Efficient Rank Regression with Wavelets Estimated Scores

by

Eddy Armand Kwessi Nyandjou

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
August 6, 2011

Keywords: Regression, Wavelets, Scores

Approved by

Asheber Abebe, Chair, Associate Professor of Mathematics and Statistics
Geraldo De Souza, Professor of Mathematics and Statistics
Nedret Billor, Associate Professor of Mathematics and Statistics
Peng Zeng, Associate Professor of Mathematics and Statistics

Abstract

Wavelets have been widely used lately in many areas such as physics, astronomy, biological sciences and recently to statistics. The main goal of this dissertation is to provide a new contribution to an important problem in statistics and particularly nonparametric statistics, namely estimating the optimal score function from the data with unknown underlying distribution. This problem naturally arises in nonparametric linear regression models and could be important in order to have a better insight on more important and actual problems in longitudinal and repeated measures analysis through mixed models. Our approach in estimating the score function is to use suitable compactly supported wavelets like the Daubechies, Symlets or Coiflets family of wavelets. The smoothness and time-frequency properties of these wavelets allow us to find an asymptotically efficient estimator of the slope parameter of the linear model. Consequently, we are also able to provide a consistent estimator of the asymptotic variance of the regression parameter. For related mixed models, asymptotic relative efficiency is also discussed.

Acknowledgments

Without your blessings Almighty God, I could never have achieved this. Thank you so much for the strength, the health, the knowledge you gave throughout these years and simply throughout my life.

I would like to dedicate this dissertation to my late parents NYANDJOU Felix and NGUEBA Marie. Even though they did not live long enough to see this achievement, every step of my journey was made possible because of the sacrifices they put in raising me and my brothers. I wish to express my immense gratitude to the following:

My wife Corine Michelle KWESSI for her patience, understanding and total support and for taking care of our wonderful children Eddy Harold KWESSI KWESSI, Alexandra Felixia KWESSI and Alyssa Francesca KWESSI.

Dr. Asheber Abebe who suggested and supervised this work. His understanding, availability and encouragement throughout this dissertation will never be forgotten.

Dr. Geraldo De Souza for helping me understand the field of Harmonic Analysis, for mentoring me like a father, for being there when I needed him.

Dr. Nedret Billor for her valuable advices and for playing other wonderful and immeasurable roles.

Drs. Michel Smith and Chris Rodger for their total support and encouragement during my short stay at Auburn University as a graduate student.

All Professors who taught during the course work for the marvelous job done, particularly Drs. Peng Zeng, Mark Carpenter, Gary Gruenhage, Paul Schmidt, Amnon J. Meir, Narenda Govil.

Pr. Charles Chidume whose encouragement and support has made my journey to the US possible.

My Brothers Blaise Tchepnda Nyandjou, Thierry Roland Nono Nyandjou, Sidoine Raoul Youtcha Nyandjou, Blaise II Tchepnda, my sisters Pelagie Njinkeu Nyandjou, Elysee Tchatchoua Nyandjou, Grace Signe Nyandjou, Christelle Nyandjou for all the time we spent together and all the support they provided to me. My course mates, Guy-vannie Miankokana, Achard Bindele, Guy Merlin Nguelefack, Sean O'Neil, Frank Sturm, Daniel Roberts, Drs. Gerald Chidume, Julian Allagan, Ngwane Fidele, Paul Alfonso, for the fun we had at Auburn. My Friends, Alain Desire Kounkeu, Ruben Bayemi, Vales Ngoule, Moses Ntam, Drs. Foutse Khomh, Etienne Non, for their support.

Table of Contents

Abstract ii

Acknowledgments iii

List of Figures vii

List of Tables viii

1 Introduction 1

1.1 Contribution of the Dissertation 2

1.2 Organization 3

2 Preliminaries: Review of Rank Based Analysis and Wavelet Theory 4

2.1 Rank Based Analysis 4

2.1.1 Estimation of regression parameters 5

2.1.2 Tests of linear hypothesis 12

2.2 Basic Wavelet Theory 13

2.2.1 A motivating example 13

2.2.2 Examples of wavelets 17

3 Estimation of the Score Function 22

3.1 Introduction 22

3.2 General Assumptions 28

3.2.1 Discussion of the assumptions 28

3.3 Estimation of the coefficients 32

3.4 Estimation of the score function 37

4 Estimation of the Slope Parameter 39

5 Discussion 43
5.1 Issues related to simulations 43
5.2 Simple Mixed Models with Dependent Error Structure 45
5.3 Adequate space for score functions 49
Bibliography 58
Appendix 62

List of Figures

2.1 Geometry of Rank Estimation 9

2.2 Geometry of Rank Tests 13

2.3 Gibbs Phenomenon 16

2.4 The Haar Wavelet 17

2.5 The Shannon Wavelet 18

2.6 The Mexican Hat Wavelet 19

2.7 The Daubechies Wavelet 20

5.1 Simulations 44

List of Tables

5.1 Comparison between the different methods 45

Chapter 1

Introduction

One of the most widely used models in statistical modeling is the linear regression model where a hyperplane that 'best' describes the relationship between a response variable Y and a vector of covariates $\mathbf{x}$ is constructed. Typically, one has a set of data measured on, say, n subjects $(Y_1, \mathbf{x}_1), \ldots, (Y_n, \mathbf{x}_n)$ and the construction of the hyperplane involves errors since not all the data points fall on a plane. These errors are assumed to be random.

The classical approach to estimating the hyperplane is using the least squares (LS) procedure where the hyperplane is taken to be the column space spanned by the columns of the matrix $(\mathbf{x}_1^T, \ldots, \mathbf{x}_n^T)^T$ that is the closest in terms of the Euclidean distance to the vector $(Y_1, \ldots, Y_n)^T$. It turns out, by the Gauss-Markov Theorem, that the LS procedure gives the model that is the best linear unbiased estimator (BLUE) if the errors have expectation zero, are uncorrelated, and have equal variances. However, the LS procedure is not robust in the presence of outliers and other violations of underlying assumptions.

There are a several number of approaches one can follow to achieve robustness. One of the more popular techniques is the technique of M-estimation (Huber, 1964) which includes the LS procedure as a subset. Another approach was the method of R-estimation which is based on ranks of the data. This was initially proposed for the simple linear model by Adichie (1967) based on simple Hodges-Lehmann type location estimators. This was later generalized for the multiple regression model by Jureckova (1971) and Jaeckel (1972). In subsequent works, Hettmansperger and McKean Naranjo and McKean (1997), and colleagues developed methods for testing general linear hypotheses, construction of confidence intervals, etc. to make R-estimation of linear models a complete treatment (Hettmansperger and McKean, 1998). As a result, the approach is also known as the Jaeckel-Hettmansperger-McKean approach of model fitting (Hollander and Wolfe, 1999).

1.1 Contribution of the Dissertation

One of the difficulties of using M- and R-estimators is that they depend on an unknown function (a score function) that needs to be chosen by the investigator. This is just about all the control the investigator has so it is a very critical activity. However, it is usually not very clear which score functions to use. A common approach is to choose the function on the basis of a heuristic investigation of a fit based on a chosen (usually simple such as linear) score function. Another common approach is to go for robustness by sacrificing efficiency and use score functions that contain some form of trimming or Winsorization.

It is of interest to choose score functions that maximize the efficiency of the resulting estimator. The efficiency of an estimator depends on the underlying distribution of the random error terms and this distribution is unknown. One approach is to use certain density estimators (e.g. kernel density estimator) to estimate the score function and use the estimated score function to estimate the regression parameters. Such estimators, known as adaptive estimators, are discussed in Stone (1975) and Koul and Susarla (1983) among others.

A different approach that uses the underlying structure of R-estimators to determine the score function that maximizes efficiency was given by Dionne (1981) and Naranjo and McKean (1997). They employed Fourier series approximation based on a first order Taylor expansion to determine the score function that maximizes the efficiency of the R-estimator. However, this lacked the flexibility that is required for certain underlying distribution. This was especially evident at the two extremes of the domain of the score function.

In this dissertation, we propose a Wavelet based approximation of the score function based on a second order Taylor series approximation. As it turns out the second order Taylor approximation is ideal under the smoothness considerations of the score function. We will provide a theoretical investigation of the optimality of this approach as well as establish the asymptotic properties of the resulting estimator. We will also consider the mixed model and investigate how our approach can be used to estimate the fixed-effects parameters. Moreover, we will take on the task of determining the function space that is most suited for such Wavelet-based approximation. This leads us to consider the problem with greater generality from the perspective of harmonic analysis.

1.2 Organization

This dissertation is organized as follows. Chapter 2 contains a brief review of rank based analysis of linear models and wavelet theory. In Chapter 3, we consider the problem of estimation of score functions and give the main results of the dissertation. Chapter 4 gives an adaptive estimator of the slope parameter as well as the scale parameter. These are then used to construct Wald type tests for general linear hypotheses on the slope parameter. Chapter 5 provides a brief discussion on some issues related to computations and includes proposals to include dependent error structure as well as the spaces of score functions that are suitable for wavelet approximation.

Chapter 2

Preliminaries: Review of Rank Based Analysis and Wavelet Theory

In this chapter, we review basic notions of rank based analysis for linear models and basic facts about wavelet theory.

2.1 Rank Based Analysis

We consider the following model:

$$Y_i = \mathbf{x}_i^T \boldsymbol{\beta} + e_i^*, \quad 1 \leq i \leq n, \tag{2.1.1}$$

where Y_i denotes the ith response and $\mathbf{x}_i$ denote a $p \times 1$ vector of explanatory variables, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown regression parameters and e_i^* is the ith random error with distribution F.

Our interest is to estimate $\boldsymbol{\beta}$ and to test linear hypotheses about it. Note that we could also have a model with intercept parameter. Indeed, let $\mu = L(e_i^*)$ be a location functional and let $e_i = e_i^* - \mu$. Then $L(e_i) = 0$ and the model can be written as

$$Y_i = \mu + \mathbf{x}_i^T \boldsymbol{\beta} + e_i, \quad 1 \leq i \leq n, \tag{2.1.2}$$

Remark 2.1.1. $\boldsymbol{\beta}$ does not depend on the location functional used if F is a member of the location family of distributions.

Proof. Consider any location functional L of the distribution of e_i and let $\mu = L(F)$ where F is the common distribution of the e_i's for $1 \leq i \leq n$. Then e_i^* has distribution

$F^*(x) = F(x - \mu)$ and $L(F^*) = 0$. We then have that $F(x) = F^*(x - \mu)$, so $L(F) = \mu$ is the location functional for x_i.

Furthermore, Y_i has a distribution $H(x) = F(x - (\mathbf{x}_i^T \boldsymbol{\beta} + \mu))$. Thus $L(H) = \mathbf{x}_i^T \boldsymbol{\beta} + \mu$ is a location functional for Y_i and consequently $\boldsymbol{\beta} = (\mathbf{x}_i \mathbf{x}_i^T)^{-1}(L(H) - L(F))$. $\square$

Definition 2.1.2. Let $\mathbf{Y} = (Y_1, \cdots, Y_n)^T$ denote the vector of observations, let $\mathbf{X}$ denote the $n \times p$ matrix whose ith row is $\mathbf{x}_i^T$ and let $e = (e_1, \cdots, e_n)$. The model in (2.1.2) can be expressed as

$$\mathbf{Y} = 1\mu + \mathbf{X}\boldsymbol{\beta} + e, \tag{2.1.3}$$

where 1 is a $n \times 1$ vector of ones, 1μ is a vector of reals representing the intercept and $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown regression parameters.

Note that the model

$$\mathbf{Y} = 1\mu + \boldsymbol{\eta} + e \tag{2.1.4}$$

is called **Coordinate Free Model**, where $\eta = \mathbf{X}\boldsymbol{\beta} \in \Omega_F$ with Ω_F being the column space spanned by the columns of $\mathbf{X}$.

In addition to estimating the parameter of interest $\boldsymbol{\beta}$, we will also be interested in general linear tests of the form

$$H_0 : \mathbf{M}\boldsymbol{\beta} = 0 \quad versus \quad H_A : \mathbf{M}\boldsymbol{\beta} \neq 0, \tag{2.1.5}$$

where $\mathbf{M}$ is $q \times p$ matrix of full row rank.

2.1.1 Estimation of regression parameters

Definition 2.1.3. An operator $||\cdot||_*$ is called a **pseudo-norm** if it satisfies the following conditions

1. $||x||_* \geq 0$, for all $x \in \mathbb{R}^n$

2. $||x||_* = 0$ if and only if $x_1 = \cdots = x_n$

3. $||\alpha x||_* = |\alpha|||x||_*$, for all $\alpha \in \mathbb{R}, x \in \mathbb{R}^n$

4. $||x + y||_* \leq ||x||_* + ||y||_*$, for all $x, y \in \mathbb{R}^n$.

Consider the function

$$||x||_* = \sum_{i=1}^{n} a(R(x_i))x_i, \tag{2.1.6}$$

where $x = (x_1, \cdots, x_n)$ is a vector in $\mathbb{R}^n$, the $a(i)$ are called **scores** and are such that $a(1) \leq \cdots \leq a(n)$, $\sum_{i=1}^{n} a(i) = 0$ and $a(i) = -a(n+1-i)$, $R(x_i)$ is the rank of x_i among $x_1, \cdots, x_n$.

Theorem 2.1.4. *The function* $|| \cdot ||_*$ *in* (2.1.6) *is a pseudo norm.*

Proof. 1. Positivity.

Using the connection between ranks and order statistics, we can write

$$||x||_* = \sum_{i=1}^{n} a(i)x_{(i)}.$$

Suppose that $x_{(i_0)}$ is the last order statistics with a negative score. Since $\sum_{i=1}^{n} a(i) = 0$, we have

$$\begin{aligned} ||x||_* &= \sum_{i=1}^{n} a(i)(x_{(i)} - x_{(i_0)}) \\ &= \sum_{i \leq i_0} a(i)(x_{(i)} - x_{(i_0)}) + \sum_{i \geq i_0} a(i)(x_{(i)} - x_{(i_0)}). \end{aligned}$$

Since both terms on the right are nonnegative, we have $||x||_* \geq 0$.

2. Furthermore, if $||x||_* = 0$, then both terms in the last equality must be zero. Since $a(1) < a(n)$ and $a(1) = -a(n)$, we must have $a(1) < 0$ and $a(n) > 0$. Therefore the first term on the right can be written as

$$\sum_{1 \leq i \leq i_0} a(i)(x_{(i)} - x_{(i_0)}).$$

Thus,

$$\sum_{1 \leq i \leq i_0} a(i)(x_{(i)} - x_{(i_0)}) = 0 \quad \text{implies } x_{(1)} = x_{(2)} = \cdots = x_{(i_0)}.$$

Likewise, we have $x_{(i_0)} = x_{(i_0+1)} = \cdots = x_{(n)}$.

This shows that $||x||_* = 0$ implies that $x_1 = \cdots = x_n$.

3. Homogeneity.

For some positive real α, we know that $R(\alpha x_{(i)}) = R(x_{(i)})$ and $R(-x_{(i)}) = R(x_{(n+1-i)})$. Clearly, for a positive real α, one has:

$$\begin{aligned} ||\alpha x||_* &= \sum_{i=1}^{n} a(R(\alpha x_i)(\alpha x_i) \\ &= \alpha \sum_{i=1}^{n} a(R(x_i))x_i \\ &= |\alpha| ||x||_*. \end{aligned}$$

If $\alpha < 0$, then

$$\begin{aligned}
||\alpha x||_* &= \sum_{i=1}^{n} a(R(\alpha x_i))(\alpha x_i) \\
&= \sum_{i=1}^{n} a(R(-(-\alpha x_i))(-(-\alpha x_i)) \\
&= -\alpha \sum_{i=1}^{n} a(R(x_{n+1-i}))(-x_i) \\
&= |\alpha| \sum_{i=1}^{n} -a(n+1-i)(x_{(i)}) \\
&= |\alpha| \sum_{i=1}^{n} a(i)x_{(i)} \\
&= |\alpha|||x||_*.
\end{aligned}$$

4. Triangle inequality.

$$\begin{aligned}
||x+y||_* &= \sum_{i=1}^{n} a(R(x_i+y_i))(x_i+y_i) \\
&= \sum_{i=1}^{n} a(R(x_i+y_i))x_i + \sum_{i=1}^{n} a(R(x_i+y_i))y_i \\
&\leq \sum_{i=1}^{n} a(i)x_{(i)} + \sum_{i=1}^{n} a(i)y_{(i)} \quad \text{by Hardy's Tauberian Theorem} \\
&= ||x||_* + ||y||_*.
\end{aligned}$$

□

We now suppose that the scores are generated as $a(i) = h(i/(n+1))$ for some nondecreasing function h defined on the interval $(0,1)$ and such that

$$\int_0^1 h(u)du = 0, \int_0^1 h^2(u)du < \infty.$$

Consider the model in (2.1.4). Rewriting the pseudo-norm above as $||x||_h = \sum_{i=1}^{n} a(R(x_i))x_i$, a **Rank-estimate** for $\boldsymbol{\eta}$ is a vector $\widehat{\mathbf{Y}}_h$ such that

$$D_h(\boldsymbol{\beta}) = \text{Distance}(\mathbf{Y}, \Omega_F) = ||\mathbf{Y} - \widehat{\mathbf{Y}}_h||_h = \min_{\boldsymbol{\eta} \in \Omega_F} ||\mathbf{Y} - \boldsymbol{\eta}||_h.$$

The function $D_h(\boldsymbol{\beta})$ is also called the **dispersion function**.

Once $\boldsymbol{\eta}$ has been estimated, $\boldsymbol{\beta}$ can be estimated by solving the equation $\mathbf{X}\boldsymbol{\beta} = \widehat{\mathbf{Y}}_h$, hence the Rank-estimate of $\boldsymbol{\beta}$ is $\widehat{\boldsymbol{\beta}}_R = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\widehat{\mathbf{Y}}_h$. The intercept μ can be estimated by a location estimate based on the residuals $\widehat{e} = \mathbf{Y} - \widehat{\mathbf{Y}}_h$. We could use the median of residuals denoted by $\widehat{\mu}_S = \underset{i=1,\cdots,n}{\text{med}} \{\mathbf{Y}_i - \mathbf{x}_i\widehat{\boldsymbol{\beta}}_R\}$. Geometrically, the Rank-estimate of $\boldsymbol{\eta}$ is a vector that minimizes the normed difference between $\mathbf{Y}$ and Ω_F as shown in Figure 2.1.

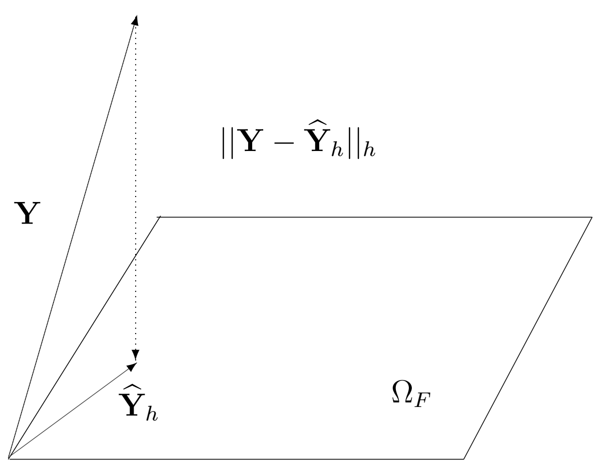

Figure 2.1: Geometry of Rank Estimation

The following result justifies the existence of the rank estimation and is due to Jaekel Jaeckel (1972).

Remark 2.1.5. The dispersion function $D_h(\boldsymbol{\beta})$ is a continuous, convex, almost everywhere differentiable function.

Proof. Continuity follows from the the inequality $|D_h(\boldsymbol{\beta}_1) - D_h(\boldsymbol{\beta}_1)| \leq ||\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2||_p ||x||_h$, for all $x \in \mathbb{R}^n$, for all $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2 \in \mathbb{R}^p$.

Convexity follows from the equalities and inequality

$$\begin{aligned} D_h(\alpha\boldsymbol{\beta}_1 + (1-\alpha)\boldsymbol{\beta}_2) &= ||\mathbf{Y} - X(\alpha\boldsymbol{\beta}_1 + (a - \alpha\boldsymbol{\beta}_2))||_h \\ &= ||\alpha\mathbf{Y} + (1-\alpha)\mathbf{Y} + \alpha\boldsymbol{\beta}_1 + (1-\alpha)\boldsymbol{\beta}_2|| \\ &\leq \alpha D_h(\boldsymbol{\beta}_1) + (1-\alpha)D_h(\boldsymbol{\beta}_2), \quad \text{for some } \alpha \in (0,1). \end{aligned}$$

Differentiability follows from the equality $\nabla D_h(\boldsymbol{\beta}) = -\sum_{i=1}^{n} x_i a(R(Y_i - x_i^T\boldsymbol{\beta}))$. □

Remark 2.1.6. The Rank-estimate $\widehat{\boldsymbol{\beta}}_R$ of $\boldsymbol{\beta}$ is location and scale **equivariant**, that is $\widehat{\boldsymbol{\beta}}_R(k\mathbf{Y}) = k\widehat{\boldsymbol{\beta}}_R(\mathbf{Y})$ and $\widehat{\boldsymbol{\beta}}_R(\mathbf{Y} + \mathbf{X}\delta) = \widehat{\boldsymbol{\beta}}_R(\mathbf{Y}) + \delta$, for $k \in \mathbb{R}$ and for $\delta \in \mathbb{R}^p$.

Proof. $D_h(\widehat{\boldsymbol{\beta}}_R(\mathbf{Y})) = \text{Dist}(\mathbf{Y}, \Omega_F) = ||\mathbf{Y} - \widehat{\mathbf{Y}}_h||_h$. Therefore,

$$\begin{aligned} D_h(\widehat{\boldsymbol{\beta}}_R(\mathbf{Y} + \mathbf{X}\delta)) &= \text{Dist}(\mathbf{Y} + \mathbf{X}\delta, \Omega_F) \\ &= ||\mathbf{Y} + \mathbf{X}\delta - \widehat{\mathbf{Z}}_h||_h \\ &= ||\mathbf{Y} - (\widehat{\mathbf{Z}}_h - \mathbf{X}\delta)||_h \\ &= \text{Dist}(\mathbf{Y}, \Omega_F). \end{aligned}$$

So we must have $\widehat{\mathbf{Z}}_h = \widehat{\mathbf{Y}}_h + \mathbf{X}\delta$. Thus,

$$\begin{aligned} \widehat{\boldsymbol{\beta}}_R(\mathbf{Y} + \mathbf{X}\delta) &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T[\widehat{\mathbf{Y}}_h + \mathbf{X}\delta] \\ &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\widehat{\mathbf{Y}}_h + \delta \\ &= \widehat{\boldsymbol{\beta}}_R(\mathbf{Y}) + \delta. \end{aligned}$$

In the same way, for any $k \neq 0$,

$$\begin{aligned} D_h(\widehat{\boldsymbol{\beta}}_R(k\mathbf{Y})) &= \text{Dist}(k\mathbf{Y}, \Omega_F) \\ &= ||k\mathbf{Y} - \widehat{\mathbf{W}}_h||_h \\ &= |k|||\mathbf{Y} - \frac{1}{k}\widehat{\mathbf{W}}_h||_h \\ &= |k|\text{Dist}(Y, \Omega_F) = |k|D_h(\widehat{\boldsymbol{\beta}}_R(\mathbf{Y}). \end{aligned}$$

So we must have $\frac{1}{k}\widehat{\mathbf{W}}_h = \widehat{\mathbf{Y}}_h$.

Thus,

$$\begin{aligned} \widehat{\boldsymbol{\beta}}_R(k\mathbf{Y}) &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\widehat{\mathbf{W}}_h \\ &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(k\widehat{\mathbf{Y}}_h) \\ &= k(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\widehat{\mathbf{Y}}_h) \\ &= k\widehat{\boldsymbol{\beta}}_R(\mathbf{Y}). \end{aligned}$$

□

As a consequence, without loss of generality, the theory we will develop will be established under the assumption that the true $\boldsymbol{\beta}$ is 0 for simplicity.

We end this section with the following theorem proved in the appendix, on the asymptotic properties of the Rank-estimator $\widehat{\boldsymbol{\beta}}_R$ of $\boldsymbol{\beta}$.

Theorem 2.1.7.

$$\begin{pmatrix} \widehat{\mu}_S \\ \widehat{\boldsymbol{\beta}}_R \end{pmatrix} \approx N_{p+1}\left(\begin{pmatrix} \mu \\ \boldsymbol{\beta} \end{pmatrix}, \begin{bmatrix} n^{-1}\tau_S^2 & 0 \\ 0 & \tau_h^2(\mathbf{X}^T\mathbf{X})^{-1} \end{bmatrix} \right)$$

where τ_h and τ_S will be defined later.

2.1.2 Tests of linear hypothesis

Let's consider the model (2.1.2). Note that Distance$(\mathbf{Y}, \Omega_F)$ is the amount of residual dispersion not accounted for in the model (2.1.2). Let Ω_F^R be the subspace subject to H_0, that is, $\Omega_F^R = \{\boldsymbol{\eta} \in \Omega_F : \boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}, \quad \text{for some } \boldsymbol{\beta} \text{ such that } \mathbf{M}\boldsymbol{\beta} = 0\}$.

Clearly, Ω_F^R is a subspace of Ω_F and since $\mathbf{MX} = 0$ is a system of q equations with p unknowns, then $\text{Dim}(\Omega_F^R) = p - q$.

If $\widehat{\mathbf{Y}}_h^R$ is the Rank-estimate when the reduced model is fitted, then the nonnegative quantity

$$RD_h = \text{Dist}(Y, \Omega_F^R) - \text{Dist}(\mathbf{Y}, \Omega_F) \qquad (2.1.7)$$

represents the **Reduction in residual dispersion** when we pass from the reduced model to the full model as shown in Figure 2.2.

Thus, large values of RD_h indicate H_A while small values support H_0. If RD_h is standardized, the ensuing asymptotic distribution theory suggests that $F_h = \dfrac{RD_h/q}{\widehat{\tau}_h/2}$ should be compared with the F-critical values with q and $n-(p+1)$ degrees of freedom, at least for small sample studies, where $\widehat{\tau}_h$ is a consistent estimator of τ_h.

The rank analogue of Wald's test for the full model is given by

$$W_h = \frac{(\mathbf{M}\widehat{\boldsymbol{\beta}}_R)[M(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{M}^T]^{-1}(\mathbf{M}\widehat{\boldsymbol{\beta}}_R)/q}{\widehat{\tau}_h^2}. \qquad (2.1.8)$$

It can be proved (Hettmansperger and McKean, 1983) that W_h has an asymptotic $\chi^2(q)$ distribution.

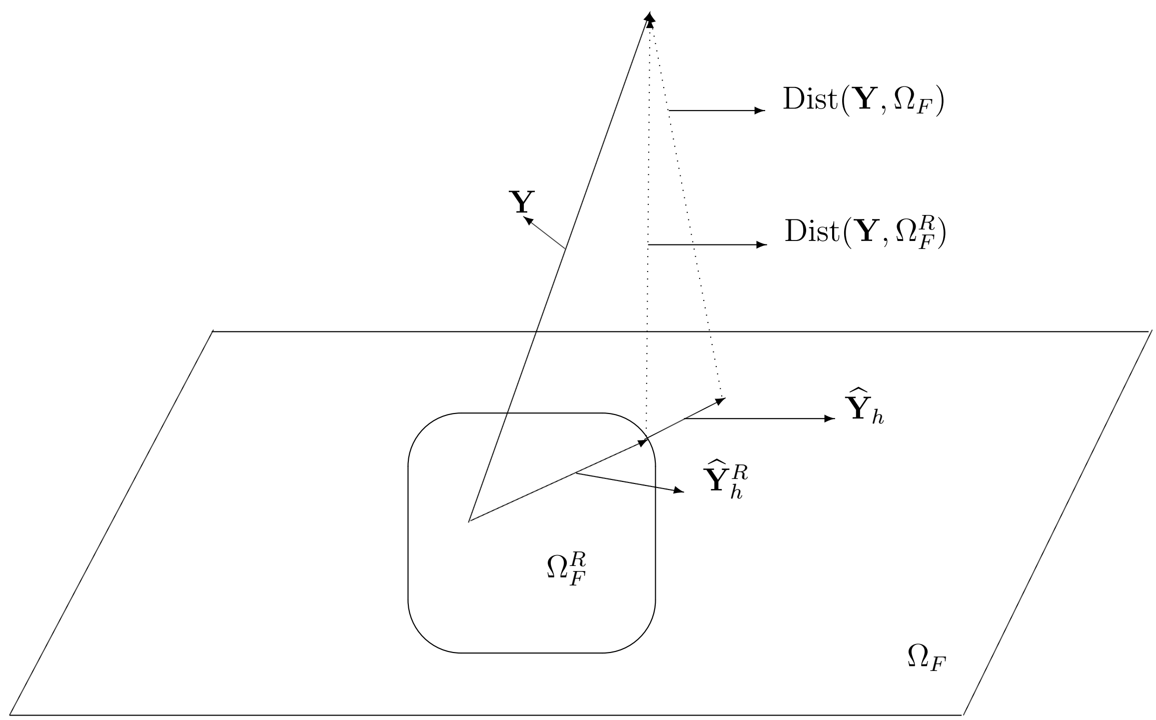

Figure 2.2: Geometry of Rank Tests

2.2 Basic Wavelet Theory

In this section, we introduce the basics of wavelet and provide the motivation for their invention.

2.2.1 A motivating example

The Fourier series of a square integrable function f can be obtained by dilating the orthonormal basis $\{e^{-ikx}\}_{k\in\mathbb{Z}}$. However, each element of the basis $\{e^{-ik(\frac{\pi}{\sigma})x}\}_{k\in\mathbb{Z}}$ obtained by dilation is a complex sinusoidal wave which is global in x, hence the Fourier coefficients do not provide information on the local behavior of the function f. For instance, consider the 2π-periodic function

$$f(x) = \begin{cases} \pi/2, & x \in (0, \pi) \\ 0, & x = 0 \\ -\pi/2, & x \in (-\pi, 0). \end{cases}$$

After computation of the Fourier coefficients, we have

$$f(x) = 2\sum_{n=1}^{\infty} \frac{\sin(2n-1)x}{2n-1}, \quad x \in (-\pi, \pi).$$

Clearly for any finite interval (a, b), the behavior of the function

$$f_{loc}(x) = \begin{cases} f(x), & x \in (a, b) \\ 0, & \text{otherwise} \end{cases}$$

cannot be directly obtained via the Fourier coefficients since f is local in the areas $(-\pi, 0)$ and $(0, \pi)$.

Another shortcoming of the Fourier theory is the **Gibbs' Phenomenon**. Indeed, consider the previous function f. Since $2\sum_{n=1}^{\infty} \frac{\sin(2n-1)x}{2n-1}\Big|_{x=0} = 0$, the Fourier series of f also converges to $f(x)$ at $x = 0$.

Let $S_n(f, x) = 2\sum_{k=1}^{n} \frac{\sin(2k-1)x}{2k-1}$. The jump of amplitude of $f(x)$ at $x = 0$ is

$$f(0^+) - f(0^-) = \pi.$$

Moreover,

$$\lim_{n\to\infty} S_n\Big(f, \frac{\pi}{2n}\Big) = \int_0^{\pi} \frac{\sin t}{t} dt \approx 1.85193706$$

and

$$\lim_{n\to\infty} S_n\Big(f, -\frac{\pi}{2n}\Big) = -\int_0^\pi \frac{\sin t}{t} dt \approx -1.85193706.$$

Thus,

$$\begin{aligned} \lim_{n\to\infty} \Big| S_n\Big(f, \frac{\pi}{2n}\Big) - S_n\Big(f, -\frac{\pi}{2n}\Big)\Big| &\approx 3.70387412 \\ &\approx 1.179\pi. \end{aligned}$$

The latter means that the amplitude of $S_n(f, x)$ around 0 is at least 1.179 multiple of the jump of f at 0. This is the Gibbs' phenomenon as seen in Figure 2.3.

Finally, another shortcoming of the Fourier analysis that is worth mentioning is the problem of convergence. Indeed, in 1873, Paul Du Bois-Reymond constructed a 2π-periodic function whose Fourier series diverges at a given point.

Therefore, there is a natural need for an orthogonal system for which the local behavior of a function can be recognized from its coefficients, for which the Gibbs' Phenomenon can be avoided or at least dealt with and for which the phenomenon discovered by Du Bois-Reymond cannot happen. Wavelets provide an answer to those concerns.

Definition 2.2.1. A function ψ is called a **wavelet function** if $\{2^{j/2}\psi(2^j x - k)\}_{j,k\in\mathbb{Z}}$ is an orthonormal basis of $L^2(\mathbb{R})$.

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

$$f(x) = \sum_{j\in\mathbb{Z}} \sum_{k\in\mathbb{Z}} C_{jk}\psi_{jk}(x). \tag{2.2.1}$$

Note that (2.2.1) is called the **homogeneous wavelet expansion** of the function f. This is to say that there exists an **inhomogeneous wavelet expansion** for f defined

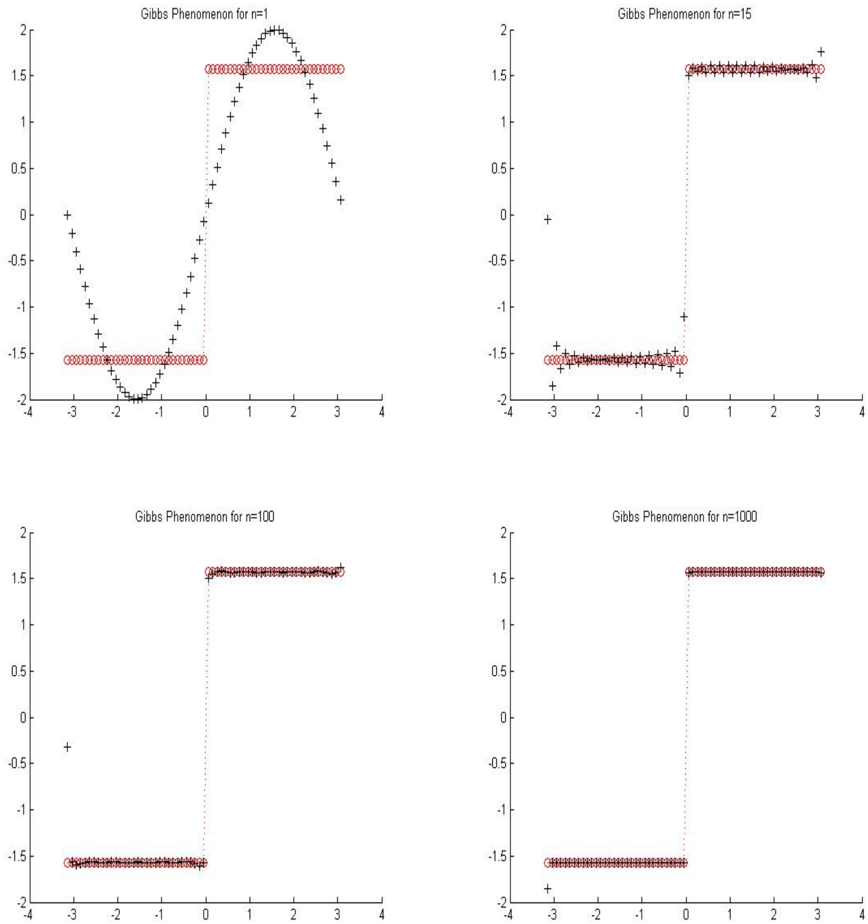

Figure 2.3: Gibbs Phenomenon

as

$$f(x) = \sum_{k\in\mathbb{Z}} C_{0k}\phi_{0k}(x) + \sum_{j=0}^{\infty}\sum_{k\in\mathbb{Z}} C_{jk}\psi_{jk}(x), \tag{2.2.2}$$

where ϕ and ψ are respectively called "**father wavelet or scaling function**" and "**mother wavelet**". Besides, the "mother wavelet" can be obtained from the "father wavelet" through the relation $\psi(x) = \sqrt{2}\sum_{k\in\mathbb{Z}}\lambda_k\phi(2x-k)$, where the λ_k's are some carefully chosen coefficients. There are many examples of scaling functions.

2.2.2 Examples of wavelets

1. **Haar Wavelet**

   It's the very first wavelet constructed. Its "father wavelet" and "mother wavelet" are respectively defined as

$$\phi(x) = \begin{cases} 1, & x \in [0,1) \\ 0, & \text{otherwise} \end{cases} \qquad \psi(x) = \begin{cases} 1, & x \in [0,1/2) \\ -1, & x \in [1/2,1) \\ 0, & \text{otherwise.} \end{cases}$$

   The functions $\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k), \quad j,k \in \mathbb{Z}$ are called "daughters wavelets" and are compactly locally supported in diadic interval $I_n = [k2^{-j}, (k+1)2^{-j}]$. The left and right panels in Figure 2.4 respectively depict ϕ and ψ for the Haar wavelet system.

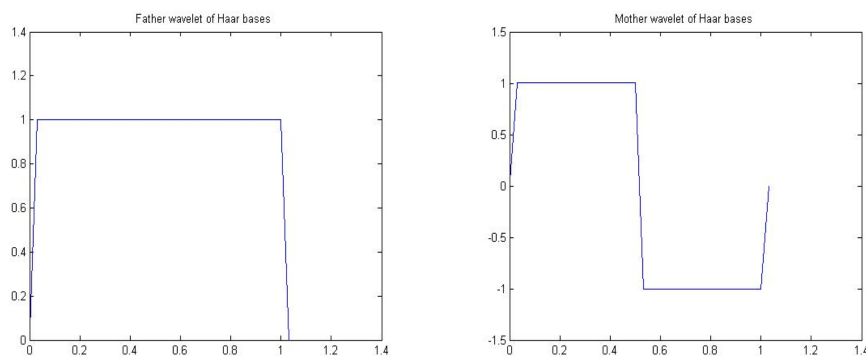

Figure 2.4: The Haar Wavelet

2. **Shannon wavelet**

   Consider the space $V_0^{Sh} = \{f \in L^2(\mathbb{R}) : \text{Support}(\mathcal{F}f)(\xi) \subset [-\pi, \pi]\}$.

   Then,

$$\forall f \in V_0^{Sh}, f(x) = \sum_{k\in\mathbb{Z}} f(k)\frac{\sin \pi(x-k)}{\pi(x-k)}. \tag{2.2.3}$$

This result is also known as the **Sampling Theorem** (see Hong et al. (2005)) since the function f can be recovered from its sample values $f(k)$. It follows that the Shannon "father wavelet" and "mother wavelet" are respectively given by

$$\psi(x) = \frac{\sin \pi x}{\pi x}, \quad \psi(x) = \frac{\sin \pi(x-1/2) - \sin 2\pi(x-1/2)}{\pi(x-1/2)}.$$

These are shown in Figure 2.5.

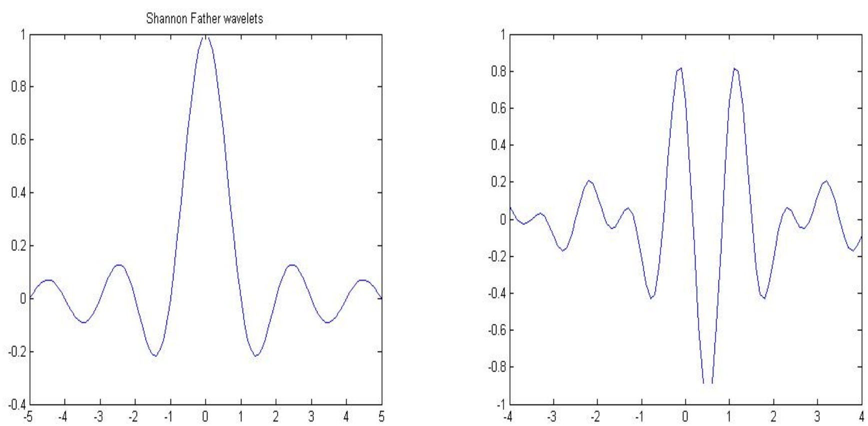

Figure 2.5: The Shannon Wavelet

3. **Mexican Hat Wavelet**

This is by far the most used wavelet in practice for its simplicity. Its name comes from the resemblance of the graph of its "mother wavelet" to a Mexican hat. The "father wavelet" and "mother wavelet" are given respectively by:

$$\phi(x) = e^{-x^2}, \quad \psi(x) = \frac{2}{\sqrt{3}}\pi^{-1/4}(1-x^2)e^{-x^2}.$$

These are shown in Figure 2.6.

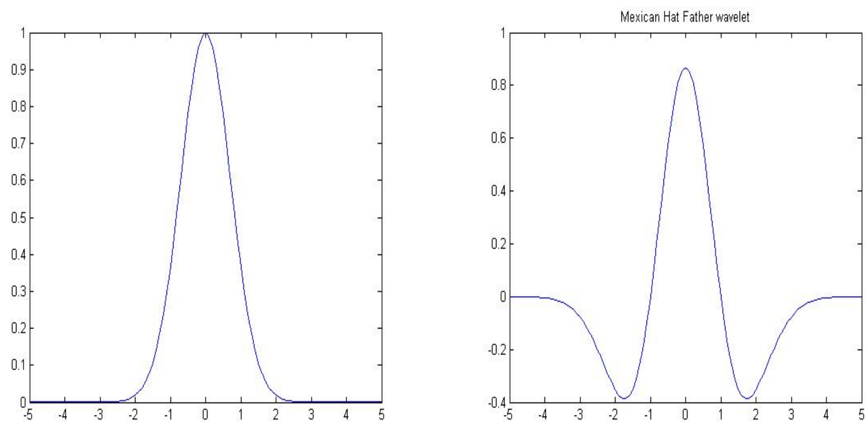

Figure 2.6: The Mexican Hat Wavelet

4. **Daubechies Wavelet**

Ingrid Daubechies Daubechies. (1992) was the first to introduce continuous compactly supported wavelets. Indeed, she proved that there exists a scaling function $\phi \in L^2(\mathbb{R})$ such that

$$\phi(x) = \sqrt{2}\sum_{k\in\mathbb{Z}} h_k \phi(2x-k),$$

where $\{h_k\}_{k\in\mathbb{Z}}$ is a sequence of real numbers such that

$$h_k = \sqrt{2}\int \phi(x)\overline{\phi(2x-k)}dx, \quad \sum_{k\in\mathbb{Z}} |h_k|^2 < \infty,$$

where $\overline{\phi}$ is the complex conjugate of ϕ. Daubechies wavelets are classified as DN for $N = 2, \cdots, 20$ where N is even and represents the number of coefficients and

$N/2$ the number of vanishing moments. Vanishing moments denote the ability of the wavelets to encode a polynomial function or a signal. For example, $D2$ has one vanishing moment so it's good for encoding constant signals. Note that $D2$ coincides with the Haar wavelet. It's actually the only explicit Daubechies wavelet since the others do not have scaling functions that can be expressed in a closed form. Daubechies wavelets can also be defined on any interval $[a, b]$ and more information can be found in Andersson et al. (1994) and Cohen et al. (1993).

These are shown in Figure 2.7.

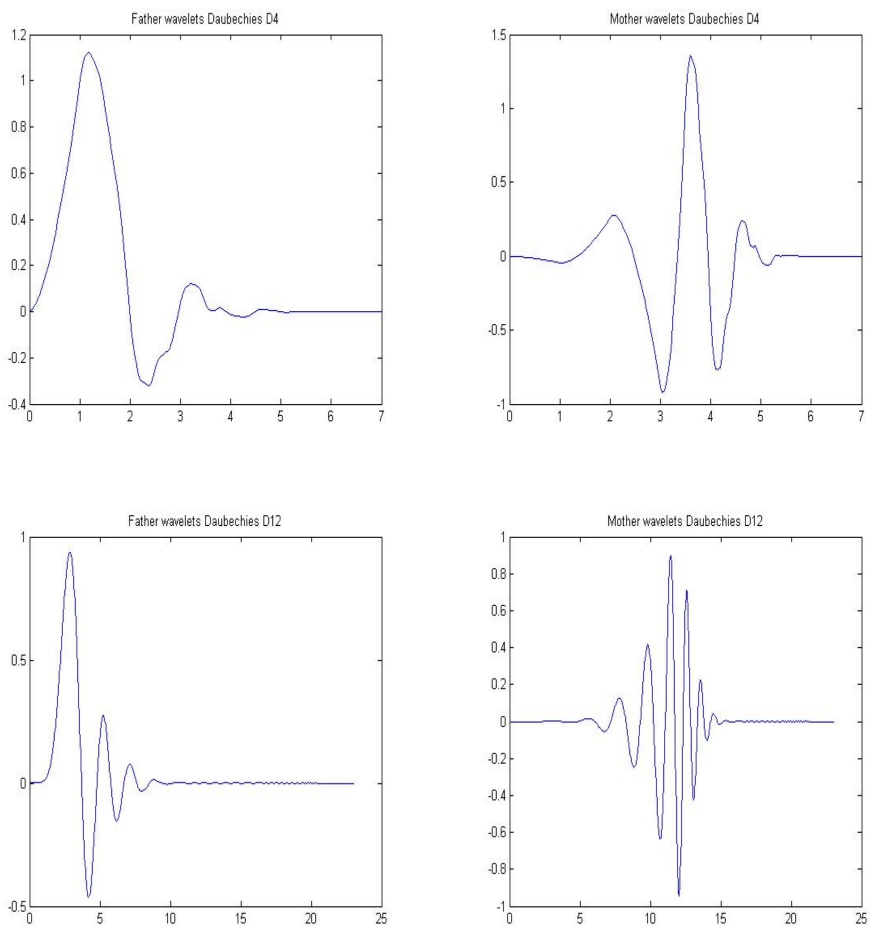

Figure 2.7: The Daubechies Wavelet

We end this section by noting that one key difference between wavelet approximation and Fourier approximation is that the Fourier approximation uses one function (called **window function**) that is translated over the interval of definition whereas its wavelet counterpart uses a function that is translated and dilated to adapt itself to the local properties of the function.

Chapter 3

Estimation of the Score Function

3.1 Introduction

In this chapter, we introduce the problem to be solved and put it into its historical context.

Consider the linear model

$$Y_i = \mu + \mathbf{x}_i^T \boldsymbol{\beta} + e_i^* \quad 1 \leq i \leq n, \tag{3.1.1}$$

where $e_1, \cdots, e_n$ are independent random variables with distribution function F and density f.

It's known that the least squares estimate $\widehat{\boldsymbol{\beta}}_{LS}$ of $\boldsymbol{\beta}$ is the one that minimizes the square distance $||\mathbf{Y} - \mathbf{x}\boldsymbol{\beta}||^2_{\mathbb{R}^n}$ but fails to be robust, in the sense that it's very sensitive to departure from normality and to perturbations. Robust alternatives to the least squares method include the **M-estimation** and **Z-estimation**.

M-estimation consists of finding an estimate $\widehat{\boldsymbol{\beta}}_M$ of $\boldsymbol{\beta}$ that **maximizes** (hence the term M-estimation) a criterion function $M_n(\boldsymbol{\beta}) = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} m_{\boldsymbol{\beta}}(X_i)$ where $m_{\boldsymbol{\beta}} : \mathcal{X} \mapsto \mathbb{R}$ are known functions. Z-estimation consists of finding an estimate $\widehat{\boldsymbol{\beta}}_Z$ of $\boldsymbol{\beta}$ that almost maximizes the criterion function $M_n(\boldsymbol{\beta})$ or is one of its near **Zeros** (hence the term Z-estimation). Note that popular choices for the criterion function include the so called **Huber function** and the **biweight function** given respectively by

$$M(u) = \begin{cases} \frac{1}{2}u^2, & |u| \le k \\ k|u| - \frac{1}{2}k^2, & |u| > k \end{cases}, \quad M(u) = \begin{cases} \frac{1}{6}\left[1 - \left\{1 - (u/k)^2\right\}^3\right], & |u| \le k \\ \frac{1}{6}k^2, & |u| > k, \quad k \in \mathbb{R}. \end{cases}$$

Although the aforementioned methods solve the problem of robustness, they have their own shortcomings. First, it may be hard to find zeros of $(M_n(\beta))'$. Second, the existence of zeros near the boundary of the parameter set may make the estimation problem become ill-posed. Third, consistency and uniqueness are not guaranteed.

Consider the estimator $\widehat{\boldsymbol{\beta}}_R$ of $\boldsymbol{\beta} \in \mathbb{R}^p$ that minimizes the dispersion function

$$D_h(\boldsymbol{\beta}) = \sum_{i=1}^{n} a(R(Y_i - \mathbf{x}_i^T\boldsymbol{\beta}))(Y_i - \mathbf{x}_i^T\boldsymbol{\beta}), \tag{3.1.2}$$

where $R(y_i - \mathbf{x}_i^T\boldsymbol{\beta})$ is the rank of $y_i - \mathbf{x}_i^T\boldsymbol{\beta}$ among $y_1 - \mathbf{x}_1^T\boldsymbol{\beta}, \cdots, y_n - \mathbf{x}_n^T\boldsymbol{\beta}$ and $a(1) \le \cdots \le a(n)$ are some scores. The scores are usually chosen as $a(j) = h(j/(n+1))$, where $h : (0,1) \to \mathbb{R}^+$ is a nondecreasing score function.

Definition 3.1.1. The **score function** is the gradient with respect to some parameter θ of the logarithm of the likelihood function, that is $h(u) = \dfrac{\partial}{\partial\theta} \log f(\theta; u)$.

Examples of score functions

1. Sign score function: $h(u) = sign(u - 1/2)$.

2. Logistic score function: $h(u) = 2u - 1$.

3. Normal score function: $h(u) = \Phi^{-1}(u)$ where Φ is the standard normal distribution.

Remark 3.1.2. The mean of the score function given a parameter θ is zero, that is, $E(h|\theta) = 0$. This entails that the variance of the score function $I(f)$, which is called the **Fisher Information**, is given by $I(f) = \int_{-\infty}^{\infty} \left(\frac{f'(u)}{f(u)}\right)^2 f(u)du$.

Proof.

$$\begin{aligned}
E(h|\theta) &= \int_{-\infty}^{\infty} \frac{\partial \ln f(u;\theta)}{\partial\theta} f(u;\theta)du \\
&= \int_{-\infty}^{\infty} \frac{\partial f(u;\theta)}{\partial\theta} du \\
&= \frac{\partial}{\partial\theta} \int_{-\infty}^{\infty} f(u;\theta)du \\
&= \frac{\partial 1}{\partial\theta} = 0.
\end{aligned}$$

□

Definition 3.1.3. Given a score function h, we define the **scale parameter** as $\tau_h = 1/\gamma$ where

$$\gamma = \int_0^1 h(u)h_F(u)du \quad \text{and} \quad h_F(u) = -\frac{f'(F^{-1}(u))}{f(F^{-1}(u))}.$$

Suppose that an estimate T_{n_ν} of $T(\boldsymbol{\beta})$ based on n observations is such that as $\nu \to \infty$,

$$\sqrt{n_\nu}(T_{n_\nu} - T(\boldsymbol{\beta})) \rightsquigarrow N(0, \sigma^2(\boldsymbol{\beta})). \tag{3.1.3}$$

Definition 3.1.4. Given two estimators $T_{n_{\nu,1}}$ and $T_{n_{\nu,2}}$ of $T(\boldsymbol{\beta})$, let $n_{\nu,1}, n_{\nu,2}$ be the number of observations needed to meet (3.1.3). Then the **(Pitman) Asymptotic Relative Efficiency(ARE)** of the two estimators is defined as

$$\text{ARE}(T_{n_\nu,2}, T_{n_{\nu,1}}) = \lim_{\nu\to\infty} \frac{n_{\nu,1}}{n_{\nu,2}} = \frac{\sigma_1^2(\boldsymbol{\beta})}{\sigma_2^2(\boldsymbol{\beta})}. \tag{3.1.4}$$

It follows from the above definition that an estimator $T_{n_\nu,1}$ of $T(\boldsymbol{\beta})$ is said to be as asymptotically relatively efficient as an estimator $T_{n_\nu,2}$ of $T(\boldsymbol{\beta})$ if $\text{ARE}(T_{n_\nu,2}, T_{n_\nu,1}) \to 1$.

Theorem 3.1 (Asymptotic Relative Efficiency)**.** If the score function h is such that $h(u) = h_F(u)$, then the resulting R-estimate $\widehat{\boldsymbol{\beta}}_R$ of $\boldsymbol{\beta}$ in the linear model (3.1.1) is asymptotically relatively as efficient as the least squares estimate $\widehat{\boldsymbol{\beta}}_{LS}$.

Proof. Since $\text{var}(\widehat{\boldsymbol{\beta}}_{LS}) = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}$ where σ^2 is the variance of the underlying normal distribution and $\text{var}(\widehat{\boldsymbol{\beta}}_R) = \tau_h^2(\mathbf{X}^T\mathbf{X})^{-1}$ by theorem 2.1.7, we have

$$\begin{aligned} \text{ARE}(\widehat{\boldsymbol{\beta}}_R, \widehat{\boldsymbol{\beta}}_{LS}) = \frac{\sigma^2}{\tau_h^2} &= \sigma^2\left(\int_0^1 h(u)h_F(u))du\right)^2 \\ &= \sigma^2\sqrt{I(f)}\sqrt{\text{Var}(h_F)}\big[\text{corr}(h(u), h_F(u))\big]^2 \\ &= \sigma^2 I(f)\big[\text{corr}(h(u), h_F(u))\big]^2 \end{aligned}$$

So optimality is obtained when $\text{corr}(h(u), h_F(u)) = 1$ that is, $h(u) = h_F(u)$. □

In the sequel, we assume the scores are chosen so that $a(j) = h_F(j/(n+1))$, where

$$h_F(u) = -\frac{f'(F^{-1}(u))}{f(F^{-1}(u))}. \tag{3.1.5}$$

In this case, the estimator $\widehat{\boldsymbol{\beta}}_R$ of $\boldsymbol{\beta}$ is **asymptotically efficient**, that is, it is as efficient as the least squares estimator $\widehat{\boldsymbol{\beta}}_{LS}$. The choice of the least squares estimator for comparison is that by Gauss-Markov theorem, it achieves the uniform minimum variance among all linear unbiased estimators.

Lemma 3.1.5. *If the Fisher information $I(f)$ is finite, then $h_F \in L^2(0,1)$.*

Proof.

$$\begin{aligned} I(f) &= \int_{-\infty}^{\infty} \left[\frac{f'(x)}{f(x)}\right]^2 f(x)dx \\ &= \int_0^1 \left[\frac{f'(F^{-1})(u)}{f(F^{-1})(u)}\right]^2 du \quad \text{by change of variable} \quad u = F(x) \\ &= \int_0^1 h_F^2(u)du. \end{aligned}$$

Hence, $I(f) < \infty$ implies $h_F \in L^2(0,1)$. □

Therefore, under $I(f) < \infty$, there exist coefficients C_{jk}, such that

$$h_F(t) = \sum_{j=-\infty}^{\infty} \sum_{k \in \mathbb{Z}} C_{jk} \psi_{jk}(t), \tag{3.1.6}$$

where $\{\psi_{jk}\}_{j,k\in\mathbb{Z},}$ is an orthonormal system in $L^2(0,1)$ with $\psi_{jk}(t) = 2^{j/2}\psi(2^{j/2}t - k)$ for some function ψ and

$$C_{jk} = \int_0^1 h_F(s)\psi_{jk}(s)ds\ . \tag{3.1.7}$$

An asymptotically efficient estimate of the coefficients C_{jk} will yield an asymptotically efficient estimate of h_F.

A common approach in rank regression is to fix the score function apriori on the basis of robustness or simplicity considerations. However, for efficient results, a good approximation of h_F based on an approximate knowledge of F from a sample is of some value. To that end, Van Eeden (1970) proposed an asymptotically efficient estimate of location parameters using an estimate of h_F based on a subset of the data. Dionne (1981) used a similar subset-based technique to develop estimators of linear model parameters. Beran (1974), for the location model, and Naranjo and McKean (1997), for the linear model, provided Fourier series estimators of h_F based on the whole sample. A different approach to aforementioned methods that uses density estimation and based on quantile regression was proposed by Koenker and Basset (1978).

The estimator proposed in this dissertation also uses the whole sample to estimate h_F. Our approach differs from that of Naranjo and McKean (1997) in that:

1. we develop estimates of h_F that provide asymptotically efficient estimators $\widehat{\boldsymbol{\beta}}_R$ based on a large class of orthonormal basis in $L^2(0,1)$,

2. we develop estimates based on second order approximations (Beran (1974) and Naranjo and McKean (1997) used first order approximations),

3. we eliminate restrictive assumptions on the data such as those in assumption $(A6)$ of Naranjo and McKean (1997) by using second order approximations, and

4. we provide a consistent, wavelet-based, estimator of the asymptotic variance of the estimator $\widehat{\boldsymbol{\beta}}_R$.

Zygmund (1945) pointed out that second differences of functions are much more useful than first differences in estimating smooth functions. This motivates our use of second order approximations. Also, the use of the second derivative gives us expressions of coefficients that are easier to manipulate than the ones in Naranjo and McKean (1997). This allows us to avoid making restrictive distributional assumptions such as assumption $(A6)$ of Naranjo and McKean (1997) that asserts that the first derivative of $(\phi(F))'F^{-1}$ be bounded, where $\phi(t) = \exp(-2\pi ikt)$ and k is an integer. This excludes a wide range of distributions such as the normal and the logistic.

In the following, we provide an asymptotically efficient estimate of the score function. We begin by laying out the assumptions and discussing their consequences .

3.2 General Assumptions

We will assume without loss of generality that $\mu = 0$, $\boldsymbol{\beta} = \mathbf{0}$ in view of remark 2.1.6, and that the $\mathbf{x}_i$'s are centered to have mean $\mathbf{0}$ in (3.1.1). We assume the following conditions:

(H_1) ψ has compact support in $(0, 1)$ and is three times differentiable with bounded derivatives.

(H_2) $\frac{1}{\sqrt{n}} \max_{1 \le i \le n} \|\mathbf{x}_i\|_p = o(1)$.

(H_3) $\frac{1}{n} \sum_{i=1}^{n} \|\mathbf{x}_i\|_p^2 = O(1)$.

(H_4) f is absolutely continuous with $I(f) < \infty$ and $\frac{f'}{f}$ monotone.

(H_5) There exists a sequence $\{\widehat{\boldsymbol{\beta}}_n\}$ in $\mathbb{R}^p$ such that $\sqrt{n}\widehat{\boldsymbol{\beta}}_n = O_p(1)$.

(H_6) $\lim_{n \to \infty} n^{-1} \mathbf{X}^T \mathbf{X} = \Sigma$.

3.2.1 Discussion of the assumptions

(H_1) assumes that ψ is the mother wavelet of a wavelet system with compact support. There exist many such systems of wavelets satisfying (H_1) among which the Daubechies wavelets, Coiflets and Symlets . This assumption implies that the "daughter" wavelets ψ_{jk} satisfy

$$\psi_{jk}^{(l)} = O(2^{2j}) \quad \forall k \in \mathbb{N} \text{ and } l = 0, 1, 2, 3. \tag{3.2.1}$$

(H_2) and (H_3) guarantee that we can apply the **Lindeberg-Feller Central Limit Theorem**. To see how that's possible, let's recall the Lindeberg-Feller Central Limit theorem.

Theorem 3.2.1 (Lindeberg-Feller Central Limit Theorem)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\mathbf{x}_i : \Omega \to \mathbb{R}, \quad i \in \mathbb{N}$ be independent random variables defined on that probability space. Assume that $E(\mathbf{x}_i) = \mu_i$ and $Var(\mathbf{x}_i)$ exist and are finite. Let $S_n^2 = \sum_{i=1}^{n} Var(\mathbf{x}_i)$. If*

$$\lim_{n\to\infty} \frac{1}{S_n^2} \sum_{i=1}^{n} \int_{\{|\mathbf{x}_i-\mu_i|>\epsilon S_n\}} (\mathbf{x}_i - \mu_i)^2 d\mathbb{P} = 0 \quad \forall \epsilon > 0,$$

then $Z_n = \dfrac{\sum_{i=1}^{n}(\mathbf{x}_i - \mu_i)}{S_n}$ converges in distribution as $n \to \infty$ to the standard normal distribution.

Indeed for any $\epsilon > 0$, we have

$$\int_{\{|\mathbf{x}_i-\mu_i|>\epsilon S_n\}} (\mathbf{x}_i - \mu_i)^2 d\mathbb{P} \leq \max_{1\leq i\leq n} |\mathbf{x}_i - \mu_i|^2 \int_{\{|\mathbf{x}_i-\mu_i|>\epsilon S_n\}} d\mathbb{P}.$$

Applying Tchebychev inequality to the integral on the right, we have

$$\int_{\{|\mathbf{x}_i-\mu_i|>\epsilon S_n\}} d\mathbb{P} \leq \frac{\mathrm{Var}(\mathbf{x}_i)}{\epsilon^2 S_n^2}.$$

Hence, by applying the latter to the Lebesgue integral over $\mathbb{R}^p$ and considering the $\mathbf{x}_i$'s to be centered, we have

$$\frac{1}{S_n^2} \sum_{i=1}^{n} \int_{\{|\mathbf{x}_i-\mu_i|>\epsilon S_n\}} (\mathbf{x}_i - \mu_i)^2 d\mathbb{P} \leq \frac{1}{\epsilon^2} \frac{\max_{1\leq i\leq n} ||\mathbf{x}_i||_p^2}{\sum_{i=1}^{n} ||\mathbf{x}_i||_p^2} = o(1).$$

by (H_2) and (H_3). Thus, we have **Lindeberg Condition**

$$\lim_{n\to\infty} \frac{1}{S_n^2} \sum_{i=1}^{n} \int_{\{|\mathbf{x}_i-\mu_i|>\epsilon S_n\}} (\mathbf{x}_i - \mu_i)^2 d\mathbb{P} = 0, \quad \forall \epsilon > 0.$$

For practical applications, this assumption means that the contribution of any individual random variable $\mathbf{x}_i$ for $1 \leq i \leq n$ to the variance S_n^2 is arbitrarily small, for sufficiently large values of n.

Assumption (H_4) implies that f is uniformly bounded, uniformly continuous and square integrable on $\mathbb{R}$.

Uniform boundedness.

$$\begin{aligned}
f(x) &= |f(x)| = \left| \int_{-\infty}^{x} f(t)dt \right| \quad \text{by absolute continuity} \\
&= \left| \int_{-\infty}^{x} \frac{f'(t)}{\sqrt{f(t)}} \sqrt{f(t)} dt \right| \\
&\leq \sqrt{\int_{-\infty}^{x} \left(\frac{f'(t)}{\sqrt{f(t)}} \right)^2 dt} \sqrt{\int_{-\infty}^{x} f(t)dt} \quad \text{by Cauchy-Schwartz inequality} \\
&\leq \sqrt{\int_{-\infty}^{\infty} \left(\frac{f'(t)}{\sqrt{f(t)}} \right)^2 dt} \\
&= \sqrt{I(f)}.
\end{aligned}$$

Uniform continuity.

$$\begin{aligned}
|f(x) - f(y)| &= \left| \int_{x}^{y} f'(t)dt \right| \\
&\leq \sqrt{\int_{x}^{y} \left(\frac{f'(t)}{\sqrt{f(t)}} \right)^2 dt} \sqrt{\int_{x}^{y} f(t)dt} \\
&\leq \sqrt{I(f)} \sqrt{|F(x) - F(y)|} \quad \text{by the Mean Value Theorem} \\
&\leq \sqrt{I(f)} \sqrt{f\xi |x-y|} \quad \text{where } \xi \text{ is between } x \text{ and } y \\
&\leq I(f)^{3/4} \sqrt{|x-y|} \quad \text{by uniform boundedness.}
\end{aligned}$$

Thus for any $\epsilon > 0$,

$$|x - y| < \frac{\epsilon^2}{[I(f)]^{3/2}} \Rightarrow |f(x) - f(y)| < \epsilon.$$

Square integrability.

$$\begin{aligned}
\int_{-\infty}^{\infty} f^2(x)dx &= \int_{-\infty}^{\infty}\left(\int_{\infty}^{x} f'(t)dt\right)f(x)dx \\
&\leq \sqrt{\int_{-\infty}^{\infty}\left(\int_{-\infty}^{x} f'(t)dt\right)^2}\sqrt{\int_{-\infty}^{\infty} f^2(x)dx} \quad \text{by Cauchy-Schwartz inequality} \\
&\leq \sqrt{\int_{-\infty}^{\infty}\int_{\infty}^{x}\left(\frac{f'(t)}{\sqrt{f(t)}}\right)^2 dt}\sqrt{\int_{-\infty}^{\infty}\int_{-\infty}^{x} f(t)dt}\sqrt{\int_{-\infty}^{\infty} f^2(x)dx} \\
&\leq \sqrt{I(f)}\sqrt{\int_{-\infty}^{\infty} f^2(x)dx}.
\end{aligned}$$

Thus,

$$\int_{-\infty}^{\infty} f^2(x)dx \leq I(f).$$

Remark 3.2.2. Even if a function f is uniformly bounded, uniformly continuous, positive almost everywhere, it is not guaranteed that $I(f)$ is finite. An example is given by the function

$$f(x) = \begin{cases} \frac{1-x}{2}, & 0 \leq x \leq 1 \\ \frac{x-2j+1}{2^{j+2}}, & 2j-1 \leq x \leq 2j \\ \frac{2j+1-x}{2^{j+2}}, & 2j \leq x \leq 2j+1, j \geq 1 \\ f(-x), & x \leq 0. \end{cases}$$

Note that there are numerous estimators that satisfy assumption (H_5) including the least squares estimator and the general rank estimator with a specified score function h, see Jureckova (1971); Jaeckel (1972).

We end this discussion by noticing that (H_6) requires the design matrix $\mathbf{X}$ to be such that the sample sizes go to infinity at the same rate.

3.3 Estimation of the coefficients

The following lemma provides an alternative representation of the orthonormal basis coefficients C_{jk} in the expansion of h_F.

Lemma 3.3.1. *Assume that* (H_1) *and* (H_4) *hold.Then*

$$\int_0^1 h_F(t)\psi(t)dt = -\int_{-\infty}^{\infty} \frac{d^2}{dz^2}\Big[\psi\big(F(z)\big)\Big] F(z)dz \; . \tag{3.3.1}$$

Proof. Eq. (5) of Naranjo and McKean (1997) gives

$$\int_0^1 h_F(t)\psi(t)dt = \int_{-\infty}^{\infty} \frac{d}{dz}\left[\psi(F(z))\right] dF(z) \; . \tag{3.3.2}$$

Integrating by parts the right-hand side of equation (3.3.2), we have

$$\int_{-\infty}^{\infty} \frac{d}{dz}\left[\psi(F(z))\right] dF(z) = \left[F(z)\frac{d}{dz}\left[\psi\left(F(z)\right)\right]\right]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \frac{d^2}{dz^2}\left[\psi(F(z))\right] F(z)dz \; .$$

We can write

$$\left[F(z)\frac{d}{dz}\left[\psi\left(F(z)\right)\right]\right]_{-\infty}^{\infty} = \lim_{z\to\infty}\left[F(z)f(z)\psi'(F(z)) - F(-z)f(-z)\psi'(F(-z))\right] \; .$$

But

$$\lim_{z\to\infty} F(z) = 1, \quad \lim_{z\to-\infty} F(z) = 0$$

and

$$\lim_{z\to\infty} \psi'(F(z)) = \psi'(1) = 0 = \lim_{z\to-\infty} \psi'(F(z)) = \psi'(0), \quad \text{since } \psi \text{ is compactly supported.}$$

The Lemma follows. □

Remark 3.3.2. If we replace assumption (H_1) in Lemma 3.3.1 by the assumption that f is absolutely continuous and $\int_{-\infty}^{\infty} |f'(x)|dx < \infty$, then (3.3.1) still holds. In fact, these two conditions insure that both $\lim_{z \to -\infty} f(z)$ and $\lim_{z \to \infty} f(z)$ exist and are both equal to zero. They are restrictive though since they require a well behaved source for the sample data. In comparison, we have numerous functions ψ satisfying (H_1) such as the Daubechies base functions, Symlets and coiflets.

In the remainder of the dissertation, for $\boldsymbol{\alpha} \in \mathbb{R}^p$, we will let $F_n(\cdot; \boldsymbol{\alpha})$ represent the empirical distribution function of $y_1 - \mathbf{x}_1^T \boldsymbol{\alpha}, \ldots, y_n - \mathbf{x}_n^T \boldsymbol{\alpha}$; that is,

$$F_n(z; \boldsymbol{\alpha}) = \frac{1}{n} \sum_{i=1}^{n} I(y_i - \mathbf{x}_i^T \boldsymbol{\alpha} \leq z).$$

The following lemma is a combination of Lemma 1 and Lemma 2 of Naranjo and McKean (1997) and gives the asymptotic linearity of $F_n(w; \boldsymbol{\alpha}_n)$ for $\boldsymbol{\alpha}_n$ converging to $\mathbf{0}$ at a suitable rate. The proof follows from Section 2.3 of Koul (1992) and is given in the appendices.

Lemma 3.3.3. *Assume* $(H_1) - (H_5)$. *Then*

$$\sup_{z \in \mathbb{R}} \sqrt{n} |F_n(z; \widehat{\boldsymbol{\beta}}_n) - F(z)| = \mathrm{O}_p(1)\ .$$

Characterizations of orthonormal basis systems of $L^2(0,1)$ can be found in Meyer (1991), Cohen et al. (1993) and Andersson et al. (1994). If the scaling function φ satisfies conditions given in Theorem 9.6 of Härdle et al. (1998) (for example certain Daubechies wavelets), then h_F belongs to the Besov Space $B_2^{sq}(\mathbb{R})$. Besov spaces can be characterized using wavelet coefficients; thus, they are the natural spaces for wavelet estimation of functions. Moreover, in some Besov spaces, wavelet coefficients decay faster than Fourier coefficients. For instance, it is shown in Zygmund (2002) that if a function belongs to the Zygmund space $B^{1\infty}_{\infty}(0,1)$, then its Fourier coefficients C_n are $O(n^{-1})$. It was proved in Meyer (1990) that the wavelet coefficients W_{jk} of such a function are $O(2^{-3j/2})$.

We now start the estimation process of the score function h_F. The strategy consists of first estimating the coefficients C_{jk} by some $\widehat{C^n_{jk}}$ in the expansion (3.1.6) of h_F and then use them to provide an estimate $\widehat{h^n_F}$ of h_F, for some fixed n .

Let $\{\theta_n\}_{n\in\mathbb{N}}$ and $\{M_n\}_{n\in\mathbb{N}}$ be sequences of real numbers such that $M_n = O(n^{\alpha}), 0 < \alpha < 1/4$ and $\frac{M_n}{\sqrt{n}\theta_n^2} \to 0, M_n\theta_n^2 \to 0$ as $n \to \infty$. This means that $\theta_n = O(n^{\lambda})$ where $\lambda = \alpha/2 - 1/4 - \gamma$ and $\alpha - 1/4 < \gamma < 0$. Given a scaling function φ with corresponding "mother wavelet" ψ and a $\sqrt{n}$-consistent estimator $\widehat{\boldsymbol{\beta}}_n$ of $\boldsymbol{\beta}$, equation (3.3.1) in Lemma 3.3.1, using second differences, suggests an estimator

$$\widehat{C^n_{jk}} := \frac{1}{\theta_n^2}\int_{-\infty}^{\infty}\Big[2\psi_{jk}\big(F_n(z;\widehat{\boldsymbol{\beta}}_n)\big) - \psi_{jk}\big(F_n(z+\theta_n;\widehat{\boldsymbol{\beta}}_n)\big) - \psi_{jk}\big(F_n(z-\theta_n;\widehat{\boldsymbol{\beta}}_n)\big)\Big]F_n(z;\widehat{\boldsymbol{\beta}}_n)dz$$

of $C_{jk} = \int_0^1 h_F(s)\psi_{jk}(s)ds.$

Remark 3.3.4. Note that $\widehat{C^n_{jk}}$ can be computed from the data as

$$\widehat{C^n_{jk}} = \frac{2}{n\theta_n}\sum_{i=1}^{n} i\cdot\Big[2\phi\big(F_n(e_i;\hat{\boldsymbol{\beta}}_n)\big) - \phi\big(F_n(e_i+\theta_n;\hat{\boldsymbol{\beta}}_n)\big) - \phi\big(F_n(e_i-\theta_n;\hat{\boldsymbol{\beta}}_n)\big)\Big]\,, \quad (3.3.3)$$

where $e_i = y_i - \mathbf{x}_i^T\widehat{\boldsymbol{\beta}}_n$

Proof. In fact, let $W(z) = \displaystyle\int_{\eta}^{z} F(s)ds$, where η a zero of W. Then

$$-\int_{-\infty}^{\infty}\frac{d^2}{dz^2}\Big[\phi(F(z))\Big]F(z)dz = -\sum_{i=1}^{n}\int_{e_i-\theta_n}^{e_i+\theta_n}\frac{d^2}{dz^2}\Big[\phi(F_n(z;\hat{\boldsymbol{\beta}}_n))\Big]dW(z)\,. \quad (3.3.4)$$

For θ_n small enough so that there is only one e_j between $e_i - \theta_n$ and $e_i + \theta_n$, namely e_i, an approximation of the opposite of the integral on the right-hand side of equation

(3.3.4) is

$$\left[\frac{\phi(F_n(e_i+\theta_n;\widehat{\boldsymbol{\beta}}_n)) - 2\phi(F_n(e_i;\widehat{\boldsymbol{\beta}}_n)) + \phi(F_n(e_i-\theta_n;\widehat{\boldsymbol{\beta}}_n))}{\theta_n^2}\right]\left[W(e_i+\theta_n) - W(e_i-\theta_n)\right]$$

$$= \Big[\frac{\phi(F_n(e_i+\theta_n;\widehat{\boldsymbol{\beta}}_n)) - 2\phi(F_n(e_i;\widehat{\boldsymbol{\beta}}_n)) + \phi(F_n(e_i-\theta_n;\widehat{\boldsymbol{\beta}}_n))}{\theta_n^2}\Big]\left[\int_{e_i-\theta_n}^{e_i+\theta_n} F_n(z;\widehat{\boldsymbol{\beta}}_n)dz\right]$$

$$= \frac{2}{n\theta_n}\big[\phi(F_n(e_i+\theta_n;\widehat{\boldsymbol{\beta}}_n)) - 2\phi(F_n(e_i;\widehat{\boldsymbol{\beta}}_n)) + \phi(F_n(e_i-\theta_n;\widehat{\boldsymbol{\beta}}_n))\big]\Big[\sum_{j=1}^{n} I(e_j \le \xi_i)\Big] ,$$

where $\xi_i \in [e_i - \theta_n, e_i + \theta_n]$. Since we can replace e_i by their order statistics, the form of $\widehat{C^n_{jk}}$ proposed in equation (3.3.3) follows. □

The following lemma establishes the consistency of $\widehat{C^n_{jk}}$.

Lemma 3.3.5. *Suppose that $(H_1)-(H_5)$ are satisfied. Then* $|M_n(\widehat{C^n_{jk}} - C_{jk})| = o_p(1)$

Proof. Define $\widetilde{C^n_{jk}} = -\int_{-\infty}^{\infty} \frac{d^2}{dz^2}\Big[\psi_{jk}\big(F(z)\big)\Big]F_n(z;\widehat{\boldsymbol{\beta}}_n)dz$. Now

$$\widetilde{C^n_{jk}} = \frac{1}{\theta_n^2}\int_{-\infty}^{\infty}\left[2\psi_{jk}\left(F(z)\right) - \psi_{jk}\left(F(z+\theta_n)\right) - \psi_{jk}\left(F(z-\theta_n)\right)\right]F_n(z;\widehat{\boldsymbol{\beta}}_n)dz + \mathrm{O}(M_n\theta_n^2) .$$

Taking the difference $\widehat{C^n_{jk}} - \widetilde{C^n_{jk}}$ and expanding $\psi_{jk}(F_n)$ about $\psi_{jk}(F)$, we have

$$\begin{aligned} M_n(\widehat{C^n_{jk}} - \widetilde{C^n_{jk}}) \;=\; & \frac{2M_n}{\theta_n^2\sqrt{n}}\int_{-\infty}^{\infty}\Big[\sqrt{n}(F_n(z;\widehat{\boldsymbol{\beta}}_n) - F(z))\Big]\psi'_{jk}(\xi_{1,n}(z))F_n(z;\widehat{\boldsymbol{\beta}}_n)dz \\ & - \frac{M_n}{\theta_n^2\sqrt{n}}\int_{-\infty}^{\infty}\Big[\sqrt{n}(F_n(z+\theta_n;\widehat{\boldsymbol{\beta}}_n) - F(z+\theta_n))\Big]\psi'_{jk}(\xi_{2,n}(z))F_n(z;\widehat{\boldsymbol{\beta}}_n)dz \\ & - \frac{M_n}{\theta_n^2\sqrt{n}}\int_{-\infty}^{\infty}\Big[\sqrt{n}(F_n(z-\theta_n;\widehat{\boldsymbol{\beta}}_n) - F(z-\theta_n))\Big]\psi'_{jk}(\xi_{3,n}(z))F_n(z;\widehat{\boldsymbol{\beta}}_n)dz \\ & + O(M_n\theta_n^2) , \end{aligned}$$

where $\xi_{1,n}(z)$ is between $F_n(z;\widehat{\boldsymbol{\beta}}_n)$ and $F(z)$, $\xi_{2,n}(z)$ is between $F_n(z+\theta_n;\widehat{\boldsymbol{\beta}}_n)$ and $F(z+\theta_n)$, $\xi_{3,n}(z)$ is between $F_n(z-\theta_n;\widehat{\boldsymbol{\beta}}_n)$ and $F(z-\theta_n)$. Since ψ'_{jk} and F_n are bounded with respect to n and $\sup_z|\sqrt{n}(F_n(z;\widehat{\boldsymbol{\beta}}_n) - F(z))| = \mathrm{O}_p(1)$, we have

$$|M_n(\widehat{C^n_{jk}} - \widetilde{C^n_{jk}})| = O_p(M_n/\theta_n^2\sqrt{n}) + O(M_n\theta_n^2). \tag{3.3.5}$$

On the other hand,

$$\begin{aligned} M_n(C^n_{jk} - \widetilde{C^n_{jk}}) &= -\frac{M_n}{\theta_n^2}\int_{-\infty}^{\infty}(-\theta_n^2)(\psi_{jk}(F))''(z)F_n(z;\widehat{\boldsymbol{\beta}}_n)dz - \int_{-\infty}^{\infty}(\psi_{jk}(F))''(z)F(z)dz \\ &\quad - \frac{M_n}{\theta_n^2}\int_{-\infty}^{\infty}\left(-\frac{\theta_n^3}{6}\right)\left[(\psi_{jk}(F))'''(\kappa_1(z)) + (\psi_{jk}(F))'''(\kappa_2(z))\right]F_n(z;\widehat{\boldsymbol{\beta}}_n)dz \\ &\quad + O(M_n\theta_n^2) \\ &= M_n\int_{-\infty}^{\infty}\Big[F_n(z;\widehat{\boldsymbol{\beta}}_n) - F(z)\Big](\psi_{jk}(F))''(z)dz + O_p(M_n\theta_n) + O(M_n\theta_n^2);, \end{aligned}$$

where $\kappa_1(z) \in (z-\theta_n, z)$ and $\kappa_2(z) \in (z, z+\theta_n)$.

Thus we have

$$M_n(C^n_{jk} - \widetilde{C^n_{jk}}) = \frac{M_n}{\sqrt{n}}\int_{-\infty}^{\infty}\Big[\sqrt{n}(F_n(z;\widehat{\boldsymbol{\beta}}_n) - F(z))\Big](\psi_{jk}(F))''(z)dz + O_p(M_n\theta_n) + O(M_n\theta_n^2)\ .$$

But

$$\int_{-\infty}^{\infty}(\psi_{jk}(F))''(z)dz = \int_{-\infty}^{\infty}f'(z)\psi''_{jk}(F(z))dz + \int_{-\infty}^{\infty}f^2(z)\psi'_{jk}(F(z))dz)\ .$$

The two integrals on the right are bounded since f is absolutely continuous, $f \in L^2(\mathbb{R})$, and ψ_{jk} has bounded derivatives. Therefore we have

$$|M_n(C^n_{jk} - \widetilde{C^n_{jk}})| = O_p(M_n/\sqrt{n}) + O_p(M_n\theta_n) + O(M_n\theta_n^2). \tag{3.3.6}$$

Equations (3.3.5) and (3.3.6) imply that

$$|M_n(C^n_{jk} - \widehat{C^n_{jk}})| = O_p(M_n/\theta_n^2\sqrt{n}) + O_p(M_n/\sqrt{n}) + O(M_n\theta_n^2) = o_p(1)\ . \tag{3.3.7}$$

$\square$

Remark 3.3.6. In view of (3.2.1), this means that there is a constant $L > 0$ such that

$$|\widehat{C^n_{jk}} - C_{jk}| \leq L\frac{2^{2j}}{M_n} , \quad \forall k \in \mathbb{N} .$$

3.4 Estimation of the score function

Now define the wavelet estimator of h_F as

$$\widehat{h^n_F}(t) = \sum_{j=0}^{j_1} \sum_k \widehat{C^n_{jk}} \psi_{jk}(t) ,$$

where j_1 is some chosen resolution level in $\mathbb{N} \cup \{0\}$. Note that since we are using compactly supported wavelets, the sum over k contains only a finite number of terms for a given value of t (see Remark 10.1 on p. 127 of Härdle et al. (1998)).

Theorem 3.4.1. *Under $(H_1) - (H_5)$, we have $E\|h_F - \widehat{h^n_F}\|_2^2 = o(1)$.*

Proof. Let

$$h_F(t) = \sum_{j=0}^{j_1} \sum_k \widehat{C_{jk}} \psi_{jk}(t) + \sum_{j>j_1} \sum_k C_{jk}\psi_{jk}(t) ,$$

where the convergence is absolute in $L^2(0, 1)$. Thus

$$E\|h_F - \widehat{h^n_F}\|_2^2 \leq 2E\left(\int_0^1 \left|\sum_{j=0}^{j_1} \sum_k (C_{jk} - \widehat{C^n_{jk}})\psi_{jk}(t)\right|^2 dt\right) + 2E\left(\int_0^1 \left|\sum_{j>j_1} \sum_k C_{jk}\psi_{jk}(t)\right|^2 dt\right) .$$

The second term on the right is $o(1)$ by absolute convergence in $L^2(0, 1)$.

For the first term, we have

$$\int_0^1 \left| \sum_{j=0}^{j_1} \sum_k (C_{jk} - \widehat{C^n_{jk}}) \psi_{jk}(t) \right|^2 dt \le \int_0^1 \sum_{j=0}^{j_1} (j_1+1) \left(\sum_k |C_{jk} - \widehat{C^n_{jk}}| \; |\psi_{jk}(t)| \right)^2 dt \; .$$

But there is a positive constant L such that $|C_{jk} - \widehat{C^n_{jk}}| \le L \frac{2^{2j}}{M_n}$ (see Remark 3.3.6). Thus

$$\int_0^1 \left| \sum_{j=0}^{j_1} \sum_k (C_{jk} - \widehat{C^n_{jk}}) \psi_{jk}(t) \right|^2 dt \le L^2 (j_1+1) \sum_{j=0}^{j_1} \frac{2^{4j}}{M_n^2} \int_0^1 \left(\sum_k |\psi_{jk}(t)| \right)^2 dt \; .$$

Since, by Proposition 8.3 of Härdle et al. (1998), the integral on the right hand side is uniformly bounded in j, there is a constant $C > 0$ such that

$$\int_0^1 \left| \sum_{j=0}^{j_1} \sum_k (C_{jk} - \widehat{C^n_{jk}}) \psi_{jk}(t) \right|^2 dt \le \frac{C}{M_n^2} (j_1+1) \sum_{j=0}^{j_1} 2^{4j} \le \frac{L}{M_n^2} (j_1+1) 2^{4j_1+1}.$$

Choosing j_1 such that $(j_1+1)2^{4j_1+1} < M_n^2$ completes the proof. □

Chapter 4

Estimation of the Slope Parameter

Define

$$\mathbf{U}_n(\boldsymbol{\beta}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{x}_i h_F \left(\frac{R(y_i - \mathbf{x}_i'\boldsymbol{\beta})}{n+1} \right) .$$

Let $\widehat{\mathbf{U}}_n(\boldsymbol{\beta})$ denote the same expression with h_F replaced by $\widehat{h_F^n}$. Let $\mathbf{X}$ be the $n \times p$ matrix with $\mathbf{x}_i'$ as its ith row. Without loss of generality, we will assume that the design matrix $\mathbf{X}$ is centered; i.e, $\sum_{i=1}^{n} \mathbf{x}_i = \mathbf{0}$.

Theorem 4.0.2. *Under* (H_1) - (H_5),

$$\widehat{\mathbf{U}}_n(\mathbf{0}) \sim AN(\mathbf{0}, \Sigma) ,$$

where $\Sigma = \lim_{n\to\infty}(1/n)\mathbf{X}'\mathbf{X}$.

Proof. Heiler and Willers (1988) have shown that $\mathbf{U}_n(\mathbf{0}) \sim AN(\mathbf{0}, \Sigma)$. We will have $\widehat{\mathbf{U}}_n(\mathbf{0}) \sim AN(\mathbf{0}, \Sigma)$, if it can be shown that $\widehat{\mathbf{U}}_n(\mathbf{0}) - \mathbf{U}_n(\mathbf{0}) = o_p(1)$. Our approach follows that of Naranjo and McKean (1997) closely with minor modifications to suit wavelets.

It is enough to show that $\widehat{\mathbf{U}}_n(\mathbf{0}) - \mathbf{U}_n(\mathbf{0}) = o_p(1)$ elementwise; so, assume that $\mathbf{U}_n \equiv U_n$ is scalar. Suppose j_1 is such that $2^{4j_1+1} < M_n$. Note that $2^{4j_1+1} < M_n$ implies that $(j_1+1)2^{4j_1+1} < M_n^2$ as required by Theorem 3.4.1. Let t be a threshold such that

$C_{jk} = C_{jk} I(\max_k |C_{jk}| < t)$. We have

$$\begin{aligned}\widehat{U}_n(0) - U_n(0) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_i \left[\widehat{h_F^n}\left(\frac{R(y_i)}{n+1}\right) - h_F\left(\frac{R(y_i)}{n+1}\right)\right] \\ &= \sum_{j=0}^{j_1} \sum_k (C_{jk} - \widehat{C_{jk}^n}) \frac{1}{n} \sum_{i=1}^{n} \frac{x_i}{\sqrt{n}} \psi_{jk}\left(\frac{R(y_i)}{n+1}\right)\end{aligned}$$

Since $|C_{jk} - \widehat{C_{jk}^n}| = o_p(2^{2j}/M_n)$, it suffices to show that

$$\sum_{i=1}^{n} \frac{x_i}{\sqrt{n}} \sum_k \psi_{jk}\left(\frac{R(y_i)}{n+1}\right) = O_p(2^{2j}) \ .$$

This follows from Chebychev's inequality if

$$E\left[\sum_{i=1}^{n} \frac{x_i}{\sqrt{n}} \sum_k \psi_{jk}\left(\frac{R(y_i)}{n+1}\right)\right]^2 = O(2^{4j}) \ .$$

To that end, letting $K_{ij} = \sum_k \psi_{jk}\left\{R(y_i)/(n+1)\right\}$, we have

$$E\left[\sum_{i=1}^{n} \frac{x_i}{\sqrt{n}} \sum_k \psi_{jk}\left(\frac{R(y_i)}{n+1}\right)\right]^2 = \sum_i (x_i^2/n) E(K_{ij}^2) + \sum_{r \neq s} (x_r x_s/n) E(K_{rj} K_{sj}) \ .$$

But by Theorem 9.6 of Härdle et al. (1998), we have $K_{ij} = O(2^{2j})$ and because $(\sum_i x_i)^2 = 0$, we have by (H_3) that $\sum_i (x_i^2/n) = \sum_{r \neq s} (x_r x_s/n) = O(1)$. The proof is complete. □

Given an initial estimator $\widehat{\boldsymbol{\beta}}_n$, define the one-step estimator as

$$\widehat{\boldsymbol{\beta}}_R^* = \widehat{\boldsymbol{\beta}}_n + \tau\sqrt{n}(\mathbf{X}'\mathbf{X})^{-1} \widehat{\mathbf{U}}(\widehat{\boldsymbol{\beta}}_n) \ , \qquad (4.0.1)$$

where $\tau^{-1} = \displaystyle\int_0^1 |h_F(t)|^2 dt$. The estimators $\widehat{\boldsymbol{\beta}}_R^*$ and $\widehat{\boldsymbol{\beta}}_R$ have the same asymptotic distribution as given in the following theorem. The proof is direct and will not be given here for the sake of brevity.

Theorem 4.0.3. *If* (H_1)- (H_5) *are satisfied, then* $\widehat{\boldsymbol{\beta}}_R^* \sim AN(0, \tau^2\Sigma^{-1})$.

For practical applications of Theorem 4.0.3, one needs a consistent estimator of τ^{-1}. Koul et al. (1987) have given a consistent estimator of τ^{-1} for the case where the score function is known. Their estimator is based on a kernel density estimator of the density of the errors based on the residuals of the model. The following theorem gives a consistent estimator of τ^{-1} for the case of estimated scores.

Theorem 4.0.4. *Define* $(\widehat{\tau_F^n})^{-1} = \displaystyle\int_0^1 |\widehat{h_F^n}(t)|^2 dt$. *Then, under* (H_1) - (H_5),

$$(\widehat{\tau_F^n})^{-1} - \tau^{-1} \xrightarrow{\mathcal{P}} 0 \ .$$

Proof. Note that

$$\left| \int_0^1 |\widehat{h_F^n}(t)|^2 - |h_F(t)|^2 dt \right| \le \int_0^1 \left| |\widehat{h_F^n}(t)|^2 - |h_F(t)|^2 \right| dt \le \int_0^1 \left| \widehat{h_F^n}(t) - h_F(t) \right|^2 dt \ .$$

Thus

$$P\left(\left| (\widehat{\tau_h^n})^{-1} - \tau^{-1} \right| > \epsilon \right) \ \le \ P\left(\int_0^1 \left| \widehat{h_F^n}(t) - h_F(t) \right|^2 dt > \epsilon \right) \ ,$$

which is bounded by $\epsilon^{-1} E\|h_F - \widehat{h_F^n}\|_2^2$ by Markov's inequality. The desired result follows from Theorem 3.4.1. □

One may estimate $(\widehat{\tau_h^n})^{-1}$ using numerical integration methods (such as Gaussian quadrature) using some grid on $(0, 1)$ since the $\widehat{h_F^n}(t)$ can be computed for any given $t \in (0, 1)$.

As one application, consider testing the general linear hypothesis

$$H_0 : \mathbf{M}\boldsymbol{\beta} = \mathbf{0} \qquad \text{versus} \qquad H_1 : \mathbf{M}\boldsymbol{\beta} \neq \mathbf{0} \;,$$

where $\mathbf{M}$ is a $q \times p$ matrix of full row rank forming linear constraints. Under H_0, the quantity

$$B_{\mathbf{M}} = \frac{\left(\mathbf{M}\widehat{\boldsymbol{\beta}}_R^*\right)' \left[\mathbf{M}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{M}'\right]^{-1} \left(\mathbf{M}\widehat{\boldsymbol{\beta}}_R^*\right)}{q(\widehat{\tau_h^n})^2} \;.$$

is asymptotically $\chi^2(q)$ by Theorem 4.0.3, Theorem 4.0.4, and Slutsky's Lemma. Thus a level-α Wald test rejects H_0 if $B_{\mathbf{M}}$ exceeds the $(1-\alpha)$ quantile of the $\chi^2(q)$ distribution.

Chapter 5

Discussion

In this dissertation, we developed an asymptotically efficient rank estimator based on score functions estimated using wavelets. A consistent estimator for τ is given for the asymptotic variance of the rank estimator. This can be used in constructing Wald tests of general linear hypotheses.

5.1 Issues related to simulations

We consider the logistic density function $f(x) = \dfrac{e^{-x}}{(1+e^{-x})^2}$ Its cumulative density function F is given by $F(x) = \frac{1}{1+e^{-x}}$ and its score function for the logistic distribution is given by: $h_F(u) = 2u - 1, u \in (0, 1)$. This is density function satisfies all the assumptions $(H_1) - (H_5)$. We then generate $N = 10, 15, 75, 100$ random points from this distribution. We will apply respectively the the classical Fourier approach denoted by (CF), the first order estimate with Fourier basis functions denoted by (FOF) proposed by Naranjo and McKean to estimate its score function. Our approach that uses the second order estimations and compact supported wavelets is denoted by (SOW). Our method, even though theoretical results show it is superior to both approaches in terms of flexibility and precision has a shortcoming of its own. It's very difficult to apply with the current algorithms. The main reason for this complexity is the implicit nature of the compactly supported wavelets used. An idea of the complexity of the use of these wavelets can be seem in the cascade algorithm by Mallat (1989) which is the most used in applications of compactly supported wavelets. Indeed this algorithm by Mallat (1989) requires to have a sample of points from the function to be estimated whereas our method does not require such a sample, but is still very complex. We are working on an algorithm that will address this issue in the future. The Figure 5.1 below show the issues related to the Naranjo and McKean's approach. Indeed the continuous curve for different values of N seems to have less precision that the other one with + symbols. The dotted line is the original score function.

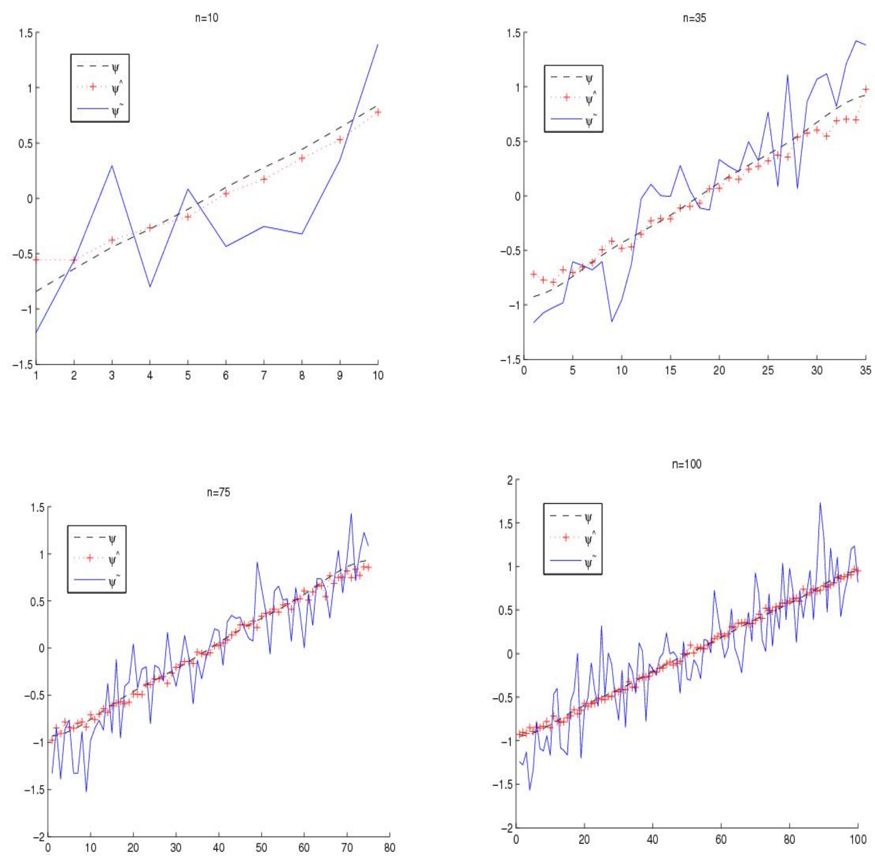

Figure 5.1: Simulations

The Table 5.1 below represents a comparison of the different methods to approximate score functions in terms of convergence, flexibility, applicability, Gibbs phenomenon, precision.

	CF	FOF	SOW
Convergence	Yes	Yes	Yes
Flexibility	No	No	Yes
Applicability	Good	Good	No yet
Gibbs	No	Yes	No
Precision	Good	Fair	Best

Table 5.1: Comparison between the different methods

5.2 Simple Mixed Models with Dependent Error Structure

In our treatment we assumed that the errors are independent and identically distributed with a cdf F that has pdf f. Estimating the score function to maximize efficiency is generally a very difficult problem for dependent error models. In simple dependent data problems, however, this may be tractable using some of the methodology developed earlier.

Consider the model

$$\mathbf{Y}_k = 1_{n_k}\mu + \mathbf{X}_k^T\boldsymbol{\beta} + e_k \quad 1 \le k \le m \quad N = \sum_{k=1}^{m} n_k. \tag{5.2.1}$$

Suppose that $e_k = 1_{n_k}b_k + \varepsilon_k$ where the ε_k are iid and b_k is a continuous random variable, independent of ε_k. We also assume the random effects $b_1, \cdots, b_m$ are iid. Then the errors e_k are exchangeable with the same marginal distribution F.

Then the asymptotic R-estimate of $\boldsymbol{\beta}$ is

$$\sqrt{N}\hat{\boldsymbol{\beta}}_R = \tau_h N(\mathbf{X}^T\mathbf{X})^{-1}U_N(\boldsymbol{\beta}) + o_p(1),$$

and from Brunner and Denker (1994), it follows that

$$\widehat{\boldsymbol{\beta}}_R \sim N_p(\boldsymbol{\beta}, V_h)$$

where

$$V_h = \tau_h^2(\mathbf{X}^T\mathbf{X})^{-1}\left(\sum_{k=1}^{m}\mathbf{X}_k^T C_{h,k}\mathbf{X}_k\right)(\mathbf{X}^T\mathbf{X})^{-1}$$

and

$$C_{h,k} = (1-\rho_h)I_{n_k} + \rho_h J_{n_k} \qquad \rho_h = \text{cov}\{h(G(e_{11})), h(G(e_{12}))\}$$

where I_{n_k} and J_{n_k} are respectively the $n_k \times n_k$ identity matrix and the $n_k \times n_k$ matrix of ones. Assuming that $\mathbf{X}$ is centered, the asymptotic relative efficiency of the rank estimator versus the least squares estimator is given by

$$\text{ARE}(\widehat{\boldsymbol{\beta}}_R, \widehat{\boldsymbol{\beta}}_{LS}) = \frac{\sigma^2(1-\rho)}{\tau_h^2(1-\rho_h)} \; ,$$

where

$$\rho = \text{Corr}(\varepsilon_1, \varepsilon_2) \quad \sigma^2 = \text{Var}(\varepsilon_1) \quad \rho_h = \text{Corr}[h(F(\varepsilon_1)), h(F(\varepsilon_2))],$$

and

$$\tau_h^{-1} = \int_0^1 h(u)\left\{-\frac{f'(F^{-1}(u))}{f(F^{-1}(u))}\right\} du.$$

We would like to find h that maximizes the ARE. This amounts to minimizing τ_h and maximizing ρ_h. However, the function that minimizes τ_h does not necessarily maximize ρ_h. Analytically finding h that would simultaneously do both is a difficult problem in calculus of variations.

Simple case: the random error vector has a multivariate normal distribution

Kloke and McKean-2009 proved that if $h(u) = \sqrt{2}(u - 1/2)$ (Wilcoxon score), then:

$$\text{ARE}(\widehat{\boldsymbol{\beta}}_R, \widehat{\boldsymbol{\beta}}_{LS}) = 12\sigma^2\left(\int f^2(t)dt\right)^2 \frac{(1-\rho)}{(1-\rho_h)},$$

where σ^2 is the variance of the underlying normal distribution, ρ is the intra-class correlation and ρ_h is the **Spearman Correlation** within each class. Hence

1. ARE$(\widehat{\boldsymbol{\beta}}_R, \widehat{\boldsymbol{\beta}}_{LS}) \in [0.8660, 0.9549]$ if $0 < \rho < 1$

2. ARE$(\widehat{\boldsymbol{\beta}}_R, \widehat{\boldsymbol{\beta}}_{LS}) \in [0.9549, 0.9662]$ if $-1 < \rho < 0$.

However, for general score function , estimating the optimal score function is the following problem in calculus of variations:

$$\widehat{h} = \operatorname*{Sup}_h \left\{ \frac{\tau_h^{-2}}{1-\rho_h} \right\} = \operatorname*{Sup}_h \left\{ \frac{\left[\int_0^1 h(u)\{ -\frac{f'(F^{-1}(u))}{f(F^{-1}(u))} \} du \right]^2}{1 - \int\int_{\mathbb{R}^2} h(F(x))h(F(y))f(x,y)dxdy} \right\} \tag{5.2.2}$$

The wavelet method developed earlier provides an approximation of the minimizer of τ_h. The maximizer of ρ_h can also be found using the techniques in this dissertation since the wavelet basis of $L^2(\mathbb{R}^2)$ can be given as a product of wavelet basis in $L^2(\mathbb{R})$. Indeed, Consider the optimization problem

$$\operatorname*{Sup}_h \int\int_{\mathbb{R}^2} h(F(x))h(F(y))f(x,y)dxdy \tag{5.2.3}$$

and let $J(h) = \displaystyle\int\int_{\mathbb{R}^2} h(F(x))h(F(y))f(x,y)dxdy - 2\int_{\mathbb{R}} g(y)h(F(y))dy$ where g is some continuous function on $\mathbb{R}$.

Then for any integrable function $\zeta \neq 0$,

$$\begin{aligned} J(h+\alpha\zeta) - J(h) &= \alpha \int_{\mathbb{R}} \zeta(F(y)) \Big(\int_{\mathbb{R}} f(x,y)h(F(x))dx \Big) dy \\ &+ \alpha \int_{\mathbb{R}} \zeta(F(x)) \Big(\int_{\mathbb{R}} f(x,y)h(F(y))dy \Big) dx \\ &+ \alpha^2 \int\int_{\mathbb{R}^2} f(x,y)\zeta(F(x))\zeta(F(y))dxdy \\ &+ -2\alpha \int_{\mathbb{R}} g(y)h(F(y))dy - 2\alpha \int_{\mathbb{R}} g(y)\zeta(F(y))dy. \end{aligned}$$

Thus, if f is a symmetric function of x and y, we have

$$\begin{aligned}\frac{J(h+\alpha\zeta)-J(h)}{\alpha} &= 2\alpha\int_{\mathbb{R}}\zeta(F(y))\left[\int_{\mathbb{R}} f(x,y)h(F(x))dx - g(y)\right]dy \\ &+ \alpha\int\int_{\mathbb{R}^2} f(x,y)\zeta(F(x))\zeta(F(y))dxdy.\end{aligned}$$

Hence,

$$\lim_{\alpha\to 0}\frac{J(h+\alpha\zeta)-J(h)}{\alpha} = 0 \Rightarrow \int_{\mathbb{R}} f(x,y)h(F(x))dx = g(y).$$

Therefore, if $f \in L^2(\mathbb{R}^2)$, then (5.2.3) becomes the Homogeneous **Fredholm Integral Equation**

$$\int_{\mathbb{R}} h(F(x))f(x,y)dx = g(y), \quad \text{for some continuous function } g. \tag{5.2.4}$$

1. If $f(x,y) = k(y-x)$, for some function k, then (5.2.4) has a solution

$$h(F(x)) = \int_{-\infty}^{\infty}\frac{\mathcal{F}_y[g(y)](u)}{\mathcal{F}_y[k(y-x)](u)}e^{2i\pi uy}du$$

   where $\mathcal{F}$ is the Fourier transform of F.

2. In general, the solution to (5.2.4) can be written as

$$h(x) = \sum_{i=1}^{\infty}\frac{< g(y), h_i(y) >}{a_i}l_i(x)$$

   where a_i is a decreasing sequence of reals, h_i, l_i are basis functions (Could be wavelets) in $L^2(\mathbb{R})$ and $<,>$ is the scalar product in $L^2(\mathbb{R})$.

Thus, we have two approximations of h. The compromise is to use the estimated score in one in the estimation of the other. This could be iterated until the difference between the two approximations is below a specified level of tolerance. Either one or the average of the two approximations can be taken as the final approximation of h as suggested by the algorithm below.

- ✓ Get residuals from an initial fit. Use them to get an estimate of F, say $\hat{F}^0$.
- ✓ Use wavelets and $\hat{F}^0$ to estimate the maximizer of τ_h, say $\hat{h}_F^\tau$.
- ✓ Fit model using $\hat{h}_F^\tau$ and get new residuals and a new estimate of F, say $\hat{F}^\tau$.
- ✓ Use $\hat{F}^\tau$ to estimate a maximizer of ρ_h, say $\hat{h}_F^\rho$.
- ✓ Fit model using $\hat{h}_F^\rho$ and get new residuals and a new estimate of F, say $\hat{F}_F^\rho$.
- ✓ Set $\hat{F}^0 = \hat{F}_F^\rho$ and go back to step 2 until $\dfrac{||\hat{h}_F^\tau - \hat{h}_F^\rho||_2}{\frac{1}{2}(||\hat{h}_F^\tau||_2 + ||\hat{h}_F^\rho||)} < \epsilon$. Otherwise stop.
- ✓ Take $\dfrac{\hat{h}_F^\tau + \hat{h}_F^\rho}{2}$ as an estimate of h.

5.3 Adequate space for score functions

In the previous sections, it was assumed that the score function belongs to the space of square integrable functions. Though this property of the score function is guaranteed if its Fisher information is finite, it is worth mentioning that this space if very "big". In fact, pinning down the adequate space where score functions could be approximated by wavelets is of some value. The space of square integrable functions is contained in the space of continuous functions which are nice for practical applications but rare in reality. This space contains the Sobolev space which is nice for theory but also rare in reality. So the compromise could be a space lying in between the continuous functions and the Sobolev space, and contained in the space of square integrable functions. The Besov space is such a space. Besov spaces can completely be characterized in terms of wavelets in the sense that any function in this space has a wavelet decomposition and any function decomposition in terms of wavelets coefficients entails a function belonging to a Besov space. As any space, Besov space have "nice" and "bad" functions. By "bad" function, we mean an analytic function that cannot be continued outside its disk of convergence. They are also called **lacunary functions**. Though the results we found were a partial answer to the problem of adequate space for score function, the techniques used made it worthwhile.

Characterization of lacunary function in Bergman-Besov-Lipchitz Spaces

The space B^ρ has been studied at length by various authors for various purposes. This space first appears in its simplest form in De Souza (1980) where it was denoted by B^1. This was later generalized to a weighted version B^ρ in De Souza (1985) and De Souza (1983).It was shown in these papers that B^ρ is the boundary value of those functions F for which $\int_0^1 \int_0^{2\pi} |F'(re^{i\theta})|(1-r)^{\frac{1}{q}-1} d\theta dr < \infty$ for the weight function $\rho(t) = t^{1/q}$. It was shown in Bloom and De Souza (1989) that B^ρ, for a general weight function ρ, is a real characterization of analytic functions in the unit disc for which $\int_0^1 \int_0^{2\pi} |F'(re^{i\theta})| \frac{\rho(1-r)}{1-r} d\theta dr < \infty$, generalizing the results obtained in De Souza (1985) and De Souza (1983).The main result here is the analytic characterization of lacunary functions in the spaces B^ρ for ρ belonging to a class $\mathfrak{G}$ of weights satisfying some conditions that will be stated in the sequel.

Preliminaries

Definition 5.3.1. Lacunary functions are analytical functions $F(z) = \sum_{k=1}^{\infty} a_k z^{n_k}$ for which $\lambda = \inf_k \frac{n_{k+1}}{n_k} > 1$.

Definition 5.3.2. We define

$$B^\rho = \left\{ F : \mathbb{D} \to \mathbb{D}, F \text{ analytic and } \quad \int_0^1 \int_0^{2\pi} |F'(re^{i\theta})| \frac{\rho(1-r)}{1-r} d\theta dr < \infty \right\}$$

and

$$b^\rho = \left\{ F : \mathbb{D} \to \mathbb{D}, F(z) = \sum_{n=0}^{\infty} a_n z^n, \quad \sum_{n=0}^{\infty} 2^n K(n,\rho) \left(\sum_{k \in I_n} |a_k|^2 \right)^{1/2} < \infty \right\} ,$$

where $I_n = \{k \in \mathbb{N} : 2^{n-1} \leq k < 2^n\}$.

Note that B^ρ and b^ρ are endowed with norms $\|F\|_{B^\rho} = \int_0^1 \int_0^{2\pi} |F'(re^{i\theta})| \frac{\rho(1-r)}{1-r} d\theta dr$ and $\|F\|_{b^\rho} = \sum_{n=0}^{\infty} 2^n K(n,\rho) \left(\sum_{k \in I_n} |a_k|^2 \right)^{1/2}$, respectively.

Notation: If $\rho(t) = t^{1/q}$, $q \geq 1$, in Definition 5.3.2, then we denote B^ρ by B^q and b^ρ by b^q.

Definition 5.3.3. We say the weight function $\rho : [0,1] \to [0,\infty)$ belongs to the class $\mathfrak{S}$ if $\rho(0) = 0$, ρ is nondecreasing, and there are positive constants $C_1, C_2, K(n,\rho)$ satisfying

$$\int_0^1 r^{2^{n-1}-1} \frac{\rho(1-r)}{1-r} dr \leq C_1 K(n,\rho), \quad \forall n \geq 1 \tag{5.3.1}$$

and

$$\int_{1-2^{-(n-1)}}^{1-2^{-n}} r^{2^n - 1} \frac{\rho(1-r)}{1-r} dr \geq C_2 K(n,\rho), \quad \forall n \geq 2 \ . \tag{5.3.2}$$

Hereafter c and C denote generic positive constants and when there is no ambiguity, we shall name all constants by c and C. Similar weight function classes can be found in Mateljević and Pavlović (1984) where they were used to characterize weighted Hardy spaces.

Lemma 5.3.4. *The class $\mathfrak{S}$ is not empty.*

Proof. Consider the family of weights defined by $\mathfrak{U} = \{\rho : \rho(t) = t^{1/q}, 1 \leq q < \infty\}$ so that $\dfrac{\rho(1-r)}{1-r} = (1-r)^{\frac{1}{q}-1}$. Clearly $\rho(0) = 0$ and ρ is a nondecreasing function of t so that the weight function ρ satisfies the conditions stated in Bloom and De Souza (1989). We will show that $\mathfrak{U} \subset \mathfrak{S}$.

Take $\rho \in \mathfrak{U}$. Using a result of Alzer (2001) and the facts that $1/q \leq 1$ and $2^n \geq 1$ we have

$$\begin{aligned} \int_0^1 r^{2^{(n-1)}-1}(1-r)^{\frac{1}{q}-1}dr &= B\left(2^{(n-1)}, \frac{1}{q}\right) \\ &\leq \frac{1}{q^{-1}2^{(n-1)}} \\ &\leq 2 \times 2^{-n/q}, \quad n \geq 1, \end{aligned} \tag{5.3.3}$$

where $B(\cdot,\cdot)$ is the beta function defined by $B(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}dt$.
Also for $1 - 2^{-(n-1)} \leq r \leq 1 - 2^{-n}$, we have $2^{n-1}2^{(1-n)/q} \leq (1-r)^{\frac{1}{q}-1} \leq 2^n 2^{-(n-1)/q}$. Thus, since $2^{1/q} > 1$ and $0 \leq r \leq 1$, we obtain

$$\begin{aligned} \int_{1-2^{-(n-1)}}^{1-2^{-n}} r^{2^n-1}(1-r)^{\frac{1}{q}-1}dr &\geq 2^{n-1}2^{-n/q}\int_{1-2^{-(n-1)}}^{1-2^{-n}} r^{2^n-1}dr \\ &\geq 2^{n-1}2^{-n/q}\int_{1-2^{-(n-1)}}^{1-2^{-n}} r^{2^n}dr \\ &= \frac{2^{n-1}2^{-n/q}}{2^n+1}\left((1-2^{-n})^{2^n+1} - (1-2^{-(n-1)})^{2^n+1}\right) \\ &= \frac{2^{n-1}2^{-n/q}}{2^n+1}(2^{-(n-1)} - 2^{-n})\sum_{k=0}^{2^n}(1-2^{-n})^{2^n-k}(1-2^{-(n-1)})^k \end{aligned}$$

But $(1-2^{-n})^{2^n-1}(1-2^{-(n-1)}) \geq (1-2^{-(n-1)})^{2^n}$. So for $0 \leq k \leq 2^n$, $(1-2^{-n})^{2^n-k}(1-2^{-(n-1)})^k \geq (1-2^{-(n-1)})^{2^n}$. Thus

$$\sum_{k=0}^{2^n}(1-2^{-n})^{2^n-k}(1-2^{-(n-1)})^k \geq (2^n+1)(1-2^{-(n-1)})^{2^n} .$$

Therefore

$$\begin{aligned} \int_{1-2^{-(n-1)}}^{1-2^{-n}} r^{2^n-1}(1-r)^{\frac{1}{q}-1} r\,dr &\geq 2^{n-1}2^{-n/q}2^{-n}(1-2^{-(n-1)})^{2^n} \\ &\geq 2^{-5}2^{-n/q}, \quad n \geq 2. \end{aligned} \tag{5.3.4}$$

Taking $K(n,\rho) = 2^{-n/q}$, (5.3.3) and (5.3.4) imply that $\rho \in \mathfrak{S}$. □

Theorem 5.3.5. *Suppose $\rho \in \mathfrak{S}$. Then b^ρ is a Banach space. Moreover for any function $F(z) = \sum_{n=1}^{\infty} a_n z^n$ belonging to B^ρ, there is a constant $C > 0$ such that*

$$\|F\|_{B^\rho} \leq C\|F\|_{b^\rho} .$$

If $F(z) = \sum_{k=1}^{\infty} a_k z^{n_k}$ is lacunary and belongs to B^ρ, then there is a constant $c > 0$ such that

$$\|F\|_{B^\rho} \geq c\|F\|_{b^\rho} .$$

Proof. We will first show that b^ρ is a Banach space. First we show that b^ρ is a linear space. Let α and β be two complex numbers and let $f, g \in b^\rho$. By Minkowski's inequality we have

$$\left(\sum_{k\in I_n} |\alpha a_k + \beta b_k|^2\right)^{1/2} \leq |\alpha| \left(\sum_{k\in I_n} |a_k|^2\right)^{1/2} + |\beta| \left(\sum_{k\in I_n} |b_k|^2\right)^{1/2} .$$

Thus $\alpha f + \beta g \in b^\rho$.

Now we show that b^ρ is complete. To that end let $\{F^s\}_{s\in\mathbb{N}}$ be a Cauchy sequence in $(b^\rho, \|\cdot\|_{b^\rho})$, where $F^s(z) = \sum_{n=0}^\infty a_n^s z^n$. We will show that there is some $F \in b^\rho$ such that $F^s \to F$ in b^ρ. Given $\epsilon > 0$, there is some $S \in \mathbb{N}$ such that for $s, t \geq S$ we have $\|F^s - F^t\|_{b^\rho} < \epsilon$; that is,

$$\sum_{n=0}^{\infty} 2^n K(n,\rho) \left(\sum_{k\in I_n} |a_k^s - a_k^t|^2 \right)^{1/2} < \epsilon$$

for $s, t \geq S$. This implies that for all $n \in \mathbb{N}$,

$$\sum_{k\in I_n} |a_k^s - a_k^t|^2 < \frac{\epsilon^2}{2^n K(n,\rho)}$$

for $s, t \geq S$. Therefore for all $n \in \mathbb{N}, \{a_n^s\}_{s\in\mathbb{N}}$ is a Cauchy sequence in $\mathbb{R}$, a complete metric space. This implies that for all $n \in \mathbb{N}$ there exists some $a_n \in \mathbb{R}$ such that $a_n^s \to a_n$ in $\mathbb{R}$. Now let $F(z) = \sum_{n=0}^\infty a_n z^n$. We shall show that $F \in b^\rho$ and that $F^s \to F$ in b^ρ. To that end, we will prove that $F - F^s \in b^\rho$ and use the linearity of b^ρ to conclude that $(F - F^s) + F^s = F \in b^\rho$.

Given $\epsilon > 0$, there exists $S \in \mathbb{N}$ such that

$$\sum_{n=0}^{\infty} 2^n K(n,\rho) \left(\sum_{k\in I_n} |a_k^t - a_k^s|^2 \right)^{1/2} < \epsilon$$

whenever $s, t \geq S$. Now let $M > 0$ be arbitrary. Then for $s, t \geq S$

$$\sum_{n=0}^{M} 2^n K(n,\rho) \left(\sum_{k\in I_n} |a_k^t - a_k^s|^2 \right)^{1/2} < \epsilon\ .$$

Fixing $s \geq S$ and letting $t \to \infty$, we have

$$\lim_{t\to\infty} \sum_{n=0}^{M} 2^n K(n,\rho) \left(\sum_{k\in I_n} |a_k^t - a_k^s|^2 \right)^{1/2} = \sum_{n=0}^{M} 2^n K(n,\rho) \left(\sum_{k\in I_n} |a_k - a_k^t|^2 \right)^{1/2} < \epsilon \ .$$

M being arbitrary, this shows that for a given $\epsilon > 0$, there exists $S \in \mathbb{N}$ such that

$$\sum_{n=0}^{\infty} 2^n K(n,\rho) \left(\sum_{k\in I_n} |a_k - a_k^s|^2 \right)^{1/2} < \epsilon$$

whenever $s \geq S$. That is $F - F^s \in b^p$. This also proves that given $\epsilon > 0$, $\|F - F^s\|_{b^\rho} < \epsilon$, for $s \geq S$ and hence $F^s \to F$ in b^ρ.

Let us now prove that there are constants $C, c > 0$ such that

$$\|F\|_{B^\rho} \leq C\|F\|_{b^\rho}$$

and, for lacunary functions,

$$\|F\|_{B^\rho} \geq c\|F\|_{b^\rho} \ .$$

Let

$$J = \int_0^1 \int_0^{2\pi} \frac{1}{2\pi} |F'(re^{i\theta})| \frac{\rho(1-r)}{1-r} d\theta dr \ .$$

Note that $F'(re^{i\theta}) = \sum_{n=1}^{\infty} n a_n r^{n-1} e^{in\theta}$. By the Cauchy-Schwarz inequality and Parseval's identity we have

$$J \leq C \int_0^1 \left(\sum_{n=1}^{\infty} |n a_n r^{n-1}|^2 \right)^{1/2} \frac{\rho(1-r)}{1-r} dr \ .$$

But

$$\int_0^1 \left(\sum_{n=1}^{\infty} |na_n r^{n-1}|^2\right)^{1/2} \frac{\rho(1-r)}{1-r} dr = \int_0^1 \left(\sum_{n=1}^{\infty}\sum_{k\in I_n} k^2|a_k|^2 r^{2(k-1)}\right)^{1/2} \frac{\rho(1-r)}{1-r} dr$$

and since $0 < r < 1$ and $2^{n-1} \leq k < 2^n$

$$J \leq C \int_0^1 \left(\sum_{n=1}^{\infty}\sum_{k\in I_n} k^2|a_k|^2 r^{2(2^{n-1}-1)}\right)^{1/2} \frac{\rho(1-r)}{1-r} dr \ .$$

Applying the Hardy-Littlewood Inequality to the right hand side and noting that $2^{n-1} \leq k < 2^n$ we have

$$J \leq C \sum_{n=1}^{\infty} 2^n \left(\sum_{k\in I_n} |a_k|^2\right)^{1/2} \int_0^1 r^{(2^{n-1}-1)} \frac{\rho(1-r)}{1-r} dr \ .$$

Since $\rho \in \mathfrak{S}$ we have

$$J \leq C \sum_{n=1}^{\infty} 2^n K(n,\rho) \left(\sum_{k\in I_n} |a_k|^2\right)^{1/2} \ .$$

On the other hand, from Zygmund (2002) we get

$$\begin{aligned} J &\geq c \int_0^1 \left(\sum_{n=1}^{\infty} n^2|a_n|^2 r^{2(n-1)}\right)^{1/2} \frac{\rho(1-r)}{1-r} dr \\ &\geq c \sum_{n=1}^{\infty} 2^n \int_{1-2^{-(n-1)}}^{1-2^{-n}} \left(\sum_{k\in I_n} |a_k|^2\right)^{1/2} r^{2^n-1} \frac{\rho(1-r)}{1-r} dr \ , \end{aligned}$$

where the last inequality is because $0 < r < 1$ and $2^{n-1} \leq k < 2^n$. Therefore

$$J \geq c \sum_{n=1}^{\infty} 2^n K(n,\rho) \left(\sum_{k\in I_n} |a_k|^2\right)^{1/2}$$

since $\rho \in \mathfrak{S}$. $\square$

Remark 5.3.6. Note that to obtain the upper bound in the main theorem, lacunary sequences are not necessary. The inequality from Zygmund (2002) that allowed us to obtain the lower bounds requires the sequence to be lacunary.

Remark 5.3.7. In the second part of the proof of the main theorem, our approach of writing the interval $[0, 1]$ as the union of non-overlapping intervals $[1 - 2^{n-1}, 1 - 2^{-n})$, $n \geq 1$, is similar to that in Blasco (2001).

Remark 5.3.8. Since the weight function $\rho(t) = t^{1/q}$ is in the class $\mathfrak{S}$, the main theorem remains true for the spaces B^q and b^q. This space has been studied extensively in Bloom and De Souza (1989) for analytic functions on the unit complex disk.

Remark 5.3.9. A particular case of the main result of this paper can be obtained using the results of Mateljević and Pavlović (1983) and Mateljević and Pavlović (1984). In this case, we will be required to use lacunarity of the series to establish both sides of the inequality. Such a result also appears in Proposition 1.7 of Jevtic and Pavlovic (1998).

Remark 5.3.10. Similar inequalities for the weight function $\rho(t) = t^{1/q}$ can be found in Zhao (1996) but the weight functions used in there were quite different from the one we used.

Bibliography

Adichie, J. N. (1967) *Estimates of regression parameters based on rank tests.* Ann. Math. Statist. 38:894–904.

Andersson, L., Hall, N., Jawerth, B., and Peters, G. (1994). *Wavelets on closed subsets of the real line.* Recent advances in wavelet analysis, volume 3 of *Wavelet Anal. Appl.*, pages 1–61. Academic Press, Boston, MA.

Alzer, H. *Sharp inequalities for the beta function*. Indag. Mathem., **12** (2001), 15–21.

Beran, R. (1974). *Asymptotically efficient adaptive rank estimates in location models.* Ann. Statist., 2:63–74.

Blasco, O. *Multipliers on spaces of analytic functions* Canad. J. Math. **47** (2001), 44–64.

Bloom, S. and G. De Souza, *Atomic decomposition of generalized Lipschitz spaces*, Ill. J. Math., **33** (1989), 181–209.

Cohen, A., Daubechies, I., and Vial, P. (1993). *Wavelets on the interval and fast wavelet transforms.* Appl. Comput. Harmon. Anal., 1(1):54–81.

Daubechies, I. (1992). Ten Lectures on Wavelets. *SIAM*

De Souza, G. (1980) *Spaces formed by special atoms*, PhD Dissertation, SUNY at Albany.

De Souza, G. (1985) *The atomic decomposition of Besov-Bergman-Lipschitz Spaces*, Proc. Amer. Math. Soc.**94** , 682–686.

De Souza, G. and Sampson, G.(1983) *A real characterization of the pre-dual of Bloch functions*, J. London Math. Soc(2), **27** , 267–276.

Dionne, L. (1981). *Efficient nonparametric estimators of parameters in the general linear hypothesis.* Ann. Statist., 9(2):457–464.

Härdle, W., Kerkyacharian, G., Picard, D., and Tsybakov, A. (1998). *Wavelets, approximation, and statistical applications*, volume 129 of Lecture Notes in Statistics. Springer-Verlag, New York.

Hong, D., Wang, J. and Gardner, R. (2005). *Real Analysis with an introduction to Wavelets and Applications.* Eslevier Academic Press.

Heiler, S. and Willers, R. (1988). *Asymptotic normality of R-estimates in the linear model.* Statistics, 19(2):173–184.

Hettmansperger, T. P. and McKean, J. W. (1983) *A geometric interpretation of inferences based on ranks in the linear model.* J. Amer. Statist. Assoc., 78:885–893.

Hettmansperger, T. P. and McKean, J. W. (1998). *Robust nonparametric statistical methods*, volume 5 of *Kendall's Library of Statistics.* Edward Arnold, London.

Hollander, M. and Wolfe, D. A. (1999) *Nonparametric statistical methods. Second edition.* John Wiley & Sons, Inc., New York.

Huber, P. J. (1964) *Robust estimation of a location parameter*, Ann. Math. Statist., 35:73–101.

Jaeckel, L. A. (1972). *Estimating regression coefficients by minimizing the dispersion of the residuals.* Ann. Math. Statist., 43(5):1449–1458.

Jevtić, M. and Pavlović, M.(1998) *Coefficient multipliers on spaces of analytic functions.* Acta Sci. Math. (Szeged)**64 , 531–545.**

Jureckova, J. (1971). *Nonparametric estimate of regression coefficients.* Ann. Math. Statist., 42(4):1328–1338.

Koenker, R., Basset, G. (1987). *Regression Quantiles.* Econometrica., 46:33–50.

Koul, H. L. (1992). *Weighted empiricals and linear models.* Institute of Mathematical Statistics Lecture Notes—Monograph Series, 21. Institute of Mathematical Statistics, Hayward, CA.

Koul, H. L., Sievers, G. L., and McKean, J. (1987). *An estimator of the scale parameter for the rank analysis of linear models under general score functions.* Scand. J. Statist., 14(2):131–141.

Koul, H. L. and Susarla, V. (1983) *Adaptive estimation in linear regression.* Statist. Decisions 1:379–400.

Mallat, S. (1989) *A Theory for Multiresolution Signal Decomposition: The Wavelet Representation* IEEE Transactions on PAMI, 7(11), 674-693

Mateljević M. and M. Pavlović, M. (1983) *L^p-Behavior of power series with positive coefficients and Hardy spaces,* Proc. Amer. Math. Soc.**87** , 309–316.

Mateljević, M. and Pavlović, M. (1984) *L^p-behaviour of the integral means of analytic functions.* Studia Math.**77** , 219–237.

Meyer, Y. (1990). *Ondelettes et opérateurs. II.* Actualités Mathématiques. [Current Mathematical Topics]. Hermann, Paris. Opérateurs de Calderón-Zygmund. [Calderón-Zygmund operators].

Meyer, Y. (1991). *Ondelettes sur l'intervalle.* Rev. Mat. Iberoamericana, 7(2):115–133.

Naranjo, J. D. and McKean, J. W. (1997). *Rank regression with estimated scores.* Statist. Probab. Lett., 33(2):209–216.

Stone, C. J. (1975) *Adaptive maximum likelihood estimators of a location parameter.* Ann. Statist. 3:267–284.

Van Eeden, C. (1970). *Efficiency-robust estimation of location.* Ann. Math. Statist., 41:172–181.

Zhao, R. (1996) *On a General family of Function Spaces*, Ann. Acad. Sci. Fenn. Math. Diss., **105**, 1–56.

Zygmund, A. (1945). *Smooth functions.* Duke Math. J., 12:47–76.

Zygmund, A. (2002). *Trigonometric series. Vol. I, II.* Cambridge Mathematical Library. Cambridge University Press, Cambridge, third edition. With a foreword by Robert A. Fefferman.

Appendix

Proof of Theorem 2.1.7

We will assume without loss of generality that the true parameter $\boldsymbol{\beta}$ is 0. It is then easier to work with the vector $T_n = (\tau_S^{-1}\sqrt{n}\widehat{\mu}_S, \sqrt{n}(\tau_h^{-1}(n^{-1}\mathbf{X}^T\mathbf{X})\widehat{\boldsymbol{\beta}}_h)^T)^T$, where $\tau_S = (2f(\theta_e))^{-1}$ and θ_e denotes the median of the error distribution, that is, $\theta_e = F^{-1}(1/2)$.

Let $t = (t_1, t_2^T)^T$ be an arbitrary, non zero vector in $\mathbb{R}^{p+1}$. We need only to show that $Z_n = t^T T_n$ has an asymptotic univariate normal distribution. Based on the asymptotic representations of $\widehat{\mu}_S$ given by

$$n^{1/2}(\widehat{\mu}_S - \mu) = \tau_S n^{-1/2} \sum_{i=1}^{n} \text{sgn}(\mathbf{Y}_i - \mu) + o_p(1),$$

and that of $\widehat{\boldsymbol{\beta}}_h$ given by

$$n^{1/2}(\widehat{\boldsymbol{\beta}}_h - \boldsymbol{\beta}) = \tau_h(n^{-1}\mathbf{X}^T\mathbf{X})^{-1}n^{-1/2}\mathbf{X}^T h(F(\mathbf{Y})) + o_p(1),$$

we have

$$Z_n = n^{-1/2}\sum_{k=1}^{n}(t_1\text{sgn}\mathbf{Y}_k) + (t_2^T\mathbf{x}_k)h(F(\mathbf{Y}_k)) + o_p(1). \tag{5.3.5}$$

Denoting the right side of 5.3.5 as Z_n^*, we need only to show that Z_n^* converges in distribution to a univariate normal distribution. Let Z_{nk}^* be the kth summand of Z_n^*. We will use the Lindenberg-Feller central Limit Theorem. First note that $E(Z_n^*) = 0$.
Let $B_n^2 = \text{Var}(Z_n^*)$. Since the individual summands are independent, $\mathbf{Y}_k$ are identically distributed and the design is centered, then B_n^2 simplifies to

$$
\begin{aligned}
B_n^2 &= n^{-1}\big(\sum_{k=1}^n t_1^2 + \sum_{k=1}^n (t_2^T \mathbf{x}_k)^2 + 2t_1 \text{cov}(\text{sgn}(\mathbf{Y}_1), h(F(\mathbf{Y}_1)))t_2^T \sum_{k=1}^n \mathbf{x}_k\big) \\
&= t_1^2 + t_2^T(n^{-1}\mathbf{X}^T\mathbf{X})t_2 + 0.
\end{aligned}
$$

Hence by (H_6), we have

$$
\lim_{n\to\infty} B_n^2 = t_1^2 + t_2^T \Sigma t_2,
$$

which is a positive number. To satisfy the Lindeberg-Feller condition, we need to show that for any $\epsilon > 0$

$$
\lim_{n\to\infty} B_n^{-2} \sum_{k=1}^n E[Z_{nk}^{*2} I(|Z_{nk}^*| > \epsilon B_n)] = 0. \tag{5.3.6}
$$

Since B_n^2 converges to a positive constant, we need only to show that the sum converges to 0. By definition of Z_{nk}^* and Cauchy-Schwartz Inequality, we have

$$
|Z_{nk}^*| \le n^{-1/2}|t_1| + N^{-1/2}|t_2^T \mathbf{x}_k||h(F(\mathbf{Y}_k))|.
$$

Hence

$$
I(|Z_{nk}^*| > \epsilon B_n) \le I(n^{-1/2}|t_1| + n^{-1/2}|t_2^T \mathbf{x}_k||h(F(\mathbf{Y}_k))| > \epsilon B_n).
$$

By Cauchy-Schwartz inequality,

$$
\begin{aligned}
n^{-1/2}|t_2^T \mathbf{x}_k| &\le n^{-1/2}||\mathbf{x}_k||_n ||t||_{p+1} \\
&= \left[n^{-1}\sum_{j=1}^p \mathbf{x}_{kj}^2\right]^{1/2} ||t||_{p+1} \\
&\le \left[p \max_j n^{-1}\mathbf{x}_{kj}^2\right]^{1/2} ||t||_{p+1}
\end{aligned}
$$

Let $K_n = \left[p \max_j n^{-1} \mathbf{x}_{kj}^2 \right]^{1/2}$. K_n is independent of k and converges to 0. Therefore we have

$$I\left(|f(F(\mathbf{Y}_k))| > \frac{\epsilon B_n - n^{-1/2} t_1}{K_n} \right) \geq I(n^{-1/2}|t_1| + n^{-1/2}|t_2^T \mathbf{x}_k||h(F(\mathbf{Y}_k))| > \epsilon B_n)$$

Finally, we also have:

$$\sum_{k=1}^{n} E\left[Z_{nk}^* I\left(|f(F(\mathbf{Y}_k))| > \frac{\epsilon B_n - n^{-1/2} t_1}{K_n} \right) \right] = t_1 E\left[I\left(|f(F(\mathbf{Y}_1))| > \frac{\epsilon B_n - n^{-1/2} t_1}{K_n} \right) \right] +$$
$$(2/n) E\left[\operatorname{sgn}(\mathbf{Y}_1) h(F(\mathbf{Y}_1)) I\left(|f(F(\mathbf{Y}_1))| > \frac{\epsilon B_n - n^{-1/2} t_1}{K_n} \right) \right] t_2^T \sum_{k=1}^{n} \mathbf{x}_k +$$
$$E\left[h^2(F(Y_1)) I\left(|f(F(\mathbf{Y}_1))| > \frac{\epsilon B_n - n^{-1/2} t_1}{K_n} \right) \right] (1/n) \sum_{k=1}^{n} (t_2^T \mathbf{x}_k)^2.$$

Note that the sum in (5.3.6) is less than or equal to the expression above. The design matrix being centered, the middle term on the right side is 0.

In the expression $\frac{\epsilon B_n - n^{-1/2} t_1}{K_n}$, the numerator converges to a positive constant as the denominator converges to 0; hence the expression goes to ∞. Since h is bounded, the indicator converges to 0. Using again the boundedness of h, the limit and the expectation can be interchanged by the Lebesgue Dominated Convergence Theorem. This show that (5.3.6) is true and hence Z_n^* converges in distribution to a univariate normal distribution. Therefore, T_n converges to a multivariate normal distribution.