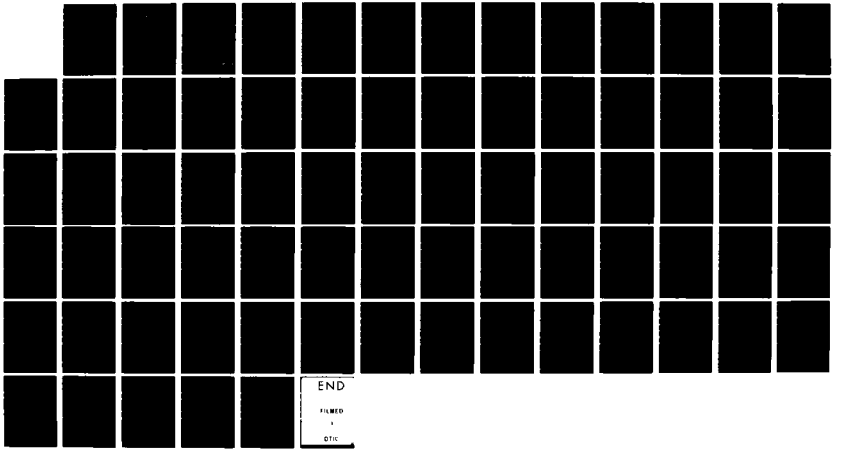
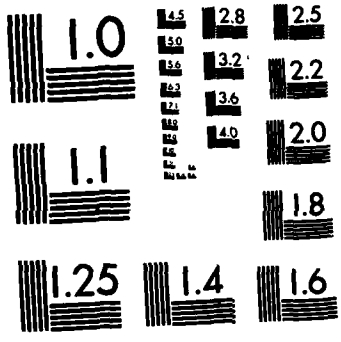
AD-A124 678 ROBUST MULTIPLE LINEAR REGRESSION(U) AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH SCHOOL OF ENGINEERING A M SULTAN DEC 82 AFIT/GOR/MA/82D-3 1/1

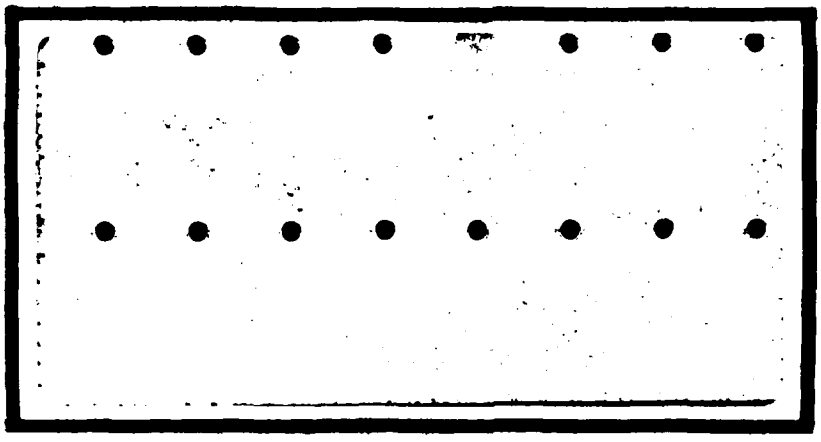
UNCLASSIFIED F/G 12/1 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AD A124678



This document has been approved for public release and sale; its distribution is unlimited.

DTIC
ELECTE
FEB 2 2 1983
S D
A

DEPARTMENT OF THE AIR FORCE

AIR UNIVERSITY (ATC)

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

DTIC FILE COPY

ROBUST MULTIPLE

LINEAR REGRESSION

THESIS

AFIT/GOR/MA/82D-3 Ahmed Mohamed M. Sultan
Major Egyptian AF

DTIC
ELECTE
FEB 2 2 1983
S D
A

This document has been approved for public release and sale; its distribution is unlimited.

AFIT/GOR/MA/82D-3

ROBUST MULTIPLE LINEAR REGRESSION

THESIS

Presented to the Faculty of the School of Engineering

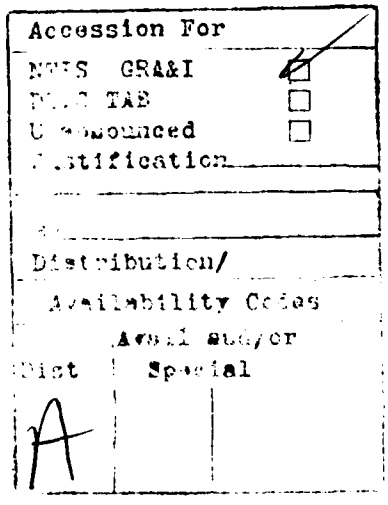
of the Air Force Institute of Technology

Air University

in Partial Fulfillment of the

Requirements for the Degree of

Master of Science



by

Ahmed Mohamed M. Sultan

Major Egyptian AF

Graduate Operations Research

December 1982

Approved for public release; distribution unlimited

Table of Contents

Preface

The linear regression model is one of the most widely used quantitative tools of the applied social sciences and many of the physical sciences. The most common used techniques in this kind of model, is the ordinary least squares because of its low computational costs, its intuitive plausibility in a wide variety of circumstances, and its support by a broad and sophisticated body of statistical inference. The least squares tool could be used on 3-basic levels:

1. It can be applied mechanically, or descriptively, as a means of curve fitting.

2. It enables us to perform hypothesis testing.

3. It gives a reasonable way of understanding complex physical and social phenomena.

Let us now denote the regression model by

$$Y = X\beta + \epsilon \text{ , where}$$

X - is an $N \times k + 1$ matrix,

β - is an $k + 1 \times 1$ vector,

ϵ - is an $n \times 1$ vector, and

Y - is an $N \times 1$ vector.

The assumptions for the least square method are:

1. $E(\epsilon) = 0$ i.e.

The expected value of the error term ϵ is zero

2. $E(\epsilon - E(\epsilon))^2 = \sigma^2 I$: i.e.

All error terms have constance variance σ^2 and they are independent.

3. The X matrix is nonstochastic with rank $\rho(x) = k + 1$ i.e. none of the columns of x is a linear combination of other columns.

The estimators for the coefficients vector β which are given as $\underline{\hat{\beta}}$:

$$\underline{\hat{\beta}} = (X'X)^{-1} X'\underline{Y}$$

These estimators have the properties:

1. $\hat{\beta}$ is a linear function of Y.

2. $E(\hat{\beta}) = \beta$ i.e. unbiasedness

3. $V(\hat{\beta}) = E(\underline{\hat{\beta}} - E(\underline{\hat{\beta}})\ (\underline{\hat{\beta}} - E(\underline{\hat{\beta}})' = \sigma^2\ (X'X)^{-1}$

and the estimate for σ^2 is given by S^2 where:

$$S^2 = \frac{\text{Error sum of squares}}{N-(k+1)}$$

4. The basic and most important assumption for that model is the assumption of normality. The confidence interval and testing procedures are all based on the normality assumption. It is true that normality assumption is an important case and that it can sometimes be justified by the central limit theorem, but it is equally true that the assumption is made in many cases in which it does not really hold. There are two basic questions arising in this case:

   1) How serious are the consequences?

   2) To what extent is a test "robust"?

i.e. To what extent is a test insensitive to departures from the assumption under which it is derived?

In that concern appears two basic issues: - First: Tests which concern first moments (such as t-tests for elements of the parameter vector β of the expectation $X\beta$ in the standard linear model, are relatively insensitive to departures from normality.

Second: Tests concerning second moments such as F-tests are much less robust (see Kendall and Stuart 1967, pp. 455). Thus our search here will be bascially for a robust technique that could be applied for estimating parameters of the linear model $Y = X\beta + \epsilon$

Foreward

During the course "Linear Statistical Models" given in AFIT, I started to be interested in the regression models due to their wide use in management, management sciences, and social sciences. These models are successfully used in real life applications basically because of the sound understanding of both the underlying theory and the practical applications themselves.

Robust linear regression model is an area of greater interest since in many sets of data, there are fairly large percentages of "Outliers" due to heavy tailed models of errors in collecting and recording. Due to the fact that these outliers have an unusually great influence on "least squares" estimators (or generalized least square estimators), robust procedure attempts to modify those schemes. During a course by Dr. A. H. Moore, Professor in the Department of Mathematics, Air Force Institute of Technology, School of Engineering in robust statistics, I became interested in the area of robust regression. After talking to Dr. Moore about my interest in robust regression, we decided to make a search in robust multiple linear regression.

I wish to express my thanks to Dr. A. H. Moore, my thesis advisor, for his valuable remarks, directions for search and his aid in the accomplishment of my thesis. I also wish to thank Dr. J. P. Cain, my reader, to whom I am especially indebted for learning a lot about multiple linear regression.

Finally, I owe my wife, Azza, my son Mohamed and my lovely American daughter, Dina, a great debt of love and best wishes for their patience and encouragement during my study at AFIT.

AHMED MOHAMED M. SULTAN

List of Tables

Abstract

An extensive Monte Carlo analysis is conducted to determine the performance of robust linear regression techniques with and without outliers. Thirteen methods of regression are compared including least squares and minimum absolute deviation. The classical robust techniques of Huber, Hampel were studied and robust techniques using the Q-statistic as a discriminant were introduced.

The model studied contained eleven variables with 27 observations. The error distributions considered were uniformly normally, double exponentially distributed.

Least squares gave the best fit without outliers. In the presence of gross outliers a rejection of outliers technique gave the best fit.

I. Introduction

Problem Statement

Regression analysis is a statistical technique for expressing the relationship between variables in a mathematical form. Moreover it is considered one of the most widely used statistical techniques due to its large applications in almost every field. An earlier search has been done by James E. Flanagan GOR/81-D to examine the use of Lp-norms and distance estimation. Due to computer and algorithm limitations it was only possible to examine the following linear models:

$$y = \beta_o + \beta_1 X_1 + \varepsilon$$

and

$$y = \beta_o + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

However, the application envisioned is to try to improve the "predictive" operations and maintenance cost model (ALPOS model) developed for Air Force Avionics Laboratory Systems Evaluation group. However, their linear model used 20 independent variables. The earlier search demonstrated the feasibility of a generalized approach to the regression problem but was unable to handle many independent variables.

This thesis envisions using a different approach (Adaptive) so that many independent variables (up to 100) can generally be handled.

Verification of the model can be made by comparing its prediction capability with the prediction capability of the ALPOS model.

Review of Applicable Literature

The possible existence of non-normal error distribution having infinite variance or with large tails, has led the statistician to a search for estimators that are more "robust" than least squares (L.S.)

estimators. By "robust" here one means a reasonably efficient estimator regardless of the form of the underlying error distribution. When the errors are i.i.d., normal random variables, L.S. estimator are efficient, and so, the search is for estimators that are not much worse than L.S. when the errors are normally distributed but are really better for non-normal errors.

A large number of estimators, were suggested in a considerable body of literature. For example, the surveys of Huber (Ref 43:1041) in which a selective review on robust statistics, centering on estimates of location and extending into other estimation and testing problems. In 1973 Huber (Ref 45:799) defined the maximum likelihood type robust estimates of regression, and investigated their asymptotic properties both theortically and empirically. Koenker and Bassett (Ref 55:33) introduced a new class of linear model called "regression quantiles", which is a simple minimization problem yielding the ordinary sample quantiles in location model. This model generalizes naturally to the linear model. The estimator which minimizes the sum of absolute residuals is an important case. Estimators were suggested, which have comparable efficiency to least squares for normal linear model while substantially out-performing the least squares estimator over a wide class of non-normal error distributions. Another study was made by McKean, J and Hettmansperger, Thomas for the general linear model based on one step R-estimates (Ref 60:571). One step iterations based on a second derivative approximation to the surface was proposed. These estimates, can be obtained quickly from initial estimates. Further the analysis resulting from these estimates is asymptotically equivalent to the minimum dispersion analysis. Thus it can be recommended for large data sets. In addition Maddala

(Ref 57:308) surveyed the work done by Huber and Anscombe for minimizing

$$\Sigma f (y_i - \sum_k X_{ik} \beta_k)$$

with different definition of f for each of them. Then a discussion of least absolute deviation minimization was discussed. Also a relevant part of Mosteller (Ref 64:105) discussed different suggestions for solution of non-normal error linear models. Finally, Narula (Ref 66:185) suggested the minimization of the sum of relative errors (MSRE) as an alternative to least squares. The problem is formulated as a linear programming problem and a solution procedure is given.

Model Selected

The model selected is

$$y = \beta_o + \beta_1 X_1 + \text{--------} + \beta X_{11} + \varepsilon$$

for the problem of property valuation. The objective is to predict y, the sale price of a home for known value, of the variables X_1 through X_{11} which represent (taxes, number of baths, lot size, ---- , lot size, number of fireplaces). The data, 27 observations on variables (y_1 X_1, --- X_{11}) were obtained from Multiple Listing, Vol. 87 for area 12 (Erie, PA).

Choice of Error Models

In order to see the behavior of the proposed adaptive technique, it was necessary to add different error distributions to an exact fit of data. The way it is done here is through getting an estimation for the value of β as $\hat{\beta}_o$ and generating exact values for the y by multiplying X by $\hat{\beta}_o$:

$$y = X \hat{\beta}_o$$

The choice of non-normal error distribution is basically dependent on the

tail length of the distribution. For the uniform case, it has smaller tails, while for the double exponential it has thicker tails relative to the normal distribution.

II. Methods of Estimation

As in the general decision problem, there is no single, best procedure for estimating the parameters of a distribution. In a given case under study, it may be advisable to use the method of moments, Bayes estimates, minimax estimates, or maximum likelihood estimates.

Methods of Moments

This method is oldest method of estimating parameters, which was devised by K. Pearson about 1894. If there are K parameters to be estimated, the method consists of expressing the first K population moments in terms of these K-parameters, equating them to the corresponding sample moments and taking the solutions of the resulting equations as estimates of the parameters. The method usually leads to relatively simple estimates.

The estimates obtained in this way are clearly functions of the sample moments. Since the sample moments are consistent estimates of population moments, the parameter estimates will generally be consistent.

Although the asymptotic efficiency of estimates obtained by the method of moments is often less than 1, such estimates may conveniently be used as first approximation from which more efficient estimates may be obtained by other means.

Bayes Estimates

In the methods of point estimation the assumption is that the random sample came from density $f(.;\phi)$, where the function $f(.;\phi)$ is assumed to be known. Moreover ϕ was some fixed, though unknown, point. In some real world situations which the density $f(.;\phi)$ represents, there is often additional information about ϕ i.e. ϕ itself may act as a random variable

for which one could postulate a realistic density function.

It has been seen that the Bayes action for a given observation $Z = z$ is that which minimizes the expected value of the loss with respect to the posterior distribution. This expected loss, assuming a quadratic loss function $(\phi - \alpha)^2$, is

$$E_H (\phi - \alpha)^2 = \int_{-\infty}^{\infty} (\phi - \alpha)^2 dH(\phi)$$

where $H(\phi)$ is the distribution function for the posterior distribution. Since this expected loss is a second moment of a distribution, it is minimized when taken about the mean of the distribution. That is, the minimizing action and hence the Bayes estimate of ϕ is

$$E_H (\phi) = \int_{-\infty}^{\infty} \phi dH(\phi)$$

Maximum Likelihood Estimates

We shall suppose first that the population of interest is discrete, so that it is meaningful to speak of the probability that $X = \chi$, where X denotes a sample $(X_1,\ldots,X_n)$ and χ a possible realization $(\chi_1,\ldots,\chi_n)$. This probability that $X = \chi$ depends on χ, of course, but it also depends on the state of nature ϕ which governs. As a function of ϕ for given χ, it is called the likelihood function.

$$L(\phi) = P_\phi (X = \chi)$$

Thinking of a state of nature as a possible "explanation" of observed data, the maximum likelihood considers the "best" explanation to be the state of nature $\hat{\phi}$ that maximizes the likelihood function - that maximizes the probability of getting what was actually observed. A maximum likelihood procedure is then one that is best when the state of nature is the maximum likelihood state, $\hat{\phi}$. This is determined from the loss function as the action that minimizes the loss function as a function of

$\hat{\phi}$ and a (i.e. the loss resulting from an action α when the state of nature is take as $\hat{\phi}$).

The best explanation $\hat{\phi}$ of a given observation $X = \chi$ depends on χ, and so defines a function of χ or a statistic. The rule that says take the action that minimizes $\ell(\hat{\phi}, \alpha)$, where ℓ is the loss function, assigns this action to the χ that leads to $\hat{\phi}$, and so the maximum likelihood principle defines a decision function, called the maximum likelihood decision function.

Thus a maximum likelihood estimate is a value of ϕ that maximizes the likelihood function. If ϕ is multidimensional, so is $\hat{\phi}$, and the components are said to be joint maximum likelihood estimates of the corresponding components of Φ.

Some Other Techniques

A brief mention will be made in this port of certain other techniques for obtaining estimators involving somewhat more mathematical preparation than has been provided or assumed. As in general, a decision procedure can be replaced by one based on a sufficient statistic, so in estimating a parameter an estimator can be replaced by a function of a sufficient statistic without deterioration of the risk. In particular, given an unbiased estimate U of the parameter $h(\phi)$, an unbiased estimate based on the sufficient statistic T can be constructed whose variance is not greater than that of U. In some instances the method yields an unbiased estimate of minimum variance.

Given the statistic U, then, consider the function

$$g(t) = E(U|T = t)$$

If T is sufficient, the conditional distribution of X, and therefore that

of the statistic U, are independent of the state ϕ. The function g(t) really depends, then, only on t, as the notation implies. It defines a statistic

$$V = g(T),$$

Whose mean is the same as that of U:

$$E(V) = E(E(U|T))$$

$$= E(U)$$

Consequently, if U is an unbiased estimate of $h(\phi)$, so is V.

The variance of U can be expressed as follows:

$$Var(U) = E((U - E(V))^2)$$

$$= Var\ (V) + E((U - V)^2) + 2E\ ((U - V)\ (V - E(v)))$$

The assertion that Var (U) $\geq$ var (V) will be established as soon as it is shown that the cross product term vanishes. So, Consider

$E((U - V)\ (V - E(V))) = \int_{-\infty}^{\infty} E((U - V)\ (V - E(V))|\ T = t)\ D\ F_T(t)$, where $F_T(t)$ is the distribution function of T. Now,

$$E\ (V - U\ |\ T = t) = E\ (V|T = t) - E\ (U|T = t)$$

$$= g(t) - g(t)$$

$$= 0$$

and

$$E\ ((U - V)\ (V - E(V))|T = t) = E\ ((U - V)\ (g(t) - h(\phi))|T = t)$$

$$= (g(t) - h(\phi))\ E(U - V|T = t)$$

$$= 0$$

Thus the above integral vanishes, and Var (u) $\geq$ Var (V). The variance of V is actually smaller if U does not depend on the data through the value of T only, and so one can do better using V than using U. Clearly, any estimator that is unbiased and has a smaller variance than does g(T) would also have to be a function of the sufficient statistic T (since

otherwise the preceding technique would yield a function of T that does at least as well). But if there is such a function, K(T), also unbiased in estimating $h(\phi)$, then

$$EK(T) = h(\phi)$$
$$= Eg(T)$$

for all ϕ. Frequently the family of densities for T has the property of completeness, which says that if

$$\int_{-\infty}^{\infty} K(t)\, dF_T(t) = \int_{-\infty}^{\infty} d(t) dF_T(t)$$

for all ϕ, then K(t) is essentially the same function as g(t). In this event g(T) is actually an unbiased estimate of $h(\phi)$ with minimim variance.

Thus, although maximum likelihood estimates are known to be consistent, asymptotically efficient, and asymptotically normal, there are usually other estimates that have these properties and which would then appear to serve just as well for large samples (they might even be better for small samples). Such estimates are called best asymptotically normal, or BAN, and can be obtained in various ways.

One class of BAN estimates consists of certain "Minimum Chi-square" estimates, defined as follows: Consider a sample $X_1, ---, X_n$, from a vector valued population X with mean vector $\mu(\phi)$ and covariance matrix $M(\phi)$, ϕ being the parameter to be estimated (it could be multidimensional). The quadratic expression

$$\chi^2 = \frac{1}{n} (\bar{X} - \mu(\phi))^1 [M(\phi)]^{-1} (\bar{X} - \mu(\phi))$$

is minimized as a function of ϕ for given $X_1, ----, X_n$. The minimizing value $\phi(X_1, ---, X_n)$ is called minimum Chi-square estimate of ϕ. It is known to be BAN when X has a distribution belonging to the exponential

family. Various modifications of the minimum Chi-square method also yield BAN estimates.

III. Robust Procedures

General

A mathematical model is basically based upon a set of assumptions. These assumptions are not supposed to be exactly true - they are mathematically convenient rationalizations of an often fuzzy knowledge or belief. These rationalizations or simplifications are vital, and one justifies their use by appealing to a vague continuity or stability principle. This principle states that "A minor error in the mathematical model should cause only a small error in the final conclusions.

A statistical inference model being a branch of the mathematical model should be consistent with the stated principle for a mathematical model. In the simplest cases there are implicit and explicit assumptions about randomness and independence, about distributional models, perhaps prior distributions for some unknown parameters and so on.

During the last decade a "robust" procedures have been introduced to solve the conflict between the model assumptions and the real system being studies to get insensitivity to small deviations from assumptions. Basically, we consider the distributional robustness which means that the true underlying distribution deviates slightly from the assumed model (usually the Gaussian law).

As an example for that Tukey (Ref: 78) introduced a case of a contaminated normal distribution with contamination factor ϵ from two normal distributions $N(\mu,\sigma^2)$ and $N(\mu,9\sigma^2)$. So the observations χ_i will be independent, identically distributed with common underlying distribution $F(x)$ where:

$$F(x) = (1 - \epsilon)\,\Phi\left(\frac{x - \mu}{\sigma}\right) + \epsilon\,\Phi\left(\frac{x-\mu}{3\sigma}\right)$$

where

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2}\, dy \quad \text{is } N(0,1)$$

Two measures of scatter are the mean absolute deviation

$$dn = \frac{1}{n} \Sigma \mid X_i - \bar{X} \mid$$

and the mean square deviation.

$$S_n = [\frac{1}{n} \Sigma (X_i - \bar{X})^2]^{1/2}$$

These two measures indicate different characteristics of the error distribution. The performance of these two measures is summarized by Huber (Ref:46) according to their asymptotically relative efficiency (ARE) of S_n relative to dn versus the contamination factor given in the following table.

$$ARE(\epsilon) = \lim_{n \to \infty} \frac{Var\ (S_n) \mid (E\ (S_n)^2}{Var\ (d_n) \mid (E\ (d_n)^2}$$

ϵ	ARE (ϵ)
0	0.876
0.001	0.948
0.002	1.016
0.005	1.198
0.01	1.439
0.02	1.752
0.05	2.035
0.10	1.903
0.15	1.689
0.25	1.371
0.5	1.017
1.0	0.876

From this Huber concluded that:

1. The above does not imply that we advocate the use of the mean absolute deviation (There are still better estimates of scale).

2. The contaminating observations could be considered as outliers and on treating them one can get a better estimate of the mean square error.

Till this point it seems reasonably to clear the data by rejecting the outliers and then using classical estimation and testing procedures for the remainder one can end with a better estimating model. In reality this approach faces three basic pitfalls in application:

1) It is difficult to identify the real outliers unless one uses a robus estimating model (case multiple linear regression).

2) Even if the original set of observations consists of normal with some gross erros, the cleaned data will not be normal, and the situation is even worse with a non-normal distribution.

3) As an empirical fact the best rejection procedure do not quite reach the performance of the best robust procedure. Because robust procedures make a smooth transition between full acceptance and full rejection of an observation.

Thus a robust procedure should have the following features:

1) It should have a reasonably good (optimal or near optimal) efficiency at the assumed model.

2) Small deviations from the model assumptions should affect the model performance only slightly.

3) Relatively larger deviations from the model should not completely spoil the behavior of the model.

Basic Types of Robust Estimators

The basic types of robust estimators are

1) M-Estimator

(The maximum likelihood tpes estimates)

2) L-Estimaotr

(The linear combinations of order statistic estimator).

3) R-Estimator

(The estimator derived from r and k tests)

1. The M-Estimator

This kind of estimates is the most flexible one, and it generalizes straight forwardly to multiparameter problems, even though (or, perhaps because) it is not automatically scale invariant and has to be supplemented for practical applications by an auxiliary estimate of scale.

Definition: Any estimate T_n, defined by a minimization problem of the form

$$\Sigma \rho (X_i, Tn) = \min$$

or by an implicit equation

$$\Sigma \frac{\partial}{\partial \phi} \rho (X_i, T_n) = 0$$

i.e.

$$\Sigma \psi (X_i, T_n) = 0$$

Where $\psi (X_i, T_n) = \frac{\partial}{\partial \phi} \rho(X_i, T_n)$ is called an M-estimate. (This estimate is the orginary M.L.E. if

$$\rho (X;\phi) = -\log f (X;\phi)$$

In the linear model we have

$$y = X\beta + \epsilon$$

and we are interested in the expected value of the response

$$E(y) = E(X\beta + \epsilon)$$

$$= E(X\beta) + E(\epsilon)$$

$$= XE(\beta) + E(\epsilon)$$

So in case of $E(\epsilon) = 0$, we get

$$E(y) = XE(\beta)$$

i.e. we basically will be interested in the location parameter. Thus assuming

$$\rho(X_i - T_n) = \rho(X_i - T_n), \text{ then}$$

$$\Sigma\, \rho(X_i - T_n) = \min$$

or

$$\Sigma\, \psi(X_i - T_n) = 0$$

Assuming

$$W_i = \frac{\psi(X_i - T_n)}{X_i - T_n} \qquad \text{then}$$

$$\Sigma\ W_i\,(X_i - T_n) = 0$$

$$T_n = \frac{\Sigma\, X_i W_i}{\Sigma\, W_i}$$

Where the weights are dependent on the sample.

For the functional form of

$$\Sigma\, \rho\,(X_i;\, T_n) = 0$$

if it is not possible to generally define T(F) to be a value of t which minimizes

$$\int \rho\,(X;t)\, F(dx)$$

For example the median corresponds to

$$\rho\,(X;t) = |X-t| \text{ while}$$

$$\int |X-t|\, F(dX) \equiv \infty$$

identically in t unless F has a finite first absolute moment. A simple solution to that is obtained by replacing $\rho(X;t)$ by $\rho(X;t)- \rho(X;t_o)$ for some fixed to i.e. in case of the median minimize

$$\int(|x-t|-|x|)\ F\ (d\chi)$$

In a similar way the functional form of $\psi(X_i,t)$ is

$$\int \psi\ (X;T(F))F(d\chi) = 0,$$

This form of $\psi(X,t)$ does not suffer from the previous difficulty, but it might have more solutions corresponding to local minima.

Influence Function of M-Estimates

The influence function describes the effect of adding one more observation with value x to a very large sample on the value of an estimate or test statistic $T(F_n)$ where F_n is the empirical distribution function.

In case of M-Estimates the influence function was found to be proportional to ψ and given as

$$IC(x,F,T) = \frac{\psi(X,T(F))}{-\int(\frac{\partial}{\partial\sigma})\ \psi\ (X_i T(F)\ F(dx)}$$

and in case if $\psi(X;0) = \psi(x - 0)$ we obtain

$$IC(X\ F,T) = \frac{\psi(X - T(F))}{\int\psi'[X - T(F)]F(dx)}$$

2. The L-Estimates

Consider a statistic that is a linear combination of order statistics, or more generally, of some function h of them:

$$T_n = \sum_{i=1}^{n} a_{ni}\ h(X_{(i)})$$

We assume that the weights are generated by a (signed) measure M on (u,1) interval:

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

(This choice of the weights preserves the total mass, $\Sigma a_{ni} = M\{(0,1)\}$, and symmetry of the coefficients, if M is symmetric about $t = \frac{1}{2}$)

Then $T_n = T(F_n)$ derives from the functional $T(F) = \int h\,(F^{-1}(s))\,M\,(ds)$ and this gives exact equality $T_n = T(F_n)$ if the integral is regularized at its discontinuity points and will be equal to

$$\frac{1}{2}\, h\,(F_n^{-1}(s - 0) + \frac{1}{2}\, h(F_n^{-1}\,(S + 0)),$$

where the inverse of any distribution function F is defined in the usual way as

$$F^{-1}(s) = \inf\{\, x \mid F(x) \geq s\} \qquad 0 < s < 1$$

Influence Function of L-Estimates

In a similar way like that for the M-Estimate we can find the influence function of Ts where $Ts = F_t^{-1}(s)$

$$IC(X;F,Ts) = \frac{s-1}{f(F^{-1}(s))}, \text{ for } X<F^{-1}(s)$$

$$= \frac{s}{f(F^{-1}(s))}, \text{ for } X>F^{-1}(s)$$

It is worthwhile to note here that the influence function has a value only if F has a non-zero finite derivative f at $F^{-1}(s)$.

Using the chain rule for differentiation, the influence function of $h(T_s)$ is

$$IC(X,F,h(T_s)) = IC(X,F,T_s)\; h'(Ts)$$

and thus the influence function of the estimator T itself will be

$$IC(X,F,T) = \int IC(X,F,h(T_s))\; M\; (ds)$$

3. R-Estimates

R estimation is a procedure based on ranks. To illustrate the general procedure, consider replacing one factor in the least squares objective function ($\sum_{i=1}^{n} (Y_i - X_i' \beta)^2$) by its rank. Thus if R_i is the rank of $Y_i - X_i' \beta$, then we wish to minimize $\sum_{i=1}^{n} (Y_i - X_i' \beta) R_i$ (Ref 1:894)

Now consider a two sample rank test for shift: let $X_1, ---, X_m$ and $Y_1, ---, Y_n$ be two independent samples from the distributions $F(x)$ and $G(x) = F(X - \Delta)$, respectively merge the two samples into one of size $m + n$ and let R_i be the rank of X_i in the combined sample. Let $a_i = a(i)$, $1 \leq i \leq m + n$, be some given scores; then base a test of $\Delta = 0$ against $\Delta > 0$ on the test statistic

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

Usually, we assume that the scores a_i are generated by some function J as follows

$$a_i = J\left(\frac{i}{m + n + 1}\right)$$

In case of the Wilcoxon test, $J(t) = t - \frac{1}{2}$.

Estimates of shift Δ_n and of location T_n can be derived from such rank test:

(1) In the two sample cases, adjust Δ_n such that $S_{n,n} = 0$ when computed from $(X_1, ---, X_n)$ and $(Y_1 - \Delta_n, ---, Y_n - \Delta_n)$.

(2) In the one sample case, adjust T_n such that $S_{n,n} = 0$ when computed from $(X_1, ---, X_n)$ and $(2T_n - X_1, ---, 2T_n - X_n)$. So a mirror image of the first sample is used as a second sample.

Influence Function of R-estimates

The influence function in this case is given as

$$IC(X,F,T) = \frac{U(x) - \int U(x)f(x)dx}{\int U'(x)f(x)dx}$$

where

$$U(x) = \int J' \left\{ \frac{1}{2} \left[F(x) + 1 - F\,(2T(F) - X) \right] \right\} \cdot f(2T(F) - x)\,dx$$

For symmetric F this can be simplified, since $U(x) = J(F(x))$, then

$$IC(X,F,T) = \frac{J(F(x))}{\int J'(F(x)))f(x)^2 dx}$$

IV. Multiple Linear Regression

A regression model that involves more than one regressor variable is called a multiple regression model. Here we are going to discuss the fit and analysis of this model and some lightspot on the measures of adequacy that are useful in multiple regression.

Multiple Regression and Least Squares

Suppose that we have a certain response y which may be related to K regressor variables by the model

$$y = \beta_o + \beta_1 X_1 + \beta_2 X_2 + \text{---} + \beta_k X_k + \varepsilon$$

This model is called a multiple linear regression model with k-regressors. The parameters β_j, $j = 0,1,\text{--},k$ are called the regression coefficients. This model describes a hyperplane in the k-dimensional space of the regressor variables X_j. The parameter β_j represents the expected change in the response y per unit change in X_j when all the remaining regressor variables X_i $(i \neq j)$ are held constant. For this reason the parameters β_j, $j = 1,2,\text{---},k$ are often called partial regression coefficients.

Multiple linear regression models are often used as approximating functions. That is, the true functional relationship between y and $X_1, X_2, \text{---}, X_k$ is unknown, but over certain ranges of the regressor variables the linear regression model is an adequate approximation.

Models that are more complex in structure may often still be analyzed by multiple linear regression techniques. For instance the polynomial model of degree k in one variable which has the form:

$$y = \sum_{i=0}^{k} \beta_i X^i$$

can be easily modeled by using the substitution

$$X_i^i = X_j \text{ and } \beta_i = \beta_j \text{ with } X_o = 1$$

Thus the model will be the original linear model

$$y = \beta_o + \beta_1 X_1 + \text{---} + \beta_k X_k$$

Similar transformations could transform the model under consideration into the general form of the linear model, keeping in mind that the linearity of the model means linearity in the β coefficients and not in the independent variables.

So, the basic idea behind multiple linear regression model is to find a linear relation that can adequately approximate an unknown relation between a set of independent variables (k independent variables) and a certain respons y.

Mathematical Model

A scientific model is a representation of some subject of inquiry (such as objects, events, processes, systems) and is used basically for prediction and control. This scientific model is basically divided into three basic types:

1. Iconic model: which pictorially or visually represents certain aspects of a system (as does a photograph or model airplane).

2. Analogue model: which employes one set of properties to represent some other set of properties which the system being studied possesses.

3. Mathematical (or symbolic model): which employs symbols to designate properties of the system under study (by means of a mathematical equation or set of equations).

Consequently the mathematical model often used by scientists has three main types:

1. The function model,

2. The control model,

3. The predictive model

1. The functional model

This kind of model exists if the true functional relationship between a response and the independent variables in a problem is known, so the response could be easily understood, controlled, and predicted. In practice there are few cases which can be easily modeled by a functional model. Even though those models turn to be very complicated, difficult to interpret and usually of nonlinear form. In this kind of models, the the linear regression procedure do not apply or else linear models can be used only as approximations to the correct models in itterative estimation procedures.

2. The control model

Even if it is known completely, the functional model is not always suitable for controlling a response variable. For example if the model contains the ambient temperature as an independent variable in the model, this temperature is not controllable in the sense that other variables in the model are controllable. Thus a model which contains variables under the control of the experimenter is essential for control of a response.

A useful control model can sometimes be constructed by multiple regression techniques, but they should be used carefully because they are very dangerous if improperly used or interpreted.

3. The predictive model

When the functional model is very complex and when the ability to obtain independent estimates of the effects of the control variables is limited, one can often obtain a linear predictive model which, though it may be some senses unrealistic, at least reproduces the main features

of the behavior of the response under study. This type of model is very useful and under certain conditions can lead to real insight into the process or the system under study. It is in the construction of this type of predictive model that multiple regression techniques have their greatest contribution to make. These problems are usually referred to as "problems with messy data". That is, data in which much intercorrelation exists. The predictive model is not necessarily functional and need not be useful for control purposes. This, of course, does not make it useless. If nothing else, it can and does provide guidelines for further experimentation, it pinpoints important variables, and it is a very useful variable screening device.

Non-Normal Error Distribution

1. Consequences of Non-normal Disturbances

Here I'll discuss the violation of the normality assumption of the error term in the regression model:

$$y = X\beta + \varepsilon$$

The discussion will be made in two phases, according if the variance of the error has a finite or infinite variance.

a. Finite Variance Case

In this case the basic definitions and assumptions of the model are exactly the same i.e.

1) Y is called the response, β is the vector of coefficient, x is the independent variables matrix, and ε is the error term.

2) X is nonstochastic of rank (P)

3) The $\lim_{N \to \infty} N^{-1} X'X$ is a finite nonsingular matrix, and

4) The random vector ε is such that

$E(\varepsilon) = 0$ and

$E(\varepsilon \varepsilon') = \sigma^2 I$, σ^2 is finite.

Furthermore if ϵ is normally distributed

1) The L.S. estimator $b = (x'x)^{-1} X'y$ is unbiased minimum variance among the class of unbiased estimators, assymptotically efficient and consistent.

2) The variance estimator

$\hat{\sigma}^2 = (y - xb)' (y - xb)/(N-P)$ is best quadratic unbiased, i.e. it has minimum variance of all estimators of σ^2 that are unbiased and quadratic in y, in addition it is asymptotically efficient and consistent.

3) $b \sim$ normally,

$(N-P)\ \hat{\sigma}^2/\sigma^2 \sim \chi^2_{(N-P)}$

and they are independent

4) The F-test (for $R\beta=r$) and t-test (for the individual coefficients) are valid in finite samples.

On the other hand if ϵ is not normally distributed, we shall have:

1) b is unbiased minimum variance among the class of linear unbiased estimators, and consistent.

2) $\hat{\sigma}^2$ is unbiased and consistent.

3) b and $\hat{\sigma}^2$ are no longer efficient or asymptotically efficient. If the form of error distribution is known, we can use the likelihood function of y to estimate β and σ^2. In this case the estimator for β will be nonlinear in general and, under appropriate regularity conditions $\hat{\beta}$, $\hat{\sigma}^2$ will be asymptotically efficient. Otherwise it is better to use nonlinear robust estimators.

4) b will not be normal and $(N-P)\ \hat{\sigma}^2/\sigma 2$ also will not be χ^2. This means that the F- and t-test for β are not necessarily valid in finite samples.

b. Infinite Variance Case

In this case the error distribution has an infinite variance. As an example for this case take the Pareto distribution

$$f(\varepsilon) = C(\varepsilon - \varepsilon_o)^{-\alpha-1}, C, \varepsilon_o, \alpha \quad \text{are constants}$$

For $\alpha > 2$ the variance does not exist.

Due to the fact that infinite variance distribution has "thicktails", so outliers will frequently occur. As an implementation of these outliers the L.s technique will no longer lead to sensitive estimation of β i.e. β will considerably vary in repeated samples. Also, it will be impossible to get a meaningful estimate for σ^2 and $\hat{\beta}$ will no longer have the minimum variance property which in addition means that F- and t-test will be misleading.

Malinvaud (Ref 58:308) mentioned that, in practice, one can assume that the error distribution is bounded and this will lead to a finite variance. However this will not solve the problem and in case of relatively large number of outliers $\hat{\sigma}^2$ will be unstable in repeated samples and the estimates will behave as if the variance is infinite.

The Double Exponential and L_1 Technique to Estimate β Coefficients

To demonstrate why it may be desirable to use an alternative to least square when the observations are double exponential, consider the simple linear model

$$y_i = \beta_o + \beta_i X_i + \varepsilon_i, \; i = 1.2, ---, n$$

Where the error terms are independent random variables that follow the double exponential distribution.

$$f(\varepsilon_i) = \frac{1}{2\sigma} \; e^{-|\varepsilon_i|/\sigma}, \quad -\infty < \varepsilon_i < \infty$$

The double exponential distribution is more pointed in the middle than the normal and tails go to zero as $|\varepsilon_i|$ goes to infinity. However, since the density function goes to zero as $e^{-|\varepsilon_i|}$ goes to zero, and the normal density function goes to zero as $e^{-\varepsilon_i^2}$ goes to zero, so the double exponential distribution has heavier tails than the normal.

Here, we shall use the method of maximum likelihood to estimate β_o and β_1. The likelihood function is

$$L(\beta_o,\beta_1) = \prod_{i=1}^{n} \frac{1}{2\sigma} e^{-|\epsilon_i|/\sigma}$$

$$= \frac{1}{(2\sigma)^n} e^{-\sum_{i=1}^{n} |\epsilon_i|/\sigma}$$

So to maximize $L(\beta_o,\beta_1)$ is the same as maximizing the exponent $-\sum_{i=1}^{n}|\epsilon_i|/\sigma$ or minimizing $\sum_{i=1}^{n}|\epsilon_i|$, the sum of the absolute errors. Knowing that the method of maximum likelihood applied to the regression model with normal errors leads to the least squares criterion. Thus the assumption of an error distribution with heavier tails than the normal implies that the method of least squares is no longer an optimal estimation technique. However the absolute error criterion would weight outliers far less than would least squares (ϵ_i^2 is much greater than $|\epsilon_i|$ in case of outliers). Minimizing the sum of the absolute errors is often called the L_1-norm regression problem. The least squares is the L_2-norm regressing problem.

The L_1-norm regression problem can be formulated as a linear programming (LP) problem.

Now let X_{ij}, i = 1,2, --- ,n, and j = 1,2, --- ,K denote the set of n observational measurements on k independent variables, and y_i, i =1, --- ,n, denote the associated measurement on the dependent variable (response). The L_1 technique wishes to find the regression coefficient $\hat{\beta}_j$ that:

$$\text{Minimize } \sum_i \left| \sum_j X_{ij} \hat{\beta}_j - y_i \right|$$

Chranes, Cooper and Ferguson (Ref 16) introduced a reduction which can transform the problem into

$$\text{Minimize } \sum_i \epsilon_{1i}^{+} \sum_i \epsilon_{2i}$$

Subject to

$$\Sigma\, X_{ij}\ \hat{\beta}_j + \epsilon_{1i} - \epsilon_{2i} = y_i\ ,\ i = 1,2,\text{---}\ ,n,$$

$\hat{\beta}_j$ is unrestricted,

$$\epsilon_{1i},\ \epsilon_{2i} \geq 0$$

Where ϵ_{1i} is the vertical deviation above the fitted line and ϵ_{2i} is the vertical deviation below the fitted line for i^{th} observation. Thus $\epsilon_{1i} * \epsilon_{2i}$ will be the absolute deviation between the fit $\sum_j X_{ij}\ \hat{\beta}_0$ and y_i. By the nature of the linear programming model, ϵ_{1i} and ϵ_{2i} cannot both be strictly positive in an optimal solution. So, the problem is formulated as L.P problem of the form:

$$\text{Minimize } C_1 Z_1 + \text{----} + C_k\ Z_k$$

subject to

$$Z_1\ a_{1\ell} + \text{--} + Z_k\ a_{\ell k} \begin{cases} \geq d_\ell & \text{if } \ell \in N_1 \\ = d_\ell & \text{if } \ell \in N_2 \end{cases}$$

and

$$Z_h \begin{cases} \geq 0\ \forall\ h \in M_1 \\ \text{Unrestricted } \forall\ h \in M_2 \end{cases}$$

where

M_1, M_2 is a partitioning of the linear relations (mutually exclusive and completely exhaustive partitioning) and similarly N_1, N_2 partions for the set of the variables.

The solution of the model in our case will be

$$X\ \underline{\hat{\beta}} = y$$

Where $\underline{\beta}$ is the vector $(\beta_o, \beta_1, \text{---}, \beta_k')$

It worths here to mention that if the number of observations is large enough, the present model will be somewhat computationally difficult and it will be better to use the dual problem for determining $\hat{\beta}$.

The dual model is still large since it contains k + 2n relations. To reduce it, let

$$f_i = D_i + 1 \qquad i = 1,2,---,n$$

and the dual model will be equivalent to

$$\text{Maximize} \quad \sum_i y_i f_i - \sum_i y_i$$

Subject to

$$\sum X_{ij} f_i \begin{cases} \leq X_{ij} & j \in M_1 \\ = \sum X_{ij} & j \in M_2 \end{cases}$$

$$0 \leq f_i \leq 2 \qquad i = 1,2,---,n$$

Which will give a model with k linear relations and n non-negative bounded variables. This final model could be solved quite rapidly for k (<10) by simplex algorithm for bounded variables problems. On solving this model we can determine the values for $\hat{\underline{\beta}}$.

The Uniform Dist. and Minimax Criterian

Again we shall consider that the error term is distributed uniformally with mean equal to zero and standard deviation equal to unity i.e. $U(-\sqrt{3\sigma}\ ,\ \sqrt{3\sigma}\ )$. Now consider also the case of a simple linear model

$$y_i = \beta_0 + \beta_1 X_i + \epsilon_i\ ,\ i = 1,2,---,n$$

where

$$\epsilon_i \sim U\ (-\sqrt{3\sigma}\ ,\sqrt{3\sigma}\ ),\ \text{then}$$

$$f(\epsilon_i) = \frac{1}{2\sqrt{3\sigma}} \left(I_{(-\sqrt{3\sigma},\sqrt{3\sigma})}(\epsilon_i) \right)$$

where

$I_{(-\sqrt{3\sigma},\sqrt{3\sigma})}(\epsilon_i)$ is the indicator function.

The maximum likelihood function as function of the coefficients

β_0, β_1 is given as

$$L(\beta_o,\beta_1) = \prod_{i=1}^{n} \frac{1}{2\sqrt{3\sigma}} \left(I(\epsilon_i)_{(-\sqrt{3\sigma},\sqrt{3\sigma})} \right)$$

$$= \left(\frac{1}{2\sqrt{3\sigma}}\right)^n \left(\prod_{i=1}^{n} I(\epsilon_i)_{(-\sqrt{3\sigma},\sqrt{3\sigma})} \right)$$

This function will achieve its maximum when the difference between the first and last order statistic will be minimum i.e. the criterion for obtaining a maximum likelihood estimators for β_0 and β_1 will be by minimizing the difference $(\epsilon_{(n)} - \epsilon_{(1)})$ in other words by minimizing the maximum difference of ϵ_i (in absolute value) or equivalently will be to

$$\text{Minimize } \{\text{maximum } |\epsilon_i|\}$$

or in a general multiple linear model will be

$$\underset{\hat{\beta}_j}{\text{Minimize}}\{ \underset{i}{\text{maximum}} \left|\sum_j X_{ij} \hat{\beta}_j - y_i\right|\}$$

Paralleling Kelley transformed this problem into an L.P model:

$$\text{Minimize } \delta \ , \delta \geq 0$$

subject to

$$-\delta \leq \sum_j X_{ij} \hat{\beta}_j - y_i \leq \delta, \ i = 1,2,---,n$$

Where δ is the minimized value of the maximum absolute deviation $\left|\sum_j X_{ij} \hat{\beta}_j - y_i\right|$. Using the same approach as in the case of minimizing the sum of absolute deviation briefly discussed in the double exponential case, the model formulation will be:

$$\text{Minimize } \delta$$

subject to

$$-\sum_j X_{ij} \beta_j + \delta \geq -y_i \ , \ i = 1,2, \ --- \ , \ n$$

$$\sum_j X_{ij} \hat{\beta}_j + \delta \geq y_i \quad i = 1,2,---,n$$

$$\hat{\beta}_j \begin{cases} \geq 0 & \forall\ j \in M_1 \\ \text{unrestricted} & \forall\ j \in M_2 \end{cases}$$

$$\delta \geq 0 \quad ,$$

and the dual formulation will be

$$\text{Maximize} \; - \sum_i y_i \, d_{1i} + \sum_i y_i \, d_{2i}$$

subject to

$$- \sum_i X_{ij} \, d_{1i} + \sum_i X_{ij} d_{2i} \begin{cases} \leq 0 & \forall\ j \in M_1 \\ = 0 & \forall\ j \in M_2 \end{cases}$$

$$\sum_i d_{1i} + \sum_i d_{2i} \leq 1$$

$$d_{1i}, \; d_{2i} \geq 0$$

This dual model is a regular L.P problem in k + 1 relations and could be solved by a standard simplex algorithm. If d_{1i} (d_{2i}) is positive in the optimal solution of the dual problem, then the maximum deviation occurs for the i^{th} point and this point will lie above (below) the fitted line. Thus the solution of this L.P model will give the value of $\hat{\beta}$ as the estimated value of β in our multiple linear regression model.

V. Results

In this part of my study I'm going to summarize the research done to find some technique that could handle the model of 27 observation in 11 independent variables.

The model chosen was obtained from Multiple Listing, Vol. 87 for area 12 (Erie, PA). To search for a technique that will handle such types of multiple linear regression, it was necessary to find some real hyperplane (fit) to take as reference for how good the assumed technique is. In order to do that a least squares regression was performed for the observed Y and X. The coefficient vector β_1 from this model (L.S.) was multiplied by the X matrix after being augmented by a vector of 1's to give vector Y_t which was considered as a real value of Y which gives an exact fit

$$Y_t = \beta_1 X$$

The values of the matrix X and the vector β_1 are shown in Table I.

Description of Methods

The basic idea that was used at the very beginning of the study was to use the Q - statistic introduced by Hogg (Ref 40) and defined as:

$$Q = [\bar{U}(.05) - \bar{L}(.05)]/[\bar{U}(.5) - \bar{L}(.5)]$$

where $\bar{U}(\beta)$ is the average of the largest $n\beta$ order statistics (fractional items are used if $n\beta$ is not an integer) and where $\bar{L}(\beta)$ has a similar definition using the smallest items. The Q statistic was basically used as a discriminator for the error distribution tail length. The reason for choosing Q to be used as a discriminator was due to its convergence properties which are much better than those of the Kurtosis, since Q is a ratio of two linear functions of order statistic. In addition it is

easy to see some similarity between Q and the following measure of tail length of the distribution function f:

$$[F^{-1}(.975) - F^{-1}(.025)]/[F^{-1}).75) - F^{-1}(.25)]$$

The next step was to choose some value for the Q statistic upon which it will be possible to determine the tail length of the error distribution. As a matter of fact the basic idea was to come up with a comparative study between the three known regression techniques discussed earlier: least squares (L_2), Minimization of the absolute deviation (L_1), and Minimization of the maximum error (L_∞). As it was pointed out these three techniques will give maximum likelihood estimators for normal, double exponential and uniform error distributions. Thus these estimators will be of desirable properties expressing the unknown relation. To do that a set of random deviates was generated from the three distributions and added to Y_t in succession to give a new value of Y which is considered as the observed value for Y_t i.e.,

$$Y = Y_t + \varepsilon$$

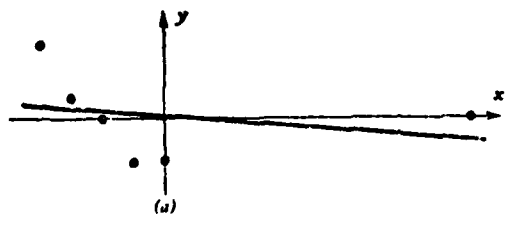
Trying different values for Q statistic to get reasonable bounds (Q_L, Q_U) to discriminate the tail length of the distribution, it turns out to use $Q_L = 2.21$ and $Q_U = 2.81$ i.e., if $Q \leq 2.21$ then we can say that the distribution is uniform, if $2.21 < Q < 2.81$ the distribution is normal, while if $Q \geq 2.81$ the distribution will be double exponential. The number of times these bounds will discriminate the distribution for the known three underlying ones and for monte carl of size 1000 is as follows:

1.	For Uniform:	872 times	uniform
		127 times	normal
		1 time	D.E.
2.	For Normal:	135	uniform
		619	normal
		246	D.E.
3.	For D.E.:	17	uniform
		229	normal
		754	D.E.

As a start for knowing the distribution of the residuals through the use of the Q-statistic, a linear least squares fit was performed for each of 1000 different cases of added error vector from the three considered distributions. Addition of an outlier to one of the observations at multiple values of standard deviations is also considered during the start of the search. The results from this step is shown in Table A-1 for the number of times Q will discriminate each of the residuals distriubtion when the underlying distribution is known. While Table B-1 exhibits the average error sum of squares which is defined as:

$$ESS_{av} = \sum_{i=1}^{1000} (Y_{ti} - Y_i)^2/1000$$

for the different cases discussed above. The steady increase in the values of ESS_{av} with the outlier location with respect to the real line prevail the effect of the so called leverage point effect on the fit which can be demonstrated by the following graph:



Using the residuals from L.S. and making a decision on using L_∞, L_2, or L_1 according to the Q values ($Q_L = 2.21$, $Q_U = 2.81$) is shown in Table B-2. It is clear from this table that ESS_{av} is still steadily increasing since the Q statistic is discriminating the residual most of the times (Table A-1) as normal due to the previously mentioned effect by leverage points. So it seemed to be a better notion to use L_1 instead of using L_2. The way how Q discriminates the distribution for this case is displayed in Table A-2 and ESS_{av} for L_1 is in Table B-3. Using the residuals from L_1 and with two limits again for Q the decision was taken for the choice between L_∞, L_2 or L_1. Table B-4 shows ESS_{av} for this case. It seemed to be a reasonable idea to use only one limit for Q to discriminate between thick tail (D.E.) and thin tail (uniform and normal) distriubutions directly. The resulting ESS_{av} is shown in Table B-5 which improves the values of ESS_{av}.

The previous approaches for taking the problem led to the notion of using one of the robust iterative techniques for handling leverage points and as a result will give what could be called as robust Q that will give a better discrimination for the distribution without being effected by the outliers. As starting step for this approach Huber's function defined by:

$$\Psi(Z) = \begin{cases} Z & \text{if } |Z| \leq 2 \\ 2 \text{ sign } (Z) & |Z| > 2 \end{cases}$$

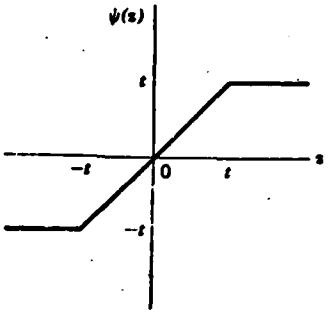
and calculating weight matrix

$$W_{io} = \begin{cases} \dfrac{\Psi[(Y_i - X_i'\hat{\beta}_o)/s]}{(Y_i - X_i'\hat{\beta}_o)/s} & \text{if } Y_i \neq X_i'\hat{\beta}_o \\ 1 & \text{if } Y_i = X_i'\hat{\beta}_o \end{cases}$$

which will give the coefficient vector as

$$\hat{\beta} = (X' W_o X)^{-1} X' W_o Y$$

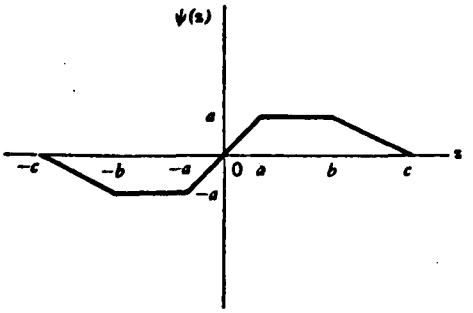
This Huber's function has an influence function which will get rid of the effect of outlier by weighting them with constant weights. The influence function for this case is as hown in the following figure.



The iterative technique for robust regression needs an initial value to start iterations with. In this context a comparison was done between using L.S. or L_1 as initial estimation. The ESS_{av} for these two cases are shown in Table B-5 and Table B-6 respectively. While Table A-3 and Table A-4 show the number of times Q discriminates each distribution. In this case only two iterations were used. Using the residuals from Huber, ESS_{av} is calculated again and displayed in Table B-7. Till this point an improvement in the values for ESS_{av} for outliers at more than 100 S.D. is achieved over using the robust technique alone but still very high value of ESS_{av}. Trying some other robust techniques we ended with using Hampel function defined as:

$$\begin{aligned}
\Psi(Z) &= Z && , |Z| \leq .7 \\
&= 1.7 \text{ sign } (Z) && , 1.7 < |Z| \leq 3.4 \\
&= \frac{1.7 \text{ sign } (Z)\ (8.5 - |Z|)}{5.1} && , 3.4 < |Z| \leq 8.5 \\
&= 0 && , |Z| > 8.5
\end{aligned}$$

which has influence function as shown in the following figure



The ESS_{av} from Hampel with L.S. and L_1 as initial estimation is shown in Table B-8 and Table B-9 respectively. In Table B-10 the resulting ESS_{av} from using Hampel's residual is shown.

Coming to this point we started to search for a different approach to handle our problem. This search basically took 3-phases. Each phase is based on using the residual themselves as our tool to make the decision:

a. Phase I:

Using the residual from L_1 and testing if its greater than 3 S.D. then use L_1 technique, if not use L_2 (least squares) Table B-11 shows ESS_{av} from this phase.

b. Phase II:

In this phase the residual from Hampel iterative technique was used and choice of technique was done as in Phase I. Table B-12 shows the resulting ESS_{av} from this phase.

c. Phase III:

This is really a different approach which gives the nearest fit to the real line throughout our study. The idea is to perform an initial fit and by replacement of all points that are more than 3 S.D. apart from this initial fit back to the initial line. Then by redoing the fit

a lower ESS_{av} could be easily obtained. Table B-13 shows the resulting ESS_{av} from thise phase.

Conclusions

1. Presence of outlier's mode discrimination of distribution outliers. (Tables A-1 - A-4) difficult.
2. With no outlier's Least Squares gave the best fit.
3. Iterating Robust Estimators resulted in no improvement.
4. The Hampel Robust Estimator did not provide outlier protection.
5. The technique of detecting outlier and using L1 if it is greater than 3D and least squares otherwise was the best method of handling the outliers without modifying the data (B-12).
6. The method of mapping the outlier back onto the regression line if residual is greater than 3SD and using L1 gave the best fit using all data points.
7. Alternatively the best fit is obtained by rejecting the points whose residuals are greater than 3SD and repeating the L.S. fit. See 0 line of B-1.

Table 1

The values of independent variables and the calculated $\hat{\beta}$ coefficient used to generate the real line.

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}
	4.9176	1.0	3.4720	0.9980	1.0	7	4	42	3	1	0
	5.0208	1.0	3.5310	1.5000	2.0	7	4	62	1	1	0
	4.5429	2.0	2.2750	1.1750	1.0	6	3	40	2	1	0
	4.5573	1.0	4.0500	1.2320	1.0	6	3	54	4	1	0
	5.0597	1.0	4.4550	1.1210	1.0	6	3	42	3	1	0
	3.8910	1.0	4.4550	0.9880	1.0	6	3	56	2	1	0
	5.8980	1.0	5.8500	1.2400	1.0	7	3	51	2	1	1
	5.6039	1.0	9.5200	1.5010	0.0	6	3	32	1	1	0
	15.4202	2.5	9.800	3.420	2.0	10	5	42	2	1	1
	14.4598	2.5	12.800	3.0000	2.0	9	5	14	4	1	1
	5.8282	1.0	6.4350	1.2250	2.0	6	3	32	1	1	0
	5.3003	1.0	4.9883	1.5520	1.0	6	3	30	1	2	0
	6.2712	1.0	5.5200	0.9750	1.0	6	2	30	1	2	0
	5.9592	1.0	6.6660	1.1210	2.0	6	3	32	2	1	0
X =	5.0500	1.0	5.0000	1.0200	0.0	5	2	46	4	4	1
	8.2464	1.5	5.1500	1.6640	2.0	8	4	50	4	1	0
	6.6969	1.5	6.9020	1.4880	1.5	7	3	22	1	1	1
	7.7841	1.5	7.1020	1.3760	1.0	6	3	17	2	1	0
	9.0384	1.0	7.8000	1.5000	1.5	7	3	23	3	3	0
	5.9894	1.0	5.5200	1.2560	2.0	6	3	40	4	1	1
	7.5422	1.5	4.0000	1.6900	1.0	6	3	22	1	1	0
	8.7951	1.5	9.8900	1.8200	2.0	8	4	50	1	1	1
	6.0931	1.5	6.7265	1.6520	1.0	6	3	44	4	1	0
	8.3607	1.5	9.1500	1.7770	2.0	8	4	48	1	1	1
	8.1400	1.0	8.0000	1.5040	2.0	7	3	3	1	3	0
	9.1416	1.5	7.3262	1.8310	1.5	8	4	31	4	1	0
	12.000	1.5	5.0000	1.2000	2.0	6	3	30	3	1	1

The $\hat{\beta}$ coefficient used as real fit

$$B = \begin{matrix} 3.2621860 \\ .84373136 \\ 8.2369984 \\ .25660890 \\ 14.035590 \\ 1.6223667 \\ -1.0604545 \\ -.32560404 \\ -.074490869 \\ .96740379 \\ 1.0447037 \\ 2.6899793 \end{matrix}$$

Table A-1

Number Q discriminates the tail length for uniform, normal, and double exponential error distributions after performing least squares fit.

NS.D	Underlying Dist. Uniform UR	Uniform N	Uniform D.E	Normal UR	Normal N	Normal D.E	D.E. UR	D.E. N	D.E. D.E
0	222	633	145	126	596	278	61	450	489
1	231	623	146	131	593	276	79	435	486
3	247	600	153	151	584	265	98	458	444
6	283	584	133	206	608	186	162	527	311
9	319	578	103	268	606	126	234	574	192
100	8	992	0	12	988	0	9	991	0
1000	0	1000	0	0	1000	0	0	1000	0
10000	0	1000	0	0	1000	0	0	1000	0

Table A-2

Number of times Q discriminates the tail length for uniform, normal and double exponential error distribution after performing L_1.

NS.D	Underlying Dist.								
	Uniform			Normal			D.E		
	UR	N	D.E	UR	N	D.E	UR	N	D.E
0	246	595	159	136	550	314	64	446	490
1	218	554	228	115	506	379	64	437	499
3	63	327	610	27	252	721	10	139	851
6	23	94	883	14	88	898	3	73	924
9	22	93	885	15	86	899	3	71	926
100	22	93	885	15	86	899	3	70	927
1000	22	93	885	15	86	899	3	70	927
10000	22	93	885	15	86	899	3	70	927

Table A-3

Number of times Q discriminates the tail length for uniform, normal and double exponential error distribution after performing Huber robust technique with L.S as initial estimation.

	Underlying Dist.								
	Uniform			Normal			D.E		
NS.D	UR	N	D.E	UR	N	D.E	UR	N	D.E
0	217	535	248	122	476	402	58	320	622
1	223	515	262	127	461	412	75	299	626
3	239	511	250	147	470	383	98	338	564
6	274	503	223	202	500	298	160	414	426
9	309	510	181	258	515	227	222	479	299
100	3	753	244	10	722	268	4	723	273
1000	0	671	329	0	681	319	0	656	344
10000	0	867	133	0	869	131	0	888	112

Table A-4

Number of times Q discriminates the tail length for uniform, normal, and double exponential error distribution after performing Huber robust technique with L_1 as initial estimation using only one limit Q = 2.81.

NS.D	Underlying Dist. Uniform		Normal		D.E	
	N	D.E	N	D.E	N	D.E
0	752	248	598	402	378	622
1	738	262	588	412	374	626
3	750	250	617	383	436	564
6	777	223	702	298	574	426
9	819	181	773	227	701	299
100	756	244	732	268	727	273
1000	671	329	681	319	656	344
10000	867	133	869	131	888	112

Table B-1

Ess_{av} from L.S with an outlier at N*S D (at multiples of standard deviation) for monte carlo of size 1000.

N	UNIFORM	NORMAL	D.E.
0	12.03	11.69	12.0
1	12.96	12.49	12.9
3	19.86	19.13	19.73
6	42.8	41.68	42.57
9	80.85	79.34	80.51
16	22.84	22.60	22.78
100	84.15E2	84.02E2	84.12E2
1000	83.95E4	83.94E4	83.95E4
10000	83.94E6	83.94E6	83.94E6

Table B-2

ESS_{av} for using L.S and calculating Q from its residuals and choose between L_∞, L_2 or L_1 according to the value of Q (Q_2 = 2.21, Q_U = 2.81) with throwing an outlier at N S.D (multiple of S.D)

N	UNIFORM	NORMAL	D.E
0	14.86	14.03	13.03
1	15.33	14.92	14.40
3	22.83	22.82	21.95
6	46.58	96.63	45.35
9	84.17	81.84	80.15
100	63.91E2	61.90E2	61.44E2
1000	56.34E4	57.17E4	55.67E4
10000	72.78E6	72.94E6	74.54E6

Table B-3

ESS_{av} from L_1 with an outlier at N*S.D (at multiple of standard deviation) for monte carlo of size 1000

N	UNIFORM	NORMAL	D.E
0	20.06	17.22	13.73
1	21.23	18.35	14.95
3	29.87	26.98	23.53
6	52.66	48.53	43.86
9	73.02	67.68	59.09
16	84.93	80.51	70.06
100	85.13	81.09	70.65
1000	85.13	81.09	70.65
10000	85.13	81.09	70.65

Table B-4

ESS_{av} for making decision according to Q statistic calculated from L_1 residuals and using L_∞, L.S or L_1 according to Q value (Q_2 = 2,21m Q_U = 2.81)

N	UNIFORM	NORMAL	D.E
0	13.49	13.67	13.02
1	14.92	14.89	13.93
3	26.25	25.08	23.16
6	54.03	50.05	45.34
9	78.53	72.72	63.20
100	10.41E2	91.75E1	67.68E1
1000	96.53E3	84.76E3	61.29E3
10000	84.77E5	96.53E5	61.27E5

Table B-4 (cont.)

ESS_{av} for using L_1 with Q for making the decision and with only one limit for Q (Q = 2.81) and using either L.S or L_1.

N	UNIFORM	NORMAL	D.E
0	13.73	13.46	12.80
1	14.87	14.49	13.91
3	21.92	21.69	21.56
6	45.67	44.40	44.57
9	81.52	79.42	77.11
100	63.87E2	61.75E2	61.38E2
1000	56.34E4	57.17E4	55.07E4
10000	71.78E6	72.94E6	74.54E6

Table B-5

ESS_{av} from Huber with an outlier at N*S.D (at multiple of standard deviation) for monte carlo of size 1000 using L.S as initial estimation.

1. One Iteration

N	UNIFORM	NORMAL	D.E
0	12.28	11.79	11.53
1	13.22	12.60	12.44
3	20.16	19.31	19.43
6	43.27	42.05	42.67
9	81.52	79.91	80.88
100	81.12E2	80.78E2	80.90E2
1000	78.98E4	78.98E4	78.97E4
10000	78.75E6	78.75E6	78.75E6

2. Two Iterations

N	UNIFORM	NORMAL	D.E
0	12.37	11.85	11.46
1	13.31	12.67	12.37
3	20.26	19.40	19.40
6	43.43	42.18	42.74
9	81.75	80.14	81.06
100	78.37E2	77.82E2	77.97E2
1000	74.48E4	74.49E4	74.47E4
10000	74.78E6	72.94E6	74.54E6

Table B-6

ESS_{av} from Huber with an outlier at N*S.D (at multiple of standard deviation for monte carlo of size 1000 using L_1 as initial estimation

1. One Iteration

N	UNIFORM	NORMAL	D.E
0	12.19	11.72	11.52
1	13.13	12.53	12.43
3	20.03	19.22	19.43
6	43.06	41.91	42.57
9	81.26	79.73	80.68
100	81.12E2	80.78E2	80.90E2
1000	78.98E4	78.93E4	78.97E4
10000	78.75E6	78.75E6	78.75E6

2. Two Iterations

N	UNIFORM	NORMAL	D.E
0	12.25	11.75	11.44
1	13.19	12.57	12.36
3	20.09	19.27	19.40
6	43.16	42.0	42.64
9	81.42	79.89	80.80
100	78.37E2	77.82E2	77.97E2
1000	74.48E4	74.49E4	74.47 E4
10000	74,02E6	74.03E6	74.03E6

Table B-7

ESS_{av} for the decision according to Q calculated from the residual from Huber with L.S as initial estimation and using $Q_L = 2.21$, $Q_U = 2.81$ i.e. using L , L_S or L_1 according to the value of Q (Adaptive)

N	UNIFORM	NORMAL	D.E
0	14.86	14.03	13.03
1	15.87	15.04	14.34
3	22.83	22.82	21.95
6	46.58	96.63	45.35
9	84.17	81.84	80.15
100	63.91E2	61.90E2	61.44E2
1000	56.34E4	57.17E4	55.07E4
10000	72.78E6	72.94E6	74.54E6

Table B-8

ESS_{av} from Hampel with an outlier at N*S.D (at multiple of standard deviation) for monte carlo of size 1000 using L.S as initial estimation.

1. One Iteration

N	UNIFORM	NORMAL	D.E
0	21.50	11.86	11.40
1	13.44	12.68	12.32
3	20.44	19.49	19.41
6	43.68	42.40	42.85
9	82.01	80.29	81.02
16	21:86E1	21.16E1	20.50E1
100	77.78E2	77.49E2	77.53E2
1000	75.52E4	75.56E4	75.56E4
10000	75.39E6	75.39E6	75.39E6

2. Two Iterations

N	UNIFORM	NORMAL	D.E
0	12.68	12.01	11.35
1	13.63	12.84	12.30
3	20.66	19.68	19.46
6	44.01	42.70	43.11
9	82.53	80.76	81.38
16	20.52E1	19.17E1	17.24E1
100	71.52E2	70.97E2	71.02E2
1000	67.70E4	67.61E4	78.60E4
10000	67.29E5	67.29E5	67.29E5

Table B-9

ESS_{av} from Hampel with an outlier at NS.D for monte carlo of size 1000 using L_1 as initial estimation.

1. One iteration

N	UNIFORM	NORMAL	D.E
0	12.23	11.71	11.30
1	13.16	12.53	12.25
3	20.09	19.30	19.35
6	43.09	41.91	42.42
9	80.91	78.87	79.17
16	21.86E1	21.16E1	20.50E1
10000	75.39E6	75.39E6	75.39E6

2. Two Iterations

N	UNIFORM	NORMAL	D.E
0	12.31	11.80	11.24
1	13.25	12.61	12.20
3	20.19	19.43	19.41
6	43.19	41.98	42.50
9	80.65	78.06	77.65
16	20.52E1	19.17E1	17.24E1
100	13.57E2	24.45E2	64.17E2
10000	67.29E6	67.30E6	67.29E6

Table B-10

ESS_{av} from using L_1 or L_2 after making decision according to Q calculated from residuals of Hampel (L.S as initial estimation)

N	UNIFORM	NORMAL	D.E
0	14.45	13.85	12.90
1	15.43	14.88	14.02
3	22.78	22.25	21.85
6	46.50	44.94	44.56
9	81.80	79.11	75.95
100	30.91E2	28.81E2	28.80E2
1000	85.13	81.09	70.65

Table B-11

ESS_{av} using residual from L_1 and making decision to use L_1 or L_2 according if the residual is greater than 3 S.D or not.

N	UNIFORM	NORMAL	D.E
0	13.49	13.67	13.02
1	14.92	14.89	13.93
3	26.25	25.08	23.16
6	54.03	50.05	45.34
9	78.53	72.72	63.20
100	10.41E2	91.75E1	67.68E1
1000	96.53E3	84.76E3	61.29E3
10000	96.53E5	84.77E5	61.27E5

Table B-12

ESS_{av} from using residual from Hampel to make the decision. If residual is greater than 3 S.D use L_1 if not use L.S

N	UNIFORM	NORMAL	D.E
0	12.03	11.69	12.00
1	12.96	12.49	12.90
3	19.86	19.11	19.65
6	42.48	41.04	41.45
9	78.59	74.96	71.66
16	17.23E1	15.31E1	12.89E1
100	10.99E1	13.14E1	12.10E1

Table B-13

ESS_{av} from Phase III (replacement of outliers).

N	UNIFORM	NORMAL	D.E
0	20.06	17.22	13.72
1	21.23	18.35	14.95
3	29.87	26.98	23.53
6	52.66	48.53	43.87
9	73.01	67.66	65.32
16	84.93	80.49	70.02
100	85.08	81.08	70.59

SUPPLEMENTAL BIBLIOGRAPHY

1. Adichie, J. N. "Estimates of Regression Parameters Based on Rank Tests." Annals of Mathematical Statistics, 38: 894-904 (1967).

2. Agee, W.S. and R.H. Turner, "Robust Regression: Some New Methods and Imrpovement of Old Methods", Tech Report, White Sands Missile Range (1978).

3. Andrews, D.F., "A Robust Method for Multiple Linear Regression", Technometrics, 16, 523-531 (1974).

4. Andrews, D.F. et al., Robust Estimates of Location Survey and Advances. Princeton: Princeton University Press, 1972.

5. Anscombe, F.J. "Topics in the Investigation of Linear Relations Fitted by the Method of Least Squares." Journal of the Royal Statistical Society, B, 29: 1-52 (1967).

6. Ashar, V.G. and Wallace, T.D. "A Sampling Study of Minimum Absolute Deviations Estimators," Operations Research, 11: 747-758 (September-October 1963).

7. Barrodale, Ian and Young, Andrew. "Algorithms for Best L_1 and L_∞ Linear Approximations on a Discrete Set." Numerische Mathematik, 8: 295-306 (1966).

8. Bartels, Richard H. and Golub, Gene H. "Stable Numerical Methods for Obtaining the Chebyshev Solution to an Overdetermined System of Equations." ACM Communications, 11: 401-406 (June 1968).

9. -------, "Algorithm 328 Chebyshev Solution to an Overdetermined Linear System (F4)." ACM Communications, 11: 428-430 (June 1968).

10. Belsley, David A., Edwin Kuh, and Roy E. Welsch. Regression Identifying Influential Data and Sources of Collinearity. New York John Wiley & Sons, Inc., 1980.

11. Bourdon, Gerard A. A Monte Carlo Sampling Study for Further Testing of the Robust Regression Procedure Based Upon the Kurtosis of the Least Squares Residuals. MS Thesis. Wright-Patterson AFB OH: Air Force Institute of Technology, 1974.

12. Box, G.E.P. "Non-Normality and TEsts on Variances," Biometrika, 40: 318-335 (1953).

13. ----- and George C. Tiao. "A Further Look at Robustness via Baye's Theorem," Biometrika, 49: 419-431 (1962).

14. Box, G.E.P. and Muller, Mervin E. "A Note on the Generation of Random Normal Deviates." Annals of Mathematical Statistics, 29: 610-611 (1958).

15. Caso, John Robust Estimation Techniques for Location Parameter Estimation of Symmetric Distributions. WPAFB, Ohio: AFIT, March 1972 (Thesis)

16. Charnes, A., Cooper, W.W., and Ferguson, R.O., "Optimal Estimation of Executive Compensation by Linear Programming." Management Science, 1: 138-151 (1955).

17. Clelland, Richard C., John S. deCani, and Francis E. Brown. Basic Statistics with Business Applications. New York: John Wiley & Sons, Inc., 1973.

18. Cramer, Harold. Mathematical Methods of Statistics. Princeton: Princeton University Press, 1946.

19. Crocker, Douglas C. "Linear Programming Techniques in Regression Analysis: the Hidden Danger." AIIE Transactions, 1: 112-126 (June 1969).

20. Crowder, George E., Jr. Adaptive Estimation Based on a Family of Generalized Exponential Power Distribution. MS Thesis. Wright-Patterson AFB OH: Air Force Institute of Technology, 1977.

21. Daniels, Tony E. Robust Estimation of the Generalized T Distribution Using Minimum Distance Estimation. MS Thesis. Wright-Patterson AFB OH: Air Force Institute of Technology, 1980.

22. Denly, L. and W. Larsen, Robust Regression Estimators Compared via Monte Carlo", Comm. in Statistics, Vol. A6, No. 4, 335-362 (1977).

23. Dufton, A. F. "Correlation." Nature, 121: 866 (June 1928).

24. Emshoff, James R., and Sisson, Rogert L. Design and Use of Computer Simulation Models. New York: Macmilland Co., 1970.

25. Feltus, Erasmus E. Predictive Operations and Maintenance Cost Model. Volume II. AFAL-TR-78-49. Wright-Patterson AFB OH: Air Force Avionics Laboratory, 1978.

26. -----, Predictive Operation and Maintenance Cost Model, Volume II. AFAL-TR-79-1120. Wright-Patterson AFB OH: Air Force Institute of Technology, 1979.

27. Fisher, Walter D. "A Note on Curve Fitting with Minimum Deviations by Linear Programming," Journal of the American Statistical Association, 56: 359-362 (June 1961).

28. Forsythe, Alan B. "Robust Estimation of Straight Line Regression Coefficients by Minimizing the p^{th} Power Deviations." Technometrics, 14: 159-166 (1972).

29. Forth, Charles R. Robust Estimation Techniques for Population Parameters and Regression Coefficients. MS Thesis. Wright-Patterson AFB OH: Air Force Institute of Technology, 1974.

30. Fraser, D. "Necessary Analysis/Adaptive Inference," JASA, 71: 99-110 (March 1976)

31. Glahe, F.R. and Hunt, J.G. "The Small Sample Properties of Simulataneous Equation Least Absolute Estimators vis-a-vis Least Squares Estimators." Econometrica, 38: 742-753 (September 1970).

32. Harter, H. Leon. The Method of Least Squares and Some Alternatives. ARL-72-0129. Wright-Patterson AFB OH: Aerospace Research Laboratories, 1972. (AD 75221).

33. -----. More on Robust Estimators of Location, Scale and Regression Parameters (Preliminary Report). Paper presented at the Central Region Meeting of the Institute of Mathematical Statistics, Bowling Green OH, 10-12 June 1974. (Abstract, IMS Bulletin, 3: 120).

34. -----. "Simple Robust Estimators of Location, Scale, and Regression Parameters (Preliminary report)." Presented at the New York joint statistics meetings, December, 1973. (Abstract, IMS Bulletin, 2 214-215).

35. Heltmansperger, T.P. and J.W. McKean, "A Robust Alternative Based on Ranks to Least Squares in Analyzing Linear Models", Technometrics, 19, 275-284 (1979).

36. Hill, B.M. "Foundations for the Theory of Least Squares." Journal of the Royal Statistical Society, B, 31: 89-97 (1969).

37. Hill, R. "Robust Regression When There are Outliers in the Carriers", PhD Dissertation, Department of Statistics, Harvard University (1977).

38. Hillier, Frederick S., and Gerald J. Lieberman. Introduction to Operations Research. San Francisco: Holden-Day, Inc., 1980.

39. Hogg, Robert V. "More Light on Kurtosis and Related Statistics." Journal of the American Statistical Association, 67: 422-424 (1972).

40. ----. Adaptive Robust Procedures: A Partial Review and Some Suggestions for Future Applications and Theory. The University of Iowa, Department of Statistics: February, 1974. (Tech Report No 31).

41. -----. "Some Observations on Robust Estimation," Journal of American Statistical Association, 62: 1179-1186 (December 1967).

42. Huber, Peter J. "Robust Estimation of a Location Parameter," Annals of Mathematical Statistics, 35: 73-101 (1964).

43. -----. "Robust Statistics: A Review, "Annals of Mathematical Statistics, 43: 1041-1067 (1964).

44. -----, "Studentizing Robust Estimates," Nonparametric Techniques in Statistical Inference, edited by M.L. Puri, Cambridge, England: Cambridge University Press, 1970.

45. -----, "Robust Regression: Asympototics, Conjectures and Monte Carlo", Annals of Statistics, 1, No. 5, 799-821 (1973).

46. -----, "Robust Statistical Procedure" Regional Conference Series in Applied Mathematics, 27 (1977).

47. Jackson, Dunham. "Note on the Median of a Set of Numbers," Bulletin of the American Mathematical Society, 27: 160-164 (1921).

48. Jaeckel, L. A. Robust Estimates of Location. MS Thesis. Berkeley CA: University of California, 1969.

49. Jaeckel, Louis A. "Estimating Regression Coefficients by Minimizing the Dispersion of the Residuals." Annals of Mathematical Statistics, 43: 1449-1458, (1972).

50. Johnston, J. Econometric Methods (Second Edition). New York: McGraw-Hill Co., 1972.

51. Jorgenson, Loren W. Robust Estimation of Location and Scale Parameters. Wright-Patterson Air Force Base, Ohio: Air Force Institute of Technology (AU), June 1973. (Thesis)

52. Jureckova, Jana. "Nonparametric Estimate of the Regression Coefficients." Annals of Mathematical Statistics, 42: 1328-1338. (1971).

53. Karst, Otta J. "Linear Curve Fitting Using Least Diviations." Journal of the American Statistical Assocation, 53: 118-132 (March 1958).

54. Kelley, J.E., "An Application of Linear Programming to Curve Fitting." Journal of Industrial and Applied Mathematics, 6: 15-22 (1958).

55. Keonker, R. and Bassett B. "Regression Quantiles" Econometrica, 46 33-50 (1978).

56. Launder, Robert L., and Graham N. Wilkinson, Robustness in Statistics. New York: Academic Press, 1979.

57. Maddala, G. "Econometrics" New York: McGraw-Hill Co., 1977.

58. Malinvaud, E. "Statistical Method in Econometrics" Amsterdam: North Holland, 1970.

59. Maronna, R.A. and V.J. Yohai, "Robust M-estimators for Regression With Contaminated Independent Variables" (1977)

60. McKean, J. and Hettmansperger, T. "A Robust Analysis of the General Linear Model Based on One Step R-estimates" Biometrica, 65, 571-579, 1978.

61. Mendenhall, W., and Scheaffer L., "Mathematical Statistics with Applications. Norta Scituate MA: Duxbury Press, 1973.

62. Moberg, Thomas F., and John S. Ramberg, and Ronald Randles. "An Adaptive Multiple Regression Procedure Based on M-Estimators," Technometrics, 22: 213-224 (1980).

63. Mood, Alexander M. et al. Introduction to the Theory of Statistics, Third Edition. New York: McGraw-Hill, Inc., 1974.

64. Mosteller, F. and Tukey, J.W. Data Analysis and Regression. Mass: Addison Wesely, 1977.

65. Mulzer, Wayne J. Adaptive Estimation of Life Distributions Based on a Family of Half Generalized Exponential Power Distributions. MS Thesis. Wright-Patterson AFB OH: Air Force Institite of Technology, 1977.

66. Narula, S. and Wellington J. "Prediction, Lineary Regression and Minimum Sum of Relative Errors" Technometrics, 19: 185-190, 1977.

67. Parr, William C. Minimum Distance and Robust Estimation. Ph.D. Dissertation. Department of Statistics, Southern Methodist University, 1978.

68. Relles, D.A., "Robust Regression by Modified Least Squares", PhD Dissertation, Department of Statistics, Yale University (1968).

69. Rice, John R., and John S. White. "Norms for Smoothing and Estimation," SIAM Review, 6: 243-256 (July 1964).

70. Romanowski, M. "On the Modified Normal Distributions of Accidental Errors," Metrologia, 1: 74 (1965).

71. Schryer, Normal L. "Certification of Algorithm 328 (F4), Chebyshev Solution to an Overdetermined Linear System." ACM Communications 12: 326 (June 1969).

72. Sen, Pranab Kumar. "Robust Statistical Procedures in Problems of Linear Regression with Special Reference to Quantitative Bioassays, I," Review of the International Statistical Institute, 39: 21-38 (1971).

73. Stigler, Stephen. "Do Robust Estimators Work With Real Data?," Annals of Statistics, 5: 1055-1098 (1977).

74. -----, Simon Newcomb, Percy Daniell and the History of Robust Estimation, 1880-1920. Technical Report No. 319. Department of Statistics, University of Wisconsin, 1972. (AD 757026).

75. -----. "Simon Newcomb, Percy Daniell, and the History of Robust Estimation, 1885-1920, "Journal of the American Statistical Association, 68: 872-879 (1973).

76. Student. "Errors of Routine Analysis," Biometrika, 19: 151-164 (1927).

77. Thiel, Henri. Principles of Econometrics. New York: John Wiley and Sons, Inc., 1972.

78. Tukey, John W. The Future of Data Analysis. Technical Report No. 43. Princeton NJ: Princeton University, Department of Mathematics, 1961. (AD 268852).

79. ----. "A Survey of Sampling from Contaminated Distributions in: Contributions to Probability and Statistics, Calif, Standford University Press., 1960.

80. Wagner, Harvey M. "Linear Programming Techniques for Regression Analysis." Journal of the American Statistical Association, 54: 206-212 (March 1959).

81. Wolfowitz, J. "Estimation by the Minimum Distance Method," Annals of Mathematical Statistics, 5: 9-23 (1953).

Vita

Ahmed Mohamed M. Sultan was born on 22 May 1952 in Cairo, Egypt. After graduating from Military Technical College (MTC) in Cairo as electrical engineer in 1975, he joined the Air Force as a first lieutenant. In 1979 he received an OR Diploma from Cairo University. Then he enrolled in a one year program for preparation to register for a M.S. in OR at the Institute of Statistical Studies and Research of Cairo University. In summer 1981 he began a M.S. Program in OR at Air Force Institute of Technology.

Permament Address: 94 Wassef St.
Ein Shams
Ahmed Essmat St.
Cairo, Egypt

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFIT/GOR/MA/82D-3	2. GOVT ACCESSION NO. AD-A124 678	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) ROBUST MULTIPLE LINEAR REGRESSION		5. TYPE OF REPORT & PERIOD COVERED MS Thesis 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Ahmed Mohamed M. Sultan Maj., Egyptian AF		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Air Force Institute of Technology AFIT/EN Wright-Patterson AFB, OH 45433		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS		12. REPORT DATE December 1982 13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

Approved for public release; IAW AFR 190-17
Frederick C. Lynch, Major, USAF
Director of Public Affairs

18. SUPPLEMENTARY NOTES

Approved for public release: IAW AFR 190-17.
LYNN E. WOLAVER
Dean for Research and Professional Development
Air Force Institute of Technology (ATC)
Wright-Patterson AFB OH 45433

4 JAN 983

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Multiple Linear Regression
Robust Procedures
Robust Estimators
Robust Regression
Least Squares
L_1-norm

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

An extensive Monte Carlo Analysis is conducted to determine the performance of robust linear regression techniques with and without outliers. Thirteen methods of regression are compared including least squares and minimum absolute deviation. The classical robust techniques of Huber, Hampel were studied and robust techniques using the Q-statistic as discriminant were introduced.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

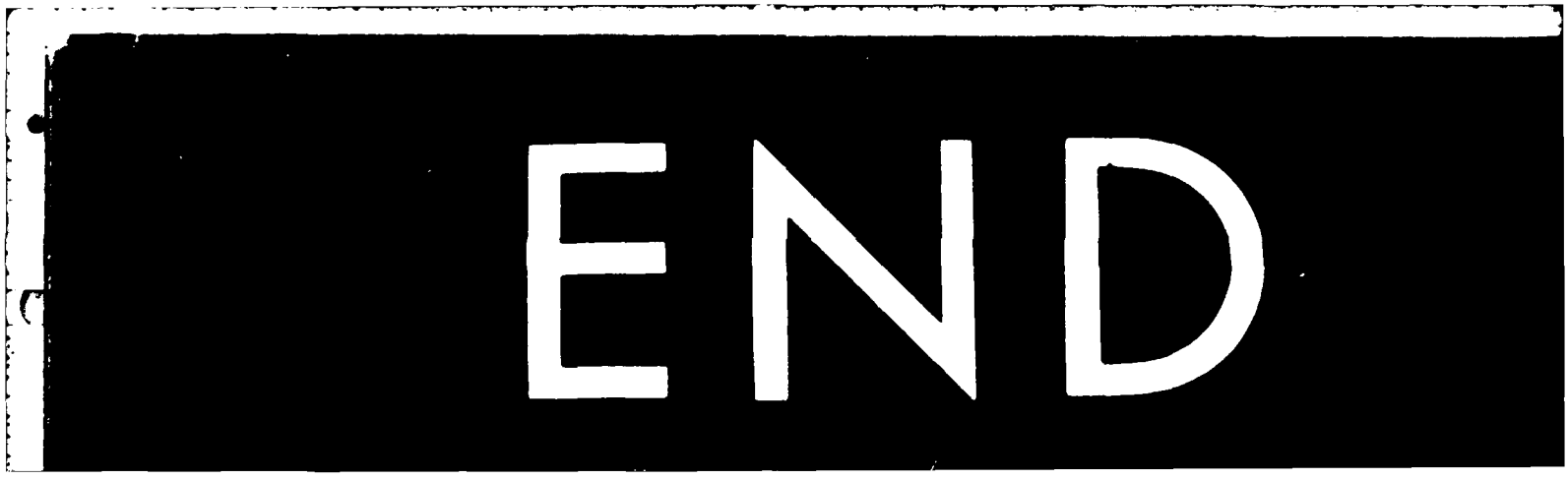
UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Block #20

The model studied contained eleven variables with 27 observations. The error distributions considered were uniofrmly, normally and double exponentially distributed.

Least squares gave the best fit without outliers. In the presence of gross outliers a rejection of outlier technique gave the best fit.


END