Generalized Signed-Rank Estimator for Certain Complex Regression Models

by

Brice Merlin Nguelifack

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
May 9, 2015

Keywords: Multidimensional indices, Nonlinear models,Two-phase, Linear models, Compound poisson process, fix jump size, ϕ-mixing dependence, Generalized signed-rank, Strong consistency, Asymptotic normality, Random fields.

Copyright 2015 by Brice Merlin Nguelifack

Approved by

Ash Abebe, Chair, Associate Professor of Mathematics & Statistics
Peng Zeng, Associate Professor of Mathematics & Statistics
Xiaoyu Li, Assistant Professor of Mathematics & Statistics
Geraldo S. De Souza, Professor of Mathematics & Statistics

Abstract

This dissertation is mainly concerned with the Generalized signed-rank estimation of model parameters in complex regression models, specifically nonlinear models with multi-dimensional indices and two-phase linear models. First we consider a nonlinear regression model when the index variable is multidimensional. Such models are useful in signal processing, texture modeling, and spatio-temporal data analysis. A generalized form of the signed-rank estimator of the nonlinear regression coefficients is proposed. This general form of the signed-rank estimator includes L_p estimators and hybrid variants. Sufficient conditions for strong consistency and asymptotically normality of the estimator are given. It is shown that the rate of convergence to normality can be different from $\sqrt{n}$. The sufficient conditions are weak in the sense that they are satisfied by harmonic type functions for which results in the current literature are not applicable. A simulation study shows that the certain generalized signed-rank estimators (eg. signed-rank) perform better in recovering signals than others (eg. least squares) when the error distribution is contaminated or is heavy-tailed. For two-phase regression models, we consider two-phase random design linear models with arbitrary error densities and where the regression function has a fixed jump at the true change-point. We establish the consistency and the limiting distributions of signed-rank estimators of the model parameters. The left end point of the minimizing interval with respect to the change-point, herein called the signed-rank estimator $\hat{r}_n$ of the change-point parameter r, is shown to be n-consistent and the underlying process, of the standardized change-point parameter, is shown to converge weakly to a compound poisson process. This process obtains maximum over a bounded interval and $n(\hat{r}_n - r)$ converges weakly to the left end point of this interval.

Acknowledgments

Let me start by thanking God for the strength, the health and the knowledge given to me throughout all these years and simply throughout my life. Without all your blessings, I could never have accomplished this long journey.

This research project would not have been possible without the support of many people. I wish to express my sincere gratitude to my supervisor, Dr. Ash Abebe who was abundantly helpful and offered invaluable assistance, support and guidance. Deepest gratitude are also due to the members of the supervisory committee, Drs. Geraldo De souza, Peng Zeng, and Xiaoyu Li without whose knowledge and assistance this study would not have been successful.

Big dedication of this dissertation to my parents Fonguila Jean and Matibedon Regine who have always been supportive and have made a lot of sacrifices for the great success of my life.

Special thanks to my sisters and brothers, Kenfack Guilleine, Aounfack Edith, Djifack Elodie, Zeinse Fonguila Simplice, and Fonguila Loic who always stand up for me for any educational purposes. Another special thanks to Desouza's family, Abebe's family and Eddy's family who represent the key point of my success by providing all necessary advices and encouragement to my education life.

I would like also to express my profound gratitude to the Department Chair and the graduate Chair of the Mathematics and Statistics department Drs. Tin-Yau Tam, Michel Smith, Chris Rodger, Paul Schmidt and Ulrich Albrecht for all their support and encouragements during my graduate student life at Auburn University. At the same time, I would like to thank our department staff: Lori Bell, Gwen Kirk and Carolyn Donegan for their tremendous help provided when needed. I wish to express my gratitude to the following:

Prof. Geraldo S. De souza and Prof. Tin Yau Tam whose support, advice and encouragement has made this journey possible in such an amazing way.

All Professors who taught me during the course work for the marvelous job done, particularly, Drs. Michel Smith, Tin-Yau Tam, Bertram Zinner, Ming Liao, Pate, Ash Abebe, Nedret Billor, Mark Carpenter, Trevor Park and Xiaoyu Li. Also all beloved faculty members such as Drs. Greg Harris, Overtoun Jenda, Peter Johson, Olav Kallenberg, Amnon Meir, Andreas Bezdek, Frank Uhlig and Narendra Govil.

My course mates and friends, Eddy Kwessi, Bertrand Sedar Ngoma, Guy-Vanie Miankokana, Sunday Asogwa, Eze Nweze, Edith Ngouanfo, Salako, Katy Perry, Dawit Denu, Dawit Tadesse, Italo and his wife Carolina, Beny Huluka, Bill Ferguson, Matthew Ricther, Kyle Kimel, Edward Demetz, Meredith Ricther, Austin Haisten, Caroline Trible, Abouh Coulibaly, Adam Coulibaly, Mallory Mathew, Jade Withaker, Erica Handerson, Tanner Vaugha, Nick Amidon, Hannah Marxen, Kelly Joyner, Jessica Busler, Margo Kaestner, Julian Allagan, Fidele Ngwane's family, Nar Rawal, Sean O'Neil, Frank Sturn, Serge Phanzu, Jebessa Mijena, Melody Donhere, Omer Tadjion, Seth Kwame Kermaussuor, Jassep Panu, Jessica Godwin, Lilly Kristin, Moses Tam, Fortuné Massamba, Amin Bahmanian, for the good time and fun we have had.

My brothers, sisters and family members, Folifack Aloys, Tefiang Jeanne D'arc, Tsafack Eleonore, Elo's family, Nguelifack's family, Zeinse's family, Dongmo's family, Tefiang's family, Ebene's family, Kengne's family, for their support.

Finally many thanks to the Auburn community for their amazing life style and the fun things they have offered me during my study at Auburn University.

Table of Contents

Abstract ii

Acknowledgments iii

List of Figures vii

List of Tables viii

List of Abbreviations ix

1 Introduction 1

1.1 Background 1

1.2 Contribution 3

2 Generalized Signed-Rank Estimator for Nonlinear Models with Multidimensional Indices 6

2.1 Introduction 6

2.2 Generalized Signed-Rank Estimators 11

2.3 Strong Consistency 13

2.4 Asymptotic Normlity 19

2.5 Simulation Study 25

2.6 Conclusion 30

3 Signed-Rank Estimator for Two-Phase Linear Models 32

3.1 Introduction 32

3.2 Model Definition and Signed-Rank Estimation 34

3.3 Consistency 36

3.4 Asymptotic Normality 44

3.5 Simulation Study 52

3.5.1 An Application: gasoline mileage data 54

3.5.2 Application: Grade Inflation at a Large Southeastern University from 1981-2011 57
3.6 Conclusion 58

List of Figures

2.1 Signal plot of $y(m,n)$ with Gaussian $N(0,0.25)$ field perturbations 9

2.2 Signal plot of $y(m,n)$ with Cauchy$(0,0.25)$ field perturbations 9

2.3 Signal plot of $y(m,n)$ with t_5 field perturbations 10

2.4 Image plot of the signal. 11

2.5 Residual Plots 29

2.6 LS, LAD, and SR Density Profiling Plots 31

3.1 function $k_\theta(x)$: d=vertical jump, r=break point 35

3.2 MSE Plots 55

3.3 Scatter plot with the four fitted curves. 56

3.4 Scatter plot with the four fitted curves. 58

List of Tables

2.1 Estimated relative efficiencies versus LS 28

3.1 Signed-Rank estimates and estimated standard deviation for each piece 57

3.2 Parameter estimates and estimated standard deviation for each piece 59

3.3 Estimated relative efficiencies versus LS 59

3.4 Mean, and MSE, of LAD, LS and SR for $(0.5, -1, -0.7, 1, 0.0)$ 60

List of Abbreviations

ARE	Asymptotic Relative Efficiency
DDP	Density Profiling Plots
ERE	Estimate Relative Efficiency
GSR	Generalized Signed-Rank
GSRE	Generalized Signed-Rank Estimation
LAD	Least Absolute Deviation
LADE	Least Absolute Deviation Estimation
LS	Least Squares
LSE	Least Squares Estimation
MLE	Maximum Likelihood Estimation
MSE	Mean Squares Error
SR	Signed-Rank
SRE	Signed-Rank Estimation
WW	Weigthed Wilcoxon

Chapter 1

Introduction

1.1 Background

The historical approach to fitting linear and nonlinear models of the form:

$$y_{\mathbf{t}} = f(x_{\mathbf{t}}, \theta) + \epsilon_{\mathbf{t}}, \quad \mathbf{t} \leq \mathbf{n},$$

for some generic function f (linear or nonlinear) where $\mathbf{t} = (t_1, t_2, \ldots, t_k)$ and $\mathbf{n} = (n_1, n_2, \ldots, n_k) \in \mathbb{N}^k$ which is the set of k dimensional non-negative integer values; $\leq$ denotes the partial ordering, that is, for $\mathbf{m} = (m_1, m_2, \ldots, m_k) \in \mathbb{N}^k$ and $\mathbf{n} = (n_1, n_2, \ldots, n_k) \in \mathbb{N}^k$, $\mathbf{m} \leq \mathbf{n}$ if $m_i \leq n_i$ for $i = 1, 2, \ldots, k$, proceeds by finding coefficient estimate $\hat{\theta}$ of θ_0 that minimizes the sum of squared errors: $\sum_{\mathbf{t} \leq \mathbf{n}} (y_{\mathbf{t}} - f(x_{\mathbf{t}}, \theta))^2$. Such estimators, known as least squares estimators, are computationally simple and possess general optimality properties was recently developed by Bansal et al. (1999), Kundu (1993) and others. However, the optimality can be lost due to the existence of even a single extreme outlying data point. This problem is seen with the sample mean, $\overline{y}$, which is itself the least squares solution to the model $y_{\mathbf{t}} = \theta_0 + \epsilon_{\mathbf{t}}$. To overcome this problem, we briefly survey a few approaches that have been taken to develop estimators of the θ coefficients that are not as easily affected as the least-squares estimators. Two methods have been investigated: one is a generalization of least-squares estimation called M-estimation, and the other is the so-called generalized rank based estimation methods referred to as generalized signed-rank regression methods. For each of these two approaches a first method was developed to downweight outlier data points, but was later shown to be susceptible to high leverage points (outliers in the x space) in regression problems, and newer methods have emerged to address both outlier and leverage problems.

One approach that has been used to lessen the impact of outliers in linear and nonlinear models is to use the least absolute deviation also known as the L_1 regression, that is, finding coefficient estimates $\hat{\theta}$ that minimizes $\sum_{\mathbf{t}\leq\mathbf{n}} |y_{\mathbf{t}} - f(x_{\mathbf{t}}, \theta)|$. A further generalization to this, was made by Huber (1964). He obtained the so-called M-estimators by minimizing $\sum_{\mathbf{t}\leq\mathbf{n}} \rho\Big(\frac{y_{\mathbf{t}}-f(x_{\mathbf{t}},\theta)}{\hat{\sigma}_{\mathbf{t}}}\Big)$, where $\rho(\cdot)$ is a symmetric function and $\hat{\sigma}_{\mathbf{t}}$ is an estimate of the standard devaition of the errors $\epsilon_{\mathbf{t}}$. It was shown that these M-estimators have the advantage of downweighting outliers while retaining efficiency when compared to least squares estimators. However, the original M-estimators can be affected by leverage points (outliers in the x space) in regression problems. A type of M-estimator developed to protect against outliers and leverage points, is the least trimmed squares estimator that minimizes $\sum_{\mathbf{t}\leq\mathbf{m}} (y_{\mathbf{t}} - f(x_{\mathbf{t}}, \theta))^2_{(|\mathbf{t}|)}$, where $\mathbf{m} \leq \mathbf{n}$. This estimator ignores the largest $\mathbf{n} - \mathbf{m}$ residuals. However, the fact that it does not use the entire data results in a loss of efficiency. More recently, Yohai (1987) and others have developed extensions of these methods, called MM estimators, that protect against both outliers and leverage points while retaining efficiency.

At the time when Huber (1964) and others were developing the theory of M-estimators, methods based on ranking were known as R-estimation and were used for simple problems such as estimating location and scale or making location comparisons for two-sample problems. They were not considered to be generalizable to linear and nonlinear models. Later Abebe et al. (2012), Bindele and Abebe (2012), and others showed that generalized signed-rank estimators could also be cast as estimators obtained by minimizing, $\sum_{\mathbf{t}\leq\mathbf{n}} a\big(R(z_{\mathbf{t}}(\theta))\big)\rho\big(z_{\mathbf{t}}(\theta)\big)$, where $R(z_{\mathbf{t}}(\theta))$ is the rank of $z_{\mathbf{t}}(\theta) = y_{\mathbf{t}} - f(x_{\mathbf{t}}, \theta)$, $a(\cdot)$ some score function and $\rho(\cdot)$ is a positive convex function. The generalized signed-rank estimators, can be used for any general linear model, and have been shown to have high efficiency compared to least squares estimators. However, these generalized signed-rank estimators, can also be affected by leverage points in regression problems. A weighted Wilcoxon (WW) method were later developed to take care of leverage points and shown to possess highly efficient.

1.2 Contribution

The first part of this Ph.D dissertation is concerned with the study of conditions sufficient for strong consistency and asymptotic normality of a class of estimators of parameters of nonlinear regression models with multidimensional indices. The study consider an extension of the work done by Abebe et al. (2012), and Bindele and Abebe (2012) on the consistency of a certain class of estimator of nonlinear regression models and Bounded influence nonlinear signed-rank regression . In the sense that the study considers continuous functions depending on a vector of parameters and a set of random regressors. A new definition of the generalized form of the signed-rank norm is given using the multidimensional indices approach. The estimators chosen are minimizers of a generalized form of the signed-rank norm, that is, minimizing $\frac{1}{|\mathbf{n}|}\sum_{\mathbf{t}\leq\mathbf{n}} a\big(\mathbf{t}\big)\rho\big(|z(\theta)|_{(|\mathbf{t}|)}\big)$, where $z_{\mathbf{t}}(\theta) = y_{\mathbf{t}} - f(x_{\mathbf{t}}, \theta)$ and $|z(\theta)|_{(|\mathbf{t}|)}$ is the $|\mathbf{t}|^{th}$ ordered value among $|z_{\mathbf{1}}(\theta)|, \ldots, |z_{\mathbf{n}}(\theta)|$. The function $\rho : \mathbb{R}^+ \to \mathbb{R}^+$ is continuous, convex and strictly increasing. The numbers $a_{\mathbf{n}}(\mathbf{t})$ are scores generated as $a_{\mathbf{n}}(\mathbf{t}) = \varphi^+(|\mathbf{t}|/(|\mathbf{n}|+1))$, for some bounded nondecreasing score function $\varphi^+ : (0,1) \to \mathbb{R}^+$ that has at most a finite number of discontinuities. The generalization allows to make consistency statements about minimizers of a wide variety of norms including the L_1 and L_2 norms. By extending their results to the case on multidimensional indices, it is shown that the sufficient conditions in estimations are weak in the sense that they are satisfied by harmonic type functions for which results in the current literature are not applicable. Examples and Monte Carlo simulation experiments demonstrate the robustness and efficiency of the proposed estimator.

The second part of this dissertation, considers a two-phase random design linear models with arbitrary error densities and where the regression function has a fixed jump at the true change-point. This may also be considered as a type of pathological nonlinear model due to the non-zero jump in the model. In this particular case, important work has been done in the estimation of the regression parameters involving the least squares method and eventually the M-estimation method in the literature. As pointed out by many authors, the statistical inference in such models is heavily influenced by the continuity or discontinuity of the regression function at the change-point. Here we consider the case where we have a discontinuity of the regression function at the change-point defined as for $j = 1, 2$ and for $\theta_j = (\theta_j^0, \theta_j^1) \in \mathbb{R}^2$ the linear function $f_{\theta_j}(x) = \theta_j^0 + \theta_j^1 x$ then we consider the following model of two-phase linear regression function

$$k_\theta(x) = f_{\theta_1}(x) \cdot \mathbb{I}_{[-\infty,r]}(x) + f_{\theta_2}(x) \cdot \mathbb{I}_{(r,\infty]}(x),$$

where $x \in \mathbb{R}$, $\theta := (\theta^\star, r) = (\theta_1, \theta_2, r) \in \Theta = K \times \overline{\mathbb{R}}$, for a compact set K of $\mathbb{R}^4$, and $\overline{\mathbb{R}}$ represents the compactification of the real line. Here $\mathbb{I}_A$ represents the characteristic function of the set A. For a set of independent observations (X_i, Y_i), $i = 1, 2, \ldots, n$, and for some unknown $\theta \in \mathbb{R}^5$, we let

$$Y_i = k_\theta(X_i) + e_i, \qquad i = 1, 2, \ldots, n$$

where the e_i, $i = 1, 2, \ldots, n$ are independent identically distributed (i.i.d.) random variables together with the assumption that the two line segments are different and that

$$\mathrm{d} \equiv \theta_2^0 - \theta_1^0 + r(\theta_2^1 - \theta_2^1) \text{ is fixed and non-zero.}$$

The least squares method or the M-estimation method may not provide suitable estimators in the existence of even a single extreme outlier data point. To this end, we consider a particular case of the generalized signed-rank estimation called the signed rank estimation to provide suitable estimators which is obtained by minimizing the objective function

$$D_n(\theta) := \frac{1}{n} \sum_{i=1}^{n} a_n(R(|e_i(\theta)|))|e_i(\theta)|, \quad \theta \in \Theta,$$

where $R(|e_i(\theta)|)$ is the rank of $|e_i(\theta)|$ among $|e_1(\theta)|, \ldots, |e_n(\theta)|$. The numbers $a_n(i)$ are scores generated as $a_n(i) = \varphi(i/(n+1))$, for some bounded and nondecreasing score function $\varphi : (0,1) \to \mathbb{R}^+$ that has at most a finite number of discontinuities. To illustrate this case a simulation study is conducted and shows that the rank estimator perform better than the least squares estimator when dealing with the existence of a single extreme outlier data point. Also, under some suitable conditions, we obtain the consistency, and the limiting distributions of signed-rank estimators of the underlying parameters in these models. The left end point of the minimizing interval with respect to the change point, herein called the R-estimator $\hat{r}_n$ of the change-point parameter r is shown to be n-consistent and the underlying R-process, as a process in the standardized change-point parameter, is shown to converge weakly to a compound poisson process. This process obtains maximum over a bounded interval and $n(\hat{r}_n - r)$ converges weakly to the left end point of this interval. These results are different from those available in the literature for the case of two-phase linear regression models when jump sizes tends to zero as n tends to infinity.

Chapter 2

Generalized Signed-Rank Estimator for Nonlinear Models with Multidimensional Indices

2.1 Introduction

Models with multidimensional indices play an important role for spatial or spatio-temporal modeling, signal processing Rao et al. (1994); McClellan (1982), and texture modeling Francos et al. (1993); Yuan and Rao (1992); Zhang and Mandrekar (2001). An example of a spatio-temporal model would be the one with three dimensional indices that can be used to study deforestation in a specific area over time. Most models in these areas are nonlinear and there has been extensive work in the literature on nonlinear regression models with one dimensional index Wu (1981); Jennrich (1969); Gallant (1987). A classic nonlinear regression model with one dimensional index can be defined as follows:

$$y_t = f(\mathbf{x}_t, \boldsymbol{\theta}) + \epsilon_t, \quad t = 1, 2, \ldots, n, \tag{2.1}$$

where the observed data are $\{(y_t, x_t),\ t = 1, 2, \ldots, n\}$ with $y_t \in \mathbb{R}$ is random and $\mathbf{x}_t \in \mathbb{R}^m$ are constant vectors, $\Theta \subset \mathbb{R}^p$ is the parameter set with $\boldsymbol{\theta} \in \Theta$ an unknown parameter vector, the ϵ_t for $t = 1, 2, \ldots, n$ are unobserved random errors, and f is a known function. Jennrich (1969) and Wu (1981) gave some sufficient conditions based on the function f, the design x, and the error distribution to establish certain asymptotic properties of the least squares estimators of $\boldsymbol{\theta}$. As discussed in Kundu (1993), however, some of the Lipschitz type conditions on f imposed in Jennrich (1969) and Wu (1981) are not satisfied if the function f is of harmonic type.

In this paper we consider the extension of model (2.1) to the one with multidimensional indices given by

$$y_{\mathbf{t}} = f(\mathbf{x}_{\mathbf{t}}, \boldsymbol{\theta}) + \epsilon_{\mathbf{t}}, \quad \mathbf{t} \leq \mathbf{n}, \tag{2.2}$$

where $\mathbf{t} = (t_1, t_2, \ldots, t_k)$ and $\mathbf{n} = (n_1, n_2, \ldots, n_k) \in \mathbb{N}^k$ which is the space of k dimensional non-negative integer values; $\leq$ denotes the partial ordering, that is, for $\mathbf{m} = (m_1, m_2, \ldots, m_k) \in \mathbb{N}^k$ and $\mathbf{n} = (n_1, n_2, \ldots, n_k) \in \mathbb{N}^k$, $\mathbf{m} \leq \mathbf{n}$ if $m_i \leq n_i$ for $i = 1, 2, \ldots, k$. The set $\{\epsilon_{\mathbf{t}}, \mathbf{t} \in \mathbb{N}^k\}$ is a ϕ-mixing (a weakly dependent) field of random variables with mean 0, $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p$ is a parameter vector, $\{\mathbf{x}_{\mathbf{t}}, \mathbf{t} \in \mathbb{N}^k\}$ a set of known field of constant vectors, and f is a known nonlinear function. For signal processing models (see Rao et al. (1994) for an example), $y_{\mathbf{t}}$, $f(\cdot, \cdot)$ and $\epsilon_{\mathbf{t}}$ may be real or complex-valued. For notational convenience we assume them to be real-valued.

We are interested in robust estimators of $\boldsymbol{\theta}$. Our goal is to find sufficient conditions on the function $f(\cdot, \cdot)$ so that these estimators are strongly consistent and asymptotically normal as $|\mathbf{n}| = \prod_{i=1}^{k} n_i \to \infty$. Our results will also be of interest in the one dimensional case since they can be applied to harmonic type functions. This is not the case for Wu's and Jennrich's sufficient conditions (see Kundu (1993)). Bindele and Abebe (2012) were able to establish the asymptotic and robustness properties of the generalized signed-rank estimator (GSR) for nonlinear regression models with iid errors without imposing such Lipschitz type sufficient conditions. Thus the GSR estimator will be a good candidate for dealing with multidimensionally indexed nonlinear models that are of the harmonic variety. We should note that asymptotic normality of the signed rank estimator for the classic nonlinear models with dependent errors was recently established by Bindele (2014).

Example 1. *Consider the model*

$$y_t = \cos(2\pi t\theta) + \varepsilon_t, \quad \textit{where} \quad t = 1, 2, \ldots, n, \quad \theta \in [0, 1/2]. \tag{2.3}$$

This is a nonlinear model with important applications in modeling time series data. Note that the function is harmonic for $\theta \in (0, 1/2)$. *See, for example, Kundu (1993), Rice and Rosenblatt (1988), or Hannan (1973).*

Example 2. *The following nonlinear regression model involves superimposed sinusoidal signals:*

$$y_{\boldsymbol{t}} = \sum_{k=1}^{m} \alpha_k \cos(t_1\lambda_{1k} + t_2\lambda_{2k}) + \varepsilon_t, \tag{2.4}$$

where $\boldsymbol{t} = (t_1, t_2), \alpha_k's$ *are real unknown parameters, and* $\lambda_{1k}, \lambda_{2k}$ *are unknown parameters in* $[0, \pi]$. *This model can be used to model textures as discussed by Rao et al. (1994) and Zhang and Mandrekar (2001). Focusing on the case* $m = 1$, *we will illustrate the strong consistency and the asymptotic normality of the GSR estimator for* (2.4).

Example 3. *A more general form of the model* (2.3) *is the frequency model*

$$y(m, n) \;\; = \;\; \sum_{i=1}^{p} [A_k \cos(m\lambda_k + n\mu_k) + B_k \sin(m\lambda_k + n\mu_k)] + X(m, n) \;, \tag{2.5}$$

where $A_k, B_k \in \mathbb{R}$, $\lambda_k, \mu_k \in [0, \pi]$ *are unknown and* $X(m, n)$ *is a random field, possibly non-Gaussian. Given the number of components* p *and a sample* $\{y(m, n) : m = 1, \ldots, M;\ n = 1, \ldots, N\}$, *the problem of interest is to estimate* $A_k, B_k, \lambda_k, \mu_k$ *and recover the signal. Once again, this function is harmonic if* $\lambda_k, \mu_k \in (0, \pi)$. *Such models (2.5) are useful in signal processing Kay (1999) and time series analysis Brillinger (1986). For model* (2.5) *with Gaussian random field errors, consistency and asymptotic normality of the LS estimator were established by Rao and Zhao (1993). More recently, Nandi (2012) established the asymptotic properties of the LS estimator for symmetric* α*-stable errors.*

Below are examples of 3D image plot of $y(m, n)$ *in model* (2.5) *for* $p = 1$ *when there noise are assumed the come from a normal distribution, Cauchy distribution with location parameter* 0 *and scale parameter* 0.025 *and a t-distribution with* 5 *degree of freedom.*

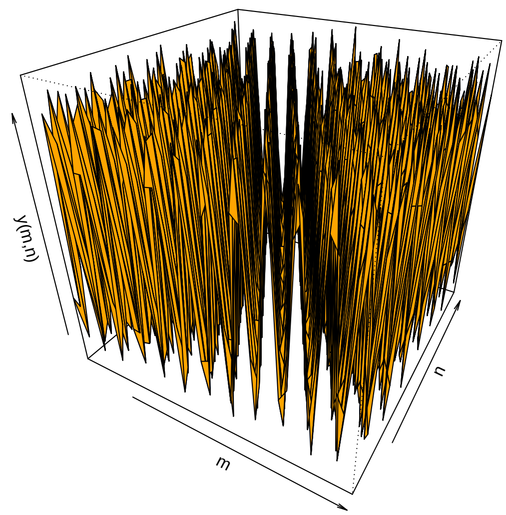

Figure 2.1: Signal plot of $y(m, n)$ with Gaussian $N(0, 0.25)$ field perturbations

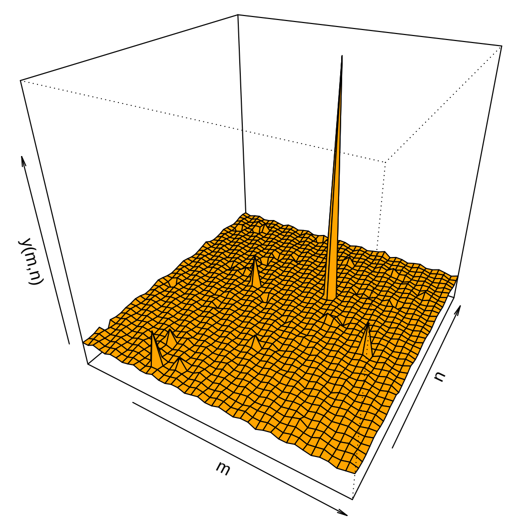

Figure 2.2: Signal plot of $y(m, n)$ with Cauchy$(0, 0.25)$ field perturbations

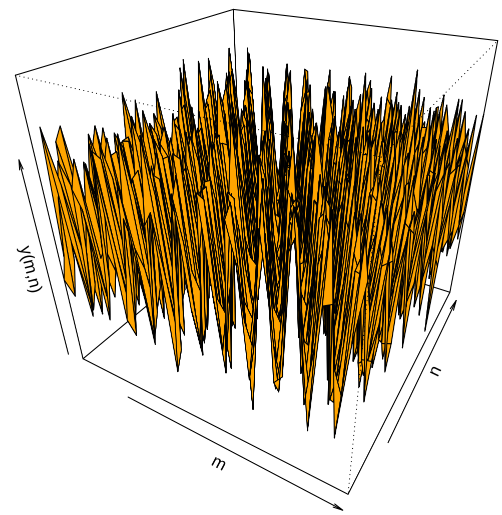

Figure 2.3: Signal plot of $y(m,n)$ with t_5 field perturbations

Figure 2.4 below represents the two dimensional image plot of $y(m,n)$ *in model* (2.5) *for* $p = 1$ *when there is no noise. We will use this model in our Monte Carlo simulation experiments to study the robustness of the GSR estimator.*

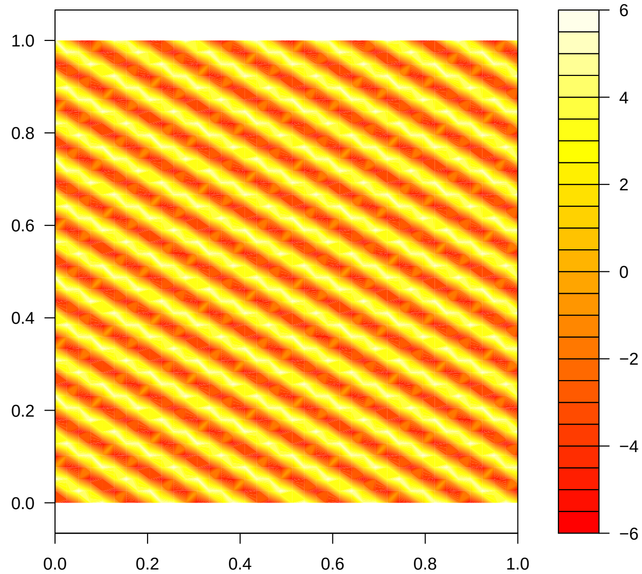

Figure 2.4: Image plot of the signal.

Several variants of the aforementioned harmonic regression models have been applied recently for modeling water quality (Autin and Edwards, 2010), for economic recessions (Bujosa et al., 2013), and chirp signals (Lahiri et al., 2013, 2014). For such models, Mitra et al. (2011) developed robust M-estimates for sequential estimation of sinusoidal signals while Lahiri et al. (2014) considered least absolute deviations estimation of one-dimensional chirp signals. Other notable works in this area include Lee and Haberman (2013), Nandi et al. (2013); Nandi and Kundu (2013), as well as the book by Kundu and Nandi (2012). As discussed above, we will develop GSR estimation for the general nonlinear regression model with multidimensional indices, including harmonic regression models.

2.2 Generalized Signed-Rank Estimators

Consider model (2.2). We shall assume that $\boldsymbol{\theta}$ is in the parameter space Θ, $\boldsymbol{\theta}_0$ is the true value of $\boldsymbol{\theta}$ which is an interior point of Θ, and $x \in \mathbb{R}^p$. We define the GSR estimator of

$\boldsymbol{\theta}_0$ to be any vector $\widehat{\boldsymbol{\theta}}_{\mathbf{n}}$ minimizing

$$D_{\mathbf{n}}(\boldsymbol{\theta}) = \frac{1}{|\mathbf{n}|} \sum_{\mathbf{t} \leq \mathbf{n}} a_{\mathbf{n}}(\mathbf{t}) \rho(|z(\boldsymbol{\theta})|_{(|\mathbf{t}|)}) \tag{2.6}$$

where $z_{\mathbf{t}}(\boldsymbol{\theta}) = y_{\mathbf{t}} - f(\mathbf{x}_{\mathbf{t}}, \boldsymbol{\theta})$ and $|z(\boldsymbol{\theta})|_{(|\mathbf{t}|)}$ is the $\mathbf{t}^{th}$ ordered value among $|z_{\mathbf{1}}(\boldsymbol{\theta})|, \ldots, |z_{\mathbf{n}}(\boldsymbol{\theta})|$. The function $\rho : \mathbb{R}^+ \to \mathbb{R}^+$ is continuous, convex and strictly increasing. The numbers $a_{\mathbf{n}}(\mathbf{t})$ are scores generated as $a_{\mathbf{n}}(\mathbf{t}) = \varphi^+(|\mathbf{t}|/(|\mathbf{n}|+1))$, for some bounded nondecreasing score generating function $\varphi^+ : (0,1) \to \mathbb{R}^+$ that has at most a finite number of discontinuities. Since $D_{\mathbf{n}}(\boldsymbol{\theta})$ is continuous in θ, Lemma 2 in Jennrich (1969) implies the existence of a minimizer of $D_{\mathbf{n}}(\boldsymbol{\theta})$. Because φ^+ is positive and nondecreasing, $D_{\mathbf{n}}$ defines a norm on $\mathbb{R}^{|\mathbf{n}|}$ by Theorem 2.1 of McKean and Schrader (1980). Boundedness of φ^+ simplifies our consistency argument; however, it can be weakened to an L_p integrability condition (see Remark 2 below). As shown in Bindele and Abebe (2012), boundedness of φ^+ is also one of the sufficient conditions for bounded influence function of the GSR estimator.

It is clear that the least squares (LS) and the least absolute deviation (LAD) estimators are particular cases of GSR estimators. In fact the LS estimator is obtained by taking $\varphi^+ \equiv 1$ and $\rho(t) = t^2, t \geq 0$ while the LAD estimator is obtained by taking $\varphi^+ \equiv 1$ and $\rho(t) = t$. The LS case has been discussed by Bansal et al. (1999) and Nandi (2012) among others.

It is perhaps much more intuitive to consider an equivalent representation of (2.6) by switching the ordering from ρ to $a_{\mathbf{n}}$ and joint ranking of the absolute residuals $|z_{\mathbf{t}}(\boldsymbol{\theta})|$ in one dimension. This is given by

$$D_{\mathbf{n}}(\theta) = \frac{1}{|\mathbf{n}|} \sum_{\mathbf{t} \leq \mathbf{n}} a_{\mathbf{n}}(R(|z_{\mathbf{t}}(\boldsymbol{\theta})|) \rho(|z_{\mathbf{t}}(\boldsymbol{\theta})|) \ ,$$

where $R(|z_{\mathbf{t}}(\boldsymbol{\theta})|)$ is the rank of $|z_{\mathbf{t}}(\boldsymbol{\theta})|$ among all $|\mathbf{n}|$ univariate residuals.

2.3 Strong Consistency

Let $\boldsymbol{\theta}_0$ be the true parameter vector. Our objective is to prove the strong consistency of $\widehat{\boldsymbol{\theta}}_{\mathbf{n}}$ in the sense that $\widehat{\boldsymbol{\theta}}_{\mathbf{n}} \to \boldsymbol{\theta}_0$ *a.s.* as $|\mathbf{n}| \to \infty$. Note that if $\{y_{\mathbf{t}}, \mathbf{1} \le \mathbf{t} \le \mathbf{n}\}$ is the observed data, where $\mathbf{1} = (1, 1, \ldots, 1) \in \mathbb{N}^k$, then the total number of observations is $|\mathbf{n}|$. For $\mathbf{t} \le \mathbf{n}$ assume that $x_{\mathbf{t}}$ and $\epsilon_{\mathbf{t}} = y_{\mathbf{t}} - f(\mathbf{x}_{\mathbf{t}}, \boldsymbol{\theta})$ are independent random variables with distributions H and G, respectively. Let $\widetilde{G}_{\boldsymbol{\theta}}$ denote the distribution of $|z(\boldsymbol{\theta})|$ and define $\tilde{G}_{\theta}^{-1}(s) = \inf\{y : \tilde{G}_{\theta}(y) \ge s\}$ as the quantile function corresponding to the distribution function $\tilde{G}_{\theta}$. Let $\xi_{\mathbf{1}}, \ldots, \xi_{\mathbf{n}}$ be a sequence of $|\mathbf{n}|$ iid random variables uniformly distributed on $[0, 1]$ and let $\xi_{(1)} \le \cdots \le \xi_{(|\mathbf{n}|)}$ be their order statistics. Then the joint distributions of the random vectors $(|z(\theta)|_{(1)}, \ldots, (|z(\theta)|_{(|\mathbf{n}|)})$ and $(\tilde{G}_{\theta}^{-1}(\xi_{(1)}), \ldots, (\tilde{G}_{\theta}^{-1}(\xi_{(|\mathbf{n}|)}))$ coincide. Thus we can rewrite $D_{\mathbf{n}}(\boldsymbol{\theta})$ in equation (2.6) as

$$D_{\mathbf{n}}(\boldsymbol{\theta}) = \frac{1}{|\mathbf{n}|} \sum_{\mathbf{t} \le \mathbf{n}} a_{\mathbf{n}}(\mathbf{t})(\rho \circ \tilde{G}_{\boldsymbol{\theta}}^{-1})(\xi_{(|\mathbf{t}|)}) \tag{2.7}$$

and $\widehat{\boldsymbol{\theta}}_{\mathbf{n}} = \underset{\boldsymbol{\theta} \in \Theta}{\text{Argmin}}\, D_{\mathbf{n}}(\boldsymbol{\theta})$. To prove the strong consistency, we will use Lemma 2.1 in Bansal et al. (1999) and an extension to random fields of Lemma 1 in Wu (1981). We first make the following assumptions on the function $f(\cdot, \cdot)$, the parameter space Θ, and the distribution G:

A1 : The parameter space Θ is compact and the function $f(\cdot, \cdot)$ is continuous with continuous derivatives.

A2 : G has Lebesgue density g that is symmetric about 0 and strictly decreasing on $\mathbb{R}^+$.

A3 : $P(f(\mathbf{x}, \boldsymbol{\theta}) = f(\mathbf{x}, \boldsymbol{\theta}_0)) < 1$ for any $\theta \ne \theta_0$.

A4 : For $1 < q < \infty$, there is an integrable function h not depending on $\boldsymbol{\theta}$ such that $|\rho(\tilde{G}_{\boldsymbol{\theta}}^{-1}(v))| \le h(v)$, for all $\boldsymbol{\theta} \in \Theta$ with $E[h^q(Y)] < \infty$.

Remark 1. *Assumption **A3** is needed for* $\boldsymbol{\theta}_0$ *to be identified. In our proof all we need is that the space defined by*

$$\Omega^{\varphi^+}_{\boldsymbol{\theta},\boldsymbol{\theta}_0} = \{s \in (0,1) : \tilde{G}_{\boldsymbol{\theta}}(s) \neq \tilde{G}_{\boldsymbol{\theta}_0}(s) \text{ and } \varphi^+(s) > 0\},$$

has positive measure for $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$*.*

Remark 2. *Since* φ^+ *is bounded, we can always find a* p *such that* $\|\varphi^+\|_p < \infty$*, where* $1/p + 1/q = 1$ *and* $1 < q < \infty$*. Then **A4** and Holder's inequality ensure that the product* $(\varphi^+)(\rho \circ \tilde{G}^{-1}_{\boldsymbol{\theta}})$ *is integrable. Furthermore, since* ρ *is a convex function, the Minkowski's inequality yields*

$$\{E[\rho(|z_{\boldsymbol{t}}(\boldsymbol{\theta})|)]^q\}^{1/q} \leq \{E[\rho(|\epsilon_{\boldsymbol{t}}|)]^q\}^{1/q} + \{E[\rho(|f(\mathbf{x}_{\boldsymbol{t}}, \boldsymbol{\theta}) - f(\mathbf{x}_{\boldsymbol{t}}, \boldsymbol{\theta}_0)|)]^q\}^{1/q}.$$

Thus separate conditions on ϵ *and* f *are sufficient for* $E[\rho(|z(\boldsymbol{\theta})|)]^q < \infty$*.*

Remark 3. *Assumption **A2** admits a wide variety of error distribution examples of which are the normal, the logistic, and the Cauchy distributions with location parameter equal to* 0 *which we are going to explore in Section 3.5.*

Remark 4. *Under assumptions **A1**–**A3**, we can use a similar strategy as in Hossjer (1994) to show that for any* $s > 0$*, for* $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ *and for all* $\boldsymbol{t} \leq \boldsymbol{n}$*,*

$$\tilde{G}_{\theta}(s) = P(|\epsilon_{\boldsymbol{t}} - \{f(\mathbf{x}_{\boldsymbol{t}}, \boldsymbol{\theta}) - f(\mathbf{x}_{\boldsymbol{t}}, \boldsymbol{\theta}_0)\}| \leq s) < E_X\{P_{\epsilon_t}(|\epsilon_{\boldsymbol{t}}| \leq s)\} = \tilde{G}_{\boldsymbol{\theta}_0}(s).$$

Definition 1. *A random field* $\{X_{\boldsymbol{n}}, \boldsymbol{n} \in \mathbb{N}^k\}$ *is said to be uniformly integrable if*

$$\lim_{c \to \infty} \sup_{\boldsymbol{n} \in \mathbb{N}^k} E[|X_{\boldsymbol{n}}| \mathbb{I}_{\{|X_{\boldsymbol{n}}| \geq c\}}] = 0,$$

where $\mathbb{I}_V$ *is the indicator function of the set* V*.*

Let us consider the random field $\{C_{\mathbf{n}}(U), \mathbf{n} \in \mathbb{N}^k\}$, with, $C_{\mathbf{n}}(U) = \sum_{\mathbf{t} \leq \mathbf{n}} a_{\mathbf{n}}(\mathbf{t}) \mathbb{I}_{B_{\mathbf{n},\mathbf{t}}}(U)$ for $\mathbf{t} \leq \mathbf{n}$, and where U is a random variable that is uniformly distributed on $[0, 1]$, and $B_{\mathbf{n},\mathbf{t}}$ is defined as:

$$B_{\mathbf{n},\mathbf{t}} = \left\{ s : \frac{(|\mathbf{t}| - 1)}{|\mathbf{n}|} < s \leq \frac{|\mathbf{t}|}{|\mathbf{n}|} \right\},$$

and $a_{\mathbf{n}}(\mathbf{t}) = \varphi^+(|\mathbf{t}|/(|\mathbf{n}| + 1))$. Uniform integrability of $\{C_{\mathbf{n}}(U)\}$ will be used in our proof of the consistency of $\widehat{\boldsymbol{\theta}}_{\mathbf{n}}$. To that end, it is easy to see that $C_{\mathbf{n}}(U)$ converges a.s. pointwise to $\varphi^+(U)$ and that $C_{\mathbf{n}}$ is a step function. Thus, the condition for uniform integrability of $\{C_{\mathbf{n}}(U)\}$ becomes

$$\lim_{c \to \infty} \sup_{\mathbf{n} \in \mathbb{N}^k} \frac{1}{|\mathbf{n}|} \sum_{\mathbf{t} \in A_c} |\varphi^+(|\mathbf{t}|/(|\mathbf{n}| + 1))| = 0,$$

where $A_c = \{\mathbf{t} : |\varphi^+(|\mathbf{t}|/(|\mathbf{n}| + 1))| > c\}$. This condition is satisfied since we have convergence a.s. in $L^1(0, 1)$ of $C_{\mathbf{n}}(U)$. Moreover, for $1 \leq p < \infty$, the quantity $\sup_{\mathbf{n} \in \mathbb{N}^k} E[\|C_{\mathbf{n}}(U)\|_p]$ is finite, that is

$$\sup_{\mathbf{n} \in \mathbb{N}^k} \left\{ |\mathbf{n}|^{-1} \sum_{\mathbf{t} \leq \mathbf{n}} |\varphi^+(|\mathbf{t}|/(|\mathbf{n}| + 1))|^p \right\}^{1/p} < \infty \, . \tag{2.8}$$

Remark 5. *Assumption **A4**, equation* (2.8)*, the uniform integrability of* $\{C_{\boldsymbol{n}}(U)\}$*, and the fact that* $\rho \circ \tilde{G}_{\boldsymbol{\theta}}^{-1}$ *is continuous on* $[0, 1]$ *guarantee that* $D_{\mathbf{n}}(\boldsymbol{\theta}) - \mu_{\mathbf{n}}(\boldsymbol{\theta}) \to 0$ *pointwise a.s. for all* $\boldsymbol{\theta} \in \Theta$*, where*

$$\mu_{\boldsymbol{n}}(\boldsymbol{\theta}) = \int_0^1 C_{\mathbf{n}}(u) \rho \circ \tilde{G}_{\boldsymbol{\theta}}^{-1}(u) dF(u) < \infty.$$

and $F(u) = u$ *is the distribution function of* U*.*

Since the score generating function φ^+ is bounded, the a.s. convergence of $C_{\mathbf{n}}$ to φ^+ is uniform on $[0, 1]$ as $|\mathbf{n}| \to \infty$. Thus, from Theorem 1 of Baklanov (2006), we obtain, $\mu_{\mathbf{n}} \to \mu \quad a.s. \quad \text{as} \quad |\mathbf{n}| \to \infty$, where $\mu : \Theta \to \mathbb{R}$ is a function defined as

$$\mu(\boldsymbol{\theta}) = \int_0^1 \varphi^+(s) \rho \circ \tilde{G}_{\boldsymbol{\theta}}^{-1}(s) ds \, . \tag{2.9}$$

The lemma below contains an extension to multidimensional indices of Theorem 1 of Baklanov (2006).

Lemma 1. Under **A1**–**A4**, we have that for all $\boldsymbol{\theta} \in \Theta$

$$D_{\mathbf{n}}(\boldsymbol{\theta}) \to \mu(\boldsymbol{\theta}) \quad a.s. \quad \text{as } |\mathbf{n}| \to \infty \ ,$$

where μ satisfies

$$\inf_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\| \geq \delta} \mu(\boldsymbol{\theta}) > \mu(\boldsymbol{\theta}_0) \quad \text{for every } \delta > 0 \ .$$

Proof. Under **A1**–**A4**, the a.s. pointwise convergence of $D_{\mathbf{n}}(\boldsymbol{\theta})$ follows the same approach as the proof of Theorem 1 in Baklanov (2006), expression (2.7), and Remark 5 which also provides the function given by equation (2.9).

To establish the last part of Lemma 1, we take in consideration the fact that the function ρ is strictly increasing and positive on $\mathbb{R}^+$, and Remark 12 above. That is, for $\boldsymbol{\theta} \in \{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \geq \delta\}$,

$$\begin{aligned} \mu(\boldsymbol{\theta}) - \mu(\boldsymbol{\theta}_0) &= \int \varphi^+(s)[\rho \circ \tilde{G}_{\boldsymbol{\theta}}^{-1}(s) - \rho \circ \tilde{G}_{\boldsymbol{\theta}_0}^{-1}(s)]ds \\ &= \int \varphi^+(s)[\rho(\tilde{G}_{\boldsymbol{\theta}}^{-1}(s)) - \rho(\tilde{G}_{\boldsymbol{\theta}_0}^{-1}(s))]ds \\ &> 0, \end{aligned} \tag{2.10}$$

where inequality (2.10) follows from the strictly increasing property of the function ρ, and Remark 12. It then follows that $\mu(\boldsymbol{\theta}) > \mu(\boldsymbol{\theta}_0)$ whenever $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$. Thus from the compactness of $\Theta^* = \{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \geq \delta\}$, we obtain that

$$\inf_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\| \geq \delta} \mu(\boldsymbol{\theta}) > \mu(\boldsymbol{\theta}_0) \quad \text{for every } \delta > 0.$$

□

The theorem below gives the strong consistency of GSR estimators.

Theorem 2.1. Under **A1**–**A4**, we have

$$\widehat{\boldsymbol{\theta}}_{\mathbf{n}} \to \boldsymbol{\theta}_0 \quad a.s. \text{ as } \quad |\mathbf{n}| \to \infty.$$

Proof. To establish the proof of Theorem 2.1, we follow the same strategy as in the proof of Theorem 1 in Abebe et al. (2012). Thus technical details can be found in that paper. By Lemma 1 of Wu (1981), to establish the consistency of $\widehat{\theta}_{\mathbf{n}}$, it is sufficient to show that

$$\liminf_{|\mathbf{n}|\to\infty} \inf_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\geq\delta} [D_{\mathbf{n}}(\boldsymbol{\theta}) - D_{\mathbf{n}}(\boldsymbol{\theta}_0)] > 0 \quad a.s. \tag{2.11}$$

But observe that

$$\inf_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\geq\delta} [D_{\mathbf{n}}(\boldsymbol{\theta}) - D_{\mathbf{n}}(\boldsymbol{\theta}_0)] \geq \inf_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\geq\delta} (A_{\mathbf{n}}(\boldsymbol{\theta})) + \inf_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\geq\delta} (B(\boldsymbol{\theta}, \boldsymbol{\theta}_0)) + Q_{\mathbf{n}}(\boldsymbol{\theta}_0),$$

where $A_{\mathbf{n}}(\boldsymbol{\theta}) = D_{\mathbf{n}}(\boldsymbol{\theta}) - \mu(\boldsymbol{\theta})$, $B(\boldsymbol{\theta}, \boldsymbol{\theta}_0) = \mu(\boldsymbol{\theta}) - \mu(\boldsymbol{\theta}_0)$, and $Q_{\mathbf{n}}(\boldsymbol{\theta}_0) = \mu(\boldsymbol{\theta}_0) - D_{\mathbf{n}}(\boldsymbol{\theta}_0)$. As a result of Remark 5 and Lemma 1

$$\liminf_{|\mathbf{n}|\to\infty} \inf_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\geq\delta} Q_{\mathbf{n}}(\boldsymbol{\theta}_0) = 0 \quad a.s.$$

Due to the second part of Lemma 1, we have

$$\inf_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\geq\delta} B(\boldsymbol{\theta}, \boldsymbol{\theta}_0) > 0.$$

For the statement given in (2.11) to hold, it suffices to show that

$$\liminf_{|\mathbf{n}|\to\infty} \inf_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\geq\delta} A_{\mathbf{n}}(\boldsymbol{\theta}) = 0 \quad a.s.$$

By Lemma 1, we have $A_{\mathbf{n}}(\boldsymbol{\theta}) \to 0$ a.s. uniformly for all $\boldsymbol{\theta} \in \Theta$ such that $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0| \geq \delta$. $A_{\mathbf{n}}(\boldsymbol{\theta})$ being uniformly convergent and continuous on the compact set $\Theta^* = \{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \geq \delta\}$, it follows from Lemma 3 in Abebe et al. (2012) that $A_{\mathbf{n}}(\boldsymbol{\theta})$ is equicontinuous on Θ^*. It follows that

$$\liminf_{|\mathbf{n}|\to\infty} \inf_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\geq\delta} A_{\mathbf{n}}(\boldsymbol{\theta}) = 0 \quad a.s.,$$

and the proof is complete. □

We now illustrate the sufficient conditions for strong consistency using a simple example.

Example 4. *Consider GSR estimators of the parameter* θ *in model* (2.3) *given in Example 1. For simplicity, we will assume that the errors are i.i.d. from* $N(0, \sigma^2)$ *with* $\sigma < \infty$ *and* $\rho(s) = s$. *From Theorem* 2.1, *showing strong consistency amounts to showing that* ***A1***–***A4*** *are satisfied under model* (2.3). ***A1*** *is satisfied because the parameter space* $[0, \frac{1}{2}]$ *is compact.* ***A2*** *is immediate since* G *is normal with mean 0. To verify* ***A3***, *we let* θ *and* θ_0 *be any two points in* $\Theta = [0, \frac{1}{2}]$ *such that* $\theta_0 \neq \theta$. ***A3*** *follows since we can always find a* $t = 1, 2, \ldots, n$ *such that* $|\cos(2\pi\theta t) - \cos(2\pi\theta_0 t)| > 0$. *Since* $x_t = t$ *is fixed, assumption* ***A4*** *holds if* $E(|\epsilon|^q) < \infty$ *for* $1 < q < \infty$. *Then we can show that*

$$E(|\epsilon|^q) \leq E(|\epsilon|^{\lceil q \rceil}) = \frac{\sigma^{\lceil q \rceil} 2^{\lceil q \rceil/2} \Gamma\left(\frac{\lceil q \rceil + 1}{2}\right)}{\sqrt{\pi}} < \infty\ , \tag{2.12}$$

where $\lceil \cdot \rceil$ *is the ceiling function. Thus the GSR estimators for model* (2.3) *are strongly consistent.*

Example 5. *Under assumptions* ***A1***–***A4***, *GSR estimators for model* (2.4) *given in Example 2 are strongly consistent. For notational convenience, we assume that* $m = 1$ *and deal only with*

$$y_{\boldsymbol{t}} = \alpha \cos(\lambda_1 t_1 + \lambda_2 t_2) + \epsilon_{\boldsymbol{t}}, \quad \boldsymbol{1} \leq \boldsymbol{t} \leq \boldsymbol{n}, \tag{2.13}$$

where $\lambda_1 \in [-\pi, \pi]$ *and* $\lambda_2 \in [0, \pi]$. *Further, we assume that* $m \leq |\alpha| \leq M < \infty$, *for some* $M > m > 0$. *This is a reasonable assumption since* α *represents the amplitude of the waves. From Theorem* 2.1, *showing strong consistency amounts to showing that* ***A1***–***A4*** *are satisfied* *under model* (2.13).

The parameter space Θ *is*

$$\{(\alpha, \lambda_1, \lambda_2) : m \le \alpha \le M, \lambda_1 \in [-\pi, \pi], \lambda_2 \in [0, \pi]\}.$$

Clearly ***A1*** *is satisfied. Remark 3 ensures that for a chosen distribution of the errors* $\epsilon_{\boldsymbol{t}}$ *for all* $\boldsymbol{t} \le \boldsymbol{n}$*, assumption* ***A2*** *can be easily verified. To verify* ***A3****, we let* $\theta = (\alpha, \lambda_1, \lambda_2)^T$ *and* $\theta_0 = (\alpha_0, \lambda_{10}, \lambda_{20})^T$ *be any two points in* Θ *such that* $\theta_0 \ne \theta$*. Then we can always find a* $\boldsymbol{t} = (t_1, t_2)$ *such that*

$$|\alpha \cos(\lambda_1 t_1 + \lambda_2 t_2) - \alpha_0 \cos(\lambda_{10} t_1 + \lambda_{20} t_2)| > 0.$$

Thus ***A3*** *is immediate. Remark 2 can be used to easily verify assumption* ***A4****. Thus the GSR estimators for model* (2.4) *are strongly consistent.*

2.4 Asymptotic Normlity

Put $\Gamma_{\boldsymbol{\theta}}(s) = \rho[\tilde{G}_{\boldsymbol{\theta}}^{-1}(s)]$ for $s \in [0,1]$ and $\lambda_{\mathbf{t}} = a_{\mathbf{n}}(R(\xi_{|\mathbf{t}|}))$, where $R(\xi_{|\mathbf{t}|})$ is the rank of $\xi_{|\mathbf{t}|}$ among $\xi_{\mathbf{1}}, \ldots, \xi_{\mathbf{n}}$. Then (2.6) can be rewritten as

$$D_{\mathbf{n}}(\boldsymbol{\theta}) = \frac{1}{|\mathbf{n}|} \sum_{\mathbf{t} \le \mathbf{n}} a_{\mathbf{n}}(\mathbf{t}) (\rho \circ \tilde{G}_{\theta}^{-1})(\xi_{|\mathbf{t}|}) = \frac{1}{|\mathbf{n}|} \sum_{\mathbf{t} \le \mathbf{n}} \lambda_{\mathbf{t}} \Gamma_{\theta}(\xi_{|\mathbf{t}|}).$$

A4 shows that $\|\lambda_{\mathbf{t}}\| < \infty$. Set $S_{\mathbf{n}}(\boldsymbol{\theta}) = \mathcal{D}_{\boldsymbol{\theta}}^{\beta} D_{\mathbf{n}}(\boldsymbol{\theta})$ and $\Phi_{\boldsymbol{\theta}}(s) = \mathcal{D}_{\boldsymbol{\theta}}^{\beta} \Gamma_{\boldsymbol{\theta}}(s)$ for $|\beta| = 1$, where $\mathcal{D}_{\boldsymbol{\theta}}^{\beta}$ is the differential operator defined by $\mathcal{D}_{\boldsymbol{\theta}}^{\beta} = \frac{\partial^{|\beta|}}{\partial\theta_1^{\beta_1} \ldots \partial\theta_p^{\beta_p}}$ with $\beta = (\beta_1, \ldots, \beta_p) \in \mathbb{N}_0^n, \mathbb{N}_0 = \mathbb{N} \cup \{0\}$ the multi-index and $|\beta| = \sum_{i=1}^{p} \beta_i$. Let the $|\mathbf{n}| \times p$ matrix $\boldsymbol{X}^*$ be the matrix of $\Phi_{\boldsymbol{\theta}}$ evaluated at all $|\mathbf{n}|$ residuals $z(\boldsymbol{\theta})$ and $h_{\mathbf{tt}}^{\mathbf{n}}$ be the $|\mathbf{t}|th$ diagonal element of the hat-matrix

$\boldsymbol{X}^*(\boldsymbol{X}^{*T}\boldsymbol{X}^*)^{-1}\boldsymbol{X}^{*T}$. It follows that $\widehat{\theta}_{\mathbf{n}}$ is a zero of

$$S_{\mathbf{n}}(\boldsymbol{\theta}) = \frac{1}{|\mathbf{n}|}\sum_{\mathbf{t}\le\mathbf{n}} \lambda_{\mathbf{t}}\Phi_\theta(\xi_{|\mathbf{t}|}). \tag{2.14}$$

In the discussion to follow, as in Bindele and Abebe (2012), we will let $W^{m,p}(\Omega)$ be the usual Sobolev space on an open neighborhood Ω of $\mathbb{R}^{|\mathbf{n}|}$ defined as

$$W^{m,p}(\Omega) = \left\{\Gamma \in L^p(\Omega):\ D^\beta_{\boldsymbol{\theta}}\Gamma \in L^p(\Omega) \text{ with } |\beta| \le m\right\} .$$

To establish the asymptotic normality of $\widehat{\boldsymbol{\theta}}_{\mathbf{n}}$, we will need the following regularity conditions in addition to assumptions **A1**– **A4**:

A5 : Let $\{M_{\mathbf{n}}, \mathbf{n} \in \mathbb{N}^k\}$ be a field of $k \times k$ non-singular matrices such that

$$\frac{1}{|\mathbf{n}|} M^T_{\mathbf{n}} \sum_{\mathbf{t}\le\mathbf{n}} \{\nabla_{\boldsymbol{\theta}} f(\mathbf{x}_{\mathbf{t}}, \boldsymbol{\theta})\}\{\nabla_{\boldsymbol{\theta}} f(\mathbf{x}_{\mathbf{t}}, \boldsymbol{\theta})\}^T M_{\mathbf{n}}$$

converges to a positive definite matrix $\Sigma_{\boldsymbol{\theta}_0}$ uniformly as $|\mathbf{n}| \to \infty$ and $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \to 0$.

A6 : $\lim\limits_{|\mathbf{n}|\to\infty} \max\limits_{\mathbf{1}\le\mathbf{t}\le\mathbf{n}} h^{\mathbf{n}}_{\mathbf{tt}} = 0$.

A7 : $\boldsymbol{\theta} \to \Gamma_{\boldsymbol{\theta}}(t)$ is a map in $W^{3,p}(B)$, where B is a neighborhood of $\boldsymbol{\theta}_0$ for every fixed t.

A8 : $A_{\boldsymbol{\theta}_0} = E[\varphi^+(\xi)[\mathcal{D}^\beta_{\boldsymbol{\theta}}\Phi_{\boldsymbol{\theta}}(\xi)]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}]$, where $\xi \sim U(0,1)$, is a positive definite matrix for $|\beta| = 1$.

A9 : There exist functions $\psi_\beta \in W^{2,p}(\mathbb{R})$ independent of θ such that $|\mathcal{D}^\beta_{\boldsymbol{\theta}}\Phi_{\boldsymbol{\theta}}(s)| \le \psi_\beta(s)$ for every $\boldsymbol{\theta} \in B$ and $|\beta| \le 2$.

Note that **A6**–**A9** are generalizations of assumption 4 and 5 in Bansal et al. (1999). So for our example in this section we are going to show that the assumptions **A6**–**A9** are satisfied. More details about these general assumptions can be found in Bindele and Abebe (2012).

Remark 6. *Under **A2**– **A4**, Lemma 1 gives the pointwise almost sure convergence of* $D_{\boldsymbol{n}}(\boldsymbol{\theta})$ *for any* $\boldsymbol{\theta} \in \Theta$. *If in addition **A9** holds, then we have* $[\mathcal{D}^{\beta}_{\boldsymbol{\theta}} S_{\boldsymbol{n}}(\boldsymbol{\theta})]_{\theta=\boldsymbol{\theta}_0} \to A_{\boldsymbol{\theta}_0}$ *a.s., where* $A_{\boldsymbol{\theta}_0} \equiv E[\varphi^{+}(\xi)[\mathcal{D}^{\beta}_{\boldsymbol{\theta}} \Phi_{\boldsymbol{\theta}}(\xi)]_{\boldsymbol{\theta}=\theta_0}]$ *for any* β *such that* $|\beta| \leq 2$.

The following theorem gives the asymptotic normality of $\widehat{\boldsymbol{\theta}}_{\mathbf{n}}$. After proper scaling, the proof can be formulated along the lines as in Bindele and Abebe (2012) and Bindele (2014). Thus we will focus on the unique aspects of the proof. The interested reader can find the details in the aforementioned papers.

Theorem 2.2. Under assumptions **A1**–**A9**,

$$\sqrt{|\mathbf{n}|}(M_{\mathbf{n}})^{-1}(\widehat{\boldsymbol{\theta}}_{\mathbf{n}} - \boldsymbol{\theta}_0) \to_{\mathcal{D}} \mathcal{N}_k(0, A^{-1}_{\boldsymbol{\theta}_0} \Sigma_{\boldsymbol{\theta}_0} A^{-1}_{\boldsymbol{\theta}_0}),$$

where $\Sigma_{\boldsymbol{\theta}_0} = E[\varphi^{+}(\xi)\Phi_{\boldsymbol{\theta}_0}(\xi)(\Phi_{\boldsymbol{\theta}_0}(\xi))^T]$.

Proof. The argument proceeds by the Taylor expansion at $\boldsymbol{\theta}_0$ of $S_{\mathbf{n}}(\boldsymbol{\theta})$, we get

$$\begin{aligned} 0 = S_{\mathbf{n}}(\widehat{\boldsymbol{\theta}}_{\mathbf{n}}) &= S_{\mathbf{n}}(\boldsymbol{\theta}_0) + \dot{S}_{\mathbf{n}}(\boldsymbol{\theta}_0)(\widehat{\boldsymbol{\theta}}_{\mathbf{n}} - \boldsymbol{\theta}_0) \\ &+ \frac{1}{2}(\widehat{\boldsymbol{\theta}}_{\mathbf{n}} - \boldsymbol{\theta}_0)' \ddot{S}_{\mathbf{n}}(\gamma_{\mathbf{n}})(\widehat{\boldsymbol{\theta}}_{\mathbf{n}} - \boldsymbol{\theta}_0), \end{aligned}$$

where $\boldsymbol{\gamma}_{\mathbf{n}}$ is a point between $\boldsymbol{\theta}_0$ and $\widehat{\boldsymbol{\theta}}_{\mathbf{n}}$,

$$\dot{S}_{\mathbf{n}}(\boldsymbol{\theta}_0) = [\mathcal{D}^{\beta}_{\boldsymbol{\theta}} S_{\mathbf{n}}(\boldsymbol{\theta})]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}, \quad \text{for } |\beta| = 1$$

and

$$\ddot{S}_{\mathbf{n}}(\boldsymbol{\theta}_0) = [\mathcal{D}^{\beta}_{\boldsymbol{\theta}} S_{\mathbf{n}}(\boldsymbol{\theta})]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \quad \text{for } |\beta| = 2.$$

Now using **A5**, **A6**, and Theorem 3.5.4 of Hettmansperger and McKean (2011) with $\mathbf{X}$ replaced by $M_n f'(x, \boldsymbol{\theta}_0)$, we have

$$\sqrt{|\mathbf{n}|} M^{-1}_{\mathbf{n}} S_{\mathbf{n}}(\boldsymbol{\theta}_0) \to N_p\left(0, \Sigma_{\boldsymbol{\theta}_0}\right)$$

in distribution. Also $\dot{S}_{\mathbf{n}}(\boldsymbol{\theta}_0)$ converges almost surely to $A_{\boldsymbol{\theta}_0}$ and hence in probability. By **A7** and Theorem 2.1, $\lim_{|\mathbf{n}|\to\infty} P(\{\boldsymbol{\gamma}_{\mathbf{n}} \in B\}) = 1$, where B is neighborhood containing $\boldsymbol{\theta}_0$. So under the event $\{\boldsymbol{\gamma}_{\mathbf{n}} \in B\}$,

$$\|\ddot{S}_{\mathbf{n}}(\boldsymbol{\gamma}_{\mathbf{n}})\| \le C\frac{1}{|\mathbf{n}|}\sum_{\mathbf{t}\le\mathbf{n}} \psi(\xi_{|\mathbf{t}|}),$$

where C stands for the bound on the score function. The right hand side of the above inequality is bounded in probability by the law of large numbers for $|\mathbf{n}|$ sufficiently large. These and the consistency of $\widehat{\boldsymbol{\theta}}_{\mathbf{n}}$ for $\boldsymbol{\theta}_0$ give

$$\begin{aligned} -S_{\mathbf{n}}(\boldsymbol{\theta}_0) &= \Big[A_{\boldsymbol{\theta}_0} o_p(1) + \frac{1}{2}(\widehat{\boldsymbol{\theta}}_{\mathbf{n}} - \boldsymbol{\theta}_0)O_p(1)\Big](\widehat{\boldsymbol{\theta}}_{\mathbf{n}} - \boldsymbol{\theta}_0) \\ &= (A_{\boldsymbol{\theta}_0} + o_p(1))(\widehat{\boldsymbol{\theta}}_{\mathbf{n}} - \boldsymbol{\theta}_0), \end{aligned}$$

since $(\widehat{\boldsymbol{\theta}}_{\mathbf{n}} - \boldsymbol{\theta}_0)O_p(1) = o_p(1)O_p(1) \to 0$ in probability. Also with probability tending to 1, the matrix $A_{\boldsymbol{\theta}_0} + o_p(1)$ is invertible. Thus

$$\sqrt{|\mathbf{n}|}M_{\mathbf{n}}^{-1}(\widehat{\boldsymbol{\theta}}_{\mathbf{n}} - \boldsymbol{\theta}_0) = -\sqrt{|\mathbf{n}|}M_{\mathbf{n}}^{-1}A_{\boldsymbol{\theta}_0}^{-1}S_{\mathbf{n}}(\boldsymbol{\theta}_0) + o_p(1)\ .$$

Application of Slutsky's lemma, noting that $\sqrt{|\mathbf{n}|}M_{\mathbf{n}}^{-1}S_{\mathbf{n}}(\boldsymbol{\theta}_0)$ is asymptotically normal, completes the proof. □

Remark 7. *Assumption **A5** is the same as assumption A5(i) of Bansal et al. (1999). As discussed in Bansal et al. (1999), if assumption **A5** holds only for a subsequence of* $\{\boldsymbol{n}, \boldsymbol{n} \in \mathbb{N}^k\}$ *such that* $|\boldsymbol{n}| \to \infty$, *then the result of Theorem* 2.2 *holds for that subsequence.*

We will look at a couple of examples to illustrate the above asymptotic normality result.

Example 6. *Let us reconsider the situation in Example 4 and establish conditions for asymptotic normality of the GSR estimator. Let* $\widehat{\theta}_n$ *be a GSR estimator of the true parameter vector*

θ_0. *In order to apply Theorem* 2.2, *we need to verify assumptions* ***A5***– ***A9***. *Since* f *is continuous with continuous derivative* $f_t'(\theta) = -2\pi t \sin(2\pi\theta t)$, *we may take* $\theta = \theta_0$ *and determine* M_n *so that* ***A5*** *is true. Consider*

$$\frac{1}{n}\sum_{t=1}^{n}[f_t'(\theta_0)][f_t'(\theta_0)]^T = \frac{1}{n}\sum_{t=1}^{n}[f_t'(\theta_0)]^2 = \frac{1}{n}\sum_{t=1}^{n}4\pi^2 t^2 \sin^2(2\pi\theta_0 t) = O(n^2)\ .$$

Thus, if we take $M_n = n^{-1}$, *we will have*

$$\frac{1}{n}M_n^T\sum_{t=1}^{n}[f_t'(\theta_0)][f_t'(\theta_0)]^T M_n = \frac{1}{n^3}\sum_{t=1}^{n}4\pi^2 t^2 \sin^2(2\pi\theta_0 t)\ \to\ 2/3$$

as $n \to \infty$. *Thus,* ***A5*** *is satisfied with* $M_n = n^{-1}$ *and* $\Sigma_{\theta_0} = 2/3$. *To verify assumption* ***A6***, *let us take* $\varphi^+ \equiv 1$ *so that we do not deal with complicated derivatives that add nothing to the argument. Since* h_{tt}^n *is the* t^{th} *diagonal element of the hat-matrix* $X^*(X^{*T}X^*)^{-1}X^{*T}$, *where we denote the* $n\times p$ *matrix* X^* *by* $X^* = (\Phi_{\theta_0}(x_{\mathbf{1}}), \ldots, \Phi_{\theta_0}(x_n))$ *and* $\Phi_\theta(t) = D_\theta^\beta \Gamma_\theta(t)$ *for* $|\beta| = 1$, *with* $\Gamma_\theta(t) = \tilde{G}_\theta^{-1}(t)$. *So if* x_{tt} *are the diagonals entries of* X^*, *then* $h_{tt}^n = x_{tt}^2/\sum_{t=1}^n x_{tt}^2$, *where* $x_{tt} = sgn(\epsilon_t) \times 2\pi t \sin(2\pi\theta_0 t)$. *Thus we have that*

$$h_{tt}^n = \frac{4\pi^2 t^2 \sin^2(2\pi\theta_0 t)}{\sum_{t=1}^n 4\pi^2 t^2 \sin^2(2\pi\theta_0 t)} = \frac{t^2 \sin^2(2\pi\theta_0 t)}{\sum_{t=1}^n t^2 \sin^2(2\pi\theta_0 t)}.$$

Therefore,

$$\max_{1\le t\le n} h_{tt}^n = \max_{1\le t\le n} \frac{t^2 \sin^2(2\pi\theta_0 t)}{\sum_{t=1}^n t^2 \sin^2(2\pi\theta_0 t)} \le \frac{n^2}{\sum_{t=1}^n t^2 \sin^2(2\pi\theta_0 t)} \to 0 \quad as \quad n\to\infty,$$

since the denominator is $O(n^3)$ *as we have seen above. We conclude that*

$$\lim_{n\to\infty}\max_{1\le t\le n} h_{tt}^n = 0.$$

*As before, to simplify computational matters while studying assumptions **A7**, **A8**, and **A9**, we will consider the LAD case; that is,* ρ *is the identity function and* $\varphi^+ \equiv 1$. *We observe that, for every* $1 \leq p < \infty$, $|\cos(2\pi\theta t)|^p$, $|2\pi t\sin(2\pi\theta t)|^p$, $|4\pi^2 t^2 \cos(2\pi\theta t)|^p$, *and* $|8\pi^3 t^3 \sin(2\pi\theta t)|^p$ *are all integrable on* $B = (0, 1/2)$, *an open neighborhood of* θ_0 *in* $\mathbb{R}$. *Thus the function* $\theta \to \Gamma_\theta(t)$ *is a map in* $W^{3,p}(B)$ *for every fixed* t*; assumption **A7** is therefore satisfied. Moreover, **A7**, implies that* Φ_θ *is well defined since* Φ_θ *is the weak derivative of a function on* $W^{3,p}(B)$. *Likewise, the weak derivatives of* Φ_θ *is also well defined. Thus,* $A_{\theta_0} = 2G'(\theta_0)$ *as in page* 25 *in Hettmansperger and McKean (2011). Hence assumption **A8** is satisfied. We have that* $|\mathcal{D}_\theta^\beta \Phi_\theta(t)| \leq (2\pi t)^\beta = \psi_\beta(t)$ *for every* $\theta \in B$ *and* $|\beta| \leq 2$, *with* $\psi_\beta \in W^{2,p}(K)$, *where* $K = (0, T]$. *Thus assumption **A9** is also satisfied. Since* $A_{\theta_0}^{-1}\Sigma_{\theta_0}A_{\theta_0}^{-1} = \{6(G'(\theta_0))^2\}^{-1}$, *we can then conclude from Theorem 2.2 that,*

$$n^{3/2}(\widehat{\theta}_n - \theta_0) \text{ converges in distribution to } \mathcal{N}\left(0, \frac{1}{6(G'(\theta_0))^2}\right) \text{ as } n \to \infty\,.$$

Example 7. *We now consider asymptotic normality of the GSR estimator for model* (2.4) *considered in Bansal et al. (1999) which is also given in Example 2. Dealing with model* (2.4)*, we will show that the parameters of the asymptotic normal distribution depend on subsequences of* $\boldsymbol{n}$.

Let $\widehat{\theta}_{\boldsymbol{n}} = (\hat{\alpha}_{\boldsymbol{n}}, \hat{\lambda}_{1\boldsymbol{n}}, \hat{\lambda}_{2\boldsymbol{n}})^T$ *be a GSR estimator of the true parameter vector* $\theta_0 = (\alpha_0, \lambda_{10}, \lambda_{20})^T$. *Let* $M_{\boldsymbol{n}} = diag(1, n_1^{-1}, n_2^{-1})$. *In order to apply Theorem* 2.2*, we need to verify assumptions **A5**–**A9**. A detailed computation verifying Assumption **A5** is given in Bansal et al. (1999) where it was shown that*

$$\Sigma_{\theta_0} = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & \alpha_0^2/6 & \alpha_0^2/8 \\ 0 & \alpha_0^2/8 & \alpha_0^2/6 \end{bmatrix} \quad \text{if } \min(n_1, n_2) \to \infty\,.$$

Two more versions of Σ_{θ_0} *are given in Bansal et al. (1999) corresponding to the cases where* n_1 *is fixed but* $n_2 \to \infty$ *and vice versa. They will not be given here. Also the verification of assumption* ***A6*** *follows the same approach as the one in Example 6 above and for sake of brevity will not be given here. To simplify computational matters while studying assumptions* ***A7****,* ***A8****, and* ***A9****, we will consider the LAD case; that is,* ρ *is the identity function and* $\varphi^+ \equiv 1$*. We observe that, for every* $1 \le p < \infty$*,* $|\cos(\lambda_1 t_1 + \lambda_2 t_2)|^p$*,* $|\sin(\lambda_1 t_1 + \lambda_2 t_2)|^p$*,* $|\alpha t_i \sin(\lambda_1 t_1 + \lambda_2 t_2)|^p$*,* $|\alpha t_i^2 \cos(\lambda_1 t_1 + \lambda_2 t_2)|^p$*, and* $|\alpha t_i^3 \sin(\lambda_1 t_1 + \lambda_2 t_2)|^p$ *are all integrable on* $B = (m, M) \times (-\pi, \pi) \times (0, \pi)$*, an open neighborhood of* θ_0 *in* $\mathbb{R}^3$ *and for all* $i = 1, 2$*. Thus the function* $\theta \to \Gamma_\theta(t)$ *is a map in* $W^{3,p}(B)$ *for every fixed* $\boldsymbol{t}$*; assumption* ***A7*** *is therefore satisfied. As in the previous example* $A_{\theta_0} = 2G'(\theta_0)$ *which verifies assumption* ***A8****. Moreover,* $|\mathcal{D}_\theta^\beta \Phi_\theta(t)| \le \alpha(t)^\beta = \psi_\beta(t)$ *for every* $\theta \in B$ *and* $|\beta| \le 2$*, with* $\psi_\beta \in W^{2,p}(K)$*, where* K *is a compact subset of* $\mathbb{R}$*. Thus assumption* ***A9*** *is also satisfied. Thus we can conclude from Theorem 2.2 that,* $\sqrt{n_1 n_2}\Big((\hat{\alpha} - \alpha_0), n_1(\hat{\lambda} - \lambda_{10}), n_2(\hat{\lambda}_2 - \lambda_{20})\Big)$ *converges in distribution to* $\mathcal{N}_3(0, A_{\theta_0}^{-1} \Sigma_{\theta_0} A_{\theta_0}^{-1})$ *as either* $\min(n_1, n_2) \to \infty$ *or* $n_2 \to \infty$ *while* n_1 *is held fixed, or* $n_1 \to \infty$ *while* n_2 *is held fixed, where* Σ_{θ_0} *is defined as above.*

2.5 Simulation Study

In this section, we provide results of some numerical experiments based on simulation to see how the proposed estimator works for finite samples. We use the same simulation settings as in Kundu and Nandi (2003) where they used the two dimensional model with stationary random field errors

$$y(m, n) = A\cos(m\lambda + n\mu) + B\sin(m\lambda + n\mu) + X(m, n) \ ,$$

where $A_k, B_k \in \mathbb{R}$, $\lambda_k, \mu_k \in (0, \pi)$ are unknown and $X(m,n)$ is a random field. In the simulation, $X(m,n)$ is generated as

$$X(m,n) = e(m,n) + 0.25e(m-1,n) + 0.25e(m+1,n) + 0.25e(m,n-1) + 0.25e(m,n+1),$$

where $e(m,n)$ are i.i.d. from normal, logistic, and Cauchy random variables with location zero and scale $\sigma = 0.25$ and also the Student t random variable with degrees of freedom 2 and 5. We used $m, n = 1, \ldots, 40$ and the true parameter value $\theta = (A, B, \lambda, \mu) = (4, 4, 1.886, 1.1)$. We replicated the procedure 1000 times and calculated the average the mean squared error (MSE) of the least squares (LS), signed-rank (SR), and least absolute deviations (LAD) estimators of the unknown parameters over these replications. These are obtained by taking $[\phi^+(u) = 1, \rho(t) = t^2]$, $[\phi^+(u) = u, \rho(t) = t]$, and $[\phi^+(u) = 1, \rho(t) = t]$ in (2.6), respectively. The results are reported in the tables below for different values of σ and using the student t-distribution with different degrees of freedom 2 and 5, and where N, L, and C stand for Normal, Logistic and Cauchy distribution (Dist.) respectively.

Scale	Dist		LS				SR				LAD			
			A	B	λ	μ	A	B	λ	μ	A	B	λ	μ
	N	Est	4.001	4.000	1.886	1.100	4.001	3.999	1.886	1.100	4.001	3.999	1.886	1.100
		MSE	3.803e-4	3.640e-4	2.187e-8	2.147e-8	4.092e-4	3.873e-4	2.294e-8	2.276e-8	5.792e-4	5.394e-4	3.340e-8	3.323e-8
$\sigma = .25$	L	Est	3.999	4.001	1.886	1.100	3.999	4.001	1.886	1.100	4.000	4.000	1.886	1.100
		MSE	1.247e-3	1.153e-3	6.920e-8	6.930e-8	1.122e-3	1.048e-3	6.501e-8	6.387e-8	1.387e-3	1.307e-3	8.575e-8	8.864e-8
	C	Est	3.984	3.847	1.787	1.123	3.993	3.993	1.885	1.100	3.995	3.991	1.885	1.100
		MSE	65.777	47.724	0.157	0.106	2.245e-2	4.275e-2	3.002e-4	2.303e-7	1.660e-2	4.321e-2	2.994e-4	8.084e-7
	N	Est	3.998	4.001	1.886	1.100	3.999	4.001	1.886	1.100	3.999	4.000	1.886	1.100
		MSE	1.414e-3	1.489e-3	8.621e-8	8.281e-8	1.425e-3	1.496e-3	8.899e-8	8.559e-8	1.831e-3	2.029e-3	1.222e-7	1.241e-7
$\sigma = .5$	L	Est	4.000	3.998	1.886	1.100	3.999	3.998	1.886	1.100	3.997	4.001	1.886	1.100
		MSE	4.918e-3	4.498e-3	2.790e-7	2.629e-7	4.388e-3	3.966e-3	2.583e-7	2.393e-7	5.359e-3	5.141e-3	3.457e-7	3.148e-7
	C	Est	5.536	6.288	1.737	1.126	3.982	3.975	1.884	1.100	3.987	3.981	1.884	1.100
		MSE	1616.191	1754.342	0.390	0.156	5.841e-2	1.116e-1	7.411e-4	7.890e-7	4.814e-2	7.004e-2	8.007e-4	9.756e-5
	N	Est	3.999	4.001	1.886	1.100	4.000	4.000	1.886	1.100	4.000	4.000	1.886	1.100
		MSE	6.594e-3	5.860e-3	3.473e-7	3.174e-7	6.392e-3	5.740e-3	3.410e-7	3.259e-7	8.495e-3	8.426e-3	5.333e-7	4.891e-7
$\sigma = 1$	L	Est	3.998	3.995	1.886	1.100	4.001	3.993	1.886	1.100	3.997	3.994	1.886	1.100
		MSE	3.153e-2	4.091e-2	1.476e-4	1.201e-6	2.664e-2	3.673e-2	1.470e-4	1.041e-6	2.817e-2	4.342e-2	1.486e-4	1.504e-6
	C	Est	7.262	7.901	1.737	1.085	3.906	3.882	1.874	1.101	3.953	3.933	1.880	1.100
		MSE	1287.652	2893.719	0.400	0.277	0.310	0.480	6.030e-3	3.271e-4	0.160	0.290	2.759e-3	2.489e-5
	t_2	Est	3.960	3.927	1.880	1.101	3.994	3.993	1.886	1.100	3.995	3.999	1.886	1.100
		MSE	0.255	0.304	3.633e-3	1.211e-3	2.238e-2	3.367e-2	1.463e-4	7.075e-7	1.512e-2	1.385e-2	8.930e-7	8.939e-7
	t_5	Est	3.996	4.003	1.886	1.100	3.996	4.004	1.886	1.100	3.993	4.008	1.886	1.100
		MSE	1.009e-2	9.315e-3	5.664e-7	5.583e-7	7.464e-3	6.881e-3	4.601e-7	4.409e-7	9.740e-3	9.207e-3	6.541e-7	5.620e-7

Some of the points are very clear from the entries in the table. It is observed that as σ increases the MSEs and biases of all methods increase. For our method we also observed that biases are quite small and when $\sigma = 1$ the SRs performs better than the LS and the LAD. Looking at the Student t distribution we can see SR performs better than LS for both cases considered and better than the LAD for 2 degree of freedom. The above table tells us also that based on the 1000 repetitions, under errors from the Normal, Logistic, and Cauchy

Table 2.1: Estimated relative efficiencies versus LS

Method	Distribution	A	B	λ	μ
LAD	Normal	0.656	0.675	0.655	0.646
	Logistic	0.899	0.882	0.807	0.782
	t_5	1.036	1.012	0.866	0.993
SR	Normal	0.929	0.940	0.953	0.943
	Logistic	1.112	1.100	1.064	1.085
	t_5	1.352	1.354	1.231	1.266

distribution, LS has larger spread (in terms of the MSE) and bias than the SR while for the Student t distribution errors with degrees of freedom 5, it appears that for very few of the amplitude, the SR has much larger variability bias than the LS. We also notice that for heavy-tailed distributions, the efficiency gain by the SR is considerable.

In Table 2.1 we report the estimated relative efficiencies versus LS of SR and LAD estimators. These are calculated by taking the ratio of the LS MSE to SR and LAD MSEs, respectively, obtained from our simulation experiment. For brevity, we only report the results for the normal, logistic, and t_5 distributions. However, we note that LS performs very poorly in the case of the heavy-tailed Cauchy and t_2 distributions as expected. Following our discussion, we will report bias results related to Cauchy and t_2 in Figure 2.5.

The relative efficiency results reported in Table 2.1 are close to the theoretical asymptotic relative efficiency (ARE) results for the normal and logistic distributions for the simple location problem under iid errors. The AREs of SR to LS under normal, logistic, and t_5 error distributions are 0.955, 1.096, and 1.241, respectively. On the other hand, the AREs of LAD to LS under normal, logistic, and t_5 error distributions are 0.637, 0.822, and 0.961, respectively. For more on the ARE, please see Hettmansperger and McKean (2011).

As mentioned above, we also studied the absolute bias in estimated response. For this, we let $\tilde{\theta}$ be the average estimated parameter over the 1000 replications. Let $\tilde{y}$ be the estimated value of y at the average. The plots given in Figure 2.5 give values of $|y - \tilde{y}|$ for normal, Cauchy, and t_2 distributions. A dark shade indicates small bias and light shade indicated large biases. It is clear that SR and LAD recover the signal much more accurately, on average, than LS when the noise distribution is heavy-tailed.

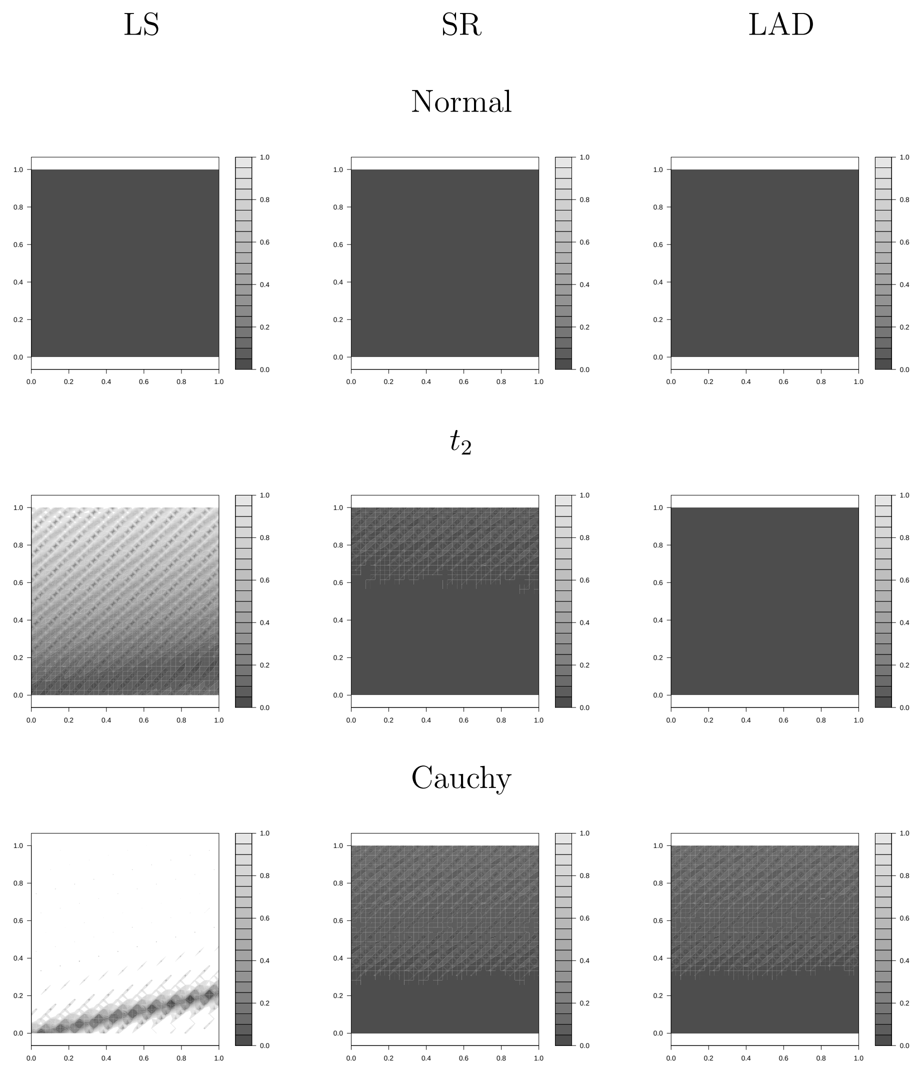

Figure 2.5: Residual Plots

We can also compare the estimates using their Density Profiling Plot (DPP). The DPP displays a two-dimensional graph of the intensities of the pixels along a rectangular section. The x-axis represents the distance along the lines and the y-axis represents the pixels intensity. We use ImageJ (see [23]) to obtain the density profiling plots (DPP). Figure 2.6 below are respectively the LS, LAD, and SR corresponding DPP. From these DPPs, it is clear the the SR method produces a better pixel intensity, particularly after the a distance of 400 pixels than the LS and the LAD methods. Therefore, we can infer that our proposed method SR extracts better signal than the other two. Other methods for comparing these images exist such as measuring sharpness, the noise, the contrast, distortion and resolution. For brevity we choose to compare them using DPP.

2.6 Conclusion

In this chapter we considered the multidimensional indices nonlinear model and for application purposes we considered the general two dimensional frequency model which was originally discussed in Zhang and Mandrekar (2001) and Kundu (1993). We considered the generalized signed-rank (GSR) estimation of the unknown parameters under the assumption of additive stationary errors. Our approach is quite general including least squares and least absolute deviations methods and our assumptions are different from those of Kundu (1993) and Zhang and Mandrekar (2001) and they are generalizations to the multidimensional case of the assumptions in Bindele and Abebe (2012). We observe through our simulation that small sample efficiency results closely mimic the asymptotic efficiency of the proposed estimator. Our results can be considered generalizations of those in Bansal et al. (1999); Kundu (1993). Note that we have not considered the estimation of the number of components present in superimposed sinusoidal signals. In practice, this is an important problem can be found by studying the plot of the periodogram function as discussed in Kundu and Nandi (2003).

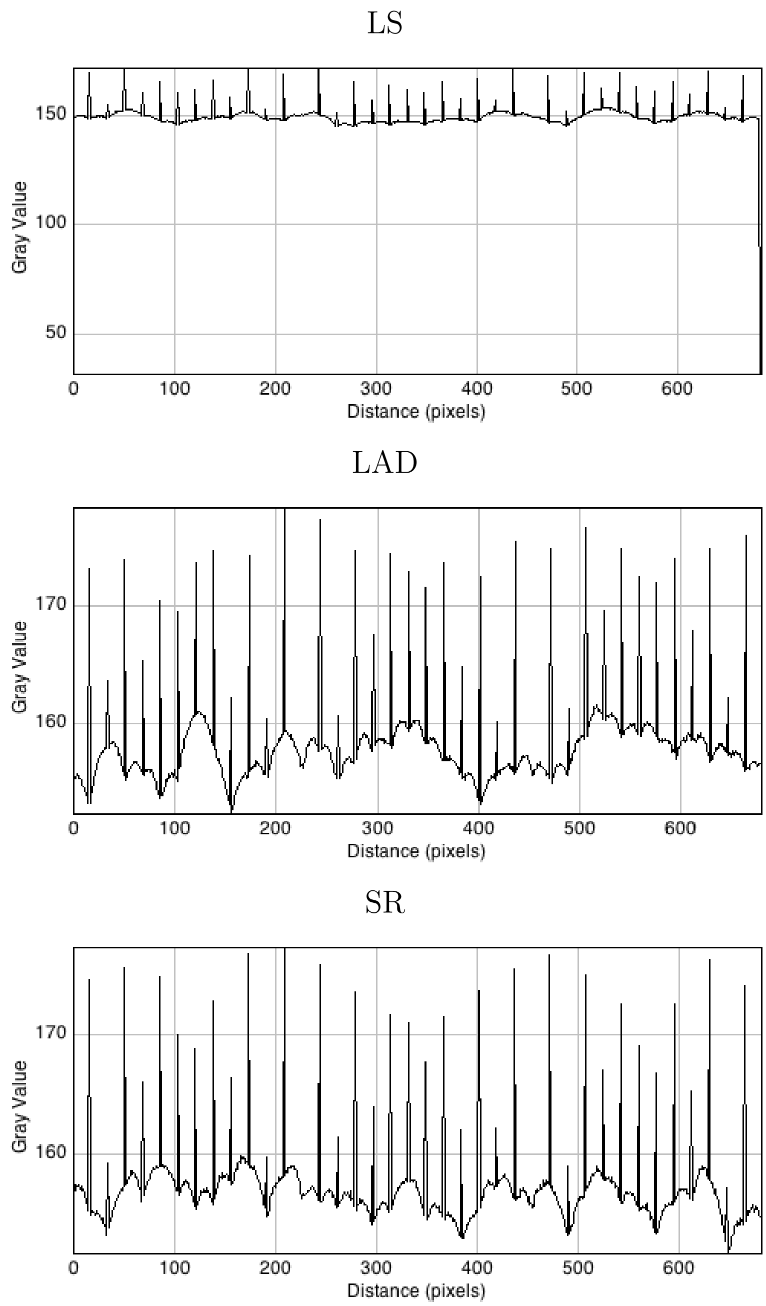

Figure 2.6: LS, LAD, and SR Density Profiling Plots

Chapter 3

Signed-Rank Estimator for Two-Phase Linear Models

3.1 Introduction

A two-phase linear regression model is piecewise linear over different domains of the design variable, each segment possibly being a different straight line. There are two types of such models, called *restricted* and *unrestricted.* In the *restricted* case, the regression function is continuous at the change-point but not differentiable while in the *unrestricted* case, it is discontinuous. The statistical inference in such models is heavily influenced by the continuity or discontinuity of the regression function at the change-point. Most of the inference about the change-point has been developed when the regression function is continuous and the design is *non-random*: Hudson (1996) gave a concise method for calculating the least-squares estimator (LS) for the change-point while Hinkley (1969) and Hinkley (1971) derived asymptotic results for maximum likelihood estimate (MLE) of the point of intersection for the special case of two line segments under normally distributed errors. Under continuity and suitable identifiability assumptions, Feder (1974a) derived the asymptotic distributions of the LS and the log likelihood ratio statistic for the two-phase non-random design regression model with the Gaussian errors. Schulze (1987) provides a collection of existing methods mainly focusing on the least-squares estimation, testing of hypotheses and testing of model stability for analyzing data using multiphase regression models.

Examples of important applications of these models in various scientific fields are discussed by numerous researchers. Anderson and Nelson (1975) used a special type of the restricted case of a two-phase regression model Sprent (1961), called linear-plateau model, to predict crop yield based on the amount of nitrogen in the soil.Eubank (1984) gave examples of a variety of applications where the regression function is difficult or impossible to specify, but can be approximated by simpler segmented models. Some other important examples are listed in paper of M*ü*ller and Stadtm*ü*ller (1999) and references therein.

The review paper of Bhattacharya (1994) discusses various aspects of change-point analysis including testing of hypothesis of no change, interval and point estimation of a change-point, changes in non-parametric models, changes in regression, and detection of a change in distribution of sequentially observed data. In particular, he discusses MLE of the change-point of a discontinuous two-phase linear regression model and the limiting behavior of the log-likelihood ratio process when the errors are Gaussian. In both problems the jump size at the change-point is assumed to tend to zero at the rate slower than $n^{-1/2}$ as the sample size n tends to infinity. See also Cs*ö*rgo and Horv*á*th (1988), van de Geer (1988) and Bhattacharya (1990) for many other results in this case. Most of the above literature deals with the parametric setup. M*ü*ller (1992), Wu and Chu (1993), Loader (1996) and M*ü*ller and Song (1997), among others, use non-parametric curve estimation methods to construct estimates of the change-point in non-random design regression models. Although M-type robust estimators exist for such models; to our knowledge, there is no result available on the limiting distribution of the Signed-Rank Estimator of the change-point or on the limiting behavior of the dispersion function for a two-phase *random* design linear regression model with a *fixed jump size* at the change-point.

We obtain the consistency and the limiting distribution of the the Signed-Rank Estimator of the underlying parameters in the discontinuous case with fixed jump size of the regression function in the two-phase linear regression models with random designs variables and general error distributions. The SR-estimator $\hat{r}_n$ of the change-point parameter r is shown to be n-consistent and the underlying dispersion function, as a process in the standardized change-point parameter, is shown to converge weakly to a compound Poisson process. This process attains maximum over a bounded random interval and as a result $n(\hat{r}_n - r)$ converges weakly to the left end point of this interval. These findings are thus different from those in the case when the jump size tends to zero as $n \to \infty$.

The rest of this is section is organized as follows. Section 3.2 describes the model and and a computational scheme for SR-estimators. Section 3.3 proves the consistency of the SR-estimator and n-consistency of the change-point estimator. Section 3.4 derives the limiting properties of the coefficient parameter estimator and the limiting distribution of the change-point estimator. Section 3.5 reports a simulation study and applications.

3.2 Model Definition and Signed-Rank Estimation

Define for $j = 1, 2$ and for $\theta_j = (\theta_j^0, \theta_j^1) \in \mathbb{R}^2$ the linear function $f_{\theta_j}(x) = \theta_j^0 + \theta_j^1 x$. We consider the following model of two-phase linear regression function

$$k_\theta(x) = f_{\theta_1}(x) \cdot \mathbb{I}_{[-\infty, r]}(x) + f_{\theta_2}(x) \cdot \mathbb{I}_{(r,\infty]}(x),$$

where $x \in \mathbb{R}$, $\theta := (\theta^\star, r) = (\theta_1, \theta_2, r) \in \Theta = K \times \overline{\mathbb{R}}$, for a compact set K of $\mathbb{R}^4$, and $\overline{\mathbb{R}}$ represents the compactification of the real line. Here $\mathbb{I}_A$ represents the characteristic function of the set A. For a set of independent observations (X_i, Y_i), $i = 1, 2, \ldots, n$, and for some unknown $\theta \in \mathbb{R}^5$, we let

$$Y_i = k_\theta(X_i) + e_i, \qquad i = 1, 2, \ldots, n \tag{3.1}$$

where the e_i, $i = 1, 2, \ldots, n$ are independent identically distributed (i.i.d.) random variables.

Remark 8. *Note that model* (3.1) *has been considered by Koul and Qian (2002), Koul et al. (2003), and Ciuperca (2008) for nonlinear functions $f_{\theta_i}(x)$ and multiple-phase changes. We will assume that the two line segments are different and that*

$$d \equiv \theta_2^0 - \theta_1^0 + r(\theta_2^1 - \theta_2^1) \textit{ is fixed and non-zero.} \tag{3.2}$$

This identifiability condition will imply that for all regression parameters, the function $k_\theta(x)$ is not a continuous function of r at the true break point. This essentially means that the function $k_\theta(x)$ is either lower-semicontinuous or upper-semicontinuous at the true beak point. In the case of lower-semicontinuity at the true beak point, we have one of the two cases in Figure 3.1 below. We only need to design methods that minimize lower-semicontinuous

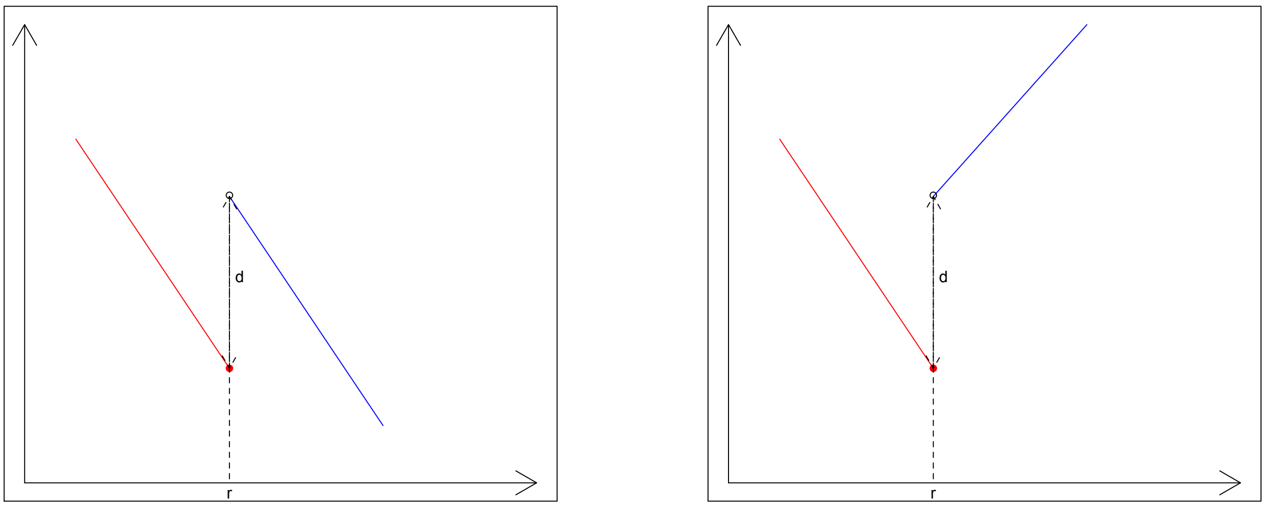

Figure 3.1: function $k_\theta(x)$: d=vertical jump, r=break point

dispersion functions at the true break point since, in the case of upper-semicontinuity, we can use the fact that for a function f

$$\limsup(-f) = -\liminf f.$$

Let $X = (X_1, X_2, \cdots, X_n)$ and $Y = (Y_1, Y_2 \cdots, Y_n)$. We define for $1 \leq i \leq n$, $e_i(\theta) = Y_i - k_\theta(X_i)$ and let $e(\theta) := Y - k_\theta(X)$. The dispersion function D_n is then defined as

$$D_n(\theta) := \frac{1}{n} \sum_{i=1}^{n} a_n(R(|e_i(\theta)|))|e_i(\theta)|, \quad \theta \in \Theta, \tag{3.3}$$

where $R(|e_i(\theta)|)$ is the rank of $|e_i(\theta)|$ among $|e_1(\theta)|, \ldots, |e_n(\theta)|$. The numbers $a_n(i)$ are scores generated as $a_n(i) = \varphi(i/(n+1))$, for some bounded and nondecreasing score function φ :

$(0,1) \to \mathbb{R}^+$ that has at most a finite number of discontinuities. An estimator $\hat{\theta}_n = \hat{\theta}_n(X,Y)$ is called a signed rank estimator for θ if $\hat{\theta}_n = \arg\min_{\theta\in\Theta} D_n(\theta), \quad a.s.$

Remark 9. *We note, following McKean and Schrader (1980), that the expression* $nD_n(e) = \sum_{i=1}^n a_n(R(|e_i|))|e_i|$ *defines a pseudo-norm on* $\mathbb{R}$.

Remark 10. *Since* $D_n(\theta)$ *is piecewise linear and monotone, so there is an interval around the point* r *such that* $D_n(\theta)$ *is monotone continuous on that interval. Thus* $D_n(\theta)$ *is lower semicontinuous on that interval.*

Remark 11. *Because* $\Theta\times\mathbb{X}$ *is compact and from Remark 10,* $D_n(\theta)$ *is lower semicontinuous. It results from Theorem 10.19 in Stakgold and Holst (2011) that a minimizer of* $D_n(\theta)$ *exists.*

To construct the signed-rank estimator, we will use the same approach as in Koul et al. (2003), and Ciuperca (2008): for a given $r \in \mathbb{R}$, we find $\hat{\theta}_n^\star(r) = \arg\min_{(\theta_1,\theta_2)\in K} D_n(\theta)$; since by Remark 8, the change-point is fixed, the estimator $\hat{\theta}_n^\star(r)$ is constant in r over any interval of two consecutive ordered $X_i's$. At the second stage, compute the minimizer $\hat{r}_n := \arg\min_{r\in\mathbb{X}} D_n(\hat{\theta}_n^\star(r))$, where $\mathbb{X} = \{X_{(i)}, 1 \le i \le n\}$. Then the estimator $\hat{\theta}_n = (\hat{\theta}_n^\star(\hat{r}_n), \hat{r}_n)$ is the signed-rank estimator of the underlying parameter θ.

3.3 Consistency

To begin with we shall state the needed assumptions: the $z_i = (x_i, y_i)$ are i.i.d. random vectors such that x_i and $e_i = y_i - k_\theta(x_i)$ are independent random variables with distributions G and F respectively, and denote the distribution of z_i by K. We also assume the following.

B1 : The score-generating function φ is non-negative, non-decreasing and bounded with a finite number of discontinuities.

B2 : F has a density f that is even and strictly decreasing for positive values of e and $E_F[|e|^r] < \infty$ for some $r \ge 1$.

B3 : G has Lebesgue density g that is continuous and positive at r and $E[X^2] < \infty$.

B4 : $P\left(k_\theta(X) = k_{\theta_0}(X)\right) < 1 \quad \text{for any} \quad \theta \neq \theta_0$.

B5 : For $1 \leq \alpha \leq \infty$, assume there exist a function h such that $|F_\theta^{-1}(y)| \leq h(y), \forall \theta \in \Theta$ with $E[h^\alpha(Y)] < \infty$.

Comments on the assumptions

Assumption **B1** is a classical assumption in rank-estimation theory meant to assure that dispersion function defined by Equation (3.3) is a pseudo-norm.
Assumption **B2** admits a wide variety of error distributions, examples of which include the normal, the double exponential, and the Cauchy with location parameter equal to 0 which we are going to explore in Section 3.5.
Assumption **B4** is needed for θ_0 to be identified.

Since $\|\varphi\|_\beta < \infty$ for β such that $1/\alpha + 1/\beta = 1$, Assumption **B5** puts h and φ in conjugate spaces when $\beta \in (1, \infty)$. Holder's inequality ensures that the product $\varphi\tilde{F}_\theta^{-1}$ is integrable, where $\tilde{F}_\theta$ is the distribution function of $|e_i(\theta)|$.

Remark 12. *Under assumptions **B1**–**B4**, we can use a similar strategy as in Nguelifack et al. (2015) to show that for any $t > 0$ and $\theta \neq \theta_0$,*

$$\tilde{F}_\theta(t) = P_K(|e - k_\theta(X)| \leq t) = E_G P_F(|e - k_\theta(X)| \leq t|X) < E_G P_F(|e| \leq t) = \tilde{F}_{\theta_0}(t).$$

We let $\|\cdot\|$ denote a norm on $\mathbb{R}^4$. For $\gamma = (\gamma^\star, s) \in \Theta$, we define, for a positive real δ, a compact neighborhood of γ in Θ by

$$N_\delta(\gamma) = \{\theta = (\theta^\star, r) \in \Theta : \|\theta^\star - \gamma^\star\| \geq \delta, |r - s| \geq \delta\}.$$

Moreover, the dispersion function in (3.3) can be re-written as

$$D_n(\theta) := \frac{1}{n}\sum_{i=1}^{n} a_n(i)\tilde{F}_\theta^{-1}(\xi_{(i)}), \quad \theta \in \Theta, \tag{3.4}$$

where $\xi_{(i)}$ are order statistics from the uniform $U(0,1)$ distribution, and $\tilde{F}_\theta^{-1}(t) = \inf\{y : \tilde{F}_\theta(y) \geq t\}$ is the quantile function corresponding to the distribution function $\tilde{F}_\theta$. Note that the two notations of $D_n(\theta)$ are equivalent since the joint distribution of the random vectors $|e(\theta)|_{(i)}$ and $(\tilde{F}_\theta^{-1}(\xi_{(i)})),\ 1 \leq i \leq n$ coincide, where $|e(\theta)|_{(i)}$ represents the i-th ordered value among $|e_1(\theta)|, |e_2(\theta)|, \cdots, |e_n(\theta)|$.

We define for $i \leq n$ and $\theta \in \Theta$ the quantities

$$\mu_n(\theta) = \int_{(i-1)/n}^{i/n} \varphi(t/(n+1))\tilde{F}_\theta^{-1}(\xi_{(t)})dt$$

and

$$\mu(\theta) = \int_0^1 \varphi(u)\tilde{F}_\theta^{-1}(u)du$$

Consistency of the signed-rank estimator is given by the following results.

Theorem 3.1. *Given the model* (3.1), *suppose equation* (3.2) *holds, and that assumptions* ***B1***–***B5*** *hold, then* $\hat{\theta}_n$ *is a strongly consistent estimator of* θ_0. *That is,*

$$\hat{\theta}_n \to \theta_0, \quad \text{a.s. as } n \to \infty.$$

To establish the proof of this theorem, we need the following definitions and results:

Definition 2. *Let* $d \geq 1$ *be a positive integer. A family of mollifiers* $\{\psi_\epsilon : \mathbb{R}^d \to \mathbb{R}_+, \epsilon \in \mathbb{R}_+\}$ *is family of functions satisfying*

1. $\int_{\mathbb{R}^d} \psi_\epsilon(z)dz = 1.$
2. *supp*$\psi_\epsilon := \{z \in \mathbb{R}^d : \psi_\epsilon(z) > 0\} \subset B_{\rho_\epsilon}(0)$, *where* $\rho_\epsilon \to 0$ *as* $\epsilon \to 0$, *and* $B_{\rho_\epsilon}(0)$ *is the ball of* $\mathbb{R}^d$ *centered at* 0 *with radius* ρ_ϵ.

Definition 3. *For a locally integrable function $f : \mathbb{R}^d \to \mathbb{R}$, we define the family $\{f_\epsilon : \epsilon \in \mathbb{R}_+\}$ of averaged (or convolution) functions by*

$$f_\epsilon(x) := \int_{\mathbb{R}^d} f(x-z)\psi_\epsilon(z)dz = \int_{\mathbb{R}^d} f(z)\psi_\epsilon(x-z)dz.$$

Definition 4. *A sequence $f_n : \mathbb{R}^d \to \mathbb{R}$ of functions is said to epi-converge to a function f at x if the following are satisfied:*

1. $\liminf_{n\to\infty} f_n(x_n) \geq f(x)$ *for all $x_n \to x$ as $n \to \infty$.*
2. $\liminf_{n\to\infty} f_n(x_n) = f(x)$ *for some sequence $x_n \to x$ as $n \to \infty$.*

Remark 13. *It is well known that, if f is lower semi-continuous at x, then there is sequence f_n of average functions that epi-converge to f.(see Theorem 3.7 in Ermoliev et al. (1995))*

Proof of Theorem 3.1 .

We note that $k_\theta(X)$ as a function of r is not continuous at r because of the identifiability condition (3.2). But without loss of generality, we will assume that $k_\theta(X)$ is lower semi-continuous at r.

Thus we can conclude from Remark 10 that, $D_n(\theta) = D_n(\theta^\star, s)$ is lower semicontinuous at the true break point r, Hence from Remark 13, there is a sequence $D_{n,k} := D_{n,\psi_k}$ of average functions that epi-converges to D_n (as $k \to \infty$) obtained by convolution with a sequence of smooth mollifiers $\{\psi_k := \psi_{\epsilon_k} : \mathbb{R} \to \mathbb{R}\}$, where $\epsilon_k \to 0$ as $k \to \infty$. That is, there is a sequence $s_k \to s$ such that $\liminf_{k\to\infty} \tilde{D}_{n,k}(\theta^\star, s_k) \geq \tilde{D}_n(\theta^\star, s)$. Let θ_0 be a point in Θ and consider a compact neighborhood $N_\delta(\theta_0)$ of θ_0 in Θ. Given $\theta \in N_\delta(\theta_0)$. For $n \geq 1$ we write

$$D_n(\theta) - D_n(\theta_0) = \underbrace{[D_n(\theta) - D_{n,k}(\theta)]}_{A_{n,k}(\theta)} + \underbrace{[D_{n,k}(\theta) - D_{n,k}(\theta_0)]}_{B_{n,k}(,\theta,\theta_0)} + \underbrace{[D_{n,k}(\theta_0) - D_n(\theta_0)]}_{-A_{n,k}(\theta_0)}, \quad (3.5)$$

By definition of epi-convergence, we have

$$\liminf_{k\to 0} A_{n,k}(\theta) = 0, \quad \liminf_{k\to\infty} [-A_{n,k}(\theta_0)] \geq 0.$$

Since by mollifying $D_n(\theta)$ by $D_{n,k}(\theta)$ at r, we obtain a smooth function of r, we conclude using Theorem 11 in Nguelifack et al. (2015) that, for each $k \geq 1$,

$$\liminf_{n\to\infty} \lim \inf_{\theta\in N_\delta(\theta_0)} [D_{n,k}(\theta) - D_{n,k}(\theta_0)] > 0.$$

Taking the lim inf three times on the right hand side of Equation (3.5) respectively as $k \to \infty$, as $\theta \in N_\delta(\theta_0)$, and as $n \to \infty$, we have

$$\liminf_{n\to\infty} \lim \inf_{\theta\in N_\delta(\theta_0)} [D_n(\theta) - D_n(\theta_0)] > 0.$$

□

The next result gives the $n-$ and $n^{1/2}-$consistency of the estimators $\hat{r}_n$ and $\hat{\theta}_n^*$ respectively.

Theorem 3.2. *Given the model* (3.1), *suppose equation* (1.1) *holds, and that assumptions* ***B1***–***B5*** *hold, then*

(i) $\|\sqrt{n}(\hat{\theta}_n^* - \theta_0^*)\| = O_p(1)$,

(ii) $|n(\hat{r}_n - r)| = O_p(1)$.

The prove of Theorem 3.2 (i), is just a consequence of the following preliminaries and follows along with some results in Hettmansperger and McKean (2011) and Koul et al. (2003). We start with the following notation that we will use throughout the rest of the paper. It is convenient to write $k_\theta(x) = k_s(x, \theta^*)$, for $\theta = (\theta^*, s)$ with $\theta^* \in \mathbb{R}^4$, $s \in \mathbb{R}$, and refer to θ^* and s as the coefficient and change point parameter, respectively. We remark that for a fix jump point, $s \in \mathbb{R}$, we have that

$$\dot{k}_s(x) \equiv (\partial/\partial\theta^*)(k_s(x, \theta^*)) = (I(x \leq s), xI(x \leq s), I(x > s), xI(x > s))^T,$$

with $s \in \mathbb{R}$ and $x \in \mathbb{R}$, denote the vector of partial derivatives of $k_s(x, \theta^*)$ with respect to θ^*. It is clear that

$$k_\theta(x) = \theta^{*T} \dot{k}_s(x), \quad \text{and} \quad \|\dot{k}_s(x)\| = \sqrt{1 + x^2} \leq 1 + |x|. \tag{3.6}$$

Thus (3.3) can be written as: For a fix s,

$$D_n^s(\theta^*) := \frac{1}{n} \sum_{i=1}^{n} \varphi\Big(\frac{R_i}{n+1}\Big) |e_i(\theta^*)|, \tag{3.7}$$

where $e_i = Y_i - \theta^{*T} \dot{k}_s(x)$, φ is the score generating function, and R_i is the rank of $|e_i(\theta^*)|$ among $|e_1(\theta^*)|, \ldots, |e_n(\theta^*)|$.

We define the gradient process $S_n(\theta^*)$ of $D_n^s(\theta^*)$, as $S_n(\theta^*) = - \bigtriangledown_{\theta^*} D_n^s(\theta^*)$. Thus using (3.7), we have that,

$$S_n(\theta^*) = \frac{1}{n} \sum_{i=1}^{n} \varphi\Big(\frac{R_i}{n+1}\Big) \dot{k}_s(x) sgn(e_i(\theta^*)). \tag{3.8}$$

Hence

$$S_n(\theta^*) - S_n(\theta_0^*) = \frac{1}{n} \sum_{i=1}^{n} \dot{k}_s(X_i) \left\{ \varphi\Big(\frac{R_i}{n+1}\Big) sgn(e_i(\theta^*)) - \varphi\Big(\frac{R_i^*}{n+1}\Big) sgn(e_i(\theta_0^*)) \right\} \tag{3.9}$$

But since φ is continuous with continuous derivatives, assuming without loss of generality that $R_i^* > R_i$, using the mean value theorem, there is $\alpha_i \in (R_i, R_i^*)$ such that,

$$\varphi\Big(\frac{R_i}{n+1}\Big) - \varphi\Big(\frac{R_i^*}{n+1}\Big) = \dot{\varphi}(\alpha_i) \Big(\frac{R_i - R_i^*}{n+1}\Big). \tag{3.10}$$

Using (3.9) and (3.10) we obtain:

$$\begin{aligned} n^{-1/2}\big(S_n(\theta^*) - S_n(\theta_0^*)\big) &= \frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\dot{k}_s(X_i)\varphi\Big(\frac{R_i}{n+1}\Big)\left\{sgn(e_i(\theta^*)) - sgn(e_i(\theta_0^*))\right\} \\ &\quad + \frac{1}{n(n+1)\sqrt{n}}\sum_{i=1}^{n}\dot{k}_s(X_i)\dot{\varphi}(\alpha_i)\big(R_i - R_i^*\big)sgn(e_i(\theta_0^*)) \\ &= T_{1n} + T_{2n}, \end{aligned} \tag{3.11}$$

where

$$\begin{aligned} T_{1n} &\equiv \frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\dot{k}_s(X_i)\varphi\Big(\frac{R_i}{n+1}\Big)\left\{sgn(e_i(\theta^*)) - sgn(e_i(\theta_0^*))\right\} \\ T_{2n} &\equiv \frac{1}{n(n+1)\sqrt{n}}\sum_{i=1}^{n}\dot{k}_s(X_i)\dot{\varphi}(\alpha_i)\big(R_i - R_i^*\big)sgn(e_i(\theta_0^*)) \end{aligned}$$

Using the triangle inequality on (3.11), we obtain that:

$$\sup_{\theta^*\in N_\delta(\gamma)} |n^{-1/2}\big(S_n(\theta^*) - S_n(\theta_0^*)\big)| \leq \sup_{\theta^*\in N_\delta(\gamma)} |T_{1n}| + \sup_{\theta^*\in N_\delta(\gamma)} |T_{2n}|. \tag{3.12}$$

For the first term on the right hand side of (3.11), Using the Strong Law of Large Numbers for functions of order statistics (see Hajek and Sidak, 1967), we have that as $n \to \infty$,

$$T_{1n} \to \begin{cases} 0, & \text{if} \quad sgn(e_i(\theta^*)) = sgn(e_i(\theta_0^*)) \\ \pm 2E[S_n(\theta_0^*)] = 0, & \text{if} \quad sgn(e_i(\theta^*)) = -sgn(e_i(\theta_0^*)) \end{cases}$$

Hence

$$\sup_{\theta^*\in N_\delta(\gamma)} |T_{1n}| = O_p(1). \tag{3.13}$$

On the other hand, using (3.6), we have that,

$$\begin{aligned}
\sup_{\theta^* \in N_\delta(\gamma)} |T_{2n}| &\leq \frac{1}{n(n+1)\sqrt{n}} \sum_{i=1}^{n} \|\dot{k}_s(X_i)\dot{\varphi}(\alpha_i)\big(R_i - R_i^*\big)\| \\
&\leq \frac{M}{n(n+1)\sqrt{n}} \sum_{i=1}^{n} \|\dot{k}_s(X_i)\| \times |\big(R_i - R_i^*\big)| \\
&\leq \frac{M}{n(n+1)\sqrt{n}} \sum_{i=1}^{n} V_i(1 + |X_i|) \qquad (3.14)
\end{aligned}$$

where $|\dot{\varphi}(\alpha_i)| \leq M$ and $|\big(R_i - R_i^*\big)| = V_i$ for all $i = 1, 2, \ldots, n$. The right hand side of (3.14) can be decomposed as

$$\begin{aligned}
\frac{M}{n(n+1)\sqrt{n}} \sum_{i=1}^{n} V_i(1 + |X_i|) &= \frac{M}{n\sqrt{n}} \sum_{i=1}^{n} \frac{V_i}{n+1} + \frac{M}{n(n+1)\sqrt{n}} \sum_{i=1}^{n} |X_i| \\
&\leq \frac{M}{n\sqrt{n}} + \frac{M}{(n+1)\sqrt{n}} \times \left\{ \frac{1}{n} \sum_{i=1}^{n} |X_i| \right\}
\end{aligned}$$

By the Strong Law of Large Number and assumption **B3**, we have that

$$\frac{1}{n} \sum_{i=1}^{n} |X_i| \to E[|X|] < \infty \quad \text{as} \quad n \to \infty \quad \text{a.s.}$$

Therefore

$$\frac{M}{n(n+1)\sqrt{n}} \sum_{i=1}^{n} V_i(1 + |X_i|) \to 0 \quad \text{as} \quad n \to \infty \quad \text{a.s.}$$

Which implies that

$$\sup_{\theta^* \in N_\delta(\gamma)} |T_{2n}| = O_p(1). \qquad (3.15)$$

Using (3.13) and (3.15), we can conclude that

$$\sup_{\theta^* \in N_\delta(\gamma)} |n^{-1/2}\big(S_n(\theta^*) - S_n(\theta_0^*)\big)| = O_p(1). \qquad (3.16)$$

Now let's construct a prove for Theorem 3.2 (ii). Let for $\delta > 0$, $0 < b < \infty$,

$$\Omega_b = \{\theta^* \in N_\delta(\gamma) : |s - r| > b/n\}.$$

So it suffices to show that for every $\xi > 0$, $0 < c < \infty$, there exist a $0 < b < \infty$, and an $N \in \mathbb{N}$ such that

$$P\Big(\inf_{\theta^* \in \Omega_b} M_n(\theta^*, s) > c\Big) > 1 - \xi \quad \forall n > N,$$

where $M_n(\theta^*, s) = D_n(\theta^*, s) - D_n(\theta_1^*, r)$, with $\theta^* \in \mathbb{R}^4$, and $r \in \mathbb{R}$. The prove uses the consistency of the parameter θ^* and for a fix parameter θ^*, the rest of the prove follows the same line as the prove of Theorem 3.2 (i), with the gradient process is this case defined as the derivative of the dispersion function $D_n^s(\theta^*)$ as s for a fix θ^*.

3.4 Asymptotic Normality

This section gives the limiting distribution of the signed-rank estimator. So to prove the asymptotic normality of the signed-rank estimator, we impose some additional regularity conditions that, together with assumptions made in section 3, will be used throughout this section. We start with the following notation that we will use throughout this section. It is convenient to write $k_\theta(x) = k_s(x, \theta^*)$, for $\theta = (\theta^*, s)$ with $\theta^* \in \mathbb{R}^4$, $s \in \mathbb{R}$, and refer to θ^* and s as the coefficient and change point parameter, respectively. Let

$$\dot{k}_s(x) \equiv (\partial/\partial\theta^*)(k_s(x, \theta^*)) = (I(x \leq s), xI(x \leq s), I(x > s), xI(x > s))^T,$$

with $s \in \mathbb{R}$ and $x \in \mathbb{R}$, denote the vector of partial derivatives of $k_s(x, \theta^*)$ with respect to θ^*. Observe that,

$$k_\theta(x) = \theta^{*T}\dot{k}_s(x),$$

where X^T denote the transpose of a given vector X. Consider the following assumptions:

B6 : The function φ is absolutely continuous on $(0,1)$ with $\|\varphi\|_\infty < \infty$. Moreover, φ' has at most a finite number of discontinuities, outside which φ'' exists, is continuous, and is bounded.

B7 : The pdf of the error distribution f is absolutely continuous with finite Fisher information $0 < I(f) = \int_{-\infty}^{\infty} f'(x)^2/f(x)dx < \infty$, and its derivative f' is bounded.

B8 : $\lim_{n\to\infty} \max_{1\le i\le n} h_{ii}^n = 0$

B9 : X is a centered matrix and $\lim_{n\to\infty} n^{-1}XX^T \to \Sigma$, where Σ is a $p \times p$ positive definite matrix.

B10 : $E_G[|X|^4] < \infty$.

Comments on the assumptions

Assumption **B6** is formulated so as to balance the non-smoothness of of the dispersion function and its gradient process with the smoothness of the error distribution or vice versa. Assumption **B7** is a major assumption on the error density function f for much of the rank-based analysis. Finally, assumption **B8** and **B9** are common assumption known classically as Noether's condition which means that the contribution of the maximum value of the design matrix to the variance becomes arbitrarily small as the sample size increases. This will ensure the asymptotic normality of the derivative of the dispersion function $D_n(\theta)$ with respect to θ.

Lemma 2. *Suppose **B1**–**B6** hold and $E[X^2] < \infty$. Then, $\forall B, \quad 0 < B < \infty$,*

$$\sup_{|s-r|\le B/n} |\hat{\theta}_n^*(s) - \theta^*| = o_P(1).$$

The proof of this lemma is facilitated by the fact that under the assumed conditions, $\forall \theta^* \in K$,

$$E\Big\{ \sup_{s\in\bar{\mathbb{R}}, \theta_1^* \in U_{\theta^*}(\eta)} |D_n(Y_i - \theta_1^* \dot{k}_s(X_i)) - D_n(Y_i - \theta^* \dot{k}_s(X_i))| \Big\} \to 0, \quad \text{as} \quad \eta \to 0,$$

where $U_{\theta^*}(\eta) = \{\theta_1^* \in K : |\theta_1^* - \theta^*| < \eta\}$, $\eta > 0$.
We define the scale parameter

$$\tau_\varphi^{-1} = \int \varphi(u)\varphi_f(u)du,$$

where

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

Under Assumptions **B4** and **B6** the scale parameter τ_φ is well defined.

The main result of this section is given by the below theorem. To facilitate its statement we need the following notation.

$$w_n := n^{1/2}(\hat{\theta}_n^* - \theta^*), \quad t_n = n(\hat{r}_n - r), \quad \mathscr{Z}_n := n^{-1/2} \sum_{i=1}^{n} \dot{k}_r(X_i) S_n(\epsilon_i),$$

$$\Gamma_r := E[\dot{k}_r(X)\dot{k}_r(X)^T].$$

Under **B6**–**B9**, and assuming the model in (3.1), we obtain the following linearity result for the gradient process $S_n(\theta_n)$ of $D_n(\theta_n)$:

$$\frac{1}{\sqrt{n}} S_n(\theta_n^*) = \frac{1}{\sqrt{n}} S_n(\theta^*) - \tau_\varphi^{-1} \Gamma_r \sqrt{n}(\theta_n^* - \theta^*) + o_P(1).$$

We also need the following asymptotic uniform quadraticity result in the coefficient parameter. But before we go ahead a present the asymptotic uniform quadraticity result, we need to draw our attention to the following procedure on the jump point given by the following remark.

Remark 14. *Since r is fixed, the below quadratic function involves only the coefficient parameters. From the above result we obtain a locally smooth approximation of the dispersion function $D_n(\theta_n^*)$ which is given by the following quadratic function:*

$$\begin{aligned} Q_n(Y - \theta_n^{*T}\dot{k}_s(X)) &= (2\tau_\varphi)^{-1}(\theta_n^* - \theta^*)^T E[\dot{k}_s(X)^T \dot{k}_s(X)](\theta_n^* - \theta^*) \\ &\quad -(\theta_n^* - \theta^*)^T S_n(Y - \theta^{*T}\dot{k}_s(X)) + D_n(Y - \theta^{*T}\dot{k}_s(X)) \end{aligned} \tag{3.17}$$

The following theorem shows that Q_n provides a local approximation to the dispersion function D_n

Theorem 3.3. Under the model in (3.1), **B6**, **B7**, **B8**, and **B9**, for any $\epsilon > 0$ and $c > 0$,

$$P\Big[\max_{\|\theta_n^* - \theta^*\| < c/\sqrt{n}} |D_n(Y - \theta_n^{*T}\dot{k}_s(X)) - Q_n(Y - \theta_n^{*T}\dot{k}_s(X))| \geq \epsilon\Big] \to 0, \quad \text{as} \quad n \to \infty.$$

The above Theorem 3.3 is used to obtain the asymptotic distribution of the signed-rank estimator.

Remark 15. *If we assume without loss of generality that the true parameter $\theta^* = 0$, then we can write*

$$Q_n(Y - \theta_n^{*T}\dot{k}_s(X)) = (2\tau_\varphi)^{-1}\theta_n^* E[\dot{k}_s(X)^T \dot{k}_s(X)]\theta_n^* - \theta_n^{*T} S_n(Y) + D_n(Y).$$

Because Q_n is a quadratic function it follows from differentiation that it is minimized by

$$\tilde{\theta}_n^* = \tau_\varphi E[\dot{k}_s(X)^T \dot{k}_s(X)]^{-1} S_n(Y).$$

We therefore have the following theorem

Theorem 3.4. Under the model in (3.1) and **B6**, **B7**, **B8**, and **B9**, we have that

$$\sqrt{n}(\tilde{\theta}_n^* - \theta^*) \to_{\mathcal{D}} \mathcal{N}_4(0, \tau_\varphi^2 \Gamma_r^{-1})$$

where $\Gamma_r := E[\dot{k}_r(X)\dot{k}_r(X)^T]$.

Since Q_n is a local approximation to D_n the next result shows that it would seem like their minimizing values are close also.

Theorem 3.5. Under the model in (3.1) and **B6**, **B7**, **B8**, and **B9**, we have that

$$\sqrt{n}(\hat{\theta}_n^* - \tilde{\theta}_n^*) \to_P 0.$$

The proof of the above Theorem 3.3 and Theorem 3.4 uses the same approach as the proof of Theorem A.3.9 in Hettmansperger and McKean (2011).

Proof. Choose $\epsilon > 0$ and $\delta > 0$. Since $\sqrt{n}\tilde{\theta}_n^*$ converges in distribution to a constant we have therefore the convergence in probability. That is, there exists a c_0 such that

$$P\left[\|\tilde{\theta}_n^*\| \geq c_0/\sqrt{n}\right] < \delta/2, \quad \text{for a sufficiently large} \quad n. \tag{3.18}$$

Let

$$\begin{aligned} m &= \min\{Q_{n1}(Y - \theta_n^{*T}\dot{k}_s(X)) : \|\theta^* - \tilde{\theta}_n^*\| = \epsilon/\sqrt{n}\} \\ &\quad -Q_{n1}(Y - \tilde{\theta}_n^*\dot{k}_s(X)). \end{aligned} \tag{3.19}$$

Since $\tilde{\theta}_n^*$ is the unique minimizer of Q_n, $m > 0$; hence, by asymptotic quadraticity we have that

$$P\left[\max_{\|\theta_n^*\| < (c_0+\epsilon)/\sqrt{n}} |D_n(Y - \theta_n^{*T}\dot{k}_s(X)) - Q_n(Y - \theta_n^*\dot{k}_s(X))| \geq m/2\right] \;\leq\; \delta/2, \tag{3.20}$$

for sufficiently large n. By (3.18) and (3.20) we can assert with probability greater than $1-\delta$ that for sufficiently large n,

$$\begin{aligned} |Q_n(Y-\tilde{\theta}_n^*\dot{k}_s(X)) - D_n(Y-\tilde{\theta}_n^*\dot{k}_s(X))| &< (m/2), \\ \text{and} \qquad \|\tilde{\theta}_n^*\| &< c_0/\sqrt{n}. \end{aligned} \tag{3.21}$$

This implies with probability greater than $1-\delta$ that for sufficiently large n,

$$\begin{aligned} D_n(Y-\tilde{\theta}_n^*\dot{k}_s(X)) &< Q_n(Y-\tilde{\theta}_n^*\dot{k}_s(X)) + m/2 \\ \text{and} \qquad \|\tilde{\theta}_n^*\| &< c_0/\sqrt{n}. \end{aligned} \tag{3.22}$$

Using the fact that $\|\tilde{\theta}_n^*\| < c_0/\sqrt{n}$ and $\|\theta^* - \tilde{\theta}_n^*\| = \epsilon/\sqrt{n}$ it follows that $\|\theta_n^*\| \le (c_0+\epsilon)/\sqrt{n}$. Now arguing as above, we have with probability greater than $1-\delta$ that $D_n(Y-\theta_n^{*T}\dot{k}_s(X)) > Q_n(Y-\theta_n^{*T}\dot{k}_s(X)) + m/2$, for sufficiently large n. From this, (3.19), and (3.21) we therefore get the following inequalities:

$$\begin{aligned} D_n(Y-\theta_n^{*T}\dot{k}_s(X)) &> Q_n(Y-\theta_n^{*T}\dot{k}_s(X)) + m/2 \\ &\ge \min\{Q_n(Y-\theta_n^{*T}\dot{k}_s(X)) : \|\theta^* - \tilde{\theta}_n^*\| = \epsilon/\sqrt{n}\} \\ &= m + Q_n(Y-\tilde{\theta}_n^*\dot{k}_s(X)) - m/2 \\ &= m/2 + Q_n(Y-\tilde{\theta}_n^*\dot{k}_s(X)) \\ &> D_n(Y-\tilde{\theta}_n^*\dot{k}_s(X)). \end{aligned} \tag{3.23}$$

Thus, $D_n(Y-\theta_n^{*T}\dot{k}_s(X)) > D_n(Y-\tilde{\theta}_n^*\dot{k}_s(X))$, for $\|\theta^* - \tilde{\theta}_n^*\| = \epsilon/\sqrt{n}$. Since D_n is convex, we must also have $D_n(Y-\theta_n^*\dot{k}_s(X)) > D_n(Y-\tilde{\theta}_n^*\dot{k}_s(X))$, for $\|\theta^* - \tilde{\theta}_n^*\| \ge \epsilon/\sqrt{n}$. But

$$D_n(Y-\tilde{\theta}_n^*\dot{k}_s(X)) \ge \min D_n(Y-\theta_n^{*T}\dot{k}_s(X)) = D_n(Y-\hat{\theta}_n^*\dot{k}_s(X)).$$

hence $\hat{\theta}_n^*$ must lies inside the disk $\|\theta^* - \tilde{\theta}_n^*\| = \epsilon/\sqrt{n}$ with probability of at least $1-2\delta$; that is,

$$P\Big[\|\hat{\theta}_n^* - \tilde{\theta}_n^*\| < \epsilon/\sqrt{n}\Big] > 1-2\delta.$$

This yields the result. □

Combining Theorem 3.4 and Theorem 3.5, we get the next corollary which gives the asymptotic distribution of the w_n.

Corollary 1. *Under the model in* (3.1) *and* ***B6****,* ***B7****,* ***B8****, and* ***B9****, we have that*

$$\sqrt{n}(\hat{\theta}_n^* - \theta^*) \to_{\mathcal{D}} \mathcal{N}_4(0, \tau_\varphi^2 \Gamma_r^{-1})$$

where $\Gamma_r := E[\dot{k}_r(X)\dot{k}_r(X)^T]$.

The next Theorem Gives us the limiting distribution of $t_n = n(\hat{r}_n - r)$, together with the joint limiting distribution of the signed-rank estimator.

Theorem 3.6. Under the model (3.1), and the fact that (3.2) and using **B6**, **B7**, **B8**, and **B9**, we have that

$$w_n = \tau_\varphi \Gamma_r^{-1} \mathcal{Z}_n + o_P(1).$$

Moreover, $(w_n, t_n) \to_{\mathcal{D}} (\mathcal{Z}, \pi_-)$, where $\mathcal{Z}$ is a $\mathcal{N}_4(0, \tau_\varphi^2 \Gamma_r^{-1})$ random variable, independent of π_-, the smallest minimizer of the process Π. Here,

$$\Pi(t) = \mathcal{P}_1(t) I(t \geq 0) + \mathcal{P}_2(-t) I(t \leq 0),$$

$\mathcal{P}_1, \mathcal{P}_2$ are two compound Poisson processes on $[0,\infty)$, with $\mathcal{P}_1(0) = 0 = \mathcal{P}_2(0)$, both having the common rate $g(r)$, and their jumps having the same distribution as that of $D_n(\epsilon+d) - D_n(\epsilon)$, $D_n(\epsilon+d) - D_n(\epsilon)$, respectively. Moreover, the processes $\mathcal{P}_1(t), t > 0$ and $\mathcal{P}_2(-t), t < 0$ are independent.

Recall that $\theta = (\theta^{*T}, r)^T$, now, write $D_n^1(\theta^*, s)$ for $D_n^1(\theta_1^* + n^{-1/2}\theta^*, r + n^{-1/2}s)$, for convenience. Theorem 3.3 gives an approximation of $D_n(\theta^*)$ by quadratic form in θ^*.

We shall next obtain an approximation for D_n^1. The details below are given for $s \geq 0$, they being similar for $s \leq 0$. The analysis here is relatively intricate because it involves the discontinuity point.

We obtain the following asymptotic uniform quadraticity result in the jump point.

Theorem 3.7. Suppose (3.1), (3.2), **B1–B5** hold. Then, for every $0 < b < \infty$,

$$Q_n(\theta^*) = D_n(\theta_1^*, r) - n^{-1/2}\theta^{*T}\sum_{i=1}^{n}\dot{k}_r(X_i)S_n(\epsilon_i) - (2\tau_\varphi)^{-1}\theta^{*T}\Gamma_r\theta^* + D_n^1(0, s) + u_p(1) \quad (3.24)$$

Where $u_p(1)$ is a sequence of stochastic processes converging to zero uniformly over the set $\|\theta^*\| \leq b, |t| \leq b$, in probability.

In view of Theorem 3.2 and (3.24) we obtain that

$$D_n(\hat{\theta}_n^*, \hat{r}_n) = Q_n(\theta_n) + D_n^1(0, s) + o_p(1),$$

$$Q_n := D_n(\theta_n^*, r) - n^{-1/2}\theta^{*T}\sum_{i=1}^{n}\dot{k}_r(X_i)S_n(\epsilon_i) - (2\tau_\varphi)^{-1}\theta^{*T}\Gamma_r\theta^* \quad (3.25)$$

Consequently, asymptotically the standardized minimizers w_n and t_n behave in a singular fashion in the sense that a minimizer of $D_n(\theta + n^{-1/2}\theta_n^*, r + n^{-1}t_n)$ with respect to θ^* is asymptotically equivalent to a minimizer of $Q_n(\theta_n^*)$ with respect to θ^* and does not depend on t_n which exactly what we have shown using (3.17). Similarly, a minimizer t_n of $D_n(\theta + n^{-1/2}\theta_n^*, r + n^{-1}t_n)$ with respect to s is asymptotically equivalent to a minimizer of $D_n^1(0, s)$ with respect to s and does not depend on θ_n^*.

In order to obtain the joint weak limit of (w_n, t_n), we need to obtain the joint weak limit of $(\mathcal{Z}_n, D_n^1)$.

But before we state the below corollary we need the following: Consider the class $\mathcal{D}(-\infty,\infty) = \mathcal{D}_0$ of functions $\gamma(u)$ without discontinuities of the second kind defined on $\mathbb{R}^+$ and such that $\lim_{|u|\to\infty}\gamma(u) = 0$. We shall assume that at discontinuity points $\gamma(u) = \gamma(u+0)$. Define the mapping $dist : \mathcal{D}_0 \times \mathcal{D}_0 \to \mathbb{R}^+$ by

$$dist(\gamma,\psi) = \inf_{\alpha}\Big[\sup_{\mathbb{R}^+}|\gamma(u) - \psi(\alpha(u))| + \sup_{\mathbb{R}^+}|u - \alpha(u)|\Big],$$

where the lower bound is taken over all the monotonic continuous one-to-one mappings $\alpha : \mathbb{R}^+ \to \mathbb{R}^+$. One verifies that $dist$ is a metric on $\mathcal{D}_0$, which transforms $\mathcal{D}_0$ into a complete metric separable space.

Corollary 2. Under (3.1), (3.2), **B1** and **B3**, $(\mathcal{Z}_n, \{D_n^1(t), t \geq 0\})$ converges weakly to $(\mathcal{Z}, \mathcal{P}_1)$ in $\mathbb{R}^4 \times \mathcal{D}[0,\infty)$, where $\mathcal{Z} \sim \mathcal{N}_4(0, \tau_\varphi^2\Gamma_r)$, $\mathcal{P}_1$ is a compound Poisson process on $[0,\infty)$, independent of $\mathcal{Z}$, with the rate $g(r)$, $\mathcal{P}_1(0) = 0$, and the distribution whose jumps is the same as that of $h(r,\epsilon) = D_n(\epsilon + d) - D_n(\epsilon)$.

Corollary 3. *Under assumption of Corollary* 2, *we have that* $(\mathcal{Z}_n, D_n^1)$ *converges weakly to* $(\mathcal{Z}, \Pi)$ *in* $\mathbb{R}^4 \times \mathcal{D}(-\infty,\infty)$, *where* $\mathcal{Z}$ *is as in Corollary* 2, *and independent of* Π.

The next Lemma complete the proof of Theorem 3.6. the proof and details of the below lemma can be found in Koul et al. (2003).

Lemma 3. *Under the assumptions of Theorem* 3.6, $n(\hat{r}_n - r)$ *converges weakly to the smallest minimizer* π_- *of the process* Π.
Moreover, $n(\hat{r} - r)$ *is asymptotically independent of* $n^{1/2}(\hat{\theta}_n^* - \theta^*)$.

3.5 Simulation Study

In this section, we shall report results of a simulation study and an application to an automobile gas mileage-weight data. In both, we are going to compare our method to the LAD and the LS estimators.

The result reported in this section focus on the performance of the SR, LAD and LS estimators for various error densities. We replicated the procedure 1000 times and calculated the average the mean squared error (MSE) of the least squares (LS), signed-rank (SR), and least absolute deviations (LAD) estimators of the unknown parameters over these replications.

The sample were generated from the following simple model:

$$Y_i = (0.5 - X_i)I(X_i \leq 0) + (-0.7 + X_i)I(X_i > 0) + \varepsilon_i, \quad i = 1, 2, \ldots, n,$$

where $\{X_i\}$ is a random sample from the standard normal distribution and the error $\{\varepsilon_i\}$ densities considered are the double exponential, standard normal, and the student t with degrees of freedom 4. In other words, we took $a_0 = 0.5$, $a_1 = -1$, $b_0 = -0.7$, $b_1 = 1$, $r = 0.0$ in (3.1). The error densities considered are the double exponential, standard normal, and the student t with degree of freedom 4. The sample sizes used are 100 and 200.

Estimators are computed using the method described in Section 2 above. Table 3.4 gives the Monte Carlo means (Mean), the mean squares error (MSE) of the LS, LAD, and SR estimators, based on 1000 repetitions. One observes that under the normal errors, SR (Signed-Rank Estimates) and LAD (Least Absolute Deviation Estimates) has larger bias than the LS, while at the double exponential errors, the LS has much larger variability and relatively larger bias than the SR and LAD. For example, for $n = 200$, the Monte Carlo MSE of the LS of r are, respectively, about 2.57 and 4.5 times larger than the SR and LAD of r at the double exponential errors. Figure 3.2 gives a comparison of the accompanying mean squared error (MSE) for the simulation setting which shows that the MSE goes to zero as the sample size increases for both LAD (Least Absolute Deviation) and SR (Signed-Rank). This illustrates our theoretical result that the proposed estimator converges to the true parameter. It is also observed that, in the presence of heavy-tail distributions errors (a gross outlier), the LS (Least Squares) loses both its accuracy and its precision whereas the proposed SR (Signed-Rank) estimator remains almost unaffected.

We also notice that for other heavy-tail distributions, the efficiency gain by SR and LAD is considerable. The estimated relative efficiencies versus LS of SR and LAD estimators. These are calculated by taking the ratio of the LS (Least Squares) MSE to SR (Signed-Rank) and LAD (Least Absolute Deviation) MSEs, respectively, obtained from our simulation experiment. The results are reported in Table 3.3. For brevity, we only report the results for the normal, double exponential, and student t distributions with degree of freedom 10 only. However, we note that LS performs very poorly in the case of the heavy-tailed double exponential and student t distributions with degree of freedom 10 as expected while the LAD and the SR are performing quite well.

3.5.1 An Application: gasoline mileage data

The data, from Koul et al. (2003), originally reported by Henderson and Velleman (1981), consists of gas mileage and weight of 38 automobiles. The original data set with several other variables were collected by Consumer Reports and by 1974 Motor Trend magazine. As mentioned in Koul et al. (2003), it was used by Henderson and Velleman (1981) to investigate the various aspects of automobiles design and performance.

As in Koul et al. (2003), in this section, we only model the relationship between MPG and weights of automobiles. These 38 cars were from the model year 1978-1979. Their weights (in units of 1000 pounds) and fuel efficiencies MPG (miles per gallon) were recorded. illustrates the scatter plot of MPG against weight which appears a pattern of two-phase linear rather than a simple linear regression. As Henderson and Velleman (1981) pointed out, the linear regression is not an appropriate model. They suggested possible other models such as quadratic regression. our case, we suggest the two-phase linear regression model since the scatter plot shows the two-phase pattern.

Thus, we use the two-phase linear regression model (3.1) to fit the data (Weight, MPG). The three estimators used in the two-phase linear regression modeling are LS, LAD and SR. For the sake of a comparison, we also fit the data by simple linear regression models using

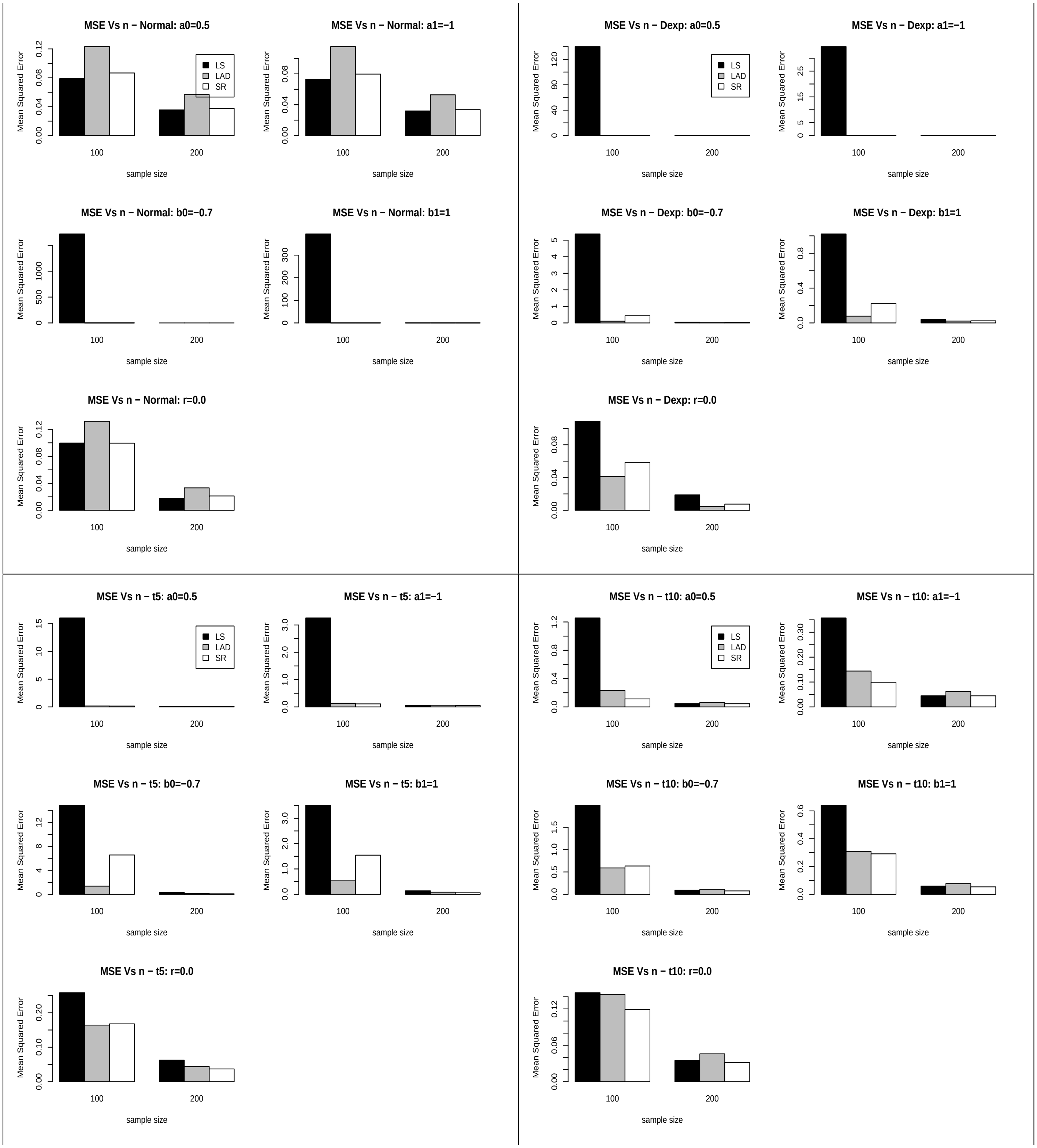

Figure 3.2: MSE Plots

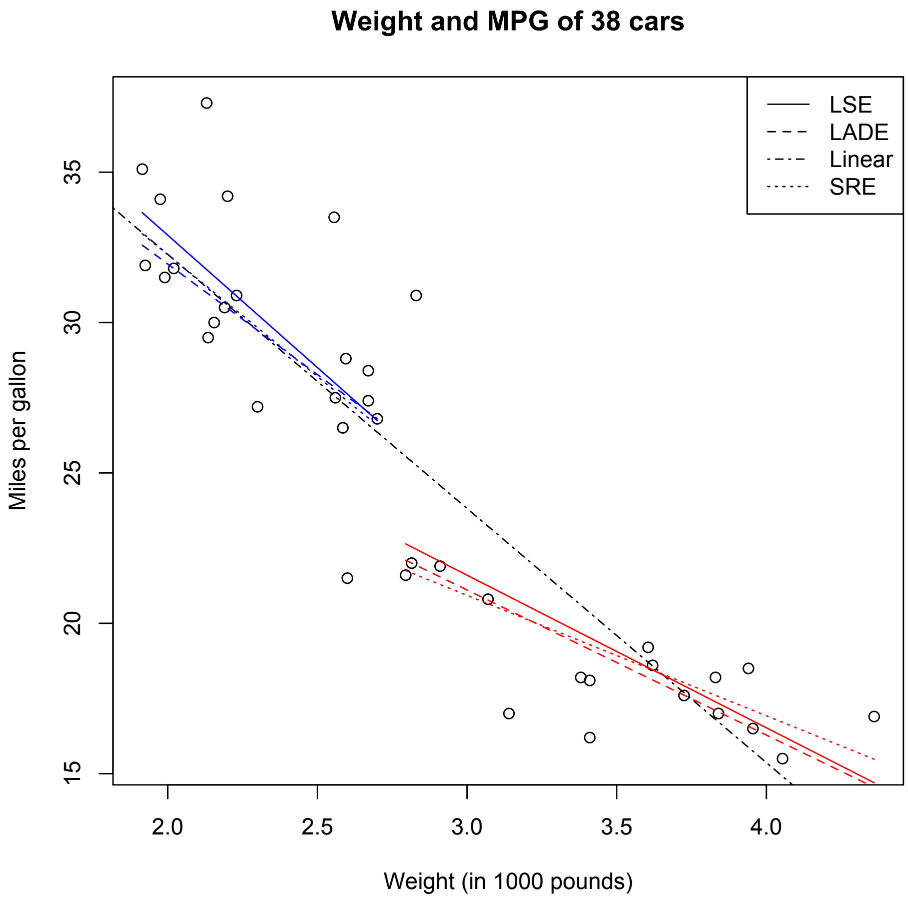

Figure 3.3: Scatter plot with the four fitted curves.

LS. Figure 3.3 shows the fits for all four estimators for the three types of regression models. As in Koul et al. (2003) both methods yield the same value for $\widehat{r} = 2.7$. since values of the parameters in the case of LAD and LS has been reported in Koul et al. (2003), for the sake of brevity we will only report the estimated parameters and the the estimated standard deviation in the case of the signed-rank. Table 3.1 below report the estimated parameters and the the estimated standard deviation in the case of the signed-rank where both s_1 and s_2 are smaller using two-phase regression than the single square-root mean square error 2.85 using simple linear regression. Here as in Koul et al. (2003),

$$s_1 = \sqrt{\frac{1}{20} \sum (Y_i - \widehat{a}_0 - \widehat{a}_1 X_i)^2 I(X_i \leq 2.7)}$$

Table 3.1: Signed-Rank estimates and estimated standard deviation for each piece

Parameter	$\widehat{a}_0$	$\widehat{a}_1$	$\widehat{b}_0$	$\widehat{b}_1$	$\widehat{r}$	s_1	s_2
SR	48.474	-8.108	32.947	-4.005	2.7	2.678	2.555

and

$$s_2 = \sqrt{\frac{1}{18}\sum(Y_i - \widehat{b}_0 - \widehat{b}_1 X_i)^2 I(X_i > 2.7)}$$

3.5.2 Application: Grade Inflation at a Large Southeastern University from 1981-2011

Since the scatter plot shows the two-phase pattern, we use the two-phase linear regression model (3.1) to fit the data (Year, Avg. Fall GPA) where Avg stands here for Average. The three estimators used in the two-phase linear regression modeling are LS, LAD and SR. For the sake of a comparison, we also fit the data by simple linear regression models using LS. Figure 3.4 shows the fits for all four estimators for the three types of regression models.

Table 3.2 lists the parameter estimators for the two-phase linear regression models for LS , LAD, and SR methods. Both methods yield the same value for $\widehat{r}$ while the estimates of the other parameters are different. It appears there is a change at year $\widehat{r} = 2000$. Out of the 30 years, there are 19 years below year 2000 and 11 years above year 2000.

It is also useful to look at the estimated standard deviation for each piece given by

$$s_1 = \sqrt{\frac{1}{19}\sum(Y_i - \widehat{a}_0 - \widehat{a}_1 X_i)^2 I(X_i \leq 2000)}$$

and

$$s_2 = \sqrt{\frac{1}{11}\sum(Y_i - \widehat{b}_0 - \widehat{b}_1 X_i)^2 I(X_i > 2000)}$$

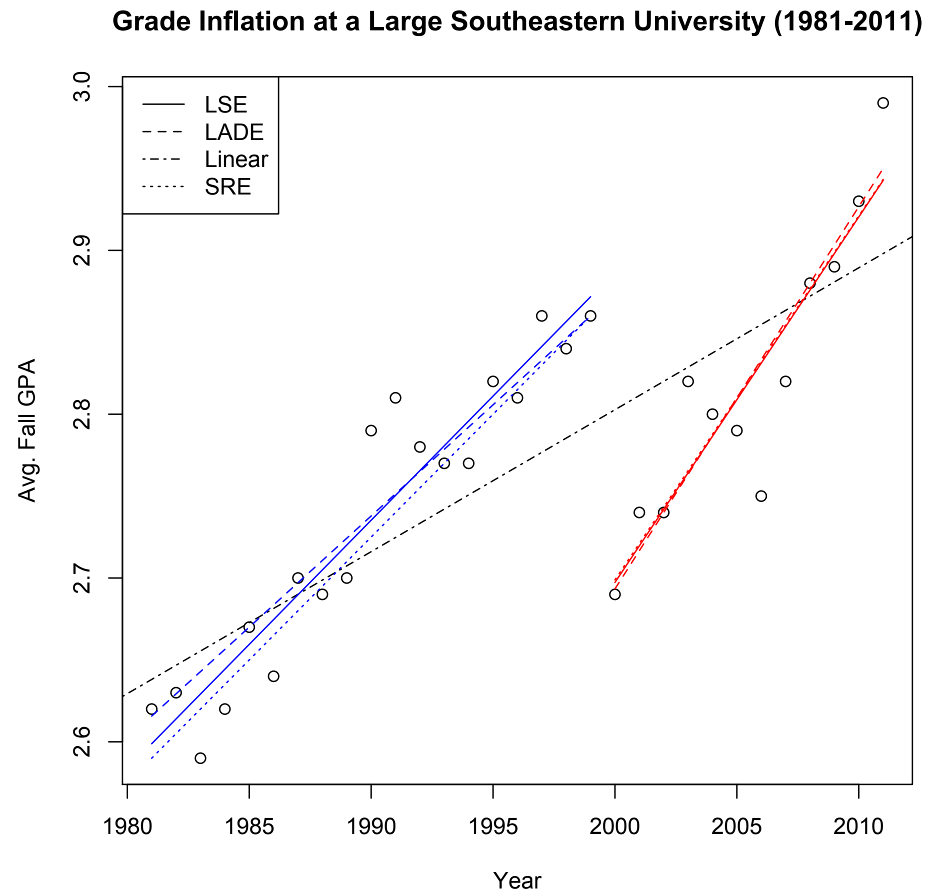

Figure 3.4: Scatter plot with the four fitted curves.

3.6 Conclusion

A rank based estimator for two phase linear model has been developed, with its performance evaluated in comparison with the LS and LAD for both simulated and real data. Our estimation procedure produces through minimization of a rank based objective function, yielding estimators that are robust in the response space. Which means estimator that are robust in the y direction. As such, the method developed in this section is ideal for data from designed experiments where the x's are controlled. There is no guarantee that our procedure results in robust estimates for uncontrolled studies. It will also be interesting to extend our procedure to the case of multiphase nonlinear regression.

Table 3.2: Parameter estimates and estimated standard deviation for each piece

Parameter	$\widehat{a}_0$	$\widehat{a}_1$	$\widehat{b}_0$	$\widehat{b}_1$	$\widehat{r}$	s_1	s_2
SR	-27.12	0.0150	-41.74	0.0222	2000	0.0285	0.0359
LAD	-24.26	0.0135	-43.97	0.0233	2000	0.0281	0.0362
LS	-27.42	0.0151	-41.91	0.0223	2000	0.0266	0.0359

Table 3.3: Estimated relative efficiencies versus LS

Method	Distribution	a_0	a_1	b_0	b_1	r
LAD	Normal	0.625	0.603	0.636	0.619	0.540
	D exp	1.813	1.665	2.646	1.811	4.093
	t_{10}	0.751	0.718	0.808	0.770	0.758
SR	Normal	0.942	0.948	0.951	0.935	0.843
	D exp	1.559	1.515	1.935	1.592	2.483
	t_{10}	1.049	1.002	1.191	1.110	1.099

Table 3.4: Mean, and MSE, of LAD, LS and SR for $(0.5, -1, -0.7, 1, 0.0)$

		Error distribution					
Method	Est.	N(0,1)		Dexp(0,1)		t(5)	
		Mean	MSE	Mean	MSE	Mean	MSE
				$n = 200$			
LAD	$\hat{a}_0$	0.510	0.056	0.506	0.0195	0.521	0.066
	$\hat{a}_1$	-0.995	0.052	-0.991	0.019	-0.980	0.062
	$\hat{b}_0$	-0.716	0.080	-0.710	0.019	-0.704	0.118
	$\hat{b}_1$	1.025	0.060	1.006	0.021	1.013	0.083
	$\hat{r}$	0.036	0.033	-0.002	0.004	0.048	0.044
LS	$\hat{a}_0$	0.523	0.035	0.515	0.035	0.527	0.065
	$\hat{a}_1$	-0.984	0.031	-0.986	0.033	-0.979	0.061
	$\hat{b}_0$	-0.720	0.051	-0.708	0.052	-0.670	0.299
	$\hat{b}_1$	1.024	0.037	1.005	0.039	1.002	0.135
	$\hat{r}$	0.014	0.017	0.008	0.018	0.064	0.062
SR	$\hat{a}_0$	0.523	0.037	0.513	0.022	0.527	0.054
	$\hat{a}_1$	-0.984	0.033	-0.985	0.021	-0.979	0.050
	$\hat{b}_0$	-0.717	0.054	-0.711	0.026	-0.719	0.075
	$\hat{b}_1$	1.023	0.040	1.006	0.024	1.026	0.059
	$\hat{r}$	0.017	0.021	0.001	0.007	0.039	0.036
				$n = 100$			
LAD	$\hat{a}_0$	0.519	0.123	0.505	0.048	0.512	0.175
	$\hat{a}_1$	-0.987	0.115	-0.994	0.050	-0.995	0.133
	$\hat{b}_0$	-0.586	0.484	-0.690	0.106	-0.554	1.360
	$\hat{b}_1$	0.936	0.247	0.998	0.077	0.929	0.558
	$\hat{r}$	0.118	0.131	0.030	0.041	0.146	0.164
LS	$\hat{a}_0$	0.528	0.078	0.902	140.08	0.655	16.03
	$\hat{a}_1$	-0.975	0.073	-0.789	34.49	-0.922	3.257
	$\hat{b}_0$	-1.954	0.001	-0.677	5.374	-0.275	14.83
	$\hat{b}_1$	1.611	0.039	0.988	1.021	0.784	3.510
	$\hat{r}$	0.090	0.099	0.085	0.108	0.172	0.258
SR	$\hat{a}_0$	0.531	0.086	0.515	0.054	0.541	0.158
	$\hat{a}_1$	-0.974	0.079	-0.987	0.055	-0.972	0.111
	$\hat{b}_0$	-0.618	0.447	-0.656	0.436	-0.393	6.557
	$\hat{b}_1$	0.964	0.217	0.978	0.222	0.849	1.544
	$\hat{r}$	0.087	0.099	0.049	0.058	0.138	0.167

Bibliography

Abebe, A., McKean, J. W. and Bindele, H. F. *On the consistency of a class of nonlinear regression estimator*, Pakistan Journal of Statistics and Operations Research, Vol. 8, pp. 543–555, 2012.

Anderson, R. L. and Nelson, L. A. *A family of models involving intersecting straight lines and concomitant experimental designs useful in evaluating response to fertilizer nutrients*, Biometrics, Vol. 31, pp. 303–318, 1975.

Autin, M. A. and Edwards, D. *Nonparametric harmonic regression for estuarine water quality data*, Environmetrics, Vol. 21, pp. 588–605, 2010.

Baklanov, E. A. *The Strong Law Of Large Numbers for L-statistics with Dependent Data*, Siberian Mathematical Journal, Vol. 47, No. 6, pp. 975–979, 2006.

Bansal, N. K., Hamedani, G. G. and Zhang, H. *Non-linear regression with multidimensional indices*, Statistics and Probability Letters, Vol. 45, pp. 175–186, 1999.

Bhattacharya, P. K. *Some aspects of change-point analysis*, IMS Lecture Notes-Monograph Series, Vol. 23 Hayward, CA, pp. 28–56, 1994.

Bhattacharya, P. K. *Weak convergence of the log-likelihood process in the two-phase linear regression problem*, In:Bahadur,R.R (Ed.),Probability, Statistics and Design of Experiments. Wiley Eastern, New Delhi, pp. 145–156, 1990.

Bindele, H. F. *Asymptotics of the Signed-Rank Estimator under Dependent Observations*, Journal of Statistical Planning and Inference, Vol. 146, pp. 49–55, 2014.

Bindele, H. F. and Abebe, A. *Bounded influence nonlinear signed-rank regression*, The Canadian Journal of Statistics, Vol. 40, No. 1, pp. 172–189, 2012.

Brillinger, D. R. Fitting Cosines: Some Procedures and Some Physical Examples in I. B. MacNeill, G. J. Umphrey, D. R. Bellhouse, R. J. Kulperger and D. R. Brillinger (Eds) *Advances in the Statistical Sciences: Applied Probability, Stochastic Processes, and Sampling Theory*. Springer Netherlands, 1986.

Bujosa, M., García-Ferrer, A. and de Juan, A., *Predicting recessions with factor linear dynamic harmonic regressions*, J. Forecast., Vol. 32, pp. 481–499, 2013.

Ciuperca, G. *The M-estimation in a multi-phase random linear model*, Stat. Prob. Lett., Vol 79, pp. 573–580, 2008.

Ermoliev, Y., Norkin, V. and Wets, R. *The Minimization of Semicontinuous Functions Mollifier Subgradients* , SIAM J. Control and Optimization, . Vol. 33 (1), pp. 149–167, (1995).

Eubank, R. L. *Approximate regression models and splines*, Comm. Statist.-Theory Methods Ser. Vol 13, pp. 433–484, 1984.

Francos, J. M., Meiri, A. Z. and Porat, B. *A united texture model based on a 2-D Wold like decomposition*, IEEE Transactions on Signal Processing, Vol. 41, pp. 2665–2678, 1993.

Feder, P. I. *On asymptotic distribution theory in segmented regression problems-identified case*, Ann. Statist. Vol 3, pp. 49–83, 1974a.

Feder, P. I. *The log likelihood ratio in segmented regression*, Ann. Statist. Vol. 3, pp. 84–97, 1975b. Gikhman, I.I., Skorokhod, A.V., 1969. Introduction to the Theory of Random Processes. W.B. Saunders Company, PhiLADlphia.

Gallant, A. R. *Asymptotic Theory of Nonlinear least Square Estimator*, Annals of Statistics, Vol. 9, No. 3, pp. 501-513, 1981.

Hannan, E. J. *The estimation of frequency*, Journal of Applied Probability, vol. 10, pp. 510–519, 1973.

Henderson, H. V. and Velleman, P. F. *Building multiple regression models interactively*, Biometrics, Vol. 37, pp. 391–411, 1981.

Hettmansperger, T. P. and McKean, J. W. *Robust Nonparametric Statistical Methods*, Second edition, CRC Press, Boca Raton, 2011.

Hinkley, D. V. *Inference about the intersection in two-phase regression*, Biometrika, Vol. 56, pp. 495–504, 1969.

Hinkley, D. V. *Inference in two-phase regression*, J. Amer. Statist. Assoc. Vol. 66, pp. 736–743, 1971.

Hossjer, O. *Rank-Based Estimates in the Linear Model With High Breakdown Point*, Journal of American Statistical Association. March, Vol. 89, No. 425, 1994.

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

Hudson, D. J. *Fitting segmented curves whose join points have to be estimated*, J. Amer. Statist. Assoc., Vol. 61, pp. 1097–1129, 1996.

Stakgold, I. and Holst, M. *Green's Function and Boundary Value Problem*, Third edition, John Wiley and Sons, Mar 1, 2011.

Jennrich, R. I. *Asymptotic Theory of Nonlinear least Square Estimator*, Annals of Statistics, Vol. 9, No. 3, pp. 501-513, 1969.

Kay, S. *Modern Spectral Estimation: Theory and Application*, Prentice Hall, NJ, 1999.

Koul, H. and Qian, L. *Asymptotics of maximum likelihood estimators in a two-phase linear regression model*, J. Statistics Planning and Inference., Vol. 108, pp. 99–119, 2002.

Koul, H., Qian, L. and Surgailis, D. *Asymptotics of M-estimators in a Two-Phase Linear Regression Model*, Stochastic Processes and Their Applications, Vol. 103, pp. 123–154, 2003.

Kundu, D. *Asymptotic theory of least squares estimator of a particular nonlinear regression model*, Statistics and Probability Letters, Vol. 18, pp. 13–17, 1993.

Kundu, D. and Nandi, S., *Determination of Discrete Spectrum in a Random Field*, Statistica Neerlandica, Vol. 57, No. 2, pp. 258–283, 2003.

Kundu, D. and Nandi, S., *Statistical signal processing*, Frequency estimation, Springer Briefs in Statistics, Springer, New Delhi, 2012.

Lahiri, A., Kundu, D. and Mitra, A., *Efficient algorithm for estimating the parameters of two dimensional chirp signal*, Sankhya B, Vol. 75, pp. 65–89, 2013.

Lahiri, A., Kundu, D., and Mitra, A., *On least absolute deviation estimators for one-dimensional chirp model*, Statistics, Vol. 48, pp. 405–420, 2014.

Lee, Y.H., and Haberman, S.J., *Harmonic regression and scale stability*, Psychometrika, Vol. 78, pp. 815–829, 2013.

McClellan, J. H. *Multidimensional spectral estimation*, Proc. IEEE, Vol. 70, pp. 1029–1039, 1982.

McKean, J. W. and Schrader, R. M. *The geometry of robust procedures in linear models*, Journal of the Royal Statistical Society. Series B. Vol. 42, pp. 366–371, 1980.

Mitra, S., Mitra, A., and Kundu, D. *Genetic algorithm and M-estimator based robust sequential estimation of parameters of nonlinear sinusoidal signals*, Commun. Nonlinear Sci. Numer. Simul., Vol. 16, pp. 2796–2809, 2011.

Nandi, S. *Estimation of parameters of two-dimensional sinusoidal signal in heavy-tailed errors*, Journal of Statistical Planning and Inference, Vol. 142, pp. 2799–2808, 2012.

Nandi, S., and Kundu, D., *Estimation of parameters of partially sinusoidal frequency model*, Statistics, Vol. 47, pp. 45–60, 2013.

Nandi, S., Kundu, D., and Srivastava, R.K., *Noise space decomposition method for two-dimensional sinusoidal model*, Comput. Statist. Data Anal., Vol. 58, pp. 147–161, 2013.

Nguelifack, B., Kwessi, E. and Abebe, A. *Generalized Signed-Rank Estimation for Nonlinear Models with Multidimensional Indices*, Journal of Nonparametric Statistics, in press.

Rao, C. R. and Zhao, L. C. *Asymptotic behavior of maximum likelihood estimates of superimposed exponential signals*, IEEE Transactions on Signal Processing, Vol. 41, pp. 1461–1464, 1993.

Rao, C. R., Zhao, L. C. and Zhou, B. *Maximum likelihood estimation of 2-D superimposed exponential signals*, IEEE Transactions on Signal Processing, Vol. 42, pp. 1795–1802, 1994.

Rice, J. A. and Rosenblatt, M. *On frequency estimation*, Biometrika, Vol. 75, pp. 477–484, 1988.

Schulze, U. *Multiphase Regression: Stability Testing, Estimation, Hypothesis Testing*, Akademie-Verlag, Berlin. ISBN 3-05-500224-5, 1987.

Sprent, P. *Some hypotheses concerning two-phase regression lines*, Biometrics, Vol. 17, pp. 634–645, 1961. van de Geer, S., 1988. Regression analysis and empirical processes. CWI Tract. 45.

Wu, J. S. and Chu, C. K. *Kernel-type estimators of jump points and values of a regression function*, Ann. Statist. Vol. 21, pp. 1545–1566, 1993.

Wu, C.-F. *Asymptotic Theory of Nonlinear Least Squares Estimator*, Annals of Statistics, Vol. 9, No. 3, pp. 501–513, 1981.

Yohai, V. J. *High breakdown-point and high efficiency robust estimates for regression*, Annals Statistics, Vol. 15, 642–656, 1987.

Yuan, T. and Rao, T. S. *Spectrum estimation for random fields with applications to Markov modeling and texture classification*, in Markov Random Fields: Theory and Applications, R. Chellappa and A.K. Jain, eds., American Mathematical Society, 1992.

Zhang, H. and Mandrekar, V. *Estimation of hidden frequencies for 2D stationary processes*, Journal of Time Series Analysis, Vol. 22, pp. 613–629, 2001.