National Library of Canada
Acquisitions and Bibliographic Services Branch
395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale du Canada
Direction des acquisitions et des services bibliographiques
395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

Canada

UNIVERSITY OF ALBERTA

Bounded-Influence R-Estimators in the Linear Model

BY

Julie Zhou

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE

OF MASTER OF SCIENCE

DEPARTMENT OF STATISTICS AND APPLIED PROBABILITY

EDMONTON, ALBERTA

FALL 1992

ISBN 0-315-77105-4

UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR: **Julie Zhou**

TITLE OF THESIS:

Bounded-Influence R-Estimators

in the Linear Model

DEGREE: **Master of Science**

YEAR THIS DEGREE GRANTED: **Fall 1992**

(Signed) Julie Zhou

Permanent Address:
Department of Statistics
and Applied Probability
University of Alberta
Edmonton, Alberta
Canada T6G 2G1

Date: July 22, 1992

UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

THE UNDERSIGNED CERTIFY THAT THEY HAVE READ, AND RECOMMEND TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH FOR ACCEPTANCE, A THESIS

ENTITLED *BOUNDED - INFLUENCE R-ESTIMATORS IN THE LINEAR MODEL*

SUBMITTED BY *JULIE ZHOU*

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE.

DOUGLAS P. WIENS

JEROME N. SHEAHAN

MAZIAR SHIRVANI

Date: May 20, 1992

To:

David

Simon

FengYa

ABSTRACT

Bounded influence M-estimators in the linear regression have been investigated for many years. But bounded R-estimators were only studied recently by Tableman in a special case which uses Wilcoxon rank scores. The object of this thesis is to study the general form of bounded influence R-estimator. Using influence function as a tool, we study the asymptotic behaviors of the R-estimator which include the asymptotic uniqueness and the asymptotic covariance. Based on spherical distribution of regressors $\mathbf{X}$, optimality problems are raised according to different constraints. Mainly Hampel type optimality problems and Huber type minimax problem are solved here under certain assumptions on the error distribution. Finally, the strategy of computing this R-estimator based on optimal functions is discussed and a **SAS** program is developed . Some numerical results are given to compare this R-estimator with other estimators.

ACKNOWLEDGEMENTS

I wish to thank my supervisor, Dr. Douglas P. Wiens, with my deepest appreciation, for his invaluable guidance and friendly support.

I also wish to thank Dr. J.N. Sheahan and Dr. M. Shirvani for their careful consideration and helpful suggestions.

A special thanks goes to Ms. Christine Fischer for her accurate and beautiful typing.

I thank the Department of Statistics and Applied Probability for providing a congenial research environment. I am also grateful to Dr. Douglas P. Wiens and the University of Alberta for their financial support.

TABLE OF CONTENTS

CHAPTER 1
INTRODUCTION

1.1 Linear Model

We consider the estimation for the unknown parameters θ_0 and $\boldsymbol{\theta}$ in the multiple linear regression model:

$$y_i = \theta_0 + \mathbf{x}_i^T\boldsymbol{\theta} + \varepsilon_i,\ i = 1, 2, \ldots, n\ , \tag{1.1.1}$$

where $y_i \in \mathbb{R}$ is the i^{th} observation of the response variable; $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ik})^T \in \mathbb{R}^k$ is the i^{th} observation of the independent variables; $\theta_0 \in \mathbb{R}$ is the unknown constant term; $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)^T \in \mathbb{R}^k$ is an unknown vector; and $\varepsilon_i \in \mathbb{R}$ is the i^{th} error.

We assume (1) $\varepsilon_1, \ldots, \varepsilon_n$ are independently identically distributed according to some symmetric distribution G, (2) $\varepsilon_1, \ldots, \varepsilon_n$ are independent of $\mathbf{x}_1, \ldots, \mathbf{x}_n$. The pairs $(\mathbf{x}_i, y_i)$ $(i = 1, \ldots, n)$ are viewed as a random sample from some joint distribution $F(\mathbf{x}, y)$. If H denotes the marginal distribution of $\mathbf{x}$, then $F(\mathbf{x}_i, y_i) = H(\mathbf{x}_i) \cdot G(y_i - \theta_0 - \mathbf{x}_i^T\boldsymbol{\theta})$. Correspondingly, f, g and h denote the probability density functions for $(\mathbf{x}, y)$, ε and $\mathbf{x}$ respectively.

If G is the normal distribution with mean zero, i.e. $N(0, \sigma^2)$ with $\sigma^2 > 0$, then the least squares estimator for θ_0 and $\boldsymbol{\theta}$ in the model (1.1.1) is asymptotically efficient. However, in the presence of heavy-tailed errors and/or anomalous data, the efficiency of the least squares estimator is remarkably reduced.

Since 1960's several robust methods have been introduced and developed. Robust regression methods can be almost as efficient as the least squares method when G is $N(0, \sigma^2)$. Under the situation of heavy-tailed errors, robust regression methods are much better than the least squares method (see Montgomery and Peck (1982)). In general, there are three types of robust methods: Maximum likelihood type or M-estimation method, the rank-based or R-estimation method, and linear combinations of order statistics type or L-estimation method.

Jaeckel (1972) introduced a class of rank-estimators to estimate $\boldsymbol{\theta}$ in the model (1.1.1). First he defined a function of $y - \mathbf{x}^T\boldsymbol{\theta}$, given by

$$D_J(y - \mathbf{x}^T\boldsymbol{\theta}) = \sum_{i=1}^{n} a(R_i(y - \mathbf{x}^T\boldsymbol{\theta}))(y_i - \mathbf{x}_i^T\boldsymbol{\theta}), \tag{1.1.2}$$

where $R_i(y - \mathbf{x}^T\boldsymbol{\theta})$ is the rank of $y_i - \mathbf{x}_i^T\boldsymbol{\theta}$, and $a(i)$ is a score function. Then the solution to the minimization problem

$$\min_{\boldsymbol{\theta}\in\Omega}\{D_J(y - \mathbf{x}^T\boldsymbol{\theta})\}\ ,\ \Omega \subseteq \mathbb{R}^k$$

is the R-estimator of $\boldsymbol{\theta}$.

In order to analyse the properties of Jaeckel's R-estimator, we first discuss the most useful heuristic tool of robust statistics - the influence curve or the influence function which was introduced by Hampel (1968, 1974).

Definition 1.1.1: Influence function (Huber (1981))

Let's add one more observation with value x to a very large sample. Its suitably normed limiting influence on the value of an estimate $\theta(F_n)$ can be expressed as a function of x, given by

$$\mathrm{IF}(x, F, \theta) = \lim_{\varepsilon\to 0} \frac{\theta((1-\varepsilon)F + \varepsilon\delta_x) - \theta(F)}{\varepsilon},$$

where δ_x denotes the pointmass 1 at x. $\mathrm{IF}(x, F, \theta)$ is called influence function for $\theta(F)$.

If $\theta(F)$ is sufficiently regular, it can be linearized near F in terms of $\mathrm{IF}(x, F, \theta)$; if G is near F, then the leading terms of a Taylor expansion are

$$\theta(G) = \theta(F) + \int \mathrm{IF}(x, F, \theta)[G(dx) - F(dx)] + \text{ Rem}\,.$$

We know (see Huber (1981))

$$\int \mathrm{IF}(x, F, \theta)F(dx) = 0\,,$$

and if we substitute the empirical distribution F_n for G in the above expansion, we obtain

$$\sqrt{n}(\theta(F_n) - \theta(F)) = \sqrt{n} \int \mathbb{IF}(x, F, \theta) F_n(dx) + \sqrt{n}\text{Rem}$$
$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbb{IF}(x_i, F, \theta) + \sqrt{n}\text{Rem}.$$

From the central limit theorem, the leading term on the right-hand side is asymptotically normal with mean zero, if the x_i are independent with common distribution F. Since it is often true (but not easy to prove) that the remaining terms are asymptotically negligible, $\sqrt{n}[\theta(F_n) - \theta(F)]$ is then asymptotically normal with mean 0 and variance

$$V(F, \theta) = \int [\mathbb{IF}(x, F, \theta)]^2 F(dx).$$

Therefore the influence function has two main uses (From Huber (1981)).

First it allows us to assess the relative influence of individual observations toward the value of an estimate. If it is unbounded, then an outlier might cause trouble.

Second, the influence function allows an immediate and simple assessment of the asymptotic properties of an estimate, since it allows us to guess an explicit formula

$$V(F, \theta) = \int [\mathbb{IF}(x, F, \theta)]^2 F(dx)$$

for the asymptotic variance.

Now let's go back to Jaeckel's procedure. We find out that the influence function for Jaeckel's R-estimator is bounded only in the y-space but not in the x-space, so a leverage point will have large effect on the estimate of θ. In order to have influence function bounded both in the x-space and in the y-space, Tableman (1990) recently introduced a bounded-influence rank estimator based

on Wilcoxon scores. She let a weighted Wilcoxon signed rank statistic be a function of the form

$$\overline{S}^{+}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} w(\mathbf{x}_i, y_i - \theta_0 - \mathbf{x}_i^T\boldsymbol{\theta}) \cdot \frac{R_i^+}{n+1}\text{sign}(y_i - \theta_0 - \mathbf{x}_i^T\boldsymbol{\theta}) \begin{pmatrix} 1 \\ \mathbf{x}_i \end{pmatrix} ,$$

here R_i^+ is the rank of $|y_i - \theta_o - \mathbf{x}_i^T\boldsymbol{\theta}|$, and w is a scalar "weight" function. The solution to $\overline{S}^{+}(\boldsymbol{\theta}) = \mathbf{0}$ is Tableman's estimator for θ_0 and $\boldsymbol{\theta}$. This estimator has bounded influence function, but we don't know whether the asymptotic variance for the estimator reaches minimum or not. So our interest is to obtain a general form of bounded-influence rank estimators. Using influence function as our tool, we derive the best estimator under certain assumptions (we will give these assumptions in later chapter).

Next section, using Jaeckel's and Tableman's ideas, we will obtain a class of bounded influence R-estimators (denoted by BRE). In Chapter 2, we develop some asymptotic results which include asymptotic equation for BRE and asymptotic uniqueness of BRE. Chapter 3 deals with the specific choice of the optimal functions. This is our primary goal. We will discuss the optimality criteria, raise the optimality problems corresponding to different requirements, and derive solutions to those problems respectively. Chapter 4 discusses computation strategy. A SAS program is given to compute estimate based on optimal functions. In section 4.2 and 4.3 we do some comparison with other estimators. At the end we will give conclusion.

1.2 Motivation of bounded influence R-estimators

Now we generalize Jaeckel (1972) and Tableman (1990) methods, for the model

$$y_i = \mathbf{x}_i^T\boldsymbol{\theta} + \varepsilon_i \, , \; i = 1, \ldots, n \, . \tag{1.2.1}$$

Denote

$$e_i = y_i - \mathbf{x}_i^T \boldsymbol{\theta},$$

$$R_i = \text{ rank of } e_i,$$

and

$$R_i^+ = \text{ rank of } |e_i|.$$

Let $\rho(\mathbf{x}, e)$ satisfy the following conditions:

i) $\rho(\mathbf{x}, e) \geq 0$, for each $\mathbf{x}$ and e,

ii) $\rho(\mathbf{x}, e)$ is convex in e, for each $\mathbf{x}$,

iii) $\rho(\mathbf{x}, e)$ is minimized at $e = 0$, for each $\mathbf{x}$,

iv) $\rho(\mathbf{x}, e)$ is an even function of e, for each $\mathbf{x}$,

v) $\rho(\mathbf{x}, e)$ is twice differentiable with respect to e almost everywhere.

Put $\eta(\mathbf{x}, e) = \frac{\partial}{\partial e}\rho(\mathbf{x}, e)$, so η is odd and nondecreasing function in e for each $\mathbf{x}$.

In least squares method, we want to

$$\min_{\boldsymbol{\theta}} \sum_{i=1}^{n} e_i^2(\boldsymbol{\theta}),$$

i.e.

$$\min_{\boldsymbol{\theta}} \sum_{i=1}^{n} e_i(\boldsymbol{\theta}) \cdot e_i(\boldsymbol{\theta}) \,. \tag{1.2.2}$$

Jaeckel replaced one $e_i(\boldsymbol{\theta})$ in (1.2.2) by a function $a(i)$ of its rank, i.e. $a(i) = a(R_i)$, to get

$$D_J(\boldsymbol{\theta}) = \sum_{i=1}^{n} a(i) e_{(i)} = \sum_{i=1}^{n} a(R_i) e_i, \tag{1.2.3}$$

where the $e_{(i)}$ are the ordered residuals. Minimizing $D_J(\boldsymbol{\theta})$, we get Jaeckel's rank-estimators.

We notice that if $\rho(\mathbf{x}, e) = e^2/2$ as in least squares, then

$$e_{(i)} = \text{sign}(e_{(i)}) \cdot \sqrt{2\rho(\mathbf{x}_i, e_{(i)})}. \tag{1.2.4}$$

So we replace one $e_{(i)}$ by $a(i)$ and another by (1.2.4) in (1.2.2), get

$$D(\boldsymbol{\theta}) = \sum_{i=1}^{n} a(i)\text{sign}(e_{(i)}) \cdot \sqrt{2\rho(\mathbf{x}_i, e_{(i)})}$$
$$= \sum_{i=1}^{n} a(R_i)\text{sign}(e_i) \cdot \sqrt{2\rho(\mathbf{x}_i, e_i)} \,. \tag{1.2.5}$$

Let

$$w_1(\mathbf{x}_i, e) = \frac{\partial}{\partial e}\sqrt{2\rho(\mathbf{x}_i, e)}$$
$$= \eta(\mathbf{x}_i, e)/\sqrt{2\rho(\mathbf{x}_i, e)} \,,$$

then

$$S(\boldsymbol{\theta}) = -\frac{\partial D(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$$
$$\approx \sum_{i=1}^{n} a(R_i) w_1(\mathbf{x}_i, e_i)\text{sign}(e_i)\mathbf{x}_i \,, \tag{1.2.6}$$

here this approximation is in the sense of equality except for a finite number of points of $\boldsymbol{\theta}$. In (1.2.6), we can replace $a(R_i)$ by $a(R_i^+)\text{sign}(e_i)$, and put

$$w(\mathbf{x}_i, e_i) = w_1(\mathbf{x}_i, e_i) \cdot \text{sign}(e_i)$$

to get

$$S(\boldsymbol{\theta}) = \sum_{i=1}^{n} a(R_i^+) \cdot \text{sign}(e_i) \cdot w(\mathbf{x}_i, e_i)\mathbf{x}_i \,. \tag{1.2.7}$$

If $\boldsymbol{\theta}$ minimizes $D(\boldsymbol{\theta})$, it should satisfy

$$-D'(\boldsymbol{\theta}) = S(\boldsymbol{\theta}) = \mathbf{0} \,.$$

In addition, we have to get some information about $w(\mathbf{x}, e)$ before we work out the expression of $S(\boldsymbol{\theta})$.

1. $w(\mathbf{x}, e) \geq 0$.

Since $\rho(\mathbf{x}, e)$ is even and convex function of e, $\eta(\mathbf{x}, e)$ is odd and nondecreasing function of e, i.e.

$$\eta(\mathbf{x}, e) \cdot \text{sign}(e) \geq 0.$$

Therefore

$$w(\mathbf{x}, e) = \frac{\eta(\mathbf{x}, e)}{\sqrt{2\rho(\mathbf{x}, e)}} \cdot \text{sign}(e) \geq 0 .$$

2. $w(\mathbf{x}, e)$ will give less weight on both large residuals and x-outliers for some particular choice of ρ. For example, we consider

$$\rho(\mathbf{x}, e) = v^2(\mathbf{x}) \rho_c(\frac{e}{v(\mathbf{x})}) \tag{1.2.8}$$

with

$$\rho_c(t) = \begin{cases} \frac{1}{2}t^2 , & \text{for } |t| \leq c , \\ c|t| - \frac{1}{2}c^2 , & \text{for } |t| > c , \end{cases} \tag{1.2.9}$$

and

$$v(\mathbf{x}) = 1/d(\mathbf{x}),$$

here $d(\mathbf{x})$ is a "distance" function of $\mathbf{x}$. Hopefully, we can find out some suitable $d(\mathbf{x})$ which gives large values for high leverage points. Then

$$\begin{aligned} w(\mathbf{x}, e) &= \frac{\partial}{\partial e} \sqrt{2\rho(\mathbf{x}, e)} \cdot \text{sign}(e) \\ &= \begin{cases} 1 , & \text{for } |\frac{e}{v(\mathbf{x})}| \leq c , \\ \frac{cv(\mathbf{x})}{v(\mathbf{x})\sqrt{2c|\frac{e}{v(\mathbf{x})}| - c^2}} , & \text{for } |\frac{e}{v(\mathbf{x})}| > c , \end{cases} \end{aligned} \tag{1.2.10}$$

and notice that when $|\frac{e}{v(\mathbf{x})}| > c$,

$$w(\mathbf{x}, e) = \frac{cv(\mathbf{x})}{v(\mathbf{x})\sqrt{2c \cdot |\frac{e}{v(\mathbf{x})}| - c^2}} < \frac{cv(\mathbf{x})}{v(\mathbf{x})\sqrt{2c \cdot c \cdot -c^2}} = 1.$$

Thus

$$w(\mathbf{x}, e) = \begin{cases} 1 , & \text{for } |\frac{e}{v(\mathbf{x})}| \leq c , \\ < 1 , & \text{for } |\frac{e}{v(\mathbf{x})}| > c . \end{cases} \tag{1.2.11}$$

3. Let's consider the following function,

$$\rho(\mathbf{x}_i, e) = \begin{cases} \frac{1}{2}e^2, & \text{for } \frac{a(R_i^+)}{v(\mathbf{x}_i)} \leq c , \\ \frac{1}{2} \frac{c^2 v^2(\mathbf{x}_i)}{a^2(R_i^+)} e^2 , & \text{for } \frac{a(R_i^+)}{v(\mathbf{x}_i)} > c , \end{cases} \tag{1.2.12}$$

where $a(R_i^+) = \frac{R_i^+}{n+1}$, so $a(R_i^+)\cdot\text{sign}(e_i)$ is Wilcoxon score function. Then

$$w(\mathbf{x}, e) = \frac{\partial}{\partial e}\sqrt{2\rho(\mathbf{x}, e)}\cdot \text{sign}(e)$$
$$= \begin{cases} 1\,, & \text{for } \frac{a(R_i^+)}{v(\mathbf{x}_i)} \le c\,, \\ \frac{cv(\mathbf{x}_i)}{a(R_i^+)}\,, & \text{for } \frac{a(R_i^+)}{v(\mathbf{x}_i)} > c\,. \end{cases} \tag{1.2.13}$$

These are Tableman's weights.

If we use

$$\psi_c(t) = \min(t, c)\,, \quad t > 0,$$

and put (1.2.13) into (1.2.7), we get

$$S(\boldsymbol{\theta}) = \sum_{i=1}^{n} v(\mathbf{x}_i)\psi_c[\frac{a(R_i^+)}{v(\mathbf{x}_i)}]\cdot \text{sign}(e_i)\mathbf{x}_i\,. \tag{1.2.14}$$

In (1.2.14), we now let ψ_c be an arbitrary function ψ. Rather than using Wilcoxon weights, we put

$$a(i) = \phi^+(\frac{i}{n+1})\,,$$

where ϕ^+ is an arbitrary function. Also we let v be another arbitrary function. Further ψ, ϕ^+ and v satisfy the following assumptions:

i) $v(\mathbf{x}_i) > 0$ for all $\mathbf{x}_i$,

ii) $\phi^+(u)$, $(0 < u < 1)$, is nonnegative, nondecreasing and absolutely continuous,

iii) $\psi(u)$, $(u > 0)$, is nonnegative, bounded, nondecreasing and absolutely continuous.

Finally, we define that our general bounded influence R-estimator is the solution to

$$S(\boldsymbol{\theta}) = \mathbf{0}\,, \tag{1.2.15}$$

where

$$S(\boldsymbol{\theta}) = \sum_{i=1}^{n} v(\mathbf{x}_i)\psi[\frac{\phi^+(R_i^+/(n+1))}{v(\mathbf{x}_i)}]\cdot \text{sign}(y_i - \mathbf{x}_i^T\boldsymbol{\theta})\mathbf{x}_i \tag{1.2.16}$$

with ψ, ϕ^+ and v satisfying assumptions i)-iii).

For the particular choices of (v, ψ, ϕ^+), (1.2.16) produces Tableman's estimator and Jaeckel's estimator. In fact, if

$$\psi(t) = t,$$
$$\phi^+(R_i^+/n+1) = a(R_i^+),$$

where $a(i)$ is a score function, then (1.2.16) gives

$$S(\boldsymbol{\theta}) = \sum_{i=1}^{n} a(R_i^+)\text{sign}(y_i - \mathbf{x}_i^T\boldsymbol{\theta})\mathbf{x}_i,$$

i.e.

$$S(\boldsymbol{\theta}) = \sum_{i=1}^{n} a(R_i)\mathbf{x}_i. \tag{1.2.17}$$

$S(\boldsymbol{\theta}) = \mathbf{0}$ defines the Jaeckel's estimator.

If

$$\psi(t) = \psi_c(t) = \min(t, c), \quad t > 0, \text{ for given } c,$$
$$\phi^+(R_i^+/n+1) = a(R_i^+),$$

here $a(R_i^+) \cdot \text{sign}(y_i - \mathbf{x}_i^T\boldsymbol{\theta})$ is Wilcoxon score function, then $S(\boldsymbol{\theta}) = \mathbf{0}$ gives Tableman's estimator.

CHAPTER 2
SOME ASYMPTOTIC RESULTS

When we introduce a new estimator, we always want to know its asymptotic behavior or properties such as: the limiting distribution of $\sqrt{n}(\boldsymbol{\theta}(F_n) - \boldsymbol{\theta}(F))$, asymptotic variance and asymptotic efficiency. For this reason, we have to get an asymptotically equivalent equation for our bounded influence R-estimator. Based on the asymptotic equation, some properties are obtained. In section one, an equivalently asymptotic equation is derived and functional $\boldsymbol{\theta}(F)$ is defined. Section two proves the asymptotic uniqueness of the estimator. In section three we will get influence function and asymptotic variance for the estimator.

2.1 Asymptotic Equation for the Bounded Influence R-estimator

In the previous chapter, we derived the general form for the bounded influence R-estimator which is the solution to (1.2.15), i.e. $S(\boldsymbol{\theta}) = \mathbf{0}$, where

$$S(\boldsymbol{\theta}) = \sum_{i=1}^{n} v(\mathbf{x}_i)\psi[\frac{\phi^+(R_i^+/(n+1))}{v(\mathbf{x}_i)}]\text{sign}(y_i - \mathbf{x}_i^T\boldsymbol{\theta})\mathbf{x}_i \, .$$

We try to express $S(\boldsymbol{\theta})$ according to some empirical distributions.

Let G_n be the empirical distribution of $y_1 - \mathbf{x}_1^T\boldsymbol{\theta}, \ldots, y_n - \mathbf{x}_n^T\boldsymbol{\theta}$, then

$$R_i^+ = \{nG_n(y_i - \mathbf{x}_i^T\boldsymbol{\theta}) - nG_n[-(y_i - \mathbf{x}_i^T\boldsymbol{\theta})]\} \cdot \text{sign}(y_i - \mathbf{x}_i^T\boldsymbol{\theta}),$$
$$i = 1, \ldots, n \, , \qquad (2.1.1)$$

and

$$\phi^+(\frac{R_i^+}{n+1}) = \phi^+\{[\frac{n}{n+1}G_n(y_i - \mathbf{x}_i^T\boldsymbol{\theta}) - \frac{n}{n+1}G_n(-y_i + \mathbf{x}_i^T\boldsymbol{\theta})] \cdot \text{sign}(y_i - \mathbf{x}_i^T\boldsymbol{\theta})\},$$
$$i = 1, \ldots, n. \qquad (2.1.2)$$

Let H_n be the design measure, i.e. the distribution function defined by

$$H_n(V) = \frac{\#\text{ of } \mathbf{x}_i\text{'s in } V \subset \mathbb{R}^k}{n} \, , \qquad (2.1.3)$$

then the empirical distribution of $(\mathbf{x}_i, y_i)$ is given by $F_n(\mathbf{x}_i, y_i) = (H_n \times G_n)(\mathbf{x}_i, y_i - \mathbf{x}_i^T\boldsymbol{\theta})$. Now we put (2.1.2) and (2.1.3) in (1.2.15), get

$$S(\boldsymbol{\theta}) = \int \mathbf{x}v(\mathbf{x})\psi\{\frac{1}{v(\mathbf{x})}\phi^+[(\frac{n}{n+1}G_n(y - \mathbf{x}^T\boldsymbol{\theta}) - \frac{n}{n+1}G_n(-y + \mathbf{x}^T\boldsymbol{\theta})) \cdot \text{sign}(y - \mathbf{x}^T\boldsymbol{\theta})]\} \cdot \text{sign}(y - \mathbf{x}^T\boldsymbol{\theta})dF_n(\mathbf{x}, y) = \mathbf{0}\,. \tag{2.1.4}$$

If

$$\int \psi\{\frac{1}{v(\mathbf{x})}\phi^+[(G(e) - G(-e)) \cdot \text{sign}(e)]\} \cdot \text{sign}(e)dG(e) = 0 \tag{2.1.5}$$

for all $\mathbf{x}$, then (2.1.4) holds if F_n is replaced by F, and we have

$$\int \mathbf{x}v(\mathbf{x})\psi\{\frac{1}{v(\mathbf{x})}\phi^+[(G(y - \mathbf{x}^T\boldsymbol{\theta}) - G(-y + \mathbf{x}^T\boldsymbol{\theta})) \cdot \text{sign}(y - \mathbf{x}^T\boldsymbol{\theta})]\} \cdot \text{sign}(y - \mathbf{x}^T\boldsymbol{\theta})dF(\mathbf{x}, y) = \mathbf{0}. \tag{2.1.6}$$

Thus (2.1.6) defines the functional $\boldsymbol{\theta}(F)$ that extends estimator (2.1.4) to general distribution F, such an estimator will be Fisher consistent for $\boldsymbol{\theta}$ provided (2.1.5) holds. Under mild regularity conditions, $\boldsymbol{\theta}(F)$ is weakly continuous, therefore $\boldsymbol{\theta}(F)$ is consistent for $\boldsymbol{\theta}$ (see Huber (1981) for more details about consistency and Fisher consistency).

If G is a symmetric distribution, then (2.1.5) and (2.1.6) become

$$\int \psi\{\frac{1}{v(\mathbf{x})}\phi^+[(2G(e) - 1) \cdot \text{sign}(e)]\} \cdot \text{sign}(e)dG(e) = 0, \tag{2.1.7}$$

and

$$\int \mathbf{x}v(\mathbf{x})\psi\{\frac{1}{v(\mathbf{x})}\phi^+[(2G(y - \mathbf{x}^T\boldsymbol{\theta}) - 1) \cdot \text{sign}(y - \mathbf{x}^T\boldsymbol{\theta})]\} \cdot \text{sign}(y - \mathbf{x}^T\boldsymbol{\theta})dF(\mathbf{x}, y) = \mathbf{0} \tag{2.1.8}$$

respectively. For this case, (2.1.7) always holds, since $(2G(e) - 1) \cdot \text{sign}(e)$ is a even function of e when $G(e)$ is symmetric around 0 and sign(e) is odd. That means $\boldsymbol{\theta}(F)$ is always Fisher consistent for $\boldsymbol{\theta}$ as long as the error is distributed symmetrically. This is a very good property.

Suppose G is not symmetric, then $\boldsymbol{\theta}(F)$ might be biased, i.e.

$$\int\int \mathbf{x}v(\mathbf{x})\psi\{\frac{1}{v(\mathbf{x})}\phi^{+}[(G(e)-G(-e))\text{sign}(e)]\}\cdot\text{sign}(e)dG(e)dH(\mathbf{x})\neq\mathbf{0}.$$

But if we can allow some bias in the intercept, then we can choose a such that (not proved)

$$\int\int \mathbf{x}(v(\mathbf{x})\psi\{\frac{1}{v(\mathbf{x})}o^{+}[(G(e+a)-G(-e-a))\cdot\text{sign}(e+a)]\}$$
$$\cdot\text{sign}(e+a)dG(e)dH(\mathbf{x})=\mathbf{0},$$

here a is independent of $(x_1,x_2,\ldots,x_k)^T$. This means that the root $\boldsymbol{\theta}$ satisfies

$$\begin{aligned} y-\mathbf{x}^T\boldsymbol{\theta} &= e+\mathbf{x}^T(\boldsymbol{\theta}_*-\boldsymbol{\theta}), \quad \text{(here } \boldsymbol{\theta}_* \text{ is the true value)} \\ &= e+(\theta_{0*}-\theta_0)+x_1(\theta_{1*}-\theta_1)+\cdots+x_k(\theta_{k*}-\theta_k) \\ &= e+a; \end{aligned}$$

i.e. $a=\theta_{0*}-\theta_0$, $\theta_{1*}-\theta_1=0,\ldots,\theta_{k*}-\theta_k=0$, since a is independent of $(x_1,x_2,\ldots,x_k)^T$. Therefore the bias is all in the intercept.

Further, let $\boldsymbol{\theta}_*$ be the true value of θ in model (1.2.1) under the joint distribution F_0 of $(\mathbf{x},e)$, i.e.

$$y_i=\mathbf{x}_i^T\boldsymbol{\theta}_*+e_i\,,\quad i=1,2,\ldots n, \tag{2.1.9}$$

and

$$\boldsymbol{\theta}_*=\boldsymbol{\theta}(F_0)\,. \tag{2.1.10}$$

So

$$y_i-\mathbf{x}_i^T\boldsymbol{\theta}=e_i+\mathbf{x}_i^T(\boldsymbol{\theta}_*-\boldsymbol{\theta})\,,\quad i=1,\ldots,n. \tag{2.1.11}$$

Combining (2.1.11) and (2.1.8), we get

$$S(\boldsymbol{\theta})=\int \mathbf{x}v(\mathbf{x})\psi\{\frac{1}{v(\mathbf{x})}\phi^{+}[(2G(e+\mathbf{x}^T(\boldsymbol{\theta}_*-\boldsymbol{\theta}))-1)\cdot\text{sign}(e+\mathbf{x}^T(\boldsymbol{\theta}_*-\boldsymbol{\theta}))]\}$$
$$\cdot\text{sign}(e+\mathbf{x}^T(\boldsymbol{\theta}_*-\boldsymbol{\theta}))dF(\mathbf{x},e)=\mathbf{0}\,. \tag{2.1.12}$$

This is the asymptotic equation for the bounded influence R-estimator.

Write

$$\eta_G(\mathbf{x}, e) = v(\mathbf{x})\psi[\frac{\phi^+(|G(e) - G(-e)|)}{v(\mathbf{x})}]\text{sign}(e),$$

then the asymptotic definition (2.1.12) may be written in a more compact form as

$$S(\boldsymbol{\theta}) = E_F[\mathbf{x}\eta_G(\mathbf{x}, e + \mathbf{x}^T(\boldsymbol{\theta}_* - \boldsymbol{\theta}))] = \mathbf{0}.$$

From now on we always assume that G is a symmetric distribution.

2.2 Uniqueness

In this section we rewrite the linear model in this form

$$y = \mathbf{x}^T\boldsymbol{\theta} + \varepsilon = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_k x_k + \varepsilon, \tag{2.2.1}$$

and true value of θ in (2.2.1) under the joint distribution F_0 of $(\mathbf{x}, y)$ is $\boldsymbol{\theta}_* = (\theta_{0*}, \theta_{1*}, \ldots \theta_{k*})^T$. From previous section, we get that the asymptotic equation for the bounded influence R-estimator is (2.1.12). It is obvious that $\boldsymbol{\theta} = \boldsymbol{\theta}_*$ is a solution to (2.1.12) by (2.1.7). The question is whether the solution to (2.1.12) is unique or not. The following theorem gives us the answer.

THEOREM 1. *(Uniqueness)*

Assuming $v(\mathbf{x})$ is an even function of every x_j, i.e.

$$v((x_1, \ldots, x_j, \ldots, x_k)^T) = v((x_1, \ldots, -x_j, \ldots, x_k)^T)\ ,\ j = 1, \ldots, k \tag{2.2.2}$$

then the solution to (2.1.12) is unique, and equals $\boldsymbol{\theta}_*$ under the joint distribution F_0 which satisfies the following assumptions: ($F_0 = G_0 \times H_0$)

a) The distribution G_0 of e is symmetric,

b) The probability density function $h_0((x_1, \ldots, x_k)^T)$ of $\mathbf{x}$ is an even function of every x_j, i.e.

$$h_0((x_1, \ldots, x_j, \ldots, x_k)^T) = h_0((x_1, \ldots, -x_j, \ldots, x_k)^T), \quad j = 1, \ldots, k.$$

We will prove this theorem later. First let's discuss some properties about $A(\mathbf{x}, u)$, here

$$A(\mathbf{x}, u) = \int \psi\{\frac{1}{v(\mathbf{x})}\phi^{+}[(2G_0(e+u)-1)\cdot \mathrm{sign}(e+u)]\}\cdot \mathrm{sign}(e+u)dG_0(e), \quad (2.2.4)$$

with

$$u = \mathbf{x}^T(\boldsymbol{\theta}_* - \boldsymbol{\theta}). \quad (2.2.5)$$

The solution to (2.1.12) satisfies

$$S(\boldsymbol{\theta}) = \int \mathbf{x}v(\mathbf{x})A(\mathbf{x}, \mathbf{x}^T(\boldsymbol{\theta}_* - \boldsymbol{\theta}))dH_0(\mathbf{x}) = \mathbf{0}.$$

Property 1. $A(\mathbf{x}, -u) = -A(\mathbf{x}, u)$.

PROOF:

$$\begin{aligned}
A(\mathbf{x}, -u) &= \int \psi[\frac{1}{v(\mathbf{x})}\phi^{+}\{(2G_0(e-u)-1)\cdot \mathrm{sign}(e-u)\}] \\
&\quad \cdot \mathrm{sign}(e-u)dG_0(e) \\
&= \int \psi[\frac{1}{v(\mathbf{x})}\phi^{+}\{(2G_0(-e-u)-1)\cdot \mathrm{sign}(-e-u)\}] \\
&\quad \cdot \mathrm{sign}(-e-u)dG_0(e) \quad \text{(by symmetry of } G_0\text{)} \\
&= -\int \psi[\frac{1}{v(\mathbf{x})}\phi^{+}\{(2G_0(e+u)-1)\cdot \mathrm{sign}(e+u)\}] \\
&\quad \cdot \mathrm{sign}(e+u)dG_0(e) \quad \text{(since } G_0(-e-u) = 1 - G_0(e+u)\text{)} \\
&= -A(\mathbf{x}, u).
\end{aligned}$$

□

Property 2.

$$A(\mathbf{x}, u) \begin{cases} 0, & \text{for } u = 0, \\ > 0, & \text{for } u > 0, \\ < 0, & \text{for } u < 0. \end{cases}$$

PROOF: For $u = 0$,

$$A(\mathbf{x}, 0) = \int \psi[\frac{1}{v(\mathbf{x})}\phi^+\{(2G_0(e) - 1) \cdot \text{sign}(e)\}] \cdot \text{sign}(e) dG_0(e)$$
$$= 0 \quad \text{(by (2.1.7))}.$$

For $u > 0$, since ϕ^+ and ψ are all nondecreasing functions,

$$\psi[\frac{1}{v(\mathbf{x})}\phi^+\{(2G_0(e + u) - 1) \cdot \text{sign}(e + u)\}] \cdot \text{sign}(e + u)$$
$$> \psi[\frac{1}{v(\mathbf{x})}o^+\{(2G_0(e) - 1) \cdot \text{sign}(e)\}] \cdot \text{sign}(e).$$

Hence $A(\mathbf{x}, u) > A(\mathbf{x}, 0) = 0$ when $u > 0$. By Property 1 now we have $A(\mathbf{x}, u) < 0$ when $u < 0$. □

Property 3. If $|u_1| > |u_2|$, then $|A(\mathbf{x}, u_1)| > |A(\mathbf{x}, u_2)|$.

This proof is similar to the proof of Property 2.

Now we can prove Theorem 1. The proof consists of two parts. Part 1 proves the result in special cases that the linear response in (2.2.1) goes through the origin, and Part 2 proves the general result.

Part 1 of the proof. Let's consider the linear model (2.2.1) without constant term θ_0, i.e.

$$y = x_1\theta_1 + \cdots + x_k\theta_k + e\,, \tag{2.2.8}$$

and true value

$$\boldsymbol{\theta}_* = (\theta_{1*}, \ldots, \theta_{k*})^T\,. \tag{2.2.9}$$

From (2.2.5),

$$u = \mathbf{x}^T(\boldsymbol{\theta}_* - \boldsymbol{\theta}) = x_1(\theta_{1*} - \theta_1) + \cdots + x_k(\theta_{k*} - \theta_k). \tag{2.2.10}$$

Using $A(\mathbf{x}, u)$, we can rewrite (2.1.12) as

$$S(\boldsymbol{\theta}) = \int \mathbf{x} v(\mathbf{x}) A(\mathbf{x}, u) dH_0(\mathbf{x}) \tag{2.2.11}$$

$$= \int \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{pmatrix} v(\mathbf{x}) A(\mathbf{x}, u) dH_0(\mathbf{x})$$

$$= (S_1(\boldsymbol{\theta}), S_2(\boldsymbol{\theta}), \ldots, S_k(\boldsymbol{\theta}))^T , \tag{2.2.12}$$

where

$$S_i(\boldsymbol{\theta}) = \int x_i v(\mathbf{x}) A(\mathbf{x}, u) dH_0(\mathbf{x}) , \quad i = 1, \ldots, k . \tag{2.2.13}$$

We will show, from $S_i(\boldsymbol{\theta}) = 0$, that θ_i is uniquely determined and equals θ_{i*}, $i = 1, \ldots, k$.

Let's start with equation $S_1(\boldsymbol{\theta}) = 0$.

Denote

$$a_1 = \theta_{1*} - \theta_1,$$

$$a_2 = x_2(\theta_{2*} - \theta_2) + \cdots + x_k(\theta_{k*} - \theta_k). \tag{2.2.14}$$

Using a_1 and a_2, (2.2.10) can be rewritten as

$$u = a_1 x_1 + a_2. \tag{2.2.15}$$

We claim that

$$S_1(\boldsymbol{\theta}) = \begin{cases} 0 , & \text{for } a_1 = 0 , \\ > 0 , & \text{for } a_1 > 0 , \\ < 0 , & \text{for } a_1 < 0 . \end{cases}$$

This means, from $S_1(\boldsymbol{\theta}) = 0$, that we must have $a_1 = 0$, i.e. $\theta_1 = \theta_{1*}$. Well, for $a_1 > 0$,

$$S_1(\boldsymbol{\theta}) = \int x_1 v(\mathbf{x}) A(\mathbf{x}, u) dH_0(\mathbf{x})$$

$$= \int_{\{x_2, \ldots, x_k\}} \cdots \int \{ \int_{-\infty}^{\infty} x_1 v(\mathbf{x}) A(\mathbf{x}, u) h_0((x_1, \ldots, x_k)^T) dx_1 \} dx_2, \ldots dx_k$$

$$= T_1 + T_2 + T_3,$$

where

$$T_i = \int_{c_i} \cdots \int \{\int_{-\infty}^{\infty} x_1 v(\mathbf{x}) A(\mathbf{x}, u) h_0((x_1, \ldots, x_k)^T) dx_1\} dx_2, \ldots, dx_k, \quad i = 1, 2, 3.$$

with

$$c_1 = \{(x_2, \ldots, x_k) | a_2 < 0\},$$
$$c_2 = \{(x_2, \ldots, x_k) | a_2 > 0\},$$
$$c_3 = \{(x_2, \ldots, x_k) | a_2 = 0\}.$$

We want to show that each $T_i > 0$, $i = 1, 2, 3$. Then $S_1(\theta) = T_1 + T_2 + T_3 > 0$ when $a_1 > 0$.

Now, in T_3, $a_1 > 0$, $a_2 = 0$, so $u = a_1 x_1 + a_2 = a_1 x_1$, and

$$\begin{aligned} A((-x_1, x_2, \ldots, x_k)^T, -a_1 x_1) &= A((x_1, x_2, \ldots, x_k)^T, -a_1 x_1) \\ &= -A((x_1, x_2, \ldots, x_k)^T, a_1 x_1) \end{aligned}$$

by Property 1. Thus $A(\mathbf{x}, u)$ is an odd function of x_1 in T_3. So we have

$$\begin{aligned} &\int_{-\infty}^{\infty} x_1 v(\mathbf{x}) A(\mathbf{x}, u) h_0((x_1, \ldots, x_k)^T) dx_1 \\ &\quad = 2 \int_0^{\infty} x_1 v(\mathbf{x}) A(\mathbf{x}, u) h_0((x_1, \ldots, x_k)^T) dx_1 > 0, \end{aligned} \tag{2.2.17}$$

since v and h_0 are even functions of x_1 by the assumptions of Theorem 1, and $A(\mathbf{x}, u) > 0$ when $u = a_1 x_1 > 0$ by property 2. Hence

$$\begin{aligned} T_3 &= \int_{c_3} \cdots \int \{\int_{-\infty}^{\infty} x_1 v(\mathbf{x}) A(\mathbf{x}, u) h_0((x_1, \ldots, x_k)^T) dx_1\} dx_2 \ldots dx_k \\ &> 0. \quad \text{(by (2.2.17))} \end{aligned}$$

Next we deal with T_1 and T_2. After making transformation:

$$(x_1, x_2, \ldots, x_k) \to (-x_1, -x_2, \ldots, -x_k),$$

we find $c_1 \to c_2$ and $T_1 = T_2$ since $-x_1 A((-\mathbf{x}), -u) = -x_1 A(\mathbf{x}, -u) = -x_1 \cdot (-A(\mathbf{x}, u)) = x_1 A(\mathbf{x}, u)$.

In T_2, let's consider the integration

$$\int_{-\infty}^{\infty} x_1 v(\mathbf{x}) A(\mathbf{x}, u) h_0((x_1, \ldots, x_k)^T dx_1 = L_1 + L_2 + L_3,$$

where

$$L_1 = \int_{-\infty}^{-\frac{a_2}{a_1}} x_1 v(\mathbf{x}) A(\mathbf{x}, u) h_0((x_1, \ldots, x_k)^T) dx_1,$$

$$L_2 = \int_{-\frac{a_2}{a_1}}^{\frac{a_2}{a_1}} x_1 v(\mathbf{x}) A(\mathbf{x}, u) h_0((x_1, \ldots, x_k)^T) dx_1,$$

and

$$L_3 = \int_{\frac{a_2}{a_1}}^{\infty} x_1 u(\mathbf{x}) A(\mathbf{x}, u) h_0((x_1, \ldots, x_k)^T) dx_1.$$

For L_1, $u = x_1 a_1 + a_2 < 0$, so $A(\mathbf{x}, u) < 0$ by property 2. Then $x_1 \cdot A(\mathbf{x}, u) > 0$ and $L_1 > 0$. For L_3, $u = x_1 a_1 + a_2 > 0$, then $A(\mathbf{x}, u) > 0$. So $L_3 > 0$. For L_2,

$$\begin{aligned} L_2 &= \left(\int_{-\frac{a_2}{a_1}}^{0} + \int_{0}^{\frac{a_2}{a_1}}\right)(x_1 \cdot v(\mathbf{x}) A(\mathbf{x}, u) h_0((x_1, \ldots, x_k)^T)) dx_1 \\ &= \int_0^{\frac{a_2}{a_1}} x_1 v(\mathbf{x}) [A(\mathbf{x}, u) - A(\mathbf{x}, -x_1 a_1 + a_2)] h_0((x_1, \ldots, x_k)^T) dx_1, \end{aligned}$$

Since $u = a_1 x_1 + a_2 > -a_1 x_1 + a_2 > 0$ (for $0 < x_1 < \frac{a_2}{a_1}, a_1 > 0$ and $a_2 > 0$), $A(\mathbf{x}, u) > A(\mathbf{x}, -x_1 a_1 + a_2) > 0$ (by Property 3). Thus $L_2 > 0$.

Therefore

$$T_1 = T_2 = \int_{c_2} \cdots \int (L_1 + L_2 + L_3) dx_2 \ldots dx_k > 0.$$

From what has been shown, we have $S_1(\boldsymbol{\theta}) > 0$ when $a_1 > 0$. Similarly we can get $S_1(\boldsymbol{\theta}) < 0$ when $a_1 < 0$. When $a_1 = 0$, $S_1(\boldsymbol{\theta}) = 0$ is obvious. In conclusion, $\theta_1 = \theta_{1*}$ is uniquely determined by $S_1(\boldsymbol{\theta}) = 0$. In a similar way, $\theta_i = \theta_{i*}$ ($i =$

$2, \ldots, k)$ is uniquely determined by $S_i(\boldsymbol{\theta}) = 0$. On the whole $\boldsymbol{\theta} = \boldsymbol{\theta}^*$ is the unique solution to $S(\boldsymbol{\theta}) = \mathbf{0}$, completing the proof of part 1.

Part 2 of the proof , we consider the general model (2.2.1) with constant terms. $S(\boldsymbol{\theta})$ in (2.2.12) can be written as

$$S(\boldsymbol{\theta}) = \begin{pmatrix} S_0(\boldsymbol{\theta}) \\ S_1(\boldsymbol{\theta}) \\ \vdots \\ S_k(\boldsymbol{\theta}) \end{pmatrix} \tag{2.2.18}$$

where

$$S_0(\boldsymbol{\theta}) = \int v(\mathbf{x}) A(\mathbf{x}, u) dH_0(\mathbf{x}), \tag{2.2.19}$$

and $S_i(\boldsymbol{\theta})$ $(i = 1, \ldots, k)$ is the same as (2.2.13) with

$$u = \theta_{0*} - \theta_0 + x_1(\theta_{1*} - \theta_1) + \cdots + x_k(\theta_{k*} - \theta_k)\,. \tag{2.2.20}$$

In the present case, we only have to show, from $S_0(\boldsymbol{\theta}) = 0$, that θ_0 is uniquely determined and equals to θ_{0*}, because all the $S_i(\boldsymbol{\theta})$ $(i = 1, \ldots, k)$ will be exactly the same as in previous case if $\theta_0 = \theta_{0*}$ and from part 1 of the proof $\theta_i = \theta_{i*}$ is uniquely determined by $S_i(\boldsymbol{\theta}) = 0$ $(i = 1, \ldots, k)$.

Now, let $a_3 = \theta_{0*} - \theta_0$ and $a_4 = x_1(\theta_{1*} - \theta_1) + \cdots + x_k(\theta_{k*} - \theta_k)$,then

$$\begin{aligned} u &= \theta_{0*} + \theta_0 + x_1(\theta_{1*} - \theta_1) + \cdots + x_k(\theta_{k*} - \theta_k) \\ &= a_3 + a_4\,. \end{aligned} \tag{2.2.21}$$

Denote

$$\begin{aligned} c_1 &= \{(x_1, \ldots, x_k) | a_4 < 0\}\,, \\ c_2 &= \{(x_1, \ldots, x_k) | a_4 > 0\}\,, \\ c_3 &= \{(x_1, \ldots, x_k) | a_4 = 0\}\,. \end{aligned}$$

Using c_1, c_2 and c_3, $S_0(\boldsymbol{\theta})$ can be expressed as

$$\begin{aligned} S_0(\boldsymbol{\theta}) &= \int v(\mathbf{x}) A(\mathbf{x}, u) dH_0(\mathbf{x}) \\ &= I_1 + I_2 + I_3 \,, \end{aligned} \tag{2.2.22}$$

with

$$I_i = \int_{c_i} v(\mathbf{x}) A(\mathbf{x}, u) dH_0(\mathbf{x}). \tag{2.2.23}$$

We claim

$$S_0(\boldsymbol{\theta}) = \begin{cases} > 0 \,, & \text{for } a_3 > 0 \,, \\ < 0 \,, & \text{for } a_3 < 0 \,, \\ = 0 \,, & \text{for } a_3 = 0 \,. \end{cases}$$

This means, from $S_0(\boldsymbol{\theta}) = 0$, that $a_3 = 0$, i.e. $\theta_0 = \theta_{0*}$.

Suppose $a_3 > 0$. In I_3, $u = a_3 + a_4 = a_3 > 0$, then $A(\mathbf{x}, u) > 0$, so $I_3 > 0$. For I_1 and I_2, we have

$$I_1 + I_2 = \int_{c_2} v(\mathbf{x}) [A(\mathbf{x}, u) + A(\mathbf{x}, a_3 - a_4)] dH_0(\mathbf{x}), \tag{2.2.24}$$

and on c_2, $u = a_3 + a_4 > |a_3 - a_4|$, since $a_3 > 0$ and $a_4 > 0$. So

$$A(\mathbf{x}, u) > |A(\mathbf{x}, a_3 - a_4)| \,, \quad \text{by Property 3.}$$

Thus

$$A(\mathbf{x}, u) + A(\mathbf{x}, a_3 - a_4) > 0,$$

when $a_3 > 0$ consequently $S_0(\boldsymbol{\theta}) = I_1 + I_1 + I_3 > 0$. Similarly we can show $S_0(\boldsymbol{\theta}) < 0$ when $a_3 < 0$. Therefore $a_3 = 0$, i.e. $\theta_0 = \theta_{0*}$, is uniquely determined by $S_0(\boldsymbol{\theta}) = 0$. Combining $\theta_0 = \theta_{0*}$ with the result in part 1, we conclude that $\boldsymbol{\theta} = \boldsymbol{\theta}_*$ is the unique solution to $S(\boldsymbol{\theta}) = \mathbf{0}$ in general case. This is the end of proving Theorem 1.

Next section will calculate the influence function for the estimator.

2.3 Influence function

In section 1.1, we gave the definition of influence function when $\theta,\ x \in \mathbb{R}$. Here we will extend that definition when $\boldsymbol{\theta},\ \mathbf{z} \in \mathbb{R}^k$. Let $\delta\mathbf{z}$ be the distribution which puts mass 1 at the point $\mathbf{z} \in \mathcal{X}$ (where $\mathcal{X}$ is the sample space of $\mathbf{z}$), then the influence function for a functional $\boldsymbol{\theta}(F)$ at a distribution F_0 is given by

$$\mathbb{IF}(\mathbf{z},\boldsymbol{\theta},F_0) = \lim_{\varepsilon\to 0}\{(\boldsymbol{\theta}((1-\varepsilon)F_0 + \varepsilon\delta\mathbf{z}) - \boldsymbol{\theta}(F_0))/\varepsilon\}.$$

The above influence function usually can be obtained by simple calculation as follows:

$$\mathbb{IF}(\mathbf{z},\boldsymbol{\theta},F_0) = \frac{d}{d\varepsilon}\boldsymbol{\theta}((1-\varepsilon)F_0 + \varepsilon G)|_{\substack{\varepsilon=0\\ G=\delta\mathbf{z}}}\ . \tag{2.3.1}$$

Under some regularity conditions, we have

$$\int \mathbb{IF}(\mathbf{z},\boldsymbol{\theta},F_0)dF_0(\mathbf{z}) = \mathbf{0} \tag{2.3.2}$$

and

$$\sqrt{n}(\boldsymbol{\theta}(F_n) - \boldsymbol{\theta}(F_0)) \xrightarrow{\mathcal{L}} N(0, V(\boldsymbol{\theta},F_0))\,, \tag{2.3.3}$$

where

$$V(\boldsymbol{\theta},F_0) = \int \mathbb{IF}(\mathbf{z},\boldsymbol{\theta},F_0)\cdot\mathbb{IF}(\mathbf{z},\boldsymbol{\theta},F_0)^T dF_0(\mathbf{z}). \tag{2.3.4}$$

In section 2.1, (2.1.12) defines the functional $\boldsymbol{\theta}(F)$. In order to get the influence function for $\boldsymbol{\theta}(F)$ at F_0, we replace F by $F_\lambda = (1-\lambda)F_0 + \lambda F_1$ in (2.1.12), then $\boldsymbol{\theta}(F_\lambda) = \boldsymbol{\theta}_\lambda$ is defined implicitly. Accordingly, the influence function for $\boldsymbol{\theta}(F)$ can be derived by

$$\mathbb{IF}(e,\mathbf{x},G_0,H_0,v,\phi^+,\psi) = \frac{d}{d\lambda}\boldsymbol{\theta}_\lambda|_{\substack{\lambda=0\\ G_1=\delta e, H_1=\delta\mathbf{x}}}.$$

For F_λ we have

$$F_\lambda(\mathbf{x},e) = (1-\lambda)H_0(\mathbf{x})G_0(e) + \lambda(H_1(\mathbf{x})G_1(e).$$

Let $\mathbf{x} \to +\infty$ (by which we mean each component of $\mathbf{x}$ goes to positive infinity) on both side, i.e.

$$\lim_{\mathbf{x}\to+\infty} F_\lambda(\mathbf{x}, e) = \lim_{x\to+\infty}(1-\lambda)H_0(\mathbf{x})G_0(e) + \lim_{x\to+\infty}\lambda H_1(\mathbf{x})G_1(e).$$

Since

$$\lim_{\mathbf{x}\to+\infty} F_\lambda(\mathbf{x}, e) = G_\lambda(e) \text{ and } \lim_{\mathbf{x}\to+\infty} H_i(\mathbf{x}) = 1 \quad (i = 0, 1)$$

we have that

$$G_\lambda(e) = (1-\lambda)G_0(e) + \lambda G_1(e).$$

Similarly, we ahve

$$H_\lambda(\mathbf{x}) = (1-\lambda)H_0(\mathbf{x}) + \lambda H_1(\mathbf{x}).$$

Denote

$$\ell(e, \mathbf{x}, G) = [2G(e + \mathbf{x}^T(\boldsymbol{\theta}_* - \boldsymbol{\theta})) - 1] \cdot \text{sign}(e + \mathbf{x}^T(\boldsymbol{\theta}_* - \boldsymbol{\theta})),$$

and

$$\begin{aligned}\ell(e, \mathbf{x}, G_\lambda) = (1-\lambda)&[2G_0(e + \mathbf{x}^T(\boldsymbol{\theta}_* - \boldsymbol{\theta}_\lambda)) - 1] \cdot \text{sign}(e + \mathbf{x}^T(\boldsymbol{\theta}_* - \boldsymbol{\theta}_\lambda)) \\ &+ \lambda[2G_1(e + \mathbf{x}^T(\boldsymbol{\theta}_* - \boldsymbol{\theta}_\lambda)) - 1] \cdot \text{sign}(e + \mathbf{x}^T(\boldsymbol{\theta}_* - \boldsymbol{\theta}_\lambda)),\end{aligned}$$

then

$$\begin{aligned}\frac{d}{d\lambda}\ell(e, \mathbf{x}, G_\lambda)|_{\lambda=0} &= -[2G_0(e) - 1] \cdot \text{sign}(e) + [2G_1(e) - 1] \cdot \text{sign}(e) \\ &\quad + 2g_0(e)(-\mathbf{x}^T\dot{\boldsymbol{\theta}})\text{sign}(e) \\ &= [2G_1(e) - 2G_0(e) - 2g_0(e) \cdot \mathbf{x}^T\dot{\boldsymbol{\theta}}] \cdot \text{sign}(e) \qquad (2.3.6)\end{aligned}$$

here

$$\dot{\boldsymbol{\theta}} = \frac{d}{d\lambda}\boldsymbol{\theta}_\lambda|_{\lambda=0}.$$

In (2.1.12) replacing F by F_λ, taking the derivative with respect to λ and evaluating at $\lambda = 0$, we get

$$\begin{aligned}
\mathbf{0} = & \int \mathbf{x} v(\mathbf{x}) \psi'[\frac{1}{v(\mathbf{x})}\phi^+(\ell(e,\mathbf{x},G_\lambda))]|_{\lambda=0} \cdot \frac{1}{v(\mathbf{x})}\phi^{+\prime}(\ell(e,\mathbf{x},G_\lambda))|_{\lambda=0} \\
& \cdot \frac{d}{d\lambda}\ell(e,\mathbf{x},G_\lambda)|_{\lambda=0} \cdot \operatorname{sign}(e) dF_0(\mathbf{x},e) \\
& + \int \mathbf{x} v(\mathbf{x}) \psi[\frac{1}{v(\mathbf{x})}\phi^+(\ell(e,\mathbf{x},G_\lambda))]|_{\lambda=0} \cdot \operatorname{sign}(e) d(F_1 - F_0)(\mathbf{x},e) \\
= & \int \mathbf{x}\psi'[\frac{1}{v(\mathbf{x})}\phi^+((2G_0(e)-1)\cdot\operatorname{sign}(e))]\phi^{+\prime}((2G_0(e)-1)\cdot\operatorname{sign}(e)) \\
& \cdot (2G_1(e) - 2G_0(e) - 2g_0(e)\mathbf{x}^T \cdot \dot{\boldsymbol{\theta}}) dF_0(\mathbf{x},e) \\
& + \int \mathbf{x} v(\mathbf{x}) \psi[\frac{1}{v(\mathbf{x})}\phi^+((2G_0(e)-1)\cdot\operatorname{sign}(e))]\cdot\operatorname{sign}(e) dF_1(\mathbf{x},e).
\end{aligned} \tag{2.3.7}$$

In addition, we assume that the model (1.2.1) goes through the origin and $v(\mathbf{x})$ is an even function of every x_i $(i = 1,\ldots,k)$. Then

$$\begin{aligned}
\int \mathbf{x}\psi'[\frac{1}{v(\mathbf{x})}\phi^+((2G_0(e)-1)\cdot\operatorname{sign}(e))] \cdot \phi^{+\prime}((2G_0(e)-1)\operatorname{sign}(e)) \\
\cdot (2G_1(e) - 2G_0(e)) dF_0(\mathbf{x},e) = 0\,.
\end{aligned} \tag{2.3.8}$$

Combining (2.3.8) with (2.3.7), we come out

$$\begin{aligned}
& -\int 2\mathbf{x}\psi'[\frac{1}{v(\mathbf{x})}\phi^{+\prime}(B(e))] \cdot \phi^{+\prime}(B(e)) \cdot g_0(e)\mathbf{x}^T\dot{\boldsymbol{\theta}} dF_0(\mathbf{x},e) \\
& \quad + \int \mathbf{x} v(\mathbf{x}) \psi[\frac{1}{v(\mathbf{x})}\phi^+(B(e))] \cdot \operatorname{sign}(e) dF_1(\mathbf{x},e) = \mathbf{0}\,,
\end{aligned} \tag{2.3.9}$$

where

$$\begin{aligned}
B(e) &= (2G_0(e)-1)\cdot\operatorname{sign}(e) \\
&= |G_0(e) - G_0(-e)|.
\end{aligned} \tag{2.3.10}$$

For simplicity, we denote $\mathbb{IF}(e,\mathbf{x}) = \mathbb{IF}(e,\mathbf{x},G_0,H_0,v,\phi^+,\psi)$. From (2.3.9), solving for $\dot{\boldsymbol{\theta}}$ and evaluating at $G_1 = \delta e$ and $H_1 = \delta\mathbf{x}$, we derive the influence function for $\boldsymbol{\theta}(F)$ at distribution F_0, given by

$$\begin{aligned}\mathbb{IF}(e,\mathbf{x}) &= \dot{\boldsymbol{\theta}}|_{\substack{G_1=\delta e \\ H_1=\delta\mathbf{x}}} \\ &= D^{-1}\{\mathbf{x}v(\mathbf{x})\psi[\frac{1}{v(\mathbf{x})}\phi^+(B(e))]\cdot\mathrm{sign}(e)\},\end{aligned} \tag{2.3.11}$$

where

$$D = \int 2\mathbf{x}\psi'[\frac{1}{v(\mathbf{x})}\phi^+(B(e))]\cdot\phi^{+\prime}(B(e))g_0(e)\mathbf{x}^T dF_0(\mathbf{x},e). \tag{2.3.12}$$

Correspondingly, we can verify (2.3.2), i.e. $\int \mathbb{IF}(e,\mathbf{x})dF_0(\mathbf{x},e) = \mathbf{0}$. Well

$$\begin{aligned}\int \mathbb{IF}(e,\mathbf{x})dF_0(\mathbf{x},e) &= D^{-1}\int \mathbf{x}v(\mathbf{x})\psi[\frac{1}{v(\mathbf{x})}\phi^+(B(e))]\cdot\mathrm{sign}(e)dF_0(\mathbf{x},e) \\ &= D^{-1}\cdot\mathbf{0} \quad \text{(by (2.1.7))} \\ &= \mathbf{0}.\end{aligned}$$

The formal asymptotic covariance matrix $V(G_0,H_0,v,\phi^+,\psi)$ (hereafter denoted by V for the sake of simplicity) is given by

$$\begin{aligned}V &= \int \mathbb{IF}(e,\mathbf{x})\cdot\mathbb{IF}(e,\mathbf{x})^T dF_0(\mathbf{x},e) \\ &= D^{-1}\{\int v^2(\mathbf{x})\psi^2[\frac{1}{v(\mathbf{x})}\phi^+(B(e))]\mathbf{x}\cdot\mathbf{x}^T dF_0(\mathbf{x},e)\}\cdot D^{-1} \\ &= D^{-1}MD^{-1}\end{aligned} \tag{2.3.13}$$

with

$$M = \int v^2(\mathbf{x})\psi^2[\frac{1}{v(\mathbf{x})}\phi^+(B(e))]\mathbf{x}\cdot\mathbf{x}^T dF_0(\mathbf{x},e). \tag{2.3.14}$$

From the beginning of this section, we know that $\sqrt{n}(\boldsymbol{\theta}-\boldsymbol{\theta}_*)$ is asymptotically normal with the covariance matrix $V = D^{-1}MD^{-1}$. Next chapter will discuss some optimality problems such that V is minimum or minimax (the definitions of minimum and minimax will be given in the next chapter.)

CHAPTER 3

OPTIMAL FUNCTIONS

In this chapter, the spherically symmetric distribution and its properties are discussed. Based on the spherically symmetric distribution, our optimality problems become much simpler. From section 2.3, we know that the asymptotic covariance matrix V depends on the distribution F_0 as well as the functions ψ, ϕ^+ and v. If we fix a distribution F_0, then we can choose optimal functions ψ, ϕ^+ and v such that V is a minimum (which is defined in section 3.1) and the influence function is bounded. This is called a Hampel-type optimality problem. If the distribution F_0 varies in a class of distributions, we can choose functions ψ, ϕ^+ and v such that V is a minimax (defined in section 3.4). That is called a Huber-type optimality problem. The above two problems are both solved in this chapter. In section one, the spherically symmetric distribution is introduced and the definition of minimum matrix among a group of matrices is given. Optimal functions in general case are derived in section two. In section three, we solve Hampel-type optimality problem. In section four we obtain the solution to Huber-type optimality problem.

3.1 Optimality criteria

Considering the distribution H_0 of $\mathbf{X}$ is spherically symmetric, then the probability density function h_0 of $\mathbf{X}$ can be written as

$$h_0(\mathbf{x}) = h_1(||\mathbf{x}||) \cdot I(\frac{\mathbf{x}}{||\mathbf{x}||}), \quad ||\cdot|| \text{ means norm.} \tag{3.1.1}$$

where $I(\mathbf{u})$ is the uniform distribution density on S_k ($S_k = \{\mathbf{u} | \, ||\mathbf{u}|| = 1, \mathbf{u} \in \mathbb{R}^k\}$) and $h_1(||\mathbf{x}||)$ is the probability density function of $||\mathbf{x}||$.

From Eaton (1981), we know $Z = ||\mathbf{X}||$ and $\mathbf{u} = \frac{\mathbf{X}}{||\mathbf{x}||}$ are independent random variables, so

$$dH_0(\mathbf{x}) = h_1(z) I(\mathbf{u}) dz d\mathbf{u}, \tag{3.1.2}$$

with $z = ||\mathbf{x}||$ and $\mathbf{u} = \mathbf{x}/||\mathbf{x}||$.

We also assume that the weight function $v(\mathbf{x})$ depends on $\mathbf{x}$ only through $||\mathbf{x}||$, say

$$v(z) = v(\mathbf{x}) . \tag{3.1.3}$$

In the remaining part of this chapter we confine our attention to the following case:

a) the distribution H_0 of $\mathbf{X}$ is spherically symmetric,

b) $v(\mathbf{x}) = v(z)$,

c) the density function g_0 is absolutely continuous.

In this case, the influence function (2.3.11) and the covariance matrix (2.3.13) become much simpler. Let's observe (2.3.12) and (2.3.14) first,

$$\begin{aligned} D = & \int 2\mathbf{x}\psi'[\frac{1}{v(\mathbf{x})}\phi^+(B(e))]\phi^{+\prime}(B(e))g_0(e)\mathbf{x}^T dF_0(\mathbf{x},e) \\ & \int\int\int 2\frac{\mathbf{x}}{||\mathbf{x}||}\cdot\frac{\mathbf{x}^T}{||\mathbf{x}||}\cdot||\mathbf{x}||^2\psi'[\frac{1}{v(||\mathbf{x}||)}\phi^+(B(e))]\cdot\phi^{+\prime}(B(e))g_0(e) \\ & \cdot h_1(||\mathbf{x}||)\cdot I(\frac{\mathbf{x}}{||\mathbf{x}||})dG_0(e)dzd\mathbf{u} \\ = & \int \mathbf{u}\cdot\mathbf{u}^T I(\mathbf{u})d\mathbf{u}\cdot\int\int 2\psi'[\frac{1}{v(z)}\phi^+(B(e))]\phi^{+\prime}(B(e))g_0(e) \\ & \cdot z^2\cdot h_1(z)dG_0(e)dz. \end{aligned} \tag{3.1.4}$$

Denoting

$$D_1 = E(\mathbf{u}\cdot\mathbf{u}^T), \tag{3.1.5}$$

and

$$K_1 = \int\int 2\psi'[\frac{1}{v(z)}\phi^+(B(e))]\cdot\phi^{+\prime}(B(e))g_0(e)\cdot z^2 h_1(z)dG_0(e)dz, \tag{3.1.6}$$

(3.1.4) becomes

$$D = D_1\cdot K_1 . \tag{3.1.7}$$

Similarly, from (2.3.14),

$$M = \int v^2(\mathbf{x})\psi^2[\frac{1}{v(\mathbf{x})}\phi^+(B(e))]\mathbf{x}\cdot\mathbf{x}^T dF_0(\mathbf{x},e)$$
$$= D_1 \cdot K_2 \,. \tag{3.1.8}$$

here

$$K_2 = \int\int \psi^2[\frac{1}{v(z)}\phi^+(B(e))]\cdot z^2 \cdot v^2(z) dG_0(e) dH_1(z). \tag{3.1.9}$$

Combining (3.1.7) with (2.3.11), we have

$$\mathbb{F}(e,\mathbf{x}) = D_1^{-1}\frac{1}{K_1}\{\frac{\mathbf{x}}{||\mathbf{x}||}\cdot||\mathbf{x}||v(||\mathbf{x}||)\psi[\frac{1}{v(||\mathbf{x}||)}\phi^+(B(e))]\cdot \text{sign}(e)\}$$
$$= D_1^{-1}\mathbf{u}\cdot\{\frac{zv(z)\psi[\frac{1}{v(z)}\phi^+(B(e))]\cdot\text{sign}(e)}{K_1}\}$$
$$= D_1^{-1}\mathbf{u}\cdot\frac{\eta(e,z)}{K_1}, \tag{3.1.10}$$

with

$$\eta(e,z) = zv(z)\psi[\frac{1}{v(z)}\cdot\phi^+(B(e))]\cdot\text{sign}(e). \tag{3.1.11}$$

Replacing D with (3.1.7) and M with (3.1.8) in (2.3.13), we get

$$V = D^{-1}MD^{-1} = \frac{1}{K_1}D_1^{-1}D_1K_2\cdot D_1^{-1}\frac{1}{K_1}$$
$$= \frac{K_2}{K_1^2}D_1^{-1}\,. \tag{3.1.12}$$

It is important to notice that $D_1 = E(\mathbf{u}\cdot\mathbf{u}^T)$ is independent of the density of $\mathbf{x}$.

$$D_1 = E[\frac{\mathbf{x}\mathbf{x}^T}{||\mathbf{x}||^2}] = aI_k \quad \text{(by symmetry)},$$

where

$$a = E(\frac{x_i^2}{||\mathbf{x}||^2})\,, \quad i = 1,\ldots,k\,,$$

so

$$ka = \sum_{i=1}^{k} E(\frac{x_i^2}{||\mathbf{x}||^2}) = E[\frac{\Sigma x_i^2}{||\mathbf{x}||^2}] = E(1) = 1,$$

hence $a = 1/k$ and $D_1 = \frac{1}{k} I_k$. In fact D_1 is a fixed positive definite matrix. Thus the covariance matrix V varies only with K_2/K_1^2. Before discussing any optimality problem, we give the optimality criterion for V.

Definition 3.1.1: Minimum matrix.

Let $V(G_0, H_0)$ be the group of matrices containing all $V(G_0, H_0, v, \psi, \phi^+)$ for all functions v, ψ and ϕ^+. Given any two matrices $V_1, V_2 \in V(G_0, H_0)$, we say $V_1 < V_2$ if $V_2 - V_1$ is positive semidefinite. V is said to be the minimum in $V(G_0, H_0)$, if $V < W$ for any $W \in V(G_0, H_0)$.

Since $V(G_0, H_0, v, \psi, \phi^+) = \frac{K_2}{K_1^2} D_1^{-1}$, where D_1^{-1} is a fixed positive definite matrix and $\frac{K_2}{K_1^2}$ is a function of v, ϕ^+, v, G_0 and H_0, minimum matrix V is obtained by choosing ψ, ϕ^+ and v such that $\frac{K_2}{K_1^2}$ is minimized. Consequently those ψ, ϕ^+ and v are called optimal functions.

3.2. Optimal functions in general

In this section, the optimality problem without any constraints, i.e. $\min V(G_0, H_0, v, \psi, \phi^+)$, is studied. From previous section, $\min V(G_0, H_0, v, \psi, \phi^+)$ is equivalent to

$$\min \frac{K_2}{K_1^2}, \tag{3.2.1}$$

or

$$\max \frac{K_1^2}{K_2}. \tag{3.2.2}$$

Using $\eta(e, z)$, we can rewrite K_1 and K_2 as

$$K_1 = E_{e,z}\{z \frac{\partial}{\partial e} \eta(e, z)\}, \tag{3.2.3}$$

and

$$K_2 = E_{e,z}\{\eta^2(e, z)\}, \tag{3.2.4}$$

where $\eta(e, z) = z v(z) \psi[\frac{1}{v(z)} \phi^+(B(e))] \cdot \text{sign}(e)$.

Remark: $E_{e,z}$ means expectation with respect to random variables e and z.

From (3.2.3),

$$\begin{aligned} K_1 &= \int\int z\frac{\partial}{\partial e}\eta(e,z)dG_0(e)dH_1(z) \\ &= \int z\{\int \frac{\partial}{\partial e}\eta(e,z)g_0(e)de\}dH_1(z) \\ &= \int z\{\int -\eta(e,z)g_0'(e)de\}dH_1(z) \quad \text{(integration by parts)} \\ &= \int z\{\int -\eta(e,z)\frac{g_0'(e)}{g_0(e)}dG_0(e)\}dH_1(z) \\ &= E_{e,z}\{-\eta(e,z)z\frac{g_0'(e)}{g_0(e)}\}. \end{aligned} \tag{3.2.5}$$

using Cauchy-Schwarz inequality, we have

$$\begin{aligned} K_1^2 &\le E_{e,z}\{\eta^2(e,z)\}\cdot E_{e,z}\{z^2(\frac{g_0'(e)}{g_0(e)})^2\} \\ &= K_2\cdot I(G_0)\cdot J(H_1)\,, \end{aligned} \tag{3.2.6}$$

where

$$I(G_0) = \int (\frac{g_0'(e)}{g_0(e)})^2 dG_0(e) \quad \text{is Fisher information,}$$

and

$$J(H_1) = \int z^2 dH_1(z).$$

Therefore

$$\frac{K_1^2}{K_2} \le I(G_0)\cdot J(H_1), \tag{3.2.7}$$

and furthermore, when $\eta(e,z)$ is in the form of $-c_1 z\frac{g_0'(e)}{g_0(e)}$ for any $c_1 \neq 0$, $\frac{K_1^2}{K_2}$ reaches the maximum. Considering the case $c_1 = 1$, and by the definition of $\eta(e,z)$, we have

$$\psi[\frac{1}{v(z)}\phi^+(B(e))]zv(z)\text{sign}(e) = -z\frac{g_0'(e)}{g_0(e)}\,. \tag{3.2.8}$$

Solving (3.2.8), we get

$$i) \qquad \psi(t) = t, \text{ for } t > 0, \tag{3.2.9}$$

and

$$ii) \qquad \phi^+(B(e)) = -\frac{g_0'(e)}{g_0(e)}\text{sign}(e), \tag{3.2.10}$$

i.e. $\phi^+(u) = -\frac{g_0'}{g_0}(G_0^{-1}(\frac{u+1}{2}))$, $\qquad 0 \le u \le 1$.

From (3.2.10), some of the common scores can be obtained by choosing different distributions G_0. For instance, if G_0 = Exponential distribution,

$$g_0(e) = exp(-e),$$

$$g_0'(e) = -exp(-e),$$

so

$$\phi^+(B(e)) = -\frac{g_0'(e)}{g_0(e)}\text{sign}(e) = \text{sign}(e),$$

this is the sign scores.

If

$$G_0 = \text{ Logistic distribution },$$

$$g_0(e) = \exp(-e)/[1+\exp(-e)]^2,$$

$$G_0(e) = 1/[1+\exp[(-e)],$$

$$\frac{-g_0'(e)}{g_0(e)} = 2G_0(e) - 1,$$

$$\phi^+(B(e)) = \frac{-g_0'(e)}{g_0(e)} \cdot \text{sign}(e) = (2G_0(e) - 1) \cdot \text{sign}(e),$$

i.e. $\phi^+(u) = u$, $0 \le u \le 1$, that is the Wilcoxon scores.

If G_0 = Normal distribution,

$$\frac{-g_0'(e)}{g_0(e)} = e,$$

$$\phi^+(B(e)) = e \cdot \text{sign}(e) = |e|,$$

that is the normal score.

These functions in (3.2.9) and (3.2.10) are the optimal functions if G_0 is specified exactly, i.e. correctly, and no outlying $\mathbf{x}$'s. However, in this case we don't have any information about $v(z)$ and the influence function is unbounded. If we define the gross error sensitivity ν^* (Huber (1981), p.14) to be the maximum norm of the influence function, i.e.

$$\nu^* = \sup_{\mathbf{x},e} ||\mathbb{IF}(e, \mathbf{x})||,$$

then the gross error sensitivity is infinite for the above case according to (3.1.10), because $\eta(e, z)$ is infinite in this case and K_1 can always be assumed finite.

In another case, the optimality problem is a constrained minimization problem, i.e.

$$\min \frac{K_2}{K_1^2} \text{ , subject to } \nu^* \leq b_1,$$

where b_1 is a given constant ($\nu^* \leq b_1$ means that the gross error sensitivity is finite). This is called a Hampel-type problem, and is going to be studied in section 3.3.

The final case considers Huber-type minimax problem in which the distribution G_0 varies in a neighborhood of G_0 , say $\mathcal{F}(G_0)$. This case is going to be solved in section 3.4 under the condition that G_0 is normal and $\mathcal{F}(G_0)$ is the ε-contaminated neighborhood $\mathcal{F}_\varepsilon(G_0)$, i.e.

$$\mathcal{F}_\varepsilon(G_0) = \{(1-\varepsilon)G_0 + \varepsilon H \ , \quad 0 < \varepsilon < 1\},$$

here H is any symmetric distribution.

3.3 Optimal functions for Hampel-type optimality problem

Observe the influence function (3.1.10),

$$\mathbb{IF}(e, \mathbf{x}) = D_1^{-1}\mathbf{u}\frac{\eta(e, z)}{K_1},$$

where D_1^{-1} is a fixed matrix, $||\mathbf{u}|| = ||\frac{\mathbf{x}}{||\mathbf{x}||}|| = 1$ and $z = ||\mathbf{x}||$. The gross error sensitivity (defined in section 3.2) is given by

$$\begin{aligned}
\nu^* &= \sup_{e,\mathbf{x}} ||\mathbb{F}(e,\mathbf{x})|| \\
&= \sup_{e,\mathbf{x}} ||D_1^{-1}\mathbf{u}\frac{\eta(e,z)}{K_1}|| \\
&= \sup_{e,\mathbf{x}}\{|\frac{\eta(e,z)}{K_1}| \cdot ||D_1^{-1}\mathbf{u}||\} \\
&= \sup_{e,\mathbf{x}}\{|\frac{\eta(e,z)}{K_1}| \cdot ||(\frac{1}{k}I_k)^{-1}\mathbf{u}||\} \\
&= \sup_{e,z}\{|\frac{\eta(e,z)}{K_1}| \cdot k||\mathbf{u}||\} \\
&= k \cdot \sup_{e,z}|\frac{\eta(e,z)}{K_1}| .
\end{aligned}$$

Therefore, for a given b_1, $\nu^* \leq b_1$ is equivalent to

$$\sup_{e,z}|\frac{\eta(e,z)}{K_1}| \leq b_2$$

for some constant b_2.

Now Hampel-type optimality problem can be stated as follows: (For fixed G_0 and H_0)

$$\min \frac{K_2}{K_1^2}, \tag{3.3.1}$$

$$s.t. \sup_{e,z}|\frac{\eta(e,z)}{K_1}| \leq b_2, \quad \text{for a given } b_2 . \tag{3.3.2}$$

In order to solve ψ, ϕ^+ and v from (3.3.1) and (3.3.2), we standardize $\eta(e,z)$ by (3.3.5) below , then the optimality problem becomes: (For fixed G_0 and H_0)

$$\min K_2 = \min E_{e,z}\{\eta^2(e,z)\}, \tag{3.3.3}$$

$$s.t. \sup_{e,z}|(\eta(e,z)| \leq b, \quad \text{for a given } b, \tag{3.3.4}$$

$$s.t. K_1 = E_{e,z}\{-\eta(e,z)\cdot z\cdot \frac{g_0'(e)}{g_0(e)}\} = S. \tag{3.3.5}$$

For mathematical convenience we choose

$$S = \iint\limits_{|z\frac{g_0'(e)}{g_0(e)}|\le b} z^2(\frac{g_0'(e)}{g_0(e)})^2 dH_1(z)dG_0(e) + \iint\limits_{|z\frac{g_0'(e)}{g_0(e)}|>b} b|z\frac{g_0'(e)}{g_0(e)}|dH_1(z)dG_0(e)\,. \tag{3.3.6}$$

Our next theorem gives the solution to optimality problem (3.3.3)-(3.3.5).

THEOREM 2. *The optimal function $\eta(e,z)$ for (3.3.3)-(3.3.5) is*

$$\eta(e,z) = \begin{cases} b, & \text{for } -z\frac{g_0'(e)}{g_0(e)} > b\,, \\ -z\frac{g_0'(e)}{g_0(e)}, & \text{for } |z\frac{g_0'(e)}{g_0(e)}| \le b\,, \\ -b, & \text{for } -z\frac{g_0'(e)}{g_0(e)} < -b\,. \end{cases} \tag{3.3.7}$$

PROOF: For any $\eta(e,z)$ satisfying constraint (3.3.5), we have

$$\begin{aligned} &\int\int(\eta(e,z) + z\frac{g_0'(e)}{g_0(e)})^2 dG_0(e)dH_1(z) \\ &\quad = \int\int \eta^2(e,z)dG_0(e)dH_1(z) - 2\cdot K_1 + \int z^2(\frac{g_0'(e)}{g_0(e)})^2 dG_0(e)dH_1(z) \\ &\quad = K_2 + I(G_0)J(H_1) - 2\cdot S, (K_1 = S \text{ by (3.3.5)})\,. \end{aligned} \tag{3.3.8}$$

So minimizingK_2 is equivalent to

$$\text{minimizing} \int\int(\eta(e,z) + z\frac{g_0'(e)}{g_0(e)})^2 dG_0(e)dH_1(z), \tag{3.3.9}$$

since, for fixed G_0 and H_0, $I(G_0)$, $J(H_1)$ and S in (3.3.8) are constant. Observe

$$\begin{aligned} &\int\int(\eta(e,z) + z\frac{g_0'(e)}{g_0(e)})^2 dG_0(e)dH_1(z) \\ &\quad = \iint\limits_{\{-z\frac{g_0'(e)}{g_0(e)}>b\}} [\eta(e,z) - (-z\frac{g_0'(e)}{g_0(e)})]^2 dG_0(e)dH_1(z) \\ &\qquad + \iint\limits_{\{-z\frac{g_0'(e)}{g_0(e)}<-b\}} [\eta(e,z) - (-z\frac{g_0'(e)}{g_0(e)})]^2 dG_0(e)dH_1(z) \\ &\qquad + \iint\limits_{\{|-z\frac{g_0'(e)}{g_0(e)}|\le b\}} [\eta(e,z) - (-z\frac{g_0'(e)}{g_0(e)})]^2 dG_0(e)dH_1(z)\,, \end{aligned} \tag{3.3.10}$$

with $\eta(e,z)$ satisfying constraints (3.3.4) and (3.3.5), (3.3.10) is minimized by $\eta(e,z)$ in (3.3.7), since each term in (3.3.10) is minimized. Hence $\eta(e,z)$ in (3.3.7) is the optimal function for (3.3.3)-(3.3.5). □

Replacing $\eta(e,z)$ in (3.3.7) by its definition in terms of ψ, ϕ^+ and v in (3.1.11),

$$\eta(e,z) = zv(z)\psi[\frac{1}{v(z)}\phi^+B(e))] \cdot \text{sign}(e),$$

solving for ψ, ϕ^+ and v from (3.3.7), we obtain the optimal functions ψ, ϕ^+ and v (given in following theorem) for Hampel-type problem.

THEOREM 3. *The optimal functions ψ, ϕ^+ and v for Hampel-type problem are*

$$\psi(t) = \begin{cases} t, & \text{for } t \le b, \\ b, & \text{for } t > b, \end{cases} \tag{3.3.11}$$

$$\phi^+(B(e)) = -\frac{g_0'(e)}{g_0(e)} \cdot sign(e), \tag{3.3.12}$$

i.e. $\phi^+(u) = -\frac{g_0'}{g_0}(G_0^{-1}(\frac{u+1}{2}))$, $\quad 0 \le u \le 1$,

$$v(z) = \frac{1}{z}. \tag{3.3.13}$$

PROOF: 1) When $|z\frac{g_0'(e)}{g_0(e)}| \le b$, from (3.3.12) and (3.3.13), we have

$$|\frac{\phi^+(B(e))}{v(z)}| = |-z\frac{g_0'(e)}{g_0(e)} \cdot \text{sign}(e)| = |z\frac{g_0'(e)}{g_0(e)}| \le b,$$

by (3.3.11),

$$\psi(\frac{\phi^+(B(e))}{v(z)}) = \frac{\phi^+(B(e))}{v(z)} = -z\frac{g_0'(e)}{g_0(e)} \cdot \text{sign}(e),$$

so

$$\begin{aligned} \eta(e,z) &= zv(z)\psi[\frac{\phi^+(B(e))}{v(z)}] \cdot \text{sign}(e) \\ &= z \cdot \frac{1}{z} \cdot (-z\frac{g_0'(e)}{g_0(e)}\text{sign}(e)) \cdot \text{sign}(e) \\ &= -z\frac{g_0'(e)}{g_0(e)}, \quad \text{(3.3.7) holds.} \end{aligned}$$

2) when $-z\frac{g_0'(e)}{g_0(e)} > b$, from (3.3.12) and (3.3.13) we have

$$\frac{\phi^+(B(e))}{v(z)} = -z\frac{g_0'(e)}{g_0(e)}\text{sign}(e) > b\,,$$

by (3.3.11),

$$\psi(\frac{\phi^+(B(e))}{v(z)}) = b\text{sign}(e),$$

so

$$\eta(e,z) = z\cdot\frac{1}{z}b\cdot\text{sign}(e)\cdot\text{sign}(e) = b,$$

(3.3.7) holds.

3) Similarly, when $-z\frac{g_0'(e)}{g_0(e)} < -b$, (3.3.7) holds.

Therefore, on the whole, ψ, ϕ^+ and v in (3.3.11)-(3.3.13) satisfy (3.3.7). □

3.4. Optimal functions in Huber-type minimax problem

Huber (1983) developed an exact finite-sample minimax theory of robust M-estimation for regression and solved various finite-sample and asymptotic minimax problems. Applying Huber's idea, we will study minimax aspects of bounded influence R-estimator for regression. Similarly the finite-sample case is considered, $\mathbf{x}_i$ is fixed and at each $\mathbf{x}_i$ we have m observations y_i. Assume that $(y_i - \boldsymbol{\theta}^T\mathbf{x}_i)$'s are distributed according to an ε_i-contaminated normal distribution $\mathcal{F}_{\varepsilon_i}$, i.e.

$$\mathcal{F}_{\varepsilon_i} = \{(1-\varepsilon_i)\Phi + \varepsilon_i H,\ 0 < \varepsilon_i < 1\}\,,$$

where Φ is the normal distribution and H is any arbitrary distribution. For $\mathcal{F}_{\varepsilon_i}$, the least informative density g_i which minimizes Fisher information is (see Huber 1981)

$$g_i(e) = \begin{cases} \frac{1-\varepsilon_i}{\sqrt{2\pi}}exp(-e^2/2), & \text{for } |e| \le k_i\,, \\ \frac{1-\varepsilon_i}{\sqrt{2\pi}}exp(k_i^2/2 - k_i|e|), & \text{for } |e| > k_i, \end{cases} \tag{3.4.1}$$

here k_i and ε_i are related by $\int g_i(e)de = 1$, i.e.

$$\frac{2\varphi(k_i)}{k_i} - 2\Phi(-k_i) = \frac{\varepsilon_i}{1-\varepsilon_i}\,, \tag{3.4.2}$$

where φ is the normal density function.

Now, for each fixed $\mathbf{x}_i$, we let

$$z_i = ||\mathbf{x}_i||,$$
$$c_i = v(z_i), \tag{3.4.3}$$

and

$$\eta_i(e) = v[\frac{\phi^+(B(e))}{c_i}]c_i z_i \text{sign}(e) . \tag{3.4.4}$$

From section 3.1, $\min V$ (covariance matrix) is equivalent to

$$\min \frac{K_2}{K_1^2} . \tag{3.4.5}$$

Here we focus on $\min \frac{K_2}{K_1^2}$ at each fixed $\mathbf{x}_i$ with a number of observations y_i's, correspondingly K_1 and K_2 are given by

$$\begin{aligned} K_1 &= E_{e,z}\{z\frac{\partial}{\partial e}\eta(e,z)\} \quad \text{(from (3.2.3))} \\ &= E_e\{z_i\frac{\partial}{\partial e}\eta_i(e)\}, \end{aligned}$$

since, at each fixed $\mathbf{x}_i$, z_i is constant, e is distributed according to $g(e)$, and $\eta(e,z)$ becomes $\eta_i(e)$. Integrating by parts, K_1 becomes

$$K_1 = E_e\{-\eta_i(e)z_i\frac{g'(e)}{g(e)}\} . \tag{3.4.6}$$

And similarly,

$$\begin{aligned} K_2 &= E_{e,z}\{\eta^2(e,z)\} \quad \text{(from (3.2.4))} \\ &= E_e\{\eta_i^2(e)\} . \end{aligned} \tag{3.4.7}$$

Minimizing problem (3.4.5), with (3.4.6) and (3.4.7) is called Huber-type optimality problem. First, we find out the solution $\eta_i(e)$ to this problem when e

is distributed as $g_i(e)$, then we will show that $\eta_i(e)$ is the minimax solution for $\min \frac{K_2}{K_1^2}$ over $\mathcal{F}_{e_i}$.

THEOREM 4. *Provided e has p.d.f. $g_i(e)$, the solution to (3.4.5) is*

$$\begin{aligned}\eta_i(e) &= -z_i \frac{g_i'(e)}{g_i(e)} \\ &= \begin{cases} z_i e, & \text{for } |e| \le k_i\,, \\ z_i k_i sign(e), & \text{for } |e| > k_i\,. \end{cases}\end{aligned} \tag{3.4.8}$$

PROOF: $\min \frac{K_2}{K_1^2}$ is equivalent to $\max \frac{K_1^2}{K_2}$. Using Cauchy-Schwarz inequality, we have

$$\begin{aligned} K_1^2 &= (E_e\{z_i \frac{\partial}{\partial e}\eta_i(e)\})^2 \\ &= (E_e\{-\eta_i(e) z_i \frac{g_i'(e)}{g_i(e)}\})^2 \\ &\le E_e(\eta_i^2(e)) \cdot E_e\{(-z_i \frac{g_i'(e)}{g_i(e)})^2\} \\ &= K_2 \cdot z_i^2 \cdot I(G_i(e)). \end{aligned}$$

Thus

$$\frac{K_1^2}{K_2} \le z_i^2 I(G_i(e)), \tag{3.4.9}$$

with equality when $\eta_i(e)$ is in the form of $-cz_i \frac{g_i'(e)}{g_i(e)}$, for any $c \neq 0$. For example, taking $c = 1$, we get

$$\eta_i(e) = -z_i \frac{g_i'(e)}{g_i(e)},$$

this is (3.4.8). Furthermore, from (3.4.1),

$$g_i'(e) = \begin{cases} \frac{1-\epsilon_i}{\sqrt{2\pi}}(-e) exp(-e^2/2) & \text{for } |e| \le k_i\,, \\ \frac{1-\epsilon_i}{\sqrt{2\pi}}(-k_i \text{sign}(e)) exp(k_i^2/2 - k_i|e|)\,, & \text{for } |e| > k_i\,, \end{cases}$$

and

$$\frac{g_i'(e)}{g_i(e)} = \begin{cases} -e\,, & \text{for } |e| \le k_i\,, \\ -k_i \text{sign}(e)\,, & \text{for } |e| > k_i\,. \end{cases} \tag{3.4.10}$$

$\eta_i(e)$ can be written as

$$\eta_i(e) = \begin{cases} z_i e\,, & \text{for } |e| \le k_i\,, \\ z_i k_i \text{sign}(\epsilon)\,, & \text{for } |e| > k_i. \end{cases}$$

The proof is completed. □

The following theorem is our main result for minimax theory.

THEOREM 5. *The $\eta_i(e)$ in Theorem 4 is also the minimax solution to* $\min \frac{K_2}{K_1^2}$ *over $\mathcal{F}_{\epsilon_i}$ (i.e. the distribution G of ϵ can also vary over $\mathcal{F}_{\epsilon_i}$) provided H is also symmetric.*

PROOF: In order to prove that $\eta_i(\epsilon)$ in (3.4.8) is the minimax solution, two things have to be shown.

1) For fixed $g_i \in \mathcal{F}_{\epsilon_i}$, $\eta_i(e)$ minimizes $\frac{K_2}{K_1^2}$. This is done in Theorem 4.

2) When this $\eta_i(e)$ in (3.4.8) is used, $\frac{K_2}{K_1^2}$ is maximized over $\mathcal{F}_{\epsilon_i}$, by g_i in (3.4.1).

By 1) and 2), we have

$$\frac{K_2}{K_1^2}(\eta_i, g) \le \frac{K_2}{K_1^2}(\eta_i, g_i) \le \frac{K_2}{K_1^2}(\eta, g_i)\,,$$

for all $g \in \mathcal{F}_{\epsilon_i}$ and any η, then

$$\begin{aligned} \sup_g \frac{K_2}{K_1^2}(\eta_i, g) &= \frac{K_2}{K_1^2}(\eta_i, g_i) \\ &\le \frac{K_2}{K_1^2}(\eta, g_i) \\ &\le \sup_g \frac{K_2}{K_1^2}(\eta, g). \end{aligned}$$

So η_i is the minimax solution.

To see 2), using the same method as in Jaeckel (1971), we recall the definition of $\eta_i(e)$ in terms of ψ and ϕ^+, i.e.

$$\begin{aligned}\eta_i(e) &= \psi[\frac{\phi^+(B(e))}{c_i}]c_i z_i \text{sign}(e) \\ &= \psi[\frac{\phi^+((2G(e)-1)\text{sign}(e))}{c_i}]c_i z_i \text{sign}(e),\end{aligned}$$

and put

$$\begin{aligned}\tilde{\mathbb{F}}(e, \mathbf{x}_i) &= \frac{\eta_i(e)}{K_1} \\ &= \frac{\eta_i(e)}{z_i \cdot \int_{-\infty}^{\infty} \frac{\partial}{\partial e}\eta_i(e)g(e)de}.\end{aligned}$$

Let $u = (2G(e)-1)\cdot \text{sign}(e)$, then

$$\begin{aligned}e &= G^{-1}(\frac{1+u}{2}), \text{ for } e \geq 0, \\ e &= G^{-1}(\frac{1-u}{2}), \text{ for } e < 0.\end{aligned}$$

Denote

$$J(u) = \psi[\frac{\phi^+(u)}{c_i}]c_i z_i, \quad \text{for } e \geq 0,$$

so, when $e \geq 0$,

$$\eta_i(e) = J(u),$$

and

$$\begin{aligned}\eta_i'(e) &= J'(u)\cdot\frac{\partial u}{\partial e} \\ &= J'(u)\cdot 2g(e).\end{aligned}$$

The influence function at $e = G^{-1}(\frac{1\pm u}{2})$ becomes

$$\begin{aligned}|\tilde{\mathbb{F}}(G^{-1}(\frac{1\pm u}{2}), \mathbf{x}_i)| &= \frac{J(u)}{2z_i \int_0^{\infty} \frac{\partial}{\partial e}\eta_i(e)g(e)de} \\ &= \frac{J(u)}{2z_i \int_0^1 J'(u)\cdot g(G^{-1}(\frac{u\pm 1}{2}))du}.\end{aligned}$$

Since $J'(u) = 0$ outside of the interval $(0, 2G(c_i) - 1)$ (notice that $c_i = k_i$ from below) and in this interval

$$g(G^{-1}(\frac{1+u}{2})) \geq g_i(G^{-1}(\frac{1+u}{2})) \geq g_i(G_i^{-1}(\frac{1+u}{2}))$$

according to (3.4.1), $\int_0^1 J'(u)g(G^{-1}(\frac{1+u}{2}))du$ is minimized by G_i. So

$$|\tilde{\mathbb{F}}(G_i^{-1}(\frac{1 \pm u}{2}), \mathbf{x}_i)| \geq |\tilde{\mathbb{F}}(G^{-1}(\frac{1 \pm u}{2}), \mathbf{x}_i)|,$$

and

$$\frac{K_2}{K_1^2}(\eta_i, g_i) \geq \frac{K_2}{K_1^2}(\eta_i, g), \text{for all } g \in \mathcal{F}_{\epsilon_i},$$

because

$$\frac{K_2}{K_1^2}(\eta_i, g) = E\{[\tilde{\mathbb{F}}(G^{-1}(\frac{1 \pm u}{2}), \mathbf{x}_i]^2\}.$$

□

From (3.4.4) and (3.4.8), ψ, ϕ^+ and c_i are obtained as

$$\psi(t) = \begin{cases} t\,, & \text{for } 0 \leq t \leq 1\,, \\ 1\,, & \text{for } t > 1\,, \end{cases} \tag{3.4.11}$$

$$\phi^+(u) = G_i^{-1}(\frac{u+1}{2})\,, 0 \leq u \leq 1, \tag{3.4.12}$$

$$c_i = k_i\,. \tag{3.4.13}$$

Those are the optimal functions for Huber-type minimax problem.

Remark: The only special property of the normal distribution used here is strong unimodality - i.e. that $-\frac{g_0'}{g_0}(e)$ is increasing. The results then extend to any strongly unimodal distribution G_0.

CHAPTER 4

COMPUTATION

After studying the bounded influence R-estimator theoretically in the previous two chapters, we want to investigate some numerical aspects in this chapter. In order to compute this R-estimator, iterative methods should be used. In particular, if we have a good starting value for the estimator, then the one-step method can be used and is effective. Usually k-step ($k \geq 2$) iteration should be done before a satisfactory estimate is obtained. In section 4.1, one-step method is explained, and a SAS program is developed for the R-estimator which is based on optimal functions obtained in section 3.3. In section 4.2, an example is given to compare the bounded influence R-estimator with other estimators (Huber (1973), Krasker and Welsch (1982) and Tableman (1990)) numerically. Simulation is done in section 4.3. The results show that the bounded influence R-estimator is very effective to deal with large leverage points and easy to apply in practice.

4.1 A SAS program to compute the bounded influence R-estimator

From the previous chapter, the optimal functions for Hampel-type optimality problem were derived and given by

$$\psi(t) = \begin{cases} t, & \text{for } t \leq b\,, \\ b, & \text{for } t > b\,, \end{cases} \tag{4.1.1}$$

$$\phi^+(B(e)) = -\frac{g_0'(e)}{g_0(e)} \cdot \text{sign}(e)\,, \tag{4.1.2}$$

$$v(z) = \frac{1}{z}\,. \tag{4.1.3}$$

Assuming $g_0(e)$ is the normal density in this chapter, so (4.1.2) is simplified as $\phi^+(B(e)) = |e|$, also we notice that $B(e) = (2G_0(e) - 1) \cdot \text{sign}(e)$, then function ϕ^+ can be written as

$$\begin{aligned} \phi^+(t) &= G_0^{-1}(\frac{t+1}{2}) \\ &= \Phi^{-1}(\frac{t+1}{2})\,, \quad 0 \leq t \leq 1, \end{aligned} \tag{4.1.4}$$

here Φ is the standard normal distribution. Back to Chapter 1, (1.2.16) gives

$$S(\boldsymbol{\theta}) = \sum_{i=1}^{n} v(x_i)\psi[\frac{\phi^+(R_i^+/(n+1))}{v(\mathbf{x}_i)}] \cdot \text{sign}(y_i - \mathbf{x}_i^T\boldsymbol{\theta})\mathbf{x}_i.$$

While doing computation, we also assume $\mathbf{x}_i$'s are observed values and valued from spherical distribution. If $\mathbf{x}_i$'s are not centered at the origin, a transformation should be made before applying the following procedure. A transformation could be like this: $\mathbf{x}_i' = (\mathbf{x}_i - \text{median}(\mathbf{x}))/Std(\mathbf{x}_i)$, where $Std(\mathbf{x}_i)$ is standard deviation. After transformation, it is reasonable to use $v(z_i) = v(\mathbf{x}_i)$, where $z_i = ||\mathbf{x}_i||$. Now $S(\boldsymbol{\theta})$ can be expressed as

$$S(\boldsymbol{\theta}) = \sum_{i=1}^{n} v(z_i)\psi[\frac{1}{v(z_i)} \cdot \Phi^{-1}(\frac{R_i^+}{2(n+1)} + \frac{1}{2})] \cdot \text{sign}(y_i - \mathbf{x}_i^T\boldsymbol{\theta})\mathbf{x}_i, \qquad (4.1.5)$$

with R_i^+ = rank of $|y_i - \mathbf{x}_i^T\boldsymbol{\theta}|$. Our estimate $\hat{\boldsymbol{\theta}}$ is the solution to $S(\boldsymbol{\theta}) = \mathbf{0}$. The asymptotic equation for (4.1.5) is

$$S(\boldsymbol{\theta}) = n \int v(\mathbf{x})\psi(\frac{|e|}{v(\mathbf{x})}) \cdot \text{sign}(e)\mathbf{x}dG_0(e)dH_0(\mathbf{x}) \,. \qquad (4.1.6)$$

In (4.1.6), taking derivative with respect to $\boldsymbol{\theta}$, we have

$$\frac{\partial S(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \doteq -n \int_{|e|<bv(\mathbf{x})} \mathbf{x}\mathbf{x}^T dG_0(e)dH_0(\mathbf{x}) \,, \qquad (4.1.7)$$

here the approximation is in the sense of equality except for a finite number of points of $\boldsymbol{\theta}$. Now, let

$$\tilde{B} = diag\{\int_{|e|<bv(\mathbf{x}_i)} dG_0(e)\} \,, \qquad (4.1.8)$$

and

$$X = \begin{pmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \mathbf{x}_n^T \end{pmatrix}$$

being $(n \times k)$ design matrix, then $-X'\tilde{B}X$ is the method of moments estimate of (4.1.7).

One-step method was introduced by Bickel (1975) to calculate Huber's M-estimator in the linear model, and also used by Tableman (1990) to compute R-estimator. It uses initial estimate $\boldsymbol{\theta}_0$ to do one iteration according to some equations. The result, say $\boldsymbol{\theta}_1$, is called a one-step estimate. The initial estimator $\boldsymbol{\theta}_0$ should be consistent for the true parameter $\boldsymbol{\theta}_*$ and satisfy $||\boldsymbol{\theta}_0 - \boldsymbol{\theta}_*|| = \mathbb{O}_p(n^{-\frac{1}{2}})$.

Here, one-step method can be developed as following:

$$\begin{aligned}\boldsymbol{\theta}_1 &= \boldsymbol{\theta}_0 - [\frac{\partial S(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}]^{-1} S(\boldsymbol{\theta}_0) \\ &\approx \boldsymbol{\theta}_0 + (X'\tilde{B}X)^{-1}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} S(\boldsymbol{\theta}_0)\,, \end{aligned} \tag{4.1.9}$$

where $\boldsymbol{\theta}_0$ is an initial estimate of true value $\boldsymbol{\theta}_*$ and satisfies $||\boldsymbol{\theta}_0 - \boldsymbol{\theta}_*|| = O_p(n^{-\frac{1}{2}})$. The covariance matrix of $\boldsymbol{\theta}_1$ is then given by

$$V = (X'\tilde{B}X)^{-1}(X'MX)(X'\tilde{B}X)^{-1}, \tag{4.1.10}$$

with

$$M = diag\{E_e(v^2(\mathbf{x}_i)\psi^2[\frac{1}{v(\mathbf{x}_i)}\phi^+(B(e))]\}. \tag{4.1.11}$$

The motivation behind this method is expanding $S(\boldsymbol{\theta})$ at $S(\boldsymbol{\theta}_0)$, i.e.

$$S(\boldsymbol{\theta}) \approx S(\boldsymbol{\theta}_0) + (\frac{\partial S(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}})_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}(\boldsymbol{\theta} - \boldsymbol{\theta}_0),$$

set $S(\boldsymbol{\theta}) = \mathbf{0}$, solve for $\boldsymbol{\theta}$ from the above equation and get

$$\boldsymbol{\theta} = \boldsymbol{\theta}_0 - [(\frac{\partial S(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}})_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}]^{-1} S(\boldsymbol{\theta}_0),$$

this is (4.1.9). And the asymptotic covariance matrix for $\boldsymbol{\theta}$ is

$$E[(X'\tilde{B}X)^{-1}S(\boldsymbol{\theta})[S(\boldsymbol{\theta})]^T(X'\tilde{B}X)^{-1}] = (X'\tilde{B}X)^{-1}(X'MX)(X'\tilde{B}X)^{-1}.$$

Remark: While applying the one-step method, we have to be very careful. If $||S(\boldsymbol{\theta}_1)||$ is not smaller than $||S(\boldsymbol{\theta}_0)||$, then a proper multiplier λ should be used, usually $\lambda = 0.5$, and iteration step (4.1.9) becomes

$$\boldsymbol{\theta}_1 = \boldsymbol{\theta}_0 + \lambda(X'\tilde{B}X)^{-1}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} S(\boldsymbol{\theta}_0)\,.$$

In practice, the true value $\boldsymbol{\theta}_*$ is unknown, so there is no good method to check $||\boldsymbol{\theta}_0 - \boldsymbol{\theta}_*|| = O_p(n^{-\frac{1}{2}})$. In this case, we recommend to use iterative method, when $\boldsymbol{\theta}_k$ is computed, $k \geq 1$,

$$\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k + \lambda(X'\tilde{B}X)^{-1}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_k} S(\boldsymbol{\theta}_k),$$

until $||(S(\boldsymbol{\theta}_{k+1})||$ is small. Generally, three to four iterations will give pretty good results regardless of the starting value. This gives us more confidence to apply the R-estimator to real problems.

The main steps to compute R-estimator iteratively,

1. Choose an initial estimate $\boldsymbol{\theta}_0$,
2. Calculate

$$z_i = ||\mathbf{x}_i||,$$
$$v(z_i) = \frac{1}{z_i},$$

where $\mathbf{x}_i$'s are standardized.

3. Compute R_i^+ = rank of $|y_i - \mathbf{x}_i^T\boldsymbol{\theta}_0|$,
4. Calculate

$$\phi^+(\frac{R_i^+}{n+1}) = \Phi^{-1}(\frac{R_i^+ + n + 1}{2(n+1)})$$

where Φ is the normal distribution.

5. Weights computation

$$w_i = \begin{cases} 1\,, & \text{if} \phi^+(\frac{R_i^+}{n+1}) \leq bv(z_i)\,, \\ bv(z_i)/\phi^+(\frac{R_i^+}{n+1})\,, & \text{if} \phi^+(\frac{R_i^+}{n+1}) > bv(z_i)\,, \end{cases}$$

where b is determined by average weights, say $\frac{1}{n}\sum_{i=1}^{n} w_i = .95$.

6. Compute

$$S(\boldsymbol{\theta}_0) = \sum_{i=1}^{n} w_i \phi^+(\frac{R_i^+}{n+1})\text{sign}(e_i)\mathbf{x}_i,$$

7. Let $\hat{B}$ be a consistent estimate of $\tilde{B}$, computed by

$$\hat{B} = \text{diag}\{(\sum_{|e_j|<bv(\mathbf{x}_i)} 1/n)\},$$

8. Estimate M by $\widehat{M}$

$$\widehat{M} = \text{diag}\{(\sum_{j} v^2(\mathbf{x}_i)\psi^2[\frac{1}{v(\mathbf{x}_i)}\phi^+(\frac{R_j^+}{n+1})])\},$$

9. $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_0 + (X'\hat{B}X)^{-1}S(\boldsymbol{\theta}_0)$,

10. $V = (X'\hat{B}X)^{-1}(X'\widehat{M}X)(X'\hat{B}X)^{-1}$,

11. If $||S(\boldsymbol{\theta}_1)|| \leq a$ ("a" a small number), then stop, else $\boldsymbol{\theta}_0 = \boldsymbol{\theta}_1$ and go to step 2. □

The program is written in SAS.

4.2. An Example

We use the same mortality data described in Henderson and Velleman (1981) as in Krasker and Welsch (1982). The original data (McDonald and Ayers 1973) consist of age-adjusted mortality in 60 United States standard metropolitan statistical area (the response) and 15 variables measuring socioeconomic, climatological, and air pollution features. After considerable analysis, Henderson and Velleman settle on four nonpollution explanatory variables (percent nonwhite, average years of education, population per square mile and precipitation) and a pollution variable related to sulfur dioxide and oxides of nitrogen. Based on an analysis of partial regression plots we choose to use as our fifth explanatory variable just the logarithm of sulfur dioxide potential. We also omit the four

dummy variables on Lancaster, York, and Miami and New Orleans that they used. We feel that a good bounded-influence regression method would probably give less weight to these observations since they appear to be overly influential in the Henderson-Velleman analysis.

For the above data set, the linear regression model is

$$MORT = \theta_0 + \theta_1 RAIN + \theta_2 EDUC + \theta_3 POPDEN + \theta_4 NONW + \theta_5 LOGSO_2 + \varepsilon \tag{4.2.1}$$

where

$$\begin{aligned} MORT &= \text{mortality rate,} \\ RAIN &= \text{percipitation,} \\ EDUC &= \text{education,} \\ POPDEN &= \text{population/mile}^2, \\ NONW &= \%\ \text{non-white,} \\ LOGSO_2 &= \text{logarithm of } SO_2 \text{ potential.} \end{aligned}$$

Since the data are not centered at the origin, a transformation is made to standardize these variables. The transformation is $x_i' = (x_i$ - median $(x))/std(x)$, here x can be any one of $MORT$, $RAIN$, $EDUC$, $POPDEN$, $NONW$ and $LOGSO_2$, median (x) is the median of x for the sample, Std is the deviation of x for the sample, and x_i is the ith observation. Then the iteration program (in the previous section) is executed based on transformed data.

Two starting values are chosen to run the iteration program. One is the least squares estimate (hereafter denoted as LS), another is the Wilcoxon rank estimate (denoted by W^* hereafter) which can be obtained using the rank-regression

procedure *RREGRESS* in Minitab. These two starting values are also transformed according to the standardized data, i.e.

$$\theta_1' = \theta_1 \cdot Std(RAIN)/Std(MORT),$$

$$\theta_2' = \theta_2 \cdot Std(EDUC)/Std(MORT),$$

$$\theta_3' = \theta_3 \cdot Std(POPDEN)/Std(MORT),$$

$$\theta_4' = \theta_4 \cdot Std(NONW)/Std(MORT),$$

$$\theta_5' = \theta_5 \cdot Std(LOGSO_2)/Std(MORT),$$

$$\theta_0' = (-Med(MORT) + \theta_0 + \theta_1 \cdot Med(RAIN) + \theta_2 \cdot Med(EDUC)$$
$$+ \theta_3 \cdot Med(POPDEN) + \theta_4 \cdot Med(NONW) + \theta_5 \cdot Med(LOGSO_2))/Std(MORT),$$

here Med and Std stand for the sample median and standard deviation respectively. After we get the final estimate from the iteration program, the reverse transformation should be used in order to get the θ_i's in the original linear model.

For this example, $b = 3.5$ yields average weights of $.951 - .955$ in the following steps, $LS-1, LS-2$, $LS-3$, W^*-1, W^*-2 and W^*-3. Where $LS-1$ (2, or 3) stands for 1st (2nd or 3rd) iteration using LS as initial value, and W^*-1 (2, or 3) stands for 1st (2nd or 3rd) iteration using W^* as initial value. Table 1 gives the estimated coefficients and standard errors to compare with other procedures: Huber (1973) (M-estimator), Krasker and Welsch (KW) (1982) (Bounded influence M-estimator), and Tableman (1990) (Bounded influence R-estimator based on Wilcoxon scores). $T1$ and $T3$ stand for Tableman's one-step and three-step using LS as initial value respectively, and TW^*1 stands for Tableman's one-step using W^* as initial value.

Table 1 Coefficient estimates

	NONW	*EDUC*	*POPDEN*	*RAIN*	$LOGSO_2$	*CONST*
LS	3.35	-13.296	.00283	1.637	13.778	930.28
	(.59)	(6.97)	(.00376)	(.62)	(3.82)	(96.15)
Huber	2.84	-12.932	.00395	1.868	14.902	915.657
	(.48)	(5.72)	(.00308)	(.51)	(3.13)	(78.87)
KW	2.601	-13.67	.00713	2.01	13.611	915.229
	(.67)	(6.12)	(.00468)	(.44)	(4.10)	(80.97)
W^*	(2.93)	-13.55	.0043	1.897	14.6	919.74
	(.506)	(5.99)	(.0032)	(.53)	(3.28)	(82.73)
$T1$	2.88	-12.427	.006	2.16	14.534	891.674
	(.54)	(6.44)	(.0036)	(.57)	(3.53)	(88.66)
$T3$	2.905	-13.918	.007	2.027	13.569	913.676
	(.4946)	(5.01)	(.0033)	(.53)	(3.24)	(81.38)
TW^*1	2.887	-13.05	.007	2.045	14.16	902.62
	(.495)	(5.92)	(.0033)	(.52)	(3.25)	(81.47)
$LS-1$	2.94	-12.29	.0048	1.996	14.894	900.39
	(.446)	(5.40)	(.0028)	(.47)	(2.85)	(80.23)
$LS-2$	2.86	-13.03	.0057	1.958	14.450	909.00
	(.448)	(5.43)	(.0028)	(.47)	(2.88)	(80.56)
$LS-3$	2.80	-13.15	.0060	1.964	14.172	910.77
	(.448)	(5.42)	(.0028)	(.47)	(2.87)	(79.80)
W^*-1	2.84	-12.78	.0057	1.955	14.497	906.34
	(.448)	(5.42)	(.0028)	(.47)	(2.87)	(79.79)
W^*-2	2.80	-13.14	.0060	1.954	14.152	910.96
	(.448)	(5.42)	(.0028)	(.47)	(2.87)	(79.80)
W^*-3	2.79	-13.46	.0063	1.945	14.01	914.43
	(.449)	(5.42)	(.0028)	(.47)	(2.87)	(79.85)

From Table 1, we find that LS estimate is not close to true value, therefore iterative method is used. $\lambda = 0.5$ yields quick convergence, let's observe $S(\boldsymbol{\theta})$,

$$S(\boldsymbol{\theta}_0) = \begin{pmatrix} -3.69 \\ -2.29 \\ 6.35 \\ 2.44 \\ 3.65 \\ 1.92 \end{pmatrix}, S(\boldsymbol{\theta}_1) = \begin{pmatrix} -1.36 \\ -0.69 \\ 1.73 \\ -0.49 \\ 0.14 \\ -0.12 \end{pmatrix}, S(\boldsymbol{\theta}_2) = \begin{pmatrix} -0.84 \\ -0.08 \\ 0.66 \\ -0.17 \\ 0.43 \\ -0.66 \end{pmatrix}.$$

$||S(\boldsymbol{\theta})||$ is decreasing. Coefficient estimates from first three iterations are in Table 1. Two-step and three-step give very good estimates with smaller estimated standard error than other procedures. Although W^* is near true value $\boldsymbol{\theta}_*$, we don't know if it satisfies $||\boldsymbol{\theta}_0 - \boldsymbol{\theta}_*|| = O_p(n^{-\frac{1}{2}})$ or not. So one-step is not enough to give good estimate. Three steps are done. $S(\boldsymbol{\theta})$ from each step is

$$S(\boldsymbol{\theta}_0) = \begin{pmatrix} -0.49 \\ 0.01 \\ 3.54 \\ -0.26 \\ 0.65 \\ 2.18 \end{pmatrix}, S(\boldsymbol{\theta}_1) = \begin{pmatrix} -0.47 \\ -0.42 \\ 0.67 \\ 0.04 \\ -0.48 \\ 0.68 \end{pmatrix}, S(\boldsymbol{\theta}_2) = \begin{pmatrix} -0.18 \\ -0.41 \\ 0.54 \\ 0.05 \\ 0.06 \\ -0.08 \end{pmatrix}.$$

These two procedures (one with initial value = LS and another with initial value = W^*) perform very well, results are almost the same (comparing $LS - 3$ with $W^* - 3$).

Final weight on observations (cities) where at least one method give a weight less than 1 are reported in Table 2.

Table 2 shows that our procedure does put less weight on Lancaster, Miami, New Orleans and York as we expected in the beginning of this example.

Table 2 Final Weights

City method	*Huber*	*KW*	*T*1	*T*3	*TW**1	*LS* − 1	*LS* − 3	*W* * − 1
Albany	.54	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Birmingham	1.00	.89	.67	.68	.70	.63	.68	.68
Buffalo	.80	.99	.83	.92	.87	.88	.96	.88
Lancaster	.41	.38	.78	.78	.79	.71	.71	.71
Los Angeles	1.00	1.00	.85	.87	.93	.92	1.00	1.00
Memphis	1.00	.95	.73	.73	.76	.68	.73	.73
Miami	.56	.29	.50	.50	.50	.55	.55	.58
Nashville	.97	1.00	1.00	1.00	1.00	1.00	1.00	1.00
New Orleans	.35	.22	.49	.50	.50	.35	.35	.35
St. Louis	1.00	1.00	1.00	.90	.99	1.00	1.00	1.00
San Jose	1.00	1.00	.85	1.00	1.00	.94	1.00	1.00
Toledo	.85	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Washington	.86	1.00	.79	.73	.75	1.00	1.00	1.00
York	.55	.19	.37	.37	.37	.40	.35	.35

4.3 Simulation

In this section, the linear model is

$$y = \beta_0 + \beta_1 x + \varepsilon \tag{4.3.1}$$

Random observations are generated as following:

· Total number of observations is 40,

· x is distributed as $(1-\varepsilon)\ N(5, 2^2) + \varepsilon \cdot H$, $\varepsilon = 0.05$,

where H is any distribution. Then 38 observations are generated from $N(5, 2^2)$ and 2 observations from other distribution H,

· True relation between y and x is $y = 1 - x + \varepsilon$, here error ε is distributed as $N(0, 1)$.

Two situations are considered here. 1) observations include outliers in Y-space, 2) observations include high leverage points. For the two situations, some estimates react differently.

Comparing Huber's M-estimator (1973) with Bounded influence R-estimator (BRE) which uses optimal score functions, we have results in Table 3.

Table 3 Simulation result

		Case 1			Case 2		
		est.	s.e	MSE	est	s.e	MSE
LS	(θ_0)	0.578	0.487	0.415	-2.248	0.26	10.26
	(θ_1)	-0.748	0.089	0.072	-0.250	0.04	0.564
Huber	(θ_0)	0.912	0.496	0.253	-2.193	0.25	10.26
	(θ_1)	-0.950	0.094	0.011	-0.260	0.04	0.549
BRE	(θ_0)	0.932	0.491	0.246	0.957	0.275	0.078
	(θ_1)	-0.952	0.095	0.011	-0.967	0.068	0.0046

In Table 3, est. is estimate of (θ_0, θ_1) which is obtained by

$$\bar{\hat{\theta}}_0 = \frac{1}{N}\sum_{i=1}^{N}\hat{\theta}_0^i,\ \bar{\hat{\theta}}_1 = \frac{1}{N}\sum_{i=1}^{N}\hat{\theta}_1^i,$$

where $\hat{\theta}_0^i$ and $\hat{\theta}_1^i$ are the estimate on the i^{th} run (letting error ε be generated by different normal seeds). s.e. is calculated by

$$s.e.(\theta_j) = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(\hat{\theta}_j^i - \bar{\hat{\theta}}_j)^2}\ ,\quad j = 0, 1.$$

And MSE is the mean squared error, given by

$$MSE(\theta_j) = [s.e.(\theta_j)]^2 + (\theta_{*j} - \bar{\hat{\theta}}_j)^2\ ,\quad j = 0, 1,$$

here $\boldsymbol{\theta}_* = (1, -1)$ is the true value. $N = 35$.

Case 1: 38 observations of x from $N(5,2^2)$, e from $N(0,1)$ and $y = 1 - x + e$. 2 observations from $x = 5.5$, e is $N(0,1)$ and $y = 10 + e$. So y is outlier for last two points. However x is not leverage point.

Case 2: 38 observations of x from $N(5,2^2)$, e from $N(0,1)$ and $y = 1 - x + e$. 2 observations from $x = 18$, e is $N(0,1)$ and $y = -4 + e$. Therefore x is leverage point for the last two observations. Iterative method is used for both estimators (Huber and BRE).

Conclusion: 1) BRE and Huber's estimate are reacting very well on Case 1, they are very close to true value.

2) BRE is also good on Case 2, it gives true value and has small MSE. But Huber's estimate is far away from true value.

In Section 4.2, the data set from that example doesn't have large leverage points since least squares estimate is not too bad and Huber's estimate is close to true value, so all those methods in Table 1 have good results. When there are large leverage points in data set, BRE will give good estimate. Tableman's one-step estimate is also good provided starting value satisfying $||\boldsymbol{\theta}_0 - \boldsymbol{\theta}_*|| = O_p(n^{-\frac{1}{4}})$. However when starting value is not close to true value, Tableman's one-step method is not effective, and three-step iteration converges very slowly.

Remark:

1) Since Bounded influence R-estimator is not invariant under linear transformations, standardized transformation is used before applying iterative program. We require the transformation be robust.

2. How do we choose λ in the iterative method? First we can try $\lambda = 1$, if $||S(\boldsymbol{\theta}_{k+1})|| \geq ||S(\boldsymbol{\theta}_k)||$ then $\lambda = 0.5$. If $||S(\boldsymbol{\theta}_{k+1})|| \geq ||S(\boldsymbol{\theta}_k)||$ when $\lambda = 0.5$, then $\lambda = 0.25$ and so on. This method will guarantee the convergence of the estimate.

References

1. Bickel, P.J. (1975). "One-step Huber estimates in the linear model", Journal of the American Statistical Association, 20, 428-434.

2. Eaton, M.L. (1981). "On the projections of isotropic distributions", Annals of Statistics, 9, 391-400.

3. Hampel, F.R. (1968). "Contributions to the theory of robust estimation", Ph.D. Thesis, University of California, Berkeley.

4. Hampel, F.R. (1974). "The influence curve and its role in robust estimation", Journal of the American Statistical Association, 62, 1179-1186.

5. Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A. (1986), "Robust Statistics: The approach based on influence functions", Wiley, New York.

6. Henderson, H.V. and Velleman, P.F. (1981). "Building multiple regression models interactively", Biometrics, 37, 391-411.

7. Huber, P.J. (1973). "Robust regression: asymptotics, conjectures, and Monte Carlo", The Annals of Statistics, 1, 799-821.

8. Huber, P.J. (1981). "Robust Statistics", Wiley, New York.

9. Huber, P.J. (1983). "Minimax aspects of bounded influence regression", Journal of the American Statistical Association, 78, 66-72.

10. Jaeckel, L.A. (1971). "Robust estimates of location: Symmetry and asymmetric contamination", The Annals Mathemathical Statistics, 42, 1020-1034.

11. Jaeckel, L.A. (1972), "Estimating regression coefficients by minimizing the dispersion of the residuals", The Annals of Mathematical Statistics, 43, 1449-1458.

12. Krasker, W.S. and Welsch, R.E. (1982). "Efficient bounded-influence regression estimation", Journal of the American Statistical Association, 77, 595-604.

13. Maronna, R.A. and Yohai, V.J. (1981). "Asymptotic behavior of general M-estimates for regression and scale with random carriers", Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 58, 7-20.

14. McDonald, G.C. and Ayers, J.A. (1978). "Some applications of the 'Chernoff faces': A technique for graphically representing multivariate data", In Graphical Representation of Multivariate Data, Ed. P.C.C. Wang, New York, Academic Press. pp. 183-197.

15. Mckean, J.W. (1978). "A robust analysis of the general linear model based on one step R-estimates", Biometrika, 65, 571-579.

16. Mckean, J.W. and Hettmansperger, T.P. (1976). "Tests of hypotheses based on ranks in the general linear model", Communications in Statistics, part A, 5, 693-709.

17. Montgomery, D.C. and Peck, E.A. (1982). "Introduction to linear regression analysis". John Wiley & Sons.

18. Tableman, M. (1990). "Bounded-influence rank regression: A one-step estimator based on Wilcoxon scores", Journal of the American Statistical Association 85, 508-513.

19. Wang, Y. (1990). "Optimal, robust, rank-based estimators in the linear model", University of Alberta, Master's Thesis.

20. Wang, Y. and Wiens, D.P. (1991). "Optimal, robust, R-estimators and test statistics in the linear model", University of Alberta, Technical Report 91-04. To appear in Statistics and Probability Letters.

21. Wiens, D.P. (1990). "Minimax-variance L- and R-estimators of location", The Canadian Journal of Statistics. Vol. 18, No. 1, 1990, 47-57.