Robust Regression by Trimmed Least-Squares Estimation

by

David Ruppert

and

Raymond J. Carroll

Koenker and Bassett (1978) introduced regression quantiles and suggested a method of trimmed least-squares estimation based upon them. We develop asymptotic representations of regression quantiles and trimmed least-squares estimators. The latter are robust, easily computed estimators for the linear model. Moreover, robust confidence ellipsoids and tests of general linear hypotheses based on trimmed least squares are available and are computationally similar to those based on ordinary least squares.

David Ruppert is Assistant Professor and Raymond J. Carroll is Associate Professor, both at the Department of Statistics, the University of North Carolina at Chapel Hill. This research was supported by the National Science Foundation Grant NSF MCS78-01240 and the Air Force Office of Scientific Research under contract AFOSR-75-2796.

The authors wish to thank Shiv K. Aggarwal for his programming assistance.

1. INTRODUCTION

We will consider the linear model

$$\underset{\sim}{y} = X \underset{\sim}{\beta} + \underset{\sim}{e} \tag{1.1}$$

where $\underset{\sim}{y} = (y_1,\ldots,y_n)'$, X is a $n\times p$ matrix of known constants, $\underset{\sim}{\beta} = (\beta_1,\ldots,\beta_p)'$ is a vector of unknown parameters, and $\underset{\sim}{e} = (e_1,\ldots,e_n)$ where $e_1,\ldots,e_n$ are i.i.d. with distribution F. Recently there has been much interest in estimators of $\underset{\sim}{\beta}$ which do not have two serious drawbacks of the least-squares estimator, inefficiency when the errors have a distribution with heavier tails than the Gaussian and great sensitivity to a few outlying observations. In general, such methods, which fall under the rubic of robust regression, are extensions of techniques originally introduced for estimating location parameters. In the location model three broad classes of estimators, M,L and R estimators are available. See Huber (1977) for an introduction to these classes.

For the linear model, M and R estimators have been studied extensively. Until recently only Bickel (1973) has studied regression analogs of L-estimators. His estimates have attractive asymptotic efficiencies, but they are computationally complex. Moreover, they are not equivariant to reparametrization (see remarks after his Theorem 3.1).

There are compelling reasons for extending L-estimators to the linear model. For the location problem, L-estimators, particularly trimmed means, are attractive to those working with real data. Stigler (1977) recently applied robusts estimators to historical data and concluded that "the 10% trimmed mean (the smallest nonzero trimming percentage included in the study) emerges as the recommended estimator."

Jaeckel (1971) has shown that if F is symmetric then for each L-estimator of location there are asymptotically equivalent M and R estimators. However,

without knowledge of F it is not possible to match up a L-estimator with its corresponding M and R estimators. For example, trimmed means are asymptotically equivalent to Huber's M-estimate which is the solution b to

$$\sum_{i=1}^{n} \rho((X_i-b)/s_n) = \min!$$

where

$$\rho(x) = x^2/2 \quad \text{if} \quad |x| \leq k$$

$$= k|x| \quad \text{if} \quad |x| > k \ .$$

The value of k is determined by the trimming proportion α of the trimmed mean, F, and the choice of s_n. In the scale non-invariant case ($s_n \equiv 1$) $k = F^{-1}(1-\alpha)$. The practicing statistician, who known only his data, may find his intuition of more assistance when choosing α compared with k.

Recently Koenker and Bassett (1978) have extended the concept of quantiles to the linear model. Let $0 < \theta < 1$. Define

$$\phi_\theta(x) = \theta \quad \text{if} \quad x \geq 0 \qquad (1.2)$$

$$= \theta - 1 \quad \text{if} \quad x < 0, \text{ and}$$

$$\rho_\theta(x) = x\ \phi_\theta(x) \ .$$

Then a θth regression quantile, $\hat{\beta}(\theta)$, is any value of $\underset{\sim}{\rho}$ which solves

$$\sum_{i=1}^{n} \rho_\theta(y_i - \underset{\sim}{x}_i\ \underset{\sim}{b}) = \min! \qquad (1.3)$$

Their Theorem 4.2 shows that regression quantiles have asymptotic behavior similar to sample quantiles in the location problem. Therefore, L-estimates consisting of linear combinations of a fixed number of order statistics, for example the median, trimean, and Gastwirth's estimator, are easily extended

to the linear model and have the same asymptotic efficiencies as in the location model. Moreover, as they point out, regression quantiles can be easily computed by linear programming techniques. They also suggest the following trimmed least-squares estimators, call it $\hat{\underset{\sim}{\beta}}_T(\alpha)$: remove from the sample any observations whose residual from $\hat{\underset{\sim}{\beta}}(\alpha)$ is negative or whose residual from $\hat{\underset{\sim}{\beta}}(1-\alpha)$ is positive and calculate the least-squares estimator using the remaining observations. They conjecture that if $\lim_{n\to\infty} n^{-1}(X'X) = Q$(positive definite), then the variance of $\hat{\underset{\sim}{\beta}}_T(\alpha)$ is $n^{-1}\sigma^2(\alpha,F)Q^{-1}$, where $n^{-1}\sigma^2(\alpha,F)$ is the variance of an α-trimmed mean from a population with distribution F.

In this paper we develop asymptotic expansions for $\hat{\underset{\sim}{\beta}}(\theta)$ and $\hat{\underset{\sim}{\beta}}_T(\alpha)$ which provide simple proofs of Koenker and Bassett's Theorem 4.2 and their conjecture about the asymptotic covariance of $\hat{\underset{\sim}{\beta}}_T(\alpha)$.

In the location model, if F is asymmetric then there is no natural parameter to estimate. In the linear model, if the design matrix is chosen so one column, say the first, consists entirely of ones and the remaining columns each sum to zero, then our expansions show that for each $0 < \alpha < \frac{1}{2}$

$$n^{\frac{1}{2}}(\hat{\underset{\sim}{\beta}}_T(\alpha) - \underset{\sim}{\beta} - \underset{\sim}{\delta}(\alpha)) \xrightarrow{L} N(0, Q^{-1}\sigma^2(\alpha,F))$$

where $\underset{\sim}{\delta}(\alpha)$ is a vector whose components are all zero except for the first. Therefore, the ambiguity about the parameter being estimated involves only the intercept and none of the slope parameters.

Additionally we present a large sample theory of confidence ellipsoids and general linear hypothesis testing, which is quite similar to that of least squares estimation with Gaussian errors.

The close analogy between the asymptotic distributions of trimmed means of our trimmed least-squares estimator, $\hat{\underset{\sim}{\beta}}_T(\alpha)$ is remarkable. Other reasonable definitions of a trimmed least squares estimator do not have this property.

For example, define an estimator of β, call it K50, as follows: compute the residuals from $\hat{\underset{\sim}{\beta}}(.5)$ and after removing the observations with $k = [\alpha n]$ smallest and k largest residuals compute the least squares estimator. The asymptotic behavior of K50, which is complicated and is not identical to that of $\hat{\underset{\sim}{\beta}}_T(\alpha)$, will be reported elsewhere.

Section 2 presents the formal model being considered, the main results are in Section 3, and several examples are found in Section 4. Proofs are in the appendix.

2. NOTATION AND ASSUMPTIONS

Recall the form (1.1) of our model. Although $\underset{\sim}{y}$, X, and $\underset{\sim}{e}$ will depend on n we will not make that explicit in the notation. Let $\underset{\sim}{x}_i = (x_{i1},\ldots,x_{ip})$ be the i^{th} row of X. Assume $x_{i1} = 1$, $i = 1,\ldots,n$,

$$\lim_{n\to\infty} \left(\max_{j\le p;\, i\le n} (n^{-\frac{1}{2}}|x_{ij}|)\right) = 0, \tag{2.1}$$

and there exists positive definite Q such that

$$\lim_{n\to\infty} n^{-1}(X'X) = Q\ . \tag{2.2}$$

Without loss of generality, we assume $F^{-1}(\frac{1}{2}) = 0$. Let $N_p(\underset{\sim}{\mu},\Sigma)$ denote the p-variate Gaussian distribution with mean $\underset{\sim}{\mu}$ and covariance Σ. Assume $0 < \theta < 1$, $0 < \theta_1 < \ldots < \theta_m < 1$, and $0 < \alpha_1 < \frac{1}{2} < \alpha_2 < 1$. For $0 < p < 1$, let $\xi_p = F^{-1}(p)$ and define $\xi_1 = \xi_{\alpha_1}$ and $\xi_2 = \xi_{\alpha_2}$. Assume F is everywhere continuous and has a continuous positive density in neighborhoods of ξ_θ, $\xi_{\theta_1},\ldots,\xi_{\theta_n}$, ξ_1, and ξ_2. Whenever r a number, $\underset{\sim}{r}$ is a p-dimensional vector with $(\underset{\sim}{r})_1 = r$ and $(\underset{\sim}{r})_j = 0$, $j = 2,\ldots,p$. Let I_p be the p×p identity matrix.

3. MAIN RESULTS

In Section 1 we defined a θth regression quantile to be any value of $\underset{\sim}{b}$ which solves (1.3). We now assume that some rule has been imposed which for each θ selects a single solution $\underset{\sim}{b}$ which we denote by $\hat{\underset{\sim}{\beta}}(\theta)$. Asymptotic results do not depend on the rule used.

Theorem 1: *The solution* $\hat{\underset{\sim}{\beta}}(\theta)$ *of* (1.3) *satisfies*

$$n^{-\frac{1}{2}} \sum_{i=1}^{n} \underset{\sim}{x}_i \psi_\theta(y_i - \underset{\sim}{x}_i \hat{\underset{\sim}{\beta}}(\theta)) \xrightarrow{p} \underset{\sim}{0} .$$

Define $\underset{\sim}{\beta}(\theta) = \underset{\sim}{\beta} + \underset{\sim}{\xi}_\theta$. The next theorem shows that $\hat{\underset{\sim}{\beta}}(\theta) - \underset{\sim}{\beta}(\theta)$ is essentially a sum of i.i.d. random variables.

Theorem 2: *The following representation holds:*

$$n^{\frac{1}{2}}(\hat{\underset{\sim}{\beta}}(\theta) - \underset{\sim}{\beta}(\theta)) = n^{-\frac{1}{2}}(f(\xi(\theta))^{-1} Q^{-1} \sum_{i=1}^{n} \underset{\sim}{x}_i \psi_\theta(e_i - \xi(\theta)) + o_p(1).$$

Theorem 2 provides an easy proof of Theorem 4.2 of Koenker and Bassett (1978) which we state as a corollary.

Corollary 1: *Let* $\Omega = \Omega(\theta_1, \ldots, \theta_m; F)$ *be the symmetric* m×m *matrix defined by*

$$\Omega_{ij} = \frac{\theta_i(1-\theta_j)}{f(\xi(\theta_i))f(\xi(\theta_j))} \qquad 1 \leq i \leq j \leq m .$$

Then

$$n^{\frac{1}{2}}(\underset{\sim}{\beta}(\theta_1) - \underset{\sim}{\beta}(\theta_1), \ldots, \underset{\sim}{\beta}(\theta_m) - \underset{\sim}{\beta}(\theta_m)) \xrightarrow{L} N_{mp}(\underset{\sim}{0}, \Omega \otimes Q^{-1}) . \qquad (3.1)$$

We now define trimmed least-squares estimators. Since F is not assumed to be symmetric it is natural to allow asymmetric trimming. Let $\underset{\sim}{\alpha} = (\alpha_1, \alpha_2)$ and define $\hat{\underset{\sim}{\beta}}_T(\underset{\sim}{\alpha})$ to be a least-squares estimator calculated after removing all observations satisfying

$$y_i - \underset{\sim}{x}_i \hat{\underset{\sim}{\beta}}(\alpha_2) \geq 0 \quad \text{or} \quad y_i - \underset{\sim}{x}_i \hat{\underset{\sim}{\beta}}(\alpha_1) \leq 0 \tag{3.2}$$

(Asymptotic results are unaffected by requiring strict inequalities in (3.2), which is Koenker and Bassett's suggestion.) Let $a_i = 0$ or 1 according as i satisfies (3.2) or not and let A be the n×n diagonal matrix with $A_{ii} = a_i$. Then

$$\hat{\underset{\sim}{\beta}}_T(\underset{\sim}{\alpha}) = (X'AX)^{-}(X'A\underset{\sim}{y})$$

where $(X'AX)^{-}$ is a generalized inverse of (X'AX). (For n sufficiently large X'AX will be invertible.) Let

$$\begin{aligned} \phi(x) &= \xi_1/(\alpha_2-\alpha_1) \quad \text{if} \quad x < \xi_1 \\ &= x/(\alpha_2-\alpha_1) \quad \text{if} \quad \xi_1 \leq x \leq \xi_2 \\ &= \xi_2/(\alpha_2-\alpha_1) \quad \text{if} \quad \xi_2 < x \ . \end{aligned} \tag{3.3}$$

Define

$$\underset{\sim}{\delta}(\underset{\sim}{\alpha}) = (\alpha_2-\alpha_1)^{-1}\int_{\xi_1}^{\xi_2} x \, dF(x), \quad \underset{\sim}{\beta}_T(\underset{\sim}{\alpha}) = \underset{\sim}{\beta} + \underset{\sim}{\delta}(\underset{\sim}{\alpha}) \ ,$$

and

$$\sigma^2(\underset{\sim}{\alpha},F) = (\alpha_2-\alpha_1)^{-2}\left(\int_{\xi_1}^{\xi_2} (x-\delta(\underset{\sim}{\alpha}))^2 dF(x) + \alpha_1\xi_1^2 + (1-\alpha_2)\xi_2^2 - ((1-\alpha_2)\xi_2+\alpha_1\xi_1)^2\right) .$$

By deWet and Venter (1974, equation (6)), $\sigma^2(\underset{\sim}{\alpha},F)$ is the asymptotic variance of a trimmed mean with trimming proportions α_1 and $1-\alpha_2$ from a population with distribution F.

Theorem 3: *We have the asymptotic expansion*

$$n^{\frac{1}{2}}(\hat{\underset{\sim}{\beta}}_T(\underset{\sim}{\alpha}) - \underset{\sim}{\beta}) = Q^{-1}n^{-\frac{1}{2}}\{\sum_{i=1}^{n} \underset{\sim}{x}_i'(\phi(e_i) - E\phi(e_i) + \delta(\underset{\sim}{\alpha})\} + o_p(1) \ . \quad (3.4)$$

Therefore, if

$$\sum_{i=1}^{n} x_{ij} = 0 \quad \textit{for} \quad j = 2,\ldots,p \ , \quad (3.5)$$

then

$$\sqrt{n} \ \ (\hat{\underset{\sim}{\beta}}_T(\underset{\sim}{\alpha}) - \underset{\sim}{\beta}_T(\underset{\sim}{\alpha})) \xrightarrow{L} N_p(\underset{\sim}{0}, \ \sigma^2(\underset{\sim}{\alpha},F)Q^{-1}) \ . \quad (3.6)$$

Expression (3.4) is similar to a result of deWet and Venter (1974, equation (5)). Note that (3.6) verifies Koenker and Bassett's hypothesis on the covariance of $\hat{\underset{\sim}{\beta}}_T(\underset{\sim}{\alpha})$. Moreover, since $\underset{\sim}{\beta}_T(\underset{\sim}{\alpha})_i = \underset{\sim}{\beta}_i$ for $i = 2,\ldots,p$ the bias of $\hat{\underset{\sim}{\beta}}_T(\underset{\sim}{\alpha})$ for $\underset{\sim}{\beta}$ involves only the intercept, $\underset{\sim}{\beta}_1$, and not the slopes. Also $\underset{\sim}{\beta}_T(\underset{\sim}{\alpha})_1 = \underset{\sim}{\beta}$, if F is symmetric.

We will now show that $\sigma^2(\alpha,F)$ can be estimated consistently.

Theorem 4: *Let* S *be the sum of squares for residuals calculated from the trimmed sample, i.e.*

$$S = \underset{\sim}{y}'A(I_p - X(X'AX)^{-}X')A\underset{\sim}{y} \ .$$

Let $a_j = (\hat{\underset{\sim}{\beta}}(\alpha_j) - \hat{\underset{\sim}{\beta}}(\underset{\sim}{\alpha}_j))_1$ *for* $j = 1,2$, *and*

$$s^2(\underset{\sim}{\alpha},F) = (n-p)^{-1}(\alpha_2-\alpha_1)^{-2} \ (S + \alpha_1 a_1^2 + (1 - \alpha_2)a_2^2 - (\alpha_1 a_1 + (1 - \alpha_2)a_2)^2).$$

Then

$$s^2(\underset{\sim}{\alpha},F) \xrightarrow{P} \sigma^2(\underset{\sim}{\alpha},F) \ .$$

We have seen that applying least squares to the trimmed sample procedures a robust point estimator. We will now see that one can also construct confidence

ellipsoids and test general linear hypotheses by applying simple modifications of least-squares methods to the trimmed sample.

Theorem 5: *Suppose* m *is the number of observations which have been removed by trimming. For* $0 < \epsilon < 1$, *let* $F(n_1,n_2,\epsilon)$ *denote the* $(1-\epsilon)$ *quantile of the* F *distribution with* n_1 and n_2 *degrees of freedom and let*

$$d(n_1,n_2,\epsilon) = (\alpha_2-\alpha_1)^{-1} S^2(\underset{\sim}{\alpha},F)\ n_1\ F(n_1,n_2,\epsilon) \ .$$

Suppose for some integer ℓ, K *and* $\underset{\sim}{c}$ *are matrices of size* $\ell\times p$ *and* $\ell\times 1$, *respectively, and that* K *has rank* ℓ. *If* $K'\ \underset{\sim}{\beta}_T(\underset{\sim}{\alpha}) = \underset{\sim}{c}$, *then*

$$\lim_{n\to\infty} P\{(K'\hat{\underset{\sim}{\beta}}_T(\underset{\sim}{\alpha})-\underset{\sim}{c})'[K'(X'AX)^{-1}K]^{-1}\ (K'\hat{\underset{\sim}{\beta}}_T(\underset{\sim}{\alpha})-\underset{\sim}{c}) \geq d(\ell,\ n-m-p,\epsilon)\} = \epsilon \qquad (3.7)$$

Letting $K = I_p$ and $\underset{\sim}{c} = \underset{\sim}{\beta}_T(\underset{\sim}{\alpha})$ the confidence ellipsoid

$$(\hat{\underset{\sim}{\beta}}_T(\underset{\sim}{\alpha})-\underset{\sim}{\beta}_T(\underset{\sim}{\alpha}))'(X'AX)(\hat{\underset{\sim}{\beta}}_T(\underset{\sim}{\alpha})-\underset{\sim}{\beta}_T(\underset{\sim}{\alpha})) \leq d(p,\ n-m-p,\epsilon) \qquad (3.8)$$

for $\underset{\sim}{\beta}_T(\underset{\sim}{\alpha})$ has an asymptotic confidence coefficient of $(1-\epsilon)$. See Scheffé (1959) for a discussion of confidence ellipsoids and their use. Moreover, if we test

$$H_0:\ K'\ \underset{\sim}{\beta}_T(\underset{\sim}{\alpha}) = \underset{\sim}{c}$$

versus (3.9)

$$H_1:\ K'\ \underset{\sim}{\beta}_T(\underset{\sim}{\alpha}) \neq \underset{\sim}{c}$$

by rejecting H_0 whenever

$$(K'\ \hat{\underset{\sim}{\beta}}_T(\underset{\sim}{\alpha})-\underset{\sim}{c})'\ [K'(X'AX)^{-1}K]^{-1}\ (K'(\hat{\underset{\sim}{\beta}}_T(\underset{\sim}{\alpha})-\underset{\sim}{c}) \geq d(\ell,\ n-m-p,\ \epsilon) \qquad (3.10)$$

then the asymptotic size of our test is ϵ. Most hypotheses of interest in regression analysis are special cases of (3.9). See Searle (1971, section 3.6)

for further discussion.

Of course, in the special cases where $\alpha_1 = 0$, $\alpha_2 = 1$ (so m = 0 and A = I) and F is Gaussian, (3.8) is an exact 1-ϵ confidence ellipsoid and (3.10) is and exact size ϵ test.

Finally let us demonstrate that $\hat{\beta}(\theta)$ and $\hat{\beta}_T(\alpha)$ are equivariant under the transformation which maps β to $A^{-1}\beta$ and X to XA, where A is any p×p nonsingular matrix. The equivariance of $\hat{\beta}(\theta)$ is (iv) of Koenker and Bassett's Theorem 3.2 (1978). Thus the residuals are invariant, which implies that the observations trimmed are the same for the original model and the transformed model. Then the equivariance of $\hat{\beta}_T(\alpha)$ follows from the equivariance of the least squares estimator.

4. EXAMPLES

By presenting a plot of regression quantiles for a simple linear regression example and comparing trimmed least squares with several other estimation techniques in two cases of multiple regression, we hope to further illustrate the utility of regression quantiles. All computations where performed on the IBM 360 at the University of North Carolina at Chapel Hill. Regression quantiles were computed using the linear programming routine LPMPS described by McKeown and Rubin (1977). The SAS package was used for least-squares calculations.

For any function ψ, following Huber (1977) we say b is a scale invariant M-estimate corresponding to ψ if (b,σ) is a solution to the equations

$$\sum_{i=1}^{n} \psi((y_i - x_i\, b)/\sigma) = 0 \quad \text{and}$$

$$\sum_{i=1}^{n} \psi^2((y_i - x_i\, b)/\sigma) = (n-p) \int_{-\infty}^{\infty} \psi^2(x)\, d\Phi(x)$$

where Φ is the standard Gaussian distribution function. We will consider two specific M-estimates, the Huber where $\psi(x) = \max(-1.5, \min(x, 1.5))$ and the Hampel where

$$\begin{aligned} \psi(x) &= x && \text{if} \quad 0 \leq x \leq 1.5 \\ &= 1.5 && \text{if} \quad 1.5 \leq x \leq 3.5 \\ &= (8-x)/3 && \text{if} \quad 3.5 \leq x \leq 8 \\ &= 0 && \text{if} \quad 8 \leq x \end{aligned} \tag{4.1}$$

and $\psi(-x) = -\psi(x)$. The Huber and Hampel estimators were computed iteratively using a program written by one of the authors (RJC) and adapted from work of Lenth (1976).

The dependent variable of the first example is average biweekly water temperature in the shrimp nursery areas of the Pamlico Sound, North Carolina, from March to May of 1972 to 1977 with several biweekly periods missing. The single independent variable is average air temperature at Cape Hatteras for the same biweekly periods. Figure A is a scatter plot of the data with the regression quantile lines for $\theta = .05$, .50, and .95. (These and all regression quantiles mentioned in this section are unique.) There appears to be no serious outliers in this data set and the least squares, Huber, Hampel, and trimmed least squares estimates are very similar for both slope and intercept.

In the second example, average biweekly water salinity in Pamlico Sound is the dependent variable. The three independent variables are a linear time trend, water discharged by two rivers emptying into the Sound, and salinity in the Sound lagged one month. The data is from the shrimp harvest seasons of 1972 to 1977.

The last example is the stackloss data of Brownlee (1965, section 13.12), which has been further studied by Daniel and Wood (1971, chapter 5), and

Andrews (1974), who reviews the earlier work. We will compare several estimators which are described in table 1. For each estimation, let $\{r_i\}$ be the residuals, let m be the median of $\{r_i\}$, and then define MAD to be the median of $\{|r_i-m|\}$. Of course, as is usual with real data it is impossible to compare estimators with respect to estimation error since the true parameters are unknown. Instead, we rank the estimators according to their MAD and thus compare their relative success at fitting the data.

Table 2 presents the MAD's and their ranks for both examples. Least squares performs relatively poorly as do Huber and Hampel, and K15 does about as well as Huber and Hampel. In the introduction, it was mentioned that the performance of M-estimators depends upon, among other factors, the method of scale estimation. This should be reemphasized. The Hampel and Huber estimators may perform well on this data set, if scale is estimated in a manner other than ours.

The K05 estimator does somewhat better than Huber and Hampel. Since in these two cases (and often where n is small), K05 is least squares after removing 2p (here 8), it might be expected that K05 would fair poorly since the proportion of observations removed is large. Yet K05 does well.

The K50 estimator's performance makes it worthy of further study.

The Andrews estimator is admirably suited to the stackloss data. However, this data is somewhat extreme in having four apparent outliers in 21 observations; here LAD outperforms all estimators except Andrews. Since we did not program a routine for computations of the Andrews estimator, it is not available for the shrimp harvest data. LAD ranks third there. Perhaps LAD should be more widely used despite its relative inefficiency in the Gaussian model; it works well with some "dirty" data sets.

APPENDIX

Lemma A.1: *With probability one there exists no vector,* $\underset{\sim}{b}$, *and* p+1 *rows of* X, $\underset{\sim}{x}_{i(1)},\ldots,\underset{\sim}{x}_{i(p+1)}$, *such that* $y_i = \underset{\sim}{x}_{i(j)}\underset{\sim}{b}$ *for* $j = 1,\ldots,p+1$.

Proof: Routine. Use the continuity of F.

Proof of Theorem 1: Let $G(\underset{\sim}{b})$ denote the left hand side of (1.3) and $\underset{\sim}{f}_i$ be the p-vector with $(\underset{\sim}{f}_i)_j = 0$ or 1 according as $i \neq j$ or $i=j$. Define

$$G_j'(\underset{\sim}{b}) = \lim_{h\to 0} \frac{G(\underset{\sim}{b} + (1-\theta)h\ \underset{\sim}{f}_j) - G(\underset{\sim}{b} - \theta\ h\ \underset{\sim}{f}_j)}{h} .$$

Notice that the limit always exists. Let $H_j(a) = -G_j'(\hat{\underset{\sim}{\beta}}(\theta) + a\ \underset{\sim}{f}_j)$. Clearly,

$$H_j(0) = \sum_{i=1}^{n} x_{ij}\ \psi_\theta(y_i - \underset{\sim}{x}_i\ \hat{\underset{\sim}{\beta}}(\theta)) .$$

For $\epsilon > 0$, we have

$$H_j(-\epsilon) \leq H_j(0) \leq H_j(\epsilon)$$

and

$$H_j(\epsilon) \geq 0 \text{ and } H_j(-\epsilon) \leq 0 .$$

Thus, for all $\epsilon > 0$,

$$|H_j(0)| \leq H_j(\epsilon) - H_j(-\epsilon) \tag{A.1}$$

$$= \sum_{i=1}^{n} x_{ij}\ \{\psi_\theta(y_k - \underset{\sim}{x}_i\ \hat{\underset{\sim}{\beta}}(\theta) + \epsilon\ x_{ij}) - \psi_\theta(y_i - \underset{\sim}{x}_i\ \hat{\underset{\sim}{\beta}}(\theta) - \epsilon\ x_{ij})\} .$$

Letting $\epsilon \to 0$ in (A.1), we see that

$$|H_j(0)| \leq \sum_{i=1}^{n} |x_{ij}|\ I(y_i - \underset{\sim}{x}_i\ \hat{\underset{\sim}{\beta}}(\theta) = 0) .$$

Theorem 1 now follows from Lemma A.1.

Lemma A.2: *For* $\underset{\sim}{\Delta} \in R^p$, *define*

$$M(\underset{\sim}{\Delta}) = n^{-\frac{1}{2}} \sum_{i=1}^{n} \underset{\sim}{x}_i \psi_\theta(e_i - \underset{\sim}{x}_i \underset{\sim}{\Delta} n^{-\frac{1}{2}} - \xi_\theta)$$

and

$$U(\underset{\sim}{\Delta}) = M(\underset{\sim}{\Delta}) - M(\underset{\sim}{0}) - E(M(\underset{\sim}{\Delta}) - M(\underset{\sim}{0})) .$$

For all $L > 0$

$$\sup_{0 \le ||\underset{\sim}{\Delta}|| \le L} ||U(\underset{\sim}{\Delta})|| = o_p(1) \tag{A.2}$$

and

$$\sup_{0 \le ||\underset{\sim}{\Delta}|| \le L} ||M(\underset{\sim}{\Delta}) - M(\underset{\sim}{0}) + f(\xi_\theta)Q\underset{\sim}{\Delta}|| = o_p(1) . \tag{A.3}$$

Proof: Equation (A.2) follows from Lemma 4.1 of Bickel (1975) since F is continuous near ξ_θ. Equation (A.3) is verified by noting that as $h \to \infty$

$$E(M(\underset{\sim}{\Delta}) - M(\underset{\sim}{0})) \to -f(\xi_\theta)Q\underset{\sim}{\Delta} .$$

Remark: Equation (A.2) is the conclusion of Theorem 4.1 of Jurecková (1977) which she proves under conditions different from ours. Her C_{ji} is our $x_{ij} n^{-\frac{1}{2}}$.

Lemma A.3: *For all* $\epsilon > 0$, *there exists* $K > 0$, $\eta > 0$, *and an integer* n_0 *such that if* $n > n_0$, *then*

$$P\{\inf_{||\underset{\sim}{\Delta}|| \ge K} ||M(\underset{\sim}{\Delta})|| < \eta\} < \epsilon .$$

Proof: This is shown in exactly the same manner as Lemma 5.2 of Jurecková (1977).

Proof of Theorem 2: In $M(\underset{\sim}{\Delta})$ replace $\underset{\sim}{\Delta}$ by $\sqrt{n}\ (\hat{\underset{\sim}{\beta}}(\theta) - \underset{\sim}{\beta}(\theta))$. By theorem 1

$$M(\sqrt{n}\ (\hat{\underset{\sim}{\beta}}(\theta) - \underset{\sim}{\beta}(\theta)) = n^{-\frac{1}{2}} \sum_{i=1}^{n} \underset{\sim i}{x}' \psi_\theta(y_i - \underset{\sim i}{x}\ \hat{\underset{\sim}{\beta}}(\theta)) = o_p(1). \qquad (A.4)$$

From Lemma A.3, $\sqrt{n}\ (\hat{\underset{\sim}{\beta}}(\theta) - \underset{\sim}{\beta}(\theta)) = 0_p(1)$. Theorem 2 follows from (A.3) and (A.4).

Proof of Corollary 1: Let $\underset{\sim}{\gamma}(\theta_i) = \hat{\underset{\sim}{\beta}}(\theta_i) - \underset{\sim}{\beta}(\theta_i)$ and $\gamma' = (\underset{\sim}{\gamma}(\theta_1)', \ldots, \underset{\sim}{\gamma}(\theta_m)')$. We need only show that if $\underset{\sim j}{c} \in R^p$ for $j=1,\ldots,m$ and $\underset{\sim}{c}' = (\underset{\sim 1}{c}', \ldots, \underset{\sim m}{c}')$ then

$$\sqrt{n}\ \underset{\sim}{c}'\ \underset{\sim}{\gamma} \xrightarrow{\mathcal{D}} N(0, \underset{\sim}{c}'(\Omega \otimes Q^{-1})\underset{\sim}{c}). \qquad (A.5)$$

Define

$$\eta_{ij} = f(\xi_{\theta_j})^{-1}\ \underset{\sim j}{c}'\ Q^{-1}\ \underset{\sim i}{x}\ \psi_{\theta_j}(e_i - \xi_{\theta_j})$$

and

$$\eta_{i\cdot} = \sum_{j=1}^{m} \eta_{ij}\ .$$

By Theorem 2,

$$\sqrt{n}\ \ \underset{\sim}{c}'\underset{\sim}{\gamma} = n^{-\frac{1}{2}} \sum_{i=1}^{n} \eta_i + o_p(1)\ .$$

Then (A.5) follows since routine calculations show that $\eta_{1\cdot}, \eta_{2\cdot}, \ldots$ satisfy the conditions of Lindeberg's Central Limit Theorem. (See for example, Loeve (1963, section 20.2, Theorem B).

For A any matrix, let $|A| = \max_{i,j} |A_{ij}|$.

Lemma A.4: *Let* D_{in} $(= D_i)$ *be a* r×c *matrix. Suppose*

$$\sup_n\ (n^{-1} \sum_{i=1}^{n} |D_i|^2) < \infty\ . \qquad (A.6)$$

Let I be an open interval containing ξ_1 and ξ_2 and let the function $g(x)$ be defined for all x and Lipschitz continuous on I. For $\underset{\sim}{\Delta}_1$, $\underset{\sim}{\Delta}_2$, and $\underset{\sim}{\Delta}_3$ in R^p and $\underset{\sim}{\Delta} = (\underset{\sim}{\Delta}_1, \underset{\sim}{\Delta}_2, \underset{\sim}{\Delta}_3)$ define

$$T^*(\underset{\sim}{\Delta}) = n^{-\frac{1}{2}} \sum_{i=1}^{n} D_i \, g(e_i + \underset{\sim}{\Delta}_3 \, \underset{\sim}{x}_i n^{-\frac{1}{2}}) \, I\{\xi_1 + \underset{\sim}{x}_i \, \underset{\sim}{\Delta}_1 n^{-\frac{1}{2}} < e_i < \xi_2 + \underset{\sim}{x}_i \underset{\sim}{\Delta}_2 \, n^{-\frac{1}{2}}\}$$

$$T(\underset{\sim}{\Delta}) = T^*(\underset{\sim}{\Delta}) - E\, T^*(\underset{\sim}{\Delta}),$$

$$S^*(\underset{\sim}{\Delta}) = T^*(\underset{\sim}{\Delta}) - T^*(0), \tag{A.7}$$

$$S(\underset{\sim}{\Delta}) = S^*(\underset{\sim}{\Delta}) - E\, S^*(\underset{\sim}{\Delta}) \, .$$

Then for all $M > 0$,

$$\sup_{0 \le |\underset{\sim}{\Delta}| \le M} |S(\underset{\sim}{\Delta})| = o_p(1).$$

Proof: Here we follow Bickel (1975, Lemma 4.1) closely. For convenience assume M=1. Define for $\ell = 1,2$

$$I_{i\ell}(\underset{\sim}{\Delta}) = I(e_i \le \xi_\ell + \underset{\sim}{x}_i \, \underset{\sim}{\Delta}_\ell \, n^{-\frac{1}{2}})$$

and let

$$b_i(\underset{\sim}{\Delta}) = g(e_i + \underset{\sim}{x}_i \, \Delta_3 \, n^{-\frac{1}{2}}) \, .$$

Then for all n large enough

$$\mathrm{Var}|S(\underset{\sim}{\Delta})| \le n^{-1} \sum_{i=1}^{n} |D_i|^2 \, \mathrm{Var}|b_i(\underset{\sim}{\Delta}) - b_i(0)| I\{e_i \in I \text{ and } e_i + \underset{\sim}{x}_i \Delta_3 \, n^{-\frac{1}{2}} \in I\}$$

$$+ n^{-1} \sum_{i=1}^{n} |D_i|^2 \, \mathrm{Var}\{b_i(0)(I_{i1}(\underset{\sim}{\Delta}) - I_{i1}(0)$$

$$+ I_{i2}(0) - I_{i2}(\underset{\sim}{\Delta}))\} = R_1 + R_2, \text{ say.}$$

Since g is Lipschitz on I

$$R_1 \le O(n^{-1} \sum_{i=1}^{n} |D_i|^2 \, |\underset{\sim}{x}_i| \, n^{-\frac{1}{2}}) = o(1)$$

by (A.6) and (2.1). Since F is continuously differentiable in neighborhoods of ξ_1 and ξ_2

$$R_2 \leq O(n^{-1} \sum_{i=1}^{n} |D_i|^2 \, |x_{\sim i}| \, n^{-\frac{1}{2}}) = o(1) \ .$$

Therefore for any fixed $\underset{\sim}{\Delta}$,

$$S(\underset{\sim}{\Delta}) \xrightarrow{P} 0 \ . \tag{A.8}$$

Choose $\delta > 0$. Now cover the p-dimensional cube, $[-1,1]^p$, with a union of cubes having vertices on the grid $J(\delta) = \{(j_1\delta, j_2\delta, \ldots, j_p\delta);\ j_i = 0, \pm 1, \pm 2, \ldots,$ or $\pm ([1/\delta] + 1)\}$. If $|\underset{\sim}{\Delta}| \leq 1$, then for $\ell = 1,2,3$ let $V_\ell(\underset{\sim}{\Delta})$ be the lowest vertex of the cube containing $\underset{\sim\ell}{\Delta}$ and let $V(\underset{\sim}{\Delta}) = (V_1(\underset{\sim}{\Delta}), V_2(\underset{\sim}{\Delta}), V_3(\underset{\sim}{\Delta}))$. Then straightforward calculations show that for some C

$$\begin{aligned} |S^*(\underset{\sim}{\Delta}) - S^*(V(\underset{\sim}{\Delta}))| &\leq n^{-\frac{1}{2}} \sum_{i=1}^{n} |D_i| \, |b_i(\underset{\sim}{\Delta}) - b_i(V(\underset{\sim}{\Delta}))| \, I(e_i \pm |x_{\sim i}| \, n^{-\frac{1}{2}} \in I) \\ &\quad + n^{-\frac{1}{2}} \sum_{i=1}^{n} |D_i| \, |b_i(V(\underset{\sim}{\Delta}))| \, [I\{-a_i \leq P_{1i} \leq a_i\} + I\{-a_i \leq P_{2i} \leq a_i\}] \\ &= W_1(\delta, V(\underset{\sim}{\Delta})) + W_2(\delta, V(\underset{\sim}{\Delta})), \text{ say, where} \end{aligned}$$

$$P_{\ell i} = e_i - \xi_\ell - x_{\sim i} \, V_\ell(\underset{\sim}{\Delta}) n^{-\frac{1}{2}}) \text{ and } a_i = |x_i| \, \delta n^{-\frac{1}{2}} \ .$$

Now since g is Lipschitz on I and F has a continuous derivative in neighborhoods of ξ_1 and ξ_2 there is a constant K_1 such that for m=1,2

$$E \, W_m(\delta, V(\underset{\sim}{\Delta})) \leq K_1 \, \delta \, n^{-1} \sum_{i=1}^{n} |D_i| \, |x_{\sim i}|$$

for all $\underset{\sim}{\Delta} \in (J(\delta))^3$. Then by (2.2), (A.6) and the Cauchy-Schwarz inequality for some K_2

$$\max_{\underset{\sim}{\Delta} \in (J(\delta))^3} E(W_m(\delta, \underset{\sim}{\Delta})) \leq K_2 \, \delta.$$

Exactly as in the argument leading to (A.8) we have (letting $W_m(\underset{\sim}{\Delta}) = W_m(\delta, \underset{\sim}{\Delta}_1)$

$$\max_{\underset{\sim}{\Delta} \in (J(\delta))^3} [W_m(\underset{\sim}{\Delta}) - E\, W_m(\underset{\sim}{\Delta})] = o_p(1) .$$

Thus, for all $\epsilon > 0$ and $\delta > 0$, there exists n_0 such that if $n \geq n_0$, then for m=1,2

$$P\{(\max_{\underset{\sim}{\Delta}_o \in J^3} W_m(\Delta_o)) > \epsilon + K_1\, \delta\} < \epsilon,$$

whence

$$P\{\sup_{\underset{\sim}{\Delta} \in [0,1]^{3p}} |S^*(\underset{\sim}{\Delta}) - S^*(V(\underset{\sim}{\Delta}))| > \epsilon + K_1\delta\} < \epsilon \quad .$$

Note that there exists a constant K_3 such that

$$\sup_{\underset{\sim}{\Delta} \in [0,1]^{3p}} |E(S^*(\underset{\sim}{\Delta}) - S^*(V(\underset{\sim}{\Delta})))| \leq K_3\delta .$$

Choosing $\delta = \epsilon/\max(K_1, K_3)$ we have that for all $\epsilon > 0$, there exists n_0 such that

$$P(\sup_{\underset{\sim}{\Delta} \in [0,1]^{2p}} |S(\underset{\sim}{\Delta}) - S(V(\underset{\sim}{\Delta})| > 3\epsilon) < \epsilon \quad . \tag{A.9}$$

By (A.8), for this δ and ϵ there exists n_1 such that if $n \geq n_1$

$$P(\sup_{\underset{\sim}{\Delta}_o \in J(\delta)^3} |S(\underset{\sim}{\Delta}_o)| > \epsilon) < \epsilon . \tag{A.10}$$

Inequalities (A.9) and (A.10) prove the lemma.

Proof of Theorem 3. By Theorem 2 we have for $\ell = 1,2$ that

$$\hat{\beta}_T(\alpha_\ell) = \beta + \xi_\ell + n^{-1/2} H(\alpha_\ell), \text{ where}$$

$$f(\xi_\ell)H(\alpha_\ell) = Q^{-1} n^{-1/2} \sum_{i=1}^{n} x_i' \psi_{\alpha_\ell}(e_i - \xi_\ell) + o_p(1) . \tag{A.11}$$

Therefore,

$$y_i - x_i \hat{\beta}(\alpha_\ell) = e_i - \xi_\ell - x_i H(\alpha_\ell) n^{-1/2} .$$

For $\Delta_1, \Delta_2 \in R^p$ and $\Delta = (\Delta_1, \Delta_2)$ we set

$$U^*(\Delta) = n^{-1} \sum_{i=1}^{n} x_i' x_i I\{\xi_1 + x_i \Delta_1 n^{-1/2} < e_i < \xi_2 + x_i \Delta_2 n^{-1/2}\}$$

$$T^*(\Delta) = n^{-1/2} \sum_{i=1}^{n} x_i' e_i I\{\xi_1 + x_i \Delta_1 n^{-1/2} < e_i < \xi_2 + x_i \Delta_2 n^{-1/2}\} .$$

Then

$$n^{-1}(X'AX) = U^*(H(\alpha_1), H(\alpha_2)) \text{ and } n^{-1/2} X'A(y - AX\beta) = T^*(H(\alpha_1), H(\alpha_2)) .$$

For $M > 0$, if we apply lemma A.4 with $D_i = x_i$ and $g(x) = x$ we obtain

$$\sup_{0 \le |\Delta| \le M} |T^*(\Delta) - T^*(0) - E(T^*(\Delta) - T^*(0))| = o_p(1) . \tag{A.12}$$

A Taylor expansion which uses the continuity of f about ξ_1 and ξ_2 shows that

$$E(T^*(\Delta) - T^*(0)) = Q[\Delta_2 \xi_2 f(\xi_2) - \Delta_1 \xi_1 f(\xi_1)] + o_p(1) . \tag{A.13}$$

Thus by (A.11), (A.12) and (A.13),

$$n^{-\frac{1}{2}} X'A(\underset{\sim}{y} - AX\ \underset{\sim}{\beta})$$

$$= n^{-\frac{1}{2}} \sum_{i=1}^{n} \underset{\sim}{x}_i' \, [e_i \, I(\xi_1 \leq e_i \leq \xi_2)$$

$$+ \xi_2(I(e_i > \xi_2) - \alpha_2) + \xi_1(I(e_i < \xi_1) - \alpha_1)] + o_p(1)$$

$$= n^{-\frac{1}{2}} (\alpha_2 - \alpha_1) \sum_{i=1}^{n} \underset{\sim}{x}_i' [\phi(e_i) - E\,\phi(e_1) + \delta(\underset{\sim}{\alpha})] + o_p(1) \ . \qquad (A.14)$$

In a similar manner, $n^{-1}(X'AX) = (\alpha_2 - \alpha_1)Q + o_p(1)$ which, together with (A.14) proves (3.4). Then (3.5) implies that $n^{-\frac{1}{2}} Q^{-1} \sum_{i=1}^{n} \underset{\sim}{x}_i' \, \delta(\underset{\sim}{\alpha}) = n^{\frac{1}{2}} \, \underset{\sim}{\delta}(\underset{\sim}{\alpha})$ and (3.6) follows by routine calculations.

Proof of Theorem 4: For Δ_1, Δ_2, Δ_3 in R^p define $\Delta = (\Delta_1, \Delta_2, \Delta_3)$ and

$$V(\Delta) = n^{-1} \sum_{i=1}^{n} (e_i - \underset{\sim}{x}_i \, \Delta_1 n^{-\frac{1}{2}} - \delta(\underset{\sim}{\alpha}))^2 \, I(\xi_1 + \underset{\sim}{x}_i' \, \Delta_2 n^{-\frac{1}{2}} \leq e_i \leq \xi_2 + \underset{\sim}{x}_i' \, \Delta_3 n^{-\frac{1}{2}}) \ .$$

We see that

$$S = nV(\sqrt{n} \ (\underset{\sim}{\hat{\beta}}_T(\underset{\sim}{\alpha}) - \underset{\sim}{\beta}_T(\underset{\sim}{\alpha})), \ \sqrt{n} \ (\underset{\sim}{\hat{\beta}}(\alpha_1) - \underset{\sim}{\beta}(\alpha_1)), \ \sqrt{n} \ (\underset{\sim}{\hat{\beta}}(\alpha_2) - \underset{\sim}{\beta}(\alpha_2)) \ .$$

Applying Lemma A.4 with $g(x) = x^2$ and $D_i = 1$ we have for $M > 0$,

$$\sup_{0 \leq |\Delta| \leq M} |V(\Delta) - V(0) - E(V(\Delta) - V(0))| = o_p(1) \ .$$

By a Taylor expansion of F and additional simple calculations

$$E(V(\Delta) - V(0)) \to 0, \text{ whence}$$

$$\sup_{0 \leq |\Delta| \leq M} |V(\Delta) - V(0)| = o_p(1) \ .$$

Therefore by (3.1) and (3.6) we have

$$S = V(0) + o_p(1) \ .$$

By Chebyshev's inequality $\mathrm{Var}(V(0)) \to 0$ so

$$S = EV(0) + o_p(1)$$

$$= E(e_i - \delta(\underset{\sim}{\alpha}))^2 \, I(\xi_1 \leq e_i \leq \xi_2) + o_p(1) \ .$$

Furthermore for $j = 1,2$

$$a_j = \xi_j - \delta(\underset{\sim}{\alpha}) + o_p(1)$$

by (3.1) and (3.6), and Theorem 4 follows.

Proofs of Theorems 5 and 6: These follow in a straightforward manner from (3.6), Theorem 4, and Theorem 4.4 of Billingsley (1968).

1. Regression Estimators

Name	Description
LS	Ordinary least squares.
Huber	M-estimator with $\psi(x) = \max(-1.5, \min(x, 1.5))$.
Hampel	M-estimate with ψ as in (4.1).
K15	Least squares with observations having positive residuals from the $\theta = .85$ quantile hyperplane or negative residuals from the $\theta = .15$ hyperplane removed.
K05	Least squares after deletion of observations having non-negative residuals from the $\theta = .95$ hyperplane or non-positive residuals from the $\theta = .05$ hyperplane.
Andrews	M-estimate using $\psi(x) = \sin(x/1.5)$ if $\|x\| \leq 1.5$ and $\psi(x) = 0$, otherwise and reported by Andrews (1974).
LAD	The $\theta = .50$ regression quantile.
K50	Least squares after removing those observations with the kth smallest or kth largest residual from the $\theta = .50$ hyperplane for k=1,...,[.05n] (n is the sample size).

2. MAD and Rank of MAD for Estimators in Table 1

Estimator	Data Set			
	Shrimp Harvest (n=32)		Stackloss (n=21)	
	MAD	Rank	MAD	Rank
LS	0.726	8	1.867	6
Huber	0.553	6	1.926	7
Hampel	0.551	5	1.928	8
Andrews	—	—	0.87	1
K15	0.699	7	1.407	4
K05	0.460	1.5	1.463	5
K50	0.460	1.5	1.204	3
LAD	0.508	3	1.182	2

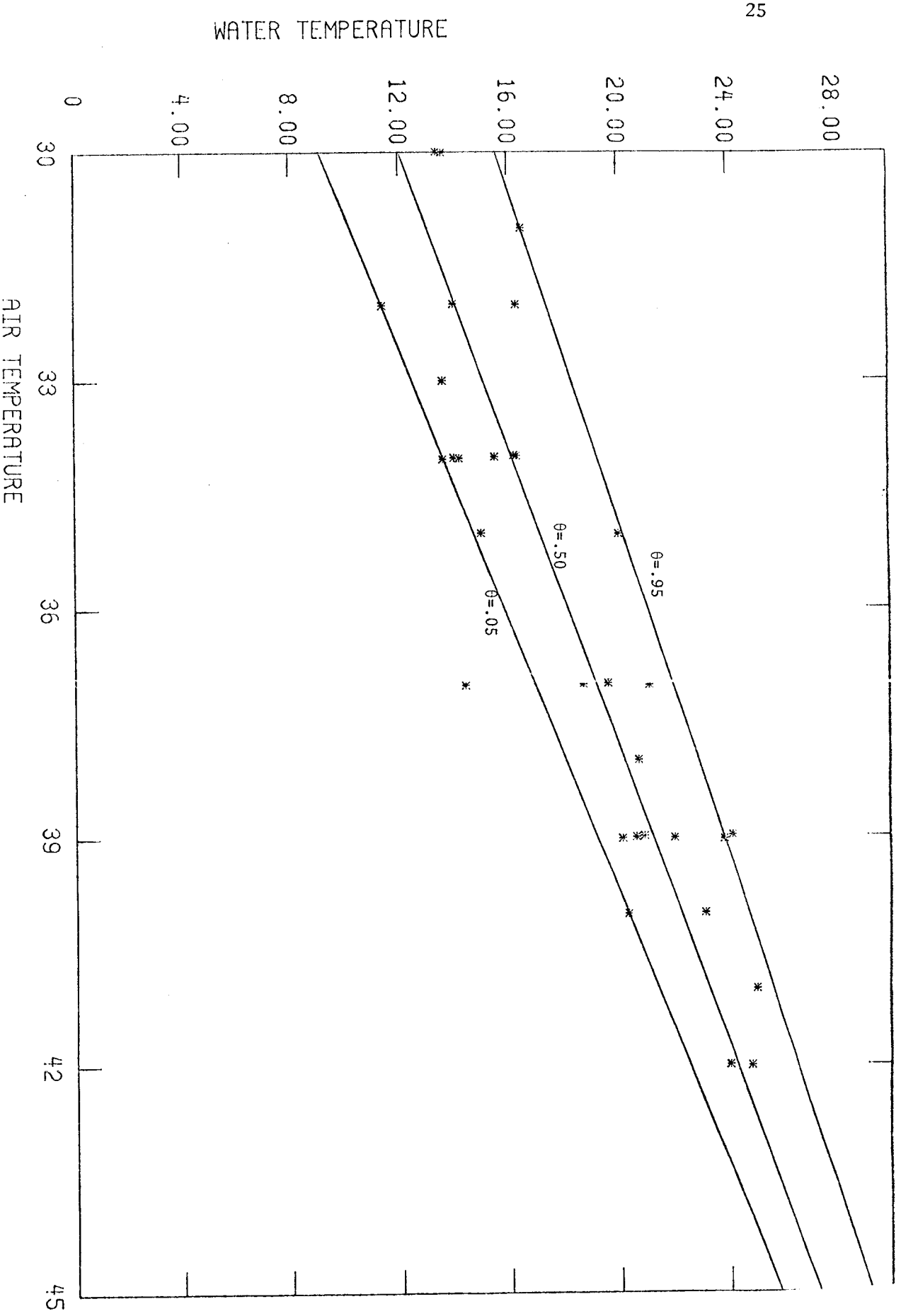

Figure A - Scatter Plot and Three Regression Quantile Lines for a Simple Linear Regression Example

REFERENCES

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

Bickel, Peter J. (1973), "On Some Analogues to Linear Combinations of Order Statistics in the Linear Model," *The Annals of Statistics*, 1, 597-616.

Bickel, Peter J. (1975), "One-Step Huber Estimates in the Linear Model," *Journal of the American Statistical Association*, 70, 428-433.

Billingsley, Patrick (1968), *Convergence of Probability Measures*, New York: John Wiley and Sons.

Brownlee, K. A. (1965), *Statistical Theory and Methodology in Science and Engineering, 2nd Edition*, New York: John Wiley.

deWet, T. and Venter, J.H. (1974), "*An Asymptotic Representation of Trimmed Means with Applications*," South African Statistics Journal, 8, 127-134.

Daniel, Cuthbert and Wood, Fred S. (1971). *Fitting Equations to Data*, New York: John Wiley.

Huber, Peter J. (1977), *Robust Statistical Procedures*, Philadelphia: SIAM.

Jaeckel, Louis A. (1971), "Robust Estimates of Location: Symmetry and Asymmetric Contamination," *The Annals of Mathematical Statistics*, 42, 1020-1034.

Jurecková, Jana (1977), "Asymptotic Relations of M-estimates and R-estimates in Linear Regression Model", *The Annals of Statistics*, 5, 464-472.

Koenker, Roger and Bassett, Gilbert, Jr. (1978), "Regression Quantiles," *Econometrica*, 46, 33-50.

Lenth, Russell V. (1976), "A Computational Procedure For Robust Multiple Regression," Tech. Report 53, University of Iowa.

Loève, Michel (1963), *Probability Theory*, New York: Van Nostrand.

McKeown, P.G. and Rubin, D.S. (1977), "A Student Oriented Preprocessor for MPS/360," *Computers and Operations Research*, 4, 227-229.

Searle, Shayle R. (1971), *Linear Models*, New York: John Wiley and Sons.

Scheffé, Henry (1959), *The Analysis of Variance*, New York: John Wiley and Sons.

Stigler, Stephen M. (1977), "Do Robust Estimators Work with Real Data?" *The Annals of Statistics*, 5, 1055-1098.