A Nonparametric Regression Spectrum: Estimation, Asymptotic Properties and Data Analysis

Dissertation

zur Erlangung des akademischen Grades
des Doktors der Naturwissenschaften (Dr. rer. nat.)
an der
Universität Konstanz
Mathematisch-Naturwissenschaftliche Sektion
Fachbereich Mathematik und Statistik

vorgelegt von

Mark Heiler

Tag der mündlichen Prüfung: 22.05.2007

Referent: Prof. Dr. Jan Beran, Universität Konstanz

Referent: Prof. Dr. Michael Junk, Universität Konstanz

Foreword

This work was financially supported by the research unit *Price-, Liquidity- and Credit Risk: Measurement and Allocation* funded by the German Research Foundation and the *Centre of Finance and Econometrics* at the University of Konstanz.
My special thanks go to my supervisor Prof. Dr. Jan Beran. I have greatly profited from his hints and suggestions, generously lavished during numerous conversations. I am very grateful to Prof. Dr. Michael Junk for refereeing my PhD thesis.

In particular, I would like to thank my family and Vera Niedermeyer whose support enabled a succesful completion of the present work.

Vorwort

Die vorliegende Arbeit wurde durch die DFG-Forschergruppe *Preis-, Liquiditäts- und Kreditrisiken: Messung und Verteilung* sowie durch das *Zentrum für Finanzen und Ökonometrie* an der Universität Konstanz finanziert.
Mein besonderer Dank gilt Herrn Prof. Dr. Jan Beran für die Betreuung meiner Dissertation, für vielfältige Ideen und Ratschläge sowie für zahlreiche anregende Gespräche. Herrn Professor Dr. Michael Junk danke ich für die Übernahme des Koreferats.

Bedanken möchte ich mich bei meiner Familie und Vera Niedermeyer, deren fortwährende Unterstützung wesentlich zum Gelingen dieser Arbeit beigetragen haben.

Abstract

Classical spectral analysis in statistics considers decomposition of stationary time series into sinusoidal components. The autocovariance and the spectrum are fundamental elements for analyzing a given time series both in time and frequency domain. However, in practice one frequently observes nonstationary time series. In order to apply spectral analysis to these processes, an extension of the classical spectral theory to more general situations is required.

This thesis investigates dependence structures in multivariate time series that are characterized by deterministic trends. Here, we extend the theory of stationary processes to deterministic nonparametric trend functions. In a nonparametric regression setting these functions are usually unknown and have to be estimated. Estimation of the trend function will be performed by applying wavelet thresholding, a simple but yet efficient way to recover a signal of unknown regularity from some noisy data.

Chapter 2 presents a review about wavelets and their use in statistics. This involves construction of compactly supported wavelet bases, wavelet transformation of a square integrable function and the application in linear and nonlinear function estimation. An extensive review of the literature on wavelet thresholding is presented and some asymptotic results are derived.

In chapter 3, we consider dependence structures in multivariate time series that are due to similarities in underlying deterministic trends. Results from spectral analysis for stationary processes are extended to deterministic trend

functions. A regression cross covariance and spectrum are defined. Estimation of these quantities is based on wavelet thresholding. An algorithm is presented that automatically estimates common frequency components and possible "lead-lag" effects in multivariate time series. A small simulation study illustrates the theoretical results.

Chapter 4 is concerned with a direct estimation of the regression spectrum based on the classical periodogram. We present a modification of the periodogram that results in consistent estimation of the regression spectrum. In contrast to spectral estimation of stationary time series no further smoothing of the periodogram over neighboring frequencies is required. Furthermore, asymptotic normality of the modified periodogram is proved and the results are compared to those of chapter 3.

Chapter 5 investigates lag estimation in time domain by analyzing the maximizing lag of the regression cross covariance. Asymptotic properties of the estimate are derived and illustrated with a small simulation study. This method proves to be particularly useful in situations where application of the cross spectrum seems to be inadequate, e.g. if the amount of common frequencies in a multivariate signal is small or the contribution of the noise relative to the signal in the time series is too large.

Several examples in biology, climatology and finance are given in chapter 6. They emphasize the practical importance of the present work for various fields of research. We finish this thesis with some concluding remarks in chapter 7.

Zusammenfassung

In der Statistik befasst sich die klassische Spektralanalyse mit der frequenzabhängigen Zerlegung von stationären Prozessen. Die Autokovarianzfunktion sowie das Spektrum sind wesentliche Elemente, um eine gegebene Zeitreihe sowohl im Zeit- als auch im Frequenzbereich zu analysieren. In praktischen Anwendung hingegen werden häufig nichtstationäre Zeitreihen beobachtet. Um die Methoden der Spektralschätzung auch für diese Fälle nutzbar zu machen, ist es notwendig, eine Verallgemeinerung der vorhandenen Theorie zu entwickeln.

In dieser Arbeit werden Abhängigkeitsstrukturen in multivariaten Zeitreihen betrachtet, die auf den Einfluss deterministischer Komponenten zurückzuführen sind. Wesentliches Ziel dabei ist es, die Theorie der klassischen Spektralschätzung auf nichtparametrische deterministische Trendfunktionen auszuweiten. In der nichtparametrischen Regression sind diese Funktionen normalerweise unbekannt und müssen geschätzt werden. Dafür wenden wir *Wavelet Thresholding* an, eine einfache und dennoch effiziente Methode, um ein Signal von unbekannter Regularität aus einer durch Störungen überlagerten Zeitreihe zu schätzen.

Kapitel 2 bietet einen Überblick über Wavelets und ihre Anwendungen in der Statistik. Dies beinhaltet die Konstruktion orthonormaler Basen von Waveletfunktionen mit kompaktem Träger, die Wavelettransformation einer quadratisch integrierbaren Funktion und die Anwendung von Wavelets in der

linearen und nichtlinearen Schätzung einer Mittelwertsfunktion. Im Anschluss diskutieren wir einige Fragen zu Konvergenzgeschwindigkeiten in der linearen und nichtlinearen Trendschätzung und geben einen umfangreichen Überblick über die Literatur im Bereich des *Wavelet Thresholding*.

Kapitel 3 befasst sich mit Abhängigkeitsstrukturen in multivariaten Zeitreihen, die auf die Existenz von deterministischen Trendkomponenten zurückzuführen sind. Dabei werden Resultate aus dem Bereich der Spektralanalyse für stationäre Zeitreihen auf die Theorie der deterministischen Funktionen ausgeweitet. Wir definieren eine Regressionskovarianz sowie ein Regressionsspektrum und schätzen diese Größen basierend auf Wavelets.

In Kapitel 4 befassen wir uns mit einer alternativen Schätzmethode für das Regressionsspektrum, die auf dem klassischen Periodogramm beruht. Wir definieren ein modifiziertes Periodogramm, dass im Unterschied zur Theorie stationärer Prozesse ohne eine Glättung über benachbarte Frequenzen eine konsistente Schätzung des Spektrums erlaubt. Eine Herleitung der asymptotischen Eigenschaften sowie ein Vergleich zur vorherigen Schätzmethode beschließen das Kapitel.

Kapitel 5 untersucht die Schätzung von zeitlichen Verschiebungen in multivariaten deterministischen Funktionen. Die Benutzung der Regressionskreuzkovarianz erweist sich dabei als hilfreiche Alternative zur Schätzung von Abhängigkeitsstrukturen, wenn z.B. durch eine zu geringe Anzahl von gemeinsamen Frequenzen im Signal die Verwendung des Spektrums als unmöglich erscheint.

Verschiedene Beispiele aus den Bereichen Biologie, Medizin, Klimaforschung sowie der Finanzwissenschaft werden in Kapitel 6 vorgestellt. Diese illustrieren die praktische Bedeutung und Anwendbarkeit der theoretisch entwickelten Methoden. Einige abschließende Bemerkungen finden sich in Kapitel 7.

Contents

List of Figures

Chapter 1

Introduction

Classical spectral analysis in statistics is concerned with stationary time series. The autocovariance and its spectral representation are common concepts to researchers for estimating dependence structures in multivariate time series both in time and frequency domain. However, due to changing means, variances or dependence structures one frequently observes time series whose properties deviate from the stationary case. Consider e.g. recordings of heart beat or encephalographic signals, weather conditions or unemployment rates. Due to seasonal effects, biological processes or business cycles, these time series are expected to contain deterministic (periodic) components such that the classical theory is not applicable.

In a nonparametric regression setting these time series are generally modeled in the form

$$Y(i) = f(t_i) + \epsilon(i), \tag{1.1}$$

where $(Y(1), \ldots, Y(n))^T$ is a vector of observations that is assumed to be additively composed of a deterministic component f, measured on an increasingly fine grid $t_i = \frac{i}{n}$ $(i = 1, \ldots, n)$, and a sequence of innovations $\epsilon(i)$.

If one assumes model (1.1) to hold, spectral properties of $Y(i)$ are determined

by the combined influence of the deterministic and stochastic component and a precise statistical analysis of the time series requires a modification of the classical spectral theory for stationary processes to the current situation.
In this thesis, we investigate dependence structures in multivariate time series that are due to similarities in underlying deterministic trends. We extend the theory of spectral analysis of stationary time series to deterministic *nonparametric* trend functions. This requires estimation of the unknown function f. Several different methods for estimating an unobservable f from some noisy data $Y(i)$ are discussed in literature. Linear methods involving Fourier techniques, splines, kernel estimates or local polynomials can be found in all fields of research. However, as will be presented in this article, linear estimators are of limited use because they achieve optimal rates of convergence to the true f only under very tight restrictions on the regularity of the estimand.
A relatively new nonlinear way of estimating deterministic functions of unknown regularity that is simple but yet efficient is given by wavelet thresholding. Wavelets constitute an orthonormal basis in $L^2(\mathbb{R})$ that are well-localized both in time and frequency. The wavelet representation of a function with respect to this basis is therefore given by a linear combination where each component incorporates information about certain time-frequency properties of f. A further advantage of applying wavelets is their parsimony in the number of used coefficients, i.e. very complicated functions may be described in a concise manner by using only a few wavelet coefficients. Especially when modeling discontinous functions, the local definition of wavelets allows for representing a jump by using only a few wavelet coefficients. In contrast, application of Fourier series expansions for modeling a discontinuous function involves a large amount of compensating frequencies in the neighborhood of the jump. Furthermore, numerical calculation of the wavelet representation is simple and fast. Mallat's pyramid algorithm calculates wavelet coefficients in $O(n)$ operations and thus allows for efficiently handling of large data sets. By replacing the coefficients in the wavelet representation by certain almost consistent es-

timates and by using a truncation for the infinite series of wavelet coefficients, we obtain a *wavelet thresholding estimate* of the function f. Various different types of thresholding estimates will be presented and theoretically discussed throughout the thesis. Applications of wavelets and wavelet thresholding are found in all areas of research including physics, engineering, medicine, biology and economics (see e.g. Aldroubi and Unser (1996), Capobianco (1999), Ramsey (2002), Ramsey and Lampart (1998) or Prokoph and Barthelmes (1996) amongst many others). For a comprehensive overview of wavelets in mathematics and their statistical applications see the monographs of Daubechies (1992), Vidakovic (1999), Antoniadis and Oppenheim (1995) and Percival and Walden (2000).

This thesis is structured as follows. In chapter 2 we give a short introduction to wavelets. This contains existence and construction of compactly supported wavelet bases and the wavelet representation of a square integrable function. Next, linear function estimation based on wavelets is presented. It will be shown that linear estimates are in general not optimal for the recovery of an unknown function f in the sense that they do not achieve asymptotic minimax rates of convergence. Nevertheless, simple modifications involving shrinkage of empirical wavelet coefficients perform a lot better. We extensively present the current state of literature in the field of nonlinear wavelet regression involving the most common thresholding techniques. This chapter concludes with some asymptotic properties of thresholding estimates involving the bias, mean integrated squared error and asymptotic distribution required in the subsequent chapters. Chapter 3 contains the main part of the thesis. Spectral analysis of stationary time series is extended to multivariate nonparametric trend functions that are estimated by wavelet thresholding. A regression cross covariance (correlation) and regression spectrum are defined. Asymptotic properties of the estimates of these quantities are derived. An algorithm to automatically detect common frequency components as well as possible lag structures in multivariate deterministic functions is presented. The chapter concludes with

a small simulation study.

Chapter 4 investigates application and properties of the classical periodogram in case that the observed time series contains deterministic components. An adjusted periodogram is defined that constitutes a consistent estimate for the regression spectrum and the asymptotic distribution is derived. The results are then compared to those of chapter 3.

Chapter 5 develops an alternative procedure for estimating lags in multivariate time series based on the maximum of the regression cross correlation. If dependence between time series may be restricted to a small number of frequencies or if the influence of the noise in model (1.1) is too large, the regression spectrum may fail to reveal information about the true dependence structure in the signal. The regression cross correlation may then be used as an alternative estimation procedure in time domain that overcomes some of the problems in frequency domain. Chapter 6 presents a variety of applications for the methods theoretically developed in the previous chapters. Examples from biology, medicine, physiology and finance illustrate the good practical performance of the theoretical results. An extensive data analysis investigating the trading behavior of stock market participants at the New York Stock Exchange will end this section. We finish the thesis with some concluding remarks in chapter 7.

Chapter 2

Wavelets and Statistics

In this chapter we briefly review the basic concepts of wavelet analysis and provide an overview of the current research in linear and nonlinear wavelet estimation. The chapter is structured as follows. Section 2.1 is concerned with a basic definition of wavelets and the concept of multiresolution analysis. Nonparametric trend estimation in time series via a linear wavelet approach is presented in section 2.2. Rates of convergence of wavelet estimates are described in case that the true trend function belongs to a certain class of functions. For various combinations of risk measure and regularity of the unknown trend component linear estimates do not achieve optimal rates of convergence and are outperformed by surprisingly simple nonlinear techniques. Section 2.3 presents nonlinear wavelet estimation and discusses several methods of applying wavelet thresholding. The chapter finishes with some asymptotic properties including the bias, mean integrated squared error and asymptotic distribution of wavelet thresholding estimates that will be needed in the subsequent chapters.

2.1 Wavelets and Multiresolution Analysis

The following section summarizes some of the main ideas about wavelets and multiresolution analysis as given e.g. in Daubechies (1992), Vidakovic (1999), Härdle et al. (1998) or Steeb (1998).
Denote by $L^2(\mathbb{R})$ the space of square integrable complex-valued functions on the real line. This space is equipped with the scalar product

$$< f, g >= \int_{-\infty}^{\infty} f(x)\overline{g(x)}dx,$$

where $f, g \in L^2(\mathbb{R})$ and $\overline{g(x)}$ denotes the complex conjugate of $g(x)$. This induces the norm

$$||f||_2 =< f, f >^{1/2}= \left(\int_{-\infty}^{\infty} |f(x)|^2 dx\right)^{1/2}$$

which is finite for all $f \in L^2(\mathbb{R})$. As usually, we refer to orthogonality of $f, g \in L^2(\mathbb{R})$ whenever $< f, g >= 0$.

2.1.1 Basic definitions

Assume we are given a function $\phi \in L^2(\mathbb{R})$ such that $\{\phi_{0k},\, k \in \mathbb{Z}\} = \{\phi(\cdot - k),\, k \in \mathbb{Z}\}$ is a set of orthonormal functions. Denote by V_0 the subspace of $L^2(\mathbb{R})$ generated by $\{\phi_{0k}\}$. Then, there exist coefficients $\alpha_k \in \mathbb{C}$, $k \in \mathbb{Z}$, such that every $f \in V_0$ can be written as

$$f(x) = \sum_{k\in\mathbb{Z}} \alpha_k \phi(x - k)$$

almost everywhere with

$$< \phi_{0j}, \phi_{0k} >= \delta_{j,k},$$

where

$$\delta_{j,k} = \begin{cases} 1, & j = k, \\ 0, & \text{else.} \end{cases}$$

Due to the orthonormality of $\{\phi_{0k},\, k \in \mathbb{Z}\}$ the coefficients α_k are given by

$$\alpha_k =< f, \phi_{0k} >$$

with $\sum_k |\alpha_k|^2 < \infty$. A function $\phi \in L^2(\mathbb{R})$ such that $\{\phi(\cdot - k), k \in \mathbb{Z}\}$ generates an orthonormal family will be called **scaling function**. Define a set of functions $\{\phi_{j,k}\}$ by

$$\phi_{j,k}(x) := 2^{j/2}\phi(2^j x - k),\ j, k \in \mathbb{Z}, \tag{2.1}$$

i.e., we consider integer translations and dyadic dilations of the scaling function. Denote the corresponding sequence of subspaces $\{V_j,\ j \in \mathbb{Z}\}$ by

$$V_j \ = \ \{f(2^j x) \ : \ f \in V_0\},\ j \in \mathbb{Z}. \tag{2.2}$$

Remark 2.1. *According to construction, for each j, $\{\phi_{j,k}, k \in \mathbb{Z}\}$ constitutes a basis for V_j.*

Definition 2.1. *(Multiresolution Analysis) If the function ϕ is chosen such that V_j in (2.2) satisfies*

1. $\ldots V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \ldots$,
2. $\overline{\cup_{j\in\mathbb{Z}} V_j} = L^2(\mathbb{R})$,
3. $\cap_{j\in\mathbb{Z}} V_j = \{0\}$,

then the sequence of subspaces $\{V_j,\ j \in \mathbb{Z}\}$ is called a **multiresolution analysis** *(MRA) of $L^2(\mathbb{R})$.*

Definition 2.2. *If a MRA of $L^2(\mathbb{R})$ is generated by a function ϕ, then ϕ is called a* **father wavelet.**

Though the union over all nested subspaces V_j in a MRA is dense in $L^2(\mathbb{R})$, the resulting system of functions in (2.1) is generally not orthonormal. To construct an orthonormal basis consider the orthogonal complement W_j of V_j in V_{j+1},

$$W_j = V_{j+1} \ominus V_j, \ j \in \mathbb{Z}.$$

It follows that

$$W_j \perp W_k, \ j \neq k.$$

For some $l \in \mathbb{Z}$, each subspace V_j $(j > l)$ of the MRA can be then written as

$$V_j = V_{j-1} \oplus W_{j-1} = V_{j-2} \oplus W_{j-2} \oplus W_{j-1} = \cdots = V_l \oplus \bigoplus_{k=1}^{j-l} W_{j-k}. \tag{2.3}$$

Hence, by remembering 2. of definition 2.1., we may decompose $L^2(\mathbb{R})$ for an arbitrary $l \in \mathbb{Z}$ into orthogonal subspaces,

$$L^2(\mathbb{R}) = V_l \oplus \bigoplus_{k=0}^{\infty} W_{l+k}. \tag{2.4}$$

Therefore, any $f \in L^2(\mathbb{R})$ may be represented as

$$f(x) = \sum_{k \in \mathbb{Z}} \alpha_{l,k} \phi_{l,k}(x) + \sum_{j \geq l, k \in \mathbb{Z}} \beta_{j,k} \psi_{j,k}(x), \tag{2.5}$$

where $\{\psi_{j,k}\}$, $k \in \mathbb{Z}$, is an orthonormal basis for W_j and $\alpha_{l,k}$, $\beta_{j,k}$ are the corresponding coefficients,

$$\alpha_{l,k} = < f, \phi_{l,k} >= \int f(x) \overline{\phi_{l,k}(x)} dx, \tag{2.6}$$

$$\beta_{j,k} = < f, \psi_{j,k} >= \int f(x) \overline{\psi_{j,k}(x)} dx, \tag{2.7}$$

where

$$\sum_k |\alpha_{l,k}|^2 + \sum_{j \geq l,k} |\beta_{j,k}|^2 < \infty.$$

For (2.5) to be called a wavelet expansion of f we need the further requirement that each $\psi_{j,k}$ is of the form

$$\psi_{j,k}(x) := 2^{j/2} \psi(2^j x - k), \ j, k \in \mathbb{Z}, \tag{2.8}$$

for some function $\psi \in L^2(\mathbb{R})$.

Definition 2.3. *A function $\psi \in L^2(\mathbb{R})$ generating the family $\{\psi_{j,k}\}$ in (2.8) is called a* **mother wavelet.**

Definition 2.4. *Representation (2.5) with a basis of the form (2.1) and (2.8) is called* **wavelet series expansion** *of f. The corresponding coefficients $\alpha_{l,k}$ and $\beta_{j,k}$ are called* **wavelet coefficients** *of f (relative to the wavelet basis $\{\phi_{l,k}, \psi_{j,k}, l \in \mathbb{Z}, j \geq l, k \in \mathbb{Z}\}$).*

One notices that the mother wavelet ψ - associated with ϕ - is not uniquely determined. General conditions for existence and construction of ϕ and ψ will be given in the next section. From (2.5) it follows that the wavelet series expansion represents functions by successive approximation starting on a coarse level by using linear combinations of the translated and dilated father and mother wavelet (level l). Finer levels are then obtained by zooming into the time series for increasing numbers of j ($j \geq l$).

2.1.2 Construction of a mother wavelet with given father

The previous section theoretically describes a pair of orthonormal wavelets as the basic building blocks of a multiresolution analysis. Construction of the mother and father wavelet is presented in this section. The following ideas and lemmas can be found e.g. in Daubechies (1992), Härdle et al. (1998), Pinsky (2002), Strichartz (1994) and Vidakovic (1999). For $f \in L^1(\mathbb{R})$ let

$$f^*(\xi) = \int_{-\infty}^{\infty} f(x) e^{-ix\xi} dx \tag{2.9}$$

denote the Fourier transform of f. Furthermore, assume we are given a compactly supported function $\phi \in L^2(\mathbb{R})$. Then, orthonormality of its integer translates is discussed in

Lemma 2.1. *Assume that $\phi \in L^2(\mathbb{R})$. A necessary and sufficient condition for $\{\phi(\cdot - k),\, k \in \mathbb{Z}\}$ to constitute an orthonormal system in $L^2(\mathbb{R})$ is that*

$$\sum_k |\phi^*(\xi + 2\pi k)|^2 = 1 \tag{2.10}$$

almost everywhere.

Let the corresponding subspace generated by $\{\phi(\cdot - k),\, k \in \mathbb{Z}\}$ be denoted by V_0 and let V_j be defined as in (2.2). In case that the subspaces V_j are nested, the scaling function can be written as

$$\begin{aligned} \phi(x) &= \sum_n \alpha_n \phi_{1,n}(x) \\ &= \sum_n \alpha_n 2^{1/2} \phi(2x - n), \end{aligned} \tag{2.11}$$

where the sum contains only a finite amount of coefficients α_n. The family $\{\phi_{1,n}, n \in \mathbb{Z}\}$ is an orthonormal basis in V_1 such that $\alpha_n =< \phi, \phi_{1,n} >$. Inserting (2.11) in (2.9) gives

$$\begin{aligned} \phi^*(\xi) &= \int \phi(x) e^{-ix\xi} dx \\ &= \frac{1}{\sqrt{2}} \sum \alpha_n e^{-i\frac{\xi}{2}n} \phi^*\left(\frac{\xi}{2}\right) \\ &= m_0\left(\frac{\xi}{2}\right) \phi^*\left(\frac{\xi}{2}\right), \end{aligned}$$

where m_0 is a trigonometric polynomial with coefficients α_n,

$$m_0(\xi) = \frac{1}{\sqrt{2}} \sum \alpha_n e^{-i\xi n}. \tag{2.12}$$

This is the intuition for

Lemma 2.2. *A necessary and sufficient condition for $V_j \subset V_{j+1}$, $j \in \mathbb{Z}$, according to definition 2.1. is that there exists a 2π-periodic function $m_0 \in L^2([0, 2\pi])$ such that*

$$\phi^*(\xi) = m_0\left(\frac{\xi}{2}\right) \phi^*\left(\frac{\xi}{2}\right) \tag{2.13}$$

almost everywhere.

Remark 2.2. *If $\phi \in L^2(\mathbb{R})$ is compactly supported, it follows that $\phi \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ and ϕ^* is continous. Furthermore, m_0 is a trigonometric polynomial with a finite amount of coefficients α_n. Thus m_0 is continous as well and (2.13) holds for all ξ.*

One can show that every solution of (2.13) satisfies

$$|m_0(\xi)|^2 + |m_0(\xi + \pi)|^2 = 1 \tag{2.14}$$

almost everywhere and that usage of reasonable scaling functions (i.e. $\int \phi(x)dx = 1$) implies $m_0(0) = 1$ (see e.g. Härdle et al. (1998, p. 37 and p. 92)). Equation (2.14) will be used further down to construct a scaling function ϕ. Notice that the function m_0 needs not to be unique.
The polynomial m_0 is now suitable to develop an associated mother wavelet ψ, i.e. $\psi \in W_0$ such that the family $\{\psi(\cdot - k),\, k \in \mathbb{Z}\}$ is an orthonormal basis in W_0.

Lemma 2.3. *Let the assumptions of lemma 2.2. hold. Define $m_1(\xi) = \overline{m_0(\xi + \pi)}e^{-i\xi}$ and set*

$$\psi^*(\xi) = m_1\left(\frac{\xi}{2}\right)\phi^*\left(\frac{\xi}{2}\right).$$

Then, the inverse Fourier transform of ψ^* is a mother wavelet.

Starting with a function $\phi \in L^2(\mathbb{R})$, lemmas 2.1. and 2.2. provide methods to check whether ϕ may be used as a father wavelet. Furthermore, lemma 2.3. introduces a way to construct a mother wavelet. In remains to show that 2. and 3. of definition 2.1. are satisfied. Daubechies (1992, p. 141) shows that for all $\phi \in L^2(\mathbb{R})$ satisfying (2.10) the resulting sequence of subspaces has trivial intersection. For 2. to be satisfied, a mild condition of the integrability of ϕ is given in (see Härdle et al. (1998, p. 95))

Lemma 2.4. *Assume that there exists a function* Φ *that is monotonely decreasing and bounded with*

$$\int \Phi(|u|)du < \infty.$$

Let $\phi \in L^2(\mathbb{R})$ *satisfy (2.10) and (2.13). Then, if* $|\phi(u)| \leq \Phi(|u|)$ *almost everywhere,*

$$\overline{\bigcup V_j} = L^2(\mathbb{R}).$$

2.1.3 Construction of a father wavelet

The statements of the previous lemmas are based on the presence of a suitable father wavelet. Construction of a suitable scaling function will be described in this section. First remember the previously developed properties of ϕ:

1. $\sum_k |\phi^*(\xi + 2\pi k)|^2 = 1$ almost everywhere,
2. $\phi^*(\xi) = m_0\left(\frac{\xi}{2}\right)\phi^*\left(\frac{\xi}{2}\right)$ almost everywhere, with m_0 defined as in lemma 2.2., and
3. $|m_0(\xi)|^2 + |m_0(\xi + \pi)|^2 = 1$ and $m_0(0) = 1$.

As a consequence,

$$|m_0(\pi)|^2 = 1 - |m_0(0)|^2 = 0,$$

such that (cf. lemma 2.3.)

$$\int \psi(x)dx = 0. \tag{2.15}$$

Remark 2.3. *Many applied papers and textbooks use (2.15) as a defining property for mother wavelets (see e.g. Percival and Walden (2000), Bruce and Gao (1996)).*

According to lemma 2.2. the sequence of subspaces V_j is nested if and only if

$$\begin{aligned}\phi^*(\xi) &= m_0\left(\frac{\xi}{2}\right)\phi^*\left(\frac{\xi}{2}\right) \\ &= m_0\left(\frac{\xi}{2}\right)m_0\left(\frac{\xi}{4}\right)\phi^*\left(\frac{\xi}{4}\right) = \cdots = \prod_{j=1}^{\infty} m_0\left(\frac{\xi}{2^j}\right), \qquad (2.16)\end{aligned}$$

where it is assumed that $\phi^*(0) = 1$. If this product properly defines a function $\phi \in L^2(\mathbb{R})$ such that it generates an orthonormal system in $L^2(\mathbb{R})$, we get a family of orthonormal basis function associated with a multiresolution analysis. Pointwise convergence of (2.16) is discussed in

Lemma 2.5. *Let $m_0(\xi)$ satisfy condition 3. above with $m_0(\xi)$ being Lipschitz continous. Then,*

$$\prod_{j=1}^{\infty} m_0\left(\frac{\xi}{2^j}\right) \qquad (2.17)$$

converges uniformly on every compact subspace of $\mathbb{R}$.

Remark 2.4. *Assume that m_0 is of the form (2.12) with the number of non-zero coefficients being finite. Then, lemma 2.5. surely holds. Daubechies (1992) mentiones a generalization to infinite sets of coefficients in case that the decay is sufficiently fast.*

The next lemma is due to Mallat (1989a) and shows that the resulting limit function as well as its Fourier transform are in $L^2(\mathbb{R})$.

Lemma 2.6. *(Mallat) Let m_0 satisfy the assumptions of lemma 2.5., then $\phi, \phi^* \in L^2(\mathbb{R})$ with $||\phi||_{L^2} \leq 1$.*

Finally, for the so constructed ϕ we have to investigate whether the resulting family $\{\phi(\cdot - k), k \in \mathbb{Z}\}$ generates an orthonormal system in $L^2(\mathbb{R})$. The next

theorem is due to Cohen and presents the results for the most popular case of polynomials m_0 to construct a compactly supported father wavelet ϕ. It can be found in Härdle et al. (p. 56). For a proof refer to Cohen and Ryan (1995).

Theorem 2.1. *(Cohen) Assume that the function $m_0(\xi)$ is a trigonometric polynomial of the form*

$$m_0(\xi) = \frac{1}{\sqrt{2}} \sum_{k=N_0}^{N_1} \alpha_k e^{-ik\xi},$$

where N_0, $N_1 \in \mathbb{Z}$ and

$$\frac{1}{\sqrt{2}} \sum_{k=N_0}^{N_1} \alpha_k = 1.$$

Assume further that there exists a compact set $\mathcal{K}$ in $\mathbb{R}$, containing a neighborhood of 0, such that

1. $\sum_k \mathbf{1}_{\{\xi+2k\pi \in \mathcal{K}\}} = 1$ *almost everywhere,*
2. $m_0(2^{-j}\xi) \neq 0$, *for all $\xi \in \mathcal{K}$ and for all $j \geq 1$.*

Then, the function $\phi^*(\xi)$ is the Fourier transform of a function $\phi \in L^2(\mathbb{R})$ such that

(i) *supp $\phi \subseteq [N_0, N_1]$,*
and

(ii) *$\{\phi(\cdot - k), k \in \mathbb{Z}\}$ is an orthonormal system in $L^2(\mathbb{R})$.*

This section developed a method to construct compactly supported wavelets. For further information on the regularity of the wavelet basis or construction of a MRA starting from a Riesz basis see e.g. the monographs of Daubechies and Härdle et al.

2.2 Linear wavelet trend estimation

This section is concerned with nonparametric estimation of a deterministic signal given some noisy data. Let an observed time series $Y(1), \ldots, Y(n)$ be of the form

$$Y(i) = f(t_i) + \epsilon(i), \tag{2.18}$$

where $t_i = i/n$ $(i = 1, \ldots, n)$ and $\epsilon(i)$ is independent and identically distributed with mean 0 and variance σ^2 (to be relaxed in the subsequent chapters). The only knowledge about f is that it belongs to some known set $\mathcal{F}$ of functions. Define the L^p risk of an estimator $\hat{f}$ of f by

$$E||\hat{f} - f||_p^p, \tag{2.19}$$

$1 \leq p < \infty$. The performance in the worst case over the set $\mathcal{F}$ is given by

$$\sup_{f \in \mathcal{F}} E||\hat{f} - f||_p^p.$$

By taking the infimum over all possible estimates $\hat{f}$, the L^p minimax risk is obtained,

$$R_n(\mathcal{F}, p) = \inf_{\hat{f}} \sup_{f \in \mathcal{F}} E||\hat{f} - f||_p^p.$$

Furthermore, we write R_n^{lin} in case that the corresponding risk is restricted to linear estimators. Estimation of f will be now performed by minimizing L^p risk (2.19) over a certain set of functions. We therefore introduce the concept of Besov spaces.

2.2.1 Approximation in Besov spaces

The following definition and characterization of Besov spaces follows the representation of Devore and Popov (1988) in connection to Donoho and Johnstone (1998). See also Donoho and Johnstone (1995) and Härdle et al. (1998).

Assume that $f \in L^p(\mathbb{R})$, $1 \leq p \leq \infty$. Define the rth order difference by

$$\Delta_h^{(r)} f = \sum_{k=0}^{r} \binom{r}{k} (-1)^k f(t + kh)$$

and denote the modulus of smoothness of order r of $f \in L^p([0,1])$ by

$$w_{r,p}(f,h) := ||\Delta_h^{(r)} f||_{L^p([0,1-rh])}. \tag{2.20}$$

For $\alpha > r$, we say that $f : [0,1] \to \mathbb{R}$ satisfying $f \in L^p(\mathbb{R})$ is in the **Besov space** $B_{p,q}^{\alpha}$ whenever

$$|f|_{B_{p,q}^{\alpha}} = \begin{cases} \left(\int_0^1 \left(\frac{w_{r,p}(f,h)}{h^{\alpha}} \right)^q \frac{dh}{h} \right)^{\frac{1}{q}}, & q < \infty, \\ \sup_{0<h<1} \frac{w_{r,p}(f,h)}{h^{\alpha}}, & q = \infty, \end{cases} \tag{2.21}$$

is finite. We will refer to a **Besov ball** of radius L as the set

$$B_{p,q}^{\alpha}(L) = \{f : f \in B_{p,q}^{\alpha},\ |f|_{B_{p,q}^{\alpha}} \leq L\}.$$

Remark 2.5. *Denote $s = \alpha + \frac{1}{2} + \frac{1}{p}$ and let the wavelet basis functions be compactly supported having r vanishing moments. Define*

$$|f|_{b_{p,q}^{\alpha}} = \left(\sum_{k \in \mathbb{Z}} |\alpha_{l,k}|^p \right)^{\frac{1}{p}} + \left[\sum_{j \geq l}^{\infty} \left(2^{js} \left(\sum_{k \in \mathbb{Z}} |\beta_{j,k}|^p \right)^{\frac{1}{p}} \right)^q \right]^{\frac{1}{q}}, \tag{2.22}$$

with the usual adjustments if p or q are ∞. Then, in case $1 \leq p, q \leq \infty$, (2.21) and (2.22) are equivalent in the sense that

$$(||f||_{L^p} + |f|_{B_{p,q}^{\alpha}}) \asymp |f|_{b_{p,q}^{\alpha}},$$

for every $L^p([0,1])$, where "$\asymp$" means that the ratio of both sides is bounded by two constants. The Besov norm in function space is thus equivalent to the sequence norm of the wavelet coefficients of f. For additional results on this topic see Härdle et al. (1998).

Asymptotic minimax properties

The next theorem is due to Donoho and Johnstone (1998) and shows that linear estimates only achieve an optimal rate of convergence in very special cases of parameter constellation.

Theorem 2.2. *(Donoho and Johnstone) Let $\mathcal{F}$ be a Besov ball $B^{\alpha}_{p,q}(L)$ with $\alpha > \frac{1}{p}$ and $1 \leq p, q \leq \infty$. Then,*

$$\begin{aligned} R_n(\mathcal{F}, 2) &= E||\hat{f} - f||_2^2 \asymp n^{-r}, \ \ n \to \infty, \\ R_n^{lin}(\mathcal{F}, 2) &= E||\hat{f}^{lin} - f||_2^2 \asymp n^{-r'}, \ \ n \to \infty, \end{aligned}$$

with rate exponents

$$r = \frac{2\alpha}{2\alpha + 1} \tag{2.23}$$

$$r' = \frac{\alpha + \left(\frac{1}{p_-} - \frac{1}{p}\right)}{\alpha + \frac{1}{2} + \left(\frac{1}{p_-} - \frac{1}{p}\right)}, \tag{2.24}$$

where $p_- = \max(p, 2)$. The same conlusion holds for Besov balls $\alpha = 1$ and $p = q = 1$, and also for Bounded Variation Balls, with parameters set to $\alpha = 1$ and $p = 1$.

This theorem shows that linear estimates attain an optimal rate of convergence in case $p \geq 2$ whereas for $p < 2$ it does not. Consider e.g. the case $\alpha = 1$ and $p = 1$. We then notice a remarkable difference between $r = 2/3$ and $r' = 1/2$.

Other authors have already analyzed the suboptimality of linear estimators. In restricting considerations to Sobolev classes, Nemirovskii et al. (1983, 1985) investigate linear and arbitrary estimates and give lower bounds for the rate of convergence. In particular they show that estimators of maximum likelihood type outperform the rate of convergence of any linear estimator.

Consider now the wavelet representation of $f \in L^2(\mathbb{R})$,

$$f(t) = \sum_{k=-\infty}^{\infty} \alpha_{l,k}\phi_{l,k}(t) + \sum_{j\geq l}^{\infty}\sum_{k=-\infty}^{\infty} \beta_{j,k}\psi_{j,k}(t) \ (l \in \mathbb{Z}). \tag{2.25}$$

Application of (2.25) for statistical purposes requires consideration of only a finite amount of wavelet coefficients. To truncate the infinite series in (2.25) an upper bound for the degree of approximation J_n (dependent on n) is introduced. Furthermore, the theoretical wavelet coefficients $\alpha_{l,k}$ and $\beta_{j,k}$ are replaced by almost consistent estimates. An estimator for f is therefore given by

$$\hat{f}_{J_n}(t) = \sum_{k\in\mathbb{Z}} \hat{\alpha}_{l,k}\phi_{l,k}(t) + \sum_{j\geq l}^{J_n}\sum_{k\in\mathbb{Z}} \hat{\beta}_{j,k}\psi_{j,k}(t), \tag{2.26}$$

where

$$\hat{\alpha}_{l,k} = \frac{1}{n}\sum_{i=1}^{n} \phi_{l,k}(t_i)Y(i) \tag{2.27}$$

and

$$\hat{\beta}_{j,k} = \frac{1}{n}\sum_{i=1}^{n} \psi_{j,k}(t_i)Y(i), \tag{2.28}$$

for some J_n. Notice that for compactly supported wavelets and f defined on $[0,1]$ the sum over k on each level j is finite.

Several authors investigated nonparametric estimation of an unknown f with respect to L^p-loss. In most cases the results are gained in case that the exponent of the loss function corresponds to the regularity of the unknown function f. Härdle et al. (1998) call this the case of matched assumptions. Examples can be found e.g. in Bretagnolle and Huber (1979) or Ibragimov and Hasminskii (1980, 1981). In the context of density estimation, a generalization can be found in Donoho et. al (1996). The authors give lower bounds for linear and nonlinear estimates in case that f belongs to the some Besov class of functions. Note that in the case treated below, the parameters for the regularity of f ($\in L^r(\mathbb{R})$) and the risk (L^p-risk) do not necessarily match.

Theorem 2.3. *(Donoho, Johnstone, Kerkyacharian, Picard)*
Let $f \in B^{\alpha}_{r,q}(L)$ be compactly supported where $1 \leq r \leq \infty$, $1 \leq q \leq \infty$, $\alpha > \frac{1}{r}$, $1 \leq p < \infty$. A lower bound for the L^p minimax risk is then given by

$$R_n(B^{\alpha}_{r,q}(L), p) \geq C r_n(\alpha, r, p, q), \tag{2.29}$$

where

$$r_n(\alpha, r, p, q) = \begin{cases} n^{-\alpha_1 p}, & \alpha_1 = \frac{\alpha}{2\alpha+1}, \quad \text{if } r > \frac{p}{2\alpha+1}, \\ \left(\frac{\log n}{n}\right)^{\alpha_2 p}, & \alpha_2 = \frac{\alpha - \frac{1}{r} + \frac{1}{p}}{2\left(\alpha - \frac{1}{r}\right)+1}, \quad \text{if } r \leq \frac{p}{2\alpha+1}, \end{cases}$$

where C is a finite positive constant. Moreover, if $\alpha' = \alpha - \left(\frac{1}{r} - \frac{1}{p}\right)_+$, then

$$R^{lin}_n(B^{\alpha}_{r,q}(L), p) \asymp n^{\frac{\alpha' p}{2\alpha'+1}}. \tag{2.30}$$

Comparing the lower bound for the minimax risk of general estimates in (2.29) and the asymptotic rate of convergence of linear estimates in (2.30) we see that the theorem devides the set of all (r, p)-combinations for given α into two parts. For $r \leq \frac{p}{2\alpha+1}$ linear estimators are not able to converge with the minimax rate to the true function f. The same holds for $r > \frac{p}{2\alpha+1}$ if $r < p$. In this case $\alpha' < \alpha$ such that the actual convergence is slower as in the theorem above. Linear estimators only achieve the optimal rate $n^{-\frac{\alpha}{2\alpha+1}}$ in case that $r \geq p$, where $\alpha' = \alpha$ (compare with theorem 2.2.). The results are summarized in table 2.1. Alternatively, one may interprete the results dependent on the smoothness of f. For given r and p the regularity α needs to be sufficiently high such that a linear estimator attains the optimal rate of convergence. One notices, that for small α the optimal rate of convergence contains an additional logarithmic factor. We conclude this section with some results from Härdle et al. (p. 129) concerning the structure of (2.30).

Remark 2.6. *Let $2^{J_n} \leq n$. Then,*

(i) an upper bound for the bias error is given by

$$\sup_{f \in B^{\alpha}_{p,q}(L)} ||E\hat{f}_{J_n} - f||^p_p \leq C_1 2^{-J_n \alpha p}.$$

	$r \leq \frac{p}{2\alpha+1}$	$r > \frac{p}{2\alpha+1}$
$r < p$	linear estimates not optimal, optimal rate slower than $n^{-\frac{\alpha}{2\alpha+1}}$	linear estimates not optimal, optimal rate $n^{-\frac{\alpha}{2\alpha+1}}$
$r \geq p$	linear estimates not optimal, optimal rate slower than $n^{-\frac{\alpha}{2\alpha+1}}$	linear estimates optimal, optimal rate $n^{-\frac{\alpha}{2\alpha+1}}$

Table 2.1: Optimal rates of convergence for various parameter constellations

(ii) An upper bound for the stochastic error is given by

$$\sup_{f \in B^{\alpha}_{p,q}(L)} E||\hat{f}_{J_n} - E(\hat{f}_{J_n})||_p^p \leq C_2 \left(\frac{2^{J_n}}{n}\right)^{p/2}.$$

With increasing J_n we get a better approximation in terms of the bias at the cost of an increased variance of the estimator. If J_n is such that

$$2^{J_n} \asymp n^{\frac{1}{2\alpha+1}}, \tag{2.31}$$

bias and stochastic errors are balanced and one obtains the optimal linear risk (2.30). Furthermore, (2.31) remains valid even if the parameter p in the loss function is changed.

2.2.2 Asymptotic properties of linear wavelet estimates

Assume that model (2.18) holds. Asymptotic properties of the linear wavelet estimate (2.26) may be derived in case of very general parameter settings. Additionally, in view of

$$\begin{aligned} \hat{\alpha}_{l,k} &= \frac{1}{n}\sum_{i=1}^{n} f(t_i)\,\phi(t_i) + \frac{1}{n}\sum_{i=1}^{n} \epsilon(i)\phi(t_i) \\ &\approx \int f(x)\phi(x)dx + \frac{1}{n}\sum_{i=1}^{n} \epsilon_i \phi(t_i), \end{aligned} \tag{2.32}$$

the assumption about independence of the $\epsilon(i)$ is by far too restrictive for deriving asymptotic properties. In this section we investigate functions of bounded variation in connection with stationary error terms. The results are given in the following theorem. It is a collection of the essential assumptions and results in Brillinger (1994 and 1996). The following assumptions are needed.

(A1) The father and the mother wavelet $\phi(\cdot)$ and $\psi(\cdot)$ generate an orthonormal basis for a finite interval containing $[0, 1]$. Both functions are compactly supported and are of bounded variation.

(A2) The unknown deterministic function f is bounded and of bounded variation on the unit interval and vanishes elsewhere.

Define

$$c_m(u_1, \dots u_{m-1}) = cum\{\epsilon(t + u_1), \dots, \epsilon(t + u_m), \epsilon(t)\}$$

for $m \in \mathbb{N}$, $t, u_i \in \mathbb{Z}$, $i \in \{1, \dots, m\}$, as the mth order cumulant of $\epsilon(t)$.

(A3) The cumulants of $\epsilon(t)$ exist and are absolutely summable, i.e.

$$C_m = \sum_{u_1,\dots,u_m} |c_m(u_1, \dots u_{m-1})| < \infty,$$

and for the autocovariances $\gamma_\epsilon(h) = \text{cov}\,(\epsilon(t + h), \epsilon(t))$, $h \in \mathbb{Z}$, we have

$$\sum_h |h||\gamma_\epsilon(h)| < \infty.$$

Furthermore, the spectral density of $\epsilon(i)$, denoted by h_ϵ, does not vanish at the origin.

Theorem 2.4. *(Brillinger) Assume model (2.18) with zero mean stationary $\epsilon(i)$. Further, let the assumptions (A1)-(A3) hold. Then,*

a)

$$\begin{aligned}
E(\hat{\alpha}_{l,k} - \alpha_{l,k}) &= O(2^{l/2}n^{-1}), \\
var\,(2^{-\frac{l}{2}}\hat{\alpha}_{l,k}) &= 2\pi h_\epsilon(0)2^{-l}n^{-1} + O(n^{-2}), \\
cov\,(2^{-\frac{l}{2}}\hat{\alpha}_{l,k}, 2^{-\frac{l}{2}}\hat{\alpha}_{l,k'}) &= O(n^{-2}),\ k \neq k', \\
E(\hat{\beta}_{j,k} - \beta_{j,k}) &= O(2^{j/2}n^{-1}), \\
var\,(2^{-\frac{j}{2}}\hat{\beta}_{j,k}) &= 2\pi h_\epsilon(0)2^{-j}n^{-1} + O(n^{-2}), \\
cov\,(2^{-\frac{j}{2}}\hat{\beta}_{j,k}, 2^{-\frac{l}{2}}\hat{\beta}_{j',k'}) &= O(n^{-2}),\ (j,k) \neq (j',k'), \\
&\textit{and} \\
cov\,(2^{-\frac{j}{2}}\hat{\alpha}_{l,k}, 2^{-\frac{l}{2}}\hat{\beta}_{j,k'}) &= O(n^{-2}),\ \textit{for all } j,k,k',l.
\end{aligned}$$

where $h_\epsilon(\cdot)$ denotes the spectral density of $\epsilon(i)$. The error terms are uniform in j, j', k, k', l.

b) Finite samples of empirical wavelet coefficients are asymptotically jointly normal with the indicated first and second moments.

c) Let J_n be such that for all integers $m > 3$

$$2^{m-1}J_n n^{-m+1}/(var(\hat{f}(x))^{m/2}) \to 0.$$

Then, for finite samples in x, $\hat{f}(x)$ given in (2.26) is asymptotically normal with mean $f(x)$ and

$$\begin{aligned}
&cov(\hat{f}(x), \hat{f}(y)) \\
&= \frac{2\pi h_\epsilon(0)}{n}\left[\sum_k \phi_{l,k}(x)\phi_{l,k}(y) + \sum_{j\geq l}^{J_n}\sum_k \psi_{j,k}(x)\psi_{j,k}(y)\right] + O(2^{2J_n}n^{-2}).
\end{aligned}$$

2.3 Wavelet Thresholding

In the previous sections we have shown that linear estimators attain the optimal rate of convergence only in very special cases of parameter constellation.

Remark 2.6. indicates that this is due to an overproportional increase in the variance. In the early 1990s, Donoho and Johnstone first introduced wavelet thresholding to statistical literature. In wavelet thresholding, noise is removed by shrinking wavelet coefficients towards zero at a suitable rate (see Donoho and Johnstone (1994a, 1995)). Hereby, the variance of the wavelet estimator is reduced at the cost of some increased bias. Let

$$Y(i) = f(t_i) + \epsilon(i),$$

where $\epsilon(i)$ is independent and identically normal distributed with mean 0 and variance σ^2. The discrete wavelet transform (DWT) calculates empirical wavelet coefficients by pre-multiplying an orthogonal matrix $\mathcal{W}$ to the vector $Y = (Y(1), \ldots, Y(n))^T$ such that the vector of coefficients d is given by

$$\begin{aligned} d &= \mathcal{W}f + \mathcal{W}\epsilon \\ &= \theta + \epsilon'. \end{aligned} \tag{2.33}$$

Due to orthonormality of $\mathcal{W}$, ϵ' is $iid\, \mathcal{N}(0, \sigma^2)$ as well. See Mallat (1989b) for further information on the computation of wavelet coefficients by using the pyramid algorithm and the structure of $\mathcal{W}$ and d.

Define **soft thresholding** of a wavelet coefficient by

$$\hat{\theta}_i^S := \delta_S(d_i, \lambda) = sgn(d_i)(|d_i| - \lambda)\mathbf{1}_{\{|d_i|>\lambda\}}, \tag{2.34}$$

and let

$$\hat{\theta}_i^H = \delta_H(d_i, \lambda) = d_i\mathbf{1}_{\{|d_i|>\lambda\}} \tag{2.35}$$

denote the **hard thresholding** rule, where λ is a threshold to be specified. Application of both rules results in deletion of coefficients that are smaller than the threshold. Whereas hard thresholding leaves the remaining coefficients untouched, soft shrinkage reduces all coefficient for the same amount λ. An estimate $\hat{f}$ is now gained by replacing the vector $\hat{\theta} = (\hat{\alpha}_{l,k}, \hat{\beta}_{j,k},\, l \in \mathbb{Z}, j \geq l, k \in \mathbb{Z})^T$ in the linear estimate (2.26) by $\theta^* = (\alpha^*_{l,k}, \beta^*_{j,k},\, l \in \mathbb{Z}, j \geq l, k \in \mathbb{Z})$,

where θ^* denotes either soft shrinkage or hard thresholding applied to $\hat{\theta}$. With the usual notations, the wavelet thresholding estimate is therefore given by

$$\hat{f}(t) := \sum_k \alpha^*_{l,k}\phi_{l,k}(t) + \sum_{j\geq l}^{J_n}\sum_k \beta^*_{j,k}\psi_{j,k}(t). \tag{2.36}$$

2.3.1 Choice of threshold

According to (2.34) and (2.35) the choice of λ is crucial for the resulting properties of the estimate $\hat{f}$. A small threshold is likely to retain many coefficients in (2.36) that are due to noise. In contrast, a threshold that is chosen too large might even shrink coefficients that contribute to signal. A large variety of methods about threshold selection has been proposed in literature. The most important ones are presented in the following section.
Note in advance that the subsequent procedures only involve those values of f that are observed on an increasingly fine grid on $[0, 1]$. Risk is therefore adjusted to the current situation by considering l^2-loss of an estimator $\hat{f}$ of f by

$$R_n(\hat{f}, f) = n^{-1}\sum_{i=1}^{n} E[\hat{f}(t_i) - f(t_i)]^2.$$

We correspondingly adjust our notation of the minimax risk to

$$\mathcal{R}_n(\mathcal{L}(Y), \mathcal{F}) = \inf_{\hat{f}}\sup_{f} R_n(\hat{f}, f), \tag{2.37}$$

where $\mathcal{L}(Y)$ denotes the procedure for selecting the parameter λ that is dependent on the data alone.

Ideal wavelet estimation

As a point of reference, Donoho and Johnstone (1994a) study ideal wavelet thresholding, i.e., the best possible performance one might achieve by using (2.36). It is therefore assumed that an oracle delivers information about

the coefficients that contribute to signal. More explicitly, a data-based oracle knows σ^2 and tells which empirical wavelet coefficient exceeds the noise level. Donoho and Johnstone call this approach *selective wavelet reconstruction (SW)*. According to the oracle, let θ_i be estimated by $\hat{\theta}_i = d_i \mathbf{1}_{\{|d_i|>\sigma\}}$. Then, straightforward calculations show

$$E[\hat{\theta}_i - \theta_i]^2 = \begin{cases} \theta_i^2 & |d_i| \leq \sigma, \\ \sigma^2, & |d_i| > \sigma. \end{cases} \tag{2.38}$$

Summing over all coefficients and combining the previous results gives an optimal minimax risk of

$$\mathcal{R}_n(SW, f) = \sum_i \min(\theta_i^2, \sigma^2).$$

Of course, σ^2 is unknown in practice such that it is impossible for an estimator based on wavelet thresholding to attain the optimal risk.

Minimax threshold

Even if wavelet estimates rely on the data alone one might achieve risks that are close to $\mathcal{R}_n(SW, f)$. The following results are due to Donoho and Johnstone (1994a). According to (2.33), $d_i \sim \mathcal{N}(\theta_i, \sigma^2)$. Assume without loss of generality that $\sigma^2 = 1$ and let λ_n^* be the largest threshold λ such that

$$\Lambda_n^* = \inf_{\lambda} \sup_{\theta_i} \frac{E[\delta^S(d_i, \lambda) - \theta_i]^2}{n^{-1} + \min(\theta_i^2, 1)} \tag{2.39}$$

is attained. Then, the minimax estimate

$$\hat{\theta}_i^* = \delta_S(\hat{\theta}_i, \lambda_n^* \sigma) \;\; (i = 1, \ldots, n)$$

satisfies

$$R(\hat{f}, f) \leq \Lambda_n^* \left(\frac{\sigma^2}{n} + \mathcal{R}_n(SW, f) \right) \tag{2.40}$$

for all $\theta \in \mathbb{R}^n$.

Remark 2.7. *For large n, Λ_n^* and λ_n^* are proportional to $2\log n$ and $(2\log n)^{1/2}$ respectively, i.e., the rate of convergence using minimax estimation is within a logarithmic factor of ideal estimation (SW).*

RiskShrink

Minimax thresholding is equally applied to all coefficients in (2.36). As mentioned in the previous sections, coefficients $\beta_{j,k}$ represent detail information and are expected to cluster around 0 (see (2.7) and (2.15)). In contrast, father wavelets are closely linked to the mean value of the function f. With this intuition, Donoho and Johnstone (1994a) define the RiskShrink procedure as a modification of the minimax estimate, where only detail coefficients $\hat{\beta}_{j,k}$ are due to shrinkage. Let λ_n^* denote the minimax threshold solving (2.39) and let

$$\beta_{j,k}^* = \delta_S(\hat{\beta}_{j,k}, \lambda_n^*\sigma) \tag{2.41}$$

with corresponding estimate $\tilde{f}^*$. Then,

$$R(\tilde{f}^*, f) \leq \Lambda_n^* \left((2^l + 1)\frac{\sigma^2}{n} + \mathcal{R}_n(SW, f) \right), \tag{2.42}$$

where l is the coarsest level of approximation in (2.36).

Remark 2.8. *Due to construction, RiskShrink performs very similar to the minimax estimate. However, one might expect better performance if the underlying function f has nonzero mean. Furthermore, in view of (2.42) we notice that the loss of RiskShrink asymptotically is within a factor $2\log n$ of ideal risk.*

The mean squared error contains a variance and a bias component, but does not account for the smoothness of the underlying function. Both previously defined estimators may involve coefficients from very high resolution levels such that noise artifacts are frequently observed in applications. This effect is mitigated by the concept of universal thresholding (VisuShrink).

VisuShrink

Donoho and Johnstone (1994a) propose a universal threshold for all levels in the wavelet decomposition resulting in particularly smooth trend estimates. Let the minimax threshold in RiskShrink be replaced by a universal threshold $\lambda = (2 \log n)^{1/2}$ such that the soft thresholding rule changes to

$$\beta^*_{j,k} = \delta_S(\hat{\beta}_{j,k}, (2 \log n)^{1/2} \sigma). \tag{2.43}$$

Denote the corresponding trend estimate by $\hat{f}^v$. Then,

$$R(\hat{f}^v, f) \leq (2 \log n + 1) \left((2^l + 1) \frac{\sigma^2}{n} + \mathcal{R}_n(SW, f) \right). \tag{2.44}$$

Therefore, application of a universal threshold attains asymptotically the same risk as the minimax estimate (see remark 2.7.). For finite samples the factor Λ^*_n has to be adjusted to $2 \log n + 1$. Table 2.2 compares minimax and universal threshold for different sample sizes. This table is a part of results of Donoho and Johnstone (1994a). One notices the relative difference for small sample sizes between universal and minimax threshold as well as between the constants.

Well known results from literature show that application of a universal threshold (2.43) deletes all coefficients that are due to pure noise with high probability (see e.g. Pickands (1967)[1]). The resulting trend estimate is therefore relatively smooth compared to the minimax estimate but shows a tendency to oversmoothing in cases of locally inhomogenous trend components. In contrast, visible performance of VisuShrink is superior if the underlying f is sufficiently smooth. This feature has been observed by many authors (see e.g. Cai (1999)). Other interesting results for universal thresholding estimates

[1]Pickands shows that for a stationary $\epsilon(i) \sim \mathcal{N}(0,1)$ with $\lim_{k \to \infty} E(\epsilon(i+k)\epsilon(i)) = 0$,

$$\max\{\epsilon(i)\} / \sqrt{2 \log n} \to 1$$

almost surely as $n \to \infty$.

n	λ_m	$\sqrt{2\log n}$	Λ_n^*	$2\log n + 1$
64	1.474	2.884	3.124	8.318
256	1.669	3.330	4.442	11.090
1024	2.232	3.723	5.976	13.863
4096	2.594	4.079	7.728	16.635
16384	2.952	4.405	9.715	19.498

Table 2.2: Comparison of minimax and universal thresholding (source: Donoho and Johnstone (1994a, p.438))

show that $\hat{f}^v$ is near-optimal in pointwise convergence (Donoho and Johnstone (1992)) and, with probability tending to 1, as smooth as f (Donoho (1992)).

Alternatively, application of the universal threshold may be considered as an inference problem of testing n statistical hypotheses $H_0 : \theta_i = 0$ vs. $H_1 : \theta_i \neq 0$, where

$$P(|d_i| > \sqrt{2\log n}\sigma)$$

denotes the error of first kind (in case that H_0 holds). Because this error decreases with increasing n the power of the test decreases simultaneously. Asymptotically this leads to acceptance of H_0 in almost all cases resulting in significant underfitting of the model. In this context Abramovich and Benjamini (1995, 1996) proposed a thresholding approach based on the False Discovery Rate (FDR) of Benjamini and Hochberg (1995) that uses the proportion of coefficients in the wavelet representation that are incorrectly included. They show that their estimate is robust in mean square sense over a wide range of test functions. For further reference on testing multiple hypotheses in wavelet thresholding see Lavrik et al. (2005).

SureShrink

Donoho and Johnstone (1995) develop a method to select a threshold by minimizing Stein's unbiased estimator of risk (Stein (1981)). The main difference is that the choice of threshold is performed levelwise.

Stein showed the following: If we have an estimator $\hat{\theta}$ based on the vector of coefficients d, then the risk $||\hat{\theta}-\theta||^2$ may be estimated without bias. Therefore, write

$$\hat{\theta}(d) = d + g(d),$$

where $g = (g_i)_{i=1}^n$ is a function from $\mathbb{R}^n$ into $\mathbb{R}^n$. Then, weak differentiability of $g(d)$ implies

$$E_\theta||\hat{\theta}(d) - \theta||^2 = n + E\left(||g(x)||^2 + 2\nabla \cdot g(x)\right), \tag{2.45}$$

where

$$\nabla \cdot g \equiv \sum_i \frac{\partial}{\partial \theta_i} g_i.$$

Because the soft thresholding estimator is weakly differentiable, Donoho and Johnstone (1995) show that an unbiased estimate for the squared error loss is given by

$$SURE(d, \lambda)) = n - 2 \cdot \#\{i : |x_i| \leq \lambda\} + \sum_{i=1}^{d} (|x_i| \wedge \lambda)^2,$$

where

$$E(SURE(d, \lambda)) = E_\theta||\hat{\theta}(d) - \theta||^2.$$

The optimal threshold is then given by

$$\lambda^S = \text{argmin}_{0 \leq t \leq \sqrt{2 \log n}} SURE(d, \lambda).$$

Cross-validation

Wavelet thresholding based on cross-validation was developed by Nason (1995, 1996). Traditional cross-validation partitions a data set into subsamples where

an initial statistical analysis is performed on one of the subsamples. Results are afterwards compared with the remaining data. For standard reference see Burman (1989). Nason (1996) assumes model (2.33) with $n = 2^J$, $J \in \mathbb{N}$, and develops an adjusted cross-validation method[2]:

First, the data set is partitioned into groups of even- and odd-indexed observations. After re-indexing each group from $j = 1, \ldots, 2^{J-1}$, each subsample contributes an estimate for f by applying the universal threshold in (2.43) (denoted by $\hat{f}^E_\lambda(j)$ and $\hat{f}^O_\lambda(j)$). For reasons of comparison, an interpolated version of each estimate is used, i.e.

$$\overline{f}^E_\lambda(j) = \frac{1}{2}(\hat{f}^E_\lambda(j+1) + \hat{f}^E_\lambda(j))\,(j = 1, \ldots, \frac{n}{2}), \tag{2.46}$$

where $\overline{f}^E_\lambda(\frac{n}{2}+1) = \overline{f}^E_\lambda(1)$, because f is assumed to be periodic. Let the corresponding estimate for the odd-indexed observations be denoted by $\overline{f}^O_\lambda(j)$. Then,

$$\hat{M}(\lambda) = \sum_{j=1}^{n/2} \left((\overline{f}^E_\lambda(j) - Y(2j+1))^2 + (\overline{f}^O_\lambda(j) - Y(2j))^2 \right) \tag{2.47}$$

is a measure for the integrated squared error (ISE) of $\hat{f}$. Minimizing this over all λ gives the optimal threshold λ_{UV}. Because each initial set contains only one half of the data set, an adjustment of the threshold of the form

$$\lambda_{UV}(n) \approx \lambda_{UV}(n/2) \cdot \left(1 - \frac{\log 2}{\log n}\right)^{-1/2}$$

is applied.

The author shows that this estimate works best under independent normal or Student's t distribution. However, in case of serial correlation performance of other estimators such as VisuShrink or SureShrink is by far better. For further reference on problems with cross-validation in correlated samples see e.g. Altman (1990), Diggle (1990) or Hart (1994). Extensions to more generalized cross-validation methods including the case of correlated noise can be found

[2]The restricted sample size is due to application of *Mallat's* algorithm. Nason (1994) also mentions a possible extension to sample sizes deviating from 2^J.

in Jansen and Bultheel (1997) and for non-equidistant observations in Jansen and Bultheel (1998).

Bayesian thresholding

A Bayesian approach to wavelet thresholding is introduced by Vidakovic (1994). Under model (2.33), $d_i \sim \mathcal{N}(\theta_i, \sigma^2)$ with θ_i, σ^2 unknown. The Bayesian idea is to impose priors for θ_i and σ^2 and to derive implications for the resulting estimates. *Bayesian factor thresholding* (Vidakovic) assumes σ^2 to follow an exponential distribution with parameter μ,

$$\sigma^2 \sim \mathcal{E}(\mu).$$

A double exponential is imposed on d given θ, i.e.,

$$d|\theta \sim \mathcal{DE}(\theta, \frac{1}{\sqrt{2\mu}}),$$

and to allow for heavy tails, θ is t-distributed with mean 0, parameter τ and n degrees of freedom,

$$\theta \sim t_n(0, \tau),$$

The next theorem summarizes the results of Vidakovic (1994).

Theorem 2.5. *Assume that condition (2.33) holds and let the prior of θ be given by*

$$\pi(\theta) = \pi_0\delta_0 + \pi_1\zeta(\theta),$$

where δ_0 is the Dirac mass at 0, π_0 is the probability for $\theta = 0$, $\pi_1 = 1 - \pi_0$ and $\zeta(\theta)$ is the prior distribution (Student's t) of θ for $\theta \neq 0$. Let the Laplace transformations of $\zeta(\theta - d)$ and $\zeta(\theta + d)$ be given by Π_1 and Π_2, then the resulting thresholding estimate of θ_i is

$$\hat{\theta}_i = d_i \mathbf{1}_{\left\{\frac{\pi_0 e^{-c|d_i|}}{\pi_0 e^{-c|d_i|} + \pi_1(\Pi_1(c) + \Pi_2(c))} < \frac{1}{2}\right\}}, \tag{2.48}$$

where $c = 2\mu$.

This approach results in a simple representation of the shrinkage rule that is nevertheless effective.

In the discussion of this work, several other papers on this topic have been published. Amongst others, Chipman et al. (1997) assume a linear combination of normal distributions for the $\theta_{j,k}$ (level-dependent) of the form

$$\theta_{j,k}|\gamma_{j,k} \sim \gamma_{j,k}\mathcal{N}(0, c_j^2\tau_j^2) + (1-\gamma_{j,k})\mathcal{N}(0, \tau_j^2),$$

where $P(\gamma_{j,k}=1) = 1 - P(\gamma_{j,k}=0) = p_j$ such that

$$d_{j,k}|\theta_{j,k}, \sigma^2 \sim \mathcal{N}(\theta_{j,k}, \sigma^2).$$

Ruggeri and Vidakovic (1999) show that in case of an accurate prior the MSE of a threshold can be smaller than in one of the previously mentioned threshold selection procedures. De Candiis and Vidakovic (2001) assume the joint prior to be normal-inverse gamma distributed and use the obtained results to develop a block thresholding method. See also Abramovich, Sapatinas and Silverman (1998) for a connection between the parameters in the prior and that of the Besov space of realizations of the prior. Finally, Ruggeri and Vidakovic (2001) give a comprehensive overview over the current literature on Bayesian modeling in wavelet regression.

Block Thresholding

To our knowledge Efromovich (1985) for estimators based on general orthogonal series and Hall et al. (1995) for wavelets first introduce the idea of block thresholding into nonparametric regression. Block thresholding is based on a grouping of wavelet coefficients. With the aim of improved accuracy in function estimation, a threshold is then applied to each member of a group simultaneously. Hall et al. (1997) propose to partition neighboring coefficients on each level j into groups b of length $L = L(n)$. An estimate for the average

of squared wavelet coefficients in group b given level j,

$$B_{jb} = L^{-1} \sum_{(b)} \beta_{j,k}^2, \tag{2.49}$$

is given by

$$\hat{B}_{jb} = L^{-1} \sum_{(b)} \hat{\beta}_{j,k}^*,$$

with $\hat{\beta}_{j,k}^*$ being an estimator of the form (2.28) adjusted to show particularly low bias (compare with Hall et al. (1997)). A block thresholding estimator is then given by

$$\sum_k \hat{\alpha}_{l,k}\phi_{l,k} + \sum_{j \geq l}^{J_n} \sum_b \left(\sum_{(b)} \hat{\beta}_{j,k}\psi_{j,k} \right) \mathbf{1}_{\{\hat{B}_{j,k} > \lambda^2\}}, \tag{2.50}$$

It can be shown, that the optimal threshold in sense of the mean integrated squared error (MISE) is given by $\lambda = \sigma^2/n$ and thus differs from the universal threshold by a logarithmic factor. This implies that the corresponding risk is of order $O(n^{-1})$ as well. More explicitly, block thresholding estimates are able to obtain optimal minimax rates in contrast to the near-optimality of the previous techniques (involving a logarithmic factor). Due to this logarithmic factor universal thresholds tend to overfit the function and thus lack of local adaptivity in spatially inhomogenous regions of the function.

In a similar attempt to Donoho and Johnstone (1994), Cai (1999) analyzes block thresholding furnitured with an oracle based on grouping as in Hall et al. (1997) and defines a shrinkage estimate based on a James-Stein-type rule called BlockJS (see also Vidakovic (1999, p. 206) for a concise description). Essential results are summarized in

Theorem 2.6. *(Cai) Let ψ be a wavelet with r vanishing moments and r continous derivatives and let $f \in B_{pq}^{\alpha}(M)$, where M is the upper bound of the corresponding Besov norm. Then BlockJS satisfies*

$$\sup_{f \in B_{p,q}^{\alpha}(M)} E||\hat{f}_n - f||^2 \leq C n^{-\frac{2\alpha}{1+\alpha}} \tag{2.51}$$

for all $M \in (0, \infty)$, $\alpha \in (0, r), q \in [1, \infty]$ and $p \in [2, \infty]$.

Hence, the block thresholding estimate BlockJS attains the minimax rate over a range of Besov spaces. In a later study, Hall et al. (1999) prove that the block thresholding estimate in (2.50) also attains optimal rates of convergence in minimax sense as in (2.51).

In the last years several other block thresholding approaches have been proposed. Hall et all. (1998) consider block thresholding both in kernel and wavelet estimation and derive optimal rates of convergence. Cai (2002) analyzes the block length in relation to global and pointwise adaptivity. It is shown that local and global adaptivity are antagonistic aims in view of the optimal individual block lengths. Cai and Silverman (2001) extend the idea to block thresholding with overlapping blocks and prove the minimax risk for the resulting estimate. Cai and Zou (2005) provide a data-driven procedure based on Stein's unbiased estimate of risk. In a recent study Chicken and Cai (2005) apply block thresholding in the context of density estimation. They show that a block length of $\log n$ results in an optimal minimax rate of convergence over a range of Besov spaces. The estimator attains the minimax rate both in global sense over Besov spaces and in local (pointwise) sense. For further reference on block thresholding methods see Cai and Low (2005).

2.3.2 Further methods and applications

Early publications on statistical applications of wavelets were concerned with density estimation (see e.g. Doukhan (1988)). Kerkyacharian and Picard (1992, 1993) considered estimation if the unknown density belongs to some Besov space and asymptotic properties of thresholding estimates are investigated in Donoho et al. (1996). See also Donoho (1993), Antoniadis and Carmona (1991), Gao (1996, 1997), Huang (1999) amongst many other authors.

Silverman (1999) considers several extensions to the basic model $Y(i) =$

$f(t_i) + \epsilon(i)$ including correlated errors, non-equispaced data and heavy-tailed noise. Generalizations to non-equispaced samples have been investigated by many authors. Sardy et al. (1998) consider four different approaches to modeling non-equispaced time series in case of the Haar basis. First, time points are analyzed as if they were equidistant which yields a simple representation of risk. Second, modification of the basis to a nonorthogonal transformation is considered. The third approach uses interpolation of time points and the forth method is based on the continous wavelet transform, thus allowing the sample size to deviate from a power of 2. Interpolation is also considered in Vanraes et al. (2002). In applications they show the good performance of their approach. Cai and Brown (1998) model the non-equidistant character by using a cumulative density function for time t_i on $[0, 1]$. They show that their shrinkage method attains a risk that is within a logarithmic multiple of minimax risk. Another way of dealing with non-equispaced time series is described in Pensky and Vidakovic (1998). They consider the time series as data (x_i, y_i) $(i = 1, \ldots, n)$ and propose a probabilistic model for the x_i to estimate $E(Y|X = x)$, i.e. the regression function.

Extensions to the non-$i.i.d$ case are considered e.g. in Neumann and von Sachs (1995). It is shown, how inequalities of the form (2.51) can be developed in case of dependent errors. In case of correlated errors, Johnstone and Silverman (1997) present several thresholds and demonstrate the practical performance in a few applications. For wavelet shrinkage involving long-memory see Wang (1996).

Selection of a basis that is chosen from a library of bases is addressed in Donoho (1993) and Donoho and Johnstone (1994b). This is the natural extension of the classical shrinkage approach where a basis is given and the optimization is performed by selecting the threshold λ.

Many other topics in the context of wavelet thresholding have been published. They contain shinkage based on extreme value theory, minimum description length or on likelihood ration tests. For a more comprehensive overview see

Vidakovic (1999).

Asymptotic distribution of shrunken wavelet coefficients and function estimates is described in Brillinger (1994, 1996). Due to the importance for further considerations in the subsequent chapters, these results are more thoroughfully presented in the following section.

2.3.3 Asymptotic distribution of thresholding estimates

In this section we consider asymptotic properties of thresholding estimates in presence of a level-dependent threshold. A hard thresholding rule will be applied with the threshold specified as in the following lemma. We introduce the following assumption.

(A4) The wavelet representation of the true trend function f contains only a finite number of nonzero wavelet coefficients.

According to (A4), denote J_0 the largest j in the wavelet representation of f such that there exists at least one pair (j,k) with $\beta_{j,k} \neq 0$. We then have

Lemma 2.7. *Assume that assumptions (A1)-(A4) hold and let the threshold* λ_j *be chosen such that* $\lambda_j^* = \sqrt{var(\hat{\beta}_{j,k})}\lambda_j = O(n^{-\delta})$ *(*$n \to \infty$*) for some* $0 < \delta < 1$*. Then, for* n *large such that* $J_n \geq J_0$*,*

$$E[\int (\hat{f}(x) - f(x))^2 dx] \;\; \leq \;\; O(n^{-1}(J_n + C(n))),$$

where $C(n)$ *is an increasing function of* n *depending on the choice of the threshold* λ_j^**.*

Proof:

$$\begin{aligned} E[\int (\hat{f}(x) - f(x))^2 dx] &= \sum_k E((\hat{\alpha}_{l,k} - \alpha_{l,k})^2) \\ &\quad + \sum_{j \le J_0} E((\hat{w}_{j,k}\hat{\beta}_{j,k} - \beta_{j,k})^2) + \sum_{j > J_0} E((\hat{w}_{j,k}\hat{\beta}_{j,k})^2) \\ &= S_1 + S_2 + S_3, \end{aligned}$$

where the first equality follows from the orthonormality of wavelet functions. Use earlier results (cf. theorem 2.4.) to get

$$\sum_k E((\hat{\alpha}_{l,k} - \alpha_{l,k})^2) = O(n^{-1}).$$

First, assume $\beta_{j,k} \le \lambda^*$. We estimate

$$(\hat{w}_{j,k}\hat{\beta}_{j,k} - \beta_{j,k})^2 \le (\hat{\beta}_{j,k} - \beta_{j,k})^2 + \beta_{j,k}^2$$

such that

$$E[(\hat{w}_{j,k}\hat{\beta}_{j,k} - \beta_{j,k})^2] = O(n^{-1}) \ \ (j > J_0).$$

Therefore,

$$S_3 = O(J_n n^{-1}).$$

For $\beta_{j,k} > \lambda_j^*$, note that

$$\mathbf{1}_{\{|\hat{\beta}_{j,k}| \le \lambda_j^*\}} \le \mathbf{1}_{\{|\hat{\beta}_{j,k} - \beta_{j,k}| > \lambda_j^*/2\}} + \mathbf{1}_{\{|\hat{\beta}_{j,k}| \le \lambda_j^*, |\beta_{j,k}| \le 2\lambda_j^*\}},$$

where the inequality is obtained by noticing that $|\beta_{j,k}| > 2\lambda_j^*$ together with $|\hat{\beta}_{j,k}| \le \lambda_j^*$ implies $|\hat{\beta}_{j,k} - \beta_{j,k}| > \frac{\lambda_j^*}{2}$ and $|\hat{\beta}_{j,k}| \le \frac{|\beta_{j,k}|}{2}$. Then, $|\hat{\beta}_{j,k} - \beta_{j,k}| \ge |\beta_{j,k}| - |\hat{\beta}_{j,k}| \ge |\beta_{j,k}|/2$. We summarize

$$\mathbf{1}_{\{|\hat{\beta}_{j,k}| \le \lambda_j^*\}} \le \mathbf{1}_{\{|\hat{\beta}_{j,k} - \beta_{j,k}| > \lambda_j^*/2\}} + \mathbf{1}_{\{|\beta_{j,k}| \le 2\lambda_j^*\}}. \tag{2.52}$$

If $\beta_{j,k} > 0$ and $\lambda^* \to 0$ $(n \to \infty)$, then there exists $N_0 \in \mathbb{N}$ such that for all $n \ge N_0$ the last indicator function in (2.52) vanishes. For those large n we conclude

$$E[(\hat{w}_{j,k}\hat{\beta}_{j,k} - \beta_{j,k})^2] \ \le \ E(\hat{\beta}_{j,k} - \beta_{j,k})^2 + \beta_{j,k}^2 P(|\hat{\beta}_{j,k} - \beta_{j,k}| > \lambda_j^*/2).$$

Applying Chebyshev's inequality gives

$$P(|\hat{\beta}_{j,k} - \beta_{j,k}| > \lambda_j^*/2) \leq 4\frac{\text{var}\,(\hat{\beta}_{j,k})}{\lambda_j^{*2}},$$

where the variance is of order $O(n^{-1})$. This proves the lemma.

■

Remark 2.9. *Note that for small sample sizes ($J_n < J_0$) an additional term taking account of the bias of $\hat{f}$ has to be introduced in lemma 2.7.*

Asymptotic distribution of thresholding estimates is discussed in Brillinger (1996). Instead of presenting a univariate limit theorem, we directly consider multivariate extensions to Brillinger's results.
Assume that the observed time series $\mathbf{Y}(i) = (Y_1(i), ..., Y_p(i))^T$ is of the form

$$\mathbf{Y}(i) = \mathbf{f}(t_i) + \boldsymbol{\epsilon}(i), \tag{2.53}$$

where $t_i = i/n$ $(i = 1, \ldots, n)$, $\mathbf{f}(t) = (f_1(t), ..., f_p(t))^T \in \mathbb{C}^p$ $(t \in \mathbb{R})$ is a multivariate deterministic trend function and $\boldsymbol{\epsilon}(i)$ is zero mean stationary noise.

(A5) Let multivariate versions of assumptions (A2)-(A4) hold (i.e. f, γ_ϵ, c_m and C_m defined accordingly). Assume further that the accumulated cumulant function C_m satisfies

$$\sum_m C_m z^m/m! < \infty$$

in a small neighborhood of $z = 0$.

Then, the results of Brillinger are easily extended to a multivariate context.

Lemma 2.8. *Let the assumptions (A1)-(A5) hold. For* $1 \leq r \leq p$, *let*

$$\hat{\alpha}_{l,k}^{(r)} = \frac{1}{n}\sum_{i=1}^{n} \phi_{l,k}(t_i) Y_r(i), \tag{2.54}$$

$$\hat{\beta}_{j,k}^{(r)} = \frac{1}{n}\sum_{i=1}^{n} \psi_{j,k}(t_i) Y_r(i), \tag{2.55}$$

$$\hat{w}_{j,k}^{(r)} := \mathbf{1}_{\{|\hat{\beta}_{j,k}^{(r)}| \geq \sqrt{\mathrm{var}\,(\hat{\beta}_{j,k}^{(r)})}\lambda_j\}},$$

$J_n \to \infty$, $n2^{-J_n/2} \to \infty$, λ_j *such that* $2^{\frac{j}{2}}\lambda_j = o(n^{1/2})$ $(j = l, l+1, \ldots, J_n)$ *and*

$$\sum_{j>l}^{J_n} 2^{\frac{j}{2}} \exp(-\lambda_j^2/(1+\eta)2) = o(1)$$

for some $\eta > 0$. *Define*

$$\hat{f}_r(t) := \sum_k \hat{\alpha}_{l,k}^{(r)} \phi_{l,k}(t) + \sum_{j\geq l}^{J_n}\sum_k \hat{w}_{j,k}^{(r)} \hat{\beta}_{j,k}^{(r)} \psi_{j,k}(t), \tag{2.56}$$

$1 \leq r \leq p$. *Also denote by* $\mathbf{h}_\epsilon = (h_{\epsilon;rs})_{1\leq r,s\leq p}$ *the matrix of cross spectral densities between* $\{\epsilon_r(i), i \in \mathbb{N}\}$ *and* $\{\epsilon_s(i), i \in \mathbb{N}\}$. *Then, almost everywhere in* $t \in [0,1]$, *finite collections of* $\hat{f}_r(t)$ *are asymptotically normal with mean* $f_r(t)$ *and covariance function*

$$\gamma_f(x,y;r,s) = \mathrm{cov}\,(\hat{f}_r(x), \hat{f}_s(y)) = \frac{2\pi h_{\epsilon;rs}(0)}{n}\sum_k \phi_{l,k}(x)\phi_{l,k}(y)$$
$$+ \frac{2\pi h_{\epsilon;rs}(0)}{n}\sum_{j,k}^{J_0} w_{j,k}^{(r)} w_{j,k}^{(s)} \psi_{j,k}(x)\psi_{j,k}(y) + r(n),$$

where $w_{j,k}^{(i)} = \mathbf{1}_{\{\beta_{j,k}^{(i)} \neq 0\}}$, $J_0 = J_0(r,s)$ *is the largest common integer such that* $w_{j,k}^{(r)} w_{j,k}^{(s)} \neq 0$ *for some* $j = J_0$, *and* $r(n) = O(2^{2J_n} n^{-2})$.

Proof: The proof essentially follows directly from Theorem 3 in Brillinger (1994). To see the uniform convergence of $r(n)$ consider

$$cov(\hat{f}_r(x), \hat{f}_s(y)) = \sum_{k,k'} cov(\hat{\alpha}_{l,k}, \hat{\alpha}_{l,k'})\phi_{l,k}(x)\phi_{l,k'}(y) \tag{2.57}$$

$$+ \sum_{j,k,j',k'} cov\,(\hat{w}^{(r)}_{j,k}\hat{\beta}^{(r)}_{j,k}, \hat{w}^{(s)}_{j',k'}\hat{\beta}^{(s)}_{j',k'})\psi_{j,k}(x)\psi_{j',k'}(y) \tag{2.58}$$

$$+ \sum_{k,j',k'} cov\,(\hat{\alpha}_{l,k}, \hat{w}^{(s)}_{j',k'}\hat{\beta}^{(s)}_{j',k'})\phi_{l,k}(x)\psi_{j',k'}(y) \tag{2.59}$$

$$+ \sum_{j,k,k'} cov\,(\hat{w}^{(r)}_{j,k}\hat{\beta}^{(r)}_{j,k}, \hat{\alpha}_{l,k'})\psi_{j,k}(x)\phi_{l,k'}(y). \tag{2.60}$$

The remainder $r(n)$ essentially consists of terms with $(j,k) \neq (j',k')$. Separate e.g. (2.58) into

$$\sum_{j,j' \leq J_0} \sum_{k,k'} cov\,(\hat{w}^{(r)}_{j,k}\hat{\beta}^{(r)}_{j,k}, \hat{w}^{(s)}_{j',k'}\hat{\beta}^{(s)}_{j',k'})\psi_{j,k}(x)\psi_{j',k'}(y)$$

$$+ \sum_{j > J_0 \vee j' > J_0} \sum_{k,k'} cov\,(\hat{w}^{(r)}_{j,k}\hat{\beta}^{(r)}_{j,k}, \hat{w}^{(s)}_{j',k'}\hat{\beta}^{(s)}_{j',k'})\psi_{j,k}(x)\psi_{j',k'}(y). \tag{2.61}$$

For $j, j' \leq J_0$ and suitable constants A_1 and A_2, we have

$$\begin{aligned}
&|\sum_{j \neq j'}^{J_0} \sum_{k \neq k'} \text{cov}\ (\hat{w}^{(r)}_{j,k}\hat{\beta}^{(r)}_{j,k}, \hat{w}^{(s)}_{j',k'}\hat{\beta}^{(s)}_{j',k'})\psi_{j,k}(x)\psi_{j',k'}(y)| \\
&\leq \sum_{j \neq j'}^{J_0} \sum_{k \neq k'} |\text{cov}\ (\hat{w}^{(r)}_{j,k}\hat{\beta}^{(r)}_{j,k}, \hat{w}^{(s)}_{j',k'}\hat{\beta}^{(s)}_{j',k'})| \cdot 2^{\frac{j}{2}} 2^{\frac{j'}{2}} \sup_x \{\psi(x)^2\} \\
&\leq A_1 \sum_{j \neq j'}^{J_0} \sum_{k \neq k'} |\text{cov}\ (\hat{\beta}^{(r)}_{j,k}, \hat{\beta}^{(s)}_{j',k'})| 2^{\frac{j}{2}} 2^{\frac{j'}{2}} \\
&\leq A_2 \sum_{j \neq j'}^{J_0} \sum_{k \neq k'} n^{-2} 2^j 2^{j'} = O(n^{-2}).
\end{aligned}$$

Now for $j > J_0$ or $j' > J_0$, we have, recalling that $|cov\,(\hat{\beta}^{(r)}_{j,k}, \hat{\beta}^{(s)}_{j',k'})| = O(2^{j/2}2^{j'/2}n^{-2})$ for $(j,k) \neq (j',k')$,

$$\begin{aligned}
&\left| \sum_{j>J_0 \vee j'>J_0} \sum_{k,k'} \mathrm{cov}\,(\hat{w}^{(r)}_{j,k}\hat{\beta}^{(r)}_{j,k}, \hat{w}^{(s)}_{j',k'}\hat{\beta}^{(s)}_{j',k'})\psi_{j,k}(x)\psi_{j',k'}(y) \right| \\
&\leq \sum_{j>J_0 \vee j'>J_0} \sum_{k,k'} \left| \mathrm{cov}\,(\hat{w}^{(r)}_{j,k}\hat{\beta}^{(r)}_{j,k}, \hat{w}^{(s)}_{j',k'}\hat{\beta}^{(s)}_{j',k'}) \right| |\psi_{j,k}(x)\psi_{j',k'}(y)| \\
&\leq \sum_{j>J_0 \vee j'>J_0} \sum_{k,k'} \left| \mathrm{cov}\,(\hat{\beta}^{(r)}_{j,k}, \hat{\beta}^{(s)}_{j',k'}) \right| \cdot A_1 2^{j/2}2^{j'/2} \\
&\leq A_2 n^{-2} \sum_{j>J_0 \vee j'>J_0} 2^j 2^{j'} \sup_x\{\psi(x)^2\} \\
&= O(2^{2J_n}n^{-2}), \qquad (2.62)
\end{aligned}$$

uniformly in x and y where A_1, A_2 are suitable constants. ■

Remark 2.10. *One possible threshold that satisfies the assumptions of lemma 2.8. is given by* $\lambda_j = \sqrt{2\log(2^{-j}n)}$.

Remark 2.11. *Note that additional asymptotic properties for the wavelet coefficients are also easily carried over to the multivariate case. For instance,* $\mathrm{cov}\ (\hat{\beta}^{(r)}_{j,k}, \hat{\beta}^{(s)}_{j',k'}) = O(n^{-2})$ *for* $(j,k) \neq (j',k')$.

Remark 2.12. *If* $\mathrm{var}(\hat{\beta}^{(r)}_{j,k})$ *is unknown, then the variance has to be estimated. See Brillinger (1994) for consistent estimation of the variance and asymptotic properties.*

Remark 2.13. *Lemma 2.8. implies* $\mathrm{var}\,(\hat{f}_r(t)) = O(n^{-1})$, *uniformly in* t, *so that* $\hat{f}_r(t)$ *is a weakly consistent estimator of* $f_r(t)$, *i.e.* $\hat{f}_r(t) \to f_r(t)$ *in probability for almost all* t *in* $[0,1]$.

Chapter 3

A nonparametric regression spectrum for multivariate time series

3.1 Introduction

Here, we consider dependence structures in multivariate time series that are due to similarities in underlying deterministic trends. Suppose that we observe a multivariate time series $\mathbf{Y}(i) = (Y_1(i), ..., Y_p(i))^T$, $(i = 1, \ldots, n)$. In classical spectral analysis, a time series and its autocorrelations are decomposed into sinusoidal components. Grenander and Rosenblatt (1957) extended the idea of spectral decomposition to parametric regression with deterministic explanatory variables. They show that consistency and efficiency of least squares estimates depend on the regression spectrum. For further results on the interplay between regression spectrum and spectral properties of the stochastic part see e.g. Yajima (1988).

In this chapter, spectral analysis of regression functions is extended to multi-

variate *nonparametric* trend functions that are estimated by wavelet thresholding. The definitions of the regression cross covariance and regression cross spectrum are adapted to this context. Asymptotic properties of estimates of these quantities are derived. Specifically, this chapter is organized as follows. Basic definitions are given in section 3.2. Estimation of the regression spectrum, the regression cross covariances and their asymptotic distribution are considered in section 3.3. Alternative estimation involving Fourier coefficients is presented in section 3.4. Algorithmic issues and a data example are discussed in section 3.5. Final remarks in section 3.6 conclude the chapter.

3.2 Definition of the regression cross covariance and spectrum

3.2.1 Cross covariance and correlation

Assume that the observed time series $\mathbf{Y}(i) = (Y_1(i), ..., Y_p(i))^T$ is of the form

$$\mathbf{Y}(i) = \mathbf{f}(t_i) + \boldsymbol{\epsilon}(i), \tag{3.1}$$

where $t_i = i/n$ $(i = 1, \ldots, n)$, $\mathbf{f}(t) = (f_1(t), ..., f_p(t))^T \in \mathbb{C}^p$ $(t \in \mathbb{R})$ is a multivariate deterministic trend function and $\boldsymbol{\epsilon}(i)$ is zero mean stationary noise. We will assume that $f_r(t)$ is Lebesque measurable and $\int_0^1 |f_r(t)|^2 dt < \infty$, i.e. $f_r \in L^2$ where $L^2 = L^2(\mathbb{C})$ denotes the space of complex-valued functions that are square integrable on $[0, 1]$. For $f_r, f_s \in L^2$, we define

$$< f_r, f_s >= \int_0^1 f_r(t)\overline{f_s(t)}dt$$

Note that $<,>$ is a nonnegative sesquilinear form. If we restrict attention to functions that are periodic with period 1, then $<,>$ is a scalar product and the corresponding space $L^2[0, 1]$ is a Hilbert space. Also note that, since

$||f_r||^2 =< f_r, f_r >< \infty$, the mean vector

$$\mathbf{m}(f) = \int_0^1 \mathbf{f}(t)dt$$

is well defined and finite. Without loss of generality (and since $\mathbf{m}(f)$ can easily be estimated and subtracted from the data), we will from now on assume $\mathbf{m}(f) = 0$. The corresponding space of functions will be denoted by $L_o^2[0,1] = \{f \in L^2[0,1] : \mathbf{m}(f) = 0\}$.

We first define the autocorrelation function of $\mathbf{f}(t) \in L_o^2$. Grenander and Rosenblatt (1957) introduced a definition of cross- and autocorrelation $\rho_{rs}(u)$ between parametric (i.e. known) regression functions $\phi_r(t), \phi_s(t)$ with $t \in \mathbb{N}$ (also see Priestley (1989), chapter 7). In contrast, here, we consider completely unknown regression functions $f_j(t) \in L_o^2$ that are observed on an increasingly fine grid $t_1, t_2, ..., t_n$ of $t-$values in $[0,1]$. A natural modification of the definition by Grenander and Rosenblatt definition is therefore

$$\gamma_{rs}(u) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f_r(t_i + u)\overline{f_s(t_i)}. \ (u \in [-1,1])$$

and

$$\rho_{rs}(u) = \frac{\gamma_{rs}(u)}{\sqrt{\gamma_r(0)\gamma_s(0)}}.$$

This leads to

Definition 3.1. *Let* $\mathbf{f}(t) = (f_1(t), ..., f_p(t))^T \in \mathbb{C}^p$ $(t \in \mathbb{R})$ *be a p-dimensional deterministic function as defined above, and such that, for each* $r \in \{1, 2, ..., p\}$, $f_r \in L_o^2$ *and* $||f_r(t)|| > 0$. *Then the regression (cross-) covariance function* $\mathbf{\Gamma}(u) = [\gamma_{rs}(u)]_{r,s=1,...,p}$ *and the regression (cross-) correlation function* $\mathbf{R}(u) = [\rho_{rs}(u)]_{r,s=1,...,p}$ *of* $\mathbf{f}(t)$ *are defined by*

$$\gamma_{rs}(u) =< f_r(\cdot + u), f_s >= \int_0^1 f_r(t + u)\overline{f_s(t)}dt$$

and

$$\rho_{rs}(u) = \frac{\gamma_{rs}(u)}{\sqrt{\gamma_r(0)\gamma_s(0)}}. \tag{3.2}$$

The Hermitian property and non-negative definiteness of $\rho_{rs}(\cdot)$ are obtained in the following

Proposition 3.1. *The regression cross correlation function $\rho_{rs}(u)$ defined in (3.2) is Hermitian and non-negative definite.*

Proof: For $u \geq 0$,

$$\rho_{rs}(-u) = \frac{< f_r(\cdot - u), f_s >}{||f_r|| \cdot ||f_s||} = \frac{\overline{< f_s(\cdot + u), f_r >}}{||f_r|| \cdot ||f_s||} = \overline{\rho_{sr}(u)}.$$

Hence ρ_{rs} is Hermitian. Consider now ρ_{rs}, $u_1, \ldots, u_n \in [0, 1]$, and let $\theta_1, \ldots, \theta_n$ be arbitrary coefficients in $\mathbb{C}$. Then

$$\begin{aligned} &\sum_{i,j=1}^{n} \sum_{r,s=1}^{p} \theta_i \rho_{rs}(u_i - u_j) \overline{\theta_j} \\ &= \sum_{i,j} \sum_{r,s} \theta_i \frac{\int f_r(x + u_i - u_j) \overline{f_s(x)} dx}{\sqrt{\int |f_r(x)|^2 dx \int |f_s(x)|^2 dx}} \overline{\theta_j}. \end{aligned}$$

Since the denominator does not depend on i and j, non-negative definiteness follows from

$$\begin{aligned} &\int \sum_r \sum_i \theta_i f_r(x + u_i) \sum_s \sum_j \overline{\theta_j f_s(x + u_j)} dx \\ &= \int \left| \sum_r \sum_i \theta_i f_r(x + u_i) \right|^2 dx \geq 0. \end{aligned}$$

■

In practice, γ_{rs} and ρ_{rs} have to be estimated, since they depend on the unobservable function $\mathbf{f}(t)$.

3.2.2 The spectrum

The function $\mathbf{f}(t)$ is observed for time points in the interval $[0,1]$ only. To extrapolate $\mathbf{f}$ beyond the unit interval, we will assume that $\mathbf{f}(t)$ can be decomposed into a nonperiodic "long-term" trend component $\boldsymbol{\mu}(t)$ ($t \in \mathbb{R}$) and a component $\boldsymbol{\alpha}(t)$ with $\boldsymbol{\alpha}(\mathrm{t+T}) = \boldsymbol{\alpha}(t)$ for some $T \leq \frac{1}{2}$. To simplify presentation, it will be assumed throughout the paper that $\boldsymbol{\mu} \equiv 0$, or $\boldsymbol{\mu}$ has been estimated and removed from the data. Thus,

$$\mathbf{f}(t) = \boldsymbol{\alpha}(t)$$

with $\boldsymbol{\alpha}$ periodic with period $T \leq \frac{1}{2}$.

Remark 3.1. *The assumption of strict periodicity can be replaced by local periodicity, allowing the periodic shape of $\boldsymbol{\alpha}$ to change smoothly in time (see e.g. Heiler and Feng (2000)). In this case, $\gamma_{rs}(u)$ can be approximated by extrapolating $\boldsymbol{\alpha}$, e.g. using local trigonometric polynomials (Heiler and Feng (2000)) together with nonparametric extrapolation (see e.g. Beran and Ocker (1999)).*

Consider now the characterization of $\mathbf{f}$ in the frequency domain. For $t \in [0,1]$ and $f_r \in L_o^2$ such that $\int_0^1 |f_r(t)|dt < \infty$, we may write

$$\mathbf{f}(t) = \sum_{j=-\infty}^{\infty} \mathbf{a}(j) e^{i2\pi jt}$$

where $\mathbf{a}(j) = (a_1(j), a_2(j), ..., a_p(j))^T \in \mathbb{C}^p$ are given by

$$a_r(j) =< f_r, e^{i2\pi j \cdot} >= \int_0^1 f_r(t) e^{-i2\pi jt} dt.$$

Note that

$$\mathbf{m}(f) = \mathbf{a}(0) = 0$$

and Parseval's equation yields

$$||f_r||^2 = \sum_{j=-\infty}^{\infty} |a_r(j)|^2.$$

For the covariance function we then have

$$\mathbf{\Gamma}(u) = \mathbf{\Gamma}_\alpha(u)$$

where $\mathbf{\Gamma}_\alpha = [\gamma_{\alpha;rs}]_{r,s=1,\ldots,p}$ is a $p \times p$ matrix with

$$\gamma_{\alpha;rs}(u) =< \alpha_r(\cdot + u), \alpha_s >$$

More explicitly, we have

$$\begin{aligned}\mathbf{\Gamma}_\alpha(u) &= \int_0^1 \boldsymbol{\alpha}(t+u)\overline{\boldsymbol{\alpha}^T(t)}dt \\ &= \sum_{j=-\infty}^{\infty} e^{i2\pi ju}\mathbf{a}(j)\overline{\mathbf{a}^T(j)}\end{aligned}$$

and we may introduce the following

Definition 3.2. *The sequence of $p \times p$ matrices $\mathbf{H}(j) = [h_{rs}(j)]_{r,s=1,\ldots,p}$ ($j \in \mathbb{Z}$) defined by*

$$\mathbf{H}(j) = \mathbf{a}(j)\overline{\mathbf{a}^T(j)}$$

is called regression spectrum of $\boldsymbol{\alpha}$.

By definition we have the following relationship between regression spectrum and covariances:

$$\mathbf{\Gamma}_\alpha(u) = \sum_{j=-\infty}^{\infty} \mathbf{H}(j)e^{i2\pi ju}$$

and

$$\mathbf{H}(j) = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{-i2\pi ju}\mathbf{\Gamma}_\alpha(u)du.$$

Writing $\mathbf{H}(j)$ in polar representation, the contribution of frequency j can also be expressed in relative terms as follows.

Definition 3.3. *Let $\mathbf{H}(j) = [h_{rs}(j)]_{r,s=1,\ldots,p}$ be defined as above. Then, $\tilde{\mathbf{H}} = [\tilde{h}_{rs}]_{r,s=1,\ldots,p}$ with*

$$\tilde{h}_{rs}(j) = \frac{h_{rs}(j)}{\sqrt{\gamma_{\alpha;rr}(0)\cdot\gamma_{\alpha;ss}(0)}} = \frac{|a_r(j)\overline{a_s(j)}|}{\sqrt{\gamma_{\alpha;rr}(0)\cdot\gamma_{\alpha;ss}(0)}}\exp(i\phi_{rs}(j))$$

is called the standardized regression spectrum of $\boldsymbol{\alpha}$*,*

$$\kappa_{rs}(j) = \frac{|h_{rs}(j)|}{\sqrt{\gamma_{\alpha;rr}(0) \cdot \gamma_{\alpha;ss}(0)}} = \frac{|h_{rs}(j)|}{||\alpha_r|| \cdot ||\alpha_s||}$$
$$= \frac{|a_r(j)\overline{a_s(j)}|}{\sqrt{\sum_l |a_r(l)|^2 \sum_m |a_s(m)|^2}}$$

is the relative spectral modulus and $\phi_{rs}(j)$ *the phase shift at frequency* j*.*

Remark 3.2. *Note that, in contrast to coherence for stochastic processes, the standardization* $\sqrt{\gamma_{\alpha;rr}(0) \cdot \gamma_{\alpha;ss}(0)}$ *is not frequency dependent. The reason is that, for a deterministic signal, no frequency dependent variances can be observed. Alternatively, one may consider* $|h_{rs}(j)|^2/(h_{rr}(j)h_{ss}(j))$*. However, this quantity is either* 0 *or* 1*. In contrast,* $\kappa_{rs}(j)$ *defined above can assume any number between* 0 *and* 1*, thus giving a relative measure of the contribution of frequency* j *to the cross covariance.*

Example 3.1. *Suppose that* f_s *is a shifted version of* f_r*, i.e.*

$$f_s(t) = c \cdot f_r(t + \Delta)$$

for some $\Delta, c \in \mathbb{R}$*. Then*

$$f_s(t) = \sum a_s(j) \exp(i2\pi jt) = c \sum a_r(j) \exp(i2\pi j(t + \Delta))$$

with

$$a_s(j) = ca_r(j) \exp(i2\pi j\Delta).$$

Hence

$$\gamma_{rs}(u) = c \sum |a_r(j)|^2 \exp(i2\pi j(u - \Delta)).$$

and

$$\tilde{h}_{rs}(j) = \frac{|a_r(j)|^2}{\sum_l |a_r(l)|^2} \cdot \exp(-i2\pi\Delta j) \ (j \in \mathbb{Z} \setminus \{0\})$$

Hence, for all integer frequencies $j \neq 0$

$$|h_{rs}(j)| = \frac{|a_r(j)|^2}{\sum_l |a_r(l)|^2}$$

and the phase-shift

$$\phi_{rs}(j) = -2\pi\Delta j$$

is a linear function of the shifting parameter Δ*. Thus,* $|h_{rs}(j)|$ *is equal to the relative contribution of frequency* j *to total energy* $||f_r||^2$ *of* f_r*.*

3.3 Estimation

3.3.1 General considerations

The definitions above suggest the following approach to analyzing an observed multivariate time series $\mathbf{Y}(i) = \mathbf{f}(t_i) + \boldsymbol{\epsilon}(i)$. In a first step, the function $\mathbf{f}$ is estimated by a suitable nonparametric method. In a second step, $\boldsymbol{\alpha}$ is estimated by eliminating the mean $\mathbf{m}$ and the nonperiodic component $\boldsymbol{\mu}(t)$. The regression spectrum of $\boldsymbol{\alpha}(t)$ can now be analyzed based on the resulting series

$$\tilde{\mathbf{Y}}(i) = \mathbf{Y}(i) - \hat{\mathbf{m}} - \hat{\boldsymbol{\mu}}(t).$$

The issue of estimating $\boldsymbol{\mu}(t)$ optimally in the given context is beyond the scope of this paper and will be considered elsewhere. Note, however, that in some applications, the size of $\boldsymbol{\mu}$ is negligible compared to $\boldsymbol{\alpha}$. For instance, for high frequency financial data, the dominating feature in the deterministic part is likely to be a (local) seasonal periodicity of one day. Similar comments apply to physiological time series such as the heart beat data considered in chapter 6.

In this section we consider estimation of $\mathbf{f}$ and the resulting estimation of the regression cross covariance $\boldsymbol{\Gamma}_\alpha$ and the regression spectrum $\mathbf{H}$.

3.3.2 Trend estimation

Since $\mathbf{\Gamma}_\alpha$ and the regression spectrum $\mathbf{H}$ are functionals of $\mathbf{f}$ (which we assume to be equal to $\boldsymbol{\alpha}$ - see remark at the end of the previous section), we first consider nonparametric estimation of the trend function. As mentioned in the last chapter, methods based on wavelets are known to have attractive features, such as general applicability to L^2-functions and localization in time and frequency. Wavelet thresholding is now applied to estimate the unknown function $\mathbf{f}$.

Let $\phi(\cdot)$ and $\psi(\cdot)$ be a father and mother wavelet, i.e. $\phi(\cdot), \psi(\cdot) \in L^2(\mathbb{R})$ and the set of functions

$$\{\phi_{l,k}(x),\ \ \psi_{j,k}(x),\ l \in \mathbb{Z},\ j \geq l, k \in \mathbb{Z}\}$$

with

$$\begin{aligned}\phi_{l,k}(x) &= 2^{\frac{l}{2}}\phi(2^l x - k),\\ \psi_{j,k}(x) &= 2^{\frac{j}{2}}\psi(2^j x - k),\ \ k, j \in \mathbb{Z},\end{aligned}$$

form a basis in $L^2(\mathbb{R})$. Estimation of $\mathbf{f}$ in model (3.1) may be performed for some J_n by applying a hard thresholding rule of the form

$$\hat{\mathbf{f}}(t) = \sum_k \hat{\boldsymbol{\alpha}}_{l,k}\phi_{l,k}(t) + \sum_{j\geq l}^{J_n}\sum_k \hat{\boldsymbol{\beta}}^*_{j,k}\psi_{j,k}(t), \tag{3.3}$$

where $\hat{\boldsymbol{\beta}}^*_{j,k} = diag(\hat{\mathbf{w}}_{j,k}\hat{\boldsymbol{\beta}}^T_{j,k})$ with

$$\hat{\boldsymbol{\alpha}}_{l,k} = \frac{1}{n}\sum_{i=1}^n \phi_{l,k}(t_i)\mathbf{Y}(i), \tag{3.4}$$

$$\hat{\boldsymbol{\beta}}_{j,k} = \frac{1}{n}\sum_{i=1}^n \psi_{j,k}(t_i)\mathbf{Y}(i), \tag{3.5}$$

$$\hat{\mathbf{w}}_{j,k} = (\hat{w}^{(1)}_{j,k}, \ldots, \hat{w}^{(p)}_{j,k})^T, \tag{3.6}$$

and $\hat{w}^{(r)}_{j,k} = \mathbf{1}_{\{|\hat{\beta}^{(r)}_{j,k}| \geq \sqrt{\text{var }(\hat{\beta}^{(r)}_{j,k})}\lambda_j\}}$, $1 \leq r \leq p$. The following assumption (C) summarizes the main issues of the last chapter and is needed in the subsequent considerations.

(*C*) Let the assumptions (A1) to (A5) of chapter 2 hold and assume wavelet thresholding is performed according to lemma 2.8.

First, we derive asymptotic properties of the regression cross covariance.

3.3.3 Estimation of the regression cross covariance

Consistency

By assumption $\mathbf{m}(f) =< \mathbf{f}, 1 >= 0$, and $\boldsymbol{\mu}(t) = 0$ so that $\boldsymbol{\Gamma} = \boldsymbol{\Gamma}_\alpha$. Given $\hat{\mathbf{f}}$, the cross-covariance $\boldsymbol{\Gamma}_\alpha$ can be estimated by

$$\hat{\boldsymbol{\Gamma}}_\alpha(u) = \hat{\boldsymbol{\Gamma}}(u) = \int_0^1 \hat{\mathbf{f}}(t+u)\overline{\hat{\mathbf{f}}^T(t)}dt \tag{3.7}$$

and the regression spectrum by

$$\hat{\mathbf{H}}(j) = \int_{-1/2}^{1/2} \exp(-i2\pi ju)\hat{\boldsymbol{\Gamma}}_\alpha(u)du.$$

Consistency of $\hat{\boldsymbol{\Gamma}}(u)$ is given by

Theorem 3.1. *Under the assumption (C),* $\hat{\boldsymbol{\Gamma}}(u)$ *converges in probability to* $\boldsymbol{\Gamma}(u)$ *such that* $\hat{\boldsymbol{\Gamma}}(u) - \boldsymbol{\Gamma}(u) = O_p(n^{-1/2})$.

The proof is based on the following

Lemma 3.1. *Let* f_r *be estimated according to (C), and denote the estimation error by*

$$\widetilde{\epsilon}_n^{(r)}(t) = \hat{f}_r(t) - f_r(t).$$

Then,

$$\int_{[0,1]} \widetilde{\epsilon}_n^{(r)}(t)dt \to 0 \tag{3.8}$$

and

$$\int_{[0,1]} |\widetilde{\epsilon}_n^{(r)}(t)|dt \to 0 \tag{3.9}$$

almost surely, and both terms are of order $O_p(n^{-1/2})$.

Proof: As mentioned earlier, Brillinger shows that for each r, $1 \leq r \leq p$, the variance of the thresholding estimates, $var\,(\hat{f}_r(x))$, is of order n^{-1}. To indicate its dependence on Ω, we write $\widetilde{\epsilon}_n^{(r)}(\omega, x)$ instead of $\widetilde{\epsilon}_n^{(r)}(x)$. Define $\mathcal{X} := [0, 1]$, let μ be the Lebesgue measure on $(\mathcal{X}, \mathcal{F}_\mathcal{X})$ and denote by $\mathcal{F}_\Omega$ the σ-algebra generated by the open subsets of Ω. We consider the product space $(\Omega \times \mathcal{X})$ with σ-algebra $\mathcal{F} = \mathcal{F}_\Omega \otimes \mathcal{F}_\mathcal{X}$ and corresponding measure $P \otimes \mu$. Due to the uniform convergence of $var\,[\hat{f}_r(x)]$,

$$\int_\mathcal{X} \int_\Omega (\widetilde{\epsilon}_n^{(r)}(\omega, x) - E[\widetilde{\epsilon}_n^{(r)}(\omega, x)])^2 dP dx \to 0$$

and

$$n \int_\mathcal{X} \int_\Omega (\widetilde{\epsilon}_n^{(r)}(\omega, x) - E[\widetilde{\epsilon}_n^{(r)}(\omega, x)])^2 dP dx = O(1).$$

This implies

$$\int_\mathcal{X} \int_\Omega \widetilde{\epsilon}_n^{(r)}(\omega, x)^2 dP dx = O(n^{-1}). \tag{3.10}$$

Note that $\mathcal{X}$ is a finite interval so that $(P \otimes \mu)(\Omega \times \mathcal{X}) < \infty$. Square integrability in finite measure spaces implies that

$$\int_\mathcal{X} \int_\Omega |\widetilde{\epsilon}_n^{(r)}(\omega, x)| \; dP\, dx = O(n^{-1/2}).$$

Due to the existence of both integrals we can apply Fubini's theorem so that

$$\int_\Omega \int_\mathcal{X} |\widetilde{\epsilon}_n^{(r)}(\omega, x)| \; dx\, dP \to 0, \;\; n \to \infty.$$

Define the random variable $z_n = \int_\mathcal{X} |\widetilde{\epsilon}_n^{(r)}(x)| \; dx$. Then $z_n \geq 0$ for all n and $E[z_n] \to 0$. Hence,

$$\int_{[0,1]} \widetilde{\epsilon}_n^{(r)}(\omega, x) \; dx \to 0$$

almost surely. ■

Proof: (of theorem 3.1.) Without loss of generality it suffices to show that the theorem is true for real-valued functions f_r and non-negative values of u. By adding and subtracting terms, we split the regression covariance into four terms

$$\begin{aligned}(I) &= \int_0^1 \left(\hat{f}_r(x+u) - f_r(x+u)\right)\left(\hat{f}_s(x) - f_s(x)\right)dx \\ (II) &= \int_0^1 f_r(x+u)\left(\hat{f}_s(x) - f_s(x)\right)dx \\ (III) &= \int_0^1 \left(\hat{f}_r(x+u) - f_r(x+u)\right) f_s(x)dx \\ (IV) &= \int_0^1 f_r(x+u)f_s(x)dx.\end{aligned}$$

The functions f_r and f_s are bounded and of bounded variation. Then

$$|(III)| \leq A \int |\widetilde{\epsilon}_n^{(r)}(x+u)|dx,$$

which tends to 0 almost surely according to lemma 3.1. The same holds for term (II). Moreover, the Cauchy-Schwarz inequality implies

$$\begin{aligned}&\left|\int \left(\hat{f}_r(x+u) - f_r(x+u)\right)\left(\hat{f}_s(x) - f_s(x)\right)dx\right| \\ &\leq \left(\int (\hat{f}_r(x+u) - f_r(x+u))^2 dx\right)^{\frac{1}{2}} \left(\int (\hat{f}_s(x) - f_s(x))^2 dx\right)^{\frac{1}{2}}.\end{aligned} \tag{3.11}$$

The proof of lemma 3.1. shows that both factors converge to zero almost surely. Therefore,

$$\hat{\gamma}_{rs}(u) \xrightarrow{P} \int_0^1 f_r(x+u)f_s(x)dx = \gamma_{rs}(u).$$

■

Asymptotic normality

First, we derive the asymptotic covariance function of $\hat{\gamma}_{rs}(u)$.

Theorem 3.2. *For $u, v \in [-1, 1]$,*

$$\lim_{n\to\infty} n cov(\hat{\gamma}_{rs}(u), \hat{\gamma}_{rs}(v)) = \sigma_{rs}(u, v)$$

where σ_{rs} is finite and can be written as

$$\begin{aligned}
\sigma_{rs}(u,v) &= 2\pi h_{\epsilon;ss}(0) \left[\sum_k \int f_r(x+u)\phi_{l,k}(x)dx \int f_r(y+v)\phi_{l,k}(y)dy \right. \\
&\left. + \sum_{l,k} w_{l,k}^{(s)} \int f_r(x+u)\psi_{l,k}(x)dx \int f_r(y+v)\psi_{l,k}(y)dy \right] \\
&+ 2\pi h_{\epsilon;rs}(0) \left[\sum_k \int f_s(x)\phi_{l,k}(x+u)dx \int f_r(y+v)\phi_{l,k}(y)dy \right. \\
&\left. + \sum_{l,k} w_{l,k}^{(r)} w_{l,k}^{(s)} \int f_s(x)\psi_{l,k}(x+u)dx \int f_r(y+v)\psi_{l,k}(y)dy \right] \\
&+ 2\pi h_{\epsilon;sr}(0) \left[\sum_k \int f_r(x+u)\phi_{l,k}(x)dx \int f_s(y)\phi_{l,k}(y+v)dy \right. \\
&\left. + \sum_{l,k} w_{l,k}^{(r)} w_{l,k}^{(s)} \int f_r(x+u)\psi_{l,k}(x)dx \int f_s(y)\psi_{l,k}(y+v)dy \right] \\
&+ 2\pi h_{\epsilon;rr}(0) \left[\sum_k \int f_s(x)\phi_{l,k}(x+u)dx \int f_s(y)\phi_{l,k}(y+v)dy \right. \\
&\left. + \sum_{l,k} w_{l,k}^{(r)} \int f_s(x)\psi_{l,k}(x+u)dx \int f_s(y)\psi_{l,k}(y+v)dy \right] + O(n^{-1/2}).
\end{aligned}$$

Proof: Recall, that

$$\hat{\gamma}_{rs}(u) - \gamma_{rs}(u) = \int [f_r(x+u)\widetilde{\epsilon}_n^{(s)}(x) + \widetilde{\epsilon}_n^{(r)}(x+u)f_s(x) + \widetilde{\epsilon}_n^{(r)}(x+u)\widetilde{\epsilon}_n^{(s)}(x)]dx.$$

Hence,

$$\begin{aligned}
&\lim_{n\to\infty} n\text{cov}\ (\hat{\gamma}_{rs}(u_i), \hat{\gamma}_{rs}(u_j)) \\
&= \lim_{n\to\infty} nE[(\int f_r(x+u_i)\widetilde{\epsilon}_n^{(s)}(x) + f_s(x)\widetilde{\epsilon}_n^{(r)}(x+u_i) + \widetilde{\epsilon}_n^{(r)}(x+u_i)\widetilde{\epsilon}_n^{(s)}(x)dx) \\
&\qquad \cdot (\int f_r(y+u_j)\widetilde{\epsilon}_n^{(s)}(y) + f_s(y)\widetilde{\epsilon}_n^{(r)}(y+u_j) + \widetilde{\epsilon}_n^{(r)}(y+u_j)\widetilde{\epsilon}_n^{(s)}(y)dy)] \\
&=: (I) + (II) + (III),
\end{aligned}$$

where

$$\begin{aligned}
(I) \quad = \begin{cases} \lim_{n\to\infty} & n\Big\{E[\int\int f_r(x+u_i)f_r(y+u_j)\widetilde{\epsilon}_n^{(s)}(x)\widetilde{\epsilon}_n^{(s)}(y)dx\,dy] \\ & +E[\int\int f_s(x)f_r(y+u_j)\widetilde{\epsilon}_n^{(r)}(x+u_i)\widetilde{\epsilon}_n^{(s)}(y)dx\,dy] \\ & +E[\int\int f_r(x+u_i)f_s(y)\widetilde{\epsilon}_n^{(s)}(x)\widetilde{\epsilon}_n^{(r)}(y+u_j)dx\,dy] \\ & +E[\int\int f_s(x)f_s(y)\widetilde{\epsilon}_n^{(r)}(x+u_i)\widetilde{\epsilon}_n^{(r)}(y+u_j)dx\,dy]\Big\}, \end{cases} \\
(II) \quad = \begin{cases} \lim_{n\to\infty} & n\Big\{E[\int\int f_r(x+u_i)\widetilde{\epsilon}_n^{(s)}(x)\widetilde{\epsilon}_n^{(r)}(y+u_j)\widetilde{\epsilon}_n^{(s)}(y)dx\,dy] \\ & +E[\int\int f_s(x)\widetilde{\epsilon}_n^{(r)}(x+u_i)\widetilde{\epsilon}_n^{(r)}(y+u_j)\widetilde{\epsilon}_n^{(s)}(y)dx\,dy] \\ & +E[\int\int f_r(y+u_j)\widetilde{\epsilon}_n^{(s)}(y)\widetilde{\epsilon}_n^{(r)}(x+u_i)\widetilde{\epsilon}_n^{(s)}(x)dx\,dy] \\ & +E[\int\int f_s(y)\widetilde{\epsilon}_n^{(r)}(y+u_j)\widetilde{\epsilon}_n^{(r)}(x+u_i)\widetilde{\epsilon}_n^{(s)}(x)\,dxdy]\Big\}, \end{cases} \\
(III) = \quad \lim_{n\to\infty} \quad & n\left\{E[\int\int \widetilde{\epsilon}_n^{(r)}(x+u_i)\widetilde{\epsilon}_n^{(s)}(x)\widetilde{\epsilon}_n^{(r)}(y+u_j)\widetilde{\epsilon}_n^{(s)}(y)dx\,dy]\right\}.
\end{aligned}$$

Consider just one part of the sum in part (I). Due to Fubini's theorem

$$\begin{aligned}
&\lim_{n\to\infty} nE[\int\int f_r(x+u_i)f_r(y+u_j)\widetilde{\epsilon}_n^{(s)}(x)\widetilde{\epsilon}_n^{(s)}(y)]\,dx\,dy \\
&= \lim_{n\to\infty} n\int\int f_r(x+u_i)f_r(y+u_j)E[\widetilde{\epsilon}_n^{(s)}(x)\widetilde{\epsilon}_n^{(s)}(y)]\,dx\,dy \\
&= \int\int f_r(x+u_i)f_r(y+u_j)\lim_{n\to\infty} nE[\widetilde{\epsilon}_n^{(s)}(x)\widetilde{\epsilon}_n^{(s)}(y)]\,dx\,dy, \qquad (3.12)
\end{aligned}$$

where the second equation is due to Lebesgue's theorem. By applying (C),

$$\begin{aligned}
(3.12) &= \int\int f_r(x+u_i)f_r(y+u_j)\lim_{n\to\infty} n\text{cov}\left(\hat{f}_s(x), \hat{f}_s(y)\right)\,dx\,dy \\
&= \int\int f_r(x+u_i)f_r(y+u_j)\left[2\pi h_{\epsilon;ss}(0)\sum_k \phi_{l,k}(x)\phi_{l,k}(y)\right. \\
&\quad \left. + 2\pi h_{\epsilon;ss}(0)\sum_{l,k} w_{j,k}^{(s)}\psi_{j,k}(x)\psi_{j,k}(y) + O(n^{-1})\right]\,dx\,dy.
\end{aligned}$$

Rearrangig the sums and integrals results in

$$
\begin{aligned}
&2\pi h_{\epsilon;ss}(0)\sum_k \int f_r(x+u_i)\phi_{l,k}(x)dx \int f_r(y+u_j)\phi_{l,k}(y)dy \\
&+ 2\pi h_{\epsilon;ss}(0)\sum_{j,k} w_{j,k}^{(s)} \int f_r(x+u_i)\psi_{j,k}(x)dx \int f_r(y+u_j)\psi_{j,k}(y)dy \\
&+ O(n^{-1}).
\end{aligned}
$$

Analogous results for the other parts of (I) yield $\sigma_{rs}(u_i, u_j)$. It remains to show that the remaining parts converge to 0. By lemma 3.1. $\int \widetilde{\epsilon}_n^{(r)}(x)dx \to 0$ almost surely with rate $n^{-1/2}$. Then (3.10) together with the Cauchy-Schwarz inequality yield $\int \widetilde{\epsilon}_n^{(r)}(y+u_j)\widetilde{\epsilon}_n^{(s)}(y)dy = O_p(n^{-1})$. Hence,

$$
\begin{aligned}
&\lim_{n\to\infty} n\left| E[\int\int f_r(x+u_i)\widetilde{\epsilon}_n^{(s)}(y)\widetilde{\epsilon}_n^{(r)}(y+u_j)\widetilde{\epsilon}_n^{(s)}(x)dx\,dy]\right| \\
&\leq \lim_{n\to\infty} n\cdot AE[\underbrace{\int |\widetilde{\epsilon}_n^{(s)}(x)|dx}_{=O_p(n^{-1/2})}\underbrace{\left|\int \widetilde{\epsilon}_n^{(r)}(y+u_j)\widetilde{\epsilon}_n^{(s)}(y)dy\right|}_{=O_p(n^{-1})}] \\
&= 0.
\end{aligned}
$$

Analogous considerations imply that (III) converges to 0. ■

In order to obtain the asymptotic distribution of $\hat{\mathbf{\Gamma}}$ the following additional condition on $\boldsymbol{\epsilon}(i)$ will be used.

(A6) For $1 \leq r, s \leq p$, define $\epsilon^*(i) := \epsilon_r(i)$, $i = 1, \ldots, n$, and $\epsilon^*(i) = \epsilon_s(i-n)$, $i = n+1, \ldots, 2n$. Let $\{\mathcal{F}_i,\ i = 1, \ldots, n\}$ be a non-decreasing sequence of σ-fields of $\mathcal{F}$ sets and let the sequence $(\epsilon^*(i), \mathcal{F}_i, i = 1, \ldots, n)$ be a square-integrable martingale difference array with constant variance and $E(\epsilon^{*2}(i)|\mathcal{F}_{i-1}) = E(\epsilon^{*2}(0))$.

Theorem 3.3. *Let*

$$\xi_n = n^{\frac{1}{2}}[\hat{\gamma}_{rs}(u_1) - \gamma_{rs}(u_1), \hat{\gamma}_{rs}(u_2) - \gamma_{rs}(u_2), ..., \hat{\gamma}_{rs}(u_k) - \gamma_{rs}(u_k)]^T$$

Then, under (A6) and the assumptions of theorem 3.1.,

$$\xi_n \xrightarrow{d} N(0, \Sigma)$$

where

$$\Sigma = [\sigma_{rs}(u_i, u_j)]_{r,s=1,...,p}$$

and "$\xrightarrow{d}$" denotes convergence in distribution.

The proof is based on the following lemmas.

Lemma 3.2. *Assume model (3.1) and the assumptions of theorem 3.1. Then, for each $s \in \{1, \ldots, p\}$ and for almost all $x \in [0, 1]$,*

$$\begin{aligned}(\hat{f}_s - f_s)(x) = &\frac{1}{n}\sum_{u=1}^{n} \epsilon_s(u) \sum_k \phi_{l,k}(u/n)\phi_{l,k}(x) \\ &+ \frac{1}{n}\sum_{u=1}^{n} \epsilon_s(u) \sum_{j\geq l}^{J_n} \sum_k \hat{w}_{j,k}^{(s)} \psi_{j,k}(u/n)\psi_{j,k}(x) \\ &+ r^{(s)}(n),\end{aligned}$$

where $r^{(s)}(n) = O_p(n^{-1})$.

Proof: Consider the shrinkage estimator

$$\hat{f}_s(x) = \sum_k \hat{\alpha}_{l,k}^{(s)} \phi_{l,k}(x) + \sum_{j\geq l}^{J_n} \sum_k \hat{w}_{j,k}^{(s)} \hat{\beta}_{j,k}^{(s)} \psi_{j,k}(x). \tag{3.13}$$

As above, let J_0 be the largest number such that the corresponding resolution component of f_s contains at least one nonzero coefficient.

Hence,

$$\begin{aligned}(\hat{f}_s - f_s)(x) &= \underbrace{\sum_k \hat{\alpha}_{l,k}^{(s)}\phi_{l,k}(x) - \sum_k \alpha_{l,k}^{(s)}\phi_{l,k}(x)}_{=:S(x)} \\ &\quad + \underbrace{\sum_{j\geq l}^{J_n}\sum_k \hat{w}_{j,k}^{(s)}\hat{\beta}_{j,k}^{(s)}\psi_{j,k}(x) - \sum_{j\geq l}^{J_0}\sum_k \beta_{j,k}^{(s)}\psi_{j,k}(x)}_{=:D(x)} \\ &= S(x) + D(x).\end{aligned}$$

Inserting the empirical wavelet coefficients yields

$$\begin{aligned}S(x) &= \sum_k [\frac{1}{n}\sum_{u=1}^n Y_s(u)\phi_{l,k}(u/n) - \alpha_{l,k}^{(s)}]\phi_{l,k}(x) \\ &= \sum_k [\frac{1}{n}\sum_{u=1}^n (Y_s(u) - E[Y_s(u)] + E[Y_s(u)])\phi_{l,k}(u/n) - \alpha_{l,k}^{(s)}]\phi_{l,k}(x) \\ &= \sum_k [\frac{1}{n}\sum_{u=1}^n (\epsilon_s(u)\phi_{l,k}(u/n) + f_s(u/n)\phi_{l,k}(u/n)) - \alpha_{l,k}^{(s)}]\phi_{l,k}(x) \\ &= \frac{1}{n}\sum_{u=1}^n \epsilon_s(u)\sum_k \phi_{l,k}(u/n)\phi_{l,k}(x) + \sum_k [\frac{1}{n}\sum_{u=1}^n f_s(u/n)\phi_{l,k}(u/n) - \alpha_{l,k}^{(s)}]\phi_{l,k}(x).\end{aligned}$$

Now

$$\begin{aligned}&\left|\frac{1}{n}\sum_{u=1}^n f_s(u/n)\phi_{l,k}(u/n) - \int_0^1 f_s(x)\phi_{l,k}(x)dx\right| \\ &= \left|\frac{1}{n}\sum_{u=1}^n f_s(u/n)\phi_{l,k}(u/n) - \alpha_{l,k}^{(s)}\right| \leq \frac{V(f_s\phi_{l,k})}{n}, \qquad (3.14)\end{aligned}$$

where $V(\cdot)$ denotes total variation (see e.g. Polya and Szegö (1964)). Hence,

$$S(x) = \frac{1}{n}\sum_{u=1}^n \epsilon_s(u)\sum_k \phi_{l,k}(u/n)\phi_{l,k}(x) + O(n^{-1}). \qquad (3.15)$$

For n large enough such that $J_n > J_0$:

$$\begin{aligned}D(x) &= \sum_j^{J_n}\sum_k \hat{w}_{j,k}^{(s)}\hat{\beta}_{j,k}^{(s)}\psi_{j,k}(x) - \sum_j^{J_0}\sum_k \beta_{j,k}^{(s)}\psi_{j,k}(x) \\ &= \sum_j^{J_0}\sum_k (\hat{w}_{j,k}^{(s)}\hat{\beta}_{j,k}^{(s)} - \beta_{j,k}^{(s)})\psi_{j,k}(x) + \sum_{j=J_0+1}^{J_n}\sum_k \hat{w}_{j,k}^{(s)}\hat{\beta}_{j,k}^{(s)}\psi_{j,k}(x). \qquad (3.16)\end{aligned}$$

Now,

$$\sum_{j}^{J_0}\sum_{k}(\hat{w}_{j,k}^{(s)}\hat{\beta}_{j,k}^{(s)}-\beta_{j,k}^{(s)})\psi_{j,k}(x)$$

$$=\sum_{j,k}(\hat{w}_{j,k}^{(s)}\frac{1}{n}\sum_{u=1}^{n}\psi_{j,k}(u/n)Y_s(u)-\beta_{j,k}^{(s)})\psi_{j,k}(x)$$

$$=\sum_{j,k}\{\hat{w}_{j,k}^{(s)}[\frac{1}{n}\sum_{u=1}^{n}\psi_{j,k}(u/n)(Y_s(u)-E[Y_s(u)]+E[Y_s(u)])]-\beta_{j,k}^{(s)}\}\psi_{j,k}(x)$$

$$=\sum_{j,k}\{\hat{w}_{j,k}^{(s)}\frac{1}{n}\sum_{u=1}^{n}\psi_{j,k}(u/n)\epsilon_s(u)+\hat{w}_{j,k}^{(s)}\frac{1}{n}\sum_{u=1}^{n}\psi_{j,k}(u/n)f_s(u/n)-\beta_{j,k}^{(s)}\}\psi_{j,k}(x)$$

$$=\sum_{j,k}\hat{w}_{j,k}^{(s)}\frac{1}{n}\sum_{u=1}^{n}\psi_{j,k}(u/n)\epsilon_s(u)\psi_{j,k}(x) \tag{3.17}$$

$$+\sum_{j,k}[\hat{w}_{j,k}^{(s)}\frac{1}{n}\sum_{u=1}^{n}\psi_{j,k}(u/n)f_s(u/n)-\beta_{j,k}^{(s)}]\psi_{j,k}(x). \tag{3.18}$$

For the term in (3.18), we have

$$|\frac{1}{n}\sum_{u=1}^{n}\psi_{j,k}(u/n)f_s(u/n)-\beta_{j,k}^{(s)}|\leq\frac{V(\psi_{j,k}f_s)}{n},$$

where the total variation is $A\cdot 2^{\frac{j}{2}}$, with A a suitable constant. Consider now the binary random variable $\hat{w}_{j,k}^{(s)}$. We distinguish between the cases $\beta_{j,k}^{(s)}=0$ and $\beta_{j,k}^{(s)}\neq 0$. In the first case, $n^{-1}\sum_{u=1}^{n}\psi_{j,k}(u/n)f_s(u/n)=O(2^{\frac{j}{2}}n^{-1})$. Recall that $j\leq J_0$. This implies that (3.18) converges in probability to 0 and is of order $O_p(n^{-1})$. If $\beta_{j,k}^{(s)}\neq 0$, consider first the case $\hat{w}_{j,k}^{(s)}=1$. Then, the rate of convergence in equation (3.18) is n^{-1}. Moreover,

$$P(\hat{w}_{j,k}^{(s)}=1)=P(|\hat{\beta}_{j,k}^{(s)}|>\lambda_j\cdot\sqrt{\text{var }(\hat{\beta}_{j,k}^{(s)})})\rightarrow 1.$$

This implies that there exists a sequence of subspaces $\Omega_n\uparrow\Omega$ such that $P(\Omega_n)\rightarrow 1$ and

$$P(\hat{w}_{j,k}^{(s)}=1|\omega\in\Omega_n)=1.$$

Define $\Omega_n^c = \Omega - \Omega_n$. Then, for a suitable constant A,

$$
\begin{aligned}
& P\left(|\hat{w}_{j,k}^{(s)} \frac{1}{n} \sum_{u=1}^{n} \psi_{j,k}(u/n) f_s(u/n) - \beta_{j,k}^{(s)}| \leq \frac{V(\psi_{j,k} f_s)}{n}\right) \\
& = P\left(|\hat{w}_{j,k}^{(s)} \frac{1}{n} \sum_{u=1}^{n} \psi_{j,k}(u/n) f_s(u/n) - \beta_{j,k}^{(s)}| \leq \frac{V(\psi_{j,k} f_s)}{n} \;\middle|\; \Omega_n\right) P(\Omega_n) \\
& + P\left(|\hat{w}_{j,k}^{(s)} \frac{1}{n} \sum_{u=1}^{n} \psi_{j,k}(u/n) f_s(u/n) - \beta_{j,k}^{(s)}| \leq \frac{V(\psi_{j,k} f_s)}{n} \;\middle|\; \Omega_n^c\right) P(\Omega_n^c) \\
& \to 1.
\end{aligned}
$$

Term (3.18) contains a finite amount of coefficients $\beta_{j,k}^{(s)} \neq 0$ so that convergence is uniform ($= O_p(n^{-1})$). The second part of (3.16) is given by

$$\sum_{j=J_0+1}^{J_n} \sum_{k} \hat{w}_{j,k}^{(s)} \hat{\beta}_{j,k}^{(s)} \psi_{j,k}(x). \tag{3.19}$$

For all $j, k \in \mathbb{Z}$, $\hat{\beta}_{j,k}^{(s)}$ is a root-n consistent estimator for $\beta_{j,k}^{(s)}$ $(s = 1, \ldots, p)$ and for all $j > J_0$, $\beta_{j,k}^{(s)} = 0$. Consider $\hat{w}_{j,k}^{(s)}$ in the case $w_{j,k}^{(s)} = 0$.

$$
\begin{aligned}
P(\hat{w}_{j,k}^{(s)} = 1) &= P(|\hat{\beta}_{j,k}^{(s)}| > \lambda_j \cdot \sqrt{\text{var } (\hat{\beta}_{j,k}^{(s)})}) \\
&= P(\sqrt{n}|\hat{\beta}_{j,k}^{(s)}| > \lambda_j \sqrt{n} \cdot \sqrt{\text{var } (\hat{\beta}_{j,k}^{(s)})}) \\
&\leq P(\sqrt{n}|\hat{\beta}_{j,k}^{(s)}| > A\lambda_j) \\
&\leq P(\sqrt{n}|\hat{\beta}_{j,k}^{(s)}| > \eta) \to 0
\end{aligned}
$$

for a suitable constant A and $\eta > 0$. Convergence is uniform in j so that (3.19) can be neglected asymptotically. Rearranging (3.17) and combining this with (3.15) yields the desired result.

■

The univariate asymptotic distribution of the regression cross covariance is now derived in

Lemma 3.3. *Under the assumptions of theorem 3.3., we have, for each pair*

(r,s), $1 \le r,s \le p$, and each $u \in [-1,1]$

$$\sqrt{n}(\hat{\gamma}_{rs}(u) - \gamma_{rs}(u)) \xrightarrow{d} \mathcal{N}(0, \sigma_{rs}(u,u)),$$

where $\sigma_{rs}(u,u)$ is given in theorem 3.2.

Proof: It is sufficient to show the assertion for positive values of u.

$$\sqrt{n}(\hat{\gamma}_{rs}(u) - \gamma_{rs}(u)) = \sqrt{n}\int_0^1 [f_r(x+u)\widetilde{\epsilon}_n^{(s)}(x) + f_s(x)\widetilde{\epsilon}_n^{(r)}(x+u) + \widetilde{\epsilon}_n^{(r)}(x+u)\widetilde{\epsilon}_n^{(s)}(x)]dx.$$

Applying lemma 3.2., the first part leads to

$$\begin{aligned}
&\sqrt{n}\int f_r(x+u)\left[\frac{1}{n}\sum_{t=1}^n \epsilon_s(t)\sum_k \phi_{l,k}(t/n)\phi_{l,k}(x)\right.\\
&\left.+\frac{1}{n}\sum_{t=1}^n \epsilon_s(t)\sum_{j,k}\hat{w}_{j,k}^{(s)}\psi_{j,k}(t/n)\psi_{j,k}(x) + O_p(n^{-1})\right]dx\\
&= \frac{1}{\sqrt{n}}\sum_{t=1}^n \epsilon_s(t)\sum_k \phi_{l,k}(t/n)\int f_r(x+u)\phi_{l,k}(x)dx\\
&+\frac{1}{\sqrt{n}}\sum_{t=1}^n \epsilon_s(t)\sum_{j,k}\hat{w}_{j,k}^{(s)}\psi_{j,k}(t/n)\int f_r(x+u)\psi_{j,k}(x)dx + O_p(n^{-1/2})\\
&=: \sum_{t=1}^n \epsilon_s(t) w_{t,n}^{(s)}(u) + O_p(n^{-1/2}),
\end{aligned} \tag{3.20}$$

where $w_{t,n}^{(s)}(u)$ is a triangular array of partly deterministic weights given by

$$w_{t,n}^{(s)}(u) := \frac{1}{\sqrt{n}}\left(\sum_k \phi_{l,k}(t/n)\int f_r(x+u)\phi_{l,k}(x)dx + \sum_{j,k}\hat{w}_{j,k}^{(s)}\psi_{j,k}(t/n)\int f_r(x+u)\psi_{j,k}(x)dx\right).$$

Analogously,

$$\sqrt{n}\int f_s(x)\widetilde{\epsilon}_n^{(r)}(x+u) = \sum_{t=1}^n \epsilon_r(t) w_{t,n}^{(r)}(u) + O_p(n^{-1/2}),$$

where

$$w_{t,n}^{(r)}(u) \approx \frac{1}{\sqrt{n}} \sum_k \phi_{l,k}(t/n) \int f_s(x)\phi_{l,k}(x+u)dx + \frac{1}{\sqrt{n}} \sum_{j,k} w_{j,k}^{(r)} \psi_{j,k}(t/n) \int f_s(x)\psi_{j,k}(x+u)dx.$$

As a result of the proof of theorem 3.1.

$$n \int \widetilde{\epsilon}_n^{(r)}(x+u)\widetilde{\epsilon}_n^{(s)}(x)dx = O_p(1).$$

It follows that

$$\begin{aligned}\sqrt{n}(\hat{\gamma}_{rs}(u) - \gamma_{rs}(u)) &= \sum_{t=1}^{n}(w_{t,n}^{(r)}(u)\epsilon_r(t) + w_{t,n}^{(s)}(u)\epsilon_s(t)) + O_p(n^{-1/2}), \\ &= \sum \epsilon^*(t) w_{t,n}^*(u) + O_p(n^{-1/2}),\end{aligned}$$

where $\epsilon^*(t) = \epsilon_r(t),\ t = 1, \ldots, n$, and $\epsilon^*(t) = \epsilon_s(t-n),\ t = n+1, \ldots, 2n$, with the weights $w_{t,n}^*(u)$ defined accordingly. Denote by $\hat{B}_{t,n}(u)$ the random part of $w_{t,n}^{(s)}(u)$. Then,

$$w_{t,n}^{(s)}(u) := A_{t,n}(u) + \hat{B}_{t,n}(u) = A_{t,n}(u) + B_{t,n}(u) + (\hat{B}_{t,n}(u) - B_{t,n}(u)),$$

with

$$\begin{aligned}A_{t,n}(u) &= \frac{1}{\sqrt{n}} \sum_k \phi_{l,k}(t/n) \int f_r(x+u)\phi_{l,k}(x)dx, \\ \hat{B}_{t,n}(u) &= \frac{1}{\sqrt{n}} \sum_{j,k} \hat{w}_{j,k}^{(s)} \psi_{j,k}(t/n) \int f_r(x+u)\psi_{j,k}(x)dx, \\ B_{t,n}(u) &= \frac{1}{\sqrt{n}} \sum_{j,k} w_{j,k}^{(s)} \psi_{j,k}(t/n) \int f_r(x+u)\psi_{j,k}(x)dx,\end{aligned}$$

$$\sum_{t=1}^{n} \epsilon_s(t) w_{t,n}^{(s)}(u) = \sum_{t=1}^{n} \epsilon_s(t)(A_{t,n}(u) + B_{t,n}(u)) + \sum_{u=1}^{n} \epsilon_s(u)(\hat{B}_{t,n}(u) - B_{t,n}(u)) \tag{3.21}$$

and

$$\sum_{t=1}^{n} \epsilon_s(t)(\hat{B}_{t,n}(u) - B_{t,n}(u)) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} [\sum_{j,k} (\hat{w}_{j,k}^{(s)} - w_{j,k}^{(s)}) \psi_{j,k} \left(\frac{t}{n}\right) b_{j,k}^{(r)}] \epsilon_s(t), \tag{3.22}$$

where $b_{j,k}^{(r)} := \int f_r(x+u)\psi_{j,k}(x)dx$. Since

$$\hat{w}_{j,k}^{(s)} - w_{j,k}^{(s)} = \mathbf{1}_{\{|\hat{\beta}_{j,k}^{(s)}| \geq \sqrt{var\,(\hat{\beta}_{j,k}^{(s)})}\lambda_j\}} - \mathbf{1}_{\{|\beta_{j,k}^{(s)}|>0\}},$$

(3.22) only contains nonzero elements for those indices (j,k) where either $|\hat{\beta}_{j,k}^{(s)}| \geq \sqrt{var\,(\hat{\beta}_{j,k}^{(s)})}\lambda_j$ and $\beta_{j,k}^{(s)} = 0$ or $|\hat{\beta}_{j,k}^{(s)}| < \sqrt{var\,(\hat{\beta}_{j,k}^{(s)})}\lambda_j$ and $\beta_{j,k}^{(s)} \neq 0$. Consider first (j,k) with $\beta_{j,k}^{(s)} \neq 0$. Since $P(\hat{w}_{j,k}^{(s)} = 1) \to 1$, $\hat{w}_{j,k}^{(s)} - w_{j,k}^{(s)} \to 0$ in probability. Convergence is uniform in the set of (j,k) with $\beta_{j,k}^{(s)} \neq 0$, since this set is finite. For $\beta_{j,k}^{(r)} = 0$,

$$(3.22) = \sum_{t=1}^{n} \hat{B}_{t,n}(u)\epsilon_s(t)$$

and $P(\hat{w}_{j,k}^{(r)} = 1) \to 0$ uniformly in j,k. Consider $\hat{B}_{t,n}(u)$ for the case $u = k2^{-j}$ with (j,k) arbitrary. In this case there is at most a finite number of $b_{j,k}^{(r)} \neq 0$. For each $\eta > 0$ and a suitable constant A,

$$\begin{aligned}
&P(|\sqrt{n}\hat{B}_{t,n}(u)| > \eta) \\
&= P\left(\left|\sum_{j}^{J_n}\sum_{k} \hat{w}_{j,k}^{(s)}\psi_{j,k}\left(\frac{t}{n}\right) b_{j,k}^{(r)}\right| > \eta\right) \\
&\leq P\left(A\sum_{j}^{J_0}\sum_{k} |\hat{w}_{j,k}^{(s)}| > \eta\right) \to 0. \qquad (3.23)
\end{aligned}$$

Consider now $u \neq k2^{-j}$. For sufficiently large n there exists an integer k such that $|u - k\cdot 2^{-J_n}| < 2^{-J_n}$. For suitable constants A_1, A_2

$$\begin{aligned}
&\left|\int f_r(x+u)\psi_{j,k}(x)dx - \int f_r(x+k2^{-J_n})\psi_{j,k}(x)dx\right| \\
&\leq A_1\cdot 2^{j/2}\int |f_r(x+u) - f_r(x+k2^{-J_n})|dx \\
&\leq A_1\cdot 2^{j/2}\sup_x |f_r(x+u) - f_r(x+k2^{-J_n})|\cdot|u - k2^{-J_n}| \\
&\leq A_2\cdot 2^{-J_n} \to 0
\end{aligned}$$

due to assumption (A2). Therefore,

$$
\begin{aligned}
&P\left(|\sum_{j,k}\hat{w}_{j,k}^{(s)}\psi_{j,k}\left(\frac{t}{n}\right)b_{j,k}^{(r)}|>\eta\right)\\
&\leq P\left(\sum_{j,k}\left|\hat{w}_{j,k}^{(s)}\psi_{j,k}\left(\frac{t}{n}\right)\right|[\int|f_r(x+k2^{-j})\psi_{j,k}(x)dx|+A_2 2^{-J_n}]>\eta\right)\\
&\leq P\left(\sum_{j,k}\left|\hat{w}_{j,k}^{(s)}\psi_{j,k}\left(\frac{t}{n}\right)\right|\int|f_r(x+k2^{-j})\psi_{j,k}(x)dx|>\frac{\eta}{2}\right) \qquad (3.24)\\
&+P\left(A_2 2^{-J_n}\sum_{j,k}\left|\hat{w}_{j,k}^{(s)}\psi_{j,k}\left(\frac{t}{n}\right)\right|>\frac{\eta}{2}\right) \qquad (3.25)
\end{aligned}
$$

(3.24) converges to 0 in probability according to (3.23). Moreover,

$$
P\left(\left|A_2 2^{-J_n}\sum_{j,k}^{J_n}|\hat{w}_{j,k}^{(s)}\psi_{j,k}\left(\frac{t}{n}\right)\right|>\frac{\eta}{2}\right)\leq P\left(\left|A_2 2^{-J_n}\sum_{j}^{J_n}2^{j/2}\right|>\frac{\eta}{2}\right)\to 0,
$$

so that $P(\sqrt{n}\hat{B}_{t,n}(u)|>\eta)\to 0$. Since $\epsilon_s(t)$ form a square-integrable martingale difference array with constant variance, $n^{-\frac{1}{2}}\sum_{t=1}^{n}\epsilon_s(t)$ converges in distribution to a normal variable and $\sum_{t=1}^{n}\hat{B}_{t,n}(u)\epsilon_s(t)\to 0$ in probability. Combining all cases, $\sum_t\epsilon_s(t)(\hat{B}_{t,n}(u)-B_{t,n}(u))\to 0$ in probability so that asymptotically the second part of the sum in (3.21) can be neglected. Hence, $w_{t,n}^{(s)}(u)$ reduce to the deterministic weights

$$
\begin{aligned}
\tilde{w}_{t,n}^{(s)}(u)&=A_{t,n}(u)+B_{t,n}(u)\\
&=\frac{1}{\sqrt{n}}\sum_{k}\phi_{l,k}(t/n)\int f_r(x+u)\phi_{l,k}(x)dx\\
&+\frac{1}{\sqrt{n}}\sum_{j,k}w_{j,k}^{(s)}\psi_{j,k}(t/n)\int f_r(x+u)\psi_{j,k}(x)dx.
\end{aligned}
$$

Similarly, $w_{t,n}^{(r)}(u)$ can be replaced by $\tilde{w}_{t,n}^{(r)}(u)$ and hence also $w_{t,n}^{*}(u)$ by $\tilde{w}_{t,n}^{*}(u)$. Defining $\sigma_n^2:=\text{var}\,[\sum_{t=1}^{n}\tilde{w}_{t,n}^{*}(u)\epsilon^{*}(t)]$, we have $\sigma_n^2\to\sigma_{rs}(u,u)$. Since $w_{j,k}^{(s)}=0$,

$j > J_0$,

$$\begin{aligned}\max_{1\le t\le n}|w^{(s)}_{t,n}(u)| &= \max_{1\le t\le n}\left|\frac{1}{\sqrt{n}}\sum_k \phi_{l,k}(t/n)\int f_r(x+u)\phi_{l,k}(x)dx\right.\\ &\left.+\frac{1}{\sqrt{n}}\sum_{j,k} w^{(s)}_{j,k}\psi_{j,k}(t/n)\int f_r(x+u)\psi_{j,k}(x)dx\right|\\ &\le \frac{A}{\sqrt{n}}\sum_k\int |f_r(x+u)\phi_{l,k}(x)|dx\\ &+\frac{A}{\sqrt{n}}\sum_{j,k}^{J_0}\int |f_r(x+u)\psi_{j,k}(x)|dx,\\ &= O(n^{-1/2}),\end{aligned}$$

where A is a suitable constant, we have

$$\max_{1\le t\le n}|w^*_{t,n}(u)| = \max_{1\le t\le n}|w^{(r)}_{t,n}(u)+w^{(s)}_{t,n}(u)|\to 0.$$

From theorem 3.2. and the approximation of $w^*_{t,n}(u)$ by $\tilde{w}^*_{t,n}(u)$ we obtain $var\left[\sum_{t=1}^n \tilde{w}^*_{t,n}(u)\epsilon^*(t)\right] = O(1)$. Hence,

$$\max_{1\le t\le n}\frac{|w^*_{t,n}(u)|}{\sigma_n}\to 0,\ \ n\to\infty.$$

Then theorem 4 in Beran and Feng (2001) together with (A6) imply

$$\frac{\sum_t w^*_{t,n}(u)\epsilon^*(t)}{\sigma_n}\xrightarrow{d}\mathcal{N}(0,1),$$

and hence

$$\sqrt{n}(\hat{\gamma}_{rs}(u)-\gamma_{rs}(u))\xrightarrow{d}\mathcal{N}(0,\sigma_{rs}(u,u)).$$

■

By using lemma 3.2. we now prove the assertion of theorem 3.3.

Proof: (of theorem 3.3.) Applying the Cramér-Wold device, we show that for any constants θ_l, $l = 1,\dots,q,\ q\in\mathbb{N}$, not all equal to zero,

$$n^{\frac{1}{2}}\sum_{l=1}^q \theta_l(\hat{\gamma}_{rs}(u_l)-\gamma_{rs}(u_l))\xrightarrow{d}\mathcal{N}(0,\widetilde{\sigma}^2),$$

where $\widetilde{\sigma}^2 := \sum_{l=1}^{q}\sum_{m=1}^{q}\theta_l\theta_m\sigma_{rs}(u_l,u_m) > 0$. Note that

$$\begin{aligned} & n^{\frac{1}{2}}\sum_{l=1}^{q}\theta_l(\hat{\gamma}_{rs}(u_l)-\gamma_{rs}(u_l)) \\ & = n^{\frac{1}{2}}\underbrace{\int[\widetilde{\epsilon}_n^{(s)}(x)\sum_{l=1}^{q}\theta_l f_r(x+u_l)}_{=:A} + \underbrace{f_s(x)\sum_{l=1}^{q}\theta_l\widetilde{\epsilon}_n^{(r)}(x+u_l)}_{=:B} \\ & \underbrace{+\,\widetilde{\epsilon}_n^{(s)}(x)\sum_{l=1}^{q}\theta_l\widetilde{\epsilon}_n^{(r)}(x+u_l)]dx}_{=:C}\,. \end{aligned} \tag{3.26}$$

As shown earlier, the first part (A) can be written approximately as

$$\begin{aligned} \sum_{t=1}^{n}\epsilon_s(t)\Bigg[& n^{-\frac{1}{2}}\sum_{k}\phi_{l,k}(t/n)\sum_{l=1}^{q}\theta_l\int f_r(x+u_l)\phi_{l,k}(x)dx \\ & + n^{-\frac{1}{2}}\sum_{j,k}\hat{w}_{j,k}^{(s)}\psi_{j,k}(t/n)\sum_{l=1}^{q}\theta_l\int f_r(x+u_l)\psi_{j,k}(x)dx\Bigg], \end{aligned}$$

where the term in brackets is a weight function for the individual $\epsilon_s(t)$. We denote the weights by $w_{t,n}^{(s^*)}(u)$ and the analogous weights of term (B) by $w_{t,n}^{(r^*)}(u)$. As above, (C) converges to 0 almost surely.
Note that $\widetilde{\sigma} := \sqrt{var\left[\sum w_{t,n}^{(r^*)}(u)\epsilon_r(t) + \sum w_{t,n}^{(s^*)}(u)\epsilon_s(t)\right]}$. Then a simple consequence of previous arguments is that

$$\max_{1\le t\le n}\frac{|w_{t,n}^{(r^*)}(u)+w_{t,n}^{(s^*)}(u)|}{\widetilde{\sigma}} \to 0.$$

Hence, the Central Limit Theorem for the weighted sum holds such that

$$\sum w_{t,n}^{(r^*)}(u)\epsilon_r(t) + \sum w_{t,n}^{(s^*)}(u)\epsilon_s(t) \xrightarrow{d} N(0,\widetilde{\sigma}^2).$$

By applying the Cramér-Wold Device, this is equivalent to finite collections of $\hat{\gamma}_{rs}(\mathbf{u})$ being asymptotically jointly normal with mean zero and covariance matrix $(\sigma_{rs}(u_l,u_m))$, $l,m=1,\ldots,q$, i.e.

$$n^{\frac{1}{2}}(\hat{\gamma}_{rs}(\mathbf{u})-\gamma_{rs}(\mathbf{u}))\to\mathcal{N}(0,(\sigma_{rs}(u_l,u_m))_{1\le l,m\le q}). \tag{3.27}$$

■

Theorem 3.3. can be extended to a functional limit theorem.

Theorem 3.4. *Let P_n be the probability distribution of $\sqrt{n}(\hat{\gamma}_{rs}^{(n)}(u) - \gamma_{rs}(u))$ in $C[-1,1]$, where $C[-1,1]$ is equipped with the uniform topology defined by the metric $d(f,g) = \sup_{-1\leq t\leq 1} |f(t)-g(t)|$. Then P_n converges weakly in $C[-1,1]$ (in the metric d) to the probability distribution P of a Gaussian process where the finite-dimensional distributions are given in theorem 3.3.*

Proof: Let P_n be the probability distribution of $\hat{\gamma}_{rs}^{(n)}(u)$ in $C[-1,1]$ with corresponding σ-algebra $\mathcal{C}$ and let $C[-1,1]$ be given the uniform topology induced by the metric $d(f,g) = \sup_{-1\leq t\leq 1} |f(t) - g(t)|$. Then for any Borel subset B: $P_n(B) = \mathrm{P}(\hat{\gamma}_{rs}^{(n)}(u) \in B)$. It remains to prove tightness of the family P_n of probability distributions of $\hat{\gamma}_{rs}^{(n)}(u)$ (Billingsley (1968)). Define the modulus of continuity of an element $f \in C[-1,1]$ by

$$w_f(\delta) = w(f,\delta) = \sup_{|s-t|<\delta} |f(s) - f(t)|, \quad 0 < \delta \leq 2.$$

Then the sequence $\{P_n\}$ is tight if and only if

(i) for each positive η there exists a constant A such that

$$P_n(|\hat{\gamma}_{rs}^{(n)}(0)| > A) \leq \eta, \;\; n \geq 1, \tag{3.28}$$

(ii) For each positive η_1 and η_2, there exists δ $(0 < \delta < 2)$ and $n_0 \in I\!N$ such that

$$P_n(w_{\hat{\gamma}_{rs}^{(n)}}(\delta) > \eta_1) \leq \eta_2, \;\; n \geq n_0. \tag{3.29}$$

We first note that the probability distributions of $\hat{\gamma}_{rs}^{(n)}(0)$ are tight.

$$\hat{\gamma}_{rs}^{(n)}(0) = \int f_r(x) f_s(x) dx + \int f_r(x) \widetilde{\epsilon}_n^{(s)}(x) dx + \int f_s(x) \widetilde{\epsilon}_n^{(r)}(x) dx + \int \widetilde{\epsilon}_n^{(r)}(x) \widetilde{\epsilon}_n^{(s)}(x) dx.$$

Lemma 3.1. shows that all parts but $\int f_r(x) f_s(x) dx$ converge to 0 almost surely. Tightness of $\hat{\gamma}_{rs}^{(n)}(0)$ is not influenced by $\int f_r(x) f_s(x) dx$ and each of the other parts of the sum converges to 0 in probability due to lemma 3.1. This implies that for each $A > 0$ and each $\eta > 0$ there exists an $n_0 \in I\!N$ with e.g. $P_n(|\int f_s(x) \widetilde{\epsilon}_n^{(r)}(x) dx| > A) < \eta$ for all $n \geq n_0$. For the first indices $1, \ldots, n_0 - 1$ the statement is trivial since finitely many random variables are always tight. For the second condition note that $w_{\hat{\gamma}_{rs}^{(n)}}(\delta) = \sup_{|u_2 - u_1| < \delta} |\hat{\gamma}_n(u_2) - \hat{\gamma}_n(u_1)|$. The continuity of the integral together with the results of lemma 3.1. imply that there exists an n_0 such that condition (ii) is satisfied for all $\eta_1, \eta_2 > 0$. Therefore, P_n is tight such that $\hat{\gamma}_{rs}^{(n)}(u)$ converges to a Gaussian process where the finite dimensional distributions are given in equation (3.27). ■

3.3.4 Estimation of the regression cross spectrum

Asymptotic normality

Theorem 3.4. together with the continous mapping theorem (see e.g. Pollard (1984)) lead to

Theorem 3.5. *Under the assumptions of theorem 3.4., the vector*

$$\zeta_n = \sqrt{n}[\hat{h}_{rs}(j_1) - h_{rs}(j_1), ..., \hat{h}_{rs}(j_m) - h_{rs}(j_m)]^T$$

converges in distribution to an $m-$dimensional zero mean normal vector with covariance matrix

$$\lim_{n\to\infty} ncov\left(\hat{h}_{rs}(j_k), \hat{h}_{rs}(j_l)\right) = \int\int \exp\left\{-i2\pi(j_k u_1 - j_l u_2)\right\} \sigma_{rs}(u_1, u_2) du_1 du_2$$

where $\sigma_{rs}(u_1, u_2)$ is defined in theorem 3.2.

Proof: Define $X_n(u) := \sqrt{n}(\hat{\gamma}_{rs}(u) - \gamma_{rs}(u))$. The preceeding discussion has shown that X_n is a random element in $C[-1, 1]$ that converges in distribution to a stochastic process whose finite dimensional distributions are asymptotically normal. Consider

$$\sqrt{n}(\hat{h}_{rs}(j) - h_{rs}(j)) = \int_{-1/2}^{1/2} e^{-i2\pi ju} X_n(u) du,$$

and define

$$\begin{aligned} H :& C[-1, 1] \longrightarrow I\!R \\ & X_n \longmapsto \int_{-1/2}^{1/2} e^{-i2\pi ju} X_n(u) du. \end{aligned}$$

Then the Central Limit Theorem for random variables in $C[-1, 1]$ carries over to integral functionals (see e.g. Pollard (1984)), i.e.

$$\int_{-1/2}^{1/2} e^{-i2\pi ju} X_n(u) du \xrightarrow{d} \int_{-1/2}^{1/2} e^{-i2\pi ju} X(u) du.$$

The limiting random variable is a linear functional of the continous sample paths of the process X. Therefore, it is itself normally distributed with expectation 0 and asymptotic covariance matrix

$$\begin{aligned} & \lim_{n\to\infty} n\text{cov}\ (\hat{h}_{rs}(j_1), \hat{h}_{rs}(j_2)) \\ &= \lim_{n\to\infty} nE[((\hat{h}_{rs}(j_1) - h_{rs}(j_1))(\overline{\hat{h}_{rs}(j_2) - h_{rs}(j_2)}))] \\ &= \lim_{n\to\infty} \int\int e^{-i2\pi(u_1 j_1 - u_2 j_2)} E(X_n(u_1) X_n(u_2)) du_1 du_2 \\ &= \int\int e^{-i2\pi(u_1 j_1 - u_2 j_2)} \lim_{n\to\infty} n\text{cov}\ (\hat{\gamma}_{rs}(u_1), \hat{\gamma}_{rs}(u_2)) du_1 du_2. \end{aligned}$$

■

The exact calculation of the covariance of the estimated regression spectrum requires extensive notations. By using the Fourier series expansions of $\mathbf{f}$ and the wavelet basis functions, a more concise representation is given in

Remark 3.3. *Consider the Fourier series expansion of* $f_r, f_s, \phi_{l,k}, \psi_{j,k} \in L^2([0,1])$ *given by*

$$f_r(x) = \sum_{k_1} a_r(k_1)\exp(i2\pi k_1 x), \tag{3.30}$$

$$f_s(x) = \sum_{k_2} a_s(k_2)\exp(i2\pi k_2 x), \tag{3.31}$$

$$\phi_{l,k}(x) = \sum_{k_3} a^{\phi}_{l,k}(k_3)\exp(i2\pi k_3 x), \tag{3.32}$$

$$\psi_{l,k}(x) = \sum_{k_4} a^{\psi}_{j,k}(k_4)\exp(i2\pi k_4 x). \tag{3.33}$$

Using equations (3.30) to (3.33), we get that

$$\begin{aligned}
&\lim_{n\to\infty} n cov(\hat{h}_{rs}(j_1), \hat{h}_{rs}(j_2)) \\
= \;& 2\pi h_{\epsilon;ss}(0) a_r(j_1) a_r(-j_2) \left[\sum_k a^{\phi}_{l,k}(-j_1) a^{\phi}_{l,k}(j_2) + \sum_{j,k} w^{(s)}_{j,k} a^{\psi}_{j,k}(-j_1) a^{\psi}_{j,k}(j_2)\right] \\
& + 2\pi h_{\epsilon;rs}(0) a_s(-j_1) a_r(-j_2) \left[\sum_k a^{\phi}_{l,k}(j_1) a^{\phi}_{l,k}(j_2) + \sum_{j,k} w^{(r)}_{j,k} w^{(s)}_{j,k} a^{\psi}_{j,k}(j_1) a^{\psi}_{j,k}(j_2)\right] \\
& + 2\pi h_{\epsilon;sr}(0) a_r(j_1) a_s(j_2) \left[\sum_k a^{\phi}_{l,k}(-j_1) a^{\phi}_{l,k}(-j_2) + \sum_{j,k} w^{(s)}_{j,k} w^{(r)}_{j,k} a^{\psi}_{j,k}(-j_1) a^{\psi}_{j,k}(-j_2)\right] \\
& + 2\pi h_{\epsilon;rr}(0) a_s(-j_1) a_s(-j_2) \left[\sum_k a^{\phi}_{l,k}(j_1) a^{\phi}_{l,k}(j_2) + \sum_{j,k} w^{(r)}_{j,k} a^{\psi}_{j,k}(j_1) a^{\psi}_{j,k}(j_2)\right].
\end{aligned}$$

Estimation of amplitude and phase spectrum

Theorem 3.5. shows that finite vectors of the regression cross spectrum $\hat{h}_{rs}(\mathbf{j})$ converge to a complex-valued normal random variable. Let $c_{rs}(j) =$

$\mathfrak{Re}(h_{rs}(j))$ and $q_{rs}(j) = -\mathfrak{Im}(h_{rs}(j))$ so that $h_{rs}(j) = c_{rs}(j) - iq_{rs}(j)$. Then we have estimates

$$\hat{c}_{rs}(j) = \frac{1}{2}(\hat{h}_{rs}(j) + \hat{h}_{sr}(j)) \quad \text{and} \tag{3.34}$$

$$\hat{q}_{rs}(j) = -\frac{1}{2i}(\hat{h}_{rs}(j) - \hat{h}_{sr}(j)). \tag{3.35}$$

Due to theorem 3.5., the vector

$$\sqrt{n}[\hat{c}_{rs}(j) - c_{rs}(j), \hat{q}_{rs}(j) - q_{rs}(j)]^T$$

converges in distribution to a bivariate normal variable with mean 0 and asymptotic covariance matrix

$$\Sigma(j) = \begin{pmatrix} \Sigma_{cc}(j) & \Sigma_{cq}(j) \\ \Sigma_{qc}(j) & \Sigma_{qq}(j) \end{pmatrix}.$$

Remark 3.4. *Due to (3.34) and (3.35) the covariance matrix $\Sigma(j)$ follows by using remark 3.3. The hereby required covariances e.g. between $\hat{h}_{rs}(j)$ and $\overline{\hat{h}_{rs}(j)}$ are gained by changing corresponding signs of frequencies in remark 3.3.*

The asymptotic distributions of the amplitude and phase spectrum follow directly. Let $\kappa^*_{rs}(j)$ be the non-normalized spectral modulus. We have

Corollary 3.1. *Let $\kappa^*_{rs}(j) = |h_{rs}(j)| > 0$ and*

$$\hat{\kappa}^*_{rs}(j) = (\hat{c}^2_{rs}(j) + \hat{q}^2_{rs}(j))^{\frac{1}{2}}.$$

Then

$$\sqrt{n}(\hat{\kappa}^*_{rs}(j) - \kappa^*_{rs}(j)) \xrightarrow{d} \mathcal{N}(0, \sigma^2_{\kappa;rs}(j)),$$

where

$$\sigma^2_{\kappa;rs}(j) = \frac{1}{c^2_{rs}(j) + q^2_{rs}(j)} \left(c^2_{rs}(j)\Sigma_{cc}(j) + q^2_{rs}(j)\Sigma_{qq}(j) + 2c_{rs}(j)q_{rs}(j)\Sigma_{cq}(j)\right). \tag{3.36}$$

Proof: Note that $\kappa^*_{rs}(j)$ is continously differentiable with respect to $c_{rs}(j)$ and $q_{rs}(j)$ such that $\sqrt{n}(\hat{\kappa}^*_{rs}(j) - \kappa^*_{rs}(j))$ is itself normally distributed with mean 0. The asymptotic variance is now derived by

$$\begin{aligned}\operatorname{var}(\hat{\kappa}^*_{rs}(j)) &= \left(\left.\frac{\partial g}{\partial x}\right|_{(c_{rs}(j),q_{rs}(j))}\right)^2 \cdot \Sigma_{cc}(j) + \left(\left.\frac{\partial g}{\partial y}\right|_{(c_{rs}(j),q_{rs}(j))}\right)^2 \cdot \Sigma_{qq}(j) \\ &+ 2\left(\left.\frac{\partial g}{\partial x}\right|_{(c_{rs}(j),q_{rs}(j))}\right)\left(\left.\frac{\partial g}{\partial y}\right|_{(c_{rs}(j),q_{rs}(j))}\right) \cdot \Sigma_{cq}(j)\end{aligned}$$

where

$$\begin{aligned} g(x,y) &= (x^2+y^2)^{1/2}, \\ \left.\frac{\partial g}{\partial x}\right|_{(c_{rs}(j),q_{rs}(j))} &= \frac{c_{rs}(j)}{\sqrt{c^2_{rs}(j)+q^2_{rs}(j)}}, \\ \left.\frac{\partial g}{\partial y}\right|_{(c_{rs}(j),q_{rs}(j))} &= \frac{q_{rs}(j)}{\sqrt{c^2_{rs}(j)+q^2_{rs}(j)}}, \end{aligned}$$

such that (3.36) is satisfied. ■

Analogously we obtain

Corollary 3.2. *Assume* $|h_{rs}(j)| > 0$. *Let* $\phi_{rs}(j) = \arg h_{rs}(j)$ *and*

$$\hat{\phi}_{rs}(j) = \arg(\hat{c}_{rs}(j) - i\hat{q}_{rs}(j)) \in (-\pi, \pi].$$

Then

$$\sqrt{n}(\hat{\phi}_{rs}(j) - \phi_{rs}(j)) \xrightarrow{d} \mathcal{N}(0, \sigma^2_{\phi;rs}(j)),$$

where

$$\sigma^2_{\phi;rs}(j) = \frac{1}{(c^2_{rs}(j)+q^2_{rs}(j))^2}\left(q^2_{rs}(j)\Sigma_{cc}(j) + c^2_{rs}(j)\Sigma_{qq}(j) - 2c_{rs}(j)q_{rs}(j)\Sigma_{cq}(j)\right). \tag{3.37}$$

3.4 Estimation involving Fourier Coefficients

The regression spectrum at frequencies j is given by

$$h_{rs}(j) = a_r(j)\overline{a_s(j)}. \tag{3.38}$$

Previously, asymptotic distribution of the estimator of the regression spectrum was based on a preliminary estimation of the regression cross covariance. However, in view of (3.38), estimation and subsequent derivation of the asymptotic distribution might be more easily achieved by considering estimation of the Fourier coefficients directly.

Due to integration, the sample paths of $\hat{\gamma}_{rs}(u)$ are continous on $[-1, 1]$. In contrast, smoothness of $\hat{f}_r(t)$, $t \in [0, 1]$, heavily depends on the regularity of the wavelet basis. For example finite wavelet approximations involving the Haar basis results in discontinous step functions. According to condition (C), we consider functions $f_r(t)$, $1 \leq r \leq p$, and wavelet bases that are elements of $D[0, 1]$, the set of all right-continous functions on $[0, 1]$ with existing left-hand limits. The convergence of $f_r(\cdot)$ as a process on $[0, 1]$ is discussed in

Lemma 3.4. *Assume that condition (C) holds and let $\hat{f}_r(t) \in D[0, 1]$. Then,*

$$\vartheta_n(t) := \sqrt{n}(\hat{f}_r(t) - f_r(t))$$

converges weakly (in the Skorohod metric) to a Gaussian process with mean 0 where the finite-dimensional distributions are given in lemma 2.8.

Proof: Define $\mathcal{X} = [0, 1]$ and let Λ be the set of all strictly increasing, continous functions on $\mathcal{X}$ with $\lambda(0) = 0$ and $\lambda(1) = 1$ for $\lambda \in \Lambda$ such that

$$||\lambda|| = \sup_{s \neq t} \left| \log \frac{\lambda(t) - \lambda(s)}{t - s} \right|$$

is finite. For $g_1, g_2 \in D[0, 1]$ define the Skorohod metric (see e.g. Shiryaev (1996))

$$d(f, g) = \inf\{\epsilon > 0 | \exists \lambda \in \Lambda : ||\lambda|| < \epsilon \text{ and } \sup_{t \in [0,1]} |f(t) - g(\lambda(t))| \leq \epsilon\}.$$

Then $(D[0,1], d)$ is separable and complete and weak convergence of the probablility distributions of $\hat{a}_r(j)$ is equivalent to tightness of the finite dimensional distributions of $\hat{f}_r(x)$ (see e.g. Billingsley, chapters 14 and 15). For $s < t$,

$$\hat{f}_r(t) - \hat{f}_r(s) = \sum_k \hat{\alpha}_{l,k}^{(r)}(\phi_{l,k}(t) - \phi_{l,k}(s)) + \sum_{j,k} \hat{w}_{j,k}^{(r)}\hat{\beta}_{j,k}^{(r)}(\psi_{j,k}(t) - \psi_{j,k}(s)).$$

Because the wavelet basis is of bounded variation,

$$\begin{aligned} |\phi_{l,k}(t) - \phi_{l,k}(s)| &= |g(t) - h(t) - (g(s) - h(s))| \\ &\leq |g(t) - g(s)| + |h(t) - h(s)| \\ &\leq c_1(t-s), \end{aligned}$$

where g and h are two nondecreasing functions and c_1 is a suitable constant. Consequently,

$$\begin{aligned} \hat{f}_r(t) - \hat{f}_r(s) &\leq c_1(t-s)(\sum |\hat{\alpha}_{l,k}^{(r)}| + \sum |\hat{\beta}_{j,k}^{(r)}|) \\ &\leq c_2(t-s). \end{aligned}$$

Thus,

$$\begin{aligned} E[|\hat{f}_r(t) - \hat{f}_r(t_1)||\hat{f}_r(t_2) - \hat{f}_r(t)|] &\leq c_2^2(t-t_1)(t_2-t), \\ &\leq c_3(t_2-t_1)^2 \end{aligned}$$

and tightness of the family of distribution functions of $\hat{f}_r$ follows from theorem 15.6 in Billingsley. This proves the lemma. ■

The next corollary is a simple consequence of earlier results and derives the asymptotic properties of the estimated Fourier coefficients.

Corollary 3.3. *Let assumption (C) hold and let* $\hat{\mathbf{f}}$ *be a wavelet estimate of* $\mathbf{f}$. *Define*

$$\hat{a}_r(j) = \int_0^1 \hat{f}_r(t)\exp(-i2\pi jt)dt \ \ (r = 1, \ldots, p),$$

as the empirical Fourier coefficient of f_r *at frequency* j. *Then,*

1. $|\hat{a}_r(j) - a_r(j)| \xrightarrow{p} 0$,

2. *for finite samples of frequencies* $0 \leq j_1 < \cdots < j_k \leq \pi$, *the vector*

$$\sqrt{n}[\hat{a}_r(j_1) - a_r(j_1), \ldots, \hat{a}_r(j_k) - a_r(j_k)]^T$$

converges in distribution to a k-dimensional normal random variable with mean 0 and asymptotic covariance matrix

$$\begin{aligned} & \lim n\,cov(\hat{a}_r(j_p), \hat{a}_r(j_q)) \\ = \; & 2\pi h_{\epsilon;rr}(0)[\sum_k \int \phi_k(x)\exp(-i2\pi px)dx \int \phi_k(y)\exp(i2\pi qy)dy + \\ & \sum_{j,k} w_{j,k}^{(r)} \int \psi_{j,k}(x)\exp(-i2\pi px)dx \int \phi_{j,k}(y)\exp(i2\pi qy)dy]. \end{aligned}$$

Proof: Refer to the proof of lemma 3.1. to get

$$\begin{aligned} |\hat{a}_r(j) - a_r(j)| &= |\int_0^1 \exp(-i2\pi qt)(\hat{f}_r(t) - f_r(t))dt| \\ &\leq \int_0^1 |\hat{f}_r(t) - f_r(t)|dt \xrightarrow{p} 0 \; (n \to \infty). \end{aligned}$$

For the second part, write

$$\sqrt{n}(\hat{a}_r(j) - a_r(j)) = \int_0^1 \exp(-i2\pi jt)\vartheta_n(t)dt$$

and define

$$\begin{aligned} A: \; D[0,1] &\to \mathbb{R} \\ \vartheta_n &\longmapsto \int_0^1 \exp(-i2\pi jt)\vartheta_n(t)dt. \end{aligned}$$

Then, A is a linear functional on the sample paths of $\hat{f}_r(t) \in D[0,1]$ and the asymptotic distribution of $\hat{a}_r(j)$ is derived by applying a functional Central Limit Theorem. The covariances are a simple consequence of earlier consid-

erations.

$$\begin{aligned}
&\lim_{n\to\infty} n\mathrm{cov}\left(\hat{a}_r(j_1), \hat{a}_r(j_2)\right) \\
= &\iint \exp(-i2\pi(xj_1 - yj_2)) \lim_{n\to\infty} n\mathrm{cov}\left(\hat{f}_r(x), \hat{f}_r(y)\right) dx\, dy \\
= &\ 2\pi h_{\epsilon;rr}(0)[\sum_k \int_0^1 \phi_{l,k}(x)\exp(-i2\pi j_1 x)dx \int_0^1 \phi_{l,k}(y)\exp(i2\pi j_2 y)dy \\
&+ \sum_{j,k} w_{j,k}^{(r)} \int_0^1 \psi_{j,k}(x)\exp(-i2\pi j_1 x)dx \int_0^1 \psi_{j,k}(y)\exp(i2\pi j_2 y)dy].
\end{aligned}$$

For finite samples $(j_1, \ldots, j_k)$ the result follows as usual by applying the Cramer-Wold device. ■

Remark 3.5. *With the notations of remark 3.3.,*

$$\lim_{n\to\infty} ncov(\hat{a}_r(j_1), \hat{a}_r(j_2)) = 2\pi h_{\epsilon;rr}(0)\left(\sum a_{l,k}^{\phi}(j_1)\overline{a_{l,k}^{\phi}(j_2)} + \sum_{j,k} w_{j,k}^{(r)} a_{j,k}^{\psi}(j_1)\overline{a_{j,k}^{\psi}(j_2)}\right)$$

and

$$\lim_{n\to\infty} nvar\left(\hat{a}_r(j)\right) = 2\pi h_{\epsilon;rr}(0)\left(\sum |a_{l,k}^{\phi}(j)|^2 + \sum_{j,k} w_{j,k}^{(r)} |a_{j,k}^{\psi}(j_1)|^2\right).$$

Hence, the variance at frequency j is decomposed into energy contributions of the individual basis functions with corresponding nonzero wavelet coefficient.

Remark 3.6. *The lemma can be easily extended to joint distributions between $\hat{a}_r(j)$ and $\hat{a}_s(j)$ by replacing $h_{\epsilon;rr}(0)$ by the corresponding cross spectrum.*

Once the asymptotic distribution of the Fourier coefficients is developed, the asymptotic distribution of the regression spectrum follows by straightforward calculation. Techniques similar to those of corollary 3.1. imply the distribution of the spectrum estimate in (3.38). Due to the convolution formula for

Fourier transforms the distribution coincides with that in theorem 3.5. This is summarized in

Remark 3.7. *Let $h_{rs}(j)$ be estimated by*

$$\hat{h}_{rs}(j) = \int \hat{\gamma}_{rs}(u) \exp(-i2\pi ju)du.$$

Then,

$$\begin{aligned} \hat{h}_{rs}(j) &= \int_0^1 \int_0^1 \exp(-i2\pi ju)\hat{f}_r(t+u)\hat{f}_s(t)dtdu \\ &= \hat{a}_r(j)\overline{\hat{a}_s(j)} \end{aligned}$$

such that both estimators conincide.

This result is particularly useful for application and will be used in the following section.

3.5 Algorithm and data examples

3.5.1 General considerations

Consider example 3.1. with two functions that are shifted versions of each other. In this case, the phase spectrum consists of a straight line modulo 2π, with the slope being proportional to the shift. Discontinuities occur at frequencies where $\phi_{rs}(j)$ crosses $-\pi$ or π. More generally, if we are given a plot of a phase spectrum between two deterministic functions the detection of any linear or piecewise linear curve may be interpreted as a constant lag over this particular range of frequencies. However, if $\mathbf{f}$ and $\mathbf{H}(j)$ have to be estimated, we have a superposition of the linear structure of the phase spectrum with the phase spectrum of the noise component. Comparing (3.36) with (3.37) we see

that $\sigma^2_{\kappa;rs}(j)$ and $\sigma^2_{\phi;rs}(j)$ essentially differ by the factor $\kappa^*_{rs}(j)^{-2}$. Therefore, $\sigma^2_{\phi;rs}(j)$ will be relatively large compared to $\sigma^2_{\kappa;rs}(j)$ for all frequencies where the corresponding spectral modulus is small, and it will be relatively small where the spectrum modulus is large. Thus, in general, the phase spectrum will look more erratic than the amplitude spectrum, and estimation of common frequencies in a multivariate trend function might be easier than estimation of the phase shift. A pure visual inspection of the phase spectrum may not be sufficient to detect any linear structure. In the next section, we propose a simple algorithm that takes this into account. In a first step, frequencies are identified where the amplitude spectrum is significantly larger than 0. In a second step, the phase shift is estimated using theses frequencies only.

3.5.2 Algorithm

A data-driven algorithm for estimating the regression cross spectrum can be defined as follows.

1. Choose a wavelet basis $\{\phi_{l,k}(x),\ \psi_{j,k}(x),\ l \in \mathbb{Z}, j \geq l, k \in \mathbb{Z}\}$ and thresholds $\lambda_j,\ j = 1, \ldots, J_n$, according to theorem 3.1., and estimate $f_r,\ r = 1, \ldots, p$. This step can be carried out, for instance, using the function WAVESHRINK in the S-Plus wavelet module.

2. Apply the fast Fourier transform to obtain

$$\hat{a}_r(j) = \frac{1}{n} \sum_{t=1}^{n} \hat{f}_r(t/n) e^{-i2\pi jt/n}, \quad r = 1, \ldots, p,$$

and calculate the regression cross spectrum

$$\hat{h}_{rs}(j) = \hat{a}_r(j)\overline{\hat{a}_s(j)}$$

and estimates of the amplitude and phase spectrum (see (3.34) and (3.35)).

3. Estimate the cross spectrum $h_{\epsilon,rs}(0)$ of $\boldsymbol{\epsilon}(i)$ from the estimated residuals $\hat{\boldsymbol{\epsilon}} = \hat{\mathbf{Y}} - \hat{\mathbf{f}}$.

4. Use equation (3.36) to estimate $\sigma^2_{\kappa;rs}(j)$ and determine the set

$$J^* = \{j : \ \hat{\kappa}^*_{rs}(j) > c_{\kappa;rs} \cdot \sqrt{\sigma^2_{\kappa;rs}(j)}\}$$

   for a suitably chosen $c_{\kappa;rs} \in \mathbb{R}$.

5. To estimate the phase shift, apply a local robust regression to the points $\{(j, \phi_{rs}(j)) : j \in J^*\}$, taking into account possible 2π-jumps.

3.5.3 Simulated example

The application of the asymptotic results and the practical performance of the algorithm are illustrated in a small simulation study. Several applications to real data can be found in chapter 5.
The trend function $\mathbf{f}(x) = (f_1(x), f_2(x))^T$ $(x \in [0,1])$ is defined by

$$f_1(x) = -\sin(4\pi x) - \sin(10\pi x) - \sin(20\pi x) - \sin(30\pi x) - \sin(51\pi x),$$

$$f_2(x) = f_1(x + \Delta),$$

and

$$\Delta = .03125.$$

Moreover, $\epsilon_1(i)$, $\epsilon_2(i)$ are iid $N(0, \sigma^2_\epsilon)$ with $\sigma^2_\epsilon = 4$ and ϵ_1, ϵ_2 independent of each other. Figure 3.1 shows the simulated series (figures 3.1(a),(b)), the true trend functions f_1 and f_2 (figures 3.1(c),(d)), and the true regression cross spectrum in terms of the amplitude (figure 3.1(e)) and the phase spectrum (figure 3.1(f)). The sample size is $n = 2048$. For better visibility, only the lowest 80 frequencies are used in the spectral plots. The estimated spectral modulus and the phase spectrum between $\hat{f}_1$ and $\hat{f}_2$, obtained by wavelet thresholding with $s12-$wavelets (see e.g. Daubechies (1992), Bruce

and Gao (1996)) and λ_j as in remark 2.10., are displayed in figures 3.1(g) and (h). While the common frequencies are identified quite accurately, the linear structrue of the theoretical regression phase spectrum is lost almost entirely in its (unweighted) empirical counterpart. However, if we consider only values of $\hat{\phi}_{12}$ where $\hat{\kappa}^*_{12}(j)$ exceeds four times its standard deviation (solid line in figure 3.1(g)), the linear structure can be identified. In figure 3.1(h), the five frequencies corresponding to the five highest values of $\hat{\kappa}^*_{12}(j)$ are marked by black squares. The estimate of the phase line based on these frequencies is very close to the true line.

To check the accuracy of our estimates, a small simulation study was carried out. For n =256, 512, 1024 and 2048, 500 series were simulated, and the spectral density and the lag between f_1 and f_2 were estimated. Figure 3.2 shows a comparison of the simulated standard deviation of $\hat{\kappa}^*_{12}(j)$ and the values of $\sigma_{\kappa;12}$ (equation (3.36)) calculated by plugging in the true and the estimated functions f_j respectively. The standard deviations are fairly close together for $n = 256$ and almost identical for $n = 2048$. This confirms the theoretical results.

n	256	512	1024	2048
true value	-.03125	-.03125	-.03125	-.03125
median	-0.0337069	-0.0317893	-0.0312525	-0.0312118
mean	-0.0429855	-0.0319160	-0.0312776	-0.0311385
std.dev.	0.1490462	0.01350248	0.00311929	0.00208201
mse	0.02768125	0.00417189	0.00391940	0.00389665
skewness	0.78195065	0.48571213	0.01500243	0.18100068

Table 3.1: Detailed summary of the lag estimation for various sample sizes, in each case running 500 simulations.

Detailed results for the lag-estimates are listed in table 3.1. Boxplots in figure

3.2 illustrate the fast convergence of $\hat{\Delta}$ to the true value $\Delta = -.03125$. Note that for small sample sizes, the true frequencies in the amplitude spectrum might fail to show up due to insufficient accuracy of the trend estimates. In particular, high deterministic frequencies may be hidden or smoothed out by the shrinkage estimate. Moreover, the true lead-lag structure may not be detected due to effects of the noise component. In the example here, a reliable estimate of Δ is obtained for $n = 512$, whereas $n = 256$ seems to be too small.

3.6 Final remarks

We defined the regression cross covariance and cross spectrum for multivariate deterministic trend functions. This is a nonparametric multivariate extension of an analogous concept used by Grenander and Rosenblatt to obtain asymptotic results in the context of parametric regression. The usefulness of the nonparametric regression spectrum goes far beyond a purely mathematical device. It can be used as a data analytical tool to identify common frequencies and lead-lag effects in multivariate time series with strong deterministic components. The greatest challenge appears to be estimation of the phase spectrum. The question in how far more accurate methods than the algorithm proposed here can be devised, will be worth pursuing in future research. In section 6 the good performance in practice of the developed theory will be illustrated with several applications.

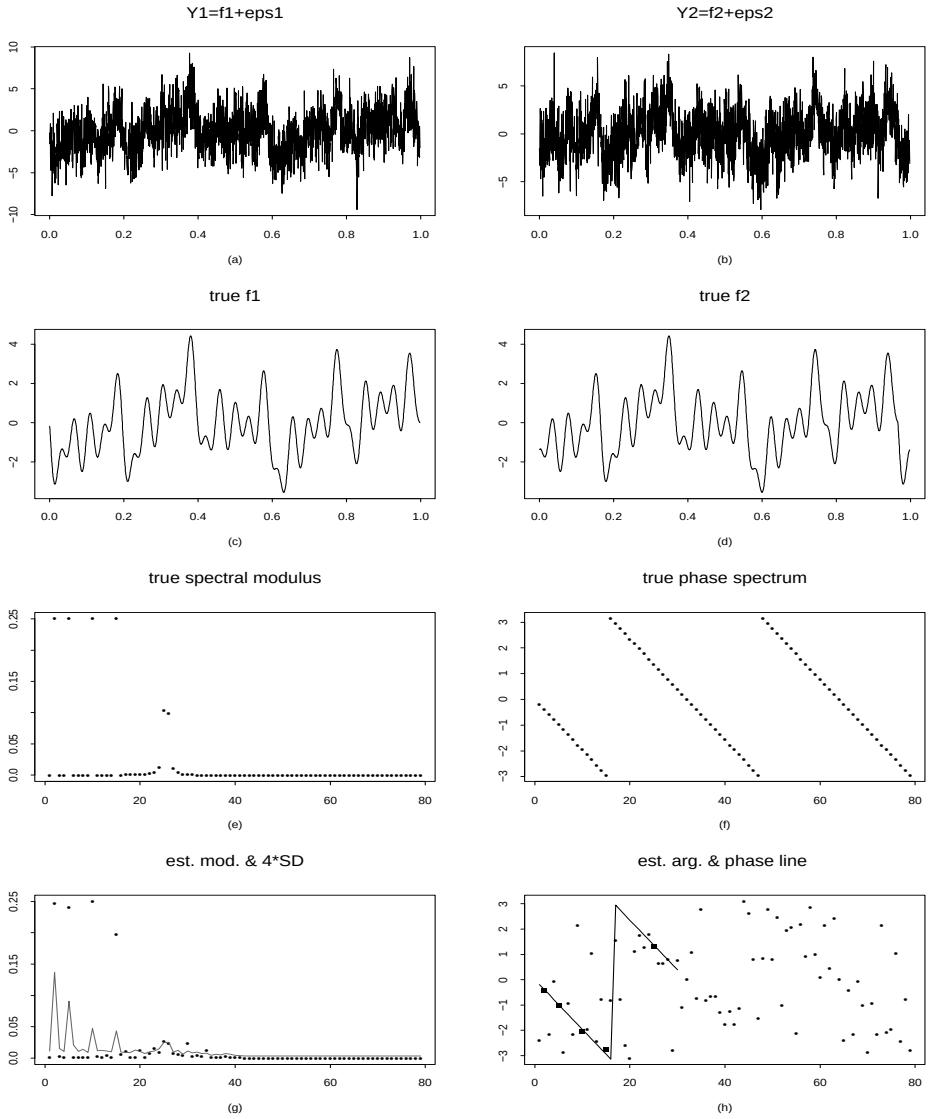

Figure 3.1: Simulated series (Y_1, Y_2) of length $n = 2048$ (a,b), true trend functions (f_1, f_2) (figures (c,d)), amplitude and phase spectrum (e,f) and estimated amplitude and phase spectrum (g,h). The solid line in figure (g) represents $4\sqrt{\text{var}(\kappa^*_{rs}(j))}$, the solid line in figure (h) corresponds to the estimated phase line.

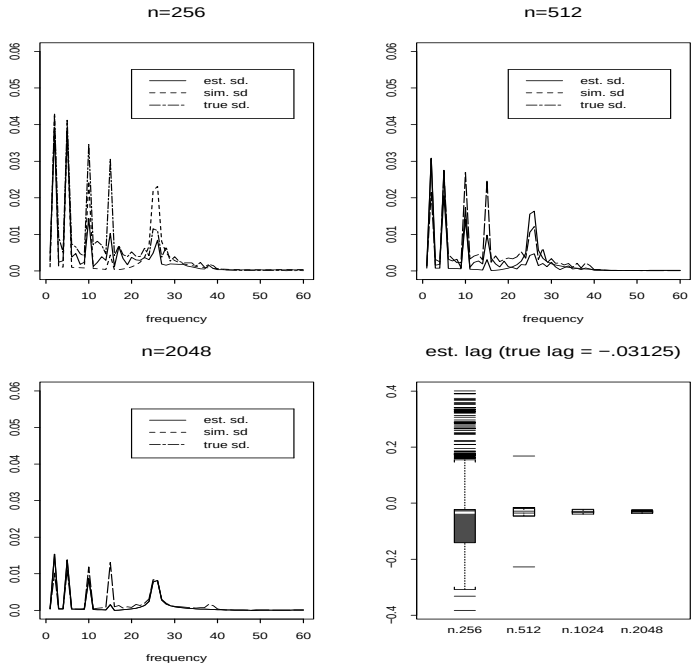

Figure 3.2: Simulated, estimated and true asymptotic standard deviation of the amplitude spectrum for various sample sizes n. Boxplots of the lag estimates are also given for $n = 256, 512, 1024$ and $n = 2048$.

Chapter 4

Periodogram Estimation

In the previous chapter we defined the regression cross covariance and spectrum $\hat{\mathbf{\Gamma}}$ and $\hat{\mathbf{H}}$ and considered estimation of these quantities based on wavelet thresholding of the unknown deterministic trend function. This approach allows for estimating common frequency components and possible lags within the deterministic component.

Traditional techniques for analyzing stationary time series deal with spectral estimation based on the periodogram. Section 4.1 presents the definition of an adjusted periodogram that accounts for the presence of a deterministic component. This modification will allow to directly estimate the regression spectrum consistently. We commence by assuming

$$\mathbf{Y}(i) = \mathbf{f}(t_i) + \boldsymbol{\epsilon}(i),$$

with $\boldsymbol{\epsilon}(i)$ being independent and identically distributed. The assumptions on $\boldsymbol{\epsilon}(i)$ will then be generalized to linear process. The essential result is given in theorem 4.2. and shows that if $\boldsymbol{\epsilon}(i)$ is generated by a generalized linear process, the periodogram estimate yields an asymptotically unbiased estimator whose distribution tends to the normal as $n \to \infty$. Section 4.2 deals with estimation of the modulus and phase of the regression spectrum. Due to the asymptotic

independence of the periodogram at different frequencies, we may derive the asymptotic distribution of the slope of the phase spectrum in case that the true underlying trend components are shifted versions of each other. We will conclude this section with a comparison of the periodogram and the regression spectrum estimate of chapter 3.

4.1 The periodogram

Given n observations of a multivariate vector $\mathbf{Y}(i)\,(i = 1, \ldots, n)$, the periodogram of $\mathbf{Y}(i)$ at frequency ω_j is defined by

$$\mathbf{I}(\omega_j) \;=\; \frac{1}{n}\left(\sum_{s=1}^{n}\mathbf{Y}(s)\exp(-i\omega_j s)\right)\left(\sum_{t=1}^{n}\mathbf{Y}(t)\exp(i\omega_j t)\right)^T$$

where $\omega_j = 2\pi j/n$, $\omega_j \in [-\pi, \pi]$. It is well-known in literature that consistency of the periodogram as an estimator of the spectrum of a stationary process requires preliminary smoothing. (see e.g. Brockwell and Davis (1987), Priestley (1989), Jenkins and Watts (1968)).
Consider again the model

$$\mathbf{Y}(i) = \mathbf{f}(t_i) + \boldsymbol{\epsilon}(i) \tag{4.1}$$

where $\mathbf{f}(t) = (f_1(t), \ldots, f_p(t))^T \in \mathbb{R}^p$ with $f_r(t) \in L_2(\mathbb{R})$ $(r = 1 \ldots, p)$. Denote by $\mathbf{A}(\omega_j) = (A_1(\omega_j), \ldots, A_p(\omega_j))^T$ and $\mathbf{B}(\omega_j) = (B_1(\omega_j), \ldots, B_p(\omega_j))^T$ the vectors

$$\mathbf{A}(\omega_j) = \sum \mathbf{f}(t_k)\exp(-i\omega_j k)$$

and

$$\mathbf{B}(\omega_j) = \sum \boldsymbol{\epsilon}(k)\exp(-i\omega_j k),$$

respectively. Then, the periodogram can be written as

$$\begin{aligned}\mathbf{I}(\omega_j) \;=\; & \frac{1}{n}\left[\mathbf{A}(\omega_j)\mathbf{A}(\omega_j)^* + \mathbf{B}(\omega_j)\mathbf{B}(\omega_j)^*\right. \\ & \left. + \mathbf{A}(\omega_j)\mathbf{B}(\omega_j)^* + \mathbf{B}(\omega_j)\mathbf{A}(\omega_j)^*\right],\end{aligned}$$

where "$*$" denotes the complex conjugate transpose.

Definition 4.1. *Assume that model (4.1) holds. The matrix* $\mathbf{I_f}(\omega_j) = [I_{f;rs}(\omega_j)]_{1\leq r,s\leq p}$ *with*

$$\mathbf{I_f}(\omega_j) = \frac{1}{n}\mathbf{A}(\omega_j)\mathbf{A}(\omega_j)^*$$

will be called the regression periodogram of the deterministic function $\mathbf{f}$*. As usually, we will refer to* $\mathbf{I}_{\boldsymbol{\epsilon}}(\omega_j) = [I_{\boldsymbol{\epsilon};rs}(\omega_j)]_{1\leq r,s\leq p}$ *with*

$$\mathbf{I}_{\boldsymbol{\epsilon}}(\omega_j) = \frac{1}{n}\mathbf{B}(\omega_j)\mathbf{B}(\omega_j)^*$$

as the periodogram of the stationary multivariate series $\boldsymbol{\epsilon}(i) = (\epsilon_1(i), \ldots, \epsilon_p(i))^T$.

For stationary $\boldsymbol{\epsilon}(i)$,

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\boldsymbol{\epsilon}(k)\exp(-i\omega_j k)$$

converges in distribution to a normal random variable which implies

$$\frac{1}{n}\mathbf{B}(\omega_j)\mathbf{B}(\omega_j)^* = O_p(1).$$

We refer again to Polya and Szegö to notice that for a function $f_r(\cdot)$ that is bounded and of bounded variation

$$\begin{aligned}|\frac{1}{n}A_r(\omega_j) - a_r(j)| &= |\frac{1}{n}\sum f_r(t_k)\exp(-i\omega_j k) - \int f_r(x)\exp(-i2\pi jx)dx| \\ &= O(n^{-1}).\end{aligned}$$

Therefore,

$$\frac{1}{n}\mathbf{A}(\omega_j)\mathbf{B}(\omega_j)^* = O_p(n^{1/2})$$

and

$$\mathbf{I_f}(\omega_j) = O(n),$$

implying that

$$\mathbf{I}(\omega_j) = O(n).$$

Hence, if model (4.1) holds, the traditional periodogram is of order $O(n)$ and thus cannot be applied as an estimate for the spectrum. However, estimation of the regression spectrum may be based on a standardized periodogram. The case of independent and identically distributed noise is presented in

Theorem 4.1. *Assume that model (4.1) holds and let* $\boldsymbol{\epsilon}(i)$ *be independent and identically distributed with mean 0 and non-singular covariance matrix* $\boldsymbol{\Sigma}$*. Assume further that the fourth moments of* $\boldsymbol{\epsilon}(i)$ *exist. Then, for each pair* (r, s) *and frequencies* $\omega_j = \frac{2\pi j}{n} \in [0, \pi]$*,*

1. $\frac{1}{n} I_{rs}(\omega_j) - h_{rs}(j) = O_p(n^{-1/2})$*,*
2. $E(\frac{1}{n} I_{rs}(\omega_j)) - h_{rs}(j) = O(n^{-1})$*,*
3. *for Fourier frequencies* $0 \leq \omega_{j_1}, \ldots, \omega_{j_k} \leq \pi$*,* $k \in \mathbb{N}$*, and for each pair* (r, s) *the vector*

$$\sqrt{n}[(\frac{1}{n} I_{rs}(\omega_{j_1}) - h_{rs}(j_1)), \ldots, (\frac{1}{n} I_{rs}(\omega_{j_k}) - h_{rs}(j_k))]^T$$

converges in distribution to a k*-dimesional normal random vector with mean* $\mathbf{0}$ *and covariance*

$$\lim_{n\to\infty} n cov(n^{-1} I_{rs}(\omega_j), n^{-1} I_{rs}(\omega_{j'})) = O(n^{-2}),$$

where $\omega_j \neq \omega_{j'}$*, and*

$$\lim_{n\to\infty} n var(n^{-1} I_{rs}(\omega_j)) \quad = \quad \Sigma_{rr}|a_s(j)|^2 + \Sigma_{ss}|a_r(j)|^2 + O(n^{-1}) \tag{4.2}$$

where Σ_{rs} *is the* (r, s)*th entry of the matrix* $\boldsymbol{\Sigma}$*.*

Proof: Remember that $h_{rs}(j) = a_r(j)\overline{a_s(j)}$, then

$$\frac{1}{n^2} A_r(\omega_j)\overline{A_s(\omega_j)} - h_{rs}(j) = O(n^{-1}). \tag{4.3}$$

Together with the results above we have

$$\begin{aligned} & \frac{1}{n} I_{rs}(\omega_j) - h_{rs}(j) \\ = \; & \frac{1}{n^2}[B_r(\omega_j)\overline{A_s(\omega_j)} + A_r(\omega_j)\overline{B_s(\omega_j)}] + O_p(n^{-1}) + O(n^{-1}) \\ = \; & O_p(n^{-1/2}). \end{aligned}$$

For the second part consider

$$\begin{aligned} E(n^{-1}I_{rs}(\omega_j)) &= \frac{1}{n^2}[A_r(\omega_j)\overline{A_s(\omega_j)} + E(B_r(\omega_j)\overline{B_s(\omega_j)})], \\ &= \big(h_{rs}(j) + O(n^{-1})\big) + n^{-1}E(I_{\epsilon;rs}(\omega_j)). \end{aligned}$$

Results from traditional spectral analysis show that $E(I_{\epsilon;rs}(\omega_j))$ converges to $2\pi h_{\epsilon;rs}(\omega_j)$ uniformly for all frequencies ω_j such that 2. is proved. Denote

$$\frac{1}{\sqrt{n}}\left(a_r(j)\overline{B_s(\omega_j)} + B_r(\omega_j)\overline{a_s(j)}\right) \;=:\; \alpha(j) + i\beta(j),$$

where

$$\alpha(j) = \frac{1}{\sqrt{n}}\sum \cos(\omega_j t)[a_r(j)\epsilon_s(t) + \overline{a_s(j)}\epsilon_r(t)], \tag{4.4}$$

$$\beta(j) = \frac{1}{\sqrt{n}}\sum \sin(\omega_j t)[a_r(j)\epsilon_s(t) - \overline{a_s(j)}\epsilon_r(t)]. \tag{4.5}$$

Consider the variance for the adjusted periodogram (the result for covariances follows similarly):

$$\begin{aligned} \text{var}\,(n^{-1}I_{rs}(\omega_j)) &= n^{-1}E[|\alpha(j) + i\beta(j)|^2] && (4.6)\\ &\quad +\text{cov}\,(n^{-1}I_{\epsilon;rs}(\omega_j), \alpha(j) + i\beta(j)) && (4.7)\\ &\quad +\text{cov}\,(\alpha(j) + i\beta(j), n^{-1}I_{\epsilon;rs}(\omega_j)) && (4.8)\\ &\quad +\text{var}\,(n^{-1}I_{\epsilon;rs}(\omega_j)). && (4.9) \end{aligned}$$

Comparison with Brockwell and Davis (p.429) gives

$$\text{cov}\,(I_{\epsilon;rs}(\omega_j), I_{\epsilon;rs}(\omega_{j'})) = \begin{cases} \Sigma_{rr}\Sigma_{ss} + O(n^{-1}); & 0 < \omega_j = \omega_{j'} < \pi, \\ \Sigma_{rr}\Sigma_{ss} + \Sigma_{rs}\Sigma_{sr} + O(n^{-1}); & \omega_j = \omega_{j'} \in \{0, \pi\}, \\ O(n^{-1}); & \omega_j \neq \omega_{j'}, \end{cases}$$

where the remainders contain the fourth order cumulants between $\epsilon_r(i)$ and $\epsilon_s(i)$. The covariances in (4.7) and (4.8) consist of terms of the form

$$\begin{aligned}\operatorname{cov}\left(I_{\epsilon;rs}(\omega_j), B_r(\omega_j)\right) &= \frac{1}{n}\sum_{t,u,v}\exp(-i\omega_j(t-u+v))E(\epsilon_r(t)\epsilon_s(u)\epsilon_r(v)) \\ &= \frac{E(\epsilon_r(1)^2\epsilon_s(1))}{n}\sum_{t=1}^{n}\exp(-i\omega_j t) = 0,\end{aligned}$$

where $E(\epsilon_r(t)^2\epsilon_s(t))$ is independent of t. Simple considerations show that the $2p$-dimensional real valued random vector

$$\mathbf{U_n}(\omega_j) = n^{-1/2}\begin{pmatrix} \sum \boldsymbol{\epsilon}(t)\cos(\omega_j t) \\ \sum \boldsymbol{\epsilon}(t)\sin(\omega_j t)\end{pmatrix} \tag{4.10}$$

is asymptotically normal with mean $\mathbf{0}$ and covariance matrix

$$E(\mathbf{U_n}(\omega_j)\mathbf{U_n}(\omega_j)^T) = \frac{1}{2}\begin{pmatrix}\boldsymbol{\Sigma} & 0 \\ 0 & \boldsymbol{\Sigma}\end{pmatrix}.$$

Furthermore, for all Fourier frequencies $\omega_j \neq \omega_{j'}$,

$$E(\mathbf{U_n}(\omega_j)\mathbf{U_n}(\omega_{j'})^T) = \mathbf{0}. \tag{4.11}$$

Therefore,

$$\begin{aligned} & n^{-1}E(B_r(\omega_j)\overline{B_s(\omega_j)}) \\ = \;& n^{-1}\operatorname{cov}\left(\sum \epsilon_r(t)(\cos(\omega_j t) - i\sin(\omega_j t)), \sum \epsilon_s(t)(\cos(\omega_j t) - i\sin(\omega_j t))\right) \\ = \;& \frac{1}{2}(\Sigma_{rs} + \Sigma_{sr}) = \Sigma_{rs}.\end{aligned}$$

Similarly, for all pairs (r, s), $1 \leq r \neq s \leq p$,

$$\begin{aligned}\operatorname{var}\left(n^{-1/2}B_r(\omega_j)\right) &= \Sigma_{rr} \quad \text{and} \\ \operatorname{cov}\left(B_r(\omega_j), \overline{B_s(\omega_j)}\right) &= \operatorname{cov}\left(\overline{B_s(\omega_j)}, B_r(\omega_j)\right) = 0.\end{aligned}$$

Notice that

$$\sum_{t=1}^{n}\cos^2(\omega_j t) = \frac{n}{2}.$$

Then, the variance in (4.6) follows by considering

$$\begin{aligned}\tau_\alpha^2(j) &:= \mathrm{var}\,(\alpha(j)) \\ &= \frac{1}{n}\sum_{t=1}^{n}\cos^2(\omega_j t)[\Sigma_{rr}|a_s(j)|^2+\Sigma_{ss}|a_r(j)|^2 \\ &\quad +\overline{a_s(j)a_r(j)}\mathrm{cov}\,(\epsilon_r(t),\epsilon_s(t))+a_r(j)a_s(j)\mathrm{cov}\,(\epsilon_s(t),\epsilon_r(t))] \\ &= \frac{1}{2}(\Sigma_{rr}|a_s(j)|^2+\Sigma_{ss}|a_r(j)|^2)+\Sigma_{rs}\mathfrak{Re}\{a_s(j)a_r(j)\}.\end{aligned}$$

Correspondingly,

$$\tau_\beta^2(j) := \mathrm{var}\,(\beta(j)) = \frac{1}{2}(\Sigma_{rr}|a_s(j)|^2+\Sigma_{ss}|a_r(j)|^2)-\Sigma_{rs}\mathfrak{Re}\{a_s(j)a_r(j)\}$$

and

$$\begin{aligned}&\tau_{\alpha,\beta}(j) := \mathrm{cov}\,(\alpha(j),\beta(j)) \\ &= \frac{1}{n}\sum_{t=1}^{n}\cos(\omega_j t)\sin(\omega_j t)\mathrm{cov}\,(\overline{a_s(j)}\epsilon_r(t)+a_r(j)\epsilon_s(t), a_r(j)\epsilon_s(t)-\overline{a_s(j)}\epsilon_r(t)),\end{aligned}$$

where the covariance is independent of t. Using the orthogonality relations for trigonometric functions we get that $\tau_{\alpha,\beta}(j)=0$.

For the asymptotic distribution, write

$$\sqrt{n}(\frac{1}{n}I(\omega_j)-h_{rs}(j)) \;=\; \alpha(j)+i\beta(j)+O_p(n^{-1/2}).$$

According to (4.10) the asymptotic distribution of the real and the imaginary part are both univariately normal. Because $\alpha(j)$ and $\beta(j)$ are uncorrelated they are asymptotically independent. Therefore,

$$\begin{pmatrix}\alpha(j)\\ \beta(j)\end{pmatrix} \xrightarrow{d} \mathcal{N}\left(\mathbf{0},\begin{bmatrix}\tau_\alpha^2(j) & 0\\ 0 & \tau_\beta^2(j)\end{bmatrix}\right),$$

and $\alpha(j)+i\beta(j)$ converges in distribution to a complex valued normal random variable with mean 0 and covariance

$$\tau_{\alpha+i\beta}^2(j)=\tau_\alpha^2(j)+\tau_\beta^2(j)=\Sigma_{rr}|a_s(j)|^2+\Sigma_{ss}|a_r(j)|^2. \tag{4.12}$$

This implies

$$\sqrt{n}(\frac{1}{n}I_{rs}(\omega_j)-h_{rs}(j)) \;\xrightarrow{d}\; \mathcal{N}_c(0,\Sigma_{rr}|a_s(j)|^2+\Sigma_{ss}|a_r(j)|^2). \tag{4.13}$$

The result for finite samples $(\omega_{j_1}, \ldots, \omega_{j_k})$ follows as usually by applying the Cramer-Wold device. ■

Remark 4.1. *The variance of the adjusted periodogram is of order $O(n^{-1})$ such that $n^{-1}I_{rs}(\omega_j)$ is a consistent estimate for $h_{rs}(j)$. In contrast, traditional periodogram estimation of the spectrum of a stationary time series requires further smoothing.*

Remark 4.2. *The asymptotic distribution in theorem 4.1. essentially differs from the stationary case, where the traditional periodogram converges to a product of normally distributed random variables.*

Remark 4.3. *Note in particular that the standardized periodogram is asymptotically independent for different frequencies. We will use this fact further down to estimate the slope of the phase line.*

Remark 4.4. *The periodogram is given by a composition of stochastic and deterministic components. This carries over to the asymptotic variance at frequency j that is given by a product of the covariance of the stationary part with the regression spectrum.*

The results of theorem 4.1. will now be extended to linear processes. The following lemma summerizes some asymptotic results of Hannan (1970) in case that $\{\boldsymbol{\epsilon}(i)\}$ is generated by a generalized linear process.

Lemma 4.1. *Assume that the sequence $\{\boldsymbol{\epsilon}(i)\}$, $i = 1, \ldots, n$, is of the form*

$$\boldsymbol{\epsilon}(i) = \sum_{j=-\infty}^{\infty} \mathbf{A}(j)\mathbf{Z}(i-j), \tag{4.14}$$

where $\mathbf{A}(j) = (A_{lk}(j))_{1\leq l,k\leq p}$ *are* $p \times p$*-matrices such that for all pairs* (l, k)*,*

$$\sum_{j\in\mathbb{Z}} |A_{lk}(j)||j|^{\frac{1}{2}} < \infty,$$

and the sequence $\mathbf{Z}(i)$ *is independent and identically distributed with mean 0, non-singular covariance matrix* $\mathbf{\Sigma}$ *and existing fourth moments. Then,*

$$\mathbf{g}(\omega_j) = n^{-1/2} \begin{pmatrix} \sum \boldsymbol{\epsilon}(t)\cos(\omega_j t) \\ \sum \boldsymbol{\epsilon}(t)\sin(\omega_j t) \end{pmatrix} = \begin{pmatrix} \mathbf{g}_1(\omega_j) \\ \mathbf{g}_2(\omega_j) \end{pmatrix}$$

with $\mathbf{g}_l(\omega_j) = (g_{l1}(\omega_j), \ldots, g_{lp}(\omega_j))^T$*,* $l \in \{1, 2\}$*, converges in distribution to a* $2p$*-dimensional random variable with mean* $\mathbf{0}$ *and asymptotic covariance matrix*

$$2\pi \cdot \frac{1}{2} \begin{pmatrix} \mathbf{C}(\omega_j) & \mathbf{Q}(\omega_j) \\ -\mathbf{Q}(\omega_j) & \mathbf{C}(\omega_j) \end{pmatrix},$$

where $\mathbf{h}_\epsilon(\omega_j) = \frac{1}{2}(\mathbf{C}(\omega_j) - i\mathbf{Q}(\omega_j))$ *with* $\mathbf{C}(\omega_j) = [c_{\epsilon;rs}(\omega_j)]_{1\leq r,s\leq p}$ *and* $\mathbf{Q}(\omega_j) = [q_{\epsilon;rs}(\omega_j)]_{1\leq r,s\leq p}$ *is the spectral density matrix of* $\boldsymbol{\epsilon}(i)$ *at Fourier frequency* $\omega_j = 2\pi j/n$*. Furthermore* $\mathbf{g}(\omega_j)$ *and* $\mathbf{g}(\omega_{j'})$ *are asymptotically independent for* $\omega_j \neq \omega_{j'}$*.*

Hannan's results may now be used to generalize theorem 4.1.

Theorem 4.2. *Assume that model (4.1) holds with* $\{\boldsymbol{\epsilon}(i)\}$ *as in (4.14). Then for each pair* (r, s)*,* $1 \leq r, s \leq p$*,*

1. $\frac{1}{n}I_{rs}(\omega_j) - h_{rs}(j) = O_p(n^{-1/2})$*,*
2. $E(\frac{1}{n}I_{rs}(\omega_j)) - h_{rs}(j) = O(n^{-1})$ *.*
3. *For Fourier frequencies* $0 \leq \omega_{j_1}, \ldots, \omega_{j_k} \leq \pi$*, the vector*

$$\sqrt{n}[(\frac{1}{n}I_{rs}(\omega_{j_1}) - h_{rs}(j_1)), \ldots, (\frac{1}{n}I(\omega_{j_k}) - h_{rs}(j_k))]^T$$

converges in distribution to a k-dimensional normal random variable with mean $\mathbf{0}$*,*

$$\lim_{n\to\infty} n cov(n^{-1}I_{rs}(\omega_j), n^{-1}I_{rs}(\omega_{j'})) \;=\; O(n^{-2}), \;\; \omega_j \neq \omega_{j'},$$

and

$$\begin{aligned} &\lim_{n\to\infty} n var(n^{-1}I_{rs}(\omega_j)) \\ =\;& 2\pi\left(h_{\epsilon;rr}(\omega_j)|a_s(j)|^2 + h_{\epsilon;ss}(\omega_j)|a_r(j)|^2\right) + O(n^{-1}). \end{aligned} \tag{4.15}$$

Proof: The first two parts of the proof are similar to those of the last theorem. For the asymptotic distribution and variance consider equations (4.6)-(4.9). Refer again to Brockwell and Davis (p. 431) to get

$$\operatorname{cov}\left(I_{\epsilon;rs}(\omega_j), I_{\epsilon;rs}(\omega_{j'})\right) = \begin{cases} (2\pi)^2 h_{\epsilon;rr}(\omega_j)h_{\epsilon;ss}(\omega_j) + O(n^{-1/2}); & \\ & 0 < \omega_j = \omega_{j'} < \pi, \\ (2\pi)^2(h_{\epsilon;rr}(\omega_j)h_{\epsilon;ss}(\omega_j) + h_{\epsilon;rs}(\omega_j)h_{\epsilon;sr}(\omega_j)) & \\ +O(n^{-1/2}); & \omega_j = \omega_{j'} \in \{0,\pi\}, \\ O(n^{-1}); & \omega_j \neq \omega_{j'}. \end{cases}$$

Using the results of lemma 4.1.,

$$\operatorname{var}\left(\frac{1}{\sqrt{n}}B_r(\omega_j)\right) \;= 2\pi h_{\epsilon;rr}(\omega_j) \;\; (1 \leq r \leq p), \tag{4.16}$$

and denote

$$h_{\epsilon;rr}(\omega_j) = c_{\epsilon;rs}(\omega_j) - iq_{\epsilon;rs}(\omega_j).$$

Then,

$$\begin{aligned} &\operatorname{cov}\left(\frac{1}{\sqrt{n}}B_r(\omega_j), \frac{1}{\sqrt{n}}\overline{B_s(\omega_j)}\right) \\ =\;& \operatorname{cov}\left(g_{1r}(\omega_j) - ig_{2r}(\omega_j), g_{1s}(\omega_j) + ig_{2s}(\omega_j)\right) \\ =\;& c_{\epsilon;rs}(\omega_j) + iq_{\epsilon;rs}(\omega_j) - iq_{\epsilon;rs}(\omega_j) - c_{\epsilon;rs}(\omega_j) = 0. \end{aligned} \tag{4.17}$$

Due to lemma 4.1. and the results of equations (4.16) and (4.17),

$$\alpha(j) + i\beta(j) \xrightarrow{d} \mathcal{N}_c\left(0, 2\pi|a_s(j)|^2 h_{\epsilon;rr}(\omega_j) + 2\pi|a_r(j)|^2 h_{\epsilon;ss}(\omega_j)\right),$$

where $\alpha(j) + i\beta(j)$ is defined in (4.4). The remaining parts (4.7) and (4.8) consist of terms of the form

$$\text{cov}\left(I_{\epsilon;rs}(\omega_j), B_r(\omega_j)\right) = \sum_{t,u,v} \exp(-i\omega_j(t-u+v))E(\epsilon_r(t)\epsilon_s(u)\epsilon_r(v)),$$

with

$$\begin{aligned} & E(\epsilon_r(t)\epsilon_s(u)\epsilon_r(v)) \\ = & \sum_{r_1,r_2,r_3=1}^{p} E(Z_{r_1}(1)Z_{r_2}(1)Z_{r_3}(1)) \sum_j A_{r,r_1}(j)A_{s,r_2}(j+u-t)A_{r,r_3}(j+v-t), \end{aligned}$$

where $A_{rr_1}(j)$ is the (r, r_1)th entry of the matrix $\mathbf{A}(j)$ and $E(Z_{r_1}(t)Z_{r_2}(t)Z_{r_3}(t))$ is the third order cumulant between the components $Z_{r_1}(t)$, $Z_{r_2}(t)$ and $Z_{r_3}(t)$ of $\mathbf{Z}(t)$. Because $\boldsymbol{\epsilon}(i)$ is of the form (4.14),

$$\left|\text{cov}\left(I_{\epsilon;rs}(\omega_j), B_r(\omega_j)\right)\right| \leq \sum_{t,u,v} |E(\epsilon_r(t)\epsilon_s(u)\epsilon_r(v))| < \infty,$$

such that $\tau_n(\omega_j) = O(n^{-1})$. ■

Remark 4.5. *Remember that $h_{rs}(j) = a_r(j)\overline{a_s(j)}$. The asymptotic variance in (4.15) consists of a mixed spectrum depending on a product of the regression and the stationary spectrum. The influence of each part is determined by their relative contribution.*

4.2 Estimation of modulus, phase spectrum and lag

In contrast to traditional spectral analysis of stationary processes, the modified periodogram $\frac{1}{n}I_{rs}(\omega_j)$ has proved to be consistent for $h_{rs}(j)$. We may

therefore consider estimation of the modulus and phase spectrum based on this estimator. Furthermore, the asymptotic distribution of the slope of the phase spectrum will be given. As mentioned earlier, this quantity constitutes an estimator for the lag between two deterministic functions. The essential result will be given in corollary 4.3. where the asymptotic distribution of the lag estimate $\hat{\Delta}$ is presented. The result of the next remark is a simple generalization of the asymptotic covariance in theorem 4.2. and will be needed in the subsequent considerations.

Remark 4.6. *Let the assumptions of theorem 4.2. hold. Then, for pairs (r,s) and (r',s'), $1 \leq r,s,r',s' \leq p$, and Fourier frequencies ω_j, $\omega_{j'}$,*

$$\begin{aligned} &\lim_{n\to\infty} n cov(n^{-1}I_{rs}(\omega_j), n^{-1}I_{r's'}(\omega_{j'})) \\ = &\begin{cases} O(n^{-2}), & j \neq j', \\ 2\pi(\overline{h_{ss'}(j)}h_{\epsilon;rr'}(\omega_j) + \overline{h_{rr'}(j)}h_{\epsilon;ss'}(\omega_j)) + O(n^{-1}), & j = j'. \end{cases} \end{aligned}$$

Let $h_{rs}(j) = a_r(j)\overline{a_s(j)}$ be estimated by $\frac{1}{n}I_{rs}(\omega_j)$ and denote the real and imaginary parts by $c_{rs}(j) = \mathfrak{Re}\{h_{rs}(j)\}$ and $q_{rs}(j) = -\mathfrak{Im}\{h_{rs}(j)\}$ respectively. Then,

$$h_{rs}(j) = c_{rs}(j) - iq_{rs}(j),$$

and the joint asymptotic distribution of the corresponding estimators is derived in

Lemma 4.2. *Define the estimates*

$$\begin{aligned} \tilde{c}_{rs}(j) &= \frac{1}{2n}(I_{rs}(\omega_j) + \overline{I_{rs}(\omega_j)}) \quad \textit{and} \\ \tilde{q}_{rs}(j) &= -\frac{1}{2ni}(I_{rs}(\omega_j) - \overline{I_{rs}(\omega_j)}). \end{aligned}$$

Then,

$$\zeta_n = \sqrt{n}(\tilde{c}_{rs}(j)) - c_{rs}(j), \tilde{q}_{rs}(j)) - q_{rs}(j))^T$$

converges in distribution to a bivariate normal random variable with mean $\mathbf{0}$. *The asymptotic variances are given by*

$$\begin{aligned}
\Sigma_{cc}(j) &= \lim_{n\to\infty} n var(\tilde{c}_{rs}(j)) \\
&= \pi\left(h_{ss}(j)h_{\epsilon;rr}(\omega_j) + h_{\epsilon;ss}(\omega_j)h_{rr}(j)\right) + 2\pi\left(c_{rs}(j)c_{\epsilon;rs}(\omega_j) - q_{rs}(j)q_{\epsilon;rs}(\omega_j)\right) \\
\Sigma_{qq}(j) &= \lim_{n\to\infty} n var(\tilde{q}_{rs}(j)) \\
&= \pi\left(h_{ss}(j)h_{\epsilon;rr}(\omega_j) + h_{\epsilon;ss}(\omega_j)h_{rr}(j)\right) - 2\pi\left(c_{rs}(j)c_{\epsilon;rs}(\omega_j) + q_{rs}(j)q_{\epsilon;rs}(\omega_j)\right)
\end{aligned}$$

and

$$\begin{aligned}
\Sigma_{cq}(j) &= \lim_{n\to\infty} n cov(\tilde{c}_{rs}(j), \tilde{q}_{rs}(j)) \\
&= 2\pi\left(q_{rs}(j)c_{\epsilon;rs}(\omega_j) + c_{rs}(j)q_{\epsilon;rs}(\omega_j)\right).
\end{aligned}$$

Proof: Note that $I_{rs}(\omega_j) = \overline{I_{sr}(\omega_j)}$. Then, straightforward calculations and the results of theorem 4.2. and remark 4.6. give

$$\begin{aligned}
\lim_{n\to\infty} n\mathrm{var}\left(\tilde{c}_{rs}(j)\right) &= \frac{1}{2n^2}\lim_{n\to\infty}\mathrm{var}\left(I_{rs}(\omega_j)\right) + \frac{1}{2n^2}\lim_{n\to\infty}\mathrm{cov}\left(I_{rs}(\omega_j), I_{sr}(\omega_j)\right) \\
&= \pi\left(h_{ss}(j)h_{\epsilon;rr}(\omega_j) + h_{rr}(j)h_{\epsilon;ss}(\omega_j)\right) + 2\pi\mathfrak{Re}\{h_{rs}(j)h_{\epsilon;rs}(\omega_j)\},
\end{aligned}$$

and analogously,

$$\lim_{n\to\infty} n\mathrm{var}\left(\tilde{q}_{rs}(j)\right) = \pi\left(h_{ss}(j)h_{\epsilon;rr}(\omega_j) + h_{rr}(j)h_{\epsilon;ss}(\omega_j)\right) - 2\pi\mathfrak{Re}\{h_{rs}(j)h_{\epsilon;rs}(\omega_j)\},$$

where

$$\mathfrak{Re}\{h_{rs}(j)h_{\epsilon;rs}(\omega_j)\} = c_{rs}(j)c_{\epsilon;rs}(\omega_j) - q_{rs}(j)q_{\epsilon;rs}(\omega_j).$$

Applying the same results gives

$$\begin{aligned}
&\lim_{n\to\infty} n\mathrm{cov}\left(\tilde{c}_{rs}(j), \tilde{q}_{rs}(j)\right) \\
&= n\lim_{n\to\infty}\mathrm{cov}\left(\frac{1}{2n^2}(I_{rs}(\omega_j) + I_{sr}(\omega_j)), -\frac{1}{2n^2 i}(I_{rs}(\omega_j) - I_{sr}(\omega_j))\right) \\
&= -2\pi\mathfrak{Im}(h_{rs}(j)h_{\epsilon;rs}(\omega_j)) = 2\pi(q_{rs}(j)c_{\epsilon;rs}(\omega_j) + c_{rs}(j)q_{\epsilon;rs}(\omega_j)).
\end{aligned}$$

■

Based on the results of lemma 4.2. we may now construct consistent estimates

of the spectral modulus and phase shift. Write $h_{rs}(j) = \kappa^*_{rs}(j)\exp(i\phi_{rs}(j))$ and estimate the spectral modulus and phase shift by

$$\tilde{\kappa}^*_{rs}(j) = (\tilde{c}^2_{rs}(j) + \tilde{q}^2_{rs}(j))^{1/2} \text{ and} \tag{4.18}$$

$$\tilde{\phi}_{rs}(j) = \arg I_{rs}(\omega_j) = \arctan\left(-\frac{\tilde{q}_{rs}(j)}{\tilde{c}_{rs}(j)}\right). \tag{4.19}$$

The next corollaries develop the asymptotic distribution of the estimators in (4.18) and (6.2). The results can be gained with similar techniques as in chapter 3.

Corollary 4.1. *Let $\kappa^*_{rs}(j) = |h_{rs}(j)|$ and $\tilde{\kappa}^*_{rs}(j)$ be given as above. Then*

$$\tilde{\kappa}_{rs}(j) - \kappa_{rs}(j) = O_p(n^{-1/2})$$

uniformly for all $j \in \mathbb{Z}$. Furthermore,

$$\sqrt{n}(\tilde{\kappa}_{rs}(j) - \kappa_{rs}(j)) \xrightarrow{d} \mathcal{N}(0, \sigma^2_{\tilde{\kappa}}(j)), \tag{4.20}$$

where

$$\sigma^2_{\tilde{\kappa}}(j) = \kappa^{*-2}_{rs}(j)\left(c^2_{rs}(j)\Sigma_{cc}(j) + q^2_{rs}(j)\Sigma_{qq}(j) + 2c_{rs}(j)q_{rs}(j)\Sigma_{cq}(j)\right),$$

and

$$\lim_{n\to\infty} n cov(\tilde{\kappa}^*_{rs}(j), \tilde{\kappa}^*_{rs}(j')) = O(n^{-2}),\ j \neq j'.$$

Corollary 4.2. *Let $\phi_{rs}(j) = \arg h_{rs}(j)$ be estimated by $\tilde{\phi}_{rs}(j)$. Then,*

$$\tilde{\phi}_{rs}(j) - \phi_{rs}(j) = O_p(n^{-1/2}) \tag{4.21}$$

uniformly for all $j \in \mathbb{Z}$. Furthermore,

$$\sqrt{n}(\tilde{\phi}_{rs}(j) - \phi_{rs}(j)) \xrightarrow{d} \mathcal{N}(0, \sigma^2_{\tilde{\phi}}(j)), \tag{4.22}$$

where

$$\sigma^2_{\tilde{\phi}}(j) = \kappa^{*-4}_{rs}(j)\left(q^2_{rs}(j)\Sigma_{cc}(j) + c^2_{rs}(j)\Sigma_{qq}(j) - 2c_{rs}(j)q_{rs}(j)\Sigma_{cq}(j)\right),$$

and

$$\lim_{n\to\infty} n cov(\tilde{\phi}_{rs}(j), \tilde{\phi}_{rs}(j')) = O(n^{-2}),\ j \neq j'.$$

We note again, that the variance of the phase spectrum essentially depends on the modulus of the regression spectrum, i.e. the variance is small whenever $\kappa^*_{rs}(j)$ is large and vice versa. In chapter 3 an algorithm for estimating the slope of the phase line was based on an appropriate average over a set of frequencies with large amplitude spectrum. A similar approach will now be presented in corollary 4.3. Assume therefore, that for some r, s $(1 \leq r, s \leq p)$

$$f_r(t) = cf_s(t + \Delta).$$

such that

$$\phi_{rs}(j) = -2\pi j\Delta$$

(see example 3.1. in chapter 3).

Corollary 4.3. *Given an (with respect to possible 2π-jumps) appropriate set of frequencies $J = (j_1, \ldots, j_{k+1})$, $k \in \mathbb{Z}$, the lag Δ may be estimated by*

$$\hat{\Delta} = \frac{1}{k}\sum_{i=1}^{k} \frac{\hat{\phi}_{rs}(j_{i+1}) - \hat{\phi}_{rs}(j_i)}{j_{i+1} - j_i}, \tag{4.23}$$

i.e. $\hat{\Delta}$ is an average of the slopes over subsequent frequencies in J. Results of corollary 4.2. imply

$$\hat{\Delta} - \Delta = O_p(n^{-1/2}).$$

Furthermore, define

$$a = \prod_{h=1}^{k}(j_{h+1} - j_h), \quad a_i = \prod_{h=1, h\neq i}^{k}(j_{h+1} - j_h).$$

Then

$$\sqrt{n}(\hat{\Delta} - \Delta) \xrightarrow{d} \mathcal{N}(0, \sigma^2_{\hat{\Delta}}),$$

where

$$\sigma^2_{\hat{\Delta}} = \frac{1}{(ka)^2}\left(a_1^2 var(\tilde{\phi}_{rs}(j_1)) + \sum_{i=1}^{k-1}(a_i - a_{i+1})^2 var(\tilde{\phi}_{rs}(j_{i+1})) + a_k^2 var(\tilde{\phi}_{rs}(j_k))\right).$$

Remark 4.7. *A particular disadvantage of $\hat{\Delta}$ in (4.23) is that it may involve values of the phase spectrum with large variance. Estimation can be improved by preliminary estimation of an appropriate set J.*

Remark 4.8. *The estimate $\hat{\Delta}$ could have been alternatively gained by linear regression. Asymptotic results for the coefficients in a linear regression model are then easily adopted to the current situation such that the distribution of $\hat{\Delta}$ follows directly.*

4.3 Comparing periodogram and shrinkage estimates

The previous results have revealed that the adjusted periodogram as well as the procedure involving wavelet thresholding both yield consistent estimates. Denote them by $\hat{h}_{rs}^{(1)}(j)$ and $\hat{h}_{rs}^{(2)}(j)$, where

$$\begin{aligned} \hat{h}_{rs}^{(1)}(j) &= \frac{1}{n} I_{rs}(\omega_j), \\ \hat{h}_{rs}^{(2)}(j) &= \hat{a}_r(j)\overline{\hat{a}_s(j)}. \end{aligned}$$

Let $b(\cdot)$ denote the bias and $MSE(\cdot)$ the mean squared error of an estimator. Because the MSE can be decomposed into a bias and a variance contribution,

$$MSE(\hat{h}_{rs}^{(l)}(j)) = \operatorname{var}(\hat{h}_{rs}^{(l)}(j)) + b(\hat{h}_{rs}^{(l)}(j))^2,\ l = 1, 2. \tag{4.24}$$

As a result of the last theorem, the bias, variance and mean squared error of $\hat{h}_{rs}^{(1)}(j)$ are given in

Corollary 4.4. *Assume again that model (4.1) where $\boldsymbol{\epsilon}(i)$ is generated by a generalized linear process of the form (4.14). Then, for each Fourier frequency $\omega_j = \frac{2\pi j}{n}$, $\omega_j \in [0, \pi]$,*

1. $b(\hat{h}_{rs}^{(1)}(j)) = E(\frac{1}{n}I_{rs}(\omega_j) - h_{rs}(j))$
   $= \frac{1}{n}\sum_{|k|<n}\left(1 - \frac{|k|}{n}\gamma_{\epsilon;rs}(k)\exp(-i\omega_j k)\right) + O(n^{-1}),$

2. $\lim_{n\to\infty} nvar(\hat{h}_{rs}^{(1)}(j)) = 2\pi h_{\epsilon;rr}(\omega_j)h_{ss}(j) + 2\pi h_{\epsilon;ss}(j)h_{rr}(j) + O(n^{-1}).$

3. $\lim_{n\to\infty} nMSE(\hat{h}_{rs}^{(1)}(j))$
   $= 2\pi h_{\epsilon;rr}(\omega_j)h_{ss}(j) + 2\pi h_{\epsilon;ss}(\omega_j)h_{rr}(j)$
   $+ \frac{1}{n}[\sum_{|k|<n}\left(1 - \frac{|k|}{n}\gamma_{\epsilon;rs}(k)\exp(-i\omega k)\right)]^2 + O(n^{-1}).$

Proof: Statement 2. is already proved in theorem 4.2. and (4.24) together with 1. imply 3. With the usual notations,

$$\begin{aligned} E(\hat{h}_{rs}^{(1)}(j)) &= n^{-1}[I_{f;rs}(\omega_j) + E(I_{\epsilon;rs}(\omega_j))] \\ &= h_{rs}(j) + O(n^{-1}) + \frac{1}{n}E(I_\epsilon(\omega_j)), \end{aligned}$$

where $E(I_\epsilon(\omega_j)) = \sum_{|k|<n} n^{-1}\sum_{t=1}^{n-|k|}\left(1 - \frac{|k|}{n}\right)E(\epsilon_r(t)\epsilon_s(t+|k|)\exp(-i\omega_j k)$ such that the result is straightforward. ■

We now compare this estimator with $\hat{h}_{rs}^{(2)}(j)$. First, note that the bias of $\hat{h}_{rs}^{(2)}(j)$ essentially depends on the bias of the wavelet thresholding estimates $\hat{f}_r(\cdot)$ and $\hat{f}_s(\cdot)$, $1 \le r, s \le p$, where the relevant results are investigated in chapter 2. For small sample sizes n such that $J_n < J_0$ the estimate based on wavelet thresholding includes a bias term which vanishes very slowly. In the wavelet case, n should be large such that $\hat{f}_r$ becomes unbiased. In the following we restrict our considerations to this case. Refer to the second estimator by

$$\hat{h}_{rs}^{(2)}(j) = \frac{1}{n^2}\sum_{k=1}^{n}\hat{f}_r(t_k)\exp(-i2\omega_j k)\sum_{l=1}^{n}\hat{f}_s(t_l)\exp(i2\omega_j l)$$

and neglect the error of approximating integrals by finite sums. The comparison of the bias of the two estimators leads us to compare

$$\sum_{k,l} E[Y_r(k)Y_s(l)] \exp(-i\omega_j(k-l)) \text{ with} \tag{4.25}$$

$$\sum_{k,l} E[\hat{f}_r(t_k)\hat{f}_s(t_l)] \exp(-i\omega_j(k-l)). \tag{4.26}$$

The bias of the wavelet estimate is relatively small compared to (4.25) whenever the bias of $\hat{f}_r(\cdot)\hat{f}_s(\cdot)$ is small. First, we have

$$E[Y_r(k)Y_s(l)] = f_r(t_k)f_s(t_l) + \underbrace{E[\epsilon_r(k)\epsilon_s(l)]}_{=\gamma_{\epsilon;rs}(k-l)}. \tag{4.27}$$

Second, similar considerations as in lemma 2.7. show

$$E[\hat{f}_r(t_k)\hat{f}_s(t_l)] = f_r(t_k)f_s(t_l) + O(n^{-1}), \tag{4.28}$$

for sufficiently large n such that $J_n \geq J_0$. Comparing (4.27) with (4.28) shows that asymptotically the difference in bias between $\hat{h}_{rs}^{(1)}(j)$ and $\hat{h}_{rs}^{(2)}(j)$ is given by a factor of order $O(n^{-1})$.

Remark 4.9. *For finite samples n such that $J_n < J_0$, the bias in wavelet thresholding might exceed the bias of the periodogram. A quantitative comparison based on the bias of both estimates requires appreciation of the particular situation. General results cannot be given.*

Consider now the variances of both estimators. The variance of $\hat{h}_{rs}^{(1)}(j)$ is

$$\lim_{n\to\infty} n\text{var}\,(\hat{h}_{rs}^{(1)}(j)) = 2\pi h_{\epsilon;rr}(\omega_j)h_{ss}(j) + 2\pi h_{\epsilon;ss}(\omega_j)h_{rr}(j) + O(n^{-1}).$$

Results from spectral analysis in chapter 3 showed that

$$\begin{aligned}
&\lim_{n\to\infty} n\mathrm{var}\,(\hat{h}_{rs}^{(2)}(j)) \\
= \;& 2\pi h_{\epsilon;ss}(0)|a_r(j)|^2\left[\sum_k |a_{l,k}^{\phi}(j)|^2 + \sum_{j,k} w_{j,k}^{(s)}|a_{j,k}^{\psi}(j)|^2\right] \\
& +2\pi h_{\epsilon;rr}(0)|a_s(j)|^2\left[\sum_k |a_{l,k}^{\phi}(j)|^2 + \sum_{j,k} w_{j,k}^{(r)}|a_{j,k}^{\psi}(j)|^2\right] \\
& +4\pi\mathfrak{Re}\left\{h_{\epsilon;rs}(0)\overline{a_r(j)a_s(j)}\left(\sum_k a_{l,k}^{\phi}(j)a_{l,k}^{\phi}(j) + \sum_{j,k} w_{j,k}^{(r)}w_{j,k}^{(s)}a_{j,k}^{\psi}(j)a_{j,k}^{\psi}(j)\right)\right\}.
\end{aligned}$$

A general result about the efficiency of $\hat{h}_{rs}^{(1)}(j)$ relative to $\hat{h}_{rs}^{(2)}(j)$ is impossible. However, the case of independent errors is presented in

Corollary 4.5. *Assume again that $\boldsymbol{\epsilon}(i)$ is independent and identically distributed with mean 0 and covariance matrix $\sigma^2\mathbf{I}_p$, where $\mathbf{I}_p$ denotes the p-dimensional identity matrix. Then,*

$$h_{\epsilon;rs}(0) = 0, \;\; h_{\epsilon;rr}(\omega_j) = h_{\epsilon;rr}(\omega_j) = \frac{\sigma^2}{2\pi},$$

and the asymptotic variance of $\hat{h}_{rs}^{(1)}(j)$ reduces to

$$\lim_{n\to\infty} nvar(\hat{h}_{rs}^{(1)}(j)) = \sigma^2\left(|a_r(j)|^2 + |a_s(j)|^2\right).$$

Notice that for each Fourier coefficient of a wavelet function we have

$$\begin{aligned}
|a_{l,k}^{\phi}(j)|^2 &= |\int \phi_{l,k}(t)\exp(-i2\pi jt)dt|^2 \\
&\leq \int \phi_{l,k}^2(t)dt = 1
\end{aligned} \tag{4.29}$$

such that

$$\sum_k |a_{l,k}^{\phi}(j)|^2 + \sum_k |a_{j,k}^{\psi}(j)|^2 \leq CJ_0, \tag{4.30}$$

where C is a suitable constant. An upper bound for the second variance is then given by

$$\lim_{n\to\infty} var(\hat{h}_{rs}^{(2)}(j)) \leq CJ_0\sigma^2\left(|a_r(j)|^2 + |a_s(j)|^2\right),$$

such that the variance of both estimators are of equal order. where the difference depends on the individual wavelet basis.

Remark 4.10. *For dependent $\boldsymbol{\epsilon}(i)$, a general statement about the relative efficiency is difficult. The result depends on the spectral density and the particular choice of wavelet basis.*

Note that the constant C in (4.30) depends on the individual wavelet basis. It is beyond the scope of this thesis to analyze the relative efficiency and bias of the two estimators above in more detail. Results may differ depending on the regularity of the unknown $\mathbf{f}$, the structure of $\boldsymbol{\epsilon}(i)$, the wavelet basis and the threshold λ. Future research will analyze these issues in more detail.

Chapter 5

Lag Estimation in Time Domain

5.1 Introduction

In the previous chapters we considered estimation of the regression cross spectrum and derived a method for estimating lead-lag effects in multivariate time series. The developed algorithm in chapter 3 is based on the presence of a set of significant common frequencies that can be used for estimating the slope of the phase line. Problems with this algorithm are expected if this set is too small, e.g. if the deterministic components oscillate with a single common frequency. These conditions are frequenctly met in many climatic of physilogical data sets. An alternative way of estimating time delays that overcomes this problem is based on the maximum of the regression cross correlation. The corresponding maximizing lag will constitute an estimate for the time delay between deterministic trend components. In the following a consistent estimate is defined and the asymptotic properties are derived. We conclude this section with a small simulation study showing the practical performance of the theoretical results. For other references on estimating the argmax of an unknown function refer to Beran and Gosh (2000).

5.2 The maximizing lag u_{rs}^{max}

Assume again that the model

$$\mathbf{Y}(i) = \mathbf{f}(t_i) + \boldsymbol{\epsilon}(i)$$

holds, where $\boldsymbol{\epsilon}(i)$ is stationary with mean $\mathbf{0}$ and let $\mathbf{f}$ be estimated by wavelet thresholding. Remember also the definition of the regression cross covariance

$$\gamma_{rs}(u) = \int_0^1 f_r(t+u)\overline{f_s(t)}dt$$

with $\hat{\gamma}_{rs}(u)$ defined accordingly. For each pair $(r,s),\ 1 \leq r,s \leq p$, we denote the set of local maxima by

$$\mathcal{M} = \{u \in [-1,1] : \gamma'_{rs}(u) = 0,\ \gamma''_{rs}(u) < 0\} \tag{5.1}$$

and

$$u_{rs}^{max} = \text{argmax}\{\gamma_{rs}(u) : u \in \mathcal{M}\}. \tag{5.2}$$

Denote $\boldsymbol{\alpha}^{(r)} = \{\alpha_{l,k}^{(r)} : l,k \in \mathbb{Z}\}$ and $\boldsymbol{\beta}^{(r)} = \{\beta_{j,k}^{(r)} : j > l,k \in \mathbb{Z}\}$ for $r = 1,\ldots,p$. Then, given the representation of $f_r(t)$ and $f_s(t)$ relatively to a particular wavelet basis, $\gamma_{rs}(u)$ depends continously on $\boldsymbol{\alpha}^{(r)}$ and $\boldsymbol{\beta}^{(r)}$. Write $\gamma_{rs}(u) = \gamma_{rs}(u,\theta_0)$ with $\theta_0 = (\boldsymbol{\alpha}^{(r)}, \boldsymbol{\beta}^{(r)}, \boldsymbol{\alpha}^{(s)}, \boldsymbol{\beta}^{(s)})$ and

$$\hat{\gamma}_{rs}(u) = \gamma_{rs}(u,\hat{\theta}).$$

Then,

$$\mathcal{M} = \mathcal{M}(\theta_0) = \{u \in [-1,1] : \gamma'_{rs}(u,\theta_0) = 0,\ \gamma''_{rs}(u,\theta_0) < 0\} \tag{5.3}$$

and

$$u_{rs}^{max} = \text{argmax}\{\gamma_{rs}(u,\theta_0) : u \in \mathcal{M}(\theta_0)\}. \tag{5.4}$$

Definition 5.1. *We will say that f_r and f_s are associated at a lag u if the regression covariance assumes its global maximum in u.*

5.3 Estimation of u_{rs}^{max}

Assume that θ has already been estimated by wavelet thresholding. Then, the lag that maximizes $\gamma_{rs}(u, \hat{\theta})$ may be estimated by

$$\hat{u}_{rs}^{max} = \text{argmax}\{\gamma_{rs}(u, \hat{\theta}) : u \in \mathcal{M}(\hat{\theta})\}. \tag{5.5}$$

To ensure existence, uniqueness and consistency of estimation we have to formulate the following assumptions.

(U1) $\gamma_{rs}(u, \theta)$ is twice continously differentiable with respect to u and θ.

(U2) The covariance $\gamma_{rs}(u, \theta_0)$ has a unique maximum at u_{rs}^{max}.

Asymptotic properties of the estimated maximizing lag are derived in

Theorem 5.1. *Let the assumption (C) together with (U1) and (U2) hold. Then,*

a) $\hat{u}_{rs}^{max} - u_{rs}^{max} = O_p(n^{-1/2})$ *(*$1 \leq r, s \leq p$*),*

b) $\sqrt{n}(\hat{u}_{rs}^{max} - u_{rs}^{max}) \xrightarrow{d} \mathcal{N}(0, \tau_{u,rs})$*,*

with

$$\tau_{u,rs}(\theta_0) = \frac{1}{(\gamma''_{rs}(u_{rs}^{max}, \theta_0))^2} \left(\frac{\partial}{\partial \theta} \gamma'_{rs}(u_{rs}^{max}, \theta_0) \right)^T var(\hat{\theta}) \left(\frac{\partial}{\partial \theta} \gamma'_{rs}(u_{rs}^{max}, \theta_0) \right).$$

Proof:

a) Earlier results proved wavelet thresholding estimates to be consistent of order $O_p(n^{-1/2})$. The regression cross covariance depends continously on θ, the vector of wavelet coefficients of f_r and f_s, such that the estimate $\hat{u}_{rs}^{max}$ continously depends onto $\hat{\theta}$. Using consistency of $\hat{\theta}$ proves consistency of u_{rs}^{max}.

b) Due to (5.4) and (5.5),

$$\gamma'_{rs}(\hat{u}^{max}_{rs}, \hat{\theta}) = 0 \text{ and } \gamma'_{rs}(u^{max}_{rs}, \theta_0) = 0.$$

Applying the mean value theorem gives

$$\gamma'_{rs}(\hat{u}^{max}_{rs}, \hat{\theta}) - \gamma'_{rs}(u^{max}_{rs}, \hat{\theta}) = \gamma''_{rs}(\tilde{u}_{rs}, \hat{\theta})(\hat{u}^{max}_{rs} - u^{max}_{rs}),$$

where $\tilde{u}$ lies on a line connecting $\hat{u}^{max}_{rs}$ and u^{max}_{rs}. Therefore,

$$\hat{u}^{max}_{rs} - u^{max}_{rs} = -\frac{\gamma'_{rs}(u^{max}_{rs}, \hat{\theta})}{\gamma''_{rs}(\tilde{u}_{rs}, \hat{\theta})}.$$

Furthermore,

$$\gamma'_{rs}(u^{max}_{rs}, \hat{\theta}) = \underbrace{\gamma'_{rs}(u^{max}_{rs}, \theta_0)}_{=0} + \left(\frac{\partial}{\partial\theta} \gamma'_{rs}(u^{max}_{rs}, \theta)\Big|_{\theta=\tilde{\theta}} \right)^T \cdot (\hat{\theta} - \theta_0)$$

for suitably chosen $\tilde{\theta}$. Together,

$$\hat{u}^{max}_{rs} - u^{max}_{rs} = -\frac{1}{\gamma''_{rs}(\tilde{u}_{rs}, \hat{\theta})} \left(\frac{\partial}{\partial\theta} \gamma'_{rs}(u^{max}_{rs}, \theta)\Big|_{\theta=\tilde{\theta}} \right)^T (\hat{\theta} - \theta_0).$$

Due to consistency of $\hat{\theta}$ and $\hat{u}^{max}_{rs}$,

$$\begin{aligned} \gamma''_{rs}(\tilde{u}_{rs}, \hat{\theta}) &= \gamma''_{rs}(u^{max}_{rs}, \theta_0) + (\gamma''_{rs}(\tilde{u}_{rs}, \hat{\theta}) - \gamma''_{rs}(u^{max}_{rs}, \hat{\theta})) \\ &\quad + (\gamma''_{rs}(u^{max}_{rs}, \hat{\theta}) - \gamma''_{rs}(u^{max}_{rs}, \theta_0)) \\ &= \gamma''_{rs}(u^{max}_{rs}, \theta_0) + O_p(n^{-1/2}), \end{aligned}$$

and

$$\frac{\partial}{\partial\theta} \gamma'_{rs}(u^{max}_{rs}, \theta)\Big|_{\theta=\tilde{\theta}} = \frac{\partial}{\partial\theta} \gamma'_{rs}(u^{max}_{rs}, \theta)\Big|_{\theta=\theta_0} + O_p(n^{-1/2}).$$

Therefore,

$$\hat{u}^{max}_{rs} - u^{max}_{rs} = -\frac{1}{\gamma''_{rs}(u^{max}_{rs}, \theta_0)} \left(\frac{\partial}{\partial\theta} \gamma'_{rs}(u^{max}_{rs}, \theta_0) \right)^T (\hat{\theta} - \theta_0)(1 + O_p(n^{-1/2})).$$

The vector $\sqrt{n}(\hat{\theta} - \theta_0)$ proved to be asymptotically normal. Hence, $\sqrt{n}(\hat{u}^{max}_{rs} - u^{max}_{rs})$ is asymptotically normal with mean 0 and covariance

$$\text{var}\,(\hat{u}^{max}_{rs}) = \frac{\left(\frac{\partial}{\partial\theta} \gamma'_{rs}(u^{max}_{rs}, \theta)\big|_{\theta=\theta_0} \right)^T \text{var}\,(\hat{\theta}) \left(\frac{\partial}{\partial\theta} \gamma'_{rs}(u^{max}_{rs}, \theta)\big|_{\theta=\theta_0} \right)}{\gamma''_{rs}(u^{max}_{rs}, \theta_0)^2}.$$

■

5.4 Simulated example

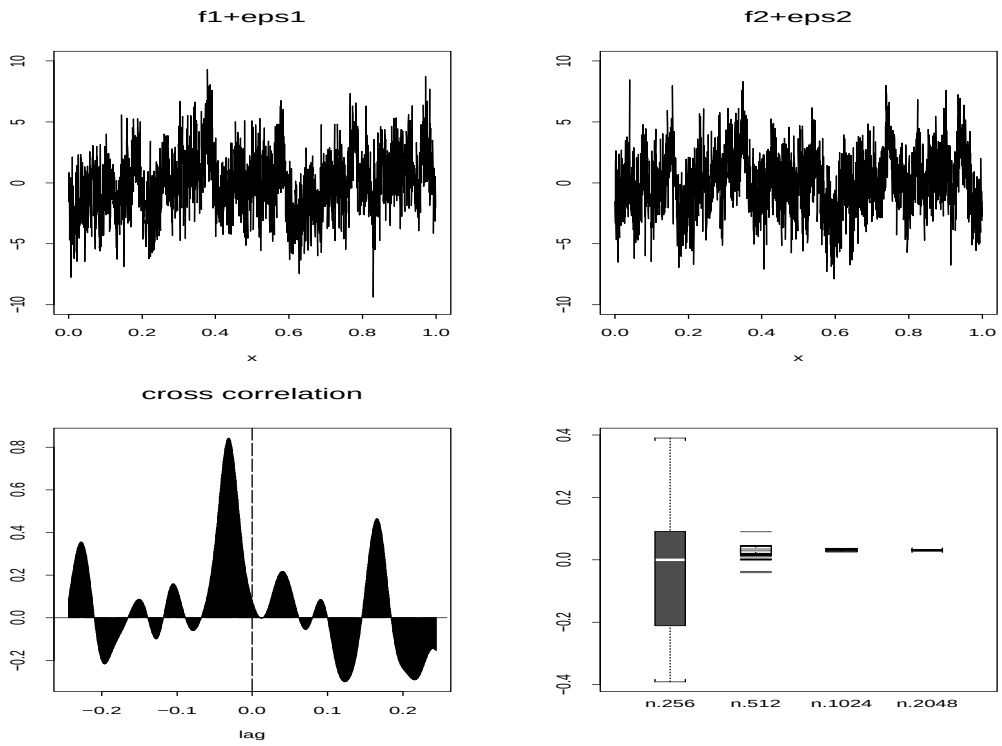

Figure 5.1: Simulated series for a sample size of $n = 2048$ together with regression cross correlation. Bottom right panel shows 500 simulations of maximizing lags for the sample sizes $n = 256, 512, 1024$ and $n = 2048$.

We consider again the example of chapter 3 with

$$f_1(x) = -\sin(4\pi x) - \sin(10\pi x) - \sin(20\pi x) - \sin(30\pi x) - \sin(51\pi x),$$

$f_2(x) = f_1(x + \Delta)$, and $\Delta = .03125$, where $\epsilon_1(i)$, $\epsilon_2(i)$ are iid $N(0, \sigma_\epsilon^2)$ with $\sigma_\epsilon^2 = 4$ and ϵ_1, ϵ_2 independent of each other. Figure 5.1 illustrates both series together with the estimated regression cross correlation between the

corresponding trend components for a sample size of $n = 2048$. In this particular case, the correlation reaches its maximum at $\hat{u}_{rs}^{max} = 0.03125$ which is equal to the true value. A small simulation study with the sample sizes $n = 256, 512, 1024$ and $n = 2048$ in accordance to chapter 3 is performed. In each case we run 500 simulations. Boxplots in the bottom right panel of figure 5.1 graphically display the results. For $n = 256$ only the mean seems to approximate the dominating lag with a large variation for individual estimates. Wavelet thresholding seems to be inappropriate for estimating trend components when the sample size is relatively small compared to the noise level. However, estimation quality improves when increasing the sample size. The detailed results are presented in table 5.1. In particular, note the proximity of

n	256	512	1024	2048
true value	-.03125	-.03125	-.03125	-.03125
median	0.00000	-0.03125	-0.03027	-0.03125
mean	0.04603	-0.03103	-0.03045	-0.03105
std.dev.	0.21586	0.00778	0.00165	0.00060
true std. dev.	0.00012	0.00012	0.00012	0.00012
mse	$5.25{\cdot}10^{-2}$	$6.04{\cdot}10^{-5}$	$3.37 \cdot 10^{-6}$	$4.01{\cdot}10^{-7}$
skewness	-0.13303	2.56569	-0.35268	0.11583

Table 5.1: Detailed summary of the estimation for various sample sizes, in each case running 500 simulations.

the mean and median to the true value Δ together with a rapidly decreasing mean squared error.

Chapter 6

Applications

In this chapter the theory developed in chapters 2 to 5 is applied to real data examples. We analyze four data sets from various fields of research including time series from biology, physiology, climatology and finance. The first example considers a bivariate time series relating warmth of the Pacific Ocean to the recruitment of new fish. Due to the presence of several common frequencies in the cross amplitude spectrum, lag estimation based on the phase line seems to be appropriate. The second example presents three physiological time series consisting of an electrocardiogram, blood pressure and stroke volume signal that are measured simultaneously. Estimation of time delays is again properly performed by usage of the phase spectrum. The third example presents rat electroencephalographic (EEG) signals revealing different structures of interdependencies that my occur in brain activity. Because the set of common frequencies is small, estimation in time domain is more adequate for detecting lag effects. We conclude this chapter with a financial example where a set of stocks traded at the New York Stock Exchange (NYSE) is investigated. The only common periodicity is a daily one representing trader's activity in the progress of a day. We will therefore use the regression cross correlation to analyze behavior of market participants and the resulting interplay between

stocks of companies of various sizes.

6.1 El Niño and recruitment of new fish

The following example investigates the interplay between sea warming and occurrence of new fish in the central Pacific Ocean. In this context we consider a bivariate time series consisting of the Southern Oscillation Index (SOI) and recruitment (amount) of new fish in a certain area. The SOI relates changes in air pressure to the temperature of the ocean at the surface. The data set can be found in Shumway and Stoffer (2000). Both time series range from 1950 to 1987 over a period of $n = 453$ months and are displayed in figures 6.1(a),(b).

At a first glance, both time series show cyclic behavior. While the SOI is dominated by a strong periodic component with a period of 12 months the second series oscillates with a lower frequency. However, one notices the 12-months cycles in some parts of this series as well. Another remarkable feature is the changing mean and amplitude of SOI over time. In literature this is refered to as the El Niño effect. The corresponding estimated trend components for both time series are displayed in figures 6.1(c),(d). Due to their smoothness and symmetry we applied s20-wavelets in the wavelet transformation. These seem to be ideal for modeling the cyclic behavior in both series. In the bottom row the residuals in each case are displayed (figures 6.1(e),(f)). Note that for better representation we changed the scale in the second panel.

The univariate regression amplitude spectra for both time series are displayed in figures 6.2(a),(b). The amplitude spectrum of SOI shows a dominating frequency at $j = 38$ indicating a period of 12 months combined with subsidiary contributions at low frequencies ($j = 2$). The second series is characterized by a superposition of several frequencies. On the one hand we notice the same 12-month cycle as in SOI. On the other hand, further large moduli can be

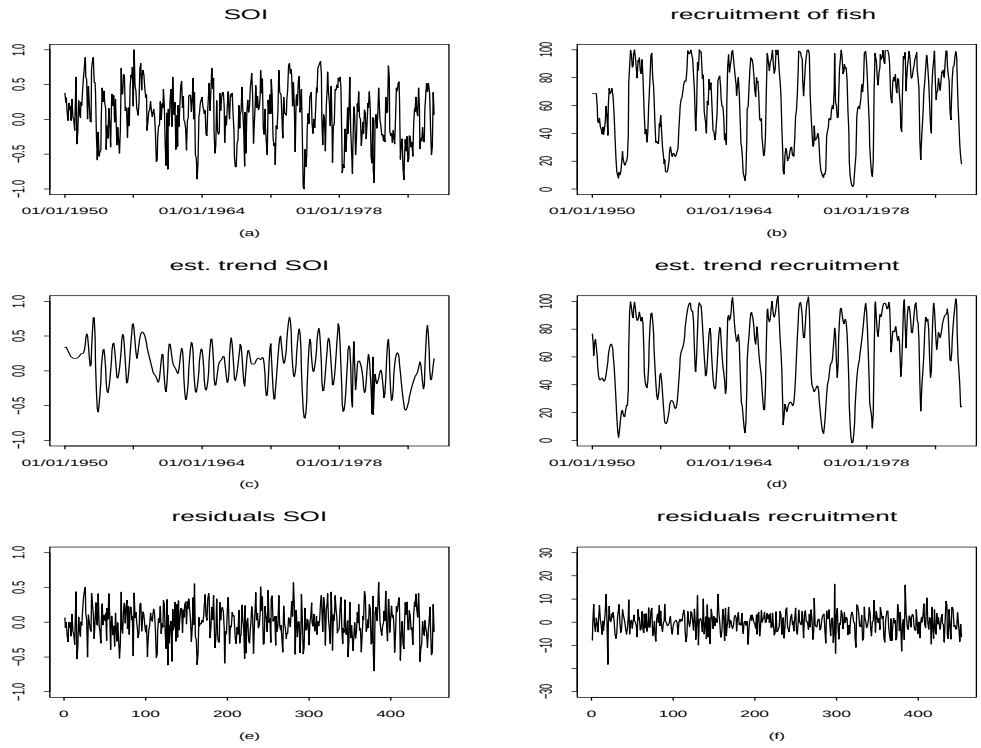

Figure 6.1: The SOI and numbers of new fish are displayed in figures (a) and (b) (1950-1987). Figures (c), (d) show the trend estimates followed by the corresponding residuals (e) and (f).

observed at lower frequencies with a dominating component at $j = 2$. These individual features now contribute to the particular shape of the modulus of regression cross spectrum (figure 6.2(c)), where we notice the common period of 12 months together with a set of important smaller frequencies. According to the results and the algorithm of chapter 3, the largest values in the amplitude spectrum are now chosen for estimation of the phase line in figure 6.2(d). The corresponding frequencies in the phase spectrum are marked by black squares. We notice a linear structure in the phase spectrum over this set of frequencies. The phase line is estimated by linear regression resulting in an estimated slope of around 8.5. This indicates that the SOI signal leads the recruitment of new fish by around eight to nine months. In other words, the variation in sea temperature indicates a change in the number of fish eight months later.

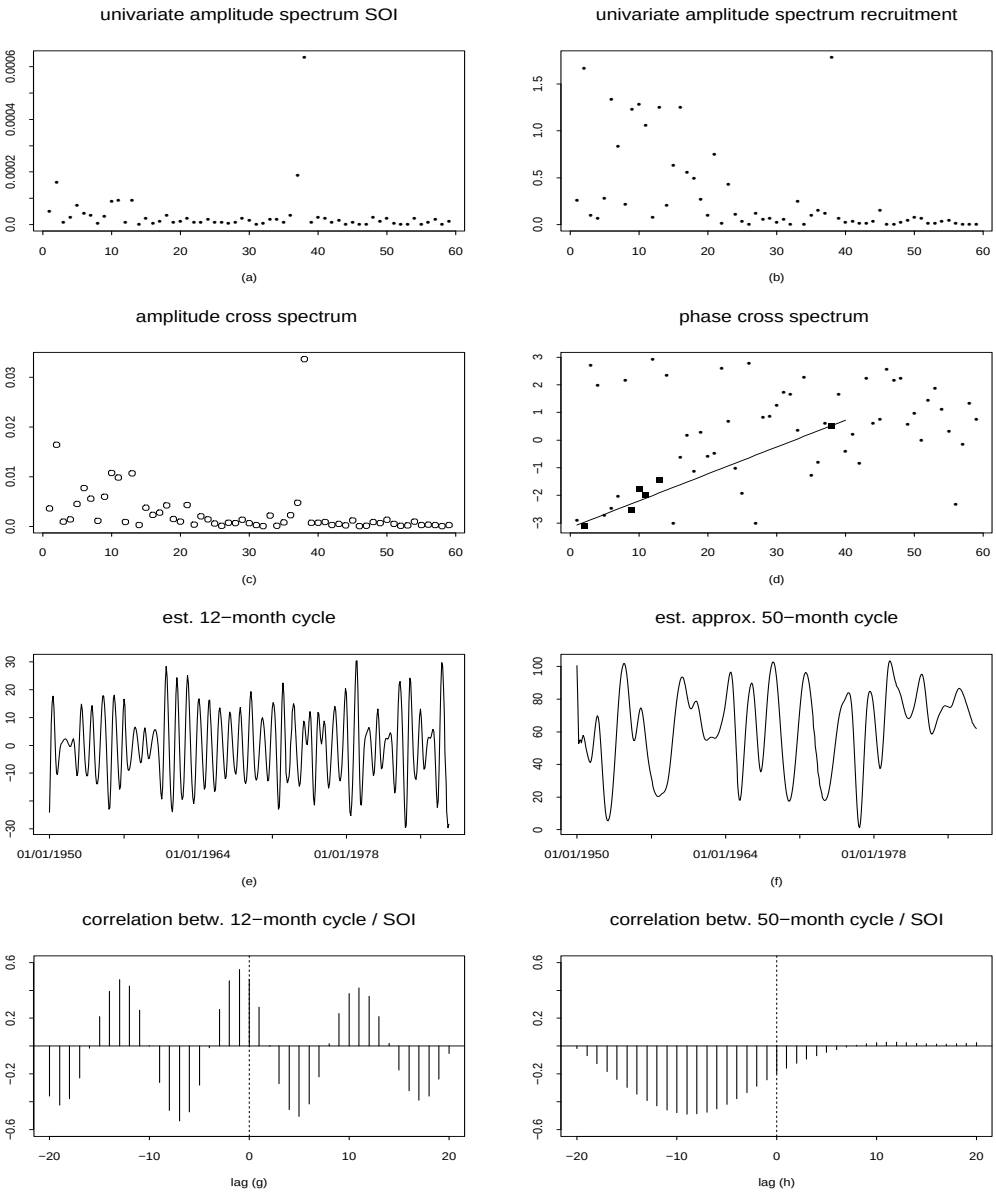

Figure 6.2: Univariate amplitude spectra are displayed in figures (a) and (b) followed by the cross spectrum in (c) and (d). Important frequencies of the amplitude spectrum are marked in the phase spectrum by black squares and an estimated phase line is drawn. Figures (e) and (f) present the decomposition of the trend in the recruitment of fish into a 12-month cycle and a long-term component. The autocovariance of each series with SOI is displayed in (g),(h).

One may initially argue that the trend component of the second series is too close to the true series. For further analysis we consider a decomposition of the estimated trend in the number of new fish in figures 6.2(e) and 6.2(f). The decomposition is the result of taking the three coarsest levels of the thresholding estimate and the fourth level separately. The first signal represents the 12-month cycle (fourth level). The coarser parts of the thresholding estimate are given in the second panel representing the four to five years cycles in warming of the Pacific Ocean. Because annual seasonality is the only common periodic oscillation when considering the first series, time delays are now estimated in time domain. The corresponding regression correlations between each part and SOI trend are now calculated and displayed in figures 6.2(g),(h). The annual seasonality of the number of fish reveals a lag over SOI of one month indicating that an increased water temperature induces an increased number of fish one month later. This effect interfers with the El Niño effect, where abnormal warming of the sea occurs every four to five years. Here, we notice a lead of nine months of the SOI indicating a negative influence of the El Niño effect on the number of fish in this particular area of the ocean.
The results about the interplay between water temperature and fish recruitment is confirmed by studies of many other authors such as Murawski (1993), Victor et al. (2001) or Shumway and Stoffer (2000).

6.2 ECG data

We consider a three-dimensional time series consisting of electrocardiogram (ECG), blood pressure (BP) and the cardiac stroke volume (SV) measurements of a sleeping patient, recorded at the Beth Israel Deaconess Medical Center in Boston (Goldberger et al. (2000)). The data are represented in units of 50mV (ECG), 50mmHG (BP) and 50ml (SV). The observational period consists of $n = 8192$ observations recorded at a rate of 200 observations

per second. A detailed description of the data set can be found in Ichimaru and Moody (1999). For better visibility, only the first 1000 observations

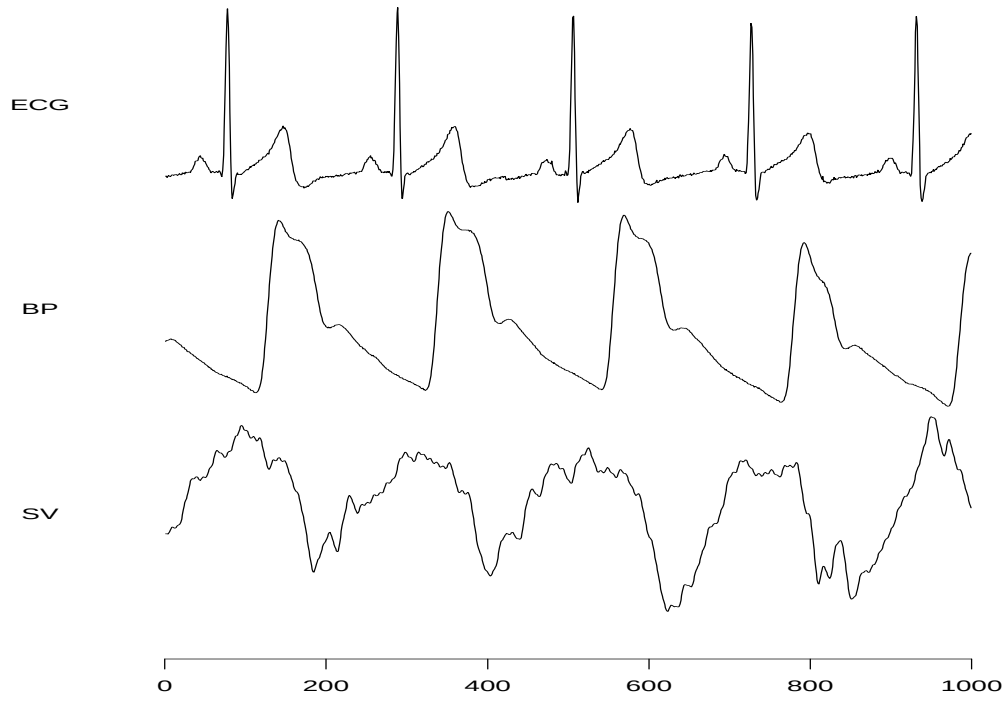

Figure 6.3: Electrocardiogram (ECG), blood pressure (BP) and cardiac stroke volume (SV) of a sleeping person over a period of 5 seconds (1000 observations).

(5 seconds) of the raw time series are displayed in figure 6.3. The analysis is carried out for all 8192 observations. As expected, all time series mainly consist of a deterministic almost periodic movement, with a period of about 200 observations, indicating a heart rate around 60 beats per minute. The ECG signal starts with a small bump indicating the atrial contraction and is followed by a sharp peak representing the contraction of the ventricles. The third peak represents the repolarization period. At the same time, contraction of the ventricles causes the stroke volume to increase up to a peak value, after which it falls back to its minimum shortly after the repolarization. The

growing amount of blood pushed into the arteries also induces a rise in blood pressure. The electrocardiogram receives information directly from the heart whereas the blood pressure is measured by a catheder in the radial artery. Therefore, the delay represents the time it takes the pressure wave, initialized by a heart beat, to reach the catheder. Consider now the regression cross spectra in figure 6.4. All three amplitude spectra reveal dominating common frequencies around $j = 40$ indicating a common period of length around $1s$ (figures 6.4(a),(c),(e)). Moreover, the phase spectrum of ECG and BP exhibits linear structures over certain ranges of frequencies (figure 6.4(b)). The phase spectrum between ECG and BP for the interval $j \in [34, 40]$ is drawn in figure 6.4(g). We see an almost perfectly linear structure in the phase spectrum and the estimated phase-shift is about 26 time units indicating a lead of BP against ECG. Note, however, that for the corresponding dominating lag the regression cross-correlation is negative (figure 6.4(h)). This effect can be noticed visually when comparing e.g. local maxima of the BP signal with minima of the ECG signal after the repolarization period (local maxima of the blood pressure occur shortly before the ECG signal reaches its minimum). Due to the physiological fact that the peaks in ECG representing contractions of the ventricals are constantly around 94 data points away from the minimum after the repolarization period, we estimate a lag between the contraction of the ventricals and the maximum of the blood pressure of around $94 - 26 = 68$ time units which is equivalent to a time delay of $340ms$. A similar analysis between ECG and SV results in a lead of the ECG signal of about 11 data points or $55ms$.

6.3 Electroencephalographic data

The following example presents three bivariate time series taken from a study of Quiroga et al. (2002). They study bivariate electroencephalographic

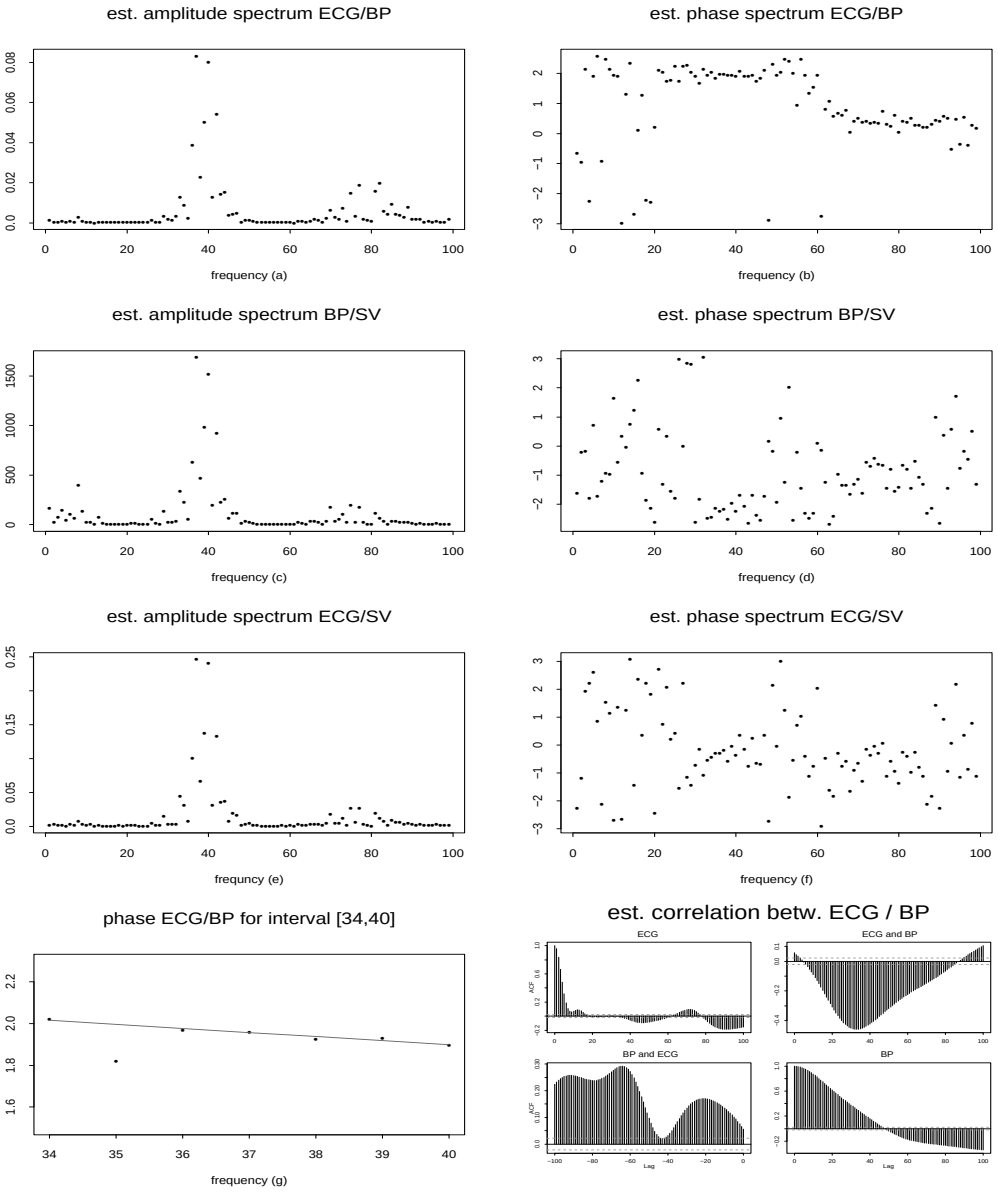

Figure 6.4: Amplitude (figures (a),(c),(e)) and phase spectrum (figures (b),(d),(f)) between the thresholding estimates of ECG, BP and SV. Figure (g) shows the phase spectrum between ECG and BP for $j \in [34, 40]$ and the cross covariance is given in (h).

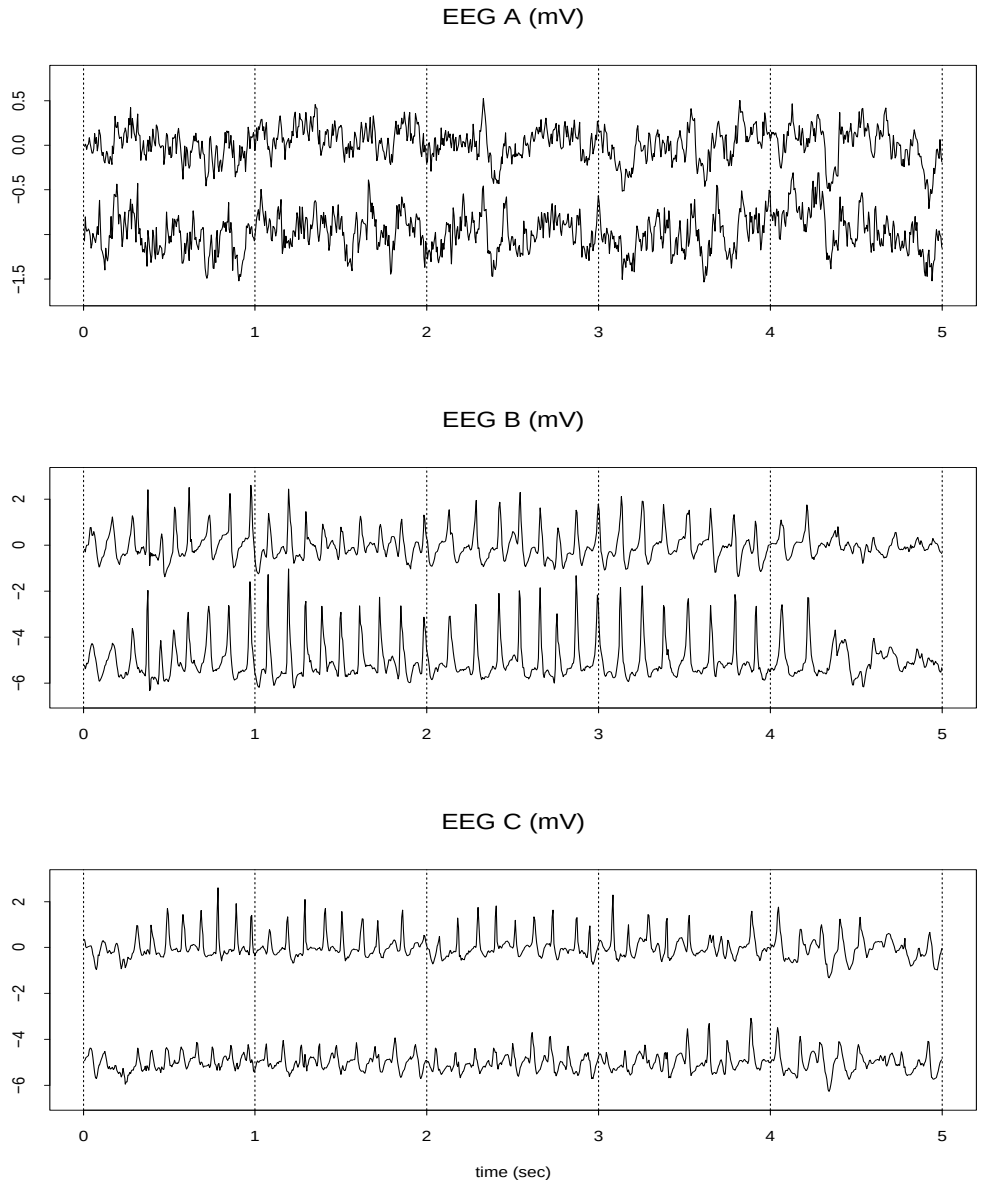

Figure 6.5: Left and right hemisphere channels of a rat EEG. **A** is a normal EEG where **B** and **C** contain abnormal spikes. Peaks and other abnormal interdependencies are commonly considered as a first hint of epileptic tendencies. An offset is added to the plot of the left channels (lower signals).

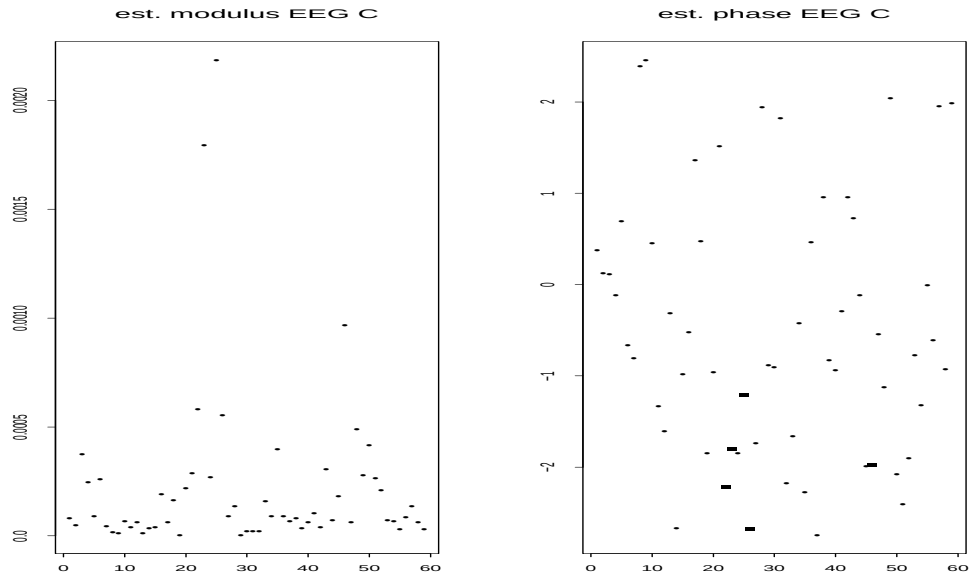

Figure 6.6: Amplitude and phase spectrum between thresholding estimates of the left and right channel of EEG C. Lag estimation involving the phase spectrum seems to be impossible.

(EEG) signals recorded from rats in order to analyze the interplay between certain parts of the brain. Each time series consists of the right and left hemisphere channels of a rat EEG recorded over a period of 5 seconds by cortical intracranial electrodes (left and right). All time series are measured at 200 Hz such that total sample size is 1000. Although all signals are taken from rats, EEG data of human beings qualitatively shows the same structure containing similar features such as periodicity of peaks. The three time series are displayed in figure 6.5. For better representation, the left channel signal is plotted with an offset. The first one (EEG A) is described as a normal bivariate EEG signal (Quiroga et al.) showing some synchronization between the two series covered with noise. The second and third series (EEG B and EEG C) present abnormal interdependencies particularly seen in the simultaneous presence of peaks occuring every tenth of a second. This may indicate epileptic behavior. Periodicity in series B and C seem to be the same, but the phase appears to be different. Peaks in the two series of signal B generally occur

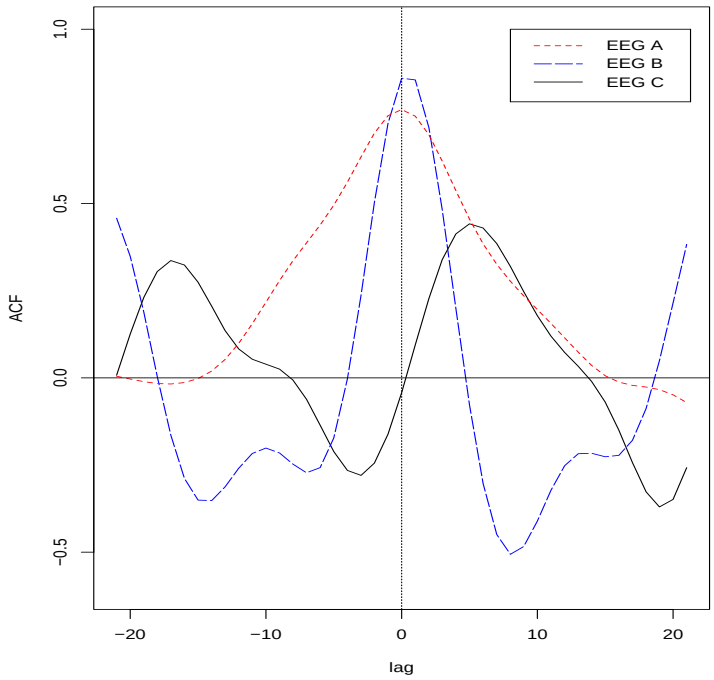

Figure 6.7: Regression cross correlation for the three bivariate EEG signals.

simultaneously, whereas the right channel of signal C shows a lag of changing amount over time when compared to the left channel. Although the three time series contain dominating periodic components lag estimation based on the phase spectrum is not feasible. We exemplary present the modulus and phase spectrum of the left and right channel of EEC C in figure 6.6. The five largest frequencies in the amplitude spectrum are marked in the phase plot by black boxes. No linear structure is observable, although a visual inspection may notice a time delay. We therefore use the techniques of chapter 5 based on the maximum of the regression cross correlation. The correlation between the left and right hemispheres of all three signals is displayed in figure 6.7. Obviously, EEG A and EEG B both reach their maximum at a lag $u = 0$, whereas the covariance of EEG C is almost 0 at this lag. EEG C takes its maximum at $u = 9$ indicating a lead of the left hemisphere of around 0.045 seconds.

6.4 Trading behavior at the NYSE

In this example trading behavior of stock market participants is investigated. Multivariate time series of stock prices from various industries traded at the New York Stock Exchange (NYSE) are analyzed by applying spectral analysis of trends functions as outlined in chapters 2 to 5. The method is of particular interest for high-frequency data, since there one typically observes pronounced seasonal behaviour (with a one-day periodicity) with rapidly changing intra-day trend functions. Also note that wavelet estimation is useful, because it allows for simultaneous investigation at different time scales. Our data consist of several stocks traded at the NYSE as recorded in the Trade and Quote database (TAQ). All time series are given as tick-by-tick data. This allows for modeling dependence structures at a very fine time resolution.

6.4.1 Data

We consider stock prices from 14 companies of various industries that are listed at the NYSE. For each individual stock all trades for the year 2001 are recorded. This includes volume and price for each single trade and the exact trading time (in seconds). The time series are not equidistant. For each individual day, all transaction of the continous trade between 09.30 am and 4 pm are recorded. The opening procedure in the morning differs from the rest of the day. For each share, the specialist collects all orders that reached the exchange after closing time the previous day and before opening in the morning. The actual price is then determined with respect to these orders and the share is released to the open trade. Since in this case, trading time, price and volume are not a result of actual trades we exclude this observation. Furthermore, we focus on companies with high intraday trading volume and small trading duration between subsequent trades. The size of a company

will be characterized by its cash flow (net cash from operating activities). For each industry, we compare a market leader (the company with the highest cash flow) with one of its market laggards. For example, in the category of conglomerates, General Electric (GE) with a cash flow of 21195 millions of dollars is the largest company. With about one seventh of the size (3078 m. \$), 3M Company (MMM) is one of its market laggards. The industries and companies, and their cash flows, are displayed in table 6.1.

6.4.2 Empirical Analysis

Univariate results

Trades at the NYSE are reported immediately once the order reached the exchange. Therefore, the original times series are non-equidistant. Here, we define equidistant series by aggregation at the time resolution of one minute. The price of a stock at time t is defined as the mean price over all trades between minute $t-1$ and t. Similarily, the volume at time t is the sum of volumes of all shares traded in the interval $(t-1,t]$. Observed volatilities are defined as follows. For a given set of stocks the time series of stock prices $\mathbf{X}(t) = (X_1(t), \ldots, X_p(t))^T$ $(t = 1, 2, ...)$ is transformed into a vector of volatilities $\mathbf{Y}(t) = (Y_1(t), \ldots, Y_p(t))^T$ by

$$Y_j(t) = |\Delta \log X_j(t)|^{1/4},\ j = 1, \ldots, p,$$

where Δ denotes the differencing operator. The power $1/4$ is used in order to obtain an approximately symmetric volatility distribution. The volatility series of General Electric, Citygroup, 3M Company and BP are displayed in figure 6.8. For reasons of illustration we restrict the time series plot to a period of one day whereas the histogram is plotted for a period of 100 trading

Selected stocks and industries (NYSE)

Industry	Market leader/ laggard	Ticker	Net cash (in m. $) (12/31/2001)
Conglomerates	*General Electric*/	GE	21195
	3M Company	MMM	3078
Money Center	*Citigroup Inc.*/	C	26578
Banks	*Wachovia Corp.*	WB	7287
Drug Manufacturers	*Pfizer Inc.*/	PFE	9263
(Major)	*Johnson & Johnson*	JNJ	8864
Aerospace/Defense	*Boeing Co.*/	BA	3814
(Major Diversified)	*Gen Dynamics Corp.*	GD	1103
Home	*Home Depot Inc.*/	HD	5963
Improvement Stores	*Lowes Companies*	LOW	1613
Major Integrated	*Exxon Mobil Inc.*/	XOM	22889
Oil & Gas	*BP plc.*	BP	17433
Wireless	*Vodafone Group plc.*/	VOD	11548
Communications	*China Mobile Limited*	CHL	6028

Table 6.1: Market leaders and laggards in 7 different industries. Last column presents net cash in millions of dollars from operating activities (source: quotemedia.com).

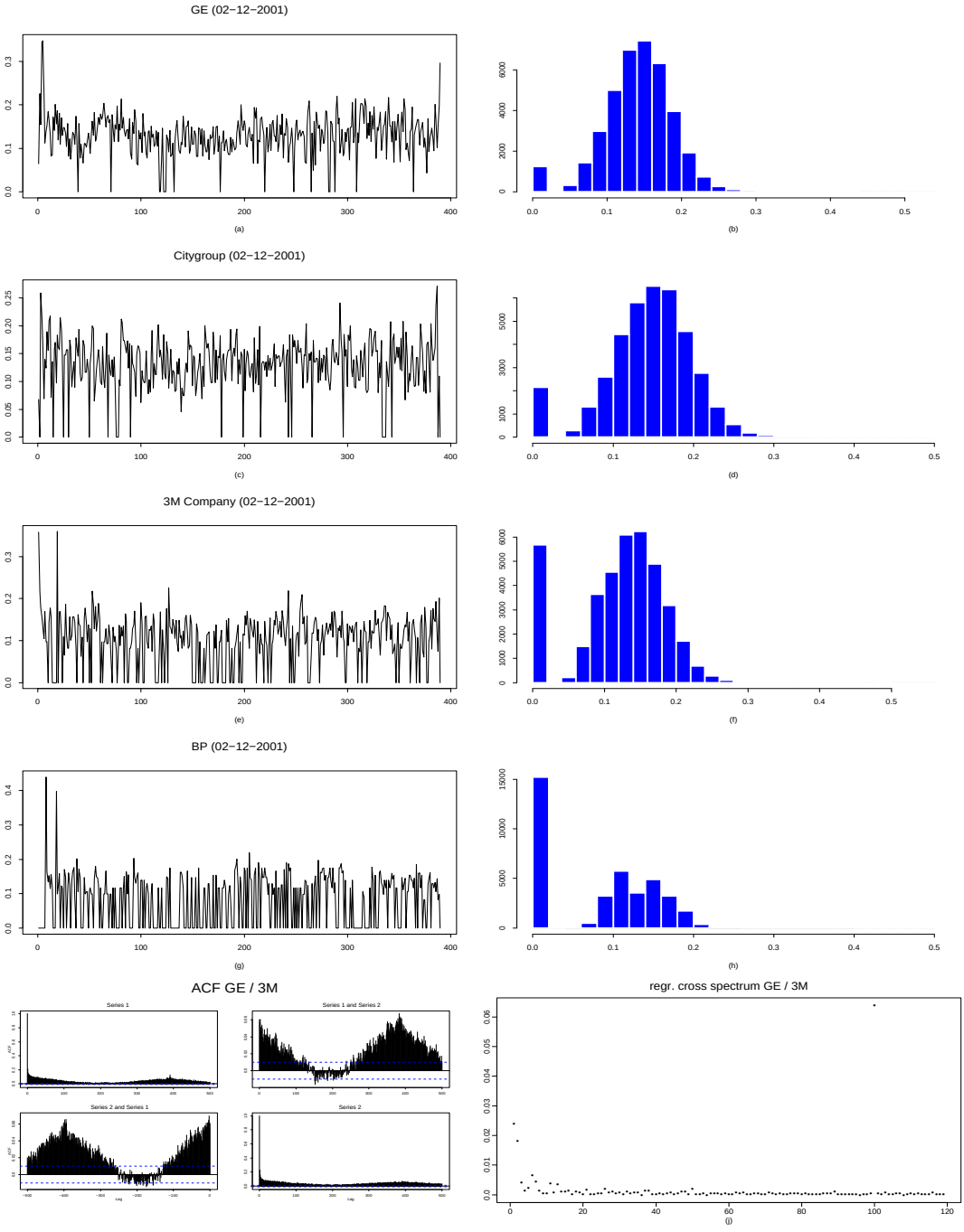

Figure 6.8: Volatilities of General Electric, Citygroup, 3M Company and BP (left column) with corresponding histograms (right column). The lower panels describe the traditional cross covariance and the regression cross spectrum between GE and MMM.

days (1 day equals 390 minutes starting at 09:30 and ending at 16:00.). For each series, there is a certain percentage of zeros. Apart from that, the distribution of volatilities is quite symmetric around the median of 0.15. The plot of empirical cross correlations between GE and 3M in figure 6.8(i) indicates a seasonal component with a period of about one day. Similar comments apply to cross correlations between the other time series. Figure 6.8(j) displays the modulus of the regression cross spectrum between GE and 3M. The peak at frequency $j = 100$ ($n = 39000$ min $\hat{=}$ 100 trading days) corresponds to a period of one day. We may conclude that there is a common daily seasonality

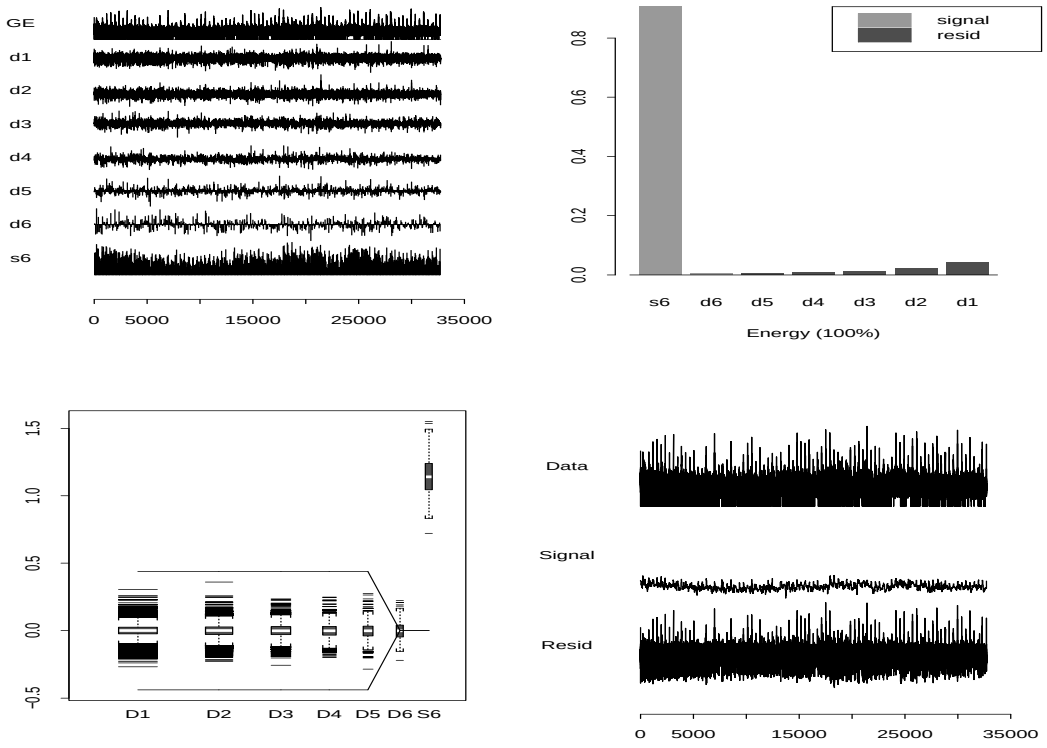

Figure 6.9: General Electric - original time series and multiresolution decomposition for the levels S6 to D5.

in the data. The intraday seasonality will now be analyzed in more detail. Figure 6.9 shows the decomposition of GE volatility into different wavelet components. Here, S8-wavelets are used, because of their smoothness and

symmetry. Coefficients at different resolution levels are shown in the upper left figure. The length of the vertical lines is proportional to the size of the corresponding wavelet coefficient. The absolute size of the wavelet coefficients is highest for the S6 component. Level 6 (S6 and D6) represents the long-term behavior of the time series. The energy distribution of the wavelet decomposition is displayed in the upper right panel. Obviously, almost all energy is concentrated in the smooth component S6. Overall, the level $J = 6$ accounts for more than 90% of the energy in the series. We therefore apply wavelet thresholding such that all coefficients at resolution level $J = 6$ are included in the estimate (i.e. the threshold is set equal to zero), whereas a high threshold is used for finer levels. The components of the wavelet representation at level 6 are given by

$$\phi_{6,k}(x) = 2^{6/2}\phi(2^6 x - k),$$
$$\psi_{6,k}(x) = 2^{6/2}\psi(2^6 x - k).$$

Each component represents $2^6 = 64$ minutes. This corresponds to an analysis at the time resolution of one hour. Boxplots of the coefficients at different levels are plotted in the lower left panel, together with horizontal lines indicating the level-dependent thresholds. The two thresholds for level 6 were chosen to be zero. For the remaining components, all coefficients are smaller (in absolute value) than the threshold and are therefore set equal to zero. Notice the parsimony of the resulting shrinkage estimate. Estimated daily trends are characterized by about 12 wavelet coefficients. The lower right panel shows the decomposition of the time series into the thresholding fit and residuals. Note that the variance of the signal is relatively small compared to the variance of the noise.

Multivariate results

In the following, we are interested in multivariate dependence structures in the trading behavior of market participants. We conjecture that, within a specific industry, shares of large companies tend to be traded slightly earlier than shares of smaller companies. Specifically, we analyze possible time delays in the volatility process for the seven industries described in table 6.1. For illustration, consider for instance volatilities of General Electric (GE), Citygroup (C), 3M Company (MMM) and BP for one single day only. Figure 6.12 presents the estimated trend components in the volatilities together with the corresponding trading volume. The time span is 390 minutes (09.30 am to 04.00 pm). During the first 30 minutes, a large portion of the day's transactions are carried out, implying highly volatile stock prices. Typically, this first peak is followed by an additional increase in volatility around about 90 to 150 minutes after opening of the stock exchange. Volatility is low between 12 am to 1 pm (150 to 210 minutes) - where traders usually take their lunch break. Finally, an increase in volatility can be observed towards the end of the day. This results in a typically U-shaped appearence of volatility series.

Similar observations have been made, for instance, by Bauwens and Giot (2001) who analyzed the duration between subsequent trades of a particular stock. They observed that duration time is relatively low at the beginning and at the end of the trading day, whereas durations around lunch time are relatively large. All four time series exhibit a second peak, indicating an increased trading activity before lunchtime. Analogous comments apply to the series of actual trading volumes. Typically, one can see an increased trading volume at the beginning and at the end of the day, with additional trading peaks between 12-12:30 (t=150-180 min). Figure 6.11 compares volatility trends of GE (black line) and 3M (red line) for 4 separate weeks with 5 successive days of trading. A visual inspection indicates that, overall, GE may have a lead over 3M. Consider for example February 12th, 2001 (upper left panel). There

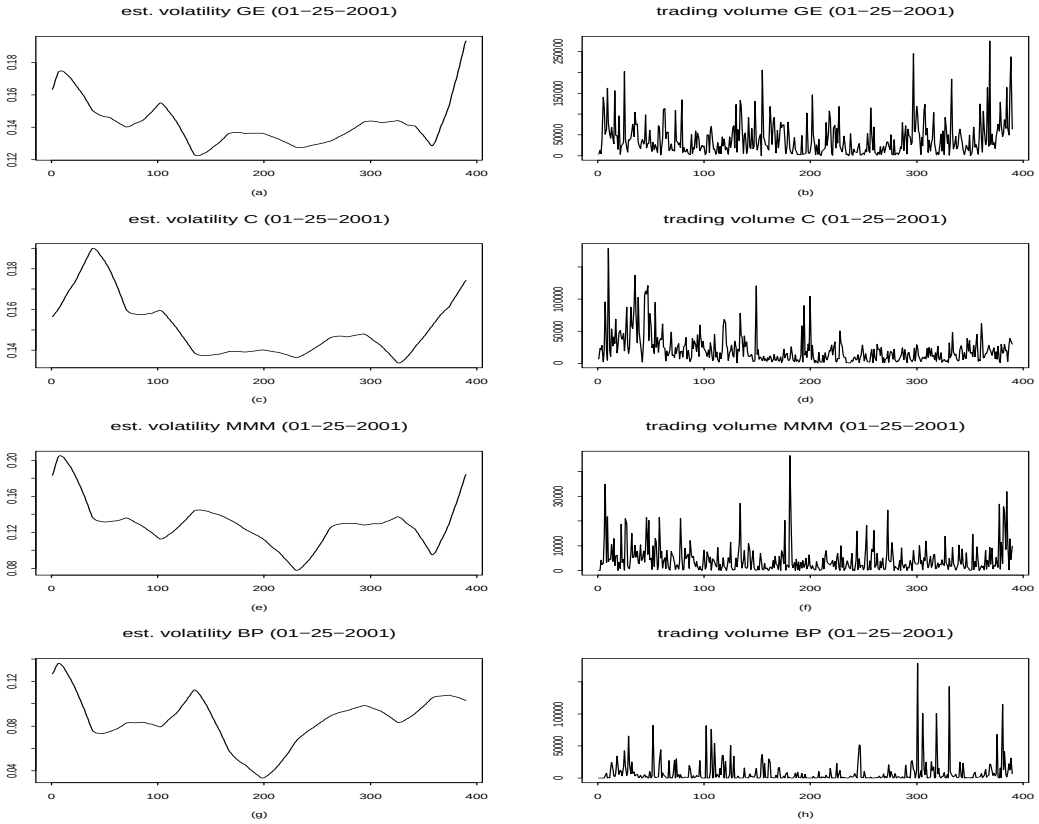

Figure 6.10: Estimated volatility trends (left column) and the actual traded volume of shares during one day for GE, Citygroup, 3M and BP.

is a clear resemblance between the two trend functions, with the red curve (MMM) being shifted to the right. The same observation can be made for most of the other days. In contrast, in the original time series, this resemblance and lead-lag relationship is not visible at all (see figure 6.12(e),(f) for February 12th, 2001).

Consider now the conjectured shift between trend components for several industries. Figure 6.12(a)-(d) exemplary displays the regression cross spectrum between GE and 3M for a single day (a,b) and for 100 consecutive trading days (c,d). Intraday activity is characterized by low frequency components. The corresponding phase plot (figure 6.12(b)) reveals an almost linear structure over this set of frequencies. In view of the upper left plot of figure 6.11, this is due to the almost perfect shift between the two trends in the shares for this

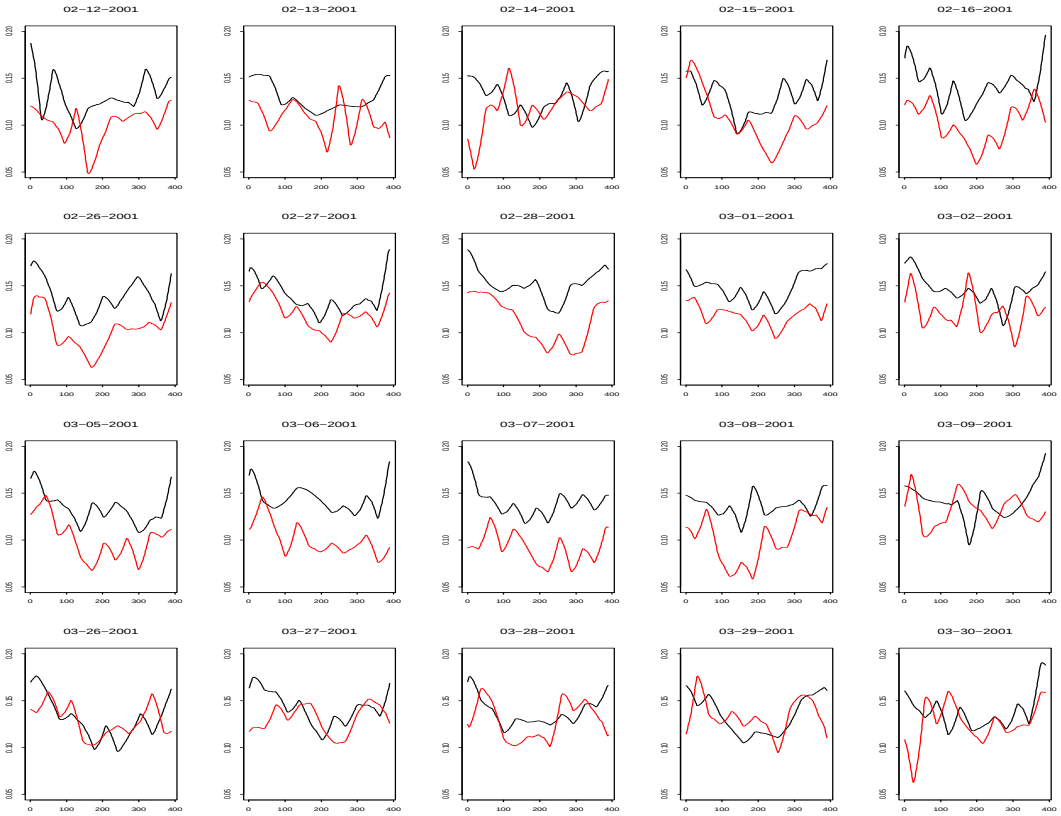

Figure 6.11: Trends in volatilities for GE (black line) and 3M (red line) for 20 trading days.

day. The estimated lag for this set of frequencies is around 80 minutes. However, the regression spectrum cannot be used for estimation over an interval of 100 days. Visual inspection of the phase spectrum figure 6.12(d) does not reveal any traceable linear structure. We therefore apply estimation involving the regression cross spectrum.

For each of the seven industries listed in table 6.1, the regression cross covariance and the lag u that maximizes $|\hat{\gamma}_{rs}(u)|$ are calculated. Since we are interested in short-term trading patterns we restrict the maximal allowed lag to a maximum of two hours. This is not a real restriction because, for all industries, the local maximum within this range turns out to be a global maximum. The results are summarized in table 6.2. The highest regression cross correlation between a market leader and laggard is equal to 0.42 for GE and

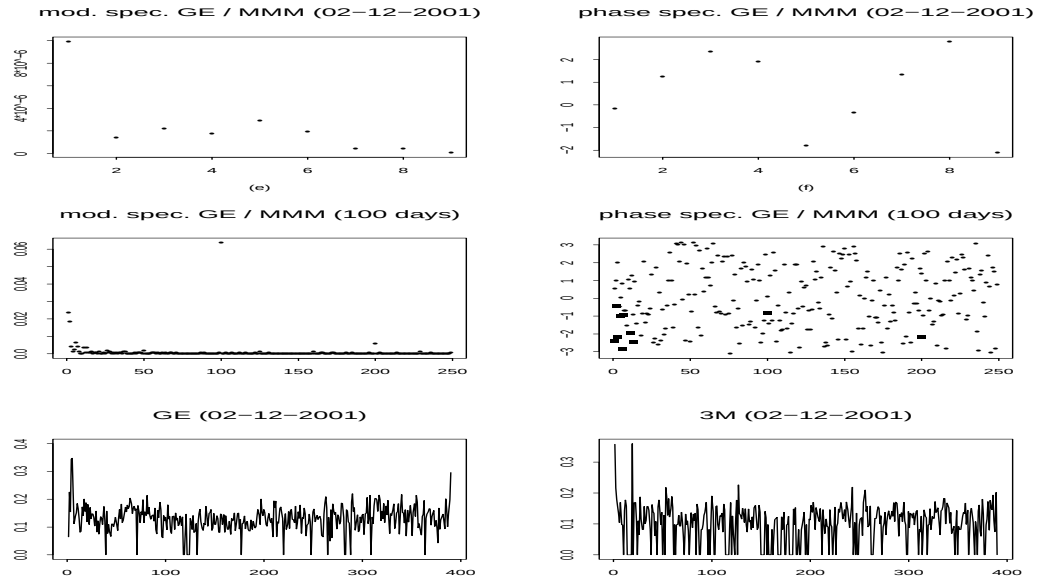

Figure 6.12: Regression cross spectrum between GE and 3M for a single day (a,b) and for a period of 100 consecutive trading days (c,d). Original volatility series of GE and 3M for February 12, 2001, are displayed in (e,f).

3M, in the industry sector "conglomerates". High cross correlations can also be observed for Citygroup and Wachovia, Exxon and BP, Pfizer and Johnson & Johnson, Boeing and Gen Dynamics, and Vodafone and China Mobile. For all industries, the estimated lag is positive, i.e. the market leader tends to have a lead over the market laggard. The only exception is Vodafone and China Mobile where the highest correlation occurs at lag 0.

Selected stocks and industries (NYSE)

industry	market leader/ laggard	estimated lag (min.)
Conglomerates	*GE/ MMM*	14 (.4176)
Money Center Banks	*C/ WB*	10 (.3297)
Drug Manufacturers	*PFE/ JNJ*	6 (.3095)
Aerospace/Defense	*BA/ GD*	23 (.3097)
Home	*HD/ LOW*	6 (.0600)
Major Integrated	*XOM/BP*	4 (.3317)
Wireless Comm.	*VOD/CHL*	0 (.3264)

Table 6.2: Estimated lag between companies in individual industries when applying the regression cross correlation.

Chapter 7

Concluding Remarks

In this paper we investigated dependence structures in multivariate time series that are due to similarities in the underlying deterministic component. Developed methods involved the regression spectrum, the periodogram and the regression cross correlation. All methods were theoretically discussed and illustrated with a sequence of applications. The usefulness of the nonparametric regression spectrum goes far beyond a purely mathematical device. It can be used as a data analytical tool to identify common frequencies and lead-lag effects in multivariate time series with strong deterministic components. The physiological and climatic series considered above are typical examples. As presented above, application of the phase spectrum heavily depends on the amount of common frequencies in the cross amplitude spectrum. In case of a relatively large number lag estimation using the phase line seems to be appropriate, whereas a small set of frequencies rather suggests to apply the regression correlation. Though in each example the decision about the method was a personal one, we obtained reasonable results. However, it is the task of future research to develop a data-driven approach that automatically choses the best method for analyzing the interdependencies in the deterministic component of a multivariate time series.

The use of wavelets for nonparametric trend estimation offers many advantages compared to other techniques such as Fourier analysis, piecewise polynomial fits or kernel smoothing. Wavelets are well-localized both in time and frequency and offer lots of optimality properties such as smoothness and adaptivity both in view of global or local loss over large classes of functions. However, there are still many open questions to be solved and the research on this topic is still in progress. Empirical studies show that estimation of deterministic and stochastic mulivariate dependence structures heavily depend on many individual choices of parameters in the wavelet approach. This complicates automated applications and might yield to objectively non-understandable results. In the following we present several possible extensions to the current paper and discuss some shortcomings that are open for future research.

- **Choice of wavelet basis.** The theoretical results of this paper were based on presence of compactly supported wavelets that are of bounded variation. These properties are shared by many bases. Therefore, performance of the methods may be improved if one allows the wavelets to be chosen from a library of wavelet bases of different length of support, symmetry and regularity. Selection of an appropriate basis may lead to significant improvements in applications, particularly in case that a priori information about the unknown f is available.

- **Threshold selection**. Chapter 2 extensively discusses possible choices of threshold λ and mentiones several different thresholding rules that perform close to minimax solutions. However, from a point of view of applications, loss measured by the minimax risk is not appropriate. Minimax risks minimize the worst possible case over very large sets of functions such that the thresholds are chosen almost independently from f. In applications the concentration on a small set of functions is realistic. This will allow to develop a data-adaptive approach to wavelet thresholding both theoretically and by giving certain algorithms. Fur-

thermore, application of minimax risks leads to the occurrence of local artifacts in trend extraction. Simulations show that these heavily influence e.g. estimation of the phase spectrum. With the target of estimating the regression spectrum new measures of performance may be chosen. One might think of a minimax condition combined with certain smoothness criteria to overcome the problem of spurious artifacts.

- **Estimation of phase shift.** In view of prediction of the future the question of estimating the phase shift is of particular importance. Most results in the previous chapters are of asymptotic nature. Estimation of the phase shift is linked with several problems. First, the phase shift is observed for frequencies in $[-\pi, \pi]$ only and inevitable 2π-jumps increase the difficulty to estimate the phase line. Lead-lag structures may be observed in regions where the phase line is almost linear because the derivative of the phase line is proportional to the time delay. However, estimation of $\phi'(j)$ in regions with 2π-jumps will be difficult. This is enhanced if lag structures change over different intervals of frequencies. In general, deterministic components are expected to show mainly low frequencies. The regression spectrum may also reveal coherence between frequencies that are too far away from each other to allow for a visible detection of the connection. In addition, in regions of high frequency it may be difficult to distinguish between signal and noise. The indefiniteness with respect to $2\pi-$jumps enhances this problem even more.
  In presence of a trend function the deterministic and stochastic components interfere in the regression spectrum such that the individual contributions are hard to distinguish. As a consequence, the influence of the noise may prevent an analyst from detecting the true regression phase spectrum.

- **Prediction**. Estimation of individual components from the cross spectrum allows to estimate lagged components in a multivariate time series.

As seen in chapter 6, one may try to predict the recruitment of new fish in the Pacific based on the water temperature at the surface in the past. Another example is to predict the share of a market laggard by considering the performance of the market leader, if a lagged relationship is actually observed. Therefore, the development of prediction intervals based on the theoretical results of chapter 3 to 5 is a straightforward extension.

- **Extension to long memory**. The existence of a local varying deterministic trend functions may explain long-memory in volatilities. The separation of trend and stochastic long-memory component have already been investigated. The implications on the regression spectrum may be considered as well.

- **Applications**. The methods of this paper are particularly useful to analyze high frequency data. Data analysis of volatilities and trading volume of certain shares on a time scale from 1 to 5 minutes have shown that extraction of the regression spectrum may show interesting multivariate dependence structures. These features were not observable in the original time series. A comprehensive study of financial data sets could explain certain interplays in economy that are still unexplained up to date. Due to the hierarchical structure of the wavelets the dependence between market macro- and microstructure may be investigated. The extent of presence of deterministic components will determine the quality of prediction.

Bibliography

[1] Abramovich, F. and Benjamini, Y. (1995). Thresholding of wavelet coefficients as a multiple hypotheses testing procedure, *in Wavelets and Statistics*, Lecture Notes in Statistcs (eds. Antoniadis, A. and Oppenheim, G.), Vol. 103, 5-14.

[2] Abramovich, F. and Benjamini, Y. (1996). Adaptive thresholding of wavelet coefficients, *Computational Statistics and Data Analysis*, Vol. 22, 351-361.

[3] Abramovich, F., Sapatinas, T. and Silverman, B.W. (1998). Wavelet thresholding via a bayesian approach, *Journal of the Royal Statistical Society*, Vol. 60, Part 4, 725-749.

[4] Aldroubi, A. and Unser, M. (1996). *Wavelets in medicine and biology*, CRC Press Inc., Boca Raton.

[5] Altman, N.S. (1990). Kernel smoothing of data with correlated errors. *Journal of the American Statistical Association*, Vol. 85, 749-759.

[6] Antoniadis, A. and Carmona, R. (1991). Multiresolution analyses and wavelets for density estimation, *Technical Report*, University of California, Irvine.

[7] Antoniadis, A. and Oppenheim, G. (eds.) (1995), *Wavelets and Statistics*, Lecture Notes in Statistics, Vol. 103, Springer, Heidelberg.

[8] Bauwens, L. and Giot, P. (2001). *Econometric modelling of stock market intraday activity*, Advanced Studies in Theoretical and Applied Econometrics, Kluwer Academic Publishers, Boston.

[9] Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing, *Journal of the Royal Statistical Society B*, Vol. 57, 289-300.

[10] Beran, J. and Feng, Y. (2001). Local polynomial estimation with a FARIMA-GARCH error process, *Bernoulli*, Vol. 7, No. 5, 730-755.

[11] Beran, J. and Gosh, S. (2000). Estimation of the dominating frequency for stationary and nonstationary fractional autoregressive models, *Journal of Time Series Analysis*, Vol. 21, No. 5, 517-533.

[12] Beran, J. and Ocker, D. (1999). SEMIFAR forecasts, with applications to foreign exchange rates. *Journal of Statistical Planning and Inference*, Vol. 80, 137-153.

[13] Billingsley, P. (1968). *Convergence of probability measures*, John Wiley & Sons Inc., New York.

[14] Bretagnolle, J. and Huber, C. (1979). Estimation des densités: risque minimax, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, Vol. 47, 119-137.

[15] Brillinger, D. (1994). Uses of cumulants in wavelet analysis, *Proc. SPIE, Advanced Signal Processing*, Vol. 2296, 2-18.

[16] Brillinger, D.R. (1996). Some uses of cumulants in wavelet analysis, *Nonparametric Statistics*, Vol. 6, 93-114.

[17] Brockwell, P.J. and Davis, R.A. (1987). *Time series: theory and methods*, Springer, New York.

[18] Bruce, A. and Gao, H.-Y. (1996). *Applied wavelet analysis with S-Plus*, Springer, New-York.

[19] Burman, P. (1989). A comparative study of ordinary cross-validation, ν-fold cross-validation and the repeated learning-testing methods, *Biometrika*, Vol. 76, 503-514.

[20] Cai, T.T. (1999). Adaptive wavelet estimation: a block thresholding and oracle inequality approach, *The Annals of Statistics*, Vol. 27, No. 3, 898-924.

[21] Cai, T.T. (2002). On block thresholding in wavelet regression: adaptivity block size and threshold level, *Statistica Sinica*, Vol. 12, 1241-1273.

[22] Cai, T.T. and Brown, L. (1998). Wavelet shrinkage for nonequispaced samples, *The Annals of Statistics*, Vol. 26, No. 5, 1783-1799.

[23] Cai, T.T. and Low, M.G. (2005). Nonparametric estimation over shrinking neighborhoods: superefficiency and adaptation, *The Annals of Statistics*, Vol. 33, No. 1, 183-213.

[24] Cai, T.T. and Silverman, B.W. (2001). Incorporating information on neigboring coefficients into wavelet estimation, *Sankhya Series B*, Vol. 63, 127-148.

[25] Cai, T.T. and Zhou, H.H. (2005). A data-driven block thresholding approach to wavelet estimation, Technical Report, Department of Statistics, University of Pennsylvania.

[26] Capobianco, E. (1999). Wavelets transforms for the statistical analysis of financial volatility, *Methods Comp. Appl. Prob. I*, No. 4, 423-443.

[27] Chicken, E. and Cai, T.T. (2005). Block thresholding for density estimation: local and global adaptivity, *Journal of Multivariate Analysis*, Vol. 95, 76-106.

[28] Chipman, H. A., Kolaczyk, E.D. and McCulloch, R. E. (1997). Adaptive bayesian wavelet shrinkage, *Journal of the American Statistical Association*, Vol. 92, No. 440, 1413-1421.

[29] Cohen, A. and Ryan, R. (1995). *Wavelets and multiscale signal processing*, Chapman & Hall, London.

[30] Daubechies, I. (1992). *Ten lectures on wavelets*, Society for Industrial and Applied Mathematics, Philadelphia.

[31] De Canditiis, D. and Vidakovic, B. (2001). *Wavelet bayesian block shrinkage via mixtures of normal-inverse gamma priors*, Gatech Tech Report G05/2001, Georgia Tech.

[32] Devore, R.A. and Popov, V.A. (1988). Interpolation of Besov spaces, *Transactions of the American Mathematical Society*, Vol. 305, No. 1, 397-414.

[33] Diggle, P.J. (1990). *Time series: a biostatistical approach*, Oxford University Press, Oxford.

[34] Donoho, D.L. (1992). De-noising via soft-thresholding, Technical Report, Department of Statistics, Stanford University.

[35] Donoho, D.L. (1993). Nonlinear wavelet methods for recovery of signals, densities, and spectra from indirect and noisy data, in *Proceedings of Symposia in Applied Mathematics*, Vol. 47, 173-205, American Mathematical Society.

[36] Donoho, D.L. and Johnstone, I.M. (1992). Non-classical minimax theorems, Technical Report, Department of Statistics, Stanford University.

[37] Donoho, D.L. and Johnstone, I.M. (1994a). Ideal spatial adaption by wavelet shrinkage, *Biometrika*, Vol. 81, No. 3.

[38] Donoho, D.L. and Johnstone, I.M. (1994b). Ideal denoising in an orthonormal basis chosen from a library of bases, *C. R. Acad. Sci. Paris Sr. A Math.*, Vol. 319, 1317-1322.

[39] Donoho, D.L. and Johnstone, I.M. (1995). Adapting to unknown smoothness via wavelet shrinkage, *Journal of the American Statistical Association*, Vol. 90, 1200-1224.

[40] Donoho, D.L. and Johnstone, I.M. (1998). Minimax estimation via wavelet shrinkage, *The Annals of Statistics*, Vol. 26, No.3, 879-921.

[41] Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. and Picard, D. (1996). Density estimation by wavelet thresholding, *The Annals of Statistics*, Vol. 24, No.2, 508-539.

[42] Doukhan, P. (1988). Déviation quadratique d'estimateurs de densité par projections orthogonales, *C. R. Acad. Sci. Paris Sr. A Math.*, Vol. 306, 663-666.

[43] Efromovich, S. (1985). Nonparametric estimation of a density of unknown smoothness, *Theoretical Probability and Applications*, Vol. 30, 557-661.

[44] Gao, H.-Y. (1996). *Spectral density estimation via wavelet shrinkage*, StatSci Division of MathSoft Inc., Seattle.

[45] Gao, H.-Y. (1997). Choice of thresholds for wavelet shrinkage estimates of the spectrum, *Journal of Time Series Analysis*, Vol. 18, No. 3, 231-251.

[46] Goldberger, A.L., Amaral, L.A.N., Glass, L., Hausdorff, J.M., Ivanov, P.C., Mark, R.G., Mietus, J.E., Moody, G.B., Peng, C.-K. and Stanley, H.E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals, *Circulation*, Vol. 101(23), e215-e220.

[47] Grenander, U. and Rosenblatt, M. (1957). *Statistical analysis of stationary time series*, John Wiley & Sons, New York.

[48] Hannan, E.J. (1970), *Multiple time series*, John Wiley & Sons Inc., New York.

[49] Hart, J.D. (1994). Automated kernel smoothing of dependent data by using time-series cross-validation. *Journal of the Royal Statistical Society B*, Vol. 56, 529-542.

[50] Härdle, W., Kerkyacharian, G., Picard, D. and Tsybakov, A. (1998). *Wavelets, approximation, and statistical applications*, Lecture Notes in Statistics, Springer, New York.

[51] Hall, P., Kerkyacharian, G. and Picard, D. (1995). *On block threshold rules for curve estimation using wavelet methods*, Manuscript.

[52] Hall, P., Kerkyacharian, G. and Picard, D. (1998). Block threshold rules for curve estimation using kernel and wavelet methods, *The Annals of Statistics*, Vol. 26, No.3, 922-942.

[53] Hall, P., Kerkyacharian, G. and Picard, D. (1999). On the minimax optimality of block thresholded wavelet estimators, *Statistica Sinica*, Vol. 9, 33-49.

[54] Hall, P., Spiridon, P. and Kerkyacharian, G. (1997). Numerical performance of block thresholded wavelet estimators, *Statistics and Computing*, Vol. 7, 115-124.

[55] Heiler, S. and Feng, Y. (2000). Data-driven decomposition of seasonal time series. *Journal of Statistical Planning and Inference*, Vol. 91, 351-363.

[56] Huang, S.-Y. (1999). Density estimation by wavelet-based reproducing kernels, *Statistica Sinica*, Vol. 9, No. 1, 137-151.

[57] Ibragimov, I.A. and Hasminskii, R.Z. (1980). On nonparametric estimation of regression, *Soviet Mathematics Doklady*, Vol. 21, 810-814.

[58] Ibragimov, I.A. and Hasminskii, R.Z. (1981). *Statistical estimation: asymptotic theory*, Springer, New York.

[59] Ichimaru, Y. and Moody G.B. (1999). Development of the polysomnographic database on CD-ROM, *Psychiatry and Clinical Neurosciences*, Vol. 53, 175-177.

[60] Jansen, M. and Bultheel, A. (1997). Multiple wavelet threshold estimation by generalized cross validation for data with correlated noise, Technical Report TW250, Department of Computer Science, Katholieke Universiteit Leuven.

[61] Jansen, M. and Bultheel, A. (1998). Smoothing non-equidistantly sampled data using wavelet and cross validation, in *Proceedings of the IEEE Benelux Signal Processing Symposium*, Leuven, Belgium, 111-114.

[62] Jenkins, G.M. and Watts, D.G. (1968). *Spectral analysis and its applications*, Holden-Day, Oakland.

[63] Johnstone, I.M. and Silverman, B.W. (1997). Wavelet threshold estimators for data with correlated noise, *Journal of the Royal Statistical Society B*, Vol. 59, No. 2, 319-351.

[64] Kerkyacharian, G. and Picard, D. (1992). Density estimation in Besov spaces, *Statist. Probab. Lett.*, Vol. 13, 15-24.

[65] Kerkyacharian, G. and Picard, D. (1993). Density estimation by kernel and wavelets method: Optimality of Besov spaces, *Statist. Probab. Lett.*, Vol. 18, 327-336.

[66] Lavrik, I., Jung, Y.Y., Ruggeri, F. and Vidakovic, B. (2005). Bayesian false discovery rate wavelet shrinkage: theory and applications, Technical Report, Department of Biomedical Engineering, GaTech, Atlanta.

[67] Mallat, S.G. (1989a). Multiresolution approximation and wavelets, *Trans. Amer. Math. Soc.*, Vol. 315, 69-88.

[68] Mallat, S.G. (1989b). A theory for multiresolution signal decomposition: the wavelet representation, *IEEE Transactions of Pattern Analysis and Machine Intelligence*, Vol. 11, No. 7, 674-693.

[69] Murawski, S.A. (1993). Climate change and marine fish distributions: forecasting from analogy, *Transactions of the American Fisheries Society*, Vol. 122, 647-658.

[70] Nason, G.P. (1995). Choice of the threshold parameter in wavelet function estimation, in: *Wavelets and Statistics*, Lecture Notes in Statistics (eds. Antoniadis, A. and Oppenheim, G.), Vol. 103, Springer, New York, 261-280.

[71] Nason, G.P. (1996). Wavelet shrinkage using cross-validation, *Journal of the Royal Statistical Society B*, Vol. 58, 463-479.

[72] Nemirowskii, A.S., Polyak, B.T. and Tsybakov, A.B. (1983). Estimators of maximum likelihood type for nonparametric regression, *Soviet Mathematics Doklady*, Vol. 28, No. 3, 788-792.

[73] Nemirowskii, A.S., Polyak, B.T. and Tsybakov, A.B. (1985). Rate of convergence of nonparametric estiamtes of maximum-likelihood type, *Problems of Information Transmission*, Vol. 21, 258-272.

[74] Neumann, M.H. and von Sachs, R. (1995). Wavelet thresholding: beyond the Gaussian I.I.D. situation, in: *Wavelets and Statistics*, Lecture Notes in Statistics (eds. Antoniadis, A. and Oppenheim, G.), Vol. 103, Springer, New York, 301-329.

[75] Pensky, M. and Vidakovic, B. (1998). On non-equally spaced wavelet regression, *Technical Report 98-06*, ISDS, Duke University.

[76] Percival, D.P. and Walden, A.T. (2000). *Wavelet methods for time series analysis*, Cambridge University Press, Cambridge.

[77] Pickands, J. (1967). Maxima of stationary Gaussian process, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, Vol. 7, 190-223.

[78] Pinsky, M.A. (2002). *Introduction to Fourier analysis and wavelets*, The Brooks / Cole Series in Advanced Mathematics, Brooks / Cole, Pacific Grove.

[79] Pollard, D. (1984). Convergence of stochastic processes, 2^{nd} Edition, Springer, New York.

[80] Polya, G. and Szegö, G. (1964). Aufgaben und Lehrsätze aus der Analysis, Erster Band, 3. Auflage, Springer, Berlin.

[81] Priestley, M.B. (1989). *Spectral analysis and time series*, 6^{th} Edition, Academic Press, London.

[82] Prokoph, A. and Barthelmes, F. (1996). Detection of nonstationarities in geological time series. Wavelets transform of chaotic and cyclic sequences, *Computer & Geosciences*, Vol. 22, 1097-1108.

[83] Quiroga, R.Q., Kraskov, A., Kreuz, T. and Grassberger, P. (2002). Performance of different synchronization measures in real data: a case study on electroencephalographic signals, *Physical Review E*, Vol. 65: 041903.

[84] Ramsey, J.B. (2002). Wavelets in economics and finance: past and future, *Studies in Nonlinear Dynamics & Econometrics*, Vol. 6, No. 3, 1-27.

[85] Ramsey, J.B. and Lampart, C. (1998). The decomposition of economic relationships by time scale using wavelets: expenditure and income, Vol. 3, No. 1, 23-42.

[86] Ruggeri, F. and Vidakovic, B. (1999). A Bayesian decision theoretic approach to the choice of thresholding parameter, *Statistica Sinica*, Vol. 9, 183-197.

[87] Ruggeri, F. and Vidakovic, B. (2001). Bayesian modeling in the wavelet domain, in: *Bayesian Thinking: Modeling and Computation*, Handbook of Statistics, Vol. 25, Elsevier North-Holland, Amsterdam, 315-338.

[88] Sardy, S., Percival, D.B., Bruce, A.G., Gao, H.-Y. and Stuetzle, W. (1998). Wavelet shrinkage for unequally spaced data, *Research Report No. 41*, MathSoft Inc.

[89] Shiryaev, A.N. (1996). *Probability*, 2^{nd} Edition, Springer, New York.

[90] Silverman, B.W. (1999). Wavelets in statistics: beyond the standard assumptions, *Phil. Trans. R. Soc. Lond. A*, Vol. 357, 2459-2473.

[91] Shumway, R.H. and Stoffer, D.S. (2000). *Time series analysis and its applications*, 2^{nd} Edition, Springer, New York.

[92] Steeb, W.-H. (1998). *Hilbert spaces, wavelets, generalised functions and modern quantum mechanics*, Kluwer Academic Publishers, Dordrecht.

[93] Stein, C. (1981). Estimation of the mean of a multivariate normal distribution, *The Annals of Statistics*, Vol. 9, 1135-1151.

[94] Strichartz, R.S. (1994). Construction of orthonormal wavelets, in: *Wavelets: Mathematics and Applications, Studies in Advanced Mathematics* (eds. Benedetto, J.J. and Frazier, M.W.), CRC Press, Boca Raton.

[95] Vanraes, E., Jansen, M. and Bultheel, A. (2002). Stabilized wavelet transforms for non-equispaced data smoothing, *Signal Processing* Vol. 82, No. 12, 1979-1990.

[96] Victor, B.C., Wellington, G.M., Robertson, D.R. and Ruttenberg, B.I. (2001). The effect of the EL Niño-Southern Oscillation event on the distribution of reef-associated labrid fishes in the eastern Pacific Ocean, *Bulletin of Marine Science*, Vol. 69, No. 1, 280-288.

[97] Vidakovic, B. (1994). Nonlinear wavelet shrinkage with bayes rules and bayes factors, Discussion Paper **94 − 24**, ISDS, Duke University.

[98] Vidakovic, B. (1998), Nonlinear wavelet shrinkage with bayes rules and bayes factors, *Journal of the American Statistical Association*, Vol. 93, No.441, 173-179.

[99] Vidakovic, B. (1999). *Statistical modeling by wavelets*, Wiley Series in Probability and Statistics, John Wiley & Sons, New York.

[100] Wang, Y. (1996), Function estimation via wavelet shrinkage for long-memory data, *The Annals of Statistics*, Vol. 24, No. 2, 466-484.

[101] Yajima, Y. (1991). Asymptotic properties of the LSE in a regression model with long-memory stationary errors, *The Annals of Statistics*, Vol. 19, No.1, 158-177.