Data adaptive wavelet methods for Gaussian long-memory processes

Dissertation

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

vorgelegt von

Yevgen Shumeyko

Tag der mündlichen Prüfung: 28. Februar 2012

Referenten
Prof. Dr. Jan Beran (Universität Konstanz)
Prof. Dr. Donald B. Percival (Universität Washington)

Acknowledgment

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 *Center of Finance and Econometrics* at the University of Konstanz.

It is a pleasure to convey my gratitude to my supervisor Prof. Dr. Jan Beran for his advice, and guidance from the first stage of this research. I would like to record my gratitude to Prof. Dr. Donald B. Percival for interesting discussions that inspired us to look at bootstrap in the context of wavelets.

It is a pleasure to pay tribute also to Prof. Dr. Alexander Shumeyko and Dr. Vsevolod Bodrikov for their advice and correction. My special thanks also go to my wife Nadiya Bodrikova, to my daughter Stefanie, to my whole family and my friends for their support and patience since the beginning of my studies.

Konstanz, November 2011 Yevgen Shumeyko

Abstract

In this thesis, we investigate some adaptive wavelet approaches for a so-called nonparametric regression model with strongly dependent Gaussian residuals. At first, we discuss data adaptive wavelet estimation of a trend function. It turns out that under certain smoothing conditions on the trend function, the asymptotic rate of the mean integrated square error ($MISE$) of a trend estimator obtained by a hard wavelet thresholding is the same as for a linear wavelet estimator. Asymptotic expressions for the optimal $MISE$ and corresponding optimal smoothing and resolution parameters are derived. Furthermore, we focus on the non-continuous trend functions and derive corresponding optimal smoothing, resolution and thresholding parameters. Due to adaptation to the properties of the underlying trend function, the approach shows very good performance for piecewise smooth trend functions while remaining competitive with minimax wavelet estimation for functions with discontinuities. It turns out that the same expression for $MISE$ still holds and the hard thresholding wavelet estimator can be understood as a combination of two components, a smoothing component consisting of a certain number of lower resolution levels where no thresholding is applied, and a higher resolution component filtered by thresholding procedure. The first component leads to good performance for smooth functions, whereas the second component is useful for modeling discontinuities. This fact is used to develop an appropriate test for the null hypothesis that the trend is continuous against the alternative that it has at least one isolated jump. The proposed test statistic is based on blockwise resampling of estimated residual variances. Asymptotic validity of the test is derived. Simulations illustrate the asymptotic results and finite sample behavior of the proposed methods.

Zusammenfassung

In der vorliegenden Arbeit werden adaptive wavelet-basierte Methoden für nicht parametrische Regressionsmodelle mit langfristig abhängigen Innovationen untersucht. Zunächst beschäftigen wir uns mit der Frage der adaptiven wavelet-basierten Trendschätzung. Dabei stellt sich heraus, dass unter gewissen Glattheitsbedingungen an die Trendfunktion die optimale asymptotische Konvergenzrate des mittleren integrierten quadratischen Fehlers des Wavelet-Trendschätzers mit *Hard Thresholding* der eines linearen Waveletschätzers entspricht. Die asymptotischen Ergebnisse für den mittleren integrierten quadratischen Fehler, sowie die entsprechenden Glättungs- und Zerlegungsparameter werden hergeleitet. Darüber hinaus konzentrieren wir uns auf nichtstetige Trendfunktionen und untersuchen in diesem Zusammenhang die dazugehörigen Parameter. Aufgrund der Anpassung an die Trendfunktion zeigt unser Schätzer gute Resultate auch für stückweise glatte Funktionen und ist mit Minimax-Waveletschätzern sicherlich vergleichbar. Es stellte sich auch heraus, dass das asymptotische Ergebnis für den mittleren integrierten quadratischen Fehler unverändert bleibt. Außerdem kann der *Hard Thresholding*-Waveletschätzer als lineare Kombination einer *Smoothing*- und *High-Resolution*-Komponente dargestellt werden. Die *Smoothing*-Komponente setzt sich aus niederfrequenten Anteilen zusammen und beinhaltet keine Anwendung der Thresholding Methode. Dagegen besteht die *High-Resolution*- Komponente aus höheren Frequenzbändern, die mit Thresholding gefiltert sind. Die erste Komponente liefert eine sehr gute Schätzung der glatten Anteile der Trendfunktion, während die zweite überwiegend zur Schätzung der Sprünge geeignet ist. Diese Eigenschaft haben wir uns zunutze gemacht, um einen geeigneten Test zu entwerfen. Dabei wird die Stetigkeit der Trendfunktion gegen die Alternative, dass die Funktion mindestens einen isolierten Sprung besitzt, getestet. Die vorgeschlagene Teststatistik basiert auf *Blockwise Resampling* der Varianzen der geschätzten Residuen. Die asymptotische Konsistenz des Tests wird bewiesen. Eine empirische Untersuchung der vorgeschlagenen Methoden illustriert deren Verhalten für endliche Stichproben.

Contents

Chapter 1

Introduction

Often financial time series seem to be stationary and exhibit strong dependencies measured by the autocovariances. In this case one speaks of long memory (or long-range dependence). By long memory, we mean that the autocovariances of a second order stationary process are not absolutely summable. For instance, let

$$\gamma(k) = cov\left(\xi_i, \xi_{i+k}\right) \underset{k\to\infty}{\sim} C_\gamma |k|^{-\alpha} \tag{1.1}$$

with constants $C_\gamma > 0$ and $\alpha \in (0,1)$ where "$\sim$" means that the ratio of the two sides converges to one, then

$$\sum_{k\in\mathbb{Z}} \gamma(k) = \infty.$$

As illustrated in the extensive literature, long memory processes occur quite frequently in different fields like hydrology or finance. In addition, it is not unusual that data contain deterministic trends. In a nonparametric regression setting this is generally modeled by

$$Y_i = g(t_i) + \xi_i, \ (i = 1, 2, \ldots, n). \tag{1.2}$$

In this thesis we consider time series data of the form (1.2) with $t_i = i/n$, $g \in L^2\left([0,1]\right)$ and ξ_i a Gaussian zero mean second order stationary process with long-range dependence, i.e. with autocovariances $\gamma(k)$ as in (1.1) with $\alpha \in (0,1)$ and $C_\gamma > 0$. For the spectral density $f(\lambda) = (2\pi)^{-1} \sum \gamma(k) \exp(-ik\lambda)$ this corresponds to a pole at the origin of the form $C_f \left|\lambda\right|^{\alpha-1}$ for a suitable constant C_f (see e.g. Beran 1994, Zygmund 1953, Samorodnitsky 2007). The issue addressed in this thesis is optimal wavelet trend estimation and as an extension to this testing whether there is at least one jump in the function g.

Nonparametric estimation of g in this context has been studied extensively in the last two decades, including kernel smoothing (Hall and Hart 1990, Csörgö and Mielniczuk 1995, Ray and Tsay 1997, Robinson 1997, Beran and Feng 2002a,b), local polynomial estimation (Beran and Feng 2002c, Beran et al. 2002) and wavelet thresholding (Beran and Shumeyko 2011a, Wang 1996, Johnstone and Silverman 1997, Kulik and Raimondo 2009). For nonparametric quantile estimation in long-memory processes also see Ghosh et al. (1997) and Ghosh and Draghicescu (2002a,b). In this thesis we take a closer look at optimal wavelet estimation of g. Wang (1996) and Johnstone and Silverman (1997) derived optimal minimax rates within general function spaces and Gaussian long-memory residuals. In particular, the minimax threshold $\sigma_i\sqrt{2\log n}$ turns out to achieve the minimax rate even under long memory. For some practical applications, the minimax approach may however be too pessimistic. It may, for instance, be known a priori that g or some derivatives of g are piecewise continuous. Li and Xiao (2007) therefore considered data adaptive selection of resolution levels. They derived an asymptotic expansion for the mean integrated squared error (MISE) under the assumption that g is piecewise smooth, and the resolution levels used for the estimation are chosen according to certain asymptotic rules (formulated in terms of the parameters J and q, as defined below). The rate of the MISE achieved this way turns out to be the same as for minimax rules. No further justification for the specific choice of J and q is given however, and no optimality result is derived. We refer to Remark 11 below for further discussion on Li and Xiao (2007).

For literature on trend estimation by wavelet thresholding in the case of iid or weakly dependence residuals see e.g. Donoho and Johnstone (1994, 1995, 1998), Donoho et al. (1995), Daubechies (1992), Brillinger (1994, 1996), Abramovich et al. (1998), Nason (1996), Johnstone and Silverman (1997), Johnstone (1999), Percival and Walden (2000), Vidakovic (1999), Hall and Patil (1995, 1996a,b), Sachs and Macgibbon (2000) and Truong and Patil (2001). Apart from Johnstone and Silverman (1997) and Wang (1996), wavelet trend estimation in the long-memory case has also been considered by Yang (2001) for random design models. As we will see in chapter 4, a natural approach to testing continuity is provided by wavelet analysis, since wavelets are likely to pick up even local deviations from smoothness. More specifically, as discussed in chapter 4, a suitable wavelet estimator of $g(t)$ can be decomposed into a low resolution part $\hat{g}_{low}(t)$ characterizing

the smooth component of g and a detailed high-resolution part $\hat{g}_{high}(t)$ that may capture jumps. This fact is used to develop a bootstrap based test for the null hypothesis that g is continuous everywhere against the alternative that g has at least one isolated jump.

In the iid and weakly dependent setting, the question of detecting changes or structural breaks in a mean or a (parametric or nonparametric) regression function has been considered by many authors (see e.g. Quandt 1960, Hinkley 1969, Hawkins 1977, 1987, Sen 1980, Andrews 1993, 2003, Andrews and Ploberger 1994, Bai 1994, 1997, Hansen 1992, 2007, MacNeill 1978). For instance, change point detection in regression was considered by Quandt (1960). Hinkley (1996) derived an approximate distribution of a corresponding likelihood ratio statistic. For an overview of the extended literature see e.g. Hackl and Westlund (1991), Csörgö and Horváth (1998), Banerjee and Urgab (2005) and references therein. An interesting approach is taken in a recent paper by Ait-Sahalia and Jacod (2009) who define an index of jump activity in a short-memory context (also see references therein for related literature). In the case of long memory, it has been recognized that structural breaks can be confused with stationary long-memory components, and standard tests do not apply (Hidalgo and Robinson 1996, Horváth and Kokoszka 1997, Wright 1998, Horvath 2001, Delgado and Hidalgo 2000, Diebold and Inoue 2001, Krämer and Sibbertsen 2002, Ray and Tsay 2002, Granger and Hyung 2004, Mikosch and Stărică 2004, Lazarová 2005, Berkes et al. 2006, Choi and Zivot 2007, Gil-Alana 2008). In spite of this insight, only a few formal tests are available for the situation where the time point of the break is unknown (Wright 1998, Krämer and Sibbertsen 2002, Lazarová 2005; for an overview and more references see e.g. Sibbertsen 2004 and Banerjee and Urgab 2005). In particular, since sample paths of stationary processes with strong long memory tend to exhibit local spurious trends, very local changes in the mean function are difficult to detect. It is therefore very difficult to decide whether a completely unspecified (and thus nonparametric) trend function is continuous or only piecewise continuous. No general test for this hypothesis seems to be available in the long-memory setting. Due to their localization in time and frequency, wavelets provide a natural approach to solve this problem. This approach will be taken here.

The thesis is organized as follows:
Chapter 2 is a short introduction to long memory, limit theorems, $FARIMA$ models and bootstrap under long memory introduced by Lahiri (1993). The references to this chapter include Beran (1994), Doukhan et al. (2003), Lahiri (1997) and Schützner (2009).

Chapter 3 is an overview about wavelets and their use in statistics. This involves multiresolution analysis, the continuous and discrete wavelet transformation, construction of wavelet bases, shrinkage, thresholding, and some well-known results on wavelet based trend estimation under long memory given by Li and Xiao (2007), Donoho and Johnstone (1997) and Wang (1997). The references to wavelets and their use in statistics include Daubechies (1992), Vidakovic (1999), Hardle et al. (1998), Steeb (1998) and Heiler (2006).

The aim of chapter 4 is to obtain concrete data adaptive rules for optimal estimation of g. In a first step, it is shown that for functions with continuous derivatives, the rate given by Li and Xiao (2007) can be achieved without thresholding by choosing optimal values of J and q. In a second step, exact constants for the MISE and asymptotic formulas for the optimal choice of J and q are derived. This is comparable to results on optimal bandwidth selection in kernel smoothing (Gasser and Müller 1984, Hall and Hart 1990, Beran and Feng 2002a,c). In a third step, additional higher resolution levels combined with thresholding are added in order to include the possibility of discontinuities. The resulting estimator shows very good performance for smooth trend functions (comparable to optimal kernel estimators) while remaining competitive with (and even superior to) minimax wavelet estimation for functions with discontinuous derivatives.

In chapter 5, a bootstrap based test is developed for the null hypothesis that g is continuous everywhere against the alternative that g has at least one jump. In terms of the decomposition of $\hat{g}$ and g respectively, this amounts to testing the null hypothesis $H_0 : g_{high} \equiv 0$ against the alternative that $g_{high}(t)$ is not zero for at least one isolated point t. Properties of adaptive wavelet estimators of g_{low} and g_{high} are used to derive a test statistic W based on bootstraped blocks of $Y_i - \hat{g}_{low}(t_i)$. Critical values of W under H_0 are obtained by suitable blockwise resampling of $Y_i - \hat{g}_{low}(t_i) - \hat{g}_{high}(t_i)$. Previous references to bootstrap in the

long-memory setting include Lahiri (1993), Hall et al. (1998, 2000), Percival et al. (2000), Davidson (2002), Franco and Reisen (2004), Arteche and Orbe (2005), Lazarova (2005), Andrews et al. (2006), Silva et al. (2006) and Poskitt (2008).

The thesis concludes with some final remarks in chapter 6. Some `S-PLUS`-programs, that were used for simulations in sections 4.3 and 5.3, are given in the appendix.

Chapter 2

Basics on Long memory and Bootstrap

In this chapter, we cite some well-known results from Beran (1994), Doukhan et al. (2003) and Lahiri (1997). This will be used in later chapters of this thesis. The chapter is structured as follows. Standard notation is introduced in section 2.1. In section 2.2 we consider functionals of Gaussian processes and show that the asymptotic behavior of partial sums essentially depends on the connection to Hermite polynomials. After introducing the fractional autoregressive integrated moving average models in section 2.3, we consider in section 2.4 Lahiri's results on the moving block bootstrap under long range dependence (Lahiri 1997). For our propose we consider this result for the Gaussian process only. The result of Lahiri (1997) gives us a proof idea for the approach in chapter 5.
As already mentioned, for more details on long memory we refer to Beran (1994) and Doukhan et al. (2003).

2.1 Basic concepts

At first we define long memory processes:

Definition 2.1. *Let $(\xi_i)_{i\in\mathbb{Z}}$ be a stochastic process.*

(a) $(\xi_i)_{i\in\mathbb{Z}}$ is called strictly stationary, if for all $(i_1,\dots,i_k)^T \in \mathbb{Z}^k$, $k \geq 1$, the joint distributions of $(\xi_{i_1+t},\dots,\xi_{i_k+t})$ do not depend on $t \in \mathbb{Z}$.

(b) Define $\mu_i = E[\xi_i]$ and $\gamma_\xi(i,j) = cov(\xi_i,\xi_j)$. Then, $(\xi_i)_{i\in\mathbb{Z}}$ is called second-order stationary, if μ_i is constant for all i, j and $\gamma_\xi(i,j)$ only depends on

$|i-j|$. *In this case, the autocovariance function* γ *of* $(\xi_i)_{i\in\mathbb{Z}}$ *is denoted by*

$$\gamma(k) := \gamma_\xi(0,k).$$

(c) $(\xi_i)_{i\in\mathbb{Z}}$ *has long memory (or long range dependence), if*

$$\sum_{k\in\mathbb{Z}} |\gamma(k)| = \infty.$$

Otherwise, $(\xi_i)_{i\in\mathbb{Z}}$ *has short memory, if* $\left|\sum_{k\in\mathbb{Z}}\gamma(k)\right| > 0$ *or* $(\xi_i)_{i\in\mathbb{Z}}$ *is antipersistent, if* $\sum_{k\in\mathbb{Z}}\gamma(k) = 0$.

The following linear process is an example of a long-memory process. Let $\zeta_i, i = 1,2,\ldots$ be a zero mean iid sequence of random variables with finite variance and define coefficients

$$b_j = cj^{d-1}, \quad j = 1,2,\ldots,$$

where c is some positive constant and $d \in \left(0, \frac{1}{2}\right)$. Then, we can define the linear process

$$\xi_i = \sum_{j=1}^{\infty} b_j \zeta_{i-j}, \quad t \in \mathbb{Z}.$$

The autocovariances are given by

$$\begin{aligned} E[\xi_0\xi_k] &= \sum_{j=1}^{\infty} b_j b_{j+k} = c^2 \sum_{j=1}^{\infty} j^{d-1}(j+k)^{d-1} \\ &= k^{2d-1}c^2 \sum_{j=1}^{\infty} \left(\frac{j}{k}\right)^{d-1} \left(1+\frac{j}{k}\right)^{d-1} \frac{1}{k} =: k^{2d-1}\Lambda(k). \end{aligned}$$

Taking in account that the following truncated sum is a Riemann approximation for the corresponding integral we have

$$\sum_{j=1}^{[nk]} \left(\frac{j}{k}\right)^{d-1} \left(1+\frac{j}{k}\right)^{d-1} \frac{1}{k} \to \int_0^n x^{d-1}(1+x)^{d-1}dx, \quad \text{as} \quad k\to\infty,$$

where convergence is uniform in n. Thus

$$\begin{aligned} \lim_{t\to\infty}\lim_{n\to\infty} \sum_{j=1}^{[nk]} \left(\frac{j}{k}\right)^{d-1} \left(1+\frac{j}{k}\right)^{d-1} \frac{1}{k} &= \lim_{n\to\infty}\lim_{t\to\infty} \sum_{j=1}^{[nk]} \left(\frac{j}{k}\right)^{d-1} \left(1+\frac{j}{k}\right)^{d-1} \frac{1}{k} \\ &= \int_0^\infty x^{d-1}(1+x)^{d-1}dx < \infty. \end{aligned}$$

Here $L(\cdot)$ is a slowly varying function, in the sense that $L(\cdot)$ is positive with $\lim_{j\to\infty} \frac{L(aj)}{L(j)} = 1,\ \forall a > 0$. By Karamata's theorem

$$\sum_{k=1}^{n} \gamma(k) = \sum_{k=1}^{n} L(k)k^{2d-1} \sim L(n)\frac{1}{2d}n^{2d} \to \infty.$$

Thus, X_t has long memory in the sense of definition 2.1.
For simplicity of presentation in chapters 4 and 5 long-range dependence will be characterized (see (1.1)) by

$$\gamma(k) = E\left[\xi_i \xi_{i+k}\right] \underset{k\to\infty}{\sim} C_\gamma |k|^{-\alpha}$$

for some constants $\alpha \in (0,1)$ and $C_\gamma > 0$. For the spectral density $f(\lambda) = (2\pi)^{-1}\sum \gamma(k)\exp(-ik\lambda)$ this corresponds to a pole at the origin of the form $C_f\,|\lambda|^{\alpha-1}$ for a suitable constant C_f.

2.2 Limit theorems

The asymptotic behavior of partial sums $\sum_{i=1}^{n} Y_i$ for stationary processes Y_i with linear long-range dependence is well known (see e.g. Rosenblatt 1961, Taqqu 1975, 1979, Dobrushin and Major 1979, Surgailis 1981, 1982, Giraitis 1983, 1985, Giraitis and Surgailis 1985, 1986, 1999, Avram and Taqqu 1987, Dehling and Taqqu 1989, Arcones and Yu 1994, Ho and Hsing 1996, 1997). Here we will describe limit theorems for $Y_i = G(\xi_i)$, where ξ_i is Gaussian and G admits a Hermite expansion. Let $(\xi_i)_{i\in\mathbb{Z}}$ be a stationary Gaussian process with $E[\xi_i] = 0$, $E[\xi_i^2] = 1$ and covariances are such that

$$\gamma(k) = E[\xi_0\xi_k] \sim k^{-\alpha}L(k), \qquad k \to \infty \tag{2.1}$$

where $L(k)$ is a slowly varying function (in fact L may take negative values as well) and $0 < \alpha < 1$. For a function G with $E[G(\xi_0)] = 0$ and $E[G(\xi_0)^2] < \infty$, we derive the asymptotic behavior of $\sum_{i=1}^{[nu]} G(\xi_i)$.
Let us define Hermite polynomials as follows:

Definition 2.2. *For $q \geq 0$, the Hermite polynomials H_q are defined by*

$$H_q(x) = e^{x^2/2}\frac{d^q}{dx^q}e^{-x^2/2}.$$

Note that $(H_q)_{q\geq 0}$ constitutes a complete orthogonal system in the space $\{G : \mathbb{R} \to \mathbb{R} : E[G(\xi_0)] = 0, E[G^2(\xi_0)] < \infty\}$ (Abramowitz and Stegun 1972). Thus, the following expansion holds:

$$G(\xi_i) = \sum_{q=0}^{\infty} c_q H_q(\xi_i)$$

where $\sum_{m=0}^{\infty} c_q^2 q! < \infty$. Since the variance of $H_q(\xi_i)$ is given by $Var(H_q(\xi_i)) = q!(\gamma(0))^q$ (see (2.2)), the series above converges in L^2. Moreover, due to the orthogonality of Hermite polynomials the coefficients can be calculated by $c_q = \frac{1}{q!}E[G(\xi_0)H_q(\xi_0)]$. The lowest integer $q^* \in \mathbb{N}$ with $c_{q^*} \neq 0$ is then called the Hermite rank of H. Denote by $\gamma_G(i)$ the covariance function of $G(\xi_i)$ and note that the covariances of Hermite polynomials are given by

$$E[H_k(\xi_0)H_j(\xi_i)] = \delta_{kj} k!(\gamma(i))^k. \tag{2.2}$$

This implies

$$\begin{aligned} \gamma_G(i) &= \sum_{k,j=q^*}^{\infty} c_k c_j E\left[H_k(\xi_i)H_j(\xi_0)\right] \\ &= \gamma(i)^{q^*} \sum_{j=q^*}^{\infty} c_j^2 j! \gamma(i)^{j-q^*}. \end{aligned} \tag{2.3}$$

If $q^* > 1/\alpha$, the covariances of $H_{q^*}(\xi_i)$ are absolutely summable, since

$$\sum_{i=0}^{\infty} |\gamma(i)|^{q^*} = \sum_{i=0}^{\infty} i^{-q^*} L^{q^*}(i) < \infty.$$

and (due to (2.3))

$$\sum_{i=0}^{\infty} |\gamma_G(i)| \leq c \sum_{i=0}^{\infty} |\gamma(i)|^{q^*},$$

where $c \geq \sum_{j=q^*}^{\infty} c_j^2 j! \gamma(i)^{j-q^*}$, $\forall i \in \mathbb{Z}^+$ (notice that $\gamma(i) \to 0$). It is not difficult to show that this implies summability of the covariances of the process $G(\xi_i)$. On the other hand, (2.3) and $\gamma(i) \to 0$ imply

$$|\gamma_G(i)| \geq |\gamma(i)|^{q^*} c_{q^*}^2 q^*!/2,$$

for i large enough. Consequently, we have the relation

$$\sum_{i=0}^{\infty} |\gamma_G(i)| < \infty \qquad \Longleftrightarrow \qquad \sum_{i=0}^{\infty} |\gamma(i)|^{q^*} < \infty,$$

meaning that a function G with Hermite rank $q^* > 1/\alpha$ leads to short memory in the process $G(\xi_i)$. Indeed, Giraitis and Surgailis (1985) have proved the following theorem:

Theorem 2.1. *Let $(\xi_i)_{i\in\mathbb{Z}}$ be a zero mean Gaussian process for which (2.1) holds. If $\sum_{i=0}^{\infty} |\gamma_G(i)| < \infty$ (i.e. $q^* > 1/\alpha$) and $\sigma^2 := \sum_{i=0}^{\infty} \gamma_G(i) > 0$, then*

$$n^{-1/2} \sum_{i=1}^{[nu]} G(\xi_i) \xrightarrow{\mathcal{D}} \sigma B(u),$$

where $B(u)$ is a standard Brownian motion and "$\xrightarrow{\mathcal{D}}$"stand as usual for convergence in distribution.

We now consider the case $q^* < 1/\alpha$. At first, we derive the divergence order of the variance of $\sum_{i=1}^{n} H_q(\xi_i)$ in the case (1.1):

$$Var\left(\sum_{i=1}^{n} H_q(\xi_i)\right) = n \cdot Var(H_q(\xi_0)) + 2n \sum_{i=1}^{n-1} \left(1 - \frac{i}{n}\right) E[H_q(\xi_0)H_q(\xi_i)]$$

$$= n \cdot Var(H_q(\xi_0)) + 2n \sum_{i=1}^{n-1} \left(1 - \frac{i}{n}\right) q!\gamma(i)^q$$

$$= n \cdot Var(H_q(\xi_0)) + 2C_\gamma n \sum_{i=1}^{n-1} q! i^{-\alpha q} - 2C_\gamma n \sum_{i=1}^{n-1} \frac{i}{n} q! i^{-\alpha q}. \tag{2.4}$$

Note that

$$\sum_{i=0}^{n-1} i^{-\alpha q} = n^{1-\alpha q} \sum_{i=0}^{n-1} \left(\frac{i}{n}\right)^{-\alpha q} \frac{1}{n} \sim n^{1-\alpha q} \int_0^1 x^{-\alpha q} dx.$$

Similarly, we derive for the term in (2.4) that

$$\sum_{i=1}^{n-1} i^{1-\alpha q} \frac{1}{n} = n^{1-\alpha q} \sum_{i=1}^{n-1} \left(\frac{i}{n}\right)^{1-\alpha q} \frac{1}{n} \sim n^{1-\alpha q} \int_0^1 x^{1-\alpha q} dx.$$

Hence, by (2.4) and the last relations yields $\lim_{n\to\infty} Var\left(\sum_{i=1}^{n} H_q(\xi_i)\right)/n^{2-\alpha q} = const.$ Consequently, since $2 - \alpha q^* > 1$, the usual central limit theorem (with standard $\sqrt{n}$-scaling) do not hold for sums of $H_{q^*}(\xi_i)$. Indeed, a limiting distribution of $\sum_{i=1}^{[nu]} G(\xi_i)$ can be non-normal, as the next theorem, due to Taqqu (1979) and Dobrushin and Major (1979), states.

Theorem 2.2. *Let q^* be the Hermite rank of G and $(\xi_i)_{i\in\mathbb{Z}}$ a zero-mean Gaussian process for which (2.1) holds with $0 < \alpha < 1/q^*$, then*

$$\frac{1}{n^{1-q^*\alpha/2}L^{q^*/2}(n)}\sum_{i=1}^{[nu]} G(\xi_i) \xrightarrow{\mathcal{D}} c_{q^*}\mathcal{H}_{q^*}(u), \quad as \quad n\to\infty,$$

where c_{q^*} is the q^*-th coefficient in the Hermite expansion of G.

Here, $\mathcal{H}_k(u)$ denotes the Hermite process of order $k \geq 1$, defined by

$$\mathcal{H}_k(u) = C_q \int_{\mathbb{R}^k}' \int_0^u \left(\prod_{j=1}^k (s-y_j)^{-(1+\alpha)/2}\right) ds dB(y_1)\cdots dB(y_k)$$

where $B(\cdot)$ is a standard Brownian motion and $C_q > 0$ (see e.g. Taqqu 1979). $\mathcal{H}_1(u)$ is fractional Brownian motion, and thus Gaussian, while $\mathcal{H}_k(u)$ has non-normal marginal distributions for $k \geq 2$.

Theorem 2.2 can be understood as a reduction principle, in the sense that the asymptotic properties of $S_{n,G}(u)$ only depend on the Hermite rank q^* of G. In particular, the sum of $G(\xi_i)$ and the sum of $c_{q^*}H_{q^*}(\xi_i)$ have the same limiting distribution.

Theorem 2.2 plays an important role for the proof of the approach in section 2.4.

2.3 Fractional $ARIMA$ models

Box and Jenkins (1970) have introduced $ARIMA$ models. Due to the simplicity and flexibility, the $ARIMA$ models became popular in time series analysis and econometrics. The theory of statistical inference for these processes is well developed. Some of the references can be found, for example in Box and Jenkins (1970), Priestley (1981) and Brockwell and Davis (1987). This section summarizes results on the well known extension of classical $ARIMA$ models, namely the so called Fractional $ARIMA$ models.

Let B denote the backshift operator defined by $BX_t = X_{t-1}$, $B^2X_t = X_{t-2}$, ... In particular, differences can be expressed in terms of B as $X_t - X_{t-1} = (1 - B)X_t$, $(X_t - X_{t-1}) - (X_{t-1} - X_{t-2}) = (1-B)^2X_t, \ldots$.

Let us now recall the definition of $ARMA$ and $ARIMA$ processes. For simplicity of notation we assume $\mu = E[X_t] = 0$. Otherwise, X_t must be replaced by $X_t - \mu$.

Let p and q be some integers. Define

$$\phi(x) = 1 - \sum_{j=1}^p \phi_j x^j$$

and

$$\psi(x) = 1 - \sum_{j=1}^{q} \psi_j x^j.$$

In following, we suppose that all solutions of $\phi(x_0) = 0$ and $\psi(x_0) = 0$ are outside of the unit circle. The process $\{X_t\}$ is said to be an $ARMA(p, q)$ process if it is a stationary solution of

$$\phi(B)X_t = \psi(B)\epsilon_t \tag{2.5}$$

where $\epsilon_t\,(t = 1, 2, \dots)$ are iid normal variables with zero expectation and variance σ_ϵ^2.
$\{X_t\}$ is called $ARIMA(p, d, q)$ process if (2.5) holds for the dth difference $(1 - B)^d X_t$. The corresponding equation is

$$\phi(B)(1 - B)^d X_t = \psi(B)\epsilon_t. \tag{2.6}$$

Note that if d is larger than or equal to 1, then the original series X_t is not stationary but the dth difference of X_t is a stationary $ARMA$ process.
Fractional $ARIMA$ models are obtained by allowing d in (2.6) to be real valued. As the gamma function $\Gamma(\cdot)$ is also defined for all numbers, the definition of the binomial coefficient can be extended to all real numbers d. Thus, the (fractional) differencing operator can be written as

$$(1 - B)^d = \sum_{k=0}^{\infty} \frac{\Gamma(d + 1)}{\Gamma(k + 1)\Gamma(d - k + 1)}(-1)^k B^k.$$

If d is an integer, then this coincides with

$$(1 - B)^d = \sum_{k=0}^{d} \binom{d}{k}(-1)^k B^k.$$

For non-integer values of d, the summation is over an infinite number of indices. Due to the preceding representation we now extend the definition (2.6) to non-integer values of d in the following way:

Definition 2.3. *Let X_t be a stationary process such that*

$$\phi(B)(1 - B)^d X_t = \psi(B)\epsilon_t \tag{2.7}$$

for some $-0.5 < d < 0.5$. Then X_t is called a fractional $ARIMA(p, d, q)$ process.

This definition was proposed by Granger and Joyeux (1980) and Hosking (1981). In the context of this thesis we are interested in stationary long-memory processes. The corresponding values of d are in the interval $(0, 0.5)$. For $d \geq 0.5$ the corresponding process is not stationary and it is not invertible if $d \in \{0.5, 1.5, 2.5, \dots\}$. The parameter d determines the long-memory property, whereas p, q, and the corresponding parameters in $\phi(B)$ and $\psi(B)$ allow for modeling short-range behavior.

Note that equation (2.7) can be interpreted as

$$(1 - B)^d X_t = \tilde{X}_t$$

where $\tilde{X}_t$ is an $ARMA$ process defined by $\tilde{X}_t = \phi^{-1}(B)\psi(B)\epsilon_t$.

Denote by

$$f_{ARMA}(\lambda) = \frac{\sigma_\epsilon^2}{2\pi} \frac{|\psi(e^{i\lambda})|^2}{|\phi(e^{i\lambda})|^2}$$

the spectral density of the $ARMA$ process $\tilde{X}_t$. Recall that if X_t is obtained from a process Y_t with spectral density f_Y by applying the linear filter $\sum a(s)Y_{t-s}$ then the spectral density of X_t is equal to $|\sum a(s)e^{is\lambda}|^2 f_Y(\lambda)$ (see e.g. Priestley 1981, p. 266). Due to invertibility of $(1 - B)^d$ for all $d \in (0, 0.5)$, the spectral density of X_t is therefore

$$f(\lambda) = |1 - e^{i\lambda}|^{-2d} f_{ARMA}(\lambda).$$

At origin we have

$$f(\lambda) \sim \frac{\sigma_\epsilon^2}{2\pi} \frac{|\psi(1)|^2}{|\phi(1)|^2} |\lambda|^{-2d} = f_{ARMA}(0)|\lambda|^{-2d}.$$

Note that for nonnegative d the spectral density has a pole at zero.

The asymptotic formula for the covariance is given by

$$\gamma(k) \sim C_\gamma(d, \phi, \psi)|k|^{2d-1},$$

where

$$C_\gamma(d, \phi, \psi) = \frac{\sigma_\epsilon^2}{2\pi} \frac{|\psi(1)|^2}{|\phi(1)|^2} \Gamma(1 - 2d) \sin \pi d$$

and $d \in (0, 0.5)$ (see e.g. Beran 1994, Zygmund 1953, Samorodnitsky 2007).

2.4 Moving block bootstrap under long range dependence

Efron (1979) introduced the bootstrap method as a general tool for estimating the unknown sampling distribution of a statistic. For any statistical inference based on some sample we need the n dimensional distribution function, which can be calculated exactly by evaluating an n dimensional integral. In most cases, the calculation of high dimensional integrals is difficult. One of the options to approximate such integrals is Monte Carlo approximation, which means that we estimate the one dimensional distribution by a simulated distribution based on a sufficiently large sample. This is the main idea of the bootstrap approach. The bootstrap method provides very accurate approximations to distributions of many commonly used statistics in the case of iid observations. However, a similar statement does not necessarily hold for dependent observations. The iid resampling scheme associated with the method fails to capture the underlying dependence in the joint distribution of the observations and as a result, the classical bootstrap fails for dependent series, even in the simple case of the sample mean. Different modifications of this method for specific dependent models has been studied extensively in the last few years (see Bose 1988, Freedman 1984, Datta and McCormick 1992). Künsch (1989) and Liu and Singh (1992) have independently formulated a moving block bootstrap procedure, that provides valid approximation to distributions for weakly-dependent observations in general. However, Lahiri (1993) has shown that, under some additional conditions, the moving block bootstrap provides valid approximation to the distribution of the correctly normalized sample mean, for a class of long-range dependent observations if and only if the normalized sample mean is asymptotically normal. For simplicity of representation we consider only Gaussian zero mean noise.

Next we give a brief description of Lahiri's moving block bootstrap procedure. Let $\{\xi_i\}$ be a Gaussian zero mean second order stationary process with long-range dependence. For blocks of size l $(1 \leq l \leq n)$ define block sums $\zeta_1, \ldots, \zeta_{n-l+1}$ by

$$\zeta_i = \xi_i + \cdots + \xi_{i+l-1} \ \ (1 \leq i \leq n - l + 1).$$

Then draw $\zeta_1^*, \ldots, \zeta_k^*$ randomly with replacement from $\{\zeta_1, \ldots, \zeta_{n-l+1}\}$ and cal-

culate

$$T_{kl}^{*} = a_l^{-1} \left(k^{-1/2} \sum_{i=1}^{k} \zeta_i^{*} \right) \tag{2.8}$$

where $a_l = C_\gamma^{1/2} l^{1-\alpha/2}$.
The limit behavior of T_{kl}^{*}, is given by the following result.

Theorem 2.3. *Assume that $l = O(n^{1-\epsilon})$ for some $0 < \epsilon < 1$. If $k^{-1} + l^{-1} = o(1)$ and $\sigma^2 = Var(\xi_i)$, then*

$$\sup_{x \in \mathbb{R}} \left| P_n(T_{kl}^{*} \leq x) - \Phi\left(\tilde{\sigma}^{-1} x\right) \right| = o_p(1)$$

where

$$\tilde{\sigma}^2 = 2\sigma^2 (1-\alpha)^{-1} (2-\alpha)^{-1}.$$

The proof follows from three lemmas:

Lemma 2.1. *Assume that $l = O(n^{1-\epsilon})$ for some $0 < \epsilon < 1$, and $l^{-1} = o(1)$. Then $a_l^{-1} E_n\left[\zeta_i^{*}\right] = o_p(1)$.*

Lemma 2.2. *Assume that the conditions of theorem 2.3 hold and that $l = O(n^{1-\epsilon})$ for some $0 < \epsilon < 1$, and $l^{-1} = o(1)$. Let $\hat{\sigma}_n^2 = a_l^{-2} E_n(\zeta_i^{*})^2$. Then $\hat{\sigma}_n^2 = 2\sigma^2 (1-\alpha)^{-1}(2-\alpha)^{-1} + o_p(1)$.*

Lemma 2.3. *Let $Z_1, Z_2, \ldots, Z_n$ be n iid random variables with $E[Z_i] = 0$, and $E[Z_i^2] = 1$. Then, for any $\eta > 0$, and every $n \geq 1$,*

$$\sup_{x \in \mathbb{R}} \left| P\left(Z_1 + \cdots + Z_n \leq \sqrt{n} x\right) - \Phi(x) \right|$$

$$\leq C \left[\left(1 + \left| 1 - n^{-1} - \delta_n(1) \right|^{-1} \right) \delta_n(1) + (\eta + \delta_n(\eta)) \left| 1 - n^{-1} - \delta_n(1) \right|^{-3/2} \right]$$

where $\delta_n(x) = E\left[Z_1^2 I(|Z_1| > x\sqrt{n})\right], \forall x > 0$, and C is a constant.

See Lahiri (1993) for more details.

Proof. (of theorem 2.3) By lemma 2.3

$$\sup_{x \in \mathbb{R}} \left| P\left(\sum_{i=1}^{k} \zeta_i^{*} \leq \sqrt{k} a_l \hat{\sigma}_n x \right) - \Phi(x) \right|$$

$$\leq C \left[\left(1 + \left| 1 - k^{-1} - \hat{\delta}_n \right|^{-1} \right) \hat{\delta}_n + \left(k^{-1/4} + \hat{\delta}_n \right) \left| 1 - k^{-1} - \hat{\delta}_n \right|^{-3/2} \right] \tag{2.9}$$

where $\hat{\delta}_n \equiv (a_l\hat{\sigma})^{-2}E_n\left[(\zeta_1^*)^2\, I\left(|\zeta_1^*| > k^{1/4}a_l\hat{\sigma}\right)\right]$.
We now show that $\hat{\delta}_n \to 0$. By lemma 2.1 and 2.2 it follows that $\hat{\sigma}^2 = \tilde{\sigma}^2 + o_p(1)$ and $\hat{\delta}_n = (a_l\tilde{\sigma})^{-2}E_n\left[(\zeta_1^*)^2\, I\left(|\zeta_1^*| > k^{1/4}a_l\tilde{\sigma}\right)\right] + o_p(1)$. Hence

$$E\left[(a_l\tilde{\sigma})^{-2}E_n\left[(\zeta_1^*)^2\, I\left(|\zeta_1^*| > k^{1/4}a_l\tilde{\sigma}\right)\right]\right] = (a_l\tilde{\sigma})^{-2}E\left[(\zeta_1)^2\, I\left(|\zeta_1| > k^{1/4}a_l\tilde{\sigma}\right)\right] = o(1).$$

Consequently, $\hat{\delta}_n = o_p(1)$. Due to (2.9) we conclude the proof. □

Chapter 3

Wavelets

In this chapter we briefly recall well-known results from Daubechies (1992), Vidakovic (1999), Härdle et al. (1998), Steeb (1998), Donoho and Johnstone (1997), Wang (1997), Li and Xiao (2007) and Heiler (2006). Up to section 3.8 we summarise the main ideas of wavelets and multiresolution analysis as given in Daubechies (1992), Vidakovic (1999), Härdle et al. (1998), Steeb (1998) or Heiler (2006). We provide an overview of the basic concepts of wavelet analysis and briefly review current research ideas in nonlinear wavelet based trend estimation under long memory. This chapter is organized as follows. After introducing the continuous wavelet transform in section 3.1, the discretization of the continuous wavelet transform is discussed in section 3.2. In section 3.3 we introduce the concept of multiresolution analysis. Sections 3.4 and 3.5 handle the problem of constructing related mother and father wavelet functions. The Daubechie's wavelets are introduced in section 3.6 and some of their properties are discussed. In section 3.7 we focus on wavelet shrinkage and derive two different thresholds that are of particular interest in the following chapters. In section 3.8 wavelet based trend estimation and its rates of convergence are discussed for certain classes of trend functions and Gaussian noise with long memory. Finally, an overview of results derived by Donoho and Johnstone (1997), Wang (1997) and Li and Xiao (2007) is given with a specific focus on the long-memory case.

3.1 Continuous wavelet transformation

The first theoretical results on wavelets stated in the early 1980s, are connected with the continuous wavelet decomposition of L^2 functions. Articles of Morlet

et al. (1982) and Grossmann and Morlet (1985) are the primary source on this subject.

Let us start with some notations. We denote by $L^2(\mathbb{R})$ the Hilbert space of square integrable complex-valued functions on the real line. Commonly, this space is equipped with the scalar product

$$\langle f, g\rangle = \int_{-\infty}^{\infty} f(x)\,\overline{g(x)}\,dx\,, \forall f, g \in L^2(\mathbb{R})$$

where $\overline{g(x)}$ denotes complex conjugation of $g(x)$. This induces the norm

$$\|g\|_{L_2} := \langle g, g\rangle^{1/2} = \left(\int_{-\infty}^{\infty} |g(x)|^2\,dx\right)^{1/2}$$

which is finite for all $g \in L^2(\mathbb{R})$. One speaks of orthogonality of $f, g \in L^2(\mathbb{R})$ whenever $\langle f, g\rangle = 0$.

Most recent practical applications of wavelets especially in statistics mainly involve discrete wavelet transform. However, an understanding for continuous wavelet transformations is important, since many of their properties have analogous discrete counterparts. For example, there is a significant body of research in probability and time series analysis where the problems are formulated in terms of continuous wavelet transformations.

Let

$$\psi^*(\omega) = \int_{-\infty}^{\infty} \psi(x) e^{-ix\omega}\,dx$$

denote the Fourier transform of ψ. Assume $\psi_{a,b},\ a \in \mathbb{R}\backslash\{0\}, b \in \mathbb{R}$ be a family of functions defined by translation and dilation of a single function $\psi(x) \in L^2(\mathbb{R})$,

$$\psi_{a,b}(x) = \frac{1}{\sqrt{|a|}}\psi\left(\frac{x-b}{a}\right).$$

The factor $|a|^{-1/2}$ ensures that $\|\psi_{a,b}(x)\|_{L_2}$ is independent of a and b. The function ψ is called the mother wavelet and is assumed to fulfill the admissibility condition,

$$0 < C_\psi = 2\pi \int_{-\infty}^{\infty} \frac{|\psi^*(\omega)|^2}{|\omega|}\,d\omega < \infty.$$

The role of this condition will soon become clear. Note that the admissibility condition implies

$$0 = \psi^*(0) = \int_{-\infty}^{\infty} \psi(x)\,dx.$$

Furthermore, we assume that the wavelet functions are normalized, i.e.
$\|\psi_{a,b}(x)\|_{L_2} = 1$.
For any function $g(x) \in L^2\mathbb{R}$, the continuous wavelet transform is defined by

$$T_g(a,b) = \langle g, \psi_{a,b} \rangle = \int_{-\infty}^{\infty} g(x)\overline{\psi_{a,b}(x)}\, dx.$$

The parameters $a \in \mathbb{R}\backslash\{0\}$ and $b \in \mathbb{R}$ denote the dilation and translation parameters respectively. A function g can be recovered from its wavelet transform as follows

Lemma 3.1. *For all $f, g \in L^2(\mathbb{R})$ we have*

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{a^2} T_f(a,b)\overline{T_g(a,b)}\, da\, db = C_\psi \langle f, g \rangle. \tag{3.1}$$

Proof. (see Daubechies 1992, proposition 2.4.1, p. 24) Note that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{a^2} T_f(a,b)\overline{T_g(a,b)}\, da\, db$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{a^2} \left[\int_{-\infty}^{\infty} f^*(\omega)|a|^{1/2} e^{-ib\omega}\overline{\psi^*(a\omega)}\, d\omega\right]$$

$$\times \left[\int_{-\infty}^{\infty} \overline{g^*(\omega')}|a|^{1/2} e^{-ib\omega'}\psi^*(a\omega')\, d\omega'\right] da\, db$$

The expression between the first pair brackets can be viewed as $(2\pi)^{1/2}$ times the Fourier transform of $F_a(\omega) = |a|^{1/2} f^* \overline{\psi^*(a\omega)}$. The expression between the second pair brackets has a similar interpretation as $(2\pi)^{1/2}$ times complex conjugate of the Fourier transform of $G_a(\omega) = |a|^{1/2} g^*(\omega)\overline{\psi^*(a\omega)}$. By the unitarity of the Fourier transform we obtain

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{a^2} T_f(a,b)\overline{T_g(a,b)}\, da\, db$$

$$= 2\pi \int_{-\infty}^{\infty} \frac{1}{a^2} \int_{-\infty}^{\infty} F_a(\omega)\overline{G_a(\omega)}\, d\omega\, da$$

$$= 2\pi \int_{-\infty}^{\infty} \frac{1}{a} \int_{-\infty}^{\infty} f^*(\omega)\overline{g^*(\omega)}\, |\psi^*(a\omega)|^2\, d\omega\, da$$

$$= 2\pi \int_{-\infty}^{\infty} f^*(\omega)\overline{g^*(\omega)}\, d\omega \int_{-\infty}^{\infty} \frac{1}{a} |\psi^*(a\omega)|^2\, d\omega\, da = C_\psi \langle f, g \rangle$$

□

It is now clear why we imposed the admissibility condition: if C_ψ is infinite, then the resolution of the equality (3.1) would not hold.
Formula (3.1) can be rewritten as

$$g(x) = C_\psi \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{a^2} T_g(a,b)\psi_{a,b}(x)\, da\, db.$$

This is called synthesizing, or reconstructing, the function. The key property of wavelet transforms is their ability to decompose and perfectly reconstruct square-integrable functions.
As an example, we consider the continuous Haar transformation. Let

$$\psi_{a,b}^{Haar}(x) = \frac{1}{\sqrt{a}} \left[I\left(x \in \left[b, \frac{a}{2} + b\right]\right) - I\left(x \in \left[\frac{a}{2} + b, a + b\right]\right)\right],\ a \in \mathbb{R}^+,\ b \in \mathbb{R}.$$

Let G be a primitive for g, i.e., $G' = g$. Then,

$$T_g^{Haar}(a,b) = \langle g, \psi_{a,b}^{Haar} \rangle = \frac{2}{\sqrt{a}} \left[G\left(\frac{a}{2} + b\right) - \frac{G(b) + G(a+b)}{2} \right].$$

3.2 Discretization of the continuous wavelet transformation

The continuous wavelet transformation of a function of one variable is a function of two variables. It is clear, that this transformation is redundant. To minimize the transformation discrete values of a and b can be selected, and we still will have a invertible transformation. Nevertheless, the sampling of a and b that preserves all information about the decomposed function cannot be coarser than the critical sampling.
It is known (see e.g. Härdle et al. 1998), that the critical sampling is given by

$$a = 2^{-j}, b = k2^{-j},\ j, k \in \mathbb{Z}. \tag{3.2}$$

Any coarser sampling will not give a unique inverse transformation of the original function. There are other possible discretization choices (see e.g. Daubechies 1992 and Heil and Walnut 1989). Further, we focus on the discrete wavelet transforms for values given in (3.2). Next, we consider Mallat's multiresolution analysis, which provides an elegant theoretical framework for critically sampled wavelet transform.

3.3 Multiresolution analysis

Assume that $\phi \in L^2(\mathbb{R})$ is such that $\{\phi_{0k}, k \in \mathbb{Z}\} = \left\{N^{1/2}\phi(N \cdot -k), k \in \mathbb{Z}\right\}$ is an orthonormal system in $L^2(\mathbb{R})$ for every fixed positive number N. We denote by V_0 the subspace of $L^2(\mathbb{R})$ generated by $\{\phi_{0k}\}$. Thus, there exist coefficients $\alpha_k \in \mathbb{C}$, $\forall k \in \mathbb{Z}$ such that every $g \in V_0$ can be written as

$$g(x) = \sum_{k\in\mathbb{Z}} \alpha_k N^{1/2}\phi(Nx - k)$$

almost everywhere with $\alpha_k = \langle g, \phi_{0k}\rangle$, $\sum_k |\alpha_k|^2 < \infty$ and

$$\langle \phi_{0j}, \phi_{0k}\rangle := \delta_{j,k} = \begin{cases} 1 & , j = k, \\ 0 & , \text{else.} \end{cases} \tag{3.3}$$

We now define integer translations and dyadic dilations of the scaling function as follows

$$\phi_{jk}(x) := 2^{j/2}N^{1/2}\phi(N2^j x - k),\ j, k \in \mathbb{Z}. \tag{3.4}$$

Due to this representation we define a set of functions $\{\phi_{jk}\}$.
The sequence of subspaces $\{V_j,\ j \in \mathbb{Z}\}$ is defined by

$$V_j = \left\{g(2^j x),\ :\ g \in V_0\right\},\ j \in \mathbb{Z}. \tag{3.5}$$

Furthermore the $\{\phi_{jk},\ k \in \mathbb{Z}\}$ constitutes a basis system of V_j and we can say that ϕ generates the sequence $\{V_j,\ j \in \mathbb{Z}\}$.
Taking account of previous notations we can introduce the underlying concept of multiresolution analysis:

Definition 3.1. *If the function ϕ is chosen such that V_j, $j \in \mathbb{Z}$ in (3.5) satisfies*

1. $\ldots V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset \ldots$,
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 of $L^2(\mathbb{R})$.

Definition 3.2. *If a multiresolution analysis of $L^2(\mathbb{R})$ is generated by a function ϕ, then ϕ is called a scaling function or a father wavelet.*

Note, that the resulting system of functions in (3.4) is in general not orthonormal. For the aim to construct an orthonormal basis we now consider the orthogonal complement W_j of V_j in V_{j+1}. Thus,

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

For $J \in \mathbb{Z}$ and for each fixed $j > J$ the subspace V_j can be represented as follows

$$V_j = V_{j-1} \oplus W_{j-1} = V_{j-2} \oplus W_{j-2} \oplus W_{j-1} = \cdots = V_J \oplus \bigoplus_{k=1}^{j-J} W_{j-k}.$$

Due to the definition 3.1 the union over all nested subspaces V_j is dense in $L^2(\mathbb{R})$. Hence, we can decompose $L^2(\mathbb{R})$ into the combination of orthogonal subspaces:

$$L^2(\mathbb{R}) = V_J \oplus \bigoplus_{k=1}^{\infty} W_{J-k}.$$

Thus, any $g \in L^2(\mathbb{R})$ can be represented by

$$g(x) = \sum_{k\in\mathbb{Z}} s_{Jk}\phi_{Jk}(x) + \sum_{j\geq 0}\sum_{k\in\mathbb{Z}} d_{jk}\psi_{jk}(x) \tag{3.6}$$

where $\{\psi_{jk}\}$, $k \in \mathbb{Z}$ denotes an orthonormal basis in W_{J+j} and s_{Jk}, d_{jk} are the corresponding coefficients given by

$$s_{Jk} := \langle g, \phi_{Jk}\rangle = \int g(x)\,\overline{\phi_{Jk}(x)}\,dx,$$

$$d_{jk} := \langle g, \psi_{jk}\rangle = \int g(x)\,\overline{\psi_{jk}(x)}\,dx.$$

Which fulfill

$$\sum_{k\in\mathbb{Z}} |s_{Jk}|^2 + \sum_{j\geq 0, k\in\mathbb{Z}} |d_{jk}|^2 < \infty.$$

To call (3.6) a wavelet expansion of g we have to introduce the following conditions: suppose there exist a function $\psi \in L^2(\mathbb{R})$ that each ψ_{jk} is of the form

$$\psi_{jk} := 2^{(J+j)/2} N^{1/2} \psi(2^{(J+j)} Nx - k),\ j, k \in \mathbb{Z}. \tag{3.7}$$

Definition 3.3. *A function $\psi \in L^2(\mathbb{R})$ generating the family $\{\psi_{jk}\}$ in (3.7) is called a mother wavelet.*

Definition 3.4. *Representation (3.6) with a basis of the form $\{\phi_{Jk},\ \psi_{jk},\ J \in \mathbb{Z},\ j \geq 0,\ k \in \mathbb{Z}\}$ is called wavelet series expansion of g. The corresponding coefficients s_{Jk} and d_{jk} are called wavelet coefficients of g.*

Notice that the mother wavelet ψ associated with ϕ, is not necessarily unique. In next section we introduce general conditions for existence and construction of ϕ and ψ. In (3.6) is shown that the wavelet series expansion represents functions by using linear combination, starting on a coarse level J, of the translated and dilated father and mother wavelet.

3.4 Construction of a mother wavelet with given father wavelet

In the previous section, we have described theoretically a pair of orthonormal wavelet functions, which serves as a basic component in the multiresolution analysis. Whereas in this section we construct the mother wavelet with a given father. Daubechies (1992), Härdle et al. (1998), Pinsky (2002), Strichartz (1994) and Vidakovic (1999) provide us with more information on this topic.
Let $\phi \in L^2(\mathbb{R})$ be scaling function with compact support. In chapters 4 and 5 we denote by N the support length of ϕ and ψ. In the current section, we assume without lost of generality that $N = 1$. Following this assumption the next lemma discuss the orthonormality of the integer translates of ϕ.

Lemma 3.2. *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*

$$\sum_{k\in\mathbb{Z}} |\phi^*(\omega + 2\pi k)|^2 = 1 \tag{3.8}$$

almost everywhere.

Proof. (see Vidakovich 1999, p. 55) By the unitarity of the Fourier transform and the 2π-periodicity of $e^{i\omega l}$ one has

$$\begin{aligned}
\delta_l &= \int_{-\infty}^{\infty} \phi(x)\overline{\phi(x-l)}\, dx \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} \phi^*(\omega)\overline{\phi^*(\omega)}e^{i\omega l}\, d\omega \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} \sum_{k\in\mathbb{Z}} |\phi^*(\omega + 2\pi k)|^2 e^{i\omega l}\, d\omega.
\end{aligned}$$

The last line is the Fourier coefficient in the Fourier series decomposition of $f(\omega) = \sum_{k\in\mathbb{Z}} |\phi^*(\omega+2\pi k)|^2$. Due to the uniqueness of the Fourier representation we conclude that $f(\omega) = 1$.

□

As in the last section we suppose that V_j defined by (3.5) and this subspaces are nested, then the scaling function can be represented as

$$\begin{aligned}\phi(x) &= \sum_{k\in\mathbb{Z}} s_{1k}\phi_{1k}(x) \\ &= \sum_{k\in\mathbb{Z}} s_{1k}2^{1/2}\phi(2x-k)\end{aligned}$$

where the family $\{\phi_{1k},\ k\in\mathbb{Z}\}$ is an orthonormal basis in V_1 with $s_{1k} = \langle\phi,\phi_{1k}\rangle$. The Fourier transform is then given by

$$\begin{aligned}\phi^*(\omega) &= \int \phi(x)e^{-ix\omega}\,dx \\ &= \frac{1}{\sqrt{2}}\sum_{k\in\mathbb{Z}} s_{1k}e^{-i\frac{\omega}{2}k}\phi^*\left(\frac{\omega}{2}\right) \\ &= m_0\left(\frac{\omega}{2}\right)\phi^*\left(\frac{\omega}{2}\right)\end{aligned}$$

where m_0 is a trigonometric polynomial with coefficients s_{1k}. Thus,

$$m_0(\omega) = \frac{1}{\sqrt{2}}\sum_{k\in\mathbb{Z}} s_{1k}e^{-i\omega k}. \tag{3.9}$$

This is the preposition for

Lemma 3.3. *Let $\{\phi_{0k}\}$ be an orthonormal system. Every 2π-periodic function m_0 satisfying*

$$\phi^*(\omega) = m_0\left(\frac{\omega}{2}\right)\phi^*\left(\frac{\omega}{2}\right) \tag{3.10}$$

almost everywhere such that $m_0 \in L_2([0,2\pi])$, also satisfies

$$|m_0(\omega)|^2 + |m_0(\omega+\pi)|^2 = 1 \tag{3.11}$$

almost everywhere.

Proof. (see Härdle et al. 1998, lemma 5.2, p. 37) By (3.10)

$$\left|\phi^*(2\omega+2\pi k)\right|^2 = \left|m_0(\omega+\pi k)\right|^2\left|\phi^*(\omega+\pi k)\right|^2.$$

Summing up in k and using the fact that $\{\phi_{0k}\}$ is an orthonormal system and m_0 is 2π-periodic we get by lemma 3.2 the following equality almost everywhere

$$\begin{aligned}
1 &= \sum_{k\in\mathbb{Z}} |m_0(\omega+\pi k)|^2 \, |\phi^*(\omega+\pi k)|^2 \\
&= \sum_{k\in\mathbb{Z}} |m_0(\omega+2\pi k)|^2 \, |\phi^*(\omega+2\pi k)|^2 \\
&+ \sum_{k\in\mathbb{Z}} |m_0(\omega+2\pi k+\pi)|^2 \, |\phi^*(\omega+2\pi k+\pi)|^2 \\
&= \sum_{k\in\mathbb{Z}} |m_0(\omega)|^2 \, |\phi^*(\omega+2\pi k)|^2 + \sum_{k\in\mathbb{Z}} |m_0(\omega+\pi)|^2 \, |\phi^*(\omega+2\pi k+\pi)|^2 \\
&= |m_0(\omega)|^2 + |m_0(\omega+\pi)|^2.
\end{aligned}$$

□

Now we are able to derive a necessary and sufficient condition for nesting of V_j in V_{j+1} (see e.g. Härdle 1998).

Lemma 3.4. *A necessary and sufficient condition for $V_j \subset V_{j+1}$, $j \in \mathbb{Z}$, according to definition 3.1 is that there exists a 2π-periodic function $m_0 \in L^2([0,2\pi])$ such that (3.10) holds almost everywhere.*

Proof. (see Härdle et al. 1998, lemma 5.1, p. 36) Note that necessary condition is already proven. Now, prove that (3.10) is a sufficient condition. A consequence of lemma 3.3 is that such a function m_0 is bounded. It is clear if we denote by V_0^* and respectively V_1^* the sets of Fourier transforms of the functions of V_0 and of V_1 we have

$$V_0^* = \{m(\omega)\phi^*(\omega) \mid m(\omega) \text{ is } 2\pi \text{ periodic}, \ m \in L_2([0,2\pi])\},$$

and

$$V_1^* = \{m(\omega/2)\phi^*(\omega/2) \mid m(\omega) \text{ is } 2\pi \text{ periodic}, \ m \in L_2([0,2\pi])\}.$$

Condition (3.10) implies that every function in V_0^* has the form $m(\omega)m_0(\omega/2)\phi^*(\omega/2)$ and belongs to V_1^*. In fact $m(2\omega)m_0(\omega)$ is a 2π-periodic function belonging to $L_2([0,2\pi])$ since $m \in L_2([0,2\pi])$, and m_0 is bounded due to previous lemma. □

Remark 1. *If $\phi \in L^2(\mathbb{R})$ has a compact support, then $\phi \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ and ϕ^* is continuous. Furthermore, m_0 is a trigonometric polynomial with a finite amount of coefficients s_{1k}. Thus m_0 is continuous as well and (3.10) holds for all ω.*

Due to the relation (3.3) we imply $m_0(0) = 1$. Equation (3.11) will be used further to construct a scaling function ϕ. Notice that the function m_0 is not necessarily unique. Recall that if $\psi \in W_0$ is a mother wavelet then the family $\{\psi(\cdot - k), k \in \mathbb{Z}\}$ is an orthonormal basis in W_0. The idea how we can define the mother wavelet ψ, corresponding to the given polynomial m_0, is derived in following lemma:

Lemma 3.5. *Let the assumption of lemma 3.4 hold. Define* $m_1(\omega) = -\overline{m_0(\omega + \pi)}e^{-i\omega}$ *and set*

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

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

For details see Härdle et al. (1998), lemma 5.3, p. 38. The same in more general form

Remark 2. *The statement of lemma 3.5 is fulfilled if we choose*

$$m_1(\omega) = \theta(\omega)\overline{m_0(\omega + \pi)}e^{-i\omega},$$

where $\theta(\omega)$ *is an arbitrary* π*-periodic function such that* $|\theta(\omega)|$*=1.*

Lemmas 3.2 and 3.4 introduce methods to verify whether ϕ can be chosen as a father wavelet. Furthermore, lemma 3.5 and remark 2 provide a way to construct a corresponding mother wavelet. It remains to show a such conditions that 2. and 3. of definition 3.1 are satisfied. Daubechies (1992, p. 141) shows that for all $\phi \in L^2(\mathbb{R})$ satisfying (3.8) the resulting sequence of subspaces has trivial intersection. Moreover, it can be shown (see Härdle et al. (1998, p. 95)) that a mild condition of the integrability of ϕ fulfill the 2. of definition 3.1. Thus,

Lemma 3.6. *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 (3.8) and (3.10). Then, if* $|\phi(u)| \leq \Phi(|u|)$ *almost everywhere,*

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

3.5 Construction of a compactly supported father wavelet

Until now we have assumed that an appropriate father wavelet is given and we derived previous lemmas and remarks under this assumption. The aim of this section is to construct a suitable scaling function. First, recall the previously described properties of ϕ:

1. $\sum_k |\phi^*(\omega + 2\pi k)|^2 = 1$ almost everywhere,
2. $\phi^*(\omega) = m_0\left(\frac{\omega}{2}\right)\phi^*\left(\frac{\omega}{2}\right)$ almost everywhere, with m_0 defined in lemma 3.4,
3. $|m_0(\omega)|^2 + |m_0(\omega+\pi)|^2 = 1$ and $m_0(0) = 1$.

As a consequence,

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

such that (see lemma 3.5)

$$\int \psi(x)\,dx = 0. \tag{3.12}$$

Remark 3. *In many applied articles and textbooks (3.12) is used as a defining property of a mother wavelet (see e.g. Percival and Walden 2000, Bruce and Gao 1996).*

According to lemma 3.4 the sequence of subspaces V_j is nested if and only if

$$\begin{aligned}\phi^*(x) &= m_0\left(\frac{\omega}{2}\right)\phi^*\left(\frac{\omega}{2}\right)\\ &= m_0\left(\frac{\omega}{2}\right)m_0\left(\frac{\omega}{4}\right)\phi^*\left(\frac{\omega}{4}\right) = \cdots = \prod_{j=1}^{\infty} m_0\left(\frac{\omega}{2^j}\right)\end{aligned} \tag{3.13}$$

with $\phi^*(0) = 1$. If (3.13) properly defines a function $\phi \in L^2(\mathbb{R})$, which generates a basis in $L^2(\mathbb{R})$, then we receive a multiresolution analysis and corresponding orthonormal basis function. For this propose we now derive the pointwise convergence of (3.13).

Lemma 3.7. *Let $m_0(\omega)$ satisfy condition 3. above with $m_0(\omega)$ being Lipschitz continuous. Then,*

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

converges uniformly to $\phi^*(x)$ on every compact subset of $\mathbb{R}$.

Proof. (see Härdle et al. 1998, lemma 6.2, p. 55) Since $m_0(0) = 1$, then for all $|\omega| \leq K$ yields

$$\prod_{j=0}^{\infty} m_0\left(\frac{\omega}{2^j}\right) = \prod_{j=0}^{\infty}\left(1 + m_0\left(\frac{\omega}{2^j}\right) - m_0(0)\right)$$

$$\leq \prod_{j=0}^{\infty}\left(1 + \left|m_0\left(\frac{\omega}{2^j}\right) - m_0(0)\right|\right) \leq \prod_{j=0}^{\infty}\left(1 + \frac{LK}{2^j}\right)$$

where L is the Lipschitz constant. Hence, the infinite product converges uniformly on every compact set. □

Remark 4. *Assume that m_0 is of the form (3.9) with a finite number of nonzero coefficients. Then, lemma 3.7. surely holds. Daubechies (1992) mentions a generalization to infinite sets of coefficients in case that the decay rate is sufficiently fast.*

The next lemma is firstly derived by Mallat (1989). It shows that the limiting function in (3.13) together with its Fourier transform is in $L^2(\mathbb{R})$ space.

Lemma 3.8. *Let m_0 satisfy the assumptions of lemma 3.7, then $\phi \in L^2(\mathbb{R})$ with $\|\phi\|_{L_2} \leq 1$.*

Moreover, we have also to show that so constructed scaling function generates an orthonormal basis in $L^2(\mathbb{R})$. This question addressed to the next theorem, which is originally introduced by Cohen and Ryan (1995), and can be found, for example, in Härdle et al. (1998, p. 56). This result is common for polynomials m_0 with associated compactly supported wavelet function, which is the most popular case in applied sciences.

Theorem 3.1. *Assume that the function $m_0(\omega)$ is a trigonometric polynomial of the form*

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

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

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

Further assume that condition 3. above is satisfied and there exists a compact set $\mathfrak{K}$ in $\mathbb{R}$, containing a neighborhood of 0, such that

1. $\sum_k I\left(\omega + 2\pi \in \mathfrak{K}\right) = 1$ *almost everywhere,*

2. $m_0(2^{-j\omega}) \neq 0$, *for all* $\omega \in K$ *and for all* $j \geq 1$.

Then, the function $\phi^*(\omega)$ *in (3.13) is the Fourier transform of a function* $\phi \in L^2(\mathbb{R})$ *such that*

(i) $\operatorname{supp}\phi \subseteq [N_0, N_1]$,

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

Notice that every solution ϕ, defined by theorem 3.1, has a compact support. This is a huge computational advantage in contrast to the basis defined on the real line. Moreover, due to the choice of the coefficient s_{1k} the father wavelet ϕ and/or the mother wavelet ψ could have a prescribed number of vanishing moments and/or have prescribed number of continuous derivatives. The following remarks (see e.g. Härdle et al. 1998, lemma 6.4 and 6.5,p. 57) describe how exactly it could happen.

Remark 5. *Let the conditions of theorem 3.1 be satisfied, and let*

$$\sum_{k=N_0}^{N_1} s_{1k} k^l = 0,\ l = 1, \ldots, n.$$

Then for ϕ *defined as the inverse Fourier transform of (3.13) we have*

$$\int_{-\infty}^{\infty} \phi(x) x^l \, dx = 0,\ l = 1, \ldots, n.$$

Remark 6. *Let the conditions of theorem 3.1 be satisfied. Then* $\psi \in L_2(\mathbb{R})$, ψ *is compactly supported, and yields*

$$\operatorname{supp}\psi \subseteq \left[\frac{1}{2}(1 - N_1 + N_0), \frac{1}{2}(1 - N_0 + N_1)\right].$$

If, in addition,

$$\sum_{k=1-N_1}^{1-N_0} (-1)^k \overline{s_{1k}} (1-k)^l = 0,\ l = 1, \ldots, n, \tag{3.14}$$

then

$$\int \psi(x) x^l \, dx = 0,\ l = 1, \ldots, n.$$

It is clear, that (3.14) can be satisfied only if $n+1$ is smaller than the degree of the polynomial $m_0(\omega)$, since (3.14) contains n equalities.
In this section we developed an approach for the construction of the compactly supported wavelets. For further information on the regularity of the wavelet basis or construction of a multiresolution analysis starting from a Riesz basis we refer to the monographs of Daubechie (1992) and Härdle et al. (1998).

3.6 Daubechies wavelets

The original construction of compactly supported wavelets is due to Daubechies (1988). In this section we briefly review her concept. However, there are also other concepts such as "Coiflets" or "Symmlets", for more details see, for example, Härdle et al. (1998). Here we sketch the main points of Daubechies theory. We are interested in defining the exact form of functions $m_0(\omega)$, which are trigonometric polynomials, with associated compact supported father and mother wavelets. Moreover, the moments of ψ up to order n should vanish. As you will see later this property is necessary to guarantee good estimation properties of corresponding wavelet expansion.
An immediate consequence of (3.14) is the following

Lemma 3.9. *Let the conditions of theorem 3.1 and (3.14) are satisfied. Then* $m_0(\omega)$ *factorizes as*

$$m_0(\omega) = \left(\frac{1+e^{-i\omega}}{2}\right)^{n+1} \mathfrak{L}(\omega)$$

where $\mathfrak{L}(\omega)$ *is a trigonometric polynomial.*

Proof. (see Härdle et al. 1998, corollary 7.1, p. 59) The relation (3.14) implies

$$m_1^{(l)}(0) = 0,\ l = 1, \ldots, n$$

which in view of the definition of $m_1(\omega)$ is equivalent to

$$m_0^{(l)}(\pi) = 0,\ l = 1, \ldots, n.$$

Also $m_0(\pi) = 0$. Hence $m_0(\omega)$ has a root of order $n+1$ at $\omega = \pi$. This mean

$$m_0(\omega) = \left(\frac{1+e^{-i\omega}}{2}\right)^{n+1} \mathfrak{L}(\omega).$$

Since m_0 is a trigonometric polynomial, then $\mathfrak{L}(\omega)$ is also a trigonometric polynomial. □

Let

$$M_0(\omega) = |m_0(\omega)|^2 = \left(\cos^2 \frac{\omega}{2}\right)^N |\mathfrak{L}(\omega)|^2$$

where $|\mathfrak{L}(\omega)|^2$ is a polynomial in $\cos\omega$, then the orthogonality condition (3.11) leads to

$$M_0(\omega) + M_0(\omega + \pi) = 1.$$

$|\mathfrak{L}(\omega)|^2$ can be rewritten as a polynomial in $\sin^2 \frac{\omega}{2}$ since $\cos\omega = 1 - 2\sin^2 \frac{\omega}{2}$. We then denote $|\mathfrak{L}(\omega)|^2$ by $P(\sin^2 \frac{\omega}{2})$. Then the condition (3.11) becomes

$$(1-y)^N P(y) + y^N P(1-y) = 1,\ \forall y \in [0,1].$$

Due to Daubechies (1992) the unique solution of last equation with degree $\deg P(y) \leq N-1$ is

$$P(y) = \sum_{k=0}^{N-1} \binom{N+k-1}{k} y^k + y^N R(1/2 - y)$$

where $R(\cdot)$ is an odd polynomial with $R(y) \geq 0,\ \forall y \in [0,1]$.
The propose of Daubechies (1988) is to take $R \equiv 0$, and she showed that under this restriction

$$|m_0(\omega)|^2 = C_N \int_\omega^\pi \sin^{2N-1} x\, dx \tag{3.15}$$

where C_N fulfills $m_0(0) = 1$.

Definition 3.5. *Wavelets fulfilling (3.15) called Daubechies wavelets and will be denoted by d2N.*

For Daubechies wavelets one can calculate corresponding coefficients $\{s_{1k}\}$. You can found them in table 3.2. Table 3.1 also provides the compact overview of the Daubechies wavelets and their properties. Notice, that the Hölder exponent there measures the smoothness of a function and is defined by $\alpha = m + \beta$, where α and β are the largest values, satisfying

$$|f^{(m)}(x) - f^{(m)}(x+t)| \leq C|t|^\beta\ , \forall x, t.$$

The figure 3.1 represents some of Daubechies wavelets.
As example, we consider the case $N = 1$. Due to (3.15)

$$|m_0(\omega)|^2 = C_N \int_\omega^\pi \sin(x)\, dx = C_N(1 + \cos(\omega)).$$

Taking $\omega = 0$ we obtain $C_N = 1/2$. Solving last equation we achive $m_0(\omega) = \frac{1}{2}(1 + e^{-i\omega})$. Taking account on (3.13), we compute function ϕ^*

$$\begin{aligned}\phi^*(\omega) &= \lim_{n\to\infty}\prod_{j=1}^{n}\frac{1}{2}\left(1+e^{-i\omega 2^{-j}}\right) = \lim_{n\to\infty}\prod_{j=1}^{n}\left(\frac{1-e^{-i\omega 2^{-j+1}}}{2(1-e^{-i\omega 2^{-j}})}\right)\\ &= \lim_{n\to\infty}\frac{1}{2^n}\left(\frac{1-e^{-i\omega}}{1-e^{-i\omega 2^{-n}}}\right) = \frac{1-e^{-i\omega}}{i\omega}.\end{aligned}$$

Then it yields

$$\phi(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i\omega x}\phi^*(\omega)\,d\omega = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i\omega x}\frac{1-e^{-i\omega}}{i\omega}\,d\omega = I(x\in[0,1]).$$

Taking account on lemma 3.5, we achieve

$$\begin{aligned}\psi^*(\omega) &= m_1\left(\frac{\omega}{2}\right)\phi^*\left(\frac{\omega}{2}\right) = -\overline{m_0\left(\frac{\omega}{2}+\pi\right)}e^{-i\omega/2}\phi^*\left(\frac{\omega}{2}\right)\\ &= -\frac{1}{2}\overline{(1+e^{-i(\pi+\omega/2)})}e^{-i\omega/2}2\frac{1-e^{-i\omega/2}}{i\omega} = \frac{(1-e^{-i\omega/2})^2}{i\omega}.\end{aligned}$$

Finally, by the inverse Fourier transform we derive

$$\psi(x) = I(x\in[0,1/2]) - I(x\in(1/2,1]).$$

Thus, the Daubechies $d2$ wavelet coincides with the Haar wavelet.

3.7 Non-linear wavelet shrinkage: thresholding

Let us first recall (3.6)

$$g(x) = \sum_{k\in\mathbb{Z}} s_{Jk}\phi_{Jk}(x) + \sum_{j\geq 0}\sum_{k\in\mathbb{Z}} d_{jk}\psi_{jk}(x)$$

for any $g \in L^2(\mathbb{R})$. The idea of the linear wavelet estimation is simple: we replace the unknown coefficients $\{s_{Jk}\}$, $\{d_{jk}\}$ in wavelet expansion (3.6) by estimates

$$\hat{s}_{Jk} = \frac{1}{n}\sum_{i=1}^{n} Y_i\phi_{Jk}(t_i),\quad \hat{d}_{jk} = \frac{1}{n}\sum_{i=1}^{n} Y_i\psi_{jk}(t_i), \tag{3.16}$$

which are based on the observed data $\{Y_i\}$. This will require a truncation of the infinite series in (3.6), since we can only deal with a finite number of coefficients and this leads to

$$\hat{g}(t) = \sum_{k\in\mathbb{Z}} \hat{s}_{Jk}\phi_{Jk}(t) + \sum_{j=0}^{q}\sum_{k\in\mathbb{Z}} \hat{d}_{jk}\psi_{jk}(t)$$

Figure 3.1: Daubechie's orthogonal wavelets: haar, d4, d6, d8 and d10. Left: father wavelets. Right: mother wavelets.

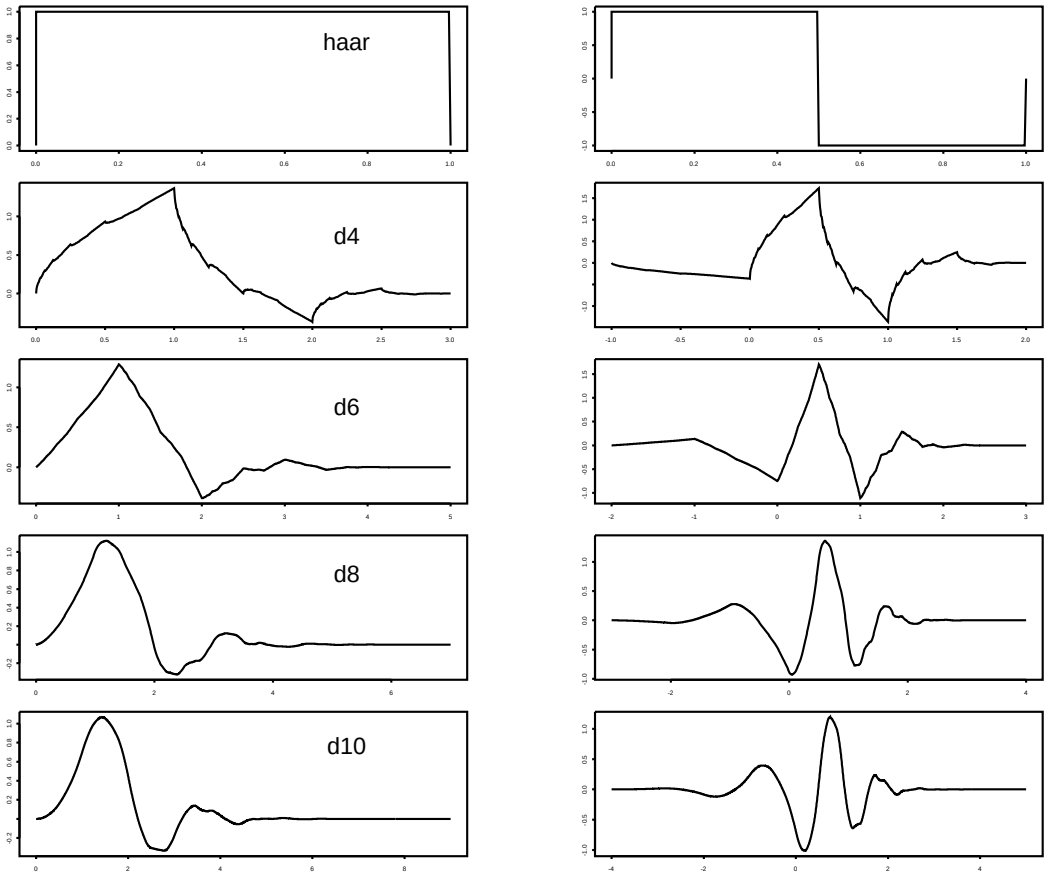


Table 3.1: Table of Daubechie's wavelets and their properties.

wavelet	support length	vanishing moments	number of derivatives	Hölder exponent
haar	1	1	0	0
d4	3	2	0	0.55
d6	5	3	1	1.09
d8	7	4	1	1.62
d10	9	5	1	1.97
d12	11	6	2	2.19
d14	13	7	2	2.46
d16	15	8	2	2.76
d18	17	9	3	3.07
d20	19	10	3	3.38

where J, q denote the decomposition level and smoothing parameter correspondingly. In case of $L^2([0,1])$ we define associated estimator by

$$\hat{g}(t) = \sum_{k=-N+1}^{N2^J-1} \hat{s}_{Jk}\phi_{Jk}(t) + \sum_{j=0}^{q} \sum_{k=-N+1}^{N2^{J+j}-1} \hat{d}_{jk}\psi_{jk}(t). \tag{3.17}$$

Further, we also assume that $\phi(t)$ and $\psi(t)$ are the father and the mother wavelets with compact support $[0, N]$ for some $N \in \mathbb{N}$.

It was pointed out by many researchers that linear methods are not efficient when signals have considerable time-inhomogeneity, such as varying degrees of smoothness, and in fact the linear wavelet estimators often include unnecessary high oscillations. Non-linear estimators can improve the efficiency and achieve better convergence rates. The simplest way to correct this disadvantages is to suppress too small coefficients by the wavelet estimation, for example, by introducing a threshold. Such an approach is called wavelet thresholding. There exist various thresholding concepts. However, in this thesis we restrict ourselves to the two of them: soft thresholding and hard thresholding. In the context of wavelets, these techniques were proposed by D.Donoho and I. Johnstone in the beginning of 1990-ies.

Table 3.2: This table presents the corresponding coefficients of Daubechies wavelets $d2N$ for $N = 1, 2, \ldots, 10$

k	$N = 1$	$N = 2$	$N = 3$	$N = 4$	$N = 5$
0	0.707106781	0.482962913	0.332670553	0.230377813	0.160102398
1	0.707106781	0.836516304	0.806891509	0.714846571	0.603829270
2		0.224143868	0.459877502	0.630880768	0.724308528
3		-0.129409523	-0.135011020	-0.027983769	0.138428146
4			-0.085441274	-0.187034812	-0.242294887
5			0.035226292	0.030841382	-0.032244870
6				0.032883012	0.077571494
7				-0.010597402	-0.006241490
8					-0.012580752
9					0.003335725
k	$N = 6$	$N = 7$	$N = 8$	$N = 9$	$N = 10$
0	0.111540743	0.077852054	0.054415842	0.038077947	0.026670058
1	0.494623890	0.396539320	0.312871591	0.243834675	0.188176800
2	0.751133908	0.729132091	0.675630736	0.604823124	0.527201189
3	0.315250352	0.469782287	0.585354684	0.657288078	0.688459040
4	-0.226264694	-0.143906004	-0.015829105	0.133197386	0.281172344
5	-0.129766868	-0.224036185	-0.284015543	-0.293273783	-0.249846424
6	0.097501606	0.071309219	4.724846E-4	-0.096840783	-0.195946274
7	0.027522866	0.080612609	0.128747427	0.148540749	0.127369340
8	-0.031582039	-0.038029937	-0.017369301	0.030725682	0.093057365
9	5.538422E-4	-0.016574542	-0.044088254	-0.067632829	-0.071394147
10	0.004777258	0.012550999	0.013981028	2.509471E-4	-0.029457537
11	-0.001077301	4.295780E-4	0.008746094	0.022361662	0.033212674
12		-0.001801641	-0.004870353	-0.004723205	0.003606554
13		3.537138E-4	-3.917404E-4	-0.004281504	-0.010733176
14			6.754494E-4	0.001847647	0.001395352
15			-1.174768E-4	2.303858E-4	0.001992405
16				-2.519632E-4	-6.858567E-4
17				3.934732E-5	-1.164669E-4
18					9.358867E-5
19					-1.326420E-5

In soft thresholding one replaces $\hat{d}_{jk}$ in (3.17) by

$$(|\hat{d}_{jk}| - \delta_j) I(|\hat{d}_{jk}| > \delta_j) sign(\hat{d}_{jk})$$

where δ_j is a threshold. In comparison, in hard thresholding one replaces $\hat{d}_{jk}$ by

$$\hat{d}_{jk} I(|\hat{d}_{jk}| > \delta_{jk}).$$

The plots of hard thresholding and soft thresholding versus estimate $\hat{d}_{jk}$ are shown in figure (3.2). The hard thresholding wavelet estimator is respectively

$$\hat{g}(t) = \sum_{k=-N+1}^{N2^J-1} \hat{s}_{Jk} \phi_{Jk}(t) + \sum_{j=0}^{q} \sum_{k=-N+1}^{N2^{J+j}-1} \hat{d}_{jk} \, I(|\hat{d}_{jk}| > \delta_j) \psi_{jk}(t). \tag{3.18}$$

(see e.g. Donoho and Johnstone 1994, 1998, Abramovich et al. 1998). For estimates without thresholding (i.e. $\delta_j \equiv 0$) also see Johnstone and Silverman (1997), Nason (1996), Brillinger (1994, 1996), among others.
For soft thresholding wavelet estimator will be additionally assumed that the decomposition level J is equal to zero and $q = \lfloor \log_2 n \rfloor$. Thus, the corresponding estimator then given by

$$\tilde{g}(t) = \sum_{k=-N+1}^{N-1} \hat{s}_{0k} \phi_{0k}(t) + \sum_{j=0}^{\lfloor \log_2 n \rfloor} \sum_{k=-N+1}^{N2^j-1} (|\hat{d}_{jk}| - \delta_j) I(|\hat{d}_{jk}| > \delta_j) sign(\hat{d}_{jk}) \psi_{jk}(t). \tag{3.19}$$

3.8 Trend estimation via wavelet shrinkage

This section is concerned with nonparametric estimation of a deterministic trend of a noisy data. Suppose that we observe time series data of the form (1.2). Thus,

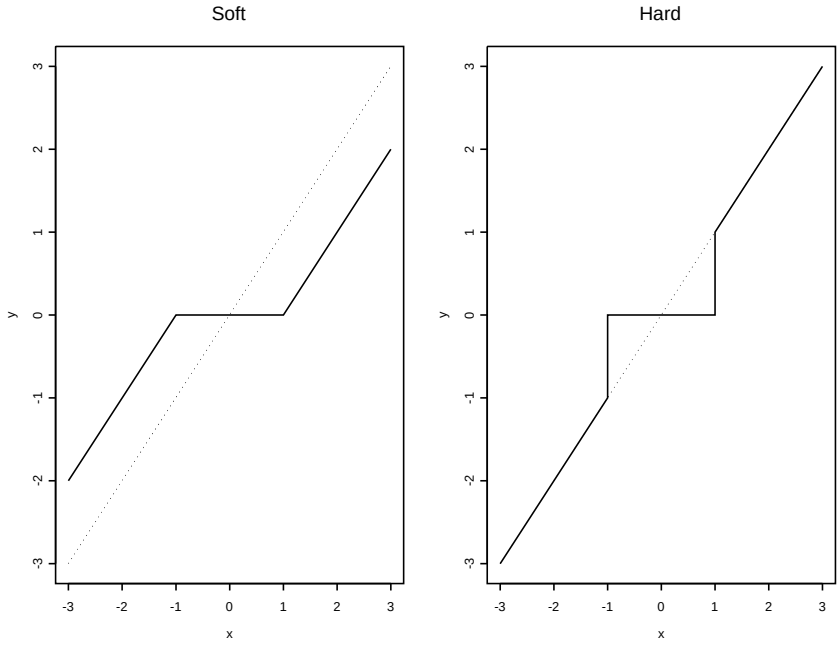
$$Y_i = g(t_i) + \xi_i, \; i = 1, 2, \ldots, n,$$

with $t_i = i/n$, $g \in L^2([0,1])$ and ξ_i a Gaussian zero mean second order stationary process with long-range dependence. Here, long-range dependence is characterized by (1.1). The only knowledge about g is that it belongs to some known set $\mathfrak{F} \subseteq L^2$ of functions. Now we define the mean integrated square error ($MISE$) of an estimator $\hat{g}$ of g by

$$E\left[\|\hat{g} - g\|_{L_2}^2\right]. \tag{3.20}$$

The aim of the next two subsections is to derive two well known results on the error minimization in some appropriate sense.

Figure 3.2: Plot of $y = (|x| - 1)I(|x| > 1)sign(x)$ (left) and $y = xI(|x| > 1)$ (right).



3.8.1 Minimax wavelet trend estimation

The performance in the worst case over the set $\mathfrak{F} \subseteq L^2$ is given by

$$\sup_{g \in \mathfrak{F}} E\left[\|\hat{g} - g\|_{L_2}^2\right].$$

We define the L^2 minimax risk by minimizing the $MISE$ in the worst case over all possible estimates $\hat{g}$. Thus,

$$R_n(\mathfrak{F}, g) = \inf_{\hat{g}} \sup_{g \in \mathfrak{F}} E\left[\|\hat{g} - g\|_{L_2}^2\right]. \tag{3.21}$$

Estimation of g we perform now by minimizing L^2 risk (3.21) over a certain set of functions. Donoho and Johnstone (1994) suggested a universal threshold by setting

$$\delta_j = \sigma\sqrt{2\ln n}$$

in case ξ_i in (1.2) is iid. The combination of a soft-thresholding rule with universal thresholding is known as *VisuSchrink*. They also show that the risk in this case is close up to a logarithmic factor to the minimax $R_n(\mathfrak{F}, g)$ over every $\mathfrak{F}$ classes taken from norms in the Besov and Triebel scales with $\sigma > 1/p$- e.g. Sobolev

balls. The fundamental property behind *VisuSchrink* is following. Let $Z_1, \ldots, Z_n$ be Gaussian iid with mean 0 and variance σ^2, then

$$P\left\{\max_{1\leq i\leq n}|Z_i| \leq \sigma\sqrt{2\ln n}\right\} \to 1.$$

Then by using the universal threshold, we are assured, that if the sample size is large enough, then the probability of the noisy coefficients asymptotically being shrunk to zero. As already mentioned, this result is an asymptotic one and does not guarantee good performance with respect to finite sample sizes. However, if the noise in the data is stationary and correlated, then the variance of the wavelet coefficients will depend on the level in the wavelet decomposition and will be constant within each level. Therefore, it will be natural to use a level dependent thresholding rule.

We define now exactly what we mean by Besov spaceses and give their appropriate characterization by the wavelet decomposition coefficients (see e.g. Donoho and Johnstone 1995, Härdle et al. 1998 and Heiler 2006).

Definition 3.6. *Let $\Delta_n^m f$ be the m-th order difference of f with $m \leq r \leq m+1$. We say that $g : [0,1] \to \mathbb{R}$ satisfying $g \in W_p^m([0,1])$ is in the Besov space $B_{p,q}^r$ whenever*

$$|g|_{B_{p,q}^r} = \begin{cases} \left(\int_0^1 \left(\frac{||\Delta_n^m g||_{L_p([0,1-rh])}}{h^r}\right)^q \frac{dh}{h}\right)^{1/q}, & q < \infty, \\ \sup_{0<h<1} \frac{||\Delta_n^m g||_{L_p([0,1-rh])}}{h^r}, & q = \infty, \end{cases}$$

is finite. The space $B_{p,q}^r(\mathbb{R})$ equipped with the norm

$$||g||_{B_{p,q}^r} = ||g||_{L_p} + ||g^{(m)}||_{L_p} + |g|_{B_{p,q}^r}$$

where $g^{(\cdot)}$ denote weak derivative.

Here we note that r, corresponding to the number of derivatives that g has in L_p (weak derivative) and if $p = q = 2$, then the Besov spaces coincide with the classical Sobolev spaces.

Donoho and Johnstone (1997) derived their results for a broad range of regression function classes, corresponding to sequence space models of its coefficients $s = \{s_{0k}\}$ and $d = \{d_{jk}\}$. A flexible scale of functional classes is given by the Besov family, which is specified in sequence space form as follows. Set $||d_{j\cdot}||_{L_p}^p = \sum_{k=-N+1}^{N2^j-1} |d_{jk}|^p$, $||s_{0\cdot}||_{L_p}^p = \sum_{k=-N+1}^{N2^j-1} |s_{0k}|^p$,

$$||g||_{b_{p,q}^r} = ||s_{0\cdot}||_{L_p} + \left[\sum_{j=0}^{\infty} 2^{jq(r+1/2-1/p)} ||d_{j\cdot}||_{L_p}^q\right]^{1/q}$$

and

$$b^r_{p,q}(C) = \left\{(d_{jk}) \,:\, ||g||_{b^r_{p,q}} \leq C\right\}.$$

For a fuller discussion on this definition and importance of this classes we refer to Donoho and Johnstone (1997) and Härdle et al. (1998).
The relation between Besov spaces and wavelet coefficients sequence spaces (see e.g. Härdle et al. 1998) given by

Remark 7. *In the case $1 \leq p, q \leq \infty$ the Besov norm is equivalent to the sequence norm of wavelet coefficients of g. Thus,*

$$||g||_{B^r_{p,q}} \asymp ||g||_{b^r_{p,q}}$$

where $''\asymp''$ means that the ratio of both sides is bounded by two constants.

For additional information on this see Heiler (2006) and Härdle et al. (1998).
We now suppose that the noise exhibits long-range dependence and introduce the concept of Donoho and Johnstone (1997):

Theorem 3.2. *If $0 < p, q \leq \infty$ and $r > 1/p$, then*

$$\sup_{d \in b^r_{p,q}(C)} E\left[||d - (|\hat{d}| - \delta)I(|\hat{d}| > \delta)sign(\hat{d})||^2_{L_2}\right] \leq C_0 C^{\frac{2\alpha}{2r+\alpha}} \log n \cdot n^{-\frac{2r\alpha}{2r+\alpha}}$$

where $\delta = \{\delta_j\} = \left\{\sigma_j\sqrt{2\ln n}\right\}$ and $\sigma_j^2 = Var(d_{jk})$, $\forall k$.

Moreover, Wang (1996) states

Theorem 3.3. *Under assumption of theorem 3.2 yields*

$$\inf_{\hat{d}} \sup_{d \in b^l_{p,q}(C)} E\left[||d - \hat{d}||^2_{L_2}\right] \sim n^{-\frac{2r\alpha}{2r+\alpha}}$$

where $\hat{d}$ all possible estimators.

Then by Donoho and Johnstone (1997) yields also following remark

Remark 8. *Since the wavelet basis is orthogonal, the risk of the soft thresholding estimator (3.19) with $\delta = \left\{\sigma_j\sqrt{2\ln n}\right\}$ achieves the optimal convergence rate up to a logarithmic factor.*

3.8.2 Hard thresholding wavelet trend estimation

Li and Xiao (2007) considered the nonparametric regression model (1.2) with long memory data, that are not necessarily Gaussian, and provided an asymptotic expansion for the MISE of hard thresholding wavelet estimators. They derived MISE expansion for the piecewise smooth underlying mean regression function. It turns out that MISE convergence rate is the same as by analogous minimax expansion (see e.g. Wang 1996). Throughout this thesis, $m_\psi \in \mathbb{N}$ will denote the number of vanishing moments of ψ, i.e.

$$\int_0^N t^k \psi(t)\, dt = 0,\ k = 0, 1, \ldots, m_\psi - 1 \tag{3.22}$$

and

$$\int_0^N t^{m_\psi} \psi(t)\, dt = \nu_{m_\psi} \neq 0. \tag{3.23}$$

We assume that both ϕ and ψ satisfy a uniform Hölder condition of exponent $1/2$ and the following conditions are satisfied:

- The errors ξ_i in (1.2) given by $\xi_i = G(\epsilon_i)$ where ϵ_i a Gaussian zero mean second order stationary process with long-range dependence and G has Hermit rank m. Here, long-range dependence is characterized by (1.1).

- The smoothing parameter q, decomposition level J and thresholding δ_j are functions of n. We assume that $J \to \infty$, $q \to \infty$ as $n \to \infty$, and for every $j = 1, 2, \ldots, q-1$, holds

  $$2^{J+j}\delta_j^2 \to 0,\ 2^{(J+j)(2r+1)}\delta_j^2 \to \infty,\quad \delta_j^2 \geq \frac{(4e)^m C_1^2 N^{-1+m\alpha}(\ln n)^{m+1}}{m^m n^{m\alpha} 2^{(J+j)(1-m\alpha)}} \tag{3.24}$$

  where

  $$C_1^2 = C_\gamma \frac{J^2(m)}{m!} \int_0^N \int_0^N |x-y|^{-m\alpha}\, \psi(x)\, \psi(y)\, dx\, dy$$

  with $J(m) = E(G(Z)H_m(Z))$ where Z Gaussian with mean 0, variance 1 and H_m is Hermite polynomial of order m. We additionally assume that $0 < m\alpha < 1$.

Under these conditions they derived

Theorem 3.4. *If, in addition to the conditions stated before, we assume that the rth derivative $g^{(r)}$ is continuous and bounded on $[0,1]$. Let $m_\psi = r$, then as*

$n \to \infty$,

$$E\left[||g-\hat{g}||_{L_2}^2\right] = C_2(n^{-1}N2^J)^{m\alpha} + (N2^J)^{-2r}(r!)^{-2}\nu_r^2(1-2^{-2r})^{-1}\int_0^1 g^{(r)^2}(t)\,dt \tag{3.25}$$

where C_2 is a positive and finite constant defined by

$$C_2^2 = C_\gamma \frac{J^2(m)}{m!}\int_0^N\int_0^N |x-y|^{-m\alpha}\,\phi(x)\,\phi(y)\,dx\,dy.$$

In this theorem, they have assumed that the mean regression function g is rth times continuously differentiable. However, this result can be generalized. In addition to last theorem, Li and Xiao (2007) showed, that if $g^{(r)}$ is only piecewise continuous, this result still holds, as given in the following:

Theorem 3.5. *If, in addition to the conditions stated before, we suppose that $g^{(r)}$ exists on $[0,1]$ except for at most a finite number of points, and, where it exists, it is piecewise continuous and bounded. Furthermore, assume that $\mathrm{supp}(g^{(r)})$ has positive Lebesgue measure and $m_\psi = r$. In particular, g itself may be only piecewise continuous. Also, assume that $2^{(2r+m\alpha)J}n^{-2rm\alpha} \to \infty$. Then (3.25) still holds.*

Furthermore, this implies

Remark 9. *For the Gaussian error special case, the Hermite rank of G is 1 (i.e., $m=1$). In this case, if the decomposition level J is chosen of size $\frac{\alpha}{2r+\alpha}\log_2 n$, then the convergence rate of $MISE$ is $n^{-\frac{2r\alpha}{(2r+\alpha)}}$, which is the same as those in Wang (1996). Moreover, Hall and Hart (1990) gave a similar asymptotic expansion for $MISE$, considering kernel estimator in fixed-design nonparametric regression when error is Gaussian long memory process and trend is smooth.*

No further justification for the specific choice of J and q was given by Li and Xiao (2007), as well as no optimality result is derived. We refer to Li and Xiao (2007) for further discussion.

Beran and Feng (2002a) in contrast to Hall and Hart (1990) considered kernel estimator in semiparametric fractional autoregressive models ($SEMIFAR$) and derived corresponding results for this case. The class of $SEMIFAR$ models

includes Gaussian $FARIMA$ models, what allows us to formulate this results in form that we using later in the empirical studies.
Before starting the result, we require following definitions. Let $K(x)$ be a symmetric, nonnegative polynomial kernels with

$$K(x) = \begin{cases} \sum_{l=0}^{r} a_l x^{2l} & |x| \leq 1, \\ 0 & |x| > 1. \end{cases} \tag{3.26}$$

where $r \in \mathbb{N} \cup \{0\}$ and the coefficients a_i fulfill $\int_{-1}^{1} K(x)\,dx = 1$ (see e.g. Gasser and Müller 1979). Suppose that we observe time series data of the form (1.2). For a given bandwidth $b > 0$ and $t \in [0, 1]$, the kernel estimate of g defined by

$$\hat{g} = \frac{1}{nb} \sum_{i=1}^{n} K\left(\frac{t - t_i}{b}\right) Y_i.$$

The following results originally derived by Beran and Feng (2002a)

Theorem 3.6. *Let $b_n > 0$ be a sequence of bandwidths with $b_n \to 0$ and $nb_n \to \infty$, and Y_i a $FARIMA$ process. Assume g be a twice differentiable, for all $x, y \in [0, 1]$*

$$\sup_{x \in (0,1)} \max_{j=0,1,2} |g^{(j)}(x)| \leq C_1 < \infty$$

and

$$|g''(x) - g''(y)| \leq C_2 |x - y|^{\beta}$$

where constants $C_1, C_2 < \infty$ and $\beta \in (2, 3]$. Moreover, assume that at least one $j \in \{0, 1\}$, $g^{(j)}$ does not vanish in $[\Delta, 1-\Delta]$ and $g^{(j)}$ achieves an absolut maximum or minimum in $[\Delta, 1 - \Delta]$. Then, we have

1. *Mean integrated squared error in $[\Delta, 1 - \Delta]$*

$$\int_{\Delta}^{1-\Delta} E\left\{[g(t) - \hat{d}(t)]^2\right\} dt = b_n^4 \frac{1}{4} \left(\int_{\Delta}^{1-\Delta} \left(g^{(2)}\right)^2 dx\right) \left(\int_{-1}^{1} x^2 K(x)\,dx\right)^2$$
$$+ (nb_n)^{2d-1} V + o\left(\max\left\{b_n^4, (nb_n)^{2d-1}\right\}\right)$$

   where

$$V = 2C_f \Gamma(1 - 2d) \sin(\pi d) \int_{-1}^{1} \int_{-1}^{1} |x - y|^{2d-1} K(x) K(y)\,dx\,dy.$$

2. *The bandwidth that minimizes the asymptotic MISE is given by*

$$b_{opt} = n^{\frac{2d-1}{5-2d}} \left[\frac{1}{(1-2d)V} \left(\int_{\Delta}^{1-\Delta} \left(g^{(2)}\right)^2 dx\right) \left(\int_{-1}^{1} x^2 K(x)\,dx\right)^2\right]^{-1/(5-2d)}.$$

Note that for $r = 0$ we obtain box-kernel and this result take following representation:

Theorem 3.7. *Let* $K(x) = \frac{1}{2} I\{x \in [-1, 1]\}$. *Define*

$$\beta(d) = \frac{2^{2d}\Gamma(1-2d)\sin(\pi d)}{d(2d+1)}.$$

Then, under the assumptions of theorem 3.6, we have

1. *Mean integrated squared error in* $[\Delta, 1-\Delta]$

$$\int_{\Delta}^{1-\Delta} E\left\{[g(t) - \hat{d}(t)]^2\right\} dt = b_n^4 \frac{1}{36}\left(\int_{\Delta}^{1-\Delta} \left(g^{(2)}\right)^2 dx\right)$$
$$+(nb_n)^{2d-1} C_f \beta(d) + o\left(\max\left\{b_n^4, (nb_n)^{2d-1}\right\}\right).$$

2. *The bandwidth that minimizes the asymptotic MISE is given by*

$$b_{opt} = n^{\frac{2d-1}{5-2d}}\left[\frac{1}{(1-2d)C_f\beta(d)}\left(\int_{\Delta}^{1-\Delta}\left(g^{(2)}\right)^2 dx\right)\right]^{-1/(5-2d)}.$$

Note that the K, defined in (3.26), is only a second order kernel. Similar results can be obtained for kernel estimates with higher order kernels. This is obviously beyond the scope of the present thesis.

Chapter 4

Asymptotically optimal wavelet estimation of trend functions

In this chapter, we consider the hard thresholding wavelet trend estimation for data of the form (1.2). In theorem 4.1 we establish rate optimality and this result is only of limited practical use, since the estimator is not exactly defined. In order to apply the result to observed data, optimal parameter choice needs to be derived. This question is addressed in theorems 4.2 and 4.3 below. The presentation in this chapter is fairly detailed whereas the main results correspond to Beran and Shumeyko (2011a).
Specifically, this chapter is organized as follows. Basic definitions and notations are introduced in section 4.1, the main results are given in section 4.2. The results are illustrated by tables and figures, which achieved by a short simulation study, in section 4.3. To verify the behavior of data adaptive wavelet estimation as outlined below, we carried out a simulation study with different test functions g and a Gaussian $FARIMA(0, d, 0)$ residual process. We also simultaneously compared hard thresholding wavelet estimators with minimax soft thresholding wavelet estimators and kernal estimators. Proofs are in appendix 4.4 will finalize the chapter.

4.1 Notations

Throughout this chapter, we are making a following assumptions. Suppose that we observe time series data of the form (1.2) with a Gaussian zero mean second

order stationary process with long-range dependence. The long-range dependence will be characterized by (1.1).
Let $\phi(t)$ and $\psi(t)$ be the father and mother wavelets respectively with compact support $[0, N]$ for some $N \in \mathbb{N}$ and such that

$$\psi(0) = \psi(N) = 0. \tag{4.1}$$

It then follows

$$\int_0^N \phi(t)dt = \int_0^N \phi^2(t)dt = \int_0^N \psi^2(t)dt = 1, \tag{4.2}$$

Note that, for the sake of generality, the support of ϕ and ψ is chosen to be $[0, N]$ instead of $[0, 1]$. By this way, it is possible to choose from a larger variety of wavelet generating functions satisfying (4.2) (see e.g. Daubechies 1992, Cohen et al. 1993). $m_\psi \in \mathbb{N}$ will denote the number of vanishing moments of ψ and ν_{m_ψ} is m_ψth moment, as it is defined by (3.22) and by (3.23) respectively.
For every trend function $g \in L^2([0, 1])$, and every decomposition level $J \geq 0$, smoothing parameter q and threshold δ_j, we define the hard thresholding wavelet estimator by (3.18).

4.2 Main results

In the context of long-memory errors, an explicit asymptotic expansion for the MISE is given in the section 3.8.2 (see e.g. Li and Xiao 2007), under specific assumptions on the decomposition level J and the smoothing parameter q. The question how to choose J and q optimally is not investigated. The following theorem establishes the optimal convergence rate of the $MISE$ when minimizing with respect to J, q and $\{\delta_j\}$.
In following ϕ and ψ will either be assumed to be piecewise differentiable or they satisfy a uniform Hölder condition with exponent 1/2, i.e.,

$$|\psi(x) - \psi(y)| \leq C|x - y|^{1/2}, \forall x, y \in [0, N]. \tag{4.3}$$

In chapter 3 (see e.g. Chap. 6 of Daubechies 1992) we provided some examples of wavelets satisfying these conditions. Moreover, throughout the paper $2^J = o(n)$ to ensure that $\hat{g}$ includes resolution levels lower than the distance between successive time points. This assumption is needed for consistency of $\hat{g}$ as considered below.

Theorem 4.1. *Suppose that* $g \in C^r[0,1]$*, the support* $\mathrm{supp}(g^{(r)}) = \{t \in [0,1] : g^{(r)}(t) \neq 0\}$ *has positive Lebesgue measure, the process* ξ_i *is Gaussian with covariance structure (1.1), and* ψ *is such that* $m_\psi = r$*. Then, minimizing the* $MISE$ *with respect to* J*,* q *and* $\{\delta_j\}$ *yields the optimal order*

$$MISE_{opt} = O(n^{-\frac{2r\alpha}{2r+\alpha}}). \tag{4.4}$$

Theorem 4.1 is of limited practical use, since only rate optimality is established. The following theorem 4.2 shows that the rate obtained in Li and Xiao (2007) can be achieved without thresholding by minimizing the MISE with respect to J and q. In order to apply the result to observed data, optimal constants need to be derived. This question is addressed in theorems 4.2 and 4.3 below. The following constants will be needed:

$$C_\phi^2 = C_\gamma \int_0^N \int_0^N |x-y|^{-\alpha}\, \phi(x)\, \phi(y)\, dx\, dy, \tag{4.5}$$

$$C_\psi^2 = C_\gamma \int_0^N \int_0^N |x-y|^{-\alpha}\, \psi(x)\, \psi(y)\, dx\, dy, \tag{4.6}$$

$$C^*(r,\alpha,\psi,g^{(r)}) = \frac{1}{2r+\alpha}\log_2\left[\frac{\int_0^N \nu_r^2\left(g^{(r)}(t)\right)^2\, dt}{C_\psi^2 (r!)^2}\right] - \log_2 N,$$

$$\Delta_n(g, C_\psi) = \frac{\alpha}{2r+\alpha}\log_2 n + C^*(r,\alpha,\psi,g^{(r)}) - \left\lfloor \frac{\alpha}{2r+\alpha}\log_2 n + C^*(r,\alpha,\psi,g^{(r)})\right\rfloor, \tag{4.7}$$

where $\lfloor x \rfloor$ denotes the largest integer less than or equal to x,

$$A_1(r,\alpha,\psi) = \left(\frac{2^{2r\Delta_n(g,C_\psi)}}{2^{2r}-1} + \frac{2^{\alpha(1-\Delta_n(g,C_\psi))}}{2^\alpha - 1}\right)\left(C_\psi^2\right)^{\frac{2r}{2r+\alpha}},$$

$$A_2(r,\alpha,\psi,g^{(r)}) = \left(\frac{\nu_r^2}{(r!)^2}\int_0^1 \left(g^{(r)}(t)\right)^2\, dt\right)^{\frac{\alpha}{2r+\alpha}},$$

$$\nu_r = \int t^r \psi(t)\, dt,$$

$$C^*(r,\alpha,\phi,g^{(r)}) = \frac{1}{2r+\alpha}\log_2\left[\frac{\int_0^1 \nu_r^2\left(g^{(r)}(t)\right)^2\, dt}{C_\phi^2(2^\alpha-1)(r!)^2}\right] - \log_2 N,$$

$$\Delta_n(g, C_\phi) = \frac{\alpha}{2r+\alpha}\log_2 n + C^*(r,\alpha,\phi,g^{(r)})$$

$$- \left\lfloor \frac{\alpha}{2r+\alpha} \log_2 n + C^*(r, \alpha, \phi, g^{(r)}) \right\rfloor, \tag{4.8}$$

$$A_3(r, \alpha, \phi) = \left(\frac{2^{2r\Delta_n(g, C_\phi)}}{2^{2r} - 1} + \frac{2^{\alpha(1-\Delta_n(g, C_\psi))}}{2^\alpha - 1} \right) \left(C_\phi^2 (2^\alpha - 1) \right)^{\frac{2r}{2r+\alpha}}.$$

For the case where no thresholding is used, exact asymptotic expressions for the MISE and an optimal solution can be given as follows.

Theorem 4.2. *Under the assumptions of theorem 4.1 and thresholds*

$$\delta_j = 0 \ (0 \le j \le q),$$

the following holds.

(i) If $(2^\alpha - 1) C_\phi^2 > C_\psi^2$*, then the asymptotic MISE is minimized by the smoothing parameter*

$$q^* = \left\lfloor \frac{\alpha}{2r+\alpha} \log_2 n + C^*(r, \alpha, \psi, g^{(r)}) \right\rfloor - J^* \tag{4.9}$$

with decomposition levels J^* *satisfying* $2^{J^*} = o\left(n^{\frac{\alpha}{2r+\alpha}} \right)$*. The optimal MISE is of the form*

$$MISE = A_1(r, \alpha, \psi) A_2(r, \alpha, \psi, g^{(r)}) \cdot n^{-\frac{2r\alpha}{2r+\alpha}} + o\left(n^{-\frac{2r\alpha}{2r+\alpha}} \right). \tag{4.10}$$

Moreover, if $\Delta_n(g, C_\psi) = 0$*, then also*

$$q^* = \left\lfloor \frac{\alpha}{2r+\alpha} \log_2 n + C^*(r, \alpha, \psi, g^{(r)}) \right\rfloor - J^* - 1$$

(and J^* *as before) minimizes the* $MISE$*.*

(ii) If $(2^\alpha - 1) C_\phi^2 < C_\psi^2$*, then minimizing the asymptotic MISE with respect to* J *and* q *yields*

$$J^* = \left\lfloor \frac{\alpha}{2r+\alpha} \log_2 n + C^*(r, \alpha, \phi, g^{(r)}) \right\rfloor + 1 \tag{4.11}$$

and

$$\hat{g}(t) = \sum_{k=-N+1}^{N2^J - 1} \hat{s}_{Jk} \phi_{Jk}(t), \tag{4.12}$$

with $J = J^*$*. The optimal MISE is of the form*

$$MISE = A_3(r, \alpha, \phi) A_2(r, \alpha, \psi, g^{(r)}) \cdot n^{-\frac{2r\alpha}{2r+\alpha}} + o\left(n^{-\frac{2r\alpha}{2r+\alpha}} \right). \tag{4.13}$$

Moreover, if $\Delta_n(g, C_\phi) = 0$*, then also*

$$J^* = \left\lfloor \frac{\alpha}{2r+\alpha} \log_2 n + C^*(r, \alpha, \phi, g^{(r)}) \right\rfloor$$

minimizes the MISE.

If higher resolution levels beyond those used in theorem 4.2 are included together with thresholding, then the value of the MISE given in (4.10) and (4.13) respectively can be attained even if $g^{(r)}$ does not exists everywhere and is only piecewise continuous:

Theorem 4.3. *Suppose that* $g^{(r)}$ *exists on* $[0,1]$ *except for at most a finite number of points, and, where it exists, it is piecewise continuous and bounded. Furthermore, assume that* $\operatorname{supp}(g^{(r)})$ *has positive Lebesgue measure,* $m_\psi = r$ *and the process* ξ_i *is Gaussian and such that (1.1) holds. Then the following holds:*

(i) If $(2^\alpha - 1)C_\phi^2 > C_\psi^2$*,* J *is such that* $2^J = o\left(n^{\frac{\alpha}{2r+\alpha}}\right)$*,* $q = \lfloor \log_2 n \rfloor - J$*,* q^* *is defined by (4.9), and* δ_j *is such that for* $0 \le j \le q^*$

$$\delta_j = 0 \tag{4.14}$$

and for $q^* < j \le q$

$$2^{J+j}\delta_j^2 \to 0,\ 2^{(J+j)(2r+1)}\delta_j^2 \to \infty, \qquad \delta_j^2 \ge \frac{4eC_\psi^2 N^{-1+\alpha}(\ln n)^2}{n^\alpha 2^{(J+j)(1-\alpha)}}, \tag{4.15}$$

then equation (4.10) holds.

(ii) If $(2^\alpha - 1)C_\phi^2 < C_\psi^2$*,* $J = J^*$ *with* J^* *defined by (17),* $q = \lfloor \log_2 n \rfloor - J$ *and* δ_j *such that*

$$2^{J+j}\delta_j^2 \to 0,\ 2^{(J+j)(2r+1)}\delta_j^2 \to \infty,\ \delta_j^2 \ge \frac{4eC_\psi^2 N^{-1+\alpha}(\ln n)^2}{n^\alpha 2^{(J+j)(1-\alpha)}}\ (0 \le j \le q), \tag{4.16}$$

then equation (4.13) holds.

Remark 10. *Li and Xiao (2007) derived an asymptotic expansion for the MISE under the assumption that* $J, q \to \infty$*,* $2^{J+j}\delta_j^2 \to 0$*,* $2^{(2r+1)(J+j)}\delta_j^2 \to \infty$ *and* δ_j^2 *are above a certain bound that depends on* j*,* n*,* g*,* α *and* J*. The question how to choose* J*,* q *and* δ_j *optimally is not considered. Here a partial solution to the optimality problem is given. Theorem 4.2 provides optimal values of* q *and* J*, and*

a corresponding formula for the optimal MISE, for estimators with no thresholding (i.e. $\delta_j \equiv 0$). This result is obtained for $r-$times continuously differentiable trend functions. Thus, jumps and other irregularities in g are excluded. In a second step, we therefore ask the question whether the asymptotic formula for the optimal MISE can be extended to more general functions. Theorem 4.3 shows that this is indeed the case, in the sense that (essentially) g does not need to be differentiable everywhere. This includes, for instance, the possibility of isolated jumps. Note that, for a given n, $q = \lfloor \log_2 n \rfloor - J$ is the highest available resolution. By adding all available higher resolution levels combined with thresholding, the same formula for the MISE applies as in theorem 4.2. The intuitive reason is that isolated discontinuities are "infinitesimally local" and can therefore be characterized best when the finest possible levels of resolution are included. At very high resolution, however, nonzero thresholds are needed in order to distinguish deterministic jumps from noise. For functions where theorem 4.2 applies, the optimal MISE in theorem 4.2 and the MISE obtained in theorem 4.3 are the same.

Remark 11. *The only quantity in (4.9) and (4.11) that depends on n is $\alpha(2r+\alpha)^{-1}\log_2 n$. The constants $C^*(r,\alpha,\psi,g^{(r)})$ and $C^*(r,\alpha,\phi,g^{(r)})$ respectively provide data adaptive adjustments to optimize the multiplicative constant in the MISE. They can be decomposed into several terms with different meanings. For instance,*

$$C^*(r,\alpha,\phi,g^{(r)}) = \frac{C_1^* + C_2^* + C_3^*}{2r+\alpha} + C_4^*$$

with

$$C_1^* = \log_2 \int_0^1 \left(g^{(r)}(t)\right)^2 dt$$

reflecting the properties of g,

$$C_2^* = \log_2 \left(\frac{\nu_r}{r!}\right)^2$$

depending on the basis function ψ,

$$C_3^* = -\log_2 \left[C_\phi^2 (2^\alpha - 1)\right]$$

characterized by the basis function ϕ and the asymptotic covariance structure (1.1) of ξ_i, and

$$C_4^* = -\log_2 N$$

defined by the length of the support of ψ and ϕ. Note that, for $N = 1$, $C_4^* = 0$.

Remark 12. *The question in how far the MISE can be optimized further with respect to freely adjustable thresholds is more difficult and subject to current research. The same comment applies to the possibility of soft thresholding. It is worth mentioning here, however, that for some classes of functions, $\delta_j = 0$ is indeed the best threshold. For instance, it can be shown that, if $g \in L^2[0,1]$ and $C < |g^{(r)}(\cdot)| \leq C2^{r+\alpha/2}$ (almost everywhere) for some finite constant C, then $\delta_j = 0$ is asymptotically optimal. This includes, for example, functions that can be represented (or approximated in an appropriate sense) by piecewise rth order polynomials.*

Remark 13. *The results in Li and Xiao (2007) are derived for residuals of the form $\xi_i = G(Z_i)$ where Z_i is a stationary Gaussian long-memory process and the transformation G has Hermite rank m_G. For simplicity of presentation, the results given here are derived for Gaussian processes only. An extension to $\xi_i = G(Z_i)$ would be possible along the same lines.*

Remark 14. *Asymptotic expressions for the MISE and formulas for optimal bandwidth selection in kernel regression with long memory are given in Hall and Hart (1990), Csörgö and Mielniczuk (1995) and Beran and Feng (2002a,c), among others. Note, however, that there, $g^{(r)}$ has to be assumed to be continuous instead of only piecewise continuous, and $r \geq 2$. In that sense, the applicability of kernel estimators (and also of local polynomials) is more limited. This is illustrated in the simulation study in the next section.*

Remark 15. *In analogy to kernel estimation, the optimal rate of convergence of wavelet estimates becomes faster the more derivatives of g exist. However, the optimal MISE can only be achieved, if the number of vanishing moments of the mother wavelet ψ is equal to r. In other words, the choice of an appropriate wavelet basis is essential. This is analogous to kernel estimation where a kernel of the appropriate order should be used (see e.g. Gasser and Müller (1984)). Consider, for instance, the case where only the first derivative of g exists (and is piecewise continuous), i.e. $r = 1$. Then, for the wavelets estimator, the optimal order of the MISE is $O(n^{-\frac{2\alpha}{2+\alpha}})$. In this case, we may use Haar wavelets (for which $m_\psi = 1$). In contrast to the wavelet estimator, the usual asymptotic expansion for the MISE of kernel estimators does not hold in this case. On the other hand, if g is twice continuously differentiable, then the optimal rate achieved by kernel estimators is at least $O(n^{-\frac{4\alpha}{4+\alpha}})$. If Haar wavelets are used, in spite of r being*

equal to 2, then the optimal rate of the wavelet estimator cannot be better than $O(n^{-\frac{2\alpha}{2+\alpha}})$ *and is thus slower than the rate achieved by kernel estimators. In order to match the rate of kernel estimators, a wavelet basis with* $m_\psi = 2$ *vanishing moments has to be used.*

Remark 16. *The optimal rate of convergence of the MISE is the same as the minimax rate obtained by Wang (1996) and Johnstone and Silverman (1997). However, for a given function, the multiplicative constant in the asymptotic expression of the MISE is essential. This is achieved here by data adaptive choice of* q *and* J*. The simulations in the next section illustrate that the data adaptive method tends to outperform the minimax solution, provided that the assumptions of theorem 4.2 or 4.3 hold.*

Remark 17. *The best smoothing parameter and decomposition level depend on the unknown parameters* α*,* C_γ *and the unknown* r*th derivative of* g*. Based on theorems 4.2 and 4.3, an iterative data adaptive algorithm along the line of Beran and Feng (2002b) can be designed. Essentially, the iteration consists of a step where* g *is estimated (using the best estimates of relevant parameters available at that stage), and a step where* α*,* C_γ *and other quantities in the asymptotic MISE formula are estimated. For the estimation of* C_γ *and* α*, see for instance, Yajima (1985), Fox and Taqqu (1986), Dahlhaus (1989), Giraitis and Surgailis (1990), Beran (1994, 1995), Beran et al. (1998), Abry and Veitch (1998). A detailed iterative algorithm is currently being developed and will be presented elsewhere. An obvious choice for estimating* α *is to use an appropriate wavelet based method such as described in Bardet et al. (2000). Note that, while the idea of the iteration is simple, a concrete implementation is far from trivial (see Beran and Feng (2002b)). In particular, in the presence of long-range dependence, small changes in the smoothing parameters can lead to considerable changes in the estimate of the long-memory parameter* α *and vice versa.*

4.3 Simulations

To study the potential benefits of data adaptive wavelet estimation as outlined above, a simulation study was carried out with four different test functions g (figure 4.1) and a Gaussian $FARIMA(0, d, 0)$ residual process ξ_i. Note that $\alpha = 1 - 2d$. The test functions are:

- Sine function: $g_1(t) = 10\sin(4\pi t)$
- Jumpsine function: $g_2(t) = 10\sin(4\pi t) + \Delta \cdot I\{\frac{5}{8} < t < \frac{7}{8}\}, \quad (\Delta > 0)$
- Sharp function: $g_3(t) = 10\left[\exp\left(tI\left\{t < 0.5\right\} + (1-t)I\{t > 0.5\}\right) - 1\right]$
- Doppler function: $g_4(t) = 10\left[t(1-t)\right]^{1/2}\sin\left[2\pi(1+0.05)/(t+0.05)\right]$

The following methods are compared:

- Wavelet estimator with hard thresholding, and q, J as in theorem 4.3 and

$$\delta_j^2 = \frac{4eC_\psi^2 N^{-1+\alpha}(\ln n)^2}{n^\alpha 2^{(J+j)(1-\alpha)}}, \quad (q^* < j \leq q).$$

  Note that for the first three functions, theorem 4.3(ii) applies whereas for the Doppler function, derivatives are not bounded. Nevertheless we carried out the simulations using a modified version of C^* (see the remarks at the end of this section).

- Wavelet estimator with soft thresholding defined by

$$sign(\hat{d}_{jk})(|\hat{d}_{jk}| - \lambda_n)I\{|\hat{d}_{jk}| > \lambda_j\}$$

  and minimax thresholds

$$\lambda_j = \sigma_j(2\log n)^{1/2},$$

  as given in remark 8 (see e.g. Johnstone and Silverman 1997).

- Kernel estimator with rectangular kernel $K(x) = \frac{1}{2}I\{x \in [-1,1]\}$ and asymptotically optimal bandwidth

$$b_{opt} = C_{opt}n^{(2d-1)/(5-2d)}$$

  where

$$C_{opt} = \left(\frac{9(1-2d)\beta(d)C_f}{I(g'')}\right)^{1/(5-2d)},$$

$$\beta(d) = \frac{2^{2d}\Gamma(1-2d)\sin(\pi d)}{d(2d+1)},$$

  as given in theorem 3.7 (see e.g. Hall and Hart 1990, Beran and Feng 2002a).

Sine: Figure 4.2 shows reasonably good agreement between the simulated and theoretical MISE of the adaptive wavelet estimator with basis "d4". Here "d4", "d6", ... denote Daubechies' wavelets with $2, 3, \ldots$ vanishing moments respectively (see Daubechies (1992)). Table 4.1 illustrates the effect of using different basis functions, for the case $d = 0.2$. Irrespective of the wavelet basis ("d4", "d6", "d8" and "d10" respectively), the agreement between the simulated MISE and the theoretic formula is very good already for $n = 256$. However, since g is infinitely continuously differentiable, the MISE can be reduced by using very smooth basis functions. This explains why the performance of "d4" is considerably worse compared to "d6", "d8" and "d10". Table 4.2 shows that, as expected, the mean squared error increases with increasing long memory (also see figure 4.2). A comparison between minimax wavelet thresholding, the data adaptive wavelet estimator and kernel smoothing is given in figures 4.3 and 4.4. Since the sine function is well behaved, optimal kernel estimation is expected to perform well. The kernel estimator does indeed outperform the minimax procedure. In contrast, the MISE of the data adaptive wavelet method is comparable to optimal kernel estimation. A typical sample path and the corresponding estimated trend functions are plotted in figure 4.5. The minimax rule leads to a rather erratic function near local minima and maxima, whereas this is not the case for the other two methods.

Jumpsine: The simulated and asymptotic MISE for the Jumpsine function are compared in table 4.3 for $d = 0.2$ and jump sizes $\Delta = 0.1, 0.5, 1, 10, 20$ and 50. The agreement between the asymptotic and simulated MISE is reasonably good, in particular for small and very large values of Δ. Figure 4.6a shows a typical sample path with $d = 0.3$ and fits obtained by the three methods. Figure 4.6b shows that, as expected from theorem 4.3(ii), almost all nonzero coefficients belong to the father wavelet. The mother wavelet functions are useful to model the two jumps. Due to thresholding almost all coefficients are eliminated except those near $t = 5/8$ and $7/8$ respectively. Similar results were obtained for other values of d. In comparison, the data adaptive wavelet method shows the best performance (figures 4.7 and 4.8) though the difference between the two wavelet methods is smaller under strong long memory. As expected, kernel estimation cannot compete with the wavelet approach.

Sharp: Quite in contrast to the Jumpsine function, for the Sharp function, the performance of the kernel estimator is comparable to the data adaptive wavelet

method (figures 4.9 and 4.10), at least when the criterion is the MISE. With respect to the visual fit, as exemplified by figure 4.11, the kernel method leads to oversmoothing of the edge in the middle.

Doppler: For the Doppler function, theorem 4.3 is not applicable and J^* in equation (4.11) is not well defined. Nevertheless it is interesting to see how well hard thresholding may work with a slight modification of (4.11). Specifically, consider

$$\tilde{J}^* = \left\lfloor \frac{\alpha}{2r+\alpha} \log_2 n + \tilde{C}^*(r, \alpha, \psi, \phi, g^{(r)}) \right\rfloor + 1$$

where

$$\tilde{C}^*(r, \alpha, \psi, \phi, g^{(r)}) = \frac{1}{2r+\alpha} \log_2 \left[\frac{\int_{0.1}^{0.95} \nu_r^2 \left(g^{(r)}(t)\right)^2 \, dt}{C_\phi^2 (2^\alpha - 1)(r!)^2} \right] - \log_2 N.$$

Note that the only change compared to C^* consists of bounding the integration limits away from 0 and 1. For moderate long memory with $d = 0.2$ the data adaptive wavelet estimator still turns out to be best (figure 4.12). For strong long memory with $d = 0.4$, the minimax approach appears to be slightly better for very long series (figure 4.13). The relatively good performance of the minimax approach is expected, because, in contrast to the data adaptive estimator, the coarser levels of resolution are not favored a priori. This way, it is easier to catch the increasingly fast oscillations towards the left of the time scale. As expected the kernel method does not work well. A typical example is shown in figure 4.14.

Figure 4.1: Trend functions used in the simulations: Sine, Jumpsine, Sharp and Doppler.

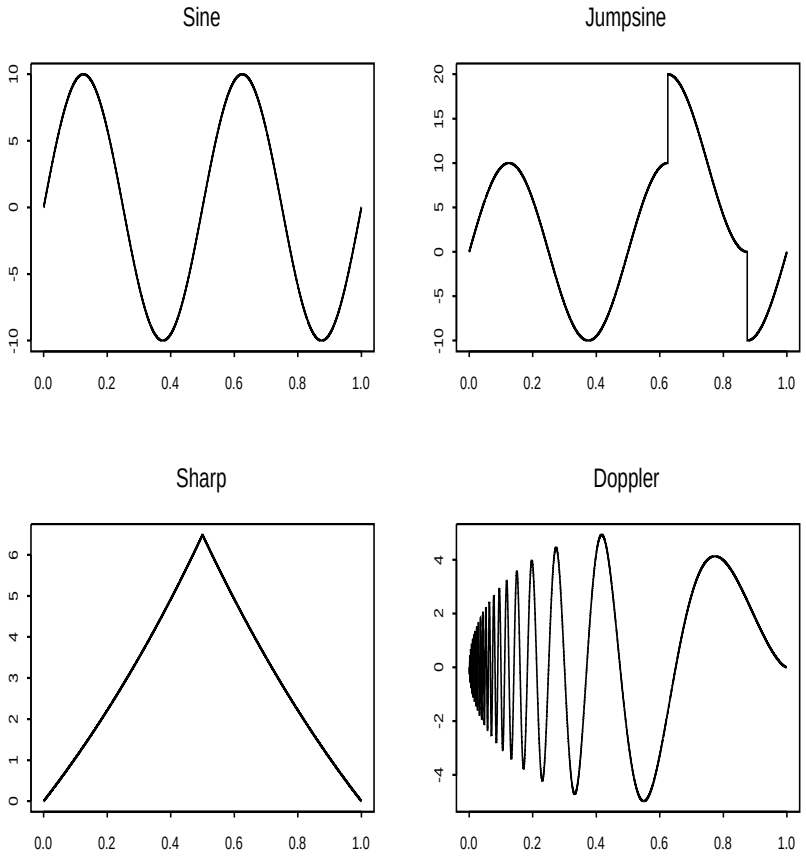


Table 4.1: Logarihms (to the base 2) of simulated values of the mean integrated squared error, $\log_2 MISE_{sim}$, as a function of n and the wavelet bases "d4", "d6", "d8" and "d10" respectively. For comparison, $\log_2 MISE_{theor}$ obtained from the asymptotic formulas in theorem 4.3 is also given. The results are based on 400 simulations of a $FARIMA(0, 0.2, 0)$ model with trend function $g_1(t) = 10\sin(4\pi t)$.

MISE Trend: Sine Noise: $FARIMA(0, 0.2, 0)$				
n	Simulation "d4"	Theor "d4"	Simulation "d6"	Theor "d6"
128	0.516420047	0.408553554	0.251744659	0.332459614
256	0.263441364	0.294451230	0.214928924	0.222321976
512	0.217604044	0.219171771	0.112951872	0.149658234
1024	0.150284851	0.150545678	0.110547951	0.101718042
2048	0.109213215	0.100879757	0.079795806	0.070089311
4096	0.061483507	0.068112469	0.049441935	0.049222131
8192	0.050871673	0.046494121	0.030814609	0.035454926
16384	0.040330363	0.032231330	0.020141994	0.026371959

MISE Trend: Sine Noise: $FARIMA(0, 0.2, 0)$				
n	Simulation "d8"	Theor "d8"	Simulation "d10"	Theor "d10"
128	0.251744659	0.290131091	0.348379471	0.251989178
256	0.214928924	0.193352318	0.20541786	0.174618829
512	0.112951872	0.129502140	0.158692616	0.123573436
1024	0.110547951	0.087376732	0.074319167	0.089896035
2048	0.079795806	0.059584328	0.061712354	0.065326166
4096	0.049441935	0.041248179	0.030175723	0.043107368
8192	0.030814609	0.029150833	0.027662929	0.028448428
16384	0.020141994	0.021169561	0.020361623	0.018777135

Table 4.2: Simulated values of the MISE for different sample sizes and values of d. The results are based on 400 simulations of model (1.2) with $FARIMA(0, d, 0)$ residuals, the sine trend function g_1 and the wavelet estimator based on theorem 4.3 with wavelet basis "d4".

MISE Trend: Sine Noise: $FARIMA(0, d, 0)$ Basis: "d4"				
n	d=0.1	d=0.2	d=0.3	d=0.4
128	0.284521469	0.516420047	0.661787865	1.104194018
256	0.210694474	0.263441364	0.537558642	1.42979724
512	0.110584545	0.217604044	0.403889173	0.927229839
1024	0.078905169	0.150284851	0.29832426	0.717419015
2048	0.041133887	0.109213215	0.228981208	0.64283222
4096	0.037871696	0.061483507	0.165045782	0.818104781
8192	0.021438157	0.050871673	0.1444763	0.505236717
16384	0.012234701	0.040330363	0.11107171	0.351823994

Table 4.3: $MISE_{sim}/MISE_{theor}$ for the Jumpsine function and $FARIMA(0, 0.2, 0)$ residuals, in dependence of the jump size Δ. The results are based on 400 simulations and a thresholding estimate according to theorem 4.3, with wavelet basis "d4".

$MISE_{sim}/MISE_{theor}$ Trend: Jumpsine Noise: $FARIMA(0, 0.2, 0)$ Basis: "d4"			
Δ	n=2048	n=4096	n=8192
0.1	1.02984365	1.000066053	0.996328962
0.5	1.044736472	1.007194657	1.004583086
1	1.10352021	1.120497921	1.096100157
10	1.635074083	1.690840646	1.563330038
20	1.301618649	1.234763386	1.207770083
50	1.222581848	1.21888936	1.115174282

Figure 4.2: Simulated values of the mean integrated squared error, $MISE_{sim}$, for different values of the fractional parameter d, plotted against the sample size ($n = 2^7, 2^8, \ldots, 2^{13}$) on log-log-scale (log to the base 2). The results are based on 400 simulations of model (1.2) with the Sine trend function and $FARIMA(0, d, 0)$ residuals with $d = 0.1, 0.2, 0.3, 0.4$. The estimates are based on theorem 4.3 and wavelet basis "d4".

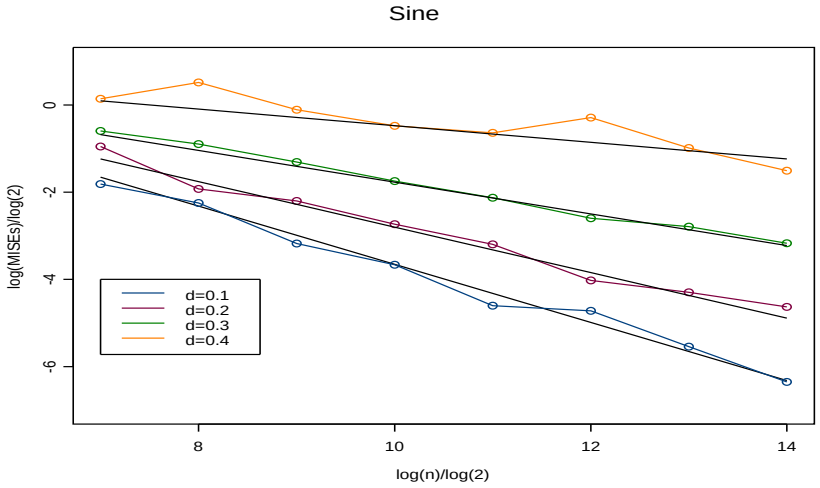


Figure 4.3: Simulated values of $\log_2 MISE_{sim}$ plotted against $\log n$ ($n = 2^7, 2^8, \ldots, 2^{13}$), for trend estimates obtained by kernel smoothing, minimax soft threshold wavelet estimation and data adaptive hard threshold wavelet estimation obtained from theorem 4.3 (both with basis "d4"). The results are based on 400 simulations of model (1.2) with the Sine trend function and $FARIMA(0, 0.2, 0)$ residuals.

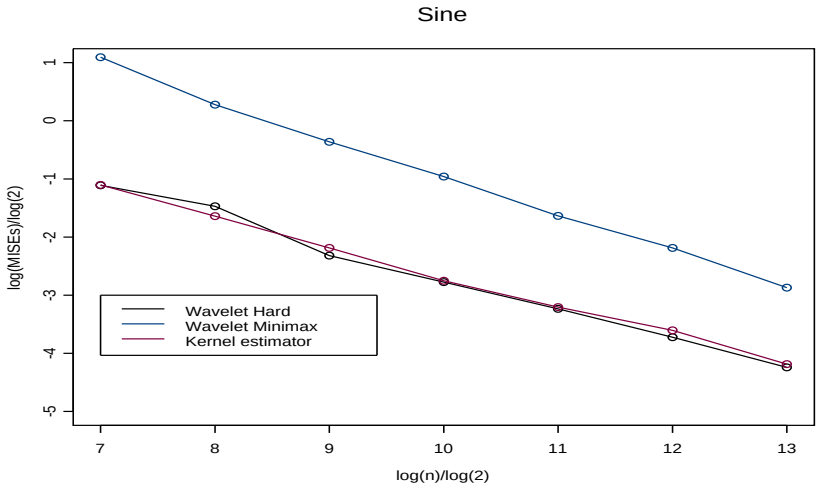


Figure 4.4: Simulated values of $\log_2 MISE_{sim}$ plotted against $\log n$ ($n = 2^7, 2^8, \ldots, 2^{13}$), for trend estimates obtained by kernel smoothing, minimax soft threshold wavelet estimation and data adaptive hard threshold wavelet estimation obtained from theorem 4.3 (both with basis "d4"). The results are based on 400 simulations of model (1.2) with the Sine trend function and $FARIMA(0, 0.4, 0)$ residuals.

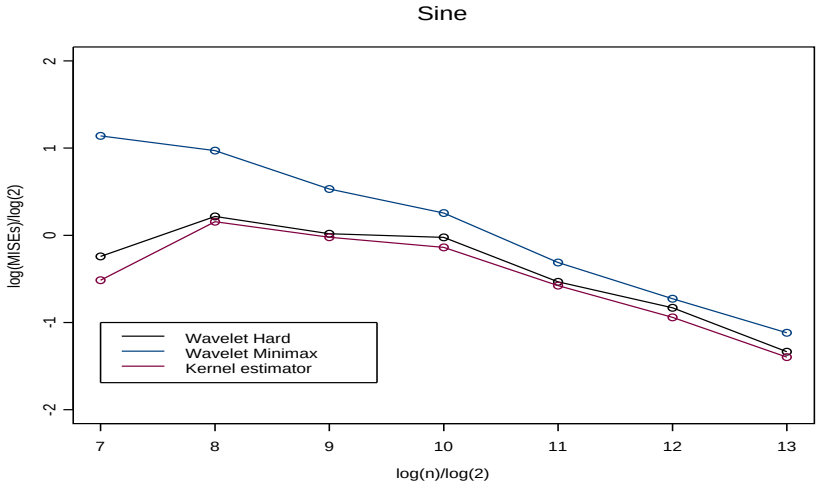


Figure 4.5: Simulated data with sine function plus $FARIMA(0, 0.3, 0)$ process, and trend estimates obtained by optimal kernel smoothing, minimax soft thresholding wavelet estimation and data adaptive hard threshold wavelet estimation according to theorem 4.3 (both with basis "d4").

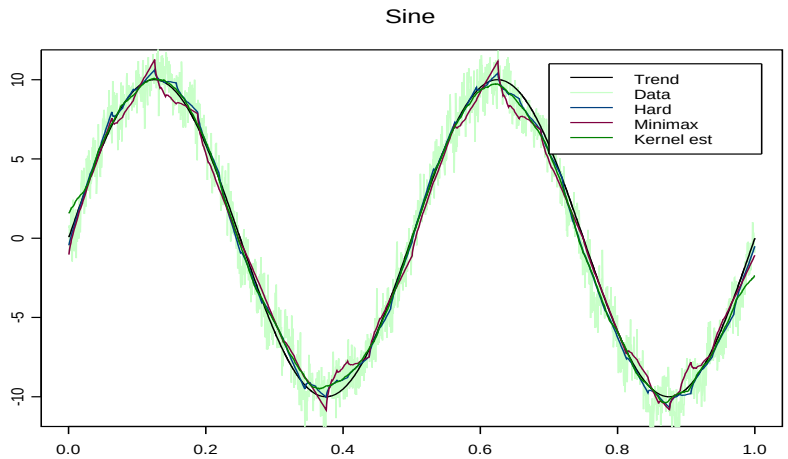


Figure 4.6: Simulated data (figure 4.6a) with Jumpsine function plus $FARIMA(0, 0.3, 0)$ process, and trend estimates obtained by kernel smoothing, minimax soft threshold wavelet estimation and data adaptive hard threshold wavelet estimation obtained from theorem 4.3 (both with basis "d4"). Figure 4.6b shows the coefficients of the data adaptive wavelet estimate.

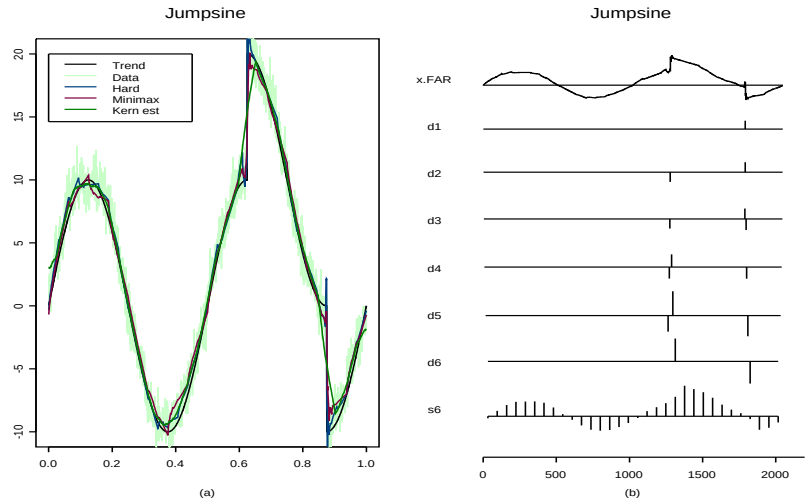


Figure 4.7: Simulated values of $\log_2 MISE_{sim}$ plotted against $\log n$ ($n = 2^7, 2^8, \ldots, 2^{13}$), for trend estimates obtained by kernel smoothing, minimax soft threshold wavelet estimation and data adaptive hard threshold wavelet estimation obtained from theorem 4.3 (both with basis "d4"). The results are based on 400 simulations of model (1.2) with the Jumpsine trend function and $FARIMA(0, 0.2, 0)$ residuals.

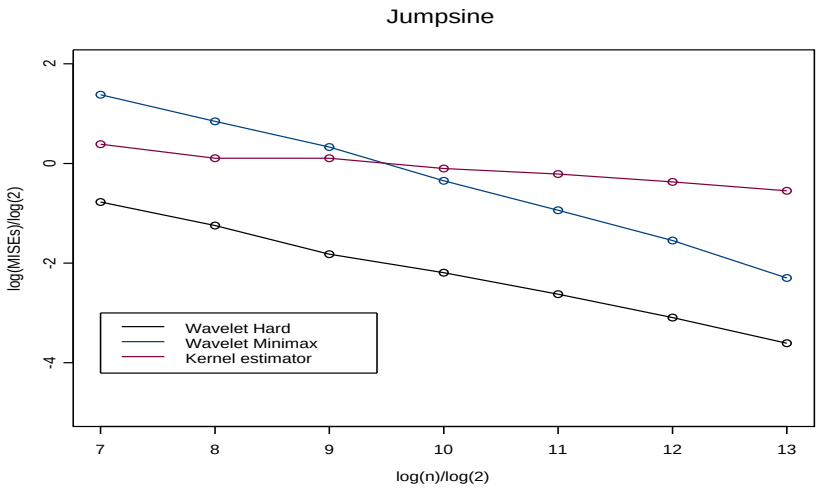


Figure 4.8: Simulated values of $\log_2 MISE_{sim}$ plotted against $\log n$ ($n = 2^7, 2^8, \ldots, 2^{13}$), for trend estimates obtained by kernel smoothing, minimax soft threshold wavelet estimation and data adaptive hard threshold wavelet estimation obtained from theorem 4.3 (both with basis "d4"). The results are based on 400 simulations of model (1.2) with the Jumpsine trend function and $FARIMA(0, 0.4, 0)$ residuals.

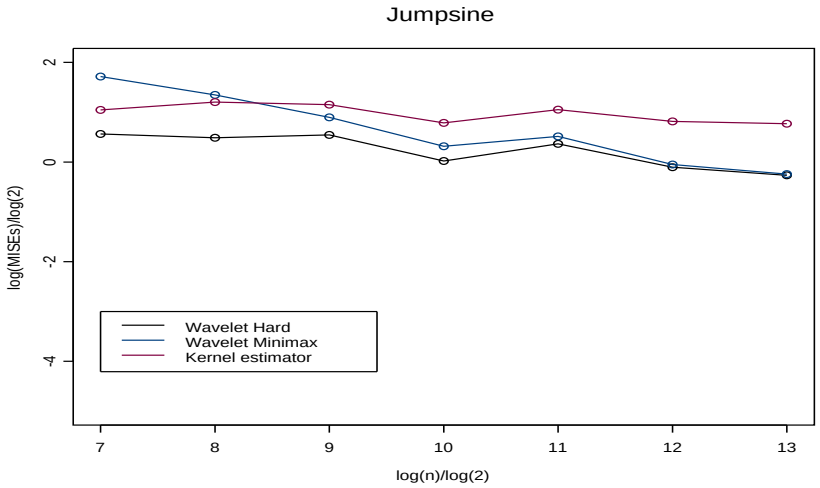


Figure 4.9: Simulated values of $\log_2 MISE_{sim}$ plotted against $\log n$ ($n = 2^7, 2^8, \ldots, 2^{13}$), for trend estimates obtained by kernel smoothing, minimax soft threshold wavelet estimation and data adaptive hard threshold wavelet estimation obtained from theorem 4.3 (both with basis "d4"). The results are based on 400 simulations of model (1.2) with the Sharp trend function and $FARIMA(0, 0.2, 0)$ residuals.

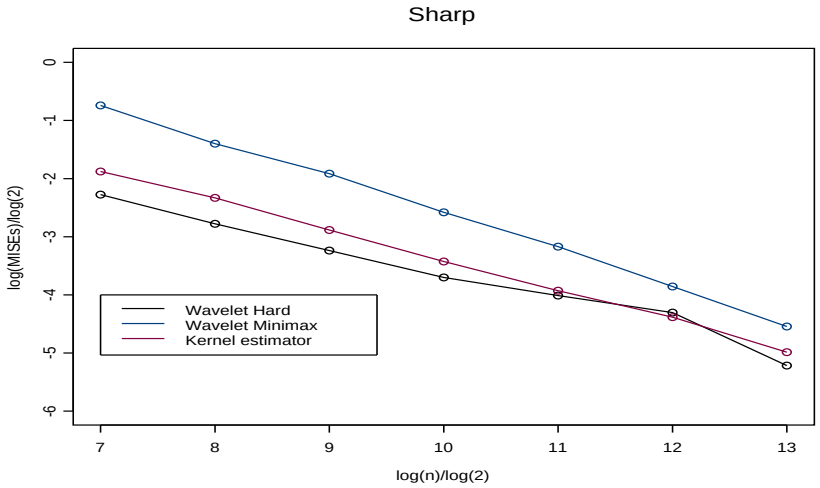


Figure 4.10: Simulated values of $\log_2 MISE_{sim}$ plotted against $\log n$ ($n = 2^7, 2^8, \ldots, 2^{13}$), for trend estimates obtained by kernel smoothing, minimax soft threshold wavelet estimation and data adaptive hard threshold wavelet estimation obtained from theorem 4.3 (both with basis "d4"). The results are based on 400 simulations of model (1.2) with the Sharp trend function and $FARIMA(0, 0.4, 0)$ residuals.

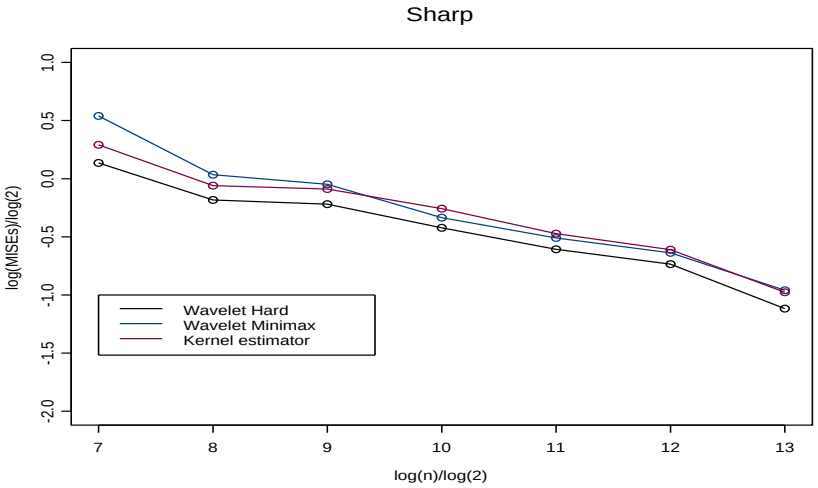


Figure 4.11: Simulated data with Sharp function plus $FARIMA(0, 0.3, 0)$ process, and trend estimates obtained by kernel smoothing, minimax soft threshold wavelet estimation and data adaptive hard threshold wavelet estimation obtained from theorem 4.3 (both with basis "d4").

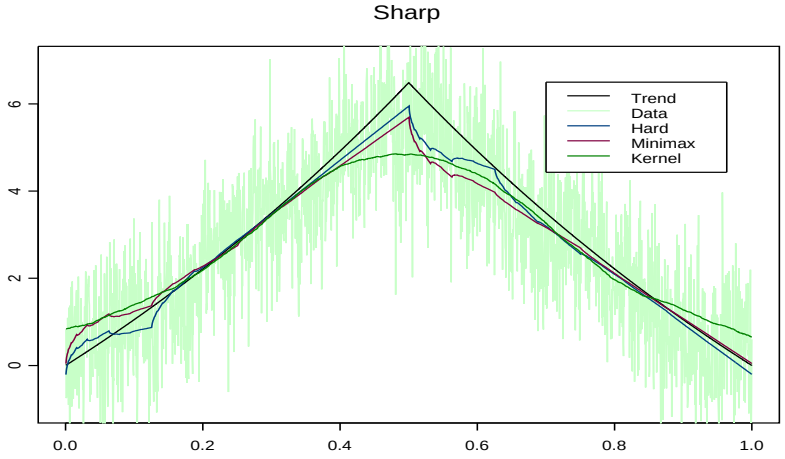


Figure 4.12: Simulated values of $\log_2 MISE_{sim}$ plotted against $\log n$ ($n = 2^7, 2^8, \ldots, 2^{13}$), for trend estimates obtained by kernel smoothing, minimax soft threshold wavelet estimation and data adaptive hard threshold wavelet estimation with $J = \tilde{J}^*$ and thresholds δ_i as in theorem 4.3 (ii) (both with basis "d4"). The results are based on 400 simulations of model (1.2) with the Doppler trend function and $FARIMA(0, 0.2, 0)$ residuals.

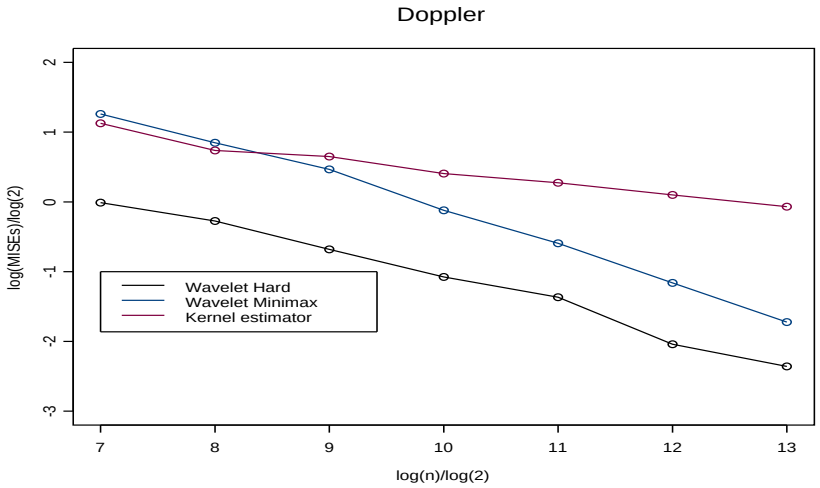


Figure 4.13: Simulated values of $\log_2 MISE_{sim}$ plotted against $\log n$ ($n = 2^7, 2^8, \ldots, 2^{13}$), for trend estimates obtained by kernel smoothing, minimax soft threshold wavelet estimation and data adaptive hard threshold wavelet estimation with $J = \tilde{J}^*$ and thresholds δ_i as in theorem 4.3 (ii) (both with basis "d4"). The results are based on 400 simulations of model (1.2) with the Doppler trend function and $FARIMA(0, 0.4, 0)$ residuals.

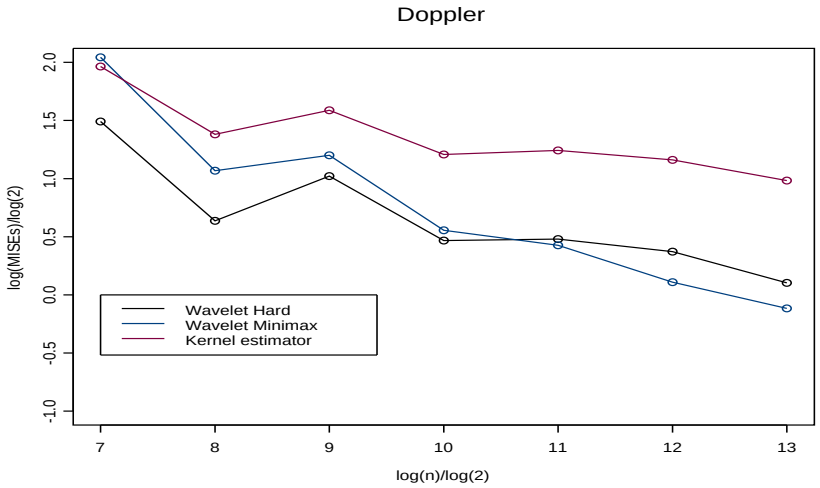
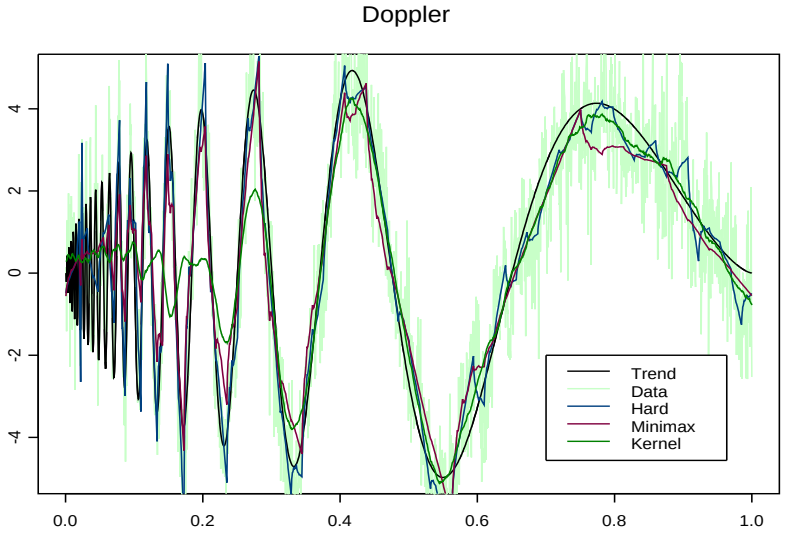


Figure 4.14: Simulated data with the Doppler function plus $FARIMA(0, 0.3, 0)$ process, and trend estimates obtained by kernel smoothing, minimax soft threshold wavelet estimation and data adaptive hard threshold wavelet estimation with $J = \tilde{J}^*$ and thresholds δ_i as in theorem 4.3 (ii) (both with basis "d4").



4.4 Appendix: Proofs

In the proofs of theorem 4.1, 4.2 and 4.3, ϕ and ψ will be assumed to be piecewise differentiable. Analogous results (apart from some expressions in the remainder terms) can be obtained even if ϕ' and ψ' do not exist anywhere, provided that both functions, ϕ and ψ, satisfy a uniform Hölder condition with exponent $1/2$ (see (4.3)). The proofs are analogous, with the difference that instead of the rectangle rule (4.19) the mean value theorem is applied.

Proof. (of theorem 4.1) Let

$$MISE = E\left[\int_0^1 (g(t) - \hat{g}(t))^2\, dt\right] \tag{4.17}$$

denote the mean integrated square error. Combining (4.17) with (3.6) and (3.18), we have

$$MISE = E\left\{\int_0^1 \left[\sum_{k=-N+1}^{N2^J-1} (s_{Jk} - \hat{s}_{Jk})\phi_{Jk}(t) \right.\right.$$

$$\left.\left. + \sum_{j=0}^{q}\sum_{k=-N+1}^{N2^{J+j}-1} \left(d_{jk} - \hat{d}_{jk}\, I(|\hat{d}_{jk}| > \delta_j)\right)\psi_{jk}(t) + \sum_{j=q+1}^{\infty}\sum_{k=-N+1}^{N2^{J+j}-1} d_{jk}\psi_{jk}\right]^2 dt\right\}.$$

Orthonormality of the basis in $L^2(\mathbb{R})$ implies

$$MISE = E\left\{\sum_{k=-N+1}^{N2^J-1} [\hat{s}_{Jk} - s_{Jk}]^2\right\} + E\left\{\sum_{j=0}^{q}\sum_{k=-N+1}^{N2^{J+j}-1} \left[\hat{d}_{jk}\, I(|\hat{d}_{jk}| > \delta_j) - d_{jk}\right]^2\right\}$$

$$+ \sum_{j=q+1}^{\infty}\sum_{k=-N+1}^{N2^{J+j}-1} d_{jk}^2 = \sum_{k=-N+1}^{N2^J-1} [E(\hat{s}_{Jk}) - s_{Jk}]^2 + \sum_{k=-N+1}^{N2^J-1} E\left\{[\hat{s}_{Jk} - E(\hat{s}_{Jk})]^2\right\}$$

$$+ \sum_{j=0}^{q}\sum_{k=-N+1}^{N2^{J+j}-1} \left\{E\left[(\hat{d}_{jk} - d_{jk})^2 I(|\hat{d}_{jk}| > \delta_j)\right] + E\left[d_{jk}^2 I(|\hat{d}_{jk}| \le \delta_j)\right]\right\}$$

$$+ \sum_{j=q+1}^{\infty}\sum_{k=-N+1}^{N2^{J+j}-1} d_{jk}^2 = \Lambda_1 + \Lambda_2 + \Lambda_3 + \Lambda_4. \tag{4.18}$$

The proof then follows from lemmas 4.1 to 4.6 given below. □

Lemma 4.1. *Suppose that the first derivatives of g and ϕ exist except for a finite number of points. Moreover, assume that g' and ϕ' (where they exist) are piecewise continuous and bounded. Then*

$$\Lambda_1 = \sum_{k=-N+1}^{N2^J-1} [E(\hat{s}_{Jk}) - s_{Jk}]^2 = O(n^{-2}2^{2J}).$$

Proof. For the expected value we have

$$E(\hat{s}_{Jk}) = E\left(\frac{1}{n}\sum_{i=1}^{n} Y_i \phi_{Jk}(t_i)\right) = \frac{N^{1/2}2^{J/2}}{n}\sum_{i=1}^{n} g\left(\frac{i}{n}\right)\phi\left(N2^J\frac{i}{n} - k\right)$$

$$= N^{1/2}2^{J/2}\sum_{i=1}^{n}\frac{1}{n}g\left(\frac{i}{n}\right)\phi\left(N2^J\frac{i}{n} - k\right).$$

Assume first that g and ϕ are continuously differentiable and recall the rectangle rule

$$\int_a^b f(t)\,dt = \frac{b-a}{n}\sum_{i=0}^{n-1} f\left(a + i\frac{(b-a)}{n}\right)$$

$$+ O\left(\sum_{i=0}^{n-1}\sup_{t\in I_i}|f'(t)|\cdot\frac{(b-a)^2}{n^2}\right) \tag{4.19}$$

with $I_i = \left[a + i\frac{(b-a)}{n}, a + (i+1)\frac{(b-a)}{n}\right]$. Noting that the support of $\phi\left(N2^Jt - k\right)$ (as a function of t) is $[kN^{-1}2^{-J}, (kN^{-1}+1)\,2^{-J}]$, we obtain

$$E(\hat{s}_{Jk}) = N^{1/2}2^{J/2}\sum_{i=i_1(k)}^{i_2(k)}\frac{1}{n}g\left(\frac{i}{n}\right)\phi\left(N2^J\frac{i}{n} - k\right),$$

with

$$i_1(k) = nkN^{-1}2^{-J}$$

and

$$i_2(k) = n\left(kN^{-1}+1\right)2^{-J}.$$

Thus, the number of nonzero terms in the sum is $n2^{-J}+1$. This together with the rectangle rule for $f(i/n) = g\,(i/n)\,\phi\left(N2^Ji/n - k\right)$ (and integration limits $a = 0$, $b = 1$) implies

$$E(\hat{s}_{Jk}) = N^{1/2}2^{J/2}\int_0^1 g(t)\phi(N2^Jt - k)\,dt + O(n^{-1}2^{J/2}) = s_{Jk} + O(n^{-1}2^{J/2}).$$

Note that here, the factor 2^J from the derivative of $\phi\left(N2^Jt - k\right)$ is compensated by the fact that the number of nonzero terms in the sum is proportional to 2^{-J}.

Assume now more generally that g' and ϕ' exist except for a finite number of points, and, where they exist, they are piecewise continuous and bounded. The result then follows by using the rectangle rule piecewise.
In summary, we have

$$E(\hat{s}_{Jk}) - s_{Jk} = O(n^{-1}2^{J/2}).$$

This implies

$$\Lambda_1 = \sum_{k=-N+1}^{N2^J-1} \left[E\left(\hat{s}_{Jk}\right) - s_{Jk}\right]^2 = O\left(\sum_{k=-N+1}^{N2^J-1} n^{-2}2^J\right) = O(n^{-2}2^{2J})$$

which concludes the proof. □

Lemma 4.2. *Suppose that the first derivative of ϕ exists on $[0, N]$ except for a finite number of points, and, where ϕ' exists, it is piecewise continuous and bounded. Let $J \geq 0$ and $-N+1 \leq k \leq N2^J - 1$. Then*

$$E\left\{\left[\hat{s}_{Jk} - E\left(\hat{s}_{Jk}\right)\right]^2\right\} = C_\phi^2 N^{-1+\alpha} n^{-\alpha} 2^{-J(1-\alpha)} + O(n^{-1})$$

and

$$\Lambda_2 = \sum_{k=-N+1}^{N2^J-1} E\left\{\left[\hat{s}_{Jk} - E\left(\hat{s}_{Jk}\right)\right]^2\right\} = C_\phi^2 n^{-\alpha} N^\alpha 2^{\alpha J} + O(n^{-1}2^J) + O\left(n^{-\alpha}2^{-J(1-\alpha)}\right),$$

where C_ϕ is the constant in (4.5).

Proof. Assume first that ϕ is continuously differentiable. Note that C_ϕ is a positive finite constant (see Li and Xiao (2007)). Now, we consider the behavior of $E\left\{\left[\hat{s}_{Jk} - E\left(\hat{s}_{Jk}\right)\right]^2\right\}$. We have

$$E\left\{\left[\hat{s}_{Jk} - E\left(\hat{s}_{Jk}\right)\right]^2\right\} = E\left\{\left[\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - E\left(Y_i\right)\right)\phi_{Jk}(t_i)\right]^2\right\}$$

$$= E\left[\left(\frac{N^{1/2}2^{J/2}}{n}\sum_{i=1}^{n}\xi_i\phi\left(N2^J\frac{i}{n} - k\right)\right)^2\right]$$

$$= Nn^{-2}2^J\sum_{i=1}^{n}\sum_{l=1}^{n}E(\xi_i\xi_l)\phi\left(N2^J\frac{i}{n} - k\right)\phi\left(N2^J\frac{l}{n} - k\right)$$

$$= Nn^{-2}2^J\sum_{i=nkN^{-1}2^{-J}}^{(1+kN^{-1})n2^{-J}}\sum_{l=nkN^{-1}2^{-J}}^{(1+kN^{-1})n2^{-J}}\gamma(l-i)\,\phi\left(N2^J\frac{i}{n} - k\right)\phi\left(N2^J\frac{l}{n} - k\right)$$

$$= Nn^{-2}2^J \sum_{\substack{i,l=nkN^{-1}2^{-J} \\ i\neq l}}^{(1+kN^{-1})n2^{-J}} \gamma(l-i)\,\phi\left(N2^J\frac{i}{n}-k\right)\phi\left(N2^J\frac{l}{n}-k\right)$$

$$+ Nn^{-2}2^J\gamma(0) \sum_{i=nkN^{-1}2^{-J}}^{(1+kN^{-1})n2^{-J}} \phi^2\left(N2^J\frac{i}{n}-k\right).$$

Equation (4.19) implies

$$Nn^{-2}2^J\gamma(0) \sum_{i=nkN^{-1}2^{-J}}^{(1+kN^{-1})n2^{-J}} \phi^2\left(N2^J\frac{i}{n}-k\right)$$

$$= n^{-1}\gamma(0)\left(\frac{N}{n2^{-J}} \sum_{i=nkN^{-1}2^{-J}}^{(1+kN^{-1})n2^{-J}} \phi^2\left(N2^J\frac{i}{n}-k\right)\right)$$

$$= n^{-1}\gamma(0)\int_0^N \phi^2(t)\,dt + o(n^{-1}).$$

Due to (4.2), this is equal to

$$n^{-1}\gamma(0) + o(n^{-1}) = O(n^{-1}).$$

Hence

$$E\left\{\left[\hat{s}_{Jk} - E\left(\hat{s}_{Jk}\right)\right]^2\right\} =$$

$$Nn^{-2}2^J \sum_{\substack{i,l=nkN^{-1}2^{-J} \\ i\neq l}}^{(1+kN^{-1})n2^{-J}} \gamma(l-i)\,\phi\left(N2^J\frac{i}{n}-k\right)\phi\left(N2^J\frac{l}{n}-k\right) + O(n^{-1}).$$

Using again formula (1.1), we obtain, by analogous arguments as e.g. in Taqqu (1975),

$$E\left\{\left[\hat{s}_{Jk} - E\left(\hat{s}_{Jk}\right)\right]^2\right\}$$

$$\sim C_\gamma Nn^{-2}2^J \sum_{\substack{i,l=nkN^{-1}2^{-J} \\ i\neq l}}^{(1+kN^{-1})n2^{-J}} |l-i|^{-\alpha}\,\phi\left(N2^J\frac{i}{n}-k\right)\phi\left(N2^J\frac{l}{n}-k\right)$$

$$= C_\gamma N^\alpha n^{-1-\alpha}2^{\alpha J} \sum_{i=nkN^{-1}2^{-J}}^{(1+kN^{-1})n2^{-J}} \phi\left(N2^J\frac{i}{n}-k\right)$$

$$\cdot\frac{N2^J}{n} \sum_{\substack{l=nkN^{-1}2^{-J} \\ l\neq i}}^{(1+kN^{-1})n2^{-J}} \left|N2^J\frac{l}{n} - N2^J\frac{i}{n}\right|^{-\alpha} \phi\left(N2^J\frac{l}{n}-k\right).$$

The function $f(x) = \left|x - \left(N2^J \frac{i}{n} - k\right)\right|^{-\alpha} \phi(x)$ is differentiable on $[0, N2^J \frac{i-1}{n} - k] \bigcup [N2^J \frac{i+1}{n} - k, N]$ for all fixed i and n. Therefore the rectangle rule implies

$$\frac{N}{n2^{-J}} \sum_{\substack{l=nkN^{-1}2^{-J} \\ l\neq i}}^{(1+kN^{-1})n2^{-J}} \left|N2^J \frac{l}{n} - N2^J \frac{i}{n}\right|^{-\alpha} \phi\left(N2^J \frac{l}{n} - k\right)$$

$$= \frac{N}{n2^{-J}} \sum_{l=nkN^{-1}2^{-J}}^{i-1} \left|\left(N2^J \frac{l}{n} - k\right) - \left(N2^J \frac{i}{n} - k\right)\right|^{-\alpha} \phi\left(N2^J \frac{l}{n} - k\right)$$

$$+ \frac{N}{n2^{-J}} \sum_{l=i+1}^{(1+kN^{-1})n2^{-J}} \left|\left(N2^J \frac{l}{n} - k\right) - \left(N2^J \frac{i}{n} - k\right)\right|^{-\alpha} \phi\left(N2^J \frac{l}{n} - k\right)$$

$$= \int_0^{N2^J \frac{i-1}{n} - k} \left|x - \left(N2^J \frac{i}{n} - k\right)\right|^{-\alpha} \phi(x)\, dx$$

$$+ \int_{N2^J \frac{i+1}{n} - k}^{N} \left|x - \left(N2^J \frac{i}{n} - k\right)\right|^{-\alpha} \phi(x)\, dx + K_{1,n} + K_{2,n},$$

where

$$K_{1,n} = O\left(\left(\frac{N}{n2^{-J}}\right)^2 \sum_{l=nkN^{-1}2^{-J}}^{i-2} \sup_{x\in I_l(k)} \left|\frac{d}{dx}\left(\left|x - \left(N2^J \frac{i}{n} - k\right)\right|^{-\alpha} \phi(x)\right)\right|\right)$$

with $I_l(k) = [N2^J l/n - k, N2^J(l+1)/n - k]$ and

$$K_{2,n} = O\left(\left(\frac{N}{n2^{-J}}\right)^2 \sum_{l=i+1}^{(1+kN^{-1})n2^{-J}-1} \sup_{x\in I_l(k)} \left|\frac{d}{dx}\left(\left|x - \left(N2^J \frac{i}{n} - k\right)\right|^{-\alpha} \phi(x)\right)\right|\right).$$

Now

$$\left|\left(\frac{N}{n2^{-J}}\right)^2 \sum_{l=nkN^{-1}2^{-J}}^{i-2} \sup_{x\in I_l(k)} \left|\frac{d}{dx}\left(\left|x - \left(N2^J \frac{i}{n} - k\right)\right|^{-\alpha} \phi(x)\right)\right|\right|$$

$$\leq \alpha N^2 \max_{x\in[0,1]} \phi(x) \cdot \left|n^{-2}2^{2J} \sum_{l=nkN^{-1}2^{-J}}^{i-2} \left(\left(N2^J \frac{i}{n} - k\right) - \left(N2^J \frac{l+1}{n} - k\right)\right)^{-1-\alpha}\right|$$

$$+ N^2 \max_{x\in[0,1]} \phi'(x) \cdot \left|n^{-2}2^{2J} \sum_{l=nkN^{-1}2^{-J}}^{i-2} \left(\left(N2^J \frac{i}{n} - k\right) - \left(N2^J \frac{l+1}{n} - k\right)\right)^{-\alpha}\right|$$

$$\leq C_1 n^{-(1-\alpha)} 2^{J(1-\alpha)} \sum_{j=1}^{i-1-nkN^{-1}2^{-J}} j^{-1-\alpha}$$

$$+C_2 n^{-(2-\alpha)} 2^{J(2-\alpha)} \sum_{j=1}^{i-1-nkN^{-1}2^{-J}} j^{-\alpha}$$

$$\leq C_1 n^{-(1-\alpha)} 2^{J(1-\alpha)} \sum_{j=1}^{\infty} j^{-1-\alpha} + C_2 n^{-(2-\alpha)} 2^{J(2-\alpha)} \sum_{j=1}^{n2^{-J}} j^{-\alpha}$$

$$\leq C_1 n^{-(1-\alpha)} 2^{J(1-\alpha)} \sum_{j=1}^{\infty} j^{-1-\alpha} + C_2^* n^{-1} 2^J.$$

Thus

$$K_{1,n} = O(n^{-(1-\alpha)} 2^{J(1-\alpha)}).$$

By analogous arguments we obtain

$$K_{2,n} = O(n^{-(1-\alpha)} 2^{J(1-\alpha)}).$$

This implies

$$E\left\{\left[\hat{s}_{Jk} - E\left(\hat{s}_{Jk}\right)\right]^2\right\}$$

$$= C_\gamma N^\alpha n^{-1-\alpha} 2^{\alpha J} \sum_{i=nkN^{-1}2^{-J}}^{(1+kN^{-1})n2^{-J}} \phi\left(N2^J \frac{i}{n} - k\right) \cdot$$

$$\left(\int_0^{N2^J \frac{i-1}{n} - k} \left|x - \left(N2^J \frac{i}{n} - k\right)\right|^{-\alpha} \phi(x)\, dx \right.$$

$$\left. + \int_{N2^J \frac{i+1}{n} - k}^{N} \left|x - \left(N2^J \frac{i}{n} - k\right)\right|^{-\alpha} \phi(x)\, dx + O(n^{-(1-\alpha)} 2^{J(1-\alpha)})\right) + O(n^{-1})$$

$$= C_\gamma N^\alpha n^{-1-\alpha} 2^{\alpha J} \sum_{i=nkN^{-1}2^{-J}}^{(1+kN^{-1})n2^{-J}} \phi\left(N2^J \frac{i}{n} - k\right) \int_0^{N2^J \frac{i-1}{n} - k} \left|x - \left(N2^J \frac{i}{n} - k\right)\right|^{-\alpha} \phi(x)\, dx$$

$$+ C_\gamma N^\alpha n^{-1-\alpha} 2^{\alpha J} \sum_{i=nkN^{-1}2^{-J}}^{(1+kN^{-1})n2^{-J}} \phi\left(N2^J \frac{i}{n} - k\right) \int_{N2^J \frac{i+1}{n} - k}^{N} \left|x - \left(N2^J \frac{i}{n} - k\right)\right|^{-\alpha} \phi(x)\, dx$$

$$+ O(n^{-1}) = A_1 + A_2 + O(n^{-1}).$$

Using again (4.19), we obtain by analogous arguments as before

$$A_1 = C_\gamma N^{-1+\alpha} n^{-\alpha} 2^{-J(1-\alpha)} \int_0^N \int_0^{y - N2^J \frac{1}{n}} |x-y|^{-\alpha} \phi(x)\, \phi(y)\, dx\, dy + O(n^{-1})$$

and

$$A_2 = C_\gamma N^{-1+\alpha} n^{-\alpha} 2^{-J(1-\alpha)} \int_0^N \int_{y + N2^J \frac{1}{n}}^N |x-y|^{-\alpha} \phi(x)\, \phi(y)\, dx\, dy + O(n^{-1}).$$

Noting that

$$\int_{y-N2^J\frac{1}{n}}^{y+N2^J\frac{1}{n}} |x-y|^{-\alpha}\,\phi(x)\,\phi(y)\,dx \le 2\max_{x\in[0,N]}\left(\phi^2(x)\right)\cdot\int_0^{N2^Jn^{-1}} z^{-\alpha}\,dy = O(n^{-(1-\alpha)}2^{J(1-\alpha)})$$

we obtain

$$\int_0^N\int_{y-N2^J\frac{1}{n}}^{y+N2^J\frac{1}{n}} |x-y|^{-\alpha}\,\phi(x)\,\phi(y)\,dx\,dy = O(n^{-(1-\alpha)}2^{J(1-\alpha)}),$$

and

$$E\left\{[\hat{s}_{Jk}-E(\hat{s}_{Jk})]^2\right\} = C_\gamma N^{-1+\alpha}n^{-\alpha}2^{-J(1-\alpha)}\int_0^N\int_0^N |x-y|^{-\alpha}\,\phi(x)\,\phi(y)\,dx\,dy + O(n^{-1})$$

$$= C_\phi^2 N^{-1+\alpha}n^{-\alpha}2^{-J(1-\alpha)} + O(n^{-1}),$$

where C_ϕ is the constant in (4.5). Hence

$$\Lambda_2 = \sum_{k=-N+1}^{N2^J-1} E\left\{[\hat{s}_{Jk}-E(\hat{s}_{Jk})]^2\right\} = \sum_{k=-N+1}^{N2^J-1}\left(C_\phi^2 N^{-1+\alpha}n^{-\alpha}2^{-J(1-\alpha)} + O(n^{-1})\right)$$

$$= C_\phi^2 n^{-\alpha}N^\alpha 2^{\alpha J} + O(n^{-1}2^J) + O\left(n^{-\alpha}2^{-J(1-\alpha)}\right).$$

In the general case where ϕ' exists except for a finite number of points, and, where it exists, it is piecewise continuous and bounded, the result follows by applying the rectangle rule piecewise. □

Lemma 4.3. *Suppose that the first derivative of ψ exists on $[0,N]$ except for a finite number of points, and, where ψ' exists, it is piecewise continuous and bounded. Let $J \ge 0$, $j \ge 0$ and $-N+1 \le k \le N2^{J+j}-1$. Then*

$$\sigma_j^2 = E\left\{\left[\hat{d}_{jk}-E(\hat{d}_{jk})\right]^2\right\}$$

$$= C_\psi^2 N^{-1+\alpha}n^{-\alpha}2^{-(J+j)(1-\alpha)} + O(n^{-1}),$$

where C_ψ is the constant in (4.6).

Proof. Noting that

$$E\left\{\left[\hat{d}_{jk}-E(\hat{d}_{jk})\right]^2\right\} = E\left\{\left[\frac{N^{1/2}2^{(J+j)/2}}{n}\sum_{i=1}^n \xi_i\psi\left(N2^{J+j}\frac{i}{n}-k\right)\right]^2\right\}$$

the proof is analogous to the proof of lemma 4.2, with the difference that ψ instead of ϕ is used and J is substituted by $J+j$. □

Lemma 4.4. *Suppose that the first r derivatives of g exist and are continuous on $[0,1]$. Then, for all $j \geq 0$ and $0 \leq k \leq N2^{J+j} - 1$,*

$$d_{jk} = \frac{\nu_r}{r!}\, g^{(r)}(kN^{-1}2^{-(J+j)})N^{-\frac{2r+1}{2}}2^{-\frac{2r+1}{2}(J+j)} + o\left(2^{-\frac{2r+1}{2}(J+j)}\right), \tag{4.20}$$

where ν_r is the rth moment of ψ (see (3.23)). Together with the assumptions of lemma 4.3 this yields

$$E(\hat{d}_{jk}) - d_{jk} = O(n^{-1}2^{(J+j)/2}).$$

Proof. Note that

$$d_{jk} = N^{1/2}2^{(J+j)/2}\int_0^1 g\,(t)\,\psi(N2^{J+j}t - k)\,dt$$

$$= N^{1/2}2^{(J+j)/2}\int_{kN^{-1}2^{-(J+j)}}^{(1+kN^{-1})2^{-(J+j)}} g\left(N^{-1}2^{-(J+j)}\left[N2^{J+j}t - k + k\right]\right)\psi(N2^{J+j}t - k)\,dt$$

$$= N^{-1/2}2^{-(J+j)/2}\int_0^N g(N^{-1}2^{-(J+j)}(y+k))\psi(y)\,dy.$$

Since g is r-times continuously differentiable, the local Taylor expansion (see e.g. Zorich (2004), pp. 225-226) of g yields

$$d_{jk} = N^{-1/2}2^{-(J+j)/2}\int_0^N \psi(y)\left[\, g(kN^{-1}2^{-(J+j)}) + N^{-1}2^{-(J+j)}g'(kN^{-1}2^{-(J+j)})y\right.$$

$$\left. + \cdots + \frac{N^{-r}2^{-r(J+j)}}{r!}g^{(r)}(kN^{-1}2^{-(J+j)})y^r\right]\,dy + o\left(2^{-\frac{2r+1}{2}(J+j)}\right).$$

The moment conditions (3.22) and (3.23) then imply

$$d_{jk} = \frac{1}{r!}\, g^{(r)}(kN^{-1}2^{-(J+j)})N^{-\frac{2r+1}{2}}2^{-\frac{2r+1}{2}(J+j)}\int_0^N y^r\psi(y)\,dy + o\left(2^{-\frac{2r+1}{2}(J+j)}\right)$$

$$= \frac{\nu_r}{r!}\, g^{(r)}(kN^{-1}2^{-(J+j)})N^{-\frac{2r+1}{2}}2^{-\frac{2r+1}{2}(J+j)} + o\left(2^{-\frac{2r+1}{2}(J+j)}\right).$$

For $E(\hat{d}_{jk})$ we have

$$E(\hat{d}_{jk}) = \frac{1}{n}\sum_{i=1}^n E\left[Y_i\psi_{jk}(t_i)\right]$$

$$= N^{1/2}2^{(J+j)/2}\sum_{i=1}^n n^{-1}g\left(\frac{i}{n}\right)\psi\left(N2^{J+j}\frac{i}{n} - k\right).$$

Using again the same arguments as in lemma 4.1 for $E(\hat{s}_{Jk})$, we obtain

$$E(\hat{d}_{jk}) = d_{jk} + O(n^{-1}2^{(J+j)/2}).$$

□

Lemma 4.5. *Under the assumptions of lemma 4.4,*

$$\Lambda_4 = \frac{1}{(r!)^2}\frac{1}{2^{2r}-1}N^{-2r}2^{-2r(J+q)}\int_0^1 \nu_r^2\left(g^{(r)}(t)\right)^2\,dt + o\left(2^{-2r(J+q)}\right).$$

Proof. Using (4.18) we have

$$\begin{aligned}\Lambda_4 &= \sum_{j=q+1}^{\infty}\sum_{k=-N+1}^{N2^{J+j}-1} d_{jk}^2 \\ &= \sum_{j=q+1}^{\infty}\sum_{k=-N+1}^{N2^{J+j}-1}\left[\frac{\nu_r}{r!}\,g^{(r)}(kN^{-1}2^{-(J+j)})\right]^2 N^{-(2r+1)}2^{-(2r+1)(J+j)} + o\left(2^{-(2r+1)(J+j)}\right).\end{aligned}$$

Note that continuity of $g^{(r)}$ implies convergence of the Riemann sum. Hence, Λ_4 is equal to

$$\begin{aligned}&\frac{1}{(r!)^2}\sum_{j=q+1}^{\infty} N^{-2r}2^{-2r(J+j)}\left\{\int_0^1 \nu_r^2\left(g^{(r)}(t)\right)^2\,dt + o(1)\right\} + o\left(2^{-2r(J+q)}\right) \\ &= \frac{1}{(r!)^2}\frac{1}{2^{2r}-1}N^{-2r}2^{-2r(J+q)}\int_0^1 \nu_r^2\left(g^{(r)}(t)\right)^2\,dt + o\left(2^{-2r(J+q)}\right).\end{aligned}$$

□

Lemma 4.6. *Let*

$$\hat{q} = \log_2 n^{\frac{\alpha}{2r+\alpha}} + \frac{1}{2r+\alpha}\log_2\left(\frac{\nu_r^2}{(r!)^2C_\psi^2N^{2r+\alpha}}\max_{t\in[0,1]}\left[g^{(r)}(t)\right]^2\right) - J + 1, \quad (4.21)$$

and

$$\lambda_{jk} = E\left[(\hat{d}_{jk}-d_{jk})^2 I(|\hat{d}_{jk}|>\delta_j)\right] + E\left[d_{jk}^2 I(|\hat{d}_{jk}|\le\delta_j)\right].$$

Then under the assumptions of lemma 4.3 and 4.4 the following holds: If $q>\hat{q}$, then for all j with $\hat{q}<j<\frac{2+\alpha}{4r+2+\alpha}\log_2 n - J$,

$$\min_{\delta_j}\lambda_{jk} = d_{jk}^2 + O\left(2^{\alpha(J+j)/2}n^{-(1+\alpha/2)}\right).$$

Proof. Defining $S_0 = 2^{-(2r+\alpha)}C_\psi^2N^{-1+\alpha}n^{-\alpha}2^{-(J+j)(1-\alpha)}$ and taking into account lemma 4.3 we have $S_{1,j} = 2^{-(2r+\alpha)}\sigma_j^2S_0^{-1} = 1 + r_{1,j}$ with $|r_{1,j}|\le r_1 = o(1)$ for all $j\ge\hat{q}$. Moreover, lemma 4.4 implies

$$S_{2,jk} = d_{jk}^2S_0^{-1} = \frac{\nu_r^2}{(r!)^2C_\psi^2N^{2r+\alpha}}\left[g^{(r)}(kN^{-1}2^{-(J+j)})\right]^2 2^{-(2r+\alpha)(J+j-1)}n^\alpha + r_{2,jk}$$

$$\leq \frac{\nu_r^2}{(r!)^2 C_\psi^2 N^{2r+\alpha}} \max_{t\in[0,1]} \left[g^{(r)}(t)\right]^2 n^\alpha \max_{j>\hat{q}} \left\{2^{-(2r+\alpha)(J+j-1)}\right\} + r_2$$

with $r_2 = o(1)$ independent of j and k. Using (4.21) we obtain

$$S_{2,jk} \leq 2^{-(2r+\alpha)} + r_2 \leq 1 + r_1 = S_{1,j}$$

for $j > \hat{q}$ and n large enough, which implies

$$\sigma_j \geq 2^{r+\alpha/2} \max_k |d_{jk}|. \tag{4.22}$$

The mean squared error λ_{jk} can be written as

$$\begin{aligned}
\lambda_{jk} &= E\left[(\hat{d}_{jk} - d_{jk})^2 I(|\hat{d}_{jk}| > \delta_j)\right] + E\left[d_{jk}^2 I(|\hat{d}_{jk}| \leq \delta_j)\right] \\
&= \frac{1}{\sqrt{2\pi}\sigma_j} \int_{|t|>\delta_j} (t - d_{jk})^2 e^{-\frac{1}{2\sigma_j^2}(t - E(\hat{d}_{jk}))^2}\, dt + d_{jk}^2 \frac{1}{\sqrt{2\pi}\sigma_j} \int_{|t|<\delta_j} e^{-\frac{(t-E(\hat{d}_{jk}))^2}{2\sigma_j^2}}\, dt \\
&= A_1 + A_2.
\end{aligned}$$

We approximate A_1 and A_2 separately. Taylor expansion of A_1 with respect to $E(\hat{d}_{jk})$ in the neighborhood of d_{jk} yields

$$\begin{aligned}
A_1 &= \frac{1}{\sqrt{2\pi}\sigma_j} \int_{|t|>\delta_j} (t - d_{jk})^2 e^{-\frac{1}{2\sigma_j^2}(t - E(\hat{d}_{jk}))^2}\, dt \\
&= \frac{1}{\sqrt{2\pi}\sigma_j} \int_{|t|>\delta_j} (t - d_{jk})^2 e^{-\frac{1}{2\sigma_j^2}(t - d_{jk})^2}\, dt \\
&\quad + \frac{E(\hat{d}_{jk}) - d_{jk}}{\sqrt{2\pi}\sigma_j} \int_{|t|>\delta_j} \frac{(t - d_{jk})^3}{\sigma_j^2} e^{-\frac{1}{2\sigma_j^2}(t - d_{jk})^2}\, dt \\
&\quad + O\left(\frac{\left[E(\hat{d}_{jk}) - d_{jk}\right]^2}{\sigma_j} \int_{|t|>\delta_j} \left(\frac{(t - d_{jk})^4}{\sigma_j^4} - \frac{(t - d_{jk})^2}{\sigma_j^2} \right) e^{-\frac{1}{2\sigma_j^2}(t - d_{jk})^2}\, dt \right).
\end{aligned}$$

If $d_{jk} \neq 0$, then lemmas 4.3 and 4.4 imply

$$\begin{aligned}
&\frac{E(\hat{d}_{jk}) - d_{jk}}{\sigma_j} \int_{|t|>\delta_j} \frac{(t - d_{jk})^3}{\sigma_j^2} e^{-\frac{1}{2\sigma_j^2}(t - d_{jk})^2}\, dt \\
&= \sigma_j \left[E(\hat{d}_{jk}) - d_{jk}\right] \int_{|t|>\delta_j/\sigma_j} \left(t - \frac{d_{jk}}{\sigma_j}\right)^3 e^{-\frac{1}{2}\left(t - \frac{d_{jk}}{\sigma_j}\right)^2}\, dt \\
&= O\left(n^{-\alpha/2} 2^{-(J+j)(1-\alpha)/2} \cdot n^{-1} 2^{(J+j)/2}\right) = O\left(2^{\alpha(J+j)/2} n^{-(1+\alpha/2)}\right).
\end{aligned} \tag{4.23}$$

If $d_{jk} = 0$, then

$$\frac{E(\hat{d}_{jk}) - d_{jk}}{\sigma_j} \int_{|t|>\delta_j} \frac{(t-d_{jk})^3}{\sigma_j^2} e^{-\frac{1}{2\sigma_j^2}(t-d_{jk})^2}\, dt = 0$$

and

$$\frac{\left[E(\hat{d}_{jk}) - d_{jk}\right]^2}{\sigma_j} \int_{|t|>\delta_j} \left(\frac{(t-d_{jk})^4}{\sigma_j^4} - \frac{(t-d_{jk})^2}{\sigma_j^2} \right) e^{-\frac{1}{2\sigma_j^2}(t-d_{jk})^2}\, dt$$

$$= \left[E(\hat{d}_{jk}) - d_{jk}\right]^2 \int_{|t|>\delta_j/\sigma_j} \left(\left(t - \frac{d_{jk}}{\sigma_j}\right)^4 - \left(t - \frac{d_{jk}}{\sigma_j}\right)^2 \right) e^{-\frac{1}{2}\left(t-\frac{d_{jk}}{\sigma_j}\right)^2}\, dt$$

$$= O\left(n^{-2} 2^{J+j}\right). \tag{4.24}$$

The condition $j + J < \frac{2+\alpha}{4r+2+\alpha} \log_2 n$ implies that $n^{-2}2^{J+j} = o(2^{\frac{\alpha}{2}(J+j)} n^{-(1+\frac{\alpha}{2})})$ so that

$$A_1 = \frac{1}{\sqrt{2\pi}\sigma_j} \int_{|t|>\delta_j} (t-d_{jk})^2 e^{-\frac{1}{2\sigma_j^2}(t-d_{jk})^2}\, dt + O\left(2^{\alpha(J+j)/2} n^{-(1+\alpha/2)}\right). \tag{4.25}$$

By analogous arguments, we have for $d_{jk} \neq 0$,

$$A_2 = d_{jk}^2 \frac{1}{\sqrt{2\pi}\sigma_j} \int_{|t|<\delta_j} e^{-\frac{(t-E(\hat{d}_{jk}))^2}{2\sigma_j^2}}\, dt$$

$$= d_{jk}^2 \frac{1}{\sqrt{2\pi}\sigma_j} \int_{|t|<\delta_j} e^{-\frac{(t-d_{jk})^2}{2\sigma_j^2}}\, dt + O\left(\frac{d_{jk}^2 \left[E(\hat{d}_{jk}) - d_{jk}\right]}{\sigma_j} \int_{|t|<\delta_j} \frac{(t-d_{jk})}{\sigma_j^2} e^{-\frac{(t-d_{jk})^2}{2\sigma_j^2}}\, dt \right)$$

with

$$\frac{d_{jk}^2 \left[E(\hat{d}_{jk}) - d_{jk}\right]}{\sigma_j} \int_{|t|<\delta_j} \frac{(t-d_{jk})}{\sigma_j^2} e^{-\frac{(t-d_{jk})^2}{2\sigma_j^2}}\, dt$$

$$= \frac{d_{jk}^2 \left[E(\hat{d}_{jk}) - d_{jk}\right]}{\sigma_j} \int_{|t|<\delta_j/\sigma_j} \left(t - \frac{d_{jk}}{\sigma_j}\right) e^{-\frac{1}{2}\left(t-\frac{d_{jk}}{\sigma_j}\right)^2}\, dt$$

$$= O\left(2^{-(2r+1)(J+j)} \cdot n^{-1} 2^{(J+j)/2} \cdot n^{\alpha/2} 2^{(J+j)(1-\alpha)/2}\right)$$

$$= O\left(n^{-(1-\alpha/2)} 2^{-(2r+\alpha/2)(J+j)}\right). \tag{4.26}$$

For $d_{jk} = 0$, we have

$$A_2 = d_{jk}^2 \frac{1}{\sqrt{2\pi}\sigma_j} \int_{|t|<\delta_j} e^{-\frac{(t-E(\hat{d}_{jk}))^2}{2\sigma_j^2}}\, dt = 0.$$

In summary we derived the approximation,

$$\lambda_{jk} = A_1 + A_2$$

$$= \sigma_j^2 \frac{1}{\sqrt{2\pi}} \int_{|t|>\delta_j/\sigma_j} \left(t - \frac{d_{jk}}{\sigma_j}\right)^2 e^{-\frac{1}{2}\left(t-\frac{d_{jk}}{\sigma_j}\right)^2} dt + d_{jk}^2 \frac{1}{\sqrt{2\pi}} \int_{|t|<\delta_j/\sigma_j} e^{-\frac{1}{2}\left(t-\frac{d_{jk}}{\sigma_j}\right)^2} dt$$
$$+ O\left(2^{\alpha(J+j)/2} n^{-(1+\alpha/2)}\right) + O\left(n^{-(1-\alpha/2)} 2^{-(2r+\alpha/2)(J+j)}\right) \quad (4.27)$$

with uniformly bounded error terms (see (4.23), (4.24) and (4.26)). It is then sufficient to show that for all k and for all j with $\hat{q} < j < \frac{2+\alpha}{4r+2+\alpha} \log_2 n - J$ we have

$$\min_{\delta_j} \hat{\lambda}_{jk} = d_{jk}^2$$

where

$$\hat{\lambda}_{jk} = \sigma_j^2 \frac{1}{\sqrt{2\pi}} \int_{|t|>\delta_j/\sigma_j} \left(t - \frac{d_{jk}}{\sigma_j}\right)^2 e^{-\frac{1}{2}\left(t-\frac{d_{jk}}{\sigma_j}\right)^2} dt$$
$$+ d_{jk}^2 \frac{1}{\sqrt{2\pi}} \int_{|t|<\delta_j/\sigma_j} e^{-\frac{1}{2}\left(t-\frac{d_{jk}}{\sigma_j}\right)^2} dt. \quad (4.28)$$

In the following we distinguish two cases: $\delta_j \le \sigma_j$ and $\delta_j > \sigma_j$.
At first let $\delta_j \le \sigma_j$. Recall that $\sigma_j \ge 2^{r+\alpha/2} d_{jk}$ for all k (see (4.22)). Then

$$\frac{1}{\sqrt{2\pi}} \int_{|t|\ge\delta_j/\sigma_j} \left(t - \frac{d_{jk}}{\sigma_j}\right)^2 e^{-\frac{1}{2}\left(t-\frac{d_{jk}}{\sigma_j}\right)^2} dt$$

$$\ge \min_{0\le x\le 2^{-(r+\alpha/2)}} \frac{1}{\sqrt{2\pi}} \int_{|t|\ge 1} (t-x)^2 e^{-\frac{1}{2}(t-x)^2} dt$$

$$\ge \min_{0\le x\le 2^{-1}} \frac{1}{\sqrt{2\pi}} \int_{|t|\ge 1} (t-x)^2 e^{-\frac{1}{2}(t-x)^2} dt > 0.57.$$

Also note that

$$\frac{1}{\sqrt{2\pi}} \int_{|t|<\delta_j/\sigma_j} e^{-\frac{1}{2}\left(t-\frac{d_{jk}}{\sigma_j}\right)^2} dt \ge 0.$$

These two inequalities and (4.28) imply that for all $j > \hat{q}$,

$$\inf_{\delta_j\le\sigma_j} \hat{\lambda}_{jk} = \inf_{\delta_j\le\sigma_j} \left\{ \sigma_j^2 \frac{1}{\sqrt{2\pi}} \int_{|t|>\delta_j/\sigma_j} \left(t - \frac{d_{jk}}{\sigma_j}\right)^2 e^{-\frac{1}{2}\left(t-\frac{d_{jk}}{\sigma_j}\right)^2} dt \right.$$
$$\left. + d_{jk}^2 \frac{1}{\sqrt{2\pi}} \int_{|t|<\delta_j/\sigma_j} e^{-\frac{1}{2}\left(t-\frac{d_{jk}}{\sigma_j}\right)^2} dt \right\}$$
$$\ge \inf_{\delta_j\le\sigma_j} \left\{ \sigma_j^2 \frac{1}{\sqrt{2\pi}} \int_{|t|>\delta_j/\sigma_j} \left(t - \frac{d_{jk}}{\sigma_j}\right)^2 e^{-\frac{1}{2}\left(t-\frac{d_{jk}}{\sigma_j}\right)^2} dt \right\} \ge 0.57\sigma_j^2.$$

For the case with $\delta_j > \sigma_j$ we need some auxiliary results. Without loss of generality let $d_{jk} \geq 0$. At first note that if $\delta_j/\sigma_j > (1 + \frac{d_{jk}}{\sigma_j})$, then

$$\frac{1}{\sqrt{2\pi}} \int_{|t|<\delta_j/\sigma_j} \left[\left(t - \frac{d_{jk}}{\sigma_j} \right)^2 - 1 \right] e^{-\frac{1}{2}\left(t-\frac{d_{jk}}{\sigma_j}\right)^2} dt$$

$$\leq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left[\left(t - \frac{d_{jk}}{\sigma_j} \right)^2 - 1 \right] e^{-\frac{1}{2}\left(t-\frac{d_{jk}}{\sigma_j}\right)^2} dt = 0$$

so that

$$\frac{1}{\sqrt{2\pi}} \int_{|t|<\delta_j/\sigma_j} \left[\left(t - \frac{d_{jk}}{\sigma_j} \right)^2 - 1 \right] e^{-\frac{1}{2}\left(t-\frac{d_{jk}}{\sigma_j}\right)^2} dt \leq 0$$

and

$$\frac{1}{\sqrt{2\pi}} \int_{|t|<\delta_j/\sigma_j} \left(t - \frac{d_{jk}}{\sigma_j} \right)^2 e^{-\frac{1}{2}\left(t-\frac{d_{jk}}{\sigma_j}\right)^2} dt \leq \frac{1}{\sqrt{2\pi}} \int_{|t|<\delta_j/\sigma_j} e^{-\frac{1}{2}\left(t-\frac{d_{jk}}{\sigma_j}\right)^2} dt.$$

Similarily, if $1 \leq \delta_j/\sigma_j \leq (1 + d_{jk}/\sigma_j)$, then

$$\frac{1}{\sqrt{2\pi}} \int_{|t|<\delta_j/\sigma_j} \left[\left(t - \frac{d_{jk}}{\sigma_j} \right)^2 - 1 \right] e^{-\frac{1}{2}\left(t-\frac{d_{jk}}{\sigma_j}\right)^2} dt$$

$$\leq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\delta_j/\sigma_j} \left[\left(t - \frac{d_{jk}}{\sigma_j} \right)^2 - 1 \right] e^{-\frac{1}{2}\left(t-\frac{d_{jk}}{\sigma_j}\right)^2} dt.$$

Moreover, since (4.22), we have $d_{jk}/\sigma_j < 1 \leq \delta_j/\sigma_j$ so that an upper bound is given by

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{d_{jk}/\sigma_j} \left[\left(t - \frac{d_{jk}}{\sigma_j} \right)^2 - 1 \right] e^{-\frac{1}{2}\left(t-\frac{d_{jk}}{\sigma_j}\right)^2} dt$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} \left[t^2 - 1 \right] e^{-\frac{1}{2}t^2} dt = 0.$$

Hence, if $\delta_j > \sigma_j$, we also have the inequality

$$\frac{1}{\sqrt{2\pi}} \int_{|t|<\delta_j/\sigma_j} \left(t - \frac{d_{jk}}{\sigma_j} \right)^2 e^{-\frac{1}{2}\left(t-\frac{d_{jk}}{\sigma_j}\right)^2} dt \leq \frac{1}{\sqrt{2\pi}} \int_{|t|<\delta_j/\sigma_j} e^{-\frac{1}{2}\left(t-\frac{d_{jk}}{\sigma_j}\right)^2} dt.$$

In summary we obtain

$$\frac{1}{\sqrt{2\pi}} \int_{|t|>\delta_j/\sigma_j} \left(t - \frac{d_{jk}}{\sigma_j} \right)^2 e^{-\frac{1}{2}\left(t-\frac{d_{jk}}{\sigma_j}\right)^2} dt$$

$$= 1 - \frac{1}{\sqrt{2\pi}} \int_{|t|<\delta_j/\sigma_j} \left(t - \frac{d_{jk}}{\sigma_j} \right)^2 e^{-\frac{1}{2}\left(t-\frac{d_{jk}}{\sigma_j}\right)^2} dt$$

$$\geq 1 - \frac{1}{\sqrt{2\pi}} \int_{|t|<\delta_j/\sigma_j} e^{-\frac{1}{2}\left(t-\frac{d_{jk}}{\sigma_j}\right)^2} dt.$$

Defining

$$\gamma = \frac{1}{\sqrt{2\pi}} \int_{|t|<\delta_j/\sigma_j} e^{-\frac{1}{2}\left(t-\frac{d_{jk}}{\sigma_j}\right)^2} dt \in [0,1],$$

this inequality together with (4.28) implies that for all $j \geq \hat{q}$, all k and n large enough, $\inf_{\delta_j>\sigma_j} \hat{\lambda}_{jk}$ is equal to

$$\inf_{\delta_j>\sigma_j} \left\{ \sigma_j^2 \frac{1}{\sqrt{2\pi}} \int_{|t|>\delta_j/\sigma_j} \left(t - \frac{d_{jk}}{\sigma_j}\right)^2 e^{-\frac{1}{2}\left(t-\frac{d_{jk}}{\sigma_j}\right)^2} dt + d_{jk}^2 \frac{1}{\sqrt{2\pi}} \int_{|t|<\delta_j/\sigma_j} e^{-\frac{1}{2}\left(t-\frac{d_{jk}}{\sigma_j}\right)^2} dt \right\}$$

so that

$$\inf_{\delta_j>\sigma_j} \hat{\lambda}_{jk} \geq \inf_{\gamma\in[0,1]} \left\{(1-\gamma)\sigma_j^2 + \gamma d_{jk}^2\right\} = d_{jk}^2.$$

Moreover note that the minimum is attained at the border. Now,

$$\inf_{\delta_j} \hat{\lambda}_{jk} = \min\left\{ \inf_{\delta_j\leq\sigma_j} \hat{\lambda}_{jk}, \inf_{\delta_j>\sigma_j} \hat{\lambda}_{jk} \right\} \geq \min\left\{0.57\sigma_j^2, d_{jk}^2\right\}$$

$$\geq \min\left\{0.57 \cdot 2^{2r+\alpha} \cdot \max d_{jk}^2, d_{jk}^2\right\} = d_{jk}^2$$

where the last inequality follows from (4.22). Clearly, the value of d_{jk}^2 is attained if and only if $\delta_j = \infty$.

Finally we obtain

$$\min_{\delta_j} \lambda_{jk}$$

$$= \min_{\delta_j} \hat{\lambda}_{jk} + O\left(2^{\alpha(J+j)/2} n^{-(1+\alpha/2)}\right) + O\left(n^{-(1-\alpha/2)} 2^{-(2r+\alpha/2)(J+j)}\right)$$

$$= d_{jk}^2 + O\left(2^{\alpha(J+j)/2} n^{-(1+\alpha/2)}\right) + O\left(n^{-(1-\alpha/2)} 2^{-(2r+\alpha/2)(J+j)}\right).$$

Now $d_{jk} = O(2^{-\frac{2r+1}{2}(J+j)})$ and the assumption

$$\hat{q} < j < \frac{2+\alpha}{4r+2+\alpha} \log_2 n - J$$

implies

$$2^{-(2r+1)(J+j)} > 2^{\alpha(J+j)/2} n^{-(1+\alpha/2)}$$

and

$$2^{\alpha(J+j)/2} n^{-(1+\alpha/2)} > n^{-(1-\alpha/2)} 2^{-(2r+\alpha/2)(J+j)}.$$

Therefore the remainder term $2^{\alpha(J+j)/2} n^{-(1+\alpha/2)}$ is of smaller order than d_{jk}^2, and $O\left(2^{\alpha(J+j)/2} n^{-(1+\alpha/2)}\right)$ dominates $O\left(n^{-(1-\alpha/2)} 2^{-(2r+\alpha/2)(J+j)}\right)$. This concludes the proof of lemma 4.6.

We now come back to the *proof of the theorem 4.1.* Suppose that ϕ and ψ piecewise differentiable. Define

$$\hat{J} = \log_2 n^{\frac{\alpha}{2r+\alpha}} + \frac{1}{2r+\alpha}\log_2\left(\frac{\nu_r^2}{(r!)^2 C_\psi^2 N^{2r+\alpha}}\max_{t\in[0,1]}\left[g^{(r)}(t)\right]^2\right) + 1$$

and let $J \geq \hat{J}$. Noting that $\Lambda_i \geq 0$ $(i = 1, 2, 3, 4)$ and taking into account lemma 4.2 we obtain for all $q \geq 0$

$$E\int_0^1 \left(g(t) - \hat{g}(t)\right)^2 dt = \Lambda_1 + \Lambda_2 + \Lambda_3 + \Lambda_4 \geq \Lambda_2$$

$$\geq C_\phi^2 n^{-\alpha} N^\alpha 2^{\alpha J} + O(n^{-1}2^J) + O\left(n^{-\alpha}2^{-J(1-\alpha)}\right) \geq C_1 n^{-\frac{2r\alpha}{2r+\alpha}}.$$

Consider now $J < \hat{J}$ and let $q \leq \hat{q}$ where $\hat{q}$ is defined as in lemma 4.6. Lemmas 4.4 and 4.5 imply

$$\sum_{j=\hat{q}+1}^{\infty}\sum_{k=-N+1}^{N2^{J+j}-1} d_{jk}^2 \geq C_2 n^{-\frac{2r\alpha}{2r+\alpha}}.$$

Since $q \leq \hat{q}$ we have

$$E\int_0^1 \left(g(t) - \hat{g}(t)\right)^2 dt = \Lambda_1 + \Lambda_2 + \Lambda_3 + \Lambda_4$$

$$\geq \Lambda_4 = \sum_{j=q+1}^{\infty}\sum_{k=-N+1}^{N2^{J+j}-1} d_{jk}^2 \geq \sum_{j=\hat{q}+1}^{\infty}\sum_{k=-N+1}^{N2^{J+j}-1} d_{jk}^2 \geq C_2 n^{-\frac{2r\alpha}{2r+\alpha}}.$$

For the other case with $q > \hat{q}$, taking into account Λ_3 in (4.18) and lemma 4.6 leads to

$$E\int_0^1 \left(g(t) - \hat{g}(t)\right)^2 dt \geq \Lambda_3$$

$$= \sum_{j=0}^{q}\sum_{k=-N+1}^{N2^{J+j}-1}\left\{E\left[(\hat{d}_{jk} - d_{jk})^2 I(|\hat{d}_{jk}| > \delta_j)\right] + E\left[d_{jk}^2 I(|\hat{d}_{jk}| \leq \delta_j)\right]\right\}$$

$$= \sum_{j=0}^{q}\sum_{k=-N+1}^{N2^{J+j}-1}\lambda_{jk} \geq \sum_{j=0}^{\hat{q}}\sum_{k=-N+1}^{N2^{J+j}-1}\lambda_{jk} + \sum_{j=\hat{q}+1}^{q}\sum_{k=-N+1}^{N2^{J+j}-1}\min_{\delta_j}\lambda_{jk}$$

$$\geq \sum_{j=\hat{q}+1}^{\hat{q}+1}\sum_{k=-N+1}^{N2^{J+j}-1}\min_{\delta_j}\lambda_{jk} = \sum_{j=\hat{q}+1}^{\hat{q}+1}\sum_{k=-N+1}^{N2^{J+j}-1} d_{jk}^2$$

$$+O\left(2^{(1+\alpha/2)(J+\hat{q})}n^{-(1+\alpha/2)}\right) \geq C_3 n^{-\frac{2r\alpha}{2r+\alpha}}.$$

In summary, we obtain a lower bound

$$\min_{\{\delta_j\},q,J} E\int_0^1 \left(g(t)-\hat{g}(t)\right)^2 dt = \min_{\{\delta_j\},q,J}\left(\Lambda_1+\Lambda_2+\Lambda_3+\Lambda_4\right) \geq Cn^{-\frac{2r\alpha}{2r+\alpha}}.$$

It is shown in the proof of theorem 4.2 that equality can indeed be achieved, by using a specific choice of δ_j, q, J and C. This concludes the proof of theorem 4.1. □

Proof. (of theorem 4.2):
Under conditions of theorem 4.2 and taking into account lemmas 4.3 and4.4 we obtain

$$\Lambda_3 = \sum_{j=0}^{q}\sum_{k=-N+1}^{N2^{J+j}-1} E\left\{\left[\hat{d}_{jk}-d_{jk}\right]^2\right\} = \sum_{j=0}^{q}\sum_{k=-N+1}^{N2^{J+j}-1}\left(\sigma_j^2+\left(E(\hat{d}_{jk})-d_{jk}\right)^2\right)$$

$$= \frac{C_\psi^2}{2^\alpha-1}\left(2^{\alpha(q+1)}-1\right)N^\alpha n^{-\alpha}2^{\alpha J}+O(n^{-1}2^{J+q})+O\left(n^{-\alpha}2^{-J(1-\alpha)}\right)+O(n^{-2}2^{2(J+q)}).$$

This and lemmas 4.1, 4.2 and 4.5 imply that the expression in (4.18) will take following form

$$\begin{aligned} MISE_g(q,J) &= \Lambda_1+\Lambda_2+\Lambda_3+\Lambda_4 \\ &= \left(C_\phi^2-\frac{C_\psi^2}{2^\alpha-1}\right)N^\alpha n^{-\alpha}2^{\alpha J}+\frac{2^\alpha C_\psi^2}{2^\alpha-1}N^\alpha n^{-\alpha}2^{\alpha(J+q)} \\ &\quad +\frac{1}{(r!)^2}\frac{1}{2^{2r}-1}N^{-2r}2^{-2r(J+q)}\int_0^1 \nu_r^2\left(g^{(r)}(t)\right)^2 dt \\ &\quad +o\left(2^{-2r(J+q)}\right)+O(n^{-1}2^{J+q})+O\left(n^{-\alpha}2^{-J(1-\alpha)}\right). \end{aligned} \tag{4.29}$$

Now let q and J be such that $MISE$ is minimal. Then (4.18), $\delta_j=0$ and

$$MISE_g(q,J)-MISE_g(q+1,J)<0$$

imply

$$MISE_g(q,J)-MISE_g(q+1,J) = \sum_{k=-N+1}^{N2^{J+q+1}-1} d_{q+1,k}^2 - \sum_{k=-N+1}^{N2^{J+q+1}-1}\sigma_{q+1}^2 < 0.$$

By analogous argument as in the proof of (4.22) the last inequality implies, together with lemmas 4.3 and 4.4, that for n large enough we have

$$C_\psi^2 n^{-\alpha}N^\alpha 2^{\alpha(J+q+1)} \geq \frac{1}{(r!)^2}N^{-2r}2^{-2r(J+q+1)}\int_0^1 \nu_r^2\left(g^{(r)}(t)\right)^2 dt$$

and

$$q \geq \log_2 n^{\frac{\alpha}{2r+\alpha}} - J - 1 + \frac{1}{2r+\alpha}\log_2\left[\frac{\int_0^1 \nu_r^2\left(g^{(r)}(t)\right)^2\,dt}{C_\psi^2(r!)^2}\right] - \log_2 N. \qquad (4.30)$$

On the other hand

$$MISE_g(q,J) - MISE_g(q-1,J) < 0$$

implies the second necessary condition

$$C_\psi^2 n^{-\alpha} N^\alpha 2^{\alpha(J+q)} \leq \frac{1}{(r!)^2} N^{-2r} 2^{-2r(J+q)} \int_0^1 \nu_r^2\left(g^{(r)}(t)\right)^2\,dt$$

so that

$$q \leq \log_2 n^{\frac{\alpha}{2r+\alpha}} + \frac{1}{2r+\alpha}\log_2\left[\frac{\int_0^1 \nu_r^2\left(g^{(r)}(t)\right)^2\,dt}{C_\psi^2(r!)^2}\right] - \log_2 N - J. \qquad (4.31)$$

Note that q and J are integers. Then the inequalities (4.30) and (4.31) imply that the value

$$q^* = \left\lfloor \log_2 n^{\frac{\alpha}{2r+\alpha}} + \frac{1}{2r+\alpha}\log_2\left[\frac{\int_0^1 \nu_r^2\left(g^{(r)}(t)\right)^2\,dt}{C_\psi^2(r!)^2}\right] - \log_2 N \right\rfloor - J \qquad (4.32)$$

asymptotically minimizes the $MISE$. Using the definition of $\Delta_n(g, C_\psi)$ in (4.7) we conclude

$$q^* = \log_2 n^{\frac{\alpha}{2r+\alpha}} + \frac{1}{2r+\alpha}\log_2\left[\frac{\int_0^1 \nu_r^2\left(g^{(r)}(t)\right)^2\,dt}{C_\psi^2(r!)^2}\right] - \log_2 N - J - \Delta_n(g, C_\psi).$$

Note that if $\Delta_n(g, C_\psi) \neq 0$, then for every fixed J there exists a unique q^* such (4.30) and (4.31) hold.

Combining these results with (4.29) yields

$$MISE_g(q^*, J) = 2^{-\alpha\Delta_n(g,C_\psi)}\left(C_\phi^2 - \frac{C_\psi^2}{2^\alpha - 1}\right) N^\alpha n^{-\alpha} 2^{\alpha J}$$

$$+ \left(\frac{2^{2r\Delta_n(g,C_\psi)}}{2^{2r}-1} + \frac{2^{\alpha(1-\Delta_n(g,C_\psi))}}{2^\alpha - 1}\right) C_\psi^{\frac{4r}{2r+\alpha}} \left(\frac{\nu_r^2}{(r!)^2}\int_0^1 \left(g^{(r)}(t)\right)^2 dt\right)^{\frac{\alpha}{2r+\alpha}} n^{-\frac{2r\alpha}{2r+\alpha}}$$

$$+ O(n^{-1}2^J) + o(n^{-\frac{2r\alpha}{2r+\alpha}}) + O\left(n^{-\alpha}2^{-J(1-\alpha)}\right). \qquad (4.33)$$

The first term is monotonically decreasing in J if $(2^\alpha - 1)C_\phi^2 < C_\psi^2$ and monotonically increasing if $(2^\alpha - 1)C_\phi^2 > C_\psi^2$. The second term does not depend of J.

Hence, if $(2^\alpha - 1)C_\phi^2 > C_\psi^2$, then the optimal decomposition level J is equal to zero. Note that the optimal decomposition level is not unique, since the same asymptotic expressions will be achieved for all integers J such that $2^J = o(n^{\frac{\alpha}{2r+\alpha}})$. Combining this with the last formulae implies that $MISE_g(q, J)$ is equal to

$$\left(\frac{2^{2r\Delta_n(g,C_\psi)}}{2^{2r}-1} + \frac{2^{\alpha(1-\Delta_n(g,C_\psi))}}{2^\alpha - 1}\right) C_\psi^{\frac{4r}{2r+\alpha}} \left(\frac{\nu_r^2}{(r!)^2}\int_0^1 \left(g^{(r)}(t)\right)^2 dt\right)^{\frac{\alpha}{2r+\alpha}} n^{-\frac{2r\alpha}{2r+\alpha}} + o\left(n^{-\frac{2r\alpha}{2r+\alpha}}\right). \tag{4.34}$$

On the other hand, suppose that $(2^\alpha - 1)C_\phi^2 < C_\psi^2$. Then, taking into account (4.32) and $q \geq 0$ (see (3.18)), we have

$$0 \leq J \leq \left\lfloor \log_2 n^{\frac{\alpha}{2r+\alpha}} + \frac{1}{2r+\alpha}\log_2\left[\frac{\int_0^1 \nu_r^2\left(g^{(r)}(t)\right)^2 dt}{C_\psi^2 (r!)^2}\right] - \log_2 N \right\rfloor.$$

Hence, the optimal choice of J is

$$J = \left\lfloor \log_2 n^{\frac{\alpha}{2r+\alpha}} + \frac{1}{2r+\alpha}\log_2\left[\frac{\int_0^1 \nu_r^2\left(g^{(r)}(t)\right)^2 dt}{C_\psi^2 (r!)^2}\right] - \log_2 N \right\rfloor. \tag{4.35}$$

Due to (4.32) this also implies $q^* = 0$.

Note that (3.18) with $q \geq 0$ and $\delta_j = 0$ always includes at least one level of mother wavelets. The case where the estimate includes father wavelets only is automatically considered in theorem 4.1, namely if $q = 0$ and $\delta_0 = \infty$. To conclude the proof we also need to compare with the estimate that includes father wavelets only. Thus consider

$$\tilde{g}(t) = \sum_{k=-N+1}^{N2^J-1} \hat{s}_{Jk}\phi_{Jk}(t),$$

and denote the corresponding mean integrated square error by $MISE_g(-1, J)$. Then

$$MISE_g(-1, J) = \sum_{k=-N+1}^{N2^J-1} \left[E\left(\hat{s}_{Jk}\right) - s_{Jk}\right]^2 + \sum_{k=-N+1}^{N2^J-1} E\left\{\left[\hat{s}_{Jk} - E(\hat{s}_{Jk})\right]^2\right\} + \sum_{j=0}^{\infty}\sum_{k=-N+1}^{N2^{J+j}-1} d_{jk}^2.$$

Let J^* be such that $MISE_g(-1, J^*)$ is minimal. Then

$$MISE_g(-1, J^*) - MISE_g(-1, J^* + 1) < 0$$

and

$$MISE_g(-1, J^*) - MISE_g(-1, J^* - 1) < 0.$$

Suppose n is large enough. Then the elementary calculations similar to those above show that the optimal decomposition level J^* is given by

$$J^* = \left\lfloor \log_2 n^{\frac{\alpha}{2r+\alpha}} + \frac{1}{2r+\alpha} \log_2 \left[\frac{\int_0^1 \nu_r^2 \left(g^{(r)}(t)\right)^2 dt}{C_\phi^2 (2^\alpha - 1)(r!)^2} \right] - \log_2 N \right\rfloor + 1. \qquad (4.36)$$

Defining $\Delta_n(g, C_\phi)$ as in (4.8), the corresponding $MISE$ is equal to

$$\left(\frac{2^{2r\Delta_n(g,C_\phi)}}{2^{2r}-1} + \frac{2^{\alpha(1-\Delta_n(g,C_\phi))}}{2^\alpha - 1} \right) \left(C_\phi^2(2^\alpha - 1)\right)^{\frac{2r}{2r+\alpha}} \left(\frac{\nu_r^2}{(r!)^2} \int_0^1 \left(g^{(r)}(t)\right)^2 dt \right)^{\frac{\alpha}{2r+\alpha}} n^{-\frac{2r\alpha}{2r+\alpha}}$$

$$+ o\left(n^{-\frac{2r\alpha}{2r+\alpha}}\right).$$

Now, let $(2^\alpha - 1)C_\phi^2 > C_\psi^2$. Suppose that J defined by (4.36) and the estimator consisting of father wavelets only minimizes the $MISE$. Now

$$MISE_g(0, J) - MISE_g(-1, J+1)$$

$$= n^{-\alpha} N^\alpha 2^{\alpha J} \left(C_\psi^2 - C_\phi^2(2^\alpha - 1)\right) + o(n^{-\frac{2r\alpha}{2r+\alpha}}).$$

so that, for n large enough,

$$MISE_g(0, J) - MISE_g(-1, J+1) < 0$$

which is a contradiction. Thus, it follows that the best J is equal to zero, q is defined by (4.32) and the $MISE$ is as in (4.34).

Now suppose that

$$C_\phi^2(2^\alpha - 1) < C_\psi^2,$$

$q = 0$ and J given by (4.35) minimizes the $MISE$. Consider

$$MISE_g(-1, J+1) - MISE_g(0, J)$$

$$= n^{-\alpha} N^\alpha 2^{\alpha J} \left(C_\phi^2(2^\alpha - 1) - C_\psi^2\right) + o(n^{-\frac{2r\alpha}{2r+\alpha}}).$$

Using the same argument as before, $MISE_g(-1, J+1) - MISE_g(0, J) < 0$ for n large enough. Thus, the best estimator includes father wavelets only and the optimal decomposition level is defined by (4.36).

In conclusion we consider the case $\Delta_n(g, C_\psi) = 0$. Suppose that

$$(2^\alpha - 1)C_\phi^2 > C_\psi^2,$$

$J = 0$ and q as in (4.32) minimizes the $MISE$. Now

$$MISE_g(q,0) - MISE_g(q-1,0)$$

$$= \left(\frac{1}{2^{2r}-1} + \frac{2^\alpha}{2^\alpha - 1} - \frac{2^{2r}}{2^{2r}-1} - \frac{1}{2^\alpha - 1}\right) C_\psi^{\frac{4r}{2r+\alpha}} \left(\frac{\nu_r^2}{(r!)^2}\int_0^1 \left(g^{(r)}(t)\right)^2 dt\right)^{\frac{\alpha}{2r+\alpha}} n^{-\frac{2r\alpha}{2r+\alpha}}$$
$$+ o\left(n^{-\frac{2r\alpha}{2r+\alpha}}\right) = o\left(n^{-\frac{2r\alpha}{2r+\alpha}}\right).$$

Then, for every fixed J there exist two smoothing parameters that minimize the $MISE$ asymptotically. The same follows also for the case $(2^\alpha - 1)C_\phi^2 < C_\psi^2$ and $\Delta_n(g, C_\phi) = 0$. This concludes the proof. □

The extension to functions with piecewise continuous r-th derivative follows from the following lemma 4.7 which will be proved in a similar manner as lemma 4.5 and lemma 4.6 in Li and Xiao (2007).

Lemma 4.7. *Suppose that the assumptions of theorem 4.3 hold. Then*

(i) If $(2^\alpha - 1)C_\phi^2 > C_\psi^2$ *then*

$$\sum_{j=q^*+1}^{q}\sum_{k=-N+1}^{N2^{J+j}-1} \lambda_{jk} + \sum_{j=q+1}^{\infty}\sum_{k=-N+1}^{N2^{J+j}-1} d_{jk}^2$$

$$= \frac{\nu_r^2}{(r!)^2}\frac{1}{2^{2r}-1} N^{-2r} 2^{-2r(J+q^*)} \int_0^1 \left(g^{(r)}(t)\right)^2 dt + o\left(2^{-2r(J+q^*)}\right).$$

(ii) If $(2^\alpha - 1)C_\phi^2 < C_\psi^2$ *then*

$$\sum_{j=0}^{q}\sum_{k=-N+1}^{N2^{J+j}-1} \lambda_{jk} + \sum_{j=q+1}^{\infty}\sum_{k=-N+1}^{N2^{J+j}-1} d_{jk}^2$$

$$= \frac{\nu_r^2}{(r!)^2}\frac{1}{2^{2r}-1} N^{-2r} 2^{-2r(J-1)} \int_0^1 \left(g^{(r)}(t)\right)^2 dt + o\left(2^{-2rJ}\right).$$

Proof. At first, recall that

$$\lambda_{jk} = E\left[(\hat{d}_{jk} - d_{jk})^2 I(|\hat{d}_{jk}| > \delta_j)\right] + E\left[d_{jk}^2 I(|\hat{d}_{jk}| \le \delta_j)\right].$$

Let $(2^\alpha - 1)C_\phi^2 > C_\psi^2$ and ω, β denote positive numbers satisfying $2\omega + \beta = 1$. Set

$$S_1 = \sum_{j=q^*+1}^{q}\sum_{k=-N+1}^{N2^{J+j}-1} E\left[(\hat{d}_{jk} - d_{jk})^2\right] I(|d_{jk}| > \omega\delta_j),$$

$$S_2 = \sum_{j=q^*+1}^{q} \sum_{k=-N+1}^{N2^{J+j}-1} E\left[(\hat{d}_{jk} - d_{jk})^2 I(|\hat{d}_{jk} - d_{jk}| > (\omega+\beta)\delta_j)\right].$$

The triangle inequality then implies

$$\sum_{j=q^*+1}^{q} \sum_{k=-N+1}^{N2^{J+j}-1} E\left[(\hat{d}_{jk} - d_{jk})^2 I(|\hat{d}_{jk}| > \delta_j)\right] \leq S_1 + S_2.$$

It follows from (4.15) and $I(|d_{jk}| > \omega\delta_j) = 1$ that

$$d_{jk} \geq \omega \cdot 2\sqrt{e} C_\psi N^{-(1-\alpha)/2} \ln n \cdot n^{-\alpha/2} 2^{-(J+j)(1-\alpha)/2}.$$

Due to lemma 4.4, there are only finitely many ks satisfying this inequality. Therefore, by using lemmas 4.3 and 4.4 we derive

$$\begin{aligned} S_1 &= \sum_{j=q^*+1}^{q} \sum_{k=-N+1}^{N2^{J+j}-1} \left\{\sigma_j^2 + E\left[(E\hat{d}_{jk} - d_{jk})^2\right]\right\} I(|d_{jk}| > \omega\delta_j) \\ &= O(n^{-\alpha} 2^{-(J+q^*)(1-\alpha)}) + O(n^{-2} 2^{J+q}) = o\left(2^{-2r(J+q)}\right). \end{aligned}$$

Note that the triangle also inequality implies

$$I\left\{|\hat{d}_{jk} - d_{jk}| > (\omega+\beta)\delta_j\right\} \leq I\left\{\frac{|\hat{d}_{jk} - E\hat{d}_{jk}|}{\sigma_j} > \frac{(\omega+\beta)\delta_j}{\sigma_j} - \frac{|E\hat{d}_{jk} - d_{jk}|}{\sigma_j}\right\}.$$

Using this, lemmas 4.3 and 4.4, the condition (4.15) and the general inequality $P(Z > \eta) \leq e^{-\frac{1}{2}\eta^2}$ for standard normal variables Z (see e.g. Pollard 1984, p. 191) leads to

$$\begin{aligned} E\left[I\left\{|\hat{d}_{jk} - d_{jk}| > (\omega+\beta)\delta_j\right\}\right] &\leq P\left(\frac{|\hat{d}_{jk} - E\hat{d}_{jk}|}{\sigma_j} > \frac{(\omega+\beta)\delta_j}{\sigma_j} - \frac{|E\hat{d}_{jk} - d_{jk}|}{\sigma_j}\right) \\ &\leq 2n^{-2(\omega+\beta)^2 e \ln n} + o\left(n^{-2(\omega+\beta)^2 e \ln n}\right). \end{aligned} \tag{4.37}$$

Due to Hölders inequality and last inequality we achieve

$$\begin{aligned} S_2 &= \sum_{j=q^*+1}^{q} \sum_{k=-N+1}^{N2^{J+j}-1} E\left[(\hat{d}_{jk} - E\hat{d}_{jk})^2 I(|\hat{d}_{jk} - d_{jk}| > (\omega+\beta)\delta_j)\right] \\ &\quad + \sum_{j=q^*+1}^{q} \sum_{k=-N+1}^{N2^{J+j}-1} (E\hat{d}_{jk} - d_{jk})^2 E\left[I(|\hat{d}_{jk} - d_{jk}| > (\omega+\beta)\delta_j)\right] \\ &\leq \sum_{j=q^*+1}^{q} \sum_{k=-N+1}^{N2^{J+j}-1} \sqrt{E\left[(\hat{d}_{jk} - E\hat{d}_{jk})^4\right] \cdot E\left[I(|\hat{d}_{jk} - d_{jk}| > (\omega+\beta)\delta_j)\right]} \end{aligned}$$

$$+O(n^{-2}2^{2(J+q)}n^{-2(\omega+\beta)^2 e\ln n}) = o\left(2^{-2r(J+q)}\right).$$

Since

$$\sum_{j=q+1}^{\infty}\sum_{k=-N+1}^{N2^{J+j}-1} d_{jk}^2 = O(2^{-(J+q)}) = o(2^{-2r(J+q^*)}),$$

then for proving this theorem, one has to show essentially that

$$\sum_{j=q^*+1}^{q}\sum_{k=-N+1}^{N2^{J+j}-1} E\left[d_{jk}^2 I(|\hat{d}_{jk}| \le \delta_j)\right]$$

$$= \frac{\nu_r^2}{(r!)^2}\frac{1}{2^{2r}-1}N^{-2r}2^{-2r(J+q^*)}\int_0^1 \left(g^{(r)}(t)\right)^2 dt + o\left(2^{-2r(J+q^*)}\right).$$

For this propose, assume that $\omega > 0$, and define

$$S_3 = \sum_{j=q^*+1}^{q}\sum_{k=-N+1}^{N2^{J+j}-1} d_{jk}^2 I(|\hat{d}_{jk}| \le \delta_j),$$

$$S_4 = \sum_{j=q^*+1}^{q}\sum_{k=-N+1}^{N2^{J+j}-1} d_{jk}^2 I(|d_{jk}| \le (1+\omega)\delta_j),$$

$$S_5 = \sum_{j=q^*+1}^{q}\sum_{k=-N+1}^{N2^{J+j}-1} d_{jk}^2 I(|d_{jk}| \le (1-\omega)\delta_j),$$

$$S_6 = \sum_{j=q^*+1}^{q}\sum_{k=-N+1}^{N2^{J+j}-1} d_{jk}^2 I(|\hat{d}_{jk} - d_{jk}| > \omega\delta_j),$$

Then the triangle inequality implies

$$S_4 - S_6 \le S_3 \le S_5 + S_6.$$

Using lemma 4.4 and the assumption (4.15), we see that $I(|d_{jk}| \le (1+\omega)\delta_j) = I(|d_{jk}| \le (1-\omega)\delta_j) = 1$ for n sufficiently large, and

$$S_4 = S_5 = \frac{\nu_r^2}{(r!)^2}\frac{1}{2^{2r}-1}N^{-2r}2^{-2r(J+q^*)}\int_0^1 \left(g^{(r)}(t)\right)^2 dt + o\left(2^{-2r(J+q^*)}\right). \quad (4.38)$$

On the other hand, due to (4.37) and lemma 4.4

$$ES_6 = \sum_{j=q^*+1}^{q}\sum_{k=-N+1}^{N2^{J+j}-1} d_{jk}^2 E\left[I(|\hat{d}_{jk} - d_{jk}| > \omega\delta_j)\right] = o\left(2^{-2r(J+q^*)}\right).$$

Combining this and (4.38), we conclude the proof for case $(2^\alpha - 1)C_\phi^2 > C_\psi^2$. The proof extension to the case $(2^\alpha - 1)C_\phi^2 < C_\psi^2$ can be provided in a similar manner. □

As already mentioned, similar results (with the difference of remainder terms) can be attained even if ϕ' and ψ' do not exist everywhere and both ϕ and ψ satisfy a uniform Hölder condition of exponent $1/2$ (see (4.3)). The proofs are analogous to the those above, with the difference that instead of rectangle rule (4.19) we use Mean-value theorem.
For example the corresponding result for lemma 4.1 will take following representation:

Lemma 4.8. *Suppose that ϕ satisfy a uniform Hölder condition of exponent $1/2$, the first derivative of g exists except for a finite number of points. Moreover, assume that g' (where it exists) are piecewise continuous and bounded.Then*

$$\Lambda_1 = \sum_{k=-N+1}^{N2^J-1} [E(\hat{s}_{Jk}) - s_{Jk}]^2 = O(n^{-1}2^J).$$

Proof. At first, we assume that g is continuously differentiable. The Mean-value theorem implies, that it exists $\hat{t}_i \in [t_i - n^{-1}, t_i]$ with

$$\frac{1}{n} g\left(\hat{t}_i\right) \phi\left(N2^J \hat{t}_i - k\right) = \int_{t_i - n^{-1}}^{t_i} g\left(t\right) \phi\left(N2^J t - k\right)\, dt.$$

This implies

$$\begin{aligned}
&\frac{1}{n} g\left(t_i\right) \phi\left(N2^J t_i - k\right) \\
&= \frac{1}{n} g\left(\hat{t}_i\right) \phi\left(N2^J \hat{t}_i - k\right) + \frac{1}{n}\left(g\left(t_i\right) \phi\left(N2^J t_i - k\right) - g\left(\hat{t}_i\right) \phi\left(N2^J \hat{t}_i - k\right)\right) \\
&= \int_{t_i - n^{-1}}^{t_i} g\left(t\right) \phi\left(N2^J t - k\right)\, dt + \frac{1}{n}\left(g\left(t_i\right) \phi\left(N2^J t_i - k\right) - g\left(t_i\right) \phi\left(N2^J \hat{t}_i - k\right)\right) \\
&\qquad + \frac{1}{n}\left(g\left(t_i\right) \phi\left(N2^J \hat{t}_i - k\right) - g\left(\hat{t}_i\right) \phi\left(N2^J \hat{t}_i - k\right)\right). \qquad (4.39)
\end{aligned}$$

Due to the differentiability of g and the boundedness of ϕ we derive

$$\begin{aligned}
&\left|\frac{1}{n} \phi\left(N2^J \hat{t}_i - k\right)\left(g\left(t_i\right) - g\left(\hat{t}_i\right)\right)\right| \\
&\leq \frac{1}{n} \max_{t \in [0,N]} \left\{|\phi(t)|\right\} \left|g\left(t_i\right) - g\left(\hat{t}_i\right)\right| \leq C n^{-2}. \qquad (4.40)
\end{aligned}$$

Furthermore, the boundedness of g and uniform Hölder condition of ϕ imply

$$\left|\frac{1}{n} g\left(t_i\right)\left(\phi\left(N2^J t_i - k\right) - \phi\left(N2^J \hat{t}_i - k\right)\right)\right|$$

$$\leq \frac{1}{n} \max_{t\in[0,1]} \{|g(t)|\} \left|\phi\left(N2^J t_i - k\right) - \phi\left(N2^J \hat{t}_i - k\right)\right| \leq C n^{-3/2} 2^{J/2}. \qquad (4.41)$$

Now, combining (4.39) with (4.40) and (4.41) we derive

$$\frac{1}{n} g\left(t_i\right) \phi\left(N2^J t_i - k\right) = \int_{t_i - n^{-1}}^{t_i} g\left(t\right) \phi\left(N2^J t - k\right)\, dt + O(n^{-2}) + O(n^{-3/2}2^{J/2}).$$

This relation implies

$$E(\hat{s}_{Jk}) = N^{1/2} 2^{J/2} \sum_{i=1}^{n} \frac{1}{n} g\left(\frac{i}{n}\right) \phi\left(N2^J \frac{i}{n} - k\right)$$

$$= N^{1/2} 2^{J/2} \int_0^1 g(t)\phi(N2^J t - k)\, dt + O(n^{-1/2}).$$

Note, that the number of nonzero terms in the previous sum is $n2^{-J} + 1$. This gives us following estimation:

$$\Lambda_1 = O(n^{-1}2^J).$$

The proof in case g' is only piecewise differentiable, follows by the same argument as before, by applying it piecewise.

□

In the following we take a look at the case ϕ and ψ satisfy a uniform Hölder condition of exponent $1/2$ and not piecewise differentiable, and derive the equivalents to the lemmas 4.2-4.6.

Lemma 4.9. *Suppose that ϕ satisfy a uniform Hölder condition of exponent $1/2$. Let $J \geq 0$ and $0 \leq k \leq N2^J - 1$, then*

$$E\left\{\left[\hat{s}_{Jk} - E\left(\hat{s}_{Jk}\right)\right]^2\right\} = C_\phi^2 N^{-1+\alpha} n^{-\alpha} 2^{-J(1-\alpha)} + O(n^{-1}) + O\left(n^{-1/2-\alpha} 2^{-J(1/2-\alpha)}\right)$$

and

$$\Lambda_2 = C_\phi^2 n^{-\alpha} N^\alpha 2^{\alpha J} + O(n^{-1}2^J) + O\left(n^{-1/2-\alpha} 2^{J(1/2+\alpha)}\right) + O\left(n^{-\alpha} 2^{-J(1-\alpha)}\right),$$

where C_ϕ is the constant in (4.5).

Lemma 4.10. *Suppose that ψ satisfy a uniform Hölder condition of exponent $1/2$. Let $J \geq 0$, $j \geq 0$ and $0 \leq k \leq N2^{J+j} - 1$. Then*

$$\sigma_j^2 = C_\psi^2 N^{-1+\alpha} n^{-\alpha} 2^{-(J+j)(1-\alpha)} + O(n^{-1}) + O\left(n^{-1/2-\alpha} 2^{-(J+j)(1/2-\alpha)}\right),$$

where C_ψ is the constant in (4.6).

Lemma 4.11. *Suppose that the first r derivatives of g exist and continuous on $[0,1]$. Then, for all $j \geq 0$ and $0 \leq k \leq N2^{J+j} - 1$,*

$$d_{jk} = \frac{\nu_r}{r!}\, g^{(r)}(kN^{-1}2^{-(J+j)})N^{-\frac{2r+1}{2}}2^{-\frac{2r+1}{2}(J+j)} + o\left(2^{-\frac{2r+1}{2}(J+j)}\right)$$

and under the assumptions of lemma 4.10 yields

$$E(\hat{d}_{jk}) - d_{jk} = O(n^{-1/2}),$$

where ν_r is rth moment of ψ (see (3.23)). Note that under assumption $J + q < (2r+1)^{-1}\log_2 n$ the remainder term $O(n^{-1/2}2^{-(J+j)/2})$ is of smaller order than d_{jk}.

Lemma 4.12. *Under the assumptions of lemma 4.11,*

$$\Lambda_4 = \frac{1}{(r!)^2}\frac{1}{2^{2r}-1}N^{-2r}2^{-2r(J+q)}\int_0^1 \nu_r^2\left(g^{(r)}(t)\right)^2\, dt + o\left(2^{-2r(J+q)}\right).$$

Lemma 4.13. *Let $\hat{q}$ as in (4.21) and*

$$\lambda_{jk} = E\left[(\hat{d}_{jk} - d_{jk})^2 I(|\hat{d}_{jk}| > \delta_j)\right] + E\left[d_{jk}^2 I(|\hat{d}_{jk}| \leq \delta_j)\right].$$

Then under the assumption of lemma 4.10 and 4.11 the following holds: If $q > \hat{q}$, then for all j with $\hat{q} < j < \frac{2+\alpha}{4r+2+\alpha}\log_2 n - J$,

$$\min_{\delta_j} \lambda_{jk} = d_{jk}^2 + O\left(n^{-(1+\alpha)/2}2^{-(J+j)(1-\alpha)/2}\right).$$

Proof. The proof is analogous to the proof of lemma 4.6, with the difference that we derive instead of (4.23)-(4.27) following relations:

$$\frac{E(\hat{d}_{jk}) - d_{jk}}{\sigma_j}\int_{|t|>\delta_j}\frac{(t-d_{jk})^3}{\sigma_j^2}e^{-\frac{1}{2\sigma_j^2}(t-d_{jk})^2}\, dt = O\left(n^{-(1+\alpha)/2}2^{-(J+j)(1-\alpha)/2}\right),$$

$$\left[E(\hat{d}_{jk}) - d_{jk}\right]^2\int_{|t|>\delta_j/\sigma_j} t^4 e^{-\frac{t^2}{2}}\, dt = O\left(n^{-1}\right),$$

$$A_1 = \frac{1}{\sqrt{2\pi}\sigma_j}\int_{|t|>\delta_j}(t-d_{jk})^2 e^{-\frac{1}{2\sigma_j^2}(t-d_{jk})^2}\, dt + O\left(n^{-(1+\alpha)/2}2^{-(J+j)(1-\alpha)/2}\right),$$

$$\frac{d_{jk}^2\left[E(\hat{d}_{jk}) - d_{jk}\right]}{\sigma_j}\int_{|t|<\delta_j}\frac{(t-d_{jk})}{\sigma_j^2}e^{-\frac{(t-d_{jk})^2}{2\sigma_j^2}}\, dt = O\left(n^{-(1-\alpha)/2}2^{-(2r+1/2+\alpha/2)(J+j)}\right),$$

and

$$\lambda_{jk} = A_1 + A_2$$

$$= \sigma_j^2 \frac{1}{\sqrt{2\pi}} \int_{|t|>\delta_j/\sigma_j} \left(t - \frac{d_{jk}}{\sigma_j}\right)^2 e^{-\frac{1}{2}\left(t-\frac{d_{jk}}{\sigma_j}\right)^2} dt + d_{jk}^2 \frac{1}{\sqrt{2\pi}} \int_{|t|<\delta_j/\sigma_j} e^{-\frac{1}{2}\left(t-\frac{d_{jk}}{\sigma_j}\right)^2} dt$$
$$+O\left(n^{-(1+\alpha)/2} 2^{-(J+j)(1-\alpha)/2}\right) + O\left(n^{-(1-\alpha)/2} 2^{-(2r+1/2+\alpha/2)(J+j)}\right).$$

□

Based on lemmas 4.8 to 4.13, theorem 4.2 and 4.3 (in analogous way as before) can be derived for the case ϕ and ψ satisfy a uniform Hölder condition of exponent 1/2.

Chapter 5

Bootstrap testing for discontinuities

The issue addressed in this chapter is testing whether there is at least one jump in the trend function g. Our approach is provided by wavelets, since this are likely to pick up even local deviations from smoothness. More specifically, as discussed in chapter 4 and in Beran and Shumeyko (2011a), an appropriate wavelet estimator $\hat{g}(t)$ consists of two estimators and the high-resolution part $\hat{g}_{high}(t)$ is able to capture jumps. This fact we have used to develop a bootstrap based test for the null hypothesis that g is continuous everywhere against the alternative that g has at least one isolated jump. As in the previous chapter, we give here the detailed presentation of the results derived by us before in Beran and Shumeyko (2011b). The chapter is organized as follows. Basic definitions are introduced in section 5.1, the main results are given in section 5.2. The results are illustrated by a simulation study in section 5.3. And finally, proofs are in appendix 5.4 conclude the chapter.

5.1 Notations

Throughout this chapter, the assumption of section 4.1 will be fulfilled. Additionally, ϕ and ψ will be assumed to be continuous and piecewise differentiable. As already mentioned in chapter 4, the estimator defined by equation (3.18) can

be understood as a combination of two components, a smoothing component

$$\hat{g}_{low}(t) = \sum_{k=-N+1}^{N2^J-1} \hat{s}_{Jk}\phi_{Jk}(t) \tag{5.1}$$

consisting of the scaling function decomposition only with no thresholding, and a higher resolution component

$$\hat{g}_{high}(t) = \sum_{j=0}^{q} \sum_{k=-N+1}^{N2^{J+j}-1} \hat{d}_{jk}\, I(|\hat{d}_{jk}| > \delta_j)\psi_{jk}(t) \tag{5.2}$$

where estimated coefficients are filtered by thresholding. The first component provides a good estimate of smooth functions, whereas the second is useful for modeling discontinuities.

We will assume, that the following conditions for decomposition level, smoothing parameter and threshold are satisfied:

$$J = J_n = \frac{\alpha}{2r+\alpha}\log_2 n + C \text{ where } C \in \mathbb{R}, \tag{5.3}$$

$$q = q_n = \log_2 n - J, \tag{5.4}$$

and

$$2^{J+j}\delta_j^2 \to 0,\ 2^{(J+j)(2r+1)}\delta_j^2 \to \infty,\ \delta_j^2\,\frac{n^\alpha 2^{(J+j)(1-\alpha)}}{\ln n} \to \infty. \tag{5.5}$$

In the following we will use the notation $\kappa(n)$ for the sequence

$$\kappa(n) = \frac{1}{4}C_\psi^{-2}N^{\alpha-1}\delta_j^2\frac{n^\alpha 2^{(J+j)(1-\alpha)}}{\ln n}. \tag{5.6}$$

Note that, by definition, $\kappa(n) \to \infty$.

Consider now the null hypothesis $H_0 : g \in C[0,1]$ against the alternative that g is only piecewise continuous, with a finite number of isolated jumps. The following statistics will be used to define a bootstrap based test of H_0. Denote by X_i the residuals after extracting the estimated trend $\hat{g}(t_i)$ given in (3.18),

$$X_i = Y_i - \hat{g}(t_i) = \xi_i + g(t_i) - \hat{g}_{low}(t_i) - \hat{g}_{high}(t_i).$$

For blocks of size l $(1 \le l \le n)$ define block sums $\zeta_1, \ldots, \zeta_{n-l+1}$ by

$$\zeta_i = X_i + \cdots + X_{i+l-1} \ \ (1 \le i \le n-l+1).$$

Then draw $\zeta_1^*, \ldots, \zeta_k^*$ randomly with replacement from $\{\zeta_1, \ldots, \zeta_{n-l+1}\}$ and calculate

$$T_{kl}^* = a_l^{-1}\left(k^{-1/2}\sum_{i=1}^{k}\zeta_i^*\right) \tag{5.7}$$

where $a_l = C_\gamma^{1/2} l^{1-\alpha/2}$. Similarily, residuals $X_{i,low}$ are defined by subtracting the low resolution component only, i.e.

$$X_{i,low} = Y_i - \hat{g}_{low}(t_i) = \xi_i + g(t_i) - \hat{g}_{low}(t_i)$$

and corresponding block sums are denoted by

$$\zeta_{i,low} = X_{i,low} + \cdots + X_{i+l-1,low} \ \ (1 \leq i \leq n-l+1).$$

Sampling randomly with replacement from $\{\zeta_{1,low}, \ldots, \zeta_{n-l+1,low}\}$ we obtain $\zeta^*_{1,low}, \ldots, \zeta^*_{k,low}$ and

$$T^*_{kl,low} = a_l^{-1} \left(k^{-1/2} \sum_{i=1}^{k} \zeta^*_{i,low} \right). \tag{5.8}$$

In the following section it is shown that under H_0, T^*_{kl} and $T^*_{kl,low}$ have approximately the same distribution whereas this is no longer the case under H_1.

5.2 Main results

5.2.1 Asymptotic distribution of T^*_{kl} and $T^*_{kl,low}$

Throughout the chapter, E_n and Var_n will denote expectation and variance with respect to the distribution generated by the resampling procedure, conditionally on the observations $Y_1, ..., Y_n$. Similarily, $P_n(A)$ will denote the conditional probability of A. For block lengths increasing like a power of n (but slower than n) and bootstrap sample size k tending to infinity, the expected value and variance of $T^*_{kl,low}$ can be approximated as follows. (Note that, since E_n, Var_n and P_n are random variables depending on $Y_1, ..., Y_n$, errors of approximations are of the form $o_p(\cdot)$ instead of $o(\cdot)$.)

Theorem 5.1. *Let $l = O\left(n^{1-\epsilon}\right)$ with $1 - 2(2r+\alpha)^{-1} < \epsilon < 1 - (2r+\alpha)^{-1}$. Assume that $g \in L^2[0,1]$, g' exists on $[0,1]$ except for a finite number of points, and, where it exists, g' is piecewise continuous and bounded. Then the following holds as $\min\{k,n\} \to \infty$.*

(a) If g has no jumps on $(0,1)$, then

$$E_n\left(T^*_{kl,low}\right) = o_p(1)$$

and

$$Var_n\left(T^*_{kl,low}\right) = \tilde{\sigma}^2 + o_p(1)$$

where

$$\tilde{\sigma}^2 = 2\sigma^2(1-\alpha)^{-1}(2-\alpha)^{-1}. \tag{5.9}$$

(b) If g has at least one jump on $(0,1)$, then

$$E_n\left(T^*_{kl,low}\right) = o_p(1)$$

and there exists a constant $C^* > 0$ such that

$$Var_n\left(T^*_{kl,low}\right) = \tilde{\sigma}^2 + v_n + o_p(1)$$

with

$$v_n = C^* n^{\alpha(1-\epsilon)-\frac{\alpha}{2r+\alpha}}.$$

Theorem 5.1 means that under H_1, the variance of $T^*_{kl,low}$ increases by the additional term v_n. Since $\epsilon < 1-(2r+\alpha)^{-1}$, v_n diverges to infinity as n increases. Note also that ϵ can be specified as required in theorem 5.1 without a priori knowledge of α, because $0 < \alpha < 1$ and $r \geq 1$ implies

$$1 - \frac{2}{2r+\alpha} < 1 - \frac{2}{2r+1} < 1 - \frac{1}{2r} < 1 - \frac{1}{2r+\alpha}.$$

We may thus choose an ϵ from the interval $[1-2(2r+1)^{-1}, 1-1/(2r)]$.
The following result shows that the standardized bootstrapped statistic $T^*_{kl,low}$ is asymptotically normal under H_0 and H_1 respectively and convergence is uniform on $\mathbb{R}$.

Theorem 5.2. *Under the assumptions of theorem 5.1 the following holds:*

(a) If g has no jumps on $(0,1)$, then

$$\sup_{x\in\mathbb{R}} \left| P_n(T^*_{kl,low} \leq x) - \Phi\left(\frac{x}{\tilde{\sigma}}\right) \right| = o_p(1).$$

(b) If g has at least one jump on $(0,1)$, then

$$\sup_{x\in\mathbb{R}} \left| P_n(v_n^{-1/2} T^*_{kl,low} \leq x) - \Phi(x) \right| = o_p(1)$$

where v_n is as in theorem 5.1b).

The next step is to compare the distributions of $T^*_{kl,low}$ and T^*_{kl}. First it is shown that under the null hypothesis both variables are asymptotically equivalent. However, in contrast to $T^*_{kl,low}$, the variance of T^*_{kl} is not affected by isolated jumps.

Theorem 5.3. *Under the assumptions of theorem 5.1 the following holds:*

(a) If g has no jumps on $(0,1)$, then

$$E_n\left(T_{kl}^*\right) - E_n\left(T_{low,kl}^*\right) = o_p(n^{\frac{1}{2}\alpha(1-\epsilon)-\frac{1}{4}\kappa(n)})$$

and

$$Var_n\left(T_{kl}^*\right) - Var_n\left(T_{low,kl}^*\right) = o_p(n^{\alpha(1-\epsilon)-\frac{1}{2}\kappa(n)}).$$

Moreover,

$$T_{kl}^* = T_{low,kl}^* + o_p(n^{\frac{1}{2}\alpha(1-\epsilon)-\frac{1}{4}\kappa(n)}).$$

(b) If g has at least one jump on $(0,1)$, then

$$E_n\left(T_{kl}^*\right) = o_p(1)$$

and

$$Var_n\left(T_{kl}^*\right) = \tilde{\sigma}^2 + o_p(1).$$

Finally, it is shown that T_{kl}^* is also asymptotically normal, however, as implied by theorem 5.3, with the same standardization under H_0 and H_1.

Theorem 5.4. *Under the assumptions of theorem 5.1 the following holds.*

(a) If g has no jumps on $(0,1)$, then

$$\sup_{x\in\mathbb{R}}\left|P_n(T_{kl}^* \leq x) - \Phi\left(\tilde{\sigma}^{-1}x\right)\right| = o_p(1)$$

and

$$\sup_{x\in\mathbb{R}}\left|P_n(T_{kl}^* \leq x) - P_n(T_{kl,low}^* \leq x)\right| = o_p(1).$$

(b) If g has at least one jump on $(0,1)$, then

$$\sup_{x\in\mathbb{R}}\left|P_n(T_{kl}^* \leq x) - \Phi\left(\tilde{\sigma}^{-1}x\right)\right| = o_p(1).$$

5.2.2 Testing for jumps

Given theorems 5.1 to 5.4, a natural approach to testing for jumps is to compare $H_0 : Var(T_{kl,low}^*) = Var(T_{kl}^*)$ with $H_1 : Var(T_{kl,low}^*) > Var(T_{kl}^*)$. This can be done as follows. In a first step, m bootstrap samples $\{\zeta_{1,low}^{*(i)}, \ldots, \zeta_{k,low}^{*(i)}\}$ $(i = 1, 2, ..., m)$ are drawn and $T_{kl,low}^{*(1)}, ..., T_{kl,low}^{*(m)}$ are computed. The test statistic is defined by

$$W_{low} = \tilde{\sigma}^{-2}\sum_{i=1}^{m}\left(T_{kl,low}^{*(i)} - \bar{T}_{kl,low}^*\right)^2,$$

with $\bar{T}^*_{kl,low} = m^{-1}\sum_{i=1}^m T^{*(i)}_{kl,low}$. In a second step, critical values for W_{low} are calculated by resampling in the same way s sets of m bootstraped values of T^*_{kl}, i.e. we have $A_1 = \left\{T^{*(1,1)}_{kl}, ..., T^{*(m,1)}_{kl}\right\}$,..., $A_s = \left\{T^{*(1,s)}_{kl}, ..., T^{*(m,s)}_{kl}\right\}$. For each A_j $(1 \leq j \leq s)$, the statistic

$$W_j = \tilde{\sigma}^{-2}\sum_{i=1}^m \left(T^{*(i,j)}_{kl} - \bar{T}^{*(j)}_{kl}\right)^2,$$

with $\bar{T}^{*(j)}_{kl} = m^{-1}\sum_{i=1}^m T^{*(i,j)}_{kl}$ is computed and critical values for W_{low} are obtained from the empirical distribution of $W_1, ..., W_s$. The null hypothesis is rejected at the level of significance α, if W_{low} is above the $(1-\alpha)-$quantile. Theorems 5.1 to 5.4 imply that asymptotically, the level of the test is correct and the power converges to one.

5.3 Simulations

The finite sample behavior of the proposed test is illustrated by the following simulations. In the main part of the simulation study, residuals are generated by a Gaussian $FARIMA(0, d, 0)$ process (note that $\alpha = 1 - 2d$). Five trend functions g (figures 5.1 and 5.8b) are considered:

a) Sine function: $g_1(t) = 10\sin(4\pi t)$ (figure 5.1a)

b) "JumpSine" function (figure 5.1b):

$$g_2(t) = 10\sin(4\pi t) + \Delta \cdot I\left\{\frac{5}{8} < t < \frac{7}{8}\right\}$$

c) "Sharp" function (figure 5.1c):

$$g_3(t) = 10\left[\exp\left(tI\left\{t < 0.5\right\} + (1-t)I\{t > 0.5\}\right) - 1\right]$$

d) "JumpSharp" function, i.e. "Sharp" function with a single jump (figure 5.1d):

$$g_4(t) = 10\left[\exp\left(tI\left\{t < 0.5\right\} + (1-t)I\{t > 0.5\}\right) - 1\right]$$
$$+\Delta I\{t > 3/8\}, \quad (\Delta > 0)$$

e) Sharp plus a local constant shift (figure 5.8b):

$$g_5(t) = 10\left[\exp\left(tI\left\{t < 0.5\right\} + (1-t)I\{t > 0.5\}\right) - 1\right]$$
$$+\Delta I\left\{\frac{20}{64} \leq t \leq \frac{21}{64}\right\}, \quad (\Delta > 0)$$

Moreover, modifications of function b) by more than one jump are also included (see below). The parameters q, J and δ_j were chosen as outlined in Beran and Shumeyko (2011a).

Simulated sample paths of Y_i with trend functions a) and b) are shown in figures 5.2a) and b). For the series in figure 5.2b), the complete estimate $\hat{g}$ and the low resolution estimate $\hat{g}_{low}$ are displayed in figures 5.3c) and d) respectively. In this case, including the high resolution component mainly leads to the detection of the second jump. Figures 5.4a) and b) show the residuals $X_{i,low} = Y_i - \hat{g}_{low}$ obtained after subtracting the low resolution estimate. The corresponding block sums $\zeta_{i,low}$ are shown in figure 5.5, and bootstrapped samples of $T^{*(j)}_{kl,low}$ (standardized by $\tilde{\sigma}$) are displayed in figures 5.6a) and b) respectively. Compared to figure 5.6a), the variability of $T^{*(j)}_{kl,low}$ appears to be higher in figure 5.6b). This is due to the discontinuity of the second trend function, as explained by theorems 5.1(b) and 5.4(b).

Table	$FARIMA$ model	ϵ	k	Δ	no. of jumps	wavelet
sine, JumpSine						
5.2 & 5.3	$d = 0.1, 0.4$	$1.5(2r+\alpha)^{-1}$	10 to 150	0, 10	2	"d6"
5.4	$d = 0.4$	$1.5(2r+\alpha)^{-1}$	50	5 to 12	2	"d6"
5.5	$d = 0.4$	$1.5(2r+\alpha)^{-1}$	50	5	1 to 4	"d6"
5.6	$d = 0.4$	$1.5(2r+\alpha)^{-1}$	50	10	2	"d6" "d8" "d10"
5.7	$d = 0.4$	$1-2(2r+1)^{-1}$, $1-1.5(2r+\alpha)^{-1}$, $1-\frac{1}{2}r^{-1}$	50	10	2	"d6"
Sharp, JumpSharp						
5.8 & 5.9	$d = 0.1, 0.4$	$1.5(2r+\alpha)^{-1}$	10 to 150	0, 2	1	"d6"
5.11 & 5.12	$d = 0.1, 0.4$, $\phi = -0.5$	$1.5(2r+\alpha)^{-1}$	10 to 150	0, 2	1	"d6"
5.13 & 5.14	$d = 0.1, 0.4$, $\phi = 0.5, \psi = 0.5$	$1.5(2r+\alpha)^{-1}$	10 to 150	0, 2	1	"d6"
Sharp with local shift						
5.10	$d = 0.3$	$1.5(2r+\alpha)^{-1}$	150	3	2	"d6"

Table 5.1: Overview of tables with simulation results.

Numerical results are summarized in tables 5.2 to 5.14. An overview of the simulated models and parameter settings is given in table 5.1. Tables 5.2 and 5.3 show simulated rejection probabilities for the sine and JumpSine function and $d = 0.1$ and 0.4 respectively. Results for different values of k are compared. The nominal levels of significance are achieved with a reasonable degree of accuracy for all cases, even for the relatively small sample size of $n = 128$ and strong long memory ($d = 0.4$). Similar results are obtained for significance levels for case c) (tables 5.8 and 5.9). As expected, the power increases with increasing values of k and n. Moreover, discontinuities are more difficult to detect under stronger long memory. This is not surprising, since a large value of d leads to sample paths that tend to have local cycles and trends resembling discontinuities in the expected value.

The influence of Δ is illustrated in table 5.4. In table 5.5, it can be seen that the power increases considerably when the number of jumps is increased. Moreover, the power is also influenced by the choice of the wavelet basis. In the case of the JumpSine function, the power improves when one replaces "d6" by "d8" or "d10" respectively (table 5.6). The effect of choosing different values of l (or ϵ) is considered in table 5.7. As a general rule l should be neither too small nor too large. Choosing l very small (ϵ large) prevents the procedure from capturing the asymptotic effect of long memory. On the other hand, if l is very large (ϵ small), then local jumps are smoothed out and the number of available blocks is small.

Simulated powers for case d) (sharp function with one jump) with $\Delta = 2$ are displayed in tables 5.8 and 5.9. Case e) is considered in table 5.10. Discontinuities of this type are very difficult to detect, since under long memory local shifts are often masked by similar spurious local trends generated by the residual process. This is illustrated in figure 5.7. Looking at the data only, without any visual guideline, it is very difficult to see whether and where a jump in the mean function may have occurred. Figure 5.8c) illustrates that the high resolution fit detects the local jump, though the estimate is more wiggly than the true trend. The reason is that for $n = 512$ and relatively strong long memory with $d = 0.3$, the extremely local shift is masked by spurious trends of the residual process. Nevertheless, it is remarkable that the location is identified correctly, in spite of several similar looking (but spurious) "discontinuities" at other time points. The simulation results in table 5.10 show a reasonably good power of the bootstrap test already for $n = 512$. However, the power appears to converge to 1 rather slowly.

Finally, tables 5.11 to 5.14 consider the same situation as in tables 5.8 and 5.9, except that short-memory parameters are included. In tables 5.11 and 5.12, we have a negative autoregressive parameter $\phi_1 = -0.5$ whereas strong positive short-range dependence is added in tables 5.13 and 5.14 with $\phi_1 = 0.5$ and $\psi = 0.5$ (table 5.14). The results indicate that a negative short-memory component improves power. This coincides with the fact that negative autocorrelations improve the accuracy of trend estimates. For the $FARIMA(1, d, 1)$ model with positive short-memory components the results are less clear. Here, the choice of k apparently plays a role.

$\Delta = 0$						
	$n = 128$		$n = 256$		$n = 512$	
$k\backslash\alpha$	0.1	0.05	0.1	0.05	0.1	0.05
10	0.1075	0.0425	0.1050	0.0475	0.1050	0.0575
20	0.1075	0.0350	0.0925	0.0550	0.0950	0.0550
50	0.1000	0.0475	0.0925	0.0575	0.1050	0.0525
100	0.1050	0.0450	0.0950	0.0525	0.1050	0.0500
150	0.0950	0.0500	0.0975	0.0525	0.1025	0.0525
$\Delta = 10$						
	$n = 128$		$n = 256$		$n = 512$	
$k\backslash\alpha$	0.1	0.05	0.1	0.05	0.9	0.05
10	0.8200	0.7575	0.9625	0.9250	0.9875	0.9650
20	0.9625	0.9050	0.9900	0.9850	1	1
50	1	0.9950	1	1	1	1
100	1	1	1	1	1	1
150	1	1	1	1	1	1

Table 5.2: Empirical rejection frequencies based on 400 simulations of a $FARIMA(0, 0.1, 0)$ process with g equal to the sine function a) (upper part of the table) and the JumpSine function b) with $\Delta = 10$ (lower part of the table). The wavelet basis "d6" was used. The levels of significance are $\alpha = 0.1$ and 0.05. For each simulation, the rejections regions were determined using a bootstrap sample of size 400.

$\Delta = 0$						
$k\backslash\alpha$	0.1	0.05	0.1	0.05	0.1	0.05
10	0.1275	0.0625	0.0950	0.0575	0.1075	0.0625
20	0.1000	0.0500	0.0975	0.0550	0.0950	0.0500
50	0.0950	0.0525	0.1025	0.0525	0.1050	0.0425
100	0.1000	0.0625	0.1050	0.0450	0.0900	0.0550
150	0.0975	0.0525	0.1025	0.0525	0.1000	0.0500
$\Delta = 10$						
$k\backslash\alpha$	0.1	0.05	0.1	0.05	0.1	0.05
10	0.3650	0.2975	0.4650	0.3775	0.6075	0.4950
20	0.5050	0.4075	0.6550	0.5600	0.6775	0.6250
50	0.7425	0.6925	0.7575	0.7125	0.7975	0.7575
100	0.8225	0.7750	0.8425	0.8050	0.8725	0.8225
150	0.8850	0.8375	0.9025	0.8750	0.9100	0.8675

Table 5.3: Empirical rejection frequencies based on 400 simulations of a $FARIMA(0, 0.4, 0)$ process with g equal to the sine function a) (upper part of the table) and the JumpSine function b) with $\Delta = 10$ (lower part of the table). The wavelet basis "d6" was used. The levels of significance are $\alpha = 0.1$ and 0.05. For each simulation, the rejections regions were determined using a bootstrap sample of size 400.

Δ	$\alpha = 0.1$	$\alpha = 0.05$
5	0.1100	0.0625
6	0.1525	0.0725
7	0.1725	0.0900
8	0.3075	0.2725
9	0.5825	0.4950
10	0.7425	0.6925
11	0.9550	0.9325
12	1	1

Table 5.4: Empirical rejection frequencies based on 400 simulations of a $FARIMA(0, 0.4, 0)$ process of length $n = 128$ and g equal to the JumpSine function b). The wavelet basis "d6" was used and k was set equal to 50. The levels of significance are $\alpha = 0.1$ and 0.05. For each simulation, the rejections regions were determined using a bootstrap sample of size 400. Jumps of sizes $\Delta \in \{5, 6, ..., 12\}$ are considered.

# Jumps	$\alpha = 0.1$	$\alpha = 0.05$
1	0.1525	0.1050
2	0.7250	0.6750
3	0.9800	0.9650
4	1	1

Table 5.5: Empirical rejection frequencies based on 400 simulations of a $FARIMA(0, 0.4, 0)$ process of length $n = 128$, and g equal to a JumpSine function with one or more jumps of size $\Delta = 5$. The wavelet basis "d6" was used and k was set equal to 50. The levels of significance are $\alpha = 0.1$ and 0.05. For each simulation, the rejections regions were determined using a bootstrap sample of size 400. The following locations of the jumps were used: 1) one jump at 7/8; 2) two jumps at 5/8 and 7/8; 3) three jumps at 3/8, 5/8 and 7/8; and 4) four jumps at 1/8, 3/8, 5/8 and 7/8.

Basis	$\alpha = 0.1$	$\alpha = 0.05$
"d6"	0.7425	0.6925
"d8"	0.9100	0.8950
"d10"	0.9200	0.9025

Table 5.6: Comparison of the wavelet basis functions "d6", "d8" and "d10". Given are empirical rejection frequencies based on 400 simulations of a $FARIMA(0, 0.4, 0)$ process of length $n = 128$, and g equal to a JumpSine function with $\Delta = 10$. The value of k was set equal to 50. The levels of significance are $\alpha = 0.1$ and 0.05. For each simulation, the rejections regions were determined using a bootstrap sample of size 400.

	$\epsilon = 1 - \frac{2}{2r+1}$		$\epsilon = 1 - \frac{1.5}{2r+\alpha}$		$\epsilon = 1 - \frac{1}{2r}$	
n	0.1	0.05	0.1	0.05	0.1	0.05
2^7	0.6450	0.5825	0.7275	0.7075	0.7025	0.6825
2^8	0.7725	0.7000	0.7450	0.7125	0.7525	0.7050
2^9	0.7750	0.7000	0.8050	0.7550	0.7775	0.7100

Table 5.7: Empirical rejection frequencies based on 400 simulations of a $FARIMA(0, 0.4, 0)$ process for different values of ϵ ($l = n^{1-\epsilon}$), and g equal to the JumpSine function b) with $\Delta = 10$. The wavelet basis "d6" was used and k was set equal to 50. The levels of significance are $\alpha = 0.1$ and 0.05. For each simulation, the rejections regions were determined using a bootstrap sample of size 400.

$\Delta = 0$						
	$n = 128$		$n = 256$		$n = 512$	
$k\backslash\alpha$	0.1	0.05	0.1	0.05	0.1	0.05
10	0.0900	0.0400	0.0950	0.0575	0.1075	0.0425
20	0.1050	0.0575	0.0950	0.0500	0.0950	0.0475
50	0.1050	0.0500	0.0975	0.0525	0.1075	0.0425
100	0.1025	0.0575	0.0950	0.0475	0.0975	0.0575
150	0.0975	0.0425	0.1075	0.0450	0.1025	0.0550
$\Delta = 2$						
	$n = 128$		$n = 256$		$n = 512$	
$k\backslash\alpha$	0.1	0.05	0.1	0.05	0.1	0.05
10	0.4750	0.3650	0.6250	0.5475	0.7225	0.6375
20	0.6175	0.5225	0.8425	0.7325	0.8925	0.8450
50	0.8150	0.7150	0.9600	0.9450	0.9825	0.9600
100	0.9225	0.8750	0.9825	0.9825	1	1
150	0.9575	0.9525	1	1	1	1

Table 5.8: Empirical rejection frequencies based on 400 simulations of a $FARIMA(0, 0.1, 0)$ process with g equal to the Sharp function c) (upper part of the table) and the JumpSharp function d) with $\Delta = 2$ (lower part of the table). The wavelet basis "d6" was used. The levels of significance are $\alpha = 0.1$ and 0.05. For each simulation, the rejections regions were determined using a bootstrap sample of size 400.

$\Delta = 0$						
$k\backslash\alpha$	0.1	0.05	0.1	0.05	0.1	0.05
10	0.0950	0.0550	0.1000	0.0475	0.0950	0.0425
20	0.0950	0.0575	0.0950	0.0425	0.0975	0.0425
50	0.0975	0.0550	0.0975	0.0425	0.1050	0.0525
100	0.0950	0.0550	0.1025	0.0425	0.1025	0.0450
150	0.1000	0.0550	0.0950	0.0500	0.0975	0.0550
$\Delta = 2$						
$k\backslash\alpha$	0.1	0.05	0.1	0.05	0.1	0.05
10	0.1350	0.1000	0.2250	0.1550	0.2650	0.1900
20	0.2100	0.1400	0.3225	0.2450	0.3125	0.2275
50	0.2550	0.2050	0.3075	0.2550	0.4450	0.3700
100	0.2700	0.2300	0.3300	0.2850	0.4325	0.3925
150	0.2775	0.2450	0.3475	0.3125	0.5200	0.4850

Table 5.9: Empirical rejection frequencies based on 400 simulations of a $FARIMA(0, 0.4, 0)$ process with g equal to the Sharp function c) (upper part of the table) and the JumpSharp function d) with $\Delta = 2$ (lower part of the table). The wavelet basis "d6" was used. The levels of significance are $\alpha = 0.1$ and 0.05. For each simulation, the rejections regions were determined using a bootstrap sample of size 400.

n	$\alpha = 0.1$	$\alpha = 0.05$
2^9	0.3400	0.2850
2^{10}	0.4050	0.3150
2^{12}	0.4175	0.3225
2^{15}	0.6450	0.4900

Table 5.10: Empirical rejection frequencies based on 400 simulations of a $FARIMA(0, 0.3, 0)$ residual process and g equal to the Sharp function with a local shift (of size $\Delta = 3$ and length 1/64) as defined in e). The value of k was set equal to 150, the wavelet basis was "d6". The levels of significance are $\alpha = 0.1$ and 0.05. For each simulation, the rejections regions were determined using a bootstrap sample of size 400.

$\Delta = 0$						
	$n = 128$		$n = 256$		$n = 512$	
$k\backslash\alpha$	0.1	0.05	0.1	0.05	0.1	0.05
10	0.0975	0.0625	0.0900	0.0725	0.0750	0.0500
20	0.1250	0.0500	0.0700	0.0375	0.0850	0.0550
50	0.1275	0.0800	0.1250	0.0400	0.0975	0.0475
100	0.1100	0.0400	0.1175	0.0450	0.0800	0.0625
150	0.1125	0.0625	0.0925	0.0550	0.0925	0.0500
$\Delta = 2$						
$k\backslash\alpha$	0.1	0.05	0.1	0.05	0.1	0.05
10	0.4150	0.3275	0.7100	0.6425	0.7800	0.6925
20	0.3425	0.2850	0.8650	0.7300	0.9150	0.8800
50	0.6025	0.5000	0.9775	0.9600	1	1
100	0.5975	0.5750	1	1	1	1
150	0.6550	0.6475	1	1	1	1

Table 5.11: Empirical rejection frequencies based on 400 simulations of a $FARIMA(1, 0.1, 0)$ process with $\phi = \{-0.5\}$, and g equal to the Sharp function c) (upper part of the table) and the JumpSharp function d) with $\Delta = 2$ (lower part of the table). The wavelet basis "d6" was used. The levels of significance are $\alpha = 0.1$ and 0.05. For each simulation, the rejections regions were determined using a bootstrap sample of size 400.

$\Delta = 0$						
	$n = 128$		$n = 256$		$n = 512$	
$k\backslash\alpha$	0.1	0.05	0.1	0.05	0.1	0.05
10	0.0975	0.0625	0.0900	0.0725	0.0750	0.0500
20	0.1250	0.0500	0.0700	0.0375	0.0850	0.0550
50	0.1275	0.0800	0.1250	0.0400	0.0975	0.0475
100	0.1100	0.0400	0.1175	0.0450	0.0800	0.0625
150	0.1125	0.0625	0.0925	0.0550	0.0925	0.0500
$\Delta = 2$						
$k\backslash\alpha$	0.1	0.05	0.1	0.05	0.1	0.05
10	0.4650	0.4125	0.5200	0.4575	0.6275	0.5400
20	0.5600	0.5125	0.6000	0.5325	0.6625	0.6325
50	0.7425	0.6600	0.7625	0.7225	0.7700	0.7225
100	0.7725	0.7550	0.8025	0.7700	0.8725	0.8450
150	0.8025	0.7700	0.8725	0.8575	0.8900	0.8675

Table 5.12: Empirical rejection frequencies based on 400 simulations of a $FARIMA(1, 0.4, 0)$ process with $\phi = \{-0.5\}$, and g equal to the Sharp function c) (upper part of the table) and the JumpSharp function d) with $\Delta = 2$ (lower part of the table). The wavelet basis "d6" was used. The levels of significance are $\alpha = 0.1$ and 0.05. For each simulation, the rejections regions were determined using a bootstrap sample of size 400.

$\Delta = 0$						
	$n = 128$		$n = 256$		$n = 512$	
$k\backslash\alpha$	0.1	0.05	0.1	0.05	0.1	0.05
10	0.1175	0.0600	0.1475	0.0600	0.0800	0.0325
20	0.1150	0.0525	0.1125	0.0625	0.1150	0.0600
50	0.0800	0.0400	0.1525	0.0475	0.1175	0.0425
100	0.1225	0.0775	0.1050	0.0575	0.1075	0.0475
150	0.1150	0.0550	0.1150	0.0550	0.1000	0.0525
$\Delta = 2$						
$k\backslash\alpha$	0.1	0.05	0.1	0.05	0.1	0.05
10	0.4300	0.3475	0.6375	0.5475	0.7175	0.6325
20	0.6150	0.5600	0.8475	0.7600	0.9125	0.8850
50	0.8225	0.7650	0.9975	0.9850	1	1
100	0.8975	0.8950	1	1	1	1
150	0.9225	0.9100	1	1	1	1

Table 5.13: Empirical rejection frequencies based on 400 simulations of a $FARIMA(1, 0.1, 1)$ process with $\phi = \{0.5\}$ and $\psi = \{0.5\}$, and g equal to the Sharp function c) (upper part of the table) and the JumpSharp function d) with $\Delta = 2$ (lower part of the the table). The wavelet basis "d6" was used. The levels of significance are $\alpha = 0.1$ and 0.05. For each simulation, the rejections regions were determined using a bootstrap sample of size 400.

$\Delta = 0$						
	$n = 128$		$n = 256$		$n = 512$	
$k\backslash\alpha$	0.1	0.05	0.1	0.05	0.1	0.05
10	0.1250	0.0800	0.0900	0.0750	0.1250	0.0500
20	0.1475	0.0875	0.1325	0.0850	0.1000	0.0275
50	0.1375	0.0650	0.0950	0.0325	0.1425	0.0425
100	0.1100	0.0875	0.1075	0.0600	0.1475	0.0800
150	0.0850	0.0675	0.1000	0.0525	0.1075	0.0625
$\Delta = 2$						
$k\backslash\alpha$	0.1	0.05	0.1	0.05	0.1	0.05
10	0.2475	0.1650	0.2300	0.1400	0.2425	0.1575
20	0.2800	0.2475	0.2325	0.1875	0.2650	0.2125
50	0.3325	0.2775	0.3625	0.2850	0.3800	0.3650
100	0.3650	0.3200	0.4025	0.3775	0.5450	0.5050
150	0.4650	0.4400	0.4150	0.3875	0.5900	0.5200

Table 5.14: Empirical rejection frequencies based on 400 simulations of a $FARIMA(1, 0.4, 1)$ process with $\phi = \{0.5\}$ and $\psi = \{0.5\}$, with g equal to the Sharp function c) (upper part of the table) and the JumpSharp function d) with $\Delta = 2$ (lower part of the table). The wavelet basis "d6" was used. The levels of significance are $\alpha = 0.1$ and 0.05. For each simulation, the rejections regions were determined using a bootstrap sample of size 400.

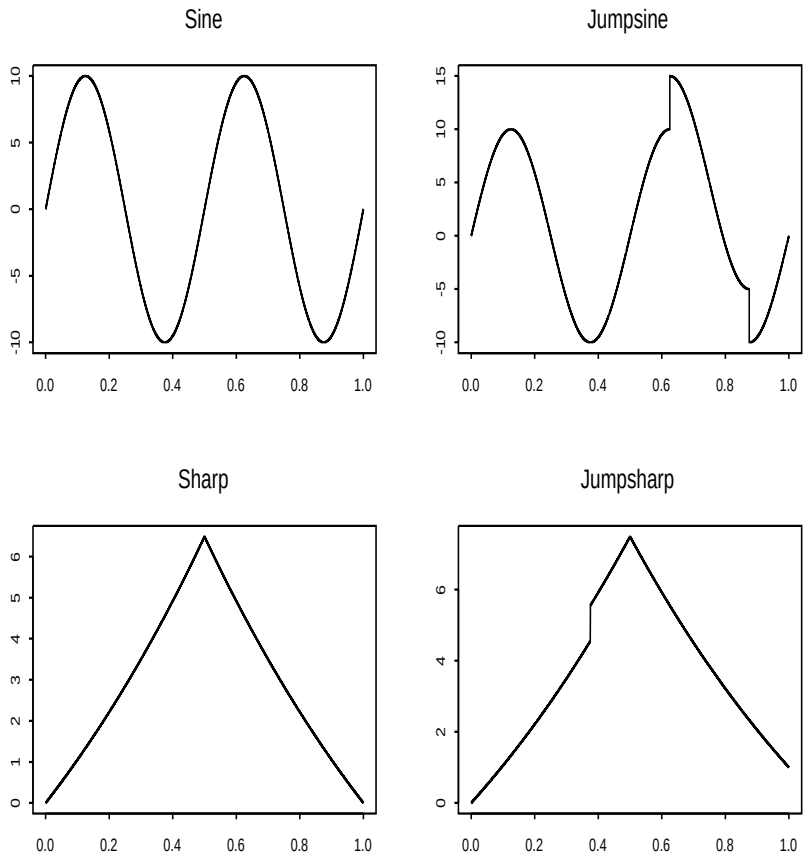


Figure 5.1: Trend functions: a) sine, b) "JumpSine", c) "Sharp", d) "Sharp" with a single jump.

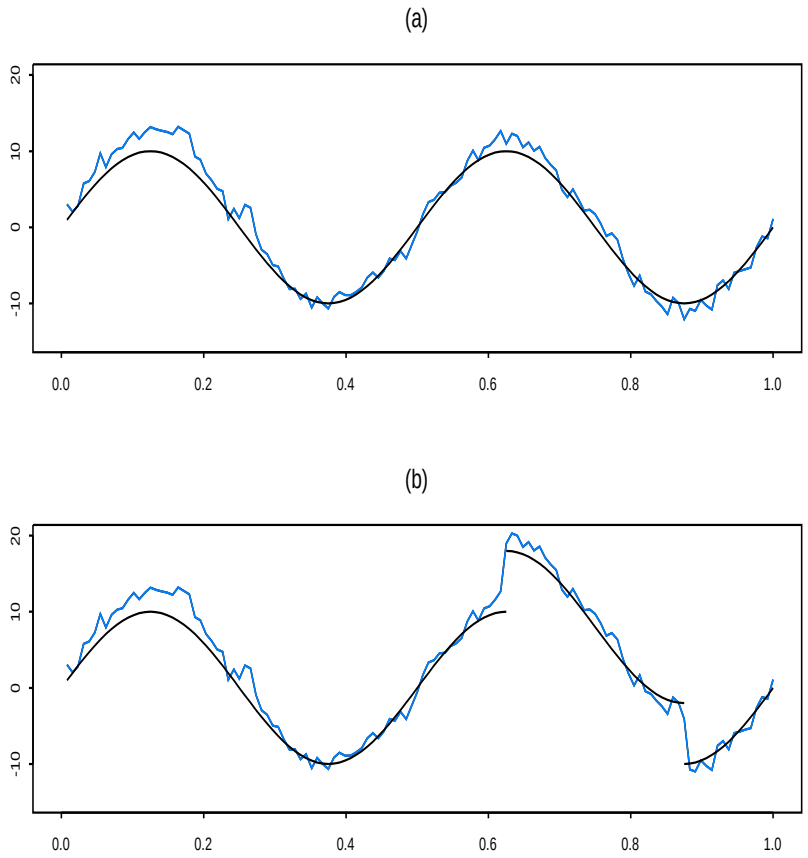


Figure 5.2: Simulated data of length $n = 128$ with the trend functions a) and b) (sine and "JumpSine") respectively and a $FARIMA(0, 0.4, 0)$ residual process. In (b), Δ is equal to 8.

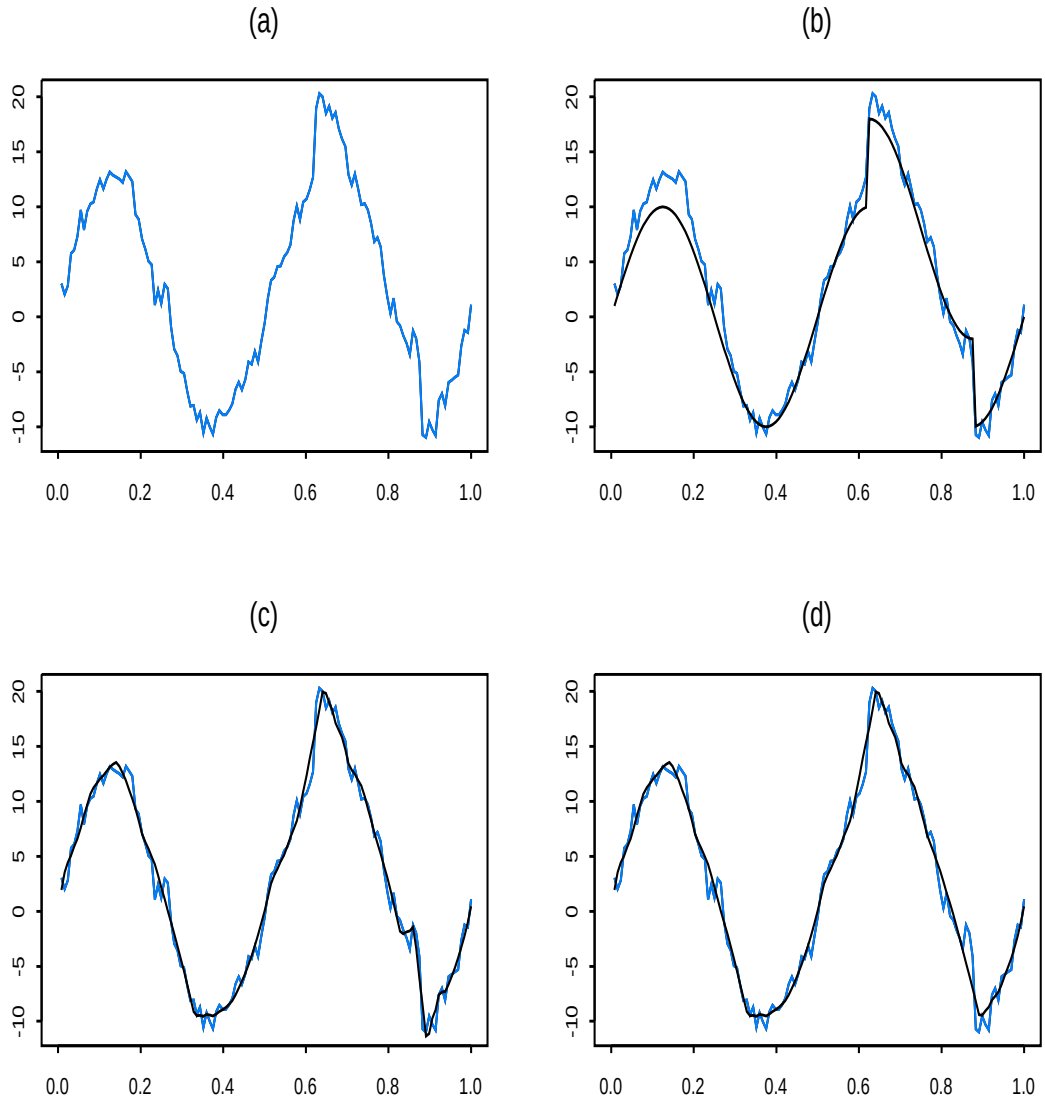


Figure 5.3: Comparison of trend estimation by the low resolution component $\hat{g}_{low}$ and by a full wavelet decomposition $\hat{g} = \hat{g}_{low} + \hat{g}_{high}$ for the simulated series in figures 5.2 a) and b) respectively. The wavelet basis "d6" was used. The figures are: (a) data, (b) data and trend function, (c) data and $\hat{g}$, (d) data and $\hat{g}_{low}$.

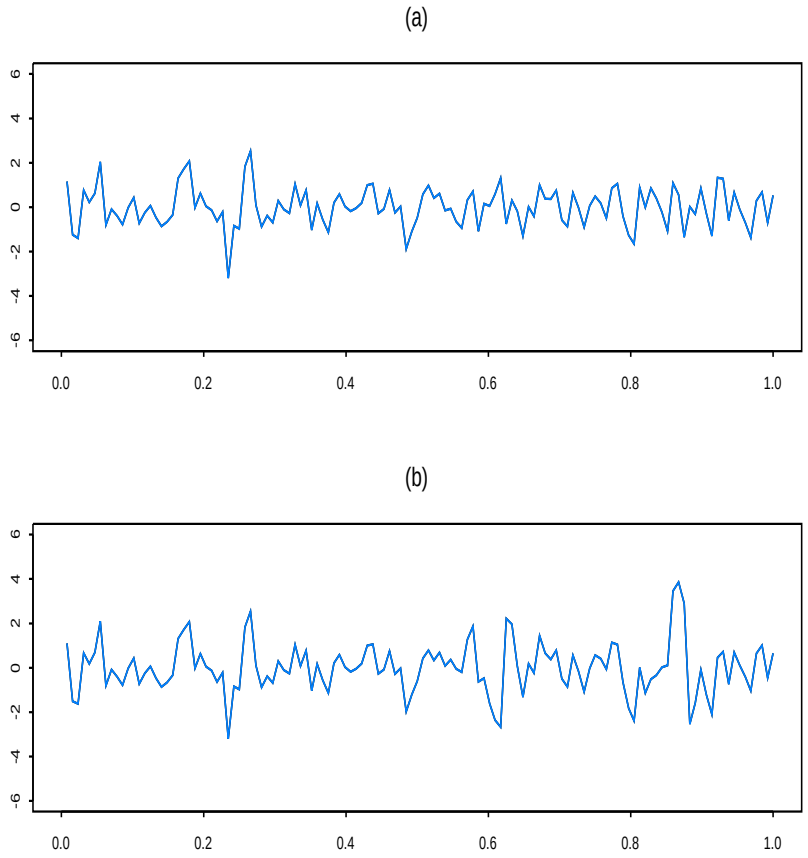


Figure 5.4: Residuals for simulated data in figures 5.2 a) and b) respectively after subtracting the low resolution estimate $\hat{g}_{low}$.

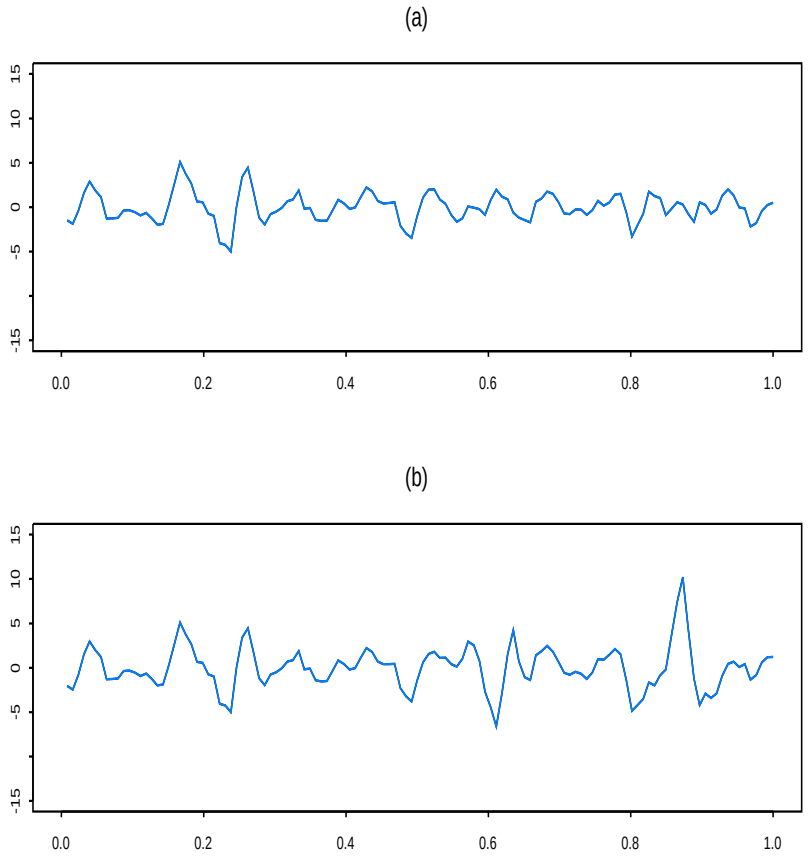


Figure 5.5: Block sums $\zeta_{i,low}$ for the simulated data in figures 5.2 a) and b) respectively.

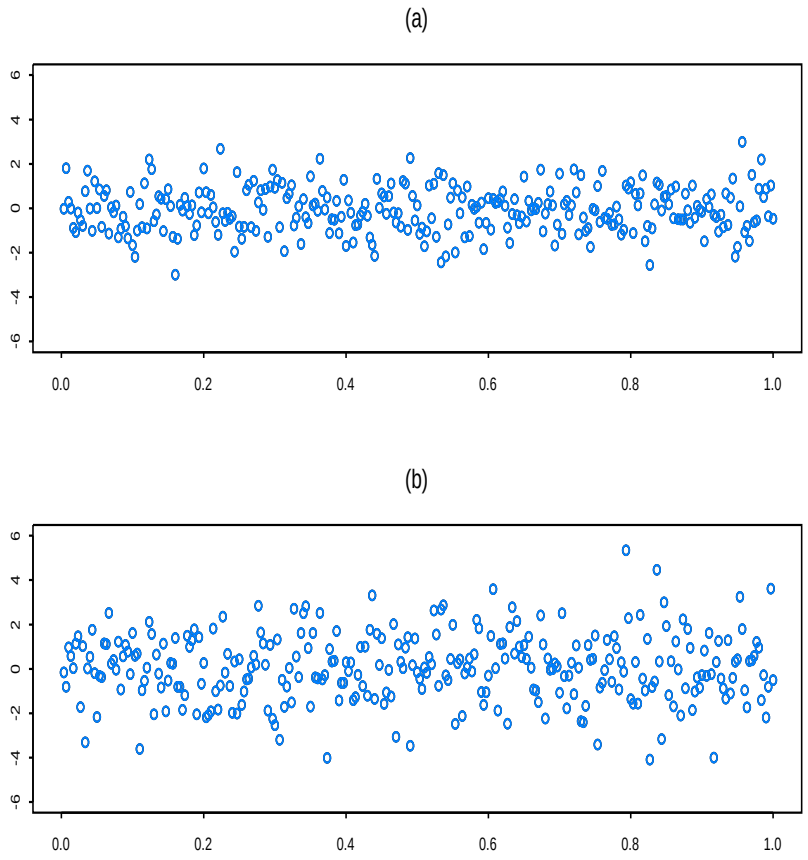


Figure 5.6: Random sample $\left\{T_{kl,low}^{*(1)}, ..., T_{kl,low}^{*(m)}\right\}$ for the simulated series in figure 5.2 a) and b) respectively, standardized by $\tilde{\sigma}$.

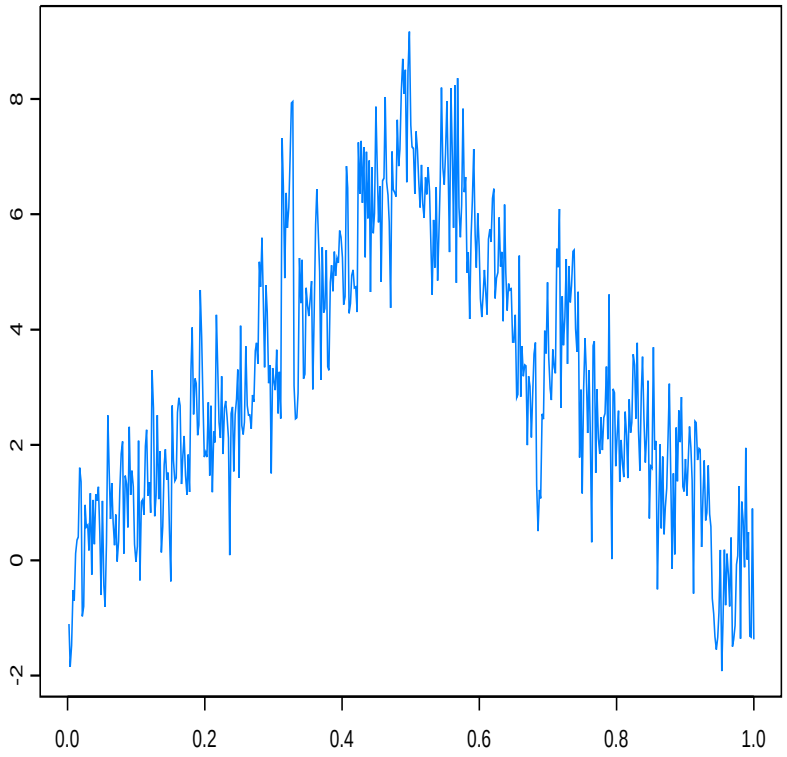


Figure 5.7: Simulated data with trend function e) ("Sharp" plus a local shift) with $\Delta = 3$) and residuals generated by a $FARIMA(0, 0.3, 0)$ process with sample size $n = 512$.

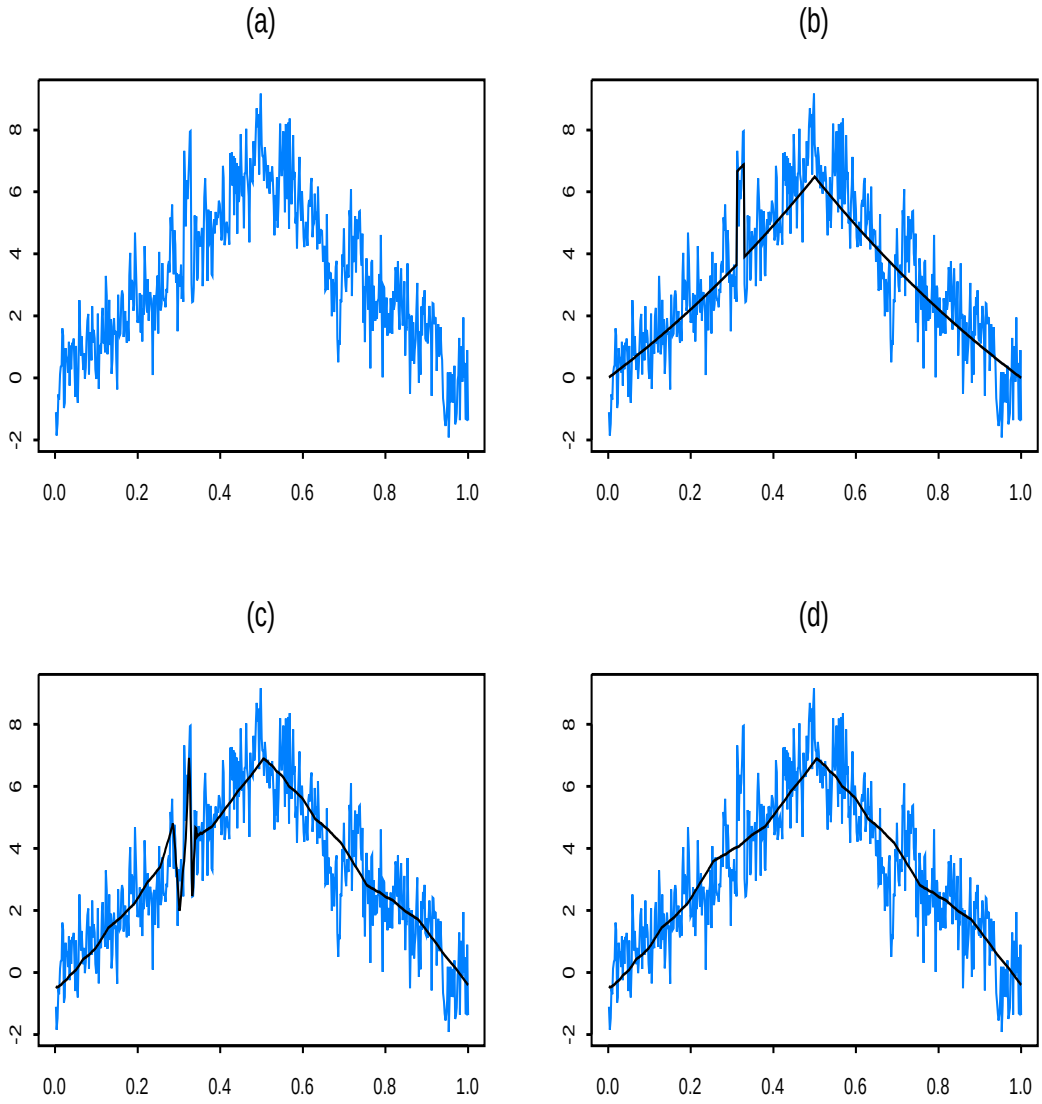


Figure 5.8: Comparison of trend estimation by $\hat{g}_{low}$ and the full wavelet decomposition $\hat{g} = \hat{g}_{low} + \hat{g}_{high}$ for the simulated data in figure 5.7. The "d6" wavelet basis was used. The figures are: (a) data, (b) data with trend function, (c) data with $\hat{g}$ and (d) data with $\hat{g}_{low}$.

5.4 Appendix: Proofs

The following assumptions will be used.

(A1) The derivatives ψ', ϕ' and $g^{(r)}$ exist except for a finite number of points and, where they exist, they are piecewise continuous and bounded.

(A2) $J, j \geq 0$ and $-(N-1) \leq k \leq N2^{J+j} - 1$.

(A3)

$$\nu_k = \int_0^N t^k \psi(t) dt = 0 \ (0 \leq k \leq r-1),$$
$$\nu_r \neq 0.$$

In the case where g is constant and estimated by the sample mean, the results under H_0 essentially follow from Lahiri (1993). In the following these results are extended to wavelet estimation of a nonconstant function g, under H_0 and H_1 respectively. At first, we cite results that follow from chapter 4:

Lemma 5.1. *Under (A1), (A2) and (A3) the following holds:*

(a)
$$s_{Jk} = O(2^{-J/2});$$

(b)
$$E(\hat{s}_{Jk}) - s_{Jk} = O(n^{-1}2^{J/2});$$

(c)
$$E\left\{\left[\hat{s}_{Jk} - E\left(\hat{s}_{Jk}\right)\right]^2\right\} = C_\phi^2 N^{-1+\alpha} n^{-\alpha} 2^{-J(1-\alpha)} + O(n^{-1});$$

(d)
$$\sigma_j^2 = E\left[\left(\hat{d}_{jk} - E\left(\hat{d}_{jk}\right)\right)^2\right] = C_\psi^2 N^{-(1-\alpha)} n^{-\alpha} 2^{-(J+j)(1-\alpha)} + O(n^{-1});$$

(e)
$$E[\hat{d}_{jk}] - d_{jk} = O(n^{-1}2^{(J+j)/2}); \tag{5.10}$$

(f) If $g^{(r)} \in C(I_{jk})$ *where* $I_{jk} = \left[kN^{-1}2^{-(J+j)}, (k+N)N^{-1}2^{-(J+j)}\right]$, *then*

$$d_{jk} = \frac{\nu_r}{r!} g^{(r)}(kN^{-1}2^{-(J+j)}) N^{-\frac{2r+1}{2}} 2^{-\frac{2r+1}{2}(J+j)} + o\left(2^{-\frac{2r+1}{2}(J+j)}\right); \tag{5.11}$$

(g) If there exists an $m \in \{0, 1, \ldots, r-1\}$ *such that* $g \in C^m(I_{jk})$, $g^{(m+1)}$ *exists almost everywhere, but* $g \notin C^{m+1}(I_{jk})$, *then*

$$d_{jk} = O\left(2^{-\frac{2(m+1)+1}{2}(J+j)}\right);$$

(h) If $g \notin C(I_{jk})$, *then*

$$d_{jk} = O\left(2^{-(J+j)/2}\right).$$

Proof. (a) through (f) follow directly from lemmas 4.1-4.7.
For (g), we have

$$\begin{aligned} d_{jk} &= N^{1/2}2^{(J+j)/2}\int_0^1 g(t)\,\psi(N2^{J+j}t-k)\,dt \\ &= N^{-1/2}2^{-(J+j)/2}\int_0^N g(N^{-1}2^{-(J+j)}(y+k))\psi(y)\,dy. \end{aligned} \tag{5.12}$$

The points where g is not differentiable are isolated so that, for n large enough, there is exactly one such point in the support of ψ_{jk}. Thus, assume that there exists exactly one y^* such that $y^*_{jk} = N^{-1}2^{-(J+j)}(y^*+k) \in supp(\psi_{jk})$ and

$$\lim_{\Delta\to 0}\left[g^{(m+1)}\left(y^*_{jk}+N^{-1}2^{-(J+j)}\Delta\right) - g^{(m+1)}\left(y^*_{jk}-N^{-1}2^{-(J+j)}\Delta\right)\right] = C_{m+1,y^*} \neq 0.$$

Since g is m-times continuously differentiable on I_{jk}, and $m+1$-times continuously differentiable on $I_{jk}\setminus\{y_{jk}\}$, a local Taylor expansion of g at y^* (see e.g. Zorich 2004, pp. 225-226) yields

$$\begin{aligned} d_{jk} = &\left(N2^{(J+j)}\right)^{-\frac{1}{2}}\sum_{u=0}^{m}\frac{\left(N2^{(J+j)}\right)^{-u}}{u!}g^{(u)}\left(y^*_{jk}\right)\int_0^N\psi(y)(y-y^*)^u dy \\ &+\frac{\left(N2^{(J+j)}\right)^{-(m+3/2)}}{(m+1)!}\left(A^- + A^+\right) + o\left(2^{-(m+3/2)(J+j)}\right) \end{aligned}$$

with

$$\begin{aligned} A^- &= g^{(m+1)}(y^*_{jk}-0)\int_0^{y^*}\psi(y)(y-y^*)^{m+1}\,dy, \\ A^+ &= g^{(m+1)}(y^*_{jk}+0)\int_{y^*}^{N}\psi(y)(y-y^*)^{m+1}\,dy. \end{aligned}$$

Taking into account condition (3.22) and $m < r$, we have

$$\begin{aligned} A^- + A^+ &= g^{(m+1)}(y^*_{jk}-0)\int_0^N\psi(y)(y-y^*)^{m+1}\,dy \\ &\quad + \left[g^{(m+1)}(y^*_{jk}+0) - g^{(m+1)}(y^*_{jk}-0)\right]\int_{s^*}^N\psi(y)(y-y^*)^{m+1}dy \\ &= g^{(m+1)}(y^*_{jk}-0)\nu_{m+1} + C_{m+1,y^*}\int_{y^*}^N\psi(y)(y-y^*)^{m+1}dy \end{aligned}$$

and hence

$$d_{jk} = 2^{-\frac{2(m+1)+1}{2}(J+j)} C_{jk} + o\left(2^{-(m+3/2)(J+j)}\right) = O\left(2^{-\frac{2(m+1)+1}{2}(J+j)}\right)$$

where C_{jk} is a finite constant.
Finally note that (h) follows directly from (5.12). □

Now we consider the expected value and the autocovariance function of $\hat{g}_{low}$ under the null hypothesis.

Lemma 5.2. *Suppose that $g \in C[0,1]$, and (A1), (A2), (A3) hold. Then*

$$Z_i = \hat{g}_{low}(t_i) - E\left[\hat{g}_{low}(t_i)\right] \sim N(0, \sigma_i^2)$$

and

$$E\left[\hat{g}_{low}(t_i)\right] = g(t_i) + O(n^{-\frac{\alpha}{2r+\alpha}})$$

where

$$\sigma_i^2 \leq C_1 n^{-\frac{2r\alpha}{2r+\alpha}}$$

and, as $|k| \to \infty$,

$$cov\left(Z_i, Z_{i+k}\right) \leq C_2 |k|^{-\alpha}$$

for some positive constants C_1 and C_2.

Proof. Taking into account (3.6), (3.16), (5.1), lemma 5.1b) and the bounded support of ϕ_{Jk} we obtain

$$\begin{aligned} g(t_i) - \hat{g}_{low}(t_i) &= g(t_i) - E[\hat{g}_{low}(t_i)] + E[\hat{g}_{low}(t_i)] - \hat{g}_{low}(t_i) \\ &= \sum_{k=-(N-1)}^{N2^J-1} \left(s_{Jk} - E\left[\hat{s}_{Jk}\right]\right) \phi_{Jk}(t_i) + \sum_{j=0}^{\infty} \sum_{k=-(N-1)}^{N2^{J+j}-1} d_{jk}\psi_{jk}(t_i) \\ &\quad + \sum_{k=-(N-1)}^{N2^J-1} \left(E\left[\hat{s}_{Jk}\right] - \hat{s}_{Jk}\right) \phi_{Jk}(t_i) \\ &= O\left(2^J n^{-1}\right) + \sum_{j=0}^{\infty} \sum_{k=-(N-1)}^{N2^{J+j}-1} d_{jk}\psi_{jk}(t_i) - n^{-1} \sum_{u=1}^{n} \xi_u \sum_{k=-(N-1)}^{N2^J-1} \phi_{Jk}(t_u)\phi_{Jk}(t_i). \end{aligned}$$

The stochastic part is normal and can be written as

$$\begin{aligned} Z_i &= n^{-1} \sum_{u=i-n2^{-J}}^{i+n2^{-J}} \xi_u \sum_{k=N2^J n^{-1} i-N}^{N2^J n^{-1} i} \phi_{Jk}(t_u)\phi_{Jk}(t_i) \\ &= N2^J n^{-1} \sum_{u=i-n2^{-J}}^{i+n2^{-J}} \xi_u \sum_{k=N2^J n^{-1} i-N}^{N2^J n^{-1} i} \phi(N2^J u n^{-1} - k)\phi(N2^J i n^{-1} - k) \\ &= N2^J n^{-1} \sum_{u=i-n2^{-J}}^{i+n2^{-J}} b_{ui}\xi_u \end{aligned}$$

where b_{ui} are some real constants. By analogous arguments as in Beran and Shumeyko (2011a) and Li and Xiao (1997) one obtains

$$Var(Z_i) = N^2 n^{-2} 2^{2J} Var\left(\sum_{u=i-n2^{-J}}^{i+n2^{-J}} b_{ui}\xi_u \right)$$

$$\begin{aligned} &\leq N^2 n^{-2} 2^{2J} \max_{u,i} \left\{ b_{ui}^2 \right\} \left[\sum_{u=i-n2^{-J}}^{i+n2^{-J}} Var(\xi_u) + \sum_{\substack{u,v=i-n2^{-J} \\ u\neq v}}^{i+n2^{-J}} |\gamma(u-v)| \right] \\ &\sim 2N^2 \max_{u,i} \left\{ b_{ui}^2 \right\} \left(n2^{-J} \right)^{-\alpha} C_\gamma \int_0^1 \int_0^1 |x-y|^{-\alpha} \, dx \, dy + O\left(n^{-1} 2^J \right). \end{aligned}$$

Equation (5.3) then implies, for n large enough,

$$Var\left(Z_i\right) \leq C_1 n^{-\frac{2r\alpha}{2r+\alpha}} \tag{5.13}$$

for all i, where C_1 is a suitable constant. Finally lemma 5.1f) implies

$$\left| \sum_{j=0}^{\infty} \sum_{k=0}^{N2^{J+j}-1} d_{jk}\psi_{jk}(t_i) \right| \leq \sum_{j=0}^{\infty} \sum_{k=N2^{J+j}n^{-1}i-N}^{N2^{J+j}n^{-1}i} \max_k |d_{jk}| \max_k |\psi_{jk}(t_i)|$$

$$\leq C_3 \sum_{j=0}^{\infty} 2^{-\frac{3}{2}(J+j)} 2^{\frac{1}{2}(J+j)} = O\left(2^{-J}\right) = O(n^{-\frac{\alpha}{2r+\alpha}})$$

where $C_3 > 0$ is a suitable constant. In summary,

$$\begin{aligned} E\left[g(t_i) - \hat{g}_{low}(t_i)\right] &= O(n^{-\frac{\alpha}{2r+\alpha}}), \\ Var\left[\hat{g}_{low}(t_s)\right] &= O\left(n^{-\frac{2r\alpha}{2r+\alpha}}\right). \end{aligned}$$

Finally,

$$cov\left(Z_i, Z_j\right) \leq \max_{u,i}\left\{b_{u,i}^2\right\} N^2 n^{-2} 2^{2J} \sum_{u=i-n2^{-J}}^{i+n2^{-J}} \sum_{v=j-n2^{-J}}^{j+n2^{-J}} \left|E(\xi_u \xi_v)\right|$$

$$\sim C_4 n^{-2} 2^{2J} \cdot \left(n2^{-J}\right)^{2-\alpha} \int_{-1}^{1}\int_{-1}^{1} \left|\frac{i-j}{n2^{-J}} + x - y\right|^{-\alpha} dx\, dy + O(n^{-(1+\alpha)} 2^{(1+\alpha)J} |i-j|^{-(1+\alpha)})$$

$$\leq C_5 \left|i-j\right|^{-\alpha}$$

for suitable constants C_4, C_5 and $|i-j| \to \infty$. □

The next lemma provides a useful approximation of $\hat{g}_{high}$.

Lemma 5.3. *Suppose that (A1), (A2), (A3) hold. Define for each j the set $D_j = \{d_{jk} \,:\, g \in C\left[supp(\psi_{jk})\right]\}$. Then*

$$\hat{g}_{high}(t_i) = \sum_{j=0}^{q} \sum_{k=-(N-1)}^{N2^{J+j}-1} \hat{d}_{jk} I\left\{d_{jk} \in \bar{D}_j\right\} \psi_{jk}(t_s) + \eta(t_i)$$

where for any fixed $M \in \mathbb{N}$ and $\omega \in (0,1)$,

$$\sup_{0 \leq m \leq M} E\left[\eta^m(t_i)\right] = O\left(n^{-\omega^2 \kappa(n)}\right).$$

Proof. Recall that

$$\hat{g}_{high}(t_i) = \sum_{j=0}^{q} \sum_{k=-N+1}^{N2^{J+j}-1} \hat{d}_{jk} I\left\{|\hat{d}_{jk}| > \delta_j\right\} \psi_{jk}(t_s).$$

The triangle inequality implies for every fixed $\omega \in (0,1)$ (see Hall and Patil (1995))

$$I\left\{|d_{jk}| > (1+\omega)\delta_j\right\} - I\left\{|\hat{d}_{jk} - d_{jk}| > \omega\delta_j\right\} \leq I\left\{|\hat{d}_{jk}| > \delta_j\right\}$$

$$\leq I\left\{|d_{jk}| > (1-\omega)\delta_j\right\} + I\left\{|\hat{d}_{jk} - d_{jk}| > \omega\delta_j\right\}.$$

Now, due to the definition of δ_j and lemma 5.1, we have for all $j \in \{0, ..., q\}$ and n large enough,

$$I\left\{|d_{jk}| > (1-\omega)\delta_j\right\} = I\left\{|d_{jk}| > (1+\omega)\delta_j\right\} = I\left\{d_{jk} \in \bar{D}_{jk}\right\},$$

so that

$$A_n(t_i) - B_n(t_i) \leq \hat{g}_{high}(t_i) \leq A_n(t_i) + B_n(t_i)$$

with

$$A_n(t_i) = \sum_{j=0}^{q} \sum_{k=-N+1}^{N2^{J+j}-1} \hat{d}_{jk}\psi_{jk}(t_i) I\left\{d_{jk} \in \bar{D}_{jk}\right\},$$

$$B_n(t_i) = \sum_{j=0}^{q} \sum_{k=-N+1}^{N2^{J+j}-1} |\hat{d}_{jk}\psi_{jk}(t_i)| I\left\{|\hat{d}_{jk} - d_{jk}| > \omega\delta_j\right\}$$

and

$$\hat{g}_{high}(t_i) = A_n(t_i) + \eta(t_i)$$

with

$$|\eta(t_i)| \leq B_n(t_i). \tag{5.14}$$

Before considering moments of $\eta(t_i)$ note that lemma 5.1 implies

$$\max_k E\left(\hat{d}_{jk}\right) = O(\sigma_j)$$

for $0 \leq j \leq \log_2 n - J$, and also, since $\hat{d}_{jk}$ is normal,

$$E\left(\hat{d}_{jk}^{2m}\right) = O(\sigma_j^{2m}) = O(n^{-m\alpha}2^{-m(J+j)(1-\alpha)}).$$

Let

$$I_{jk} = I\left\{\frac{|\hat{d}_{jk} - E\hat{d}_{jk}|}{\sigma_j} > \frac{\omega\delta_j}{\sigma_j} - \frac{|E\hat{d}_{jk} - d_{jk}|}{\sigma_j}\right\}$$

and $P_{jk} = E(I_{jk})$. The triangle inequality implies $I\left\{|\hat{d}_{jk} - d_{jk}| > \omega\delta_j\right\} \leq I_{jk}$. Using this, the compact support of wavelets, lemma 5.1c) and 5.1e), the condition $\delta_j^2 = 4\kappa(n)\ln n\,\sigma_j^2$ ((5.6) and lemma 5.1c)) and the general inequality $P(Z > \eta) \leq e^{-\frac{1}{2}\eta^2}$ for standard normal variables Z (see e.g. Pollard 1984, p. 191) leads to

$$P_{jk} \leq 2n^{-2\omega^2\kappa(n)} + o\left(n^{-2\omega^2\kappa(n)}\right).$$

Hölder's inequality together with (5.14) imply

$$\begin{aligned}|E[\eta(t_i)]| &\leq \sum_{j=0}^{q} \sum_{k=-N+1}^{N2^{J+j}-1} E\left[|\hat{d}_{jk}|I_{jk}\right] |\psi_{jk}(t_i)| \\ &\leq \sum_{j=0}^{q} \sum_{k=Nn^{-1}2^{J+j}s-N}^{Nn^{-1}2^{J+j}s} \sqrt{E\hat{d}_{jk}^2 \cdot P_{jk}}|\psi_{jk}(t_i)| = O\left(n^{-\omega^2\kappa(n)}\right).\end{aligned}$$

By analogous arguments one obtains

$$|E[\eta^m(t_i)]| \leq \left(\sum_{j=0}^{q} \sum_{k=-N+1}^{N2^{J+j}-1} \left(E\left[\hat{d}_{jk}^{2m}\right] P_{jk}\right)^{1/2m} |\psi_{jk}(t_i)|\right)^m = O\left(n^{-\omega^2\kappa(n)}\right).$$

□

The following two lemmas provide asymptotic approximations of the first two conditional moments of $\zeta^*_{1,low}$ under the null hypothesis.

Lemma 5.4. *Suppose that the assumptions of lemma 5.2 hold. Let $l = O(n^{1-\epsilon})$ for some $1 - 2(2r+\alpha)^{-1} < \epsilon < 1$. Then*

$$a_l^{-1} E_n\left(\zeta^*_{1,low}\right) = o_p(1).$$

Proof.

$$a_l^{-1} E_n\left(\zeta^*_{1,low}\right) = \frac{1}{a_l(n-l+1)} \sum_{i=1}^{n-l+1} \sum_{j=i}^{i+l-1} X_{j,low} = S_1 + S_2$$

with

$$S_1 = \frac{1}{a_l(n-l+1)} \sum_{i=1}^{n-l+1} \sum_{j=i}^{i+l-1} \left(\xi_j - Z_j\right),$$
$$S_2 = \frac{1}{a_l(n-l+1)} \sum_{i=1}^{n-l+1} \sum_{j=i}^{i+l-1} \left\{g(t_j) - E\left[\hat{g}_{low}(t_j)\right]\right\}$$

and $Z_j = \hat{g}_{low}(t_j) - E\left[\hat{g}_{low}(t_j)\right] = o_p(\xi_j)$ (see lemma 5.2). Then $S_1 = o_p(1)$ due to lemma 2.1 (see e.g. lemma 3.1 in Lahiri 1993) and lemma 5.2 together with (5.3) imply

$$S_2 = \frac{1}{a_l(n-l+1)} \sum_{i=1}^{n-l+1} \sum_{j=i}^{i+l-1} O(2^{-J}) = O\left(l a_l^{-1} 2^{-J}\right) = O\left(n^{-\frac{\alpha}{2r+\alpha} + \frac{\alpha(1-\epsilon)}{2}}\right)$$

which is $o(1)$ because $\epsilon > 1 - 2(2r+\alpha)^{-1}$. □

Lemma 5.5. *Suppose that the assumptions of lemma 5.2 hold. Let $\hat{\sigma}_n^2 = a_l^{-2} E_n\left[\zeta^{*2}_{1,low}\right]$ and $l = O(n^{1-\epsilon})$ with $1 - 2(2r+\alpha)^{-1} < \epsilon < 1$. Then*

$$\hat{\sigma}_n^2 = 2\sigma^2(1-\alpha)^{-1}(2-\alpha)^{-1} + o_p(1).$$

Proof.

$$\hat{\sigma}_n^2 = a_l^{-2} E_n\left[\zeta^{*2}_{1,low}\right] = \frac{1}{a_l^2(n-l+1)} \sum_{i=1}^{n-l+1} \left(\sum_{j=i}^{i+l-1} X_{j,low}\right)^2 = S_1 + S_2 + S_3$$

with

$$S_1 = \frac{1}{a_l^2(n-l+1)} \sum_{i=1}^{n-l+1} \left(\sum_{j=i}^{i+l-1} (\xi_j - Z_j) \right)^2,$$

$$S_2 = \frac{1}{a_l^2(n-l+1)} \sum_{i=1}^{n-l+1} \left(\sum_{j=i}^{i+l-1} (g(t_j) - E\left[\hat{g}_{low}(t_j)\right]) \right)^2,$$

$$S_3 = \frac{2}{a_l^2(n-l+1)} \sum_{i=1}^{n-l+1} \left(\sum_{j=i}^{i+l-1} (\xi_j - Z_j) \cdot \sum_{j=i}^{i+l-1} (g(t_j) - E\left[\hat{g}_{low}(t_j)\right]) \right)$$

and Z_j as before. Then

$$S_1 = \frac{2\sigma^2}{(1-\alpha)(2-\alpha)} + o_p(1).$$

Lemma 2.2 (see e.g. lemma 3.2 of Lahiri (1993)) and lemma 5.2 implies

$$S_2 = \frac{1}{a_l^2(n-l+1)} \sum_{i=1}^{n-l+1} \left(\sum_{j=i}^{i+l-1} O(2^{-J}) \right)^2 = O\left(l^2 a_l^{-2} 2^{-2J}\right) = O\left(n^{\alpha(1-\epsilon)-\frac{2\alpha}{2r+\alpha}}\right)$$

which is $o(1)$ because $\epsilon > 1 - 2(2r+\alpha)^{-1}$. Finally, noting that $E(S_3) = 0$, using Hölder´s inequality and writing

$$v(i,l) = Var\left(\sum_{j=i}^{i+l-1} (\xi_j - Z_j) \right),\ w(i,l) = \sum_{j=i}^{i+l-1} \{g(t_j) - E\left[\hat{g}_{low}(t_j)\right]\}$$

we obtain

$$Var(S_3) \leq \frac{4}{a_l^4(n-l+1)^2} \max_i w^2(i,l) \sum_{i_1,i_2=1}^{n-l+1} \sqrt{v(i_1,l)v(i_2,l)} = O\left(n^{\alpha(1-\epsilon)-\frac{2\alpha}{2r+\alpha}}\right) = o(1).$$

□

Now we consider $\zeta_{1,low}^{*2}$ under H_1. Without loss of generality it is sufficient to state the results for the case with one jump.

Lemma 5.6. *Suppose that (A1), (A2), (A3) hold. Futhermore assume that g has a single jump of size Δ at a point $s^* \in (0,1)$. Let $l = O(n^{1-\epsilon})$ with $1-2(2r+\alpha)^{-1} < \epsilon < 1$. Then*

$$a_l^{-1} E_n\left(\zeta_{1,low}^*\right) = o_p(1)$$

and

$$a_l^{-2} E_n\left(\zeta_{1,low}^{*2}\right) = \frac{2\sigma^2}{(1-\alpha)(2-\alpha)} + v_n + o_p(1)$$

with

$$v_n = \Delta^2 C n^{-\frac{\alpha}{2r+\alpha}+\alpha(1-\epsilon)}$$

and $C > 0$. *In particular, if* $\epsilon < 1 - (2r+\alpha)^{-1}$, *then* $v_n \to \infty$.

Proof. Recall that

$$g(t_s) - \hat{g}_{low}(t_s) = O\left(2^J n^{-1}\right) - Z_s + \sum_{j=0}^{\infty} \sum_{k=-N+1}^{N2^{J+j}-1} d_{jk}\psi_{jk}(t_s).$$

The compact support of the wavelets implies that for $k \notin \left[Nn^{-1}2^{J+j}s - N, Nn^{-1}2^{J+j}s\right]$, $\psi_{jk}(t_s)$ is equal to zero. Similarily, $\psi_{jk}(t_i) = 0$ if $i \notin \left[s - n2^{-(J+j)}, s + n2^{-(J+j)}\right]$ and $k \in \left[Nn^{-1}2^{J+j}s - N, Nn^{-1}2^{J+j}s\right]$. Let $i^* \in \mathbb{N}$ be such that $i^*/n \leq s^* \leq (i^*+1)/n$. Then lemma 5.2 implies that $g(t_s) - \hat{g}_{low}(t_s) = -Z_s + O(2^{-J})$ for $s \notin \left[i^* - n2^{-J}, i^* + 1 + n2^{-J}\right]$.
Also, for $s \in \left[i^* - n2^{-J}, i^* + 1 + n2^{-J}\right]$,

$$\sum_{j=0}^{\infty} \sum_{k=-N+1}^{N2^{J+j}-1} d_{jk}\psi_{jk}(t_s) = g(t_s) - \sum_{k=-N+1}^{N2^J-1} s_{Jk}\phi_{Jk}(t_s) = O(1).$$

Define $f(x) = g(x) - \Delta \cdot H(x - s^*)$ where $H(x) = I\{x \geq 0\}$ is the heaviside function, and denote by $\hat{H}_{low}(t - s^*)$ the corresponding low resolution wavelet decomposition

$$\hat{H}_{low}(t - s^*) = \sum_{k=-N+1}^{N2^J-1} \left(\frac{1}{n} \sum_{i=1}^{n} H\left(\frac{i}{n} - s^*\right) \phi_{Jk}\left(\frac{i}{n}\right) \right) \phi_{Jk}(t)$$

$$= \sum_{k=-N+1}^{N2^J-1} \frac{1}{n} \sum_{i=i^*}^{n} \phi_{Jk}\left(\frac{i}{n}\right) \phi_{Jk}(t) = \sum_{k=N2^Jt-N}^{N2^Jt} \int_{s^*}^{1} \phi_{Jk}(x)\phi_{Jk}(t)\, dx + o(n^{-1}).$$

Note that

$$\hat{f}_{low}(t) = \hat{g}_{low}(t) - \Delta \cdot \hat{H}_{low}(t - s^*)$$

and

$$\int_0^1 \hat{H}_{low}(t - s^*)\, dt = \int_0^1 H(t - s^*)\, dt + o(n^{-1}).$$

For $t > s^* + n^{-1} + 2^{-J}$ we have

$$H(t - s^*) = \hat{H}_{low}(t - s^*) + o(n^{-1})$$

and for $t < s^* - 2^{-J}$ the same equality holds exactly. Then

$$a_l^{-1} E_n\left(\zeta^*_{1,low}\right) = \frac{1}{a_l(n-l+1)} \sum_{i=1}^{n-l+1} \sum_{j=i}^{i+l-1} X_{j,low} = S_1 - S_2$$

with

$$S_1 = \frac{1}{a_l(n-l+1)} \sum_{i=1}^{n-l+1} \sum_{j=i}^{i+l-1} \left[\xi_j + f(t_j) - \hat{f}_{low}(t_j) \right],$$

$$S_2 = \frac{\Delta}{a_l(n-l+1)} \sum_{i=1}^{n-l+1} \sum_{j=i}^{i+l-1} \left[H(t_j - s^*) - \hat{H}_{low}(t_j - s^*) \right].$$

Lemma 5.4 implies $S_1 = o_p(1)$. Moreover,

$$\begin{aligned} S_2 &= \frac{\Delta \cdot l}{a_l(n-l+1)} \sum_{j=i^*-n2^{-J}}^{i^*+1+n2^{-J}} \left[H(t_j - s^*) - \hat{H}_{low}(t_j - s^*) \right] + o\left(n^{\alpha(1-\epsilon)/2-1}\right) \\ &= \frac{\Delta \cdot l \cdot n}{a_l(n-l+1)} \int_0^1 \left[H\left(t - s^*\right) - \hat{H}_{low}\left(t - s^*\right) \right] dt + O\left(l a_l^{-1} n^{-1}\right) + o\left(n^{\alpha(1-\epsilon)/2-1}\right) \\ &= O\left(n^{\alpha(1-\epsilon)/2-1}\right). \end{aligned}$$

Taking into account lemma 5.5 we have

$$\begin{aligned} a_l^{-2} E_n(\zeta_{1,low}^{*2}) &= \frac{1}{a_l^2(n-l+1)} \sum_{i=1}^{n-l+1} \left(\sum_{j=i}^{i+l-1} X_{j,low} \right)^2 \\ &= \frac{1}{a_l^2(n-l+1)} \sum_{i=1}^{n-l+1} \left(\sum_{j=i}^{i+l-1} \left(\xi_j + f(t_j) - \hat{f}_{low}(t_j) \right) \right)^2 + S_1 + S_2 \\ &= 2\sigma^2(1-\alpha)^{-1}(2-\alpha)^{-1} + S_1 + S_2 + o_p(1) \end{aligned} \tag{5.15}$$

with

$$S_1 = -\frac{2\Delta}{a_l^2(n-l+1)} \sum_{i=1}^{n-l+1} \left\{ \sum_{j=i}^{i+l-1} \left[\xi_j + f(t_j) - \hat{f}_{low}(t_j) \right] \sum_{j=i}^{i+l-1} \left[H(t_j - s^*) - \hat{H}_{low}(t_j - s^*) \right] \right\},$$

$$S_2 = \frac{\Delta^2}{a_l^2(n-l+1)} \sum_{i=1}^{n-l+1} \left\{ \sum_{j=i}^{i+l-1} \left[H(t_j - s^*) - \hat{H}_l(t_j - s^*) \right] \right\}^2.$$

Now

$$\left(\frac{\Delta^2}{a_l^2(n-l+1)} \right)^{-1} S_2 = \sum_{i=i^*-n2^{-J}-l}^{i^*+1+n2^{-J}} \left\{ \sum_{j=i}^{i+l-1} \left[H(t_j - s^*) - \hat{H}_{low}(t_j - s^*) \right] \right\}^2 + O(l^2 n^{-1})$$

$$\begin{aligned} &= \sum_{i=i^*-n2^{-J}-l}^{i^*+1+n2^{-J}} \left\{ n \int_{\frac{i}{n}}^{\frac{i}{n}+\frac{l}{n}} \left[H\left(x - s^*\right) - \hat{H}_{low}\left(x - s^*\right) \right] dx + o(l) \right\}^2 + O(l^2 n^{-1}) \\ &= n \int_{\frac{i^*}{n}-2^{-J}-\frac{l}{n}}^{\frac{i^*}{n}+2^{-J}} \left\{ n \int_y^{y+\frac{l}{n}} \left[H\left(x - s^*\right) - \hat{H}_{low}\left(x - s^*\right) \right] dx \right\}^2 dy + o(l^2 n 2^{-J}). \end{aligned}$$

Using the representation of $\hat{H}_{low}$ and $\int_y^{y+\frac{l}{n}} \phi_{Jk}(x)\,dx = \frac{l}{n}\phi_{Jk}(y) + o(ln^{-1}2^{J/2})$, the integral can be written as

$$n\int_{s^*-2^{-J}-\frac{l}{n}}^{s^*+2^{-J}} \left\{ n\int_y^{y+\frac{l}{n}} \left[H\left(x-s^*\right) - \sum_{k=-N+1}^{N2^J-1} \int_{s^*}^1 \phi_{Jk}(t)\phi_{Jk}(x)\,dt \right] dx \right\}^2 dy$$
$$= nl^2 \int_{s^*-2^{-J}-\frac{l}{n}}^{s^*+2^{-J}} \left(\tilde{H}\left(y\right) - \sum_{k=-N+1}^{N2^J-1} \int_{s^*}^1 \phi_{Jk}(t)\phi_{Jk}(y)\,dt \right)^2 dy + o(l^2 n 2^{-J})$$

with

$$\tilde{H}(t) = \begin{cases} 0 & ,\ y < s^* - ln^{-1} \\ l^{-1}ny + 1 - l^{-1}s^* & ,\ s^* - ln^{-1} \le y < s^* \\ 1 & ,\ y \ge s^* \end{cases} .$$

Noting that $\tilde{H}\left(y\right) - \sum_{k=-N+1}^{N2^J-1} \int_{s^*}^1 \phi_{Jk}(t)\phi_{Jk}(y)\,dt = 0$ for $y \notin \left[s^* - 2^{-J} - \frac{l}{n}, s^* + 2^{-J}\right]$ and using the wavelet decomposition of $\tilde{H}$ and orthonormality of ϕ_{Jk} and ψ_{jk} we obtain

$$nl^2 \int_0^1 \left(\tilde{H}(y) - H(y-s^*) + \sum_{j=0}^{q} \sum_{k=-N+1}^{N2^{J+j}-1} \int_0^1 H\left(t-s^*\right)\psi_{jk}(t)\psi_{jk}(y)\,dt \right)^2 dy$$
$$+ o(l^2 n 2^{-J}) = I_1 + I_2 + I_3 + o(l^2 n 2^{-J})$$

with

$$I_1 = nl^2 \int_0^1 \left[\tilde{H}(y) - H(y-s^*) \right]^2 dy,$$
$$I_2 = nl^2 \sum_{j=0}^{\infty} \sum_{k=-N+1}^{N2^{J+j}-1} \left[\int_0^1 H\left(t-s^*\right)\psi_{jk}(t)\,dt \right]^2,$$
$$I_3 = nl^2 \int_0^1 \left[\tilde{H}(y) - H(y-s^*) \right] \sum_{j=0}^{q} \sum_{k=-N+1}^{N2^{J+j}-1} \int_0^1 H\left(t-s^*\right)\psi_{jk}(t)\psi_{jk}(y)dtdy$$

where

$$I_1 = \frac{1}{3}l^3,$$
$$I_3 = nl^2 \int_0^1 \left[\tilde{H}(y) - H(y-s^*) \right] \left[H(y-s^*) - \hat{H}_{low}(y-s^*) \right] dy = O(l^3).$$

Since $\int_0^1 H\left(t-s^*\right)\psi_{jk}(t)\,dt = 0$ for $\forall k \notin U_j = \left[\lfloor N2^{J+j}s^* \rfloor - N, \lfloor N2^{J+j}s^* \rfloor + 1\right]$

and $N2^{J+j} - \lfloor N2^{J+j}s^* \rfloor - k \to \infty$, we obtain

$$\begin{aligned} I_2 &= nl^2 \sum_{j=0}^{\infty} \sum_{k \in U_j} \left(\int_{s^*}^{1} \psi_{jk}(t)\, dt \right)^2 \\ &= N^{-1} nl^2 \sum_{j=0}^{\infty} 2^{-(J+j)} \sum_{k=-N}^{1} \left(\int_{N2^{J+j}s^* - \lfloor N2^{J+j}s^* \rfloor - k}^{N} \psi(t)\, dt \right)^2 \\ &\geq 2N^{-1} nl^2 2^{-J} \min_{z \in (0,1)} \left[\sum_{k=-N}^{1} \left(\int_{z-k}^{N} \psi(t)\, dt \right)^2 \right] \end{aligned}$$

and the bounds $I_{lower} \leq I_2 \leq I_{upper}$, where

$$\begin{aligned} I_{lower} &= 2N^{-1} nl^2 2^{-J} \min_{z \in (0,1)} \left[\sum_{k=-N}^{1} \left(\int_{z-k}^{N} \psi(t)\, dt \right)^2 \right] + o(l^2 n 2^{-J}), \\ I_{upper} &= N^{-1} nl^2 \sum_{j=0}^{\infty} 2^{-(J+j)} \sum_{k=-N}^{1} \int_{0}^{N} \psi^2(t)\, dt = 2N^{-1} nl^2 2^{-J} (N+2). \end{aligned}$$

Thus, using the definitions of l and J, there exists a constant C such that

$$S_2 = \Delta^2 C n^{-\frac{\alpha}{2r+\alpha} + \alpha(1-\epsilon)} + o\left(n^{-\frac{\alpha}{2r+\alpha} + \alpha(1-\epsilon)} \right).$$

By similar detailed calculations of $E(S_1)$ and $Var(S_1)$ we now show that S_1 is asymptotically negligible. At first, we consider

$$\begin{aligned} ES_1 &= -\frac{2\Delta}{a_l^2 (n-l+1)} \sum_{i=1}^{n-l+1} \left(\sum_{j=i}^{i+l-1} \left(f(t_j) - E\left[\hat{f}_{low}(t_j) \right] \right) \right. \\ &\qquad \left. \cdot \sum_{j=i}^{i+l-1} \left(H(t_j - s^*) - \hat{H}_{low}(t_j - s^*) \right) \right) \\ &= O\left(\frac{l 2^{-J}}{a_l^2 (n-l+1)} \sum_{i=1}^{n-l+1} \sum_{j=i}^{i+l-1} \left(H(t_j - s^*) - \hat{H}_{low}(t_j - s^*) \right) \right) \\ &= O\left(\frac{l^2 2^{-J}}{a_l^2 (n-l+1)} \right) = O\left(n^{-1+\alpha(1-\epsilon) - \frac{\alpha}{2r+\alpha}} \right) = o(1). \end{aligned}$$

Let $U = [s^* - n2^{-J} - l, s^* + 1 + n2^{-J}]$ and

$$\Delta H_i = \sum_{j=i}^{i+l-1} \left(H(t_j - s^*) - \hat{H}_{low}(t_j - s^*) \right).$$

Hence,

$$\Delta H_i = \begin{cases} 0 & , \ i < s^* - n2^{-J} - l \\ O(l) & , \ i \in [s^* - n2^{-J} - l, s^* + 1 + n2^{-J}] \\ O(ln^{-1}) & , \ i > s^* + 1 + n2^{-J} \end{cases} .$$

Using this notation we derive

$$VarS_1^2 = \frac{4\Delta^2}{a_l^4(n-l+1)^2} E \left(\sum_{i=s^*-n2^{-J}-l}^{s^*+1+n2^{-J}} \sum_{j=i}^{i+l-1} (\xi_j - Z_j)\, \Delta H_i \right.$$

$$\left. + \sum_{i=s^*+2+n2^{-J}}^{n-l+1} \sum_{j=i}^{i+l-1} (\xi_j - Z_j)\, \Delta H_i \right)^2 = S_{11} + S_{12} + S_{13}$$

where

$$S_{11} = \frac{4\Delta^2}{a_l^4(n-l+1)^2} E \left(\sum_{i=s^*-n2^{-J}-l}^{s^*+1+n2^{-J}} \sum_{j=i}^{i+l-1} (\xi_j - Z_j)\, \Delta H_i \right)^2 ,$$

$$S_{12} = \frac{8\Delta^2}{a_l^4(n-l+1)^2} E \left(\sum_{i=s^*-n2^{-J}-l}^{s^*+1+n2^{-J}} \sum_{j=i}^{i+l-1} (\xi_j - Z_j)\, \Delta H_i \sum_{i=s^*+2+n2^{-J}}^{n-l+1} \sum_{j=i}^{i+l-1} (\xi_j - Z_j)\, \Delta H_i \right)$$

and

$$S_{13} \frac{4\Delta^2}{a_l^4(n-l+1)^2} E \left(\sum_{i=s^*+2+n2^{-J}}^{n-l+1} \sum_{j=i}^{i+l-1} (\xi_j - Z_j)\, \Delta H_i \right)^2 .$$

It then yields

$$S_{11} = \frac{4\Delta^2}{a_l^4(n-l+1)^2} E \left(\sum_{i=s^*-n2^{-J}-l}^{s^*+1+n2^{-J}} \sum_{j=i}^{i+l-1} (\xi_j - Z_j)\, \Delta H_i \right)^2$$

$$= \frac{4\Delta^2}{a_l^4(n-l+1)^2} E \left(\sum_{i=s^*-n2^{-J}-l}^{s^*+1+n2^{-J}} \sum_{j=i}^{i+l-1} (\xi_j - Z_j)\, \Delta H_i \right)^2$$

$$= \frac{4\Delta^2}{a_l^4(n-l+1)^2} E \left(\sum_{i=s^*-n2^{-J}-l}^{s^*+n2^{-J}+l} (\xi_i - Z_i) \sum_{j=i}^{i+l-1} \Delta H_j \right)^2$$

$$= \frac{4\Delta^2}{a_l^4(n-l+1)^2} \sum_{i'=s^*-n2^{-J}-l}^{s^*+l+n2^{-J}} \sum_{i=s^*-n2^{-J}-l}^{s^*+l+n2^{-J}} E\left((\xi_i - Z_i)(\xi_{i'} - Z_{i'})\right) \sum_{j=i}^{i+l-1} \Delta H_j \sum_{j'=i'}^{i'+l-1} \Delta H_{j'}$$

$$\sim \frac{4\Delta^2}{a_l^4(n-l+1)^2} \sum_{i'=s^*-n2^{-J}-l}^{s^*+l+n2^{-J}} \sum_{i=s^*-n2^{-J}-l}^{s^*+l+n2^{-J}} |i-i'|^{-\alpha} \sum_{j=i}^{i+l-1} \Delta H_j \sum_{j'=i'}^{i'+l-1} \Delta H_{j'}$$

$$\leq \max\left\{\left(\sum_{j=i}^{i+l-1}\Delta H_j\right)^2\right\}\frac{4\Delta^2}{a_l^4(n-l+1)^2}\sum_{i'=s^*-n2^{-J}-l}^{s^*+l+n2^{-J}}\sum_{i=s^*-n2^{-J}-l}^{s^*+l+n2^{-J}}|i-i'|^{-\alpha}$$

$$= O\left(\frac{l^4(n2^{-J})^{2-\alpha}}{a_l^4 n^2}\right) = O\left(n^{\alpha-\frac{(2-\alpha)\alpha}{2r+\alpha}-2\alpha\epsilon}\right).$$

This is $o(1)$ for all $\epsilon > \frac{r-1+\alpha}{2r+\alpha}$. By similar calculations of S_{12} and S_{13} one can see that

$$S_{13} = O\left(n^{-2+\alpha-2\alpha\epsilon}\right) = o(1),\ \forall\epsilon > 0$$

and due to Hölder's inequality

$$S_{12} = O\left(n^{-1+\alpha-\left(1-\frac{\alpha}{2}\right)\frac{\alpha}{2r+\alpha}-2\alpha\epsilon}\right) = o(1)\ ,\ \forall\epsilon > 0.$$

This concludes the proof. □

The following two lemmas compare the conditional expected values and conditional variances of ζ_1^* and $\zeta_{1,low}^*$ under H_0.

Lemma 5.7. *Suppose that the assumptions of lemma 5.2 hold. Let $l = O(n^{1-\epsilon})$ for some $0 < \epsilon < 1$. Then*

$$a_l^{-1}E_n\left(\zeta_1^*\right) = a_l^{-1}E_n\left(\zeta_{1,low}^*\right) + O_p\left(n^{\frac{\alpha}{2}(1-\epsilon)-\frac{1}{2}\omega^2\kappa(n)}\right).$$

Proof.

$$a_l^{-1}E_n\left(\zeta_1^*\right) = a_l^{-1}E_n\left(\zeta_{1,low}^*\right) - S$$

with

$$S = \frac{1}{a_l(n-l+1)}\sum_{i=1}^{n-l+1}\sum_{j=i}^{i+l-1}\hat{g}_{high}(t_j).$$

From lemma 5.3 we have

$$E\left(S\right) = \frac{1}{a_l(n-l+1)}\sum_{i=1}^{n-l+1}\sum_{s=i}^{i+l-1}\left(\sum_{j=0}^{q}\sum_{k=-N+1}^{N2^{J+j}-1}E\left[\hat{d}_{jk}\right]I(d_{jk}\in\bar{D}_j)\psi_{jk}(t_s) + E\left[\eta(t_s)\right]\right)$$

$$= \frac{1}{a_l(n-l+1)}\sum_{i=1}^{n-l+1}\sum_{s=i}^{i+l-1}E\left[\eta(t_s)\right] = O\left(n^{\frac{\alpha}{2}(1-\epsilon)-\omega^2\kappa(n)}\right)$$

and

$$E\left(S^2\right) = \frac{1}{a_l^2(n-l+1)^2}E\left[\left(\sum_{i=1}^{n-l+1}\sum_{s=i}^{i+l-1}\eta(t_s)\right)^2\right]$$

$$\leq \frac{1}{a_l^2(n-l+1)^2}\sum_{i_1=1}^{n-l+1}\sum_{s_1=i_1}^{i_1+l-1}\sum_{i_2=1}^{n-l+1}\sum_{s_2=i_2}^{i_2+l-1}\sqrt{E\left[\eta^2(t_{s_1})\right]E\left[\eta^2(t_{s_2})\right]} = O\left(n^{\alpha(1-\epsilon)-\omega^2\kappa(n)}\right).$$

□

Lemma 5.8. *Suppose that the assumptions of lemma 5.2 hold. Let $l = O(n^{1-\epsilon})$ for some $0 < \epsilon < 1$. Then*

$$a_l^{-2} E_n\left(\zeta_1^{*^2}\right) = a_l^{-2} E_n\left(\zeta_{1,low}^{*^2}\right) + O_p(a_l^{-2} l \cdot n^{-1/2} n^{-\frac{1}{4}\omega^2\kappa(n)}).$$

Proof.

$$\begin{aligned}\hat{\sigma}_n^2 = a_l^{-2} E_n\left(\zeta_1^{*^2}\right) &= \frac{1}{a_l^2(n-l+1)} \sum_{i=1}^{n-l+1} \left(\sum_{j=i}^{i+l-1} X_j\right)^2 \\ &= a_l^{-2} E_n\left(\zeta_{1,low}^{*^2}\right) - S_1 + S_2\end{aligned}$$

with

$$\begin{aligned}S_1 &= \frac{2}{a_l^2(n-l+1)} \sum_{i=1}^{n-l+1} \left(\sum_{j=i}^{i+l-1} \left(\xi_j + g(t_j) - \hat{g}_{low}(t_j)\right) \sum_{j=i}^{i+l-1} \hat{g}_{high}(t_j)\right), \\ S_2 &= \frac{1}{a_l^2(n-l+1)} \sum_{i=1}^{n-l+1} \left(\sum_{j=i}^{i+l-1} \hat{g}_{high}(t_j)\right)^2.\end{aligned}$$

Now

$$E\left(S_2\right) \leq \frac{1}{a_l^2(n-l+1)} \sum_{i=1}^{n-l+1} \sum_{j=i}^{i+l-1} E\left[\hat{g}_{high}^2(t_j)\right] = O\left(l a_l^{-2} n^{-\omega^2\kappa(n)}\right)$$

and

$$E\left(S_2^2\right) \leq \frac{1}{a_l^4(n-l+1)^2} \sum_{i=1}^{n-l+1} \sum_{j=i}^{i+l-1} E\left[\hat{g}_{high}^4(t_j)\right] = O\left(l a_l^{-4} n^{-1} n^{-\omega^2\kappa(n)}\right).$$

Furthermore,

$$\begin{aligned}\left|E\left(S_1\right)\right| &\leq \frac{2}{a_l^2(n-l+1)} \sum_{i=1}^{n-l+1} \sqrt{E\left[\left(\sum_{j=i}^{i+l-1} \left[\xi_j + g(t_j) - \hat{g}_{low}(t_j)\right]\right)^2\right]} \sum_{j=i}^{i+l-1} \sqrt{E\left[\hat{g}_{high}^2(t_j)\right]} \\ &= O\left(l^{2-\alpha/2} a_l^{-2} n^{-\frac{1}{2}\omega^2\kappa(n)}\right)\end{aligned}$$

and

$$\begin{aligned}E\left(S_1^2\right) &\leq \frac{4}{a_l^4(n-l+1)^2} \sum_{i=1}^{n-l+1} \sum_{j=i}^{i+l-1} \sqrt{E\left[\left(\xi_j + g(t_j) - \hat{g}_{low}(t_j)\right)^4\right]} \sum_{j'=i}^{i+l-1} \sqrt{E\left[\hat{g}_{high}^4(t_{j'})\right]} \\ &= O\left(a_l^{-4} l^2 n^{-1} n^{-\frac{1}{2}\omega^2\kappa(n)}\right).\end{aligned}$$

□

Finally, we derive asymptotic expressions for the first two conditional moments of ζ_1^* under H_1.

Lemma 5.9. *Suppose that the assumptions of lemma 5.6 hold. Let $l = O(n^{1-\epsilon})$ with $1 - 2(2r+\alpha)^{-1} < \epsilon < 1$. Then*

$$a_l^{-1} E_n\left(\zeta_1^*\right) = o_p(1)$$

and

$$a_l^{-2} E_n\left(\zeta_1^{*2}\right) = \frac{2\sigma^2}{(1-\alpha)(2-\alpha)} + o_p(1).$$

Proof. We use the same notation as in the proof of lemma 5.6. Recall that

$$\hat{g}_{high}(t_s) = \sum_{j=0}^{q} \sum_{k=-N+1}^{N2^{J+j}-1} \hat{d}_{jk} I(d_{jk} \in \bar{D}_j) \psi_{jk}(t_s) + \eta(t_s),$$

for $k \notin \left[N2^{J+j} n^{-1} i^* - N, N2^{J+j} n^{-1} i^*\right]$ we have $\psi_{jk}(s^*) = 0$ and $d_{jk} \in D_j$ and also for $k \in \left[N2^{J+j} s^* - N, N2^{J+j} s^*\right]$ and t outside the interval $\left[s^* - 2^{-(J+j)}, s^* + 2^{-(J+j)}\right]$, one has $\psi_{jk}(t) = 0$. Defining
$U = \left\{i^* - n2^{-J} - l, \ldots, i^* + n2^{-J}\right\}$ it then follows that

$$\begin{aligned} a_l^{-1} E_n\left(\zeta_1^*\right) &= \frac{1}{a_l(n-l+1)} \sum_{i=1}^{n-l+1} \sum_{j=i}^{i+l-1} \left(\xi_j + g(t_j) - \hat{g}_{low}(t_j) - \hat{g}_{high}(t_j)\right) \\ &= \frac{1}{a_l(n-l+1)} \sum_{i \notin U} \sum_{j=i}^{i+l-1} + \frac{1}{a_l(n-l+1)} \sum_{i \in U} \sum_{j=i}^{i+l-1} = S_1 + S_2. \end{aligned}$$

For S_1 the result is as before (since $\bar{U}$ does not include jumps). For S_2 consider first

$$g(t_s) - \hat{g}_{low}(t_s) - \hat{g}_{high}(t_s) = A_{1,s} + A_{2,s} + A_{3,s} + A_{4,s}$$

with

$$A_{1,s} = \sum_{k=-N+1}^{N2^J-1} \left(s_{Jk} - \hat{s}_{Jk}\right) \phi_{Jk}(t_s), \; A_{2,s} = \sum_{j=0}^{q} \sum_{k=-N+1}^{N2^{J+j}-1} d_{jk} I\left\{d_{jk} \in D_j\right\} \psi_{jk}(t_s),$$

$$A_{3,s} = \sum_{j=0}^{q} \sum_{k=-N+1}^{N2^{J+j}-1} \left(d_{jk} - \hat{d}_{jk}\right) I\left\{d_{jk} \in \bar{D}_j\right\} \psi_{jk}(t_s), \; A_{4,s} = \sum_{j=q+1}^{\infty} \sum_{k=-N+1}^{N2^{J+j}-1} d_{jk} \psi_{jk}(t_s).$$

If $s \in [i^* - 1, i^* + 1]$, then

$$A_{4,s} = g(t_s) - \sum_{k=-N+1}^{N2^{J+q}-1} s_{(J+q)k} \phi_{(J+q)k}(t_s) = O(1).$$

On the other hand, if $s \notin [i^* - 1, i^* + 1]$, then $A_{4,s} = O(n^{-1})$. Furthermore $A_{2,s} = O(2^{-J})$. Finally, for $a \in \{0, 1, \ldots, \log_2 n - J\}$, and s between $i^* - n2^{-(J+a)}$ and $i^* + n2^{-(J+a)}$ but not in the interval $\left[i^* - n2^{-(J+a+1)}, i^* + n2^{-(J+a+1)}\right]$, we have

$$E(A_{3,s}) = \sum_{j=0}^{q} \sum_{k=-N+1}^{N2^{J+j}-1} \left[d_{jk} - E\left(\hat{d}_{jk}\right)\right] I\left\{d_{jk} \in \bar{D}_j\right\} \psi_{jk}(t_s)$$
$$= \sum_{j=0}^{a} + \sum_{j=a+1}^{q} = O(n^{-1}2^{J+a})$$

and

$$Var\ (A_{3,s}) = E\left\{\left(\sum_{j=0}^{a} \sum_{k=-N+1}^{N2^{J+j}-1} \left[E\left(\hat{d}_{jk}\right) - \hat{d}_{jk}\right] I\left\{d_{jk} \in \bar{D}_j\right\} \psi_{jk}(t_s)\right)^2\right\} + O(n^{-2}2^{2(J+a)})$$

$$\leq \left(\sum_{j=0}^{a} \sum_{k=-N+1}^{N2^{J+j}-1} \sigma_j |\psi_{jk}(t_s)| I(d_{jk} \in \bar{D}_j)\right)^2 + O(n^{-2}2^{2(J+a)}) = O\left(n^{-\alpha}2^{\alpha(J+a)}\right).$$

Note that $A_{3,s}$ is normally distributed so that also $E\left(A_{3,s}^4\right) = O\left(n^{-2\alpha}2^{2\alpha(J+a)}\right)$. For S_2 we then obtain

$$|E(S_2)| = \left|E\left[\frac{1}{a_l(n-l+1)} \sum_{i \in U} \sum_{j=i}^{i+l-1} (\xi_j + Z_j + A_{3,j} + O(2^{-J}))\right]\right| + O(la_l^{-1}n^{-1})$$

$$\leq \frac{l}{a_l(n-l+1)} \sum_{j=s^*-n2^{-J}}^{s^*+n2^{-J}} |E(A_{3,j})| + o(1)$$

$$\leq \frac{l}{a_l(n-l+1)} \sum_{a=0}^{q} \left(n2^{-(J+a)} - n2^{-(J+a+1)}\right) O(n^{-1}2^{J+a}) = O(qla_l^{-1}n^{-1}) = o(1).$$

Similarily one obtains

$$E\left(S_2^2\right) = S_{21} + S_{22} + o(1)$$

with

$$S_{21} = \frac{2}{a_l^2(n-l+1)^2} E\left[\sum_{i \in U} \sum_{j=i}^{i+l-1} (\xi_j + Z_j) \sum_{i \in U} \sum_{j=i}^{i+l-1} A_{3,j}\right]$$
$$= O\left(l^2 a_l^{-2} n^{-\alpha} 2^{-(2-\alpha)J}\right) = o(1),$$
$$S_{22} = \frac{1}{a_l^2(n-l+1)^2} E\left[\left(\sum_{i \in U} \sum_{j=i}^{i+l-1} A_{3,j}\right)^2\right]$$
$$= o\left(la_l^{-1}n^{-\alpha/2}2^{-(1-\alpha/2)J}\right) = o(1).$$

Thus, the result for $a_l^{-1}E_n\left(\zeta_1^*\right)$ follows.
For the second moment we have

$$\sigma_n^2 = a_l^{-2}E_n(\zeta_1^{*2}) = \frac{2\sigma^2}{(1-\alpha)(2-\alpha)} + S_3 + o_p(1)$$

with

$$\begin{aligned} S_3 &= \frac{1}{a_l^2(n-l+1)}\sum_{i\in U}\left\{\sum_{j=i}^{i+l-1}\left[\xi_j + g(t_j) - \hat{g}_{low}(t_j) - \hat{g}_{high}(t_j)\right]\right\}^2 \\ &= \frac{1}{a_l^2(n-l+1)}\left(S_{31}+S_{32}+S_{33}+S_{34}+S_{35}+S_{36}\right) + o(1) \end{aligned}$$

and

$$S_{31} = \sum_{i\in U}\left(\sum_{j=i}^{i+l-1}\left(\xi_j + Z_j\right)\right)^2,\ S_{32} = \sum_{i\in U}\left(\sum_{j=i}^{i+l-1}A_{3,j}\right)^2, S_{33} = \sum_{i\in U}\left(\sum_{j=i}^{i+l-1}O(2^{-J})\right)^2,$$

$$S_{34} = 2\sum_{i\in U}\left[\left(\sum_{j=i}^{i+l-1}\left(\xi_j + Z_j\right)\right)\left(\sum_{j=i}^{i+l-1}A_{3,j}\right)\right],$$

$$S_{35} = 2\sum_{i\in U}\left[\left(\sum_{j=i}^{i+l-1}\left(\xi_j + Z_j\right)\right)\left(\sum_{j=i}^{i+l-1}O(2^{-J})\right)\right],$$

$$S_{36} = \sum_{i\in U}\left[\left(\sum_{j=i}^{i+l-1}A_{3,j}\right)\left(\sum_{j=i}^{i+l-1}O(2^{-J})\right)\right].$$

As you will soon discover, a detailed analysis of the six terms shows that they are asymptotically negligible so that $S_3 = o_p(1)$. Thus,

$$S_{33} = \frac{1}{a_l^2(n-l+1)}\sum_{i\in U}\left(\sum_{j=i}^{i+l-1}O(2^{-J})\right)^2 = O(l^\alpha 2^{-3J}) = o(1).$$

Due to lemma 5.4 we achieve

$$S_{35} = \frac{2}{a_l^2(n-l+1)}\sum_{i\in U}\left(\sum_{j=i}^{i+l-1}\left(\xi_j + Z_j\right)\right)\left(\sum_{j=i}^{i+l-1}O(2^{-J})\right) = o_p(1).$$

Due to lemma 2.2 (see e.g. lemma 3.2 of Lahiri 1993) we derive

$$S_{31} = 2^{-J}\cdot\frac{1}{a_l^2(n-l+1)2^{-J}}\sum_{i=s^*-n2^{-J}-l}^{s^*-n2^{-J}+1}\left(\sum_{j=i}^{i+l-1}\left(\xi_j + Z_j\right)\right)^2 = o_p(1).$$

We now turn to the next term and derive

$$ES_{36} = \frac{O(l2^{-J})}{a_l^2(n-l+1)}\sum_{i\in U}\sum_{j=i}^{i+l-1}EA_{3,j} = \frac{O(ql^2 2^{-J})}{a_l^2(n-l+1)} = o(1).$$

Whereas,

$$ES_{36}^2 = \frac{O(l^2 2^{-2J})}{a_l^4 (n-l+1)^2} E\left[\sum_{i\in U}\sum_{j=i}^{i+l-1} A_{3,j}\right]^2$$

$$\leq \frac{O(l^2 2^{-2J})}{a_l^4 (n-l+1)^2}\left(\sum_{i\in U}\sum_{j=i}^{i+l-1}\sqrt{E\left[A_{3,j}\right]^2}\right)^2 = \frac{O(l^4 2^{-2J} n^{2-\alpha} 2^{-(2-\alpha)J})}{a_l^4 (n-l+1)^2} = o(1).$$

The Hölder's inequality implies

$$ES_{32} = \frac{1}{a_l^2 (n-l+1)}\sum_{i\in U} E\left(\sum_{j=i}^{i+l-1} A_{3,j}\right)^2 \leq \frac{1}{a_l^2 (n-l+1)}\sum_{i\in U}\left(\sum_{j=i}^{i+l-1}\sqrt{E\left[A_{3,j}\right]^2}\right)^2$$

$$= \frac{1}{a_l^2 (n-l+1)}\sum_{a=0}^{q} n 2^{-(J+a+1)}\left(lO(n^{-\alpha/2}2^{\alpha(J+a)/2})\right)^2 = o(1)$$

and

$$ES_{32}^2 \leq \frac{1}{a_l^4 (n-l+1)^2}\left(\sum_{i\in U}\sqrt{E\left(\sum_{j=i}^{i+l-1} A_{3,j}\right)^2}\right)^2$$

$$\leq \frac{1}{a_l^4 (n-l+1)^2}\left(\sum_{i\in U}\sum_{j=i}^{i+l-1}\sqrt{E\left[A_{3,j}\right]^2}\right)^2 = \frac{O(l^2 n^{2-\alpha} 2^{-(2-\alpha)J})}{a_l^4 (n-l+1)^2} = o(1).$$

The rest of the proof is quite standard. Using the Hölder's inequality, we obtain that

$$ES_{34} = \frac{2}{a_l^2 (n-l+1)}\sum_{i\in U} E\left[\left(\sum_{j=i}^{i+l-1}(\xi_j + Z_j)\right)\left(\sum_{j=i}^{i+l-1} A_{3,j}\right)\right]$$

$$\leq \frac{2}{a_l^2 (n-l+1)}\sum_{i\in U}\sqrt{E\left(\sum_{j=i}^{i+l-1}(\xi_j+Z_j)\right)^2}\sqrt{E\left(\sum_{j=i}^{i+l-1} A_{3,j}\right)^2}$$

$$= \frac{O(l^{1-\alpha/2})}{a_l^2 (n-l+1)}\sum_{i\in U}\sqrt{E\left(\sum_{j=i}^{i+l-1} A_{3,j}\right)^2} \leq \frac{O(l^{1-\alpha/2})}{a_l^2 (n-l+1)}\sum_{i\in U}\sum_{j=i}^{i+l-1}\sqrt{E\left[A_{3,j}\right]^2}$$

$$= O\left(a_l^{-2} n^{-\alpha/2} l 2^{-(1-\alpha/2)J}\right) = o(1)$$

and

$$ES_{34}^2 = \frac{4}{a_l^4 (n-l+1)^2} E\left[\sum_{i\in U}\left(\sum_{j=i}^{i+l-1}(\xi_j+Z_j)\right)\left(\sum_{j=i}^{i+l-1} A_{3,j}\right)\right]^2$$

$$\leq \frac{4}{a_l^4 (n-l+1)^2}\left(\sum_{i\in U}\sqrt{E\left[\left(\sum_{j=i}^{i+l-1}(\xi_j+Z_j)\right)\left(\sum_{j=i}^{i+l-1} A_{3,j}\right)\right]^2}\right)^2$$

$$\leq \frac{4}{a_l^4(n-l+1)^2}\left(\sum_{i\in U}\left(E\left(\sum_{j=i}^{i+l-1}(\xi_j+Z_j)\right)^4 E\left(\sum_{j=i}^{i+l-1}A_{3,j}\right)^4\right)^{1/4}\right)^2$$

$$= \frac{O(l^{2-\alpha})}{a_l^4(n-l+1)^2}\left(\sum_{i\in U}\left(E\left(\sum_{j=i}^{i+l-1}A_{3,j}\right)^4\right)^{1/4}\right)^2$$

$$\leq \frac{O(l^{2-\alpha})}{a_l^4(n-l+1)^2}\left(\sum_{i\in U}\sum_{j=i}^{i+l-1}\left(EA_{3,j}^4\right)^{1/4}\right)^2$$

$$= \frac{O(l^{2-\alpha})}{a_l^4(n-l+1)^2}\left(l\sum_{a=0}^{q}n2^{-(J+a+1)}O(n^{-\alpha/2}2^{\alpha(J+a)/2})\right)^2$$

$$= O\left(l^{4-\alpha}a_l^{-4}n^{-\alpha}2^{-(2-\alpha)J}\right) = o(1).$$

□

The results of lemma 5.6 and 5.9 remain the same in the presence of any finite number of isolated jumps.

Proof. (of theorem 5.1) The results follow directly from lemmas 5.4 to 5.6 and definition (5.8). □

Proof. (of theorem 5.2) The results follow from theorem 5.1 in an analogous manner as theorem 2.3 (see e.g. theorem 2.2 in Lahiri 1993). □

Proof. (of theorem 5.3) The theorem follows directly from lemmas 5.7 to 5.9 with $\omega = 2^{-1/2}$ and the definitions (5.7), (5.8). □

Proof. (of theorem 5.4) This follows from theorem 5.3 in a similar way as theorem 5.2.

□

Chapter 6

Concluding remarks

Finally, we give some concluding remarks and mentioned a few open problems regarding the two statistical problems that have been investigated in this thesis.

(1) In chapter 4, an approach to data adaptive wavelet estimation of trend functions for long-memory time series models is proposed. The estimator can be understood as a combination of two components, a smoothing component consisting of a certain number of lower resolution levels where no thresholding is applied, and a higher resolution component filtered by thresholding. The first component leads to good performance for smooth functions whereas the second component is useful for modeling discontinuities. An open problem worth pursuing in future research is the question how much more may be gained by further optimization with respect to fully flexible thresholds δ_j. One can show that $\delta_j = 0$ is indeed the best threshold for some classes of functions. For instance, it can be shown that, if $g \in L^2[0,1]$ and $C < |g^{(r)}(\cdot)| \leq C2^{r+\alpha/2}$ (almost everywhere) for some finite constant C, then $\delta_j = 0$ is asymptotically optimal. This includes, for example, functions that can be represented (or approximated in an appropriate sense) by piecewise polynomials of order r.

    The asymptotic expressions for the tuning parameters are derived. These can be used to develop an iterative data adaptive algorithm similar to Beran and Feng (2002). However, this algorithm has not been developed yet. The assumptions and techniques in Beran and Feng (2002) are quite different so that the algorithm proposed there cannot be carried over to our situation. It is more the general idea of a plug-in algorithm that can be used. However,

this would need a detailed study and is beyond the scope of the present thesis.

(2) The test proposed in chapter 5 is fully nonparametric in the sense that the tuning constants ϵ and k can be chosen without knowledge of the underlying process. The only unknown parameters that would need to be estimated from the observed data are C_γ and α. These are needed for a correct standardization In principle, these parameters can be estimated by applying parametric or semiparametric methods to the estimated residuals $\hat{X}_i$ (see e.g. Fox and Taqqu 1986, Giraitis and Surgailis 1990, Beran 1993, 1995, Geweke and Porter-Hudak 1983, Robinson 1995a,b, Hurvich et al. 1998, Moulines and Soulier 1999). However, the optimal choice of the tuning parameters J and q for estimating g depends on α and C_γ.

(3) We did not carry out a comparative study, because there is no other procedure in the literature that would be applicable to our situation. Most of the methods discussed in the literature assume short memory (for instance Ait-Sahalia and Jacod 2009). Only a few authors deal with the case of long memory, but either consider situations that differ from ours (for instance linear regression with stochastic regressors) or mainly derive results that demonstrate failure of short- memory procedures. This appears to be the first method for testing continuity of a general nonparametric trend function with residuals exhibiting long memory.

Bibliography

[1] 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.

[2] Abramowitz, M. and Stegun, I. A. (1972). Orthogonal Polynomials. Ch. 22 in Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. *Dover Publications*, New York.

[3] Abry, P. and Veitch, D. (1998). Wavelet analysis of long-range-dependent traffic. *IEEE Transactions on Information Theory* 44, 2–15.

[4] Ait-Sahalia, Y. and Jacod, J. (2009). Estimating the degree of activity of jumps in high frequency data. *Ann. Statist.*, Vol. 37, No. 5A, 2202-2244.

[5] Andrews, D.W.K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica* 61, 821856.

[6] Andrews, D.W.K. (2003). Tests for parameter instability and structural change with unknown change point: A corrigendum. *Econometrica* 71, 395397.

[7] Andrews, D.W.K., Lieberman, O. and Marmer, V. (2006). Higher-order improvements of the parametric bootstrap for long-memory Gaussian processes. *J. Econometrics*, 133, 673-702.

[8] Andrews, D.W.K. and Ploberger, W. (1994) Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica* 62, 13831414.

[9] Arcones, M. A. and Yu, B. (1994). Limit theorems for empirical processes under dependence. In: Chaos, Expansions, Multiple Wiener-Itô-Integrals and their Applications. *CRS Press*, London.

[10] Arteche, J. and Orbe, J. (2005). Bootstrapping the log-periodogram regression. *Economics Letters*, 86, 79-85.

[11] Avram, F. and Taqqu, M. S. (1987). Noncentral limit theorems and Appell polynomials. *Ann. Probab.*, 15, 767-775.

[12] Bai, J. (1994) Least squares estimation of a shift in linear processes. *J. Time Ser. Anal.* 15, 453-72.

[13] Bai, J. (1997) Estimation of a change point in multiple regression models. *Review of Economics and Statistics* 79, 674679.

[14] Banerjee, A. and Urgab, G. (2005) Modelling structural breaks, long memory and stock market volatility: an overview. Journal of Econometrics 129, 134

[15] Bardet, J.M., Lang, G., Moulines, E. and Soulier, P. (2000). Wavelet estimator of long-range dependent processes. *Statist. Inf. Stoch. Proc.*, Vol. 3, No. 1-2, 85-99.

[16] Beran, J. (1993). Fitting long memory models by generalized regression. *Biometrika*, 80, 817-822.

[17] Beran, J. (1994). Statistics for long-memory processes. *Chapman & Hall*, London.

[18] Beran, J. (1995). Maximum likelihood estimation of the differencing parameter for invertible short- and long-memory $ARIMA$ models. *JRSS*, Series B, Vol. 57, No.4, 659-672.

[19] Beran, J., Bhansali, R.J. and Ocker, D. (1998). On unified model selection for stationary and nonstationary short- and long-memory autoregressive processes. *Biometrika*, Vol. 85, No. 4, 921-934.

[20] Beran, J. and Feng, Y. (2002a). $SEMIFAR$ models - a semiparametric framework for modelling trends, long-range dependence and nonstationarity. Computational Statistics & *Data Analysis*, Vol. 40, No. 2, 393-419.

[21] Beran, J. and Feng, Y. (2002b). Data driven bandwidth choice for $SEMIFAR$ models. *Journal of Computational and Graphical Statistics*, Vol. 11, No. 2, 690-713.

[22] Beran, J. and Feng, Y. (2002c). Local polynomial fitting with long memory, short memory and antipersistent errors. *Annals of the Institute of Statisitcal Mathematics*, Vol. 54, No. 2, 291-311.

[23] Beran, J., Feng, Y., Ghosh, S. and Sibbertsen, P. (2002) On robust local polynomial estimation with long-memory errors. *International Journal of Forecasting*, Vol. 18, 227-241.

[24] Beran, J. and Shumeyko, Y. (2011a). On asymptotically optimal wavelet estimation of trend functions under long-range dependence. *Bernoulli*, in press.

[25] Beran, J. and Shumeyko, Y. (2011b). Bootstrap testing for discontinuities under long-range dependence. *Journal of Multivariate Analysis*, under revision.

[26] Berkes,I., Horváth,L., Kokoszka, P. and Shao, Q. (2006) On discriminating between long-range dependence and changes in mean. *The Annals of Statistics*, Vol. 34, No. 3, 11401165

[27] Bose, A. (1988), Edgeworth correction by bootstrap in autoregressions, *The Annals of Statistics*, 16, 1709-1722.

[28] Box, G.E.P. and Jenkins, G.M. (1970). Time series analysis: forecasting and control. *Holden Day*, San Francisco

[29] Brillinger, D. (1994). Uses of Cumulants in Wavelet Analysis. *Proc. SPIE, Advanced Signal Processing*, Vol. 2296, 2-18.

[30] Brillinger, D. (1996). Some Uses of Cumulants in Wavelet Analysis. *Nonparametric Statistics*, Vol. 6, 93-114.

[31] Brockwell, P.J. and Devis, R.A. (1987). Time series: Theory and Methods. (first edition). *Springer*, New York.

[32] Bruce, A. and Gao, H.-Y. (1996). Applied wavelet analysis with S-Plus, *Springer, New-York.*

[33] Choi, K. and Zivot, E. (2007) Long memory and structural changes in the forward discount: An empirical investigation . *J. Int. Money and Finance*, Vol. 26, No. 3, 342-363.

[34] Cohen, A., Daubechies, I. and Vial, P. (1993). Wavelets on the interval and fast wavelet transforms.*Applied and Computational Harmonic Analysis*, 1, 54-81.

[35] Cohen, A. and Ryan, R. (1995). Wavelets and multiscale signal processing, *Chapman & Hall*, London.

[36] Csörgö, M. and Horváth, L. (1998). Limit Theorems in Change-Point Analysis. Wiley, New York.

[37] Csörgö, S. and Mielniczuk, J. (1995). Nonparametric regression under long-range dependent normal errors. *The Annals of Statistics*, 23, 1000–1014.

[38] Dahlhaus R. (1989). Efficient parameter estimation for self-similar processes. *Annals of Statistics*, 17, 1749–1766.

[39] Datta, S. and W.P. McCormick (1992), Bootstrap for a finite state Markov chain based on iid resampling, in: R. Lepage and L. Billard, eds., *Exploring the Limits of Bootstrap (Wiley, New York)*, pp. 77-98.

[40] Daubechies, I. (1988). Orthonormal bases of compactly supported wavelets, *Comm. Pure and Appl. Math.*, 41, 909-996

[41] Daubechies, I. (1992). Ten Lectures on Wavelets, *Society for Industrial and Applied Mathematics*, Philadelphia.

[42] Davidson, J. (2002). A model of fractional cointegration, and tests for cointegration using the bootstrap. *J. Econometrics*, 110, 187-212.

[43] Dehling, H. and Taqqu, M.S. (1989). The empirical process of some long-range dependent sequences with applications to U-statistics. *Ann. Statist.*, 17, No. 4, 1767-1783.

[44] Delgado, M. A. and Hidalgo, J. (2000). Nonparametric inference on structural breaks. *J. Econometrics*, Vol. 96, No. 1, 113-144.

[45] Diebold, F.X. and Inoue, A. (2001). Long memory and regime switching. *J. Econometrics*, Vol. 105, No. 1, 131-159.

[46] Dobrushin, R.L. and Major, P. (1979). Non-central limit theorems for non-linear functionals of Gaussian fields. *Z. Wahrsch. verw. Geb.*, 50, 27-52.

[47] Donoho, D.L. and Johnstone, I.M. (1994). Ideal Spatial Adaption by Wavelet Shrinkage. *Biometrika*, Vol. 81, No.3.

[48] Donoho, D.L. and Johnstone, I.M. (1995). Adapting to Unknown Smoothness Via Wavelet Shrinkage. *Journal of the American Statistical Association*, Vol. 90, No. 432, 1200-1224.

[49] Donoho, D.L. and Johnstone, I.M. (1997). Wavelet Threshold Estimators for Data with Correlated Noise. *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 59, No. 2, pp. 319-351

[50] Donoho, D.L. and Johnstone, I.M. (1998). Minimax estimation via wavelet shrinkage. *The Annals of Statistics*, 26, 879–921.

[51] Donoho, D. L., Johnstone, I. M., Kerkyacharian, G. and Picard, D. (1995). Wavelet shrinkage: asymptopia? (with discussion). *Journal of the Royal Statistical Society.* Series B. Methodological, 57, 301–369.

[52] Doukhan, P., Oppenheim, G. and Taqqu, M.S. (2003). Theory and Applications of Long-Range Dependence. *Birkhäuser*, Boston.

[53] Efron, B. (19791, Bootstrap methods: Another look at the Jackknife, *Ann. Statist.* 7, l-26.

[54] Fox, R. and Taqqu, M.S. (1986). Large sample properties of parameter estimates for strongly dependent stationary time series. *The Annals of Statistics*, 14, 517-532.

[55] Franco, G.C. and Reisen, V.A. (2004). Bootstrap techniques in semiparametric estimation methods for $ARFIMA$ models: a comparison study. *Computational Statistics*, 19, 243-259.

[56] Freedman, D.A. (1984), On bootstrapping two stage least squares estimates in stationary linear models, *The Annals of Statistics*, 12, 827-842.

[57] Gasser, T. and Müller, H.G. (1979) Kernel estimation of regression functions. In: Gasser, T., Rosenblatt, M. (Eds.), Smoothing Tehniques of Curve Estimation. Springer, New York, pp. 23-68.

[58] Gasser, T. and Müller, H.G. (1984). Estimating regression functions and their derivatives by the kernel method. *Scand. J. Statist.*, 11, 171–185.

[59] Geweke, J. and Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *J. Time Series Analysis*, 4, 221-238.

[60] Ghosh, S., Beran, J. and Innes, J. (1997). Nonparametric conditional quantile estimation in the presence of long memory. *Student*, 2, 109-117.

[61] Ghosh, S. and Draghicescu, D. (2002a). Predicting the distribution function for long-memory processes. *International Journal of Forecasting*, 18, 283-290.

[62] Ghosh, S. and Draghicescu, D. (2002b). An algorithm for optimal bandwidth selection for smooth nonparametric quantiles and distribution functions. In, Statistics in Industry and Technology: Statistical Data Analysis based

on the L1-norm and related methods, *Birkhäuser Verlag, Basel, Switzerland*, pp. 161-168.

[63] Gil-Alana, L.A.(2008) Fractional integration and structural breaks at unknown periods of time. *J. Time Ser. Analysis*, Vol. 29, No. 1, 163185.

[64] Giraitis, L. (1983). Convergence of some nonlinear transformations of a Gaussian sequence to self-similar processes. *Litovsk. Mat. Sb.*, 23, No. 1, 58-68.

[65] Giraitis, L. (1985). Central limit theorems for functionals of a linear process. *Litivskii Matematickii Sbornik*, 25, 43-57.

[66] Giraitis, L. and Surgailis, D. (1985). Central limit theorems and other limit theorems for functionals of Gaussian processes. *Z. Wahrsch. verw. Geb.*, 70, 191-212.

[67] Giraitis, L. and Surgailis, D. (1986). Multivariate Appell polynomials and the central limit theorem. Dependence in Probability and Statistics, E. Eberlein and M.S. Taqqu (eds.), *Birkhäuser, Boston*, pp. 21-27.

[68] Giraitis, L. and Surgailis, D. (1990). A central limit theorem for quadratic forms in strongly dependent linear variables and application to asymptotical normality of Whittle's estimate. *Prob. Th. Rel. Fields*, 86, 87–104.

[69] Giraitis, L. and Surgailis, D. (1999). Central limit theorem for the empirical process of a linear sequence with long memory. *J. Statist. Plann. Infer.*, 80, 81-93.

[70] Granger, C. W. J. and Hyung, N. (2004). Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns. *J. Emp. Fin.*, Vol. 11, No. 3, 399-421.

[71] Granger, C. W. J. and Joyeux, R. (1980). An introduction to long-range time series models and fractional differencing. *J. Time Ser.Anal.*, 1, 15-30

[72] Grossmann, A. and Morlet, J. (1985). Decomposition of functions into wavelets of constant shape and related transforms. In L. Streit, editor, *Mathematics and physics, lectures on recent results.* Word Scientific, River Edge, NJ.

[73] Hackl, P. and Westlund, A. (1991). Economic structural change, analysis and forecasting. *Springer*, Berlin.

[74] Härdle, W., Kerkyacharian, G., Picard, D. and Tsybakov, A. (1998). Wavelets, approximation, and statistical applications, *Lecture Notes in Statistics, Springer, New York.*

[75] Hall, P., Härdle, W., Kleinow, T. and Schmidt, P. (2000). Semiparametric bootstrap appraoch to hypothesis tests and confidence intervals for the Hurst coefficient. *Stat. Inf. Stoch. Proc.*, 3, 263-276.

[76] Hall, P. and Hart, J. D. (1990). Nonparametric regression with long-range dependence. *Stochastic Processes and Their Applications*, 36, 339–351.

[77] Hall, P., Jing, B.Y. and Lahiri, S.N. (1998). On the sampling window method for long-range dependent data. *Statistica Sinica*, 8, 1189-1204.

[78] Hall, P. and Patil, P. (1995). Formulae for mean integated squared error of non-linear wavelet-based density estimators. *The Annals of Statistics*, 23, 905–928.

[79] Hall, P. and Patil, P. (1996a). On the choice of smoothing parameter, threshold and truncation in nonparametric regression by nonlinear wavelet methods. *Journal of the Royal Statistical Society*, Series B, 58, 361–377.

[80] Hall, P. and Patil, P. (1996b). Effect of threshold rules on performance of wavelet-based curve estimators. *Statistica Sinica*, 6, 331–345.

[81] Hansen, B. E. (1992) Tests for parameter instability in regressions with $I(1)$ processes. *J. Bus. Econ. Statist.*, 10, 321-36.

[82] Hansen, B.E. (2007) Least squares model averaging. *Econometrica* 75, 11751189.

[83] Hawkins, D. L. (1977) Testing a sequence of observations for a shift in location. *J. Am. Statist. Assoc.*, 72, 180-86.

[84] Hawkins, D. L. (1987) A test for a change point in a parametric model based on a maximal Wald-type statistic. *Sankhya*, Series A, 49, 368-376.

[85] Heil, C. E. and Walnut, D. F. (1989) Continuous and discrete wavelet transforms. *SIAM Review*, 31: 628-666.

[86] Heiler, M. (2006) A Nonparametric Regression Spectrum: Estimation, Asymptotic Properties and Data Analysis. Dissertation, University of Konstanz.

[87] Hidalgo, J. and Robinson, P. M. (1996). Testing for structural change in a long memory environment. *J. Econometrics*, 70, 159-174.

[88] Hinkley, D.V. (1969). Inference about the intersection in two-phase regression. *Biometrika*, 56, 495-504.

[89] Ho, H.-C. and Hsing, T. (1996). On the asymptotic expansion of the empirical process of long-memory moving averages. *Ann. Statist.*, 24, 1725-1739.

[90] Ho, H.-C. and Hsing, T. (1997). Limit theorems for functionals of moving averages. *Ann. Probab.*, 25, no. 4, 1636–1669.

[91] Horváth, L. (2001). Change-Point Detection in Long-Memory Processes. *Journal of Multivariate Analysis.* Vol. 78, No. 2, 218-234.

[92] Horváth, L. and Kokoszka, P. (1997). The effect of long-range dependence on change-point estimators. *J. Statist. Plann. and Inf.*, Vol. 64, No. 1, 57-81,

[93] Hosking, J.R.M. (1981). Fractional differencing. *Biometrika*, 68, 165-176.

[94] Hurvich, C., Deo, R. and Brodsky, J. (1998). The mean squared error of Geweke and Porter-Hudaks estimator of a long-memory time-series. *J. Time Series Analysis*, 19, 19-46.

[95] Johnstone, I. M. (1999). Wavelet threshold estimators for correlated data and inverse problems: Adaptivity results. *Statistica Sinica*, 9, 51–83.

[96] Johnstone, I. M. and Silverman, B.W. (1997). Wavelet threshold estimators for data with correlated noise. *Journal of the Royal Statistical Society*, Series B, 59, 319-351.

[97] Krämer, W. and Sibbertsen, P. (2002). Testing for structural changes in the presence of long memory. *Int. J. Bus. Econ.*, Vol. 1, No. 3, 235-242.

[98] Kulik, R. and Raimondo, M. (2009). L^p-wavelet regression with correlated errors and inverse problems. *Statistica Sinica*, 19, 1479-1489.

[99] Künsch, H.R. (1989), The jackknife and the bootstrap for general stationary observations, *The Annals of Statistics*, 17, 1217- 1241.

[100] Lahiri S.N. (1993) On the moving block bootstrap under long range dependence. *Statistics & Probability Letters*, 18, 405-413.

[101] Lazarová, S. (2005). Testing for structural change in regression with long memory processes. *J. Econometrics*, Vol. 129, No. 1-2, 329-372.

[102] Li, L. and Xiao, Y. (2007). Mean integrated squared error of nonlinear wavelet-based estimators with long memory data. *Annals of the Institute of Statistical Mathematics*, Vol. 59, No. 2, 299-324.

[103] Liu, R.Y. and K. Singh (19921, Moving blocks jackknife and bootstrap capture weak dependence, in: R. Lepage and L. Billard, eds., *Exploring the Limits of Bootstrap (Wiley, New York)*, pp. 2255248.

[104] MacNeill, I. B. (1978). Properties of sequences of partial sums of polynomial regression residuals with applications to tests for change of regression at unknown times. *Ann. Statist.*, 6, 422-433.

[105] Mikosch, T. and Stărică, C. (2004). Nonstationarities in Financial Time Series, the Long-Range Dependence, and the $IGARCH$ Effects. *The Review of Economics and Statistics*, 86(1), 378-390.

[106] Mallat, S.G. (1989).Multiresolution approximation and wavelets, *Trans. Amer. Math. Soc.*, Vol. 315, 69-88.

[107] Morlet, J., Arens, G., Fourgeau,E. and Giard, D. (1982) Wave propagation and sampling theory. *Geophysics*,47, 203-236.

[108] Moulines, E. and Soulier, P. (1999). Broadband log-periodogram regression of time series with long-range dependence. *Ann. Statist.*, Vol. 27, No. 4, 1415-1439.

[109] Nason, G.P. (1996). Wavelet shrinkage using cross-validation. *Journal of the Royal Statistical Society*, Series B, Vol. 58, 463-479.

[110] Percival, D.B., Sardy, S. and Davison, A.C. (2000). Wavestrapping time series: adaptive wavelet-based bootstrapping. In: Nonlinear and Nonstationary Signal Processing, Fitzgerald, W.J., Smith, R.L., Walden, A.T. (eds.), *Cambridge University Press*, Cambridge, p. 442-471.

[111] Percival, D.P. and Walden, A.T. (2000). Wavelet methods for time series analysis, *Cambridge University Press*, Cambridge.

[112] Pinsky, M.A. (2002). Introduction to Fourier analysis and wavelets, *The Brooks / Cole Series in Advanced Mathematics, Brooks / Cole, Pacific Grove.*

[113] Pollard, D. (1984). Convergence of stochastic processes, *Springer-Verlag*, New York Inc.

[114] Poskitt, D.S. (2008). Properties of the sieve bootstrap for fractionally integrated and non-invertible processes. *J. Time Ser. Analysis*, Vol. 29, No. 2, 224-250.

[115] Priestley, M.B. (1981) Spectral analisys of time series. *Academic Press*, London

[116] Quandt, R. E. (1960). Test of the hypothesis that a linear regression system obeys two separate regimes. *J. Amer. Statist. Assoc.*, 55, 324-330.

[117] Ray, B. K. and Tsay, R. S. (1997). Bandwidth selection for kernel regression with long-range dependent errors. *Biometrika*, 84, 791-802.

[118] Ray, B. K. and Tsay, R. S. (2002). Bayesian methods for change-point detection in long-range dependent processes. *J. Time Ser. Analysis*, Vol. 23, No. 6, 687705.

[119] Robinson, P.M. (1995a). Log-periodogram regression of time series with long range dependence. *Ann. Statist.* 23, 1048-1072.

[120] Robinson, P.M. (1995b). Gaussian semiparametric estimation of long range dependence. *Ann. Statist.* 23, 1630-1661.

[121] Robinson, P. M. (1997). Large sample inference for nonparametric regression with dependent errors. *The Annals of Statistics*, 25, 2054–2083.

[122] Rosenblatt, M. (1961). Independence and dependence. *In: Proc. 4th Berkeley Symp., J. Neyman (ed.), Univ. California, Berkeley Univ. Press, Berkeley*, pp. 431-443.

[123] Sibbertsen, P. (2004) Long memory versus structural breaks: An overview. *Statistical Papers*, Vol. 45, No. 4, 465-515.

[124] von Sachs, R. and Macgibbon, B. (2000). Non-parametric curve estimation by wavelet thresholding with locally stationary errors. *Scandinavian Journal of Statistics*, 27, 475–499.

[125] Samorodnitsky, G. (2007). Long-range dependence. *Now Publishers Inc.*, Hanover, USA.

[126] Schützner, M. (2009) Asymptotic Statistical Theory for Long Memory Volatility Models. Dissertation, University of Konstanz.

[127] Sen, P.K. (1980). Asymptotic theory of some tests for a possible change in the regression slope occurring at an unknown time point. *Z. Wahrsch. verw. Geb.*, 52, 203218.

[128] Silva, E.M., Franco, G.C., Reisen, V.A. and Cruz, F.R.B. (2006). Local bootstrap approaches for fractional differential parameter estimation in $ARFIMA$ models. *Computational Statistics and Data Analysis*, 51, 1002-1011.

[129] Steeb, W. H. (1998). Hilbert spaces, wavelets, generalised functions and modern quantum mechanics, *Kluwer Academic Publishers, Dordrecht.*

[130] 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.

[131] Surgailis, D. (1981). Convergence of sums of nonlinear functions of moving averages to self-similar processes. *Soviet. Math.*, 23, 247-250.

[132] Surgailis, D. (1982). Zones of attraction of self-similar multiple integrals. *Lithuanian Math. J.*, 22, 327-340.

[133] Taqqu, M.S. (1975). Weak convergence to fractional brownian motion and to the Rosenblatt process. *Prob. Th. and Rel. Fields*, Vol. 31, No. 4, 287-302.

[134] Taqqu, M.S. (1979). Weak convergence of integrated processes of arbitrary Hermite rank. *Z. Wahrsch. verw. Geb.*, 50, 53-83.

[135] Truong,Y. K. and Patil, P. N. (2001). Asymptotics for wavelet based estimates of piecewise smooth regression for stationary time series. *The Annals of Institute of Statistical Mathematics*, 53, 159–178.

[136] Vidakovic, B. (1999). Statistical Modeling by Wavelets. *Wiley*, New York.

[137] Wang, Y. (1996). Function estimation via wavelet shrinkage for long-memory data. *The Annals of Statistics*, 24, 466-484.

[138] Wright, J. H. (1998). Testing for a structural break at unknown date with long-memory disturbances. *J. Time Ser. Analysis*, Vol. 19, No. 3, 369-376.

[139] Yajima, Y. (1985). On estimation of long-memory time series models. Austral. *J. Statist.*,Vol 27, No.3, 303-320.

[140] Yang, Y. (2001). Nonparametric regression with dependent errors. *Bernoulli*, Vol. 7, No. 4, 633-655.

[141] Zorich, V.A. (2004). Mathematical Analysis 1. *Springer-Verlag*, Berlin.

[142] Zygmund, A. (1953). Trigonometric Series. *Cambridge University Press*, London.

Appendix

Main S-PLUS-programs used for simulations in sections 4.3 and 5.3:

```
##########################################################################
### The function Cphi.Cpsi.Simulate calculates the
### constants C_phi and C_psi.
###
### Input:  parameters of the FARIMA model (AR, MA, d);
###         name for wavelet basis
### Output:  C_phi and C_psi
###
### Example:  Cphi.Cpsi.Simulate(0,0,0.4,"d4")
##########################################################################

Cphi.Cpsi.Simulate <- function (AR, MA, d, basis){
    module(wavelets);
    n.FAR <- 2**10;
    d.psi <- c();
    d.phi <- c();
    for(i in c(1:100))
        {

        # Simulate FARIMA(ar,d,ma)
        x.FAR <- arima.fracdiff.sim(n=n.FAR,
    model =list(d=d, ar=AR, ma=MA));

        # Wavelet estimator
        f.dwt <- dwt(x.FAR,wavelet=basis, n.levels=3);
```

```
        # choose level 3
        d.psi[i] <- var(as.numeric(f.dwt[["d3"]]));
        d.phi[i] <- var(as.numeric(f.dwt[["s3"]]));
        }

    #support
    info.wave <- summary(wavelet(basis));
    N.Gross <- info.wave$support[2]-info.wave$support[1];

    # Return C_phi, C_psi
    return(c(mean(d.phi)*2**(-2*3*d),mean(d.psi)*2**(-2*3*d)))
}


##########################################################################
### The function Wavelet.Est.Simulate calculates the simulated and
### theoretical values of log_2 MISE_sim, for trend estimates
### obtained by kernel smoothing, minimax soft threshold wavelet
### estimation and data adaptive hard threshold wavelet estimation
### obtained from theorem 4.3.  The output, which is
### saved in the file "WaveSimulate.txt", is based on "Simulation.Nr"
### simulations of our model with the sharp trend function and
### FARIMA(AR,d,MA) residuals.
###
### INPUT: Simulation.Nr
###        d
###        AR
###        MA
###        Basis (wavelet basis)
###
### OUTPUT: "WaveSimulate.txt"
###
### EXAMPLE: Wavelet.Est(400,0,0,0.2,"d4")
##########################################################################
```

```
Wavelet.Est.Simulate <- function (Simulation.Nr,AR, MA, d, basis) {
module(wavelets)
    cat(Simulation.Nr, "\n", file="WaveSimulate.txt",sep= "\t", append=T)
    cat("d", "Length" , "MISE " ,"Minimax","Kern", "Theor", "\n",
file="waveOptJun.txt",sep="\t", append=T);

    # Coef.  Cpsi, Cphi
    Cpsi <- Cphi.Cpsi.Simulate(AR, MA, d, basis)[2]
    Cphi <- Cphi.Cpsi.Simulate(AR, MA, d, basis)[1]

    #sim
    for(n.FAR.number in c(7:13))
    {
    info.wave<-summary(wavelet(basis));
    N.Gross <- info.wave$support[2]-info.wave$support[1];
    n.FAR <- 2**n.FAR.number
    Int.SQ.Error <- c(1:Simulation.Nr)
    Int.SQ.Error.MM <- c(1:Simulation.Nr)
    Int.SQ.Error.Kern <- c(1:Simulation.Nr)
    ERW.Trend <- 0*c(1:n.FAR)
    for(j in c(1:Simulation.Nr))
    {

    # Simulate FARIMA(0,d,0)
    epsilon.FAR <- arima.fracdiff.sim(n.FAR, model =list(d=d, ar=AR,
ma=MA));

    #Trend
    Trend <- 10*exp(c(1:(n.FAR))/n.FAR)-10
    Trend[c((n.FAR/2+1):n.FAR)] <- 10*exp(1-c((n.FAR/2+1):n.FAR)/n.FAR)
-10

    # Addition Trend
    x.FAR <- epsilon.FAR+Trend;
```

```
    # define m_psi
    info.wave <- summary(wavelet(basis));
    mpsi <- as.numeric(info.wave$wfilter$stats[2])+1;

    # int f^{(r)}^2
    Int.Diff.2 <- c(100*exp(1/2)^2-100,100*exp(1/2)^2
-100,100*exp(1/2)^2-100,100*exp(1/2)^2-100,100*exp(1/2)^2-100)[mpsi]

    # choose optimal decomposition level
    p.Koeff <-floor(n.FAR.number*(2*mpsi/(2*mpsi+(1-2*d)))
+(log(Cphi*(factorial(mpsi))**2
    *Int.Diff.2**(-1)*Const.nu.r(basis)**(-2)*(2**(1-2*d)-1))/log(2))
/(2*mpsi+(1-2*d)))

    # Wavelet estimator
    f.dwt <- dwt(x.FAR, wavelet=basis, n.levels=p.Koeff);

    #hard Thresholding
    for(k.Shift.moth in c(1:p.Koeff))
    {
    di <- paste("d",k.Shift.moth,sep="")
    temp.f.dwt.di <- f.dwt[[di]]
    f.dwt[[di]] <- temp.f.dwt.di*((temp.f.dwt.di)^2
> 1*n.FAR.number*2**(2*d*(k.Shift.moth)))
    }

    # wave Estimator
    # MISE
    Int.SQ.Error[j] <- mean((idwt(f.dwt)-Trend)**2)

    # Minimax Rule
    # Wavelet estimator
    f.dwt <- dwt(x.FAR, wavelet=basis, n.levels=n.FAR.number);

    #soft Thresholding
```

```
    for(k.Shift.moth in c(1:n.FAR.number))
    {
    di <- paste("d",k.Shift.moth,sep="")
    temp.f.dwt.di <- f.dwt[[di]]
    f.dwt[[di]] <- sign(temp.f.dwt.di)*(abs(temp.f.dwt.di)
-sqrt(2*n.FAR.number))*((temp.f.dwt.di)^2 > 2*n.FAR.number)
    }

    # MISE for Minimax
    Int.SQ.Error.MM[j] <- mean((idwt(f.dwt)-Trend)**2)

    #Kern estimator
    nu <- 2**(2*d)*gamma(1-2*d)*sin(pi*d)/(d*(2*d+1))
    C.f <- 1/2/pi
    C.opt<-(9*(1-2*d)*nu*C.f/(12800*pi^4))^(1/(5-2*d))
    bandwidth.opt <- C.opt*(n.FAR)^((2*d-1)/(5-2*d))

    #bandwidth.opt<-0.02208316
    Kern.smooth <- ksmooth((1:n.FAR)/n.FAR,x.FAR,
bandwidth=2*bandwidth.opt)$y

    # MISE for Kern
    Int.SQ.Error.Kern[j] <- mean((Kern.smooth - Trend)**2)
    }
    plot(f.dwt)
    plot(idwt(f.dwt), type="l",col=2)
    print(p.Koeff)
    lines(Trend)
    lines(Kern.smooth,col=3)

    # Theor MISE
    Theor.MISE <- (1/(1-2**(-2*mpsi))+1/(1-2**(-1+2*d)))
*(Cphi*(1-2**(-1+2*d)))**(2*mpsi/(2*mpsi+(1-2*d)))
*(factorial(mpsi)**(-2)*Const.nu.r(basis)**(2)
*Int.Diff.2(Trend,basis))**((1-2*d)/(2*mpsi+(1-2*d)))
```

```
*(n.FAR)**(-(2*mpsi*(1-2*d))/(2*mpsi+(1-2*d)))

    #results
    cat(d, n.FAR, mean(Int.SQ.Error),mean(Int.SQ.Error.MM),
mean(Int.SQ.Error.Kern),Theor.MISE, "\n",
file="WaveSimulate.txt", sep="", append=T)
    }     }


#########################################################################
### The function Simulations.Jump.Sharp calculates the empirical
### rejection frequencies based on 400 simulations of a
### FARIMA(AR, d, MA) process, and g equal a JumpSharp
### function with δ = 2.  The wavelet basis s6 was used.
### The levels of significance are β = 0.1, 0.05 and 0.025.
### For each simulation, the rejections regions are determined
### using a bootstrap sample of size 400.
###
### INPUT: AR
###        MA
###        d
###        n.FAR.number (= log_2 n)
###        k
###
### OUTPUT: q_0.9, q_0.95, q_0.975
###
### EXAMPLE: Simulations.Jump.Sharp(c(0.5),c(0.5),0.4,13,100)
#########################################################################

Simulations.Jump.Sharp <- function ( AR, MA, d, n.FAR.number, k)
{
    module(wavelets)
    basis <-"d6";

    # Coef.  Cpsi, Cphi
```

```
    Cpsi <- Cphi.Cpsi.Simulate(0, 0, d, basis)[2]
    Cphi <- Cphi.Cpsi.Simulate(0, 0, d, basis)[1]

    # support
    info.wave<-summary(wavelet(basis));
    N.Gross <- info.wave$support[2]-info.wave$support[1];

    #Length
    n.FAR <- 2**n.FAR.number;
    Chislo <- c();
    Chislo.n.t <- c();
    sim.nr <- 400;

    # fault number
    Fault.number09<-0
    Fault.number095<-0
    Fault.number0975<-0

    #Trend
    Trend <- 10*exp(c(1:(n.FAR))/n.FAR)-10
    Trend[c((n.FAR/2+1):n.FAR)] <- 10*exp(1-c((n.FAR/2+1):n.FAR)/n.FAR)
-10
    Trend[c((5/16*n.FAR):(n.FAR))]<-Trend[c((5/16*n.FAR):(n.FAR))]+2;

    # define m_psi
    info.wave <- summary(wavelet(basis));
    mpsi <- as.numeric(info.wave$wfilter$stats[2])+1;

    # int f^{(r)}^2
    Int.Diff.2 <- c(100*exp(1/2)^2-100,100*exp(1/2)^2
-100,100*exp(1/2)^2-100,100*exp(1/2)^2-100,
100*exp(1/2)^2-100)[mpsi]

    # choose optimal decomposition level
    p.Koeff <- floor(n.FAR.number*(2*mpsi/(2*mpsi+(1-2*d)))
```

```
+(log(Cphi*(factorial(mpsi))**2*Int.Diff.2**(-1)
*Const.nu.r(basis)**(-2)*(2**(1-2*d)-1))/log(2))/(2*mpsi+(1-2*d)))

    # Block length
    l <- floor(n.FAR**( 1.5/(2*mpsi+1-2*d)));

    # number of blocks
    Anzahl.Blocks <- n.FAR -l+1
    Tkl<-c(1:(n.FAR -l))

    # number of bootstrap
    Sample.Calcule <- c(1:k)

    # normalization coefficient
    Norm.Koef <- k**(-0.5)*l**(-1+(1-2*d)/2)
*(1/pi*gamma(1-2*d)*sin(pi*d))^(-1/2)
    for(Koll in c(1:sim.nr))
    {

    # Simulate FARIMA(0,d,0) Noise
    epsilon.FAR <- arima.fracdiff.sim(n.FAR, model =list(d=d, ar=AR,
ma=MA));

    # Addition Trend
    x.FAR <- epsilon.FAR+Trend;

    # Wavelet estimator
    f.dwt <- dwt(x.FAR, wavelet=basis, n.levels=p.Koeff);

    # hard thresholded
    for(k.Shift.moth in c(1:p.Koeff))
    {
    di <- paste("d",k.Shift.moth,sep="")
    temp.f.dwt.di <- f.dwt[[di]]
    f.dwt[[di]] <- temp.f.dwt.di*((temp.f.dwt.di)^2
```

```
> 0.25*n.FAR.number*2**(2*d*(k.Shift.moth)))
    }

    #Residuals
    Residuals.data <- x.FAR-idwt(f.dwt)

    # block sums
    for(i in c(1:Anzahl.Blocks))
    {
    Tkl[i] <- sum(Residuals.data[i:(i+l-1)])
    }

    # distribution simulation
    for(Kol in c(1:sim.nr))
    {

    # rundom pull
    for(i in c(1:k))
    {
    Sample.Calcule[i] <- sum(sample(Tkl, k,replace = T))
    }
    Sample.Calcule <- Norm.Koef*Sample.Calcule
    Chislo[Kol]<-sum(Sample.Calcule**2)/k;
    }
    q09<-quantile(Chislo, c(.90))
    q095<-quantile(Chislo, c(.95))
    q0975<-quantile(Chislo, c(.975))

    # without hard Thresholding
    for(k.Shift.moth in c(1:p.Koeff))
    {
    di <- paste("d",k.Shift.moth,sep="")
    temp.f.dwt.di <- f.dwt[[di]]
    f.dwt[[di]] <- 0*temp.f.dwt.di
```

```
}

# residuals
Residuals.data <- x.FAR-idwt(f.dwt)

# block sums
for(i in c(1:Anzahl.Blocks))
{
Tkl[i] <- sum(Residuals.data[i:(i+l-1)])
}

# random pull
for(i in c(1:k))
{
Sample.Calcule[i] <- sum(sample(Tkl, k,replace = T))
}
Sample.Calcule <- Norm.Koef*Sample.Calcule
Test.Gr <- sum(Sample.Calcule**2)/k
Fault.number09<-Fault.number09+(Test.Gr > q09);
Fault.number095<-Fault.number095+(Test.Gr > q095);
Fault.number0975<-Fault.number0975+(Test.Gr > q0975);
}
print(Fault.number09/sim.nr)
print(Fault.number095/sim.nr)
print(Fault.number0975/sim.nr)
}
```