4-1989

Coefficients of Multiple Determination Based on Rank Estimates

Lee D. Witt
Western Michigan University

COEFFICIENTS OF MULTIPLE DETERMINATION BASED ON RANK ESTIMATES

by

Lee D. Witt

A Dissertation
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Doctor of Philosophy
Department of Mathematics and Statistics

Western Michigan University
Kalamazoo, Michigan
April 1989

Order Number 9000837

Coefficients of multiple determination based on rank estimates

Witt, Lee Dean, Ph.D.

Western Michigan University, 1989

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

ACKNOWLEDGEMENTS

I wish to thank Joseph W. McKean for his support during this work . I would also like to thank my family and friends for their patience, and Jay Treiman for his help .

Lee D. Witt

TABLE OF CONTENTS

CHAPTER

Table of Contents – Continued

CHAPTER

LIST OF TABLES

CHAPTER I

INTRODUCTION

In this work we consider two problems, each of which may be approached using methods from the analysis of linear models.

In the first problem, we observe random variables $Y_1, Y_2, \ldots, Y_n$, where we assume that each Y_i is generated by a model of the form

$$Y_i = \alpha + \vec{x}_i'\vec{\beta} + e_i \tag{1.1}$$

where

1) $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n$ are fixed p dimensional vectors.

2) $e_1, e_2, \ldots, e_n$ are iid random variables

3) $\vec{\beta}$ is a fixed, unkown p dimensional vector.

By denoting by $\vec{1}$ the p vector which has all components equal to one, by X the n by p matrix whose i'th row is $\vec{x}_i$, and by $\vec{e}$ the vector of errors, we may write (1.1) as follows.

$$\vec{Y} = \vec{1}\alpha + \vec{X}\vec{\beta} + \vec{e}$$

Two common problems in the analysis of linear models are estimating $\vec{\beta}$ and testing various hypotheses about $\vec{\beta}$. We will be concerned only with the problem of testing

$$H_0 : \vec{\beta} = \vec{0}$$

The methods commonly used for handling these problems are based on least squares.

If the joint distribution of the errors is multivariate normal, then least squares techniques have many optimal properties. However, real data are often far from normally distributed. For instance, most real data have outlying observations. The use of least squares on such data often results in a poor analysis. Because of this, much has been done in the area of developing techniques which perform well in comparison to least squares when the errors are normally distributed and have more power than least squares when the errors are not normal and when outliers are present. Such techniques have been called robust procedures.

In the context of M–estimation, Huber (1973,1977) generalized his work on location estimates to the estimation of the parameters in a linear model. Subsequently, Schrader and Hettmansperger (1980) proposed a procedure for performing tests on these parameters, using M–estimates. Their motivation was the method of likelihood ratio testing. Chapters VI and VII of the book by Hampel, Rousseeuw, Ronchetti, and Stahel (1985) discuss these ideas as well as more recent methods in this area.

An alternate approach to these problems was initiated by Jaeckel in 1972. This method extends the use of rank estimates and testing from the one and two sample problems to analysis of linear models. In least squares, $\vec{\beta}$ is estimated by minimizing the sums of squares of the residuals. Jaeckel (1972) proposed minimizing a convex function of the residuals (called a dispersion function) whose definition involves the ranks of the residuals. Under some regularity conditions on the design matrix $\vec{X}$, Jaeckel (1972) obtained the asymptotic distribution of his estimate $\vec{\beta}_R$ of $\vec{\beta}$. At approximately the same time, Jureckova (1971) proposed an alternate method of estimating $\vec{\beta}$. It was shown by Jaeckel (1972) that his estimate was asymptotically equivalent to that of Jureckova, in the sense that the two estimates had the same asymptotic distribution and had a difference which converged to zero in probability as the sample size increased. McKean and Hettmansperger (1976) used the estimate of Jaeckel (1972) and his dispersion function in obtaining a testing procedure for the above hypothesis H_0. The asymptotic distribution for their statistic turned out to be chi-square, so that not only it's use but it's performance could be compared to that of least squares.

Both of these procedures just mentioned, the M–estimate based method and that based on R(ank)–estimates, utilize test statistics which are constructed in a way similar to the least squares test statistic. A goal of this work is to use rank methods to propose a quanitity which may be used the way R^2 is used in least squares. We will pattern the search on the relationship between the least squares test statistic and R^2, and see that this leads to two rank based measures. Using empirical and theoretical evidence, we will show that one of these is robust while the other is not.

So far we have considered the $\vec{x}$'s to be fixed. In the case where they are random vectors, independently distributed from the e's, we will see that the rank based quantities we proposed extend naturally to measures of multiple correlation between Y and $\vec{X}$, just as R^2 does in the least squares case. Specifically, we assume that, given

$$\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n \tag{1.3}$$

we have the linear model

$$Y_i = \alpha + \vec{\beta}'\vec{x}_i + e_i \tag{1.4}$$

With f denoting the error density and m the density of $\vec{x}$, the joint density of Y and $\vec{X}$ is

$$h(\vec{x}, y) = f(y - \alpha - \vec{\beta}'\vec{x})m(\vec{x})$$

This model will be called the correlation model. Notice that the variables Y and $\vec{X}$ are independent in this model if and only if $\vec{\beta} = \vec{0}$. We will see that the testing and estimation procedures based on ranks may be extended to the present situation from the fixed $\vec{X}$ case and may be used in exactly the same manner. In addition we will obtain influence functions of the estimate of $\vec{\beta}$ and the test statistic used in the rank analysis. Finally, a comparison will be made between the rank methods presented here, and some competing methods. For instance, Ghosh and Sen (1971) considered a model similar to our regression model. They worked only on the testing problem and not on estimation. McKean and Sievers (1988) looked at this problem from the ℓ_1 viewpoint. Finally, we will consider the performance of our rank methods against some methods based on M–estimates.

CHAPTER II

Classical Linear Model

2.1 Introduction

In this chapter we consider the classical linear model. We consider the $\vec{x}_i'$ s to be nonstochastic. First, we give a brief review of the ideas used in least squares to estimate $\vec{\beta}$, construct test statistics for testing $H_0 : \vec{\beta} = \vec{0}$, and discuss the relationship between the test statistic and the coefficient of determination R^2. Once this is done, the method of analysis of linear models using ranks is discussed. We will outline the results of rank analysis in the same order that the results for least squares were introduced. After we have discussed the rationale for the test statistic for rank analysis, we will introduce two measures of multiple determination which are related to the test statistic. The motivation for these rank based measures will be the link between the least squares test statistic and the classical measure of multiple determination. The test statistic we will use will be the one proposed by Hettmansperger and McKean (1975).

2.2 Notations and Assumptions

In this section we review the notations and assumptions used in the regression model. These will come in two parts, those which will be used throughout the chapter, and those which will be required for the least squares analysis.

We observe random variables $Y_1, Y_2, \ldots, Y_n$, which are independent and which follow the model given by :

$$Y_i = \alpha + \vec{\beta}'\vec{x}_i + e_i \qquad 1 \leq i \leq n \tag{2.1}$$

We make the following assumptions, which we will use throughout, unless stated otherwise. The errors $e_1, e_2, \ldots, e_n$ are independent, identically distributed random variables with absolutely continuous distribution function F , which has density f. We assume that f is absolutely continuous with derivative f'. The vector $\vec{\beta} \in R^p$ is unknown and must be estimated. The intercept parameter $\alpha \in R$ is unknown. If we set

$$\vec{Y} = (Y_1, Y_2, \ldots, Y_n)'$$

$$\vec{1} = (1, 1, \ldots, 1)'$$

$$\vec{e} = (e_1, e_2, \ldots, e_n)'$$

and let X denote the matrix whose i'th row is $\vec{x}_i'$, then (2.1) may be written as

$$\vec{Y} = \alpha\vec{1} + X\vec{\beta} + \vec{e} \tag{2.2}$$

For this model, the inference results based on the methods of least squares require the following assumptions. Let $\overline{X}$ denote the matrix of column means of X.

A1: E[e] = 0

A2: Var[e] $= \sigma^2$

A3: As $n \to \infty, \frac{1}{n}(X - \overline{X})(X - \overline{X}) \to \Sigma$ where we assume that Σ is positive definite

A4: (Hubers Condition) Let $V = range[\vec{1}\ X - \overline{X}]$, and denote by P_V the projection matrix onto V. We assume that $(P_V)_{ii} \to 0$ as $n \to \infty$

2.3 Least Squares Analysis

Let us now discuss the least squares analysis of the regression model 2.1. We assume that assumptions A1 through A4 hold. Since the estimation of $\vec{\beta}$ and the testing problem are closely related, these two topics will be discussed together.

To estimate $\vec{\beta}$, we must minimize the function S, which is given by

$$S(\vec{b}) = \sum_{i=1}^{n}[y_i - \overline{y} - \vec{b}'(\vec{x}_i - \overline{\vec{x}})]^2$$

Denote the minimizing value by $\vec{\beta}_{LS}$. Notice that this value may be obtained by solving the system of equations (2.4),

$$\nabla \vec{S}(\vec{\beta}_{LS}) = \vec{0} \tag{2.4}$$

where ∇ denotes the gradient operator. Let the true value of $\vec{\beta}$ be denoted $\vec{\beta}_0$. Under our assumptions A1, A2, A5, Arnold (1980) shows that

$$\sqrt{n}(\vec{\beta}_{LS} - \vec{\beta}_0) \to \mathrm{MVN}_p[\vec{0}, \sigma^2\Sigma^{-1}] \tag{2.5}$$

where $\to$ denotes convergence in distribution. In practice, Σ must be estimated and $\frac{1}{n}[(X - \overline{X})'(X - \overline{X})]$ is used in its place. To construct the least squares test statistic, set

$$\hat{\sigma}^2_{LS} = \frac{1}{n-p-1} S(\vec{\beta}_{LS}) \tag{2.6}$$

The test statistic will be denoted by F_{LS}, and is defined as

$$pF_{LS} = \frac{S(\vec{0}) - S(\vec{\beta}_{LS})}{\hat{\sigma}^2_{LS}} \tag{2.7}$$

The hypothesis that H_0 is true is rejected for values of pF_{LS} which are sufficiently large.

The statistic pF_{LS} has a pleasing geometric interpretation. It can be shown that $S(\vec{b})$ is a semi norm on R^p. Let A be given by

$$A = \{(X - \overline{X})\vec{b} : \vec{b} \in R^p\}$$

Then $S(\vec{0})$ measures the distance, according to S, between $\vec{Y}$ and the origin of R^p. Similarly, $S(\vec{\beta}_{LS})$ measures the distance between $\vec{Y}$ and the subspace A. The difference is scaled by $\hat{\sigma}^2_{LS}$, an estimate of variance. If the (scaled) difference is large, it indicates that $\vec{\beta}_{LS}$ is close to A, and offers evidence that $\vec{\beta}_{LS}$ is not the zero vector.

If the errors are normally distributed, then the exact distribution of $\vec{\beta}_{LS}$ may be obtained. Even without this requirment, the asymptotic distribution may be obtained. We will state this result for the null hypothesis and under a sequence of contiguous alternatives. To define the latter, let $\vec{d}$ denote an arbitrary non-zero vector in p dimensional space, and consider a sequence of alternative hypotheses which depend on n and converge to the null hypothesis. Such a sequence of hypthoses is given by

$$H_n : \vec{\beta} = \frac{1}{\sqrt{n}}\vec{d}$$

These hypotheses are said to be contiguous to the null hypothesis. For a discussion of contiguity, see Hájek and Sidáck(1967). With these ideas and definitions in mind, we state the following theorem:

THEOREM 2.1. *In the regresion model if assumptions A1, A2, and A4 are true, then the following results hold: (a) Under the null hypothesis, the asymptotic distribution of pF_{LS} is central chi-square with p degrees of freedom. (b) Under the sequence of contiguous alternatives, the asymptotic distribution of pF_{LS} is the noncentral chi-square distribution with p degrees of freedom and having the non-centrality parameter*

$$\delta_{LS}^2 = \frac{\vec{d}'\Sigma\vec{d}}{\sigma^2}$$

For a proof of this theorem, see Hettmansperger and McKean (1975).

2.4 Rank Analysis

While least squares procedures are optimum when the errors are normal, they are sensitive to departures from normality, and to the presence of outliers in the data. In fact, one point, sufficiently far removed from the bulk of the data can completely determine the fit. To make an attempt to get around this problem, we turn to rank based estimates and procedures. The path we follow was blazed by Jaeckel in 1972, and further extended by McKean and Hettmansperger (1976). Slightly different approaches to the problem have been proposed by Ghosh and Sen (1971), Tableman (1988), and Sievers (1979).

We first discuss the assumptions needed for the rank analysis, introduce the required notations, and procede as we did in section four in the discussion of least squares procedures. We no longer need assumption A2, and replace it with

A2' The density f of the errors is absolutely continuous, and has finite Fisher information, which is given by

$$\Im = \int_{-\infty}^{\infty} (-\frac{f'(e)}{f(e)})^2 \, dF(e)$$

Recall that in least squares, the estimation of $\vec{\beta}$ was carried out by minimizing the sum of the squares of the residuals. This function was a nonnegative convex function of the residuals. In the rank analysis of the linear model, the estimation of $\vec{\beta}$ is performed by minimizing a convex, piecewise linear,nonnegative function of the residuals. This approach was proposed by Jaeckel (1972), and the function is called a dispersion function. The dispersion function depends of the ranks of the residuals. We will need the following definitions.

DEFINITION 2.1. *Let* $z_1, z_2, \ldots, z_n$ *be n real numbers. Denote the rank of an arbitrary z by R(z). The rank of any number is simply the position of that number among the ordered numbers. For instance, the minimum z would have a rank of 1 and the maximum z would have a rank of n.*

DEFINITION 2.2. *Let* $a(1) \leq a(2) \leq \cdots \leq a(n)$ *be a sequence of real numbers, not all equal,such that* $\sum_{i=1}^{n} a(i) = 0$. *We call the a's scores.*

The scores are generally obtained by means of a score function, which we now define.

DEFINITION 2.3. *Let* φ *be a nonconstant, nondecreasing, bounded function on (0,1), and satisfying the conditions*

$$\int_0^1 \varphi(u)\, du = 0$$
$$\int_0^1 \varphi^2(u)\, du = 1$$

Then φ *is called a score function.*

We use φ to generate scores by means of the relation

$$a(i) = \varphi(\frac{i}{n+1}) \qquad 1 \leq i \leq n$$

In the least squares analysis, the quantity σ, a scale parameter, had to be estimated. In the rank analysis it is not σ that must be estimated but another scale parameter which we denote by τ. In order for this parameter to exist, we need to assume the existence of an integral which involves the error distribution and the score function. This is given in assumption 5 below.

A5 Assume that $0 < \gamma < \infty$, where γ is given by

$$\gamma = \int \varphi(u)[-\frac{f'(F^{-1}(u))}{f(F^{-1}(u)}]^2\, du$$

The parameter τ is defined to be the reciprocal of γ. An estimate of τ is given in Koul, Sievers, and McKean (1987). Some comments are in order.

(1) For the rank analysis the assumptions used are A1,A2,A3,A4,A5

(2) If φ is sufficiently differentiable, we have that

$$\gamma = \int_{-\infty}^{\infty} \varphi'[F(t)]f^2(t)\,dt$$

We will typically assume that our score functions have this degree of smoothness.

(3) If we take the score function to be $\varphi = \sqrt{12}(u - \frac{1}{2})$ then the scores generated are typically called Wilcoxon scores. In this case if the error distribution is normal, then the scale parameter τ reduces to

$$\tau = \sigma\sqrt{\frac{\pi}{3}}$$

We are now in a position to define the dispersion function. We follow the ideas in Jaeckel (1972).

DEFINITION 2.4. *For an arbitrary $a \in R, \vec{b} \in R^p$ set*

$$\begin{aligned} D(\vec{b}) &= \sum_{i=1}^{n} a[R(y_i - a - \vec{b}'\vec{x}_i)](y_i - a - \vec{b}'\vec{x}_i) \\ &= \sum_{i=1}^{n} a[R(y_i - \vec{b}'\vec{x}_i)](y_i - \vec{b}'\vec{x}_i) \end{aligned}$$

We will refer to D as (Jaeckel's) dispersion function.

In the rank analysis of linear models D plays the same role as the S function does in least squares. Jaeckel (1972) proposed estimating $\vec{\beta}$ by minimizing D as a function of $\vec{b}$. Some important properties of D are summarized below, proofs can be found in Jaeckel (1972) or McKean and Hettmansperger (1976).

THEOREM 2.2.

a) *D is nonnegative,convex,piecewise linear as a function of* $\vec{b}$.

b) *D is a differentiable function of* $\vec{b}$ *almost everywhere.*

c) *The minimum value of D is unique, but the minimizing value of* $\vec{b}$ *is not.*

d) *If* $A = \{\vec{b}^* : D(\vec{b}^*) = MIN\}$, *then*

$$\sqrt{n}\ diam(A) \rightarrow 0 \text{ in probability as } n \rightarrow \infty$$

Numerical methods must be used to minimize D. We notice that by part b of Theorem 2.2, the gradient of D exists almost everywhere, and so we may obtain the rank estimate of $\vec{\beta}$ by solving the system of equations

$$\sum_{i=1}^{n} a[R(y_i - \vec{b}'\vec{x}_i)]\vec{x}_i = 0$$

This is equivalent to solving $\nabla D(\vec{b}) = 0$. We will denote the estimate of $\vec{\beta}$ by $\vec{\beta}_R$. In the next section we outline the asymptotic behavior of the estimate and introduce the test statistic. We also note that we cannot estimate the parameter α with our rank procedure. For the problems we will be considering, this will not be a problem. For a discussion on the estimation of α, see Hettmansperger and McKean (1976).

2.5 Asymptotic Distributions

In his paper, Jaeckel (1972) obtained the asymptotic distribution of $\vec{\beta}_R$. We state it below as Theorem 2.3.

THEOREM 2.3. *Under assumptions A2',A3,A4, and A5,*

$$\sqrt{n}(\vec{\beta}_R - \vec{\beta}_0) \rightarrow MVN_p(\vec{0}, (\gamma\Sigma)^{-1})$$

Here $\rightarrow$ represents convergence in distribution, and $\vec{\beta}_0$ denotes the true value of $\vec{\beta}$.

The proof of this theorem depends on an approximation to the gradient of D. It can be shown that the gradient is, in a neighborhood of $\vec{\beta}_0$, appoximately linear. From this a quadratic approximation to the disperson function itself is formed. The minimizing value of this quadratic, and it's asymptotic distribution are easy to obtain. It can then be shown that the difference between the rank estimate and the value which minimizes the quadratic tends to zero in probability, and so the distribution of the minimizing value of the quadratic is the same as the distribution of the rank estimate. The book by Hettmansperger (1984) gives a simplified version of the proof for the special case of Wilcoxon scores.

We now move to the case of testing that $H_0 : \vec{\beta} = \vec{0}$ is true. The motivation for the test statistic comes from least squares. The testing problem was first examined by McKean and Hettmansperger (1976). They show that the dispersion function D is a semi-norm on R^p, and that D has the same class of representatives as the least squares function S. Recalling the geometric rationale behind the least squares statistic, we define the rank test statistic as

$$pF_R = \frac{[D(\vec{0}) - D(\vec{\beta}_R)]}{\frac{\hat{\tau}}{2}} \tag{2.10}$$

Here $\hat{\tau}$ is a consistent estimate of τ. Koul, Sievers, and McKean (1987) propose a consistent estimate of γ, and from this we may obtain an estimate of τ. An advantage of the estimate of Koul,Sievers, and McKean (1987) is that the assumption of symmetry need not be made for the error distribution. If we are willing to assume that the errors are symmetric, then another estimate of γ is given by Schweder (1975), who also gives the asymptotic distribution of his estimator. His result holds for a wide choice of score functions.

Just as we did for the least squares test statistic, we reject the null hypothesis when the test statistic is too large. Specifically, McKean and Hettmansperger (1976) give the following result:

THEOREM 2.4. *Under assumptions A2',A3 ,A4 , and A5, we have the following. (a) When the null hypothesis is true, the asymptotic distribution of pF_R is chi-square with p degrees of freedom. (b) Under a sequence of contiguous alternatives $H_n : \vec{\beta} = \frac{1}{\sqrt{n}}\vec{d}$, where $\vec{d} \in R^p$, the asymptotic distribution of pF_R is noncentral chi-square with p degrees of freedom, where the noncentrality parameter is*

$$\delta_R^2 = \gamma^2 \vec{d}' \Sigma \vec{d}$$

The proof of this theorem depends on the quadratic approximation to the dispersion mentioned earlier.

It is important to note the similarity between the asymptotic distribution of the rank test statistic and the least squares test statistic. The least squares statistic is a scaled reduction in variation and the rank test statistic is a scaled reduction in dispersion. Because both of these statistics have asymptotic distributions which are chi square in nature, both under the null hypothesis and under a sequence of contiguous alternatives, we may determine the asymptotic relative efficiency of the rank procedure to the least squares procedure. This is given in the following theorem.

THEOREM 2.5. *Under the assumptions of the previous theorem, the asymptotic relative efficiency of pF_R to pF_{LS} is*

$$\gamma^2\sigma^2$$

In the case where the error distribution is normal and we use Wilcoxon scores, this value reduces to .955. If we assume that the error distribution is symmetric and we use Wilcoxon scores, the efficiency is always at least .864. This generalizes a result of Hodges and Lehmann (1956) in the location case.

These results indicate that rank procedures should be fairly robust. We will address this aspect in Chapter IV where we look at the influence functions of some of our quantities. From various simulation studies these asymptotic efficiency results appear to hold for moderate sized samples.

2.6 Coefficients of Multiple Determination

We now consider the problem of constructing measures of multiple determination from our rank estimates. These measures will be rank versions of the measure R^2 discussed earlier. Recall that R^2 was defined as

$$R^2 = \frac{S(\vec{0}) - S(\vec{\beta}_{LS})}{S(\vec{0})} \tag{2.11}$$

When R^2 was introduced, we made the observation that $R^2 \times 100$ could be interpreted as the percent of variation in the data which was explained by fitting the regression. If we substitute the term dispersion for variation in this last comment, and think of our rank quantities, we see that one possible measure of multiple determination based on ranks is

$$R_{1\varphi} = \frac{D(\vec{0}) - D(\vec{\beta}_R)}{D(\vec{0})} \tag{2.12}$$

The notation is meant to indicate the fact that $R_{1\varphi}$ depends upon the choice of score function φ.

We have mentioned that the least squares procedures are susceptible to outliers. In Chapter IV we will have theoretical evidence to indicate that our rank estimates do not have this problem to as large a degree as the least squares estimates. The same may be said for the rank test statistic as compared to the least squares test statistic. We would hope that this would carry over to $R_{1\varphi}$, but it does not. In hopes of obtaining a rank measure of multiple determination that is fairly robust, we recall the relationship that exists between R^2 and the least squares test statistic, namely

$$pF_{LS} = \frac{(n-p-1)R^2}{1-R^2} \tag{2.13}$$

This sort of relationship does not hold for $R_{1\varphi}$. However, we may define a new measure $R_{2\varphi}$ by

$$pF_R = \frac{(n-p-1)R_{2\varphi}}{1-R_{2\varphi}} \tag{2.14}$$

After some simplifications we find that we may express $R_{2\varphi}$ in terms of the dispersion function as

$$R_{2\varphi} = \frac{[D(\vec{0}) - D(\vec{\beta}_R)]}{D(\vec{0}) - D(\vec{\beta}_R) + (n-p-1)\frac{\hat{\tau}}{2}} \tag{2.15}$$

Thus we have two competitors for propose to R^2, each of which is based on ranks. We have already hinted at the fact that $R_{1\varphi}$ is not as robust as we would like it to be. We shall see in Chapter IV that $R_{2\varphi}$ is fairly robust and has the same distributional properties as pF_R. In the following example, we shall see that $R_{2\varphi}$ performs fairly well in practice, and hence should be considered a viable competitor to the classical least squares value.

2.7 Examples

In this section we present two examples illustrating the uses of the coefficients $R_{1\varphi}$ and $R_{2\varphi}$, and comparing them to R^2. The first of the examples illustrates the use of these coefficients in model selection, and the second the effect that outliers can have on the value of the coefficient of multiple determination.

The data for the first example are from Hald (1952). This example actually is in two parts. We will use the data to select the best one variable model, and then repeat the procedure for the two variable case. For each task, the procedure will be carried out first for the original data and then with an outlier introduced. Best here is taken to mean the model with the highest coefficient of multiple determination. In all cases we use the Wilcoxon score function

$$\varphi(u) = \sqrt{12}(u - \frac{1}{2})$$

For the results of the first fit we examine Table I below.

Table 1

	R^2	$R_{1\varphi}$	$R_{2\varphi}$
X_1	.53	.32	.39
X_2	.67	.46	.60
X_3	.29	.18	.34
X_4	.68	.45	.65

From this we see that all 3 coefficients are in essential agreement, although the values of $R_{1\varphi}$ are smaller that those of $R_{2\varphi}$. Each one indicates that the best one variable model should have either variable X_2 or X_4 in it, and the values are so nearly equal in these cases that a choice in favor of one model or the other is not obvious. This is encouraging, because we would like our new methods to be in reasonable agreement with least squares when the data is nicely behaved. In the next part, we introduce a bad data value. We replace the value $Y_{11} = 83.8$ with $Y_{11} = 8.8$ and again try to select the best model. The results are summarized in Table II.

Table 2

	R^2	$R_{1\varphi}$	$R_{2\varphi}$
X_1	.40	.24	.41
X_2	.30	.23	.56
X_3	.45	.19	.32
X_4	.24	.19	.26

Notice that the introduction of the outlier has totally confused the coefficient $R_{1\varphi}$. The values are so close that the selection of any one model as best is very difficult. Also, the classical coefficient R^2 has changed its choice about which variable should be in the best model. Instead of selecting either X_2 or X_4 it now flags X_1 or X_3 to be in the model. However, $R_{2\varphi}$ makes the same selection now as it did with the original data.

We now consider the second part of this example. We proceed just as above, but now we are trying to select the best double variable model. The results for the fits using the original data are given below, in Table III.

Table 3

	R^2	$R_{1\varphi}$	$R_{2\varphi}$
X_1, X_2	.98	.86	.92
X_1, X_2	.55	.33	.50
X_1, X_4	.97	.84	.90
X_2, X_3	.85	.63	.74
X_2, X_4	.68	.46	.61
X_3, X_4	.94	.76	.89

Again we see that the three coefficients are in agreement as to which model to pick. If we now fit the model with the outlier in we obtain the values given in Table IV.

Table 4

	R^2	$R_{1\varphi}$	$R_{2\varphi}$
X_1, X_2	.57	.55	.92
X_1, X_3	.47	.24	.41
X_1, X_4	.52	.51	.88
X_2, X_3	.66	.46	.71
X_2, X_4	.34	.27	.56
X_3, X_4	.67	.52	.83

We observe here that the two rank coefficients behave the same as they did on the original data and the least squares coefficient has switched choices for best model, just as it did in the one variable case. This is evidence for our claim that $R_{2\varphi}$ is more robust that is R^2.

CHAPTER III

CORELATION MODEL

3.1 Introduction

In this chapter we extend the regression model studied in Chapter II. The extension is to the case where the predictor variables are random vectors. Under some conditions on the distribution of the $\vec{X}$'s, we shall see that the classical multiple correlation model becomes a special case of this model. For this reason, we call this model the correlation model. For a further discussion of this model, see Sievers (1987).

The goal of this Chapter is to show that the rank based estimates and test introduced in Chapter II may be used in the correlation model in ways exactly analagous to the regression model. This is just as may be done with least squares when going from regression to correlation. As Sievers noted, the classical correlation model is a special case of this model. Further, in the classical multiple correlation model, we shall see that our statistics are consistent estimators of functions of $\overline{R}^2$, the classical multiple correlation coefficent.

3.2 Assumptions

We now give the assumptions we need for our correlation model. We

PLEASE NOTE:

Page(s) missing in number only; text follows.
Filmed as received.

UMI

In the regression model, the null hypothesis $H_0 : \vec{\beta} = \vec{0}$ corresponded to no regression. In the correlation model, this hypothesis corresponds to Y and $\vec{X}$ being stochastically independent, hence we will be concerned with H_0 in this model also. This model has been used by Sievers (1987) who proposed a measure of multiple correlation different from ours, but which was also based on ranks. Also, Hampel et al. (1985) in their chapters on robust M estimation used this model, although they did not consider problems of correlation.

3.3 Classical Multiple Correlation Model

We have mentioned that the classical multiple correlation model is a special case of our model. This may be seen as follows. The classical multiple correlation model assumes that $\vec{X}$ and Y are jointly distributed according to the multivariate normal distribution having dimension $p + 1$. We write the covariance matrix as

$$A = \begin{pmatrix} a_{11} & \vec{a}_{12} \\ \vec{a}_{21} & A_{22} \end{pmatrix} \tag{3.7}$$

and we assume that this matrix is positive definite. Using a result on the inverse of positive definite matrices (cf Anderson,1984) we have that

$$\sigma^2 = a_{11} - \vec{a}_{12} A_{22}^{-1} \vec{a}_{21} > 0$$

Let us denote by $\vec{\beta}$ the solution to the equation

$$A_{22}\vec{\beta} = \vec{a}_{12}$$

Then we may write the covariance matrix as

(3.8)
$$A = \begin{pmatrix} \sigma^2 + \vec{\beta}' A_{22} \vec{\beta} & A_{22}\vec{\beta} \\ \vec{\beta}' A_{22} & A_{22} \end{pmatrix}$$

Using this, and results on conditional distributions of multivariate normal vectors, we see that 3.4 holds.

3.4 Rank Procedures in the Correlation Model

We now discuss properties of rank procedures in the correlation model. Assume that we have a random sample of size n from (3.4). Our first goal is to use the sample to estimate $\vec{\beta}$. As in Chapter II we have scores $a(1) \leq a(2) \leq \ldots \leq a(n)$ which are generated by a score function φ. We estimate $\vec{\beta}$ by choosing any $\vec{b}$ in R^p which minimizes $D(\vec{b})$ as a function of $\vec{b}$, where, just as in Chapter II,

(3.9)
$$D(\vec{b}) = \sum_{i=1}^{n} a[R(y_i - \vec{b}'\vec{x}_i)](y_i - \vec{b}'\vec{x}_i)$$

Equivalently we may obtain $\vec{\beta}$ by solving

(3.10)
$$\sum_{i=1}^{n} a[R(y_i - \vec{b}'\vec{x}_i)]\vec{x}_i = \vec{0}$$

In the regression model, to ensure the asymptotic behavior of $\vec{\beta}_R$ and pF_R, we had to require that Huber's condition hold for the design matrix. In the present situation, Arnold (1980) has shown that this condition holds for the design matrix we have in the correlation model. With this in mind, we may state the first theorem of this Chapter. Its proof is not given; it follows exactly as in the regression case by conditioning on the $\vec{X}$'s.

THEOREM 3.1. *In the correlation model, if the true value of $\vec{\beta}$ is $\vec{\beta}_0$, then*

$$\sqrt{n}(\vec{\beta}_R - \vec{\beta}_0) \rightarrow MVN_p(\vec{0}, (\gamma^2\Sigma)^{-1})$$

where the $\rightarrow$ indicates convergence in distribution.

In a similar manner, we may show that, when the null hypothesis is true, pF_R has, asymptotically a central chi-square distribution with p degrees of freedom. Under a sequence of contiguous alternatives, given by $H_n : \vec{\beta} = \frac{1}{\sqrt{n}}\vec{d}$, it can be shown that pF_R has an asymptotic distribution which is non-central chi-square with p degrees of freedom, with the same noncentrality parameter as in the regression case. Thus, whether we have a regression problem or a multiple correlation problem, we may use rank estimates and testing procedures the same way and use the same critical values as well. This is as it is with least squares procedures. Since our rank based procedures are more robust than the least squares are, this is welcome news. However, since in the classical multiple correlation model, R^2 is a consistent estimator of the population multiple correlation coefficient, one asks whether $R_{1\varphi}$ and $R_{2\varphi}$ estimate any meaningful parameters in our correlation model. The answer is yes, and in a special case, we shall see that these quantities estimate functions of $\overline{R}^2$, the classical multiple multiple correlation coefficient. In general, $R_{1\varphi}$ and $R_{2\varphi}$ estimate parameters which depend on the joint distribution of Y and $\vec{X}$, the marginal distribution of Y, the distribution F of the errors, and the score function φ. To obtain these parameters, we must first discuss certain functional forms of the dispersion function. This is done in the following string of lemmas.

LEMMA 3.2. *As $n \to \infty, \frac{1}{n}D(\vec{0})$ converges in probability to D1, where*

$$D_1 = \int_{-\infty}^{\infty} \varphi[G(y)]y dG(y) \tag{3.11}$$

where G is the distribution function of the random variable Y.

PROOF: This is a direct consequence of a result given in David (1970). That result states

$$\sqrt{n}(\frac{1}{n}D(\vec{0}) - D_1)$$

is asymptotically normal. The result follows.

This lemma states that $D(\vec{0})$ may be represented as a functional of the population distribution of Y. We shall see that a similar result holds for $\frac{1}{n}D(\vec{\beta}_R)$. We will use these results again when we discuss influence functions.

LEMMA 3.3. *Let $\vec{\beta}_0$ denote the true value of the parameter $\vec{\beta}$. Then as $n \to \infty, \frac{1}{n}D(\vec{\beta}_0)$ tends in probability to D_2, where*

$$D_2 = \int_{-\infty}^{\infty} \varphi[F(e)]e \, dF(e) \tag{3.12}$$

where F denotes the distribution function of e.

PROOF: Notice that

$$\frac{1}{n}D(\vec{\beta}_0) \to \int_{-\infty}^{\infty} \varphi[F(t-\alpha)]t dF(t-\alpha)$$
$$= \int_{-\infty}^{\infty} \varphi[F(e)]e \, dF(e)$$

Theorem 3.1 shows that $\vec{\beta}_R$ converges in probability to $\vec{\beta}_0$. This fact and lemma 3.3 make it reasonable to expect that $\frac{1}{n}D(\vec{\beta}_R)$ should also converge to D_2. That this is true is shown in the next lemma.

LEMMA 3.4. *Let $\vec{\beta}_R$ denote the rank estimate of $\vec{\beta}_0$ obtained by minimizing 3.1. Then*

$$\frac{1}{n}[D(\vec{\beta}_R) - D(\vec{\beta}_0)]$$

converges to 0 in probability as n goes to infinity.

PROOF: We may show this by arguing conditionally, given $\vec{X}_1, \vec{X}_2, \ldots, \vec{X}_n$ The proof goes then exactly as in Hettmansperger and McKean (1976) for the regression model. Thus, the result holds unconditionally as well.

Combining lemmas 3.3 and 3.4, and using one form of Slutsky's Theorem (Bickel and Doksum,1977) we immediately obtain the following result.

$$\frac{1}{n}D(\vec{\beta}_R) \to D_2$$

where the indicated convergence is in probability.

The previous lemmas have concerned the behavior of the dispersion function evaluated at $\vec{0}$, $\vec{\beta}_0$ and the rank estimate of beta. The following lemma is concerned with a certain integral involving φ and a distribution function F.

Notice that here F refers to an arbitrary continuous distribution function on the real line.

LEMMA 3.5. *Let φ be a nonconstant, increasing score function on (0,1). If F denotes an absolutely continuous function on the real line. Then*

$$\int_{-\infty}^{\infty} \varphi[F(t)]tdF(t)dt \geq 0$$

PROOF: Since $\int_{-\infty}^{\infty} \varphi[F(t)]dF(t) = 0$ and φ is nonconstant, the set

$$\{t : \varphi[F(t)] = 0\}$$

is not empty. Then let t_0 denote the g.l.b of this set. Then we have the following

$$\begin{aligned}\int_{-\infty}^{\infty} \varphi[F(t)]tdF(t) = &\int_{-\infty}^{t_0} \varphi[F(t)](t - t_0)\, dF(t)\\ &+ \int_{t_0}^{\infty} \varphi[F(t)](t - t_0)dF(t)\\ &+ t_0 \int_{-\infty}^{\infty} \varphi[F(t)]dF(t)\end{aligned}$$

Since the last integral is 0, and each of the first two integrals is nonnegative, we are done.

We are now in a position to examine the behavior of $R_{1\varphi}$ and $R_{2\varphi}$, and we do this in the next section.

3.5 Behavior of Rank Coefficients

In this section we consider the behavior of the statistics $R_{1\varphi}$ and $R_{2\varphi}$ in the correlation model. We shall see that these two statistics converge in probability to parameters that are zero if and only if Y and $\vec{X}$ are independent. In this sense they generalize the behavior of R^2 in the classical multiple correlation model. We shall obtain expressions for these parameters which will show that in certain cases these parameters reduce to one-to-one functions of the population value of the multiple correlation coefficient. Using the the notation,

$$RD = D(\vec{0}) - D(\vec{\beta})$$

we can write $R_{1\varphi}$ and $R_{2\varphi}$ as

$$R_{1\varphi} = \frac{RD}{D(\vec{0})} \tag{3.14}$$

$$R_{2\varphi} = \frac{RD}{RD + (n-p-1)\frac{\hat{\tau}}{2}} \tag{3.15}$$

where $\hat{\tau}$ is a consistent estimator of τ. Lemmas 3.2 to 3.4 combine to yield the following theorem.

THEOREM 3.5. *In the correlation model the statistics* $R_{1\varphi}$ *and* $R_{2\varphi}$ *are consistent estimators of* $\overline{R}_{1\varphi}$ *and* $\overline{R}_{2\varphi}$*, respectively, where*

$$\begin{aligned} \overline{R}_{1\varphi} &= \frac{[D_1 - D_2]}{D_1} \\ \overline{R}_{2\varphi} &= \frac{[D_1 - D_2]}{[D_1 - D_2] + \frac{\tau}{2}} \end{aligned} \tag{3.16}$$

PROOF: Notice that

$$R_{1\varphi} = \frac{RD}{D(\vec{0})}$$

If we divide both numerator and denominator of the above expression by n, we may use the results of lemmas 3.2 to 3.4 to obtain the desired result. Similarly, we may prove the stated result for $R_{2\varphi}$

We interpret $\overline{R}_{1\varphi}$ and $\overline{R}_{2\varphi}$ as measures of multiple correlation between Y and $\vec{X}$. This will be seen in general in the next section, but for the moment we shall show, for a special subcase that these parameters reduce to functions of $\overline{R}^2$, the classical multiple correlation coefficient, which is defined as

$$\overline{R}^2 = \frac{\vec{\beta}_0' \Sigma \vec{\beta}_0}{\sigma^2 + \vec{\beta}_0' \Sigma \vec{\beta}_0}. \tag{3.16}$$

THEOREM 3.6. *Assume that in the correlation model, there is a distribution function F_0 such that*

$$F(e) = F_0(\frac{\epsilon}{\sigma})$$

and

$$G(y) = F_0(\frac{y-\alpha}{\sigma_y})$$

where

$$\sigma_y = \sqrt{\sigma^2 + \vec{\beta}' \Sigma \vec{\beta}}$$

Then

$$\overline{R}_{1\varphi} = 1 - \sqrt{1 - \overline{R}^2}$$
$$\overline{R}_{2\varphi} = \frac{1 - \sqrt{1 - \overline{R}^2}}{(1 - \sqrt{1 - \overline{R}^2}) + \frac{\tau}{2I\sigma_y}}$$

where

$$I = \int_{-\infty}^{\infty} \varphi[F_0(t)] t \, dF_0(t)$$

PROOF: Notice that

$$\begin{aligned}
D_1 &= \int_{-\infty}^{\infty} \varphi[G(y)]y \, dG(y) \\
&= \int_{-\infty}^{\infty} \varphi[F_0(\frac{y}{\sigma})] dF_0(\frac{y}{\sigma_y}) \\
&= \sigma_y \int_{-\infty}^{\infty} \varphi[F_0(t)]t \, dF_0(t) \\
&= I\sigma_y
\end{aligned}$$

Similarly $D_2 = I\sigma$. Thus

$$RD_{1\varphi} = \frac{[D_1 - D_2]}{D_1} = \frac{(\sigma_y - \sigma)I}{I\sigma_y} = 1 - \sqrt{1 - \overline{R}^2}$$

and

$$\begin{aligned}
\overline{R}_{2\varphi} &= \frac{(\sigma_y - \sigma)}{(\sigma_y - \sigma)I + \frac{\tau}{2}} \\
&= \frac{1 - \sqrt{1 - \overline{R}^2}}{1 - \sqrt{1 - \overline{R}^2} + \frac{\tau}{I 2\sigma_y}}
\end{aligned}$$

A very important instance of the situation in Theorem 3.6 is the classical correlation model in which Y and $\vec{X}$ have a joint normal distribution.In this case we have that the integral I in the above theorem reduces to a convenient form. We state this in the following

COROLLARY. *Under the conditions of Theorem 3.6, if the error distribution is normal with mean zero and variance σ^2 then the value of thie integral I is*

$$I = \sigma\gamma$$

Proof: In this case the distribution of the errors takes the form $\Phi(\frac{e}{\sigma})$ and we have

$$\begin{aligned} I &= \int_{-\infty}^{\infty} \varphi[F_0(t)]t dF_0(t) \\ &= \int_{-\infty}^{\infty} \varphi[\Phi(\frac{t}{\sigma})]t d\Phi(\frac{t}{\sigma}) \\ &= \sigma \int_{-\infty}^{\infty} \varphi[\Phi(s)]s d\Phi(s) \\ &= \sigma \int_{-\infty}^{\infty} \varphi[\Phi(s)](-\phi'(s))ds \\ &= \sigma \int_{-\infty}^{\infty} \varphi'[\Phi(s)]\phi^2(s)ds \\ &= \sigma\gamma \end{aligned}$$

In the next section we investigate the general case and obtain some of the characteristics of $\overline{R}_{1\varphi}$ and $\overline{R}_{2\varphi}$, and see that these parameters are meaningful even when F and G do not belong to a location scale family.

3.6 Properties of New Parameters

In this section we demonstrate that the parameters

$$\overline{R}_{1\varphi}, \overline{R}_{2\varphi}$$

are zero if and only if $\vec{\beta}_0 = \vec{0}$. Since this is true if and only if Y and $\vec{X}$ are independent, our parameters are zero if and only if we have independence. In this sense they generalize the behavior of the population multiple correlation coefficient in the classical correlation model. We do this by showing that the quantity which appears in the numerator of each of the parameters is equal to zero if and only if $\vec{\beta}$ is $\vec{0}$. This is accomplished by showing that the numerator is a convex function of $\vec{\beta}$. Let $G(y)$ denote the distribution function of Y. Then

$$G(y) = \int \cdots \int F(y - \vec{\beta}'\vec{x})dM(\vec{x}) \tag{3.20}$$

Thus when the null hypothesis is true, $G = F$ for all y. Notice that we have assumed that the parameter α is equal to zero. We may do this without loss of generality, since α does not contribute to correlation between Y and $\vec{X}$.

THEOREM 3.7. *The difference $D_1 - D_2$ is a convex funtion of $\vec{\beta}$.*

PROOF: Let us denote the difference by RD, for reduction in dispersion. Then we have that

$$RD = \int \varphi[G(y)]ydG(y) - \int \varphi[F(\epsilon)]\epsilon dF(\epsilon)$$

where $G(y)$ was given in (3.20). When $\vec{\beta} = \vec{0}$, we have, by the comment in (3.20), that $RD = 0$. To show that RD is a convex function, it will be shown that the matrix of second partials with respect to $\vec{\beta}$ is positive definite for all $\vec{\beta}$, and that the difference has an extreme value for $\vec{\beta}$ equal to $\vec{0}$. To do this, notice that D_2 does not depend on $\vec{\beta}$, so that we need not consider it when we calculate our derivatives. If we differentiate RD with respect to $\vec{\beta}$, we obtain

$$\begin{aligned} \frac{\partial RD}{\partial \vec{\beta}} &= -\int \varphi[G(y)]yf(y - \vec{\beta}'\vec{x})\vec{x}\ dy\ dM \\ &= \int \varphi'[G(y)]yf(y - \vec{\beta}'\vec{u})f(y - \vec{\beta}'\vec{x})\vec{u}\ dy\ dM(\vec{u})\ dM(\vec{x}) \end{aligned} \tag{3.21}$$

Since the random vector $\vec{X}$ has zero expectation zero, we have that the above expression is zero when $\vec{\beta} = \vec{0}$. Thus RD may have an extrema at this point.

We may use the definition of G to simplify expression (3.21) as follows: Write the last integral as

$$\int \varphi'[G(y)]yf(y-\vec{\beta}'\vec{x})\int f(y-\vec{\beta}'\vec{u})\ dy\ dM(\vec{u})\ dM(\vec{x})$$

The integral taken with respect to y may be written as

$$\int \varphi'[G(y)]yf(y-\vec{\beta}'\vec{u})g(y)\ dy$$

Now integrate by parts with respect to y. We obtain the following form:

$$-\int \varphi[G(y)]f(y-\vec{\beta}'\vec{u})\ dy - \int \varphi'[G(y)]f(y-\vec{\beta}'\vec{u})\ dy \tag{3.22}$$

Substituting (3.22) into (3.21) yields

$$\frac{\partial RD}{\partial \vec{\beta}} = \int \varphi[G(y)]f(y-\vec{\beta}'\vec{u})\vec{u}\ dy\ dM \tag{3.23}$$

If we differentiate (3.23) with respect to $\vec{\beta}$, we obtain for the matrix of second partials

$$-\int \varphi[G(y)]f'(y-\vec{\beta}'\vec{u})\vec{u}\vec{u}'\ dy\ dM(\vec{u})$$
$$-\int \varphi'[G(y)]f(y-\vec{\beta}\vec{u})f(y-\vec{\beta}'\vec{x})\vec{u}\vec{x}'\ dy\ dM(\vec{u})dM(\vec{x})$$

This may be simplified also. Again we integrate with respect to y, this time using

$$\int \varphi[G(y)]f'(y-\vec{\beta}'\vec{u})\ dy$$

We obtain

$$\int \varphi'[G(y)]g(y)f(y-\vec{\beta}'\vec{u})\ dy$$

as the result. Substituting into into the previous expression the following form for the matrix of second partial derivatives is obtained.

$$+\int \varphi'[G(y)]f(y-\vec{\beta}'\vec{u})f(y-\vec{\beta}'\vec{x})\vec{u}\vec{u}'\ dy\ dM(\vec{u})\ dM(\vec{x})$$

$$-\int \varphi'[G(y)]f(y-\vec{\beta}'\vec{x})f(y-\vec{\beta}'\vec{u})\vec{u}\vec{x}'\ dy\ dM(\vec{x})\ dM(\vec{u})$$

(3.24)

$$=\int \varphi'[G(y)]f(y-\vec{\beta}'\vec{x})f(y-\vec{\beta}'\vec{u})\vec{u}(\vec{u}-\vec{x})\ dy\ dM(\vec{x})\ dM(\vec{u})$$

Now we may rewrite (3.24), using the following.(Remember that (3.24) is the matrix of second partial derivatives of RD):

$$\int \varphi'[G(y)]f(y-\vec{\beta}'\vec{x})f(y-\vec{\beta}'\vec{x})(\vec{u}-\vec{x})(\vec{u}\vec{x}')\ dy\ dM(\vec{u})\ dM(\vec{x})\ dM(\vec{u})$$

$$=\int \varphi'[G(y)]f(y-\vec{\beta}'\vec{x})f(y-\vec{\beta}'\vec{u})\vec{u}'(\vec{u}-\vec{x})'dydM(\vec{x})dM(\vec{u})$$

$$-\int \varphi'[G(y)]f(y-\vec{\beta}'\vec{x})f(y-\vec{\beta}'\vec{u})\vec{x}(\vec{u}-\vec{x})'dydM(\vec{x})dM(\vec{u})$$

$$=\int \varphi'[G(y)]f(y-\vec{\beta}'\vec{x})f(y-\vec{\beta}'\vec{u})\vec{u}(\vec{u}-\vec{x})'dydM(\vec{x})dM(\vec{u})$$

$$+\int \varphi'[G(y)]f(y-\vec{\beta}'\vec{x})f(y-\vec{\beta}'\vec{x})\vec{x}(\vec{x}-\vec{u})'dydM(\vec{x})dM(\vec{x})$$

(3.25)

$$=2\frac{\partial^2 RD}{\partial\vec{\beta}\partial\vec{\beta}'}$$

Notice that (3.25) gives that the matrix of second partials is positive definite no matter what the value of $\vec{\beta}$. To see this, notice that we may write the last integral as a sum of two by splitting the region of integration as follows.

$$\int_A \varphi'[G(y)]f(y-\vec{\beta}'\vec{u})f(y-\vec{\beta}'\vec{x})(\vec{u}-\vec{x})(\vec{u}-\vec{x})'dydM(\vec{x})dM(\vec{u})$$

$$+\int_B \varphi'[G(y)]f(y-\vec{\beta}'\vec{u})f(y-\vec{\beta}'\vec{x})(\vec{u}-\vec{x})(\vec{u}-\vec{x})'dydM(\vec{x})dM(\vec{u})$$

$$=\int_B \varphi'[G(y)]f(y-\vec{\beta}'\vec{u})f(y-\vec{\beta}'\vec{x})\vec{u}-\vec{x})(\vec{u}-\vec{x})'dydM(\vec{x})dydM(\vec{u})$$

where A denotes the set where $\vec{u} = \vec{x}$ and B denotes the complement of this set. Now,the last equality follows since the distribution function M is continuous. Now φ' is positive since φ is increasing, the densities are positive, and, if $\vec{a}$ is a vector in p-space, we have

$$\vec{a}'(\vec{u} - \vec{x})(\vec{u} - \vec{x})'\vec{a}$$

$$= ((\vec{u} - \vec{x})'\vec{a})^2 > 0$$

Since the matrix of second partials is positive definite, the function RD is convex as a function of $\vec{\beta}$, and we are done.

We also comment that the above proof shows that RD takes on its minimum value when the null hypothesis is true. We have then the following result about $\overline{R}_{1\varphi}$ and $\overline{R}_{2\varphi}$.

COROLLARY 3.7. *$\overline{R}_{1\varphi}$ and $\overline{R}_{2\varphi} = 0$ if and only if $\vec{\beta} = \vec{0}$, which is true if and only if Y and $\vec{X}$ are independent.*

The proof of the corollary follows directly from the previous theorem. It is well known that the corresponding theorem for the classical multiple correlation coefficent holds in this model as well. Thus we see that $\overline{R}_{1\varphi}$ and $\overline{R}_{2\varphi}$ simply generalize $\overline{R}^2$. Since rank procedures and estimates are more robust than their least squares counterparts are, the fact that we now have robust estimates of correlation should be pleasing.

3.7 Asymptotic Distributions of Statistics

In this section we briefly discuss the asymptotic distributions of $R_{1\varphi}$ and $R_{2\varphi}$. To do this we use the results stated concerning the asymptotic distribution of the pF_R and the convergence in probality of $\hat{\tau}$ to τ. Concerning pF_R, we recall that when the null hypothesis is true, the asymptotic distribution is central chi square with p degrees of freedom. Now, we may write $R_{1\varphi}$ as

$$
\begin{aligned}
R_{1\varphi} &= \frac{\frac{1}{n}RD}{\frac{1}{n}D(\vec{0})} \\
&= \frac{\frac{1}{n}RD}{\frac{\hat{\tau}}{2}}\frac{\frac{\hat{\tau}}{2}}{\frac{1}{n}D(\vec{0})} \qquad (3.26)
\end{aligned}
$$

We may write this as

$$
nR_{1\varphi} = (pF_R)\frac{\frac{\hat{\tau}}{2}}{\frac{1}{n}D(\vec{0})}
$$

Thus, when the null hypothesis is true, $R_{1\varphi}$converges in distribution to a constant times the central chi square distribution with p degrees of freedom. The constant is given by $\frac{\tau}{2D_1}$. We obtain the value of the constant from the behavior of the multiplier of the test statistic above. For the case of $R_{2\varphi}$, we note that we may write this in the following manner.

$$
nR_{2\varphi} = pF_R\frac{\frac{\hat{\tau}}{2}}{\frac{1}{n}[D(\vec{0}) - D(\vec{\beta}_R) + (n-p-1)\frac{\hat{\tau}}{2}]} \qquad (3.27)
$$

Under the null hypothesis we may conclude that the drop in dispersion converges in probability to zero, and so the asymptotic distribution of $R_{2\varphi}$ is the same as that of pF_R. For the case of contiguous alternatives, we may conclude that the above discussions may be repeated almost word for word, the only difference being that the central chi square distribution is now replaced by the noncentral chi square distribution, with p degrees of freedom, and noncentrality parameter

$$\delta_R^2 \tag{3.28}$$

which we have discussed previously. This follows from properties of contiguity.

Thus, we see that, just as the least squares test statistic and R^2 are linked in the sense that they share the same asymptotic distribution, so are pF_LS and $R_{2\varphi}$. This fact should appeal to those who are used to using least squares techniques, as they may use rank procedures in much the same manner, while having the extra feature of protection against outliers.

3.8 A Preliminary Result

In this section we state some results which will be used in the next Chapter when we discuss influence functions. Let $\vec{b}$ denote an arbitrary fixed vector in p dimensional space. Denote by G^* the cumulative distribution function of the random variable $Y - \vec{b}'\vec{X}$, and let H denote the joint distribution of Y and $\vec{X}$. Further, denote by G_n^* and H_n the empirical versions of these quantities. Then we may write

$$\frac{1}{n}D(\vec{b}) = \int \varphi[\frac{n}{n+1}G_n^*(y - \vec{b}'\vec{x})](y - \vec{b}'\vec{x})dH_n(\vec{x}, y)$$

As n tends to infinity, this tends in probability to

$$\int \varphi[G^*(y - \vec{b}'\vec{x})](y - \vec{b}'\vec{x})\ dH(\vec{x}, y) \tag{3.29}$$

Lemmas 3.1 and 3.2 are special cases of 3.28 corresponding to values of $\vec{b}$ which are $\vec{0}$ and $\vec{\beta}_0$, respectively. We may write 3.28 in the following manner.

$$\begin{aligned}&\int \varphi[G^*(y-\vec{b}'\vec{x})](y-\vec{b}'\vec{x})f(y-\vec{\beta}'\vec{x})\ dy\ dM(\vec{x})\\ &= \int \varphi[G^*(v-\vec{\delta}'\vec{x})](v-\vec{\delta}'\vec{x})f(v)dvdM(\vec{x})\end{aligned}$$

The last equality comes from the substition $v = y - \vec{\beta}'\vec{x}$. This means that, without loss of generality, when we are considering the functional forms of the quantities involved, we may assume that the true value of $\vec{\beta}$ is zero. Thus, without loss of generality, we may find influence functions of our quantities at the null hypothesis.

CHAPTER IV

INFLUENCE FUNCTONS

4.1 Introduction

In this chapter we obtain the influence functions of

i our estimate $\vec{\beta}_R$

ii The test statistic pF_R

iii both of $\overline{R}_{1\varphi}$ and $\overline{R}_{2\varphi}$

We assume that the correlation model holds, and as in the last section, we will also assume that the true value of $\vec{\beta}$ is zero. The influence functions will allow us to examine robustness properties of the quantities we have proposed, and indicate how improvements may be made. We shall show that the influence functions we obtain lead to the correct asymptotic distributions, and they will be compared to the influence functions of the analagous M estimates of Hampel et al. (1985). We will also obtain the influence function of the parameter γ, and use this to make an assertion about the asymptotic distribution of our estimate of γ. We will see that if we assume additionally that the distribution of the errors is symmetric about zero, and if we use Wilcoxon scores, the predicted asymptotic distribution obtained from the influence function agrees with the result given in Schweder (1975).

4.2 Definition of Influence Function

We now define the influence function, discuss some important implications and uses of it, and mention an alternate method of obtaining it. We will rely on the functional representations of D1 and D2 and RD which we obtained in Chapter III. We state the following definition for the vector case.

DEFINITION 4.1. *Let T denote a functional defined on the space of distribution functions. Let H denote a fixed distribution function in the domain of T, and let $\Delta_{(\vec{x},y)}$ denote the point mass at $\vec{x}$ and y. The influence function of T at H is defined by*

$$I\vec{F}(\vec{x}_0, y_0) = lim_{s \to 0^+} \frac{[T[(1-s)H + \Delta_{(\vec{x}_0, y_0)}] - T[H]]}{s}$$

provided the limit on the right exists.

The influence function is a useful heuristic tool. Intuitively, it measures how a statistic reacts to a point $(\vec{x}_0, y_0)$. Because of this, we hope that our statistics have influence functions which are reasonably continuous and bounded in their arguments. We may also use the influence function to obtain the asymptotic distribution of a statistic. Of course, we must be able to express the statistic we are interested in as a functional of distribution functions in order to apply the defintion, and we have done that for the quantities we are interested in the rank analysis case in Chapter III.

The influence function was introduced by Hampel (1974), who called it the influence curve. It is closely related to the idea of a differentiable statistical functional, which was first discussed by vonMises (1937), and also to the concept of Gâteux differentiability of a functional. These ideas are discussed further in Huber (1981) and also in Fernholtz (1983).

We now define the Gâteux derivative and discuss the relation it has to the influence function of a statistical functional.

DEFINITION 4.2. *Let T be a statistical functional defined in the space of distribution functions, and let H denote a fixed distributin function in the domain of T. We say that T is Gâteux differentiable at H if there exists a function Ψ_H, symmetric in its arguments, such that for any distribution function G such that $(1-s)H + sG$ is in the domain of T, we have that*

$$\lim_{s \to 0} \frac{T[(1-s)H + sG] - T[H]}{s} = \int \Psi_H \, dG$$

If we examine this definition, we see that we may interpret the limit on the left as the partial derivitive of T[(1-s)H+sG] with respect to s at s = 0. Notice that we may think of this as the directional derivative of T in the direction of G. Also, if we set G = H in the definition we obtain the following.

$$\int \Psi_H \, dG = 0 \tag{4.1}$$

This allows us to write the right hand side in the previous definition as

$$\int \Psi_H \, d(G - H) \tag{4.2}$$

From (4.2) we see that Ψ_H may be considered as the first kernel in the vonMises expansion of T –Hampel et al. (1985), or Fernholtz (1983). We have introduced the Gâteux derivative for the following reason; if we set G $=\Delta_{(\vec{x}_0,y_0)}$ in the definition, we find that

$$\Psi_H = \vec{IF}(\vec{x}_0, y_0) \tag{4.3}$$

Huber (1977) also makes this observation. With this in mind, we see that (4.1) yields the following.

$$E[\vec{IF}(\vec{x}_0, y_0)] = \vec{0} \tag{4.4}$$

We will now outline how the influence function may be used to obtain the asymptotic distribution of a statistic. Assume for simplicity that we are in the one dimensional case. Let $T_n = T(H_n)$ denote the value of the statistic when applied to the empirical distribution function. The idea of vonMises (1937) was that, if G is in some sense close to the true distribution function H, then by expanding $T[(1-s)H + sG]$ in a Taylor series in s yields

$$\begin{aligned} T(G) &= T(H) + \int_{-\infty}^{\infty} \Psi_H(x)\ d(G-H)(x) + R_1 \\ &= T(H) + \int_{-\infty}^{\infty} \Psi_H(x)\ d(G)(x) + R_1 \end{aligned} \tag{4.5}$$

Here R_1 denotes a remainder term. If we set $G = H_n$, and use the fact that Ψ_H is the influence function, we have, upon rearranging (4.5),

$$\sqrt{n}(T(H_n) - T(H)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} IF(x_i) + \sqrt{n}R_1 \tag{4.6}$$

If it may be verified that the last term on the right hand side of (4.6) goes to zero in probability as n increases, then we may use the Central Limit Theorem and Slutsky's Theorem to conclude that

$$\sqrt{n}(T(H_n) - T(H)) \rightarrow n(0, \sigma^2(T, H)) \tag{4.7}$$

where

$$\sigma^2(T, H) = Var(IF) = \int IF^2(x)\, dH(x)$$

Typically, the verification that the remainder term goes to zero at a sufficently fast rate is a formidable task. The book by Fernholtz (1983) and the paper by Fillipova (1961) both discuss the regularity conditions which need to hold in order for these results to be rigorous. In addition, further discussion may be found in Hampel (1974), Hampel et al. (1985), and Huber (1981). Typically we use the influence function to conjecture the distributional result, and use other methods to verify that the assertion is correct.

4.3 Influence Function of Rank Estimate

We are now in a position to find the influence functions of the quantities we have proposed. The method we will follow will be to find an expression for the Gâteux derivative of the appropriate functional, and obtain the influence function from that. As we commented at the end of Chapter III, we may, without loss of generality, assume that the true value of $\vec{\beta}$ is zero, so that Y and $\vec{X}$ are independent. This is equivalent to obtaining the results at the null hypothesis. If we denote the joint distribution function of Y and $\vec{X}$ in this situation by H_0, then we have that the parent distribution factors into the product of its marginal distributions, which we will denote by F_0 and M_0. We now state the assumptions we need for the work which will come, as well as some notation.

Let $\vec{T}$ denote the value of the functional estimate of $\vec{\beta}$

G^* will denote the distribution function function of $Y - \vec{b}'\vec{x}$ where $\vec{b} \in R^p$

We make the following assumptions

I1 φ is twice differentiable on (0,1)

I2 $\ell(\vec{b}) = \int \varphi[G^*(y - \vec{b}'\vec{x})](y - \vec{b}'\vec{x})\, dH(\vec{x}, y)$ is finite for $\vec{b} \in R^p$

I3 $\frac{\partial \ell}{\partial \vec{b}}$ and $\frac{\partial^2 \ell}{\partial \vec{b} \partial \vec{b}'}$ both exist for $\vec{b} \in R^p$

I4 We will denote the joint distribution of Y and $\vec{X}$ when the null distribution holds by $H_0(\vec{x}, y)$

We also let G denote an abitrary distribution function on p+1 dimensional space. In the next theorem, we use the fact that $T(H)$ may be defined as the solution to

$$\int \varphi[G^*(y - \vec{x}'\vec{T}(H))]\vec{x}dH(\vec{x}, y) = \vec{0} \tag{4.8}$$

This is the functional form of the equation

$$\sum_{i=1}^{n} a[R(y_i - \vec{\beta}_R'\vec{x}_i) = \vec{0}$$

THEOREM 4.1. *Under and I1 and I2, the influence function of $T(H)$ is*

$$\vec{IF}(\vec{x}_0, y_0 : \vec{T}, H) = \frac{1}{\gamma}\varphi[F_0(y_0)]\Sigma^{-1}\vec{x}_0$$

PROOF: Set H_s into (4.8) to obtain

$$(1 - s)\int \varphi[G_s^*(y - \vec{T}'\vec{x})]\vec{x}\, dH_0(\vec{x}, y) + s\int \varphi[G_s^*(y - \vec{T}'\vec{x})]\vec{x}\, dG(\vec{x}, y) = \vec{0}$$

We will need the derivative of G_s^* with respect to s at s=0. This is

$$-F_0(y) - f_0(y)\vec{x}'\vec{T}^{\cdot} + G_Y(y)$$

where $\vec{T}^{\cdot}$ represents the Gâteux derivative of $\vec{T}$ and G_Y denotes the marginal distribution of Y obtained from G. Using this we obtain for the Gâteux derivative of $\vec{T}$

$$\frac{1}{\gamma}\Sigma^{-1}\int \varphi[F_0(y)]\vec{x}\, dG(\vec{x},y) \tag{4.9}$$

From (4.9) we may read the influence function of $\vec{T}$.

Notice that the result of this theorem is a multivariate generalization of a result in Hettmansperger (1984), which he obtains in a slightly different manner, and for the case of Wilcoxon scores.

As long as the score function φ is bounded, the influence function of $\vec{T}$ is a bounded function of Y. This supports the idea that scores should always be generated by bounded score functions. This implies that the effects of outlying values of Y have a limited effect of the rank estimate of $\vec{\beta}$. Unfortunately, we see that the influence function of $\vec{T}$ is an unbounded function of $\vec{X}$, and so the estimate may still be adversely effected by outliers in $\vec{X}$ space. This is an improvement over the case of least squares, where the influence function of the estimate of that estimate is

$$I\vec{F}(\vec{x}_0, y_0) = y_0\Sigma^{-1}\vec{x}_0 \tag{4.10}$$

and is clearly an unbounded function of both Y and $\vec{X}$. For estimates which have influence functions which are bounded in both variables, see Hampel et al. (1985), who discuss M-estimates, and Tableman (1988) who has introduced rank estimates of $\vec{\beta}$ with bounded influence. In both of these cases, however, it was assumed that F_0 is symmetric, which is not needed in our result. Even though we have obtained our influence function of $\vec{\beta}$ in a purely mechanical manner, we will now show that, using previous results, we may show that it gives a rigorous result, in the sense that if we were to expand $\vec{\beta}_R$ in a vonMises expansion of length one, with the influence function as kernel, the remainder term would be of order $\frac{1}{\sqrt{n}}$. Before we do that, however we indicate how the influence function may be used to obtain the asymptotic distribution of the estimate. The next theorem is a direct result of the multivariate central limit theorem.

THEOREM 4.2. *When the true value of $\vec{\beta}$ is zero, we have*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \vec{IF}(\vec{x}_0, y_0 : \vec{T}(H), H_0)$$

has an asymptotic distribution which is multivariate normal with zero mean vector and covariance matrix $\frac{1}{\gamma^2}\Sigma^{-1}$

In the next theorems, we put the influence function on the solid ground hinted at earlier.

THEOREM 4.3. *Under the conditions listed in Theorem 4.2,*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} a[R(Y_i)]\vec{x}_i - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varphi[F_0(Y_i)]\vec{x}_i$$

goes to zero in probability as n tends to infinity.

PROOF: It is sufficient to prove this result in the case where the vectors $X_1, X_2, \ldots, X_n$ are nonstochastic. Further, since the result is stated for a sequence of vectors, it is sufficient to prove it for an arbitrary component, which we may take to be the first. Thus, we must show that as n goes to infinity

$$\frac{1}{\sqrt{n}}\sum a[R(Y)_i]x_{1i} - \frac{1}{\sqrt{n}}\sum \varphi[F_0(Y)_i]x_{1i} \tag{4.11}$$

converges to 0 in probability. If we condition on $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$, we may apply theorem 1.6(a) of Hájek and Šidák (1967).

LEMMA 4.4. *Let*

$$S(\vec{Y}) = \frac{1}{\sqrt{n}\gamma}\Sigma^{-1}\sum_{i=1}^{n} a[R(Y_i)]\vec{x}_i$$

Then under the conditions of Theorem 4.3,

$$\vec{S}(\vec{Y}) - \frac{1}{\sqrt{n}}\sum \vec{I}F(\vec{x}_i, y_i : \vec{T}(H_0), H_0) \tag{4.12}$$

goes to zero in probability as n goes to infinity.

PROOF: If we multiply (4.9) by the positive definite matrix $(\gamma\Sigma)^{-1}$, the result is immediate.

The next theorem uses the results just obtained to show that the error term in the vonMises expansion of $\vec{\beta}_R$ is of the correct order of magnitude.

THEOREM 4.5. *Under the assumptions of the previous theorem,*

$$\sqrt{n}\vec{\beta}_R - \frac{1}{\sqrt{n}}\sum \vec{IF}(\vec{x}, y : \vec{T}(H_0), H_0) \tag{4.13}$$

goes to zero probability as n goes to infinity.

PROOF: Again, let us condition on $X_1, X_2, \ldots, X_n$. Then the regularity conditions needed by McKean and Hettmansperger (1976) are met, so that

$$\sqrt{n}\vec{\beta}_R - \vec{S}(\vec{Y}) \tag{4.14}$$

goes to zero in probability. Using (4.13) along with (4.11) and Slutsky's Theorem yields (4.12). Thus we are done.

These last two theorems and the corollary imply that

$$\sqrt{n}\vec{\beta}_R \rightarrow MVN_p[\vec{0}, (\gamma^2\Sigma)^{-1}] \tag{4.15}$$

We have seen this result in Chapter II stated for an arbitrary value of $\vec{\beta}$. We may think of the result as obtained here as applying to the null hypothesis.

If we consider a sequence of contiguous alternatives, say $H_n = \frac{1}{\sqrt{n}}\vec{d}$, where $\vec{d}$ is a fixed nonzero p–dimensional vector, we may obtain the asymptotic distribution of $\sqrt{n}\vec{\beta}_R$ in this case as well. Since (4.14) and (4.15) hold under the null hypothesis, they continue to hold under the sequence H_n as well. We know that in this case

$$\sqrt{n}\vec{\beta}_R \rightarrow MVN_p[\gamma\Sigma\vec{d}, (\gamma^2\Sigma)^{-1}] \tag{4.16}$$

in distribution, and so once again using Slutsky's Theoremn, (4.16) holds with $\sqrt{n}\vec{\beta}_R$ replaced by $\frac{1}{\sqrt{\gamma}}\sum_{i=1}^{n} I\vec{F}(\vec{x}_i, y_i; \vec{T}(H_0), H_0)$

4.4 Influence Function of Rank test statistic

In this section we obtain the influence function of the test statistic for testing the hypothesis that the true value of $\vec{\beta}$ is zero. We will proceed as we did in the previous section, where the influence function of the estimate of $\vec{\beta}$ was obtained. In doing so, we shall see the reason that $R_{2\varphi}$ is more robust than $R_{1\varphi}$. We recall that the test statistic is given by

$$pF_R = \frac{[D(\vec{0}) - D(\vec{\beta}_R)]}{\frac{\tau}{2}} \tag{4.17}$$

Now in Chapter III, we saw that in a certain sense, the numerator of the test statistic may be represented as a functional which involves the marginal distribution of Y, the score function φ, and the distribution function of $Y - \vec{b}'\vec{X}$. If we wish to examine the behavior of this functional by means of an influence function, we need to consider this expression as a function of the distributions H,F and G^*. With this in mind, we define, for an arbitrary distribution H

$$\begin{aligned} D_1 &= \int_{-\infty}^{\infty} \varphi[F(y)]y \; dF(y) \\ D_2 &= \int_{-\infty}^{\infty} \varphi[G^*(y - \vec{x}'\vec{T}(H))](y - \vec{x}'\vec{T}(H)) \; dH(\vec{x}, y) \\ RD &= D_1 - D_2 \end{aligned}$$

where $\vec{T}(H)$ denotes the value the estimate of $\vec{\beta}$. In the case that H is the model distribution H_0, we have that $\vec{T} = \vec{0}$, $G^* = F$, and so RD is equal to zero. The functional whose influence function we want is

$$\frac{RD}{\frac{\tau}{2}} \tag{4.18}$$

To do this we proceed as follows:

1. Consider the first and second derivatives of RD to obtain the influence function of RD.
2. Divide the influence function of RD by $\frac{\tau}{2}$ to obtain the influence function of the test statistic
3. Justify treating the denominator of the test statistic as a constant rather than a functional in it's own right in parts 1 and 2.

Let G denote an arbitrary distribution function, and as we did in the previous section, form the contaminating distribution by $H_s = (1-s)H_0 + sG$. When we insert this into the functional RD,we see that RD becomes a function of s, as do F and G^*. By considering derivatives of this functional with respect to s at $s = 0$, we obtain the following theorem.

THEOREM 4.4. *In the correlation model, under the assumptions of the previous theorems, we have*

a $RD(\vec{0}) = 0$

b $\left.\dfrac{\partial RD(s)}{\partial s}\right|_{(s=0)} = \vec{0}$

c $\left.\dfrac{\partial^2 RD(s)}{\partial s^2}\right|_{(s=0)} = \vec{A}(\varphi, G)' \Sigma^{-1} \vec{A}(\varphi, G)$

where $\vec{A}(\varphi, G)$ *is equal to* $\frac{1}{\tau} \int \varphi[F_0(y)]\vec{x}\, dG(\vec{x}, y)$

PROOF: Part a follows from the discussion in Chapter III. To obtain b, we write

$$D_1(s) = (1-s)\int \varphi[F_s(y)]y dF_0(y) + s\int \varphi[F_s(y)]y dG(y)$$

and

$$D_2(s) = (1-s)\int \varphi[G_s^*(y-\vec{x}'\vec{T}_s)](y-\vec{x}'\vec{T}_s)dH_0(\vec{x},y)$$
$$+ s\int \varphi[G_s^*(y-\vec{x}'\vec{T}_s)](y-\vec{x}'\vec{T}_s)dG(\vec{x},y)$$

When $s=0$, $\vec{T}=\vec{0}$ so $G^*=F_0$ hence $RD(0)=0$. If we take the derivative with respect to s at s=0, we find that

$$\frac{\partial D_1}{\partial s}|_{(s=0)} = -\int \varphi[F_0(y)]ydF_0(y) - \int \varphi'[F_0(y)]F_0(y)ydF_0(y)$$
$$+\int \varphi'[F_0(y)]F(y)ydF_0(y) + \int \varphi[F_0(y)]ydF(y)$$
$$= \frac{\partial D_2}{\partial s}|_{(s=0)}$$

Thus we have part b of the theorem. We proceed with taking second derivatives and obtain, after some tedious algebra

$$\frac{\partial^2 RD}{\partial s^2}|_{(s=0)} = [-\int \varphi''[F_0(y)]yf_0^2(y)dF_0(y) - 2\gamma$$
$$-2\int \varphi[F_0(y)]yf_0'(y)dF_0(y)]\vec{T}^{\cdot\prime}\Sigma\vec{T}^{\cdot\prime}$$
$$+2\int \varphi'[F_0(y)]yf(y)(\vec{x}-\vec{u})'dM(\vec{u})dH_0(\vec{x},y)\vec{T}^{\cdot\prime}$$
$$+2\int \varphi'[F_0(y)]y\vec{x}'dG(\vec{x},y)\vec{T}^{\cdot\prime}$$
$$+\int \varphi[F_0(y)]\vec{x}'dG(\vec{x},y)\vec{T}^{\cdot\prime} \tag{4.19}$$

Now, integrating by parts with respect to Y, we have that

$$-\int \varphi''[F_0(y)]yf_0^2(y)dF_0(y) =$$
$$\int \varphi'[F_0(y)]f_0(y)dF_0(y) + s\int \varphi'[F_0(y)]yf_0'(y)dF_0(y)$$
$$= \gamma + 2\int \varphi'[F_0(y)]yf_0'(y)dF_0(y).$$

Furthermore, we obtain

$$2\int \varphi'[F_0(y)]yf(y)(\vec{x}-\vec{u})dM(\vec{u})dH_0$$
$$+2\int \varphi'[F_0(y)]yf_0(y)\vec{x}dG(\vec{x},y)$$
$$=2\int \varphi'[F_0(y)]yf(y)\vec{x}dM(\vec{u})dM_0(\vec{x})dF_0(y)$$
$$-2\int \varphi'[F_0(y)]yf(y)\vec{u}dM(\vec{u})dM_0(\vec{x})dF_0(y)$$
$$+2\int \varphi'[F_0(y)]yf_0(y)\vec{x}dG(\vec{x},y)$$
$$=\vec{0}-2\int \varphi'[F_0(y)]yf_0(y)\vec{u}dG(\vec{u},y)$$
$$+2\int \varphi'[F_0(y)]yf_0(y)\vec{x}dG(\vec{x},y)=\vec{0}$$

so that (4.18) reduces to

$$-\gamma\vec{T}^{\cdot\prime}\Sigma\vec{T}^{\cdot}$$

Here, $\vec{T}^{\cdot}$ denotes the Gâteux derivative of $\vec{T}$, which we obtained in the previous section. When we insert it we obtain

$$\frac{1}{\gamma}[\int \varphi[F_0(y)]\vec{x}dG(\vec{x},y)]'\Sigma^{-1}[\int \varphi[F_0(y)]\vec{x}dG(\vec{x},y)]$$

which yields part c of the theorem, and so we are done.

The result of the previous theorem will allow us to obtain the influence function of the test statistic. To do this we follow the procedure outlined in Hampel et al. (1985). We denote the functional representing the test statistic as W^2, and consider γ as a constant, so that $W^2 = 2\gamma RD$. Then the previous theorem says that the following is true :

$$\frac{\partial^2 W^2}{\partial s^2}|_{(s=0)} = 2\varphi^2[F_0(Y_0)]\vec{x}_0\Sigma^{-1}\vec{x}_0 \qquad (4.20)$$

To obtain the influence function of W, we proceed as follows. The Gâteux derivative of W is

$$
\begin{aligned}
& lim_{s\to 0}\frac{([W[(1-s)H_0+sG]-W[H_0])}{s} \\
&= [lim_{s\to 0}\frac{(W^2[(1-s)H_0+sG]-W^2[(H_0)])}{s}]^{\frac{1}{2}} \\
&= [lim_{s\to 0}\frac{W^2[(1-s)H_0+sG]}{s^2}]^{\frac{1}{2}} \\
&= \frac{1}{2}[\frac{\partial^2 W^2}{\partial s^2}|_{(s=0)}]^{\frac{1}{2}}
\end{aligned}
$$

so that the influence function of W is

$$IF(\vec{x}_0, y_0 : \vec{T}, H_0) = |\varphi|[F_0(y_0)]|\sqrt{\vec{x}_0'\Sigma^{-1}\vec{x}_0} \tag{4.21}$$

Basically, the trick here amounts to defining $\sqrt{pF_R}$ as the test statistic. The point to observe from the influence function is that, just as the influence function of $\vec{\beta}_R$, (4.21) is bounded in Y and unbounded in $\vec{X}$. The influence function of the least squares test statistic is $|y|\sqrt{\vec{x}'\Sigma^{-1}\vec{x}}$ which is unbounded in both Y and $\vec{X}$, similar to the influence function for the least squares estimate of $\vec{\beta}$. Thus, the value of the test statistic can be completely determined by one outlying data point, and this clearly is not desirable. There are test statistics for the hypothesis we are considering which have influence functions which are bounded functions of both Y and $\vec{X}$, but these test statistics have asymptotic distributions which are not chi square in form, except in certain special cases. These are based on M–estimation and are discussed in Hampel et al. (1985)

We now show that when we found the influence function of the test statistic, we were justified in treating γ as a constant instead of a functional.

We proceed as follows.

$$W^2(s) = 2\gamma(s)RD(s)$$
$$\frac{\partial W^2}{\partial s} = 2\gamma' RD(s) + 2\gamma RD'(s)$$
$$\frac{\partial^2 W^2}{\partial s^2} = 2\gamma'' RD + 4\gamma' RD' + 2\gamma RD''$$

Since, when $s = 0$, both RD and the first derivative of RD are 0, we see that the second derivative of RD involves only the parameter γ.

We now indicate how the influence function of W^2 may be used to obtain the correct asymptotic distribution of the test statistic. Recall that the first nonvanishing derivative of W^2 was the second. This means that when we expand W^2 in a vonMises expansion, we need to go to two terms, as follows

$$\begin{aligned} W^2(G) &= W^2(H_0) + \frac{\partial W^2}{\partial s}|_{(s=0)} + \\ &\frac{1}{2}\frac{\partial^2 W^2}{\partial s^2}|_{(s=0)} + R_2 \\ &= \left(\int \varphi[F_0(y)]\vec{x}dG(\vec{x},y)\right)' \\ &\Sigma^{-1}\left(\int \varphi[F_0(y)]\vec{x}dG(\vec{x},y)\right) + R_2 \end{aligned}$$

When we set $G = H_n$, the empirical distribution function, we obtain

$$\begin{aligned} nW_n^2 &= \\ &\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varphi[F_0(y_i)]\vec{x}_i\right)'\Sigma^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varphi[F_0(y_i)]\vec{x}_i\right) \\ &+ nR_2 \end{aligned}$$

We may easily show that the first term on the right hand side converges in distribution to a central chi square distribution with p degrees of freedom, and we know that this is the correct result for the test statistic. In a manner similar to that used for the estimate of $\vec{\beta}$, we may show that the error term converges to zero in probability as n tends to infinity, and so the diffence between the test statistic and the leading term on the right hand side goes to zero as n tends to infinity. This gives the asymptotic distribution at the null hypothesis and similar techniques give the correct for a sequence of contiguous alternatives.

4.5 Comparison of Coefficients

In this section we briefly indicate why the statistic $R_{2\varphi}$ is more robust than $R_{1\varphi}$. Recall that the denominator of $R_{2\varphi}$ is $\frac{1}{n}D(\vec{0})$. From the last section, we see that the influence function of this is

$$\int_{y}^{\infty} \varphi'[F_0(y)]ydF_0(y)$$

$$-\int \varphi[F_0(y)]ydF_0(y) - \int \varphi'[F_0(y)]yF_0(y)dF_0(y) + \int_{y}^{\infty} \varphi'[F_0(y)]ydF_0(y)$$

which is an unbounded function of Y.Thus, the numerator of $R_{1\varphi}$ is easily influencd by values of $\vec{x}$, and the denominator by values of Y. This makes the fraction unstable, and is the reason $R_{1\varphi}$ is not robust. On the other hand, $R_{2\varphi}$ has an influence function which is bounded in Y, and so is more robust than $R_{1\varphi}$. In the next section we will obtain the influence function of τ.

4.6 Influence Function of Gamma

In this section we obtain the influence function of γ, and use it to make a conjecture about the asymptotic distribution of the estimate of γ. This conjecture is correct for the case where the error distribution is symmetric, and Wilcoxon scores are employed. This comes from a result of Schweder(1975).

THEOREM 4.5. *The influence function of γ is*

$$IF(y_0 : \gamma, H_0) = -2\gamma + 2\varphi'[F_0(Y_0)]f_0(Y_0) - \int_{-\infty}^{\infty} \varphi''[F_0(x)]F_0(x)f^2(x)dx + \int_{y_0}^{\infty} \varphi''[F_0(x)]f_0^2(x)dx$$

PROOF: The parameter γ is given by

$$\gamma = \int \varphi'[F_0(y)]f_0(y)dF_0(y)$$

We think of F here not as the distribution of the errors, but as the distribution of the residuals $Y - \vec{T}'\vec{X}$. When $\vec{T}$ is equal to the true value of $\vec{\beta}$, then F is the error distribution. If we set

$$H_s = (1 - s)H_0(\vec{x}, y) + sG(\vec{x}, y)$$

then instead of of F_0 we must use

$$G_s^*(y - \vec{T}_s'\vec{X}) = (1 - s)G_0^*(y - \vec{T}_s'\vec{x}) + \int F_{Y|X}(y + (\vec{u} - \vec{x}'\vec{T}_s)dM(\vec{u})$$

where G_0^* denotes the distribution of $Y - \vec{T}'\vec{x}$ obtained from F_0, and $F_{Y|X}$ denotes the conditional distribution of Y given $\vec{X}$ computed from G. Notice that the density of G_s^* is given by an expression which looks like the one above with g_0^* and $f_{Y|X}$ replacing G_0^* and $F_{Y|X}$, respectively. Further when $s = 0$, we have that $G^* = F^*0$ and $g^* = f_0$. With this in mind we have

$$\gamma(s) = (1-s)\int \varphi'[G_s^*(y - \vec{T}_s'\vec{x})g_0^*(y - \vec{T}_s')\vec{x}dH_0$$

$$+ s\int \varphi'[G_s^*(y - \vec{T}_s'\vec{x})g_s^*(y - \vec{T}_s\vec{x})dG$$

$$\frac{\partial\gamma}{\partial s} = -\int \varphi'[G_s^*(y - \vec{T}_s'\vec{x})g_s^*(y - \vec{T}_s'\vec{x})dH_0$$

$$+ \int \varphi''[G_s^*(y - \vec{T}_s\vec{x})g_0^*(y - \vec{T}_s'\vec{x})\frac{\partial G_s^*}{\partial s}dH_0$$

$$+ \int \varphi'[G_s^*(y - \vec{T}_s'\vec{x})\frac{\partial g_s^*}{\partial s}dH_0$$

$$+ \int \varphi'[G_s^*(y - \vec{T}_s'\vec{x})g_s^*(y - \vec{T}_s'\vec{x})dG$$

$+\, s$ unneeded terms.

If we set s equal to 0 we obtain

$$-\int \varphi'[F_0(y)]f_0(y)dH_0 - \int \varphi''[F_0(y)]f_0(y)F_0(y)dH_0$$

$$-\int \varphi''[F_0(y)]f_0(y)\vec{x}'dH_0\vec{T}^{(')}$$

$$+\int \varphi''[F_0(y)]f(y)dH_0 + \int \varphi[F_0(y)]f_0(y)dG$$

$$= -\int \varphi'[F_0(y)]f_0(y)dF_0(y) - \int \varphi''[F_0(y)]F_0(y)f_0(y)dF_(y)$$

$$-\int \varphi'[F_0(y)]f_0(y)dF_0(y) + \int \varphi'[F_0(y)]f_0(y)dF(y)$$

$$+\int \varphi''[F_0(y)]F(y)f_0(y)dF_0(y) + \int \varphi'[F_0(y)]f_0(y)dF(y)$$

If we set $G = \Delta$, and interpret f above as dG, we obtain the influence function as stated in the theorem.

We are now in a position to state the conjecture concerning the asymptotic distribution of our estimate of γ.

CLAIM. *The asymptotic distribution of* $\sqrt{n}(\hat{\gamma}-\gamma)$ *is* $n(0,\sigma^2(\gamma,F_0))$ *where*

$$\begin{aligned}
\sigma^2(\gamma,F_0) &= Var(IF(Y;\gamma,F_0)) \\
&= 4[\int(\varphi'[F_0(t)])^2 f_0^3(t)dt-\gamma^2] \\
&+ (\int\varphi''[F_0(x)]f_0^2(x)dx)(\int\varphi''[F_0(x)]F_0(x)f_0^2(x)dx) \\
&- (\int\varphi''[F_0(x)]F_0(x)f_0^2(x)dx)^2 \\
&+ 4\gamma\int\varphi''[F_0(x)]f_0^2(x)dx \\
&- 4\iint\varphi'[F_(x)]\varphi''[F_0(y)]f_0^2(x)f_0^2(y)I[X>Y]dxdy \\
&- 4\gamma\int\varphi''[F_0(x)]F_0(x)f_0^2(x)dx.
\end{aligned}$$

This reduces to the expression given by Schweder (1975) when φ is the Wilcoxon score function $\sqrt{12}(u-\frac{1}{2})$. The general form of the asymptotic variance given by Schweder (1975) is

$$\begin{aligned}
\sigma_s^2 = 2(&\int\varphi''[F_0(x)]f_0^2(x)dx)(\int\varphi''[F_0(y)]F_0(y)f_0^2(y)dy) \\
&- (\int\varphi''[F_0(x)]F_0(x)f_0^2(x)dx)^2 \\
&- 2\iint\varphi''[F_0(x)]f_0^2(x)\varphi''[F_0(y)]F_0(y)f_0^2(y)I[y>x]dydx \\
&- 4\gamma\int\varphi[F_0(x)]f_0^2(x)dx \\
&+ 4[\int\varphi'^2[F_0(x)]f_0^3(x)dx-\gamma^2]
\end{aligned}$$

For the case of Wilcxon scores, both of these expressions reduce to

$$4[12\int_{-\infty}^{\infty}f_0^3(x)dx-\gamma^2]$$

CHAPTER V

ALTERNATIVE PROCEDURES

5.1 Introduction

In this chapter we turn attention to some procedures which could be used to obtain competitors to $R_{1\varphi}$ and $R_{2\varphi}$. When possible, we consider estimation and testing procedures based on other methods, and compare them with ours. The following books will be referenced ; the book by Hampel et al. (1985),and the book by Puri and Sen (1985). The methods discussed will be ones which are generally considered to be robust, and will consist of M–estimate based procedures and rank-based procedures.

5.2 Notations and Assumptions for M–setimates

The first procedures we consider are those discussed in Hampel et al. (1985) and are based on M– estimates. These authors use the correlation modelof Chapter III and study the problem of estimation of $\vec{\beta}_0$, and the problem of testing H_0. They make the additional assumption that the error distribution is symmetric about 0.

In the estimation problem, they propose to estimate $\vec{\beta}$ by choosing the vector in R^p which minimizes the function

$$\sum \tau(\vec{x}_i, y_i - \vec{x}_i'\vec{T}_n) \tag{5.1}$$

or, equivalently, the vector that solves the system

$$\sum \nu(\vec{x}_i, y_i - \vec{x}_i'\vec{T}_n)\vec{x}_i = \vec{0} \tag{5.2}$$

where

$$\nu(\vec{x}, r) = \frac{\partial \tau}{\partial r}$$

We have the following conditions on the function ν.

H1(i) $\nu(\vec{x}, \bullet)$ is continuous on $R \backslash C(\vec{x}, \nu)$ for all $\vec{x} \in R^p$, where $C(\vec{x}, \nu)$ is a finite set. In each point of $C(\vec{x}, \nu)$, $\nu(\vec{x}, \bullet)$ has finite left and right hand limits.

H1(ii) $\nu(\vec{x}, \bullet)$ is odd,and $\nu(\vec{x}, r) \geq 0$ for all $\vec{x} \in R^p, r \in R^+$

H2 For all $\vec{x}$ the set $D(\vec{x}, \nu)$ of points in which $\nu(\vec{x}, \bullet)$ is continuous but in which $\nu'(\vec{x}, r)$ is not defined or continuous is finite. Here $\nu'(\vec{x}, r) = \frac{\partial \nu}{\partial r}$.

In addition, we assume that each of the matrices

$$M = \int \nu'(\vec{x}, r)\vec{x}\vec{x}' d\Phi(r) dM(\vec{x})$$
$$Q = \int \nu^2(\vec{x}, r)\vec{x}\vec{x}' d\Phi(r) dM(\vec{x})$$

exists and is nonsingular.

The functional form of (5.1) is

$$\int \tau(\vec{x}, y - \vec{x}'\vec{T}(H))\, dH(\vec{x}, y) \tag{5.3}$$

and the functional form of (5.2) is

$$\int \nu(\vec{x}, y - \vec{x}'\vec{T})\vec{x}\, dH(\vec{x}, y) = \vec{0} \tag{5.4}$$

Here H represents an arbitrary distribution function on R^{p+1}. The influence function of the estimator of $\vec{\beta}$ is found to be

$$I\vec{F}(\vec{x}_0, y_0; \vec{T}_M, H_0) = \nu(\vec{x}_0, y_0)M^{-1}\vec{x}_0 \tag{5.5}$$

when the null hypothesis is true. This suggests that the covariance matrix of the asymptotic distribution is

$$M^{-1}QM^{-1} \tag{5.6}$$

For the special case that $\tau(\vec{x}, y) = \rho(y)$ we have that the influence function of $\hat{\beta}_M$ is

$$\frac{\rho'(y)}{\int \rho''(y)\, dF(y)} \Sigma^{-1}\vec{x}_0 \tag{5.7}$$

and the asymptotic covariance matrix is

$$\frac{\int [\rho(y)]^2\, dF(y)}{[\int \rho''(y)\, dF(y)]^2} \Sigma^{-1} \tag{5.8}$$

These are very similar to the corresponding quantities in the rank estimate case. This causes us to wonder if rank and M estimates may be equivalent. It was shown by Jureckova (1977) that if φ and ρ are related by

$$\varphi(t) = a\rho(F_0^{-1}(t)) + b$$

almost everywhere t, for some real b, a, then $\vec{\beta}_R$ and $\vec{\beta}_M$ are indeed asymptotically equivalent. We also note that least squares estimates are special cases of M–estimates, obtained when we take $\rho(y) = y^2$. The M–estimates obtained here are consistent and asymptotically normal with asymptotic variance given by (5.6) when the null hypothesis is true.

5.3 Tests Based on M–Estimates

In this section we turn to the problem of hypothesis testing in the correlation model, using methods other that our rank estimates. The first method discusssed will be based on M–estimates. We will concentrate on test statistics which have the same type of asymptotic distribution as our rank statistic pF_R. The test statistic we will discuss was proposed by Schrader and Hettmanspeger (1980), and is given as a special case of a class of test statistics obtained in Hampel et al. (1985).

Denote by $\Gamma(\vec{b})$ the quantity

$$\Gamma(\vec{b}) = \sum_{i=1}^{n} \rho(y_i - \vec{b}'\vec{x}_i) \tag{5.9}$$

Then the test statistic we are considering is

$$S_n = [\frac{2}{np}[\Gamma(\vec{0}) - \Gamma(\vec{\beta}_M)]^{\frac{1}{2}} \tag{5.10}$$

Large values of the test statistic are significant. Schrader and Hettmansperger (1980) show that, under both the null hupothesis and a sequence of contiguous alternatives, S_n^2 has, asymptotically, a distribution which is chi-square in form, with p degrees of freedom. Thus, the test statistic has the same limiting behavior as our pF_R, and we are able to determine the asymptotic relative efficiency of the rank procedure relative to the M estimate procedure. The noncentrality parameter for the statistic given here is

$$\delta_M^2 = \frac{B^2}{A}\vec{d}'\Sigma\vec{d}$$

where A and B are defined as follows

$$A = \int (\rho'(y)^2)\, dF(y)$$

$$B = \int \rho''(y)\, dF(y)$$

The asymptotic relative efficiency of the M procedure to the R procedure is seen to be $\frac{\tau^2 A}{B^2}$, just as in the regression case. The influence fucntion of the test statistic is very similar to the one for pF_R and is given in Hampel as

$$|\rho(y)|\sqrt{\frac{\vec{x}'\Sigma^{-1}\vec{x}}{\int \rho'(y)\; dF_0(y)}} \tag{5.11}$$

Notice that the function given here is bounded in Y but not in $\vec{X}$, just as was true for the influence function for our test statistic. Hampel discusses test statistics which have bounded influence, but these do not have asymptotic distributions which are chi square in form. For a further discussion see Chapter VII of Hampel et al. (1985).

5.4 Coefficents Based on M-estimates

In this section we consider the use of the M estimate based procedures discussed in the previous section in constructing quantities analagous to $R_{1\varphi}$ and $R_{2\varphi}$. We will construct the new quantities in the same manner as we did $R_{1\varphi}$ and $R_{2\varphi}$. Will denote the quantities by M_1 and M_2. Treating the function $\Gamma(\vec{b})$ as a dispersion function, we obtain the following expressions.

$$M_1 = \frac{\Gamma(\vec{0}) - \Gamma(\vec{\beta}_M)}{\Gamma(\vec{0})}$$
$$M_2 = \frac{2[\Gamma(\vec{0}) - \Gamma(\vec{\beta}_M)]}{2[\Gamma(\vec{0}) - \Gamma(\vec{\beta}_M)] + (n-p-1)}$$

Under the regularity conditions needed by the M estimates, each of M_1 and M_2 converge in probability to the following functionals. For notation, we set

$$RD\Gamma = \int [\tau(\vec{x}, y) - \tau(\vec{x}, y - \vec{\beta}'\vec{x})]dH$$

$$\overline{M}_1 = \frac{RD\Gamma}{\int \tau(\vec{x}, y)dH(\vec{x}, y)}$$
$$\overline{M}_2 = \frac{2RD\Gamma}{2RD\Gamma + 1}$$

where H is the distribution function of $\vec{X}$ and Y, and $\vec{T}(H)$ is the functional form of the estimate of $\vec{\beta}$. In the case that H is the model distribution, we have by results in Hampel (1985) that $\vec{T}(H) = \vec{0}$ and so both of the new measures are zero when the null hypothesis holds. Conversely, both of these are zero only when the null hypothethis holds. In this sense, they are similar to our quantities $R_{1\varphi}$ and $R_{2\varphi}$. However, they do not seem to be estimates of any easily interpretable parameters, even when the parent distribution is normal, as we shall see. We will concentrate on the case where ρ is Hubers' rho function, because of its simple form and wide use. The function is given by

$$\begin{aligned}\rho_c(y) &= \frac{1}{2}y^2 \qquad |y| \leq c \\ &= \frac{c}{2}|y| \qquad |y| > c\end{aligned}$$

If we use this choice of ρ and evaluate the integrals for $\overline{M}_1$ and $\overline{M}_2$, we obtain the following expressions.

$$\begin{aligned}\overline{M}_1 &= \frac{R^2 + 2\frac{I}{A}}{1 + 2\frac{I}{A}} \\ \overline{M}_2 &= \frac{R^2 + 2\frac{I}{A}}{1 + 2\frac{I}{A} + \frac{2}{A}}\end{aligned}$$

where

$$\begin{aligned}A &= \sigma^2 + \vec{\beta}'\Sigma\vec{\beta} \\ I &= \iint_B \{|y + \vec{\beta}'\vec{x}| - |y| - (y - \vec{\beta}'\vec{x})^2 + y^2\} f(y)dydM \\ B &= \{(\vec{x}, y) \in R^{p+1} | \quad |y| > c\}\end{aligned}$$

Note that although as $c \to \infty$ $\overline{M}_1$ tends to R^2, it does not seem that these represent any simple function of R^2. We would expect, however that these two quantities would inherit some of the robust properties of M estimates in general.

5.5 Alternate Rank-Based Procedures

In this section we consider the problem of testing the null hypothesis $H_0 : \vec{\beta} = \vec{0}$. We consider procedures which use rank methods but which are than the procedures we proposed in earlier Chapters. The first procedures we will discuss are ones which have been proposed by Ghosh and Sen (1971) and also by Puri and Sen (1985). They consider only the testing problem and do not consider, in this setting the problem of estimating the parameter $\vec{\beta}$. Two test procedures are proposed, one a pure rank statistic in which both Y's and $\vec{X}$'s are ranked and scored, and a mixed rank statistic, in which only the Y's are ranked and scored. It should also be mentioned that the procedures discussed here are actually suitable for a null hypothesis which is more general than the one of interest in the correlation model. Specifically, it may be used when the null hypothesis is $H_0 : F(y|\vec{x}) = F_0(y)$, which clearly contains the situation in the correlation as a special case. Let us first consider the mixed rank test statistic, as it has a limiting distribution which agrees with that of pF_R, both under the null and a sequence of contguous alternatives. As we did in Chapter II, we generate a sequence of scores $a(1) \leq a(2) \leq \cdots \leq a(n)$ not all of which are equal and which have a sum of 0. Next, we make the following definitions.

following definitions. Here l goes from 1 to n.

$$\bar{a} = \frac{1}{n}\sum_{i=1}^{n} a(i) = 0$$

$$\bar{x}_{ln} = \frac{1}{n}\sum x_{ln}$$

$$m_{ln}^{*} = \sum_{i=1}^{n} a[R(Y_i)][x_{li} - \bar{x}_{ln}]$$

$$v_n^1 = \sum_{i=1}^{n}[a(i)]^2$$

$$S_n = \frac{1}{n-1}\sum_{i=1}^{n}(\vec{x}_i - \bar{\vec{x}}_n)(\vec{x}_i - \bar{\vec{x}}_n)'$$

We will also need the matrix

$$V_n^{*} = v_n^1 S_n \text{ and the p vector } \vec{M}_n = (m_{nl}^{*}).$$

The test statistic is a quadratic form in $\vec{M}_n$and V_n^{*-}, which is a generalized inverse of V_n^{*-}. Specifically, the test statistic is

$$L_n^{*} = \frac{1}{n}\vec{M}_n' V_n^{*-}\vec{M}_n \qquad (5.12)$$

To obtain the asymptotic distribution of L_n^{*} Ghosh and Sen (1971) employ a conditional argument, given $\vec{x}_1 \ldots \vec{x}_n$, and apply a previous theorem on permutational convergence. The result is that L_n^{*} converges in probability to a random variable which has, under the null hypothesis, a distribution which is chi square with p degrees of freedom. If we consider a sequence of alternative hypotheses which are contiguous to the null, Ghosh and Sen (1971) obtain the result that L_n^{*} has a limiting distribution which is non central chi square with p degrees of freedom and non centrality parameter given by

$$\delta^*_{PS} = \gamma^2 \vec{d}' \Sigma \vec{d} \tag{5.13}$$

which we recognize as the noncentrality parameter of the asymptotic distribution of pF_R under the same sequence of alternatives. Thus our statistic and the one of Ghosh and Sen (1971) are asymptotically equivalent. Now we consider the pure rank statistic proposed by Ghosh and Sen (1971) for this problem. In this procedure, both variables are ranked and scored. We denote the new score generating functions by φ^*_j, and use them to generate scores by the following relation.

$$b_{nj}(i) = \varphi^*_j\left(\frac{1}{n+1}\right) \qquad 1 \le i \le n \quad 1 \le j \le p \tag{5.14}$$

These are the scores which will be applied to the ranks of the $\vec{X}$' s. We will need the following definitions.

$$\bar{b}_l = \frac{1}{n}\sum_{i=1}^{n} b_{nl}(i) \quad 1 \le l \le p$$
$$v^1_n = \frac{1}{n}\sum_{i=1}^{n} a(R_i)^2$$
$$V^2_{njl} = \frac{1}{n}\sum_{i=1}^{n} [b_{nj}(S_{ji}) - \bar{b}_j][b_{nl}(S_{li}) - \bar{b}_l]$$
$$m_{nl} = \sum_{i=1}^{n} a(R_i)[b_{nl}(S_{li}) - \bar{b}_l]$$

In the formulas above, R_i denotes the rank of Y_i among $Y_1, Y_2, \ldots, Y_n$ and S_{li} denotes the rank of X_{li} among $X_{l1}, X_{l2}, \ldots, X_{ln}$ for $1 \le l \le p$. If define the quantities

$$V^{(2)}_n = (V^2_{njl}),\ V_n = v^1_n V^2_n,\ \vec{M}_n = (m_{nl})$$

and if we denote a generalized inverse of V_n by V_n^-, the test statistic is

$$L_n = \frac{n-1}{n} \vec{M}_n' V_n^- \vec{M}_n \tag{5.15}$$

Under the null hypothesis, Ghosh and Sen (1971) show that L_n converges in probability to a random variable which has a central chi square distribution with p degrees of freedom. In this case the behavior is identical to the behavior of the mixed rank statistic. However, in the case of a sequence of contiguous alternatives, while L_n still has a chi square limiting distribution, the noncentrality parameter is much more complex. Let us denote the noncentrality parameter in this case by δ_{PS}. In order to define this, we need the following definitions.

$$\begin{aligned}
\tau^* &= (\tau^*_{(ss')})_{s,s'=1,\ldots,p} \\
\tau^*_{(ss')} &= \sum_{l=1}^{p}\sum_{l'=1}^{p} B^*_{sl} B^*_{sl'} \nu_{ll'}(F_1) \\
B^*_{ll'} &= \iint x_l \varphi^*[F_{[ll']}(x_{ll'})]\; dF_{3[ll']}(x_l, x_{l'}) \\
B^*_{ll} &= \int x_l \varphi[F_{1(l)}(x_l)]\; dF_{1(l)}(x_l) \\
\nu^{ll}[F_1] &= \int_0^1 \varphi^2(u)du = 1 \qquad 1 \le i \le p \\
\nu^{ll'}[F_1] &= \iint \varphi^*_l[F_{1(l)}(x_{1(l)})]\varphi^*_{(1l')}(x_{l'}) dF_{3(ll')}(x_l, x_{l'}) \\
\nu^{11} &= \int_0^1 \varphi^2(u)\; du = 1
\end{aligned}$$

The notation above is from Puri and Sen (1985). From the above, we obtain the non centrality parameter as $\delta_{PS} = \nu^{11}\vec{d}'\tau'\vec{d}$. For a discussion of this non centrality parameter and efficiency of L_n, see Puri and Sen (1985).

5.6 Likelihood Ratio Test Procedure

In this section we discuss one last testing procedure, the one based on the method of likelihood ratio. We assume that the forms of the distributions of Y and $\vec{X}$ are known. We also assume that the usual conditions needed for the validity of the likelihood ratio procedure met. Denote the MLE of $\vec{\beta}$ by $\vec{\beta}_L$. The likelihood ratio statistic is Λ_n and is defined as

$$\Lambda_n = \Pi_{i=1}^{n} \frac{f_0(y_i)}{f_0(y_i - \vec{\beta}'\vec{x}_i)}$$

Standard theory shows that $-2log\Lambda_n$ converges in distribution, when the null hypothesis is true, to a central chi squae distribution with p degrees of freedom. When we consider a sequence of contiguous alternatives, this statistic has an asymptotic distribution which is non central chi square with noncentrality parameter $\delta_L R$, which is given by $\Im\vec{d}'\Sigma\vec{d}$, where $\Im$ denotes the Fisher information of the distribution function F_0 of the errors. There is a simple relationship between the noncentrality parameter of the likelihood procedure and the one for our statistic. If we know the form of the error distribution, define $\varphi(u; f_0)$ as follows

$$\varphi(u; f_0) = -\frac{f_0'[F^{-1}(u)}{f_0[F_0^{-1}]}$$

Let $\varphi(u)$ denote the score function we plan to use to generate the scores. Then we have the following

$$\begin{aligned}\gamma^2 &= \int_0^1 \varphi(u)\varphi(u;f_0)du \\ &\leq [\int_0^1 \varphi^2(u)du][\int_0^1 \varphi^2(u;f_0)du] \\ &= \Im\end{aligned}$$

Thus we have the following relation between the noncentrality parameters of our statistic and the likelihood ratio statistic.

$$\delta_R = \gamma^2 \vec{d}'\Sigma\vec{d} \leq \Im\vec{d}'\Sigma\vec{d} = \delta_{LR} \tag{5.16}$$

If we choose our score function to be the function $\varphi(u;f_0)$ the inequality above may be replaced by an equality, and we see that in that case our rank procedure is just as powerful as the likelihood ratio procedure and is more robust than the likelihood ratio test. This would seem to indicate that, if we are willing to assume that we know the form of the error distribution and the form of the distribution of the $\vec{X}$'s, then we should use the rank procedure we propose instead of the likelihood ratio test. We sacrifice little or no power, and gain some robustness.

5.7 Asymptotic Efficiencies

We now list the ARE's of the various testing procedures we have discussed. We use the following notation.

A L_{LR} will denote the likelihood ratio statistic.

B L_M will denote the M-estimate based test statistic.

C L_R will denote the statistic we propose.

D L^*_{PS} will denote the mixed rank statistic of Ghosh and Sen(1971).

Then the ARE's of the various procedures are

$$e(L^*_{PS}, L_{LR}) = \frac{\gamma^2}{\Im}$$
$$e(L_{LR}, L_M) = \frac{A\Im}{B^2}$$
$$e(L^*_{PS}, L_R) = 1$$
$$e(L_R, L_{LR}) = \frac{\gamma^2}{\Im}$$
$$e(L_{PS}, L_R) = \frac{A\gamma^2}{B^2}$$
$$e(L_R, L_M) = \frac{A\gamma^2}{B^2}$$

This information shows that our statistic and the mixed rank statistic of Ghosh and Sen (1971) perform equally well. However, there does not seem to be any way to use their procedure to obtain a measure of multiple association as we may with our procedure. This could be viewed as a drawback of the Ghosh and Sen (1971) methods. If we consider the M-estimate methods we have discussed, in order to obtain test statistics which have a limiting distribution which is chi square in form, we must restrict attention to the case in which the robustness properties for the test statistic are essentially similar to those of our statistic. It is true that there are testing procedures based on M-estimates which have bounded influence functions, but the asymptotic distributions of these statistics are not simple in form. Also, the M-estimate based proocedures require that the error distribution be symmetric about zero, which is not needed for our procedures. As we have already mentioned, if we know the form of the error distribution, we may use a rank procedure which just as powerful as the likelihood ratio procedure and which is more robust than the likelihood ratio procedure. It seems, based on the material presented, that rank based procedures in the analysis of linear models posses characteristics which are similar enough to least squares to make them easy to use, and familiar in interpretation. The motivation for pF_R is the same as that for F_{LS}, and R^2 has the same relation to F_{LS} as R_2 has to pF_R. Further, R_2 is easily interpreted in either the regression model or the correlation model, just as R^2 is, while being more robust than R^2. While we may formally construct analogues of R_2 using M-estimates, these do not seem to represent any easily interpretable parameters in the correlation model, and so there use seems limited. Since the test procedures of Ghosh and Sen (1971) do not lend themselves to any estimation methods,it seems that only our proceedure is flexible enough to be used for estimating $\vec{\beta}$, testing hypotheses, and constructing measures of multiple determination.

CHAPTER VI

APPLICATION TO ELLIPTICAL DISTRIBUTIONS

6.1 Discussion of Elliptical Distributions

In this chapter we introduce a family of elliptical distributions discussed by Muirhead (1982), and consider the behavior of $\vec{\beta}_R$ and $\vec{\beta}_M$ when the parent distribution is one these elliptical distributions. We will see that these distributions do not meet the conditions of the correlation model, and so the results stated for the rank estimate of $\vec{\beta}$ will only be conjectures. However, when the elliptical distribution is question is multivariate normal, the conjectured asymptotic distribution of $\vec{\beta}_R$ reduces to the form it should have in that situation, so that the conjecture seems reasonable. The results for $\vec{\beta}_M$ are valid, however, and we shall see that this result too reduces to the usual value when the parent distribution is normal.

We now give the definitions of the quantities we need to introduce the elliptical distributions we will consider. We follow the notation of Muirhead for the most part. That author calls these distributions normal mixture distributions.

DEFINITIONS.

1) $\vec{T}$ has a $MVN_{p+1}[\vec{0}, \Sigma]$ distribution.

2) Ξ is a $(p+1)\times(p+1)$ positive definite matrix. Denote by $\Xi^{\frac{1}{2}}$ the square root of Ξ.

3) Let Z denote a positive random variable, stochastically independent of **T**, *which has distribution function K.*

4) Let q denote a positive function defined on R^+

Now define the random vector $(Y, \vec{X}')'$ as

$$(y, \vec{x}')' = q^{\frac{1}{2}}(Z)\Xi\vec{T} \tag{6.1}$$

The joint distribution of Y and $\vec{X}$ given $Z = z$ is seen to be multivariate normal with mean vector $\vec{0}$ and covarance matrix $q(z)\Xi$. Just as we did in Chapter III, we may write the matrix Ξ as

$$\Xi = \begin{pmatrix} \sigma^2 + \vec{\beta}'\Sigma\vec{\beta} & \vec{\beta}'\Sigma \\ \Sigma\vec{\beta} & \Sigma \end{pmatrix} \tag{6.2}$$

Notice that the conditional distribution of Y and $\vec{X}$ given $Z = z$ is multivariate normal, and it is this conditional case which falls in the correlation model. The joint density of Y and $\vec{X}$ is seen to be that of of multivariate normal vector with dimension $p+1$ and having for it's mean vector and covariance matrix

$$(0, \vec{0}')' \quad \Xi$$

respectively.

We now consider some special cases of these normal mixture distributions. It is seen that these distributions depend on the choice of the function q and the type of distribution that the random variable Z posesses.

1) If we choose $q(z) = z^{\frac{1}{2}}$ and choose the distribution of Z to be given by $P[Z = 1] = 1 - \epsilon$, $P[Z = \tau^2] = \epsilon$ for some $0 \leq \epsilon \leq 1$, then the distribution generated by Z is a contaminated normal distribution.

2) If we choose $q(z) = \frac{n^{\frac{1}{2}}}{z^{\frac{1}{2}}}$ and take Z as a random variable with a chi square distribution having n degrees freedom, then Y and $\vec{X}$ have a joint multivariate t distribution.

3) If we take $q(z) = z^{\frac{1}{2}}$, and $P[Z = 1] = 1$, then Y and $\vec{X}$ have a joint multivariate normal distribution. Case 3 is a special case of the correlation model.

6.2 Influence Function of Rank Estimator of Beta for Elliptical Distributions

In this section we will obtain the influence function of the rank estimate of $\vec{\beta}$ and use it to formally obtain the asymptotic distribution of the $\vec{\beta}_R$. We assume that the true value of $\vec{\beta}$ is $\vec{0}$. Using the ideas of previous sections we choose as our estimate of $\vec{\beta}$ any value of $\vec{b}$ which solves the system

$$(6.4) \qquad \sum_{i=1}^{n} a[R(Y_i - \vec{b}'\vec{x}_i)]x_i = \vec{0}$$

A functional form of 6.4 is

$$(6.5) \qquad \int \varphi[G^*(y - \vec{T}(H)'\vec{x})]\vec{x}dH(\vec{x}, y) = \vec{0}$$

Here $G^*(t)$ denotes the distribution function of $Y - \vec{T}(H)'\vec{X}$. When $\vec{T}(H) = \vec{0}$, $G^*(t) = G(t)$ which is the marginal distribution of Y. To obtain the influence function of the estimate of $\vec{\beta}$, we replace the distribution function H by the contaminated distribution function $H_s = (1-s)H + sH_1$, where H_1 is an arbitrary distribution function. By differentiating with respect to s we obtain the Gâteux derivative. Specifically we have the following theorem.

THEOREM 6.1. *If we define γ_Z to be*

$$\gamma_Z = \int_0^{\infty} [\int \varphi'[G(y)]g(y)q(z)\frac{1}{\sqrt{2q(z)\sigma^2 y}}exp[\frac{-y^2}{2q(z)\sigma^2}]dydK(z)$$

Then the influence function of $\vec{T}$ is

$$\vec{I}F(\vec{x}_0, y_0 : \vec{T}, H) = \frac{1}{\gamma_Z}\varphi[G(y)]\Sigma^{-1}\vec{x}_0 \tag{6.8}$$

PROOF: The Gâteux derivative of the estimate is equal to

$$\frac{1}{\gamma_Z}\Sigma^{-1}\int \varphi[G(y)]\vec{x}dH_1(\vec{x}, y) \tag{6.9}$$

and 6.8 follows readily from 6.9.

Notice that if $q(z) = 1$ for all z, and $P[Z = 1] = 1$, then γ_z reduces to the parameter γ discussed previously.

We now use the influence function just obtained to conjecture the asymptotic distribution of the estimate under the null hypothesis that the true value of $\vec{\beta}$ is zero. The conjecture is that $\sqrt{n}\vec{\beta}_R$ is asymptotically normally distributed with covariance matrix equal to

$$\begin{aligned} &Var(\vec{I}F(\vec{x}, y; \vec{T}, H) \\ &= \frac{1}{\gamma_Z}\Sigma^{-1}\int \varphi^2[G(y)]\vec{x}\vec{x}'dH(\vec{x}, y)\Sigma^{-1} \\ &= \frac{A_Z^2}{\gamma_Z^2}\Sigma^{-1} \end{aligned} \tag{6.10}$$

where A_Z^2 is defined to be

$$\int_0^\infty \{\int_{-\infty}^\infty \varphi^2[G(y)]\frac{1}{\sqrt{(2\pi q(z)\sigma^2)}} exp[\frac{-y^2}{2q(z)\sigma^2}]dy\}q(z)dK(z) \tag{6.11}$$

Notice that if q(z) is equal to one and $P[Z = 1] = 1$, then 6.11 gives the correct asymptotic covariance matrix for $\sqrt{n}\vec{\beta}_R$ when Y and $\vec{X}$ have a joint multivariate normal distribution.

6.3 Properties of M-estimates of Beta at Elliptical distributions

In this section we discuss the M-estimate of $\vec{\beta}$ when the joint distribution of Y and $\vec{X}$ is a normal mixture distribution. We mention that Hampel et al. (1985) state that the estimate of $\vec{\beta}$ they present are consistent and asymptotically normal for arbitrary distribution functions which satisfy their conditions.

We let H denote the joint distribution of Y and $\vec{X}$, and define two matrices M_Z and Q_Z as follows

$$M_Z(\nu, H) = \int \nu'(\vec{x}, y)\vec{x}\vec{x}' dH(\vec{x}, y)$$
$$Q_Z(\nu, H) = \int \nu^2(\vec{x}, y)\vec{x}\vec{x}' dH(\vec{x}, y)$$

In the case of least squares, where $\nu(\vec{x}, y) = y$, these matrices reduce to the following

$$M_Z = E[q(z)]\Sigma$$

$$Q_Z = \sigma^2 E[q^2(z)]\Sigma$$

The asymptotic distribution of $\sqrt{n}\vec{\beta}_M$ is multivariate normal with covariance matrix $M_Z^{-1}Q_Z M^{-1}$. In the least squares case this reduces to $\frac{E[q^2(z)]}{E^2[q(z)]}\sigma^2\Sigma^{-1}$. If we take the function $q(z) = 1$ and choose Z so that $P[Z = 1] = 1$, then this reduces to the value given for the asymptotic distribution of $\vec{\beta}_M$ given in Chapter III.

6.4 Comparison of M and R–Estimates

Finally we examine, in one case, the ARE of our rank estimate to the M-estimate. We concentrate on the situation where the normal mixture distribution of Y and $\vec{X}$ is contaminated normal, and we use Wilcoxon scores. Also, we compare the rank estimate to the least squares estimate. We choose this case because the expressions we obtain are tractable.

Recall that to generate the contaminated normal distribution, we need to have the function q to be the square root of Z, and assume Z has the distribution which gives probability $1-\epsilon$ to the value 1 and puts probability ϵ on the value τ^2. In this case the distribution function of Y is also contaminated normal with density $g(y)$ given as

$$(1-\epsilon)\frac{1}{\sigma}\phi(\frac{y}{\sigma}) + \epsilon\frac{1}{\tau\sigma}\phi(\frac{y}{\sigma\tau})$$

where ϕ denotes the standard normal density function. Now, we may write the parameter γ_Z in the following manner.

$$\gamma_Z = (1-\epsilon)\int \varphi'[G(y)]g(y)\frac{1}{\sigma}\phi(\frac{y}{\sigma})dy$$
$$+\tau^2\epsilon\int \varphi'[G(y)]g(y)\frac{1}{\sigma\tau}\phi(\frac{y}{\sigma\tau})dy$$
$$= \int \varphi'[G(y)]g^2(y)dy + \epsilon(\tau^2-1)\int \varphi'[G(y)]g(y)\frac{1}{\sigma\tau}\phi(\frac{y}{\tau\sigma})dy$$

Notice that the integral

$$\int \varphi'[G(y)]g^2(y)dy$$

is simply the parameter γ we have seen before, calculated at the distribution G. From this, we see that γ_Z is a minimum when τ is equal to 1. We notice that if this is the case, then Y and $\vec{X}$ have a joint normal distribution. We may rewrite A_Z^2 in a similar manner. We have

$$A_Z^2 = (1-\epsilon)\int \varphi^2[G(y)]\frac{1}{\sigma}\phi(\frac{y}{\sigma})dy$$
$$+\epsilon\tau^2\int \varphi^2[G(y)]\frac{1}{\tau\sigma}\phi(y\sigma\tau)dy$$
$$= \int \varphi^2[G(y)]g(y)dy + \epsilon(\tau^2-1)\times$$
$$\int \varphi^2[G(y)]\frac{1}{\tau\sigma}\phi(\frac{y}{\tau\sigma})dy$$
$$= 1 + \epsilon(\tau^2-1)\int \varphi^2[G(y)]\frac{1}{\tau\sigma}\phi(\frac{y}{\tau\sigma})dy$$

and we see that A_Z^2 takes on its minimum value when the joint distribution of Y and $\vec{X}$ is normal.

Now, if we take the M-estimate to be least squares, we obtain for the ARE the following expression

$$\frac{E[Z]\gamma_Z^2}{E^2[\sqrt{Z}]A_Z^2}$$

If we evaluate the expectations in the previous expression we obtain

$$E[Z] = 1 + (\tau^2 - 1)\epsilon$$

$$E[Z^{\frac{1}{2}}] = 1 + (\tau - 1)\epsilon$$

CHAPTER VII

SUMMARY AND CONCLUSIONS

The purpose of this work was to examine the rank analysis of linear models proposed by Hettmansperger and McKean (1975) and construct a measure of multiple determination to be used with it. The hope was that such a coefficient could be constructed in such a way that its use would be similar to that of R^2, but that it would inherit some of the robustness properties of the rank statistic and estimate used. In looking for such a coefficient we found that there were two candidates for consideration, and each was easily interpreted.

To examine which of these two new coeffiecents to propose, the influence functions of the test statistic, estimate, and each of $R_{1\varphi}$ and $R_{2\varphi}$ were obtained and examined. On the basis of these influence functions and several examples, we determined that $R_{2\varphi}$ is the statistic which posseses the qualities we are interested in. It is robust and, in the case of the multiple correlation model, estimates a 1-1 increasing function of the classical multiple correlation coefficient $\overline{R}^2$. In addition, we saw that the rank testing and estimating procedures extend to the correlation model, and so we have a set of procedures to use in correlation model which are robust, and which relate to each other just as the corresponding quantities in least squares do.

We also noted that the procedures described here are only a first step in addressing the problem. The next step is to obtain rank procedures which have influence functions which are bounded in both $\vec{x}$ and Y.

BIBLIOGRAPHY

Anderson, T.W, *An Introduction to Multivariate Analysis,* New York: John Wiley 1984

Arnold, S.F., Asymptotic Validity of F-Tests for ordinary linear models and the multiple correlation model. *JASA* 1980

Arnold, S.F., *The Theory of Linear Models and Multivariate Analysis* New York: John Wiley 1981

Bickel, Peter J., and Doksum, Kjell A., *Mathematical Statistics : Basic Ideas and Selected Topics* San Fransisco: Holden-Day 1977

David, H.A., *Order Statistics* New York: John Wiley 1970

Filipovva, A.A., Mises' Theorem on the asymptotic behavior of functionals of empirical distributions and its Statistical Application. *Theo. Prob. App., 7,* 1962, 24-57

Fernholtz, L.T., *VonMises Calculus for Differentiable Statistical Functionals* New York: Springer Lecture Notes in Statistics 1983

Ghosh, M. and Sen, P.K., On a class of rank order tests for regression with partially informed predictors. *AMS 42* 1971, 650-661

Hald, Anders. *Statistical Theory, with Engineering Applications.* New York: John Wiley 1952

Hájek, J. and Šidák, Z. *Theory of Rank Tests.* New York: Academic Press 1967

Hampel, F.R., The influence curve and its role in robust estimation *JASA 69* 1974, 383-393

Hampel, F.R., Rousseeuw, E.M., Ronchetti, E.M., and Stahel, W.A., *Robust Statistics: The Approach based on Influence Functions.* New York: John Wiley 1985

Hettmansperger, T.P., *Statistical Inference Based on Ranks.* New York: John Wiley 1984

Hettmansperger, T.P., and McKean, J.W., A robust alternative based on ranks to least squares in analyzing linear models. *TC 19* 1975, 275-284

Hettmansperger, T.P., and McKean, J.W., A geometric interpretation of inferences based on ranks in the general linear model. JASA 78 1983, 885-893

Hodges, J.L., and Lehmann, E.L., The efficiency of some nonparametric competitors to the t-test. *AMS 27* 1956 pp 324-335

Huber, P.J., Robust Regression: Asymptotics, conjectures and Monte Carlo. AS 1 1973, 799-821

Huber, P.J., Robust Statistical Procedures. *Regional Conference Series in Applied Mathematics 27* Philadelphia:SIAM 1977

Huber, P.J., *Robust Statistics.* New York: John Wiley 1981

Jaeckel, L., Estimating regression coefficients by minimizing the dispersion function. *AMS 43* 1972, 1449-1458

Jureckova, J., Nonparametric estimate of regression coefficients. *AMS 42* 1971, 1328-1338

Jureckova, J., Asymptotic relations of M-Estimates and R-Estimates in linear regression model. *AS 5* 1977, 464-472

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

McKean, J.W., and Hettmansperger, T.P., A robust analysis of the general linear model based on one-step R-Estimates. *Biom 65* 1975, 571-579

McKean, J.W., and Hettmansperger, T.P., Tests of hypothesis based on ranks in the general linear model. *Comm. St. A* 1976, 693-709

McKean, J.W., and Sievers, G.L., Coefficients of determination for least absolute deviation analysis. *Stat. Prob. L. 5* 1987. 49-54

Muirhead, R.J., *Aspects of Multivariate Statistical Analysis.* New York: John Wiley 1982

Puri, M.L., and Sen, P.K., *Nonparametric Methods in Multivariate Analysis.* New York: John Wiley 1971

Puri, M.L. and Sen, P.K., *Nonparametric Methods in General Linear Models.* New York: John Wiley 1985

Ronchetti, E., Robust Testing in Linear Models: The Infinitesimal Approach. *Phd Thesis* ETH: Zurich 1982

Schrader, R.M., and Hettmansperger, T.P., Robust analysis of variance based upon a likelihood ratio criterion. *Biom 67* 1980, 93-101

Schweder, T., Window estimation of the asymptotic variance of rank estimates of location. *SJS 2* 1975, 113-126

Sievers, G.L., A weighted dispersion function for estimation in linear models. *Technical Report No. 62, Department of Mathematics, WMU.* 1979

Sievers, G.L., A different measure of correlation based on R-Estimates. *unpublished manuscript* 1987

Tableman, Mara, Bounded Influence Rank Regression: A One-Step Estimator Based on Wilcoxon Scores. *Statistics Research Report No.88-08.* 1988

vonMises,R., On the asymptotic distribution of differentiable statistical functions. *AMS 18* 1937, 309-348