ROBUST WAVELET REGRESSION WITH AUTOMATIC BOUNDARY CORRECTION

by

ALSAIDI ALMAHDI MOHAMED ALTAHER

Thesis Submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

December 2012

ACKNOWLEDGMENTS

In the name of Allah, Most Gracious, Most Merciful

First of all, thanks to Allah Who gives me strength and patience along the accomplishment of this thesis. My first debt of gratitude goes to my supervisor, Dr. Mohd Tahir Ismail for his constant guidance, ceaseless support, patience, comments and suggestions throughout the completion of this thesis.

I am thankful to my family members, especially to my mother. Her prayers provided me a constant source of inspiration. In loving memory of my father who passed away in 1984, may Allah let us meet again in paradise.

I wish to dedicate this thesis to my beloved wife Rahma; and my children Mohamed and Ishraq. I am greatly indebted to their enthusiasm and strong support.

Finally, thanks to all friends, even more those who have provided helpful suggestions and encouragements.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

PWR	Polynomial Wavelet Regression
LPWR	Local Polynomial Wavelet Regression
MMR	Robust Estimation Based on MM-estimator with Robust Threshold
MMBR	Robust Estimation Based on MM-estimator, Bootstrap with Robust Threshold
OLSR	Robust Estimation Based on OLS with Robust Threshold
LPWR	Robust Estimation Based on Robust Local Polynomial Fitting with Robust Threshold
EbayesThresh	Empercal Bayesian Thresholding
G-EbaysThresh	Global Thresholding Using EbayesThresh
T-EbaysThresh	Term by Term Thresholding Using EbayesThresh
G-LDCV	Global Thresholding Using Cross Validation
G-LDCV	Term by Term Thresholding Using Cross validation

Regresi Wavelet Teguh dengan Pembetulan Sempadan Otomatik

ABSTRAK

Sejak kebelakangan ini, kaedah wavelet telah digunakan untuk menganggarkan fungsi yang tidak diketahui yang diperhatikan dalam kehadiran hingar. Satu kelebihan utama regresi wavelet adalah keupayaannya untuk mendedahkan aspek-aspek data yang mana tiada dalam teknik regresi klasik terlepas. Aspek seperti tren, titk perubahan, dan tidak berterusan boleh dipertimbangkan dengan kaedah wavelet. Walau bagaimanapun, penganggar regresi wavelet klasik mengalami masalah sempadan yang disebabkan oleh penggunaan penjelmaan wavelet terhadap isyarat terhingga. Akibatnya, berlakunya terpincang besar di tepi dan bergoyang tiruan. Walaupun regresi wavelet polinomial (PWR) dan regresi wavelet polinomial setempat (LPWR) berkesan mengurangkan risiko masalah ini, anggaran dari kedua-dua kaedah ini mudah dipengaruhi apabila andaian kenormalan tidak dipenuhi atau kehadiran data terpencil, maka anggaran tidak tepat akan diberikan. Semua pengiraan yang terlibat dalam melaksanakan anggaran terakhir adalah berdasarkan semata-mata atas andaian kenormalan yang diadakan. Tesis ini mencadangkan kaedah teguh yang berlainan dalam usaha untuk terus menggunakan idea PWR dan LPWR walaupun melampaui andaian yang biasa seperti titik terpencil, hingar Gaussian yang tak bersandar atau berkorelasi dan data rawak hilang. Maka, tesis ini dibahagikan kepada tiga bahagian.

Bahagian pertama memperkenalkan lima metodologi teguh untuk lanjutan pengesahan PWR dan LPWR dapat menerangkan data yang tercemar dengan titik terpencil dan hingar tak bersandar. Kaedah-kaedah ini adalah berdasarkan penuras median, penganggar-MM dan bukan OLS; langkah ambang teguh oleh Oh et al. (2009a), dan langkah ambang teguh oleh Oh et al. (2008). Bahagian kedua pula

memberi penekanan khas apabila struktur hingar berkorelasi. Kesan praktikal korelasi itu dijelaskan dengan menunjukkan pecahan beberapa kaedah popular pemilihan ambang. Kemudian, ambang tahap demi tahap telah disiasat sebagai penyelesaian yang lebih baik daripada menggunakan ambang global. Bahagian ini berakhir dengan memperkenalkan gabungan dua kaedah teguh. Kaedah-kaedah ini menggabungkan penganggar ambang Kovac dan Silverman (2000) kepada sama ada model polinomial setempat atau model polinomial menggunakan kaedah kuasa dua terkecil teritlak dan bukan kaedah kuasa dua terkecil yang biasa. Sebelum menggunakan kaedah ini cerapan terpencil dibuang menggunakan fungsi statistik yang sesuai. Di bahagian ketiga, disebabkan kedua-dua PWR dan LPWR adalah berdasarkan penjelmaan wavelet, kaedah yang dicadangkan memerlukan rekabentuk ruang titik yang sama jaraknya dengan jarak kuasa 2 ($n = 2^j$). Oleh itu, kaedah yang dicadangkan tidak boleh segera dilaksanakan dengan kehadiran cerapan rawak. Kekangan ini dihadapi dengan memperkenalkan dua kaedah imputasi yang efektif berdasarkan gabungan butstrap dan model regresi ramalan. Kaedah yang dicadangkan menggunakan anggaran awal butstrap yang akan mengulangi sama ada melalui model polinomial atau model polinomial setempat sehingga penumpuan. Pencapaian berangka kaedah yang dicadangkan telah dinilai melalui simulasi dan contoh data sebenar dengan perbandingan kaedah imputasi lain yang sedia ada.

ROBUST WAVELET REGRESSION WITH AUTOMATIC BOUNDARY CORRECTION

ABSTRACT

In recent years, wavelet methods have been applied to estimate an unknown function observed in the presence of noise. One major advantage of wavelet regression is its ability to reveal aspects of data that classical regression techniques miss. Aspects like trends, breakdown points, and discontinuities can be considered by wavelet methods. However, the classical wavelet regression estimator suffers from boundary problem caused by the application of wavelet transformations to a finite signal. As a result, large bias at the edges, and artificial wiggles occur. Although polynomial wavelet regression (PWR) and local polynomial wavelet regression (LPWR) effectively reduce the risk of this problem, the estimates from these two methods can be easily affected when the normality assumption is not met or in presence of outliers, give inaccurate estimates. All calculations involved in performing the final estimates are based solely on normality assumption to be held. This thesis proposes different robust methods in an attempt to keep using the idea of PWR and LPWR even beyond the usual assumptions of such outliers, independent or correlated non Gaussian noises and random missing data. Therefore, this thesis is divided into three parts.

The first part introduces five different robust methodologies to extend the validity of PWR and LPWR to describe data contaminated with outliers and independent noises. These methods are based on the median filter, *MM*-estimator, robust threshold procedure of Oh et al. (2009a), and robust threshold procedure of Oh et al. (2008). The second part pays special exception when the noise structure is correlated. The practical consequences of such correlation are explained; showing

the breakdown of several popular thresholding selection methods. Then, *level by level* thresholding has been investigated as a better solution than using global thresholding. This part ends up by introducing two robust combined methods. These methods combine the thresholding estimator of Kovac and Silverman (2000) with either local polynomial model or polynomial model using generalized least squares method instead of ordinary least squares. Before applying these methods the outlying observations are removed using an appropriate statistical function. In the third part, since both PWR and LPWR are based on wavelet transform, the proposed methods require equally spaced design points with length to a power of 2 ($n = 2^j$). Hence, the proposed methods can not be immediately implemented in the presence of missing data. We cope with this limitation by introducing two effective imputation methods based on a combination of bootstrap and regression prediction model. The proposed methods use initial-bootstrap estimates which will be iterated through either polynomial or local polynomial regression model until convergence. The numerical achievement of the proposed methods has been evaluated through simulation and real data examples with comparison to other existing imputation methods.

CHAPTER 1
INTRODUCTION

1.1 Overview

In almost all fields of study, finding relationships that exist in a set of variables subjected to random fluctuations is an ultimate goal in statistical analysis and curve estimation. Parametric models such as linear regression, polynomial and exponential would be preferable if there is good information about the appropriate model family. However, if there is insufficient past experience, little idea of an appropriate form for the model, then it is logical to let the data specify the function that fits them best without the constraints imposed by a parametric model. This freedom leads to a new class of regression modeling called *non parametric regression*. In the mean time, there exist several techniques for obtaining the regression function non-parametrically. While some methods are based on quite simple ideas, others are based on highly developed mathematics; and each has its own particular strength and limitations.

For the past two decades, there has been a great development of wavelet methods for estimating an unknown function, f observed in the presence of noise; following the pioneering work of Donoho and Johnstone (1994, 1995) where the concept of wavelet regression has been introduced to the statistical literature. In fact, wavelet regression possesses some key advantages as it is superior to traditional nonparametric regression methods. It can be viewed as orthonormal basis functions that are localized in both time and frequency domains, with time-widths adapted to their frequency. Besides, it has the advantages of ideal minimax property, spatial inhomogeneous adaptivity, and can be extended to high dimensions with fast algorithm. However, wavelet regression has been proven to be highly effective at estimating the unknown function as long as there are no boundary problems with f . Within boundaries; one side data information are available due to the bounded support of the underlying function. As a result, the application of the wavelet transformation to a finite signal provides a large bias at the edges, creating artificial wiggles. This problem has been recognized as a quite serious problem for easily getting boosted and may distort substantially the final estimates.

As treatment for boundary problem, two automatic boundary correction methods are used, namely: polynomial wavelet regression (PWR) in which the function f is decomposed as a combination of wavelet regression function, $f_W(x)$ and a low order polynomial model, $f_P(x)$ hoping to take the boundary problem into consideration by the polynomial model. For a noisy data set $\{(x_i, y_i, i = 1.2, ..., n\}$, the progress of polynomial wavelet regression involves two steps. First, regress the observed data $\{y_i\}_{i=1}^n$ on $\{x, ..., x^d\}$ for a fixed order, d. The second step is to apply wavelet regression to the residuals $\{e_i = y_i - \hat{f}_P(x)\}_{i=1}^n$ to obtain $\hat{f}_W(x)$. The final estimate of f will be the summation of $\hat{f}_P(x)$ and $\hat{f}_W(x)$. Another method is called local polynomial wavelet regression (LPWR). The idea is to combine wavelet regression, $f_W(x)$ with local polynomial regression model, $f_{Lp}(x)$ where the later is well known to posses excellent boundary properties.

1.2 Problem Statement

The boundary problem in wavelet regression has been studied previously by Oh et al. (2001), Lee and Oh (2004) and Oh and Kim (2008). In all previous work the Ordinary Least Squares (OLS) is usually used for getting the residuals at first step. However, the OLS estimates can be easily affected when the normality assumption is not met or in presence of outliers, correlated noise and missing data giving inaccurate estimates. Additionally, the polynomial order selection criteria use functions of the residuals from all the data to obtain the order that minimizes the discrepancy between the predicted and true models. Normality of the residuals again played a key role in deriving these criteria. The sensitivity of these estimation techniques to these underlying assumptions has been identified as a weakness that can even lead to wrong interpretations. Thus, both PWR and LPWR are very sensitive to outliers, correlated noise and missing data.

1.3 Objectives

The main objectives of this thesis are

1. to propose robust methods in such a way that extend the validity of using both PWR and LPWR for describing data with independent noise and in presence of outliers.

2. to propose robust methods in such a way that extend the validity of using both PWR and LPWR for describing data with correlated noise and in presence of outliers.

3. to propose effective imputation methods in order to relax the classical sample size condition, so the proposed methods can be applied even in the presence of missing data where the data size is no longer a power of 2.

1.4 Scope of the Study

This thesis focuses on the boundary problem of wavelet regression in the presence of outliers, independent noise, correlated noise. Noises such as mixture normal noise, heavy tailed noise and noises from autoregressive process and moving average will be considered. The presence of missing at random will also be considered to prevent increasing the bias again once it has been corrected. Only two automatic boundary correction estimators are considered i.e. PWR and LPWR for one dimension. In order to provide a convenient framework for the discussion of different thresholding methods, we only focus on equally spaced design points in which the designer chooses the predictor variable values by using a predefined scheme.

1.5 Significance of the study

Wavelet regression is a new non parametric technique with the ability to reveal unusual aspects which might be observed in noisy data. Aspects like trends, breakdown points, and discontinuities can be well taken into consideration by wavelet methods. In addition, due to the fact that wavelet regression affords a different view of data, wavelet regression is capable in denoising a signal without appreciable degradation. Because of the boundary problem and its effects on the final estimates, there is a need to keep in level with these excellent advantages when the estimation at the boundary region is greatly important. We believe that in many situations in our daily life more concern should be paid to the estimation at the boundary region; say for instance, if the data are related to poverty analysis, it will be necessary to get reliable estimates of the income distribution on the left side, close to "0" (left boundary point). Similarly, when wavelet regression is employed to describe econometric data, looking at the performance of specifically old or young people, comparing large and small companies, etc., we always focus on the boundary's estimates. In such cases data may contain outliers or they may come from populations where the normality assumption is not met or the noise is correlated. Therefore, boundary problem must be taken seriously and effective solutions in robust frame are required .

1.6 Methodology

To perform the objectives of this research we have built the following methodologies.

- We shall first use the so called *median filter* as a primary step to filter the data from outliers then; both PWR and LPWR can be used. The advantages and disadvantages of these methods will be shown.

- For both PWR and LPWR, we suggest to keep using OLS at first step but using robust threshold procedure given by Oh et al. (2009a) at second step.

- For polynomial wavelet regression (PWR):

  1. We suggest using a robust regression estimator instead of OLS; particularly we shall use *MM*-estimator at first step with robust threshold of Oh et al. (2009a) at second step.

  2. A new robust order selection criterion will be proposed. This method is based on bootstrap method by minimizing the predicted mean squared error with a parametric penalty function; with robust threshold procedure of Oh et al. (2009a).

- For LPWR, the robust local polynomial estimator of Oh et al. (2008) will be used; with robust threshold procedure of Oh et al. (2009a).

- In case of correlated noise we combine the thresholding estimator of Kovac and Silverman (2000) with either local polynomial model or polynomial model using the generalized least squares method instead of ordinary least squares. Before applying these methods the outlying observations are removed using an appropriate statistical function.

- In the presence of missing at random, two effective imputation methods will be proposed and compared with existing ones under different conditions. The proposed methods use initial-bootstrap estimates which will be iterated through either polynomial or local polynomial regression model until convergence. The numerical achievement of the proposed methods will be evaluated through simulation experiments and real data example with comparison to other existing imputation methods.

1.7 Outlines

This thesis is organized in the following manner: Chapter 2 gives a general review of wavelets and some other nonparametric regression techniques. In this chapter we also discuss several thresholding selection methods used in the later chapters; together with emphasis on the two automatic boundary correction estimators PWR and LPWR. At the end, we review some previous robust regression techniques. In chapter 3, we propose five different robust estimation methods with automatic boundary corrections to reduce the effects of outliers and independent non Gaussian noise. A performance evaluation of the proposed methods is also described here. Chapter 4 investigates the level by level thresholding methods when the noise is correlated followed by introducing two new robust-combined methods to cope with the difficulty found when PWR and LPWR are considered for recovering a signal from noisy data contaminated with outliers and correlated noise. In chapter 5, the problem of missing at random is considered. Two effective imputation methods are proposed and compared with existing ones under different conditions. Finally, conclusions and suggestions for further research are drawn in chapter 6.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter presents some of the necessary background material relevant to subsequent chapters. More details with deep mathematical distributions can be found in Daubechies (1992), Meyer (1993), Chui (1997), Mallat (1999), Burrus et al. (1998), and Walter and Shen (2001). Vital important papers concerning wavelet theory can be found in Heil and Walnut (2006).

2.2 Notations

Definition 2.1 *The $\mathbb{R}$ refers to the set of real numbers whereas $\mathbb{Z}$ refers to the set of integers.*

Definition 2.2 *The space of measurable function, $L^p(\mathbb{R}), 1 \leq P < \infty$ is defined as*

$$L^p(\mathbb{R}) = \{f : \int_{-\infty}^{+\infty} \mid f(x) \mid^p dx < +\infty\}$$

Definition 2.3 $L^2[0,1]$*: represents the space of all square-integrable functions on the closed interval* $[0,1]$.

Definition 2.4 *The inner product of two functions $f(x), g(x) \in L^2(\mathbb{R})$ is defined as*

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

2.3 Basis Vector and Basis Function

The two-dimensional vector $(x; y)$ can be written as a linear combination in terms of two vectors such (1; 0) and (0; 1): These two vectors are considered as the basis vectors for $(x; y)$ since when x is multiplied by $(1; 0)$, the result will be the vector

$(x; 0)$; and y multiplied by $(0; 1)$ gives the vector $(0; y)$: The sum is the original vector, $(x; y)$.

In similar way as shown in Graps (1995), the basis function can be explained. Assume that we have the function $f(x)$ instead of the vector $(x; y)$. The function f in the function space can be represented as a linear combination of basis functions as any vector in a vector space is represented as a linear combination of basis vectors. For example, the collection of second order polynomials with real coefficients will have $\{1, x, x^2\}$ as a basis. Therefore, every quadratic polynomial can be written as a linear combination of the basis functions $1, x$, and x^2, that is

$$f(x) = a\mathbf{1} + bx + cx^2$$

where $\mathbf{1}$ is a vector of unity (its elements are equal to 1).

One advantage of basis function is that it varies in scale by chopping up the same function using different scale sizes. For example, if the function f is defined over the interval $[0, 1]$, this function can be divided into two step functions ranging from 0 to 1/2 and 1/2 to 1. Similarly, again the original function can be divided using four step functions, ranging from 0 to 1/4, 1/4 to 1/2, 1/2 to 3/4, and 3/4 to 1 and so on. With this scheme each set of these representations codes the original signal with a particular scale.

2.4 Fourier Analysis

Definition 2.5 *Fourier theory states that the function f can be represented in terms of sum of* sine *and* cosine *functions, that is*

$$f(x) = \frac{a_0}{2\pi} + \frac{1}{\pi}\sum_{n=1}^{\infty}(a_n cos(nx) + b_n sin(nx))$$

where

$$a_n = \int_0^{2\pi} f(x)cos(nx)\,dx; \;\; n \geq 0$$
$$b_n = \int_0^{2\pi} f(x)sin(nx)\,dx; \;\; n \geq 1$$

Representations in this form is possible since the collection $\{1,\, cos(nx),\, sin(nx)\}$ constitutes an orthonormal basis of $\mathbb{L}^2([0,\, 2\pi])$; see Goodwin (2008).
The fourier transform of a function $f \in \mathbb{L}^1(\mathbb{R})$ can be defined as

$$\begin{aligned}\hat{f}(w) &= \mathcal{F}[f(w)] \\ &=< f(x), e^{iwx} > \\ &= \int_{\mathbb{R}} f(x)e^{iwx}dx\end{aligned}$$

The inverse Fourier transform is written as

$$f(x) = \mathcal{F}^{-1}[\hat{f}(w)] = \frac{1}{2\pi}\int_{\mathbb{R}} \hat{f}(w)e^{iwx}dw$$

2.5 History of Wavelets

The fundamental history of wavelet relates back to the wavelet transform coined by the mathematician Alfrd Haar in 1910. At that time there was no a clear conception of wavelet. Only in 1984 the concept was proposed by the geophysicist Jean Morlet. At that time the only orthogonal wavelet that people knew was Haar wavelet. Soon after, in 1985 the mathematician Yves Meyer constructed the second orthogonal wavelet called Meyer wavelet. Rapidly more and more scholars joined in this new field, gathering in the first international conference which was held in France in 1987. In 1988, Stephane Mallat and Meyer introduced the concept of multiresolution. In the same year, Ingrid Daubechies invented a systematical method constructs the compact support orthogonal wavelet. In 1989, the fast wavelet transform was proposed by Mallat. With the appearance of this fast algorithm to wavelet community, the wavelet transform had numerous applications in many useful applications such as data compression, signal processing, detecting features in images, and removing noise from time series. Finally, the concept of wavelet regression was introduced to the statistical literature by Donoho and Johnstone (1994).

2.6 What is a wavelet

The word *wavelet* goes back to Morlet and Grossmann in the early 1980s. The French word *ondelette*, meaning *small wave* was used. Later, the same word was translated to English by translating *onde* into *wave*, giving *wavelet.*

From mathematical point of view wavelets are functions that break up data into distinct frequency components, and then each component is studied with a resolution matched to its scale. These wavelets comprise the family of translations and dilations of a single function, denoted ψ, which is called a mother wavelet. The mother function was defined by Meyer (1993) as follows

Definition 2.6 : *Let m be a non negative integer. A function $\psi(x)$ of order m of a real variable is said to be a mother function if the following properties are held:*

1. *If $m = 0, \psi(x) \in L^{\infty}(\mathbb{R})$; if $m \geq 1$, then $\psi(x)$ and all its derivatives up to order m belong to $L^{\infty}(\mathbb{R})$.*

2. *$\psi(x)$ and all its derivatives up to order m decrease rapidly as $x \to \pm\infty$.*

3. *For each $k \in \{0, \ldots, m\}$,*

$$\int x^k \psi(x) dx = 0$$

4. The collection $\{\psi_{j,k}\}_{j,k\in\mathbb{Z}}$ is an orthonormal basis of $\mathbb{L}^2(\mathbb{R})$, the $\psi_{j,k}$ being constructed from the mother function using

$$\psi_{j,k}(x) = 2^{\frac{j}{2}}\psi(2^j x - k)$$

The definition given above reveals that the mother wavelet function is not an arbitrary function, but it should be compact support with N vanishing moments. Below is the explanation to these terminologies:

- **Compact Support**

  Support: the region of the mother wavelet where it is not equal to zero.

  Compact support: the width of the support is finite

- **Vanishing Moments**

  The vanishing moment is a numerical criterion describes the decay of a function toward infinity. A function $\psi \in L^2(\mathbb{R})$ is said to have N vanishing moments if it satisfies

  $$\int x^l \psi(x)\, dx = 0; \quad l = 0, ..., N-1.$$

  Vanishing moments are extremely useful. A wavelet with N vanishing moments tells that wavelet coefficients of any polynomial of degree m or less will be exactly zero. Thus, for any smooth function with only a little discontinuity or some other singularity aspects, the wavelet coefficients on the smooth parts will be very small or even zero. This is a key advantage of wavelet transform in that, having only a few non-zero coefficients means that there are only a few coefficients need to be estimated. In terms of information, it is much better to have data of size n to estimate a few coefficients rather than data of size n to estimate n coefficients; see Nason (2008).

2.7 Some Wavelet Bases

2.7.1 The Haar Basis

The most basic wavelet basis is the Haar basis discovered by the mathematician Alfred Haar in 1910. It is usually constructed from a scaling function ϕ and mother wavelet ψ as below. A picture of the Haar wavelet basis is depicted in Figure 2.1. One major disadvantage of Haar basis is that it is not continuously differentiable which somewhat limits its applications.

$$\phi(x) = \begin{cases} 1 & \text{if } x \in [0,1) \\ 0 & \text{otherwise} \end{cases}, \quad \psi(x) = \begin{cases} 1 & \text{if } x \in [0,\frac{1}{2}) \\ -1 & \text{if } x \in [\frac{1}{2},1) \\ 0 & \text{otherwise} \end{cases}$$

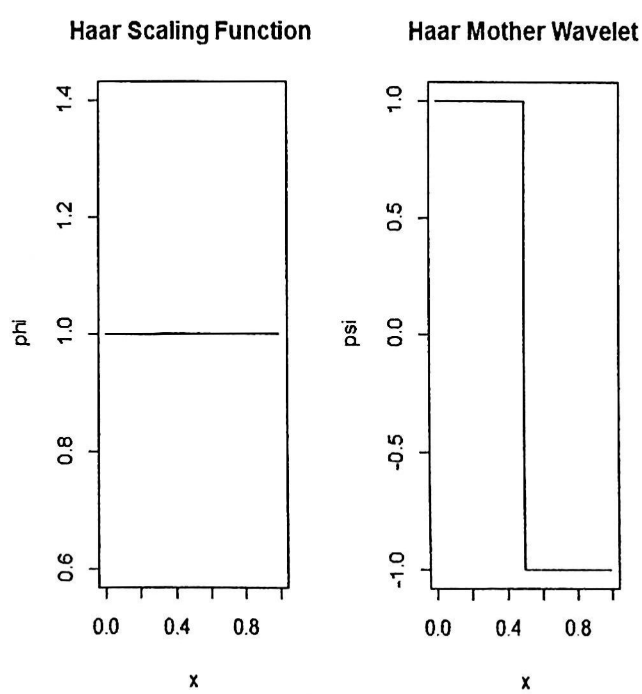

Figure 2.1: Haar Mother Wavelet and Scaling Function

2.7.2 Daubechies Wavelet

Daubechies (1992) introduced two families of compactly support wavelets namely: Daubechies *extremal phase* and *least asymmetric wavelets.* Some examples of Daubechies mother wavelet with different degree of smoothing are shown in Figure 2.2.

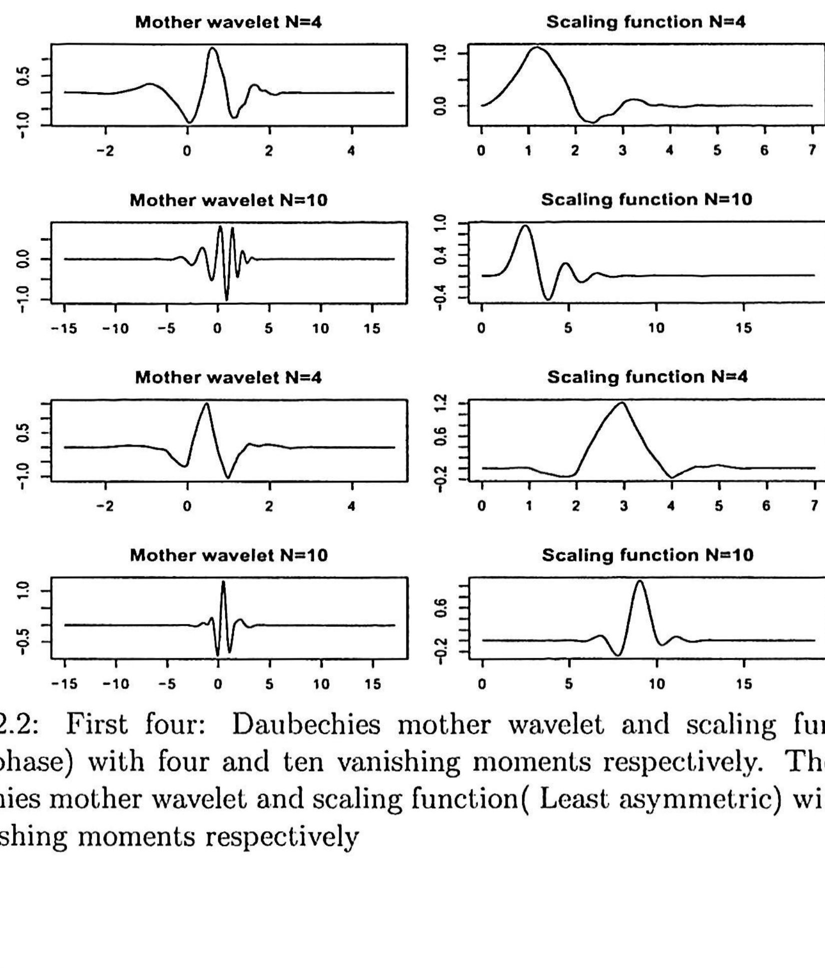

Figure 2.2: First four: Daubechies mother wavelet and scaling function(extremal phase) with four and ten vanishing moments respectively. The last four, Daubechies mother wavelet and scaling function(Least asymmetric) with four and ten vanishing moments respectively

2.8 Multiresolution Analysis

Multiresolution analysis (MRA) is a general tool for the construction of orthonormal wavelet basis. The fundamental concept of (MRA) is given in Mallat (1989) as follows:

Definition 2.7 *A multiresolution analysis is a chain of nested closed subspaces,* $\{V^j . j \in \mathbb{Z}\}$, *satisfying the following conditions:*

1. *The spaces have trivial intersection:*

$$\bigcap V^j = \{0\},\ j \in \mathbb{Z}$$

2. *The union is dense in $L^2(\mathbb{R})$:*

$$\bigcup V^j = L^2(\mathbb{R}),\ j \in \mathbb{Z}$$

3. *The following scale relations exist*

$$f(x) \in V^j \Leftrightarrow f(2x) \in V^{j+1} \tag{2.1}$$

$$f(x) \in V^0 \Leftrightarrow f(x-k) \in V^0 \tag{2.2}$$

4. *There exists a function $\phi(x) \in V^0$ such that the sequence $\{\phi(x-k), k \in \mathbb{Z}$ is an orthonormal basis of V^0.*

The subspaces V^j are called levels. One level is said to be coarser (or finer) than another level, if its index is less (or greater) than the other index. The function ϕ is called the scaling function or father wavelet.

The orthogonal complement W^j of V^j can be found such that

$$V^{j+1} = V^j \oplus W^j,$$

where the symbol $\oplus$ stands for direct sum.

In general, $V^j = V^{j-1} \oplus W^{j-1}$.

Conditions (3) above implies that $\forall j \in \mathbb{Z}$, $\{\phi_{j,k} : k \in \mathbb{Z}\}$ constitutes an orthonormal basis for V^j,where

$$\phi_{j,k}(x) = 2^{j/2}\phi(2^j x - k)$$

Daubechies (1988) introduced a projection operator P_j that projects a function into the space V^j. Since $\{\phi_{j,k}(x)\}_k$ is a basis for V^j, the projection is written as

$$f_j(x) = \sum c_{j,k}\phi_{j,k}(x) = P_j f$$

for some coefficients $\{c_{j,k}\}$, see Nason (2008).

Condition (1) implies that, as $j \to -\infty$ all details of f are lost, and $\{P_j f\}$ converges to $\{0\}$ in an L^2 space

$$\lim_{j\to-\infty} P_j f = 0$$

On the other hand, condition (2) ensures that the signal approximation converges to the original signal in the same sense

$$\lim_{j\to\infty} P_j f = f$$

2.9 Discrete Wavelet Transform (DWT)

In statistics many problems arise as a sequence of observations, thus, wavelet analysis of sequences (Discrete Wavelet Transform) is usually considered rather than functions (Continuous Discrete Wavelet Transform). The Discrete Wavelet Transform (DWT) is an efficient algorithm, proposed by Mallat (1989) to obtain the wavelet coefficients of a discrete series through either linear operator or matrix representation.

- **Linear Operator**

  The DWT can be defined in terms of two filters: a low pass filter denoted as $\mathcal{H} = \{h_k\}$ and a high pass filter denoted as $\mathcal{G} = \{g_k\}$. The $\{h_k\}$ and $\{g_k\}$ denote the coefficients of the filters.

  Let f be a function defined on equally spaced design points x_k. $k = 0, 1, \ldots, n-1$. Let $c_{j,k} = f(x_k)$. The discrete wavelet transform of $c_{j,k}$ can be obtained using the following relations:

$$c_{j-1,k} = \sum h_{n-2k}\, c_{j,n}$$

$$d_{j-1,k} = \sum g_{n-2k}\, c_{j,n}$$

Here $d_{j-1,k}$ is called wavelet coefficients (detail coefficients) while $c_{j-1,k}$ are known as father wavelet (or scaling coefficients) . Note that the term $2k$ refers to filtering the data using $\mathcal{H}$ and $\mathcal{G}$ and then selecting every even observation from the series. For more details and possible choices for the filtering coefficients, see Mallat (1989).

- **Matrix representation**

  Another way to find the DWT is to use an orthogonal matrix W associated with orthonormal wavelet basis from mother wavelet. This matrix yields a vector $\mathbf{w}$ of the wavelet coefficients of the series y via

$$\mathbf{w} = W\mathbf{y}$$

  The inverse formula can be easily written as

$$\mathbf{y} = W^T\mathbf{w}$$

To demonstrate how wavelet transform matrix is generated we here present an example using the Haar scaling and translating functions defined in Section 2.7.1 for 8 observations.

Define $\phi_{j,k} : \mathbb{R} \to \mathbb{R}$ as

$$\phi_{j,k}(x) = \sqrt{2^j}\phi(2^j x - k), \ j = 0, 1, \ldots \text{ and } k = 0, 1, \ldots, 2^j - 1 .$$

Let

$$f = (y_1, y_2, ..., y_n)$$

Define the vector space V^j as

$$V^j = span\{\phi_{j,k}\},\ k = 0, \ldots, 2^j - 1$$

Define $\psi_{j,k} : \mathbb{R} \to \mathbb{R}$ as

$$\psi_{j,k}(x) = \sqrt{2^j}\psi(2^j x - k),\ j = 0, 1, \ldots \text{ and } k = 0, 1, \ldots, 2^j - 1\ .$$

Define the vector space W^j

$$W^j = span\{\psi_{j,k}\},\ k = 0, \ldots, 2^j - 1$$

Since we have 8 observations ($n = 2^3$), we can expand f into V^3 such: $V^3 = V^0 \oplus W^0 \oplus W^1 \oplus W^2$. This can be performed in three steps.

Step 1: Expanding f into $V^2 \oplus W^2$.

The first step starts by using 8-wavelet transform matrix W_1 which can be described as

$$W_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix}.$$

The first four rows in matrix W_1 represent the basis vectors $\phi_{2,0}$, $\phi_{2,1}$, $\phi_{2,2}$ and $\phi_{2,3}$ respectively which span V^2 whereas the last four rows correspond to the basis vectors $\psi_{2,0}$, $\psi_{2,1}$, $\psi_{2,2}$ and $\psi_{2,3}$ which span W^2. The matrix W_1 expands the signal into $V^2 \oplus W^2$.

Step 2: Expanding f into $V^1 \oplus W^1 \oplus W^2$.

The coefficients for the basis vectors of W^2 have been already computed in step 1. Hence, there is no need to find these coefficients again in this step, we will keep the last four entries in the previous step and only work with the first four entries. The wavelet transform matrix in this step can be described as

$$W_2 = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

The first and the second step can be combined as W_2W_1 where

$$W_2W_1 = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}.$$

In this matrix the first two rows correspond to the basis vectors $\phi_{1,0}$ and $\phi_{1,1}$ which span V^1. The third and fourth rows correspond to the basis vectors $\psi_{1,0}$ and $\psi_{1,1}$ which span W^1. The last four rows correspond to the basis vectors of W^2.

Step 3: Expanding f into $V^0 \oplus W^0 \oplus W^1 \oplus W^2$.

The coefficients for the basis vectors of $W^1 \oplus W^2$ have been already computed. Therefore, these coefficients are kept in this step (the last six entries in W_2W_1) and we only need to get the first two entries. In this case the matrix wavelet transform can be described as

$$W_3 = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

The first, second and third step can be combined to get $W = W_3W_2W_1$ where

$$W = W_3W_2W_1 = \begin{pmatrix} \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} \\ \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}.$$

So the discrete wavelet transform coefficients at level 3 can be found as

$$\begin{pmatrix} \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} \\ \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \end{pmatrix}$$

Similarly, the discrete wavelet transform coefficients at level 2 and level 1 can be found as

$$W_2 W_1 f$$

$$W_1 f$$

respectively. Wavelet matrix can be easily found using the command function *GenW* (available in R in WaveThresh package).

2.10 Wavelet Transform vs Fourier Transform

There are some similarities and differences between Fourier transform and wavelet transform. Fourier transform (FT) and Wavelet Transform (WT) are both linear operations localized in frequency domain. The mathematical properties of the matrices involved in computing these transforms are similar as well. Both transforms can be viewed as a rotation in function space to a different domain since the inverse transform matrix for both FT and the WT is the transpose of the original matrices.

The major difference between FT and WT is that wavelet functions are localized in in time domain meanwhile Fourier functions (sine and cosine) are not. "This localization feature, along with wavelets localization of frequency,makes many functions and operators using wavelets sparse when transformed into the wavelet domain. This sparseness, in turn, results in a number of useful applications such as data compression, detecting features in images, and removing noise from time series". Graps (1995).

2.11 Classical Nonparametric Regression

This section gives a brief introduction to some common non parametric techniques such as kernel regression, local polynomial regression, orthogonal series, and smoothing spline.

2.11.1 Kernel Regression

Kernel regression in its simplest form, it is just a moving average estimator. In general, the kernel estimator $\hat{f}_h(x)$ is given by:

$$\hat{f}_h(x) = \frac{1}{nh}\sum_{j=1}^{n} K\left(\frac{x - x_j}{h}\right) y_j = \frac{1}{n}\sum w_j y_j,$$

where $w_j = K\left(\frac{x-x_j}{h}\right)/h$.

If the $x's$ are spaced very unevenly, then this estimator can give poor results. This problem is modified by the Nadaraya-Watson estimator as defined by Faraway (2006):

$$\hat{f}_h(x) = \frac{\sum w_i y_i}{\sum w_i}$$

where $K_h(x)$ is the kernel function with scale factor h. The kernel function K is usually chosen to be a continues, bounded and symmetric function satisfying:

$$\int_{-\infty}^{\infty} K(x)dx = 1$$

The most commonly used kernel functions include:

- The Gaussian kernel

$$K(x) = \frac{1}{\sqrt{2\pi}} exp\left(\frac{x^2}{2}\right)$$

- The uniform kernel

$$K(x) = \begin{cases} 1/2; & x \in [-1, 1] \\ 0; & \text{otherwise} \end{cases}$$

- Epanechnikov kernel

$$K(x) = \begin{cases} \frac{3}{4}(1 - x^2); & x \in [-1, 1] \\ 0; & \text{otherwise} \end{cases}$$

- The Triangular kernel

$$K(x) = \begin{cases} (1 - |x|); & |x| \leq 1 \\ 0; & \text{otherwise} \end{cases}$$

2.11.2 Local Polynomial Regression

Local polynomial regression (LPR) is similar to kernel estimation unless the fitted values of (LPR) are produced by locally weighted regression rather than locally weighted averaging.

Suppose that the regression function $f(x)$ is smooth enough to be approximated by Taylor's expansion such

$$f(x) \approx \sum_{j=0}^{\infty} \frac{f^{(j)}(x_0)}{j!}(x - x_0)^j = \sum_{j=0}^{\infty} \beta_j (x - x_0)^j,$$

where x is close to a point x_0. The least squares estimator is given as

$$\hat{f}(x; d; h) = e_1'(X_d' W X_d)^{-1} X_d' W Y$$

Here d refers to the order of local polynomial regression while h is smoothing parameter. W_x is diagonal matrix of weights with: $W_j \equiv K_h(x_j - x) \equiv \{K[(x_j - x)/h]\}$, e_1 is the $(d+1) \times 1$ vector having 1 in the first entry and zeros elsewhere. K is a kernel function.

if d=2

$$X = \begin{pmatrix} 1 & (x_1 - x) & (x_1 - x)^2 \\ 1 & (x_2 - x) & (x_2 - x)^2 \\ . & . & . \\ . & . & . \\ 1 & (x_n - x) & (x_n - x)^2 \end{pmatrix}$$

2.11.3 Orthogonal Series Estimation

Let ϕ_j, $j = 1, 2, ..., n$, be a sequence of orthonormal basis functions with respect to the design points $x_1, ..., x_n$, that is

$$\sum_{i=1}^{n} \phi_j(x_i)\phi_k(x_i) = \delta_{jk} = \begin{cases} 0 & \text{if } j \neq k; \\ 1 & \text{if } j = k. \end{cases}$$

The regression function can be written as

$$f(x_i) = \sum_{j=1}^{n} a_j \phi_j(x_i), \tag{2.3}$$

where $a_j = \sum_{k=1}^{n} f(x_k)\phi_j(x_k)$. The coefficients estimated can be computed as:

$$\hat{a}_j = \sum y_k \phi_j(x_k), \; j \in I \subseteq \{1, 2, ..., n\}$$

For some possible choices of basis functions; see Tarter and Lock (1993).