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.

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.

UNIVERSITY OF ALBERTA

OPTIMAL, ROBUST, RANK-BASED ESTIMATORS IN THE LINEAR MODEL

BY

YU WANG ©

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, 1990

ISBN 0-315-65107-5

UNIVERSITY OF ALBERTA

RELEASE FORM

Name of Author: Yu Wang

TITLE OF THESIS OPTIMAL, ROBUST, RANK-BASED ESTIMATORS IN THE LINEAR MODEL

DEGREE FOR WHICH THESIS WAS PRESENTED MASTER OF SCIENCE

YEAR THIS DEGREE WAS GRANTED: 1990

(signed)

Permanent Address:

#205, 10728 - 80 Avenue
Edmonton, Alberta T6E 1V8
Canada

Date: Oct. 11, 1990

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

"Optimal, robust, rank-based estimators
in the linear model"

submitted by Yu Wang in partial fulfillment of the requirements for the degree of Master of Science.

Supervisor D. Wiens

Jerome N. Sheahan

Jerome N. Sheahan

M. Shirvani

M. Shirvani

N.G.N. Prasad

Sept 5, 1990

Date

ACKNOWLEDGEMENTS

I am greatly indebted to my supervisor, Dr. D.P. Wiens for his advice and suggestions and for the considerable time and effort that he has spent during the preparation of this thesis.

I am grateful to Dr. N. Prasad, Dr. J. Sheahan, and Dr. M. Shirvani for their concerns about my thesis.

Finally, I would like to thank Mrs. Christine Fischer for her accurate typing of this thesis.

TABLE OF CONTENTS

ABSTRACT

In this thesis, we will discuss the optimal estimators with R-estimation and testing methods in linear model. First, we will review the classical analysis of linear model and introduce the results obtained previously with the R-estimation and testing methods. Then we will give definitions to influence functions for estimators and for tests and derive them. The optimality criteria will be given both for estimation and testing. After that we will give the solutions for the optimality problem in general, for Hampel-type optimality problem and for Huber-type optimality problem. Finally, we will show the numerical results for estimation and testing for several estimators.

CHAPTER 1
INTRODUCTION

1.1 Linear Model and Classical Analysis of it.

The linear model is one of the most widely used models. It has received great attention both in theory and in practice. Many statisticians have developed the theory of estimation and testing of linear hypotheses in linear models, and applied the results to solve practical problems.

Generally we consider the following linear model.

Let $\{(x_i, y_i) : i = 1, \ldots n\}$ be a sequence of independent identically distributed random variables such that

$$y_i = x_i^T \theta + e_i \quad i = 1, \ldots n \tag{1.1.1}$$

where

$y_i \in \mathbb{R}$,

$x_i \in \mathbb{R}^p$,

$\theta \in \Omega \subset \mathbb{R}^p$ is a $p-$ vector of unknown parameters,

$e_i \in \mathbb{R}$ is the ith error,

or in matrix form

$$Y = X\theta + e \tag{1.1.2}$$

where

$$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad X = \begin{pmatrix} x_1^T \\ \vdots \\ x_n^T \end{pmatrix}, \quad e = \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix}.$$

It is assumed that Ω is open and convex and that e_i is independent of x_i and has distribution function $G(e)$ and density $g(e)$ with respect to Lebesgue measure.

Let $H(x)$ be the distribution function of x_i, with density $h(x)$ with respect to Lebesgue measure. If we denote by $f_\theta(x, y)$ the joint density of (x_i, y_i) then,

$$f_\theta(x, y) = g(y - x^T\theta)h(x). \tag{1.1.3}$$

In the usual formalization, we consider a linear model in which $x_1, \dots x_n$ are fixed constants and $y_1, \dots, y_n$ are observations.

In the linear model the most important problems are the estimation of parameters and the testing of linear hypotheses.

Classical estimation and testing theories are based on the well-known method of Least Squares which was introduced by Legendre and Gauss at the beginning of the nineteenth century. The Least Squares estimate $(T_{LS})_n$ of θ is defined to be the θ which minimizes

$$D(\theta) = \sum_{i=1}^{n} (y_i - x_i^T \theta)^2 \ . \tag{1.1.4}$$

That is

$$D((T_{LS})_n) = \min\{D(\theta) | \theta \in \Omega\} \ .$$

Some of the assumptions which are made about the error e_i in the classical analysis are listed below

1) $E(e_i) = 0 \quad i = 1, \dots, n$
2) $e_1, \dots e_n$ are independent identically normally distributed i.e. $e_1, \dots, e_n \sim$ i.i.d. $N(0, \sigma^2)$

Huber (1981) and Hampel et al (1986) and many other statisticians have found that when the above assumptions are violated then the LS method will have a poor performance. This means the classical analysis is not robust against departures from these assumptions. So people prefer the robust methods which behave well under the circumstances that the above assumptions hold or are slightly violated.

In general the testing problem for linear model is to test the linear hypothesis

$$\ell_j(\theta) = 0 \quad j = q+1, \dots, p \tag{1.1.5}$$

where $\ell_{q+1}, \dots \ell_p$ are $(p-q)$ linearly independent linear estimable functions and $0 < q < p$.

Through a transformation of the parameters we can reduce this hypothesis to

$$H: \ \theta_{q+1} = \dots = \theta_p = 0 \ . \tag{1.1.6}$$

Let ω be the subspace of Ω obtained by imposing the condition H_0. Then the classical critical region for testing this hypothesis is

$$\{D^* > F_{1-\alpha,p-q,n-p}\} \tag{1.1.7}$$

where

$$D^* = \frac{\sum_{i=1}^{n}\{(y_i - x_i^T(T_\omega)_n)^2 - (y_i - x_i^T(T_\Omega)_n)^2\}/(p-q)}{\sum_{i=1}^{n}(y_i - x_i^T(T_\Omega)_n)^2/(n-p)} \tag{1.1.8}$$

and $(T_\Omega)_n$ and $(T_\omega)_n$ are the LS estimates of θ in the full and reduced model, respectively.

We can see that the classical testing procedure also depends on the Least Squares estimates in the full and reduced model. So, like Least Squares estimates, it is not robust, either.

Because of the lack of robustness of classical estimation and testing procedures people want better alternatives - robust procedures in the linear model.

Since the 1960's several robust methods have been introduced and developed. They are the maximum likelihood type or M-estimation method, the rank-based or R-estimation method, and linear combination of order statistics or L-estimation method.

Any estimator T_n, defined by a minimum problem of the form

$$\sum_{i=1}^{n} \rho(x_i, T_n) = \min! \tag{1.1.9}$$

or by an explicit equation

$$\sum_{i=1}^{n} \psi(x_i, T_n) = 0 \tag{1.1.10}$$

where ρ is an arbitrary function, $\psi(x,\theta) = \frac{\partial}{\partial\theta}\rho(x,\theta)$, is called an M-estimator. An M-estimator T_n for the linear model is defined by the equation

$$\sum_{i=1}^{n} \eta(x_i, (y_i - x_i^T T_n)/\sigma))x_i = 0 \tag{1.1.11}$$

where function η satisfies some regularity conditions.

An L-estimator T_n is of the form

$$T_n = \sum_{i=1}^{n} a_i h(x_{(i)}) \tag{1.1.12}$$

where weights $a_i = \int_{(i-1)/n}^{i/n} dM(u)$, h is some function and $x_{(i)}$ is the i-th order statistic. L-estimator T_n for linear model is defined by Bassett and Koenker (1982) to satisfy a minimum problem

$$\min_{T_n \in \mathbb{R}^p} \sum_{i=1}^{n} P_\varepsilon(y_i - x_i^T T_n) \tag{1.1.13}$$

where $P_\varepsilon(u) = \varepsilon u^+ + (1-\varepsilon)u^-$, $0 < \varepsilon < 1$,and u^+, u^- are the positive and negative parts of u, respectively.

In this article we will focus on the estimation and testing problems only with R-estimation method.

1.2 R-estimation and testing of linear hypotheses based on it.

1.2.1 R-estimation

Let $x_1, \ldots x_m$ and $y_1, \ldots, y_n$ be two independent samples from distributions $F(x)$ and $F(x - \Delta)$, respectively. Let R_i be the rank of x_i in the combined sample with size $m + n$. Let $a_i = a(i)$ $(i = 1, \ldots, m + n)$ be given scores, then a rank test of $\Delta = 0$ against $\Delta > 0$ is based on the test statistic

$$S_{m,n} = \frac{1}{m} \sum_{i=1}^{m} a(R_i)\,. \tag{1.2.1.1}$$

Generally we assume that the scores a_i are generated by a score function J as follows:

$$a_i = J(\frac{i}{m+n+1})\,, \tag{1.2.1.2}$$

or

$$a_i = J(\frac{i-\frac{1}{2}}{m+n+1})\,, \tag{1.2.1.3}$$

or

$$a_i = (m+n) \int_{(i-1)/(m+n)}^{i/(m+n)} J(t)dt\,. \tag{1.2.1.4}$$

We also assume that

$$\sum a_i = 0 \, . \tag{1.2.1.5}$$

So the corresponding requirement for J in (1.2.1.4) is

$$\int_0^1 J(t)dt = 0. \tag{1.2.1.6}$$

Definition 1.2.1:

An R-estimator of location is defined as $T_n = T_n(x_1, \ldots, x_n)$ when T_n is chosen in order that (1.2.1.1) becomes as close to zero as possible when computed from samples $x_1, \ldots x_n$ and $2T_n - x_1, \ldots, 2T_n - x_n$.

This definition is for the one-sample case. It means that in (1.2.1.1) if we let $m = n$ and replace each y_i by $2T_n - x_i$ then we can obtain the R-estimator of location by solving the equation $S_{n,n} = 0$. The idea behind this definition is this: From the original sample $x_1, \ldots, x_n$ we can construct a mirror image by replacing each x_i by $T_n - (x_i - T_n)$. We choose T_n such that the test can't detect any shift, i.e. $S_{n,n} = 0$.

Similarly the definition of R-estimator of shift $\Delta_{m,n}$ in two-sample case requires (1.2.1.1) to be as close to zero as possible when computed from samples $x_1, \ldots x_m$ and $y_1 - \Delta_{m,n}, \ldots, y_n - \Delta_{m,n}$.

1.2.2 R-estimation of parameters in linear models

In 1963 Lehmann considered the analysis of variance techniques with R-estimation method. The model he discussed is

$$y_{ij} = \mu_i + e_{ij} \left\{ \begin{array}{ll} i = 1, \ldots, I & \\ & \sum n_i = n \\ j = 1, \ldots, J & \end{array} \right\} \tag{1.2.2.1}$$

with several observations per cell. The e_{ij} are independent identically distributed continuous random variables with density f satisfying

$$\int_{-\infty}^{+\infty} f^2(x)dx < \infty \tag{1.2.2.2}$$

and $\sigma^2 = \text{Var}(e_{ij}) < \infty$. He derived a robust method to estimate the parameters and to test the linear hypothesis.

In 1971 Jurečková generalized Lehmann's results to the general linear model. Later Jaeckel (1972) simplified Jurečková's work and made the results more usable. Hettmansperger and McKean (1977) developed the testing theory with Jaeckel's estimation method.

The model considered by Jaeckel, Hettmansperger and McKean is the same as (1.1.1), but has more restrictions to error term and can be written as

$$y_i = \theta_0 + \sum_{j=1}^{p-1} x_{ij}\theta_j + e_i \quad i = 1, \ldots, n , \tag{1.2.2.3}$$

i.e.

$$y_i = \theta_0 + x_i^T\theta + e_i \quad i = 1, \ldots, n$$

where

$x_i \in \mathbb{R}^{p-1}$ are fixed constants for $i = 1, \ldots, n$

e_i are i.i.d continuous random variables with density g such that $\int_{-\infty}^{\infty} g^2 ds$ and $\sigma^2 = \text{Var}(e_i)$ are finite

$\theta \in \Omega' \subset \mathbb{R}^{p-1}$

or in matrix form

$$Y = \theta_0\mathbf{1} + X\theta + e . \tag{1.2.2.4}$$

In the Least Squares estimation we define θ_{LS} to be the $\boldsymbol{\theta} = \begin{pmatrix} \theta_0 \\ \theta \end{pmatrix} \in \Omega$ which minimizes

$$D(\boldsymbol{\theta}) = \sum_{i=1}^{n} (y_i - \theta_0 - x_i^T\theta)^2 . \tag{1.2.2.5}$$

If we denote

$$e_i(\boldsymbol{\theta}) = y_i - \theta_0 - x_i^T\theta , \tag{1.2.2.6}$$

then

$$D(\boldsymbol{\theta}) = \sum_{i=1}^{n} e_i^2(\boldsymbol{\theta}) . \tag{1.2.2.7}$$

From the robustness point of view, the trouble with the classical method is that it places too much weight on the extreme residuals when the data contains outliers or has a distribution with heavier tail than normal.

To solve this problem Jaeckel defined a function for any $Z = (z_1, \ldots z_n)^T$

$$D_J(Z) = \sum_{i=1}^{n} a(R_i(Z))z_i \tag{1.2.2.8}$$

where $R_i(Z)$ is the rank of z_i among $z_1, \ldots z_n$.

To ensure that D_J is translation invariant (i.e. $D_J(Z + C\mathbf{1}) = D_J(Z)$ for any constant C) and convex it is required that

$$\sum_{i=1}^{n} a(i) = 0 \tag{1.2.2.9}$$

and

$$a(1) \leq \ldots \leq a(n)\,. \tag{1.2.2.10}$$

With (1.2.2.9) we can easily see that $D_J(Z+C\mathbf{1}) = D_J(Z)$ holds for any constant C. So $D_J(Y - \theta_0\mathbf{1} - X\theta)$ does not depend on intercept θ_0. In later discussion we will use the model (1.2.2.3) rather than (1.1.1). D_J can also be expressed as

$$D_J(Z) = \sum_{i=1}^{n} a(R_i(Z))z_i = \sum_{i=1}^{n} a(i)z_{(i)} \tag{1.2.2.11}$$

where $z_{(i)}$ are the ordered residuals. Using $D_J(Y - X\theta)$ as a dispersion measure of residuals we get a function of θ

$$\begin{aligned} D_J(Y - X\theta) &= \sum_{i=1}^{n} a(R_i(Y - X\theta))(y_i - x_i^T\theta) \\ &= \sum_{i=1}^{n} a(i)(y_{(i)} - x_{(i)}^T\theta) \end{aligned} \tag{1.2.2.12}$$

where $R_i(Y - X\theta)$ is the rank of $y_i - x_i^T\theta$ among $y_1 - x_1^T\theta, \ldots, y_n - x_n^T\theta$ and $y_{(i)} - x_{(i)}^T\theta$ are the ordered residuals. Now with (1.2.2.9) and (1.2.2.10) we try to prove that $D_J(Z)$ is nonnegative and a convex function of θ. Let $Z = (z_1, \ldots, z_n)^T$ be a vector and ℓ be the index of the first positive $a(i)$. Then

$$D_J(Z) = \sum_{i=1}^{n} a(i) z_{(i)} = \sum_{i=1}^{n} a(i)(z_{(i)} - z_{(\ell)}) \geq 0 \qquad (1.2.2.13)$$

since every term in the sum is nonnegative. To prove $D_J(Y - X\theta)$ is a convex function of θ we need to show that for any θ', θ'' and $0 \leq t \leq 1$ the inequality

$$tD_J(Y - X\theta') + (1-t)D_J(Y - X\theta'') \geq D_J(Y - X(t\theta' + (1-t)\theta'')) \quad (1.2.2.14)$$

holds. Let $P_1(i), P_2(i)$ and $P_3(i)$ be some permutations of $1, \ldots, n$ and $Y_{P_1}, Y_{P_2}, Y_{P_3}$, $X_{P_1}, X_{P_2}, X_{P_3}$ are the matrices formed by $Y_{P_1(i)}, Y_{P_2(i)}, Y_{P_3(i)}, X_{P_1(i)}, X_{P_2(i)}$ and $X_{P_3(i)}$, respectively. Since

$$\begin{aligned}
&tD_J(Y - X\theta') + (1-t)D_J(Y - X\theta'') - D_J(Y - X(t\theta' + (1-t)\theta'')) \\
&= t\sum_{i=1}^{n} a(i)[Y_{P_1(i)} - X_{P_1(i)}^T\theta'] + (1-t)\sum_{i=1}^{n} a(i)[Y_{P_2(i)} - X_{P_2(i)}^T\theta''] \\
&\qquad - \sum_{i=1}^{n} a(i)[Y_{P_3(i)} - X_{P_3(i)}^T((1-t)\theta'')] \\
&= t\sum_{i=1}^{n} a(i)[(Y_{P_1(i)} - Y_{P_3(i)}) - (X_{P_1(i)}^T - X_{P_3(i)}^T)\theta'] \\
&\qquad + (1-t)\sum_{i=1}^{n} a(i)[(Y_{P_2(i)} - Y_{P_3(i)}) - (X_{P_2(i)}^T - X_{P_3(i)}^T)\theta''] \\
&= tD_J[(Y_{P_1} - Y_{P_3}) - (X_{P_1}^T - X_{P_3}^T)\theta'] + (1-t)D_J[(Y_{P_2} - Y_{P_3}) - (X_{P_2}^T - X_{P_3}^T)\theta''] \geq 0
\end{aligned}$$

due to the nonnegativity of $D_J(Z)$ we have the convexity of D_J. We need to mention here that the formula (1.2.2.12) is reasonable because of an easily seen fact that with probability 1 there will be no ties in the ranks.

Hettmasperger and McKean added the symmetry of the scores

$$a(i) = -a(n+1-i) \quad i = 1, \ldots, n\,.$$

This assumption is not necessary in general and is natural only when the e_i are symmetrically distributed. If not specified, we assume that the error term e_i are symmetrically distributed in our later discussion.

Let $\widehat{\theta}$ be the robust estimator of θ in the linear model with the R-estimation method. That is, $\widehat{\theta}$ is defined to be the θ which minimizes Jaeckel's dispersion measure, i.e.

$$D_J(Y - X\widehat{\theta}) = \min\{D_J(Y - X\theta)|\, \theta \in \Omega'\}$$

where $\Omega' \subset \mathbb{R}^{p-1}$. The value of $\widehat{\theta}$ can be obtained by solving a group of equations

$$\sum_{i=1}^{n}(x_{ij} - x_j)a(R_i(Y - X\theta)) = 0 \quad j = 1, \ldots, p-1$$

where $x_j = n^{-1}\sum_{i=1}^{n} x_{ij}$ and $R_i(Y - X\theta)$ is the rank of $y_i - x_i^T\theta$ among $y_1 - x_1^T\theta, \ldots, y_n - x_n^T\theta$.

The above equations are solved in the sense that the value of $\widehat{\theta}$ makes $\sum_{i=1}^{n}(x_{ij} - x_j)a(R_i(Y - X\widehat{\theta}))$ as close to zero as possible. Jaeckel (1972) showed that the solutions are not necessarily unique, but that the diameter of the solution set goes to zero in probability as $n \to \infty$. Hettmansperger and McKean (1976) have developed several successful algorithms to handle this numerical problem.

To estimate θ_0 after θ it is recommended by Hettmansperger and McKean to use

$$\widehat{\theta}_0 = med_{1 \le i \le j \le n}\{(\widehat{e}_i + \widehat{e}_j)/2\}$$

when e_i are symmetrically distributed and to use (by Aubuchon and Hettmansperger (1989))

$$\widehat{\theta}_0 = med_{1 \le i \le n}\{\widehat{e}_i\}$$

in the absence of symmetry. Here, $\widehat{e}_i = y_i - (X\widehat{\theta})_i$ in both cases.

As at (1.2.1.2),(1.2.1.4) we take scores in the form

$$a_i = J(\frac{i}{n+1}) \quad i = 1, \ldots, n\,,$$

or

$$a_i = n\int_{(i-1)/n}^{i/n} J(t)dt \quad i = 1, \ldots, n\,.$$

Very common and simple scores are the sign scores

$$a(i) = sign\left(\frac{i}{n+1} - \frac{1}{2}\right) ,$$

Wilcoxon scores

$$a(i) = \left(\frac{i}{n+1} - \frac{1}{2}\right) ,$$

and normal scores

$$a(i) = \Phi^{-1}\left(\frac{i}{n+1} - \frac{1}{2}\right) .$$

Policello and Hettmansperger (1976) proposed a mixture of sign and Wilcoxon scores in which $\varepsilon/2$ of the residuals at each end are given sign scores and the remaining $(1-\varepsilon)$ in the middle receive Wilcoxon scores.

1.2.3 Testing of linear hypotheses based on R-estimation

In 1976 McKean and Hettmansperger proposed using Jaeckel's dispersion measure to test linear hypotheses in the linear model. Since under the condition (1.2.2.9) Jaeckel's dispersion measure does not depend on intercept any more, we need only consider testing of linear hypotheses of slopes. Generally we consider testing $H_0: \theta_{2H} = \ldots = \theta_{p-1} = 0$ where $0 < q < p-1$.

Let D_J^* be as follows:

$$D_J^* = D_J(Y - X\widehat{\theta}_\omega) - D_J(Y - X\widehat{\theta}_\Omega)$$

where $\widehat{\theta}_\omega$ and $\widehat{\theta}_\Omega$ are the parameter estimates under reduced and full model with R-estimation, respectively. Then they found that

$$\frac{2rD_J^*}{A^2} \rightarrow \chi^2_{p-q-1} \text{ in distribution under } H_0$$

where

$$r = -\int_0^1 J(t)[g'(G^{-1}(t))/g(G^{-1}(t))]dt$$

$$A^2 = \int_0^1 J^2(t)dt .$$

In the case of asymmetric errors some work has been done. Koul, Sievers, and McKean (1987) and later McKean and Sievers (1989) considered the linear model with skewed errors and gave their method of estimation and testing hypotheses. Aubuchon and Hettmansperger (1989) proposed a way of using Wilcoxon scores to estimate parameters and to test linear hypotheses in the linear model.

A lot of work has been done by many statisticians on the problems of estimating parameters and testing hypotheses with R-estimation in the linear model. Some of them are Adichie (1967, 1978), Aubuchon(1982), Draper (1981, 1988), Hodges and Lehmann (1963), Puri and Sen (1969).

Some people have done their research work on the performance of some specified score functions. The problem of determining optimal score functions for general linear model still remains.

In this article we try to give our answers to the above question. Hampel et al (1986) once discussed this problem in M-estimation. Now we extend it to R-estimation. In Chapter 2, we will discuss R-estimation in the linear model and optimality criteria in estimation. In Chapter 3 we will construct a test statistic, find its asymptotic distribution and discuss the optimality criteria in testing. In Chapter 4 we will give solutions to all above optimality problems. At last in Chapter 5 we will show some numerical study results.

CHAPTER 2

ESTIMATION IN THE LINEAR MODEL

2.1 Definition of influence function for estimators

Later in this chapter we will use the influence function as a tool. Now we give its definition and some of the properties.

Consider a parametric model $(X, A, \{F_\theta : \theta \in \Omega\})$ where X is a complete separable metric space, A is the σ-algebra generated by the topology, and $\{F_\theta : \theta \in \Omega\}$ is a family of distributions on the measurable space $\{X, A\}$ and Ω is a convex subset of $\mathbb{R}^p$ for some integer p.

Let $x_1, \ldots, x_n$ be n independent and identically distributed observations in X and let $\{T_n : n \in \mathbb{N}\}$ be a sequence of estimators of θ such that:

(1) $T_n(x_1, \ldots, x_n) = T_n(F_n)$
where F_n is the empirical distribution.

(2) There exists a functional T on a certain subset of the space of all probability distributions into $\mathbb{R}^p$ such that $T_n(x_1, \ldots, x_n) \to T(F)$ in probability when $n \to \infty$ and the observations are distributed according to F. Generally, $T_n = T$.

(3) We assume that the estimator T is Fisher consistent: $T(F_\theta) = \theta \quad \forall \theta \in \Omega$.

Definition 2.1.1:

Let δ_x be the distribution which puts mass 1 at the point $x \in X$. Then, the influence function of a functional T at a distribution F is given by

$$IF(x, T, F) = \lim_{\varepsilon \searrow 0} \{(T((1-\varepsilon)F + \varepsilon\delta_x) - T(F))/\varepsilon\} .$$

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

$$\frac{d}{d\varepsilon} T((1-\varepsilon)F + \varepsilon G)|_{\substack{\varepsilon=0 \\ G=\delta_x}} .$$

Under some regularity conditions, we have

$$\int IF(x, T, F) dF(x) = 0 \tag{2.1.1}$$

and

$$\sqrt{n}(T(F_n) - T(F)) \to N(0, V(T,F)) \tag{2.1.2}$$

where $V(T,F) = \int IF(x,T,F) \cdot IF(x,T,F)^T dF(x)$. Now we outline the derivation of (2.1.1) and (2.1.2) in one dimensional case. Let $T(F)$ be defined on a class $\mathcal{F}$ of distribution functions and T be Gateaux differentiable. Then by definition there exists a function $\psi_F(z)$ such that for all $G \in \mathcal{F}$,

$$\lim_{\varepsilon \searrow 0} \frac{T((1-\varepsilon)F + \varepsilon G) - T(F)}{\varepsilon} = \int \psi_F(z) dG(z) \, . \tag{2.1.3}$$

By putting $G = \delta_x$ we have

$$IF(x,F,T) = \psi_F(x) \, . \tag{2.1.4}$$

Putting $G = F$ gives

$$\int IF(x,F,T) dF(x) = 0. \tag{2.1.5}$$

Let $\xi(\varepsilon) = T((1-\varepsilon)F + \varepsilon G)$. By Taylor expansion we have

$$\xi(1) = \xi(0) + \xi'(0)\varepsilon|_{\varepsilon=1} + \xi''(\eta)\frac{\varepsilon^2}{2}|_{\varepsilon=1} \text{ for } 0 < \eta < 1 \, , \tag{2.1.6}$$

i.e.

$$T(G) = T(F) + \int IF(x,F,T) dG(x) + \text{ remainder} \, . \tag{2.1.7}$$

By substituting the empirical distribution F_n for G we obtain

$$\begin{aligned} \sqrt{n}(T(F_n) - T(F)) &= \sqrt{n} \int IF(x,F,T) dF_n(x) + \text{ remainder} \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} IF(x_i,F,T) + \text{ remainder} \, . \end{aligned} \tag{2.1.8}$$

The first term on the right-hand side is asymptotically normal by Central Limit Theorem. It is often true that the remaining terms are asymptotically negligible. So $\sqrt{n}(T(F_n) - T(F))$ is asymptotically normal with mean 0 and variance

$$V(T,F) = \int IF(x,T,F)^2 dF(x). \tag{2.1.9}$$

In multidimensional case we need more conditions.

2.2. Influence function of slopes.

We consider the linear model (1.2.2.3), i.e.

$$y_i = \theta_0 + x_i^T\theta + e_i \quad i = 1, \ldots, n \tag{2.2.1}$$

with assumptions that x_i and e_i are independent. x_i has distribution fucntion $H(x)$ and density $h(x)$, and e_i has distribution function $G(e)$ and density $g(e)$ such that $\int_{-\infty}^{\infty} g^2 dx$ and $\sigma^2 = Var(e_i)$ are finite. So the joint density of (x_i, y_i) is

$$f_\theta(x, y) = g(y - x^T\theta - \theta_0)h(x) \tag{2.2.2}$$

In the following part of this section, we consider $x_1, \ldots, x_n$ as constants and $y_1, \ldots, y_n$ as observations. We are going to use the dispersion function D_J that Jaeckel, Hettmansperger and McKean used

$$\begin{aligned} D_J(Y - \theta_0 1 - X^T\theta) &= D_J(Y - X^T\theta) \\ &= \sum_{i=1}^{n} a(R_i(Y - X^T\theta))(y_i - x_i^T\theta) \end{aligned} \tag{2.2.3}$$

where $R_i(Y - X^T\theta)$ is the rank of $y_i - x_i^T\theta$ among $y_1 - x_1^T\theta, \ldots, y_n - x_n^T\theta$. Also, the scores need to be monotone and symmetric:

$$\begin{aligned} &a(1) \leq \ldots \leq a(n)\,, \\ &a(i) = -a(n+1-i) \quad i = 1, \ldots, n\,. \end{aligned} \tag{2.2.4}$$

The scores are chosen in the following form

$$a(i) = n\int_{(i-1)/n}^{i/n} J(t)dt \quad i = 1, \ldots, n \tag{2.2.5}$$

where J is a monotone function and we require that

$$\int_0^1 J(t)dt = 0 \tag{2.2.6}$$

be satisfied.

Let θ_{0*} and θ_* be the true value of θ_0 and θ, respectively. Then we have

$$y_i = \theta_{0*} + x_i^T\theta_* + e_i \ . \tag{2.2.7}$$

We estimate θ_* through

$$\frac{1}{n}\sum_{i=1}^{n} a(R_i(Y - X^T\theta))(x_i - \overline{x}) = 0 \ . \tag{2.2.8}$$

In this chapter finally we will use the joint influence function of slopes and intercept to discuss the optimal scores in estimation. First we need to derive the influence function of slopes.

Let $E_{n,\theta}$ be the empirical distribution of $y_1 - x_1^T\theta, \ldots, y_n - x_n^T\theta$ so that $R_i(Y - X^T\theta) = nE_{n,\theta}(y_i - x_i^T\theta)$ and then (2.2.8) becomes

$$\frac{1}{n}\sum_{i=1}^{n} n\int_{E_{n,\theta}(y_i - x_i^T\theta)-\frac{1}{n}}^{E_{n,\theta}(y_i - x_i^T\theta)} J(t)dt(x_i - \overline{x}) = 0 \ . \tag{2.2.9}$$

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

$$H_n(V) = \frac{\#\text{of } x_i\text{'s in } V \subset \mathbb{R}^{p-1}}{n} \tag{2.2.10}$$

and G_n be the empirical distribution function of e. Then we have

$$F_n(x,e) = (G_n \times H_n)(x,e) \ . \tag{2.2.11}$$

From (2.2.7) we have

$$\begin{aligned} y_i - x_i^T\theta &= \theta_{0*} + x_i^T\theta_* + e_i - x_i^T\theta \\ &= \theta_{0*} + x_i^T(\theta_* - \theta) + e_i \ . \end{aligned} \tag{2.2.12}$$

So (2.2.9) becomes

$$\int [n\int_{E_{n,\theta}(\theta_{0*} + x^T(\theta_* - \theta)+e)-\frac{1}{n}}^{E_{n,\theta}(\theta_{0*} + x^T(\theta_* - \theta)+e)} J(t)dt](x - \mu_{H_n})dF_n(x,e) = 0 \tag{2.2.13}$$

where $\mu_{H_n} = \int x dH_n(x) = \overline{x}$. By definition

$$\begin{aligned} E_{n,\theta}(t) &= \frac{1}{n} I\{y_i - x_i^T\theta \leq t\} \\ &= \frac{1}{n}\sum_{i=1}^{n} I\{\theta_{0*} + x_i^T(\theta_* - \theta) + e_i \leq t\} \\ &= \frac{1}{n}\sum_{i=1}^{n} I\{e_i \leq t - \theta_{0*} - x_i^T(\theta_* - \theta)\}\,. \end{aligned} \tag{2.2.14}$$

Note that $I\{e_i \leq t - \theta_{0*} - x_i^T(\theta_* - \theta)\}$ is a random variable with mean $G\{t - \theta_{0*} - x_i^T(\theta_* - \theta)\}$ and variance

$$G(t - \theta_{0*} - x_i^T(\theta_* - \theta))[1 - G(t - \theta_{0*} - x_i^T(\theta_* - \theta))] \leq 1/4\,,$$

so that

$$E_{n,\theta}(t) - \frac{1}{n}\sum_{i=1}^{n} G(t - \theta_{0*} - x_i^T(\theta_* - \theta)) = \frac{1}{n}\sum_{i=1}^{n} Z_i \tag{2.2.15}$$

where we have $EZ_i = 0$, $Var(Z_i) \leq 1/4$. By Law of Large Numbers

$$\frac{1}{n}\sum_{i=1}^{n} Z_i \to 0 \text{ in probability as } n \to \infty\,.$$

So (2.2.15) becomes

$$E_{n,\theta}(t) - \int G(t - \theta_{0*} - x^T(\theta_* - \theta))dH_n(x) = \frac{1}{n}\sum_{i=1}^{n} Z_i \to 0 \text{ in probabilitiy}$$

as $n \to \infty$. If $H_n \xrightarrow{W} H$ as $n \to \infty$, i.e. $H_n \to H$ in distribution, we have

$$E_{n,\theta}(t) \to \int G(t - \theta_{0*} - x^T(\theta_* - \theta))dH(x) \text{ in probability}\,. \tag{2.2.16}$$

Then

$$E_{n,\theta}(\theta_{0*} + x^T(\theta_* - \theta) + e) \to \int G(e + (x - z)^T(\theta_* - \theta))dH(z) \tag{2.2.17}$$

and the limiting form of (2.2.13) should be

$$\int J[\int G(e + (x - z)^T(\theta_* - \theta))dH(z)](x - \mu_H)dF(x, e) = 0 \tag{2.2.18}$$

with $F(x,e) = H(x)G(e)$. Under $F_0(x,e) = H_0(x)G_0(e)$, Fisher consistency requires

$$\theta_* = \theta(F_0)$$

and by (2.2.18) the defining equation for $\theta = \theta(F)$ is

$$\int J[\int G(e + (x-z)^T(\theta(F_0) - \theta(F)))dH(z)](x - \mu_H)dF(x,e) = \mathbf{0}\,. \qquad (2.2.19)$$

Put

$$\ell(e,x,F) = \int G(e + (x-z)^T(\theta(F_0) - \theta(F)))dH(z)\,. \qquad (2.2.20)$$

Then (2.2.19) is

$$\int J(\ell(e,x,F))(x - \mu_H)dF(x,e) = \mathbf{0}\,. \qquad (2.2.21)$$

To get the influence function, we replace F by

$$\begin{aligned} F_\lambda &= (1-\lambda)F_0 + \lambda F_1 \\ &= (1-\lambda)G_0H_0 + \lambda G_1H_1 \end{aligned}$$

in (2.2.21) and define $\theta(F_\lambda)$ implicitly. Then

$$IF(u, \mathrm{v}, G_0, H_0, J) = \frac{d}{d\lambda}\theta(F_\lambda)|_{\substack{\lambda=0 \\ G_1=\delta_u, H_1=\delta_v}}\,. \qquad (2.2.22)$$

We have

$$\begin{aligned} \ell(e,x,F_\lambda) = (1-\lambda)&\int G_0(e + (x-z)^T(\theta(F_0) - \theta(F_\lambda)))dH_0(z) \\ &+ \lambda\int G_1(e + (x-z)^T(\theta(F_0) - \theta(F_\lambda)))dH_1(z)\,. \qquad (2.2.23) \end{aligned}$$

Put $\dot\theta = \frac{d}{d\lambda}\theta(F_\lambda)|_{\lambda=0}$, then

$$\begin{aligned} \frac{d}{d\lambda}\ell(e,x,F_\lambda)|_{\lambda=0} = &-\int G_0(e)dH_0(z) \\ &+ \int g_0(e + (x-z)^T(\theta(F_0) - \theta(F_\lambda))|_{\lambda=0}(-(x-z)^T\dot\theta)dH_0(z) \\ &+ \int G_1(e)dH_1(z) \\ = &\,(G_1 - G_0)(e) - g_0(e)(x - \mu_{H_0})^T\dot\theta\,. \qquad (2.2.24) \end{aligned}$$

In (2.2.21) replacing F by F_λ, then taking the derivative w.r.t. λ and evaluating at $\lambda = 0$ give

$$\begin{aligned}
0 &= \int J'(\ell(e,x,F_0)]\frac{d}{d\lambda}\ell(e,x,F_\lambda)|_{\lambda=0}(x-\mu_{H_0})dF_0(x,e) \\
&\quad + \int J[\ell(e,x,F_0)](-\frac{d}{d\lambda}\mu_{H_\lambda})dF_0(x,e) \\
&\quad + \int J[\ell(e,x,F_0)](x-\mu_{H_0})d(F_1-F_0)(x,e) \\
&= \int J'[G_0(e)]\{(G_1-G_0)(e)-g_0(e)(x-\mu_{H_0})^T\dot{\theta}\}(x-\mu_{H_0})dF_0(x,e) \\
&\quad - \int J[G_0(e)](\mu_{H_1}-\mu_{H_0})dF_0(x,e) \\
&\quad + \int J[G_0(e)](x-\mu_{H_0})dF_1(x,e) - \int J(G_0(e))(x-\mu_{H_0})dF_0(x,e)\,.
\end{aligned}$$

Since $\int(x-\mu_{H_0})dH_0(x) = 0$ and $\int J(G_0(e))dG_0(e) = 0$, putting $C_0 = \int(x-\mu_{H_0})(x-\mu_{H_0})^T dH_0(x)$, we get

$$\begin{aligned}
0 &= -\int J'(G_0(e))g_0^2(e)de\cdot C_0\dot{\theta} \\
&\quad + \int J(G_0(e))dG(e)\cdot\int(x-\mu_{H_0})dH_1(x)\,,
\end{aligned}$$

so that

$$\dot{\theta} = \frac{\int J(G_0(e))dG_1(e)}{\int J'(G_0(e))g_0^2(e)de}C_0^{-1}(\mu_{H_1}-\mu_{H_0})\,. \tag{2.2.25}$$

From (2.2.22) we get the influence function of slopes

$$IF(u,\mathbf{v},G_0,H_0,J) = \frac{J(G_0(u))}{\int J'(G_0(e))g_0^2(e)de}C_0^{-1}(\mathbf{v}-\mu_{H_0}) \tag{2.2.26}$$

where

$$\begin{aligned}
\mu_{H_0} &= \int x dH_0 = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n x_i \\
C_0 &= \int(x-\mu_{H_0})(x-\mu_{H_0})^T dH_0(x) \\
&= \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n(x_i-\overline{x})(x_i-\overline{x})^T\,.
\end{aligned}$$

2.3 Joint influence function of slopes and intercept.

From the above section we know that the model is

$$y = \theta_{0*} + x^T\theta_* + e \,. \tag{2.3.1}$$

$E_{n,\theta}$ is the empirical distribution function of $y_1 - x_1^T\theta, \dots y_n - x_n^T\theta$ and

$$E_{n,\theta}(t) \to \int G(t - \theta_{0*} - z^T(\theta_* - \theta))dH(z) \text{ in probability}\,.$$

Let $G^*(t) = \int G(t - z^T(\theta_* - \theta))dH(z)$, then

$$y - x^T\theta - \theta_0 = \theta_{0*} - \theta_0 + x^T(\theta_* - \theta) + e \tag{2.3.2}$$

distributed as $G^*(t - (\theta_{0*} - \theta_0))$.

Suppose we have observations $y_1 - x_1^T\theta - \theta_0, \dots, y_n - x_n^T\theta - \theta_0$ from distribution $G^*(t - (\theta_{0*} - \theta_0))$. We construct their mirror images $2(\theta_{0*} - \theta_0) - (y_1 - x_1^T\theta - \theta_0), \dots, 2(\theta_{0*} - \theta_0) - (y_n - x_1^T\theta - \theta_0)$. Let $M_n(z)$ be the empirical distribution function of the combined samples just obtained. Then

$$M_n(z) = \frac{1}{2}[G^*(z - (\theta_{0*} - \theta_0)) + 1 - G^*(-3(\theta_{0*} - \theta_0) - z)]\,. \tag{2.3.3}$$

Let R_i = rank of $y_i - x_i^T\theta - \theta_0 = 2nM_n(y_i - x_i^T\theta - \theta_0)$. So the R -estimator of θ_0 is defined by the equation

$$\begin{aligned} 0 &= \frac{1}{n}\sum_{i=1}^{n} b(R_i) \\ &= \int b(R_i)dF_n(x,e) \\ &= \int 2n\left(\int_{M_n(y-x^T\theta-\theta_0)-\frac{1}{2n}}^{M_n(y-x^T\theta-\theta_0)} K(t)dt\right)dF_n(x,e) \end{aligned} \tag{2.3.4}$$

where $b(i) = 2n\int_{(i-1)/2n}^{1/2n} K(t)dt$ and $K(t)$ has the same properties as $J(t)$. If we select $b(i)$ as Wilcoxon score, i.e. $K(t) = (t-1)/2$, we will get the $\widehat{\theta}_0$ recommended by Hettmansperger and McKean

$$\widehat{\theta}_0 = med_{1\le i\le j\le n}\{(\widehat{e}_i + \widehat{e}_j)/2\}$$

as shown in chapter 1.

If we select $b(i)$ as sign score, i.e. $K(t) = \begin{cases} 1, & 1 \geq t > \frac{1}{2} \\ -1, & \frac{1}{2} > t \geq 0, \end{cases}$ then we get the $\widehat{\theta}_0$ proposed by Aubuchon and Hettmansperger

$$\widehat{\theta}_0 = med_{1 \leq i \leq n}\{\widehat{e}_i\} .$$

Here, $\widehat{e}_i = y_i - (X\widehat{\theta})_i$ in both cases.

Let $n \to \infty$, then $\theta_{0*} - \theta_0 \to 0$ in probability and so

$$M_n(z) = \frac{1}{2}[G^*(z-(\theta_{0*}-\theta_0))+1-G^*(-3(\theta_{0*}-\theta_0)-z)] \xrightarrow{W} \frac{1}{2}[G^*(z)+1-G^*(-z)] .$$

Thus,

$$M_n(y-x^T\theta-\theta_0) \to \frac{1}{2}[G^*(y-x^T\theta-\theta_0)+1-G^*(-y+x^T\theta+\theta_0)] \triangleq \frac{1}{2}\ell(e,x,\theta_0,\theta,F) ,$$

i.e. from the definition of $G^*(t)$

$$\begin{aligned} \ell(e,x,\theta_0,\theta,F) = &\int G(e+\theta_{0*}-\theta_0+(x-z)^T(\theta_*-\theta))dH(z)+1 \\ &-\int G(-e-\theta_{0*}+\theta_0-(x+z)^T(\theta_*-\theta))dH . \end{aligned} \tag{2.3.5}$$

Then θ_0 can be defined from the equation

$$0 = \int K[\frac{1}{2}\ell(e,x,\theta_0,\theta,F)]dF(x,e) . \tag{2.3.6}$$

To get the influence function, we replace F by

$$\begin{aligned} F_\lambda &= (1-\lambda)F_0 + \lambda F_1 \\ &= (1-\lambda)G_0H_0 + \lambda G_1H_1 . \end{aligned}$$

Then

$$\begin{aligned} \ell(e,x,\theta_0,\theta,F_\lambda) = (1-\lambda)[&\int G_0(e+\theta_0(F_0)-\theta_0(F_\lambda) \\ &+(x-z)^T(\theta(F_0)-\theta(F_\lambda)))dH_0(z)- \\ &-\int G_0(-e-\theta_0(F_0)+\theta_0(F_\lambda)-(x+z)^T(\theta(F_0)-\theta(F_\lambda)))dH_0(z)] \\ &+\lambda[\int G_1(e+\theta_0(F_0)-\theta_0(F_\lambda)+(x-z)^T(\theta(F_0)-\theta(F_\lambda))dH_1(z) \\ &-\int G_1(-e-\theta_0(F_0)+\theta_0(F_\lambda) \\ &-(x+z)^T(\theta(F_0)-\theta(F_\lambda)))dH_1(z)] . \end{aligned} \tag{2.3.7}$$

Express

$$\dot\theta_0 = \frac{d}{d\lambda}\theta_0(F_\lambda)|_{\lambda=0}$$
$$\dot\theta = \frac{d}{d\lambda}\theta(F_\lambda)|_{\lambda=0}\ .$$

By calculation we have

$$\begin{aligned}\frac{d\ell(e,x,\theta_0,\theta,F_\lambda)}{d\lambda}|_{\lambda=0} &= -(G_0(e)-G_0(-e)) + g_0(e)[-\dot\theta_0 - (x-\mu_{H_0})^T\dot\theta] \\ &\quad - g_0(-e)[-\dot\theta_0 - (x+\mu_{H_0})^T\dot\theta] + G_1(e) - G_1(-e) \\ &= G_1(e) - G_0(e) + G_0(-e) - G_1(-e) - [g_0(e)+g_0(-e)]\dot\theta_0 \\ &\quad - (g_0(e)(x-\mu_{H_0})^T + g_0(-e)(x+\mu_{H_0})^T)\dot\theta\ . \qquad (2.3.8)\end{aligned}$$

Taking derivatives w.r.t. λ in the equation

$$0 = \int K[\frac{1}{2}\ell(e,x,\theta_0,\theta,F_\lambda)]dF_\lambda(x,e) \qquad (2.3.9)$$

and evaluating at $\lambda = 0$ give

$$\begin{aligned}0 &= \int K'[\frac{1}{2}\ell(e,x,\theta_0,\theta,F_0)]\frac{d\ell}{d\lambda}|_{\lambda=0}\frac{1}{2}dF_0(x,e) \\ &\quad + \int K[\frac{1}{2}\ell(e,x,\theta_0,\theta,F_0)]d(F_1-F_0)(x,e) \\ &= \frac{1}{2}\int K'[\frac{1}{2}(G_0(e)+1-G_0-(-e))]\frac{d\ell}{d\lambda}|_{\lambda=0}dF_0(x,e) \\ &\quad + \int K[\frac{1}{2}(G_0(e)+1-G_0-(-e))]d(F_1-F_0)(x,e) \\ &= \frac{1}{2}\int K'[\frac{1}{2}(G_0(e)+1-G_0(-e))][(G_1-G_0)(e)+(G_0-G_1)(e) \\ &\quad - (g_0(e)+g_0(-e)\dot\theta_0 - 2g_0(-e)\mu_{H_0}^T\dot\theta]dG_0(e) \\ &\quad + \int K[\frac{1}{2}(G_0(e)+1-G_0(-e))]d(F_1-F_0)(x,e)\ . \qquad (2.3.10)\end{aligned}$$

If G_0 is symmetric, $G_0(e) = 1 - G_0(-e)$. From (2.3.10) we have

$$\begin{aligned}0 &= \int K'(G_0(e))[\frac{G_1(-e)+1-G_1(e)}{2}dG_0(e) - \int K'(G_0(e))G_0(e)dG_0(e) \\ &\quad - A(K,G_0)\dot\theta_0 - A(K,G_0)\mu_{H_0}^T\dot\theta + \int K(G_0(e))dG_1(e) - \int K(G_0(e))dG_0(e)\end{aligned}$$

where $A(K,G_0) = \int K'(G_0(e))g_0^2(e)de$. Putting $G_1 = \delta_u$, $H_1 = \delta_{\mathbf{v}}$, through calculation we have

$$0 = K(G_0(u)) - A(K,G_0)[IF_{\theta_0}(u,\mathbf{v},G_0,H_0,J,K) + \mu_{H_0}^T IF_\theta(u,\mathbf{v},G_0,H_0,J)] \,,$$

so

$$IF_{\theta_0}(u,\mathbf{v},G_0,H_0,J,K) = -\mu_{H_0}^T IF_\theta(u,\mathbf{v},G_0,H_0,J) + \frac{K(G_0(u))}{A(K,G_0)} \,. \tag{2.3.11}$$

The joint influence function of θ and θ_0 is

$$\begin{aligned} IF &= \binom{IF_{\theta_0}}{IF_\theta} \\ &= \begin{pmatrix} -\mu_{H_0}^T C_0^{-1}(\mathbf{v} - \mu_{H_0})\frac{J(G_0(u))}{A(J,G_0)} + \frac{K(G_0(u))}{A(K,G_0)} \\ \frac{J(G_0(u))}{A(J,G_0)} C_0^{-1}(\mathbf{v} - \mu_{H_0}) \end{pmatrix} \end{aligned} \tag{2.3.12}$$

where

$$A(J,G_0) = \int J'(G_0(e))g_0^2(e)de$$
$$A(K,G_0) = \int K'(G_0(e))g_0^2(e)de \,.$$

W.l.o.g. we can assume that

$$\mu_{H_0} = 0 \,, \tag{2.3.13}$$

then (2.3.12) is simplified to be

$$IF = \begin{pmatrix} \frac{K(G_0(u))}{A(K,G_0)} \\ \frac{J(G_0(u))}{A(J,G_0)} C_0^{-1}\mathbf{v} \end{pmatrix} \,. \tag{2.3.14}$$

2.4 Optimality criteria in estimation

In this section we will discuss the optimality criteria in linear estimation.

From section 2.1 we know that $\binom{\theta_0}{\theta}$ is asymptotically normal with the covariance matrix

$$V(G_0,H_0,J,K) = \int IF \cdot IF^T dF_0(u,\mathbf{v}) \,. \tag{2.4.1}$$

Using (2.3.14) we have

$$IF \cdot IF^T = \begin{pmatrix} IF^2_{\theta_0} & IF_{\theta_0} \cdot IF^T_\theta \\ IF_\theta \cdot IF^T_{\theta_0} & IF_\theta \cdot IF^T_\theta \end{pmatrix}$$
$$= \begin{pmatrix} \frac{K^2(G_0(u))}{A^2(K,G_0)} & \frac{K(G_0(u))J(G_0(u))}{A(J,G_0)A(K,G_0)} \mathbf{v}^T C_0^{-1} \\ \frac{K(G_0(u))J(G_0(u))}{A(J,G_0)A(K,G_0)} C_0^{-1}\mathbf{v} & \frac{J^2(G_0(u))}{A^2(J,G_0)} C_0^{-1}\mathbf{v}\mathbf{v}^T C_0^{-1} \end{pmatrix},$$

so

$$V(G_0, H_0, J, K) = \begin{pmatrix} \frac{\int_0^1 K^2(x)dx}{A^2(k,G_0)} & \mathbf{0}^T \\ \mathbf{0} & \frac{\int_0^1 J^2(t)dt}{A^2(J,G_0)} C_0^{-1} \end{pmatrix}. \tag{2.4.2}$$

Let $V(G_0, H_0)$ be the group of matrices which contains all the $V(G_0, H_0, J, K)$ for all score functions J and K. For 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. We define V to be the minimum in $V(G_0, H_0)$ if $V < W$ for any $W \in V(G_0, H_0)$. So the optimal score functions J and K are the ones which minimize $V(G_0, H_0, J, K)$ in $V(G_0, H_0)$.

Since C_0 is a positive semidefinite matrix, it is easy to see that the optimal J and K are those which minimize

$$\frac{\int_0^1 J^2(x)dx}{A^2(J,G_0)} \text{ and } \frac{\int_0^1 K^2(x)dx}{A^2(K,G_0)}, \text{ respectively },$$

or equivalently maximize

$$\frac{A^2(J,G_0)}{\int_0^1 J^2(x)dx} \text{ and } \frac{A^2(K,G_0)}{\int_0^1 K^2(x)dx}, \text{ respectively }.$$

We can also consider the optimality problems in the following two ways in which $J(t)$ and $K(t)$ result in the same.

The first one is called Hampel-type problem. It minimizes the variance subject to a bound on influence function, i.e.

$$\text{minimize } \frac{\int_0^1 J^2(x)dx}{A^2(J,G_0)}$$

under the condition

$$\sup_u \left|\frac{J(G_0(u))}{A(J,G_0)}\right| \le b$$

for given b. In the above problems it is assumed that G_0 is known.

The second is the Huber-type problem. The optimal score function will minimize the maximum variance, as the unknown G_0 varies over a given class of distribution functions. We will discuss this later in Chapter 4.

The solutions for all the three kinds of optimality problems will be given in Chapter 4.

CHAPTER 3

HYPOTHESIS TESTING IN THE LINEAR MODEL

3.1. Definition of influence function for tests

In this chapter we will derive the influence functions of tests and then use them. So at first we need to give its definition. The influence function for tests is an extension of that for estimates.

Suppose we have a hypothesis H_1 where

$$H_1 : \theta = \theta_* \, .$$

Consider a sequence of tests for testing H_1 which depend on the observations only through a sequence of test statistics $\{T_n : n \in N\}$. Assume that T_ns satisfy the conditions (1) and (2) in section 2.1. Then we can define the influence function for tests.

Definition 3.1.1:

Let δ_x be the distribution which puts mass 1 at point $x \in X$. Then, the influence function of the test defined by the test statistic T at F_{θ_*} is

$$IF_{\text{test}}(x, T, F_{\theta_*}) = \lim_{\varepsilon \searrow 0} (T((1-\varepsilon)F_{\theta_*} + \varepsilon\delta_x) - T(F_{\theta_*}))/\varepsilon \, . \tag{3.1.1}$$

3.2. Definition and influence function of test statistic

In this section we will define a test statistic with the dispersion measure D_J given by Jaeckel to test hypothesis

$$H_0 : \theta_{q+1} = \ldots = \theta_{p-1} = 0 \, . \tag{3.2.1}$$

Let ω be the subspace of $\Omega \subset \mathbb{R}^{p-1}$ obtained by imposing the condition H_0. Then we can define our test statistic.

Definition 3.2.1.

Let $(T_\Omega)_n$ and $(T_\omega)_n$ be the R-estimators of θ in the model Ω and ω respectively, i.e.

$$D_J(Y - X(T_\Omega)_n) = \min\{D(Y - X\theta) | \theta \in \Omega\}$$

$$D_J(Y - X(T_\omega)_n) = \min\{D(Y - X\theta) | \theta \in \omega\} \, .$$

A test statistic D_n is defined by

$$D_n^2(x_1,\ldots,x_n,y_1,\ldots,y_n) = \frac{2}{n}[D_J(Y - X^T(T_\omega)_n) - D_J(Y - X^T(T_\Omega)_n)] \,. \quad (3.2.2)$$

Comparing (1.1.8) and (3.2.2) we can see some similarity between classical and robust test statistic.

Notation: For any vector $x \in \mathbb{R}^{p-1}$ denote by $\tilde{x}$ the vector $(x_{(1)},\ldots,x_{(q)},0,\ldots,0)^T$.

From subsection 1.2.2 we know that $(T_\omega)_n$ and $(T_\Omega)_n$ fulfill the equations

$$\sum_{i=1}^{n} a(R_i(\tilde{e}_i))(\tilde{x}_i - \bar{\tilde{x}}_i) = 0 \quad (3.2.3)$$

$$\sum_{i=1}^{n} a(R_i(e_i))(x_i - \bar{x}) = 0. \quad (3.2.4)$$

where $\tilde{e}_i = y_i - \tilde{x}_i^T(T_\omega)_n$ and $(T_\omega)_n = (((T_\omega)_n)_{(1)},\ldots,((T_\omega)_n)_{(q)},0,\ldots,0)^T$.

Let D, T_ω , T_Ω be the functionals corresponding to D_n, $(T_\omega)_n$, and $(T_\Omega)_n$. Then from (2.2.3), (2.2.17) and (3.2.2) we have

$$D^2(F) = 2\int J(\ell(e,\tilde{x},F))(y - \tilde{x}^T T_\omega(F))dF - 2\int J(\ell(e,x,F))(y - x^T T_\Omega(F))dF$$

where

$$\ell(e,\tilde{x},F) = \int G(e + (\tilde{x} - \tilde{z})^T(T_\omega(F_\theta) - T_\omega(F)))dH(z)$$

$$\ell(e,x,F) = \int G(e + (x - z)^T(T_\Omega(F_\theta) - T_\Omega(F)))dH(z)$$

and T_ω, T_Ω fulfill

$$\int J(\ell(e,\tilde{x},F))(\tilde{x} - \tilde{\mu}_H)dF(x,e) = 0 \quad (3.2.6)$$

$$\int J(\ell(e,x,F))(x - \mu_H)dF(x,e) = 0 \,. \quad (3.2.7)$$

When we discuss Hampel-type optimality problem in testing in Chapter 4 we will try to find a kind of score function which maximizes the asymptotic power, subject to a bound on the influence function of the test statistic. So we need to calculate the influence functions of T_ω, T_Ω and D under H_0.

Theorem 3.2.1. *The influence functions of the functionals T_ω, T_Ω, and D under H_0 are as follows:*

(1) $$IF(u, \mathrm{v}, T_\omega, F_{\tilde{\theta}}, J) = \frac{J(G(u))}{A(J,G)} C^{-1}(\tilde{\mathrm{v}} - \tilde{\mu}_H) \tag{3.2.8}$$

(2) $$IF(u, \mathrm{v}, T_\Omega, F_{\tilde{\theta}}, J) = \frac{J(G(u))}{A(J,G)} C_0^{-1}(\mathrm{v} - \mu_H) \tag{3.2.9}$$

(3) $$IF(u, \mathrm{v}, D, F_{\tilde{\theta}}, J) = \frac{J(G(u))}{\sqrt{A(J,G)}}[(\mathrm{v} - \mu_H)^T(C_0^{-1} - \tilde{C}_0^{-1})(\mathrm{v} - \mu_H)]^{1/2} \tag{3.2.10}$$

where

$$C_0 = \int (x - \mu_H)(x - \mu_H)^T dH(x)$$
$$\tilde{C}_0 = \int (\tilde{x} - \tilde{\mu}_H)(\tilde{x} - \tilde{\mu}_H)^T dH(x)$$

and

$$\tilde{C}_0^{-1} = \begin{pmatrix} (\tilde{C}_0)_{11}^{-1} & 0 \\ 0 & 0 \end{pmatrix}$$

where $(\tilde{C}_0)_{11}$ is the upper-left corner of $\tilde{C}_0$, a $q \times q$ matrix and $A(J,G) = -\int J(G(u))g'(u)du$ under (3.2.12).

Notation: Under the null hypothesis $\theta = \tilde{\theta} = (\theta_1, \ldots, \theta_q, 0, \ldots, 0)^T$. $F_{\tilde{\theta}}$ is the distribution function under the null hypothesis.

In this theorem we assume that J'' exist and $J'(t)$ exist and is continuous on $(0,1)$ and that

$$\lim_{e \to \infty} J'(G(e))g^2(e)e = 0 \tag{3.2.11}$$

$$\lim_{e \to \infty} J(G(e))g(e) = 0\,. \tag{3.2.12}$$

Proof: We have for the model

$$y = \theta_{0*} + x^T\theta_* + \varepsilon\,.$$

Then from the definition of D and (3.2.6), (3.2.7) we get

$$
\begin{aligned}
D^2(F) &= 2\int J(\ell(e,\tilde{x},F))(\tilde{x}^T(T_\omega(F_{\tilde{\theta}}) - T_\omega(F)) + \varepsilon)dF \\
&\quad - 2\int J(\ell(e,x,F))(x^T(T_\omega(F_{\tilde{\theta}}) - T_\omega(F)) + \varepsilon)dF \\
&= 2\int J(\ell(e,\tilde{x},F))(\tilde{\mu}_H^T(T_\omega(F_{\tilde{\theta}}) - T_\omega(F)) + \varepsilon)dF \\
&\quad - 2\int J(\ell(e,x,F))(\mu_H^T(T_\omega(F_{\tilde{\theta}}) - T_\omega(F)) + \varepsilon)dF\,. \qquad (3.2.13)
\end{aligned}
$$

With the same method in Section 2.2 we can get

$$
IF((u,\mathrm{v},T_\omega,F_{\tilde{\theta}},J) = \frac{J(G(u))}{A(J,G)}\tilde{C}_0^{-1}(\tilde{\mathrm{v}} - \tilde{\mu}_H)
$$

$$
IF((u,\mathrm{v},T_\omega,F_{\tilde{\theta}},J) = \frac{J(G(u))}{A(J,G)}C_0^{-1}(\mathrm{v} - \mu_H)
$$

which are the results in (1) and (2).

Now we derive the influence function of D at the null hypothesis.

Let $F_\lambda = (1-\lambda)F_{\tilde{\theta}} + \lambda F_1$ and

$$
\dot{T}_\omega(F_\lambda) = \tfrac{dT_\omega(F_\lambda)}{d\lambda}\,, \qquad \ddot{T}_\omega(F_\lambda) = \tfrac{d^2T_\omega(F_\lambda)}{d\lambda^2}
$$

$$
\dot{T}_\Omega(F_\lambda) = \tfrac{dT_\Omega(F_\lambda)}{d\lambda}\,, \qquad \ddot{T}_\Omega(F_\lambda) = \tfrac{d^2T_\Omega(F_\lambda)}{d\lambda^2}
$$

Then we have

$$
\begin{aligned}
\frac{dD^2(F_\lambda)}{d\lambda} &= 2\int J'(\ell(e,\tilde{x},F_\lambda))\frac{d\ell(e,\tilde{x},F_\lambda)}{d\lambda}[\mu_{H_\lambda}^T(T_\omega(F_{\tilde{\theta}}) - T_\omega(F_\lambda)) + \varepsilon]dF_\lambda \\
&\quad - 2\int J(\ell(e,\tilde{x},F_\lambda))\mu_{H_\lambda}^T\dot{T}_\omega(F_\lambda)dF_\lambda \\
&\quad + 2\int J(\ell(e,\tilde{x},F))[\mu_{H_\lambda}^T(T_\omega(F_{\tilde{\theta}}) - T_\omega(F)) + \varepsilon]d(F_1 - F_{\tilde{\theta}}) \\
&\quad + 2\int J(\ell(e,\tilde{x},F))[(\mu_{H_1} - \mu_H)^T(T_\omega(F_{\tilde{\theta}}) - T_\omega(F_\lambda))]dF_\lambda \\
&\quad - 2\int J'(\ell(e,x,F_\lambda))\frac{d(\ell(e,x,F_\lambda)}{d\lambda}[\mu_{H_\lambda}^T(T_\omega(F_{\tilde{\theta}}) - T_\omega(F_\lambda)) + \varepsilon]dF_\lambda \\
&\quad + 2\int J(\ell(e,x,F_\lambda))\mu_{H_\lambda}^T\dot{T}_\omega(F_\lambda)dF_\lambda \\
&\quad - 2\int J(\ell(e,x,F_\lambda))[\mu_{H_\lambda}^T(T_\omega(F_{\tilde{\theta}}) - T_\omega(F)) + \varepsilon]d(F_1 - F_{\tilde{\theta}}) \\
&\quad - 2\int J(\ell(e,x,F_\lambda))[(\mu_{H_1} - \mu_H)^T(T_\omega(F_{\tilde{\theta}}) - T_\omega(F_\lambda))]dF_\lambda\,, \qquad (3.2.14)
\end{aligned}
$$

$$
\frac{dD^2(F_\lambda)}{d\lambda}|_{\lambda=0} = 0\,,
$$

and

$$\frac{d^2D^2(F_\lambda)}{d\lambda^2}|_{\lambda=0} = 2\int J''(G(e))[(\frac{d\ell(e,\tilde{x},F_\lambda)}{d\lambda})^2|_{\lambda=0} - (\frac{d\ell(e,x,F_\lambda)}{d\lambda})^2|_{\lambda=0}]edF_{\bar{\theta}}$$
$$+2\int J'(G(e))[\frac{d^2\ell(e,\tilde{x},F_\lambda)}{d\lambda^2}|_{\lambda=0} - \frac{d^2\ell(e,x,F_\lambda)}{d\lambda^2}|_{\lambda=0}]edF_{\bar{\theta}}$$
$$-4\int J'(G(e))[\frac{d\ell(e,\tilde{x},F_\lambda)}{d\lambda}|_{\lambda=0}\mu_H^T\dot{T}_\omega(F_{\bar{\theta}})$$
$$-\frac{d\ell(e,x,F_\lambda)}{d\lambda}|_{\lambda=0}\mu_H^T\dot{T}_\Omega(F_{\bar{\theta}})]edF_{\bar{\theta}}$$
$$+4\int J'(G(e))[\frac{d\ell(e,\tilde{x},F_\lambda)}{d\lambda}|_{\lambda=0} - \frac{d\ell(e,x,F_\lambda)}{d\lambda}|_{\lambda=0}]ed(F_1 - F_{\bar{\theta}})$$
$$-4\int J(G(e))\mu_H^T[\dot{T}_\omega(F_{\bar{\theta}}) - \dot{T}_\Omega(F_{\bar{\theta}})]d(F_1 - F_{\bar{\theta}})\,. \qquad (3.2.15)$$

We know

$$\ell(e,x,F_\lambda) = (1-\lambda)\int G(e+(x-z)^T(T_\lambda(F_{\bar{\theta}}) - T_\Omega(F_\lambda)))dH(z)$$
$$+\lambda\int G_1(e+(x-z)^T(T_\lambda(F_{\bar{\theta}}) - T_\Omega(F_\lambda)))dH_1(z)\,.$$

Then by calculation we get

$$\frac{d\ell(e,x,F_\lambda)}{d\lambda}|_{\lambda=0} = (G_1 - G)(e) - g(e)(x-\mu_H)^T\dot{T}_\Omega(F_{\bar{\theta}}) \qquad (3.2.16)$$

$$\frac{d^2\ell(e,x,F_\lambda)}{d\lambda^2}|_{\lambda=0} = 2g(e)(x-\mu_H)^T\dot{T}_\Omega(F_{\bar{\theta}}) + g'(e)\int((x-z)^T\dot{T}_\Omega(F_{\bar{\theta}}))^2dH(z)$$
$$-g(e)(x-\mu_H)^T\ddot{T}_\Omega(F_{\bar{\theta}})$$
$$-2g_1(e)(x-\mu_{H_1})^T\dot{T}_\Omega(F_{\bar{\theta}})\,. \qquad (3.2.17)$$

Similarly we have

$$\frac{d\ell(e,\tilde{x},F_\lambda)}{d\lambda}|_{\lambda=0} = (G_1 - G)(e) - g(e)(\tilde{x}-\tilde{\mu}_H)^T\dot{T}_\omega(F_{\bar{\theta}}) \qquad (3.2.18)$$

and

$$\frac{d^2\ell(e,\tilde{x},F_\lambda)}{d\lambda^2}|_{\lambda=0} = 2g(e)(\tilde{x}-\tilde{\mu}_H)^T\dot{T}_\omega(F_{\bar{\theta}}) + g'(e)\int((\tilde{x}-\tilde{z})^T\dot{T}_\omega(F_{\bar{\theta}}))^2dH(z)$$
$$-g(e)(\tilde{x}-\tilde{\mu}_H)^T\ddot{T}_\omega(F_{\bar{\theta}})$$
$$-2g_1(e)(\tilde{x}-\tilde{\mu}_{H_1})^T\dot{T}_\omega(F_{\bar{\theta}})\,. \qquad (3.2.19)$$

Taking (3.2.11), (3.2.12) (3.2.16) - (3.2.19) into (3.2.15) gives

$$\frac{d^2D(F_\lambda)}{d\lambda^2}|_{\lambda=0} = 2[J'(G(e))g^2(e)|_{-\infty}^{+\infty} - \int J'(G(e))g^2(e)de]\cdot$$
$$\int [((\tilde{x}-\tilde{\mu}_H)^T\dot{T}_\omega(F_{\tilde{\theta}}))^2 - ((x-\mu_H)^T\dot{T}_\Omega(F_{\tilde{\theta}}))^2]dF_{\tilde{\theta}}(x)$$
$$= 2A(J,G)\cdot\int [((x-\mu_H)^T\dot{T}_\Omega(F_{\tilde{\theta}}))^2 - ((\tilde{x}-\tilde{\mu}_H)^T\dot{T}_\Omega(F_{\tilde{\theta}}))^2]dH(x)\,.$$

By L'Hopital's rule, we get

$$\frac{dD(F_\lambda)}{d\lambda}|_{\lambda=0} = \lim_{\lambda\to 0}\frac{D(F_\lambda)-D(F_{\tilde{\theta}})}{\lambda}$$
$$= (\lim_{\lambda\to 0}\frac{D^2(F_\lambda)}{\lambda^2})^{1/2}$$
$$= [\frac{1}{2}\frac{d^2D^2(F_\lambda)}{d\lambda^2}|_{\lambda=0}]^{1/2}$$
$$= [A(J,G)\cdot\int [(x-\mu_H)^T\dot{T}_\Omega(F_{\tilde{\theta}}))^2 - ((\tilde{x}-\tilde{\mu}_H)^T\dot{T}_\omega(F_{\tilde{\theta}}))^2]dH(x)]^{1/2}\,.$$

So

$$\begin{aligned} IF(u,\mathrm{v},D,F_{\tilde{\theta}},J) &= [A(J,G)\int [(x-\mu_H)^T IF(u,\mathrm{v},T_\Omega,F_{\tilde{\theta}},J))^2 \\ &\quad - ((\tilde{x}-\tilde{\mu}_H)^T\cdot IF(u,\mathrm{v},T_\omega,F_{\tilde{\theta}},J))^2]dH(x)]^{1/2} \\ &= [\frac{J^2(G(x))}{A(J,G)}\int [((x-\mu_H)^T C_0^{-1}(\mathrm{v}-\mu_H))^2 \\ &\quad - ((\tilde{x}-\tilde{\mu}_H)^T\tilde{C}_0^{-1}(\tilde{\mathrm{v}}-\tilde{\mu}_H))^2]dH(x)]^{1/2} \\ &= \frac{|J(G(u))|}{\sqrt{A(J,G)}}[(\mathrm{v}-\mu_H)^T(C_0^{-1}-\tilde{C}_0^{-1})(\mathrm{v}-\mu_H)]^{1/2}\,. \end{aligned} \tag{3.2.20}$$

This completes the proof.

3.3. Asymptotic distribution of the test statistic

Let $F_\lambda = (1-\lambda)F_{\tilde{\theta}} + \lambda F_1$. We first perform an expansion of the functional D^2. Following von Mises (1947), we get

$$D^2(F_1) = D^2(F_{\tilde{\theta}}) + \frac{dD^2(F_\lambda)}{d\lambda}|_{\lambda=0} + \frac{1}{2}\frac{d^2D^2(F_\lambda)}{d\lambda^2}|_{\lambda=0} + \text{ remainder}\,. \tag{3.3.1}$$

With the same calculation as in the proof of Theorem 3.2.1 we obtain

$$
\begin{aligned}
D^2(F_\lambda)|_{\lambda=0} &= 0 \\
\frac{dD^2(F_\lambda)}{d\lambda}|_{\lambda=0} &= 0 \\
\frac{d^2D^2(F_\lambda)}{d\lambda^2}|_{\lambda=0} &= 2\int J'G(e))g^2(e)de\int[((x-\mu_H)^T\dot{T}_\Omega(F_{\tilde{\theta}}))^2 \\
&\quad -((\tilde{x}-\tilde{\mu}_H)^T\cdot\dot{T}_\omega(F_{\tilde{\theta}}))^2]dH \\
&= 2A(J,G)[\frac{(\int J(G(e))dG_1(e))^2}{A^2(J,G)}(\mu_{H_1}-\mu_H)^T(C_0^{-1}-\tilde{C}_0^{-1})(\mu_{H_1}-\mu_H)] \\
&= \frac{2}{A(J,G)}[\int J(G(e))(x-\mu_H)^T dF_1](C_0^{-1}-\tilde{C}_0^{-1})[\int J(G(e))(x-\mu_H)dF_1]
\end{aligned}
\tag{3.3.2}
$$

where the second equality comes from (2.2.25). So we get

$$
D^2(F_1) = \frac{1}{A(J,G)}[\int J(G(e))(x-\mu_H)^T dF_1](C_0^{-1}-\tilde{C}_0^{-1})[\int J(G(e))(x-\mu_H)dF_1] + \text{remainder}\,. \tag{3.3.3}
$$

Put $F_1 = F_n =$ empirical distribution functions. We get (under H_0)

$$
nD_n^2 = nD^2(F_n) = \frac{1}{A(J,G)}[V_n^T(\tilde{\theta})(C_0^{-1}-\tilde{C}_0^{-1})V(\tilde{\theta})] + \text{remainder} \tag{3.3.4}
$$

where $V_n(\theta) = \frac{1}{\sqrt{n}}\sum_{i=1}^n a(R_i(e))(x_i-\overline{x})$ and $e_i = y_i - x_i^T\theta$. The above derivation is not meant to be rigorous. We tried to find the main part of nD_n^2 and will show the remainder is asymptotically negligible, i.e. the main part of nD_n^2 has the same asymptotic distribution as nD_n^2.

Next theorem given by Hettmansperger and McKean (1976) shows that

$$
nA(J,G)D_n^2/\int_0^1 J^2(t)dt
$$

and

$$
V_n^T(\tilde{\theta})(C_0^{-1}-\tilde{C}_0^{-1})V_n(\tilde{\theta})/\int_0^1 J^2(t)dt
$$

have the same asymptotic distribution χ^2_{p-q-1} under H_0.

Theorem 3.3.1. *Under the null hypothesis*

$$nA(J,G)D_n^2/\int_0^1 J^2(t)dt$$

and

$$V_n^T(\tilde{\theta})(C_0^{-1}-\tilde{C}_0^{-1})V_n(\tilde{\theta})/\int_0^1 J^2(t)dt$$

are asymptotically distributed as χ^2_{p-q-1}.

Further also from Hettmansperger and McKean (1976) we have

Theorem 3.3.2. *Under the sequence of alternatives*

$$H_1:\theta_j=n^{-\frac{1}{2}}\Delta_j \quad j=q+1,\ldots,p-1$$

where $\Delta=(\Delta_1,\ldots,\Delta_{p-1})^T$ *the test statistic* $nA(J,G)D_n^2/\int_0^1 J^2(t)dt$ *has the asymptotic distribution* $\chi^2_{p-q-1}(\eta^2)$ *with noncentrality paramter*

$$\eta^2=\frac{A^2(J,G)}{2\int_0^1 J^2(t)dt}(\Delta_{q+1}\ldots\Delta_{p-1})^T C_{22,1}\begin{pmatrix}\Delta_{q+1}\\ \vdots\\ \Delta_{p-1}\end{pmatrix}$$

where $C_{22,1}$ *comes from* C_0*:*

$$C_{22,1}=C_{22}-C_{12}^T C_{11}^{-1}C_{12}\,.$$

3.4 Optimality criteria in testing

The optimality criteria for a robust test can be stated as follows:

Consider a class $\mathbb{D}$ of tests depending on the observations only through a test statistic $D_n(x_1,\ldots,x_n,y_1,\ldots,y_n)$. Then we can try to find the optimal score function which maximizes the asymptotic power within $\mathbb{D}$. If we consider the maximizing problem under the side condition of a bound on the influence function of the test statistic at the null hypothesis, it is the Hampel-type optimality problem. We can also consider the problem of maximizing the minimum power, as G_0 varies over a specified class. That is the Huber-type optimality problem.

Let the test statistic E be

$$E^2 = A(J,G)D^2 / \int_0^1 J^2(t)dt \,. \tag{3.4.1}$$

Then we have from section 3.3 that E^2 has the asymptotic distribution $\chi^2_{p-q-1}(\eta^2)$ with

$$\eta^2 = \frac{A^2(J,G)}{2\int_0^1 J^2(t)dt}(\Delta_{q+1}\dots\Delta_{p-1})C_{22,1}\begin{pmatrix}\Delta_{q+1}\\ \vdots \\ \Delta_{p-1}\end{pmatrix} \tag{3.4.2}$$

under the sequence of alternatives

$$H_1 : \theta_j = n^{-\frac{1}{2}}\Delta_j \quad j = q+1,\dots,p-1 \,.$$

Also from Theorem 3.2.1 we have

$$IF(u,\mathrm{v},E,F_{\tilde{\theta}},J) = \frac{|J(G(u))|}{\sqrt{\int_0^1 J^2(t)dt}}[(\mathrm{v}-\mu_H)^T(C_0^{-1}-\tilde{C}_0^{-1})(\mathrm{v}-\mu_H)]^{1/2} \,. \tag{3.4.3}$$

We can write IF as a product of two factors, namely the influence of residual (IR) and the influence of position (IP):

$$IF(u,\mathrm{v},E,F_{\tilde{\theta}},J) = IR(u,E,F_{\tilde{\theta}},J)IP(\mathrm{v},E,F_{\tilde{\theta}}) \tag{3.4.4}$$

where

$$IR(u,E,F_{\tilde{\theta}},J) = |J(G(u))|/(\int_0^1 J^2 dt)^{1/2}$$
$$IP(\mathrm{v},E,F_{\tilde{\theta}}) = [(\mathrm{v}-\mu_H)^T(C_0^{-1}-\tilde{C}_0^{-1})(\mathrm{v}-\mu_H)]^{1/2} \,.$$

Since the asymptotic power is a monotone increasing function of the noncentrality parameter η^2, the optimal problems become ones of maximizing η^2.

Generally we can consider

$$\text{maximize}\frac{A^2(J,G)}{\int_0^1 J^2 dt} \,. \tag{3.4.5}$$

In Hampel-type problem we try to find J which

$$\text{maximize}\frac{A^2(J,G)}{\int_0^1 J^2 dt} \tag{3.4.5}$$

under the condition

$$\sup_u \frac{|J(G(u))|}{\sqrt{\int_0^1 J^2(t)dt}} \leq b\,. \tag{3.4.6}$$

In Huber-type problem which will be discussed in Chapter 4 we want to find J which maximizes the minimum η^2 as the unknown G_0 varies over a given class of distribution functions.

CHAPTER 4

OPTIMAL SCORE FUNCTIONS

In this chapter we will give solutions to all the optimality problems mentioned above. First we will consider the general case, then Hampel-type problem in estimation, Hampel-type problem in testing, and at least Huber-type problem in both estimation and testing.

4.1. Optimal score function in general

In this case the optimality problem is to find J which

$$\text{maximize}\frac{A^2(J,G)}{\int_0^1 J^2 dt}\,. \tag{4.1.1}$$

We can standardize J so that $\int_0^1 J^2(t)dt = 1$. Then the optimality problem becomes

$$\text{maximize}(\int J(G(u))g'(u)du)^2 \tag{4.1.2}$$

under the condition

$$\int_0^1 J^2(t)dt = 1\,. \tag{4.1.3}$$

Throughout this chapter we assume that g satisfies the following condition

$$-\frac{g'(u)}{g(u)} \nearrow \text{ in } u > 0\,. \tag{4.1.4}$$

By the Cauchy-Schwarz inequality we can easily find out that

$$J(t) = \psi(G^{-1}(t))/\sqrt{I(G)} \tag{4.1.5}$$

where $\psi = -\frac{g'}{g}$ and $I(G) = \int(\frac{g'}{g})^2 dG$ is the optimal score function in this problem.

4.2. Optimal score function in Hampel-type problem in estimation

In Hample-type problem in estimation we consider the problem

$$\text{minimize}\frac{\int_0^1 J^2(t)dt}{A^2(J,G)} \tag{4.2.1}$$

under the condition

$$\sup_u |\frac{J(G(u))}{A(J,G)}| \leq b \tag{4.2.2}$$

for given b. We standardize $J(t)$ so that $A(J,G) = -\int J(G(u))\frac{g'}{g}dG = \frac{1}{2}I(G) = \int(\frac{g'}{g})^2 dG$. Then the optimality problem is to

$$\text{minimize } \int_0^1 J^2(t)dt \tag{4.2.3}$$

under the condition

$$\sup_u |J(G(u))| \leq b \tag{4.2.4}$$

and

$$A(J,G) = -\int J(G)g'du = \frac{1}{2}I(G)\,. \tag{4.2.5}$$

We claim that

$$J_E(t) = \begin{cases} b & \text{if } -c\frac{g'}{g}(G^{-1}(t)) > b \\ -c\frac{g'}{g}(G^{-1}(t)) & \text{if } |c\frac{g'}{g}(G^{-1}(t))| \leq b \\ -b & \text{if } -c\frac{g'}{g}(G^{-1}(t)) < -b \end{cases} \tag{4.2.6}$$

is the optimal score function in this problem where c is selected to satisfy condition (4.2.5).

Proof: For any J satisfying conditions (4.2.4), (4.2.5) we have

$$\begin{aligned} \int (J(G) + c\frac{g'}{g})^2 &= \int_0^1 J^2 dt + 2c\int J(G)g'du + c^2\int(\frac{g'}{g})^2 dG \\ &= \int_0^1 J^2 dt + (c^2 - c)I(G)\,. \end{aligned} \tag{4.2.7}$$

So minimizing $\int_0^1 J^2 dt$ is equivalent to minimizing $\int(J(G) + c\frac{g'}{g})^2 dG$. We can write

$$\begin{aligned} \int (J(G) + c\frac{g'}{g})^2 DG &= \int (J(G) - (-c\frac{g'}{g}))^2 dG \\ &= \int_{\{-c\frac{g'}{g} > b\}} (J(G) - (-c\frac{g'}{g}))^2 dG + \int_{\{-c\frac{g'}{g} < -b\}} (J(G) - (-c\frac{g'}{g}))^2 dG \\ &\quad + \int_{\{|c\frac{g'}{g}| \leq b\}} (J(G) - (-c\frac{g'}{g}))^2 dG\,. \end{aligned} \tag{4.2.8}$$

It can be seen that $J_E(t)$ in (4.2.6) minimizes each of the three parts at the RHS of (4.2.8). So it is the optimal score function.

4.3. Optimal score function in Hampel-type problem in testing

In this Hampel-type problem the optimal score function J will

$$\text{maximize}\frac{A^2(J,G)}{\int_0^1 J^2 dt} \tag{4.3.1}$$

under the condition

$$\sup_u \frac{|J(G(u))|}{\sqrt{\int_0^1 J^2(t)dt}} \leq b\,. \tag{4.3.2}$$

We can standardize J so that $\int_0^1 J^2(t)dt = 1$. Then the optimality problem becomes

$$\text{maximize}(-\int J(G(u))g'(u)du)^2 \tag{4.3.3}$$

under the condition

$$\sup_u |J(G(u))| \leq b \tag{4.3.4}$$

and

$$\int_0^1 J^2(t)dt = 1\,. \tag{4.3.5}$$

Since $1 = \int_0^1 J^2(t)dt \leq \int_0^1 (\sup|J(t)|)^2 dt \leq b$ we should select b properly.

Next theorem will give us the solution.

Theorem 4.1.1. *The score function*

$$J_0(t) = \begin{cases} b & t > G_1(\psi^{-1}(\frac{b\sqrt{I(G)}}{c})) \\ c\psi(G^{-1}(t))/\sqrt{I(G)} & \frac{1}{2} \leq t \leq G(\psi^{-1}(\frac{b\sqrt{I(G)}}{c})) \\ -c\psi(G^{-1}(t))/\sqrt{I(G)} & -G(\psi^{-1}(\frac{b\sqrt{I(G)}}{c})) \leq t \leq \frac{1}{2} \\ -b & t < G(\psi^{-1}((\frac{b\sqrt{I(G)}}{c})) \end{cases} \tag{4.3.6}$$

where c is determined by the equation

$$\frac{1}{I(G)}\int_0^{\psi^{-1}(\frac{b\sqrt{I(G)}}{c})} c^2(\frac{g'}{g})^2(u)dG(u) + b^2(1 - G(\psi^{-1}(\frac{b\sqrt{I(G)}}{c}))) = \frac{1}{2} \tag{4.3.7}$$

is the optimal score function in the Hampel-type problem, given by (4.3.3) - (4.3.5) above.

Proof: $J_0(t)$ can also be written as

$$J_0(G(u)) = \begin{cases} b & u > K_0 \\ c\psi(u)/\sqrt{I(G)} & 0 \le u \le K_0 \\ -c\psi(u)/\sqrt{I(G)} & -K_0 \le u \le 0 \\ -b & u < -K_0 \end{cases} \tag{4.3.8}$$

where $K_0 = \psi^{-1}(\frac{b\sqrt{I(G)}}{c})$ and c is determined by

$$\frac{I}{I(G)} \int_0^{K_0} c^2\psi^2(u)dG(u) + b^2(1 - G(K_0)) = \frac{1}{2}\,. \tag{4.3.9}$$

Since functions J and g are symmetric we only need to discuss the positive parts of these functions.

We get (4.3.8) and (4.3.9) in this way. First we predict that the optimal score function is in the form

$$J(G(u)) = \begin{cases} b & u > K_0 \\ c\psi(u)/\sqrt{I(G)} & 0 \le u \le K_0 \end{cases}$$

for some K_0. We require $J(G(u))$ be continuous and satisfy $\int_0^1 J^2(t)dt = 1$ which give $K_0 = \psi^{-1}(\frac{b\sqrt{I(G)}}{C})$ and (4.3.9). It is easy to see that when $c \to +\infty$ the LHS of (4.3.9) $\to b \ge 1$, and that when $c \to 1$, then LHS of (4.3.9) $\le \frac{1}{I(G)} \int_0^{+\infty} (\frac{g'}{g})^2 dG = \frac{1}{2}$. Thus, we can find c from (4.3.9). It is a trivial case when $b = 1$, then $J(t) \equiv 1$ because of $\int_0^1 J^2(t)dt = 1$. Then

$$-\int_0^{+\infty} J_0(G(u))g'(u)du = -\int_0^{K_0} J_0(G(u))g'du + bg(K_0) \tag{4.3.10}$$

and (4.3.9) is

$$\int_0^{K_0} J_0^2(G(u))dG(u) + b^2(1 - G(K_0)) = \frac{1}{2}\,. \tag{4.3.11}$$

Now we try to prove the optimality of J_0.

Suppose we have a $J(t)$ which reaches b at $G(K_1)$, i.e. $J(G(K_1)) = b$. First we consider the case $K_0 < K_1$. We can write J in the form of

$$J(G(u)) = \begin{cases} J_0(G(u)) + A(u) & u \in D_1 \subset (0, K_0) \\ J_0(G(u)) - B(u) & u \in D_2 \subset (0, K_0) \\ J_0(G(u)) & u \in D_3 \subset (0, K_0) \\ b - C(u) & u \in (K_0, K_1) \\ b & u \geq K_1 \end{cases} \tag{4.3.12}$$

where A, B, C are positive functions, D_1, D_2, D_3 are mutually exclusive Lebesgue measurable sets and $D_1 \cup D_2 \cup D_3 = (0, K_0)$. Then

$$\begin{aligned} -\int J(G(u))g'du = &-\int_{D_1} J_0(G(u))gdu - \int_{D_1} A(u)g'du - \int_{D_2} J_0(G(u))g'du \\ &+ \int_{D_2} B(u)g'du - \int_{D_3} J_2(G(u))g'du - b(g(K_1) - g(K_0)) \\ &+ \int_{K_0}^{K_1} C(u)g'du - b(-g(K_1)) \\ = &-\int_0^{K_1} J_0(G(u))g'(u)du + bg(K_0) - \int_{D_1} A(u)g'du \\ &\int_{D_2} B(u)g'du + \int_{K_0}^{K_1} C(u)g'du\,. \end{aligned} \tag{4.3.13}$$

From $\int_0^{+\infty} J^2(G(u))dG(u) = \frac{1}{2}$ we get

$$\begin{aligned} &\int_{D_1} (J_0^2(u) + 2AJ_0(G) + A^2)dG + \int_{D_2} (J_0^2(G) - 2BJ_0(G) + B^2)dG + \int_{D_3} J_0^2(G)dG \\ &\qquad \int_{K_0}^{K_1} (b^2 - 2bC + C^2)dG + b^2(1 - G(K)) = \frac{1}{2}\,. \end{aligned}$$

Using (4.3.11) to simplify we have

$$\begin{aligned} &\int_{D_2} 2BJ_0(G)dG + \int_{K_0}^{K_1} 2bCda - 2\int_{D_1} AJ_0(G)dG \\ &\qquad = \int_{D_1} A^2dG + \int_{D_2} B^2dG + \int_{K_0}^{K_1} C^2dG \geq 0\,. \end{aligned} \tag{4.3.14}$$

Since J_0 is a non-decreasing function and $J_0(G(K_0)) = b$ so we have

$$b \leq J_0(G(u)) \text{ for } K_0 \leq u \leq K_2\,. \tag{4.3.15}$$

Take $J_0(G(u)) = -c\frac{g'}{g}(u)$ for $0 \le u \le K_0$ into (4.3.14) and use (4.3.15) we get

$$\int_{D_2} Bg'du + \int_{K_0}^{K_1} Cg'du - \int_{D_1} Ag'du \le 0\,. \tag{4.3.16}$$

Taking (4.3.16) into (4.3.13) gives

$$0 \le (-\int_0^{+\infty} J(G(u))g'du) \le -\int_0^{+\infty} J_0(G(u))g'du\,. \tag{4.3.17}$$

Similarly we can prove that when $K_0 \ge K_1$ we still can get (4.3.17). Combining both cases gives us the result.

With the same method used in the Hampel-type problem in estimation we can prove the same result as above.

4.4. Optimal score function in Huber-type problem

In this section we will give solutions to Huber-type problem, i.e. minimax problem for estimation and maximin problem for testing.

In estimation we know that under some conditions we have

$$\sqrt{n}(\widehat{\theta}_n - \theta_*) \sim N(0, V_1(G,J)) \tag{4.4.1}$$

where

$$V_1(G,J) = \frac{\int_0^1 J^2 dt}{A^2(J,G)} C_0^{-1}$$
$$A(J,G) = -\int J(G(u))g'(u)du\,. \tag{4.4.2}$$

We want to find the minimax J, i.e. that which minimizes the maximum variance $V_1(G,J)$ in estimation, or equivalently to find the maximin J that which maximizes the minimum power in testing.

Let $\mathbb{P}$ be a set of distribution functions . Then the minimax problem for estimation, or maximin problem for testing is to find J_* and $G_* \in \mathbb{P}$ such that

$$\min_J \max_{G \in \mathbb{P}} \frac{\int_0^1 J^2 dt}{A^2(J,G)} = \frac{\int_0^1 J_*^2(t)dt}{(\int J_*(G_*(u))g_*'(u)du)^2}\,. \tag{4.4.3}$$

With the minimax variance theory developed by Wiens we can solve this problem.

Suppose $\mathbb{P}$ is a group of distribution functions with location parameter and Fisher informationfor location $I(G)$ is minimized at a member G_0, symmetric about 0. Assume that $0 < I(G_0) < +\infty$, so $G_0(u)$ has an absolutely continuous density $g_0(u)$, tending to zero as $u \to \pm\infty$ (Huber, 1981). Assume also that $g_0(u) > 0$ on $0 < G_0(u) < 1$. Put $\psi_0(u) = -\frac{g_0'}{g_0}(u)$ and assume that $\psi_0(u)$ is non-decreasing, absolutely continuous, with a piecewise continuous derivative $\psi_0'(u)$.

Choose the score function

$$J_0(t) = \psi_0(G_0^{-1}(t))/I(G_0) \tag{4.4.4}$$

and let

$$V(G,J) = \frac{\int_0^1 J^2 dt}{A^2(J,G)} \, . \tag{4.4.5}$$

By the result in section 4.1 we get

$$V(G_0, J_0) \leq V(G_0, J) \text{ for all } J \, . \tag{4.4.6}$$

If we also have

$$\sup_{G\in\mathbb{P}} V(G, J_0) \leq V(G_0, J_0) = \frac{1}{I(G_0)} \tag{4.4.7}$$

then J_0 gives the minimax estimator of J. We give a condition for later use:

$$\int_0^1 H_0(u) dG^{-1}(u) \leq 1 \text{ for all } G \in \mathbb{P} \text{ with } I(G) < \infty \, , \tag{4.4.8}$$

where $H_0(u) = \psi_0'(G_0^{-1}(u)) g_0(G_0^{-1}(u))/I(G_0)$ for $0 < u < 1$.

The following theorems given by Wiens (1990) will give us the optimal minimax estimator of J.

Theorem 4.4.1. *Assume that R-estimator $\widehat{\theta}_n$ satisfies*

$$\sqrt{n}(\widehat{\theta}_n - \theta_*) \to N(0, V_1(G, J)) \tag{4.4.9}$$

with $V_1(G,J)$ as in (4.4.1) for $G \in \mathbb{P}$. In order that (4.4.7) hold, it is sufficient that $\psi_0(u)$ be non-decreasing and that condition (4.4.8) hold.

If $\mathbb{P}$ is any of the following three contamination classes of distribution functions, the next theorem will give us the solution.

(1) ε-contamination neighborhood:

$$C_\varepsilon(F) = \{G|\ G = (1-\varepsilon)F + \varepsilon H; F \text{ symmetric and fixed }, H \text{ symmetric}\}\ .$$

(2) Kolmogorov neighborhood

$$K_\varepsilon(F) = \{G|\ \sup_u |G(u) - F(u)| \le \varepsilon, F \text{ symmetric and fixed}\}\ .$$

(3) Levy neighborhood

$$L_{\varepsilon,\delta}(F) = \{G|F(u-\delta)-\varepsilon \le G(u) \le F(u+\delta)+\varepsilon \text{ for all } u, F \text{ symmetric and fixed }\}\ .$$

Theorem 4.4.2. *Suppose that F is strictly increasing on $(-\infty,\infty)$, with $I(F) < \infty$ and $-f'/f$ twice continuously differentiable. Then (4.4.7) holds for all above three neighborhoods iff G_0 is strongly unimodal.*

If $F = \Phi$ in all neighborhoods above from the work by Jaeckel (1971), Collins (1983), and Collins and Wiens (1989) we know that G_0 is strongly unimodal. Then we can select J_* and G_* in (4.4.3) as

$$G_*(u) = G_0(u) \tag{4.4.10}$$

and

$$J_*(t) = \psi_0(G_*^{-1}(t))/I(G_*)\ . \tag{4.4.11}$$

CHAPTER 5
NUMERICAL STUDY

In the previous chapters we have discussed the estimation and testing problems with R -estimation method in the linear model. In this chapter we will show some numerical results, using simulated data.

Hettmansperger and McKean (1976) introduced an iterative algorithm to accomplish R-estimation method in the linear model. We will use their algorithm to estimate the regression coefficients, then calculate the bias, variances and MSE of the coefficients, and coverage probability of confidence intervals.

We want to attain 0.95 coverage probability, i.e. $p = 0.95$. Let Z be the value obtained from estimation

$$Z = \begin{cases} 1 & \text{if } \widehat{\theta} \in \text{confidence interval with } p = 0.95 \\ 0 & \text{if not} \end{cases} . \tag{5.1}$$

We try the same estimation n times, then we may get $Z_1, \ldots, Z_n$ having the same distribution as Z. Let

$$\widehat{p}_n = \frac{\sum_{i=1}^{n} Z_i}{n} . \tag{5.2}$$

Then $E\widehat{p}_n = p$, $Var\widehat{p}_n = \frac{p(1-p)}{n}$. By CLT $\sqrt{n}(\widehat{p}_n - p) \sim N(0, p(1-p))$ as $n \to \infty$. If we want

$$P(|\widehat{p}_n - p| < \varepsilon) = P(\frac{\sqrt{n}|\widehat{p}_n - p|}{\sqrt{p(1-p)}} < \frac{\sqrt{n}\varepsilon}{\sqrt{p(1-p)}}) \geq 1 - \delta,$$

as n large enough we have

$$P(\frac{\sqrt{n}|\widehat{p}_n - p|}{\sqrt{p(1-p)}} < \frac{\sqrt{n}\varepsilon}{\sqrt{p(1-p)}}) \approx P(|X| < \frac{\sqrt{n}\varepsilon}{\sqrt{\widehat{p}_n(1-\widehat{p}_n)}}) \geq 1 - \delta$$

where $X \sim N(0,1)$. So 95% of confidence interval of p is $(\widehat{p}_n - \sqrt{\frac{\widehat{p}_n(1-\widehat{p}_n)}{n}} \cdot X_{0.975}), \widehat{p}_n + \sqrt{\frac{\widehat{p}_n(1-\widehat{p}_n)}{n}} \cdot X_{0.975})$ where $1 - \Phi(X_{0.975}) = 0.025$

The following tables show us the R-estimation results for four different scores where the bias is estimated by

$$Bias = \begin{pmatrix} \overline{\widehat{\theta_0}} - \theta_{0*} \\ \overline{\widehat{\theta}} - \theta_* \end{pmatrix} , \tag{5.3}$$

and variance and MSE of $\begin{pmatrix} \theta_0 \\ \theta \end{pmatrix}$ are estimated by

$$Variance = \frac{1}{n-1} \begin{pmatrix} \sum_{i=1}^n (\widehat{\theta}_{0i} - \overline{\widehat{\theta}_0})^2 \\ \sum_{i=1}^n (\widehat{\theta}_i - \overline{\widehat{\theta}})^2 \end{pmatrix}$$

$$MSE = \frac{1}{n-1} \begin{pmatrix} \sum_{i=1}^n (\widehat{\theta}_{0i} - \overline{\widehat{\theta}_0})^2 & \sum_{i=1}^n (\widehat{\theta}_{0i} - \overline{\widehat{\theta}_0})(\widehat{\theta}_i - \overline{\widehat{\theta}})^T \\ \sum_{i=1}^n (\widehat{\theta}_{0i} - \overline{\widehat{\theta}_0})(\widehat{\theta}_i - \overline{\widehat{\theta}}) & \sum_{i=1}^n (\widehat{\theta}_i - \overline{\widehat{\theta}})(\widehat{\theta}_i - \overline{\widehat{\theta}})^T \end{pmatrix}$$

$$+ \begin{pmatrix} \overline{\widehat{\theta}_0} - \theta_{0*} \\ \overline{\widehat{\theta}}_i - \theta_* \end{pmatrix} \begin{pmatrix} \overline{\widehat{\theta}_0} - \theta_{0*} \\ \overline{\widehat{\theta}}_i - \theta_* \end{pmatrix}^T \tag{5.4}$$

respectively, where $\overline{\widehat{\theta}_0} = \frac{1}{n} \sum_{i=1}^n \widehat{\theta}_{0i}$ and $\overline{\widehat{\theta}} = \frac{1}{n} \sum_{i=1}^n \widehat{\theta}_i$. In (5.4) we use the unbiased estimates of variance and MSE to estimate them which is easy to be proved. We can prove an equivalent statement: let $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ be iid random samples with $Ex_i = \mu_1, Ey_i = \mu_2$, $Var(x_i) = \sigma_1^2, Var(y_i) = \sigma_2^2$, $E((x_i - \mu_1)(y_j - \mu_2)) = \delta_{ij}\rho\sigma_1\sigma_2$ where $\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$. Then $\frac{1}{n-1} \sum_{i=1}^n (x_i - \overline{x})^2$, $\frac{1}{n-1} \sum_{i=1}^n (y_i - \overline{y})^2$, and $\frac{1}{n-1} \sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})$ are the unbiased estimates of $Var(x_i)$, $Var(y_i)$, and $Cov(x_i, y_i)$, respectively. We have the result since

$$E(\frac{1}{n-1} \sum_{i=1}^n (x_i - \overline{x})^2) = E(\frac{1}{n-1} \sum_{i=1}^n x_i^2) - \frac{nE(\overline{x})^2}{n-1}$$
$$= \frac{n}{n-1}(\sigma_1^2 + \mu_1^2) - \frac{n}{n-1}(\frac{\sigma_1^2}{n} + \mu_1^2) = \sigma_1^2 ,$$

$$E(\frac{1}{n-1} \sum_{i=1}^n (y_i - \overline{y})^2) = E(\frac{1}{n-1} \sum_{i=1}^n y_i^2) - \frac{n}{n-1} E(\overline{y})^2$$
$$= \frac{n}{n-1}(\sigma_2^2 + \mu_2^2) - \frac{n}{n-1}(\frac{\sigma_2^2}{n} + \mu_2^2)$$
$$= \sigma_2^2 ,$$

$$E(\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})) = \frac{1}{n-1}\sum_{i=1}^{n}E(x_iy_i - x_i\overline{y} - \overline{x}y_i + \overline{xy})$$
$$= \frac{1}{n-1}\sum_{i=1}^{n}E(x_iy_i) - \frac{n}{n-1}E(\overline{xy})$$
$$= \frac{1}{n-1}n(\rho\sigma_1\sigma_2 + \mu_1\mu_2) - \frac{n}{n-1}\frac{1}{n^2}(n\rho\sigma_1\sigma_2 + n^2\mu_1\mu_2)$$
$$= \rho\sigma_1\sigma_2.$$

In the tables below, P_1 represents the estimated coverage probability of confidence interval obtained from asymptotic distribution. We know that, asymptotically

$$\widehat{\theta} \sim N(\theta_*, var(\widehat{\theta})). \tag{5.5}$$

So the 95% of confidence interval for θ_* is $(\widehat{\theta}-\sqrt{var(\widehat{\theta})}\cdot X_{0.975}, \widehat{\theta}+\sqrt{var(\widehat{\theta})}\cdot X_{0.975})$ where $1-\Phi(X_{0.975}) = 0.025$. With this confidence interval we can get $Z_1, \ldots, Z_n$ having the same distribution as Z in (5.1). Then we define

$$p_1 = \frac{1}{n}\sum_{i=1}^{n} Z_i\,. \tag{5.6}$$

The width of this confidence interval is $2\sqrt{Var(\widehat{\theta})}\cdot X_{0.975}$. Similarity, P_2 represent the estimated coverage probability obtained by testing. From (3.4.1) we know that E^2 has asymptotic distribution χ^2_{p-q-1} under the null hypothesis. We construct 95% confidence interval by testing like this: θ belongs to this confidence interval iff the hypothesis that $\theta - \theta_* = 0$ is accepted by testing. Similarly as above we can get $Z'_1, \ldots, Z'_n$ and define

$$p_2 = \frac{1}{n}\sum_{i=1}^{n} Z'_i\,. \tag{5.7}$$

In this case we are unable to give the general expression to calculate the width of confidence interval for θ_*.

Table 5.1 Results for model $y = 1 + x + e$ with $e \sim N(0,1)$

	P_1/95% CI	P_2	Bias	Variance	Covariance
Wilcoxon	0.95 (0.95-0.0427, 0.95+0.0427)	0.94	$\begin{pmatrix} 0.017012 \\ -0.000593 \end{pmatrix}$	$\begin{pmatrix} 0.019909 \\ 0.000388 \end{pmatrix}$	-0.000120
Sign	0.88 (0.88-0.0637, 0.88+0.0637)	0.83	$\begin{pmatrix} 0.027386 \\ -0.001106 \end{pmatrix}$	$\begin{pmatrix} 0.029995 \\ 0.000752 \end{pmatrix}$	-0.000463
Normal	0.94 (0.94-0.0465, 0.94+0.0465)	0.95	$\begin{pmatrix} 0.015515 \\ -0.000031 \end{pmatrix}$	$\begin{pmatrix} 0.018924 \\ 0.000353 \end{pmatrix}$	-0.000143
Minimax	0.97 (0.97-0.0334, 1)	0.96	$\begin{pmatrix} 0.017192 \\ -0.001066 \end{pmatrix}$	$\begin{pmatrix} 0.021356 \\ 0.000430 \end{pmatrix}$	-0.000077

Table 5.2 Results for model $y = 1 + x + e$ with $e \sim 0.8N(0,1) + 0.2N(0,9)$

	P_1/95% CI	P_2	Bias	Variance	Covariance
Wilcoxon	0.95 (0.95-0.0427, 0.95+0.0427)	0.95	$\begin{pmatrix} -0.010623 \\ 0.000252 \end{pmatrix}$	$\begin{pmatrix} 0.032515 \\ 0.000677 \end{pmatrix}$	-0.000845
Sign	0.90 (0.90-0.0588, 0.90+0.0588)	0.83	$\begin{pmatrix} 0.004162 \\ 0.000237 \end{pmatrix}$	$\begin{pmatrix} 0.038172 \\ 0.000864 \end{pmatrix}$	-0.001063
Normal	0.96 (0.96-0.0384, 0.96+0.0384)	0.95	$\begin{pmatrix} -0.016270 \\ 0.000499 \end{pmatrix}$	$\begin{pmatrix} 0.036335 \\ 0.000756 \end{pmatrix}$	-0.0000868
Minimax	0.98 (0.98-0.0274, 1)	0.98	$\begin{pmatrix} -0.008884 \\ 0.000190 \end{pmatrix}$	$\begin{pmatrix} 0.031883 \\ 0.000649 \end{pmatrix}$	-0.000781

In above tables we take $n = 100$ and use $\mathcal{E} = 0.1$ in the $\mathcal{E}$-contaminated normal case for minimax estimates. With the same formula we can give confidence interval for P_2.

The results from tables strongly support our arguments that this robust R-estimation method performs well both in the general case when $e \sim N(0,1)$ and in the case when outliers exists, i.e. when $e \sim 0.8N(0,1) + 0.2N(0,9)$. Normal and Wilcoxon scores perform well in both cases. Sign score gives the worst performance in both cases because of its discontinuity. Normal score and minimax score give the most accurate estimation results in this two cases, respectively.

The algorithm of Hettmansperger and MeKean gives much better estimation results for slope parameters than for intercept. In both tables for each score we have much larger bias (absolute value) and variance for intercept than for slope. So further improvement is needed for more accurate estimation.

References

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

Adichie, J.N. (1978). Rank test of subhypothesis in the general linear regression, Ann. Statist 6,1012-1026.

Aubuchon, J.C. (1982). Rank tests in the linear model: Asymmetric errors. Ph.D. dissertation. Dept. Statistics, Pennsylvania State Univ.

Aubuchon, J.C. and Hettmansperger, T.P. (1989). Rank-based inference for linear models: Asymmetric errors. Statistics and Probability Letters 8, 97-107.

Bassett, G., Jr.and Koenker,R. (1982). An empirical quantile function for linear models with iid errors. J. Amer. Statist. Assoc. 77, 407-415.

Collins, J.R. (1983). On the minimax property for R-estimators of location. Ann. Statist. 11, 1190-1195.

Collins, J.R. and Wiens,D.P. (1989). Minimax properties of M-, R-, and L-estimators of location in Lévy neighborhoods. Ann. Statist. 17, 327-336.

Draper, D. (1981). Rank-based robust analysis of linear models. Ph.D. dissertation, Dept. Statistics, Univ. California, Berkeley.

Draper, D. (1988). Rank-based robust analysis of linear models. I. Exposition and Review. Statistical Science Vol. 3,No. 2, 239-271.

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.

Hettmansperger, T.P. and McKean, J.W. (1976). Computational problems involved in analysis of linear models based on ranks. Proc. Statist. Comp. Sec. Amer. Statist. Assoc. 88-94.

Hettmansperger, T.P. and McKean , J.W. (1977). A robust alternative based on ranks to least squares in analyzing linear models. Technometrics 19, 275-284.

Hodges, J.L., Jr. and Lehmann, E.L. (1963). Estimates of location based on rank tests. Ann. Math. Statist. 34, 598-611.

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

Jaeckel, L.A. (1971). Robust estimates of location: Symmetric and asymmetric contamination. Ann. Math. Statist. 42, 1020-1034.

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

Jurěcková, J. (1971). Nonparametric estimation of regression coefficents. Ann. Math. Statist. 42, 1328-1338.

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

Lehmann, E.L. (1963). Robust estimation in analysis of variance. Ann. Math. Statist. 34, 957-966.

Mckean, J.W. and Hettmansperger, T.P. (1976). Tests of hypotheses based on ranks in the general linear model. Comm. Statist. A - Theory Methods 5, 693-709.

McKean, J.W. and Sievers, G.L.(1989). Rank scores suitable for analyses of linear models under aysmmetric error distributions. Technometrices Vol. 31, No. 2, 207-218.

Policello, G.E., II and Hettmansperger, T.P. (1976). Adaptive robust procedures for the one-sample location problem. J. Amer. Statist. Assoc. 71 624-633.

Puri, M.L. and Sen, P.K. (1969). A class of rank order tests for a general linear hypothesis. Ann. Math. Statist. 40, 1325-1343.

Wiens, D.P. (1990). Minimax-variance L- and R-estimators of Location. Candian J. Statist. Vol. 18, No. 1, 47-58.

von Mises, R. (1947). On the asymptotic distribution of differentiable statistical functions. Ann. Math. Statist. 37, 87-91.