau+1}\sum_{\mathbf{r}}\prod_{n=3}^{\tau+2}\frac{\left[\kappa_N[0,s](it_2)^s+\frac{s(s-1)}{2}\kappa_N[0,s-2](it_1)^2(it_2)^{s-2}\right]^{r_n}}{r_3!\cdots r_{\tau+2}!}\right],$$

where $\mathbf{r}=(r_3,\ldots,r_{\tau+2})$, $r_n \in \{0,1,\ldots\}$ and the summation extends for all the vectors $\mathbf{r}$ that satisfy the condition

$$\sum_{n=3}^{\tau+1}(n-2)r_n = j-2,$$

so each summand in j would be a polynomial in $\mathbf{t}$ with coefficients $O(\delta_{N,2}^{(j-2)/2})$ depending on the cumulants of $\mathbf{u}$.

For our purposes, we will justify a second order Edgeworth expansion, with $\tau = 2$. That is, including in $A(\mathbf{t},2)$ terms up to order $O\left(\delta_{N,2}^{1/2}\right)$ to approximate the distribution of $\mathbf{u}$ with error $o(\delta_{N,2}^{1/2})$. Applying the general formula we get

$$A(\mathbf{t},2) = \exp\left\{\frac{1}{2}||i\mathbf{t}||^2\right\}\left[1+\kappa_N[0,3](it_2)^3+3\kappa_N[2,1](it_1)^2(it_2)\right].$$

The next lemma measures the accuracy of this approximation for the characteristic function.

Lemma 2.7 *Under Assumptions 2.1, 2.2, 2.4, 2.5 and $M^{-1}+M^2N^{-p}\to 0$, there exists a positive number $\zeta_1>0$ such that, for $||\mathbf{t}||\le\zeta_1\delta_{N,2}^{-1/2}$ and a number $d_1>0$:*

$$|\psi(\mathbf{t})-A(\mathbf{t},2)| \quad\le\quad \exp\{-d_1||\mathbf{t}||^2\}F(||\mathbf{t}||)\mathcal{O}(\delta_{N,2}) \qquad (2.8)$$

where F is a polynomial in $\mathbf{t}$ with bounded coefficients.

Having approximated the characteristic function for values of $\mathbf{t}$ such that $||\mathbf{t}||\le \zeta_1\delta_{N,2}^{-1/2}$, the following step is to study the behaviour of this function in the tails. By construction of the Edgeworth expansion, based on the normal distribution, the contribution of the tails of its characteristic function can be always neglected.

Lemma 2.8 *Under Assumptions 2.1, 2.2, 2.4, 2.5 and $M^{-1} + M^2 N^{-p} \to 0$, there exists a positive constant $d_2 > 0$ such that for $\|\mathbf{t}\| > \zeta_1 \delta_{N,2}$,*

$$|\psi(t_1, t_2)| \leq \exp\left\{-d_2 \delta_{N,2}^{-1/2}\right\}. \tag{2.9}$$

The next goal is to estimate the distribution of the vector $\mathbf{u}$ from the approximation to its characteristic function. To measure the distance between the true distribution and the Edgeworth expansion distribution, we apply a Smoothing Lemma due to Bhattacharya and Rao (1975, pp. 97-98, 113) and referenced in Taniguchi (1987, p. 18).

Lemma 2.9 (Bhattacharya and Rao) *Let P and Γ be probability measures on $\Re^2$ and $\mathcal{B}^2$ the class of all Borel subsets of $\Re^2$. Let α be a positive number. Then there exists a kernel probability measure Ψ_α such that*

$$\sup_{B \in \mathcal{B}^2} |P(B) - \Gamma(B)| \;\; \leq \;\; \frac{2}{3} \|(P - \Gamma) \star \Psi_\alpha\| + \frac{4}{3} \sup_{B \in \mathcal{B}^2} \Gamma\{(\partial B)^{2\alpha}\}$$

where Ψ_α satisfies

$$\Psi_\alpha(B(0, r)^c) = O\left(\left(\frac{\alpha}{r}\right)^3\right) \tag{2.10}$$

and its Fourier transform $\widehat{\Psi}_\alpha$ satisfies

$$\widehat{\Psi}_\alpha = 0 \;\; \textit{for} \; \|\mathbf{t}\| \geq 8 \cdot 2^{4/3}/\pi^{1/3}\alpha. \tag{2.11}$$

$(\partial B)^{2\alpha}$ is a neighbourhood of radius 2α of the boundary of B, $\|\;\|$ is the variation norm of a measure in this case, and $\star$ means convolution. □

Introduce the notation, $B \in \mathcal{B}^2$,

$$\begin{aligned}
P_N\{B\} &= \text{Prob}\{\mathbf{u} \in B\} \\
\Gamma^{(2)}\{B\} &= \frac{1}{2\pi} \int_B \exp\left\{-\frac{1}{2}\|\mathbf{u}\|^2\right\} \left[1 + \frac{1}{3!}\{\kappa_N[0,3]\, H_3(u_2) + \kappa_N[2,1] H_2(u_1) H_1(u_2)\}\right] d\mathbf{u} \\
&= \int_B \phi_2(\mathbf{u}) q_N^{(2)}(\mathbf{u}) d\mathbf{u},
\end{aligned} \tag{2.12}$$

say, where $\phi_2(\mathbf{u})$ is the density of the bivariate normal distribution and $H_j(\cdot)$ are the univariate Hermite Polynomials (bivariate ones are not needed since the covariance matrix is diagonal). The measure $\Gamma^{(2)}\{\cdot\}$ corresponds to the characteristic function $A(\mathbf{t}, 2)$. As a direct application of the previous lemma, we get

Lemma 2.10 *Under Assumptions 2.1, 2.2, 2.4, 2.5 and* $M^{-1} + M^2 N^{-p} \to 0$, *for* $\alpha_N = (\delta_{N,2})^\rho$, $1/2 < \rho < 1$:

$$\sup_{B \in \mathcal{B}^2} \left| P_N(B) - \Gamma^{(2)}(B) \right| = o\left(\delta_{N,2}^{1/2}\right) + \frac{4}{3} \sup_{B \in \mathcal{B}^2} \Gamma^{(2)} \left\{ (\partial B)^{2\alpha_N} \right\}.$$

2.5 Approximation for the distribution of the studentized estimate

In this section we obtain an Edgeworth expansion for the distribution of the studentized least squares estimate, with the estimate $\widetilde{V}_N$ using the regression residuals $\widehat{X}_t$. First, we obtain a linear stochastic approximation to the studentized OLS estimate of the regression coefficient and prove that it is correct up to order $o(\delta_{N,2}^{1/2})$.

At this point it is necessary to make some additional assumption about the relation between $\delta_{N,2}^{1/2}$ and M. We will assume

$$\lim_{N \to \infty} \inf \ \delta_{N,2}^{1/2} M^q > 0. \tag{2.13}$$

This guarantees that the bias of $\widehat{V}_N$ (or $\widetilde{V}_N$) for V_N is at most of the same order than the standard deviation of $\widehat{V}_N$, $\delta_{N,2}^{1/2}$. If the above limit diverges, then the bias is negligible with respect to the standard deviation. Note that under Condition 2.2 it is possible to propose an optimal choice of M from the mean square error of the estimate $\widetilde{V}_N$ point of view, but this is not so simple under Condition 2.1. In any case we would need the condition $M^2 N^{-p} \to 0$ to be satisfied, which together with (2.13) requires under Condition 2.1,

$$p > \frac{1}{q}$$

and under Condition 2.2

$$p > \frac{2}{2q+1}.$$

Then we introduce the following assumption:

Assumption 2.6 *We assume simultaneously (2.13) and that the parameter* p *is big enough so that* $M^{-1} + M^2 N^{-p} \to 0$ *as* $N \to \infty$.

In the expression for the bias from Lemma 2.3 we can always neglect the term in a_N, since this is $O(N^{-1})$ and the standard deviation of $\widetilde{V}_N$ never decreases faster than $N^{-1/2}$.

We write the studentized estimate of the regression coefficient in the following way,

$$\begin{aligned} Y_N &\stackrel{def}{=} \widetilde{V}_N^{-1/2} d_N(\widehat{\beta} - \beta) \\ &= u_1 \left(1 + B_N + A_N + u_2 V_N^{-1} \delta_{N,2}^{1/2}\right)^{-1/2}, \end{aligned}$$

where B_N is the bias term, defining $b'_N = b_N V_N^{-1}$,

$$B_N = V_N^{-1}\left(\mathrm{E}[\widehat{V}_N] - V_N\right) = b'_N M^{-q} + O(M^{-q-\varrho} + N^{-1}),$$

and A_N is the residuals effect,

$$A_N = \xi_N V_N^{-1} \delta_{N,2}\,,$$

where the random variable ξ_N has moments of all orders from Lemma 2.5. Define

$$h_N = (1 + B_N)^{-1/2} = 1 - \frac{1}{2} b'_N M^{-q} + O(M^{-q-\varrho} + N^{-1}).$$

Setting a neighbourhood of the origin

$$\Omega_N = \left\{\mathbf{u} : |u_i| < c_i\, \delta_{N,2}^{-\mu},\ \ 0 < \mu < 1/6,\ i = 1, 2\right\},$$

where c_i are some fixed constants, we can expand the studentized estimate of the regression coefficient, Y_N, in terms of the normalized variable $\mathbf{u}$ around $\mathbf{0}$ in Ω_N, with $|\theta| \le 1$,

$$\begin{aligned} Y_N &= h_N u_1 - \frac{1}{2} h_N^3 u_1 \left[A_N + u_2 V_N^{-1} \delta_{N,2}^{1/2}\right] \\ &\quad + \frac{3}{8}\left(1 + B_N + \left[A_N\theta + V_N^{-1}\delta_{N,2}^{1/2} u_2\right]\theta\right)^{-5/2} u_1 \left[A_N + V_N^{-1} u_2 \delta_{N,2}^{1/2}\right]^2 \\ &= u_1 - \frac{1}{2} b'_N u_1 M^{-q} - \frac{1}{2} u_1 O(M^{-q-\varrho} + N^{-1}) \\ &\quad - \frac{1}{2} u_1 u_2 V_N^{-1} \delta_{N,2}^{1/2} - \frac{1}{2} u_1 A_N + u_1 \left[A_N + u_2 V_N^{-1}\delta_{N,2}^{1/2}\right] O(M^{-q}) \\ &\quad + \frac{3}{8}\left(1 + B_N + \left[A_N + V_N^{-1}\delta_{N,2}^{1/2} u_2\right]\theta\right)^{-5/2} u_1 \left[A_N + V_N^{-1} u_2 \delta_{N,2}^{1/2}\right]^2 \\ &= u_1 - \frac{1}{2} b'_N u_1 M^{-q} + W_N(1)\delta_{N,2} \\ &\quad - \frac{1}{2} u_1 u_2 V_N^{-1}\delta_{N,2}^{1/2} + W_N(2)\delta_{N,2} + W_N(3)\delta_{N,2} \\ &\quad + W_N(4)\delta_{N,2} \\ &= u_1 - \frac{1}{2} b'_N u_1 M^{-q} - \frac{1}{2} u_1 u_2 V_N^{-1}\delta_{N,2}^{1/2} + W_N \delta_{N,2} \\ &= Y_N^{\star} + W_N \delta_{N,2}, \end{aligned}$$

say, where

$$W_N = \sum_{j=1}^{4} W_N(j),$$

and the error terms are

$$\begin{aligned}
W_N(1) &= -\frac{1}{2}u_1 O(M^{-q-\varrho} + N^{-1})\delta_{N,2}^{-1} \\
W_N(2) &= -\frac{1}{2}u_1 A_N \delta_{N,2}^{-1} \\
W_N(3) &= u_1 \left[A_N + u_2 V_N^{-1}\delta_{N,2}^{1/2}\right] O(\delta_{N,2}^{-1/2}) \\
W_N(4) &= \frac{3}{8}\left(1 + B_N + A_N\theta_1 + V_N^{-1}\delta_{N,2}^{1/2}u_2\theta_2\right)^{-5/2} u_1 \left[A_N + V_N^{-1}u_2\delta_{N,2}^{1/2}\right]^2 \delta_{N,2}^{-1},
\end{aligned}$$

and

$$Y_N^\star = u_1 - \frac{1}{2}b_N' u_1 M^{-q} - \frac{1}{2}u_1 u_2 V_N^{-1}\delta_{N,2}^{1/2}.$$

In general, the bias problem could be neglected up to first order if we assume enough degree of undersmoothing (M big enough) in the nonparametric estimate $\widehat{V}_N$.

Now we use a similar lemma to Taniguchi (1987, Lemma 4), which is a direct modification of Chibisov (1972, Theorem 2), to prove that we can neglect the remainder term W_N when we approximate the distribution of Y_N by that of $Y_N^\star$.

Lemma 2.11 (Chibisov, 1972) *Let Y_T be a random variable which has a stochastic expansion as $T \to \infty$,*

$$Y_T = Y_T^{(2)} + T^{-1}\xi_T,$$

where the distribution of $Y_T^{(2)}$ has the following Edgeworth expansion:

$$P\{Y_T^{(2)} \in B\} = \int_B \phi(x)p_2^T(x)dx + o(T^{-1/2}),$$

where $B \in \mathcal{B}$, the class of Borel sets of $\mathcal{R}^1$ satisfying

$$\sup_{B\in\mathcal{B}_0} \int_{(\partial B)^{\mathcal{E}}} \phi(x)p_2^T(d)dx = O(\mathcal{E}). \tag{2.14}$$

Also ξ_T satisfies

$$P\{|\xi_T| > \rho_T\sqrt{T}\} = o(T^{-1/2}),$$

where $\rho_T \to 0$, $\rho_T T^{1/2} \to \infty$ as $T \to \infty$. Then

$$P\{Y_T \in B\} = \int_B \phi(x)p_2^T(x)dx + o(T^{-1/2}),$$

for $B \in \mathcal{B}_0$.

Lemma 2.12 *Under Assumptions 2.1, 2.2, 2.3, 2.4, 2.5 and 2.6, Y_N has the same Edgeworth expansion as $Y_N^\star$ for convex sets up to the order $\delta_{2,N}^{1/2}$.*

The next step is to calculate an Edgeworth expansion for the distribution of $Y_N^\star$ from that of $\mathbf{u}$. This expansion will be valid for Y_N up to order $O(\delta_{N,2}^{1/2})$. We follow mainly Taniguchi (1987, Section 6). Consider the transformation

$$\mathbf{s} = \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} = \begin{pmatrix} Y_N^\star(u_1, u_2) \\ u_2 \end{pmatrix} = \Psi(\mathbf{u})$$

and its inverse

$$\mathbf{u} = \Psi^{-1}(\mathbf{s}) = \begin{pmatrix} u_1^\star(s_1, s_2) \\ u_2 \end{pmatrix},$$

where we can write, using $(1+x)^{-1} = 1 - x + x^2 - x^3 + \cdots$, for $|x| < 1$, uniformly in Ω_N,

$$u_1^\star(\mathbf{s}) = s_1 \left[1 + \frac{1}{2} b_N' M^{-q} + \frac{1}{2} u_2 V_N \delta_{N,2}^{1/2}\right] + o(\delta_{N,2}^{1/2}), \tag{2.15}$$

where the truncation of the term in $s_1 s_2^2 O(\delta_{N,2})$ with error $o(\delta_{N,2}^{1/2})$ is allowed due to the definition of the set Ω_N. Writing for convex sets C

$$\text{Prob}\{Y_N \in C\} = \text{Prob}\left\{\mathbf{u} \in \Psi^{-1}(C \times \Re)\right\}$$

as in Taniguchi (1987, p. 22), it follows from Lemma 2.10 that (as Ψ is a C^∞ mapping on Ω_N),

$$\begin{aligned} &\sup_{C \in \mathcal{B}^2} \left|\text{Prob}\left\{\mathbf{u} \in \Psi^{-1}(C \times \Re)\right\} - \Gamma^{(2)}\left\{\Psi^{-1}(C \times \Re)\right\}\right| \\ &= o(\delta_{N,2}^{1/2}) + \text{const.} \sup_{C \in \mathcal{B}^2} \Gamma^{(2)}\left\{(\partial \Psi^{-1}(C \times \Re))^{2\alpha_N}\right\}, \end{aligned} \tag{2.16}$$

where $\alpha_N = (\delta_{N,2})^\rho$, $1/2 < \rho < 1$. Also, from the continuity of Ψ, we can obtain, for some $c > 0$,

$$\Gamma^{(2)}\left\{(\partial \Psi^{-1}(C \times \Re))^{2\alpha_N}\right\} \le \Gamma^{(2)}\left\{(\Psi^{-1}((\partial C)^{c\alpha_N} \times \Re))\right\} \tag{2.17}$$

and

$$\begin{aligned} \Gamma^{(2)}\left\{(\Psi^{-1}(C \times \Re))\right\} &= \int_{\Psi^{-1}(C\times\Re)} \phi_2(\mathbf{x}) q_N^{(2)}(\mathbf{x}) d\mathbf{x} \\ &= \int_{\Omega_N \cap \Psi^{-1}(C\times\Re)} \phi_2(\mathbf{x}) q_N^{(2)}(\mathbf{x}) d\mathbf{x} + o(\delta_{N,2}^{1/2}) \\ &= \int_{\Omega_N^\star \cap \{C\times\Re\}} \phi_2(\Psi^{-1}(\mathbf{s})) q_N^{(2)}(\Psi^{-1}(\mathbf{s})) |J| d\mathbf{s} + o(\delta_{N,2}^{1/2}), \end{aligned}$$

where $\phi_2(.)$ is the bivariate standard normal density, $\Omega_N^\star = \Psi(\Omega_N)$ and $|J|$ is the Jacobian of the transformation.

Then we can obtain, if $p_j(\mathrm{s})$ denote polynomials not depending on N or M

$$\begin{aligned}
\Gamma^{(2)}\Big\{\Psi^{-1}(C\times\Re)\Big\} &= \int_{\Omega_N^\star\cap\{C\times\Re\}}\phi_2(\mathrm{s})\Big[1+p_1(\mathrm{s})\delta_{N,2}^{1/2}+p_2(\mathrm{s})M^{-q}\Big]d\mathrm{s}+o(\delta_{N,2}^{1/2})\\
&= \int_C\phi(s_1)\Big\{\int_\Re\Big[1+p_1(\mathrm{s})\delta_{N,2}^{1/2}+p_2(\mathrm{s})M^{-q}\Big]\phi(s_2)ds_2\Big\}ds_1+o(\delta_{N,2}^{1/2})\\
&= \int_C\phi(s_1)\Big[1+r_1(s_1)\delta_{N,2}^{1/2}+r_2(s_1)M^{-q}\Big]ds_1+o(\delta_{N,2}^{1/2}),
\end{aligned}$$

where $r_j(s_1)$ are polynomials on s_1 independent of N. Since we have that

$$\begin{aligned}
E[s_1] &= o(\delta_{N,2}^{1/2})\\
E[s_1^2] &= E\Big[u_1^2-b_N'M^{-q}u_1^2-u_1^2u_2V_N\delta_{N,2}^{1/2}\Big]+o(\delta_{N,2}^{1/2}) \;=\; 1-b_N'M^{-q}+o(\delta_{N,2}^{1/2})\\
E[s_1^3] &= E\Big[u_1^3-\frac{3}{2}b_N'M^{-q}u_1^3-u_1^3u_2V_N\delta_{N,2}^{1/2}\Big]+o(\delta_{N,2}^{1/2}) \;=\; o(\delta_{N,2}^{1/2}),
\end{aligned}$$

it can be seen that

$$\begin{aligned}
r_1(x) &= 0\\
r_2(x) &= -b_N'\frac{x^2-1}{2}.
\end{aligned}$$

So we have obtained our main result, remembering (2.16), (2.17) and Lemma 2.12:

Theorem 2.3 *Under Assumptions 2.1, 2.2, 2.3, 2.4, 2.5 and 2.6, for convex sets $C\in\mathcal{B}$ and $\alpha_N=(\delta_{N,2})^\rho$, $1/2<\rho<1$,*

$$\begin{aligned}
&\sup_C\Big|Prob\{Y_N\in C\}-\int_C\phi(x)\left[1+r_2(x)M^{-q}\right]dx\Big|\\
&\quad= \frac{4}{3}\sup_C\int_{(\partial C)^{c\alpha_N}}\phi(x)\left[1+r_2(x)M^{-q}\right]dx+o(\delta_{N,2}^{1/2})
\end{aligned}$$

and, in particular, for the distribution function ($C=(-\infty,y]$):

$$\sup_{y\in\Re}\Big|Prob\{Y_N\le y\}-\int_{-\infty}^y\phi(x)\left[1+r_2(x)M^{-q}\right]dx\Big|=o(\delta_{N,2}^{1/2}).$$

Integrating this last expression and making a Taylor expansion of the standard normal distribution function, $\Phi(y)$, we can get uniformly in y, under the conditions of

Theorem 2.3:

$$\begin{aligned}\text{Prob}\{Y_N \leq y\} &= \Phi(y) + \frac{y}{2}\phi(y)b'_N\ M^{-q} + o(\delta_{N,2}^{1/2}) \\ &= \Phi\left(y + \frac{y}{2}\,b'_N M^{-q}\right) + o(\delta_{N,2}^{1/2}) \\ &= \Phi(y) + O(\delta_{N,2}^{1/2}).\end{aligned}$$

These expressions can be used to construct Edgeworth size corrections for tests of significance of the OLS estimates in the same spirit as Magee (1989) did for GLS estimates with a parametric AR(1) model for the disturbances.

2.6 Multiple regression

In this section we consider the extension of the previous analysis to the two regressors framework. The model is now

$$Y_t = \alpha Z_{1,t} + \beta Z_{2,t} + X_t, \qquad t = 1, \ldots, N, \tag{2.18}$$

where $\{Z_{i,t}\}$, $i = 1, 2$ are two observable non-stochastic sequences, which we assume satisfy a natural extension of Assumption 2.4, with equivalent interpretation:

Assumption 2.7 *The regressor sequences $\{Z_{i,t}\}$, $i = 1, 2$ satisfy as $N \to \infty$,*

- **1.**

$$d_{i,N}^2 \overset{def}{=} \sum_{t=1}^{N} Z_{i,t}^2 \to \infty.$$

- **2.**

$$\max_t \frac{Z_{i,t}^2}{d_{i,N}^2} = O(N^{-p_i}), \qquad p_i \in (0, 1].$$

- **3.** *For $r = 0, \pm 1, \ldots,$*

$$\begin{aligned}R_{i,j}(r) &\overset{def}{=} (d_{i,N}d_{j,N})^{-1} \sum_{1 \leq t, t+r \leq N} Z_{i,t}Z_{j,t+r} \\ &= \rho_{i,j}(r) + O(N^{-(p_i+p_j)/4} + rN^{-(p_i+p_j)/2}).\end{aligned}$$

- **4.** *The matrix $\rho(0) = \{\rho_{i,j}(0),\ i, j = 1, 2\}$ is nonsingular.*

We define the regression spectral measure $\mathbf{G}(\lambda)$ as the 2×2 Hermitian matrix satisfying,

$$\rho(j) = \int_{\pi}^{\pi} e^{ij\lambda} d\mathbf{G}(\lambda),$$

where $\mathbf{G}(\lambda)$ has nonnegative definite increments, is continuous from the right, and satisfies $\mathbf{G}(\pi) - \mathbf{G}(-\pi) = 1$.

Now $R_{i,j}(r) = R_{j,i}(-r)$ and the same holds for the ρ sequences. Therefore, the $R_{i,i}(r)$ are even, $i = 1, 2$, but $R_{i,j}(r)$ are not, $i \neq j$. The rate in Assumption 2.7.3 is also used to simplify the approximations when $\rho_{1,2}(0) = 0$.

The OLS estimate of $\widehat{\beta}$ satisfies

$$\widehat{\beta} = \frac{\sum_{t=1}^{N} (Z_{2,t} - \overline{Z}_2 Z_{1,t}) Y_t}{\sum_{t=1}^{N} (Z_{2,t} - \overline{Z}_2 Z_{1,t}) Z_{2,t}}, \tag{2.19}$$

where $\overline{Z}_2$ is the regression coefficient between the two regressors,

$$\overline{Z}_2 = \frac{1}{d_{1,N}^2} \sum_{t=1}^{N} Z_{1,t} Z_{2,t},$$

and

$$d_{2,N}(\widehat{\beta} - \beta) = \frac{\sum_{t=1}^{N} (Z_{2,t} - \overline{Z}_2 Z_{1,t}) X_t}{d_{2,N}^{-1} \sum_{t=1}^{N} (Z_{2,t} - \overline{Z}_2 Z_{1,t}) Z_{2,t}}. \tag{2.20}$$

The leading example is the regression with intercept, where $Z_{1,t} = 1, \forall t$. Then $\overline{Z}_2$ is just the sample mean of $Z_{2,t}$. We deal with the general case, but this particular example will be considered several times.

Also, we can see that the generalization to multiple regression models does not require new ideas, since in our approach we concentrate on the estimation of the variance for each coefficient estimate separately. To that end, instead of $\overline{Z}_2 Z_{1,t}$, we should write in (2.19) and (2.20) the value of the projection of $Z_{2,t}$ on the other regressors, and given Assumption 2.7.4 the analysis would follow the same lines as for the bivariate regression. The consideration of the distribution of the whole covariance matrix of a general vector of coefficient estimates would require multivariate Edgeworth expansions and it is not an immediate generalization of our analysis. One simplified approach is to work with fixed linear combinations of the vector of least squares estimates.

The denominator in (2.20) can be written as

$$d_{2,N}^{-1} \sum_{t=1}^{N} (Z_{2,t} - \overline{Z}_2 Z_{1,t}) Z_{2,t} = d_{2,N} \left(1 - \frac{\left(\sum_{t=1}^{N} Z_{1,t} Z_{2,t} \right)^2}{d_{1,N}^2 d_{2,N}^2} \right) = d_{2,N}(1 - R_{1,2}^2(0)) \to \infty,$$

since $1 - R_{1,2}^2(0) > 0$ as $N \to \infty$, from Assumption 2.7.4. Operating in the numerator of (2.20),

$$
\begin{aligned}
d_{2,N}(\widehat{\beta} - \beta) &= \frac{1}{(1 - R_{1,2}^2(0))} \frac{\sum_{t=1}^N Z_{2,t} X_t}{d_{2,N}} - \frac{R_{1,2}(0)}{(1 - R_{1,2}^2(0))} \frac{\sum_{t=1}^N Z_{1,t} X_t}{d_{1,N}} \\
&= \frac{1}{(1 - R_{1,2}^2(0))} \left[\Upsilon_{2,N} - R_{1,2}(0) \Upsilon_{1,N} \right], \qquad (2.21)
\end{aligned}
$$

say, where $\Upsilon_{i,N}$ $i = 1, 2$ are two linear forms of the vector $\mathbf{X}$ on the normalized regression vectors $\mathbf{Z}_1^\star$ and $\mathbf{Z}_2^\star$, $\mathbf{Z}_i^\star = d_{i,N}^{-1} \mathbf{Z}_i$, and the normalized OLS estimate is a linear combination of them with coefficients depending on the sample correlation between the two regressors, $R_{1,2}(0)$. Then the analysis is not different from the univariate regressor situation, the only new feature being the cross terms between the two linear forms. As we saw before for the univariate situation,

$$
\mathrm{Var}[\Upsilon_{i,N}] \stackrel{def}{=} V_{i,N} = \sum_{j=1-N}^{N-1} \gamma(j) R_{i,i}(j),
$$

and

$$
\begin{aligned}
\mathrm{Cov}[\Upsilon_{1,N}, \Upsilon_{2,N}] \stackrel{def}{=} V_{12,N} &= (d_{1,N} d_{2,N})^{-1} \sum_{t=1}^{N} \sum_{r=1}^{N} Z_{1,t} \gamma(r - t) Z_{2,r} \\
&= (d_{1,N} d_{2,N})^{-1} \sum_{j=1-N}^{N-1} \gamma(j) \sum_t Z_{1,t} Z_{2,t+j} \\
&= \sum_{j=1-N}^{N-1} \gamma(j) R_{1,2}(j) = \sum_{j=1-N}^{N-1} \gamma(j) R_{2,1}(j),
\end{aligned}
$$

since $\gamma(j)$ is even. Compiling results and using the same notation as in the previous case,

$$
\begin{aligned}
V_N &\stackrel{def}{=} \mathrm{Var}[d_{2,N}(\widehat{\beta} - \beta)] \\
&= \frac{1}{(1 - R_{1,2}^2(0))^2} \left[V_{2,N} + R_{1,2}^2(0) V_{1,N} - 2 R_{1,2}(0) V_{12,N} \right] \\
&= \frac{1}{(1 - R_{1,2}^2(0))^2} \sum_{j=1-N}^{N-1} \gamma(j) \left\{ R_{2,2}(j) + R_{1,2}^2(0) R_{1,1}(j) - 2 R_{1,2}(0) R_{1,2}(j) \right\} \\
&= \frac{1}{(1 - R_{1,2}^2(0))^2} \sum_{j=1-N}^{N-1} \gamma(j) S_N(j), \qquad (2.22)
\end{aligned}
$$

say, where $S_N(j)$ is observable and only depends on the regressor sequences. From (2.21) we may look at the normalized estimate in another useful way. Defining

$$
h_t = \frac{1}{(1 - R_{1,2}^2(0))} \left\{ \frac{Z_{2,t}}{d_{2,N}} - R_{1,2}(0) \frac{Z_{1,t}}{d_{1,N}} \right\}
$$

we have that

$$d_{2,N}(\widehat{\beta}-\beta)=\sum_{t=1}^{N}h_tX_t=\mathbf{H}'\mathbf{X},$$

with the vector $\mathbf{H}=(h_1,\ldots,h_N)'$ and h_t satisfying

$$\sum_{t=1}^{N}h_t^2=\frac{1}{1-R_{1,2}^2(0)}.$$

Then, we are in a univariate set-up, since (2.22) is valid with the normalized regressors h_t, defining for $j=0,\pm1,\ldots,$

$$S_N(j)=(1-R_{1,2}^2(0))^2\sum_{1\leq t,t+j\leq N}h_th_{t+j}=(1-R_{1,2}^2(0))\left(\sum_{t=1}^{N}h_t^2\right)^{-1}\sum_{1\leq t,t+j\leq N}h_th_{t+j}.$$

The normalization is not exact now by the factor $0<1-R_{1,2}^2(0)\leq 1$. In h_t we consider the cross effects between the different regressors. Now, defining similarly as with the functions R_N, the real periodogram of the observables h_t (up to a normalization) $\overline{S}_N(\lambda)$, we have that

$$\overline{S}_N(\lambda)=\frac{1}{2\pi}\sum_{t=1-N}^{N-1}S_N(j)e^{ij\lambda}\geq 0,$$

from Assumption 2.7.4 and $\overline{S}_N(\lambda)$ is also integrable since

$$\int_{\Pi}\overline{S}_N(\lambda)d\lambda=S_N(0)=1-R_{1,2}^2(0)\leq 1.$$

We are going to concentrate in certain special situations. The different sets of conditions we will consider are the following:

- **Case A**. Both regressors satisfy Condition 2.1. We assume also that the series $\rho_{i,j}(r)$ and $R_{i,j}(r)$ are absolutely summable.

- **Case B**. Both regressors satisfy Condition 2.2, with regression measures with null continuous part. Then $\rho_{1,2}(0)$ will only be different from zero if the regression spectra share at least one frequency. Let us denote as $\Delta_{ij}(r)$ the jump at frequency λ_r for the function $dG_{ij}(\lambda)$.

- **Case C**. The first regressor satisfies Condition 2.2 [with regression measures with null continuous part] and the second, Condition 2.1. We assume that $\rho_{i,j}(\ell)=0$, $\forall\ell$, $i\neq j$, and this limit is approached with the rate imposed in Assumption 2.7.3.

Case C is a simplified situation but interesting since it includes the case of mean correction, $Z_{1t} = 1$ with stationary $Z_{2,t}$. The coefficients $S_N(j)$ will inherit the summability properties of its components $R_N(j)$ in (2.22), in fact, the mildest ones when the regressors satisfy different sets of conditions. When both regressors are of the same type, we assume in Cases A and B that the sequences $\rho_{i,j}$, $i \neq j$ have the same properties as $\rho_{i,i}$, $i = 1, 2$.

In *Case A*, V_N defined in (2.22) tends to

$$\frac{2\pi}{(1-\rho_{1,2}^2(0))^2}\int_\Pi f(\lambda)\left\{g_{2,2}(\lambda) + \rho_{1,2}^2(0)g_{1,1}(\lambda) - 2\rho_{1,2}(0)g_{1,2}(\lambda)\right\}d\lambda.$$

In *Case B*, with both sequences satisfying Condition 2.2 and $Z_{1,t} = 1$, $\rho_{1,2}(0)$ would be only different from zero if the spectral distribution function of the second regressor Z_2 has a jump at the zero frequency, since in this case the spectral measure of Z_1 has all its power at the origin. In this concrete situation, the variance in (2.22) is now tending to, with $\lambda_{j^\star} = 0$,

$$\frac{2\pi}{(1-\rho_{1,2}^2(0))^2}\left[\sum_j \Delta_{22}(j)f(\lambda_j) + \rho_{1,2}^2(0)\Delta_{11}(j^\star)f(0) - 2\rho_{1,2}(0)\Delta_{12}(j^\star)f(0)\right]$$

and if $\rho_{1,2}(0) = 0$ this is just the variance in the univariate case.

In *Case C*, we study the case of regressors like $Z_{1,t} = 1$, when we only make a mean correction for Y_t. Since $\rho_{1,2}(0) = 0$, we are in the same situation as in the univariate case and (2.22) converges to

$$2\pi\int_\Pi f(\lambda)g_{2,2}(\lambda)d\lambda.$$

Now we can construct the two estimates of V_N we have considered previously, using S_N instead of R_N. As before, we can study first the unfeasible estimate $\widehat{V}_N$ and then approximate the distribution of $\widetilde{V}_N$. We work with exactly the same definitions for the function Q_M and its Fourier Transform $\overline{Q}_M$ and for the matrix $\mathbf{Q}_M$, but employing now S_N where we had previously R_N. Like in the one regressor framework, the different behaviours of the regressors will determine the properties of the estimates of the variance by means of the function Q_M. Again, we do not use extra information about the regressors when we compute the estimates of V_N.

A wide combination of outcomes is possible, but the general conclusion is that the general higher order results of previous sections for the univariate case go through for multiple regression set-ups with the obvious modifications. Defining for the model (2.18)

$$Y_N = \widetilde{V}_N^{-1/2} d_{2,N}(\widehat{\beta} - \beta),$$

$\widehat{\beta}$ given by (2.19) and $\widetilde{V}_N$ defined as before with $S_N(j)/(1 - R_{1,2}^2(0))^2$ instead of $R_N(j)$, and $r_N(x)$, b_N and b'_N as before, with $\overline{S}_N(\lambda)$ in the place of $\overline{R}_N(\lambda)$, we discuss the main differences and some particular cases in an Appendix (see Section 2.11). So we obtain finally:

Theorem 2.4 *Under Assumptions 2.1, 2.2, 2.3, 2.5, 2.6 and 2.7, the conclusions of Theorem 2.3 hold.*

2.7 Conclusions

From the discussion in Section 2.6, we concluded that all the previous results hold for the two regressors situation (Cases A, B and C) under the multivariate version of the Grenander conditions in Assumption 2.7. A complete multivariate version of the results could be obtained, considering fixed linear combinations of the least squares coefficients, or by means of multivariate Edgeworth expansions.

The consideration of more general set-ups than Conditions 2.1 or 2.2 should be possible, but these conditions may lead to not easy interpretable conclusions. Of specially interest are the limits for the function $\overline{Q}_M$ in these circumstances. This is related to the rates in Assumptions 2.4 and 2.7, necessary due to the presence of the smoothing parameter M tending to infinity with the sample size, and which makes impossible application of standard results on weak convergence of spectral measures.

We can make the following points:

- Although the rate of convergence of the least squares regression estimate is determined by $d_{N,2}$, this does not affect directly the rate of convergence of the estimate of the variance $\widetilde{V}_N$, $\sqrt{N}$ or $\sqrt{N/M}$, depending on the possible nonstationary properties of the regressor at hand.
- The lag number M may not affect the variance of $\widetilde{V}_N$, but will always decide the magnitude of the bias. This asymmetry under certain conditions makes unsuitable

the typical bandwidth choice criterion based on minimization of the asymptotic mean square error. The leading term in the bias depends on the derivatives of the spectral density of the noise sequence, as in many nonparametric problems, weighted by the regression spectrum.

- The error incorporated in the estimation of V_N from the residuals of the least squares estimation is of order $\delta_{N,2}$ and its asymptotic rate does not depend on $d_{N,2}$ directly (as occur with the rate of convergence of $\widetilde{V}_N$ itself).

- When more than one regressor is present, the same conclusions apply, and the rate of convergence of the estimates of the other regression coefficients has not influence in $\widetilde{V}_N$.

- In the Edgeworth expansion for the distribution of the studentized estimate of β the only factor up to order $\delta_{N,2}^{1/2}$ that matters is the bias, due to the symmetry of the distribution of X_t. Therefore, the normal approximation for the studentized estimate is correct up to order $O(\delta_{N,2}^{1/2})$, but assuming enough undersmoothing to reduce bias, errors of magnitude $o(\delta_{N,2}^{1/2})$ are possible (see condition (2.13)). Corrections for the kurtosis (of $\delta_{2,N}$ magnitude) are possible considering the residual estimation effect in the linearization of the studentized estimate. We pursue this approach further in next chapter for the sample mean and nonparametric estimates of the spectral density at the origin.

- From Taniguchi's (1991, pp. 58-61) discussion, we can conclude that the studentization with the nonparametric estimate $\widetilde{V}_N$ will affect some second-order asymptotic optimality properties of the standardized least squares estimate. Further research is desirable in order to compare the properties of studentized least squares estimates with respect to asymptotically optimal (studentized) generalized least squares ones, based on nonparametric estimates of the disturbances spectral density.

2.8 Appendix: Proofs of Section 2.3

Proof of Lemma 2.3. First, in Assumption 2.3, if $\omega^{(q)}$ is the qth derivative of ω,

$$\omega_q = \frac{\omega^{(q)}(0)}{-q!} \neq 0,$$

at the same time that all the other derivatives evaluated at the origin of smaller order than q are zero. Then, from the definition of the function K, for even q,

$$\omega^{(q)}(0) = (-1)^{q/2} \int K(x)x^q dx,$$

and

$$\int K(x)x^r dx = 0, \qquad 0 < r < q,$$

for integer r [for odd r this is implied also for the evenness of K]. The typical case will be $q = 2$, which allows a positive K (see, for example, Anderson (1971), p. 523, for a further discussion). Next,

$$V_N = \sum_{j=1-N}^{N-1} \gamma(j) R_N(j) = 2\pi \int_\Pi f(\lambda) \overline{R}_N(\lambda) d\lambda,$$

since $R_N(j) = 0$ for all $j \geq N$, and

$$\begin{aligned} E[\widehat{V}_N] &= \frac{1}{N} \sum_{r=1}^{N} \sum_{g=1}^{N} \omega\left(\frac{r-g}{M}\right) \gamma(r-g) R_N(r-g) \\ &= \sum_{j=1-N}^{N-1} \left[1 - \frac{|j|}{N}\right] \omega\left(\frac{j}{M}\right) \gamma(j) R_N(j) \\ &= \sum_{j=1-N}^{N-1} \omega\left(\frac{j}{M}\right) \gamma(j) R_N(j) + a_N, \end{aligned} \tag{2.23}$$

where the first term in (2.23) can be written as

$$2\pi \int_{\Pi^2} f(-\lambda_2 - \lambda_3) K_M(\lambda_2) \overline{R}_N(\lambda_3) d\lambda_2 d\lambda_3. \tag{2.24}$$

Now, given the properties of ω and K, uniformly in λ_3,

$$\int_\Pi K_M(\lambda_2) f(-\lambda_2 - \lambda_3) d\lambda_2 = f(\lambda_3) + M^{-q} \frac{f^{(q)}(\lambda_3)}{q!} \int x^q K(x) dx + O(M^{-q-\varrho}),$$

so using this last expression in (2.24) the lemma follows, given the expression for V_N. □

Proof of Lemma 2.4. Making a change of variable we can obtain,

$$
\begin{aligned}
&\mathrm{Trace}\,[(\mathbf{\Sigma}_N\mathbf{Q}_M)^s] \\
&= \sum_{1\le r_1,\dots,r_{2s}\le N} \gamma(r_1-r_2)\omega(\frac{r_2-r_3}{M})R_N(r_2-r_3)\cdots\gamma(r_{2s-1}-r_{2s})\omega(\frac{r_{2s}-r_1}{M})R_N(r_{2s}-r_1) \\
&= \sum_{-N<j_1,\dots,j_{2s-1}<N} \gamma(j_1)\omega(\frac{j_2}{M})R_N(j_2)\cdots\gamma(j_{2s-1})\omega(\frac{-j_1\cdots-j_{2s-1}}{M})R_N(-j_1\cdots-j_{2s-1}) \\
&\quad\times[N-H(j_1,\dots,j_{2s-1})] \qquad (2.25)
\end{aligned}
$$

where $H(j_1,\dots,j_{2s-1})$ is defined suitably and satisfies (see Taniguchi (1991), p. 17)

$$|H(j_1,\dots,j_{2s-1})|\le|j_1|+\cdots+|j_{2s-1}|.$$

Then we can write (2.25) as ($\mathbf{j}=j_1,\dots,j_{2s-1}$),

$$
\begin{aligned}
& N\sum_{-\infty<j_1,\dots,j_{2s-1}<\infty} \gamma(j_1)\omega(\frac{j_2}{M})R_N(j_2)\cdots\gamma(j_{2s-1})\omega(\frac{-j_1\cdots-j_{2s-1}}{M})R_N(-\sum_{r=1}^{2s-1}j_r) \quad (2.26)\\
&+ N\sum_{|j_1|,\dots,|j_{2s-1}|\ge N} \gamma(j_1)\omega(\frac{j_2}{M})R_N(j_2)\cdots\gamma(j_{2s-1})\omega(\frac{-j_1\cdots-j_{2s-1}}{M})R_N(-\sum_{r=1}^{2s-1}j_r) \quad (2.27)\\
&- \sum_{|j_1|,\dots,|j_{2s-1}|<N} H(\mathbf{j})\gamma(j_1)\omega(\frac{j_2}{M})R_N(j_2)\cdots\gamma(j_{2s-1})\omega(\frac{-j_1\cdots-j_{2s-1}}{M})R_N(-\sum_{r=1}^{2s-1}j_r). \quad (2.28)
\end{aligned}
$$

The leading term in the trace, (2.26), is

$$
\begin{aligned}
& N\sum_{-\infty<j_1,\dots,j_{2s-1}<\infty} \gamma(j_1)\omega(\frac{j_2}{M})R_N(j_2)\cdots\gamma(j_{2s-1})\omega(\frac{-j_1\cdots-j_{2s-1}}{M})R_N(-j_1\cdots-j_{2s-1}) \\
&= N\int_{\Pi^{2s}} f(\lambda_1)\overline{Q}_M(\lambda_2)\cdots f(\lambda_{2s-1})\overline{Q}_M(\lambda_{2s}) \\
&\qquad\times\sum_{j_1,\dots,j_{2s-1}}\exp\{i[j_1(\lambda_1-\lambda_{2s})+\ldots+j_{2s-1}(\lambda_{2s-1}-\lambda_{2s})]\}\,d\lambda_1\ldots d\lambda_{2s} \\
&= N(2\pi)^{2s-1}\int_\Pi f^s(\lambda)\overline{Q}_M^s(\lambda)d\lambda. \qquad (2.29)
\end{aligned}
$$

If *Condition 2.1* holds, from Lemma 2.4, expression (2.29) converges to

$$N(2\pi)^{2s-1}\int_\Pi f^s(\lambda)g^s(\lambda)d\lambda=O(N),$$

but under *Condition 2.2* this tends to

$$N(2\pi)^{2s-1}\int_\Pi f^s(\lambda)\sum_j\Delta_j^sK_M^s(\lambda-\lambda_j)d\lambda\sim NM^{s-1}(2\pi)^{2s-1}\textstyle\sum_j\Delta_j^sf^s(\lambda_j)\|K\|_s^s,$$

which is $O(NM^{s-1})$, since by Lemma 2.1 we can neglect a term of smaller order of magnitude due to the integrals of cross products of kernels K_M centered at different frequencies and the continuous part of G does not contribute to the leading term.

Since $\overline{Q}_M^s(\lambda)$ behaves in very different ways depending on the properties of regressors, we need to get totally different estimates of the error terms (2.27) and (2.28) in the approximation for the trace, depending on which assumptions we make. The way of taking the properties of the sequence Z_t into account is bounding the summations with indexes in the functions ω and R_N at the same time: if $G(\lambda)$ is continuous and the sequence ρ is absolute summable, in the evaluation of these errors we can add in terms of R_N, obtaining $O(1)$ bounds:

$$\sum_{j=-M}^{M} |R_N(j)| \leq \sum_{j=-\infty}^{\infty} |\rho(j)| + O\left(MN^{-p/2} + N^{-p} \sum_{j=-M}^{M} |j| \right) = O\left(\sum_j |\rho(j)|\right).$$

Otherwise, if this last bound is not possible, like with Condition 2.2, instead of adding in R_N, we have to add for those indexes in the function ω, getting bounds in terms of M.

First, under *Condition 2.2*, (2.27) is bounded by

$$\begin{aligned}
&\sum_{h=1}^{2s-1} \sum_{j_1=-\infty}^{\infty} \cdots \sum_{j_{h-1}=-\infty}^{\infty} \sum_{|j_h|\geq N} |j_h| \sum_{j_{h+1}=-\infty}^{\infty} \cdots \sum_{j_{2s-1}=-\infty}^{\infty} \\
&\quad |\gamma(j_1)| \left|\omega(\frac{j_2}{M})\right| |R_N(j_2)| \cdots |\gamma(j_{2s-1})| \left|\omega(\frac{-j_1 \cdots - j_{2s-1}}{M})\right| |R_N(-\textstyle\sum_{r=1}^{2s-1} j_r)| \qquad (2.30)\\
&\leq \omega^\star \sum_{h=1}^{2s-1} \sum_{j_1=-\infty}^{\infty} \cdots \sum_{j_{h-1}=-\infty}^{\infty} \sum_{|j_h|\geq N} |j_h| \sum_{j_{h+1}=-\infty}^{\infty} \cdots \sum_{j_{2s-1}=-\infty}^{\infty} \\
&\qquad |\gamma(j_1)| \left|\omega(\frac{j_2}{M})\right| |\gamma(j_3)| \cdots |\gamma(j_{2s-3})| \left|\omega(\frac{j_{2s-2}}{M})\right| |\gamma(j_{2s-1})| \\
&= O(N^{1-d}M^{s-1}).
\end{aligned}$$

This bound results as follows. First if the index j_h coincides with a γ function, then the sum in j_h is of order $O(N^{1-d})$ ($d \geq 2$ in Assumption 2.1). The remaining sums in γ are all $O(1)$, and since $|R_N(j)| \leq 1$, the $s-1$ sums in $\omega(\cdot)$ are each $O(M)$. In the case that the index j_h is summing in a ω function, the whole expression is zero, since in this case $|j_h|/M > 1$ for $|j_h| \geq N$, as $M/N \to 0$.

On the other hand, if we assume *Condition 2.1* we can bound (2.27) in the following alternate way: (2.30) is now not greater than

$$\omega^\star \sum_{h=1}^{2s-1} \sum_{j_1=-\infty}^{\infty} \cdots \sum_{j_{h-1}=-\infty}^{\infty} \sum_{|j_h|\geq N} |j_h| \sum_{j_{h+1}=-\infty}^{\infty} \cdots \sum_{j_{2s-1}=-\infty}^{\infty}$$
$$|\gamma(j_1)| \left|\omega(\frac{j_2}{M})\right| |R_N(j_2)||\gamma(j_3)| \cdots \left|\omega(\frac{j_{2s-2}}{M})\right| |R_N(j_{2s-2})||\gamma(j_{2s-1})|. \quad (2.31)$$

Now, when the index j_h coincides with $\omega(\frac{j_h}{M})R_N(j_h)$ we have that

$$\sum_{|j_h|\geq N} |j_h| \left|\omega(\frac{j_h}{M})\right| = 0,$$

since ω is zero for all the values of the index. Next, for the same reason and for any other index $j \neq j_h$

$$\sum_{j=-\infty}^{\infty} \left|\omega(\frac{j}{M})\right| |R_N(j)| = \sum_{j=-M}^{M} \left|\omega(\frac{j}{M})\right| |R_N(j)| \leq \omega^\star \sum_{j=-M}^{M} |R_N(j)| = O\left(\sum_{j=-\infty}^{\infty} |\rho(j)|\right),$$

using $M^2N^{-p} \to 0$ and because the last sum is finite with Condition 2.1. Therefore, as all the summations in $\gamma(j)$ (whether corresponding to j_h or not) are bounded, (2.31) is

$$O\left(\left[\sum_j |\rho(j)|\right]^{s-1}\right) = O(1).$$

Now, dealing with the two types of conditions simultaneously and noting that the sums in $(\omega\, R_N)$ are only for indexes less than M in absolute value, (2.28) is not bigger than

$$\begin{aligned} &\sum_{|j_1|,\ldots,|j_{2s-1}|<N} \{|j_1| + \cdots + |j_{2s-1}|\} \\ &\qquad \times|\gamma(j_1)||\omega(\frac{j_2}{M})||R_N(j_2)| \cdots |\gamma(j_{2s-1})||\omega(\frac{-j_1 \cdots - j_{2s-1}}{M})||R_N(-j_1 \cdots - j_{2s-1})| \\ &= O\left(\min\left\{M^s, M^2\left[\sum_j |\rho(j)|\right]^{s-1}\right\}\right), \end{aligned}$$

using exactly the same arguments as for (2.27), where the minimum in the order of magnitude has to be interpreted as $O(M^s)$ when the sum in ρ diverge. The factor M^2 show up because when the index in curly braces is summing in $(\omega\, R_N)$ we have to use a bound for $\sum |j||\omega(\frac{j}{M})|$, if we do not want to impose further restrictions on the sequence $\rho(j)$ in Condition 2.1, like $\sum_j |j||\rho(j)| < \infty$. To the bound being completely meaningful we need additionally $N^{-1}M^2 \to 0$, implied by the conditions of the lemma. □

Proof of Theorem 2.1. First, from the results about the cumulants of $\widehat{V}_N$, under *Condition 2.1* we have then that

$$
\begin{aligned}
\kappa_N[0,s] &\sim \delta_{N,2}^{-s/2}\frac{(s-1)!\,2^{s-1}(2\pi)^{2s-1}}{N^{s-1}}\int_\Pi f^s(\lambda)g^s(\lambda)d\lambda \\
&\sim N^{1-s/2}(s-1)!\,(4\pi)^{s/2-1}\int_\Pi f^s(\lambda)g^s(\lambda)d\lambda\left[\int_\Pi f^2(\lambda)g^2(\lambda)d\lambda\right]^{-s/2} \\
&\leq N^{1-s/2}(4\pi)^{s/2-1}(s-1)!\,2\pi\left[\|f\|_\infty^{-2}\|g\|_\infty^{-2}\int_\Pi f^2(\lambda)g^2(\lambda)d\lambda\right]^{-s/2} \\
&\leq H_1\, s!\,(C_1N)^{1-s/2}, \qquad (2.32)
\end{aligned}
$$

with $H_1 < \infty$ and

$$0 < C_1 = (4\pi)^{-1}(\|f\|_\infty\|g\|_\infty)^{-2}\int_\Pi f^2(\lambda)g^2(\lambda)d\lambda < \infty.$$

Equivalently, if *Condition 2.2* holds,

$$
\begin{aligned}
\kappa_N[0,s] &\sim \delta_{N,2}^{-s/2}\frac{(s-1)!\,2^{s-1}(2\pi)^{2s-1}}{N^{s-1}}\int_\Pi f^s(\lambda)\sum\Delta_j^s K_M^s(\lambda_j-\lambda)d\lambda \\
&\sim N^{1-s/2}(s-1)!(4\pi)^{s/2-1}M^{s-1}\|K\|_s^s\sum\Delta_j^s f^s(\lambda_j)\left[M\|K\|_2^2\sum\Delta_j^2 f^2(\lambda_j)\right]^{-s/2} \\
&\leq (4\pi\, M\, N^{-1})^{s/2-1}(s-1)!\left[\|K\|_s^{-2}\|f\|_\infty^{-2}\|K\|_2^2\sum\Delta_j^2 f^2(\lambda_j)\right]^{-s/2} \\
&\leq H_2\, s!\,(C_2\, N\, M^{-1})^{1-s/2}, \qquad (2.33)
\end{aligned}
$$

with $H_2 < \infty$ and

$$0 < C_2 = (4\pi)^{-1}(\|K\|_s\|f\|_\infty)^{-2}\|K\|_2^2\sum\Delta_j^2 f^2(\lambda_j) < \infty.$$

Denoting by μ_j the eigenvalues of the matrix $\mathbf{\Sigma}_N\mathbf{Q}_M$,

$$1 = \operatorname{Var}[u_2] = \frac{2}{N^2\delta_{N,2}}\operatorname{Trace}[(\mathbf{\Sigma}_N\mathbf{Q}_M)^2] = \frac{2}{N^2\delta_{N,2}}\sum_{j=1}^N\mu_j^2,$$

and therefore

$$\sum_{j=1}^N\mu_j^2 = \frac{1}{2}N^2\delta_{N,2}.$$

Also we have that

$$\max_j|\mu_j| = \sup_{\|\mathbf{a}\|=1}|(\mathbf{\Sigma}_N\mathbf{Q}_M\mathbf{a},\mathbf{a})| = \|\mathbf{\Sigma}_N\mathbf{Q}_M\| \leq \|\mathbf{\Sigma}_N\|\|\mathbf{Q}_M\|.$$

To evaluate the norm of the matrices $\mathbf{\Sigma}_N$ and $\mathbf{Q}_M$, first,

$$
\begin{aligned}
\|\mathbf{Q}_M\| &= \sup_{\|\mathbf{a}\|=1} |(\mathbf{Q}_M\mathbf{a},\mathbf{a})| \\
&= \sup_{\|\mathbf{a}\|=1} \left| \sum_{j,h} a_j a_h \int_\Pi \overline{Q}_M(\lambda) e^{i(h-j)\lambda} d\lambda \right| \\
&\leq \sup_{\|\mathbf{a}\|=1} \int_\Pi \left| \sum_1^N a_j e^{ij\lambda} \right|^2 |\overline{Q}_M(\lambda)| d\lambda \\
&\leq \sup_{\|\mathbf{a}\|=1} \sup_\lambda |\overline{Q}_M(\lambda)| \int_\Pi \left| \sum_1^N a_j e^{ij\lambda} \right|^2 d\lambda \\
&\leq c_1 N \delta_{N,2}, \qquad (2.34)
\end{aligned}
$$

where c_1 is a constant depending on $\sup_\lambda |K(\lambda)|$ or on $\sup_\lambda g(\lambda)$, in relation to which type of conditions on the regressors we take in Assumption 2.4. Because the last integral is equal to 2π, we have under Condition 2.1, $c_1 N \delta_{N,2} = 2\pi \|g\|_\infty$ and with Condition 2.2, $c_1 N \delta_{N,2} = 2\pi\, M\, \|K\|_\infty$. On the other hand

$$
\begin{aligned}
\|\mathbf{\Sigma}_N\| &= \sup_{\|\mathbf{a}\|=1} \left| \sum_{j,h} a_j a_h \int_\Pi f(\lambda) e^{i(h-j)\lambda} d\lambda \right| \\
&\leq \sup_{\|\mathbf{a}\|=1} \sup_\lambda |f(\lambda)| \int_\Pi \left| \sum_{j=1}^N a_j e^{ij\lambda} \right|^2 d\lambda \\
&= 2\pi \|f\|_\infty \stackrel{def}{=} c_2 < \infty. \qquad (2.35)
\end{aligned}
$$

Then from (2.34) and (2.35), for $c = c_1\, c_2$, $0 < c < \infty$,

$$
\max_j |\mu_j| \leq c\, N \delta_{N,2}. \qquad (2.36)
$$

Introduce the notation

$$
g_j = \mu_j c^{-1} [N\delta_{N,2}]^{-1},
$$

where $|g_j| \leq 1$, so

$$
\sum_{j=1}^N g_j^2 = c^{-2} [N\delta_{N,2}]^{-2} \sum_{j=1}^N \mu_j^2 = \frac{1}{2\,c^2} \delta_{N,2}^{-1}
$$

and

$$
|\psi(t_2)| = \prod_{j=1}^N \left(1 + 4 t_2^2 \frac{\mu_j^2}{N^2 \delta_{N,2}}\right)^{-1/4}
$$

$$
\begin{aligned}
&= \prod_{j=1}^{N}\left(1+4t_2^2\,c^2\,g_j^2\,\delta_{N,2}\right)^{-1/4} \\
&\le \prod_{j=1}^{N}\left(1+4t_2^2\,c^2\,\delta_{N,2}\right)^{-\frac{1}{4}g_j^2} \\
&= \left(1+4t_2^2\delta_{N,2}c^2\right)^{-\frac{1}{4}\left[2c^2\delta_{N,2}\right]^{-1}}, \qquad (2.37)
\end{aligned}
$$

where we have applied the inequality $(1+bt) \geq (1+t)^b$, valid for $t \geq 0$, $0 \leq b \leq 1$. Therefore denoting,

$$
\Delta = \Delta_N = \left(4c^2\delta_{N,2}\right)^{-1/2} \to \infty, \qquad (2.38)
$$

as $N \to \infty$, we have obtained that

$$
|\psi(t_2)| \leq \left(1+\frac{t_2^2}{\Delta^2}\right)^{-\frac{1}{2}\Delta^2},
$$

and recalling (2.32) and (2.33), condition (2.4) is satisfied for the same Δ of (2.38) and one $H < \infty$. Next, for any $a > 0$,

$$
\begin{aligned}
\int_{|t|>a\Delta}\left(1+\frac{t^2}{\Delta^2}\right)^{-\frac{1}{2}\Delta^2} dt &= 2\int_{a\Delta}^{\infty}\left(1+\frac{t^2}{\Delta^2}\right)^{-\frac{1}{2}\Delta^2} dt \\
&= 2\Delta^{\Delta^2}\int_{a\Delta}^{\infty}\left(\Delta^2+t^2\right)^{-\frac{1}{2}\Delta^2} dt
\end{aligned}
$$

and this is not bigger than

$$
2\Delta^{\Delta^2}\int_{a\Delta}^{\infty} t\left(\Delta^2+t^2\right)^{-\frac{1}{2}\Delta^2} dt
$$

if $a\Delta > 1$ given $\Delta \to \infty$, and making the change of variable $t^2 = z$, the previous expression is equal to

$$
\begin{aligned}
\Delta^{\Delta^2}\int_{(a\Delta)^2}^{\infty}\left(\Delta^2+z\right)^{-\frac{1}{2}\Delta^2} dz &= \Delta^{\Delta^2}\left[\frac{(\Delta^2+z)^{1-\frac{1}{2}\Delta^2}}{1-\frac{\Delta^2}{2}}\right]_{(a\Delta)^2}^{\infty} \\
&= -\Delta^{\Delta^2}\frac{2\Delta^{2-\Delta^2}(1+a^2)^{1-\frac{1}{2}\Delta^2}}{2-\Delta^2} \\
&= 2\left(\frac{\Delta^2}{\Delta^2-2}\right)(1+a^2)^{1-\frac{1}{2}\Delta^2}
\end{aligned}
$$

and this tends to 0 faster than any power of Δ^{-1} as N increases, since $a > 0$. So we can justify an Edgeworth expansion of any order for the distribution of $\widehat{V}_N$. □

Proof of Lemma 2.5. First, since we can write

$$\widehat{\mathbf{X}} = \mathbf{Y} - \widehat{\beta}\mathbf{Z} = (\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')\mathbf{X}$$

we have

$$\begin{aligned}
\widetilde{V}_N &= \widehat{V}_N - \frac{2}{N}\mathbf{X}'\mathbf{Q}_M\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X} + \frac{1}{N}\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Q}_M\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X} \\
&= \widehat{V}_N - 2\xi_N(1) + \xi_N(2),
\end{aligned}$$

say. The cumulants of $\xi_N(1)$ are

$$\text{Cumulant}_s[\xi_N(1)] = \frac{(s-1)!\,2^{s-1}}{N^s}\,\text{Trace}[(\mathbf{\Sigma}_N\mathbf{Q}_M\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')^s] \qquad s = 1,2,\ldots,$$

where

$$\text{Trace}[(\mathbf{\Sigma}_N\mathbf{Q}_M\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')^s] = d_N^{-2s}\,\text{Trace}[(\mathbf{Z}'\mathbf{\Sigma}_N\mathbf{Q}_M\mathbf{Z})^s] = \left[d_N^{-2}\mathbf{Z}'\mathbf{\Sigma}_N\mathbf{Q}_M\mathbf{Z}\right]^s.$$

Then, using Assumption 2.4.2, with $M^{-1} + N^{-p}M^2 \to 0$, with a change of variable and neglecting a similar term to (2.28),

$$\begin{aligned}
d_N^{-2}\mathbf{Z}'\mathbf{\Sigma}_N\mathbf{Q}_M\mathbf{Z} &= d_N^{-2}\sum_{1\le r_1,r_2,r_3\le N} Z_{r_1}\gamma(r_1-r_2)\omega(\frac{r_2-r_3}{M})R_N(r_2-r_3)Z_{r_3} \\
&= \sum_{1-N\le j_1,j_2\le N-1}\gamma(j_1)\omega(\frac{j_2}{M})R_N(j_2)R_N(-j_1-j_2) + O(N^{-p}M^2) \\
&\sim 2\pi\int_{\Pi} f(\lambda)\overline{Q}_M(\lambda)\overline{R}_N(\lambda)d\lambda \qquad (2.39) \\
&= O(N\,\delta_{N,2}),
\end{aligned}$$

so the s-th cumulant of $\xi_N(1)$ is $O(\delta_{N,2}^s)$, $s = 1,2,\ldots$.

Moving now to $\xi_N(2)$, $s = 1,2,\ldots,$

$$\text{Cumulant}_s[\xi_N(2)] = \frac{(s-1)!\,2^{s-1}}{N^s}\text{Trace}[(\mathbf{\Sigma}_N\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Q}_M\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')^s],$$

with

$$\begin{aligned}
\text{Trace}[(\mathbf{\Sigma}_N\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Q}_M\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')^s] &= d_N^{-2s}\,\text{Trace}[(\mathbf{Z}'\mathbf{\Sigma}_N\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Q}_M\mathbf{Z})^s] \\
&= \left[d_N^{-2}\mathbf{Z}'\mathbf{\Sigma}_N\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Q}_M\mathbf{Z}\right]^s.
\end{aligned}$$

Reasoning in the same way,

$$\begin{aligned}
d_N^{-2}\mathbf{Z}'\mathbf{\Sigma}_N\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Q}_M\mathbf{Z} &= d_N^{-2}\sum_{1\le r_1,\dots,r_4\le N} Z_{r_1}\gamma(r_1-r_2)\frac{Z_{r_2}Z_{r_3}}{d_N^2}\omega(\frac{r_3-r_4}{M})R_N(r_3-r_4)Z_{r_4} \\
&= \sum_{j_1=1-N}^{N-1}\gamma(j_1)R_N(j_1)\sum_{j_2=1-N}^{N-1}\omega(\frac{j_2}{M})R_N(j_2)R_N(-j_2) \\
&= 2\pi V_N\int_\Pi f(\lambda)\overline{Q}_M(\lambda)\overline{R}_N(\lambda)d\lambda \qquad (2.40)\\
&= O(N\,\delta_{N,2}),
\end{aligned}$$

where the bound results as before. Then the s-th cumulant of $\xi_N(2)$ is $O\left(\delta_{N,2}^s\right)$. Therefore, focusing in the bounds for the cumulants of $\xi_N(1)$ and $\xi_N(2)$ and on the special case $s=1$ for the bias, we have proved the lemma. □

2.9 Appendix: Proofs of Section 2.4

Proof of Lemma 2.6. Making a change of variable in the indexes of the sums,

$$\begin{aligned}
&\frac{1}{d_N^2}\mathbf{Z}'(\mathbf{\Sigma}_N\mathbf{Q}_M)^s\mathbf{\Sigma}_N\mathbf{Z} \\
&= \frac{1}{d_N^2}\sum_{1\le r_1,\dots,r_{2s+2}\le N} Z_{r_1}\gamma(r_1-r_2)\cdots\omega(\frac{r_{2s}-r_{2s+1}}{M})R_N(r_{2s}-r_{2s+1})\,\gamma(r_{2s+1}-r_{2s+2})Z_{r_{2s+2}} \\
&= \sum_{-N<j_1,\dots,j_{2s+1}<N}\gamma(j_1)\cdots\omega(\frac{j_{2s}}{M})R_N(j_{2s})\,\gamma(j_{2s+1})R_N(-j_1\cdots-j_{2s+1}) \\
&\qquad\times\left[1-\frac{\mathrm{Res}(j_1,\dots,j_{2s+1})}{d_N^2}\right]
\end{aligned}$$

where the term $\mathrm{Res}(j_1,\dots,j_{2s+1})$ is defined suitably as a sum of product terms of the form $Z_tZ_{t'}$, whose total number satisfies

$$\#\{\mathrm{Res}(j_1,\dots,j_{2s+1})\}\le|j_1|+\cdots+|j_{2s+1}|.$$

Next, reasoning as in the proof of Lemma 2.4,

$$\begin{aligned}
&\sum_{-N<j_1,\dots,j_{2s+1}<N}\gamma(j_1)\cdots\omega(\frac{j_{2s}}{M})R_N(j_{2s})\gamma(j_{2s+1})R_N(-j_1\cdots-j_{2s+1}) \\
&= \sum_{\infty<j_1,\dots,j_{2s+1}<\infty}\gamma(j_1)\cdots\omega(\frac{j_{2s}}{M})R_N(j_{2s})\gamma(j_{2s+1})R_N(-j_1\cdots-j_{2s+1}) \qquad (2.41)\\
&\quad+O\Big(N^{-1}\min\Big\{M^{s+1},\Big[\textstyle\sum_j|\rho(j)|\Big]^s\Big\}\Big),
\end{aligned}$$

and then (2.41) is equal to

$$(2\pi)^{2s+1}\int_{\Pi} f^{s+1}(\lambda)\overline{Q}_M^s(\lambda)\overline{R}_N(\lambda)d\lambda. \tag{2.42}$$

Therefore, under *Condition 2.1*, (2.42) is tending to

$$(2\pi)^{2s+1}\int_{\Pi} f^{s+1}(\lambda)g^{s+1}(\lambda)d\lambda < \infty$$

and under *Condition 2.2*, expression (2.42) converges to (neglecting the continuous part contributions),

$$\begin{aligned}
&(2\pi)^{2s+1}\int_{\Pi} f^{s+1}(\lambda)\left[\sum\nolimits_j \Delta_j K_M(\lambda-\lambda_j)\right]^s dG(\lambda)d\lambda \\
\sim\;& (2\pi)^{2s+1}\sum\nolimits_{j'}\Delta_{j'} f^{s+1}(\lambda_{j'})\left[\sum\nolimits_j \Delta_j K_M(\lambda_{j'}-\lambda_j)\right]^s \\
\sim\;& (2\pi)^{2s+1}\sum\nolimits_j \Delta_j^{s+1} f^{s+1}(\lambda_j)K_M^s(0) \\
=\;& M^s(2\pi)^{2s+1}K^s(0)\sum\nolimits_j \Delta_j^{s+1} f^{s+1}(\lambda_j) = O(M^s),
\end{aligned}$$

since the cross products with any kernel function evaluated at a frequency apart from the origin are always of smaller order of magnitude as $M \to \infty$.

Now using Assumption 2.4.2 and $|R_N| \le 1$ we can see that

$$\begin{aligned}
&\left|\sum_{-N<j_1,\dots,j_{2s+1}<N} \gamma(j_1)\cdots\omega(\tfrac{j_{2s}}{M})R_N(j_{2s})\gamma(j_{2s+1})R_N(-j_1\cdots-j_{2s+1})\frac{\mathrm{Res}(j_1,\dots,j_{2s+1})}{d_N^2}\right| \\
\le\;& \max_{t,t'}\frac{|Z_tZ_{t'}|}{d_N^2}\sum_{-N<j_1,\dots,j_{2s+1}<N}\{|j_1|+\cdots+|j_{2s+1}|\}\,|\gamma(j_1)|\cdots\left|\omega(\tfrac{j_{2s}}{M})\right||R_N(j_{2s})||\gamma(j_{2s+1})| \\
=\;& O\!\left(N^{-p}\min\left\{M^{s+1}, M^2\left[\sum\nolimits_j|\rho(j)|\right]^{s-1}\right\}\right),
\end{aligned}$$

using the same reasonings as in the proof of Lemma 2.4, but taking into account that now we have s sums in the function $\omega(\frac{j}{M})R_N(j)$, one of them possibly multiplied by the index $|j|$. The Lemma is proved. □

Proof of Lemma 2.7. Similarly to Feller (1971, p. 535) or Durbin (1980a, p. 325) we have for complex α and τ

$$|e^\alpha - 1 - \tau| \le e^\gamma\left\{|\alpha-\tau| + \frac{|\tau|^2}{2}\right\} \tag{2.43}$$

where $\gamma = \max\{|\alpha|,|\tau|\}$. Let's take (with $r=2$ in (2.7)):

$$\alpha = \log\psi(\mathbf{t}) - \frac{1}{2}||i\mathbf{t}||^2 = \sum_{|\mathbf{r}|=3}\frac{1}{r_1!r_2!}\kappa_N[r_1,r_2](it_1)^{r_1}(it_2)^{r_2} + R_N(2)$$

and

$$\tau = \sum_{|\mathbf{r}|=3} \frac{1}{r_1! r_2!} \kappa_N[r_1, r_2](it_1)^{r_1}(it_2)^{r_2},$$

then we have

$$|\alpha - \tau| \leq \left| R_N[0,4](it_2)^4 + R_N[2,2](it_1)^2(it_2)^2 \right| \leq F_1(||\mathbf{t}||)O(\delta_{N,2})$$

where F_1 is a polynomial of degree 4. Now

$$\frac{|\tau|^2}{2} \leq F_2(||\mathbf{t}||)O(\delta_{N,2})$$

where F_2 is a polynomial of degree 6. Then

$$|\alpha - \tau| + \frac{|\tau|^2}{2} \leq F(||\mathbf{t}||)O(\delta_{N,2}) \tag{2.44}$$

for some polynomial F. Define the normalized cumulants $\bar{\kappa}_N[i,j] = \delta_{N,2}^{2-i-j}\kappa_N[i,j] = O(1)$. To study γ, we first bound $|\tau|$ for $||\mathbf{t}|| \leq \zeta_\tau \delta_{N,2}^{-1/2}$, $\zeta_\tau > 0$

$$\begin{aligned}
|\tau| &\leq ||\mathbf{t}||^2 \left\{ \frac{||\mathbf{t}||}{3!} \left(|\kappa_N[0,3]| + 3|\kappa_N[2,1]| \right) \right\} \\
&\leq ||\mathbf{t}||^2 \left\{ \frac{\zeta_\tau}{3!} \delta_{N,2}^{-1/2} \left(|\bar{\kappa}_N[0,3]| + 3|\bar{\kappa}_N[2,1]| \right) \right\} \\
&\leq ||\mathbf{t}||^2 T_\tau
\end{aligned} \tag{2.45}$$

with $0 < T_\tau < 1/4$ if we choose ζ_τ small enough, given the previous results for the cumulants of order 3. Now for α we can choose a $\zeta_\alpha > 0$ small enough, such that, for $||\mathbf{t}|| \leq \zeta_\alpha \delta_{N,2}^{-1/2}$,

$$\begin{aligned}
|\alpha| &\leq ||\mathbf{t}||^2 \left\{ \frac{1}{3!} (|\kappa_N[0,3]| + 3|\kappa_N[2,1]|)|) \, ||\mathbf{t} + (|R_N[0,4]| + |R_N[2,2]|) \, ||\mathbf{t}||^2 \right\} \\
&\leq ||\mathbf{t}||^2 \left\{ \frac{\zeta_\alpha}{3!} \delta_{N,2}^{-1/2} (|\kappa_N[0,3]| + 3|\kappa_N[2,1]|) + \zeta_\alpha^2 \delta_{N,2}^{-1} (|R_N[0,4]| + |R_N[2,2]|) \right\} \\
&\leq \frac{||\mathbf{t}||^2}{4}.
\end{aligned} \tag{2.46}$$

From (2.45) and (2.46) we have that

$$e^\gamma \leq \exp\left\{ \frac{||\mathbf{t}||^2}{4} \right\}$$

for $||\mathbf{t}|| \leq \zeta_1 \delta_{N,2}^{-1/2}$ where $\zeta_1 = \min\{\zeta_\alpha, \zeta_\tau\}$. Then,

$$\exp\left\{ -\frac{1}{2}||\mathbf{t}||^2 + \gamma \right\} \leq \exp\left\{ -d_1 ||\mathbf{t}||^2 \right\} \tag{2.47}$$

for one $d_1 > 0$, $\|\mathbf{t}\| \leq \zeta_1 \delta_{N,2}^{-1/2}$. Since our approximation to $\psi(\mathbf{t}) = \exp\{\frac{1}{2}\|i\mathbf{t}\|^2 + \alpha\}$ is $A(\mathbf{t}, 2) = \exp\left\{\frac{1}{2}\|i\mathbf{t}\|^2\right\}[1+\tau]$, using (2.44) and (2.47) the lemma is proved. $\square$

Proof of Lemma 2.8. Since the characteristic function of u_2 appears in the joint characteristic function itself, we can follow Bentkus and Rudzkis (1982) and the proof of Theorem 2.1.

From expression (2.37), we can see that for all $\eta > 0$, and N and M big enough we have that

$$|\psi(t_2)| \leq (1 + \eta_1^2)^{-\eta_2 \delta_{N,2}^{-1}} \tag{2.48}$$

for $|t_2| > \eta \delta_{N,2}^{-1/2}$ and some $\eta_1 > 0$ and $\eta_2 > 0$ depending on η.

Returning to the bivariate characteristic function, its modulus is equal to

$$\begin{aligned} |\psi(t_1, t_2)| &= \left| \mathrm{Det}\left[\mathbf{I} - \frac{2it_2}{N\delta_{N,2}^{1/2}} \boldsymbol{\Sigma}_N \mathbf{Q}_M\right]^{-1/2} \right| \\ &\quad \times \exp\left\{ -\frac{1}{2} t_1^2 V_N^{-1} d_N^{-2} \mathbf{Z}' \mathcal{R}\left(\mathbf{I} - \frac{2it_2}{N\delta_{N,2}^{1/2}} \boldsymbol{\Sigma}_N \mathbf{Q}_M\right)^{-1} \boldsymbol{\Sigma}_N \mathbf{Z} \right\} \\ &= |\psi(t_2)| \exp\left\{ -\frac{1}{2} t_1^2 V_N^{-1} d_N^{-2} \mathbf{Z}' \mathcal{R}\left(\mathbf{I} - \frac{2it_2}{N\delta_{N,2}^{1/2}} \boldsymbol{\Sigma}_N \mathbf{Q}_M\right)^{-1} \boldsymbol{\Sigma}_N \mathbf{Z} \right\} \end{aligned}$$

where $\mathcal{R}$ stands for real part. From Anderson (1958, p. 161),

$$\mathcal{R}\left(\boldsymbol{\Sigma}_N^{-1} - \frac{2it_2}{N\delta_{N,2}^{1/2}} \mathbf{Q}_M\right)^{-1} = \mathcal{R}\left(I - \frac{2it_2}{N\delta_{N,2}^{1/2}} \boldsymbol{\Sigma}_N \mathbf{Q}_M\right)^{-1} \boldsymbol{\Sigma}_N$$

is positive definite as $t_2 \mathbf{Q}_M$ is real (for every N). Then

$$d_N^{-2} \mathbf{Z}' \mathcal{R}\left(I - \frac{2it_2}{N\delta_{N,2}^{1/2}} \boldsymbol{\Sigma}_N \mathbf{Q}_M\right)^{-1} \boldsymbol{\Sigma}_N \mathbf{Z} > 0$$

for all $t_2 \in \Re$, and for $|t_2| \leq \zeta \delta_{N,2}^{-1/2}$, $\forall \zeta > 0$,

$$d_N^{-2} \mathbf{Z}' \mathcal{R}\left(I - \frac{2it_2}{N\delta_{N,2}^{1/2}} \boldsymbol{\Sigma}_N \mathbf{Q}_M\right)^{-1} \boldsymbol{\Sigma}_N \mathbf{Z} > \epsilon$$

for some $\epsilon > 0$, since we obtained previously that, cf. (2.36),

$$\|\boldsymbol{\Sigma}_N \mathbf{Q}_M\| = O\left(N\delta_{N,2}\right).$$

Then,

$$\exp\left\{-\frac{1}{2}t_1^2V_N^{-1}d_N^{-2}\mathbf{Z}'\mathcal{R}\left(\mathbf{I}-\frac{2it_2}{N\delta_{N,2}^{1/2}}\boldsymbol{\Sigma}_N\mathbf{Q}_M\right)^{-1}\boldsymbol{\Sigma}_N\mathbf{Z}\right\} \le \exp\left\{-\frac{1}{2}t_1^2\epsilon_1\right\} \le \exp\left\{-\frac{1}{4}\epsilon_1\zeta_1^2\,\delta_{N,2}^{-1}\right\} \quad (2.49)$$

for $|t_1|\sqrt{2} > \zeta_1\delta_{N,2}^{-1/2}$ and $|t_2|\sqrt{2} \le \zeta_1\delta_{N,2}^{-1/2}$, and some $\epsilon_1 > 0$ depending on ζ_1.

Lastly, from (2.48) and (2.49), we have that for $\|\mathbf{t}\| > \zeta_1\delta_{N,2}^{-1/2}$, there exists one number $d_2 > 0$ such that

$$|\psi(t_1,t_2)| \le \exp\left\{-d_2\delta_{N,2}^{-1}\right\},$$

as $\left\{\mathbf{t} : \|\mathbf{t}\| > \zeta_1\delta_{N,2}^{-1/2}\right\} \subset B_1 \cup B_2$ where

$$\begin{aligned} B_1 &= \left\{\mathbf{t} : |t_2| > \frac{\zeta_1}{\sqrt{2}}\delta_{N,2}^{-1/2}\right\} \\ B_2 &= \left\{\mathbf{t} : |t_2| \le \frac{\zeta_1}{\sqrt{2}}\delta_{N,2}^{-1/2} \text{ and } |t_1| > \frac{\zeta_1}{\sqrt{2}}\delta_{N,2}^{-1/2}\right\}, \end{aligned}$$

and the lemma is proved. $\square$

Proof of Lemma 2.10. First,

$$\begin{aligned} \|(P_N - \Gamma^{(2)}) \star \Psi_{\alpha_N}\| &= 2\sup_{B\in\mathcal{B}^2} |(P_N - \Gamma^{(2)}) \star \Psi_{\alpha_N}| \\ &\le \sup\left[|(P_N - \Gamma^{(2)}) \star \Psi_{\alpha_N}|; B \subset B(0,r_N)^c\right] \\ &\quad + \sup\left[|(P_N - \Gamma^{(2)}) \star \Psi_{\alpha_N}|; B \subset B(0,r_N)\right], \end{aligned}$$

where $r_N = \delta_{N,2}^{-\tau}$, ($\tau > 0$ to be chosen later).

Now for $B \subset B(0,r_N)^c$ we have uniformly

$$\begin{aligned} |(P_N - \Gamma^{(2)}) \star \Psi_{\alpha_N}| &\le |P_N \star \Psi_{\alpha_N}| + |\Gamma^{(2)} \star \Psi_{\alpha_N}| \\ &\le \mathrm{Prob}\{\|\mathbf{u}\| \ge r_N/2\} \\ &\quad + 2\,\Psi_{\alpha_N}\{B(0,r_N/2)^c\} \\ &\quad + 2\,\Gamma^{(2)}\{B(0,r_N/2)^c\}. \end{aligned}$$

First,

$$\Gamma^{(2)}\{B(0,r_N/2)^c\} = o(\delta_{N,2}^{1/2}),$$

as this measure corresponds to a polynomial with bounded coefficientes times a Gaussian density. Also

$$\text{Prob}\{\|\mathbf{u}\| \geq r_N/2\} = o(\delta_{N,2}^{1/2}),$$

as $\mathbf{u}$ has finite moments of all orders. Finally

$$\Psi_{\alpha_N}\{B(0, r_N/2)^c\} = O([\alpha_N/r_N]^2) = O(\delta_{N,2}^{3(\rho+\tau)}) = o(\delta_{N,2}^{1/2}),$$

since $\rho + \tau > 1/6$.

For $B \subset B(0, r_N)$ we have by Fourier Inversion

$$|(P_N - \Gamma^{(2)}) \star \Psi_{\alpha_N}| \leq \left[\frac{1}{(2\pi)^2}\pi r_N^2\right] \int |(\widehat{P}_N - \widehat{\Gamma}^{(2)})(\mathbf{t})\widehat{\Psi}_{\alpha_N}(\mathbf{t})|dt, \tag{2.50}$$

and as we know that $\widehat{P}_N = \psi(\mathbf{t})$ and $\widehat{\Gamma}^{(2)} = A(\mathbf{t}, 2)$, using Lemma 2.7, (2.50) is bounded by

$$O\left(\delta_{N,2}^{-2\tau+1}\right) \int_{\|\mathbf{t}\| \leq \zeta_1 \delta_{N,2}^{-1/2}} \left|e^{-\frac{1}{2}\|\mathbf{t}\|^2} F_1(\|\mathbf{t}\|)\right| |\widehat{\Psi}_{\alpha_N}(\mathbf{t})| dt \tag{2.51}$$

$$+ O(\delta_{N,2}^{-2\tau}) \int_{\zeta_1 \delta_{N,2}^{-1/2} < \|\mathbf{t}\| \leq a' \delta_{N,2}^{-\rho}} |(\widehat{P}_N - \widehat{\Gamma}^{(2)})(\mathbf{t})\, \widehat{\Psi}_{\alpha_N}(\mathbf{t})| dt \tag{2.52}$$

as from (2.11), $\widehat{\Psi}$ is zero for $\|\mathbf{t}\| > a'\delta_{N,2}^{-\rho}$ and $a' = 8 \cdot 2^{4/3}\pi^{-1/3}$. Then for (2.51) to be $o(\delta_{N,2}^{1/2})$ it is necessary to chose $\tau < 1/4$. Next, (2.52) is dominated by

$$O(\delta_{N,2}^{-2\tau-2\rho}) e^{-d_2 \delta_{N,2}^{-1}} + o(\delta_{N,2}^{1/2}) = o(\delta_{N,2}^{1/2}).$$

Applying the Smoothing Lemma the proof is complete. □

2.10 Appendix: Proofs of Section 2.5

Proof of Lemma 2.12. Since convex sets satisfy (2.14), we can apply Lemma 2.11, proving that

$$\text{Prob}\left\{|W_N| > \rho_N \delta_{2,N}^{-1/2}\right\} = o(\delta_{2,N}^{1/2}) \tag{2.53}$$

for some positive sequence $\rho_N \to 0$ and $\rho_N \delta_{2,N}^{-1/2} \to \infty$. Let's choose $\rho_N = 1/\log[\delta_{2,N}^{-1}]$. Then we have

$$\text{Prob}\left\{|W_N| > \rho_N \delta_{2,N}^{-1/2}\right\} \leq \sum_{j=1}^{4} \text{Prob}\left\{|W_N(j)| > \frac{1}{4}\rho_N \delta_{2,N}^{-1/2}\right\}$$

so writing now, for some $\epsilon > 0$

$$\begin{aligned}
\delta_{N,2}^{1/2} W_N(1) &= -\frac{1}{2} u_1 O(M^{-q-\varrho} + N^{-1}) \delta_{N,2}^{-1/2} = u_1 O(\delta_{N,2}^{\epsilon}) \\
\delta_{N,2}^{1/2} W_N(2) &= -\frac{1}{2} u_1 A_N \delta_{N,2}^{-1/2} = u_1 \xi_N O(\delta_{N,2}^{\epsilon}) \\
\delta_{N,2}^{1/2} W_N(3) &= u_1 \left[A_N + u_2 V_N^{-1} \delta_{N,2}^{1/2} \right] O(1) = [u_1 \xi_N + u_1 u_2] O(\delta_{N,2}^{\epsilon})
\end{aligned}$$

and applying Chebychev's inequality, as ξ_N, u_1 and u_2 have finite moments of all orders, (2.53) is satisfied at once. Now write

$$W_N(4) = \left(1 + B_N + \left[A_N + V_N^{-1} \delta_{N,2}^{1/2} u_2\right] \theta\right)^{-5/2} R_N(2) = R_N(1)\, R_N(2),$$

say, where $R_N(2)$ is a random variable with bounded moments of all orders. As before, in order to satisfy Chibisov condition (2.53), we need

$$\text{Prob}\left\{ |R_N(1)\, R_N(2)| > \rho_N \delta_{2,N}^{-1/2} \right\} = o(\delta_{2,N}^{1/2}), \tag{2.54}$$

but the probability in the left hand side of (2.54) is less or equal than

$$\text{Prob}\left\{ |R_N(1)| \delta_{2,N}^{1/4} > \rho_N^{1/2} \right\} + \text{Prob}\left\{ |R_N(2)| \delta_{2,N}^{1/4} > \rho_N^{1/2} \right\} = P_1 + P_2,$$

say. Now $P_2 = o(\delta_{2,N}^{1/2})$ applying Chebychev inequality. For P_1, since $B_N = O(M^{-q})$,

$$\begin{aligned}
P_1 &= \text{Prob}\left\{ \left| \left(1 + B_N + \left[A_N + V_N^{-1} \delta_{N,2}^{1/2} u_2\right] \theta\right)^{-5/2} \right| \delta_{2,N}^{1/4} > \rho_N^{1/2} \right\} \\
&= \text{Prob}\left\{ \left| 1 + B_N + A_N \theta + V_N^{-1} \delta_{N,2}^{1/2} u_2 \theta \right| \delta_{2,N}^{-1/10} < \rho_N^{-1/5} \right\} \\
&\leq \text{Prob}\left\{ \left| 1 + O(M^{-q}) + R_N' \delta_{2,N}^{1/2} \right| \delta_{2,N}^{-1/10} < \rho_N^{-1/5} \right\},
\end{aligned}$$

where R_N' is a random variable with bounded moments of all orders. Now, as $N \to \infty$, for some positive constant $c > 0$, the last probability is not greater than

$$\text{Prob}\left\{ \left| c + R_N' \delta_{2,N}^{1/2} \right| < \delta_{2,N}^{1/10} \rho_N^{-1/5} \right\} \leq \text{Prob}\left\{ \left| R_N' \delta_{2,N}^{1/2} \right| > \frac{c}{2} \right\} = o(\delta_{2,N}^{1/2}),$$

applying again Chebychev inequality, since $\delta_{2,N}^{1/10} \rho_N^{-1/5} \to 0$ from the choice of ρ_N. □

2.11 Appendix: Bivariate regression

Here we analyze briefly the main differences that arise in the two regressors framework with respect to the single regressor case.

First, we need to study the behaviour of the function $\overline{Q}_M(\lambda)$ under the different sets of conditions. Then, under Assumption 2.7 and $M^{-1} + M^2 N^{-p} \to 0$, where now $p = \min\{p_1, p_2\}$, we can obtain results similar to Lemma 2.2:

- Case A. Both regressors satisfy Condition 2.1. In this case $\overline{Q}_M(\lambda)$ is tending to

$$\int_{\Pi} K_M(\lambda - \alpha) \left[g_{22}(\alpha) + \rho_{12}^2(0) g_{11}(\alpha) - \rho_{12}(0) g_{12}(\alpha) - \rho_{12}(0) g_{21}(\alpha) \right] d\alpha,$$

  and this converges to $g_{22}(\lambda) + \rho_{12}^2(0) g_{11}(\lambda) - \rho_{12}(0) g_{12}(\lambda) - \rho_{12}(0) g_{21}(\lambda)$.

- Case B. Both regressors satisfy Condition 2.2. Now there are only discrete contributions and then $\overline{Q}_M(\lambda)$ is tending to

$$\sum_j K_M(\lambda - \lambda_j) \left[\Delta_{22}(j) + \rho_{12}^2(0) \Delta_{11}(j) - \rho_{21}(0) \Delta_{21}(j) - \rho_{12}(0) \Delta_{12}(j) \right] = O(M),$$

  so in this case the estimate of the variance will have a slower rate of convergence.

- Case C. Because the first regressor satisfies Condition 2.2 and the second regressor, Condition 2.1, $\rho_{12}(0) = 0$, and $\overline{Q}_M(\lambda)$ is tending simply to $g_{22}(\lambda)$, as in the univariate case. Then, in the mean correction situation ($Z_{1t} = 1$) the variance estimate of a coefficient corresponding to a regressor with continuous spectral distribution function shows exactly the same first order asymptotic properties as when the mean correction is not performed.

We now check step by step several results about the variance estimates and the studentized least squares estimate.

2.11.1 Bias

From the previous discussion it is immediate to check that the proof for the bias of the estimate $\widehat{V}_N$ remains valid, since of course $\overline{S}_N(\lambda)$ is integrable. Then under the same set of Assumptions (2.1, 2.2, 2.3, 2.5 and 2.7, $M^{-1} + MN^{-1} \to 0$), we have in exactly the same way as in Lemma 2.3,

$$\mathrm{E}[\widehat{V}_N] - V_N = a_N N^{-1} + b_N M^{-q} + O(M^{-q-\varrho}),$$

where now

$$b_N = -\omega_q\,(-1)^{q/2}\,2\pi\int_{\Pi} f^{(d)}(\lambda)\overline{S}_N(\lambda)d\lambda,$$

and a_N has the same modification in terms of $S_N(\cdot)$.

2.11.2 Cumulants

As we have discussed previously, the sequence $S_N(r)$ will have the same properties as the sequences $R_{i,j}(r)$. Therefore, the bounds employed for the one regressor case are still valid, but taking the worst ones, in the sense that with one or more of the regressors satisfying Condition 2.2 we should always employ the bounds in terms of M, using the properties of ω.

In exactly the same way as before (see Lemma 2.4), we have for the variance of $\widehat{V}_N$ that [with Assumptions 2.1, 2.2, 2.5, 2.7 and $M^{-1}+N^{-p}M^2\to 0$]

$$\delta_{N,2} \overset{def}{=} \mathrm{Var}[\widehat{V}_N] = \frac{2}{N^2}\mathrm{Trace}\left[(\mathbf{\Sigma}_N\mathbf{Q}_M)^2\right] = \frac{2(2\pi)^3}{N}\int_{\Pi} f^2(\lambda)\overline{Q}_M^2(\lambda)d\lambda + O\left(N^{-2}M^2\right). \quad (2.55)$$

Now it can be interesting to study the limit of $\int f^2(\lambda)\overline{Q}_M^2(\lambda)d\lambda$ in the different situations we are considering.

- Case A. The integral tends to $\int f^2(\lambda)\left[g_{22}(\lambda)+\rho_{12}^2(0)g_{11}(\lambda)-2\rho_{12}(0)g_{12}(\lambda)\right]^2 d\lambda = O(1)$, so the variance of $\widehat{V}_N$ is of order N^{-1}.

- Case B. Since now the asymptotic leading term of $\overline{Q}_M(\lambda)$ is a linear combination of the kernel K_M at different frequencies, the variance now is of order M/N.

- Case C. Since $\rho_{i,j}(0)=0$, to have a variance of $\widehat{V}_N$ of order N^{-1} we need to check that in fact $R_{i,j}(0)\to 0$ fast enough, because in the variance there are typical terms like (up to a constant),

  $$\frac{1}{N^2}R_{1,2}^4(0)\sum_{\mathbf{r}}\gamma(r_2-r_1)\omega\left(\frac{r_3-r_2}{M}\right)R_{1,1}(r_3-r_2)\gamma(r_4-r_3)\omega\left(\frac{r_1-r_4}{M}\right)R_{1,1}(r_1-r_4).$$

  The sum in the previous expression tends to

  $$MN\,\|K\|_2^2\sum_j\Delta_{11}(j)\,f^2(\lambda_j),$$

and in order to have this term negligible with respect to the leading term in the variance we need

$$R^4_{1,2}(0)M \to 0, \tag{2.56}$$

but this holds since the right hand side of (2.56) is $O(N^{-(p_1+p_2)}M) = O(N^{-2p}M) = o(1)$ if $N^{-p}M^2 \to 0$.

We have to consider also other terms, with typical sums in $R^2_{1,2}(0)\sum R_{1,1}R_{2,2}$, $R^3_{1,2}(0)\sum R_{1,2}R_{2,2}$, $R^2_{1,2}(0)\sum R_{1,1}R_{1,2}$ and $R^2_{1,2}(0)\sum R_{1,2}R_{2,1}$, but those lead to terms in $\overline{Q}_M$ with at most one function K_M, so the integral in the variance $\delta_{N,2}$ is always $O(1)$ and they do not contribute to the leading term.

The same type of comments apply for all the higher order (cross) cumulants and for the joint characteristic function, so results similar to Theorem 2.1, 2.2 and 2.3 are possible under the same assumptions. The only step that is slightly different, is the residual effect in the estimate $\widetilde{V}_N$, but a result parallel to Lemma 2.5 can be obtained in the following lines.

2.11.3 Residual approximation

Using the OLS estimates of α and β, we get immediately that

$$\widehat{X}_t = X_t - Z_{1,t}\frac{\sum_{t=1}^N (Z_{1,t} - \overline{Z}_1 Z_{2,t})X_t}{\sum_{t=1}^N (Z_{1,t} - \overline{Z}_1 Z_{2,t})Z_{1,t}} - Z_{2,t}\frac{\sum_{t=1}^N (Z_{2,t} - \overline{Z}_2 Z_{1,t})X_t}{\sum_{t=1}^N (Z_{2,t} - \overline{Z}_2 Z_{1,t})Z_{2,t}},$$

and with straightforward algebra this is equal to

$$X_t - \frac{1}{(1-R^2_{1,2}(0))}\left(\frac{\sum_t Z_{1,t}X_t}{d_{1,N}}h^\star_t + \frac{\sum_t Z_{2,t}X_t}{d_{2,N}}h_t\right)$$

where $h^\star_t$ is defined as h_t but interchanging Z_1 and Z_2. Next,

$$\widetilde{V}_N = \widehat{V}_N - 2\xi_N(1) + \xi_N(2),$$

where

$$\begin{aligned}
\xi_N(1) &= \frac{1}{N(1-R^2_{1,2}(0))}\sum_{t_1,t_2} X_{t_1} Q_M(t_1-t_2)\left(\frac{\sum_r Z_{1,r}X_r}{d_{1,N}}h^\star_{t_2} + \frac{\sum_r Z_{2,r}X_r}{d_{2,N}}h_{t_2}\right)\\
&= \frac{1}{N(1-R^2_{1,2}(0))}\sum_{t_1,r} X_{t_1}\sum_{t_2} Q_M(t_1-t_2)\left(\frac{Z_{1,r}}{d_{1,N}}h^\star_{t_2} + \frac{Z_{2,r}}{d_{2,N}}h_{t_2}\right)X_r
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{N(1-R_{1,2}^2(0))}\mathbf{X}'\mathbf{Q}_M\left\{\mathbf{H}_\star\frac{\mathbf{Z}_1'}{d_{1,N}}+\mathbf{H}\frac{\mathbf{Z}_2'}{d_{2,N}}\right\}\mathbf{X} \\
&= \frac{1}{N(1-R_{1,2}^2(0))}\mathbf{X}'\mathbf{Q}_M\mathbf{J}_N\mathbf{X},
\end{aligned}
$$

say, where the vector $\mathbf{H}_\star$ is defined similarly to $\mathbf{H}$ but with $h_t^\star$. Therefore for each element of the $N \times N$ matrix $\mathbf{J}_N$, we have

$$
\max_{i,j}|\mathbf{J}_{i,j}| = O(N^{-p}).
$$

Equivalently,

$$
\begin{aligned}
\xi_N(2) &= \frac{1}{N(1-R_{1,2}^2(0))^2}\sum_{t_1,t_2}X_{t_1}\left(\frac{\sum_t Z_{1,t}X_t}{d_{1,N}}h_{t_2}^\star+\frac{\sum_t Z_{2,t}X_t}{d_{2,N}}h_{t_2}\right)Q_M(t_1-t_2) \\
&\quad\times\left(\frac{\sum_r Z_{1,r}X_r}{d_{1,N}}h_{t_2}^\star+\frac{\sum_r Z_{2,r}X_r}{d_{2,N}}h_{t_2}\right)X_{t_2} \\
&= \frac{1}{N(1-R_{1,2}^2(0))^2}\mathbf{X}'\mathbf{J}_N'\mathbf{Q}_M\mathbf{J}_N\mathbf{X},
\end{aligned}
$$

Consequently, given the structure of the matrix $\mathbf{J}_N$, we can proceed as in the proof of Lemma 2.5 and obtain that the s-th cumulants of $\xi_N(1)$ and $\xi_N(2)$ are $O\left(\delta_{N,2}^s\right)$, obtaining an equivalent approximation as for the single regressor model in Lemma 2.5, with $N^{-p}M^2 \to 0$.

Chapter 3

Edgeworth expansions for spectral density estimates and studentized sample mean

3.1 Introduction

In this chapter we analyze higher order asymptotic properties of nonparametric estimation of the variance of the sample mean for autocorrelated observations. Our aim is to investigate the effects of studentization techniques based on nonparametric weighted-autocovariance spectral density kernel estimates at the origin. This technique was proposed by Jowett (1954), and Hannan (1957) made the connection with the spectral analysis literature. As we have seen previously, these ideas extend to more general models and are of wide use in several statistical inference problems. The studentization requires only consistent estimates, and nonparametric ones are robust to different types of misspecification. However, the performance of these estimates depends crucially on the bandwidth employed and, typically, have slower rates of converge than parametric ones. Though this does not affect the asymptotic distribution of the studentized statistic, little is known about how the use of nonparametric estimates affects higher order properties of the distribution of the studentized statistic.

With respect to the previous chapter, we concentrate here in the location problem. Therefore, under Gaussianity assumptions, only conditions on the spectral density around the zero frequency are relevant. We obtain asymptotic expansions of the Edgeworth type for the distribution of the spectral estimate and the studentized mean. Several corrections for the first order approximations are found, trying to correct for possible non optimal choices of the bandwidth number, and we show how to estimate them consistently.

Our higher order theory for nonparametric spectral estimates is based on Bentkus and Rudzkis' (1982) work. They obtained asymptotic expansions and large deviation theorems for the distribution of nonparametric spectral density estimates for zero mean Gaussian sequences with bounded spectral densities. We cover also spectral densities that might have singularities due, for example, to seasonality or cycles. We express our conditions in terms of smoothness of the spectral density of the time series, but only in an interval of the zero frequency. Nonparametric estimation of functions with possibly inhomogeneous smoothness characteristics has also been considered by Lepskii and Spokoiny (1995) among others. They applied the projective adaptive procedure of Lepskii (1991) to a "signal+noise" model on degenerating intervals around the point of interest, finding a trade-off between accuracy and adaptive properties.

The flexibility on the dependence assumptions has to be compensated with strong distributional assumptions, like Gaussianity. This is a serious limitation, but allows us to distinguish clearly the effects due to the variance estimation from those related with the basic standardized statistic.

We use kernel functions for the nonparametric smoothing with finite support spectral window, in order to avoid leakage from other frequencies. However, it is possible to use a kernel function with infinite support if we are willing to make further assumptions over the integrability of the spectrum, which would imply in turn restrictions on the size of any possible singularity. Similar results to those presented here can be obtained under stronger assumptions (like the summability conditions on the covariance sequence of the previous chapter), but this would lead to restrictions on the global smoothness of the spectrum across all frequencies.

Relaxing the Gaussianity assumption with Götze and Hipp (1983) conditions, Götze and Künsch (1995) have shown the second order correctness of Künsch (1989) block bootstrap for the studentized version of time series statistics that can be expressed as sample means. The strong mixing conditions used by them imply restrictions on the serial dependence structure much stronger that ours, and these conditions do not seem to allow for expansions of the distribution of smoothed nonparametric estimates. The reason is that multivariate asymptotic expansions are only valid for fixed dimension vectors, so they do not extend to nonparametric estimates of parameters of the whole distribution of a stationary sequence.

This chapter is organized as follows. In next section we give the main definitions and assumptions which we will use throughout. We consider in Section 3.3 the distribution of the nonparametric estimate of the spectral density. In Section 3.4 we analyze the joint distribution of the variance estimate and the sample mean. In Section 3.5 we give an Edgeworth expansion for the studentized sample mean and give consistent estimates for the higher order correction terms. Finally, we extend the previous results to obtain a third order approximation in Section 3.6. All proofs are given at the end of the chapter, in three appendices.

3.2 Assumptions and Definitions

In this section we set our general framework and assumptions. Let $\{X_t\}$ be a stationary Gaussian sequence with $E[X_t] = \mu$, autocovariance function

$$\gamma(r) = E\left[(X_t - \mu)(X_{t+r} - \mu)\right],$$

and spectral density $f(\lambda)$ defined by, $\Pi = (-\pi, \pi]$,

$$\gamma(r) = \int_{\Pi} f(\lambda) e^{ir\lambda} d\lambda,$$

satisfying $0 < f(0) < \infty$.

Let $\mathbf{X} = (X_1, \ldots, X_N)'$ be a vector of N consecutive observations of X_t. Then $\mathbf{X}$ has a multivariate normal distribution $\mathcal{N}(\mu, \Sigma_N)$, where

$$[\Sigma_N]_{r,g} = \gamma(r - g), \qquad r, g = 1, \ldots, N.$$

Without loss of generality we take $\mu = 0$. Let $\overline{X} = N^{-1}\sum_{j=1}^{N} X_j$ be the sample mean of the N observations of X_t and denote

$$V_N = \mathrm{Var}[\sqrt{N}\,\overline{X}] = \sum_{j=1-N}^{N-1} \left(1 - \frac{|j|}{N}\right) \gamma(j).$$

Then for all N we have

$$\frac{\sqrt{N}\,\overline{X}}{\sqrt{V_N}} \rightsquigarrow \mathcal{N}(0,1). \tag{3.1}$$

Since V_N is the Césaro sum of the Fourier coefficients of $f(\lambda)$ at the origin, if $f(\lambda)$ is continuous at 0 then $\lim_{N\to\infty} V_N = 2\pi f(0)$. If $\widehat{f}(0)$ is a consistent estimator, such that $\widehat{f}(0) \to_P f(0)$, then

$$Y_N \stackrel{def}{=} \frac{\sqrt{N}\,\overline{X}}{\sqrt{2\pi \widehat{f}(0)}} \to_d \mathcal{N}(0,1).$$

Assume that the mean $\mu = 0$ is known. Later we drop this assumption (see Section 3.5.1). Defining the (biased) estimator of the autocovariance function as

$$\widehat{\gamma}(\ell) = \frac{1}{N} \sum_{1\le t, t+\ell \le N} X_t X_{t+\ell}, \qquad \ell = 0, \pm 1, \ldots, \pm(N-1),$$

we consider the weighted-autocovariance type nonparametric spectral estimate

$$\begin{aligned} \widehat{f}(0) &= \frac{1}{2\pi} \sum_{\ell=1-N}^{N-1} \omega\left(\frac{\ell}{M}\right) \widehat{\gamma}(\ell) \\ &= \mathbf{X}' \left(\frac{W_M}{2\pi N}\right) \mathbf{X}, \end{aligned}$$

where W_M is the $N \times N$ matrix

$$[W_M]_{r,g} = \omega\left(\frac{r-g}{M}\right) = \int_{\Pi} K_M(\lambda) e^{i(r-g)\lambda} d\lambda \tag{3.2}$$

and $K_M(\lambda)$ is a kernel function with smoothing number M, constructed in the following way. Define, for a sequence of positive integers $M = M_N$ such that $M \to \infty$ and $NM^{-1} \to \infty$ as $N \to \infty$, and for an even, integrable function K which integrates to one:

$$K_M(\lambda) = M \sum_{j=-\infty}^{\infty} K(M[\lambda + 2\pi j]),$$

so $K_M(\lambda)$ is periodic of period 2π, even, integrable and

$$\int_{-\pi}^{\pi} K_M(\lambda) d\lambda = 1. \tag{3.3}$$

Also we have $\omega(r) = \int_{-\infty}^{\infty} e^{irx} K(x)dx$ and $\omega(0) = 1$.

We introduce the following assumptions about the dependence structure of the time series X_t and about the nonparametric estimate $\widehat{f}(0)$. They are stated in terms of the spectral density f, the function K and the lag-number M.

Assumption 3.1 *$f(\lambda)$ has d continuous derivatives ($d \geq 2$) in an interval $[-\epsilon, \epsilon]$ around the origin, for any $\epsilon > 0$, and the dth derivative satisfies a Lipschitz condition of order ϱ, $0 < \varrho \leq 1$.*

Assumption 3.2 *The spectral density $f(\lambda) \in L_p$, for some $p > 1$, i.e.*

$$\|f\|_p^p \stackrel{def}{=} \int_{-\pi}^{\pi} f^p(\lambda)d\lambda < \infty. \tag{3.4}$$

Assumption 3.3 *$K(x)$ is a bounded, even, integrable function, $-\pi \leq x \leq \pi$, and zero elsewhere, with*

$$\int_{-\pi}^{\pi} K(x)dx = 1. \tag{3.5}$$

Assumption 3.4 *$K(x)$ satisfies a uniform Lipschitz condition (of order 1) in $[-\pi, \pi]$.*

Assumption 3.5 *Defining for $j = 0, 1, \dots, d$, $d \geq 2$ and $r = 1, 2, \dots$*

$$K_j^{(r)} \stackrel{def}{=} \int_{-\infty}^{\infty} x^j \left[K(x)\right]^r dx\,,$$

the function K satisfies

$$K_j^{(1)} \begin{cases} = 1 & j = 0 \\ = 0 & j < d \\ \neq 0 & j = d. \end{cases}$$

Assumption 3.6 *$M^{-1} + MN^{-1} \to 0$, as $N \to \infty$.*

This assumption is necessary for the definition of the estimate $\widehat{f}$. Sometimes we will want to make explicit the exact form of the lag number M satisfying Assumption 3.6 in terms of N:

Assumption 3.7 *$M = C \cdot N^q$, with $0 < q < 1$, and a constant C, $0 < C < \infty$.*

Assumption 3.1 is satisfied with $\varrho = 1$ if f has bounded $(d+1)$th derivative around 0. For most of the results, the continuity of the dth derivative would be enough, but in order to prove that the bias estimation does not affect the asymptotic distribution we need an explicit rate for the bias error (see Lemma 3.7). Assumption 3.1 would be implied by conditions like (see Assumption 2.1)

$$\sum_{j=-\infty}^{\infty} |j|^{d+\varrho}|\gamma(j)| < \infty.$$

However these conditions imply global smoothness conditions on the spectral density f and are too strong when we want to focus on a particular frequency, restricting considerably the class of admissible serial dependence models.

Assumption 3.2 imposes some restrictions on the density f beyond the origin. In fact a $p > 1$ arbitrarily close to 1 will be enough for all our results (see Lemma 3.5 below). We can obtain a central limit theorem for $\widehat{f}(0)$ by the method of cumulants, assuming only $p = 1$, implied by stationarity (see expression (3.10) below).

From Assumption 3.3, the function $\omega(r)$ defined by (3.2) is even and bounded. Also, Assumption 3.3 imposes a kernel function K defined in a finite support, so it is not necessary to make further assumptions about its tail behaviour (the same results could be obtained with a support of the form $[-\tau, \tau]$ for any $0 < \tau \leq \pi$). Asymptotically we only use information around $\lambda = 0$ so we do not need to impose strong conditions about f beyond this frequency. Assumption 3.4 is needed to evaluate the cumulants of the estimate $\widehat{f}(0)$ and it is satisfied for most kernels used in practice satisfying Assumption 3.3.

Examples of kernels satisfying Assumptions 3.3, 3.4 and 3.5 for $d = 2$ are the uniform and the Barlett-Priestley. For $d > 2$, possible examples are the following ones constructed from the uniform density in $[-\pi, \pi]$:

$$\begin{aligned}
\text{For } \; d=4: \quad K_4(\lambda) &= \frac{9}{8\pi} - \frac{15}{8\pi^3}\lambda^2. \\
\text{For } \; d=6: \quad K_6(\lambda) &= \frac{225}{128\pi} - \frac{525}{64\pi^3}\lambda^2 + \frac{945}{128\pi^4}\lambda^4,
\end{aligned}$$

with lag windows, respectively:

$$\begin{aligned}
\omega_4(r) &= \frac{3}{2\pi^3 r^3}\left[5\pi r \cos \pi r - (5 - \pi^2 r^2) \sin \pi r\right], \\
\omega_6(r) &= \frac{15}{8\pi^5 r^5}\left[(-189\pi r + 14\pi^3 r^3) \cos \pi r + (189 + 77\pi^2 r^2 + \pi^4 r^4) \sin \pi r\right].
\end{aligned}$$

To analyze the joint distribution of the sample mean and the spectral estimate of its variance, it is convenient to work with standardized statistics with zero mean and with the same rate of convergence. Let's introduce some more notation. Suppose that the estimate $\widehat{f}(0)$ is $\sqrt{N/M}$-consistent. Defining

$$q_1 = \sqrt{M}\frac{\overline{X}}{\sqrt{V_N}},$$

we obtain $\sqrt{N/M}q_1 \rightsquigarrow N(0,1)$, from (3.1). We can write $\sqrt{N/M}q_1$ as a linear form in the vector $\mathbf{X}$, $\xi_N'\mathbf{X}$ say, where

$$\xi_N = \frac{\mathbf{1}}{\sqrt{NV_N}},$$

and $\mathbf{1}$ is the $N \times 1$ vector $(1,\ldots,1)'$.

The denominator of Y_N can be written in terms of one statistic with the same rate of convergence, zero mean and unit variance,

$$\begin{aligned} Y_N &= \frac{\sqrt{N}\,\overline{X}}{\sqrt{V_N}}\left(1+\frac{E[2\pi\widehat{f}(0)]-V_N}{V_N}+\frac{2\pi\widehat{f}(0)-E[2\pi\widehat{f}(0)]}{V_N}\right)^{-1/2} \\ &= \sqrt{N/M}q_1\,(1+b_N+\overline{q}_2)^{-1/2} \\ &= \sqrt{N/M}q_1\,(1+b_N+\sigma_N q_2)^{-1/2} \end{aligned}$$

where

$$b_N = \frac{E[2\pi\widehat{f}(0)]-V_N}{V_N}$$

is the bias of the spectral estimate with respect to the true variance of $\sqrt{N}\,\overline{X}$, multiplied by $2\pi/V_N$. Here, σ_N^2 is the variance of the normalized spectral estimate

$$\sqrt{N/M}\,\overline{q}_2 \overset{def}{=} \sqrt{N/M}\,\frac{2\pi\widehat{f}(0)-E[2\pi\widehat{f}(0)]}{V_N},$$

in a way that $\sqrt{N/M}q_2 = \sqrt{N/M}\,\overline{q}_2/\sigma_N$ has mean zero and variance 1. Therefore $\sqrt{N/M}q_2$ is a centered quadratic form in the vector $\mathbf{X}$,

$$\begin{aligned} \sqrt{N/M}q_2 &= \mathbf{X}'\left(\frac{1}{\sqrt{NM}}\frac{W_M}{\sigma_N V_N}\right)\mathbf{X} - E\left[\mathbf{X}'\left(\frac{1}{\sqrt{NM}}\frac{W_M}{\sigma_N V_N}\right)\mathbf{X}\right] \\ &= \mathbf{X}'Q_N\mathbf{X} - E[\mathbf{X}'Q_N\mathbf{X}] \end{aligned}$$

where the matrix Q_N is given by

$$Q_N = \frac{1}{\sqrt{NM}}\frac{W_M}{\sigma_N V_N}.$$

Set the random vector $\mathbf{u} = (u_1, u_2)' = \sqrt{N/M}\mathbf{q} = \sqrt{N/M}(q_1, q_2)'$, obtaining finally

$$Y_N = u_1 \left(1 + b_N + \sigma_N u_2 \sqrt{M/N}\right)^{-1/2}.$$

As $\mathbf{X}$ is $N(0, \Sigma_N)$ distributed, the joint characteristic function of $\mathbf{u} = (u_1, u_2)'$ is

$$\psi(t_1, t_2) = |I - 2it_2\Sigma_N Q_N|^{-1/2} \exp\left\{-\frac{1}{2}t_1^2 \xi_N' (I - 2it_2\Sigma_N Q_N)^{-1} \Sigma_N \xi_N - it_2 E\right\}$$

where $E = E[\mathbf{X}'Q_N\mathbf{X}] = \text{Trace}[\Sigma_N Q_N]$. Therefore, due to the normalizations, $\mathbf{u}$ has identity covariance matrix, and from its cumulant generating function

$$\begin{aligned}\varphi(t_1, t_2) &= \log \psi(t_1, t_2) \\ &= -\frac{1}{2}\log|I - 2it_2\Sigma_N Q_N| - \frac{1}{2}t_1^2 \xi_N' (I - 2it_2\Sigma_N Q_N)^{-1} \Sigma_N \xi_N - it_2 E,\end{aligned}$$

the only bivariate cumulants different from zero are

- $\kappa[0, s] = 2^{s-1}(s-1)!\,\text{Trace}[(\Sigma_N Q_N)^s], \ s > 1.$
- $\kappa[2, s] = 2^s s!\,\xi_N'(\Sigma_N Q_N)^s \Sigma_N \xi_N, \ s > 0.$

3.3 Distribution of the nonparametric spectral estimate

In this section we analyze the asymptotic distribution of the nonparametric spectral estimate $\widehat{f}(0)$. These results constitute an extension of the work of Bentkus and Rudzkis (1982), in the sense that we do not need to assume boundedness of the spectral density at frequencies apart from the origin. In Section 3.5.1 we also evaluate the error incorporated in the estimation of $f(0)$ when we use the mean corrected series $X_t - \overline{X}$ (or, like in the previous Chapter, the least squares residuals).

First we give two lemmas about the bias of the estimate $\widehat{f}(0)$ for V_N. The first one is a standard result in Fourier analysis about the convergence of the Césaro series of functions with bounded derivative. The logarithm term could be eliminated assuming summability conditions on the autocovariance sequence of the spectral density f.

Lemma 3.1 *Under Assumption 3.1, $d = 1$, $\varrho = 0$,*

$$V_N - 2\pi f(0) = O\left(N^{-1}\log N\right).$$

Lemma 3.2 *Under Assumptions 3.1, 3.3, 3.5 and 3.6*

$$E[\widehat{f}(0)] - f(0) - \frac{f^{(d)}(0)}{d!} K_d^{(1)} M^{-d} = O\left(N^{-1} \log N + M^{-d-\varrho}\right).$$

where $f^{(d)}(0)$ is the dth derivative of $f(\lambda)$ evaluated at $\lambda = 0$.

Then from Lemmas 3.1 and 3.2 we have that under Assumptions 3.1, 3.3, 3.5 and 3.6, as M increases,

$$\begin{aligned} b_N &= \left(\frac{1}{2\pi f(0)} + O(N^{-1}\log N)\right)\left(2\pi \frac{f^{(d)}(0)}{d!} K_d^{(1)} M^{-d} + O(M^{-d-\varrho} + N^{-1}\log N)\right) \\ &= b_1 M^{-d} + O(M^{-d-\varrho} + N^{-1}\log N), \end{aligned}$$

where

$$b_1 = \frac{f^{(d)}(0)\, K_d^{(1)}}{d!\, f(0)}.$$

We will need to consider the following quantity

$$\begin{aligned} \delta_N &= (1 + b_N)^{-1/2} \\ &= 1 - \frac{1}{2} b_1 M^{-d} + O(M^{-d-\varrho} + N^{-1}\log N), \end{aligned}$$

under the same set of assumptions, since for all M big enough $1 + b_N > 0$. Also

$$\begin{aligned} \delta_N^3 &= 1 - \frac{3}{2} b_1 M^{-d} + O(M^{-d-\varrho} + N^{-1}\log N) \\ \delta_N^5 &= 1 + O(M^{-d} + N^{-1}\log N). \end{aligned}$$

Now we study the cumulants of $\widehat{f}(0)$. We obtained for $s \geq 2$

$$\kappa[0, s] = 2^{s-1}(s-1)!\,\mathrm{Trace}\left[(\Sigma_N Q_N)^s\right]$$

where we can write

$$\mathrm{Trace}\left[(\Sigma_N Q_N)^s\right] = \frac{1}{(\sigma_N V_N)^s}(MN)^{-s/2}\,\mathrm{Trace}\left[(\Sigma_N W_M)^s\right].$$

Proposition 3.1 *Under Assumptions 3.1, 3.3, 3.4, $M^{-1} + N^{-1} M \log^{2s-1} N \to 0$, for $s \geq 2$,*

$$\mathit{Trace}\left[(\Sigma_N W_M)^s\right] = N(2\pi)^{2s-1}\sum_{j=0}^{d} L_s(j) M^{s-1-j} + O\left(N M^{s-1} e_N(s) + N M^{s-1-d-\varrho}\right),$$

where $e_N(s) = [N^{-1}M\log^{2s-1}N)]^{1/2}$ *and*

$$L_s(j) = \frac{1}{j!}\left(\frac{d}{d\lambda}\right)^j f^s(\lambda)\bigg|_{\lambda=0} K_j^{(s)}$$

with $|L_s(j)| < \infty$ *and, as* $K_j^{(s)}$, *the constants* $L_s(j)$ *are only different from zero for* j *even* $(j=0,\ldots,d)$.

The proof is based on the definition of a multivariate version of Fejér kernel (see Appendix in Section 3.8) and on the fact that, given the compact support of K, asymptotically we only smooth for frequencies in a band around zero.

Therefore, under the conditions of Proposition 3.1, the normalized cumulants result as

$$\begin{aligned}\bar{\kappa}[0,s] &\stackrel{def}{=} \left(\frac{N}{M}\right)^{\frac{s-2}{2}} \kappa[0,s] \\ &= \frac{2^{s-1}(s-1)!\,(2\pi)^{2s-1}}{(\sigma_N V_N)^s}\sum_{j=0}^{d} L_s(j)M^{-j} + O\Big(e_N(s) + M^{-d-\varrho}\Big). \qquad (3.6)\end{aligned}$$

Depending on the concrete asymptotic relation between M and N, some values in the expansion can be included in the error term, since we have only assumed that $e_N(s) \to 0$ as $N \to \infty$. Now we can apply Proposition 3.1 to evaluate σ_N^2 under the same set of assumptions ($s = 2$). As σ_N^2 is the variance of $\sqrt{N/M}2\pi\widehat{f}(0)/V_N$, a quadratic form in the Gaussian vector X,

$$\begin{aligned}\sigma_N^2\frac{V_N^2}{4\pi^2} &= \frac{N}{M}\frac{2}{(2\pi N)^2}\mathrm{Trace}[(\Sigma_N W_M)^2] \\ &= 4\pi\sum_{j=0}^{d} L_2(j)M^{-j} + O\Big(e_N(2) + M^{-d-\varrho}\Big) \qquad (3.7)\end{aligned}$$

from the proof of Proposition 3.1, if now $N^{-1}M\log^3 N \to 0$. For example

$$\begin{aligned}L_2(0) &= f^2(0)K_0^{(2)} = f^2(0)\|K\|_2^2 \qquad (3.8)\\ L_2(1) &= 0 \\ L_2(2) &= \frac{1}{2}K_2^{(2)}\left[\frac{d^2}{d\lambda^2}f^2(\lambda)\right]_{\lambda=0}\end{aligned}$$

and $4\pi f^2(0)K_0^{(2)} = 4\pi f^2(0)\|K\|_2^2$ is the asymptotic variance of $\sqrt{N/M}\widehat{f}(0)$.

Now as $0 < L_2(0) < \infty$ and all $L_2(j)$ are fixed constants independent of N or M, we can write for some constants J_s

$$\left(\sigma_N \frac{V_N}{2\pi}\right)^{-s} = (4\pi)^{-s/2} \sum_{j=0}^{d} J_s(j) M^{-j} + O\left(e_N(2) + M^{-d-\varrho}\right), \tag{3.9}$$

where for example,

$$\begin{aligned} J_s(0) &= L_2(0)^{-s/2} \\ J_s(1) &= 0 \\ J_s(2) &= -\frac{s}{2} L_2(0)^{-\frac{2+s}{2}} L_2(2). \end{aligned}$$

Now, denoting $C(0,s) = (4\pi)^{\frac{s-2}{2}}(s-1)!$ we can obtain from (3.6) and (3.9) the following expansion in powers of M^{-1} for the normalized cumulants, under the conditions of Proposition 3.1,

$$\begin{aligned} \bar{\kappa}[0,s] &= C(0,s)\left[\sum_{j=0}^{d} J_s(j) M^{-j} + O\left(e_N(2) + M^{-d-\varrho}\right)\right] \\ &\quad \times \left[\sum_{j=0}^{d} L_s(j) M^{-j} + O\left(e_N(s) + M^{-d-\varrho}\right)\right] \\ &= C(0,s) \sum_{j=0}^{d} \Gamma_s(j) M^{-j} + O\left(e_N(s) + M^{-d-\varrho}\right), \end{aligned} \tag{3.10}$$

where

$$\Gamma_s(j) = \sum_{t=0}^{j} J_s(t) L_s(j-t)$$

are constants not depending on N or M, and depending only on the spectral density f and on the Kernel K. Immediately it follows that

$$\begin{aligned} \Gamma_s(0) &= J_s(0) L_s(0) \\ \Gamma_s(1) &= 0 \\ \Gamma_s(2) &= J_s(0) L_s(2) + J_s(2) L_s(0). \end{aligned}$$

We also need expressions for σ_N and its powers. First, for $j = \pm 1, \pm 2, \ldots$, from Lemma 3.1

$$\left[\frac{V_N}{2\pi}\right]^j = f^j(0) + O(N^{-1} \log N).$$

Then, as $0 < 4\pi L_2(0) < \infty$, and $\sigma_N V_N > 0$ for all M big enough, under the conditions of Proposition 3.1 and $N^{-1}M\log^3 N \to 0$, for constants $\Theta(j)$,

$$\sigma_N = \frac{\sqrt{4\pi}}{f(0)} \sum_{j=0}^{d} \Theta(j) M^{-j} + O\left(e_N(2) + M^{-d-\varrho}\right)$$

and

$$\sigma_N^2 = \frac{4\pi}{f(0)^2} \sum_{j=0}^{d} L_2(j) M^{-j} + O\left(e_N(2) + M^{-d-\varrho}\right),$$

$$\sigma_N^3 = \frac{(4\pi)^{3/2}}{f(0)^3} \Theta(0)^3 + O\left(e_N(2) + M^{-d-\varrho}\right).$$

For example,

$$\begin{aligned} \Theta(0) &= L_2(0)^{1/2} \\ \Theta(1) &= 0 \\ \Theta(2) &= \frac{1}{2} L_2(0)^{1/2} L_2(2). \end{aligned}$$

From the previous results we can justify an optimal choice of the smoothing number M in terms of the value which minimizes the Mean Square Error (MSE) of the spectral estimate, $E[(\widehat{f}(\lambda) - f(\lambda))^2]$, at $\lambda = 0$. Since in this case we are only interested in the estimation of the spectral density at the origin, it is sensible to use local rules for the choice of the bandwidth.

Assuming conditions 3.1, 3.3, 3.4, 3.5, $M^{-1} + N^{-1}M\log^3 N \to 0$, from Lemma 3.2 we obtain an expression for the bias and from (3.8)

$$\frac{N}{M} \mathrm{Var}\left[\widehat{f}(0)\right] = 4\pi f^2(0) \|K\|_2^2 + O\left(e_N(2) + M^{-2}\right).$$

Then the smoothing or lag number which minimizes asymptotically the MSE is of the form

$$M = c \cdot N^{1/(2d+1)}, \qquad 0 < c < \infty,$$

in order to make squared bias and variance of the same order of magnitude. The optimal constant $c^\star$ that provides the minimum asymptotic MSE is

$$c^\star(f, K) = \left[\frac{2d}{4\pi} \left(\frac{f^{(d)}(0)\, K_d^{(1)}}{d!\, f(0)\, \|K\|_2} \right)^2 \right]^{\frac{1}{2d+1}}. \tag{3.11}$$

Obviously this choice is unfeasible as depends on the unknown spectral density, but initial, consistent estimates of $f(0)$ and $f^{(d)}(0)$ can be plugged in. In next chapter we propose a local cross-validatory procedure that avoids this estimation.

3.4 Joint distribution of the spectral estimate and the sample mean

In this section we prove the validity of an Edgeworth expansion to approximate the distribution of the vector $\mathbf{u}$. Of course this will imply the validity of that expansion for the distribution of the spectral density estimate $\widehat{f}(0)$. First we study the cross cumulants of the form $(s>0)$:

$$\begin{aligned}
\kappa[2,s] &= 2^s s! \xi_N'(\Sigma_N Q_N)^s \Sigma_N \xi_N \\
&= 2^s s! \frac{(MN)^{-s/2}}{N} \frac{1}{V_N^{s+1}\sigma_N^s} \mathbf{1}'(\Sigma_N W_M)^s \Sigma_N \mathbf{1},
\end{aligned}$$

with similar techniques to those of Proposition 3.1.

Proposition 3.2 *Under Assumptions 3.1, 3.3, 3.4,* $M^{-1} + N^{-1}M\log^{2s+1}N \to 0$, *for* $s \geq 1$,

$$\mathbf{1}'(\Sigma_N W_M)^s \Sigma_N \mathbf{1} = N(2\pi)^{2s+1}[f(0)]^{s+1}[K_M(0)]^s + \left(M^{s+1}\log^{2s+1}N\right).$$

Then, under the conditions of Proposition 3.2 the normalized cumulants are equal to

$$\begin{aligned}
\bar{\kappa}[2,s] &\stackrel{def}{=} \left(\frac{N}{M}\right)^{s/2} \kappa[2,s] \\
&= \left[\frac{2\pi}{V_N\sigma_N}\right]^s \frac{2\pi f(0)}{V_N}(4\pi)^s s! f(0)^s K(0)^s + O\left(\frac{M}{N}\log^{2s+1}N\right),
\end{aligned}$$

as $K_M(0) = MK(0)$ given the compact support of K. Substituting the expansion for the value of $V_N\sigma_N$ and using Lemma 3.1, we obtain:

Lemma 3.3 *Under Assumptions 3.1, 3.3, 3.4,* $M^{-1} + N^{-1}M\log^{2s+3}N \to 0$, $s \geq 1$,

$$\begin{aligned}
\bar{\kappa}[2,s] &= \left[\frac{V_N\sigma_N}{2\pi}\right]^{-s}\left[1 + O(N^{-1}\log N)\right](4\pi)^s s! f(0)^s K(0)^s + O\left(\frac{M}{N}\log^{2s+1}N\right) \\
&= (4\pi)^{-s/2}\left[\sum_{j=0}^{d} J_s(j)M^{-j}\right](4\pi)^s s! f(0)^s K(0)^s + O\left(e_N(s) + M^{-d-\varrho}\right)
\end{aligned}$$

$$= \quad C(2,s)\sum_{j=0}^{d} J_s(j)M^{-j} + O\left(e_N(s) + M^{-d-\varrho}\right),$$

where $C(2,s) = (4\pi)^{s/2}s!f^s(0)K^s(0)$ *and the* $J_s(j)$ *are defined as before.*

In order to prove the validity of an Edgeworth series expansion for the joint distribution of the vector $\mathbf{u}$ we have to check that the characteristic function of the expansion approximates well the true characteristic function. We use the same two step procedure as in the previous chapter. However, here we include only the leading terms of the expansions for the cumulants of the joint distribution, incorporating the error of this approximation in the general error term for the estimate of the characteristic function.

Let's start constructing the approximation for $\psi(\mathbf{t})$. As in Taniguchi (1987, pp. 11-14) or Durbin (1980a), using the fact that only the cumulants of the form $\kappa[0,s]$ and $\kappa[2,s]$ are different from zero, we can write the generating cumulant function as

$$\log\psi(\mathbf{t}) \quad = \quad \frac{1}{2}\|i\mathbf{t}\|^2 + \sum_{s=3}^{\tau+1}\frac{(N/M)^{\frac{2-s}{2}}}{s!}\sum_{|\mathbf{r}|=s}\frac{s!}{r_1!r_2!}\bar{\kappa}[r_1,r_2](it_1)^{r_1}(it_2)^{r_2} + R_N(\tau), \tag{3.12}$$

where the vector $\mathbf{r}$ is of the form (r_1, r_2), with $r_1 \in \{0,2\}$ and $|\mathbf{r}| = r_1 + r_2$, and the remaining term R_N is of this form, if τ is even:

$$R_N(\tau) = \left(\frac{N}{M}\right)^{-\frac{\tau}{2}}\left[R_{0,\tau+2}(it_2)^{\tau+2} + R_{2,\tau}(it_1)^2(it_2)^{\tau}\right],$$

or of this other form, if τ is odd:

$$\begin{aligned} R_N(\tau) \quad &= \quad \left(\frac{N}{M}\right)^{-\frac{\tau}{2}}\frac{1}{(\tau+2)!}\left[\bar{\kappa}[0,\tau+2](it_2)^{\tau+2} + \frac{(\tau+2)(\tau+1)}{2}\bar{\kappa}[2,\tau](it_1)^2(it_2)^{\tau}\right] \\ &\quad + \left(\frac{N}{M}\right)^{-\frac{\tau+1}{2}}\left[R_{0,\tau+3}(it_2)^{\tau+3} + R_{2,\tau+1}(it_1)^2(it_2)^{\tau+1}\right], \end{aligned} \tag{3.13}$$

where the $R_{0,j}$ and $R_{2,j}$ are bounded. So we can write from (3.10) and Lemma 3.3

$$\begin{aligned} \log\psi(\mathbf{t}) \quad &= \quad \frac{1}{2}\|i\mathbf{t}\|^2 + \sum_{s=3}^{\tau+1}\frac{(N/M)^{\frac{2-s}{2}}}{s!}\left[\bar{\kappa}[0,s](it_2)^s + \frac{s(s-1)}{2}\bar{\kappa}[0,s-2](it_1)^2(it_2)^{s-2}\right] \\ &\quad + R_N(\tau) \\ &= \quad \frac{1}{2}\|i\mathbf{t}\|^2 + \sum_{s=3}^{\tau+1}\left(\frac{N}{M}\right)^{\frac{2-s}{2}}\left[B_N(s,\mathbf{t}) + \left\{(it_1)^s + (it_1)^2(it_2)^{s-2}\right\}O\left(e_N(s) + M^{-d-\varrho}\right)\right] \\ &\quad + R_N(\tau), \end{aligned}$$

where we have grouped terms in powers of M^{-1} in $B_N(s,\mathbf{t})$:

$$B_N(s,\mathbf{t}) = \frac{1}{s!}\sum_{j=0}^{d} M^{-j}\left\{C(0,s)\Gamma_s(j)(it_2)^s + \frac{s(s-1)}{2}C(2,s-2)J_{s-2}(j)(it_1)^2(it_2)^{s-2}\right\}.$$

We are interested in obtain an approximation for the characteristic function of the vector $\mathbf{u}$ based on its cumulant generating function. This approximation ($A(\mathbf{t},\tau)$, say) should have leading term $\exp\{\frac{1}{2}\|i\mathbf{t}\|^2\}$, multiplied by a polynomial in $\mathbf{t}$, depending on the cumulants of $\mathbf{u}$, N and M. For general τ, this approximation has this form

$$A(\mathbf{t},\tau) = \exp\left\{\frac{1}{2}\|i\mathbf{t}\|^2\right\}\left[1 + \sum_{j=3}^{\tau+1}\left(\frac{N}{M}\right)^{\frac{2-j}{2}}\sum_{\mathbf{r}}\prod_{n=3}^{\tau+1}[B_N(n,\mathbf{t})]^{r_n}\frac{1}{r_3!\cdots r_{\tau+1}!}\right]$$

where $\mathbf{r} = (r_3,\ldots,r_{\tau+1})$, $r_n \in \{0,1,\ldots\}$ and the summation extends for all the vectors $\mathbf{r}$ that satisfy the condition

$$\sum_{n=3}^{\tau+1}(n-2)r_n = j-2.$$

We will only need to keep terms in the expansions up to a certain power of $(N/M)^{-1/2}$. Thus, some of the terms in big powers of M^{-1} in $B_N(n,\mathbf{t})$ may be included in the general error term, without increasing its magnitude.

In the following we are going to concentrate in obtaining a second order Edgeworth expansion, $\tau = 2$, that is, including in $A(\mathbf{t},2)$ terms up to order $(N/M)^{-1/2}$. In Section 3.6 we consider a third order approximation with $\tau = 3$. Applying the general formula we would get

$$\exp\left\{\frac{1}{2}\|i\mathbf{t}\|^2\right\}\left[\ 1\ +\ B_N(3,\mathbf{t})\left(\frac{N}{M}\right)^{-1/2}\right].$$

However from the expression for $B_N(s,\mathbf{t})$ we can see that in the term of order $(N/M)^{-1/2}$ of $A(\mathbf{t},2)$ it is only necessary to keep the leading term (in M^0) in the expansion for the cumulants of order 3.

Lemma 3.4 *Under Assumptions 3.1, 3.3, 3.4, $M^{-1} + N^{-1}M\log^5 N \to 0$, there exists a positive number $\delta_1 > 0$ such that, for $\|\mathbf{t}\| \le \delta_1\sqrt{N/M}$ and a number $d_1 > 0$:*

$$|\psi(\mathbf{t}) - A(\mathbf{t},2)| \ \le\ \exp\{-d_1\|\mathbf{t}\|^2\}F(\|\mathbf{t}\|)O\left(\left(\frac{N}{M}\right)^{-1/2}\left[M^{-2} + e_N(3)\right] + \frac{M}{N}\right), \quad (3.14)$$

where F is a polynomial in $\mathbf{t}$ with bounded coefficients and

$$A(\mathbf{t},2) = \exp\left\{\frac{1}{2}\|i\mathbf{t}\|^2\right\}\left[1 + \left(\frac{M}{N}\right)^{-1/2}\frac{1}{3!}\left\{C(0,3)\Gamma_3(j)(it_2)^3 + C(2,1)J_1(j)(it_1)^2(it_2)\right\}\right].$$

Having approximated the characteristic function for values of $\mathbf{t}$ such that $||\mathbf{t}|| \leq \delta_1\sqrt{N/M}$, the following step is to study the behaviour of this function in the tails.

Lemma 3.5 *Under Assumptions 3.1, 3.2 (some $p > 1$), 3.3, 3.4, $M^{-1}+N^{-1}M\log^3 N \to 0$ there exists a positive constant $d_2 > 0$ such that for $||\mathbf{t}|| > \delta_1 m_N$,*

$$|\psi(t_1,t_2)| \leq \exp\left\{-d_2 m_N^2\right\} \tag{3.15}$$

where

$$m_N = \min\left\{\left(\frac{N}{M\log^2 N}\right)^{1/2}, N^{\frac{p-1}{p}}\right\} \to \infty \quad as \quad N \to \infty.$$

Let's introduce the notation

$$\begin{aligned}
P_N\{B\} &= \text{Prob}\{\mathbf{u} \in B\} \\
Q^{(2)}\{B\} &= \frac{1}{2\pi}\int_B \exp\left\{-\frac{1}{2}||\mathbf{u}||^2\right\} \times \\
&\quad \left[1+\left(\frac{M}{N}\right)^{-\frac{1}{2}}\frac{1}{3!}\left\{C(0,3)\Gamma_3(j)H_3(u_2)+C(2,1)J_1(j)H_2(u_1)H_1(u_2)\right\}\right]d\mathbf{u} \\
&= \int_B \phi_2(\mathbf{u})q_N^{(2)}(\mathbf{u})d\mathbf{u},
\end{aligned} \tag{3.16}$$

say, where $\phi_2(\mathbf{u})$ is the density of the bivariate normal distribution and $H_j(\cdot)$ are the univariate Hermite Polynomials of order j. The measure $Q^{(2)}\{\cdot\}$ is the Edgeworth expansion for the distribution of the vector $\mathbf{u}$, and its characteristic function is $A(\mathbf{t},2)$.

To prove the validity of the asymptotic expansion we will use the previous lemmas and the Smoothing Lemma 2.9, for which, in the following we assume condition 3.7: $M = C \cdot N^q$, with $0 < q < 1$, and some constant $0 < C < \infty$, but we do not assume yet the choice $q = 1/(1+2d)$ and/or $C = c^*$ that would minimize the MSE of $\widehat{f}(0)$. If not stated otherwise, only the condition $0 < q < 1$ is required in Assumption 3.7. This implies Assumption 3.6 for this particular M. The reason to impose this specific form of M is that Assumption 3.7 also implies that, for some $\varepsilon > 0$ depending on q and p and a positive constant

$$m_N \geq \text{const.}N^{\varepsilon}, \quad \varepsilon > 0. \tag{3.17}$$

Now we can prove that $Q^{(2)}$ is indeed a valid Edgeworth expansion for the probability P_N corresponding to the random variable $\mathbf{u}$.

Lemma 3.6 *Under Assumptions 3.1, 3.2 ($p > 1$), 3.3, 3.4, 3.7, for $\alpha_N = (N/M)^{-\rho}$, $1/2 < \rho < 1$:*

$$\sup_{B \in \mathcal{B}^2} |P_N(B) - Q^{(2)}(B)| = o\left(\left(\frac{N}{M}\right)^{-1/2}\right) + \frac{4}{3} \sup_{B \in \mathcal{B}^2} Q^{(2)}\left\{(\partial B)^{2\alpha_N}\right\}.$$

3.5 Asymptotic expansion for the distribution of the studentized mean

The distributions of $\mathbf{u} = \sqrt{N/M}\mathbf{q}$ and Y_N are functions of quantities, like σ_N, κ_{rs}, etc., which depend on the sample size and on the bandwidth employed in the spectral estimation. We have obtained expressions up to a certain degree of error for these quantities in powers of N and M. The constants of these expansions depend on the value of the spectral density and its derivatives at the origin (unknown) and on the user-chosen kernel $K(\lambda)$.

The accuracy to which we can evaluate the above quantities depending on N determine the order of the error of the feasible Edgeworth expansion for the distribution of Y_N. This accuracy depends mainly on the value of the smoothing number M. In this section we assume condition 3.7 with $q = 1/(1+2d)$, but not necessarily $C = c^\star$. Then $0 < M^{-d}/(N/M)^{-1/2} < \infty$ as $N \to \infty$, and we have that the bias correction is of the same magnitude as the correction term we obtained in $A(\mathbf{t}, 2)$, or as the standard deviation of $\widehat{f}(0)$.

We first work out a linear stochastic approximation to the function $Y_N(\mathbf{u})$ and prove that it is correct up to order $o((N/M)^{-1/2})$. Then, the asymptotic approximation for the distribution of the linear approximation is valid also for Y_N with that error. We assume conditions 3.1, 3.2 (some $p > 1$), 3.3, 3.4, 3.5 and 3.7, $q = 1/(1+2d)$, but not necessarily $C = c^\star$. Then $\widehat{f}(0)$ is $\sqrt{N/M}$ consistent and the approximation we obtained in Section 3.2 for Y_N is valid.

Set the neighbourhood of the origin

$$A_N = \left\{\mathbf{u} : |u_i| < c_i N^{\mu}, 0 < \mu < \frac{d}{3(1+2d)},\ i = 1, 2\right\},$$

where c_i are some fixed constants, and expand Y_N around $\mathbf{0}$ in A_N, with $|\theta| \leq 1$:

$$
\begin{aligned}
Y_N &= \delta_N u_1 - \frac{1}{2}\delta_N^3 \sigma_N u_1 u_2 (N/M)^{-1/2} \\
&\quad + \frac{3}{8}\left(1 + b_N + \sigma_N \theta u_2 (N/M)^{-1/2}\right)^{-5/2} \sigma_N^2 u_1 u_2^2 (N/M)^{-1},
\end{aligned}
$$

remembering that $\delta_N = (1+b_N)^{-1/2}$. Set $Z_N(1) = \frac{3}{8}\left(1 + b_N + \sigma_N \theta u_2 (N/M)^{-1/2}\right)^{-5/2}$ $\cdot \sigma_N^2 u_1 u_2^2$ and substituting for the values of σ_N and δ_N and their powers, from the results of the previous sections, we can write

$$
\begin{aligned}
Y_N &= u_1 \left[1 - \frac{1}{2} b_1 M^{-d} + O(M^{-d-\varrho} + N^{-1}\log N)\right] \\
&\quad - \frac{1}{2}\left[1 + O(M^{-d} + N^{-1}\log N)\right]\left[\frac{\sqrt{4\pi}}{f(0)}\Theta(0) + O\left(e_N(2) + M^{-2}\right)\right] u_1 u_2 (N/M)^{-1/2} \\
&\quad + Z_N(1)(N/M)^{-1} \\
&= u_1 \left[1 - \frac{1}{2} b_1 M^{-d} - \frac{1}{2}\sqrt{4\pi}\|K\|_2 u_1 u_2 (N/M)^{-1/2}\right] \\
&\quad + Z_N(1)(N/M)^{-1} \\
&\quad + u_1 O(N^{-1}\log N + M^{-d-\varrho}) \\
&\quad + u_1 u_2 O\left((N/M)^{-1/2}[M^{-2} + e_N(2)]\right). \qquad (3.18)
\end{aligned}
$$

Define

$$
Z_N = \sum_{j=1}^{3} Z_N(j),
$$

where $(N/M)^{-1} Z_N(2)$ and $(N/M)^{-1} Z_N(3)$ are the last two terms in (3.18) with leading terms in u_1 and $u_1 u_2$ respectively. Thus

$$
Y_N = Y_N' + Z_N (N/M)^{-1}
$$

where

$$
Y_N' = u_1 \left[1 - \frac{1}{2} b_1 M^{-d} - \frac{1}{2}\sqrt{4\pi}\|K\|_2 u_2 (N/M)^{-1/2}\right].
$$

Now we use Chibisov's (1972) result to prove that the error in the previous linear approximation can be neglected with error $o((M/N))^{1/2}$.

Lemma 3.7 *Under Assumptions 3.1, 3.2 ($p > 1$), 3.3, 3.4, 3.5 and 3.7, $q = 1/(1+2d)$, Y_N has the same Edgeworth expansion as Y_N' for convex sets up to the order $(N/M)^{-1/2}$.*

The next step is calculate an Edgeworth Expansion for the distribution of Y'_N from that of $\mathbf{u}$. Consider the transformation

$$\mathbf{s} = \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} = \begin{pmatrix} Y'_N(u_1, u_2) \\ u_2 \end{pmatrix} = \Psi(\mathbf{u})$$

and its inverse

$$\mathbf{u} = \Psi^{-1}(\mathbf{s}) = \begin{pmatrix} u'_1(s_1, s_2) \\ u_2 \end{pmatrix},$$

where we can write, using $(1+x)^{-1} = 1 - x + x^2 - x^3 + \cdots$, for $|x| < 1$, uniformly in A_N,

$$u'_1(\mathbf{s}) \;=\; s_1 \left[1 + \frac{1}{2} b_1 M^{-d} + \frac{1}{2}\sqrt{4\pi}\|K\|_2 u_2 (N/M)^{-1/2}\right] + o((N/M)^{-1/2}), \quad (3.19)$$

where the truncation of the term in $s_1 s_2^2 O((N/M)^{-1})$ with error $o((N/M)^{-1/2})$ is allowed due to the definition of the set A_N.

Writing for convex sets C

$$\operatorname{Prob}\{Y_N \in C\} = \operatorname{Prob}\left\{\mathbf{u} \in \Psi^{-1}(C \times \Re)\right\},$$

as in Taniguchi (1987, p. 22), it follows from Lemma 3.6 that (as Ψ is a C^∞ mapping on A_N),

$$\begin{aligned}
&\sup_{C \in \mathcal{B}^2} \left|\operatorname{Prob}\left\{\mathbf{u} \in \Psi^{-1}(C \times \Re)\right\} - Q^{(2)}\left\{\Psi^{-1}(C \times \Re)\right\}\right| \\
&\quad = o((N/M)^{-1}) + \text{const.} \sup_{C \in \mathcal{B}^2} Q^{(2)}\left\{(\partial \Psi^{-1}(C \times \Re))^{2\alpha_N}\right\}, \quad (3.20)
\end{aligned}$$

where $\alpha_N = (N/M)^{-\rho}$, $1/2 < \rho < 1$. Also, from the continuity of Ψ, we can obtain, for some $c > 0$,

$$Q^{(2)}\left\{(\partial \Psi^{-1}(C \times \Re))^{2\alpha_N}\right\} \le Q^{(2)}\left\{(\Psi^{-1}((\partial C)^{c\alpha_N} \times \Re))\right\} \quad (3.21)$$

and exactly as we did in the previous Chapter (see the lines after expression (2.17)),

$$Q^{(2)}\left\{\Psi^{-1}(C \times \Re)\right\} = \int_C \phi(s_1)\left[1 + r_1(s_1)(N/M)^{-1/2} + r_2(s_1) M^{-d}\right] ds_1 + o((N/M)^{-1/2}),$$

where $r_j(s_1)$ are polynomials on s_1 independent of N. Since we have that

$$\begin{aligned}
E[s_1] &= o((N/M)^{-1/2}) \\
E[s_1^2] &= E\left[u_1^2 - b_1 M^{-d} u_1^2 - \sqrt{4\pi}\|K\|_2 (N/M)^{-1/2} u_1^2 u_2\right] + o((N/M)^{-1/2})
\end{aligned}$$

$$
\begin{aligned}
&= 1 - b_1 M^{-d} + o((N/M)^{-1/2}) \\
E[s_1^3] &= E\left[u_1^3 - \frac{3}{2} b_1 M^{-d} u_1^3 - \frac{3}{2}\sqrt{4\pi}\|K\|_2 (N/M)^{-1/2} u_1^3 u_2\right] + o((N/M)^{-1/2}) \\
&= o((N/M)^{-1/2}),
\end{aligned}
$$

it can be seen that

$$
\begin{aligned}
r_1(x) &= 0 \\
r_2(x) &= -b_1 \frac{x^2 - 1}{2}.
\end{aligned}
$$

So we have obtained, remembering (3.20), (3.21) and Lemma 3.7:

Theorem 3.1 *Under Assumptions 3.1, 3.2 ($p > 1$), 3.3, 3.4, 3.5 and 3.7, $q = 1/(1+2d)$, for convex sets $C \in \mathcal{B}$ and $\alpha_N = (N/M)^{-\rho}$, $1/2 < \rho < 1$,*

$$
\begin{aligned}
&\sup_C \left| Prob\{Y_N \in C\} - \int_C \phi(x)\left[1 + r_2(x) M^{-d}\right] dx \right| \\
&\quad = \frac{4}{3} \sup_C \int_{(\partial C)^{c\alpha_N}} \phi(x)\left[1 + r_2(x) M^{-d}\right] dx + o((N/M)^{-1/2})
\end{aligned}
$$

and, in particular, for the distribution function ($C = (-\infty, y]$):

$$
\sup_{y \in \Re} \left| Prob\{Y_N \le y\} - \int_{-\infty}^{y} \phi(x)\left[1 + r_2(x) M^{-d}\right] dx \right| = o((N/M)^{-1/2}).
$$

Integrating this last expression, and making a Taylor expansion of the distribution function of the standard normal, $\Phi(y)$, we can get uniformly in y, under the conditions of Theorem 3.1:

$$
\begin{aligned}
\text{Prob}\{Y_N \le y\} &= \Phi(y) + \frac{y}{2}\phi(y) b_1 \, M^{-d} + o((N/M)^{-1/2}) && (3.22) \\
&= \Phi(y) + \frac{y}{2}\phi(y) \frac{f^{(d)}(0) K_d^{(1)}}{d! f(0)} \, M^{-d} + o((N/M)^{-1/2}) \\
&= \Phi\left(y + \frac{y}{2} b_1 M^{-d}\right) + o((N/M)^{-1/2}) && (3.23) \\
&= \Phi(y) + O((N/M)^{-1/2}).
\end{aligned}
$$

If we had made the 'optimal' choice $C = c^\star$ in Assumption 3.7 from (3.11), then the approximations above could be written as

$$
\begin{aligned}
\text{Prob}\{Y_N \le y\} &= \Phi(y) + y\,\phi(y)\, b_1' \, N^{-\frac{d}{1+2d}} + o\left(N^{-\frac{d}{1+2d}}\right) \\
&= \Phi\left(y + y\, b_1' \, N^{-\frac{d}{1+2d}}\right) + o\left(N^{-\frac{d}{1+2d}}\right),
\end{aligned}
$$

where

$$b_1' = \frac{b_1}{2}\left[\frac{2d}{4\pi}\left(\frac{f^{(d)}(0)K_d^{(1)}}{d!f(0)\|K\|_2}\right)^2\right]^{\frac{-d}{2d+1}}$$

or equivalently, operating with the values of b_1 and $c^\star$,

$$\begin{aligned}\mathrm{Prob}\{Y_N \le y\} &= \Phi(y) + (N/M^\star)^{-1/2}\phi(y)a_1 + o\Big((N/M^\star)^{-1/2}\Big) \qquad (3.24)\\ &= \Phi\left(y + a_1\, y\,(N/M^\star)^{-1/2}\right) + o\Big((N/M^\star)^{-1/2}\Big), \qquad (3.25)\end{aligned}$$

with

$$a_1 = \sqrt{\frac{\pi}{2d}}\|K\|_2 \;\mathrm{sign}\Big[f^{(d)}(0)\,K_d^{(1)}\Big].$$

When $d = 2$ the approximations (3.23) and (3.25) have a very intuitive interpretation. In this case we have that

$$\frac{b_1}{2} = \frac{f^{(2)}(0)}{4}K_2^{(1)} \quad \text{and} \quad a_1 = \frac{\sqrt{\pi}}{2}\|K\|_2 \mathrm{sign}\left[f^{(2)}(0)\,K_2^{(1)}\right].$$

Suppose that $K_2^{(1)} = \int x^2 K(x)dx > 0$ (e.g. if $K(x) > 0$, $\forall x$). Then if $f(\lambda)$ has a peak at $\lambda = 0$, typically $f^{(2)}(0) < 0$ and we are probably underestimating $f(0)$, as the contribution from the adjacent frequencies cannot help to resolve the peak. Then we underestimate the variance of $\overline{X}$ and the confidence interval for $\sqrt{N/V_N}\,\overline{X}$ is too narrow for Y_N, obtaining too many rejections, since the ratio Y_N tends to increase. The above approximations tend to correct this problem as in both cases we have that they approximate the distribution with $\Phi(y\,k_N)$ where $k_N \le 1$, so for the same confidence level, the critical value y obtained in this way is bigger (in absolute value) than the one obtained using the raw normal approximation. In this way the confidence interval is wider and we do not have so many rejections.

We can make the same reasoning in the other direction, when we have a trough in $f(\lambda)$ at $\lambda = 0$. For $d > 2$ the interpretation is equivalent, but in this case we have to take into account the sign of $K_d^{(1)}$, which can be negative, as it is the case for the Kernel $K_4(x)$ and $d = 4$.

The approximations (3.24) and (3.25) are more attractive, since if we believe to have chosen M in an optimal way by any means, we only need to estimate the sign of $f^{(d)}(0)$, not its value, to achieve an asymptotic second order correction.

3.5.1 Mean Correction

All the previous results have been obtained under the assumption that the mean of the time series was known. Now we drop the assumption of known mean for the sequence $\{X_t\}$. We still considering $E[X_t] = \mu = 0$, without loss of generally, and modify the definition of the estimates of the autocovariance function based on the mean corrected series $X_t - \overline{X}$:

$$\tilde{\gamma}(\ell) = \frac{1}{N} \sum_{1 \le t, t+\ell \le N} (X_t - \overline{X})(X_{t+\ell} - \overline{X}), \qquad \ell = 0, \pm 1, \ldots, \pm(N-1),$$

and the corresponding nonparametric spectral estimate

$$\begin{aligned} \tilde{f}(0) &= \frac{1}{2\pi} \sum_{\ell=1-N}^{N-1} \omega\left(\frac{\ell}{M}\right) \tilde{\gamma}(\ell) \\ &= (X - \overline{X}\mathbf{1})' \left(\frac{W_M}{2\pi N}\right) (X - \overline{X}\mathbf{1}). \end{aligned}$$

All the definitions we did before can be adapted with $\hat{f}(0)$ substituted by $\tilde{f}(0)$. To prove the validity of the previous results we study the difference between both spectral estimates. We obtain

$$\begin{aligned} \tilde{f}(0) &= \hat{f}(0) - 2\left[X'\left(\frac{W_M}{2\pi N}\right)\mathbf{1}\right]\overline{X} + (\overline{X})^2\left[\mathbf{1}'\left(\frac{W_M}{2\pi N}\right)\mathbf{1}\right] \\ &= \hat{f}(0) - 2Z_N + R_N, \end{aligned} \tag{3.26}$$

say. We study now the cumulants of the random variables Z_N and R_N. First we note that

$$Z_N = \left[X'\left(\frac{W_M}{2\pi N}\right)\mathbf{1}\right]\overline{X} = X'\left(\frac{W_M}{2\pi N^2}\mathbf{1}\,\mathbf{1}'\right)X = X'\Lambda_N X,$$

where $\Lambda_N = (2\pi N^2)^{-1} W_M \mathbf{1}\,\mathbf{1}'$ is a $N \times N$ matrix.

Lemma 3.8 *Under Assumptions 3.1, 3.3, 3.4,* $M^{-1} + N^{-1}M\log^2 N \to 0$, $s = 1, 2, \ldots$

$$\mathit{Trace}[(\Sigma_N W_M \mathbf{1}\,\mathbf{1}')^s] = (MN)^s\left[(2\pi)^2 f(0) K(0)\right]^s + O((NM)^{s-1}) M^2 \log^2 N).$$

Then we can obtain under the same set of assumptions of the Proposition 3.8 that

$$\begin{aligned} \text{Cumulant}_s[Z_N] &= c_s \text{Trace}[(\Sigma_N \Lambda_N)^s] \\ &= c_s \left(\frac{M}{N}\right)^s [2\pi f(0) K(0)]^s + O\left(\left(\frac{M}{N}\right)^{s+1} \log^2 N\right), \end{aligned}$$

where $c_s = 2^{s-1}(s-1)!$. Then $(N/M)Z_N$ has bounded moments of all orders.

Lemma 3.9 *Under Assumptions 3.3, 3.4, $M^{-1} + N^{-1}M\log N \to 0$,*

$$\frac{1}{2\pi N}1'W_M 1 = M\,K(0) + O\left(\frac{M^2}{N}\log N\right).$$

As $\overline{X} \rightsquigarrow \mathcal{N}(0, V_N/N)$ and from Lemma 3.1, under Assumption 3.1, $V_N = 2\pi f(0) + O(N^{-1}\log N)$, it follows that $(N/M)\,R_N$ has bounded moments of all orders too.

Therefore, from (3.26),

$$\sqrt{\frac{N}{M}}\tilde{f}(0) = \sqrt{\frac{N}{M}}\hat{f}(0) + \left(\frac{M}{N}\right)^{1/2}\Delta_N, \tag{3.27}$$

where Δ_N is a random variable with bounded moments of all orders, and expectation

$$-2\pi K(0)f(0)\left(\frac{M}{N}\right)^{1/2},$$

from Lemmas 3.8 and 3.9. Therefore the distribution of $\tilde{f}(0)$ is affected to a second order ($(M/N)^{1/2}$) by the mean correction. The bias term is the same as that studied by Hannan (1958) in the context of estimation of the spectral density after removal of a general type of trend. Of course, the asymptotic relationship of this bias with the smoothing bias studied in Lemma 3.2 depends on the degree of smoothing (that is, the relation between M and N). In Section 3.6 we analyze this issue further.

Denote as $Y_N^\star$ the studentized mean when we use $\tilde{f}(0)$ and equivalently $u_2^\star$ as u_2 with $\tilde{f}(0)$. Then $u_2^\star = u_2 + (N/M)^{-1/2}\Delta_N'$, where the random variable $\Delta_N' = O(1)\Delta_N$ has the same properties of Δ_N. Now

$$\begin{aligned} Y_N^\star \;=\; & u_1\left[1 - \frac{1}{2}b_1 M^{-d} - \frac{1}{2}\sqrt{2\pi}\|K\|_2 u_1 u_2 (N/M)^{-1/2}\right] \\ & + Z_N(1)(N/M)^{-1} \\ & + u_1 O(N^{-1}\log N + M^{-d-\varrho}) \\ & + u_1 u_2 O\Big((N/M)^{-1/2}[M^{-2} + e_N(2)]\Big) \\ & + \Delta_N''(N/M)^{-1}, \end{aligned}$$

where Δ_N'' depends on Δ_N, u_1 and u_2, and has moments of all orders, so it can be neglected when we approximate $Y_N^\star$ with Y_N'. Therefore the studentized sample mean with the 'mean corrected' spectral estimate has the same Edgeworth approximation for its distribution function up to order $o((N/M)^{-1/2})$ as when the mean is known. However, the expansion for the distribution of $\tilde{f}$ would differ from that for $\hat{f}$ in terms of order $(N/M)^{-1/2}$.

3.5.2 Empirical approximation

The above approximations to the distribution of the studentized mean and to the optimal bandwidth choice depend on the unknown derivative $f^{(d)}(0)$ and on $f(0)$. These two quantities can be substituted by initial estimates to obtain an empirical Edgeworth approximation. The estimate of $f(0)$ can be obtained using an initial guess for the optimal bandwidth in $\widehat{f}(0)$.

In the remaining of this section we propose nonparametric estimates of the derivatives of f at any frequency α and prove their consistency under local conditions. To estimate derivatives further smoothness restrictions on f will be necessary, but only around the frequency of interest.

First we introduce the class of Kernels (ν, r) $\nu = 0, 1, \ldots, r-1$ to estimate the ν-th derivative of the spectral density, following Gasser et al. (1985). Define the kernel V_ν of order (ν,r) as a function such that

$$\int_{-\pi}^{\pi} V_\nu(x)x^j dx = \begin{cases} 0 & j = 0, \ldots, \nu-1, \nu+1, \ldots, r-1 \\ (-1)^\nu \nu! & j = \nu \\ \vartheta \neq 0 & j = r, \end{cases}$$

with support $[-\pi, \pi]$, and satisfying a Lipschitz condition of order 1. If $\nu = 0$ then we estimate the function itself and V_0 has equivalent properties to the kernel K we used to estimate f (compare this with Assumptions 3.3, 3.4 and 3.5). We define for a sequence of integers $m_\nu = m_\nu(N)$, satisfying at least

$$\frac{1}{m_\nu} + \frac{m_\nu}{N} \to 0,$$

the function, $x \in [-\pi, \pi]$,

$$V_{m_\nu}(x) = m_\nu V_\nu(m_\nu x),$$

such that

$$\int_\Pi |V_{m_\nu}(x)|\, dx < \infty, \text{ and } \int_\Pi |x V_{m_\nu}(x)|\, dx = O(m_\nu^{-1}).$$

Then the proposed estimate $\widehat{f}_{m_\nu}^{(\nu)}(\alpha)$ for the ν-th derivative of $f(\lambda)$ at $\lambda = \alpha$ is

$$\widehat{f}_{m_\nu}^{(\nu)}(\alpha) = (m_\nu)^\nu \int_\Pi V_{m_\nu}(\lambda) I_N(\alpha - \lambda) d\lambda.$$

In the following two lemmas we summarize the results about the expectation and the variance of these estimates under the local smoothness assumptions.

Lemma 3.10 *Under Assumption 3.1, $d = \nu + a$, $\varrho = 0$, for $f(\lambda)$ around any frequency $\lambda = \alpha$, and a Kernel of order $(\nu, \nu+a)$, for some integer $a \geq 2$, and $(m_\nu)^{-1} + N^{-1}(m_\nu)^\nu \log N \to 0$ then*

$$E[\widehat{f}_{m_\nu}^{(\nu)}(\alpha)] - f^{(\nu)}(\alpha) = O\Big((m_\nu)^\nu[N^{-1}\log N + m_\nu{}^{-\nu-a}]\Big).$$

Lemma 3.11 *Under the conditions of Lemma 3.10, with If $(m_\nu)^{-1} + N^{-1}m_\nu \log^3 N \to 0$,*

$$\frac{N}{(m_\nu)^{2\nu+1}} Var[\widehat{f}_{m_\nu}^{(\nu)}(\alpha)] = 2\pi f^2(\alpha)\|V_\nu\|_2^2\,[1 + \delta_{0,\alpha}] + o(1).$$

These two lemmas justify consistent empirical Edgeworth expansions based on the above results, plugging in consistent estimates of the unknown constants. The reason is that the correction terms are of order $(M/N)^{1/2}$ and the use of consistent estimates for f and $f^{(d)}$ will introduce only an error of magnitude $o_P((M/N)^{1/2})$.

Examples of the class of Kernels (ν, r) obtained from the uniform distribution on $[-\pi, \pi]$ are,

For ν=2, r=4, $\quad V_2(x) = -\frac{15}{4\pi^3} + \frac{45}{4\pi^5}x^2$

For ν=2, r=6, $\quad V_2(x) = -\frac{525}{32\pi^3} + \frac{2205}{16\pi^5}x^2 - \frac{4725}{32\pi^7}x^4$

For ν=4, r=6, $\quad V_4(x) = \frac{2835}{16\pi^5} - \frac{14175}{8\pi^7}x^2 + \frac{33075}{16\pi^9}x^4.$

The asymmptotic distribution of $\widehat{f}_{m_\nu}^{(\nu)}(\alpha)$ can be analyzed using related techniques to those we have employed for the distribution of $\widehat{f}(0)$.

3.6 Third order approximation

In this section we concentrate on obtaining a third order approximation (that is, including terms of order M/N) for the distribution of the studentized sample mean. The previous results are not sufficient to prove the validity of such approximation when we consider the effect of the mean estimation in the nonparametric spectral estimate. In the following we will show the main modifications of the previous scheme that will allow us to prove that third order approximation.

3.6.1 Distribution of the nonparametric spectral estimate

The main idea is to work directly with the *mean corrected* nonparametric spectral estimate $\tilde{f}(0)$ instead of $\hat{f}(0)$ and analyze the effect of mean correction in all the steps of the previous development. As we have seen in Section 3.5.1, the effect of the estimation of the unknown expectation of the time series in the spectral estimate is of order $(M/N)^{1/2}$, so it will have an effect of order M/N in a third order approximation for the studentized mean.

As before, we will denote with a star superscript, $\star$, all the quantities when the estimate $\tilde{f}(0)$ is used instead of $\hat{f}(0)$. First we will study the bias. The following lemma is a simply extension of Lemma 3.2 using Lemmas 3.8 and 3.9:

Lemma 3.12 *Under Assumptions 3.1, 3.3, 3.4, 3.5, 3.6 and* $M^{-1} + N^{-1}M\log N \to 0$,

$$E[\tilde{f}(0)] - f(0) = \frac{f^{(d)}(0)}{d!}K_d^{(1)}M^{-d} - 2\pi f(0)K(0)\frac{M}{N} + O\left(\frac{\log N}{N} + M^{-d-\varrho} + \left[\frac{M}{N}\right]^2 \log^2 N\right).$$

The second term on the right hand side is due to the mean correction in $\tilde{f}(0)$.

Define now $u_2^\star$ similarly as u_2 but with $\tilde{f}(0)$ in the place of $\hat{f}(0)$. For the analysis of its cumulants, we can write the random variable $u_2^\star$ compactly as a quadratic form in the vector $\mathbf{X}$,

$$u_2^\star = \mathbf{X}'Q_N^\star\mathbf{X} - E[\mathbf{X}'Q_N^\star\mathbf{X}],$$

where the $N \times N$ matrix

$$Q_N^\star = A_N Q_N A_N, \qquad A_N = I_N - \frac{\mathbf{1}\mathbf{1}'}{N},$$

is the *mean corrected* version of Q_N. We can define equivalently $W_M^\star$ in terms of W_M. Now it is possible to obtain, under the conditions of Proposition 3.1,

$$\mathrm{Trace}\left[(\Sigma_N W_M^\star)^s\right] = NM^{s-1}(2\pi)^{2s-1}\sum_{j=0}^{d} L_s(0) + O\left(NM^{s-1}e_N(s) + NM^{s-3}\right),$$

so the cumulants $\kappa[0,s]^\star$ (of $u_2^\star$) have the same asymptotic approximation as $\kappa[0,s]$. This can be seen using the same techniques as in the proof of Proposition 3.1, since now the Fourier transform corresponding to the matrix A_N is

$$A_N(\lambda) = \frac{1}{2\pi}\left(1 - \frac{D_N(\lambda)}{N}\right),$$

where

$$D_N(\lambda) = \sum_{j=1-N}^{N-1} e^{ij\lambda}$$

is a version of Dirichlet kernel. Therefore, obtaining all possible cross products due to the function $A_N(\lambda)$ in the corresponding expression for the trace, the only one with the factor $(2\pi)^{-s}$ is equal to the trace of $(\Sigma_N W_M)^s$. Meanwhile, all the other terms (with at least one function $D_N(\lambda)$ in them) are of an M/N order of magnitude smaller for each Dirichlet kernel $D_N(\lambda)$ function present in the product (so they are incorporated in the error term $O(NM^{s-1}e_N(s))$).

Consequently, all the conclusions about the variance and optimal bandwidth assuming known mean, still go through, since the previous expressions for that quantities are valid, up to order M/N. We will see later, that estimating only the leading term of the higher order cumulants of $u_2^\star$ is sufficient to obtain the approximation with the desired error.

3.6.2 Joint distribution of the spectral estimate and the sample mean

Concerning the cross-cumulants, it is possible to obtain, under the conditions of Proposition 3.2,

$$\mathbf{1}'(\Sigma_N W_M^\star)^s \Sigma_N \mathbf{1} = N(2\pi)^{2s+1}[f(0)]^{s+1}[K_M(0)]^s + O\left(M^{s+1}\log^{2s+1} N\right),$$

and therefore the cumulants $\kappa[2,s]^\star$ have the same asymptotic behaviour as $\kappa[2,s]$, just by using the same argument for the function $A(\lambda)$ as for the cumulants of the normalized spectral density estimate.

The next goal is to approximate the joint characteristic function of $\mathbf{u}^\star = (u_1, u_2^\star)$. Define, for that approximation,

$$A^\star(\mathbf{t},3) = \exp\left\{\frac{1}{2}\|i\mathbf{t}\|^2\right\}\left[1 + B_N^\star(3,\mathbf{t})\left(\frac{M}{N}\right)^{1/2} + \left\{B_N^\star(4,\mathbf{t}) + \frac{1}{2}B_N^\star(3,\mathbf{t})^2\right\}\frac{M}{N}\right],$$

where, we include in $B_N^\star$ the correspondent cumulants, not only the leading terms of their expansions (as we did in Chapter 2),

$$B_N^\star(s,\mathbf{t}) = \frac{1}{s!}\left[\bar{\kappa}[0,s]^\star(it_2)^s + \frac{s(s-1)}{2}\bar{\kappa}[0,s-2]^\star(it_1)^2(it_2)^{s-2}\right].$$

Now we can obtain similar results to Lemmas 3.4 and 3.5:

Lemma 3.13 *Under Assumptions 3.1, 3.3, 3.4,* $M^{-1} + N^{-1}M\log^7 N \to 0$*, there exists a positive number* $\delta_1 > 0$ *such that, for* $\|\mathbf{t}\| \leq \delta_1\sqrt{N/M}$ *and a number* $d_1 > 0$*:*

$$|\psi^\star(\mathbf{t}) - A^\star(\mathbf{t},3)| \leq \exp\{-d_1\|\mathbf{t}\|^2\}F(\|\mathbf{t}\|)O\left(\left(\frac{N}{M}\right)^{-3/2}\right),$$

where F *is a polynomial in* $\mathbf{t}$ *with bounded coefficients.*

Lemma 3.14 *Under Assumptions 3.1, 3.2 (*$p > 1$*), 3.3, 3.4,* $M^{-1} + N^{-1}M\log^3 N \to 0$ *there exists a positive constant* $d_2 > 0$ *such that for* $\|\mathbf{t}\| > \delta_1 m_N$*,*

$$|\psi^\star(t_1,t_2)| \leq \exp\left\{-d_2 m_N^2\right\}.$$

The last lemma follows using the fact that the asymptotic variance of the spectral estimate is the same, with and without mean correction, and that the norms of the matrices $\Sigma_N W_N$ and $\Sigma_N W_N^\star$ have the same asymptotic bounds. Next, defining,

$$\begin{aligned}
P_N^\star\{B\} &= \text{Prob}\{\mathbf{u}^\star \in B\} \\
Q^{(3)\star}\{B\} &= \frac{1}{2\pi}\int_B \exp\left\{-\frac{1}{2}\|\mathbf{u}\|^2\right\}\Big[1 + \frac{1}{3!}\{\kappa[0,3]^\star H_3(u_2) + \kappa[2,1]^\star H_2(u_1)H_1(u_2)\} \\
&\quad + \frac{1}{2(3!)^2}\left\{(\kappa[0,3]^\star)^2 H_6(u_2) + (\kappa[2,1]^\star)^2 H_4(u_1)H_4(u_2)\right. \\
&\quad +2\kappa[0,3]^\star\kappa[2,1]^\star H_2(u_1)H_4(u_2)\} \\
&\quad + \frac{1}{4!}\{\kappa[0,4]^\star H_4(u_2) + \kappa[2,2]^\star H_2(u_1)H_2(u_2)\}\Big]d\mathbf{u} \\
&= \int_B \phi_2(\mathbf{u})q_N^{(3)\star}(\mathbf{u})d\mathbf{u}, \qquad \text{say,}
\end{aligned}$$

we can get, for the approximation of the distribution of the vector $\mathbf{u}^\star$,

Lemma 3.15 *Under Assumptions 3.1, 3.2 (*$p > 1$*), 3.3, 3.4, 3.7, for* $\alpha_N = (N/M)^{-\rho}$*,* $1/2 < \rho < 1$*:*

$$\sup_{B\in\mathcal{B}^2} |P_N^\star(B) - Q^{(3)}(B)| = o\left(\left(\frac{N}{M}\right)^{-1}\right) + \frac{4}{3}\sup_{B\in\mathcal{B}^2} Q^{(3)}\left\{(\partial B)^{2\alpha_N}\right\}.$$

3.6.3 Distribution of the studentized mean

Now we move to the studentized sample mean $Y_N^\star$ with the nonparametric estimate $\widetilde{f}(0)$, using the results concerning the Edgeworth expansion for the distribution of $\mathbf{u}^\star$. First we can obtain a linear approximation for $Y_N^\star$. The main problem here is the bias term

$$b_N^\star = b_1 M^{-d} + b_2 \frac{M}{N} + O\left(N^{-1}\log N + M^{-d-\varrho} + \left[\frac{M}{N}\right]^2 \log^2 N\right),$$

with $b_2 = -2\pi K(0)$. If we want to make the bias term $b_N^\star$ negligible up to order M/N we can not assume that M has the optimal rate that minimizes the mean square error of $\widetilde{f}(0)$. Instead, we need to impose a condition like,

$$\lim_{N\to\infty} \frac{M}{N} M^d > 0, \tag{3.28}$$

which guarantees that the bias term of order M^{-d} is at most of order M/N, and that the term $O(M^{-d-\varrho})$ will not affect the third order approximation under Assumption 3.7. Of course this implies an important undersmoothing: M has to grow much faster than $N^{1/(1+2d)}$, at least as $N^{1/(1+d)}$. Now,

$$\delta_N^\star = 1 - \frac{1}{2} b_1 M^{-d} - \frac{1}{2} b_2 \frac{M}{N} + O\left(N^{-1}\log N + M^{-d-\varrho} + \left[\frac{M}{N}\right]^2 \log^2 N\right),$$

and we can obtain

$$\sigma_N^\star = \sqrt{4\pi}\|K\|_2 + e_N,$$

where $e_N = O(M^{-2} + e_N(2))$, and we do not need its exact value. Therefore we can write

$$Y_N^\star = Y_N' + Z_N (N/M)^{-3/2},$$

where

$$Y_N' = u_1 \left[1 - \frac{1}{2} b_1 M^{-d} - \frac{1}{2} b_2 \frac{M}{N} - \frac{1}{2}\left\{\sqrt{4\pi} + e_N\right\}\|K\|_2 u_2 \left(\frac{M}{N}\right)^{1/2} + \frac{3}{8} 4\pi \|K\|_2^2 u_2^2 \frac{M}{N}\right],$$

and the stochastic approximation error $Z_N(N/M)^{-3/2}$ can be neglected in an approximation up to order M/N. Now we can use the same arguments as before to justify the Edgeworth approximation for $Y_N^\star$ in terms of that for $\mathbf{u}^\star$, since, under condition (3.28),

$$\begin{aligned}
E[s_1] &= o(M/N) \\
E[s_1^2] &= E\left[u_1^2\left(1 - b_1 M^{-d} - b_2 \frac{M}{N} - \left\{\sqrt{4\pi}\|K\|_2 + e_N\right\}\left(\frac{M}{N}\right)^{1/2} u_2 + 4\pi\|K\|_2^2 \frac{M}{N} u_2^2\right)\right] \\
&\quad + o(M/N) \\
&= 1 - b_1 M^{-d} + \frac{M}{N}\left[-b_2 + 4\pi\|K\|_2^2 - 4\pi K(0)\right] + o(M/N)
\end{aligned}$$

$$
\begin{aligned}
E[s_1^3] &= o(M/N) \\
E[s_1^4] &= E\left[u_1^4\left(1 - 2\left\{b_1 M^{-d} + b_2\frac{M}{N}\right\} - 2\left\{\sqrt{4\pi}\|K\|_2 + e_N\right\}\left(\frac{M}{N}\right)^{1/2} u_2\right.\right. \\
&\qquad \left.\left. + 12\pi\|K\|_2^2 \frac{M}{N} u_2^2\right)\right] + o(M/N) \\
&= 3 - 6b_1 M^{-d} + \frac{M}{N}\left[-6b_2 - 48\pi K(0) + 36\pi\|K\|_2^2\right] + o(M/N),
\end{aligned}
$$

defining the polynomial

$$
\begin{aligned}
r_N(x) &= \left[4\pi\|K\|_2^2 - 2\pi K(0) - b_1 N M^{-1-d}\right]\frac{x^2-1}{2} \\
&\quad + \left[12\pi\|K\|_2^2 - 24\pi K(0)\right]\frac{x^4 - 6x^2 + 3}{24},
\end{aligned}
$$

we can obtain,

Theorem 3.2 *Under Assumptions 3.1, 3.2 ($p > 1$), 3.3, 3.4, 3.5, 3.7 and (3.28), for convex sets $C \in \mathcal{B}$ and $\alpha_N = (N/M)^{-\rho}$, $1/2 < \rho < 1$,*

$$
\begin{aligned}
&\sup_C \left| Prob\{Y_N^* \in C\} - \int_C \phi(x)\left[1 + r_N(x)\frac{M}{N}\right] dx \right| \\
&\quad = \frac{4}{3}\sup_C \int_{(\partial C)^{c\alpha_N}} \phi(x)\left[1 + r_N(x)\frac{M}{N}\right] dx + o\left(\frac{M}{N}\right).
\end{aligned}
$$

It can be observed that the sign and magnitude of the constants in the polynomial $r_N(x)$ do not depend on the spectral density or any other magnitude of the distribution of the time series (only depend on the characteristics of the kernel K), except the term b_1, which is a function of the value of the spectral density and its d-th derivative at the origin. The reason for this fact is that the standard deviation of $\tilde{f}(0)$ is proportional to $f(0)$, so the normalized distribution has constant variance and higher order cumulants (up to a first order) with respect to $f(0)$. This term b_1 in the polynomial $r_N(x)$, of order of magnitude NM^{-d-1}, could be neglected assuming enough degree of undersmoothing, that is, if in expression (3.28) the limit is tending to infinity as $N \to \infty$. In that way, the expansion will not depend on any unknown parameter. Of course, the bigger the number M, the worse the approximation from the point of view of the M/N corrections.

Alternatively, more informative expansions for the bias can be obtained, using higher order derivatives of the spectral density at the origin and appropriate conditions on the kernel. In such case, a restriction like (3.28) would not be necessary in its full strength, therefore allowing the term in b_1 to be of bigger order of magnitude that M/N and also allowing mean squared error based choices of M.

Following Hall (1992, Section 2.5) and using Theorem 3.2, we can obtain an approximation of the Cornish-Fisher type for the quantiles of the distribution of the studentized mean $Y_N^\star$. Write $w_\alpha = w_\alpha(N, M)$ for the α-level quantile of $Y_N^\star$, determined by

$$w_\alpha = \inf \{x : \text{Prob}\{Y_N^\star \leq x\} \geq \alpha\},$$

and let z_α be the α-level standard normal quantile, given by $\Phi(z_\alpha) = \alpha$. Then it is immediate

Theorem 3.3 *Under Assumptions 3.1, 3.2 ($p > 1$), 3.3, 3.4, 3.5, 3.7 and (3.28),*

$$w_\alpha = z_\alpha - r_N(x)\frac{M}{N} + o\left(\frac{M}{N}\right),$$

uniformly in $\epsilon < \alpha < 1 - \alpha$ for each $\epsilon > 0$, where r_N is defined as before.

3.7 Conclusions

In addition to our all previous comments we would like to stress the following points:

- Equivalent results to those obtained here for $\hat{f}(0)$ hold, with the obvious modifications, for nonparametric estimates of the spectral density at other frequencies $\lambda \neq 0$. This, for example, covers the estimation of the spectrum for long range dependent stationary time series at *smooth* frequencies (beyond the origin).

- We have concentrated on the studentization at the zero frequency, but similar conclusions are valid for the discrete Fourier transform of X_t and the spectral estimate at any other fixed frequency (not at Fourier frequencies, since there the discrete Fourier transform has mean identically equal to zero).

- The location problem can be seen as a particular case of the more general linear regression framework of Chapter 2. In the sample mean case it is possible to specialize the assumptions about f for a particular frequency (as could be done for a simple trigonometric regression at a known frequency).

- We expect that the corrections proposed here can outperform in applications the raw normal approximation for high positively autocorrelated processes, often found in practical applications. In this particular case, the nonparametric bias is of special significance, and would have severe influence on the finite sample distribution of the studentized mean.

- We have seen that the accuracy of the asymptotic distribution will depend asymptotically on the local behaviour of the spectral density (at the origin in this case). The same comment applies for nonparametric estimation of the spectral density at any frequency. Furthermore, the choice of the lag number M for finite sample sizes N when we estimate the spectrum at one frequency should adapt to the local properties of the spectral density. We explore a possible criterion for this local choice of M in next chapter for a related estimate to $\tilde{f}$.

- The results on nonparametric estimation of the spectral density assuming only local conditions can be applied successfully to semiparametric models where we only assume a parametric model in a band of frequencies and leave the rest of the spectrum as a nuisance nonparametric function. We adopt this approach in Chapter 5 for the estimation of the memory parameter for stationary long range dependent time series.

3.8 Appendix: Proofs of Section 3.3

The structure of the proofs for this chapter is very similar to that of the previous chapter. However, here we have expressed our assumptions in the frequency domain, so the analysis of the cumulants of the distributions can not be done in terms of the autocovariance sequence. Therefore, we have to rely on spectral analysis methods.

First we introduce some functions that will arise in the following discussion, and establish some of their properties. Define the *Multiple Fejér Kernel* as in Bentkus (1972) and in the same spirit of Dahlhaus (1983) for the tapered case:

$$\Phi_N^{(n)}(x_1,\ldots,x_n) = \frac{1}{(2\pi)^{n-1}N}\frac{\sin Nx_1/2}{\sin x_1/2}\cdots\frac{\sin Nx_n/2}{\sin x_n/2}$$

$$= \frac{1}{(2\pi)^{n-1}N} \sum_{t_1,\ldots,t_n=1}^{N} \exp\left\{ i \sum_{j=1}^{n} t_j x_j \right\},$$

with $x_n \equiv -\sum_{j=1}^{n-1} x_j$. For $n = 2$ this is the Fejér Kernel. Then $\Phi_N^{(n)}(x_1,\ldots,x_n)$ has the following properties:

-
$$\sup_N \int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} \left| \Phi_N^{(n)}(x_1,\ldots,x_n) \right| dx_1 \cdots dx_{n-1} < \infty. \tag{3.29}$$

-
$$\int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} \Phi_N^{(n)}(x_1,\ldots,x_n) dx_1 \cdots dx_{n-1} = 1. \tag{3.30}$$

- For $\delta > 0$, $N \geq 1$

$$\int_{D^c} \left| \Phi_N^{(n)}(x_1,\ldots,x_n) \right| dx_1 \cdots dx_{n-1} = O\left(\frac{\log^{n-1} N}{N \sin\delta/2}, \right) \tag{3.31}$$

  where D^c is the complementary of the set $D = \{x \in \Re^{n-1} : |x_j| \leq \delta,\ j=1,\ldots,n-1\}$.

- For $j = 1,\ldots,n-1$,

$$\int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} |x_j| \left| \Phi_N^{(n)}(x_1,\ldots,x_n) \right| dx_1 \cdots dx_{n-1} = O\left(\frac{\log^{n-1} N}{N} \right). \tag{3.32}$$

- These properties follow due to

$$\left| \Phi_N^{(n)}(x_1,\ldots,x_n) \right| \leq \frac{1}{(2\pi)^{n-1}N} |\varphi_N(x_1)| \, |\varphi_N(x_2)| \cdots |\varphi_N(x_n)|, \tag{3.33}$$

  where $\varphi_N(x)$ is the *Dirichlet Kernel*, $\varphi_N(x) = \sum_{t=1}^{N} \exp\{itx\}$, which satisfies:

  -
$$|\varphi_N(x)| \leq \min\left\{ N, 2|x|^{-1} \right\} \tag{3.34}$$

  -
$$\int_{-\pi}^{\pi} |\varphi_N(x)| \, dx = O(\log N). \tag{3.35}$$

We have followed the same convention as in Keenan (1986, p. 137): although the functions $\Phi_N^{(n)}$ depend here on only $n-1$ arguments, we will use in the notation n variables, with the restriction $\sum_1^n x_j \equiv 0(\mathrm{mod} 2\pi)$.

Proof of Lemma 3.1. By standard Fourier Analysis, we can apply the Mean Value Theorem for f around the origin, for some $|\theta| \leq 1$ depending on λ,

$$\begin{aligned}
|V_N - 2\pi f(0)| &= \left|2\pi \int_\Pi f(\lambda)\Phi_N^{(2)}(\lambda)d\lambda - 2\pi f(0)\int_\Pi \Phi_N^{(2)}(\lambda)d\lambda\right| \\
&\leq 2\pi \int_\Pi |f(\lambda) - f(0)| \left|\Phi_N^{(2)}(\lambda)\right| d\lambda \\
&\leq 2\pi \int_{|\lambda|\leq\epsilon} |f(\lambda) - f(0)| \left|\Phi_N^{(2)}(\lambda)\right| d\lambda + 2\pi \int_{|\lambda|>\epsilon} |f(\lambda) - f(0)| \left|\Phi_N^{(2)}(\lambda)\right| d\lambda \\
&= O\left(\int_{|\lambda|\leq\epsilon} |\lambda||f'(\lambda\theta)||\Phi_N^{(2)}(\lambda)|d\lambda + [\|f\|_1 + f(0)]N^{-1}\right) \\
&= O\left(N^{-1}\log N\right)
\end{aligned}$$

using the integrability of f (implied by the stationarity), its differentiability around the origin and the property of $\Phi_N^{(2)}$ due to (3.33) and (3.34):

$$|\Phi_N^{(2)}(\lambda)| = O(N^{-1}), \quad \text{if } |\lambda| \geq \delta > 0. \qquad \square$$

Proof of Lemma 3.2. The spectral estimate can be written as

$$\hat{f}(0) = \int_{-\pi}^{\pi} K_M(\lambda)I_N(\lambda)d\lambda$$

where $I_N(\lambda)$ is the periodogram of the series $X_1, \ldots, X_N$, with expectation

$$E[I_N(\lambda)] = \int_{-\pi}^{\pi} \Phi_N^{(2)}(\lambda - \alpha)f(\alpha)d\alpha$$

so substituting we have

$$E[\hat{f}(0)] = \int_{-\pi}^{\pi} K_M(\lambda)\int_{-\pi}^{\pi} \Phi_N^{(2)}(\alpha)f(\lambda + \alpha)d\alpha d\lambda.$$

Then can write the bias of the spectral estimate as

$$\begin{aligned}
E[\hat{f}(0)] - f(0) - \frac{f^{(d)}(0)}{d!}\frac{K_d^{(1)}}{M^d} &= \int_{-\pi}^{\pi} K_M(\lambda)\int_{-\pi}^{\pi} \Phi_N^{(2)}(\alpha)\left[f(\lambda + \alpha) - f(\lambda)\right] d\alpha d\lambda \\
&\quad + \int_{-\pi}^{\pi} K_M(\lambda)\left[f(\lambda) - f(0) - \frac{f^{(d)}(0)}{d!}K_d^{(1)}M^{-d}\right] d\lambda \\
&= b_1 + b_2,
\end{aligned}$$

say, where we have used (3.3). Denote the sets $D = \{|\alpha|, |\lambda| \leq \epsilon/2\}$ and D^c its complementary in $[-\pi, \pi]^2$. Let b_{11} and b_{12} be the integrals in b_1 corresponding to the sets D and D^c respectively. Then, for $|\theta| \leq 1$, depending on α

$$b_{11} = \int_D K_M(\lambda)\Phi_N^{(2)}(\alpha)\left[f'(\lambda + \theta\alpha)\alpha\right] d\alpha d\lambda$$

and

$$|b_{11}| \leq \sup_{|\lambda|\leq\epsilon} |f'(\lambda)| \int_{|\lambda|\leq\epsilon/2} |K_M(\lambda)| d\lambda \int_{|\alpha|\leq\epsilon/2} |\alpha||\Phi_N^{(2)}(\alpha)| d\alpha = O\left(\frac{\log N}{N}\right). \qquad (3.36)$$

To study b_{12} we note first that $D^c \subset A_1 \cup A_2$ where

$$A_1 = \{|\alpha| > \epsilon/2\} \text{ and } A_2 = \{|\lambda| > \epsilon/2, |\alpha| \leq \epsilon/2\}.$$

Then

$$\begin{aligned}
&\left|\int\int_{A_1} K_M(\lambda)\Phi_N^{(2)}(\alpha)\left[f(\lambda+\alpha) - f(\lambda)\right] d\alpha d\lambda\right| \\
&= \left|\int_{|\alpha|>\epsilon/2}\int_\Pi K_M(\lambda)\left[f(\lambda+\alpha) - f(\lambda)\right] d\lambda \Phi_N^{(2)}(\alpha) d\alpha\right| \\
&= O\left(N^{-1}\int_\Pi\int_\Pi |K_M(\lambda)\left[f(\lambda+\alpha) - f(\lambda)\right]| d\lambda d\alpha\right) \\
&= O\left(N^{-1}\left[\int_{|\lambda|\leq\epsilon} |K_M(\lambda)f(\lambda)| d\lambda + \int_{|\lambda|>\epsilon} |K_M(\lambda)f(\lambda)| d\lambda\right]\right) \qquad (3.37) \\
&= O\left(N^{-1}\int_\Pi |K_M(\lambda)| d\lambda\right) \\
&= O\left(N^{-1}\right),
\end{aligned}$$

as the second integral in (3.37) vanishes as $M \to \infty$. On the other hand, reasoning in a similar way, if M is big enough

$$\begin{aligned}
&\left|\int\int_{A_2} K_M(\lambda)\Phi_N^{(2)}(\alpha)\left[f(\lambda+\alpha) - f(\lambda)\right] d\alpha d\lambda\right| \\
&= \left|\int_{|\lambda|>\epsilon/2}\int_{|\alpha|\leq\epsilon/2} K_M(\lambda)\Phi_N^{(2)}(\alpha)\left[f(\lambda+\alpha) - f(\lambda)\right] d\alpha d\lambda\right| = 0.
\end{aligned}$$

So finally

$$b_{12} = O\left(N^{-1}\right). \qquad (3.38)$$

Now for b_2, splitting the integral in two parts for $|\lambda| \leq \epsilon$ and $|\lambda| > \epsilon$, and denoting these two parts as b_{21} and b_{22} respectively, we have first, constructing a Taylor expansion, (with $|\theta| \leq 1$, depending on λ),

$$\begin{aligned}
b_{21} &= \int_{|\lambda|\leq\epsilon} K_M(\lambda)\left[f(\lambda) - f(0) - \frac{f^{(d)}(0)}{d!}K_d^{(1)}M^{-d}\right] d\lambda \\
&= \int_{|\lambda|\leq\epsilon} K_M(\lambda)\left[\sum_{j=1}^{d-1} f^{(j)}(0)\frac{\lambda^j}{j!} + f^{(d)}(\theta\lambda)\frac{\lambda^d}{d!} - \frac{f^{(d)}(0)}{d!}K_d^{(1)}M^{-d}\right] d\lambda
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{j=1}^{d-1} f^{(j)}(0)\frac{1}{j!}\int_{\Pi}\lambda^j K_M(\lambda)d\lambda + \int_{|\lambda|\leq\pi/M} K_M(\lambda)\left[f^{(d)}(\theta\lambda) - f^{(d)}(0)\right]\lambda^d\,d\lambda \\
&\leq O\Big(\int_{\Pi}|K_M(\lambda)||\lambda|^{d+\varrho}d\lambda\Big) \\
&= o(M^{-d-\varrho}), \qquad (3.39)
\end{aligned}
$$

as all the integration is done inside $[-\epsilon,\epsilon]$ since $M\to\infty$ and using the Lipschitz property of $f^{(d)}$. Obviously b_{22} is zero due to the compactness of the support of K.

So, using (3.36), (3.38) and (3.39), the error in estimating the bias is of order

$$
O\Big(N^{-1}\log N + M^{-d-\varrho}\Big)
$$

and the lemma is proved. □

For the proof of Proposition 3.1 we follow Bentkus (1976), although our assumptions are different. First we prove the following Lemma, which could have been taken as an Assumption, instead of giving a sufficient condition for it as Assumption 3.4. However, we will use later in the proof of Proposition 3.2 the smooth properties of the function K.

Lemma 3.16 *Under Assumptions 3.3 and 3.4 for any sequence $\delta_N = o(M^{-1})$ as $M\to\infty$,*

$$
\sup_{|h|\leq\delta_N}\int_{-\pi}^{\pi}|K_M(\lambda+h)-K_M(\lambda)|\,d\lambda = O(M\delta_N).
$$

Proof of Lemma 3.16. As K_M is symmetric and periodic we only consider $h>0$. Using the compact support of K and the definition of K_M,

$$
\begin{aligned}
&\sup_{|h|\leq\delta_N}\int_{-\pi}^{\pi}|K_M(\lambda+h)-K_M(\lambda)|\,d\lambda \\
&= \sup_{|h|\leq\delta_N} M\int_{-\pi/M}^{\pi/M}|K(M[\lambda+h])-K(M\lambda)|\,d\lambda \\
&= \sup_{|h|\leq\delta_N} M\left[\int_{-\pi/M}^{\pi/M-h}|K(M[\lambda+h])-K(M\lambda)|\,d\lambda + \int_{\pi/M-h}^{\pi/M}|K(M\lambda)|\,d\lambda\right] \\
&= \sup_{|h|\leq\delta_N}\left[\int_{-\pi}^{\pi-Mh}|K(x+hM)-K(x)|\,dx + \int_{\pi-Mh}^{\pi}|K(x)|\,dx\right] \\
&= O(M\delta_N),
\end{aligned}
$$

where for the last step we have used the Lipschitz property of K. □

Proof of Proposition 3.1. We do the proof in two steps.

First step. We study the difference:

$$A = \left|\mathrm{Trace}\left[(\Sigma_N W_M)^s\right] - N(2\pi)^{2s-1}\int_\Pi f^s(\lambda)K_M^s(\lambda)d\lambda\right|. \tag{3.40}$$

First we rewrite the trace as

$$\begin{aligned}
&\mathrm{Trace}\left[(\Sigma_N W_M)^s\right] \\
&= \sum_{1\le r_1,\dots,r_{2s}\le N} \gamma(r_1-r_2)\omega(\frac{r_2-r_3}{M})\cdots\gamma(r_{2s-1}-r_{2s})\omega(\frac{r_{2s}-r_1}{M}) \\
&= \sum_{1\le r_1,\dots,r_{2s}\le N} \int_{\Pi^{2s}} f(\lambda_1)K_M(\lambda_2)\cdots f(\lambda_{2s-1})K_M(\lambda_{2s}) \\
&\qquad \times \exp\left\{i\left[\lambda_1(r_1-r_2)+\lambda_2(r_2-r_3)+\cdots+\lambda_{2s}(r_{2s}-r_1)\right]\right\}d\lambda_1\cdots\lambda_{2s} \\
&= \sum_{1\le r_1,\dots,r_{2s}\le N} \int_{\Pi^{2s}} f(\lambda_1)K_M(\lambda_2)\cdots f(\lambda_{2s-1})K_M(\lambda_{2s})\exp\left\{i\sum_{j=1}^{2s} r_j\mu_j\right\}d\lambda_1\cdots\lambda_{2s} \\
&= N\int_{\Pi^{2s}} f(\lambda-\mu_2-\cdots-\mu_{2s})K_M(\lambda-\mu_3-\cdots-\mu_{2s})\cdots f(\lambda-\mu_{2s})K_M(\lambda) \\
&\qquad \times(2\pi)^{2s-1}\Phi_N^{(2s)}(\mu_1,\cdots,\mu_{2s})d\lambda d\mu_2\cdots d\mu_{2s}
\end{aligned}$$

where we have made the change of variable

$$\begin{cases} \mu_1 = \lambda_1 - \lambda_{2s} \\ \mu_2 = \lambda_2 - \lambda_1 \\ \dots \\ \mu_{2s} = \lambda_{2s} - \lambda_{2s-1}, \end{cases}$$

(it is immediate that $\sum_{j=1}^{2s}\mu_j = 0$), and setting $\lambda = \lambda_{2s}$, we express all the variables λ_j in terms of λ and μ_j, $j = 2,\dots,2s$:

$$\begin{cases} \lambda_{2s-1} = \lambda - \mu_{2s} \\ \lambda_{2s-2} = \lambda - \mu_{2s} - \mu_{2s-1} \\ \dots \\ \lambda_1 = \lambda - \mu_{2s} - \cdots - \mu_2 = \lambda - \mu_1. \end{cases}$$

Then we have that

$$\begin{aligned}
A \le \int_{\Pi^{2s}} &\left|f(\lambda-\mu_2\cdots-\mu_{2s})K_M(\lambda-\mu_3\cdots-\mu_{2s})\cdots f(\lambda-\mu_{2s})K_M(\lambda) - f^s(\lambda)K_M^{s-1}(\lambda)\right| \\
&\times N(2\pi)^{2s-1}\left|K_M(\lambda)\Phi_N^{(2s)}(\mu_1,\dots,\mu_{2s})\right| d\lambda d\mu_2\cdots d\mu_{2s}.
\end{aligned} \tag{3.41}$$

We split the above integral into two sets, for small and for big μ_j. For a sequence $\delta_N = o(M^{-1})$, to be chosen optimally later, define the set

$$D = \left\{\mu \in \mathcal{R}^{2s-1} : \sup_j |\mu_j| \leq \delta_N\right\}.$$

Taking into account that $|\lambda| \leq \pi/M$ due to the compactness of the support of K, in the set D all the functions f are boundedly differentiable. Then we can use the inequality

$$|A_1 \cdots A_r - B_1 \cdots B_r| \leq \sum_{q=0}^{r-1} |B_1 \cdots B_q (B_{q+1} - A_{q+1}) A_{q+2} \cdots A_r| \tag{3.42}$$

and $\sup_\lambda |K_M(\lambda)| = O(M)$ to bound the integral of (3.41) over D with

$$O(NM^{s-1})\sum_{q=0}^{s-1}\int_\Pi\int_D |f(\lambda - \mu_{2+2q} \ldots - \mu_{2s}) - f(\lambda)| \left|K_M(\lambda)\Phi_N^{(2s)}(\mu)\right| d\lambda d\mu_2 \cdots d\mu_{2s} \tag{3.43}$$

$$+O(NM^{s-1})\sum_{q=0}^{s-2}\int_\Pi\int_D |K_M(\lambda - \mu_{3+2q} \ldots - \mu_{2s}) - K_M(\lambda)| \left|\Phi_N^{(2s)}(\mu)\right| d\lambda d\mu_2 \cdots d\mu_{2s}. \tag{3.44}$$

Then, applying the Mean Value Theorem and using (3.32) we obtain that (3.43) is

$$O(NM^{s-1}) \int_\Pi |K_M(\lambda)| d\lambda \sum_{q=2}^{2s-1} \int_\Pi |\mu_q| \left|\Phi_N^{(2s)}(\mu)\right| d\mu_2 \cdots d\mu_{2s} = O(M^{s-1} \log^{2s-1} N). \tag{3.45}$$

On the other hand, applying Lemma 3.16 and using (3.29), (3.44) is of order

$$O(NM^s \delta_N). \tag{3.46}$$

Denote as D^c the complementary of D in Π^{2s-1}. The integral in A corresponding to the set D^c is then less than

$$\begin{aligned} & N(2\pi)^{2s-1}\int_\Pi\int_{D^c} |f(\lambda-\mu_2-\cdots-\mu_{2s})K_M(\lambda-\mu_3-\cdots-\mu_{2s})\cdots f(\lambda-\mu_{2s})K_M(\lambda)| \\ & \qquad \times|\Phi_N^{(2s)}(\mu_1, \cdots, \mu_{2s})| d\lambda d\mu_2 \cdots d\mu_{2s} & (3.47) \\ & + \; N(2\pi)^{2s-1} \int_\Pi |f^s(\lambda) K_M^s(\lambda)| d\lambda \int_{D^c} |\Phi_N^{(2s)}(\mu_1, \cdots, \mu_{2s})| d\mu_2 \cdots d\mu_{2s}. & (3.48) \end{aligned}$$

The expression in (3.48) is $O(M^{s-1} \log^{2s-1} N \delta^{-1})$, by (3.31) and $\int |f^s(\lambda) K_M^s(\lambda)| d\lambda = O(M^{s-1})$, which follows from the compact support of K. Now for (3.47) we have

$$\begin{aligned} & N(2\pi)^{2s-1}\int_\Pi\int_{D^c} |f(\lambda-\mu_2-\cdots-\mu_{2s})K_M(\lambda-\mu_3-\cdots-\mu_{2s})\cdots f(\lambda-\mu_{2s})K_M(\lambda)| \\ & \qquad \times|\Phi_N^{(2s)}(\mu_1, \cdots, \mu_{2s})| d\lambda d\mu_2 \cdots d\mu_{2s} \\ \leq & \int_{D^*} |f(\lambda_1) K_M(\lambda_2) \cdots f(\lambda_{2s-1}) K_M(\lambda_{2s}) \\ & \qquad \times \varphi_N(\lambda_1 - \lambda_{2s}) \varphi_N(\lambda_2 - \lambda_1) \cdots \varphi_N(\lambda_{2s} - \lambda_{2s-1})| \, d\lambda_1 \cdots \lambda_{2s}, \end{aligned}$$

where $D^\star$ is the correspondent set to D^c with the old variables λ_j, $j = 1, \ldots, 2s$, defined by the conditions

$$D^\star = \{|\lambda_2 - \lambda_1| > \delta_N\} \cup \{|\lambda_3 - \lambda_2| > \delta_N\} \cup \ldots \cup \{|\lambda_{2s} - \lambda_{2s-1}| > \delta_N\}.$$

Also the last integral is only different from zero if

$$|\lambda_2|, |\lambda_4|, \ldots, |\lambda_{2s}| \leq \frac{\pi}{M}.$$

We are going to consider only the case where just one of the events in $D^\star$ is satisfied, $|\lambda_{2j} - \lambda_{2j-1}| > \delta_N$ $(1 \leq j \leq s)$, say, the situation with an odd index or with more than one event is dealt with in a similar or simpler way.

First, if $|\lambda_{2j} - \lambda_{2j-1}| > \delta_N$, then $|\varphi_N(\lambda_{2j} - \lambda_{2j-1})| = O(\delta_N^{-1})$. Second, we can bound the integrals in λ_{2j} and λ_{2j-1} in this way:

$$\int_\Pi |\varphi_N(\lambda_{2j+1} - \lambda_{2j}) K_M(\lambda_{2j})| \, d\lambda_{2j} = O(M \log N),$$

using (3.35) and

$$\int_\Pi |\varphi_N(\lambda_{2j-1} - \lambda_{2j-2}) f(\lambda_{2j-1})| \, d\lambda_{2j-1} = \int_{|\lambda_{2j-1}| \leq \epsilon} + \int_{|\lambda_{2j-1}| > \epsilon} . \tag{3.49}$$

If $|\lambda_{2j-1}| \leq \epsilon$ then $f(\lambda_{2j-1})$ is bounded and the correspondent integral is of order $O(\log N)$. If $|\lambda_{2j-1}| > \epsilon$, as $|\lambda_{2j-2}| < \pi/M$, we obtain that $|\lambda_{2j-1} - \lambda_{2j-2}| > \epsilon/2$, say, as $M \to \infty$, and then $|\varphi_N(\lambda_{2j-1} - \lambda_{2j-2})| = O(1)$. Thus the second integral is finite due to the integrability of f. Then (3.49) is $O(\log N)$.

There are $s-1$ integrals of each type, which can be handled in the same way. The remaining integral is of this general form:

$$\int_\Pi \int_\Pi |K_M(\lambda_{2s}) f(\lambda_1) \varphi_N(\lambda_1 - \lambda_{2s})| \, d\lambda_1 d\lambda_{2s} = O(\log N),$$

since, as before, the integral in λ_1 is $O(\log N)$, and $\int |K_M(\lambda_{2s})| d\lambda_{2s}$ is $O(1)$. Finally the integral over $D^\star$ is of order

$$O(\delta_N^{-1} M^{s-1} \log^{2s-1} N), \tag{3.50}$$

and compiling results we can obtain

$$A = O\left(M^{s-1} \log^{2s-1} N + \delta_N N M^s + \delta_N^{-1} M^{s-1} \log^{2s-1} N\right) = O\left(N M^{s-1} e_N(s)\right),$$

where $e_N(s) = [N^{-1}M\log^{2s-1}N)]^{1/2} \to 0$ under the conditions of the lemma, and we have made the optimal choice $\delta_N = [(MN)^{-1})\log^{2s-1}N]^{1/2} = o(M^{-1})$.

Second step. If we define

$$
\begin{aligned}
C_M(s) &= \sum_{j=0}^{d} L_s(j)M^{s-1-j} \\
&= \sum_{j=0}^{d} \frac{1}{j!}\left(\frac{d}{d\lambda}\right)^j f^s(0)K_j^{(s)}M^{s-1-j} \\
&= \sum_{j=0}^{d} \frac{1}{j!}\left(\frac{d}{d\lambda}\right)^j f^s(0)\int_\Pi \lambda^j K_M^s(\lambda)d\lambda,
\end{aligned}
$$

then, as $M \to \infty$,

$$
\begin{aligned}
\left|\int_\Pi K_M^s(\lambda)f^s(\lambda)d\lambda - C_M(s)\right| &\leq \int_\Pi |K_M(\lambda)|^{s-1}\left|f^s(\lambda) - \sum_{j=0}^{d}\frac{1}{j!}\left(\frac{d}{d\lambda}\right)^j f^s(0)\lambda^j\right||K_M(\lambda)|d\lambda \\
&= O\left(\sup_\lambda |K_M|^{s-1}\int_\Pi |\lambda|^{d+\varrho}|K_M(\lambda)|d\lambda\right) \qquad (3.51) \\
&= O(M^{s-1-d-\varrho}),
\end{aligned}
$$

using the Lipschitz property of $f^{(d)}(\lambda)$ in the same way as in the proof of Lemma 3.2. □

3.9 Appendix: Proofs of Section 3.4

Proof of Proposition 3.2. We can write

$$
\begin{aligned}
&\mathbf{1}'(\Sigma_N W_M)^s \Sigma_N \mathbf{1} \\
&= \sum_{0\leq r_1,\ldots,r_{2s+2}\leq N-1} \gamma(r_1-r_2)\cdots\omega\left(\frac{r_{2s}-r_{2s+1}}{M}\right)\gamma(r_{2s+1}-r_{2s+2}) \\
&= \sum_{1\leq r_1,\ldots,r_{2s+2}\leq N} \int_{\Pi^{2s+1}} f(\lambda_1)K_M(\lambda_2)\cdots K_M(\lambda_{2s})f(\lambda_{2s+1}) \\
&\qquad \times \exp\{i[\lambda_1(r_1-r_2) + \lambda_2(r_2-r_3) + \cdots + \lambda_{2s+1}(r_{2s+1}-r_{2s+2})]\}\,d\lambda_1\cdots\lambda_{2s+1} \\
&= (2\pi)^{2s+1}N\int_{\Pi^{2s+1}} f(\lambda_1)K_M(\lambda_2)\cdots K_M(\lambda_{2s})f(\lambda_{2s+1}) \\
&\qquad \times \Phi_N^{(2s+2)}(\lambda_1, \lambda_2-\lambda_1, \cdots, \lambda_{2s+1}-\lambda_{2s}, -\lambda_{2s+1})d\lambda_1\cdots d\lambda_{2s+1} \\
&= (2\pi)^{2s+1}N\int_{\Pi^{2s+1}} f(\mu_1)K_M(\mu_1+\mu_2)\cdots K_M(\mu_1+\cdots+\mu_{2s})f(\mu_1+\cdots+\mu_{2s+1}) \\
&\qquad \times \Phi_N^{(2s+2)}(\mu_1, \cdots, \mu_{2s+1}, -\textstyle\sum_{j=1}^{2s+1}\mu_j)d\mu_1\cdots d\mu_{2s+1}, \qquad (3.52)
\end{aligned}
$$

where we have made the following change of variable:

$$\begin{cases} \mu_1 = \lambda_1 \\ \mu_2 = \lambda_2 - \lambda_1 \\ \dots \\ \mu_{2s+1} = \lambda_{2s+1} - \lambda_{2s} \\ \mu_{2s+2} = -\lambda_{2s+1} \end{cases}$$

with $\sum_{j=1}^{2s+2} \mu_j \equiv 0$, and

$$\begin{cases} \lambda_1 = \mu_1 \\ \lambda_2 = \mu_1 + \mu_2 \\ \dots \\ \lambda_{2s} = \mu_1 + \cdots + \mu_{2s} \\ \lambda_{2s+1} = \mu_1 + \cdots + \mu_{2s+1}. \end{cases}$$

To study the difference between the integral in (3.52) and $[f(0)]^{s+1}[K_M(0)]^s$ we divide the range of integration, Π^{2s+1}, into two sets, D and its complementary D^c, where D is defined by the condition

$$D = \{|\mu_j| \leq \pi/[M(2s+2)],\ j = 1, \dots, 2s+1\}.$$

In this case we only need the smoothness properties of K at the origin (inside D). For the difference in the set D, we can use inequality (3.42), the Lipschitz property of K and the differentiability of f:

$$\begin{aligned} & \Big| \int_D f(\mu_1) K_M(\mu_1+\mu_2) \cdots K_M(\mu_1+\cdots+\mu_{2s}) f(\mu_1+\cdots+\mu_{2s+1}) \\ & \qquad \times \Phi_N^{(2s+2)}(\mu_1, \cdots, \mu_{2s+1}, -\textstyle\sum_{j=1}^{2s+1} \mu_j) d\mu_1 \cdots d\mu_{2s+1} \\ & \quad - \int_D f^{s+1}(0) K_M^s(0) \Phi_N^{(2s+2)}(\mu_1, \cdots, \mu_{2s+1}, -\textstyle\sum_{j=1}^{2s+1} \mu_j) d\mu_1 \cdots d\mu_{2s+1} \Big| \\ \leq \ & O(M^{s+1}) \int_{\Pi^{2s+1}} \sum_{j=2}^{2s} |\mu_j| \left| \Phi_N^{(2s+2)}(\mu_1, \cdots, \mu_{2s+1}, -\textstyle\sum_{j=1}^{2s+1} \mu_j) \right| d\mu_1 \cdots d\mu_{2s+1} \\ = \ & O\Big(M^{s+1} N^{-1} \log^{2s+1} N\Big), \end{aligned} \tag{3.53}$$

using (3.32) for the last step.

Now, focusing in the integral over the set D^c, we can see in first instance that, by (3.31),

$$\Big| \int_{D^c} f(\mu_1) K_M(\mu_1+\mu_2) \cdots K_M(\mu_1+\cdots+\mu_{2s}) f(\mu_1+\cdots+\mu_{2s+1})$$

$$
\begin{aligned}
& \times \Phi_N^{(2s+2)}(\mu_1,\cdots,\mu_{2s+1},-\textstyle\sum_{j=1}^{2s+1}\mu_j)d\mu_1\cdots d\mu_{2s+1} \\
& - \displaystyle\int_{D^c} f^{s+1}(0)K_M^s(0)\Phi_N^{(2s+2)}(\mu_1,\cdots,\mu_{2s+1},-\textstyle\sum_{j=1}^{2s+1}\mu_j)d\mu_1\cdots d\mu_{2s+1}\Big| \\
\le\ & \displaystyle\int_{D^c} |f(\mu_1)K_M(\mu_1+\mu_2)\cdots K_M(\mu_1+\cdots+\mu_{2s})f(\mu_1+\cdots+\mu_{2s+1})| \\
& \times \left|\Phi_N^{(2s+2)}(\mu_1,\ldots,\mu_{2s+1},-\textstyle\sum_{j=1}^{2s+1}\mu_j)\right| d\mu_1\cdots d\mu_{2s+1} \qquad (3.54) \\
& + O\Big(M^{s+1}N^{-1}\log^{2s+1}N\Big) .
\end{aligned}
$$

As we did in the previous appendix, the integral over D^c in (3.54) is less or equal than

$$
\begin{aligned}
& \frac{1}{(2\pi)^{2s+1}N}\int_{D^\star} |f(\lambda_1)K_M(\lambda_2)\cdots K_M(\lambda_{2s})f(\lambda_{2s+1}) \\
& \times \varphi_N(\lambda_1)\varphi_N(\lambda_2-\lambda_1)\cdots\varphi_N(\lambda_{2s+1}-\lambda_{2s})\varphi_N(-\lambda_{2s+1})|\, d\lambda_1\cdots d\lambda_{2s+1}, \qquad (3.55)
\end{aligned}
$$

where $D^\star$ is the set

$$\{|\lambda_1| > \pi/[M(2s+2)]\}\cup\{|\lambda_2-\lambda_1| > \pi/[M(2s+2)]\}\cup\ldots\cup\{|\lambda_{2s-1}+\lambda_{2s}| > \pi/[M(2s+2)]\}.$$

Also, the integral in (3.55) is different from zero only if $|\lambda_2|, |\lambda_4|, \ldots, |\lambda_{2s}| \le \pi/M$.

If we are in the situation where $|\lambda_{j+1}-\lambda_j| > \pi/[M(2s+2)]$ for at least one $j \in \{1,\ldots,2s\}$ we can repeat the same process of Proposition 3.1 to obtain a bound of order $O(N^{-1}M^{s+1}\log^{2s+1}N)$ for this part of the integration in (3.55).

Let's study the case in which $|\lambda_1| > \pi/[M(2s+2)]$. First, $|\varphi_N(\lambda_1)| = O(M)$. Now truncating the integral at $|\lambda_1| = \epsilon$,

$$\int_\Pi f(\lambda_1)|\varphi_N(\lambda_2-\lambda_1)|d\lambda_1 = O(\log N),$$

as $|\lambda_2-\lambda_1| > \epsilon/2$ if $|\lambda_1| > \epsilon$ and $|\lambda_2| \le \epsilon/[M(2s+2)]$, since $M \to \infty$. Now

$$\int_\Pi |K_M(\lambda_2)\varphi_N(\lambda_3-\lambda_2)|\, d\lambda_2 = O(M\log N),$$

and the integrals with respect the rest of variables can be bounded in the same way, being (3.55) of order $O(N^{-1}M^{s+1}\log^{2s+1}N)$ again.

Therefore, from (3.53), (3.54) and the previous discussion for (3.55), we have that (note that we have dropped the N in (3.52)),

$$\mathbf{1}'(\Sigma_N W_M)^s\Sigma_N\mathbf{1} = (2\pi)^{2s+1}N[f(0)]^{s+1}[K_M(0)]^s + O\left(M^{s+1}\log^{2s+1}N\right). \quad \square$$

Proof of Lemma 3.4. Similarly to Feller (1971, p. 535) or Durbin (1980a, p. 325) we have for complex α, β

$$|e^{\alpha} - 1 - \beta| \leq e^{\gamma}\left\{|\alpha - \beta| + \frac{|\beta|^2}{2}\right\}, \tag{3.56}$$

where $\gamma = \max\{|\alpha|, |\beta|\}$. Let's take (with $\tau = 2$ in (3.12)):

$$\begin{aligned}
\alpha &= \log\varphi(\mathbf{t}) - \frac{1}{2}||i\mathbf{t}||^2 \\
&= \left(\frac{M}{N}\right)^{1/2} \sum_{|\mathbf{r}|=3} \frac{1}{r_1!r_2!}\bar{\kappa}[r_1, r_2](it_1)^{r_1}(it_2)^{r_2} + R_N(2)
\end{aligned}$$

and

$$\beta = \left(\frac{M}{N}\right)^{1/2} \frac{1}{3!}\left\{C(0,3)\Gamma_3(j)(it_2)^3 + C(2,1)J_1(j)(it_1)^2(it_2)\right\},$$

then we have, using (3.10) and Lemma 3.3 for $s = 3$,

$$\begin{aligned}
|\alpha - \beta| &\leq \left|\left(\frac{N}{M}\right)^{-1/2} O\Big(M^{-2} + e_N(3)\Big)\left[(it_2)^3 + (it_1)^2(it_2)\right] + \frac{M}{N}\left[R_{04}(it_2)^4 + R_{22}(it_1)^2(it_2)^2\right]\right| \\
&\leq F_1(||\mathbf{t}||)O\left(\left(\frac{N}{M}\right)^{-1/2}\left[M^{-2} + e_N(3)\right] + \frac{M}{N}\right),
\end{aligned}$$

where F_1 is a polynomial of degree 4. Now

$$\frac{|\beta|^2}{2} \leq F_2(||\mathbf{t}||)O\left(\frac{M}{N}\right), \tag{3.57}$$

where F_2 is a polynomial of degree 6. Then

$$|\alpha - \beta| + \frac{|\beta|^2}{2} \leq F(||\mathbf{t}||)O\left(\left(\frac{N}{M}\right)^{-1/2}\left[M^{-2} + e_N(3)\right] + \frac{M}{N}\right) \tag{3.58}$$

for some polynomial F. Now to study γ, we first bound $|\beta|$ for $||\mathbf{t}|| \leq \delta_\beta\sqrt{N/M}$, $\delta_\beta > 0$:

$$\begin{aligned}
|\beta| &\leq ||\mathbf{t}||^2\left\{\frac{1}{3!}\left(\frac{N}{M}\right)^{-1/2}\left[|C(0,3)\Gamma_3(0)| + 3|C(2,1)J_1(0)|\right]||\mathbf{t}||\right\} \\
&\leq ||\mathbf{t}||^2\left\{\frac{\delta_\beta}{3!}\left[|C(0,3)\Gamma_3(0)| + 3|C(2,1)J_1(0)|\right]\right\} \\
&\leq ||\mathbf{t}||^2 T_\beta,
\end{aligned} \tag{3.59}$$

with $0 < T_\beta < 1/4$ if we choose δ_β small enough. Now for α we can choose a $\delta_\alpha > 0$ small enough, such that, for $||\mathbf{t}|| \leq \delta_\alpha\sqrt{N/M}$,

$$|\alpha| \leq ||\mathbf{t}||^2\left\{\frac{1}{3!}\left(\frac{N}{M}\right)^{-1/2}\left[|C(0,3)\Gamma_3(0)| + 3|C(2,1)J_1(0)| + O\Big(M^{-2} + e_N(3)\Big)\right]||\mathbf{t}||\right.$$

$$
\begin{aligned}
&\quad + \frac{M}{N}\left[|R_{04}| + |R_{22}|\right] \|\mathbf{t}\|^2 \Big\} \\
&\leq \|\mathbf{t}\|^2 \left\{ \frac{\delta_\alpha}{3!} \left[|C(0,3)\Gamma_3(0)| + 3|C(2,1)J_1(0)| + O\left(M^{-2} + e_N(3)\right)\right] + \delta_\alpha^2 [|R_{04}| + |R_{22}|] \right\} \\
&\leq \|\mathbf{t}\|^2 \left\{ \frac{1}{4} + O\left(M^{-2} + e_N(3)\right) \right\}. \qquad (3.60)
\end{aligned}
$$

From (3.59) and (3.60) we have that

$$
e^\gamma \leq \exp\left\{ \|\mathbf{t}\|^2 \left[\frac{1}{4} + O\left(M^{-2} + e_N(3)\right) \right] \right\}
$$

for $\|\mathbf{t}\| \leq \delta_1 \sqrt{N/M}$ where $\delta_1 = \min\{\delta_\alpha, \delta_\beta\}$. Then,

$$
\begin{aligned}
\exp\left\{ -\frac{1}{2}\|\mathbf{t}\|^2 + \gamma \right\} &\leq \exp\left\{ \|\mathbf{t}\|^2 \left[-\frac{1}{4} + O\left(M^{-2} + e_N(3)\right) \right] \right\} \\
&\leq \exp\left\{ -d_1 \|\mathbf{t}\|^2 \right\} \qquad (3.61)
\end{aligned}
$$

for one $d_1 > 0$, $\|\mathbf{t}\| \leq \delta_1\sqrt{N/M}$. Since our approximation to $\varphi(\mathbf{t}) = \exp\{\frac{1}{2}\|i\mathbf{t}\|^2 + \alpha\}$ is $A(\mathbf{t},2) = \exp\left\{\frac{1}{2}\|i\mathbf{t}\|^2\right\}[1+\beta]$, using (3.58) and (3.61) the lemma is proved. □

Proof of Lemma 3.5. First, following Bentkus and Rudzkis (1982) we study the characteristic function of the estimate of the spectral density, which itself appears in the joint characteristic function. Define

$$
\begin{aligned}
\tau(t_2) &= E\left[\exp\left\{ it_2 \sqrt{N/M} q_2 \right\} \right] \\
&= E\left[\exp\left\{ it_2 \sqrt{N/M} \frac{2\pi(\hat{f}(0) - E[\hat{f}(0)])}{V_N \sigma_N} \right\} \right] \\
&= |I - 2it_2 \Sigma_N Q_N|^{-1/2} \exp\{-it_2 E\} \\
&= \left| I - \frac{2it_2}{\sqrt{NM}\sigma_N V_N} \Sigma_N W_M \right|^{-1/2} \exp\left\{ \frac{-it_2}{\sqrt{NM}\sigma_N V_N} \mathrm{Trace}[\Sigma_N W_M] \right\}.
\end{aligned}
$$

Now define

$$
\begin{aligned}
\tau'(t_2) &= \left| I - \frac{2it_2}{\sqrt{NM}\sigma_N V_N} \Sigma_N W_M \right|^{-1/2} \\
&= \prod_{j=1}^{N} \left(1 - 2it_2 \frac{\mu_j}{\sqrt{NM}\sigma_N V_N} \right)^{-1/2}
\end{aligned}
$$

where μ_j are the eigenvalues of the matrix $\Sigma_N W_M$. Obviously $|\tau(t_2)| = |\tau'(t_2)|$. Now as

$$
\begin{aligned}
1 = \mathrm{Var}[\sqrt{T} q_2] &= \frac{1}{MN} \frac{1}{\sigma_N^2 V_N^2} 2\mathrm{Trace}[(\Sigma_N W_M)^2] \\
&= \frac{1}{MN} \frac{2}{\sigma_N^2 V_N^2} \sum_{j=1}^{N} \mu_j^2,
\end{aligned}
$$

we can obtain

$$\sum_{j=1}^{N} \mu_j^2 = \frac{1}{2}\sigma_N^2 V_N^2 MN = O(MN).$$

Also we have that

$$\max_j |\mu_j| = \sup_{||z||=1} |(\Sigma_N W_M z, z)| = ||\Sigma_N W_M||.$$

To evaluate the norm of the matrix $\Sigma_N W_M$,

$$||\Sigma_N W_M|| = \sup_{||z||=1} \left| \sum_{j,h} z_j z_h \int_{\Pi} f(\lambda) K_M(\omega) \varphi_N(\lambda - \omega) e^{i(h\omega - j\lambda)} d\lambda d\omega \right|, \quad (3.62)$$

first set for any vector z with $||z|| = 1$,

$$Z_N(\lambda) = \sum_{j=1}^{N} z_j e^{ij\lambda}.$$

In the integral in (3.62) we have to consider only the interval $w \in [-\pi/M, \pi/M]$, with $\pi/M \leq \epsilon$ as $M \to \infty$. Denote the supremum of $f(\lambda)$ when $\lambda \in [-\epsilon, \epsilon]$ as $||f_\epsilon||_\infty$. Then we have

$$\begin{aligned}
&\sup_{||z||=1} \int_{|\lambda| \leq \epsilon} \int_{\Pi} f(\lambda) \left| K_M(\omega) \varphi_N(\lambda - \omega) Z_N(-\lambda) Z_N(\omega) \right| d\lambda d\omega \\
\leq\ & \sup_{||z||=1} M ||K||_\infty ||f_\epsilon||_\infty \int_{\Pi} \int_{\Pi} \left| \varphi_N(\lambda - \omega) Z_N(-\lambda) Z_N(\omega) \right| d\lambda d\omega \\
=\ & \sup_{||z||=1} M ||K||_\infty ||f_\epsilon||_\infty \int_{\Pi} \int_{\Pi} \left| \varphi_N(\alpha) Z_N(-\alpha - \omega) Z_N(\omega) \right| d\alpha d\omega \\
\leq\ & \sup_{||z||=1} M ||K||_\infty ||f_\epsilon||_\infty \int_{\Pi} |\varphi_N(\alpha)| \left[\int_{\Pi} |Z_N(-\alpha - \omega)|^2 d\omega \int_{\Pi} |Z_N(\omega)|^2 d\omega \right]^{1/2} d\alpha \\
\leq\ & 2\pi M ||K||_\infty ||f_\epsilon||_\infty \int_{\Pi} |\varphi_N(\alpha)| d\alpha \\
\leq\ & c(f, K) M \log N, \qquad (3.63)
\end{aligned}$$

where $c(f, K)$ is an absolute constant depending on f and K, and we have used the change of variable $\alpha = \lambda - \omega$ and the fact that $\int_{\Pi} |Z_N(\omega)|^2 d\omega = 2\pi$. For the other values of λ we can see that $|\lambda| > \epsilon$ and $|\omega| \leq \pi/M$ imply $|\lambda - \omega| > \epsilon/2$, say, as $M \to \infty$, so $|\varphi_N(\lambda - \omega)| \leq$ const. Then, for $1 < p \leq 2$ and using $\sup_z |Z_N| \leq \sqrt{N}$ and Hölder inequality,

$$\sup_{||z||=1} \int_{|\lambda| > \epsilon} \int_{\Pi} f(\lambda) \left| K_M(\omega) \varphi_N(\omega - \lambda) Z_N(-\lambda) Z_N(\omega) \right| d\lambda d\omega$$

$$
\begin{aligned}
&\leq \text{const} \sup_{\|z\|=1} \int_\Pi \int_\Pi f(\lambda) \left| K_M(\omega) Z_N(\omega) Z_N(-\lambda) \right| d\lambda d\omega \\
&\leq \text{const} \sup_{\|z\|=1} \left[\int_\Pi |K_M(\omega)|^2 \, d\omega \int_\Pi |Z_N(\omega)|^2 \, d\omega \right]^{1/2} \left[\int_\Pi f^p(\lambda) d\lambda \right]^{1/p} \\
&\quad \times \left[\int_\Pi |Z_N(-\lambda)|^{\frac{p}{p-1}} \, d\lambda \right]^{\frac{p-1}{p}} \\
&\leq \text{const} \|K\|_\infty^{1/2} \|K\|_1 \|f\|_p N^{\frac{2-p}{2p}} M^{1/2} \\
&= c'(f, K) N^{\frac{2-p}{2p}} M^{1/2}, \qquad (3.64)
\end{aligned}
$$

using $\sup_z |Z_N| \leq \sqrt{N}$ and $\int |Z_N|^2 = 2\pi$. Then from (3.63) and (3.64), and N big enough

$$
\|\Sigma_N W_M\| \leq c_1 \max \left\{ M \log N, N^{\frac{2-p}{2p}} M^{1/2} \right\} = c_1 \vartheta_N,
$$

say, where c_1 is a constant depending on f and K. Thus we can write

$$
\max_j |\mu_j| \leq c_1 \vartheta_N.
$$

Introduce now the notation

$$
g_j = \mu_j [c_1 \vartheta_N]^{-1}
$$

where $|g_j| \leq 1$. Now we have

$$
\begin{aligned}
\sum_{j=1}^N g_j^2 &= \left(\frac{1}{c_1 \vartheta_N} \right)^2 \sum_{j=1}^N \mu_j^2 \\
&= \frac{1}{2c_1^2} \sigma_N^2 V_N^2 \frac{MN}{\vartheta_N^2},
\end{aligned}
$$

and (note that $NM/\vartheta_N^2 \to \infty, \ \forall p > 1$)

$$
\begin{aligned}
|\tau(t_2)| &= \prod_{j=1}^N \left(1 + 4t_2^2 \frac{\mu_j^2}{MN \sigma_N^2 V_N^2} \right)^{-1/4} \\
&= \prod_{j=1}^N \left(1 + 4t_2^2 \frac{c_1^2 g_j^2 \vartheta_N^2}{\sigma_N^2 V_N^2 MN} \right)^{-1/4} \\
&\leq \prod_{j=1}^N \left(1 + t_2^2 \frac{\vartheta_N^2}{MN} \frac{4c_1^2}{\sigma_N^2 V_N^2} \right)^{-\frac{1}{4} g_j^2} \\
&= \left(1 + t_2^2 \frac{\vartheta_N^2}{NM} \frac{4c_1^2}{\sigma_N^2 V_N^2} \right)^{-\frac{1}{4} \frac{1}{2c_1^2} \frac{NM}{\vartheta_N^2} \sigma_N^2 V_N^2} \\
&= \left(1 + t_2^2 \frac{\vartheta_N^2}{NM} \left[c_2 + O(M^{-2} + e_N(2)) \right] \right)^{-\frac{1}{2} \left[c_2^{-1} + O(M^{-2} + e_N(2)) \right] \frac{NM}{\vartheta_N^2}},
\end{aligned}
$$

where $c_2 = c_1^2/(\pi^2 4\pi f^2(0)\|K\|_2^2)$ is a constant from the expansion for $\sigma_N^2 V_N^2$, and we have applied $(1+at) \geq (1+t)^a$, valid for $t \geq 0$, $0 \leq a \leq 1$.

So for all $\eta > 0$, as $N, M \to \infty$ we have that

$$|\tau(t_2)| \leq (1+\eta_1^2)^{-\eta_2 \frac{NM}{\vartheta_N^2}} \tag{3.65}$$

for $|t_2| > \eta\sqrt{NM}/\vartheta_N$ and for $\eta_1 > 0$ and $\eta_2 > 0$ depending on η.

Then returning to the bivariate characteristic function, its modulus is equal to

$$\begin{aligned} |\varphi(t_1,t_2)| &= |Det\,[I - 2it_2\Sigma_N Q_N]^{-1/2}|\exp\left\{-\frac{1}{2}t_1^2\xi_N'\Re(I-2it_2\Sigma_N Q_N)^{-1}\Sigma_N\xi_N\right\} \\ &= |\tau(t_2)|\exp\left\{-\frac{1}{2}t_1^2\xi_N'\Re(I-2it_2\Sigma_N Q_N)^{-1}\Sigma_N\xi_N\right\}, \end{aligned}$$

where $\Re$ stands for real part.

From Anderson (1958, p. 161) $\Re(\Sigma_N^{-1} - 2it_2Q_N)^{-1} = \Re(I - 2it_2\Sigma_N Q_N)^{-1}\Sigma_N$ is positive definite as $t_2 Q_N$ is real (for every N). Then $\xi_N'\Re(I - 2it_2\Sigma_N Q_N)^{-1}\Sigma_N\xi_N > 0$ for all $t_2 \in \Re$. And for $|t_2| \leq \delta\sqrt{NM}/\vartheta_N$, $\forall\delta > 0$,

$$\xi_N'\Re(I - 2it_2\Sigma_N Q_N)^{-1}\Sigma_N\xi_N > \epsilon$$

for some $\epsilon > 0$ fixed depending on δ, since we have that

$$\begin{aligned} \|\Sigma_N Q_N\| &= O\Big((MN)^{-1/2}\|\Sigma_N W_M\|\Big) \\ &= O\Big((MN)^{-1/2}\vartheta_N\Big), \end{aligned}$$

and since $\|\xi_N\| = 1/V_N$, with $V_N \to 2\pi f(0)$, $0 < f(0) < \infty$, as $N \to \infty$. Then,

$$\begin{aligned} \exp\left\{-\frac{1}{2}t_1^2\xi_N'\Re(I-2it_2\Sigma_N Q_N)^{-1}\Sigma_N\xi_N\right\} &\leq \exp\left\{-\frac{1}{2}t_1^2\epsilon_1\right\} \\ &\leq \exp\left\{-\frac{1}{4}\epsilon_1\delta_1^2\frac{NM}{\vartheta_N^2}\right\} \end{aligned} \tag{3.66}$$

for $|t_1|\sqrt{2} > \delta_1\sqrt{NM}/\vartheta_N$ and $|t_2|\sqrt{2} \leq \delta_1\sqrt{NM}/\vartheta_N$, and some $\epsilon_1 > 0$ depending on δ_1.

So now from (3.65) and (3.66) we have that for $\|\mathbf{t}\| > \delta_1\sqrt{NM}/\vartheta_N$, there exists one number $d_2 > 0$ such that

$$|\varphi(t_1,t_2)| \leq \exp\left\{-d_2\frac{NM}{\vartheta_N^2}\right\}$$

as $\left\{\mathbf{t}:\|\mathbf{t}\|>\delta_1\sqrt{NM}/\vartheta_N\right\}\subset B_1\cup B_2$ where

$$\begin{aligned}
B_1 &= \left\{\mathbf{t}:|t_2|>\frac{\delta_1}{\sqrt{2}}\sqrt{NM}/\vartheta_N\right\}\\
B_2 &= \left\{\mathbf{t}:|t_2|\le\frac{\delta_1}{\sqrt{2}}\sqrt{NM}/\vartheta_N \text{ and } |t_1|>\frac{\delta_1}{\sqrt{2}}\sqrt{NM}/\vartheta_N\right\}
\end{aligned}$$

and the lemma is proved, since

$$\begin{aligned}
\frac{NM}{\vartheta_N^2} &= \frac{NM}{\left(\max\left\{M\log N, N^{\frac{2-p}{2p}}M^{1/2}\right\}\right)^2}\\
&= MN\min\left\{\frac{1}{M^2\log^2 N}, N^{\frac{p-2}{p}}M^{-1}\right\}\\
&= m_N^2\to\infty.
\end{aligned}$$

It can be noted that $p>2$ in B2(p) does not give any further improvement in any bound.

□

Proof of Lemma 3.6. First,

$$\begin{aligned}
\|(P_N-Q^{(2)})\star\Psi_{\alpha_N}\| &= 2\sup_{B\in\mathcal{B}^2}|(P_N-Q^{(2)})\star\Psi_{\alpha_N}|\\
&\le \sup\left[|(P_N-Q^{(2)})\star\Psi_{\alpha_N}|; B\subset B(0,r_N)^c\right]\\
&\quad+\sup\left[|(P_N-Q^{(2)})\star\Psi_{\alpha_N}|; B\subset B(0,r_N)\right],
\end{aligned}$$

where $r_N=(N/M)^\beta$, ($\beta>0$ to be chosen later).

Now for $B\subset B(0,r_N)^c$ we have uniformly

$$\begin{aligned}
|(P_N-Q^{(2)})\star\Psi_{\alpha_N}| &\le |P_N\star\Psi_{\alpha_N}|+|Q^{(2)}\star\Psi_{\alpha_N}|\\
&\le \mathrm{Prob}\{\|\sqrt{N/M}\mathbf{q}\|\ge r_N/2\}\\
&\quad+2\Psi_{\alpha_N}\{B(0,r_N/2)^c\}\\
&\quad+2Q^{(2)}\{B(0,r_N/2)^c\}.
\end{aligned}$$

Now

$$Q^{(2)}\{B(0,r_N/2)^c\}=o((N/M)^{-1/2})$$

as this is the measure of a polynomial in Gaussian variables. Also

$$\mathrm{Prob}\{\|\sqrt{N/M}\mathbf{q}\|\ge r_N/2\}=o((N/M)^{-1/2}),$$

as $\sqrt{N/M}\mathbf{q}$ has finite moments of all orders. Finally, from (2.10)

$$\Psi_{\alpha_N}\{B(0,r_N/2)^c\} = O([\alpha_N/r_N]^3) = O((N/M)^{-3(\rho+\beta)}) = o((N/M)^{-1/2}),$$

since $\rho+\beta > 1/6$.

For $B \subset B(0,r_N)$ we have by Fourier Inversion

$$|(P_N - Q^{(2)}) \star \Psi_{\alpha_N}| \leq \left[\frac{1}{(2\pi)^2}\pi r_N^2\right] \int |(\widehat{P}_N - \widehat{Q}^{(2)})(\mathbf{t})\widehat{\Psi}_{\alpha_N}(\mathbf{t})|dt, \tag{3.67}$$

and as we know that $\widehat{P}_N = \varphi(\mathbf{t})$ and $\widehat{Q}^{(2)} = A(\mathbf{t},2)$, using Lemma 3.4, (3.67) is bounded by

$$O\left(\left(\frac{N}{M}\right)^{2\beta-1/2}\left[M^{-2}+e_N(3)\right]\right)\int_{\|\mathbf{t}\|\leq\delta_1\sqrt{N/M}}\left|e^{-d_1\|\mathbf{t}\|^2}F(\|\mathbf{t}\|)\right||\widehat{\Psi}_{\alpha_N}(\mathbf{t})|dt \tag{3.68}$$

$$+O((N/M)^{2\beta})\int_{\delta_1\sqrt{N/M}<\|\mathbf{t}\|\leq a'(N/M)^{\rho}}|(\widehat{P}_N - \widehat{Q}^{(2)})(\mathbf{t})\widehat{\Psi}_{\alpha_N}(\mathbf{t})|dt, \tag{3.69}$$

as, from (2.11), $\widehat{\Psi}$ is zero for $\|\mathbf{t}\| > a'(N/M)^{\rho}$ and $a' = 8\cdot 2^{4/3}\pi^{-1/3}$.

Then for (3.68) to be $o((N/M)^{-1/2})$ it is necessary to chose $\beta \leq 1/4$ (due to the definition of $e_N(3)$ and $\beta < q/(1-q)$.

Finally, from (3.15) in Lemma 3.5, valid for $\delta_1 m_N < \|\mathbf{t}\|$ and also for $\delta_1\sqrt{N/M} < \|\mathbf{t}\|$, since $m_N \leq \sqrt{N/M}$ for N big enough (from the first element in the minimum of the definition of m_N), we have that (3.69) is

$$O((N/M)^{2\beta})\int_{\delta_1\sqrt{N/M}<\|\mathbf{t}\|\leq a'(N/M)^{\rho}} e^{-d_2 m_N^2}d\mathbf{t} + o((N/M)^{-1/2}),$$

and then (3.69) is dominated by $O((N/M)^{2\beta+2\rho})e^{-d_2m_N^2}+o((N/M)^{-1/2})=o((N/M)^{-1/2})$, from (3.17). Applying the Smoothing Lemma the proof is complete. □

3.10 Appendix: Proofs of Section 3.5

Proof of Lemma 3.7. We are going to make direct use of the result due to Chibisov (1972, Theorem 2), proving that

$$\text{Prob}\left\{|Z_N| > \rho_N\sqrt{N/M}\right\} = o((N/M)^{-1/2}) \tag{3.70}$$

for some positive sequence $\rho_N \to 0$ and $\rho_N\sqrt{N/M} \to \infty$. Let's choose $\rho_N = 1/\log N$. Then we have

$$\text{Prob}\left\{|Z_N| > \rho_N\sqrt{N/M}\right\} \leq \sum_{j=1}^{3}\text{Prob}\left\{|Z_N(j)| > \frac{1}{3}\rho_N\sqrt{N/M}\right\},$$

so writing now

$$\begin{aligned}(N/M)^{-1/2}Z_N(2) &= u_1O((N/M)^{1/2}[N^{-1}\log N + M^{-d-\varrho}]) &(3.71)\\ (N/M)^{-1/2}Z_N(3) &= u_1u_2O(M^{-2}+e_N(2)) &(3.72)\end{aligned}$$

and applying Chebychev's inequality, as u_1 and u_2 have finite moments of all orders it is possible to see that for (3.70) to hold it is sufficient that the error terms in the right hand sides of (3.71) and (3.72) be $O((N/M)^{-\mu})$, for some $\mu > 0$, which is true due to 3.7, $q = 1/(1+2d)$.

Now write

$$Z_N(1) = \frac{3}{8}\left(1 + b_N + \sigma_N\theta u_2(N/M)^{-1/2}\right)^{-5/2}\sigma_N^2u_1u_2^2 = R_N(1)\cdot R_N(2),$$

say, where $R_N(2) = \frac{3}{8}\sigma_N u_1 u_2^2$ is a random variable with bounded moments of all orders. As before, in order to satisfy Chibisov condition (3.70), we need

$$\text{Prob}\left\{|R_N(1)\cdot R_N(2)| > \rho_N(M/N)^{-1/2}\right\} = o((M/N)^{1/2}), \tag{3.73}$$

but the probability in the left hand side of (3.73) is less or equal than

$$\text{Prob}\left\{|R_N(1)|(M/N)^{1/4} > \rho_N^{1/2}\right\} + \text{Prob}\left\{|R_N(2)|(M/N)^{1/4} > \rho_N^{1/2}\right\} = P_1 + P_2,$$

say. Now $P_2 = o((M/N)^{1/2})$ applying Chebychev inequality. For P_1, since $b_N = O(M^{-d}+ N^{-1}\log N)$,

$$\begin{aligned}P_1 &= \text{Prob}\left\{\left|\left(1 + b_N + \sigma_N\theta u_2(N/M)^{-1/2}\right)^{-5/2}\right|(M/N)^{1/4} > \rho_N^{1/2}\right\}\\ &= \text{Prob}\left\{\left|1 + b_N + \sigma_N\theta u_2(N/M)^{-1/2}\right|(M/N)^{-1/10} < \rho_N^{-1/5}\right\}\\ &\leq \text{Prob}\left\{\left|1 + O(M^{-d} + N^{-1}\log N) + R_N'(M/N)^{1/2}\right|(M/N)^{-1/10} < \rho_N^{-1/5}\right\},\end{aligned}$$

where R_N' is a random variable with bounded moments of all orders. Now, as $N \to \infty$, for some positive constant $c > 0$, this is not greater than

$$\begin{aligned}\text{Prob}\left\{\left|c + R_N'(M/N)^{1/2}\right| < (M/N)^{1/10}\rho_N^{-1/5}\right\} &\leq \text{Prob}\left\{\left|R_N'(M/N)^{1/2}\right| > \frac{c}{2}\right\}\\ &= o((M/N)^{1/2}),\end{aligned}$$

since $(M/N)^{1/10}\rho_N^{-1/5} \to 0$ and applying again Chebychev inequality. □

Proof of Lemma 3.8. First we observe that

$$\mathrm{Trace}[(\Sigma_N W_M \mathbf{1}\,\mathbf{1}')^s] = (\mathrm{Trace}[\mathbf{1}'\Sigma_N W_M \mathbf{1}])^s = (\mathbf{1}'\Sigma_N W_M \mathbf{1})^s.$$

Then

$$\begin{aligned} \mathbf{1}'\Sigma_N W_M \mathbf{1} &= (2\pi)^2 N \int_{\Pi^2} f(\lambda_1) K_M(\lambda_2) \Phi_N^{(3)}(\lambda_1, \lambda_2 - \lambda_1) d\lambda_1 d\lambda_2 \\ &= (2\pi)^2 N \int_{\Pi^2} f(\mu_1) K_M(\mu_1 + \mu_2) \Phi_N^{(3)}(\mu_1, \mu_2) d\mu_1 d\mu_2. \end{aligned} \tag{3.74}$$

Denote the set

$$D = \{|\mu_j| \le \pi/[2M],\ j = 1, 2\}.$$

Then, using Assumptions 3.4 and 3.1, $d = 1$,

$$\begin{aligned} &\Big|(2\pi)^2 N \int_D f(\mu_1) K_M(\mu_1 + \mu_2) \Phi_N^{(3)}(\mu_1, \mu_2) d\mu_1 d\mu_2 \\ &\qquad -N \left[(2\pi)^2 f(0) K_M(0)\right] \int_D \Phi_N^{(3)}(\mu_1, \mu_2) d\mu_1 d\mu_2 \Big| \\ &\le O(N) \int_D |f(\mu_1) K_M(\mu_1 + \mu_2) - f(0) K_M(0)|\, |\Phi_N^{(3)}(\mu_1, \mu_2)| d\mu_1 d\mu_2 \\ &= O(NM) \sum_{j=1,2} \int_\Pi \left|\mu_j \Phi_N^{(3)}(\mu_1, \mu_2)\right| d\mu_1 d\mu_2 \\ &\qquad + O\left(NM^2\right) \sum_{j=1,2} \int_\Pi \left|\mu_j \Phi_N^{(3)}(\mu_1, \mu_2)\right| d\mu_1 d\mu_2 \\ &= O(M^2 \log^2 N). \end{aligned}$$

The expression in (3.74) for the complementary of the set D can be seen to be of order of magnitude $O(M^2 \log^2 N)$, operating in the same way as in the proof of Proposition 3.1. □

Proof of Lemma 3.9. We have

$$\begin{aligned} \frac{1}{2\pi N} \mathbf{1}' W_M \mathbf{1} &= \frac{1}{2\pi N} \sum_{r_1} \sum_{r_2} \omega\left(\frac{r_1 - r_2}{M}\right) \\ &= \frac{1}{2\pi N} \sum_{r_1} \sum_{r_2} \int_\Pi K_M(\lambda) \exp\{i\lambda(r_1 - r_2)\} \\ &= \int_\Pi K_M(\lambda) \Phi_N^{(2)}(\lambda) d\lambda \\ &= K_M(0) + O\left(\frac{M^2}{N} \log N\right), \end{aligned} \tag{3.75}$$

using the Lipschitz property of K and the properties of the Fejér Kernel. $\square$

Proof of Lemma 3.10. Following the proof of Lemma 3.2, we have

$$E[\hat{f}_{m_\nu}^{(\nu)}(\alpha)] = (m_\nu)^\nu \int_{-\pi}^{\pi} V_{m_\nu}(\lambda) \int_{-\pi}^{\pi} \Phi_N^{(2)}(\theta) f(\alpha - \lambda - \theta) d\theta d\lambda.$$

We can write the bias of the spectral estimate as

$$\begin{aligned}
E[\hat{f}_{m_\nu}^{(\nu)}(\alpha)] - f_{m_\nu}^{(\nu)}(\alpha) &= (m_\nu)^\nu \int_{-\pi}^{\pi} V_{m_\nu}(\lambda) \int_{-\pi}^{\pi} \Phi_N^{(2)}(\theta) \left[f(\alpha - \lambda - \theta) - f(\alpha - \lambda)\right] d\theta d\lambda \\
&\quad + (m_\nu)^\nu \int_{-\pi}^{\pi} V_{m_\nu}(\lambda) \left[f(\alpha - \lambda) - \frac{\lambda^\nu}{(-1)^\nu \nu!} f^{(\nu)}(\alpha)\right] d\lambda \\
&= b_1 + b_2\,,
\end{aligned}$$

say. Then we can reproduce exactly the same methods of Lemma 3.2, using the properties of the kernel V_ν, to obtain

$$\begin{aligned}
b_1 &= O\Big((m_\nu)^\nu N^{-1} \log N\Big) \\
b_2 &= O((m_\nu)^{-a})\,,
\end{aligned}$$

and the lemma follows. $\square$

Proof of Lemma 3.11. Likewise for the discussion of the cumulants of the spectral estimate contained in Proposition 3.1 we can write

$$\begin{aligned}
&\frac{N}{(m_\nu)^{2\nu+1}} \mathrm{Var}[\hat{f}_{m_\nu}^{(\nu)}(\alpha)] \\
&= \frac{1}{2} \sum_\delta \frac{2\pi}{m_\nu} \int_{\Pi^4} f(\lambda + \delta\alpha - \mu_2 - \mu_3 - \mu_4) V_{m_\nu}(\lambda - \mu_3 - \mu_4) f(\lambda + \delta\alpha - \mu_4) V_{m_\nu}(\lambda) \\
&\qquad \times \Phi_N^{(4)}(\mu_1, \ldots, \mu_4) d\mu_1 \cdots d\mu_3 d\lambda \qquad (3.76) \\
&\quad + \frac{1}{2} \sum_\delta \frac{2\pi}{m_\nu} \int_{\Pi^4} f(\lambda + \delta\alpha - \mu_2 - \mu_3 - \mu_4) V_{m_\nu}(\lambda - \mu_3 - \mu_4) f(\lambda + \delta\alpha - \mu_4) V_{m_\nu}(\lambda + 2\delta\alpha) \\
&\qquad \times \Phi_N^{(4)}(\mu_1, \ldots, \mu_4) d\mu_1 \cdots d\mu_3 d\lambda, \qquad (3.77)
\end{aligned}$$

where the summation is over the two values $\delta \in \{-1, 1\}$. As in Proposition 3.1 we have to take care of the possibly unbounded values of f outside the origin. We can consider the set of integration

$$D = \{\mu \in [-\pi, \pi]^3 \,:\, |\mu_j| \leq \delta_N,\ j = 2, \ldots, 4\},$$

for the sequence $\delta_N = [(Nm_\nu)^{-1}\log^3 N]^{1/2} = o(m_\nu)$. Also Lemma 3.16 holds for V_ν, substituting M for m_ν. Then, the terms in the summation correspondent to (3.76) over the set D are equal to

$$\begin{aligned}
&\frac{2\pi}{m_\nu}\int_\Pi f^2(\lambda+\delta\alpha)V_{m_\nu}^2(\lambda)d\lambda + O\Big([N^{-1}m_\nu\log^3 N]^{1/2}\Big)\\
&= \frac{2\pi}{m_\nu}f^2(\alpha)\int_\Pi V_{m_\nu}^2(\lambda)d\lambda + O\Big([N^{-1}m_\nu\log^3 N]^{1/2} + m_\nu^{-1}\Big)\\
&= 2\pi f^2(\alpha)\|V_\nu\|_2^2 + o(1),
\end{aligned}$$

using the evenness of f and its differentiability around $f(\alpha)$. The integral in (3.76) over the complementary of the set D can be seen to be $O([N^{-1}m_\nu\log^3 N]^{1/2})$, using the finite support of V_ν and the properties of $\Phi_N^{(4)}$, as in the proof of Proposition 3.1.

The terms contained in (3.77) are identical to the ones in (3.76) if $\alpha = 0$ or negligible if $\alpha \neq 0$ as the two functions V_{m_ν} are centered in frequencies away for a positive quantity and they have compact support. □

Chapter 4

Local Cross Validation for Spectrum Bandwidth Choice

4.1 Introduction

In this chapter we propose an automatic method of determining a local bandwidth or smoothing number for nonparametric kernel spectral density estimates at a single frequency. This method is a modification of a cross-validation technique for global bandwidth choices of the discrete periodogram average type estimates.

Like in many nonparametric methods of inference, smoothed estimation of the spectral density of stationary time series relies on the choice of a bandwidth or lag number depending on the sample size. The properties of the estimates depend crucially on the value of this number. Asymptotic theory prescribes a rate for the lag number M with respect to the sample size N as this tends to infinity, but gives no practical guidance for the choice of M in finite samples. Different techniques have been proposed in the literature to that end. The usual criterion is the minimization of some estimate of the asymptotic mean square error of the estimator. This can be implemented by plug-in or cross-validation methods. Also, global and local choices are possible, depending on whether we are interested in the behaviour of the spectral density for all range of frequencies or in a concrete point or small interval.

The plug-in method consists in substituting the unknowns of the leading term in the asymptotic expression for the mean square error by consistent estimates, generally nonparametric, but also parametric ones based on approximate models can be used. Cross validation procedures avoid the use of those initial estimates and approximate the mean square error indirectly. They are based in estimates which do not use the information of the sample about the function of interest at each point (at each Fourier frequency in this case).

Beltrao and Bloomfield (1987) [BB hereafter] justified a method based on a cross-validated form of Whittle's frequency domain approximation to the likelihood function of a stationary Gaussian process. Robinson (1991) extended their results under more general conditions for a wider class of situations, and proved the consistency of the estimate of M.

This cross-validated method selects a global bandwidth for all the range of frequencies $[-\pi, \pi]$ or for a fixed subset of it. Here we propose a modified version of cross validation to justify a local bandwidth choice for a single frequency, following some ideas suggested in Robinson (1991, p. 1346), related with the work of Hurvich and Beltrao (1994) in a different context. For this single frequency choice, we only use local smoothness properties of the spectral density of the time series around this frequency, allowing for a broader range of dependence models. This local adaptation could lead also to some efficiency gains. There are related works for different nonparametric problems, like Brockmann et al. (1993) for kernel regression estimators and Lepskii and Spokoiny (1995) for projective estimates in a "signal+noise" model. Here the interval of estimation is split on degenerating intervals with the asymptotics and different smoothing parameters are estimated independently for each one.

Next section is devoted to the assumptions that we will use in this chapter, together with a brief introduction to the main cross validation concepts for nonparametric spectrum estimation. In Section 4.3 we analyze the mean square error for the spectral estimate at a fixed frequency. Section 4.4 introduces the local cross validation criterion and the main result of the paper, which is proved in an Appendix. Then we carry out a Monte Carlo analysis of the finite sample behaviour of the techniques proposed. Finally we give some lemmas required for the proof of the results in another Appendix.

4.2 Assumptions and definitions

In this section we will introduce some assumptions and definitions, together with some intuitions about cross validation and BB's results.

For completeness we present first the conditions assumed by BB. They are: $\{X_t\}$ is a Gaussian process and (cf. Theorem 3.1. of BB):

(i) $E[X_0] = 0$.

(ii) The autocovariance function $\gamma(r) = E[X_0 X_r]$ satisfies

$$\sum_1^\infty r|\gamma(r)| < \infty.$$

(iii) The spectral density $f(\lambda) = (2\pi)^{-1} \sum_{-\infty}^{\infty} \gamma(r) \exp\{ir\lambda\}$ is everywhere positive.

(iv) K is a non-negative, even, bounded function, and

$$\int_{-\infty}^{\infty} K(x)dx = 1, \quad \int_{-\infty}^{\infty} x^2 K(x)dx < \infty.$$

(v) $K(x) = \int w(y) \exp\{ixy\}dy$, where w is of compact support.

(vi) $h_N^{-1} = O(N^\rho)$, for some $\rho < 2/5$.

(vii) $h_N = o(1)$.

Given the observed data X_t, $t = 1, 2, \ldots, N$ we introduce the periodogram at the frequency $\lambda_j = 2\pi j/N$, j integer,

$$I(\lambda_j) \stackrel{def}{=} \frac{1}{2\pi N} \left| \sum_{t=1}^N X_t \exp\{it\lambda_j\} \right|^2.$$

and the averaged-periodogram spectral estimate with lag number $M = M_N = h^{-1}$, where h is the bandwidth of the estimate in BB's notation, and kernel (or spectral window) K [this function was denoted by W in BB, but we use this notation later for another analogous function],

$$\hat{f}_M(\lambda_j) \stackrel{def}{=} \sigma_M^{-1} \sum_k K(M\lambda_k) I(\lambda_j - \lambda_k),$$

where the summation runs for all values of k in the support of K (not including the zero frequency periodogram ordinate to account for mean correction). We now stress the dependence of $\hat{f}$ on M in the notation, since this is the *parameter* of interest.

The 'leave-two-out version' of that estimator [we leave only two frequencies out if K were actually compactly supported inside $[-\pi, \pi]$, as we will assume later on, or if we had defined its periodic version in that interval] is

$$\hat{f}_M^{\mathbf{j}}(\lambda_j) \stackrel{def}{=} \sigma_{j,M}^{-1} \sum_k{}' K(M\lambda_k) I(\lambda_j - \lambda_k), \tag{4.1}$$

where $\sum_k{}'$ runs for the same values as before, except in the set of indices of frequencies $\lambda_j - \lambda_k$ with the same periodogram ordinate as $I(\lambda_j)$, i.e., $k \in \{0, \pm N, \ldots\} \cup \{2j, 2j \pm N, \ldots\}$. Also, the normalizing numbers σ_M and $\sigma_{j,M}$ are equal to

$$\sigma_M \stackrel{def}{=} \sum_k K(M\lambda_k), \qquad \sigma_{j,M} \stackrel{def}{=} \sum_k{}' K(M\lambda_k).$$

Introduce the pseudo log-likelihood type criterion

$$L(f) \stackrel{def}{=} \sum_{j=1}^{N-1} \{\log f(\lambda_j) + I(\lambda_j)/f(\lambda_j)\}.$$

BB showed under the previous conditions that

$$L\left(\hat{f}_M^{\cdot}\right) - L(f) = \frac{N}{2}\text{IMSE}(M)$$

plus a term of smaller order in probability, where IMSE(M) is the discrete approximation to the Integrated Mean Squared Error of $\hat{f}_M$, weighted by f^{-1}:

$$\text{IMSE}(M) \stackrel{def}{=} N^{-1} \sum_{j=1}^{N-1} E\left[\left\{\hat{f}_M(\lambda_j) - f(\lambda_j)\right\} / f(\lambda_j)\right]^2.$$

Then minimizing $L\left(\hat{f}_M^{\cdot}\right)$ and IMSE(M) should be approximately equivalent, and this is the basis for the estimation of the M that minimizes the IMSE of $\hat{f}_M(\lambda)$ in $[-\pi, \pi]$.

If we are interested in the nonparametric spectral estimation at a single frequency (of special interest is the zero one) or we want to achieve possible efficiency gains using different bandwidths for each frequency we need a criterion to choose a local bandwidth. The Mean Square Error at a frequency ν,

$$\text{MSE}(\nu, M) \stackrel{def}{=} E\left[\left\{\hat{f}_M(\nu) - f(\nu)\right\} / f(\nu)\right]^2,$$

is the usual criterion employed to assess nonparametric estimates of this class at a single frequency. We consider only fixed frequencies of the form $\nu = 2\pi v/N$, where v is an integer. We restrict to $0 \le v \le N/2$, given the symmetry and periodicity of the periodogram and the spectral density. We will use the following Assumptions:

Assumption 4.1 *X_t, $t = 1, 2, \ldots$ is a Gaussian stationary time series.*

Assumption 4.2 *The spectral density of X_t has three uniformly bounded derivatives in an interval around the fixed frequency ν, with $f(\lambda) > 0$ for λ in that interval, and $f \in L_p[-\pi, \pi]$ for some $p > 5/3$.*

Assumption 4.3 *The function K is non-negative, even, bounded, zero outside $[-\pi, \pi]$, of bounded variation and*

$$\int_{-\infty}^{\infty} K(x)dx = 1, \quad \int_{-\infty}^{\infty} x^2 K(x)dx = \omega_2 < \infty.$$

Assumption 4.4 *The function K has Fourier transform $w(x) = (2\pi)^{-1}\int_{-\infty}^{\infty} K(\lambda)e^{i\lambda x}d\lambda$ satisfying*

$$\int |x||w(x)|dx < \infty.$$

Assumption 4.1 was used also in BB, but we do not need to assume zero mean since we avoid the zero frequency periodogram ordinate in the definition of our estimates. Assumption 4.2 only requires smoothness properties of f around the frequency we are interested in, allowing for a wide class of spectral densities, including ones with zeros and poles outside a neighbourhood of ν. The only requirement outside this band is an integrability condition to ensure ergodicity (with respect to second moments) of the series (see Lemma 4.5).

A compact support kernel in Assumption 4.3 is then the complementary of Assumption 4.2 in order to guarantee that we only use information in an interval around ν. The rest of conditions on K are standard, Assumption 4.4 being necessary to approximate $\widehat{f}_M$ with a weighted autocovariance type estimate in Lemma 4.5. From this lemma, both estimates have the same asymptotic distribution and mean square error, so the bandwidth choice techniques for one are valid for the other.

With Assumption 4.3, the summation in k in the definition of $\widehat{f}_M$ takes values in $\{j-N+1,\ldots,j-1\}$ due to the compact support kernel.

4.3 Mean square error of the nonparametric spectrum estimates

In this section we present and prove a result concerning the mean square error of the estimate $\widehat{f}_M(\nu)$. First we give two lemmas about the discrete Fourier transform and periodogram of the observed sequence, that we will need in the proof. The proofs are in the first appendix of this chapter.

Lemma 4.1 *Under Assumption 4.1, if f satisfies a uniform Lipschitz condition of order $0<\alpha\le 1$ in an interval around a fixed frequency ν, then for Fourier frequencies such that $\sup_{\lambda_\ell}|\nu-\lambda_\ell|\le const\cdot m^{-1}$, $\ell\in\{j,k\}$, for some positive sequence m such that $1/m+m/N\to 0$, uniformly in j and k, $(j,k\neq 0)$,*

$$E\left[d_x(\lambda_j)\overline{d_x(\lambda_k)}\right]-\delta_{jk}2\pi Nf(\lambda_j)=O(N^{1-\alpha}\log N),$$

where $d_x(\lambda_j)$ is the discrete Fourier transform of the series X_t,

$$d_x(\lambda_j)=\sum_{t=1}^{N}X_te^{-i\lambda_jt}.$$

Lemma 4.2 *Under Assumption 4.1, if f satisfies a uniform Lipschitz condition of order $0<\alpha\le 1$, in an interval around a fixed frequency ν and if $\sup_{\lambda_{j_r}}|\nu-\lambda_{j_r}|\le const\cdot m^{-1}$, $r=1,\ldots,q$, for some positive sequence m such that $1/m+m/N\to 0$, then, uniformly in $j_r\neq 0$, with $j_r\neq j_{r'}$, $r\neq r'$,*

$$E\left[\prod_{r=1}^{q}I(\lambda_{j_r})^{p_r}\right]=\prod_{r=1}^{q}p_r!f(\lambda_{j_r})^{p_r}+O(N^{-\alpha}\log N),\tag{4.2}$$

and

$$E\left[\prod_{r=1}^{q}\left(\frac{I(\lambda_{j_r})-f(\lambda_{j_r})}{f(\lambda_{j_r})}\right)\right]=O(N^{-\alpha}\log N).\tag{4.3}$$

Define $\|K\|_2^2=\int K^2(x)dx$. In the following lemma we study the asymptotic behaviour of the mean square error of the nonparametric estimate, distinguishing between estimates for frequencies close to the origin, and at any other frequency.

Lemma 4.3 *Under Assumptions 4.1, 4.2 and 4.3, if $M = const \cdot N^{1/5}$, for frequencies $\lambda_j = 2\pi j/N$ such that $|\nu - \lambda_j| \leq const \cdot m^{-1}$ for some positive sequence m such that $1/m + m/M \to 0$, then, uniformly in j, for $\underline{\nu > 0}$,*

$$MSE(\lambda_j, M) = \frac{M}{N} 2\pi \|K\|_2^2 f(\lambda_j)^2 + M^{-4}\left[\frac{w_2}{2} f^{(2)}(\lambda_j)\right]^2 + O\left(\frac{M^2}{N^2} + N^{-1}\right), \qquad (4.4)$$

and for $\underline{\nu = 0}$,

$$MSE(\lambda_j, M) = \frac{M}{N} 2\pi \|K\|_2^2 f(\lambda_j)^2 [1 + \delta_M(j)] + M^{-4}\left[\frac{w_2}{2} f^{(2)}(\lambda_j)\right]^2 + O\left(\frac{M^2}{N^2} + N^{-1}\right), \quad (4.5)$$

where $0 \leq \delta_M(j) \leq 1$ measures the degree of overlapping between different kernels K at a distant $2M\lambda_j$ apart when $\lambda_j \to 0$. For $j = 0$, $\delta_M(j) = 1$ $\forall M$, and for $\lambda_j > 2\pi/M$, $\delta_M(j) = 0$.

Proof. An equivalent lemma is evidently valid for more general choices of M, but we are specially interested in this particular case. We can take an $\epsilon > 0$ as small as we want, in such a way that in the interval $I_\nu = [\nu - \epsilon, \nu + \epsilon]$ the conditions of Assumption 4.2 are satisfied. Then for m big enough we have that $|\nu - \lambda_j| \leq \text{const} \cdot m^{-1}$ implies $\lambda_j \in I_\nu$. Therefore when $\nu > 0$ we have that for N big enough, $0 < \lambda_j \sim \nu$, so $(\lambda_j)^{-1} = O(1)$, where $a \sim b$ means $a/b \to 1$ as $N \to \infty$. We study first the bias and the variance.

Bias. Similarly to Theorem 5.6.1 of Brillinger (1975, p.147) and using now Lemma 4.1 with $\alpha = 1$, we get,

$$\begin{aligned} E[\widehat{f}_M(\lambda_j)] &= \int_{-\pi}^{\pi} K(\lambda) f(\lambda_j - \beta/M) d\beta + O(M/N) \\ &= f(\lambda_j) + \frac{w_2}{2} f^{(2)}(\lambda_j) M^{-2} + O\left(\frac{M}{N} + M^{-3}\right). \end{aligned}$$

The bounded variation condition on K and the derivability of f are used to approximate the discrete average of K and f by and integral with error $O(M/N)$, since by Assumption 4.2 and for M big enough we are only averaging inside I_ν, thanks to the compact support of K.

Variance. First, it is more convenient to write the spectral estimate using only N frequencies in this way:

$$\widehat{f}_M(\lambda_j) = \frac{\sigma_M^{-1}}{M} \sum_{k=1}^{N-1} K_M(\lambda_k - \lambda_j) I(\lambda_k),$$

where $K_M(\lambda) = \sum_j M\, K(M[\lambda + 2\pi j])$ is the periodic extension of $MK(M\lambda)$. Then we have

$$\begin{aligned}
\mathrm{Var}[\widehat{f}_M(\lambda_j)] &= \frac{\sigma_M^{-2}}{M^2}\sum_k K_M(\lambda_k - \lambda_j)^2 \mathrm{Var}[I(\lambda_k)] &(4.6)\\
&+\frac{\sigma_M^{-2}}{M^2}\sum_k \sum_{i\neq k} K_M(\lambda_k - \lambda_j)K_M(\lambda_i - \lambda_j)\mathrm{Cov}[I(\lambda_k), I(\lambda_i)]. &(4.7)
\end{aligned}$$

Then, from Lemma 4.2 we get $\mathrm{Var}[I(\lambda_k)] = f(\lambda_k)^2 + O(N^{-1}\log N)$ and, for $k \neq i$,

$$\mathrm{Cov}[I(\lambda_k), I(\lambda_i)] = \begin{cases} f(\lambda_k)^2 + O(N^{-1}\log N) & \text{if} \quad k = N - i \\ O(N^{-1}\log N) & \text{otherwise.} \end{cases}$$

Also we have that $\sigma_M = N/(2\pi M) + O(1)$. Then (4.6) is

$$\begin{aligned}
&\frac{(2\pi)^2}{N^2}\sum_k K_M(\lambda_k - \lambda_j)^2 f(\lambda_k)^2 + O\left(\frac{M}{N^2}\log N\right)\\
&= \frac{2\pi M}{N}\int_{-\pi}^{\pi} MK(M\lambda)^2 f(\lambda_j + \lambda)^2 d\lambda + O\left(\frac{M^2}{N^2}\right)\\
&= \frac{2\pi M}{N} f(\lambda_j)^2 \int_{-\pi}^{\pi} K(\lambda)^2 d\lambda + O\left(\frac{M}{N}\left[\frac{M}{N} + M^{-2}\right]\right).
\end{aligned}$$

In (4.7) we only have to consider the situation where $k = N - i$, since for the other frequencies we have a bound of $O(N^{-1}\log N)$ for the covariance from Lemma 4.2.

Then, if $\nu = 0$ and $\lambda_j = 0$, (4.7) is similar to (4.6). In general, if $\nu = 0$ and $|\lambda_j| \leq 2\pi/M$ then the two kernels in (4.7) overlap in some interval for all M. Taking into account only the frequencies $i = N - k$, for which the leading term of the covariance is also $f(\lambda_k)$ we have that (4.7) is equal to, using the periodicity of K_M,

$$\begin{aligned}
&\frac{\sigma_M^{-2}}{M^2}\sum_k K_M(\lambda_k - \lambda_j)K_M(\lambda_k + \lambda_j)\left[f(\lambda_k)^2 + O(N^{-1}\log N)\right] &(4.8)\\
&= \frac{2\pi}{N}\int_{-\pi}^{\pi} K_M(\lambda)K_M(\lambda + 2\lambda_j) f(\lambda - \lambda_j)^2 d\lambda + O\left((M/N)^2 + \log N\left[\frac{M}{N^2}\right]\right)\\
&= \delta_M(j) f(\lambda_j)^2 \frac{2\pi M}{N}\int_{-\pi}^{\pi} K(\lambda)^2 d\lambda + O\left(MN^{-1}\left[\frac{M}{N} + M^{-2}\right]\right),
\end{aligned}$$

for some $0 < \delta_M(j) \leq 1$.

If $|\lambda_j| > 2\pi/M$ then the two kernels in (4.8) do not overlap at all and the covariance terms do not contribute to the leading term in the variance of $\widehat{f}_M$, and the lemma follows. □

We would take $M = \tau N^{1/5}$, for some $0 < \tau < \infty$, to make the bias and the variance of the same order of magnitude and then MSE will be of order $M^{-4} \sim \text{const.}M/N$. From the previous lemma, the optimal constant $\tau^\star$ that minimizes the leading term of MSE of $\widehat{f}_M(\nu)$ is

$$\tau^\star = \left(\frac{\omega_2 f^{(2)}(\nu)^2}{2\pi \|K\|_2^2 f(\nu)^2} \right)^{1/5},$$

if $\nu \neq 0$ and with 4π instead of 2π for $\nu = 0$. Now it is possible to estimate the value of $\tau^\star$ using initial, pilot estimates of the spectral density and its second derivative at ν. This is the approach of several authors, like Andrews (1991), Newey and West (1994) or Bülhmann (1995), just to give some recent contributions. In the following section we adopt instead an indirect approach using a cross-validation argument.

4.4 Local cross validation

Consider for some positive sequence $m = m_N$ such that $m^{-1} + m/M \to 0$ as $N \to 0$, one form of *local* integrated mean square mean,

$$\text{IMSE}_m(\nu, M) \stackrel{def}{=} \frac{2\pi}{N} \sum_{j=1}^{N-1} W_m(\lambda_j - \nu) \text{E} \left[\left\{ \widehat{f}_M(\lambda_j) - f(\lambda_j) \right\} / f(\lambda_j) \right]^2,$$

where $W_m(\lambda) = m \sum_j W(m[\lambda + 2\pi j])$ for some appropriate kernel function W satisfying Assumption 4.3. For $W = (2\pi)^{-1} I_{[-\pi,\pi]}$ and $m = 1$, $\text{IMSE}_m(\nu, M) = \text{IMSE}(M)$ for all ν.

Then, from Lemma 4.3 and $\underline{\nu > 0}$, we can obtain under the same regularity conditions, for m big enough

$$\begin{aligned} \text{IMSE}_m(\nu, M) &= \frac{M}{N} 2\pi \|K\|_2^2 + M^{-4} \left[\frac{w_2}{2} \frac{f^{(2)}(\nu)}{f(\nu)} \right]^2 + O\left(\frac{M^2}{N^2} + \frac{1}{N} + \frac{M}{N} \left[\frac{m}{N} + \frac{1}{m} \right] \right) \\ &= \text{MSE}(\nu, M) + o(\text{MSE}(\nu, M)), \end{aligned}$$

where the errors in m come from the continuous approximation for the sum and because the ratio $f^{(2)}(\nu)/f(\nu)$ has bounded derivative. Therefore, $\text{IMSE}_m(\nu, M)$ approximates $\text{MSE}(\nu, M)$ when $\nu > 0$ as $m \to \infty$.

When $\underline{\nu = 0}$, we can see that

$$\text{IMSE}_m(0, M) = \frac{2\pi}{N} \sum_{j=1}^{N-1} W_m(\lambda_j) \left[\frac{M}{N} 2\pi \|K\|_2^2 \{1 + \delta_M(j)\} \right] \tag{4.9}$$

$$+M^{-4}\left[\frac{w_2}{2}\frac{f^{(2)}(0)}{f(0)}\right]^2 + O\left(\frac{M^2}{N^2} + N^{-1} + \frac{M}{N}\left[\frac{m}{N} + m^{-1}\right]\right).$$

Now in the summation in (4.9) we can consider the values of λ_j smaller and bigger than $2\pi/M$ in absolute value. Since $|\delta_M(j)| \leq 1\ \forall j$, $|\delta_M(j)| = 0$ if $|\lambda_j| > 2\pi/M$ (i.e. $|j| > N/M$) and $m/M \to 0$, with $\sup_{m,j}|W_m(\lambda_j)| = O(m)$,

$$\begin{aligned}
&\frac{2\pi}{N}\sum_j W_m(\lambda_j)\left[\frac{M}{N}2\pi\|K\|_2^2\{1+\delta_M(j)\}\right] \\
&= \frac{M}{N}2\pi\|K\|_2^2\frac{2\pi}{N}\sum_j W_m(\lambda_j) + \frac{M}{N}2\pi\|K\|_2^2\frac{2\pi}{N}\sum_{|j|\leq N/M} W_m(\lambda_j)\delta_M(j) \\
&= \frac{M}{N}2\pi\|K\|_2^2\frac{2\pi}{N}\sum_j W_m(\lambda_j) + O\left(\frac{m}{N}\right) \\
&= \frac{M}{N}2\pi\|K\|_2^2 + O\left(\frac{M}{N}\frac{m}{N}\right) + o\left(\frac{M}{N}\right). \qquad (4.10)
\end{aligned}$$

Therefore, when $\nu = 0$, $\text{IMSE}_m(0, M)$ only estimates half of the variance in $\text{MSE}(0, M)$.

A possible approach to obtain a consistent estimate of the optimal local bandwidth which minimizes $\text{MSE}(\nu, M)$, $M = \tau^\star N^{1/5}$, is to minimize one estimate of $\text{MSE}(\nu, M)$ or of $\text{IMSE}_m(\nu, M)$, which approaches MSE as m increases. Some adjustments may be necessary in the case $\nu = 0$. The presence of two related bandwidth parameters, m and M, seems to imply a circular argument like the one present in a plug-in method, where pilot estimates of the spectral density and its derivatives are used, depending on other bandwidths or parametric assumptions. To circumvent this problem we propose some procedures that connect both choices.

The logical cross validation argument in this case would be the minimization with respect to M of the function (remembering the definition in (4.1) of the 'leave-two-out' spectral estimate),

$$\text{CVLL}_m(\nu, M) \stackrel{def}{=} 2\pi\sum_{j=1}^{N-1} W_m(\lambda_j - \nu)\left[\log\hat{f}_M^{\mathbf{j}}(\lambda_j) + I(\lambda_j)/\hat{f}_M^{\mathbf{j}}(\lambda_j)\right],$$

which is a likelihood that tends to use only the information around ν as $m \to \infty$. As W has compact support $[-\pi, \pi]$, just about N/m frequencies around ν are used. It is likely that this procedure lead to more variability than the global one, since we are not using all information of the sample (see Brockmann et al. (1993) for a related problem in nonparametric regression).

To justify the above ideas we have the following Proposition, proved in the second appendix of this chapter.

Proposition 4.1 *Under the Assumptions 4.1, 4.2, 4.3, 4.4, W satisfying Assumption 4.3, $M = const \cdot N^{1/5}$ and $m^{-1} + m/M \to 0$,*

$$\begin{aligned} CVLL_m(\nu, M) &= 2\pi \sum_{j=1}^{N-1} W_m(\lambda_j - \nu) [\log f(\lambda_j) + I(\lambda_j)/f(\lambda_j)] \\ &\quad + \frac{N}{2} IMSE_m(\nu, M) + o_P(N \, IMSE_m), \end{aligned}$$

where $0 < c_1 < IMSE_m/IMSE < c_2 < \infty$ as $N \to \infty$, and the first term on the right hand side depends only on m (but not on M).

Then, under regularity conditions, CVLL$_m$ is a consistent estimator of IMSE$_m$ up to a constant not depending on M. From that, minimization of CVLL$_m$ should be approximately equal to minimization of IMSE$_m$. Since the latter approximates MSE(ν) under similar conditions on m, we can expect to obtain reasonable estimates of the local optimum M using the local cross-validation criterion.

BB did not require to estimate explicitly IMSE or its asymptotic rate of convergence, but in our case we need to do so as we estimate a local MSE from an IMSE calculated around a single frequency. To this end, additional stronger conditions are required for the spectrum at that frequency, but we do not need to make global assumptions for the spectral density.

4.5 Monte Carlo work

In this section we try to assess if all the asymptotic arguments given before are good approximations for reasonable finite sample sizes. We have concentrated on the special case of the estimation of the bandwidth for the nonparametric spectral estimate at the origin ($\nu = 0$). From (4.10) we know that for this frequency in particular, IMSE$_m(0, M)$ does not approach MSE$(0, M)$ due to the different variance of the spectral density estimates around the origin. Nevertheless, from Lemma 4.3, the transition from the variance of $\widehat{f}_M(0)$ to the variance of an estimate at a frequency apart from the origin (one half of the previous one) is smooth, depending on the shape of the kernel used. Then we can expect that the approximation behaves approximately well also for this case.

We have used the following equivalent version of the cross-validated log-likelihood, given the periodicity and symmetry of W_m, f and I,

$$\mathrm{CVLL}^\star_m(0,M) \stackrel{def}{=} 2\pi \sum_{j=-[N/2]}^{[N/2]} W_m(\lambda_j)\left[\log \widehat{f}^{\mathbf{j}}_M(\lambda_j) + I(\lambda_j)/\widehat{f}^{\mathbf{j}}_M(\lambda_j)\right],$$

dropping the frequency $j = 0$, (for mean correction purposes and due to the different asymptotic behaviour of $\widehat{f}_M(0)$), and we define $\mathrm{IMSE}^\star_m$ accordingly.

We have calculated both functions $\mathrm{CVLL}^\star_m$ and $\mathrm{IMSE}^\star_m$ by Monte Carlo simulation for Gaussian sequences following five different models and sample size N=256. The models considered are the following AR(3) processes,

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \alpha_3 X_{t-3} + \epsilon_t, \qquad \epsilon_t \rightsquigarrow \mathcal{N}(0,1),$$

with parameters

MODEL 1: $\alpha_1 = 0.7$, $\alpha_2 = -0.5$, $\alpha_3 = 0.4$.

MODEL 2: $\alpha_1 = 0.6$, $\alpha_2 = -0.6$, $\alpha_3 = 0.3$.

MODEL 3: $\alpha_1 = 0.6$, $\alpha_2 = -0.9$.

MODEL 4: $\alpha_1 = 0.8$, $\alpha_2 = -0.6$.

MODEL 5: $\alpha_1 = 0.8$,

and $\alpha_i = 0$ if not stated. These models are convenient because of their simplicity and the different spectra they represent. From Figure 1, Model 1's spectral density exhibits a peak at λ_0 and at λ_{63}, approximately, for this sample size. Model 2 is similar, with a less sharp peak at the origin and other peak at about the same position as before. AR(2) Models 3 and 4 have spectral densities much more flat at the origin, but with very strong peaks at frequencies about λ_{50} and λ_{40}. Finally the AR(1) spectrum of Model 5 shows the typical peak at the origin, taking small values for all the other frequencies.

With these processes we will be able to assess the performance of the approximations in situations where global bandwidths might be not be very appropriate due to the presence of special features in the spectral density at the frequency of interest or at other frequencies, which can distort global procedures.

We employ the Barlett-Priestley Kernel (for both K and W), with spectral window:

$$K(\lambda) = \begin{cases} \frac{3}{4\pi}\left[1 - \left(\frac{\lambda}{\pi}\right)^2\right], & |\lambda| \leq \pi, \\ 0, & |\lambda| \geq \pi. \end{cases}$$

This kernel does not satisfy Assumption 4.4. However this condition is only used in the proof of Lemma 4.5 in expression (4.15). Since the lag-window of the Barlett-Priestley kernel is

$$w(x) = \frac{3}{(\pi x)^2}\left\{\frac{\sin \pi x}{\pi x} - \cos \pi x\right\},$$

we have that (4.15) is $O_P(N^{-1}M^2 \log N)$, since $|w(x)| = O(x^{-2})$ uniformly, and therefore, the remaining results go through without further problems. The uniform kernel was also tried, with much less smooth results as a consequence of the non continuity in the boundaries of its support and a lag-window with tails slow decreasing to zero.

All the experiments are based on 1000 Monte Carlo replications. The tables with the simulation outcomes and the plots are given at the end of the chapter.

4.5.1 Results for IMSEm

The first goal is to check if IMSE_m estimates MSE properly and how sensitive is to the choice of m. Specially interesting are the cases with big values of m for which the IMSE_m are very close to the MSE at the frequency of interest. Due to the problems commented before we cannot expect high precision, but yes certain information about the shape of the spectral density in different intervals around the origin.

To evaluate IMSE_m, we first estimate $\text{MSE}(\lambda_j, M)$ by Monte Carlo replications for all j and a grid of $M = 1(.5)30$, which cover all reasonable M's, including the optimal values for sample size $N = 256$. The optimal values are calculated using the usual pointwise result for the MSE at a single frequency, r^*, depending on the kernel used and on the values of f and its second derivative. Then IMSE_m is evaluated for different values of m and the minimum with respect to M found. The values of m were chosen (see Table I) in terms of the number of different Fourier frequencies around λ_0 over which the kernel W averages in each case, denoted as 'band':

$$\text{band} = \frac{N}{2m}.$$

The correspondent grid is band=1(4)129, which covers all the possibilities for $N = 256$.

The results are reported in Table I and the correspondent plots are in Figures 2 to 6 (in the two-dimensional graphs each horizontal line corresponds with one value of m). From high values of m we can check that the asymptotic expression for the optimal M for $\widehat{f}_M(0)$ it is not very precise for Models 1, 2 and 5. These are the situations with a peak at $f(0)$ and where the bias for the M minimizing IMSE_m is very big and negative. Specially in Models 2, 3 and 4 there is substantial variability between the plots of IMSE_m for different values of m, accounting for the particular features of the spectral densities: when the kernel W averages for the MSE corresponding to frequencies with sharp peaks or troughs in f, the IMSE_m is inflated and the M predicted can be too high for a particular frequency. Then the approximation of MSE by IMSE_m is adequate for moderate values of m.

4.5.2 Results for CVLLm

We next estimate the function $\mathrm{CVLL}_m(0,M)$ for a grid of values of m and M. It is possible to see in the plots (Figures 7 to 11) for all models that the functions CVLL_m depend substantially on the value of m and the minima for different values of m can be far apart. Therefore, CVLL_m reflects the different characteristics of the spectral density and can be a useful means of studying local properties of the spectral density.

In Table II we give the results for the M estimated by the minimization of CVLL_m for a grid of values of m. Again the dependence on m is clear. Concentrating in the rows with m=9.8 and 18.2, we can see that for all the models the mean across replications of the M estimated is not far from the one which minimizes IMSE_m, with reasonable standard deviations. These values of m employ information from only 19 and 13 frequencies around the origin, respectively, so for the spectral densities simulated, the cross-validation uses local enough information and it is not influenced by the characteristics of f at other frequencies.

For example, Model 3 shows a peak at λ_{50} and it is very flat anywhere else. Then, when we weight for all the frequencies in CVLL_m, including the peak, the values of M found are much bigger than the optimal ones, specially when we do not use the frequencies higher than the one corresponding to the peak. However, when CVLL_m includes just the frequencies around the origin, skipping the peak, the M estimated are much smaller and are very close to the optimal value for m big enough. In the case of Model 5 it is precisely at the origin where the spectrum has its main feature. Then, when we give weight to frequencies beyond the low ones, this entails the use of information where the spectrum is very flat, hence too small M are predicted. Consequently the bias is bigger for small values of m as expected.

The variability of the estimates is relatively high, as in most of bandwidth choice methods, characterized by slow rates of convergence. Like in any nonparametric method this variance tends to increase in general with the value of m (which is proportional to the inverse of the actual bandwidth of the kernel W_m).

4.5.3 Two-step procedure

From a theoretical point of view, the choice of m has not a definitive answer. One possibility is a selection criteria depending only on the sample size. In the first three rows of Table III we explore the properties of choices like $m = N^{.1}, N^{.15}, N^{.2}$, which almost agree with the condition required by the asymptotic theory, $m = o(M)$, for $M = c \cdot N^{1/5}$. However, these values of m lead to possibly too many frequencies being used in CVLL_m for this sample size (see the column 'in. band.' or initial bandwidths in Table III), so we complement the experiment with the values $m = 6$ and 10 in the fourth and fifth rows of Table III, corresponding to bandwidths of 21 and 12 frequencies, approximately.

In the last five rows of Table III, we use for m the value of M estimated in the first five rows, in a two-step procedure. In this experiment we use the uniform kernel for W, this choice being of not decisive significance.

As expected, for the first five rows the best results from the bias point of view, correspond to the fifth one, with the largest m (sometimes also using a MSE criterion). For the two-step estimations (rows 6 to 10) the results are not uniform and depend on the concrete model. For Models 1 and 5 with a sharp peak at the origin, there are not big differences between any of the initial choices of m. For Model 2, the first two options give much bigger bias than the last three, and for Models 3 and 4 all give similar results, except the fifth one which shows considerable bias.

In conclusion, the third and fourth choices for the two-step procedure provide the best results in most of the situations, specially from the bias point of view. At the same time that the second step tends to reduce the bias, the variability across replications tends to increase when two steps are used instead of one.

4.5.4 Iterated procedure

Finally we implement two iterative methods to check if any advantage can be gained from the bias or variance point of view when the information from previous estimates is used successively. With each method we employ two initial choices for m_0, 7 and 9.84, corresponding to 'band' numbers of 19 and 13. For the selection of m in each step, Method 1 takes $m_i = 0.9 \cdot M_{i-1}$ and Method 2,

$$m_i = M_{i-1}\left(1 + \frac{M_{i-1} - M_{i-2}}{10(M_{i-1} + M_{i-2})}\right).$$

Both criteria are based fundamentally in the previous values of M found by a one-step cross validation, but the second one also takes into account the sign in the change of the last two estimates obtained. The maximum number of iterations is 5 and the procedures stop if there is change smaller than 0.1. The algorithms check for values of m out a sensible range given the sample size.

The results of the Monte Carlo experiment for the two methods (with the two initial choices of m) and the five AR(3) models are summarized in Table IV. First, there are not significant differences between any of the four combinations and in most cases there is a trade-off between bias and variance. These results are specially interesting in comparison with those of Table III. For Model 1 we have achieved in all cases a certain bias reduction, but at the cost of higher variability. Similar conclusions hold for Model 5. The iterated method for Model 2 produces a bias corresponding to the highest magnitude in Table III, but there is little reduction in the standard deviation across replications. In the case of Model 3, the bias is kept relatively small and there are some little gains in variance reduction. In Model 4 the bias is not much larger than in the Table III, but now the variance has increased.

In general, there are not evident benefits from the iteration of these two algorithms with respect to a simpler two-step scheme.

4.6 Conclusions

In this Chapter we have justified a procedure for local bandwidth choice of nonparametric spectral estimates and shown its performance in finite sample sizes. We have assumed throughout Gaussianity, but this seems not essential, except, perhaps, in the proof for the supremum of the periodogram in Lemma 4.4. However, we conjecture that this condition can be avoided assuming summability conditions on higher order cumulants as in Brillinger (1975), except for the second order ones (autocovariances), imposing then only local conditions on the (second order) spectral density.

A multivariate version on the method will be very useful in practical work, but if we want to stress the specific characteristics of each univariate time series it could be better to apply the method to each of them separately or to a fixed linear combination of the series, like in Newey and West (1994).

Further investigation seems necessary in the design of (iterative) algorithms that, linking m and M, would reduce the variability inherent to bandwidth choice procedures. Then additional finite sample evidence should be found for other models and distributions.

4.7 Appendix: Technical proofs

Proof of Lemma 4.1. This lemma is a restatement of, for example, the Lemma in p. 835 of Hannan and Nicholls (1977), assuming only local conditions on f. As in the proof of Lemma 4.3 we can fix one $\epsilon > 0$ such that, if $I_\nu = [\nu-\epsilon, \nu+\epsilon]$, $\lambda_j, \lambda_k \in I_\nu$ for N big enough. Defining the Dirichlet kernel $\varphi_N(\lambda)$,

$$\varphi_N(\lambda) = \sum_{j=1}^{N} e^{i\lambda j},$$

we have that for $j \neq k$, mod (N),

$$\int_{-\pi}^{\pi} \varphi_N(\lambda_j - \lambda)\varphi_N(\lambda - \lambda_k)d\lambda = 0.$$

Then, if $j \neq k$, mod (N),

$$\mathrm{E}\left[d_x(\lambda_j)\overline{d_x(\lambda_k)}\right] = \int_{-\pi}^{\pi} \varphi_N(\lambda_j - \lambda)\varphi_N(\lambda - \lambda_k)[f(\lambda) - f(\lambda_j)]d\lambda. \qquad (4.11)$$

Now we divide the range of integration in (4.11) in the following intervals. First,

$$\begin{aligned}\left|\int_{\lambda_j - N^{-1}}^{\lambda_j + N^{-1}} \varphi_N(\lambda_j - \lambda)\varphi_N(\lambda - \lambda_k)[f(\lambda) - f(\lambda_j)]d\lambda\right| &\leq \text{const} N \int_{\lambda_j - N^{-1}}^{\lambda_j + N^{-1}} |\lambda - \lambda_j|^{\alpha - 1}d\lambda \\ &\leq \text{const} N^{1-\alpha},\end{aligned}$$

using $\sup_{\lambda \in I_\nu} |f(\lambda) - f(\lambda_j)| \leq \text{const} \cdot |\lambda - \lambda_j|^\alpha$ in the interval considered, and

$$|\varphi_N(\lambda)| \leq \min\left\{2|\lambda|^{-1}, N\right\}.$$

Next,

$$\begin{aligned}&\left|\int_{\lambda_k - N^{-1}}^{\lambda_k + N^{-1}} \varphi_N(\lambda_j - \lambda)\varphi_N(\lambda - \lambda_k)[f(\lambda) - f(\lambda_j)]d\lambda\right| \\ &\leq \text{const} \cdot N^{-1} \sup_{|\lambda - \lambda_k| \leq N^{-1}} |\varphi(\lambda - \lambda_k)| \sup_{|\lambda - \lambda_k| \leq N^{-1}} |\lambda - \lambda_j|^{\alpha - 1} \leq \text{const} \cdot N^{1-\alpha},\end{aligned}$$

since the range of integration was of order N^{-1}. Define the set $I_\nu(k,j)$ as the interval I_ν substracting the previous two neighbourhoods of radius N^{-1} around λ_k and λ_j. Then

$$\begin{aligned}&\left|\int_{I_\nu(k,j)} \varphi_N(\lambda_j - \lambda)\varphi_N(\lambda - \lambda_k)[f(\lambda) - f(\lambda_j)]d\lambda\right| \\ &\leq \text{const} \sup_{I_\nu(k,j)} |\lambda - \lambda_j|^{\alpha - 1} \int_{-\pi}^{\pi} |\varphi_N(\lambda - \lambda_k)|d\lambda \\ &\leq \text{const} \cdot N^{1-\alpha} \log N,\end{aligned}$$

using $\int_{-\pi}^{\pi} |\varphi_N(\lambda)|d\lambda \leq \text{const} \cdot \log N$. Finally in the complementary set of I_ν,

$$\begin{aligned}&\left|\int_{I_\nu^c} \varphi_N(\lambda_j - \lambda)\varphi_N(\lambda - \lambda_k)[f(\lambda) - f(\lambda_j)]d\lambda\right| \\ &\leq \text{const} \sup_{I_\nu^c} |\varphi_N(\lambda_j - \lambda)\varphi_N(\lambda - \lambda_k)| \left[f(\lambda_j) + \int_{-\pi}^{\pi} f(\lambda)d\lambda\right] \leq \text{const},\end{aligned}$$

and the lemma follows in the case $j \neq k$ because any of the bounds depends on j or k. If $j = k$ then we can use the same methods as before together with

$$\int_{-\pi}^{\pi} |\varphi_N(\lambda)|^2 d\lambda = 2\pi N,$$

to get the desired result. □

Proof of Lemma 4.2. The proof it is immediate in the light of the Proposition in page 31 of BB and our Lemma 4.1, as by the Gaussianity of X_t only cumulants of order two of the discrete Fourier transform of X_t have to be considered.

Here the bound in (4.3) is only $O(N^{-\alpha}\log N)$ and not this bound to the power of q as in BB. The problem with their proof is the following. At the beginning of their page 33, for $k \in \nu_2$ in their notation, $\text{cum}\{d_x(\lambda_{k1}), d_x(\lambda_{k2})\}N^{-1} = O(1)$ at most, because we can have $\lambda_{k1} = \lambda_{k2}$ for all elements in one of the possible partitions. Then, the second bound in the third full paragraph formula of the same page is only $O(1)$ and the first one is $O(N^{-\alpha}\log N)$ (actually $O(N^{-1})$ under their conditions), since we have $\#\nu_1 \leq 1$. □

We give now some lemmas needed for the proof of Proposition 4.1.

Lemma 4.4 *Under Assumptions 4.1, if f satisfies a uniform Lipschitz condition of order $0 < \alpha \leq 1$, in an interval around a fixed frequency ν, $I_\nu = [\nu - \epsilon, \nu + \epsilon]$ for some $\epsilon > 0$, then for frequencies $\lambda_j = 2\pi j/N$, $j \neq 0$ such that $\sup_{\lambda_j} |\nu - \lambda_j| \leq const \cdot m^{-1}$, for some positive sequence m such that $1/m + m/N \to 0$, uniformly in $j \neq 0$,*

$$\overline{\lim_{N\to\infty}} \sup_{\lambda_j} I(\lambda_j) \leq 2\log N \sup_{\lambda\in I_\nu} f(\lambda) \quad \textit{w.p.1.}$$

Proof. We can proceed as in the proof of Theorems 4.5.1 and 5.3.2 of Brillinger (1975), taking the mean of X_t as zero, since we do not include the zero frequency. In our case, since X_t is a Gaussian series and $j \neq 0$, all the cumulants of order bigger than two are zero. From Lemma 4.1 we can obtain, uniformly in j, for m big enough,

$$\sigma_N \equiv \text{Var}[\text{Re}\, d_x(\lambda_j)] = \frac{N}{2} 2\pi f(\lambda_j) + O(N^{1-\alpha}\log N).$$

Then, for $\lambda_j \in I_\nu$ and any θ and one ϵ as small as we want, from Gaussianity, as $N \to \infty$,

$$\text{E}\left[\exp\left\{\theta \text{Re}\, d_x(\lambda_j)\right\}\right] \leq \exp\left\{\theta^2 2\pi N f(\lambda_j)(1+\epsilon)/4\right\}.$$

Next,

$$\text{E}\exp\left\{\theta \sup_{\lambda_j} |\text{Re}\, d_x(\lambda_j)|\right\} \quad \leq \quad \sum_{\lambda_j \in I_\nu} \text{E}\exp\left\{\theta\, |\text{Re}\, d_x(\lambda_j)|\right\}$$

$$
\begin{aligned}
&\leq \sum_{\lambda_j \in I_\nu} \exp\left\{\theta^2 2\pi N f(\lambda_j)(1+\epsilon)/4\right\} \\
&\leq 2\exp\left\{\log N + \theta^2 2\pi N \sup_{\lambda \in I_\nu} f(\lambda)(1+\epsilon)/4\right\}.
\end{aligned}
$$

Now define, for $\delta > 0$

$$
a^2 = 2\pi(1+\epsilon)(2+\delta)N\log N \sup_{\lambda \in I_\nu} f(\lambda).
$$

Then

$$
\operatorname{Prob}\left\{\sup_{\lambda_j} |\operatorname{Re} d_x(\lambda_j)| \geq a\right\} \leq \exp\{-\theta a\} 2\exp\left\{\log N + \theta^2 2\pi N \sup_{\lambda \in I_\nu} f(\lambda)(1+\epsilon)/2\right\}.
$$

Taking

$$
\theta = a\left[2\pi N(1+\epsilon)\sup_{\lambda \in I_\nu} f(\lambda)\right]^{-1},
$$

this is less or equal than

$$
2\exp\left\{-a^2\left[2\pi N \sup_{\lambda \in I_\nu} f(\lambda)(1+\epsilon)\right]^{-1}\right\}\exp\{\log N\} \leq \text{const} \cdot N^{-1-\delta}.
$$

Using this last line and the Borel-Cantelli Lemma, as ϵ and δ were arbitrary, we obtain that

$$
\overline{\lim_{N\to\infty}} \sup_{\lambda_j} |\operatorname{Re} d_x(\lambda_j)| / [2\pi N \log N]^{1/2} \leq \left[\sup_{\lambda \in I_\nu} f(\lambda)\right]^{1/2} \quad \text{w.p.1.}
$$

A similar result is possible for the imaginary part of d_x and then the lemma follows from

$$
|d_x(\lambda_j)| \leq |\operatorname{Re} d_x(\lambda_j)| + |\operatorname{Im} d_x(\lambda_j)|
$$

and $I(\lambda_j) = (2\pi N)^{-1} |d_x(\lambda_j)|^2$. □

Lemma 4.5 *Under Assumptions 4.1, 4.2, 4.3, 4.4, for frequencies $\lambda_j = 2\pi j/N$ such that $\sup_{\lambda_j} |\nu - \lambda_j| \leq const \cdot m^{-1}$, for some positive sequence m such that $1/m + m/N \to 0$, uniformly in j,*

$$
\sup_{\lambda_j}\left|\frac{\widehat{f}_M(\lambda_j) - f(\lambda_j)}{f(\lambda_j)}\right| = O_P\left(N^{-1}M^2 + N^{\frac{1-p}{2p}}M + N^{-1}\log N + M^{-1}\right) = o_P(1).
$$

Proof. Define the weighted autocovariance spectral estimate corresponding to the continuous average in $\widehat{f}_M$, when the mean of X_t is known, (and assumed to be 0 without loss of generality),

$$
\widehat{f}_M^C(\lambda_j) = \int_{-\pi}^{\pi} K_M(\lambda) I(\lambda_j + \lambda) d\lambda = \frac{1}{2\pi}\sum_{r=1-N}^{N-1} w\left(\frac{r}{M}\right)\widehat{\gamma}(r)\cos r\lambda_j,
$$

where $K_M(\cdot) = MK(M\cdot)$ periodically extended and

$$\widehat{\gamma}(k) = N^{-1} \sum_{1 \le t, t+k \le N} X_t X_{t+k}.$$

This estimate is unfeasible if the mean of the series is unknown, but we only need its definition for the proofs. Now we have, proceeding as in the proof of Theorem 2.1 of Robinson (1991),

$$\begin{aligned}
\sup_{\lambda_j} \left| \widehat{f}_M(\lambda_j) - f(\lambda_j) \right| \;&\le\; \sup_{\lambda_j} \left| \widehat{f}_M(\lambda_j) - \widehat{f}^C_M(\lambda_j) \right| && (4.12)\\
&\quad + \sup_{\lambda_j} \left| \widehat{f}^C_M(\lambda_j) - E\left[\widehat{f}^C_M(\lambda_j)\right] \right| && (4.13)\\
&\quad + \sup_{\lambda_j} \left| E\left[\widehat{f}^C_M(\lambda_j)\right] - f(\lambda_j) \right|. && (4.14)
\end{aligned}$$

Now (4.12) is less or equal than (see Robinson, 1991, p. 1353),

$$\begin{aligned}
(2\pi)^{-1} \sum_{1-N}^{N-1} \left| w\left(\frac{r}{M}\right) \right| |\widehat{\gamma}(N-r)| \;&=\; O_P\left(N^{-1} \sum_{1-N}^{N-1} \left| w\left(\frac{r}{M}\right) \right| |r| \right) && (4.15)\\
&=\; O_P\left(N^{-1} M^2 \right),
\end{aligned}$$

using Assumption 4.4 and the fact that $\widehat{\gamma}(N-r)$ is a sum of r terms whose mean exists and is uniformly bounded. Next (4.13) is not bigger than

$$\begin{aligned}
(2\pi)^{-1} \sum_{1-N}^{N-1} \left| w\left(\frac{r}{M}\right) \right| |\widehat{\gamma}(r) - E[\widehat{\gamma}(r)]| \;&=\; O_P\left(\sum_{1-N}^{N-1} \left| w\left(\frac{r}{M}\right) \right| N^{\frac{1-p}{2p}} \right)\\
&=\; O_P\left(N^{\frac{1-p}{2p}} M \right) = o_P(1),
\end{aligned}$$

because Assumptions 4.2 and 4.4, and Lemma 4.7 bellow. Finally (4.14) is bounded by

$$\begin{aligned}
&\sup_{\lambda_j} \left| \int_{-\pi}^{\pi} K_M(\lambda_j - \omega) \left\{ E[I(\omega) - f(\omega)] \right\} d\omega \right| && (4.16)\\
&\quad + \sup_{\lambda_j} \left| \int_{\mathcal{R}} K(\omega) \left\{ f(\lambda_j - \omega/M) - f(\omega) \right\} d\omega \right|. && (4.17)
\end{aligned}$$

Denote by Φ_N Fejér Kernel $\Phi_N(\lambda) = (2\pi N)^{-1} |\varphi_N(\lambda)|^2$. Similarly to Lemma 4.1, we have that in (4.16) ω lies in the interior of I_ν as $M \to \infty$ due to the compact support of K, and for fixed $\delta > 0$ small enough,

$$\sup_{\omega \in I_\nu} |E[I(\omega)] - f(\omega)| \;\le\; \sup_{\omega \in I_\nu} \left| \int_{-\pi}^{\pi} \Phi_N(\alpha - \omega)[f(\alpha) - f(\omega)] d\alpha \right|$$

$$
\begin{aligned}
&\le \sup_{\omega\in I_\nu} |f'(\omega)| \int_{|\omega-\alpha|\le\delta} |\Phi_N(\alpha-\omega)||\alpha-\omega| d\alpha \\
&\quad + \sup_{\omega\in I_\nu} \int_{|\omega-\alpha|>\delta} |\Phi_N(\alpha-\omega)|[f(\alpha)+f(\omega)] d\alpha \\
&= O(N^{-1}\log N) + O(N^{-1}) \\
&= O(N^{-1}\log N),
\end{aligned}
$$

uniformly in $\omega \in I_\nu$, so (4.16) is $O(N^{-1}\log N)$, since $\int |K_M(\alpha)| d\alpha < \infty$. Next, as $M \to \infty$, (4.17) is bounded by (denoting by λ^* a value between λ_j and $\lambda_j - \omega/M$),

$$
\sup_{\lambda_j} \int K(\omega)\,[f(\lambda_j - \omega/M) - f(\lambda_j)]\,d\omega \le \sup_{\lambda_j} \int |K(\omega)||f'(\lambda^*)| \left|\frac{\omega}{M}\right| d\omega = O(M^{-1}),
$$

using the compact support of K and that of f' is bounded in I_ν. □

Lemma 4.6 *Under the Assumptions of Lemma 4.5, uniformly in j,*

$$
\sup_{\lambda_j} \left| \frac{\widehat{f}^{\mathbf{j}}_M(\lambda_j) - f(\lambda_j)}{f(\lambda_j)} \right| = O_P\left(N^{-1}M^2 + N^{\frac{1-p}{2p}} M + N^{-1}\log N + M^{-1}\right).
$$

Proof. The proof is exactly the same as how is done in Lemma 4 of BB, using now our Lemma 4.5. □

Lemma 4.7 *Under Assumptions 4.1 and 4.2, uniformly in r, $p > 1$,*

$$
Var[\widehat{\gamma}(r)] = O\left(N^{\frac{1-p}{p}}\right).
$$

where $\widehat{\gamma}(r)$ is the (biased) estimate of the lag-r autocovariance $\gamma(r)$ when the expectation of X_t is known,

$$
\widehat{\gamma}(r) = \frac{1}{N} \sum_{1\le t, t+r \le N} (X_t - E[X_1])(X_{t+r} - E[X_1]).
$$

Proof. From e.g. Anderson (1971, p.452), denoting as before the Fejér kernel by Φ_N,

$$
N\,\mathrm{Var}\,[\widehat{\gamma}(r)] = \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \Phi_N(\alpha-\omega)(1+e^{-i(\alpha+\omega)r}) f(\alpha) f(\omega) d\alpha d\omega,
$$

and now the result follows applying Hölder inequality twice, with $|\Phi_N(\omega)| = O(N)$ uniformly in ω, $|1 + e^{-i(\alpha+\omega)r}| \le 2$ uniformly in r and $\int_{-\pi}^{\pi} f^p < \infty$ by Assumption 4.2. □

4.8 Appendix: Proof of Proposition 4.1

From the proof of Theorem 3.1 in BB the Lemma will follow, using their definitions, if we show

$$\begin{aligned} N^{-1}T_i &= o_P(\mathrm{IMSE}_m) \qquad i=1,2 \\ N^{-1}T_3 &= \mathrm{IMSE}_m + o_P(\mathrm{IMSE}_m)\,. \end{aligned}$$

First we have, denoting now $\sigma_j = \sigma_{j,M}$, from the last steps in the proofs of BB,

$$E[T_1] = 2\pi \sum_{j=1}^{N} W_m(\lambda_j - \nu)\sigma_j^{-1} {\sum_k}' K(M\lambda_k) O(N^{-1}) = O(1),$$

and, denoting as $\mathrm{IMSE}'_m(\nu, M)$ the IMSE_m calculated from the modified spectral estimate (4.1),

$$\begin{aligned} E[T_1^2] &= N\,\mathrm{MISE}'_m + 2\pi \sum_j W_m(\lambda_j-\nu)^2 \sigma_j^{-2} {\sum_k}'{\sum_n}' K(M\lambda_k)K(M\lambda_n)O(N^{-1}) \\ &\quad + 2\pi \sum_j \sum_{j\neq i} W_m(\lambda_j-\nu)W_m(\lambda_i-\nu)\sigma_j^{-1}\sigma_i^{-1} {\sum_k}'{\sum_n}' K(M\lambda_k)K(M\lambda_n)O(N^{-2}) \\ &\quad + 2\pi \sum_j W_m(\lambda_j-\nu)^2\sigma_j^{-2} {\sum_k}'{\sum_n}' K(M\lambda_k)^2 \\ &= N\,\mathrm{IMSE}'_m + O(m) + O(m) + O(m\,N\,\mathrm{IMSE}_m) \\ &= O(m\,N\,\mathrm{IMSE}_m), \end{aligned}$$

since $\sup_{\lambda,m} |W_m(\lambda)| = O(m)$. Then using $\mathrm{IMSE}_m = O(M/N)$ we can obtain

$$T_1 = O_P(\mathrm{IMSE}_m\,[m/M]^{1/2}) = o_P(\mathrm{IMSE}_m),$$

because $m/M \to 0$.

Now, in a similar fashion,

$$E[T_2] = 2\pi \sum_{j=1}^{N} W_m(\lambda_j - \nu)\sigma_j^{-2} {\sum_k}'{\sum_n}' K(M\lambda_k)K(M\lambda_n)O(N^{-1}) = O(1),$$

and as before

$$E[T_2] = O(m\,N\,\mathrm{IMSE}_m^2).$$

(Note that in BB's expression they have N^{-1} instead of N in the correspondent formula, although in the statements in the main part of their paper they give the right bounds). Then $N^{-1}T_2 = O_P(\mathrm{IMSE}_m\,[m/N]^{1/2}) = o_P(\mathrm{IMSE}_m)$.

Next,

$$
\begin{aligned}
E[T_3] &= N\,\mathrm{IMSE}_m + O([N/M]^{-1}\,N\,\mathrm{IMSE}_m) + O([N/M]^{-1}) + O([N/M]^{-1}\,M) \\
&= N\,\mathrm{IMSE}_m + O([N/M]^{-1}\,N\,\mathrm{IMSE}_m),
\end{aligned}
$$

and reasoning in the same way as before,

$$\mathrm{Var}[T_3] = O(m\,N\,\mathrm{IMSE}_m).$$

Then $N^{-1}T_3 = \mathrm{IMSE}_m + O_P(\mathrm{IMSE}_m[m/M]^{1/2}) = \mathrm{IMSE}_m + o_P(\mathrm{IMSE}_m)$.

The proof for the remainder term in BB's expression (3.2) continues the same here, using now our Lemmas 4.1, 4.4 and 4.5 instead of their references, since the bound for the third term in the expansion still holds for the modified (local) cross-validation.

TABLE I M minimizing IMSEm estimated by Monte Carlo

SAMPLE SIZE: 256

REPLICATIONS: 1000

AR(3)		MODEL 1		MODEL 2		MODEL 3		MODEL 4		MODEL 5	
Coeff.=		0.70 -0.50 0.40		0.60 -0.60 0.30		0.60 -0.90 .00		0.80 -0.60 .00		0.80 .00 .00	
M*=		15.9		7.2		7.1		7.7		20.3	
m	*band*	*M*	*bias*	*M*	*bias*	*M*	*bias*	*M*	*bias*	*M*	*bias*
1	129	9.00	-6.9	9.00	1.84	24.00	16.9	9.50	1.83	9.00	-11.34
1.024	125	9.00	-6.9	9.00	1.84	24.50	17.4	10.00	2.33	10.00	-10.34
1.058	121	9.00	-6.9	9.00	1.84	24.50	17.4	10.00	2.33	10.00	-10.34
1.094	117	9.00	-6.9	9.00	1.84	24.50	17.4	10.00	2.33	10.00	-10.34
1.133	113	9.00	-6.9	9.00	1.84	25.00	17.9	10.00	2.33	10.00	-10.34
1.174	109	9.00	-6.9	9.00	1.84	25.00	17.9	10.00	2.33	10.00	-10.34
1.219	105	9.00	-6.9	9.00	1.84	27.50	20.4	10.00	2.33	10.00	-10.34
1.267	101	9.00	-6.9	9.00	1.84	27.50	20.4	10.00	2.33	10.00	-10.34
1.32	97	9.00	-6.9	9.00	1.84	27.50	20.4	10.50	2.83	10.00	-10.34
1.376	93	9.00	-6.9	9.00	1.84	27.50	20.4	10.50	2.83	10.00	-10.34
1.438	89	9.00	-6.9	8.50	1.34	27.50	20.4	10.50	2.83	10.00	-10.34
1.506	85	9.00	-6.9	8.50	1.34	28.00	20.9	10.50	2.83	10.50	-9.84
1.58	81	8.50	-7.4	8.00	0.84	28.00	20.9	10.50	2.83	10.50	-9.84
1.662	77	8.50	-7.4	7.00	-0.16	28.00	20.9	10.50	2.83	10.50	-9.84
1.753	73	8.50	-7.4	7.00	-0.16	28.50	21.4	10.50	2.83	10.50	-9.84
1.855	69	8.50	-7.4	7.00	-0.16	28.50	21.4	10.50	2.83	10.50	-9.84
1.969	65	8.50	-7.4	7.00	-0.16	28.50	21.4	10.50	2.83	11.00	-9.34
2.098	61	8.50	-7.4	6.00	-1.16	24.00	16.9	9.50	1.83	11.50	-8.84
2.246	57	8.50	-7.4	6.00	-1.16	24.00	16.9	9.50	1.83	11.50	-8.84
2.415	53	9.00	-6.9	6.00	-1.16	23.00	15.9	9.00	1.33	11.50	-8.84
2.612	49	9.00	-6.9	6.00	-1.16	23.00	15.9	9.00	1.33	11.50	-8.84
2.844	45	9.00	-6.9	6.00	-1.16	22.50	15.4	9.50	1.83	12.00	-8.34
3.122	41	9.00	-6.9	5.50	-1.66	15.00	7.87	9.00	1.33	12.00	-8.34
3.459	37	9.00	-6.9	5.50	-1.66	13.50	6.37	9.00	1.33	12.00	-8.34
3.879	33	9.00	-6.9	5.00	-2.16	12.00	4.87	9.00	1.33	12.50	-7.84
4.414	29	9.00	-6.9	4.50	-2.66	10.00	2.87	9.00	1.33	12.50	-7.84
5.12	25	9.00	-6.9	3.50	-3.66	9.00	1.87	8.50	0.83	12.50	-7.84
6.095	21	8.50	-7.4	2.00	-5.16	8.50	1.37	8.00	0.33	12.00	-8.34
7.529	17	8.00	-7.9	2.00	-5.16	8.00	0.87	8.00	0.33	11.50	-8.84
9.846	13	8.00	-7.9	2.00	-5.16	7.50	0.37	7.50	-0.17	10.50	-9.84
14.22	9	8.50	-7.4	2.00	-5.16	7.00	-0.13	7.00	-0.67	10.50	-9.84
25.6	5	8.50	-7.4	2.00	-5.16	6.50	-0.63	6.50	-1.17	11.00	-9.34
128	1	9.00	-6.9	2.00	-5.16	6.50	-0.63	6.00	-1.67	11.00	-9.34

TABLE II **M minimizing CVLLm**

SAMPLE SIZE: 256
REPLICATIONS: 1000

AR(3)		MODEL 1			MODEL 2			MODEL 3			MODEL 4			MODEL 5		
Coeff. =		-0.70	0.50	0.40	-0.60	0.60	0.30	0.60	-0.90	00	-0.80	0.60	0.0	0.80	00 .00	0.00
M optimal:		15.9044			7.1582			7.1313			7.6765			20.3404		
m	*band*	*bias*	*sd*	*mse*	*bias*	*sd*	*mse*	*bias*	*sd*	*mse*	*bias*	*sd*	*mse*	*bias*	*sd*	*mse*
1.1130	115	-5.7347	4.2256	50.7429	3.4179	4.2028	29.3461	9.5363	4.5366	111.5222	1.4702	3.4220	13.8718	-10.9760	4.0756	137.0838
1.1743	109	-5.6041	4.2929	49.8351	3.4658	4.2195	29.8161	9.7627	4.5883	116.3635	1.5720	3.5742	15.2457	-10.8475	4.2369	135.6194
1.2427	103	-5.5431	4.2450	48.7461	3.4853	4.1938	29.7353	9.9614	4.5717	120.1307	1.9109	3.9794	19.4871	-10.7431	4.3208	134.0830
1.3196	97	-5.5121	4.3242	49.0821	3.5214	4.2704	30.6369	10.1111	4.6696	124.0395	2.0393	5.3365	32.6367	-10.7395	4.2656	133.5331
1.4066	91	-5.4037	4.4323	48.8454	3.6778	4.3772	32.6856	10.5429	4.8515	134.6894	2.0031	5.0676	29.6935	-10.5524	4.3692	130.4432
1.5059	85	-5.3072	4.4142	47.6524	3.8019	4.3616	33.4782	11.0105	5.0425	146.6572	2.1934	5.0871	30.6892	-10.4085	4.4552	128.1851
1.6203	79	-5.2542	4.4281	47.2154	3.8310	4.3600	33.6859	11.5546	5.5167	163.9421	2.1636	5.1391	31.0910	-10.3393	4.2701	125.1359
1.7534	73	-5.2212	4.3238	45.9555	3.9054	4.4545	35.0946	11.8803	5.5563	172.0136	2.5442	5.2694	34.2399	-10.2190	4.3819	123.6284
1.9104	67	-5.2831	4.5622	48.7246	3.9159	4.6323	36.7926	12.3251	5.6733	184.0950	2.6846	5.3070	35.3714	-10.0506	4.5354	121.5846
2.0984	61	-5.5326	4.7588	53.2557	3.6130	5.0502	38.5588	12.8032	6.1161	201.3294	2.6156	3.4356	18.6445	-9.9045	4.5193	118.5226
2.3273	55	-5.6068	5.0999	57.4451	3.0805	5.6247	41.1269	13.0760	6.5479	213.8564	2.7709	5.2225	34.9527	-9.7337	4.6224	116.1121
2.6122	49	-5.4980	5.3163	58.4910	2.8014	6.1766	45.9976	10.8304	5.9349	152.5214	2.8512	3.4704	20.1731	-9.4131	4.9357	112.9684
2.9767	43	-5.3557	5.4418	58.2968	2.8498	6.1333	45.7392	9.9504	5.3064	127.1676	2.8067	3.9345	23.3579	-9.1663	5.0796	109.8234
3.4595	37	-5.1626	5.6488	58.5615	2.9835	6.2250	47.6519	6.2626	4.6367	60.7188	1.5185	4.3205	20.9724	-8.6592	5.5261	105.5207
4.1290	31	-5.0566	5.9339	60.7806	2.9191	6.2713	47.8507	4.3187	4.4768	38.6931	1.6265	5.2705	30.4240	-8.5783	5.5709	104.6233
5.1200	25	-4.4755	6.4985	62.2608	3.0200	6.8305	55.7757	2.9687	5.1594	35.4326	1.9333	4.5559	24.4937	-7.8403	6.1015	98.6986
6.7368	19	-6.3533	5.4734	70.3223	-3.5810	3.4400	24.6574	0.8741	4.1158	17.7040	.045	3.3079	10.9443	-8.2622	6.0502	104.8678
9.8462	13	-5.1687	6.2085	65.2603	-2.1784	3.8172	19.3167	-0.1370	4.0081	16.0833	-0.8112	3.0549	9.9903	-6.6447	7.1031	94.6061
18.2857	7	-5.3244	7.0861	78.5618	0.3782	4.6491	21.7571	.0613	5.5129	30.3962	-0.8962	3.2507	11.3703	-6.8540	8.2943	115.7717
128.0	1	0.8368	10.2568	105.9012	9.1475	10.3880	191.5881	4.0041	9.4054	104.4943	4.1955	7.1068	68.1096	-3.9746	9.7834	111.5118

TABLE III. M minimizing CVLLm (One or two steps)
SAMPLE SIZE 256
REPLICATIONS: 1000

AR(3)		MODEL 1			MODEL 2			MODEL 3			MODEL 4			MODEL 5		
Coeff. =		.70	-0.5	0.40	.30	-0.6	0.30	.60	-0.9	.00	.80	-0.6	.00	.80	.00	.00
M optimal =		15.9044			5.0027			7.1313			7.6765			20.34		
	in. band.	*bias*	*sd*	*mse*	*bias*	*sd*	*mse*	*bias*	*sd*	*mse*	*bias*	*sd*	*mse*	*bias*	*sd*	*mse*
1	73.5167	-7.4608	3.0699	65.0875	5.0688	3.6586	39.0782	11.8405	5.5821	171.3574	1.5077	3.3717	13.6417	-12.2643	3.1808	160.5296
2	55.7152	-8.3296	3.8143	83.9314	3.4675	3.7525	26.1044	12.6728	6.1371	198.2639	2.2025	4.0805	21.5012	-11.7260	3.5404	150.0345
3	42.2243	-8.0190	4.2081	82.0128	2.3919	3.7072	19.4646	9.9881	5.6038	131.1655	1.5955	4.2409	20.5312	-11.0842	4.0881	139.5715
4	21.3333	-7.0055	5.0544	74.6239	-0.6236	3.3794	11.8092	1.5389	4.2894	20.7674	1.0255	4.4797	21.1194	-8.8957	5.6586	111.1531
5	12.8000	-6.4002	5.4968	71.1769	-0.5553	3.4884	12.4769	-0.0177	4.0179	16.1435	-0.3503	4.4327	19.7713	-8.2941	5.9674	104.4029
6		-6.1724	6.0389	74.5669	-0.2459	4.0659	16.5917	0.1169	5.3848	29.0093	-0.1250	4.2616	18.1765	-8.1885	6.5430	109.8613
7		-6.2867	5.9513	74.9413	-0.1840	4.0064	16.0851	0.2037	5.6205	31.6316	-0.1065	4.5131	20.3791	-8.0138	6.5751	107.4537
8		-6.2920	5.8987	74.3840	-0.2820	3.6636	13.5017	0.1864	5.4737	29.9963	-0.1141	4.4898	20.1717	-8.0135	6.8935	111.7364
9		-6.3485	5.9847	76.1195	1.7977	3.6270	16.3865	0.3649	4.0747	16.7361	-0.0180	4.4489	19.7931	-7.6091	7.5891	115.4934
10		-6.2322	6.1396	76.5342	2.2112	4.3895	24.1574	2.6026	4.9888	31.6618	1.1668	4.9722	26.0845	-7.7459	7.4518	115.5277

Table IV. M minimizing CVLLm Iterating
SAMPLE SIZE: 256
REPLICATIONS: 1000

Method		*bias*	*var*	*mse*	*bias*	*var*	*mse*	*bias*	*var*	*mse*	*bias*	*var*	*mse*	*bias*	*var*	*mse*
initial=7	1	-5.8181	6.1809	72.0534	-3.6862	3.4772	25.6786	0.8595	4.169	18.1191	0.6705	4.943	24.8827	-7.1909	6.6827	96.3673
	2	-5.4958	6.6031	73.8048	-3.4473	3.8295	26.5493	0.8615	4.2236	18.5814	0.6458	5.2504	27.9837	-6.8425	6.9557	95.2017
initial=9.84	1	-5.4299	6.5005	71.7397	-2.9269	3.9415	24.1024	1.0322	4.4533	20.8968	0.805	4.9801	25.449	-6.9862	6.7258	94.0435
	2	-5.4216	6.3673	69.9362	-2.6903	4.2291	25.1228	1.0801	4.562	21.9785	0.9054	5.5208	31.2992	-6.796	6.8886	93.6373

FIGURE 1. Model 1. Spectral density AR(3) .7 / −.5 / .4

Model 2. Spectral density AR(3) .6 / −.6 / .3

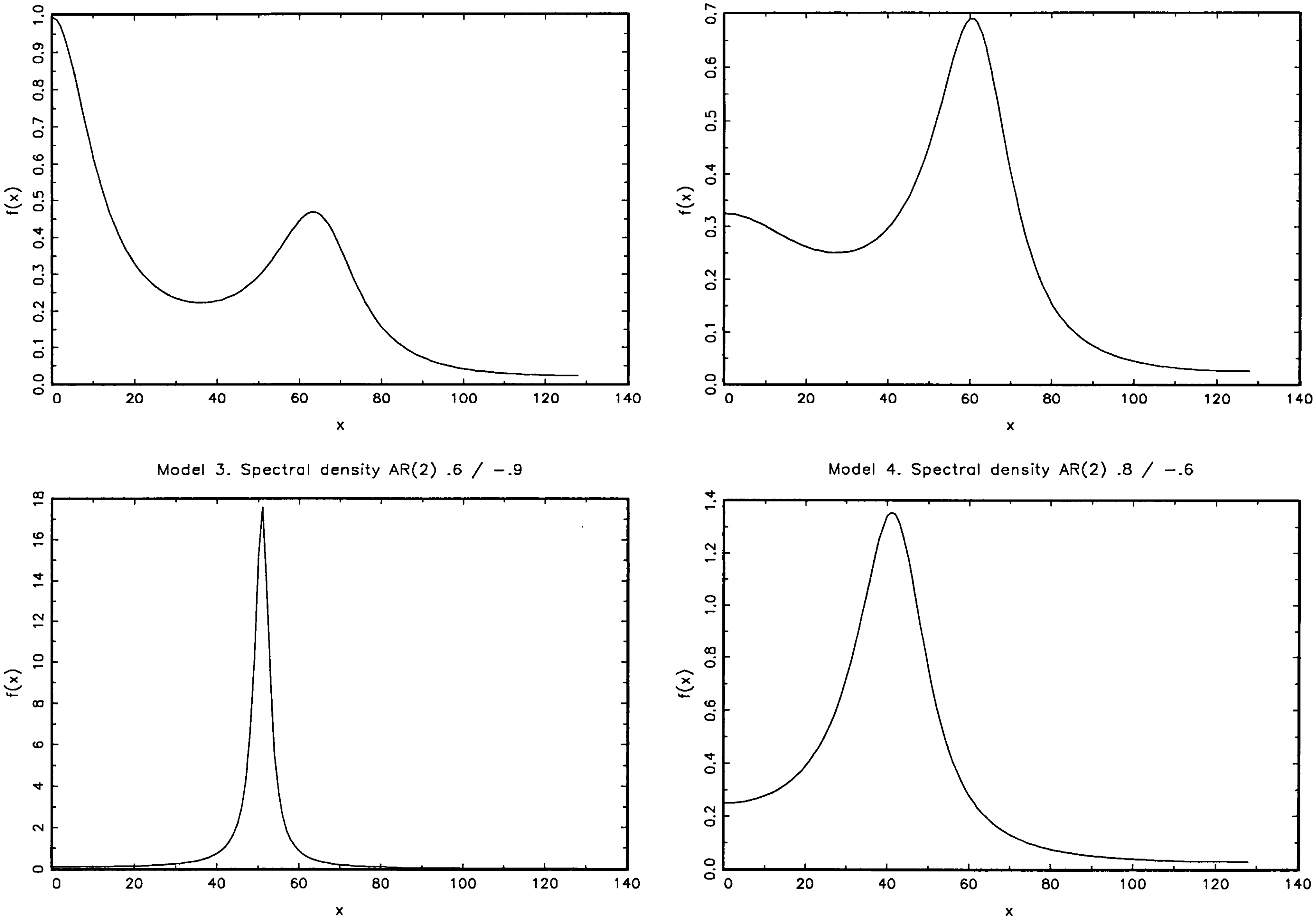
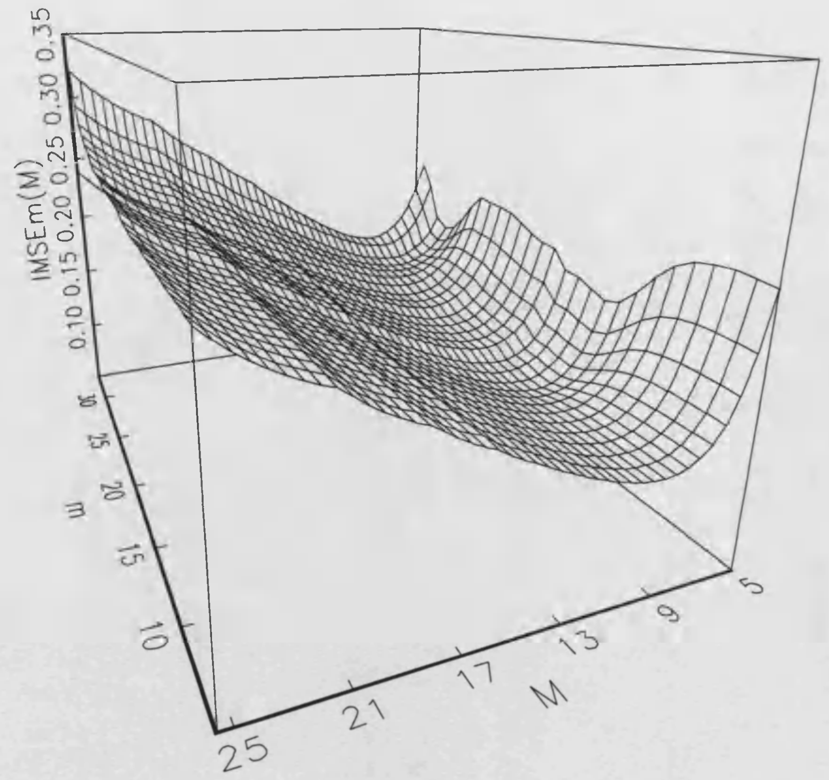


FIGURE 2. Model 1. Estimated IMSEm AR(3) .7 / −.5 / .4



Model 1. Estimated IMSEm AR(3) .7 / −.5 / .4

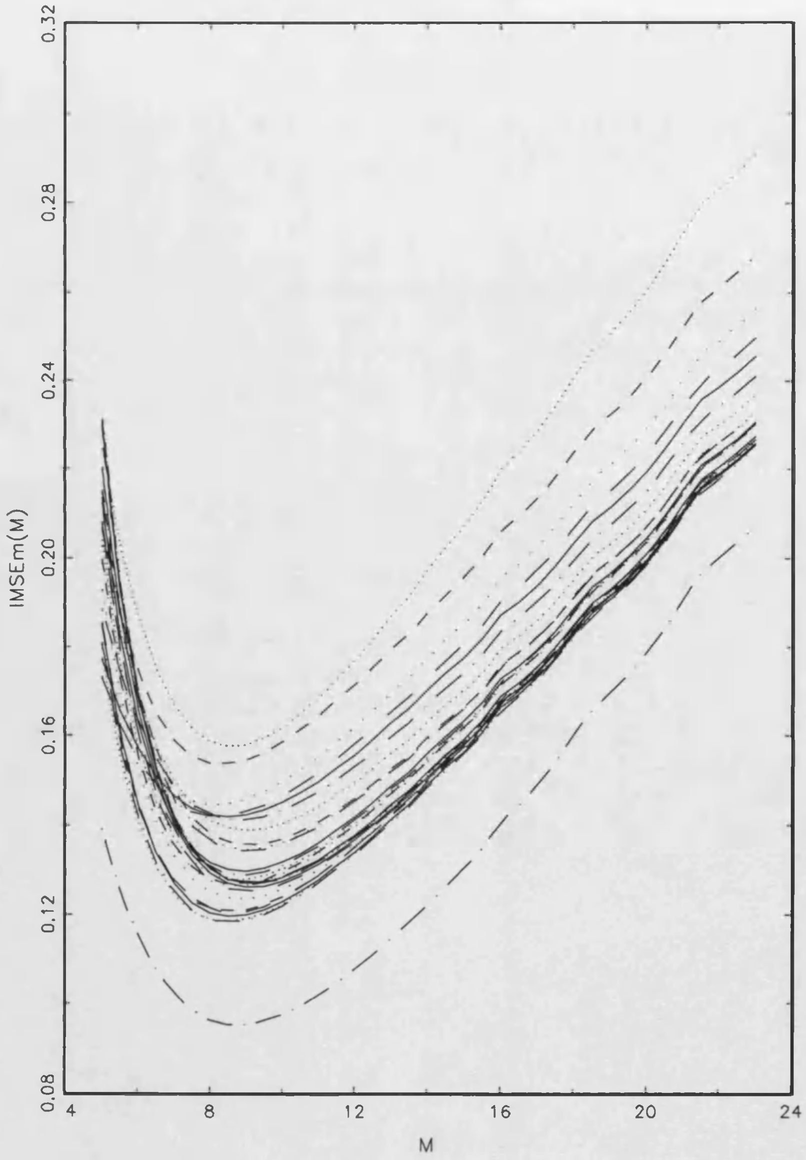
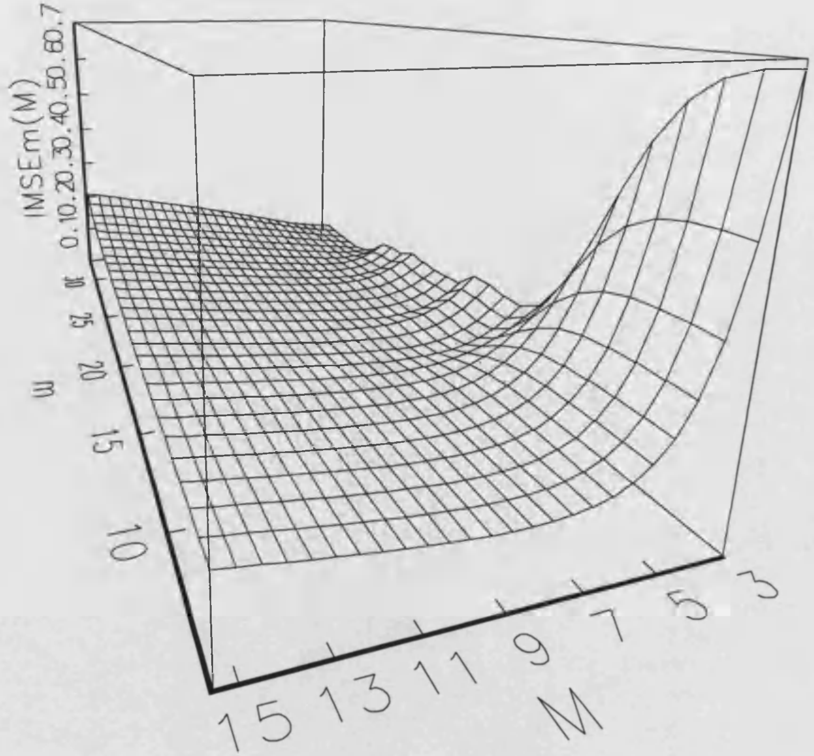


FIGURE 3. Model 2. Estimated IMSEm AR(3) .6 / −.6 / .3



Model 2. Estimated IMSEm AR(3) .6 / −.6 / .3

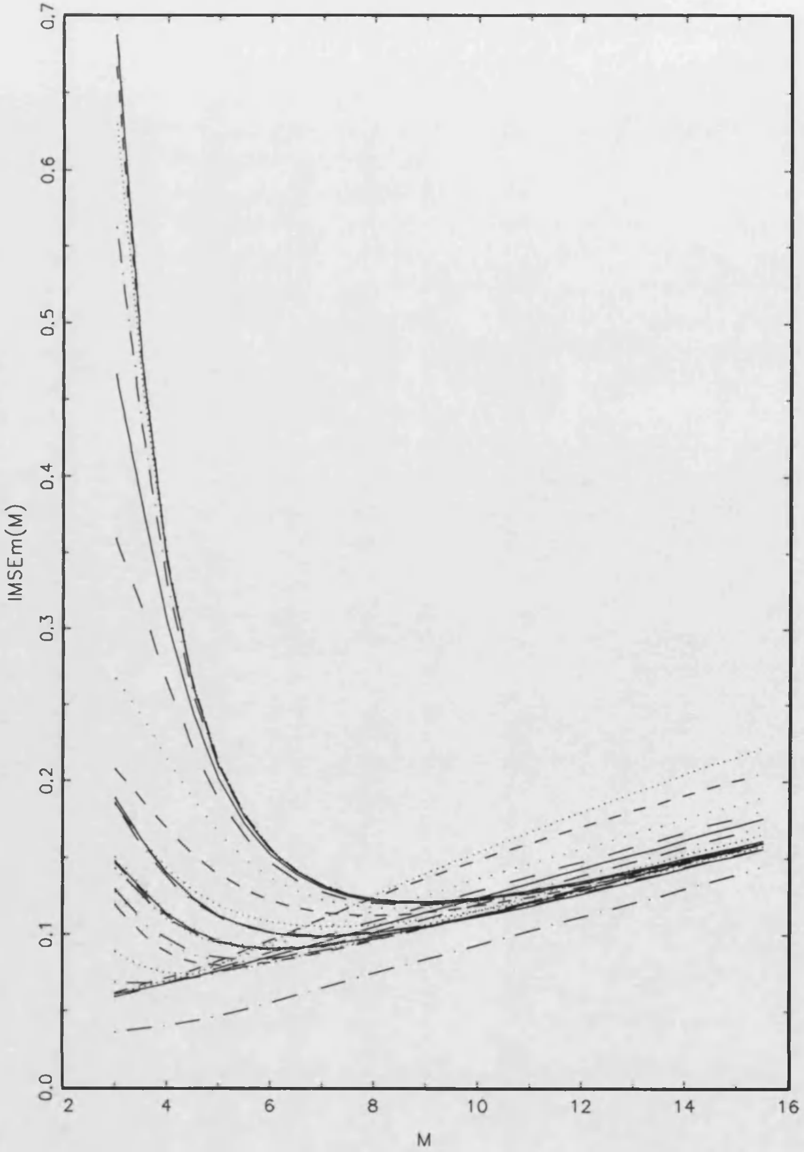
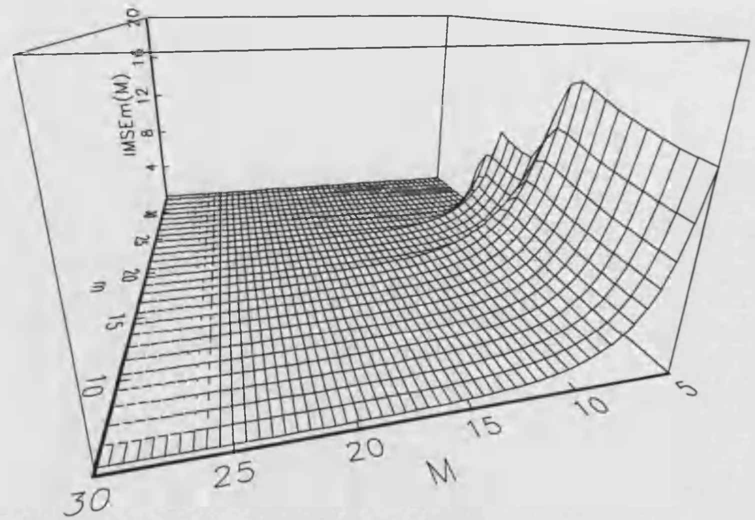


FIGURE 4. Model 3. Estimated IMSEm AR(2) .6 / −.9



Model 3. Estimated IMSEm AR(2) .6 / −.9

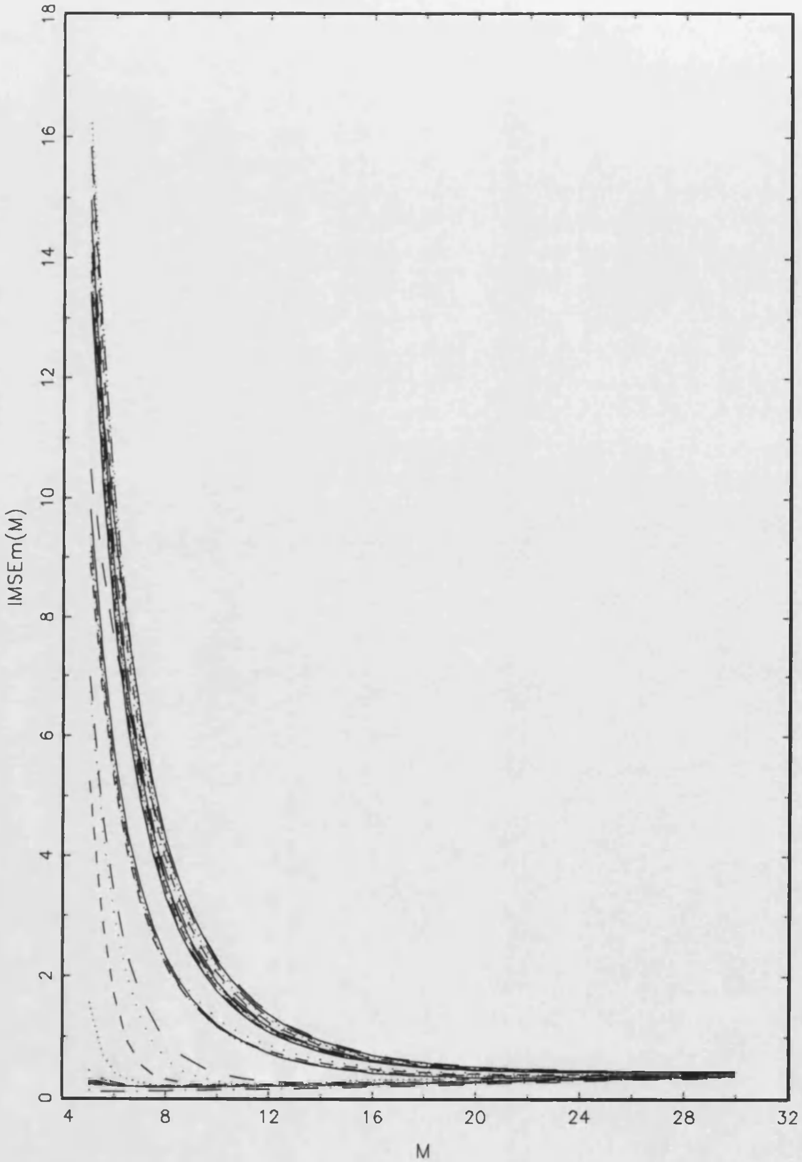
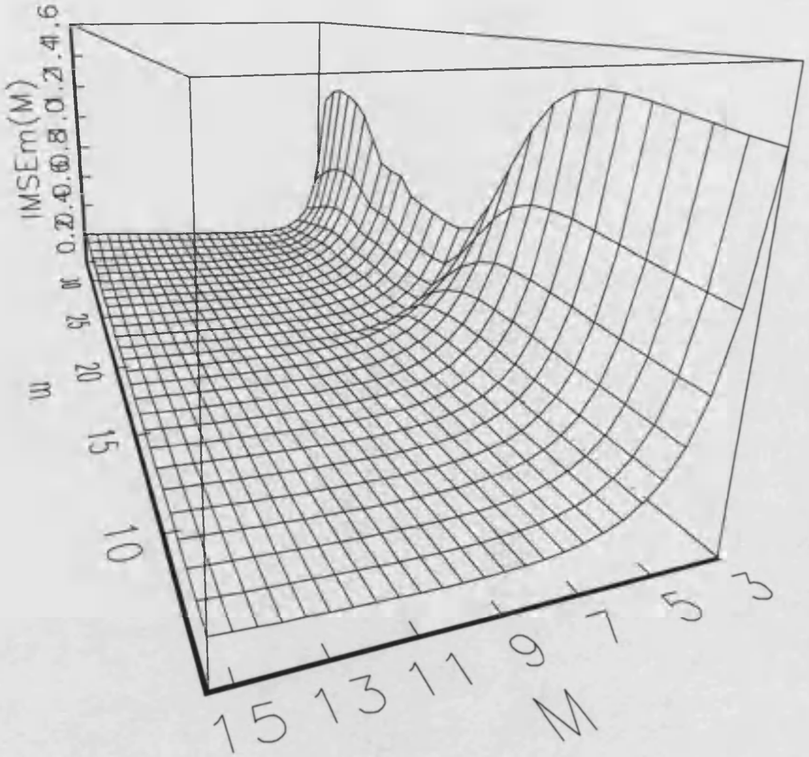


FIGURE 5. Model 4. Estimated IMSEm AR(2) .8 / -.6



Model 4. Estimated IMSEm AR(2) .8 / -.6

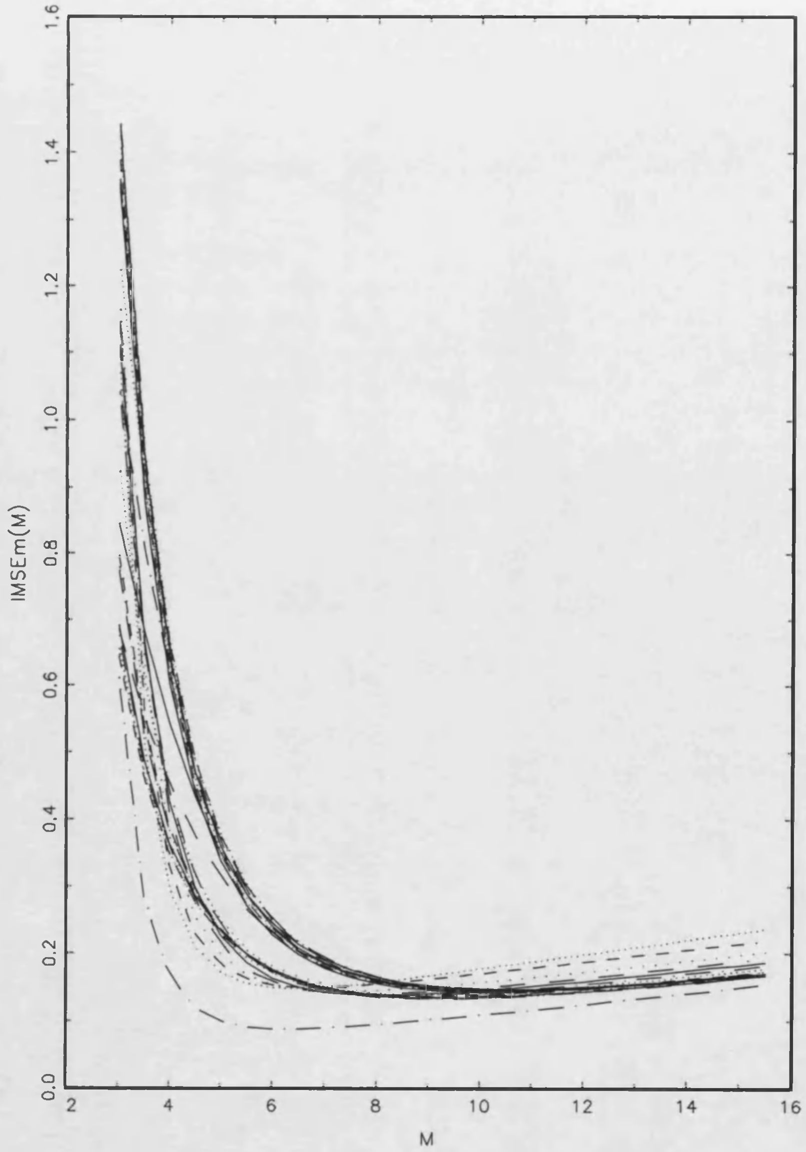
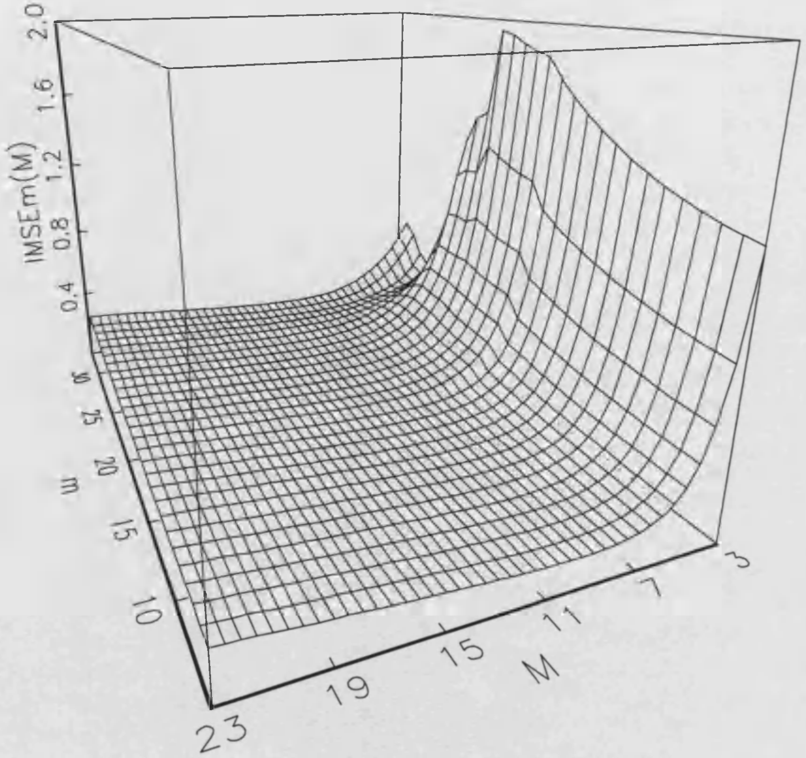


FIGURE 6. Model 5. Estimated IMSEm AR(1) .8



Model 5. Estimated IMSEm AR(1) .8

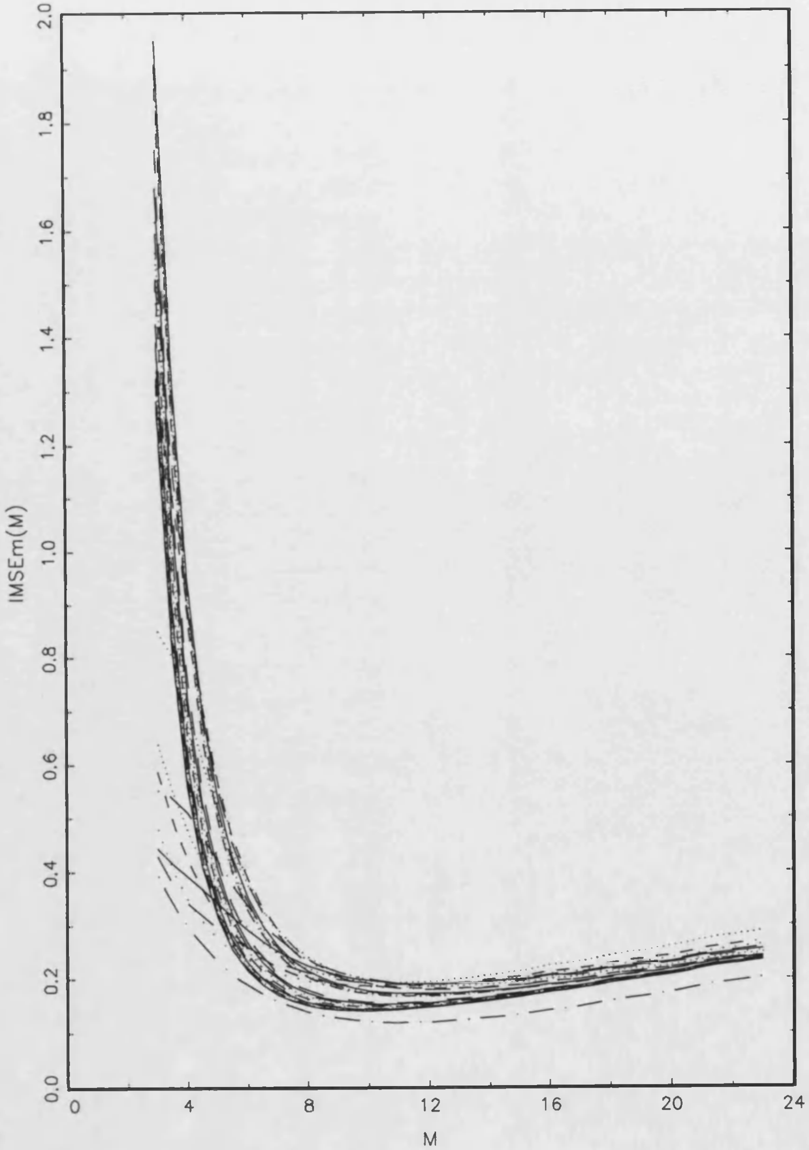
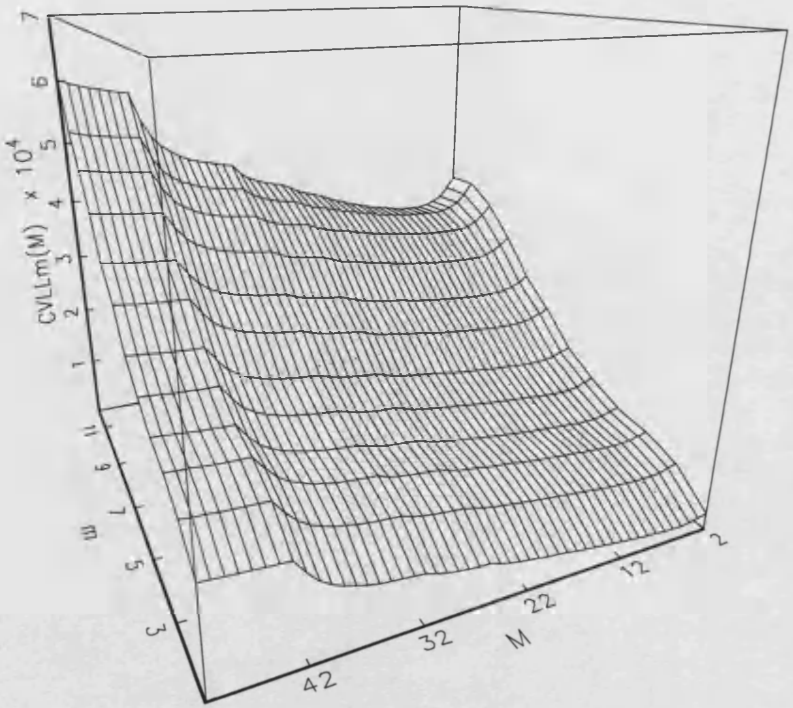


FIGURE 7. Model 1. CROSS VALIDATED LIKELD. AR(3) .7 / −.5 /.4



Model 1. CROSS VALIDATED LIKELIHOOD AR(3) .7 / −.5 /.4

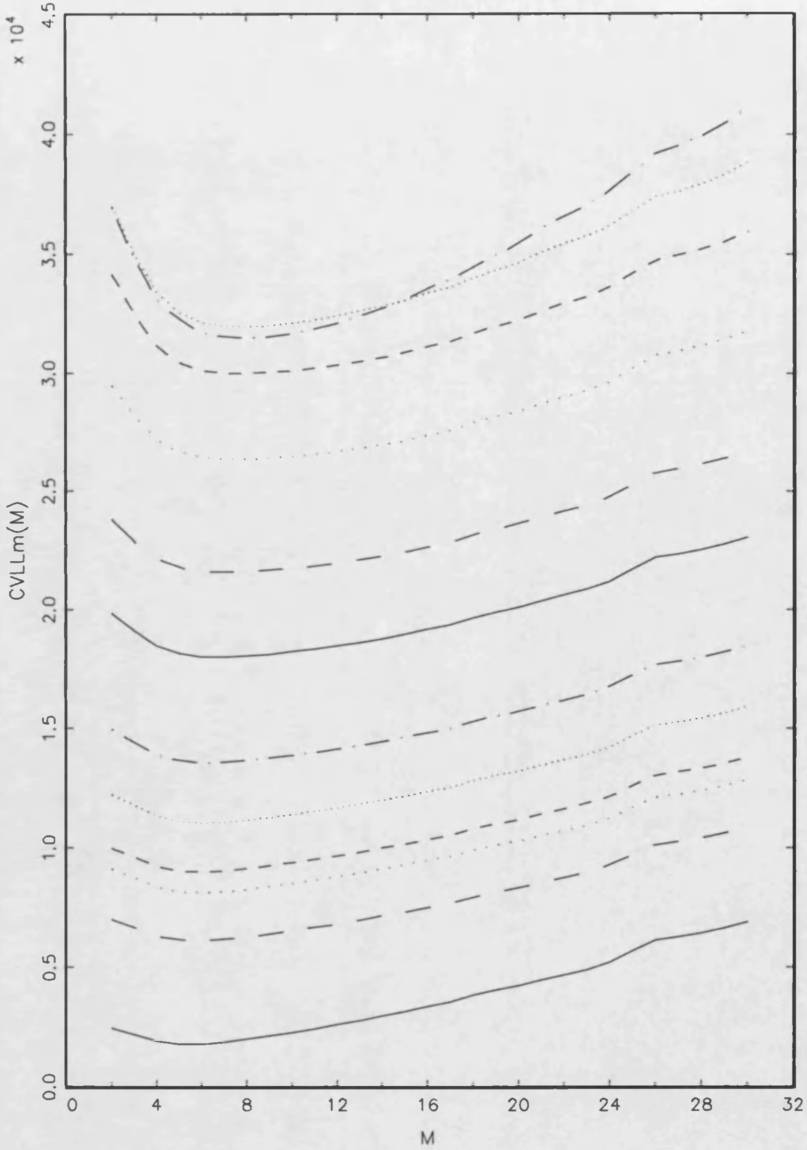
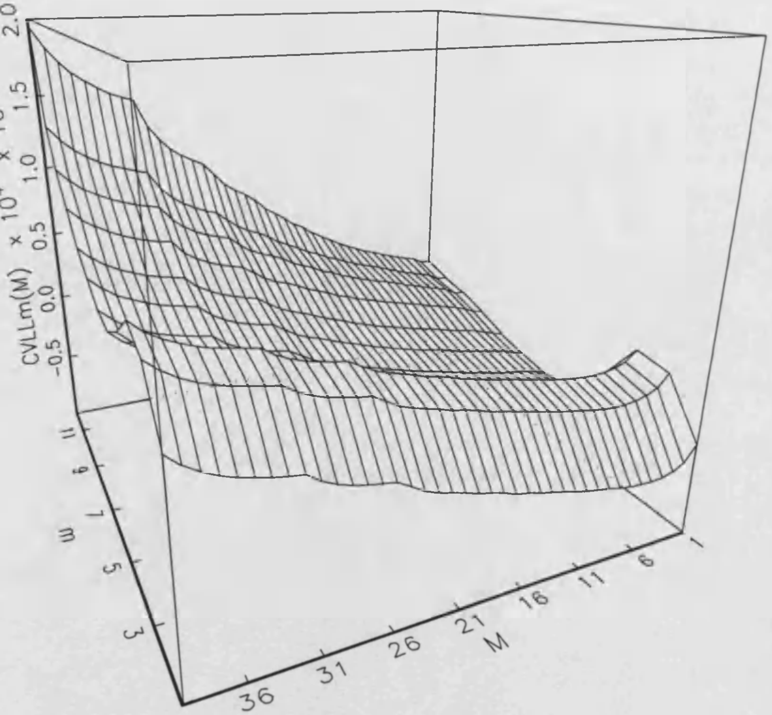


FIGURE 8. Model 2. CROSS VALIDATED LIKELD. AR(3) .6 / −.6 /.3



Model 2. CROSS VALIDATED LIKELIHOOD AR(3) .6 / −.6 /.3

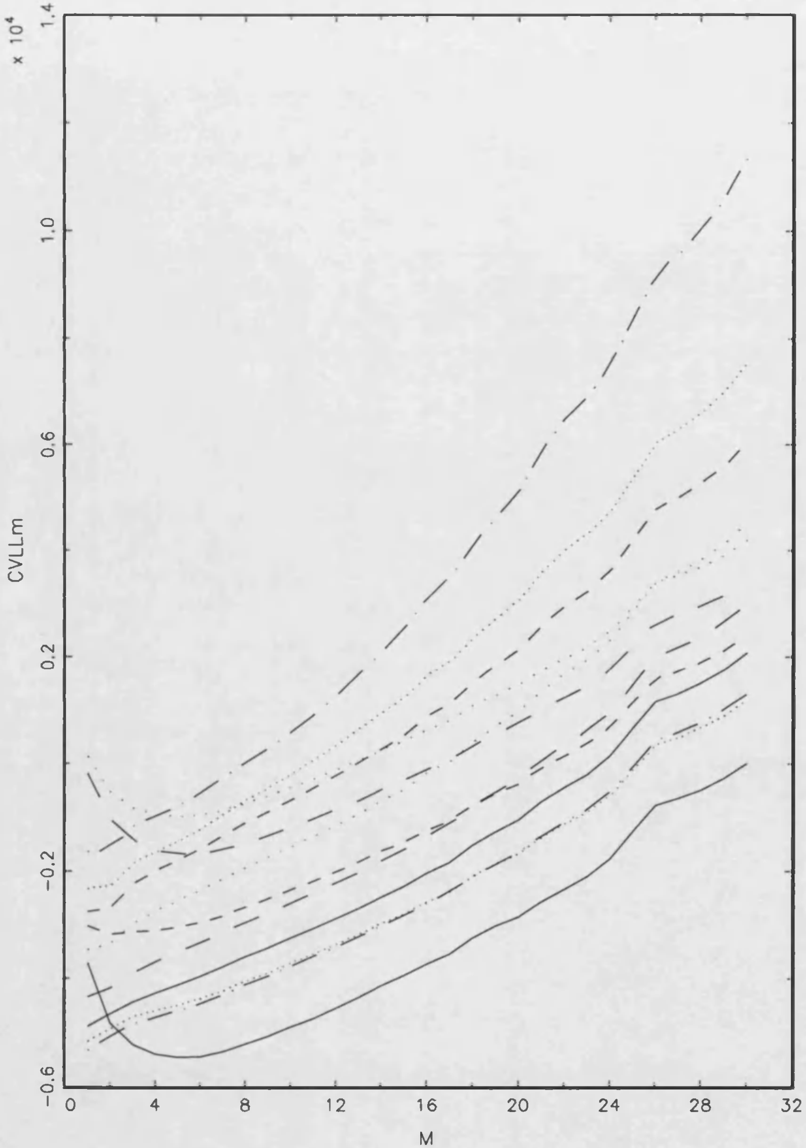
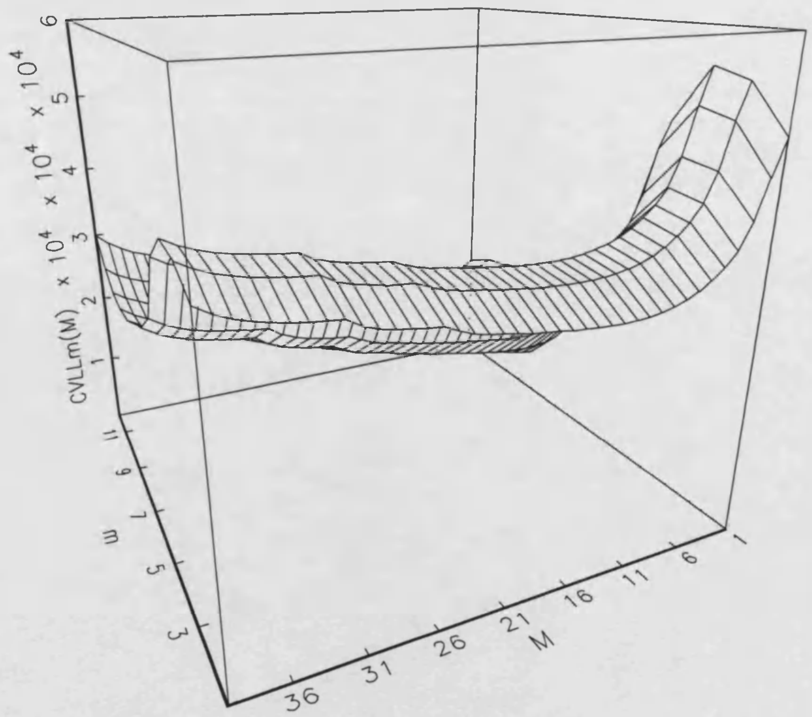


FIGURE 9. Model 3. CROSS VALIDATED LIKELIHOOD AR(3) .6 / -.9



Model 3. CROSS VALIDATED LIKELIHOOD AR(3) .6 / -.9

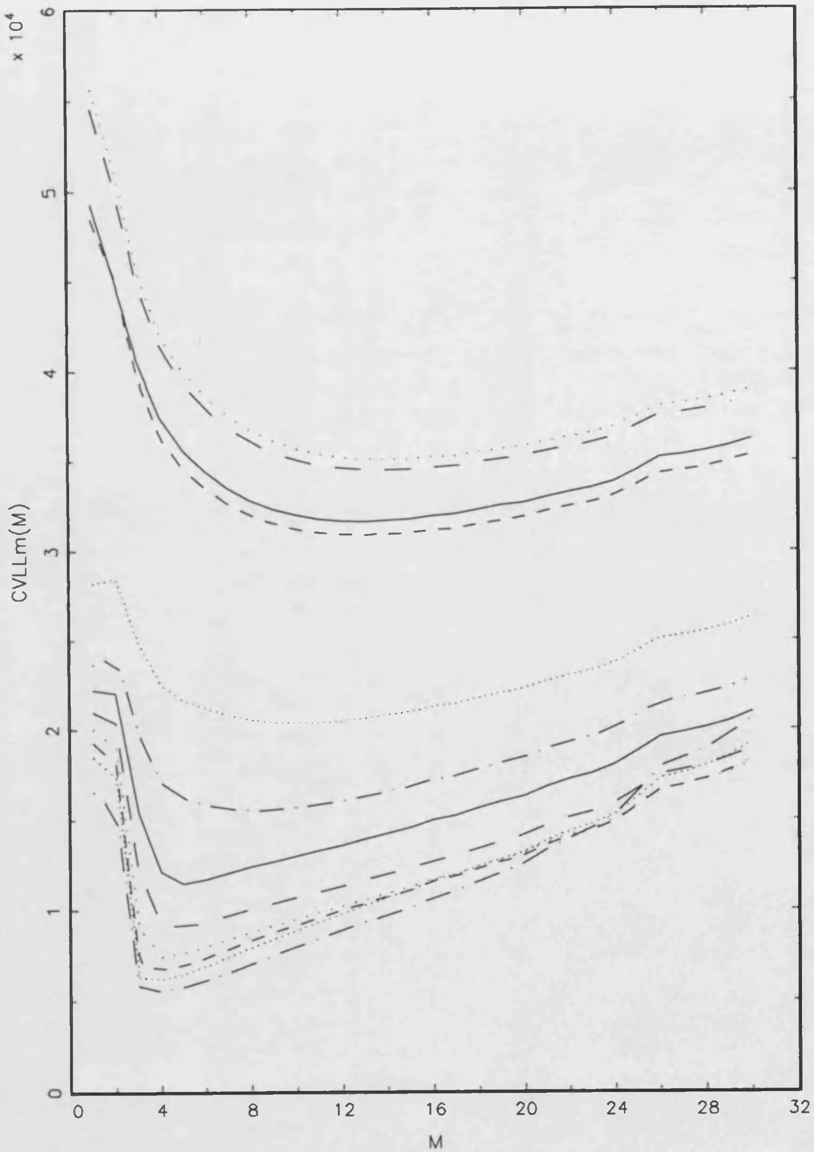
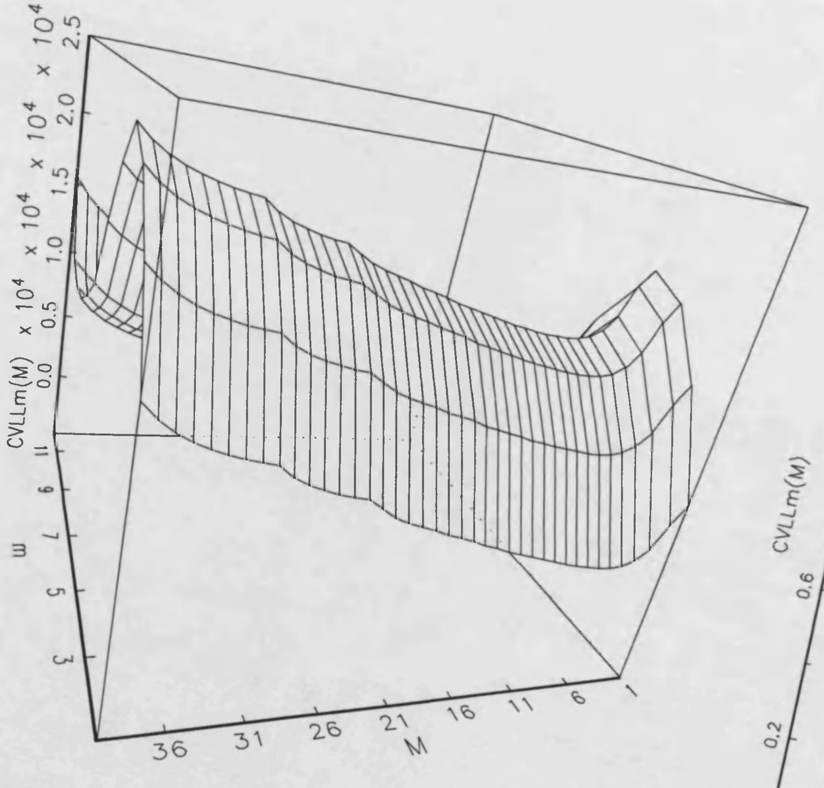


FIGURE 10. Model 4. CROSS VALIDATED LIKELIHOOD AR(2) .8 / -.6



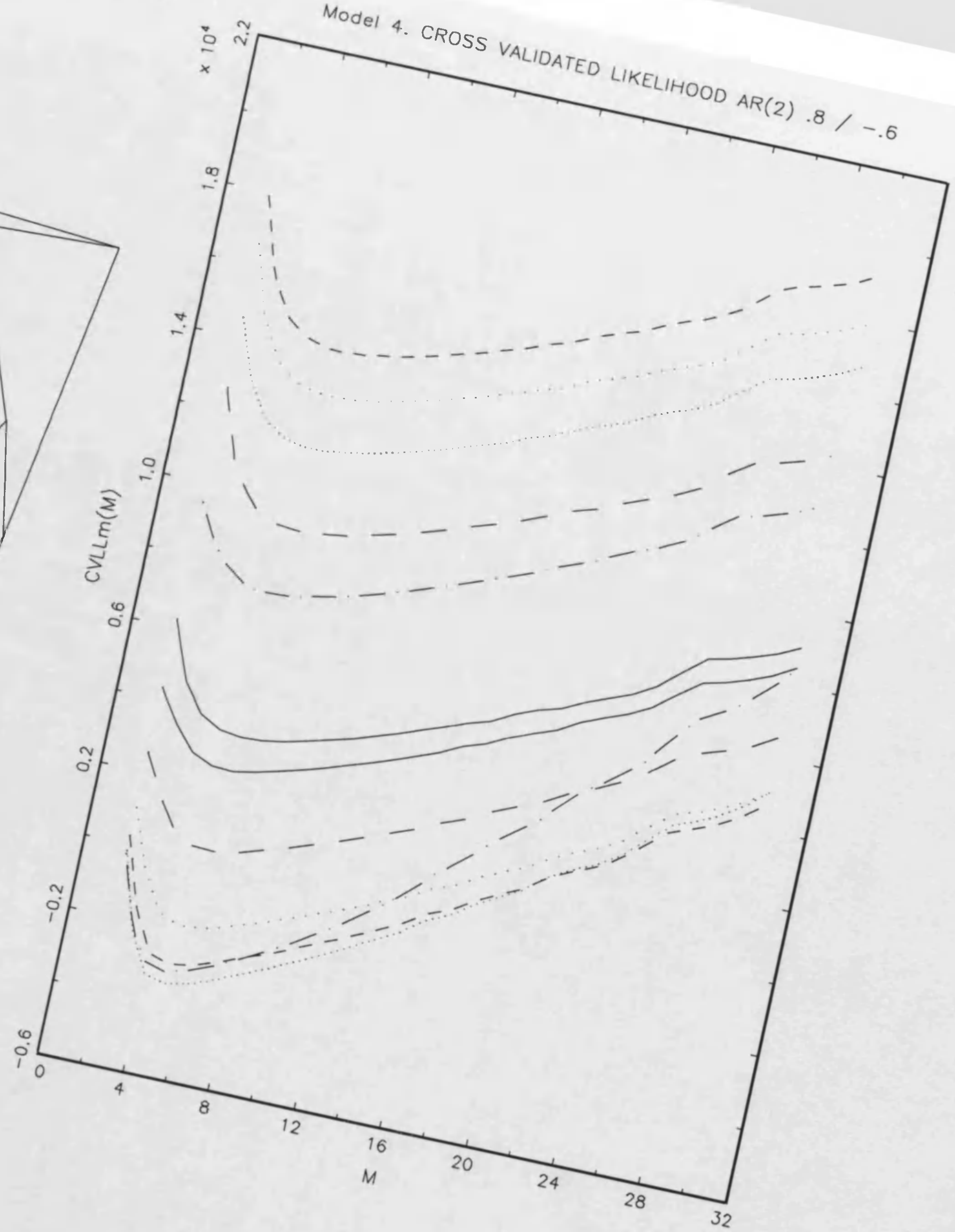
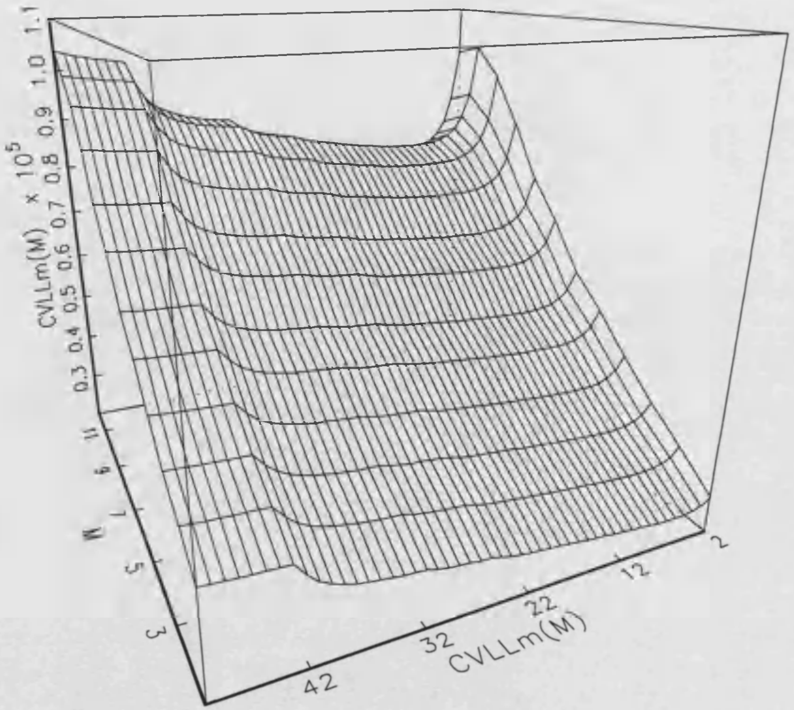
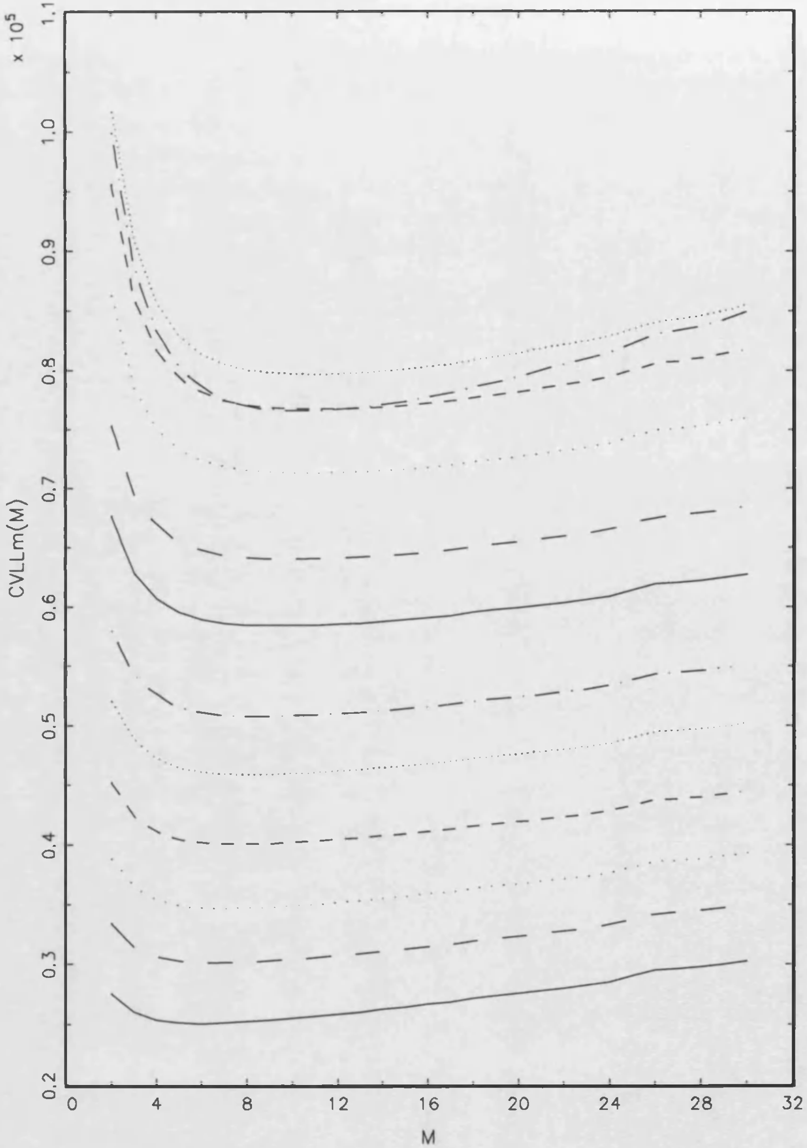


FIGURE 11. Model 5. CROSS VALIDATED LIKELIHOOD AR(1) .8



Model 5. CROSS VALIDATED LIKELIHOOD AR(1) .8



Chapter 5

Log-periodogram regression for long range dependent time series

5.1 Introduction

In this chapter we consider statistical inference for long range dependence time series. In particular, we concentrate on the semiparametric estimate of the long memory parameter based on the regression on the logarithm of the periodogram at Fourier frequencies close to the origin.

Properties of maximum likelihood methods have been analyzed extensively for parametric models of long range dependence (see, for example, Fox and Taqqu (1986) and Dahlhaus (1989)). However, if we are only interested in the estimation of long range dependence characteristics, semiparametric and nonparametric models are more robust against any sort of misspecification of the short term behaviour of the time series.

Semiparametric models focus on some properties of the autocovariance sequence (hyperbolic decay) or of the spectral density (singularity at the zero frequency). They are semiparametric because they do not make explicit assumptions on the behaviour of the autocovariances at short lags or on the spectral density apart from the origin.

We set our conditions in the frequency domain in terms of the spectral density since they are much more neat and cover a broader range of possibilities. We will assume that the spectral density satisfies

$$f(\lambda) \sim G\lambda^{-2d} \qquad \text{as } \lambda \to 0^+, \tag{5.1}$$

where $d \in (0, 1/2)$, is the self similar parameter that governs the degree of strong dependence of the series. This is the interval of values of d for which the series exhibits long range dependence and is stationary. Expression (5.1) reflects a linear relationship between the spectral density and the frequency in log-log coordinates, with slope $-2d$, and this is the basis for the log-periodogram estimate.

Robinson (1994, 1995b and 1995c) and Lobato and Robinson (1994) have used similar assumptions to the ones we employ here to study the asymptotic behaviour of several semiparametric estimates of d. Robinson (1995b) justified a modified version of a procedure proposed by Geweke and Porter-Hudak (1983) based on the regression of the logarithm of a pooled periodogram on the logarithm of the Fourier frequencies close to the origin. He proved the consistency and asymptotic normality of the estimate for Gaussian vector time series. In this chapter we extend his consistency results for linear processes not necessarily Gaussian.

Let $\{X_t,\ t = 1, 2, \ldots\}$ be a stationary process with spectral density satisfying (5.1). Given an observable sequence X_t, $t = 1, \ldots, N$, we introduce the discrete Fourier transform at the frequency $\lambda_j = 2\pi j/N$,

$$d_x(\lambda_j) = \sum_{t=1}^{N} X_t e^{it\lambda_j},$$

and the periodogram

$$I(\lambda_j) = (2\pi N)^{-1} \left| d_x(\lambda_j) \right|^2.$$

Define for $J = 1, 2, \ldots$, fixed, (assuming $(m - \ell)/J$ integer),

$$Y_k^{(J)} = \log\left(\sum_{j=1}^{J} I(\lambda_{k+j-J})\right) \qquad k = \ell + J, \ell + 2J, \ldots, m.$$

The estimate considered in Robinson (1995b) is

$$\hat{d} = \left(\sum_k \Lambda_k^2\right)^{-1} \left(\sum_k \Lambda_k Y_k^{(J)}\right),$$

where $\Lambda_k = z_k - \overline{z}$, $\overline{z} = \frac{J}{m-\ell}\sum_k z_k$ and $z_k = -2\log\lambda_k$. Here m is an integer smaller than N and ℓ is a user-chosen trimming number. In the asymptotics both numbers tend to infinity with the sample size N, but more slowly.

In the first part of the chapter, we analyze the possibility of substituting the finite averages (for J fixed) of the periodogram in $Y_k^{(J)}$, by consistent estimates of the spectral density $f(\lambda_k)$. We will consider estimates of the weighted-autocovariance type, which can be written as a continuous average of the periodogram as in Chapter 3, $\Pi = [-\pi, \pi]$,

$$\widehat{f}_M(\lambda_k) = \int_\Pi K_M(\lambda - \lambda_k) I(\lambda) d\lambda = \frac{1}{2\pi} \sum_{j=1-N}^{N-1} \omega\left(\frac{j}{M}\right) \widehat{\gamma}(j) \cos j\lambda_k,$$

where $\widehat{\gamma}(j)$ is the (biased) estimate of the j-th autocovariance (assuming now zero mean),

$$\widehat{\gamma}(j) = \frac{1}{N} \sum_{1 \le t, t+j, \le N} X_t X_{t+j}, \qquad j = 0, \pm 1, \ldots, \pm(N-1),$$

and $M = M_N$ is a lag-number, tending to infinity with the sample size, but at a slower rate. We will assume

$$K_M(\lambda) = M \cdot K(M\,\lambda),$$

where $K(\cdot)$ is an even function, that integrates to one, with compact support to avoid smooth the periodogram for frequencies very close to the origin and leakage from the pole of $f(\lambda)$ at $\lambda = 0$. The function $\omega(\cdot)$ or 'lag-window' is the Fourier transform of K,

$$\omega(r) = \int K(\lambda) e^{ir\lambda} d\lambda.$$

In the analysis of this class of estimates of d, the main problem resides in the proof of the consistency of the nonparametric estimates $\widehat{f}_M(\lambda_k)$, after suitable normalization, for frequencies that are tending to zero, but not including the origin itself. Gaussianity is not needed for that, but when we assume that X_t is Gaussian, we are able to prove a central limit for these estimates of f. Then, the consistency of the appropriate modification of $\widehat{d}$ follows easily from the consistency of $\widehat{f}_M(\lambda_k)$. Regrettably, the properties of $\widehat{f}_M(\lambda_k)$ can not be deduced from the standard results on nonparametric estimation of the spectrum. Typically, the conditions assumed in the literature include at least boundedness of the spectral density, and it has not been considered the case of estimation around a singularity.

The situation is quite different when we consider fixed averages of the periodogram. We have to deal now with the logarithm of the periodogram, which is a random variable and not a consistent estimate of anything. Non-linear functions (the logarithm in particular) of the periodogram of stationary sequences have been considered under different set-ups (see e.g. Hannan and Nicholls (1977), Taniguchi (1979), Chen and Hannan (1980), von Sachs (1994a), Janas and von Sachs (1993), Robinson (1995b) and Comte and Hardouin, (1995a and 1995b), and the references cited there). These works assume Gaussianity to obtain the main results, except Chen and Hannan, and Janas and von Sachs, who work with a linear process condition.

These last two references use higher order properties of the asymptotic distribution of the periodogram. Janas and von Sachs apply the results for weakly dependent sequences of Götze and Hipp (1983), making almost impossible to relax their assumptions for long range dependence situations. Instead, the approach of Chen and Hannan (1980) is based in the factorization of the periodogram of the observable sequence in the transfer function of the linear filter, times the periodogram of the i.i.d. innovations, plus a stochastic error term. The magnitude of this error depends on the smoothness of the spectral density and on the number of moments assumed for the innovations. Obviously the conditions they assumed, ($\sum |j|^{\delta}|w_j| < \infty$, $\delta > 1/2$, see Assumption 5.7 below), rule out any long memory behaviour or any singularity in the spectral density of X_t, but their results are based mainly on the properties of the periodogram of the i.i.d. innovations sequence, for which we assume the same set of conditions as in their Theorem 2 (see Assumption 5.9 below).

A related approach is used by Comte and Hardouin in a long-memory environment but assuming Gaussianity. We use one idea of them to avoid a modification of the estimate of d in the same spirit as the one Chen and Hannan (1980) did to account for *too small* values of $I(\lambda_j)/f(\lambda_j)$. Here, instead of redefining the periodogram with a truncation, we use an average of periodogram ordinates. Then we can use their higher order asymptotic results and the long range dependence results of Robinson (1995c) to approximate the periodogram of X_t by that of the linear i.i.d. innovations times the long memory transfer function.

The chapter is organized as follows. In next section we give the basic results about

$\widehat{f}_M$ when Gaussianity is assumed, and then we use $\widehat{f}_M$ to estimate d in Section 5.3. In Section 5.4 we show the robustness of the previous results for non Gaussian linear processes. Finally in Section 5.5 we give sufficient conditions for the consistency of $\widehat{d}$ when finite averages of the periodogram are used. All the proofs and auxiliary lemmas are given in appendices in the last three sections of the chapter.

5.2 Nonparametric estimates of the spectral density

In this section we consider asymptotic properties of nonparametric kernel estimates of the spectral density of long range dependent series for frequencies tending to zero as the sample size increases. We need to impose some restrictions on the semiparametric model for the spectral density, the form of the kernel function K and on the smoothing number M and the frequencies λ_j included in the estimate of the long memory parameter d.

We first introduce two conditions about the spectral window K, equivalent to the ones used in Chapter 3:

Assumption 5.1 *$K(x)$ is a bounded, even function, $-\pi \leq x \leq \pi$, and zero elsewhere, with*

$$\int_{\Pi} K(x)dx = 1.$$

Assumption 5.2 *$K(x)$ satisfies a uniform Lipschitz condition (of order 1) in $[-\pi, \pi]$.*

Now we introduce some assumptions about the behaviour of the spectral density around the origin, following Robinson (1995b and 1995c), but not considering negative values of d,

Assumption 5.3 *X_t is covariance stationary and as $\lambda \to 0^+$,*

$$f(\lambda) = G\lambda^{-2d} + O(\lambda^{\beta-2d}),$$

with $d \in [0, 1/2)$, $\beta \in (0, 2]$ and $0 < G < \infty$.

Assumption 5.4 *In a neighbourhood of the origin f is differentiable and*

$$\frac{d}{d\lambda} \log f(\lambda) = O(\lambda^{-1}).$$

Finally we present one assumption about the frequencies $\lambda_j = 2\pi j/N$, allowing j to increase with N, for which we will consider the nonparametric estimates $\widehat{f}_M(\lambda_j)$, and about the lag-number M,

Assumption 5.5 *As* $N \to \infty$,

$$\frac{M}{N} + \frac{1}{M} + \frac{j}{N} + \frac{N}{j\,M} \to 0. \tag{5.2}$$

Assumptions 5.1 and 5.2 guarantee a well-behaved kernel K and as in Chapter 3 the compact support is intended to avoid leakage from other frequencies, in the long memory case specially from zero frequency. We chose to split them in two conditions, because Assumption 5.2 is not necessary to analyze the expectation of $\widehat{f}_M$. They are satisfied by several kernels used in practice, like the uniform and Barlett-Priestley kernels. It would be equivalent to define the function K in any compact support $[-\tau, \tau]$ for finite $\tau > 0$, and all the results are valid with the obvious modifications in the proofs. We can also assume that $K(\alpha)$ is nonnegative to avoid meaningless negative estimates of f or problems with the logarithm function.

No assumptions are imposed on f outside a neighbourhood of the origin, apart from integrability, due to stationarity. They are satisfied for long range dependence parametric models like the fractional ARIMA (1.31) and the fractional noise (1.32) with $\beta = 2$. See Robinson (1995b) for a general account about these conditions and their connection with the closely related (and more restrictive) condition, $0 < g < \infty$,

$$\gamma(r) = \mathrm{Cov}[X_t, X_{t+r}] \sim g\, r^{2d-1}, \qquad \text{as} \quad r \to \infty.$$

The first two conditions in Assumption 5.5 are standard in nonparametric estimation of the spectral density. The third condition implies that $\lambda_j \to 0$ as $N \to \infty$. The last condition in Assumption 5.5 is used, jointly with the compact support kernel K_M (cf. Assumption 5.1), to avoid averaging for periodogram ordinates too near to the origin, since necessarily $j \to \infty$ as $N \to 0$. If, for example, j/N tends to 0 very fast, we need the kernel K_M to close around λ_j even faster. This has also implications from the bias of the estimate point of view. Then, for $|\lambda| \le \pi/M$ and $j > 0$,

$$\inf_{|\lambda| \le \pi/M} \lambda + \lambda_j \ge 2\pi\frac{j}{N} - \frac{\pi}{M} = \pi\left[\frac{2j}{N} - \frac{1}{M}\right] = \pi M^{-1}\left[2\,\frac{j\,M}{N} - 1\right] > cM^{-1}, \tag{5.3}$$

as $N \to \infty$, for some constant $c > 0$. Hence we can take $\lambda + \lambda_j > 0$ under Assumption 5.5, $|\lambda| \leq \pi/M$ and $j > 0$. The condition

$$\lim_{N\to\infty} \sup \frac{jM}{N} > \frac{1}{2}$$

would imply also (5.3), but it is not sufficient due to bias problems (see Lemma 5.1).

We now present several results concerning the asymptotic behaviour of $\widehat{f}_M(\lambda_j)$ when we assume that X_t is a Gaussian stationary sequence. They are not very different from the standard results on nonparametric estimation of smooth spectral densities, after appropriate normalization.

Lemma 5.1 *If X_t is a zero mean covariance stationary sequence, then, under Assumptions 5.1, 5.3, 5.4, 5.5 and additionally $j^{-1}\log N \to 0$,*

$$\frac{E[\widehat{f}_M(\lambda_j)] - f(\lambda_j)}{f(\lambda_j)} = O\left(\frac{\log N}{j} + \frac{N}{jM}\right).$$

Define

$$||K||_2^2 = \int_\Pi K^2(x)dx.$$

Lemma 5.2 *If X_t is a zero mean Gaussian stationary sequence, under Assumptions 5.1, 5.2, 5.3, 5.4, 5.5, and $MN^{-1}\log^3 N \to 0$,*

$$\begin{aligned}
\frac{N}{M}\frac{Var[\widehat{f}_M(\lambda_j)]}{f(\lambda_j)^2} &= 2\pi||K||_2^2 + O\left(\left[\frac{M}{N}\log^3 N\right]^{1/2} + j^{-1}\log^3 N + \frac{N}{jM}\right) \\
&= 2\pi||K||_2^2 + o(1).
\end{aligned}$$

Note that Assumption 5.5 and $MN^{-1}\log^3 N \to 0$ imply $\log^3 N/j \to 0$ as $N \to \infty$. Similar techniques as the ones used in the proof of this lemma can be utilized to estimate the covariance of two spectral estimates for frequencies apart more than M^{-1}.

In the previous lemmas we have assumed that the mean of the time series was known and equal to zero without loss of generality. Now we investigate the consequences of the estimation of the expectation of the series using the sample mean. As it is well known, under long memory conditions the rate of convergence of the sample mean is slower than under weak dependence conditions. Assuming zero mean w.l.o.g., Assumption 5.3

and covariance stationarity, we can get (see Adenstedt (1974)), that the variance of the sample mean, $\overline{X} = N^{-1}\sum_{t=1}^{N} X_t$, satisfies

$$\mathrm{Var}[\overline{X}] = O(N^{2d-1}).$$

We now consider a continuous average of the periodogram of the mean corrected series, $I^{\star}(\lambda)$,

$$\widetilde{f}(\alpha) \quad = \quad \int_{\pi}^{\pi} K_M(\lambda-\alpha) I^{\star}(\lambda) d\lambda$$

where

$$I^{\star}(\lambda) = \frac{1}{2\pi N}\left|\sum_{t=1}^{N}(X_t - \overline{X})e^{it\lambda}\right|^2.$$

In the next lemma we estimate the error introduced by the estimation of the expectation of X_t by $\overline{X}$. This error is negligible with respect to the standard deviation and bias of the estimates, so in the following, we will work with the estimate $\widehat{f}$, bearing in mind that all the results will go through for $\widetilde{f}$ as well.

Lemma 5.3 *If X_t is a covariance stationary sequence, under Assumptions 5.1, 5.3, 5.5 and $j^{-1}\log^2 N \to 0$,*

$$\sqrt{\frac{N}{M}}\left(\frac{\widetilde{f}(\lambda_j)}{f(\lambda_j)} - \frac{\widehat{f}(\lambda_j)}{f(\lambda_j)}\right) = O_P\left(\frac{N^{1/2}}{jM^{1/2}} + j^{-1/2}\log N\right) = o_P(1).$$

Finally we give a central limit theorem for the centred and normalized estimate.

Theorem 5.1 *If X_t is a Gaussian stationary sequence and under Assumptions 5.1, 5.2, 5.3, 5.4, 5.5 and $MN^{-1}\log^{2s-1} N \to 0$, for all $s \geq 2$,*

$$\sqrt{\frac{N}{M}}\left(\frac{\widehat{f}_M(\lambda_j) - E[\widehat{f}_M(\lambda_j)]}{f(\lambda_j)}\right) \to_{\mathcal{D}} \mathcal{N}(0, 2\pi\|K\|_2^2).$$

5.3 Estimation of the long memory parameter based on nonparametric spectrum estimates

The estimates $\widehat{f}_M(\lambda_j)$ can be used to construct a semiparametric estimate of the parameter d that governs the long range behaviour of the time series. We should use those estimates for frequencies tending to zero, $j/N \to 0$, for which Assumption 5.3 is reasonable, and in a number tending to infinity to achieve consistency. This idea has been used for strongly dependent Gaussian times series previously by Hassler (1993), Chen et al. (1994) and Reisen (1994). However there are problems in their proofs similar to those in Geweke and Porter-Hudak's (1983), as was pointed out by Robinson (1995b). We consider first the Gaussian case in this section. Then, in next section we extend these results to non-Gaussian series under a linear process condition and mild assumptions on the fourth-order cumulants.

We follow a similar approach to that of Robinson (1995b). We modify his estimate $\widehat{d}$ with regressors based on $\widehat{f}_M(\lambda_j)$, assuming now K is positive and defining

$$\widehat{d}_M = \left(\sum_{k=\ell+1}^{m} \Lambda_k \log \widehat{f}_M(\lambda_k) \right) \left(\sum_{k=\ell+1}^{m} \Lambda_k^2 \right)^{-1} .$$

Equivalently, we could have constructed the estimate with the sum in k running for values $k = \ell + \left[\frac{2N}{M}\right], \ell + 2\left[\frac{2N}{M}\right], \ldots, m$ (we assume that $(m-\ell)\frac{M}{2N}$ is integer). In this way we would have divided the spectral band close to the origin in non-overlapping intervals of width $\frac{2N}{M}$, so that the *regressors* in $\widehat{d}_M$ were asymptotically uncorrelated. Then, in Robinson's notation, $J \sim \frac{2N}{M}$, and now we would be using about $\frac{m}{J} \sim m\frac{M}{2N}$ regressors instead of $m - \ell$. The proofs for this situation are not very different from the case we study in detail here.

To avoid any problem with the logarithm in the definition of the estimate, we include in Assumption 5.1 that K is positive, so is $\widehat{f}_M$ for all N.

There are two main differences between $\widehat{d}$ and $\widehat{d}_M$:

- The estimate $\widehat{d}_M$ uses continuous averages of the periodogram ($\widehat{f}_M$ is a weighted autocovariance type nonparametric estimate), meanwhile $\widehat{d}$ is based on a discrete average of periodogram ordinates.
- In the expression for $\widehat{d}$, $Y_k^{(J)}$ is a sum of a fixed number of periodogram ordinates, so it is not a consistent estimate of $f(\lambda_k)$, but $\widehat{f}_M(\lambda_k)$ in $\widehat{d}_M$ is consistent under appropriate conditions.

We set the following assumption about the sequences ℓ, M, m and N, for the consistency of $\widehat{d}_M$:

Assumption 5.6 *As $N \to \infty$,*

$$\frac{\ell\,(\log N)^2}{m} + \frac{N}{M\,\ell} + \frac{M\,\log^3 N}{N} + \frac{N\,\log N\,\log m}{M\,m} + \frac{m}{N} \to 0.$$

Mainly, Assumption 5.6 imposes an important lower bound on the rate of increase of M and ℓ with N. For example, for the choice $m = N^{1/2}$, recommended by Geweke and Porter-Hudak (1983) for their estimate, M has to grow faster than $N^{1/2}$. Then, it will be possible to find ℓ sequences satisfying the first two conditions in Assumption 5.6. This issue is related with a bias problem: we cannot expect to estimate f properly around the origin if we use a very broad band for the kernel (i.e. M too small), or if we are too close to 0 (i.e. ℓ too small), since f is very steep this region. Then m has to increase, basically, slower than N but faster than N/M.

Theorem 5.2 *Under Assumptions 5.1, 5.2, 5.3, 5.4 and 5.6, $\widehat{d}_M \to_P d$.*

Condition 5.6 could have been written down in terms of the difference $m-\ell$ instead of in terms of the ratio m/ℓ. In the case were only $m\frac{M}{2N}$ regressors are used, the consistency of $\widehat{d}_M$ will follow substituting Assumption 5.6 by

$$\frac{\ell\,(\log N)^2}{m} + \frac{N}{M\,\ell} + \frac{M\,\log^3 N}{N} + \left(\frac{N}{M}\right)^2 \frac{\log N\,\log m}{m} + \frac{m}{N} \to 0.$$

The $\log N$-consistency of $\widehat{d}_M$ for studentization purposes will follow from the next condition, slightly stronger that Assumption 5.6:

$$\frac{\ell\,(\log N)^2}{m} + \frac{N}{M\,\ell} + \frac{M\,\log^4 N}{N} + \frac{N\,\log^2 N\,\log m}{M\,m} + \left(\frac{m}{N}\right)^\beta \log N \to 0.$$

In order to obtain the asymptotic normality of $\widehat{d}_M$, it seems difficult to derive the cumulants or moments of $\log \widehat{f}_M$ from those of $\widehat{f}_M$. One approach could be to prove an Edgeworth expansion for the density of $\widehat{f}_M$, which will allow for the estimation of the moments of $\log \widehat{f}_M$ (similar to the one in Chen and Hannan (1980) for $I(\lambda_k)$).

5.4 Robustness to non-Gaussianity

In this section we propose conditions on the sequence X_t to obtain the consistency of $\widehat{f}_M(\lambda_j)$ and of $\widehat{d}_M$ when X_t is no longer assumed Normal distributed. This, basically, will require conditions on the fourth cumulant structure of the series. A bounded fourth order spectral density condition would do the job, but we allow for long memory also in the fourth cumulants via a linear process condition.

Instead of Gaussianity we introduce a fourth order stationary linear process condition, with filter coefficients compatible with Assumptions 5.3 and 5.4:

Assumption 5.7 *X_t satisfies*

$$X_t = \sum_{j=0}^{\infty} w_j \epsilon_{t-j}, \qquad \sum_{j=0}^{\infty} w_j^2 < \infty,$$

where $E[\epsilon_t^4] < \infty$.

Assumption 5.8 *In Assumption 5.7 ϵ_t is a fourth-order stationary process with fourth cumulant $\kappa_4^\epsilon(t_1, t_2, t_3) = Cumulant(\epsilon_{t_1+t}, \epsilon_{t_2+t}, \epsilon_{t_3+t}, \epsilon_t)$, $\forall t$, satisfying*

$$\sum_{t_1,t_2,t_3=-\infty}^{\infty} |\kappa_4^\epsilon(t_1, t_2, t_3)| < \infty,$$

and second-order spectral density satisfying

$$0 < f^\epsilon(\lambda) < \infty$$

for values of λ in an interval around the origin.

Under Assumptions 5.7 and 5.8, X_t is fourth-order stationary and it is immediate that ϵ_t has uniformly bounded fourth-order spectral density defined by

$$f_4^\epsilon(\nu_1, \nu_2, \nu_3) = \frac{1}{(2\pi)^3} \sum_{t_1,t_2,t_3=-\infty}^{\infty} \kappa_4^\epsilon(t_1, t_2, t_3) \exp\{-i(\nu_1 t_1 + \nu_2 t_2 + \nu_3 t_3)\}.$$

Setting,

$$w(\lambda) = \sum_{j=0}^{\infty} w_j \exp\{i\lambda j\},$$

we have that $f(\lambda) = |w(\lambda)|^2 f^\epsilon(\lambda)$, with

$$|w(\lambda)| = O(f^{1/2}(\lambda)) \quad \text{as} \quad \lambda \to 0^+.$$

Note that we do not need to assume that the ϵ_t are i.i.d., uncorrelated or independent up to four moments: we only require the ϵ_t's to be short-memory in the fourth-cumulants (bounded fourth-order spectral density) and with no long range dependence in the second moments (bounded second-order spectral density around the origin). We need the boundedness of f^ϵ away from zero for small frequencies just for compatibility with Assumption 5.3 about f.

These conditions imply that the long range dependence properties of X_t at $\lambda = 0$ come only from the linear filter $\{w_j\}$ and not from the innovations ϵ_t. But X_t's second order (for frequencies different from zero) and fourth order dependence characteristics for $\lambda \neq 0$ are freely determined by the shape of $w(\lambda)$ and $f^\epsilon(\lambda)$.

To establish the consistency of $\widehat{f}_M$ we only need to estimate its first two moments. In the evaluation of the bias we have presented before we only used the second-order stationarity of X_t and no specific Gaussian properties. Therefore Lemma 5.1 is valid under Assumptions 5.7 and 5.8 instead of Gaussianity. The same applies for Lemma 5.3.

We now establish our first result about the variance of the nonparametric estimate of the spectral density for frequencies close to the origin, which is exactly the same as Lemma 5.2:

Lemma 5.4 *Under Assumptions 5.1, 5.2, 5.3, 5.4, 5.5, 5.7, 5.8 and* $MN^{-1}\log^3 N \to 0$, *the conclusions of Lemma 5.2 hold.*

Then, since $\widehat{f}_M(\lambda_j)$ has the same asymptotic variance and bias under Assumption 5.8 that under Gaussianity, we have the equivalent to Theorem 5.2:

Theorem 5.3 *Under Assumptions 5.1, 5.2, 5.3, 5.4, 5.6, 5.7 and 5.8,* $\widehat{d}_M \to_P d$.

5.5 Estimation of the long memory parameter based on finite averages of the periodogram

In this section we will consider Robinson's (1995b) estimate $\widehat{d}$ when finite averages (for J fixed) of the periodogram of X_t are used under the linear process condition in Assumption 5.7. As we commented in the Introduction, the consistency proof is based on the approximation of the logarithm of the periodogram of X_t by that of ϵ_t, times the transfer function $|\omega(\lambda)|^2$ multiplied by 2π. This approximation will depend on the properties of the filter $\{\omega_j\}$ and on the distribution of the linear innovations ϵ_t. Special care is needed because of the singularity of the logarithm function at the origin.

We introduce the next assumption following Chen and Hannan (1980) to analyze that stochastic approximation:

Assumption 5.9 *The ϵ_t in Assumption 5.7 are i.i.d., with characteristic function $\widehat{Q}(\theta) = E[e^{i\theta\epsilon_t}]$ satisfying*

$$\sup_{|\theta|\geq\theta_0} \left|\widehat{Q}(\theta)\right| = \delta(\theta_0) < 1, \quad \forall\theta_0 > 0, \quad \textit{and}$$

$$\int_{-\infty}^{\infty} |\widehat{Q}(\theta)|^p d\theta < \infty, \qquad \textit{for some integer } \; p > 0.$$

Note that Assumption 5.9 implies Assumption 5.8. Here we do not need to take care of the mean of the series, since we are omitting the periodogram at zero frequency in the definition of $\widehat{d}$.

The conditions of Assumption 5.9 are needed to prove the validity of an approximation in Lemma 5.10 for the density of the Fourier Transform of the innovations ϵ_t. The first line is a Cramér condition. The second condition is used to approximate the density and it would not be necessary to approximate the distribution function. It implies that the distribution of ϵ_t has a bounded continuous density (see, for example, Theorem 3 in p. 509 of Feller (1971)).

In the proofs we are able only to deal with the case $J \geq 2$. The reason is the following. The average of J periodogram ordinates of an i.i.d sequence will be asymptotically distributed as a χ^2_{2J} (up to constants). But for the approximation between the periodograms of X_t and of ϵ_t we need to consider the inverse moments of the periodogram of ϵ_t. The trick is that if $Z \sim \chi^2_{2J}$, $E[Z^{-1}] < \infty$ for $J \geq 2$ (see Lemma 5.11 below). Of course, to approximate the moments of a random variable we need something more that its asymptotic distribution. That is why we approximate the density of the Fourier transform of ϵ_t and the regularity conditions on Assumption 5.9. We conjecture that a related argument to the truncation of Chen and Hannan (1980) can be used to construct a proof for $J = 1$. These results in form of Lemmas 5.10 and 5.11 are postponed to the Appendix in Section 5.9.

We will make direct use of some results of Robinson (1995b and 1995c) to analyze the characteristics of the linear filter $w(\lambda)$ under Assumptions 5.3 and 5.4. We introduce some more notation. Write $I_j = I(\lambda_j)$ and $f_j = f(\lambda_j)$, and for the periodogram of the ϵ_t sequence, $I_{\epsilon j} = I_\epsilon(\lambda_j)$. Let J be a given, fixed, integer greater than or equal to 2. Define

$$\overline{I}_k = \sum_{j=1}^{J} I_{k+j-J}, \qquad k = \ell + J, \ell + 2J, \dots, m,$$

and equivalently

$$\overline{I}_{\epsilon k} = \sum_{j=1}^{J} I_{\epsilon,k+j-J}, \qquad k = \ell + J, \ell + 2J, \dots, m.$$

We suppress the dependence on J in the notation $\overline{I}_k$ and $\overline{I}_{\epsilon k}$. Then we can write, following Comte and Hardouin, (1995a and 1995b),

$$\log \overline{I}_k = \log f_k + \log 2\pi \overline{I}_{\epsilon k} + \log(1 + F_k) + \log\left(1 + \frac{\delta_k}{H_k}\right), \tag{5.4}$$

where

$$\begin{aligned} F_k &= \frac{\sum_{j=1}^{J} I_{\epsilon,k+j-J}\,[f_{k+j-J} - f_k]}{\overline{I}_{\epsilon k}\, f_k} \\ H_k &= \sum_{j=1}^{J} 2\pi I_{\epsilon,k+j-J} f_{k+j-J} \\ \delta_k &= \overline{I}_k - H_k. \end{aligned}$$

We are interested in bound in probability the last two terms in equation (5.4), but first we prove a lemma that allows us to take logs of $\overline{I}_{\epsilon k}$ (and therefore of $\overline{I}_k$) and to divide by $\overline{I}_{\epsilon k}$.

Lemma 5.5 *Under Assumption 5.9, $J \geq 2$, $k \neq 0$,*

$$\overline{I}_{\epsilon k} > 0 \qquad w.p.1.$$

Lemma 5.6 *Under Assumptions 5.3, 5.4, 5.7 and 5.9,*

$$\log(1 + F_k) = O_P(k^{-1}), \qquad k = \ell + J, \ell + 2J, \dots, m.$$

Lemma 5.7 *Under Assumptions 5.3, 5.4, 5.7 and 5.9, $J \geq 2$,*

$$\log\left(1 + \frac{\delta_k}{H_k}\right) = O_P\left(\left[\frac{\log m}{k}\right]^{1/2}\right), \qquad k = \ell + J, \ell + 2J, \dots, m.$$

After these results, we are in conditions of proving the consistency of $\widehat{d}$ for non-Gaussian series under conditions 5.7 and 5.9. First, we introduce the following condition on the bandwidth numbers.

Assumption 5.10 *As $N \to \infty$,*

$$\frac{\ell\,(\log N)^2}{m} + \frac{\log m}{\ell} + \frac{\log m \, \log^2 N}{m} + \frac{m}{N} \to 0.$$

This assumption is almost minimal given the structure of the estimate $\widehat{d}$. Then our main result is

Theorem 5.4 *Under Assumptions 5.3, 5.4, 5.7, 5.9 and 5.10, with $J \geq 2$, $\widehat{d} \to_P d$.*

Again, the $\log N$-consistency of $\widehat{d}$ will follow under a mild condition: as $N \to \infty$,

$$\frac{\ell\,(\log N)^2}{m} + \frac{\log m}{\ell} + \frac{\log m \, \log^4 N}{m} + \left(\frac{m}{N}\right)^{\beta} \log N \to 0.$$

Note that we do not need to strengthen all the conditions in the same way, and that some are left unmodified. This is straightforward from the proof of Theorem 5.4. For example, $\log m/\ell \to 0$ is only used to obtain uniform bounds for the transfer function approximation, but it does not appear itself in the approximation of d by $\widehat{d}$.

To derive the asymptotic normality of the estimate is of evident interest. First, it is necessary to improve the approximation results between the periodogram ordinates of the observables and of the innovations, and obtain the asymptotic rate of the mean square error of $\widehat{d}$. Then, a central limit theorem has to be proved for the random variable ξ_N in the proof of Theorem 5.4.

5.6 Conclusions

Although we have given no asymptotic distribution for the semiparametric estimates of d considered, they are consistent under very mild conditions on both the distribution and dependence structures of the observed time series. Therefore, these estimates can be used in studentization problems and to initialize iterative procedures to approximate more efficient estimates. Further work could be done on some of the following points:

- Related arguments can be used for the analysis of both estimates in the antipersistent case $d \in \left(-\frac{1}{2}, 0\right)$, where $f(0) = 0$. This can arise in practice as a consequence of overdifferencing nonstationary time series.

- It is still unknown if the trimming, due to the non-standard behaviour of the periodogram for low frequencies, is necessary to achieve consistency of the estimates of d. In any case, it could be a good practice to skip in the regressions the very first frequencies to avoid serious bias problems.

- Straightforward modifications of the results presented here could be used to obtain the properties of the estimate of d proposed by Parzen (1986). His estimate is based as well on the logarithm of consistent estimates of the spectral density, and its asymptotic distribution can be deduced from Theorem 5.1.

- The asymptotic normality of $\widehat{f}_M$ could be obtained in the non-Gaussian case under a linear process condition where the innovations have bounded spectral densities of all orders, using related methods to those of Theorem 5.1.

5.7 Appendix: Proofs of Section 5.2

We will use the same definitions and properties as in Section 3.8.

Proof of Lemma 5.1. Writing

$$E[\widehat{f}_M(\lambda_j)] = \int_{-\pi}^{\pi} K_M(\lambda - \lambda_j) \int_{-\pi}^{\pi} \Phi_N^{(2)}(\theta) f(\lambda + \theta) d\theta d\lambda.$$

we have

$$\begin{aligned} E[\widehat{f}_M(\lambda_j)] - f(\lambda_j) &= \int_{-\pi}^{\pi} K_M(\lambda) \int_{-\pi}^{\pi} \Phi_N^{(2)}(\theta) \left[f(\lambda + \theta + \lambda_j) - f(\lambda + \lambda_j)\right] d\theta d\lambda \\ &\quad + \int_{-\pi}^{\pi} K_M(\lambda) \left[f(\lambda + \lambda_j) - f(\lambda_j)\right] d\lambda \\ &= b_1 + b_2 , \end{aligned}$$

say. We start bounding the integral

$$\sup_{|\lambda| \leq \pi/M} \int_{-\pi}^{\pi} \Phi_N^{(2)}(\theta) \left[f(\lambda + \theta + \lambda_j) - f(\lambda + \lambda_j)\right] d\theta \tag{5.5}$$

for values $|\lambda| \leq \pi/M$. We can now consider under the assumptions of the lemma a fixed value $\epsilon > 0$, small enough such that $f(\lambda)$ satisfies $f(\lambda) \leq C_\epsilon |\lambda|^{-2d}$ and $|f'(\lambda)| \leq C'_\epsilon |\lambda|^{-2d-1}$, for $\lambda \in [-\epsilon, \epsilon]$ and some constant $0 < C'_\epsilon < \infty$. Then, for N large enough $2|\lambda + \lambda_j| = 2[\lambda + \lambda_j] < \epsilon$ for $|\lambda| \leq \pi/M$ and using (5.2). Now we split the interval of integration in (5.5), proceeding as in the proofs of Theorems 5 to 8 in Robinson (1995b):

$$
\begin{aligned}
\left| \int_{-\pi}^{-\epsilon} + \int_{\epsilon}^{\pi} \right| &\leq \sup_{|\theta| \geq \epsilon} |\Phi_N^{(2)}(\theta)| \int_{\Pi} |f(\lambda + \lambda_j + \theta) - f(\lambda + \lambda_j)| d\theta \\
&= O\left(N^{-1} \left[1 + |\lambda + \lambda_j|^{-2d} \right] \right) \\
&= O\left(N^{-1} [\lambda + \lambda_j]^{-2d} \right),
\end{aligned}
$$

since $d \geq 0$. Next, as $1 - 2d \geq 0$ and $|\Phi_N^{(2)}(\theta)| \leq \text{const.} N^{-1} \theta^{-2}$,

$$
\begin{aligned}
\left| \int_{-\epsilon}^{-2|\lambda+\lambda_j|} \right| &\leq \left\{ \sup_{\theta \in [-\epsilon, -2|\lambda+\lambda_j|]} \frac{|f(\lambda + \lambda_j + \theta)|}{|\theta|^{(1-2d)/2}} \right\} \int_{2|\lambda+\lambda_j|}^{\pi} \theta^{(1-2d)/2} \Phi_N^{(2)}(\theta) d\theta \\
&\quad + f(\lambda + \lambda_j) \int_{2|\lambda+\lambda_j|}^{\pi} \Phi_N^{(2)}(\theta) d\theta \\
&= O\left(\left\{ |\lambda + \lambda_j|^{-d-1/2} \right\} \frac{1}{N} \int_{|\lambda+\lambda_j|}^{\infty} \theta^{-(3+2d)/2} d\theta + N^{-1} |\lambda + \lambda_j|^{-2d} \int_{|\lambda+\lambda_j|}^{\pi} \theta^{-2} d\theta \right) \\
&= O\left(|\lambda + \lambda_j|^{-(1+2d)} N^{-1} \right).
\end{aligned}
$$

Further, in the same way, se can get

$$
\left| \int_{|\lambda+\lambda_j|/2}^{\epsilon} \right| = O\left(|\lambda + \lambda_j|^{-(1+2d)} N^{-1} \right).
$$

Now

$$
\begin{aligned}
\left| \int_{-|\lambda+\lambda_j|/2}^{|\lambda+\lambda_j|/2} \right| &\leq \left\{ \sup_{\theta \in [-|\lambda+\lambda_j|/2, |\lambda+\lambda_j|/2]} |f'(\lambda + \lambda_j + \theta)| \right\} \int_{-|\lambda+\lambda_j|/2}^{|\lambda+\lambda_j|/2} |\theta| \Phi_N^{(2)}(\theta) d\theta \\
&= O\left(|\lambda + \lambda_j|^{-(1+2d)} N^{-1} \int_{-|\lambda+\lambda_j|/2}^{|\lambda+\lambda_j|/2} |\varphi_N(\theta)| d\theta \right) \\
&= O\left(|\lambda + \lambda_j|^{-(1+2d)} N^{-1} \log N \right).
\end{aligned}
$$

Finally,

$$
\begin{aligned}
&\left| \int_{-2|\lambda+\lambda_j|}^{-|\lambda+\lambda_j|/2} \right| \\
&\leq \left\{ \sup_{\theta \in [-2|\lambda+\lambda_j|, -|\lambda+\lambda_j|/2]} |\Phi_N^{(2)}(\theta)| \right\} \int_{-2|\lambda+\lambda_j|}^{-|\lambda+\lambda_j|/2} \{ |f(\lambda + \lambda_j + \theta)| + |f(\lambda + \lambda_j)| \} d\theta \\
&= O\left(|\lambda + \lambda_j|^{-2} N^{-1} \left[\int_{-2|\lambda+\lambda_j|}^{-|\lambda+\lambda_j|/2} (\lambda + \lambda_j + \theta)^{-2d} d\theta + |\lambda + \lambda_j|^{1-2d} \right] \right) \\
&= O\left(|\lambda + \lambda_j|^{-(1+2d)} N^{-1} \right).
\end{aligned}
$$

Now, for $|\lambda| \leq \pi/M$ and with (5.2)

$$\begin{aligned} \sup_{|\lambda|\leq\pi/M} |\lambda+\lambda_j|^{-1} &\leq \left|2\pi\frac{j}{N}-\frac{\pi}{M}\right|^{-1} = \left(2\pi\frac{j}{N}\right)^{-1}\left[1-\frac{N}{2jM}\right]^{-1} \\ &= \left(2\pi\frac{j}{N}\right)^{-1}[1-o(1)]^{-1} = \left(2\pi\frac{j}{N}\right)^{-1}[1+o(1)] = O\left(\frac{N}{j}\right). \end{aligned} \tag{5.6}$$

Then we have, using $\int |K_M(\lambda)|d\lambda < \infty$ and (5.6),

$$b_1 = O\left(N^{2d} j^{-(1+2d)} \log N\right).$$

Consider now b_2:

$$\begin{aligned} b_2 &\leq \sup_{|\lambda|\leq\pi/M} |f'(\lambda_j+\lambda)| \int_{|\lambda|\leq\pi/M} |\lambda||K_M(\lambda)|d\lambda \\ &= O\left(\sup_{|\lambda|\leq\pi/M} |\lambda_j+\lambda|^{-1-2d} M^{-1}\right) \\ &= O\left((N/j)^{1+2d} M^{-1}\right) \\ &= O\left((N/j)^{2d}\frac{N}{jM}\right). \end{aligned}$$

Then, using $f(\lambda_j) = O((j/N)^{-2d})$, the lemma follows. □

Prior to study the variance and asymptotic distribution of the estimates $\widehat{f}_M$ we obtain a general expression for its cumulants under the Gaussianity assumption. Let $X = (X_1, \ldots, X_N)'$ be the vector of N consecutive observations of X_t. Then X has a multivariate Normal distribution $\mathcal{N}(\underline{\mu}, \Sigma_N)$, where

$$[\Sigma_N]_{r,g} = \gamma(r-g), \quad r,g = 1,\ldots,N.$$

Then we can write,

$$\widehat{f}_M(\alpha) = \frac{1}{2\pi N} X' W_M(\alpha) X,$$

where $W_M(\alpha)$ is the matrix

$$\begin{aligned} [W_M(\alpha)]_{r,g} &= e^{i(r-g)\alpha} \int_\Pi K_M(\lambda) e^{i(r-g)\lambda} d\lambda \\ &= \cos(r-g)\alpha \int_\Pi K_M(\lambda)\cos(r-g)\lambda d\lambda, \end{aligned}$$

$r, g = 1, \ldots, N$. Therefore the characteristic function of $\widehat{f}_M(\alpha)$ is

$$\varphi(t) = \left|I - \frac{2it}{2\pi N}\Sigma_N W_M(\alpha)\right|^{-1/2}.$$

Now the s-th cumulant of $\widehat{f}_M(\alpha)$ is given by

$$\kappa_s = \frac{2^{s-1}(s-1)!}{(2\pi)^s N^s}\,\mathrm{Trace}[(\Sigma_N W_M(\alpha))^s].$$

We can write for the trace

$$\begin{aligned}
&\mathrm{Trace}\,[(\Sigma_N W_M(\alpha))^s] \\
=\ & \sum_{1\le r_1,\dots,r_{2s}\le N} \gamma(r_1-r_2)\omega(\frac{r_2-r_3}{M})\cos(r_2-r_3)\alpha\cdots\gamma(r_{2s-1}-r_{2s})\omega(\frac{r_{2s}-r_1}{M})\cos(r_{2s}-r_1)\alpha \\
=\ & \frac{1}{2^s}\sum_{1\le r_1,\dots,r_{2s}\le N}\int_{\Pi^{2s}} f(\alpha_1)K_M(\alpha_2)\cdots f(\alpha_{2s-1})K_M(\alpha_{2s}) \\
& \times[\exp\{i(r_2-r_3)\alpha\}+\exp\{-i(r_2-r_3)\alpha\}]\cdots[\exp\{i(r_{2s}-r_1)\alpha\}+\exp\{-i(r_{2s}-r_1)\alpha\}] \\
& \times\exp\{i\,[\alpha_1(r_1-r_2)+\alpha_2(r_2-r_3)+\cdots+\alpha_{2s}(r_{2s}-r_1)]\}\,d\alpha_1\cdots d\alpha_{2s} \\
=\ & \sum\nolimits^{\star}\frac{1}{2^s}\sum_{1\le r_1,\dots,r_{2s}\le N}\int_{\Pi^{2s}} f(\alpha_1)K_M(\alpha_2)\cdots f(\alpha_{2s-1})K_M(\alpha_{2s}) \\
& \times\exp\{i\,[r_1(\alpha_1-\alpha_{2s}+\alpha\delta(2s,1)) \\
&\quad +r_2(\alpha_2-\alpha_1-\alpha\delta(3,2)) \\
&\quad +r_3(\alpha_3-\alpha_2+\alpha\delta(3,2)) \\
&\quad \cdots \\
&\quad +r_{2s}(\alpha_{2s}-\alpha_{2s-1}-\alpha\delta(2s,1))]\}\,d\alpha_1\cdots\alpha_{2s} \\
=\ & (2\pi)^{2s-1}\frac{N}{2^s}\sum\nolimits^{\star}\int_{\Pi^{2s}} f(\lambda-\alpha\delta(2s,1)-\mu_2-\cdots-\mu_{2s}) \\
& \times K_M(\lambda-\alpha[\delta(2s,1)-\delta(3,2)]-\mu_3-\cdots-\mu_{2s}) \\
& \times\cdots f(\lambda-\alpha\delta(2s,1)-\mu_{2s})K_M(\lambda)\Phi_N^{(2s)}(\mu_1,\cdots,\mu_{2s})d\lambda d\mu_2\cdots d\mu_{2s}, \qquad (5.7)
\end{aligned}$$

where each $\delta(j+1,j)$ is equal to $+1$ or -1 and the summation $\sum^{\star}$ adds for all the 2^s possible combinations of the s different $\delta(\cdot,\cdot)$'s. In the last equality we have done this change of variable,

$$\begin{cases}
\mu_1 = \alpha_1-\alpha_{2s}+\alpha\delta(2s,1) \\
\mu_2 = \alpha_2-\alpha_1-\alpha\delta(3,2) \\
\mu_3 = \alpha_3-\alpha_2+\alpha\delta(3,2) \\
\cdots \\
\mu_{2s} = \alpha_{2s}-\alpha_{2s-1}-\alpha\delta(3,2),
\end{cases}$$

obtaining $\sum_{j=1}^{2s}\mu_j \equiv 0$, and setting $\lambda = \alpha_{2s}$, we can get

$$\begin{cases} \alpha_{2s-1} = \lambda - \mu_{2s} - \alpha\delta(2s,1) \\ \alpha_{2s-2} = \lambda - \mu_{2s} - \mu_{2s-1} - \alpha[\delta(2s,1) - \delta(3,2)] \\ \cdots \\ \alpha_1 = \lambda - \mu_{2s} - \cdots - \mu_2 - \alpha\delta(2s,1). \end{cases}$$

Proof of Lemma 5.2. Using (5.7) we can write the variance of $\widehat{f}_M(\lambda_j)$ as, $i \in \{1,2\}$,

$$\frac{2\pi}{2}\frac{1}{N}\sum\nolimits^{\star}\int_{\Pi^4} f(\lambda-\delta_1\lambda_j-\mu_2-\cdots-\mu_4)K_M(\lambda-\lambda_j[\delta_1-\delta_2]\mu_3-\mu_4)f(\lambda-\delta_2\lambda_j-\mu_4)K_M(\lambda)$$
$$\times\Phi_N^{(4)}(\mu_1,\cdots,\mu_4)d\lambda d\mu_2\cdots d\mu_4, \tag{5.8}$$

where the summation $\sum^{\star}$ runs for all the combinations $\delta_i \in \{-1,1\}$.

Let's analyze first the case where $\underline{\delta_1 = \delta_2 = \delta = \pm 1}$. We are interested in study the difference between the integral in (5.8) and

$$\int_{\Pi} f(\lambda - \delta\lambda_j)^2 K_M(\lambda)^2 d\lambda. \tag{5.9}$$

Consider a sequence $\rho_N \to 0$, $\rho_N = o(M^{-1})$, (to be chosen later) and the sets

$$D = \left\{\mu \in \mathcal{R}^3 : \sup_j|\mu_j| \le \rho_N\right\}$$

and D^c, its complementary in $\mathcal{R}^3$. Then we have that in the set D this difference is not greater than [we take $\delta = -1$ for simplification, w.l.o.g],

$$\int_{|\lambda|\le\frac{\pi}{M}}\int_D\left|f(\delta\lambda_j+\lambda-\mu_2\cdots-\mu_4)K_M(\lambda-\mu_3-\mu_4)f(\delta\lambda_j+\lambda-\mu_4)-f(\lambda+\delta\lambda_j)^2K_M(\lambda)\right|$$
$$\times|K_M(\lambda)||\Phi_N^{(4)}(\mu_1,\cdots,\mu_4)|d\lambda d\mu_2\cdots d\mu_4$$
$$\le \left\{\sup_{|\lambda|\le\pi/M}\sup_{\mu\in D}|f(\lambda+\lambda_j+\mu_2\cdots-\mu_4)|^2\right\}\left\{\sup_{\lambda}|K_M(\lambda)|\right\}$$
$$\times\sup_{\mu\in D}\int_{|\lambda|\le\pi/M}|K_M(\lambda-\mu_3-\mu_4)-K_M(\lambda)|\,d\lambda\int_{\Pi^3}|\Phi_N^{(4)}(\mu_1,\cdots,\mu_4)|d\mu_2\cdots d\mu_4 \tag{5.10}$$
$$+\left\{\sup_{|\lambda|\le\pi/M}\sup_{\mu\in D}|f(\lambda+\lambda_j+\mu_2\cdots-\mu_4)|\right\}\left\{\sup_{|\lambda|\le\pi/M}\sup_{\mu\in D}|f'(\lambda+\lambda_j+\mu_2\cdots-\mu_4)|\right\}$$
$$\times\left\{\sup_{|\lambda|}|K_M(\lambda)|\right\}\int|K_M(\lambda)|d\lambda\sum_{j=2}^{4}\int_{\Pi^3}|\mu_j||\Phi_N^{(4)}(\mu_1,\cdots,\mu_4)|d\mu_2\cdots d\mu_4. \tag{5.11}$$

Using Lemma 3.16 we have that (5.10) is

$$O\left(M^2\rho_N|\lambda+\lambda_j|^{-4d}\right) = O\left(M^2\rho_N(j/N)^{-4d}\right),$$

and (5.11) is

$$O\Big(MN^{-1}|\lambda+\lambda_j|^{-4d-1}\log^3 N\Big) = O\Big(M(j/N)^{-4d}j^{-1}\log^3 N\Big).$$

Now we consider the contribution from the set D^c. First we consider the contribution from the integral in (5.9)

$$\begin{aligned}
&\int_{|\lambda|\le\pi/M} f(\lambda+\delta\lambda_j)^2 K_M(\lambda)^2 d\lambda \int_{D^c} |\Phi_N^{(4)}(\mu_1,\cdots,\mu_4)| d\mu_2\cdots d\mu_4 \\
&= O\left(\sup_{|\lambda|\le\pi/M} f^2(\lambda+\lambda_j)\sup_\lambda |K_M(\lambda)| \int |K_M(\lambda)| d\lambda \,(N\rho_N)^{-1}\log^3 N\right) \\
&= O\Big(M(N\rho_N)^{-1}(j/N)^{-4d}\log^3 N\Big),
\end{aligned}$$

using the properties of the Fejér Kernel inside D^c.

Now we have that for the contribution of (5.8)

$$\begin{aligned}
&\int_{|\lambda|\le\pi/M}\int_{D^c} |f(\delta\lambda_j+\lambda-\mu_2-\cdots-\mu_4)K_M(\lambda-\mu_3-\mu_4)f(\delta\lambda_j+\lambda-\mu_4)K_M(\lambda)| \\
&\qquad \times|\Phi_N^{(4)}(\mu_1,\cdots,\mu_4)| d\lambda d\mu_2\cdots d\mu_4 \\
&\le \frac{1}{(2\pi)^3 N}\int_{D^\star} |f(\lambda_j+\alpha_1)K_M(\alpha_2)f(\lambda_j+\alpha_3)K_M(\alpha_4) \\
&\qquad \times \varphi_N(\alpha_1-\alpha_4)\varphi_N(\alpha_2-\alpha_1)\varphi_N(\alpha_3-\alpha_2)\varphi_N(\alpha_4-\alpha_3)|\, d\alpha_1\cdots\alpha_4,
\end{aligned} \tag{5.12}$$

where, as in the proof of Proposition 1, $D^\star$ is the correspondent set to D^c with the variables α_j, $j=1,\ldots,4$, i.e.

$$D^\star = \{|\alpha_2-\alpha_1|>\rho_N\}\cup\{|\alpha_3-\alpha_2|>\rho_N\}\cup\{|\alpha_4-\alpha_3|>\rho_N\},$$

and the last integral is only different from zero if

$$|\alpha_2|, |\alpha_4| \le \frac{\pi}{M}.$$

We are going to consider only the case where just one of the conditions that define the set $D^\star$ is satisfied, $\underline{|\alpha_2-\alpha_1|>\rho_N}$, say. Then we have,

$$N^{-1}|\varphi_N(\alpha_2-\alpha_1)| = O(\,(N\,\rho_N)^{-1}) \tag{5.13}$$

and, because $\sup|K_M(\lambda)=O(M)|$ and (3.35),

$$\int_{|\alpha_2|\le\pi/M} |\varphi_N(\alpha_3-\alpha_2)K_M(\alpha_2)| d\alpha_2 = O(M\log N). \tag{5.14}$$

Next, defining $h_N = (\lambda_j + \pi/M)/2$, we have $h_N = O(\lambda_j)$ and $h_N^{-1} = O(\lambda_j^{-1})$. Then we can write

$$\int_\Pi |\varphi_N(\alpha_1 - \alpha_4) f(\alpha_1 + \lambda_j)| d\alpha_1 = \int_{-\pi}^{-\epsilon} + \int_{-\epsilon}^{-4h_N} + \int_{-4h_N}^{-h_N} + \int_{-h_N}^{\epsilon} + \int_{\epsilon}^{\pi} \tag{5.15}$$

for ϵ fixed small enough in a way that $f(\lambda_j + \alpha_1)$ satisfies $f(\lambda_j + \alpha_1) \le C_\epsilon |\lambda_j + \alpha_1|^{-2d}$ for $\alpha_1 \in [-\epsilon, \epsilon]$. Now

$$\int_{|\alpha_1| > \epsilon} |\varphi_N| f \, d\alpha_1 = O(1),$$

since in this case $\sup_{|\alpha_1| \ge \epsilon} |\varphi_N(\alpha_1 + \lambda_j)| = O(1)$ as $\lambda_j \to 0$. Then,

$$\begin{aligned}
\int_{-\epsilon}^{-4h_N} |\varphi_N| f \, d\alpha_1 &\le \left\{ \sup_{-\epsilon \le \alpha_1 \le -4h_N} f(\alpha_1 + \lambda_j) \right\} \int_{-\epsilon}^{-4h_N} |\varphi_N(\alpha_1 - \alpha_4)| d\alpha_1 \\
&= O\left(\sup_{\alpha_1} |\alpha_1 + \lambda_j|^{-2d} \int_{-\pi}^{\pi} |\varphi_N(\alpha_1 - \alpha_4)| d\alpha_1 \right) \\
&= O\left(\lambda_j^{-2d} \log N \right) = O\left((j/N)^{-2d} \log N \right).
\end{aligned}$$

Next,

$$\begin{aligned}
\int_{-4h_N}^{-h_N} |\varphi_N| f \, d\alpha_1 &\le \sup_{\substack{-4h_N \le \alpha_1 \le -h_N \\ |\alpha_4| \le \pi/M}} |\varphi_N(\alpha_1 - \alpha_4)| \int_{-4h_N}^{-h_N} f(\alpha_1 + \lambda_j) d\alpha_1 \\
&= O\left(\sup_{\alpha_1, \alpha_4} |\alpha_1 - \alpha_4|^{-1} \int_{-4h_N}^{-h_N} |\alpha_1 + \lambda_j|^{-2d} d\alpha_1 \right) \\
&= O\left(\lambda_j^{-1} \, \lambda_j^{-2d+1} \right) = O\left((j/N)^{-2d} \right),
\end{aligned}$$

as $h_N = \lambda_j (1 + o(1))/2$, using (5.2). Also

$$\begin{aligned}
&\int_{-h_N}^{\epsilon} |\varphi_N| f \, d\alpha_1 \le \sup_{\alpha_1 > -h_N} f(\lambda_j + \alpha_1) \int_N |\varphi_N| d\alpha_1 \\
&= O((\lambda_j + \pi/M)^{-2d} \log N) = O((j/N)^{-2d} \log N).
\end{aligned}$$

Then (5.15) is

$$O((j/N)^{-2d} \log N). \tag{5.16}$$

The remaining integral in $D^\star$ is:

$$\int_\Pi \int_\Pi |K_M(\alpha_4) f(\lambda_j + \alpha_3) \varphi_N(\alpha_4 - \alpha_3)| \, d\alpha_3 d\alpha_4 = O((j/N)^{-2d} \log N), \tag{5.17}$$

reasoning for α_3 as for the integral in α_1 and using $\int |K_M(\alpha_4)| d\alpha_4 = O(1)$.

Then from (5.13) to (5.17) we have that (5.12) is

$$O\Big((\rho_N N)^{-1} M (j/N)^{-4d} \log^3 N\Big).$$

Compiling results we have that the difference between the two integrals in (5.8) and (5.9) is of order

$$O\Big(M^2 \rho_N (j/N)^{-4d} + M(j/N)^{-4d} j^{-1} \log^3 N + (N\,\rho_N)^{-1} M (j/N)^{-4d} \log^3 N\Big),$$

so the optimal choice is

$$\rho_N^2 = \frac{M \log^3 N}{N\,M^2} = \frac{\log^3 N}{N\,M},$$

which gives an error of

$$O\Big((j/N)^{-4d}\left[M^{3/2} N^{-1/2} \log^{3/2} N + M j^{-1} \log^3 N\right]\Big). \tag{5.18}$$

For this choice of ρ_N to be $o(M^{-1})$ we need

$$\lim_{N\to\infty} M\rho_N = \lim_{N\to\infty} \left(\frac{M}{N}\right)^{1/2} \log^{3/2} N = 0,$$

which is satisfied under the conditions of the lemma.

Now we have, from (5.9), operating as with b_1 in the proof of the previous lemma,

$$\begin{aligned}
&\left|\int_\Pi f(\lambda+\delta\lambda_j)^2 K_M(\lambda)^2 d\lambda \;-\; f(\delta\lambda_j)^2 \int_\Pi K_M(\lambda)^2 d\lambda\right| \\
\leq\;& 2 \sup_{|\lambda|\leq \pi/M} |f(\delta\lambda_j+\lambda) f'(\delta\lambda_j+\lambda)| \sup_\lambda |K_M(\lambda)| \int_\Pi |\lambda||K_M(\lambda)| d\lambda \\
=\;& O\left(\sup_{|\lambda|\leq\pi/M} |\lambda_j+\lambda|^{-1-4d}\right) = O\Big((j/N)^{-1-4d}\Big),
\end{aligned} \tag{5.19}$$

using $\int |\lambda||K_M| d\lambda = O(M^{-1})$.

Now, since $\int K_M(\lambda)^2 d\lambda = M\|K\|_2^2$, we have obtained the leading term for the variance, taking into account that in (5.8) the integral for $\delta = 1$ and $\delta = -1$ is multiplied by $2\pi/2N$ and f is even. The errors are obtained multiplying (5.18) and (5.19) by $(M\,f(\lambda_j)^2)^{-1} = O(M^{-1}(j/N)^{4d})$.

It only remains to check that we can neglect the contribution from the two terms in (5.8) with $\underline{\delta_1 = -\delta_2 = \delta = \pm 1}$. Proceeding in the same way as before we can see that these terms converge to the integral

$$\int_\Pi f(\lambda-\delta\lambda_j)^2 K_M(\lambda) K_M(\lambda - 2\delta\lambda_j) d\lambda, \tag{5.20}$$

with the same errors as before. In (5.20) we have two functions K_M centered in frequencies $2\lambda_j$ away. Since $j/N \to 0$ slower than $M^{-1} \to 0$, from (5.2), we have that the two kernels do not overlap and (5.20) is zero for N big enough. $\square$

Proof of Lemma 5.3. As before, we can express $\sqrt{N/M}\widetilde{f}(\lambda_j)$ as a quadratic form in the vector $X - \mathbf{1}\overline{X}$, where $\mathbf{1}$ is the $N \times 1$ vector $(1, \ldots, 1)'$:

$$\begin{aligned}
\sqrt{\frac{N}{M}}\frac{\widetilde{f}(\lambda_j)}{f(\lambda_j)} &= (X - \mathbf{1}\overline{X})'\left(\frac{1}{2\pi\sqrt{NM}f(\lambda_j)}W_M(\lambda_j)\right)(X - \mathbf{1}\overline{X}) \\
&= \sqrt{\frac{N}{M}}\frac{\widehat{f}(\lambda_j)}{f(\lambda_j)} - 2\frac{\overline{X}\,[X'W_N(\lambda_j)\mathbf{1}]}{2\pi\sqrt{MN}f(\lambda_j)} + \frac{\overline{X}^2\,[\mathbf{1}'W_N(\lambda_j)\mathbf{1}]}{2\pi\sqrt{MN}f(\lambda_j)} \\
&= \sqrt{\frac{N}{M}}\frac{\widehat{f}(\lambda_j)}{f(\lambda_j)} - 2\Delta_1(\lambda_j) + \Delta_2(\lambda_j),
\end{aligned}$$

say, using the same definitions as before. Let's study Δ_1 first. The variance of $X'W_N(\lambda_j)\mathbf{1}$ is the equal to

$$\begin{aligned}
&\mathbf{1}'W_M(\lambda_j)\Sigma_N W_M(\lambda_j)\mathbf{1} \\
&= \sum_{r_1\cdots r_4=1}^{N} \omega\left(\frac{r_1 - r_2}{M}\right)\gamma(r_2 - r_3)\,\omega\left(\frac{r_3 - r_4}{M}\right)\cos\lambda_j(r_1 - r_2)\cos\lambda_j(r_3 - r_4) \\
&= (2\pi)^3\frac{N}{4}\sum_{\delta}\int_{\Pi^3} K_M(\alpha_1)f(\alpha_2)K_M(\alpha_3) \\
&\quad \times\Phi_N^{(4)}(\alpha_1 + \lambda_j\delta_1, \alpha_2 - \alpha_1 - \lambda_j\delta_1, \alpha_3 - \alpha_2 + \lambda_j\delta_2, -\alpha_3 - \lambda_j\delta_2)d\lambda, \qquad (5.21)
\end{aligned}$$

where the summation runs for the all the combinations $\delta_j \in \{-1, 1\}$. Then making a change of variable, the modulus of (5.21) is not bigger than

$$\begin{aligned}
&(2\pi)^3\frac{N}{4}\sum_{\delta}\int_{\Pi^3}\left|K_M(\mu_1 - \lambda_j\delta_1)f(\mu_1 + \mu_2)K_M(\mu_1 + \mu_2 + \mu_3 - \lambda_j\delta_2)\Phi_N^{(4)}(\mu)\right| d\mu \\
\leq\ &\frac{1}{4}\sum_{\delta}\int_{\Pi^3}\Big|K_M(\mu_1 - \lambda_j\delta_1)f(\mu_1 + \mu_2)K_M(\mu_1 + \mu_2 + \mu_3 - \lambda_j\delta_2) \\
&\quad\times\varphi_N(\mu_1)\,\varphi_N(\mu_2)\,\varphi_N(\mu_3)\,\varphi_N(-\textstyle\sum_{j=1}^{3}\mu_j)\Big|\, d\mu. \qquad (5.22)
\end{aligned}$$

We consider the case $\delta_1 = \delta_2 = 1$. Now, given the compact support of K the previous expression is only different from zero if

$$\mu_1,\ \mu_1 + \mu_2 + \mu_3 \in \left[\lambda_j - \frac{\pi}{M}, \lambda_j + \frac{\pi}{M}\right],$$

with μ_1, $\mu_1 + \mu_2 + \mu_3 \sim \lambda_j$, given Assumption 5.5. Therefore, using the properties of φ_N,

$$\left|\varphi_N(\mu_1)\,\varphi_N(-\textstyle\sum_{j=1}^{3}\mu_j)\right| = O(\lambda_j^2),$$

and also,

$$|K_M(\mu_1 + \mu_2 + \mu_3 - \lambda_j)| = O(M).$$

Then,

$$\int_{\Pi} |\varphi_N(\mu_3)| d\mu_3 = O(\log N).$$

Now we claim that

$$\int_{\Pi} f(\mu_1 + \mu_2)|\varphi_N(\mu_2)| d\mu_2 = O(\lambda_j^{-2d} \log N), \tag{5.23}$$

and since

$$\int_{\Pi} |K_M(\mu_1 - \lambda_j)| d\mu_1 = O(1),$$

the variance of $X'W_N(\lambda_j)\mathbf{1}$ is

$$O\left(\lambda_j^{-2(1+d)} M \log^2 N\right).$$

Consequently, $\Delta_1(\lambda_j)$ is, using the rate of convergence of the variance of the sample mean,

$$O_P\left(\frac{\lambda_j^{2d}}{\sqrt{MN}} \lambda_j^{-1-d} \sqrt{M} \log N N^{d-1/2}\right) = O_P\left(j^{d-1} \log N\right) = o_P\left(j^{-1/2} \log N\right) = o_P(1),$$

since $d < 1/2$. Now we prove our claim in (5.23). Proceeding as in the proof of Lemma 5.1, we can split the integral in (5.23) in the following intervals, for some $\epsilon > 0$ fixed,

$$\int_{|\mu_2| > \epsilon} = O(1),$$

and since $\mu_1 \sim \lambda_j$,

$$\begin{aligned}
\int_{-\epsilon}^{-\lambda_j/2} &= O\left(\sup_{\mu_2 \in [-\epsilon, -\lambda_j/2]} |\varphi_N(\mu_2)| \int_{-\epsilon}^{-\lambda_j/2} |\mu_1 + \mu_2|^{-2d} d\mu_2\right) \\
&= O\left(\lambda_j^{-1} \sup_{\mu_2 \in [-\epsilon, -\lambda_j/2]} |\mu_1 + \mu_2|^{1-2d}\right) = O(\lambda_j^{-2d}).
\end{aligned}$$

Finally,

$$\int_{-\lambda_j/2}^{\epsilon} = O\left(\sup_{\mu_2 \in [-\lambda_j/2, \epsilon]} f(\mu_1 + \mu_2)| \int_{-\lambda_j/2}^{\epsilon} |\varphi_N(\mu_2)| d\mu_2\right) = O\left(\lambda_j^{-2d} \log N\right).$$

The analysis of $\Delta_2(\lambda_j)$ is simpler. We can see that

$$\mathbf{1}'W_M(\lambda_j)\mathbf{1} = O\left(\int_{\Pi} K_M(\lambda_j - \lambda)|\varphi_N(\lambda)|^2 d\lambda\right),$$

and this expression is only different from zero if $|\lambda_j - \lambda| \leq \pi/M$, so we have to consider only the values of λ such that $\lambda \sim \lambda_j$ given Assumption 5.5. Then $|\varphi_N(\lambda)|^2 = O(\lambda_j^2)$, and the same bound applies for the whole integral. Therefore, $\Delta_2(\lambda_j)$ is

$$O_P\left(N^{2d-1}\frac{1}{\sqrt{NM}}\lambda_j^{2(d-1)}\right) = O_P\left(j^{-2(1-d)}\frac{N^{1/2}}{M^{1/2}}\right) = O_P\left(\frac{N^{1/2}}{jM^{1/2}}\right) = o_P(1),$$

using $d < 1/2$. □

Proof of Theorem 5.1. From Lemma 5.2, it only remains to check that all the higher order normalized cumulants converge to zero. We shall prove that for $s = 3, 4, ..$

$$\begin{aligned}
&\text{Cumulant}_s\left\{\sqrt{\frac{N}{M}}\frac{\widehat{f}_M(\lambda_j)}{f(\lambda_j)}\right\} \\
&= \left(\frac{M}{N}\right)^{\frac{s-1}{2}}(s-1)(2\pi)^{s-1}\|K\|_s^s + O\left(\left(\frac{M}{N}\right)^{\frac{s-1}{2}}\left\{\left[\frac{\log^{2s-1}N}{NM}\right]^{1/2} + \frac{N}{jM} + \frac{\log^{2s-1}N}{j}\right\}\right) \\
&= \left(\frac{M}{N}\right)^{\frac{s-1}{2}}(s-1)(2\pi)^{s-1}\|K\|_s^s + o\left(\left(\frac{M}{N}\right)^{\frac{s-1}{2}}\right).
\end{aligned}$$

Then, the Theorem will follow because $M/N \to 0$ as $N \to \infty$.

Similarly as in the proof Lemma 5.2, we have to study the difference

$$\begin{aligned}
\frac{N}{f^s(\lambda_j)}\Bigg|\int_{\Pi^{2s}} & f(\lambda - \delta_1\lambda_j - \mu_2 - \cdots - \mu_{2s})K_M(\lambda - [\delta_1 - \delta_2]\lambda_j - \mu_3 - \cdots - \mu_{2s})\cdots \\
&\times f(\lambda - \delta_s\lambda_j - \mu_{2s})K_M(\lambda)\Phi_N^{(2s)}(\mu_1, \cdots, \mu_{2s})d\lambda d\mu_2\cdots d\mu_{2s} \\
&- 2\int_{\Pi} f^s(\lambda_j + \lambda)K_M^s(\lambda)d\lambda\Bigg| \qquad (5.24)
\end{aligned}$$

for all combinations $\delta_i \in \{1, -1\}$. The reason is because, as with the variance, only the situations where all the δ_i are equal in $\sum^\star$ contribute to the main term of the correspondent cumulant, since, otherwise, we have at least two kernels centered in two frequencies away $|2\lambda_j|$, which is bigger than the width of $K_M(\lambda)$.

Consider the set, for some sequence $\rho_N \to 0$, $\rho_N = o(M^{-1})$,

$$D = \left\{\mu \in \mathcal{R}^{2s-1} : \sup_j|\mu_j| \leq \rho_N\right\}.$$

Assume then that $\delta_1 = \delta_2 = \ldots = \delta_s = \pm 1$. Now the contribution to the difference in (5.24) of the set D is, for $|\lambda| \leq \pi/M$,

$$O(NM^{s-1}f^{-1}(\lambda_j))$$

$$\times \sum_{q=0}^{s-1} \int_\Pi \int_D |f(\pm\lambda_j + \lambda - \mu_{2+2q} \ldots - \mu_{2s}) - f(\pm\lambda_j + \lambda)| \left| K_M(\lambda) \Phi_N^{(2s)}(\mu) \right| d\lambda d\mu \qquad (5.25)$$

$$+ O(NM^{s-1}) \sum_{q=0}^{s-1} \int_\Pi \int_D |K_M(\lambda - \mu_{3+2q} \ldots - \mu_{2s}) - K_M(\lambda)| \left| \Phi_N^{(2s)}(\mu) \right| d\lambda d\mu_2 \cdots d\mu_{2s}, \qquad (5.26)$$

taking out $\sup_{\lambda,\mu \in D} f(\lambda - \sum_r \mu_r + \lambda_j) = O(f(\lambda_j))$. Then, applying the Mean Value Theorem and using (3.32) we obtain that (5.25) is

$$\begin{aligned} & O(NM^{s-1}(j/N)^{-1}) \int_\Pi |K_M(\lambda)| d\lambda \sum_{q=2}^{2s-1} \int_\Pi |\mu_q| \left| \Phi_N^{(2s)}(\mu) \right| d\mu_2 \cdots d\mu_{2s} \\ = \ & O(M^{s-1}(j/N)^{-1} \log^{2s-1} N). \end{aligned} \qquad (5.27)$$

On the other hand, applying Lemma 3.16 and using (3.29), (5.26) is of order

$$O(NM^s \delta_N). \qquad (5.28)$$

For the contribution of the complementary of the set D in Π^{2s-1}, denoted as D^c. The integral corresponding to the set D^c is then less or equal than

$$\frac{N}{f^s(\lambda_j)} \int_\Pi \int_{D^c} |f(\pm\lambda_j + \lambda - \mu_2 \cdots - \mu_{2s}) K_M(\lambda - \mu_3 \cdots - \mu_{2s}) \cdots f(\pm\lambda_j + \lambda - \mu_{2s}) K_M(\lambda)|$$

$$\times |\Phi_N^{(2s)}(\mu_1, \cdots, \mu_{2s})| d\lambda d\mu_2 \cdots d\mu_{2s} \qquad (5.29)$$

$$+ \ \frac{N}{f^s(\lambda_j)} \int_\Pi |f^s(\pm\lambda_j + \lambda) K_M^s(\lambda)| d\lambda \int_{D^c} |\Phi_N^{(2s)}(\mu_1, \cdots, \mu_{2s})| d\mu_2 \cdots d\mu_{2s}. \qquad (5.30)$$

The expression in (5.30) is $O(M^{s-1} \log^{2s-1} N \delta^{-1})$, using (3.29) and $f^{-s}(\lambda_j) \int |f^s(\pm\lambda_j + \lambda) K_M^s(\lambda)| d\lambda = O(M^{s-1})$, which follows from the compact support of K. Now for (5.29) we have

$$\begin{aligned} & \frac{N}{f^s(\lambda_j)} \int_\Pi \int_{D^c} |f(\pm\lambda_j + \lambda - \mu_2 - \cdots - \mu_{2s}) K_M(\lambda - \mu_3 - \cdots - \mu_{2s}) \cdots \\ & \qquad \times f(\pm\lambda_j + \lambda - \mu_{2s}) K_M(\lambda)| \, |\Phi_N^{(2s)}(\mu_1, \cdots, \mu_{2s})| d\lambda d\mu_2 \cdots d\mu_{2s} \\ \leq \ & \int_{D^\star} |f(\lambda_j + \alpha_1) K_M(\alpha_2) \cdots f(\lambda_j + \alpha_{2s-1}) K_M(\alpha_{2s}) \\ & \qquad \times \varphi_N(\alpha_1 - \alpha_{2s}) \varphi_N(\alpha_2 - \alpha_1) \cdots \varphi_N(\alpha_{2s} - \alpha_{2s-1})| \, d\alpha_1 \cdots \alpha_{2s}, \end{aligned}$$

where $D^\star$ is the correspondent set to D^c with the old variables α_j, $j = 1, \ldots, 2s$, i.e.

$$D^\star = \{|\alpha_2 - \alpha_1| > \delta_N\} \cup \{|\alpha_3 - \alpha_2| > \delta_N\} \cup \ldots \cup \{|\alpha_{2s} - \alpha_{2s-1}| > \delta_N\}.$$

Also the last integral is only different from zero if

$$|\alpha_2|, |\alpha_4|, \ldots, |\alpha_{2s}| \leq \frac{\pi}{M}.$$

We are going to consider only the case where just one of the events in D^* is satisfied, $|\alpha_{2j} - \alpha_{2j-1}| > \delta_N$ $(1 \le j \le s)$, say, the situation with an odd index or with more than one event is dealt with in a similar or simpler way.

First, if $|\alpha_{2j} - \alpha_{2j-1}| > \delta_N$, then $|\varphi_N(\alpha_{2j} - \alpha_{2j-1})| = O(\delta_N^{-1})$. Second, we can bound the integrals in α_{2j} and α_{2j-1} in this way,

$$\int_\Pi |\varphi_N(\alpha_{2j+1} - \alpha_{2j}) K_M(\alpha_{2j})| \, d\alpha_{2j} = O(M \log N),$$

using (3.35) and

$$\int_\Pi |\varphi_N(\alpha_{2j-1} - \alpha_{2j-2}) f(\lambda_j + \alpha_{2j-1})| \, d\alpha_{2j-1} = O((j/N))^{-2d} \log N), \tag{5.31}$$

as with (5.15). There are $s-1$ integrals of each type, which can be handled in the same way. The remaining integral is of this general form:

$$\int_\Pi \int_\Pi |K_M(\alpha_{2s}) f(\pm\lambda_j + \alpha_1) \varphi_N(\alpha_1 - \alpha_{2s})| \, d\alpha_1 d\alpha_{2s} = O((j/N)^{-2d} \log N),$$

as we did before for (5.17).

Finally the contribution from the integral over D^c is of order

$$O\left(\rho_N^{-1} M^{s-1} \log^{2s-1} N\right), \tag{5.32}$$

and therefore, the optimal choice for ρ_N is,

$$\rho_N^2 = \frac{\log^{2s-1} N}{N M}.$$

From (5.27), (5.28) and (5.32), we can see that (5.24) is

$$O\left(N M^{s-1} \left[\frac{\log^{2s-1} N}{N M} \right]^{1/2} \right).$$

Then, since

$$2N \int_\Pi f^s(\lambda_j + \lambda) K_M^s(\lambda) d\lambda = 2N f^s(\lambda_j) \int_\Pi K_M^s(\lambda) d\lambda + O\left(f^s(\lambda_j)(j/N)^{-1} N M^{s-2}\right)$$

the estimate for the cumulant follows and the Theorem is proved. □

5.8 Appendix: Proofs of Sections 5.3 and 5.4

Proof of Theorem 5.2. First, write f_k and $\widehat{f}_k$ for $f(\lambda_k)$ and $\widehat{f}_M(\lambda_k)$, respectively. Using the results for the bias and the variance of $\widehat{f}_k$, we have,

$$\begin{aligned}\frac{\widehat{f}_k}{f_k} &= 1+O\left(\frac{\log N}{k}+\frac{N}{k\,M}\right)+O_P\left(\left[\frac{M}{N}\right]^{1/2}\right), \qquad k=\ell+1,\ell+2,\ldots,m,\\ &= 1+O\left(\frac{\log N}{\ell}+\frac{N}{\ell\,M}\right)+O_P\left(\left[\frac{M}{N}\right]^{1/2}\right) = 1+o_P(1),\end{aligned}$$

uniformly in k, under Assumption 5.6, because $N/M\,\ell{+}M\,\log^3 N/N \to 0$ imply $\log N/\ell \to 0$. Then we can write as $N\to\infty$, using $|\log(1+x)|\le 2|x|$ for $|x|\le 1/2$,

$$\log\widehat{f}_k = \log f_k + O_P\left(\frac{\log N}{k}+\frac{N}{k\,M}+\left[\frac{M}{N}\right]^{1/2}\right).$$

From Robinson (1995b) and the definition for the summation in k, we can obtain

$$\begin{aligned}&\sum_k \Lambda_k^2 = 4\,m(1+o(1)),\\ &\sup_{\ell\le k\le m}|\Lambda_k| = O(\log N),\\ &\sum_k |\Lambda_k|^p = O(m), \quad p\ge 1.\end{aligned} \tag{5.33}$$

Then we have

$$\begin{aligned}\widehat{d}_M &= \left(\sum_k \Lambda_k^2\right)^{-1}\left[\sum_k \Lambda_k\log f_k+\sum_k \Lambda_k O_P\left(\frac{\log N}{k}+\frac{N}{k\,M}+\left[\frac{M}{N}\right]^{1/2}\right)\right]\\ &= A_1\,+\,A_2,\end{aligned}$$

say. First we have, as $\sum_{k=1}^m k^{-1}=O(\log m)$,

$$A_2 = O_P\left(\frac{\log^2 N\,\log m}{m}+\frac{N\,\log N\,\log m}{m\,M}+\left[\frac{M}{N}\right]^{1/2}\right) = o_P(1), \tag{5.34}$$

using Assumption 5.6, since it follows that $\log^2 N\,\log m/m \to 0$ from the first three conditions on Assumption 5.6.

Next, with Assumption 5.3, $k=\ell+1,\ell+2,\ldots,m$

$$\log f_k - \log G\lambda_k^{-2d} = O(\lambda_k^{\beta}) = O\left([m/N]^{\beta}\right),$$

uniformly in k. Since $\sum_k \Lambda_k \log G\lambda_k^{-2d} = d\sum_k \Lambda_k^2$, by the definition of Λ_k, we have that

$$A_1 = d+O\left(\left[\frac{m}{N}\right]^{\beta}\right) = d+o(1), \tag{5.35}$$

from the last condition in Assumption 5.6, and the Theorem follows from (5.35) and (5.34). □

For the estimation of the variance of $\widehat{f}_M$ under a linear process condition we state two lemmas needed later:

Lemma 5.8 (Hosoya and Taniguchi, Lemma A2.1, 1982) *Under Assumptions 5.7 and 5.8, the process X_t has fourth order spectral density, denoted by $f_4^X(\nu_1,\nu_2,\nu_3)$, satisfying*

$$f_4^X(\nu_1,\nu_2,\nu_3) = w(\nu_1+\nu_2+\nu_3)w(-\nu_1)w(-\nu_2)w(-\nu_3)f_4^\epsilon(\nu_1+\nu_2+\nu_3,\nu_2,\nu_3) \quad a.e.$$

Lemma 5.9 (Bentkus, (8.1) and (8.2), 1976) *Under Assumptions 5.1, 5.7 and 5.8,*

$$Cov\left[\widehat{f}_M(\alpha_a),\widehat{f}_M(\alpha_b)\right] = \frac{2\pi}{N}\int_{\Pi^3} G(\mu_1,\mu_2,\mu_3)\Phi_N^{(4)}(\mu_1,\mu_2,\mu_3)d\mu,$$

where,

$$\begin{aligned} G(\mu) &= \int_{\Pi^2} K_M(\nu_1-\alpha_a)K_M(\nu_2+\alpha_b)f_4^X(\mu_1+\nu_1,\mu_2-\nu_1,\mu_3+\nu_2)d\nu_1 d\nu_2 && (5.36)\\ &\quad+\int_{\Pi} K_M(\nu-\alpha_a)K_M(\nu+\mu_1+\mu_2-\alpha_b)f(\mu_1+\nu)f(\mu_3-\nu)d\nu && (5.37)\\ &\quad+\int_{\Pi} K_M(\nu-\alpha_a)K_M(\nu+\mu_1+\mu_2+\alpha_b)f(\mu_1+\nu)f(\mu_3-\nu)d\nu. && (5.38) \end{aligned}$$

Proof of Lemma 5.4. Using Lemma 5.9 we can see that (5.37) and (5.38) are just the *Gaussian* terms of the variance (take $\alpha_a=\alpha_b=\lambda_j$), which give the result for the variance of $\widehat{f}_M$ when X_t is Gaussian. So we have to study the fourth cumulant contribution of (5.36). Now, employing Lemma 5.8, we can write the contribution from (5.36) to the variance of the spectral estimate as

$$\begin{aligned} &\frac{2\pi}{N}\int_{\Pi^2}\int_{\Pi^3} K_M(\nu_1)K_M(\nu_2)\,w(\mu_1+\mu_2+\mu_3+\nu_2-\lambda_j)\\ &\qquad\times w(-\mu_1-\nu_1-\lambda_j)\,w(-\mu_2+\nu_1+\lambda_j)\,w(-\mu_3-\nu_2+\lambda_j)\\ &\qquad\times f_4^\epsilon(\mu_1+\mu_2+\mu_3+\nu_2-\lambda_j,-\mu_1-\nu_1-\lambda_j,-\mu_2+\nu_1+\lambda_j)\Phi_N^{(4)}(\mu)d\mu\,d\nu_1\,d\nu_2\\ &= O\Big(\sup|f_4^\epsilon|\frac{1}{N}\int_{\Pi^2}\int_{\Pi^3}|K_M(\nu_1)K_M(\nu_2)|\,f^{1/2}(\mu_1+\mu_2+\mu_3+\nu_2-\lambda_j)\\ &\qquad\times f^{1/2}(-\mu_1-\nu_1-\lambda_j)f^{1/2}(-\mu_2+\nu_1+\lambda_j)f^{1/2}(-\mu_3-\nu_2+\lambda_j)\\ &\qquad\times|\Phi_N^{(4)}(\mu)|d\mu\,d\nu_1\,d\nu_2\Big), && (5.39) \end{aligned}$$

since $|w(\lambda)|^2 = f(\lambda)/f^\epsilon(\lambda) = O(f(\lambda))$ as $\lambda \to 0$ and f_4^ϵ is uniformly bounded in that interval. Define the set D by

$$D = \{\sup_j |\mu_j| \leq M^{-1}\}.$$

Then, taking into account that $|\nu_j| \leq \pi M^{-1}$, and that

$$\sup_D f^{1/2}(\mu_1 + \mu_2 + \mu_3 + \nu_2 - \lambda_j) = O(f^{1/2}(\lambda_j)),$$

and similarly for the other functions $f^{1/2}$ in (5.39), all with bound $O(f^{1/2}(\lambda_j))$ inside D, we can see that the integral in D in (5.39) is of order

$$O\Big(f^2(\lambda_j)N^{-1}\int_{\Pi^2}|K_M(\nu_1)K_M(\nu_2)|\,d\nu_1 d\nu_2\int_{\Pi^3}|\Phi_N^{(4)}(\mu)|d\mu\Big) = O\Big(f^2(\lambda_j)N^{-1}\Big), \quad (5.40)$$

because all the integrals are bounded, using the properties of the multiple Fejér Kernel $\Phi_N^{(4)}(\mu)$ and $\int|K_M(\lambda)|d\lambda < \infty$.

Now, for the contribution in the complementary of the set D in Π^3, D^c say, we can write the integral in (5.39) in the following way, up to finite constants,

$$\begin{aligned} N^{-2}\int_{\Pi^2}\int_{D^c} &|K_M(\nu_1)K_M(\nu_2)|\, f^{1/2}(\alpha_3 + \nu_2 - \lambda_j) \\ &\times f^{1/2}(-\alpha_1 - \nu_1 - \lambda_j)\, f^{1/2}(-\alpha_2 + \alpha_1 + \nu_1 + \lambda_j)\, f^{1/2}(-\alpha_3 + \alpha_2 - \nu_2 + \lambda_j) \\ &\times |\phi_N(\alpha_1)\,\phi_N(\alpha_2 - \alpha_1)\,\phi_N(\alpha_3 - \alpha_2)\,\phi_N(-\alpha_3)|\, d\alpha d\nu, \end{aligned} \quad (5.41)$$

where we have made the change of variable

$$\begin{cases} \alpha_1 = \mu_1 \\ \alpha_2 = \mu_1 + \mu_2 \\ \alpha_3 = \mu_1 + \mu_2 + \mu_3, \end{cases}$$

and then we have

$$D^c = \{|\alpha_1| > M^{-1}\} \cup \{|\alpha_2 - \alpha_1| > M^{-1}\} \cup \{|\alpha_3 - \alpha_2| > M^{-1}\}.$$

We are now going to bound all the integrals in (5.41) under the assumption that only one of the conditions that define D^c, $\underline{|\alpha_1| > M^{-1}}$, say, is satisfied. The procedure is exactly the same in the other situations. First we have

$$|\phi_N(\alpha_1)| = O(M). \quad (5.42)$$

Next, using the periodicity of f and ϕ_N, setting $x = -\alpha_2 + \alpha_1 + \nu_1$,

$$\begin{aligned}\int_\Pi f^{1/2}(-\alpha_2 + \alpha_1 + \nu_1 + \lambda_j)|\phi_N(\alpha_2 - \alpha_1)|d\alpha_1 &= \int_\Pi f^{1/2}(x + \lambda_j)|\phi_N(-x + \nu_1)|dx \\ &= O(f^{1/2}(\lambda_j)\log N), \qquad (5.43)\end{aligned}$$

proceeding in exactly the same way as in the bound for expression (5.15), replacing f by $f^{1/2}$, since $|\nu_1| \leq \pi/M$. Using the same argument we have that

$$\int_\Pi f^{1/2}(-\alpha_3 + \alpha_2 - \nu_2 + \lambda_j)|\phi_N(\alpha_3 - \alpha_2)|d\alpha_2 = O(f^{1/2}(\lambda_j)\log N), \quad (5.44)$$

$$\int_\Pi f^{1/2}(\alpha_3 + \nu_2 - \lambda_j)|\phi_N(-\alpha_3)|d\alpha_3 = O(f^{1/2}(\lambda_j)\log N). \quad (5.45)$$

Now the remaining term can be dealt with in this way:

$$\begin{aligned}&\int_{\Pi^2} |K_M(\nu_1)K_M(\nu_2)|\, f^{1/2}(-\alpha_1 - \nu_1 - \lambda_j)d\nu_1 d\nu_2 \\ &\leq \text{const.} \int_\Pi |K_M(\nu_1)|\, f^{1/2}(-\alpha_1 - \nu_1 - \lambda_j)d\nu_1 \\ &\leq \text{const.} \left[\int_\Pi |K_M(\nu_1)|^2 d\nu_1 \int_\Pi f(-\alpha_1 - \nu_1 - \lambda_j)d\nu_1\right]^{1/2} \\ &= O(M^{1/2}), \qquad (5.46)\end{aligned}$$

since $\sup|K_M(\lambda)| = O(M)$ and f is integrable by stationarity. From (5.42) to (5.46) we obtain that (5.41) is

$$O\Big(f^{3/2}(\lambda_j)N^{-2}M^{3/2}\log^3 N\Big). \quad (5.47)$$

Then, with the bounds in (5.40) and (5.47), we have that the contribution to the variance of $\widehat{f}_M(\lambda_j)/f(\lambda_j)$ from the fourth cumulant term (5.36) is

$$\begin{aligned}O\Big(N^{-1} + N^{-2}M^{3/2}\log^3 N\Big) &= O\Big(\frac{M}{N}\left[M^{-1} + N^{-1}M^{1/2}\log^3 N\right]\Big) \\ &= o\left(\frac{M}{N}\left\{\left[\frac{M}{N}\log^3 N\right]^{1/2} + j^{-1}\log^3 N + \frac{N}{jM}\right\}\right),\end{aligned}$$

under the condition $jN^{-1} + MN^{-1}\log^3 N \to 0$, and then the Lemma follows. □

5.9 Appendix: Proofs of Section 5.5

Lemma 5.10 (Chen and Hannan (1980)) *Under Assumption 5.9 , the distribution function Q_N of the vector*

$$W_N = N^{-1/2}\sum_{t=1}^{N} Y_t,$$

where

$$Y_t' = Y'(j(1),\ldots,j(k)) = \sqrt{2}\epsilon_t(\cos t\lambda_{j(1)}, \sin t\lambda_{j(1)},\ldots,\cos t\lambda_{j(k)}, \sin t\lambda_{j(k)}),$$

has density q_N for all sufficiently large N and

$$\sup_{\mathbf{y}\in\mathcal{R}^k}(1+\|\mathbf{y}\|^4)\left|q_N(\mathbf{y}) - \sum_{r=0}^{1} N^{-r/2}P_r(-\phi:\overline{\chi}_{\nu,N})(\mathbf{y})\right| = O(N^{-1}), \tag{5.48}$$

where P_r are polynomials in the average of the joint cumulants of Y_t $(1\leq t\leq N)$ of order $\nu=(\nu_1,\ldots,\nu_{2k})$, $\overline{\chi}_{\nu,N}$, multiplied by the $2k$th multivariate Normal density ϕ and where $P_0(\mathbf{y}) = \phi(\mathbf{y})$.

This is a simplified version of Chen and Hannan's Lemma 2, where we only use the first two terms of an Edgeworth expansion for the density of the Fourier transform of ϵ_t, so only four bounded moments are required.

Lemma 5.11 *Under Assumption 5.9 , for $J\geq 2$,*

$$E\left[(\overline{I}_{\epsilon k})^{-1}\right] < \infty.$$

Proof of Lemma 5.11. First, from Lemma 5.10, $2\pi\overline{I}_{\epsilon k}$ has the density of a $\frac{1}{2}\chi^2_{2J}$ distribution with error (using only P_0) of order $O((1+\|\mathbf{y}\|^4)^{-1}N^{-1/2})$. Also the density of a χ^2_{2J} is

$$\phi_{\chi^2_{2J}}(x) = \frac{x^{J-1}\,e^{-x/2}}{(J-1)!\,2^J} \qquad 0\leq x<\infty.$$

Then it is clear that if $X\sim\chi^2_{2J}$ then $E[X^{-1}]<\infty$ if $J\geq 2$. Then we would need only to check that the error in the evaluation of the inverse moment of $\overline{I}_{\epsilon k}$ using the Lemma 5.10 is bounded. If we write

$$\overline{I}_{\epsilon k} = \sum_{j=1}^{J} I_\epsilon(\lambda_{k+j-J}) = \sum_{j=1}^{J}(y^2_{aj}+y^2_{bj})$$

we need

$$N^{-1/2}\int_{\mathcal{R}^{2J}}(1+\|\mathbf{y}\|^4)^{-1}\left(\sum(y^2_{aj}+y^2_{bj})\right)^{-1}d\mathbf{y} \;<\; \infty. \tag{5.49}$$

First, defining the sets $A=[-1,1]^{2J}$ and A^c its complementary in $\mathcal{R}^{2J}$,

$$\begin{aligned}\int_{\mathcal{R}^{2J}}(1+\|\mathbf{y}\|^4)^{-1}\left(\sum(y^2_{aj}+y^2_{bj})\right)^{-1}d\mathbf{y} \;&\leq\; \text{const.}\int_A\left(\sum(y^2_{aj}+y^2_{bj})\right)^{-1}d\mathbf{y}\\ &\quad+\text{const.}\int_{A^c}(1+\|\mathbf{y}\|^4)^{-1}d\mathbf{y}\\ &\leq\; \text{const.}\int_A\left(\sum(y^2_{aj}+y^2_{bj})\right)^{-1}d\mathbf{y}+\text{const.},\end{aligned}$$

since $(1+\|\mathbf{y}\|^4)^{-1}$ and $\left(\sum(y_{aj}^2+y_{bj}^2)\right)^{-1}$ are bounded from above in A and A^c, respectively. Next, to bound the remaining integral, if $\phi(.)$ denote the densities of the correspondent distributions, we have

$$\begin{aligned}
\infty > \int_0^\infty x^{-1}\phi_{\chi^2_{2J}}(x)dx &= \int_{\mathcal{R}^{2J}} \left(\sum\nolimits_j (y_{aj}^2+y_{bj}^2)\right)^{-1} \phi_{\mathcal{N}(0,I_{2J})}(\mathbf{y})d\mathbf{y} \\
&> \int_A \left(\sum(y_{aj}^2+y_{bj}^2)\right)^{-1} \phi_{\mathcal{N}(0,I_{2J})}(\mathbf{y})d\mathbf{y} \\
&\geq \text{const.} \int_A \left(\sum(y_{aj}^2+y_{bj}^2)\right)^{-1} d\mathbf{y},
\end{aligned}$$

as the Normal density is bounded from below in A. Hence, we have got

$$\int_A \left(\sum(y_{aj}^2+y_{bj}^2)\right)^{-1} d\mathbf{y} \leq \text{const.} < \infty.$$

Then, the l.h.s. of (5.49) is $O(N^{-1/2})$, and the Lemma follows.

An alternate way of checking that the error is actually $O(N^{-1/2})$ is bounding directly the integrals

$$\int_{\mathcal{R}^{2J}} (1+\|\mathbf{y}\|^4)^{-1} \left(\sum(y_{aj}^2+y_{bj}^2)\right)^{-1} d\mathbf{y} \leq \text{const.} \int_A \left(\sum(y_{aj}^2+y_{bj}^2)\right)^{-1} d\mathbf{y} + \text{const.}.$$

This can be done as follows. In parenthesis appears the code number of the integral used from Gradshteyn and Ryzhik (1980). In each step we denote as p the sum of the squares of the remaining variables with respect to we are not integrating. Then, using (3.241.4),

$$\int_0^\infty (x^2+p)^{-1}dx = \frac{\pi}{2}p^{-1/2},$$

and from (2.271.5)

$$\int_0^1 (x^2+p)^{-1/2}dx = \log(1+\sqrt{p+1}) - \frac{1}{2}\log p.$$

Now we have

$$\int_0^1 \left[\log(1+\sqrt{x+1}) - \frac{1}{2}\log x\right] dx < \infty,$$

and the global integral is bounded; to make this process we have needed 3 integrals: that is the reason why we need $J \geq 2$ to get at least 4 degrees of freedom. □

Proof of Lemma 5.5. Using Lemma 5.11 we have that, for any $\theta > 0$,

$$\begin{aligned}
P\{\overline{I}_{\epsilon j} = 0\} &\leq P\{\overline{I}_{\epsilon k} \leq N^{-\theta}\} \\
&\leq P\{\overline{I}_{\epsilon j}^{-1} \geq N^{\theta}\} \\
&= O(N^{-\theta}),
\end{aligned}$$

and the lemma follows from the Borel-Cantelli Lemma, choosing $\theta > 1$. □

Proof of Lemma 5.6. For $j = 1 + k - J, \ldots, k$ we have

$$\begin{aligned}
\max_j |f_j - f_k| &\leq \max_j \sup_{\lambda \in [\lambda_k, \lambda_j]} |f'(\lambda)| \, |\lambda_j - \lambda_k| \\
&= O(f_k \, \lambda_k^{-1} N^{-1}) \\
&= O(f_k \, k^{-1}).
\end{aligned}$$

Then we have, since $I_{\epsilon j} \geq 0$,

$$|F_k| \leq \frac{\max_j |f_j - f_k| \, \overline{I}_{\epsilon k}}{f_k \overline{I}_{\epsilon k}} = O(k^{-1}).$$

Then $F_k = O_P(\ell^{-1})$, uniformly in k, and using $|\log(1 + x)| \leq 2|x|$ for $|x| \leq 1/2$, we obtain as $\ell \to \infty$

$$|\log(1 + F_k)| \leq 2|F_k| = O_P(k^{-1}). \quad \square$$

Proof of Lemma 5.7. First, for summations running from $j = 1 + k - J$ to $j = k$,

$$\begin{aligned}
E[(H_k)^{-1}] &= E\left[\left(2\pi \sum\nolimits_j f_j I_{\epsilon j}\right)^{-1}\right] \\
&\leq (2\pi)^{-1} \left\{\max_j f_j^{-1}\right\} E\left[(\overline{I}_{\epsilon k})^{-1}\right] \\
&= O\left(f_k^{-1}\right),
\end{aligned}$$

using Lemma 5.11. Now, from Robinson (1995c, expression (3.17)),

$$\begin{aligned}
E[|\delta_k|] &= E\left[\left|\sum_j I_j - 2\pi I_{\epsilon j} f_j\right|\right] \\
&\leq \sum_j E\,|I_j - 2\pi I_{\epsilon j} f_j| \\
&= O\left(\max_j f_j \left[\frac{\log j}{j}\right]^{1/2}\right) \\
&= O\left(f_k \left[\frac{\log k}{k}\right]^{1/2}\right).
\end{aligned}$$

Then the result follows from $\delta_k = O_P(f_k \left[\frac{\log k}{k}\right]^{1/2})$ and $(H_k)^{-1} = O_P(f_k^{-1})$, and the same reasoning of the previous lemma, since $|\delta_k / H_k| = o_P(1)$, uniformly in k. □

Proof of Theorem 5.4. As we did before in the proof of Theorem 5.2, under Assumption 5.3,

$$\left(\sum_k \Lambda_k^2\right)^{-1} \sum_k \Lambda_k \log f_k = d + O\left(\left[\frac{m}{N}\right]^{\beta}\right). \tag{5.50}$$

Now, from Lemmas 5.6 and 5.7 we have that

$$\log \overline{I}_k = \log f_k + \log \overline{I}_{\epsilon k} + O_P\left(\left[\frac{\log m}{k}\right]^{1/2}\right). \tag{5.51}$$

Substituting in the definition of $\widehat{d}$ and using (5.50), $\sum_k \Lambda_k^2 = 4(m-\ell)/J + O(\log N \log m)$ and $\sup|\Lambda_k| = O(\log N)$,

$$\begin{aligned}
\widehat{d} &= \left(\sum_k \Lambda_k^2\right)^{-1}\left(\sum_k \Lambda_k \log \overline{I}_k\right)\\
&= \left(\sum_k \Lambda_k^2\right)^{-1}\left(\sum_k \Lambda_k \left[\log f_k + \log 2\pi \overline{I}_{\epsilon k} + O_P\left(\left[k^{-1}\log m\right]^{1/2}\right)\right]\right)\\
&= d + \left(\sum_k \Lambda_k^2\right)^{-1}\left(\sum_k \Lambda_k \log 2\pi \overline{I}_{\epsilon k}\right) + O_P\left(\frac{\log N}{m}(\log m)^{1/2}\sum_k k^{-1/2}\right)\\
&\quad + O([mN^{-1}]^{\beta})\\
&= d + \xi_N + O_P\left(\log N (\log m)^{1/2} m^{-1/2} + [mN^{-1}]^{\beta}\right), \quad \text{say,}\\
&= d + \xi_N + o_P(1),
\end{aligned}$$

where the last line follows from Assumption 5.10.

To prove the consistency of the estimate $\widehat{d}$ we only need to calculate the first two moments of

$$\xi_N = \left(\sum_k \Lambda_k^2\right)^{-1}\left(\sum_k \Lambda_k \log 2\pi \overline{I}_{\epsilon k}\right).$$

To evaluate the moments of $\overline{I}_{\epsilon k}$, we approximate the density of the Fourier transform, $q_N(\mathbf{y})$, using Chen and Hannan's Lemma 5.10. This result uses some results in Bhattacharya and Rao (1975) to approximate the density of the Fourier transform $d_\epsilon(\lambda)$ of the sequence ϵ_t. They employed a finite fifth moment of ϵ_t to get a stronger result. For our purposes with Lemma 5.10 is enough.

Set $4\pi\overline{I}_{\epsilon k} = \sum_{j=1}^{J}(y_{aj}^2 + y_{bj}^2)$ were y_{aj} and y_{bj} correspond to the sine and cosine summations, respectively of $4\pi I_{\epsilon j}$. Now, from Chen and Hannan (1980), (see Lemma 5.10),

$$P_1(-\phi : \overline{\chi}_{\nu,N})(\mathbf{y}) = \sum_{|\nu|=3} \frac{\overline{\chi}_{\nu,N}}{v!} \prod_{j=1}^{2k}\left(\frac{\partial}{\partial y_i}\right)^{\nu_j} \phi(\mathbf{y}), \quad v! = \prod_j v_j!.$$

As $|\nu| = 3$ the terms in P_1 are of one of the following types when we are considering the joint distribution in $\mathcal{R}^{4J}$ of $4\pi\overline{I}_{\epsilon k}$ and $4\pi\overline{I}_{\epsilon k'}$, $k \neq k'$ (up to constants):

1. $H_3(y_s)\phi(\mathbf{y})$, where H_i are the Hermite polynomials of order i and $s \in \{1, \ldots, 4J\}$. Then this term is odd in the component y_s of $\mathbf{y}$ (since H_3 is odd and ϕ is even).

2. $H_2(y_s)H_1(y_r)\phi(\mathbf{y})$, $y_r \neq y_s$ and $r, s \in \{1, \ldots, 4J\}$. Then this term is odd in the component y_r.

3. $H_1(y_s)H_1(y_r)H_1(y_u)\phi(\mathbf{y})$, y_r, y_s, y_u all different. Then this term is odd in the components y_s, y_r and y_u.

If $k = k'$, we consider only a distribution in $\mathcal{R}^{2J}$ and the typical terms of P_1 are:

1. $H_3(y_s)\phi(\mathbf{y})$, where $s \in \{1, \ldots, 2J\}$. Then this term is odd in the component y_s of $\mathbf{y}$.

2. $H_2(y_s)H_1(y_r)\phi(\mathbf{y})$, $r \neq s$ and $r, s \in \{1, \ldots, 2\}$. Then this term is odd in the component y_r.

Then we have

$$\begin{aligned}
E\left[\log 2\pi\overline{I}_{\epsilon k}\right] + \log 2 &= \int_{\mathcal{R}^{2J}} \log\left(\sum\nolimits_j (y_{aj}^2 + y_{bj}^2)\right) q_N(\mathbf{y})d\mathbf{y} \\
&= \int_{\mathcal{R}^{2J}} \log\left(\sum\nolimits_j (y_{aj}^2 + y_{bj}^2)\right) \left[\phi(\mathbf{y}) + N^{-1/2}P_1(\mathbf{y})\right] d\mathbf{y} + O(N^{-1}) \\
&= \psi(J) + \log 2 + O(\frac{1}{N}),
\end{aligned}$$

since $\int_0^\infty (\log x)^h/(1+x^4)dx < \infty$ and $\int_0^\infty (x\log x)^h e^{-x}dx < \infty$, for all $h \geq 0$. $\psi(z) = d/dz \log\Gamma[z]$ is the Psi function. The contribution from $P_1(\mathbf{y})$ is 0 since the interval of integration is $(-\infty, \infty)$ and P_1 is always odd in one component of $\mathbf{y}$ and the log term is even in all the components.

Consider now the Covariance terms. Denote $E_k = E[\log 2\pi\overline{I}_{\epsilon k}]$. Then $(k \neq k')$

$$\begin{aligned}
&\mathrm{Cov}\left[\log 2\pi\overline{I}_{\epsilon k}, \log 2\pi\overline{I}_{\epsilon k'}\right] \\
&= \int_{\mathcal{R}^{4J}} \left[\log\left(\sum\nolimits_j (y_{aj}^2 + y_{bj}^2)\right) - E_k\right]\left[\log\left(\sum\nolimits_{j'} (y_{aj'}^2 + y_{bj'}^2)\right) - E_{k'}\right] q_N(\mathbf{y})d\mathbf{y} \\
&= \int_{\mathcal{R}^{4J}} \left[\log\left(\sum\nolimits_j (y_{aj}^2 + y_{bj}^2)\right) - E_k\right]\left[\log\left(\sum\nolimits_{j'} (y_{aj'}^2 + y_{bj'}^2)\right) - E_{k'}\right]
\end{aligned}$$

$$
\begin{aligned}
&\times \left[\phi(\mathbf{y}) + N^{-1/2}P_1(\mathbf{y})\right] d\mathbf{y} + O(N^{-1}) \\
&= N^{-1/2}\int_{\mathcal{R}^{4J}}\left[\log\left(\sum_j (y_{aj}^2 + y_{bj}^2)\right) - E_k\right]\left[\log\left(\sum_{j'} (y_{aj'}^2 + y_{bj'}^2)\right) - E_{k'}\right]P_1(\mathbf{y}) + O(\frac{1}{N}) \\
&= O(N^{-1}),
\end{aligned}
$$

as $\phi(\mathbf{y})$ is the density of the standard Normal density in $\mathcal{R}^{4J}$ (with uncorrelated components!), and since the contribution from P_1 cancel out by the same argument as before.

Now for the variance we have:

$$
\begin{aligned}
\mathrm{Var}\left[\log 2\pi \overline{I}_{\epsilon k}\right] &= \int_{\mathcal{R}^{2J}}\left[\log\left(\sum_j (y_{aj}^2 + y_{bj}^2)\right) - E_k\right]^2 q_N(\mathbf{y})d\mathbf{y} \\
&= \int_{\mathcal{R}^{2J}}\left[\log\left(\sum_j (y_{aj}^2 + y_{bj}^2)\right) - E_k\right]^2 \left[\phi(\mathbf{y}) + N^{-1/2}P_1(\mathbf{y})\right] d\mathbf{y} + O(\frac{1}{N}) \\
&= \psi'(J) + O(N^{-1}),
\end{aligned}
$$

reasoning as before.

Then it is immediate that

$$
E[\xi_N] = O(N^{-1}\log N),
$$

and that

$$
\mathrm{Var}[\xi_N] = \frac{J}{4m}\psi'(J) + o(m^{-1}) + O\left(\frac{1}{N}\right) \sim \frac{J}{4m}\psi'(J).
$$

Therefore $\xi_N = o_P(1)$ with Assumption 5.10 and the theorem is proved. □

Chapter 6

Conclusions

In this thesis we have studied different aspects of nonparametric estimation for time series analysis. Nonparametric techniques are relevant to avoid misspecification problems in the modelling of the serial dependence of the observations and are very appropriate for the design of autocorrelation robust techniques under mild conditions.

Generally, the properties of nonparametric statistics can only be analyzed by means of asymptotic techniques, as the sample size, or other index, increases. Due to the presence of a smoothing number, nonparametric estimates typically present slower rates of convergence than parametric competitors, being inefficient compared to them.

Given the previous limitations, there is a special interest on assessing how well asymptotic properties describe the actual performance of nonparametric techniques in finite samples. Also, it is necessary the construction of objective means of selection of the bandwidth number and to find alternatives for situations where these methods are no longer consistent, like with long range dependence time series models.

In the first part of this thesis we have made extensive use of higher order asymptotic theory to analyze nonparametric variance estimates of least squares estimates of linear models. Variance estimates are required to carry out valid inferences based on central limit theorems, and nonparametric ones are predominant among practitioners in the last years.

We have obtained Edgeworth expansions for the distribution of nonparametric variance estimates of least squares estimates in linear regression under wide conditions on the nonstochastic regressors. Then we have shown that, although the asymptotic distribution of the least squares estimates is not affected by the nonparametric studentization, higher order terms are. However, when we have stationary regressors, the series expansions are constructed only in terms of the sample size. This parallels the well-known fact that the rate of convergence of nonparametric variance estimates in those situations is still square root of N, not being affected by the lag number M. Nevertheless, the bias is always affected by the degree of smoothing, so the problem of bandwidth choice in that situation is still open.

Then we have concentrated in the location problem. The variance of the sample mean of weak dependence observations is proportional to the spectral density at the origin, so we have to consider nonparametric spectrum estimates at this point. Since here only the behaviour of the time series at a single frequency is relevant, we have found more natural to impose conditions on the serial dependence in the frequency domain. We have developed Edgeworth approximations for both the spectral density estimate and the studentized sample mean, extending previous results. We have proposed higher order corrections in terms of the bandwidth number used in the nonparametric estimation and shown how to approximate them using nonparametric estimates of the derivatives of the spectral density.

The performance of nonparametric estimates of the spectral density relies on the choice of the smoothing number. In Chapter 4 we have suggested an automatic method for such choice when we are only interested in a particular frequency, like zero frequency in the case of the studentization of the sample mean. Again, we have concentrated on local conditions on the spectral density to modify a cross validation procedure in the frequency domain. In a limited Monte Carlo experiment we found that our proposal captures the features of the spectrum we are mostly interested in, so it can improve on global choice methods.

Finally, we have considered time series inference problems in a different set-up. We have been assuming that the observed time series satisfied weak dependence conditions. However, for long range dependent time series these conditions do not hold and consequently most of the techniques analyzed before are no longer appropriate. The main interest is now centred on the estimation of the memory parameter. Semiparametric procedures are again pertinent on robustness grounds, and one of the most popular in practice is the log-periodogram regression.

Inference using the log-periodogram estimate of the long range dependence parameter has only been justified rigorously for Gaussian sequences. We have shown that this estimate is consistent under much broader conditions if we make a pooling of contributions from adjacent periodogram ordinates as it had been proposed earlier in the literature. This can be done with finite averages of periodogram ordinates or with consistent estimates of the spectral density similar to those considered under weak dependence conditions.

Bibliography

[1] Adenstedt, R.K. (1974). On large-sample estimation for the mean of a stationary random sequence. Annals of Statistics **2**, 1095-1107.

[2] Akahira, M. and K. Takeuchi (1981). Asymptotic Efficiency of Statistical Estimators: Concepts and Higher Order Asymptotic Efficiency. Springer-Verlag, New York.

[3] Albers, W. (1978). Testing the mean of a normal population under dependence. Annals of Statistics **6**, 817-858.

[4] Anderson, T.W. (1958). *An Introduction to Multivariate Statistical Analysis.* Wiley, New York.

[5] Anderson, T.W. (1971). *Statistical Analysis of Time Series.* Wiley, New York.

[6] Andrews, D.W.K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* **59**, 817-858.

[7] Barndorff-Nielsen, O. and D.R. Cox (1979). Edgeworth and saddle-point approximations with statistical applications. *Journal of the Royal Statistical Society, Series B,* **41**, 279-312.

[8] Barndorff-Nielsen, O. and D.R. Cox (1989). *Asymptotic Techniques for Use in Statistics.* Chapman and Hall, London.

[9] Beltrao, K.I. and P. Bloomfield (1987). Determining the bandwidth of a kernel spectrum estimate. *Journal of Time Series Analysis* **8**, 21-38.

[10] Bentkus, R.Y. (1972). On the error of the estimate of the spectral function of a stationary process. *Lithuanian Mathematical Journal* **12**, 55-71.

[11] Bentkus, R.Y. (1976). On semi-invariants of estimates of the spectrum of a stationary sequence. *Lithuanian Mathematical Journal* **16**, 37-61.

[12] Bentkus, R. (1985). On the mean square minimax asymptotic risk of the statistical estimators of the spectral density in the space L_2. *Lithuanian Mathematical Journal* **25**, 23-42.

[13] Bentkus, V., F. Götze and A. Tikhomirov (1995). Berry-Essen bounds for statistics of weakly dependent samples. Preprint University of Bielefeld.

[14] Bentkus, R.Y. and R.A. Rudzkis (1982). On the distribution of some statistical estimates of spectral density. *Theory of Probability and its Applications* **27**, 795-814.

[15] Beran, J. (1989). A test of location for data with slowing decaying serial correlation. *Biometrika* **76**, 261-269.

[16] Beran, J. (1994). Statistics for Long-Memory Processes. Chapman and Hall, London.

[17] Bertail, P., D.N. Politis and J.P. Romano (1995). On subsampling estimators with unknown rate of convergence. Preprint.

[18] Bhattacharya, R.N. (1987). Some aspects of Edgeworth expansions in statistics and probability. In *New Perspectives in Theoretical and Applied Statistics*, M.L. Puri, J. Pérez Vilaplana, and W. Wertz, eds., pp. 157-170. Wiley, New York.

[19] Bhattacharya, R.N. and J.K. Ghosh (1978). On the validity of the formal Edgeworth expansion. *Annals of Statistics* **6**, 434-451.

[20] Bhattacharya, R.N. and J.K. Ghosh (1989). On moment conditions for valid formal Edgeworth expansions. In *Multivariate Statistics and Probability: Essays in Memory of P.R. Krishnaiah,* C.R. Rao and M.M. Rao, eds., pp. 68-79. Academic, New York.

[21] Bhattacharya, R.N. and M. Qumsiyeh (1989). Second order L^p comparisons between the bootstrap and empirical Edgeworth expansion methodologies. *Annals of Statistics* **17**, 160-169.

[22] Bhattacharya, R.N. and R.R. Rao (1975). *Normal Approximation and Asymptotic Expansions.* Wiley, New York.

[23] Bickel, P.J. (1974). Edgeworth expansions in nonparametric statistics. *Annals of Statistics* **2**, 1-20.

[24] Bickel, P.J., F. Götze and W.R. van Zwet (1986). The Edgeworth expansion for U-statistics of degree two. *Annals of Statistics* **14**, 1463-1484.

[25] Bose, A. (1988). Higher order approximations for autocovariances from linear processes with applications. *Statistics* **19**, 259-269.

[26] Brillinger, D.R. (1973). Estimation of the mean of a stationary time series by sampling. *Journal of Applied Probability* **10**, 419-431.

[27] Brillinger, D.R. (1975). *Time Series, Data Analysis and Theory.* Holt, Rinehart and Winston, New York.

[28] Brillinger, D.R. (1979). Confidence intervals for the crosscovariance function. *Selecta Statistica Canadiana* **5**, 3-16.

[29] Brockmann, M., T. Gasser and E. Herrmann (1993). Locally adaptive bandwidth choice for kernel regression estimators. *Journal of the American Statistical Association* **88**, 1302-1309.

[30] Bülhmann, P. (1995). Locally adaptive lag-window spectral estimation. Technical Report, Statistics Department, University of California, Berkeley.

[31] Burg, J.P. (1975). Maximum entropy spectral analysis. Ph.D. dissertation, Stanford University, Stanford, California.

[32] Carlstein, E. (1986). The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *Annals of Statistics* **14**, 1171-1179.

[33] Chambers, J.M. (1967). On methods of asymptotic approximation for multivariate distributions. *Biometrica* **54**, 367-383.

[34] Chen, G., B. Abraham and S. Peiris (1994). Lag window estimation of the degree of differencing in fractionally integrated time series models. *Journal of Time Series Analysis* **15**, 473-487.

[35] Chen, Z.-G. and E.J. Hannan (1980). The distribution of periodogram ordinates. *Journal of Time Series Analysis* **1**, 73-82.

[36] Chibisov, D.M. (1972). An asymptotic expansion for the distribution of a statistic admitting an asymptotic expansion. *Theory of Probability and its Applications* **17**, 621-630.

[37] Cochrane, D. and G.H. Orcutt (1949). Application of least squares regression to relationships containing autocorrelated error terms. *Journal of the American Statistical Association* **44**, 32-61.

[38] Comte, F. and C. Hardouin (1995a). Regression on log-regularized periodogram under assumption on bounded spectral densities: the non fractional and the fractional cases. Working Paper *CREST*.

[39] Comte, F. and C. Hardouin (1995b). Regression on log-regularized periodogram for fractional models at low frequencies. Working Paper *CREST*.

[40] Dahlhaus, R. (1983). Spectral analysis with tapered data. *Journal of Time Series Analysis* **4**, 163-175.

[41] Dahlhaus, R. (1989). Efficient parameter estimation for self-similar processes. *Annals of Statistics* **17**, 1749-66

[42] Daniels, H.E. (1956). The approximate distribution of serial correlation coefficients. *Biometrika* **43**, 169-185.

[43] Durbin, J. (1980a). Approximations for densities of sufficient estimators. *Biometrika* **67**, 311-333.

[44] Durbin, J. (1980b). The approximate distribution of partial serial correlation coefficients calculated from residuals from regression on Fourier series. *Biometrika* **67**, 335-349.

[45] Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics* **7**, 1-26.

[46] Eicker, F. (1967). Limit theorems for regression with unequal and dependent errors. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 59-82, University of California Press, Berkeley, California.

[47] Feller, W. (1971). *An Introduction to Probability Theory and its Applications*, Vol. 2, 2nd Edition. Wiley, New York.

[48] Fox, R. and M.S. Taqqu (1986). Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Annals of Statistics* **14**, 517-532.

[49] Franke, J. and W. Härdle (1992). On bootstrapping kernel spectral estimates. *Annals of Statistics* **20**, 121-145.

[50] Gasser, T., H-G. Muller and V. Mammitzsch (1985). Kernels for Nonparametric Curve Estimation. *Journal of the Royal Statistical Society, Series B*, **47**, 238-252.

[51] Geweke, J. and S. Porter-Hudak (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis* **4**, 221-238.

[52] Götze, F. and C. Hipp (1983). Asymptotic expansions for sums of weakly dependent random vectors. *Z. Wahrsch. verw. Gebiete* **64**, 211-239.

[53] Götze, F. and C. Hipp (1994). Asymptotic distribution of statistics in time series. *Annals of Statistics* **22**, 2062-2088.

[54] Götze, F. and H.R. Künsch (1995). Second-order correctness of the blockwise bootstrap for stationary observations. Preprint.

[55] Gradshteyn, I.S. and I.M. Ryzhik (1980). *Table of integrals, series and products.* Academic Press, Orlando.

[56] Grenander, U. (1954). On the estimation of regression coefficients in the case of an autocorrelated disturbance. *Annals of Mathematical Statistics* **25**, 252-272.

[57] Grenander, U. and M. Rosenblatt (1957). *Statistical Analysis of Stationary Time Series.* Wiley, New York.

[58] Hall, P. (1991). Edgeworth expansions for nonparametric density estimators, with applications. *Statistics* **22**, 215-232.

[59] Hall, P. (1992). *The Bootstrap and Edgeworth Expansion.* Springer-Verlag, New York.

[60] Hannan, E.J. (1957). The variance of the mean of a stationary process. *Journal Royal Statistical Society, Series B*, **19**, 282-285.

[61] Hannan, E.J. (1958). The estimation of the spectral density after trend removal. *Journal of the Royal Statistical Society, Series B*, **20**, 323-333.

[62] Hannan, E.J. (1963a). Regression for time series with errors of measurement. *Biometrika* **50**, 293-302.

[63] Hannan, E.J. (1963b). Regression for time series. (Proc. Symp. on) *Time Series Analysis*, (ed. M. Rosenblatt), 17-37. Wiley, New York.

[64] Hannan, E.J. (1970). *Multiple Time Series.* Wiley, New York.

[65] Hannan, E.J. (1971). Non-linear time series regression. *Journal of Applied Probability* **8**, 767-780.

[66] Hannan, E.J. (1979). The central limit theorem for time series regression. *Stochastic Processes and their Applications* **9**, 281-289.

[67] Hannan, E.J. (1987). Rational transfer function approximation. *Statistical Science* **2**, 135-161.

[68] Hannan, E.J. and D.F. Nicholls (1977). The estimation of the prediction error variance. *Journal of the American Statistical Association* **72**, 834-840.

[69] Hannan, E.J. and P.M. Robinson (1973). Lagged regression with unknown lags. *Journal of the Royal Statistical Society, Series B,* **35**, 252-267.

[70] Hansen, L.P. (1982). Large sample properties of generalized method of moments estimates. *Econometrica* **50**, 1029-1059.

[71] Hassler, U. (1993). Regression of spectral estimators with fractionally integrated time series. *Journal of Time Series Analysis* **14**, 369-380.

[72] Hidalgo, J. (1994). Spectral analysis for bivariate time series with long memory. Preprint, London School of Economics.

[73] Hodges, J.L. Jr and E.L. Lehmann (1970). Deficiency. *Annals of Mathematical Statistics* **41**, 783-801.

[74] Hosoya, Y. and M. Taniguchi (1982). A central limit theorem for stationary processes and the parameter estimation of linear processes. *Annals of Statistics* **10**, 132-153.

[75] Hurvich, C.M. (1985). Data driven choice of a spectrum estimate: extending the applicability of cross-validation methods. *Journal of the American Statistical Association* **80**, 933-950.

[76] Hurvich, C.M. and K.I. Beltrao (1994). Automatic semiparametric estimation of the memory parameter of a long-memory time series. *Journal of Time Series Analysis* **15**, 285-302.

[77] Ibragimov I.A. and Y.V. Linnik (1971). *Independent and Stationary Sequences of Random Variables.* Wolters-Noordhoff, Groningen.

[78] Janas, D. (1994). Edgeworth expansions for spectral mean estimates with applications to Whittle estimates. *Annals of the Institute of Statistical Mathematics* **46**, 667-682.

[79] Janas, D. and R. von Sachs (1993). Consistency for non-linear functions of the periodogram of tapered data. *Journal of Time Series Analysis* **16**, 585-606.

[80] Jenkins, G.M. and D.G. Watts (1968). *Spectrum Analysis and its Applications.* Holden-Day, San Francisco.

[81] Jensen, J.L. (1995). *Saddlepoint Approximation.* Oxford Statistical Science Series - 16. Clarendon Press, Oxford.

[82] Jowet, G.H. (1954). The comparison of means of sets of observations from sections of independent stochastic series. *Journal Royal Statistical Society, Series B,* **17**, 208-227.

[83] Kakizawa, Y. (1993). Valid Edgeworth expansions for some estimators and Bootstrap confidence intervals in the first order autoregression. Preprint Osaka University.

[84] Kawata, T. (1972). *Fourier Analysis in Probability Theory.* Academic Press, New York.

[85] Keenan, D.M. (1986). Limiting behavior of functionals of higher-order sample cumulant spectra. *Annals of Statistics* **14**, 134-151.

[86] Kendall, M. G. and A. Stuart (1969). *The Advanced Theory of Statistics,* Volume I. Charles Griffin, London.

[87] Künsch, H.R. (1986). Discrimination between monotonic trends and long-range dependence. *Journal of Applied Probability* **23**, 1025-1030.

[88] Künsch, H.R. (1989). The jackknife and the bootstrap for general stationary observations. *Annals of Statistics* **17**, 1217-1241.

[89] Lahiri, S.N. (1991). Second order optimality of stationary bootstrap. *Statistics and Probability Letters* **11**, 335-341.

[90] Lahiri, S.N. (1993). Refinements in asymptotic expansions for sums of weakly dependent random vectors. *Annals of Probability* **21**, 791-799.

[91] Lahiri, S.N. (1994). On Edgeworth expansion and moving block bootstrap for studentized M-estimators in multiple linear regression models. Preprint, Iowa State University.

[92] Lepskii, O.V. (1991). Asymptotically minimax adaptive estimation I: Upper bounds. Optimally adaptive estimates. *Theory of Probability and its Applications* **36**, 682-697.

[93] Lepskii, O.V. and V.G. Spokoiny (1995). Local adaptivity to inhomogeneous smoothness.1. Resolution Level. Preprint.

[94] Linton, O.B. (1996). Second order approximation in a linear regression with heteroskedasticity of unknown form. *Econometric Reviews* **15**, 1-32.

[95] Lobato, I. and P.M. Robinson (1994). Averaged periodogram estimation of long memory. *Journal of Econometrics*, forthcoming.

[96] Lomnicki, Z.A. and S.K. Zaremba (1957). On estimating the spectral density of a stochastic process. *Journal of the Royal Statistical Society, Series B*, **19**, 13-37.

[97] Magee, L. (1989). An Edgeworth test size correction for the linear model with AR(1) errors. *Econometrica* **57**, 661-674.

[98] Newey, W.K. and K.D. West (1987). A simple positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* **55**, 703-708.

[99] Newey, W.K. and K.D. West (1994). Automatic lag selection in covariance matrix estimation. *Review of Economic Studies* **61**, 631-653.

[100] Parzen, E. (1962). Spectral analysis of asymptotically stationary time series. *Bulletin of the International Statistical Institute* **39**, 87-103.

[101] Parzen, E. (1969). Multiple time series modelling. In: P.R. Krishnaiah, ed., *Multivariate Analysis - II*, 383-509. Academic Press, New York.

[102] Parzen, E. (1986). Quantile spectral analysis and long memory time series. In *Essays in time series and allied processes: papers in honour of E.J. Hannan*, J. Gani and M.B. Priestley, Eds. Allied Probability Trust, Sheffield.

[103] Petrov, V.V. (1975). *Sums of Independent Random Variables.* Springer-Verlag, Berlin.

[104] Pfanzagl, J. (1973). Asymptotic expansions related to minimum contrast estimators. *Annals of Statistics* **1**, 993-1026.

[105] Phillips, P.C.B. (1977a). Approximations to some finite sample distributions associated with a first order stochastic difference equation. *Econometrica* **45**, 463-485; erratum, **50**, 274.

[106] Phillips, P.C.B. (1977b). A general theorem in the theory of asymptotic expansions as approximations to the finite sample distributions of econometric estimators. *Econometrica* **45**, 1517-1534.

[107] Phillips, P.C.B. (1980). Finite sample theory and the distributions of alternative estimators of the marginal propensity to consume. *Review of Economic Studies* **47**, 183-224.

[108] Politis, D.N. and J.P. Romano (1992). A general resampling scheme for triangular arrays of α-mixing random variables with application to the problem of spectral density estimation. *Annals of Statistics* **20**, 1985-2007.

[109] Politis, D.N. and J.P. Romano (1995). A general theory for large sample confidence regions based on subsamples under minimal assumptions. *Annals of Statistics*, forthcoming.

[110] Press, H. and J.W. Tukey (1956). Power spectral methods of analysis and their applications to problems in airplane dynamics *Bell Systems Monograph* 2606.

[111] Priestley, M.B. (1981). *Spectral Analysis and Time Series.* Academic Press, London.

[112] Qumsiyeh, M.B. (1990). Edgeworth expansion in regression models. *Journal of Multivariate Analysis* **35**, 86-101.

[113] Reid, N. (1988). Saddlepoint methods and statistical inference (with discussion). *Statistical Science* **3**, 213-238.

[114] Reid, N. (1991). Approximations and asymptotics. In *Statistical Theory and Modelling*, eds. D.V. Hinkley, N. Reid and E.J. Snell. Chapman and Hall, London.

[115] Reisen, V.A. (1994). Estimation of the fractional difference parameter in the ARIMA(p.d.q) model using the smoothed periodogram. *Journal of Time Series Analysis* **15**, 335-350.

[116] Robinson, P.M. (1972). Non-linear regression for multiple time-series. *Journal of Applied Probability* **9**, 758-768.

[117] Robinson, P.M. (1976). The estimation of linear differential equations with constant coefficients. *Econometrica* **44**, 751-764.

[118] Robinson, P.M. (1983). Review of various approaches to power spectrum estimation. In *Handbook of Statistics, Volume 3*, D.R. Brillinger and P.R. Krishnaiah eds., pp. 343-368. Elsevier Science Publishers, New York.

[119] Robinson, P.M. (1991). Automatic frequency domain inference on semiparametric and nonparametric models. *Econometrica* **59**, 1329-1363.

[120] Robinson, P.M. (1994a). Time series with strong dependence. In *Advances in Econometrics. Sixth World Congress, Vol. I*, C.A. Sims ed., pp. 47-95. Cambridge University Press, Cambridge.

[121] Robinson, P.M. (1994b). Semiparametric analysis of long-memory time series. *Annals of Statistics* **22**, 515-539.

[122] Robinson, P.M. (1995a). The normal approximation for semiparametric averaged derivatives. *Econometrica* **63**, 667-680.

[123] Robinson, P.M. (1995b). Log-periodogram regression of time series with long range dependence. *Annals of Statistics* **23**, 1048-1072.

[124] Robinson, P.M. (1995c). Gaussian semiparametric estimation of long range dependence. *Annals of Statistics* **23**, 1630-1661.

[125] Robinson, P.M. and C. Velasco (1996). Autocorrelation robust inference. Forthcoming in *Handbook of Statistics.*

[126] Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proceedings National Academy of Sciences USA* **42**, 4-47.

[127] Rothenberg, T. (1984a). Approximating the distributions of econometric estimators and test statistics. Ch. 14 in *Handbook of Econometrics, Vol. 2*, ed. by Z. Griliches and M. Intriligator. North Holland, New York.

[128] Rothenberg, T. (1984b). Approximate normality of generalized least squares estimates. *Econometrica* **52**, 811-825.

[129] Rudzkis, R. (1985). On the distribution of the maximum deviation of the Gaussian stationary time series spectral density estimate. *Lithuanian Mathematical Journal* **25**, 118-130.

[130] von Sachs, R. (1994a). Estimating non-linear functions of the spectral density, using a data-taper. *Annals of the Institute of Statistical Mathematics* **46**, 453-474.

[131] von Sachs, R. (1994b). Peak-insensitive nonparametric spectrum estimation. *Journal of Time Series Analysis* **15**, 429-452.

[132] Samarov, A.M. (1987). Robust spectral regression. *Annals of Statistics* **15**, 99-111.

[133] Samarov, A. and M.S. Taqqu (1988). On the efficiency of the sample mean in long memory noise. *Journal of Time Series Analysis* **9**, 191-200.

[134] Sargan, J.D. (1976). Econometric estimators and the Edgeworth approximation. *Econometrica* **4**, 421-448; erratum **45**, 272.

[135] Singh, K. (1981). On the asymptotic accuracy of Efron's bootstrap. *Annals of Statistics* **9**, 1187-1195.

[136] Song, W.T. and B.W. Schmeiser (1992). Variance of the sample mean: properties and graphs of quadratic-form estimators. *Operations Research* **41**, 501-517.

[137] Soulier, P. (1993). Pointwise estimation of the spectral density of a strongly dependent stationary Gaussian field. Preprint, University of Paris, Orsay Cedex.

[138] Taniguchi, M. (1979). On estimation of parameters of Gaussian stationary processes. *Journal of Applied Probability* **16**, 575-591.

[139] Taniguchi, M. (1983). On the second order asymptotic efficiency of estimators of Gaussian ARMA Processes. *Annals of Statistics* **11**, 157-169.

[140] Taniguchi, M. (1984). Validity of Edgeworth expansions for statistics of time series. *Journal of Time Series Analysis* **5**, 37-51.

[141] Taniguchi, M. (1986). Third order asymptotic properties of maximum likelihood estimator for Gaussian ARMA Processes. *Journal of Multivariate Analysis* **18**, 1-31.

[142] Taniguchi, M. (1987). Validity of Edgeworth expansions of minimum contrast estimators for Gaussian ARMA processes. *Journal of Multivariate Analysis* **21**, 1-28.

[143] Taniguchi, M. (1991). *Higher Order Asymptotic Theory for Time Series Analysis.* Lecture Notes in Statistics, 68. Springer-Verlag, Berlin.

[144] Toyooka, Y. (1986). Second-order risk structure of GLSE and MLE in a regression with a linear process. *Annals of Statistics* **14**, 1214-1225.

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

[146] Wallace, D.L. (1958). Asymptotic approximations to distributions. *Annals of Mathematical Statistics* **29**, 635-654.

[147] White, H. and I. Domowitz (1984). Nonlinear regression with dependent observations. *Econometrica* **52**, 143-161.

[148] Zhurbenko, I.G. (1984). On the efficiency of spectral density estimates of a stationary process. II. *Theory of Probability and its Applications* **29**, 409-419.

[149] Zygmund, A. (1977). *Trigonometric Series.* Cambridge University Press, Cambridge.