Open access • Journal Article • DOI:10.1080/01621459.1980.10477560

Trimmed Least Squares Estimation in the Linear Model — Source link

David Ruppert, Raymond J. Carroll

Institutions: University of North Carolina at Chapel Hill

Published on: 01 Dec 1980 - Journal of the American Statistical Association (Taylor & Francis Group)

Topics: Trimmed estimator, Least trimmed squares, Estimator, Truncated mean and Generalized least squares

TRIMMED LEAST SQUARES ESTIMATION IN THE LINEAR MODEL

by

David Ruppert and Raymond J. Carroll

Abstract

We consider two methods of defining a regression analogue to a trimmed mean. The first was suggested by Koenker and Bassett and uses their concept of regression quantiles. Its asymptotic behavior is completely analogous to that of a trimmed mean. The second method uses residuals from a preliminary estimator. Its asymptotic behavior depends heavily on the preliminary estimate; it behaves, in general, quite differently than the estimator proposed by Koenker and Bassett, and it can be rather inefficient at the normal model even if the percent trimming is small. However, if the preliminary estimator is the average of the two regression quantiles used with Koenker and Bassett's estimator, then the first and second methods are asymptotically equivalent for symmetric error distributions.

Key Words and Phrases: regression analogue, trimmed mean, regression quantile, preliminary estimator, linear model, trimmed least squares

---

David Ruppert is an Assistant Professor and Raymond J. Carroll an 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 MCS78-01240 and the Air Force Office of Scientific Research under Contracts AFOSR-75-2796 and AFOSR-80-0080.

1. Introduction.

This paper is concerned with the linear model

$$\underline{y} = X\underline{\beta} + \underline{Z} , \tag{1.1}$$

where $\underline{y}' = (y_1,\ldots,y_n)$, X is a n×p matrix of known constants whose i-th row is $\underline{x}_i$, $\underline{\beta}' = (\beta_1,\ldots,\beta_p)$ is a vector of unknown parameters, and $Z' = (Z_1,\ldots,Z_n)$ is a vector of independent identically distributed random variables with unknown distribution function F. Despite the advantages, including efficiency when F is normal, of the least squares estimator of $\underline{\beta}$, this estimator is inefficient when F has heavier tails than the Gaussian distribution and possesses high sensitivity to spurious observations. This inefficiency to heavy-tailed F has been amply demonstrated for the location submodel by a Monte-Carlo study (Andrews (1972)) and by asymptotics, e.g., Table 1 of this paper. The presence of spurious data can be modelled by letting F be a mixture of the distribution function of the "good" data, say standard normal, and that of the "bad" data, say normal with variance exceeding 1. Such an F will have heavier tails than a normal distribution, and inefficiency with heavy-tailed F appears to be closely related to sensitivity to outliers. Huber (1977, p. 3) states that "for most practical purposes, 'distributional robust' and 'outlier resistant' are interchangeable." For the location model, three classes of estimators have been proposed as alternatives to the sample mean, M, L, and R estimators; see Huber (1977) for an introduction. Among the L-estimates, the trimmed mean is particularly attractive because it is easy to compute and is rather efficient under a variety of circumstances.

As with M-estimates, trimmed means can be used to form confidence intervals. Let $TM(\alpha)$ be the α-trimmed mean, let $Y_{(i)}$ be the i-th order statistic, and with $k = [n\alpha]$, define

$$S^2(\alpha) = n\{k(Y_{(k+1)} - TM(\alpha))^2 + k(Y_{(n-k)} - TM(\alpha))^2 + \sum_{i=k+1}^{n-k} (Y_{(i)} - TM(\alpha))^2\}/\{(n-2k)(n-2k-1)\} .$$

If F is symmetric about μ, then $n^{\frac{1}{2}}(TM(\alpha)-\mu)/S(\alpha)$ is asymptotically N(0,1). (This can be easily seen from Theorems 1 and 2 of deWet and Venter (1974).) Therefore if we define z_γ to be the $(1-\gamma)$th quantile of the standard normal distribution,

$$TM(\alpha) \pm z_{\gamma/2}\, S(\alpha)$$

is a large sample confidence interval. Tukey and McLaughlin (1963) suggest replacing $z_{\gamma/2}$ by the $(1-\gamma/2)$th quantile of the t distribution with (n-2k-1) degrees of freedom. Huber (1970) uses a heuristic argument to justify a different choice of degrees of freedom, which is somewhat too complex to give here. Gross's (1976) Monte-Carlo study of the distribution of $TM(\alpha)/S(\alpha)$ indicates that the validity (agreement of nominal and actual significance level) of these confidence intervals will not be wholly satisfactory if n is small (he studies n = 10 and 20), but with F non-normal they appear to be as valid as the standard confidence interval based on the sample mean and standard deviation and the t distribution with n-1 degrees of freedom. Gross also suggests a more conservative interval procedure.

Hogg (1974) favors trimmed means for the above reasons, and because they can serve as a basis for adaptive estimators. Stigler (1977) applied robust estimators to data from 18th and 19th century experiments designed to measure basic physical constants. He concluded that "the 10% trimmed mean (the smallest nonzero trimming percentage included in the study) emerges as the recommended estimator."

One might argue, of course, that although L-estimates have desirable properties, they really offer no advantages over other estimators. After all, 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 an 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

$$\sum_{i=1}^{n} \rho((X_i-b)/s_n) = \min! \tag{1.2}$$

where

$$\begin{aligned} \rho(x) &= x^2/2 \quad \text{if } |x| \le k \\ &= k(|x| - k/2) \text{ if } |x| > k . \end{aligned} \tag{1.3}$$

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 knows only his data may find his intuition of more assistance when choosing α compared with k.

We do not believe that trimmed means are always preferable to M-estimates, but rather that they are worthwhile alternatives to M-estimates, particularly to Huber's M-estimate.

For the linear model, Bickel (1973) has proposed a class of one-step L-estimators depending on a preliminary estimate of $\hat{\underline{\beta}}$, but, while these have good asymptotic efficiencies, they are computationally complex and are apparently not invariant to reparameterization.

In this paper we consider two other methods of defining a regression analogue to the trimmed mean. In the location problem, both estimates reduce to the trimmed mean. The first, which we call $\hat{\underline{\beta}}_{PE}(\alpha)$ for $0 < \alpha < \frac{1}{2}$, requires a preliminary estimate, which is denoted by $\hat{\underline{\beta}}_0$. Suppose that the residuals from $\hat{\underline{\beta}}_0$ are calculated, and those observations corresponding to the $[n\alpha]$ smallest and $[n\alpha]$ largest observations are removed. Then $\hat{\underline{\beta}}_{PE} (= \hat{\underline{\beta}}_{PE}(\alpha))$ is defined to be the least squares estimator calculated from the remaining observations.

The definition of $\hat{\underline{\beta}}_{PE}$ was motivated by the applied statisticians' practice of examining the residuals from a least squares fit, removing the points with large (absolute) residuals, and recalculating the least squares solution with the remaining observations. Generally, there is no formal rule for deciding which points to remove, but $\hat{\underline{\beta}}_{PE}$ is at least similar to this practice.

The second method of defining an analogue to the trimmed mean was proposed by Koenker and Bassett (1978), who extend the concept of quantiles to the linear model. Let $0 < \theta < 1$. Define

$$\psi_\theta(x) = \theta - I(x<0) \tag{1.4}$$

and

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

Then they call $\hat{\underline{\beta}}(\theta)$, any value of $\underline{b}$ which solves

$$\sum_{i=1}^{n} \rho_\theta(y_i - \underline{x}_i \underline{b}) = \min! , \tag{1.5}$$

a θth regression quantile. (Recall that $\underline{x}_i$ is the i-th row of X.) Koenker and Bassett's Theorem 4.2 states that regression quantiles have asymptotic behavior similar to sample quantiles in the location problem. (Since $\hat{\underline{\beta}}(\theta)$ is an M-estimate its large sample behavior can also be deduced from standard M-estimate theory, as we show later.) 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. As they point out, regression quantiles can be computed by standard linear programming techniques. They also suggest the following trimmed least squares estimators, call it $\hat{\underline{\beta}}_{KB}$: remove from the sample any observations whose residual from $\hat{\underline{\beta}}(\alpha)$ is negative or whose residual from $\hat{\underline{\beta}}(1-\alpha)$ 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 covariance of $\hat{\underline{\beta}}_{KB}(\alpha)$ 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 $\hat{\underline{\beta}}(\theta)$ and $\hat{\underline{\beta}}_{KB}(\alpha)$ which provide simple proofs of Koenker and Bassett's Theorem 4.2 and their conjecture about the asymptotic covariance of $\hat{\underline{\beta}}_{KB}(\alpha)$.

The close analogy between the asymptotic distributions of trimmed means and the trimmed least squares estimator $\hat{\underline{\beta}}_{KB}(\alpha)$ is remarkable. A result that is perhaps even more surprising is that the distribution of the estimator $\hat{\underline{\beta}}_{PE}$ depends heavily on that of the preliminary estimator $\hat{\underline{\beta}}_0$. In particular, using least squares or least absolute deviations as the preliminary estimator results in versions of $\hat{\underline{\beta}}_{PE}$ which are inefficient at the normal model and which are not regression analogues to the trimmed mean (as is $\hat{\underline{\beta}}_{KB}$).

(By a version of $\hat{\underline{\beta}}_{PE}$ we mean $\hat{\underline{\beta}}_{PE}$ for a particular $\hat{\underline{\beta}}_0$.)

Our results are such that we are able to find a version of $\hat{\underline{\beta}}_{PE}$ which corresponds to a trimmed mean when the error distribution F is symmetric. The "right choice" for $\hat{\underline{\beta}}_0$ is the average of the αth and $(1-\alpha)$th regression quantiles, i.e., $\hat{\underline{\beta}}_0 = \frac{1}{2}(\hat{\underline{\beta}}(\alpha)+\hat{\underline{\beta}}(1-\alpha))$.

Hogg (1974, p. 917) mentions that adaptive estimators can be constructed from estimators similar or identical to $\hat{\underline{\beta}}_{PE}(\alpha)$ with α a function of the residuals from $\hat{\underline{\beta}}_0$. The advantage of this class of adaptive estimators, he feels, is that they "would correspond more to the trimmed means for which we can find an error structure." However, from the above results, we can conclude that even if α is non-stochastic, estimators of the type suggested by Hogg will not necessarily have error structures which correspond to the trimmed mean.

Besides its nice asymptotic covariance, $\hat{\underline{\beta}}_{KB}$ has another desirable property. 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 centered 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 $0 < \alpha < \frac{1}{2}$

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

where $\underline{\delta}(\alpha)$ 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. However, this is also true for M-estimates (see, e.g., Carroll and Ruppert (1979) or Carroll (1979)).

We will present a large sample theory of confidence ellipsoids and general linear hypothesis testing, which is quite similar to that of least squares estimation. The same theory holds for $\underline{\hat{\beta}}_{PE}$ when $\underline{\hat{\beta}}_0 = (\underline{\hat{\beta}}(\alpha)+\underline{\hat{\beta}}(1-\alpha))/2$, but only if F is symmetric.

The methods of this paper can be applied to other estimators. For example, let $\underline{\hat{\beta}}_A(\alpha)(=\underline{\hat{\beta}}_A)$ be the least squares estimate after the points with the $[2\alpha N]$ largest absolute residuals from $\underline{\hat{\beta}}_0$ are removed. In section 6 we state results for $\underline{\hat{\beta}}_A$. Their proofs are omitted, but are similar to the proofs of analogous results for $\underline{\hat{\beta}}_{PE}$.

In section 2 we give notation and assumptions. In section 3, asymptotic representations of $\underline{\hat{\beta}}_{PE}$ are developed, and their significance is discussed in section 4. Section 5 contains asymptotic results for $\underline{\hat{\beta}}_{KB}$, and section 6 discusses conditions under which $\underline{\hat{\beta}}_{KB}$- $\underline{\hat{\beta}}_{PE}$ are asymptotically equivalent. In section 7, we compare the asymptotic behavior of $\underline{\hat{\beta}}_{PE}$ for several choices of $\underline{\hat{\beta}}_0$. Large sample inference is the subject of section 8. Several examples using real data are considered in section 9. All proofs are found in the appendix.

2. Notation and Assumptions.

Although $\underline{y}$, X and $\underline{Z}$ in (1.1) depend upon n, this will not be made explicit in the notation. Let $\underline{e}' = (1,0,\ldots,0)(1\times p)$ and let I_p be the p×p identity matrix. Whenever r is a scalar, $\underline{r} = r\underline{e}$. For $0 < p < 1$, define $\xi_p = F^{-1}(p)$. Also, suppose $0 < \alpha_1 < \frac{1}{2} < \alpha_2 < 1$, and define $\xi_1 = \xi_{\alpha_1}$ and $\xi_2 = \xi_{\alpha_2}$. Let $N_p(\underline{\mu}, \Sigma)$ denote the p-variate Gaussian distribution with mean $\underline{\mu}$ and covariance Σ. We will make the following assumptions throughout.

C1. F has a continuous density f which is positive on the support of F.

C2. Letting $(x_{i1},\ldots,x_{ip})' = \underline{x}_i$ be the i-th row of X, $x_{i1} = 1$ for $i = 1,\ldots,n$ and

$$\sum_{i=1}^{n} x_{ij} = 0 \text{ for } j = 2,\ldots,p \ .$$

C3. $\lim_{n\to\infty}\left(\max_{j\le p \text{ and } i\le n} (n^{-\frac{1}{2}}|x_{ij}|)\right) = 0$.

C4. There exists positive definite Q such that

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

C5. $(\hat{\underline{\beta}}_0 - \underline{\beta} - \theta\underline{e}) = O_p(n^{-\frac{1}{2}})$ for some constant θ.

We will assume that $\xi_{\frac{1}{2}} = 0$. By C2, this involves no loss in generality. Assumption C5 is satisfied by many estimators, including the LAD (least absolute deviation or median regression) (see Corollary 5.1) and, if $Ee_1^2 < \infty$, the LS (least squares) estimators.

The residuals from the preliminary estimate $\hat{\underline{\beta}}_0$ are

$$r_i = y_i - \underline{x}_i'\hat{\underline{\beta}}_0 = Z_i - \underline{x}_i'(\hat{\underline{\beta}}_0 - \underline{\beta}) \ .$$

Let r_{1n} and r_{2n} be the $[n\alpha]$th and $[n(1-\alpha)]$th ordered residuals, respectively.

Then the estimate $\hat{\underline{\beta}}_{PE}$ is a least squares (LS) estimate calculated after removing all observations satisfying

$$r_i \leq r_{1n} \text{ or } r_i \geq r_{2n} \ . \tag{2.1}$$

Because of C1, asymptotic results are unaffected by requiring strict inequalities in (2.1). Let $a_i = 0$ or 1 according as i satisfies (2.1) or not, and let A be the n×n diagonal matrix with $A_{ii} = a_i$. The matrix A indicates which observations are not trimmed. Thus

$$\hat{\underline{\beta}}_{PE}(\alpha) = (X'AX)^{-} X'A\underline{y} \ ,$$

where $(X'AX)^{-}$ is a generalized inverse for X'AX. (Later we show that $n^{-1}(X'AX) \overset{P}{\to} (1-2\alpha)Q$, whence P (X'AX is invertible) $\to$ 1.)

Since $\hat{\underline{\beta}}_{KB}$ behaves similarly to a trimmed mean, even for asymmetric F and for asymmetric trimming, we will not restrict ourselves to symmetric trimming when defining $\hat{\underline{\beta}}_{KB}$.

Let $\underline{\alpha} = (\alpha_1, \alpha_2)$ and define $\hat{\underline{\beta}}_{KB}(\underline{\alpha})$ to be a least squares estimator calculated after removing all observations satisfying

$$y_i - \underline{x}_i \hat{\underline{\beta}}(\alpha_2) \geq 0 \text{ or } y_i - \underline{x}_i \hat{\underline{\beta}}(\alpha_1) \leq 0 \ . \tag{2.2}$$

(Again asymptotic results are unaffected by requiring strict inequalities in (2.2), which is Koenker and Bassett's suggestion.) Let $b_i = 0$ or 1 according as i satisfies (2.2) or not, and let B be the n×n diagonal matrix with $B_{ii} = b_i$. Then

$$\hat{\underline{\beta}}_{KB}(\underline{\alpha}) = (X'BX)^{-} (X'B\underline{y}) \ ,$$

where $(X'BX)^-$ is a generalized inverse of $(X'BX)$. (Again, for n sufficiently large $X'BX$ will be invertible.) Let

$$\begin{aligned}\phi(x) &= \xi_1/(\alpha_2-\alpha_1) \text{ if } x < \xi_1 \\ &= x/(\alpha_2-\alpha_1) \text{ if } \xi_1 \le x \le \xi_2 \\ &= \xi_2/(\alpha_2-\alpha_1) \text{ if } \xi_2 < x .\end{aligned} \tag{2.3}$$

Define

$$\delta(\underline{\alpha}) = (\alpha_2-\alpha_1)^{-1} \int_{\xi_1}^{\xi_2} x dF(x) ,$$

and letting $\eta_j = (\xi_j - \delta(\underline{\alpha}))$ define

$$\sigma^2(\underline{\alpha},F) = (\alpha_2-\alpha_1)^{-2} \left(\int_{\xi_1}^{\xi_2} (x-\delta(\underline{\alpha}))^2 \, dF(x) + \alpha_1\eta_1^2 + (1-\alpha_2)\eta_2^2 - ((1-\alpha_2)\eta_2 + \alpha_1\eta_1)^2\right) .$$

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

3. Main Results for $\hat{\underline{\beta}}_{PE}$.

First we will find relationships of the form

$$n^{\frac{1}{2}}(\hat{\underline{\beta}}_{PE} - \underline{\beta}) \approx n^{-\frac{1}{2}} \sum_{i=1}^{n} G(\underline{x}_i, Z_i) + n^{\frac{1}{2}} H(\hat{\underline{\beta}}_0 - \underline{\beta}) , \tag{3.1}$$

where G and H are given functions. We then show that in many special cases (including LS and LAD) the latter term in (3.1) can be further expanded so that

$$n^{\frac{1}{2}}(\underline{\hat{\beta}}_{PE} - \underline{\beta}) \approx n^{-\frac{1}{2}} \sum_{i=1}^{n} G(\underline{x}_i, Z_i) + n^{-\frac{1}{2}} \sum_{i=1}^{n} H^*(\underline{x}_i, Z_i) \tag{3.2}$$

for some function H*. It is then a simple matter to obtain the limit distribution of $n^{\frac{1}{2}}(\underline{\hat{\beta}}_{PE} - \underline{\beta})$ from (3.2).

In this section we only consider symmetric trimming, so we assume $\alpha_1 = 1 - \alpha_2 = \alpha$. Now define $a = \xi_2 f(\xi_2) - \xi_1 f(\xi_1)$ and $\underline{c}_i = (I-ee')\underline{x}_i = (0, x_{i2}, \ldots, x_{ip})'$. The specific relationship of form (3.1) is:

<u>Theorem 3.1</u>. *As* $n \to \infty$,

$$n^{\frac{1}{2}}(\underline{\hat{\beta}}_{PE} - \underline{\beta}) = (1-2\alpha)^{-1} n^{-\frac{1}{2}} \sum_{i=1}^{n} Q^{-1} \underline{c}_i Z_i I(\xi_1 \le Z_i \le \xi_2)$$

$$+ (1-2\alpha)^{-1} a n^{\frac{1}{2}} (I - \underline{e}\,\underline{e}')(\underline{\hat{\beta}}_0 - \underline{\beta}) + n^{-\frac{1}{2}} \sum_{i=1}^{n} \underline{e}\phi(Z_i) + o_p(1) . \tag{3.3}$$

We will call the first entry of $\underline{\beta}$ the intercept and the remaining entries will be called the slopes. Since premultiplication of a vector by $(I-\underline{e}\,\underline{e}')$ simply replaces the first coordinate by 0, the first two terms on the RHS of (3.3) represent the slope estimates. Note the similarity (and the difference!) between the first term and a representation of the (untrimmed) LS estimate, $\underline{\hat{\beta}}$; since

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

and

$$(n^{-1} X' X) \to Q$$

it follows that

$$n^{\frac{1}{2}}(\underline{\hat{\beta}} - \underline{\beta}) = n^{-\frac{1}{2}} \sum_{i=1}^{n} Q^{-1} \underline{x}_i Z_i + o_p(1) .$$

The second term indicates the contribution of the preliminary estimate to the trimmed LS estimate; this contribution is only to the slope estimates. Since only the first coordinate of $\underline{e}$ is non-zero, the third term on the RHS of (3.4) is a representation of the intercept estimate and is identical to a representation of the trimmed mean in the location model (cf. Corollary 3.1).

To specify the relationship of form (3.2) we make the assumption:

C6. For some function g,

$$n^{\frac{1}{2}}(\underline{\hat{\beta}}_0 - \underline{\beta}) = n^{-\frac{1}{2}} \sum_{i=1}^{n} Q^{-1} \underline{x}_i g(Z_i) + o_p(1) .$$

As indicated above, C6 holds with $g(x) = x$ if $\underline{\hat{\beta}}_0$ is the LS estimate. By Theorem 5.3, C6 holds with $g(x) = (f(0))^{-1} (\frac{1}{2} - I(x<0))$ if $\underline{\hat{\beta}}_0$ is the LAD estimate. As an immediate consequence of Theorem 3.1, we have our main result.

<u>Theorem 3.2</u>.

$$n^{\frac{1}{2}}(\underline{\hat{\beta}}_{PE} - \underline{\beta}) = (1-2\alpha)^{-1} n^{-\frac{1}{2}} \sum_{i=1}^{n} Q^{-1} c_i\{Z_i I(\xi_1 \le Z_i \le \xi_2) + a g(Z_i)\}$$
$$+ n^{-\frac{1}{2}} \sum_{i=1}^{n} \underline{e}\phi(Z_i) + o_p(1) . \tag{3.4}$$

In the next section, limit distributions are obtained from (3.4) for various special cases. Both (3.3) and (3.4) show how the preliminary estimate influences the asymptotic behavior of $\hat{\underline{\beta}}_{PE}$.

As a special case of Theorem 3.2 we obtain:

Corollary 3.1. *In the location model* ($p = 1$ *and* $x_i = 1$ *for all* i)

$$n^{\frac{1}{2}}(\hat{\underline{\beta}}_{PE} - \underline{\beta}) = n^{-\frac{1}{2}} \sum_{i=1}^{n} \phi(Z_i) + o_p(1) .$$

The key technical step in the proofs is an "asymptotic linearity" result for ordered residuals, which generalizes work of Bahadur (1966) and Ghosh (1971) for the location model.

Lemma 3.1. *For* $0 < \theta < 1$, *let* $r_{\theta n}$ *be the* $[n\theta]$*th ordered residual from* $\hat{\underline{\beta}}_0$. *Then*

$$n^{\frac{1}{2}}(r_{\theta n} - \xi_\theta) = f(\xi_\theta)^{-1}\, n^{-\frac{1}{2}} \sum_{i=1}^{n} \psi_\theta(Z_i - \xi_\theta) - \underline{e}'\, n^{\frac{1}{2}}(\hat{\underline{\beta}}_0 - \underline{\beta}) + o_p(1) .$$

(Recall that $\psi_\theta(x) = \theta - I(x<0)$.)

4. Asymptotic Behavior of $\hat{\underline{\beta}}_{PE}$.

In this section we show that Theorem 3.2 leads to these basic conclusions about $\hat{\underline{\beta}}_{PE}$:

1) The intercept estimate is asymptotically unbiased if F is symmetric.
2) The slope estimates are asymptotically unbiased even if F is asymmetric.

3) The asymptotic variance of the intercept, which does not depend upon the choice of $\hat{\underline{\beta}}_0$, is that of the trimmed mean in the location model.

4) The asymptotic covariance matrix of the slopes depends upon $\hat{\underline{\beta}}_0$ and, in general, will be difficult to estimate.

Let $\underline{0}$ be a (p-1)×1 vector of zeroes. By C2, there is a $\tilde{Q}$ such that

$$Q = \begin{bmatrix} 1 & \underline{0}' \\ \underline{0} & \tilde{Q} \end{bmatrix} \text{ and } Q^{-1} = \begin{bmatrix} 1 & \underline{0}' \\ \underline{0} & \tilde{Q}^{-1} \end{bmatrix} .$$

Moreover,

$$\lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \underline{c}_i \underline{c}_i' = \begin{bmatrix} 0 & \underline{0}' \\ \underline{0} & \tilde{Q} \end{bmatrix}$$

and

$$Q \sum_{i=1}^{n} \underline{c}_i = \begin{bmatrix} 0 \\ \underline{0} \end{bmatrix} . \qquad (4.1)$$

If we estimate $\underline{\beta}$ with $\hat{\underline{\beta}}_{PE}$ then the asymptotic bias of the intercept is

$$E\phi(Z_1) = (1-2\alpha)^{-1} \int_{\xi_1}^{\xi_2} x dF(x) ,$$

which is zero if F is symmetric about zero. By (3.4) and (4.1) the slope estimates are asymptotically unbiased, even if F is asymmetric. The asymptotic variance of the normalized intercept is

$$\sigma^2(\alpha,F) = \text{Var } \phi(Z_1) ,$$

the asymptotic variance of the normalized α-trimmed mean in the location model. The intercept is asymptotically uncorrelated with the slopes, and the asymptotic covariance matrix of the normalized slopes is $\tilde{Q}^{-1}\sigma^2(\alpha,g,F)$ where

$$\sigma^2(\alpha,g,F) = (1-2\alpha)^{-2} \operatorname{Var}(Z_1 I(\xi_1 \le Z_1 \le \xi_2) + ag(Z_1)) .$$

We see that the asymptotic distribution of the intercept estimate does not depend upon the choice of $\hat{\underline{\beta}}_0$ provided $(\hat{\underline{\beta}}_0 - \underline{\beta}) = O_p(n^{-\frac{1}{2}})$. On the other hand, we see from (3.4) that the slope estimates depend upon $\hat{\underline{\beta}}_0$, since the unusual situation where $a = 0$ is ruled out by assumption C1. Using the Lindeberg central limit theorem and Theorem 3.2, it is easy to show that under C3 and C4, $n^{\frac{1}{2}}(\hat{\underline{\beta}}_{PE} - \underline{\beta} - \underline{e}E\phi(Z_1))$ converges in distribution to a normal law.

In general, large sample statistical inference based on $\hat{\underline{\beta}}_{PE}$ will be a challenging problem because of the difficulties of estimating $a = (\xi_2 f(\xi_2) - \xi_1 f(\xi_1))$. Obtaining reasonably good estimates of the density f might take very large sample sizes.

5. Main Results for $\hat{\underline{\beta}}_{KB}$.

In section 1, we defined a θth regression quantile to be any value of $\underline{b}$ which solves (1.5). There may be multiple solutions, though in our few examples we found that the solution was always unique. However, the asymptotic results we present do not depend upon the rule used to select one. We suppose, then, that a definite rule has been used, and we denote this solution by $\hat{\underline{\beta}}(\theta)$.

For $\hat{\underline{\beta}}_{KB}$ we obtain an asymptotic representation which is similar to those for $\hat{\underline{\beta}}_{PE}$ but perhaps simpler.

<u>Theorem 5.1</u>. *The estimator* $\hat{\underline{\beta}}_{KB}$ *satisfies*

$$n^{\frac{1}{2}}(\hat{\underline{\beta}}_{KB}(\underline{\alpha}) - \underline{\beta}) = Q^{-1}n^{-\frac{1}{2}}\{\sum_{i=1}^{N} \underline{x}_i'(\phi(Z_i) - E\phi(Z_i)) + \delta(\underline{\alpha})\} + o_p(1) , \tag{5.1}$$

and therefore

$$n^{\frac{1}{2}}(\hat{\underline{\beta}}_{KB}(\alpha) - \underline{\beta} - \underline{\delta}(\alpha)) \xrightarrow{L} N_p(\underline{0}, \sigma^2(\alpha, F)Q^{-1}) . \tag{5.2}$$

Expression (5.1) is similar to a result of deWet and Venter (1974, equation (5)). Note that (5.2) verifies Koenker and Bassett's hypothesis on the covariance of $\hat{\underline{\beta}}_{KB}$. Moreover, the bias of $\hat{\underline{\beta}}_{KB}$ for $\underline{\beta}$ involves only the intercept, $\underline{\beta}_1$, and not the slopes. Also, $\hat{\underline{\beta}}_{KB}$ is asymptotically unbiased if F is symmetric.

Theorem 5.1 requires the next result on regression quantiles. Define $\underline{\beta}(\theta) = \underline{\beta} + \underline{\xi}_\theta$. The next theorem, which is a special case of a general result for M-estimators, shows that $(\hat{\underline{\beta}}(\theta) - \underline{\beta}(\theta))$ is essentially a sum of independent random variables.

<u>Theorem 5.2</u>. *The following representation holds:*

$$n^{\frac{1}{2}}(\hat{\underline{\beta}}(\theta) - \underline{\beta}(\theta)) = n^{-\frac{1}{2}}(f(\xi_\theta))^{-1} Q^{-1} \sum_{i=1}^{n} \underline{x}_i'\psi_\theta(Z_i - \xi_\theta) + o_p(1) .$$

Theorem 5.2 and the Lindeberg central limit theorem provide an easy proof of Theorem 4.2 of Koenker and Bassett (1978), which we state as a corollary.

Corollary 5.1. *Let* $\Omega = \Omega(\theta_1,\ldots,\theta_m;F)$ *be the symmetric* m×m *matrix defined by*

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

Then

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

6. A Choice of $\hat{\underline{\beta}}_0$ For Which $\hat{\underline{\beta}}_{KB}$ and $\hat{\underline{\beta}}_{PE}$ Are Asymptotically Equivalent.

We have seen that $\hat{\underline{\beta}}_{KB}$ is a close analogue to the trimmed mean, but the behavior of $\hat{\underline{\beta}}_{PE}$ depends upon $\hat{\underline{\beta}}_0$ and is not similar, in general, to that of a trimmed mean. One might ask whether $\hat{\underline{\beta}}_0$ can be chosen so that $\hat{\underline{\beta}}_{PE}$ has the same asymptotic distribution as $\hat{\underline{\beta}}_{KB}$. The answer is yes, provided F is symmetric and we allow only symmetric trimming.

Let $\hat{\underline{\beta}}_{PE}(RQ,\alpha)$ $(=\hat{\underline{\beta}}_{PE}(RQ))$ be $\hat{\underline{\beta}}_{PE}$ when $\hat{\underline{\beta}}_0$ is the average of the αth and $(1-\alpha)$th regression quantiles, i.e., $\hat{\underline{\beta}}_0 = ((\hat{\underline{\beta}}(\alpha)+\hat{\underline{\beta}}(1-\alpha))/2$. Then, by Theorem 5.2, this $\hat{\underline{\beta}}_0$ satisfies C6 with

$$g(x) = (2f(\xi_1))^{-1}\,\psi_\alpha(x-\xi_1) + (2f(\xi_2))^{-1}\,\psi_{1-\alpha}(x-\xi_2) .$$

If F is symmetric, then $\xi_1 = -\xi_2$, $f(\xi_1) = f(\xi_2)$, and therefore

$$ag(x) = \xi_1 I(x\le\xi_1) + \xi_2 I(x\ge\xi) . \tag{6.1}$$

By (3.4) and (6.1),

$$n^{\frac{1}{2}}(\hat{\underline{\beta}}_{PE}(RQ) - \underline{\beta}) = n^{-\frac{1}{2}} \sum_{i=1}^{n} Q^{-1} \underline{x}_i \phi(Z_i) + o_p(1) ,$$

and therefore, since $\underline{\delta}(\underline{\alpha}) = 0$, (5.1) implies

$$n^{\frac{1}{2}}(\hat{\underline{\beta}}_{KB} - \hat{\underline{\beta}}_{PE}(RQ)) \xrightarrow{P} \underline{0} , \tag{6.2}$$

so that asymptotically there is no difference between trimming with this preliminary estimate and using Koenker and Bassett's (1978) proposal. (However, (6.2) does not necessarily hold if F is asymmetric.)

Notice that (5.1) and (6.2) imply that

$$n^{\frac{1}{2}}(\hat{\underline{\beta}}_{PE}(RQ) - \underline{\beta}) \xrightarrow{L} N(0, Q^{-1}\sigma^2(\alpha, F)) .$$

7. Comparisons of Several Choices of $\hat{\underline{\beta}}_0$.

The choice of $\hat{\underline{\beta}}_0$ should be based on the efficiency of the resulting $\hat{\underline{\beta}}_{PE}$, not its similarity to $\hat{\underline{\beta}}_{KB}$. In this section we find further support for using $\hat{\underline{\beta}}_{PE}(RQ)$ by comparing $\hat{\underline{\beta}}_{PE}(RQ)$ with two other estimators, $\hat{\underline{\beta}}_{PE}(LS)$ and $\hat{\underline{\beta}}_{PE}(LAD)$, which are $\hat{\underline{\beta}}_{PE}$ with $\hat{\underline{\beta}}_0$ equal to the least squares estimator and $\hat{\underline{\beta}}(.5)$, respectively. Comparisons are made within the family of contaminated normal distributions, which was introduced by Tukey (1960) to study the behavior of statistical procedures under heavy-tailed distributions. These distributions have the form

$$F(x) = (1-\varepsilon)\Phi(x) + \varepsilon\Phi(x/b) ,$$

where $0 < \varepsilon < 1$ and Φ is standard normal distribution. Typically, $b > 1$ and $\Phi(x/b)$ is the distribution of the "bad" data, while ε is the proportion of "bad" observations. Recall that the asymptotic variance of the intercept does not depend upon $\hat{\underline{\beta}}_0$, and that the asymptotic covariance matrix of the slopes is $\tilde{Q}^{-1}\sigma^2(\alpha,g,F)$, where $\tilde{Q}^{-1}$ depends only on the sequence of design matrices. Therefore, we can compare the estimators by using only $\sigma^2(\alpha,g,F)$. Table 1 displays $\sigma^2(\alpha,g,F)$ for several choices of α, ε, and b, and for g corresponding to $\hat{\underline{\beta}}_{PE}(LS)$, $\hat{\underline{\beta}}_{PE}(LAD)$, and $\hat{\underline{\beta}}_{PE}(RQ)$. For comparison, we include the standardized asymptotic variance (that is σ^2 where $\sigma^2 Q^{-1}$ is the asymptotic covariance matrix) for the LS estimate and two M-estimates, a Huber and a Hampel. Both of the M-estimates use Huber's Proposal 2 to obtain scale equivariance. The Huber uses

$$\psi(x) = \min(2,|x|)\,\mathrm{sign}(x) \;,$$

and the Hampel uses

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

and $\psi(-x) = -\psi(x)$. For discussion of Huber's Proposal 2 see Carroll and Ruppert (1979). Several conclusions emerge from Table 1.

1) $\hat{\underline{\beta}}_{PE}(LS)$ and $\hat{\underline{\beta}}_{PE}(LAD)$ are rather inefficient at the normal distribution.
2) $\hat{\underline{\beta}}_{PE}(RQ)$ is quite efficient at the normal model.
3) Under heavy contamination (b large or ε large) $\hat{\underline{\beta}}_{PE}(LS)$, $\hat{\underline{\beta}}_{PE}(LAD)$, and $\hat{\underline{\beta}}_{PE}(RQ)$ are relatively efficient compared with LS. Also $\hat{\underline{\beta}}_{PE}(RQ)$ and $\hat{\underline{\beta}}_{PE}(LAD)$ compare well against the M-estimates, but $\hat{\underline{\beta}}_{PE}(LS)$ does poorly compared to the M-estimates if $\varepsilon = .25$, $b = 10$, and $\alpha = .25$.

(Intuitively, one can expect that when $\alpha = .25$, $\hat{\underline{\beta}}_{PE}(LS)$ will be heavily influenced by its preliminary estimate, which estimates $\underline{\beta}$ poorly for these b and ε.)

Because of 1) and 3), the practice of fitting by least squares or LAD, removing points corresponding to extreme residuals, and computing the least squares estimate from the trimmed sample is not an adequate substitute for robust methods of estimation.

If, instead of removing those observations with the $[n\alpha]$ smallest and $[n\alpha]$ largest residuals from $\hat{\underline{\beta}}_0$, we remove those observations with the $[2n\alpha]$ largest absolute residuals, then the asymptotic variance of the intercept is the same as that of the slopes. Specially, let $\hat{\underline{\beta}}_A(\alpha)(=\hat{\underline{\beta}}_A)$ be the estimate formed in this manner. Then, if F is symmetric,

$$(1-2\alpha)n^{\frac{1}{2}}(\hat{\underline{\beta}}_A - \underline{\beta}) = n^{-\frac{1}{2}} \sum_{i=1}^{n} Q^{-1} \underline{x}_i\{Z_i I(\xi_1 \leq Z_i \leq \xi_2) + a(\hat{\underline{\beta}}_0 - \underline{\beta})\} + o_p(1) ,$$

and if C6 holds, then

$$(1-2\alpha)n^{\frac{1}{2}}(\hat{\underline{\beta}}_A - \underline{\beta}) = n^{-\frac{1}{2}} \sum_{i=1}^{n} Q^{-1} \underline{x}_i\{Z_i I(\xi_1 \leq Z_i \leq \xi_2) + ag(Z_i)\} + o_p(1) ,$$

which in the location case reduces to

$$(1-2\alpha)n^{\frac{1}{2}}(\hat{\underline{\beta}}_A - \underline{\beta}) = n^{-\frac{1}{2}} \sum_{i=1}^{n} \{Z_i I(\xi_1 \leq Z_i \leq \xi_2) + ag(Z_i)\} + o_p(1) .$$

The proofs are similar to those of Theorems 3.1 and 3.2 and are omitted.

Since $\hat{\underline{\beta}}_A$ is particularly easy to compute in the location model, it is very suitable for Monte-Carlo studies. It is hoped that such studies will indicate the degree of agreement between the asymptotic and finite sample variances of $\hat{\underline{\beta}}_{PE}$ as well as $\hat{\underline{\beta}}_A$. Table 1 displays the variance of $\hat{\underline{\beta}}_A(LS)$, i.e., $\hat{\underline{\beta}}_A$ with $\hat{\underline{\beta}}_0$ the LS estimate, for sample sizes of n = 50, 100, 200, 300, and 400. The Monte-Carlo swindle (Gross (1973)) was employed as a variance reduction technique. One sees from this table that convergence of the variance to its asymptotic value can be extremely slow for some distributions, e.g., b = 10 and ε = .10 or .25.

8. Large Sample Inference.

Here we sketch a large sample methodology of confidence ellipsoids and hypothesis testing based upon $\hat{\underline{\beta}}_{KB}$. For symmetric trimming and symmetric F, the theory is applicable to $\hat{\underline{\beta}}_{PE}(RQ)$ as well. The asymptotic covariance matrix, $\sigma^2(\alpha,F)Q^{-1}$, can be consistently estimated since $n^{-1}X'X \to Q$, and a consistent estimate of $\sigma^2(\alpha,F)$ is provided by the next theorem.

Theorem 8.1. *Let* S *be the sum of squares for residuals calculated from the trimmed sample, i.e.*

$$S = \underline{y}'B(I_p - X(X'BX)^- X')B\underline{y} .$$

Let $c_j = \underline{e}'[\hat{\underline{\beta}}(\alpha_j) - \hat{\underline{\beta}}_{KB}(\underline{\alpha})]$ *for* j = 1, 2, *and*

$$s^2(\underline{\alpha},F) = (\alpha_2 - \alpha_1)^{-2}\left((n-p)^{-1}S + \alpha_1 c_1^2 + (1-\alpha_2)c_2^2 - (\alpha_1 c_1 + (1-\alpha_2)c_2)^2\right) .$$

Then

$$s^2(\underline{\alpha},F) \xrightarrow{p} \sigma^2(\underline{\alpha},F) .$$

Theorem 8.2. *Suppose* m *is the number of observations which have been removed by trimming. For* $0 < \varepsilon < 1$, *let* $F(n_1,n_2,\varepsilon)$ *denote the* $(1-\varepsilon)$ *quantile of the* F *distribution with* n_1 *and* n_2 *degrees of freedom and let*

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

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

$$\lim_{n\to\infty} P\{(K'\hat{\underline{\beta}}_{KB}(\underline{\alpha})-\underline{c})' \; [K'(X'AX)^{-1} K]^{-1} \; (K'\hat{\underline{\beta}}_{KB}(\underline{\alpha})-\underline{c}) \ge d(\ell , n-m-p , \varepsilon)\} = \varepsilon .$$

Letting $K = I_p$ and $\underline{c} = \underline{\beta} - \underline{\delta}(\underline{\alpha})$, the confidence ellipsoid

$$(\hat{\underline{\beta}}_{KB}(\underline{\alpha}) - \underline{\beta} - \underline{\delta}(\underline{\alpha}))' \; (X'AX) (\hat{\underline{\beta}}_{KB}(\underline{\alpha}) - \underline{\beta} - \underline{\delta}(\underline{\alpha})) \le d(\ell, n-m-p , \varepsilon) \qquad (8.1)$$

for $\underline{\beta} + \underline{\delta}(\underline{\alpha})$ has an asymptotic confidence coefficient of $(1-\varepsilon)$. Moreover, if we test

$$H_0: K'(\underline{\beta} + \underline{\delta}(\underline{\alpha})) = \underline{c}$$

versus

$$H_1: K'(\underline{\beta} + \underline{\delta}(\underline{\alpha})) \ne \underline{c}$$

by rejecting H_0 whenever

$$(K'\hat{\underline{\beta}}_{KB}(\underline{\alpha})-\underline{c})' \; [K'(X'AX)^{-1} K]^{-1} \; (K'(\hat{\underline{\beta}}_{KB}(\underline{\alpha})-\underline{c}) \ge d(\ell, n-m-p , \varepsilon) \qquad (8.2)$$

then the asymptotic size of our test is ε.

Of course, in the special cases where $\alpha_1 = 0$, $\alpha_2 = 1$ (so $m = 0$ and $A = I$) and F is Gaussian, (8.1) is an exact $1-\varepsilon$ confidence ellipsoid and (8.2) is an exact size ε test.

9. Examples.

In this section we contrast the results obtained for different estimates when applied to two data sets: (i) the *stackloss* data set given by Andrews (1974) and (ii) a set of measurements of *water salinity* and river discharge taken in North Carolina's Pamlico Sound (see Table 2). The estimates we consider are listed in Table 2. Both HUBER and ANDREW are M-estimates and are calculated by the iterative solution to

$$\sum_{i=1}^{N} \psi((y_i - x_i\beta)/s)x_i = 0 ,$$

where s = MAD/C, C is a constant, and MAD = median of the absolute values of the residuals. For Huber, C = .6745 and $\psi(Z) = \max(-1.25, \min(Z, 1.25))$. This choice of ψ should give results for normal data similar to those for the regression analogues of a 10% trimmed mean. The estimate ANDREW uses $C = 1$ and $\psi(Z) = \text{sine}(Z)I(|Z| \le \pi)$.

We defined $\hat{\underline{\beta}}_{KB}$ a bit differently than in section 2. Both data sets have four independent variables and each regression quantile hyperplane passes through four observations. Therefore, if one defines $\hat{\underline{\beta}}_{KB}$ as in section 2, at least eight observations are trimmed. Instead, we defined $\hat{\underline{\beta}}_{KB}$ by requiring strict inequality in (2.2). If $\underline{\alpha} = (.1,.9)$, this leads to no trimming for the stackloss data, and only two observations trimmed for the salinity data, so we use $\underline{\alpha} = (.15,.85)$. Then observations 4, 9, and 21 are trimmed in the stackloss data and observations 1, 13, 15, and 17 in the salinity data.

An important advantage of $\hat{\underline{\beta}}_{PE}(RQ)$ over $\hat{\underline{\beta}}_{KB}$ is that residuals from a preliminary estimate are rarely tied (at least in these data sets), and with $\hat{\underline{\beta}}_{PE}(RQ)$ one can have the actual percent trimming close to any specified α. The observations deleted when calculating $\hat{\underline{\beta}}_{PE}(RQ,.10)$ are 1, 3, 9, and 21 for the stackloss data and 1, 11, 13, 15, 16, and 17 for the salinity data.

Since both data sets have outliers, asymptotic theory and Monte-Carlo studies for the location problem (Andrews et al. (1972)) lead us to expect that LSE will be worst, ANDREW will do very well, and HUBER, $\hat{\underline{\beta}}_{PE}(RQ)$, and $\hat{\underline{\beta}}_{KB}$ will have roughly comparable performances. Of course, with these data the true parameters are unknown, and we can only measure performance by closeness of fit to the bulk of the observations, say with MAD or IQR (= interquartile range of the residuals). Using either MAD or IQR as criteria, our study does seem to agree with our expectations. The redescending M-estimator (ANDREWS) appears to be best overall.

Also, we have included $\hat{\underline{\beta}}(.5)$, the least absolute deviation estimate. Its performance was quite good here, but of course it is known to have rather poor efficiency at the normal model.

In Table 3, we list the regression coefficients, MAD, and IQR for each estimator. Figures 1 and 2 are box plots of the residuals and were obtained from the SAS package.

Least squares computations were performed on SAS. Regression quantiles were computed using MPS/360, a linear programming package, and LPMPS, a preprocessor for MPS/360 (McKeown and Rubin (1977)).

10. Summary.

We have considered two methods of defining a trimmed least squares estimator: $\hat{\underline{\beta}}_{KB}$, which uses Koenker and Bassett's (1978) regression quantiles, and $\hat{\underline{\beta}}_{PE}$, which uses a preliminary estimate.

Despite its intuitive appeal, $\hat{\underline{\beta}}_{PE}$ based on an arbitrary preliminary estimate will not be very satisfactory. Its behavior will depend heavily upon the choice of the preliminary estimate. Some choices (e.g. median regression) result in very inefficient trimmed estimates at the normal distribution, even if the trimming proportion is small. Other choices (e.g. least squares) can lead to low efficiency for heavy-tailed distributions, especially if the trimming proportion is high. Moreover, the contribution of the preliminary estimate to the variance of $\hat{\underline{\beta}}_{PE}$ depends on the density of the error distribution and might be difficult to estimate in practical situations.

The estimate $\hat{\underline{\beta}}_{KB}$ behaves analogously to a trimmed mean. Also, for a particular choice of preliminary estimate, the average of two regression quantiles, $\hat{\underline{\beta}}_{PE}$ (which for this preliminary estimate we call $\hat{\underline{\beta}}_{PE}(RQ)$), is asymptotically equivalent to $\hat{\underline{\beta}}_{KB}$, provided the error distribution is symmetric.

For moderately-sized data sets, $\hat{\underline{\beta}}_{PE}(RQ)$ has one major advantage over $\hat{\underline{\beta}}_{KB}$; with $\hat{\underline{\beta}}_{PE}(RQ)$ the proportion of observations rejected can be made quite close to any specified α. Since the number of observations lying on a regression quantile hyperplane is typically equal to the number of independent variables, $\hat{\underline{\beta}}_{KB}$ does not share this property with $\hat{\underline{\beta}}_{PE}(RQ)$.

The trimmed estimates $\hat{\underline{\beta}}_{KB}$ and $\hat{\underline{\beta}}_{PE}(RQ)$ seem to be worthwhile alternatives to M-estimates based on Huber's ψ, but perhaps not to redescending M-estimates.

Appendix

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

Proof. Routine. Use the continuity of F. □

Lemma A.2. *Let* $r_1,\ldots,r_n$ *be the residuals from* $\hat{\underline{\beta}}_0$, *suppose* $0 < \theta < 1$, *and let* μ_n *be a sequence of solutions to*

$$\sum_{i=1}^{n} \rho_\theta(r_i - \mu_n) = \min .$$

Then

$$n^{-\frac{1}{2}} \sum_{i=1}^{n} \psi_\theta(r_i - \mu_n) \to 0 \ a.s. \tag{A1}$$

In addition, the sequence of solutions $\hat{\underline{\beta}}(\theta)$ *of* (1.5) *satisfies*

$$n^{-\frac{1}{2}} \sum_{i=1}^{n} \underline{x}_i \psi_\theta(y_i - \underline{x}_i \hat{\underline{\beta}}(\theta)) \to \underline{0} \ a.s. \tag{A2}$$

Proof. We will proof only (A2) because (A1) can be demonstrated in a similar manner.

Let $\{\underline{e}_j\}_{j=1}^{p}$ be the standard basis of R^p. Define

$$G_j(a) = \sum_{i=1}^{n} \rho_\theta(y_i - \underline{x}_i(\hat{\underline{\beta}}(\theta) + a\underline{e}_j))$$

and let $H_j(t)$ be the derivative from the right of G_j so that

$$H_j(a) = \sum_{i=1}^{n} x_{ij}\psi_\theta(y_i - \underline{x}_i(\underline{\hat{\beta}}(\theta) + a\underline{e}_j)) .$$

Notice that $H_j(a)$ is non-decreasing. Therefore, for $\varepsilon > 0$

$$H_j(-\varepsilon) \le H_j(0) \le H_j(\varepsilon) ,$$

and because $G_j(a)$ achieves its minimum at $a = 0$,

$$H_j(-\varepsilon) \le 0 \text{ and } H_j(\varepsilon) \ge 0 .$$

Consequently,

$$|H_j(0)| \le H_j(\varepsilon) - H_j(-\varepsilon) . \tag{A3}$$

Letting $\varepsilon \to 0$ in (A3), we see that

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

Now (A2) follows from Lemma A.1. □

Lemma A.3. *For* $\underline{\Delta} \in R^p$, *define*

$$M(\underline{\Delta}) = n^{-\frac{1}{2}} \sum_{i=1}^{n} \underline{x}_i' \psi_\theta(Z_i - \underline{x}_i \underline{\Delta} n^{-\frac{1}{2}} - \xi_\theta) .$$

Then for all $L > 0$

$$\sup_{0 \le ||\underline{\Delta}|| \le L} ||M(\underline{\Delta}) - M(\underline{0}) + f(\xi_\theta) Q \underline{\Delta}|| = o_p(1) . \tag{A4}$$

Proof. The result follows from Lemma 4.1 of Bickel (1975) because

$$E(M_n(\underline{\Delta}) - M_n(\underline{0})) \to -f(\xi_\theta)\underline{e}'\underline{\Delta} \ . \qquad \square$$

Remark. Equation (A4) is a special case of the conclusion of Jurecková's (1977) Theorem 4.1, which she proves under conditions different from ours. Her C_{ji} is our $x_{ij}n^{-\frac{1}{2}}$.

Proof of Lemma 3.1. Since $\mu = r_{\theta n}$ is a solution to

$$\sum_{i=1}^{n} \rho_\theta(r_i - \mu) = \min \ ,$$

(A1) implies that

$$n^{-\frac{1}{2}} \sum_{i=1}^{n} \psi_\theta(Z_i - \xi_\theta - \underline{x}_i((\hat{\underline{\beta}}_0 - \underline{\beta}) + \underline{e}(r_{\theta n} - \xi_\theta))) \to 0 \text{ a.s.} \qquad \text{(A5)}$$

Define $V(\underline{\Delta}) = n^{-\frac{1}{2}} \sum_{i=1}^{n} \psi_\theta(Z_i - \underline{x}_i\underline{\Delta}n^{-\frac{1}{2}} - \xi_\theta)$. Using the method of Jurecková (1977, proof of Lemma 5.2) and (A4), we can show that for all $\varepsilon > 0$ there exists η, K, and n_0 such that

$$P(\inf_{|\underline{e}'\Delta| > K} |V(\underline{\Delta})| < \eta) < \varepsilon \text{ for } n \geq n_0 \ . \qquad \text{(A6)}$$

Next, (A5) and (A6) allow us to conclude that

$$n^{\frac{1}{2}}(\underline{e}'(\hat{\underline{\beta}}_0 - \underline{\beta}) + r_{\theta n} - \xi_\theta) = O_p(1) \ . \qquad \text{(A7)}$$

By (A7) and C5 we may substitute $n^{\frac{1}{2}}(\hat{\underline{\beta}}_0 - \underline{\beta} + \underline{e}(r_{\theta n} - \xi_\theta))$ for $\underline{\Delta}$ in (A4) and complete the proof by examining first coordinates, using C2. $\square$

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

$$\sup_n (n^{-1} \sum_{i=1}^{n} ||D_i||^2) < \infty ,$$

where $||D_i||^2 = \text{Tr } D_i' D_i$ *is the Euclidean norm of* D_i. *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* $\underline{\Delta}_1$, $\underline{\Delta}_2$, *and* $\underline{\Delta}_3$ *in* R^p *and* $\underline{\Delta} = (\underline{\Delta}_1, \underline{\Delta}_2, \underline{\Delta}_3)$ *define*

$$T(\underline{\Delta}) = n^{-\frac{1}{2}} \sum_{i=1}^{n} D_i g(Z_i + \underline{\Delta}_3 \underline{x}_i n^{-\frac{1}{2}}) I\{\xi_1 + \underline{x}_i \underline{\Delta}_i n^{-\frac{1}{2}} < Z_i < \xi_2 + \underline{x}_i \underline{\Delta}_2 n^{-\frac{1}{2}}\} .$$

Then, for all M > 0,

$$\sup_{||\Delta|| \le M} |T(\underline{\Delta}) - T(\underline{0}) - E(T(\underline{\Delta}) - T(\underline{0}))| = o_p(1) .$$

Proof. The proof is very similar to that of Bickel's (1975) Lemma 4.1 and is omitted here, but it can be found in Ruppert and Carroll (1978). □

Proof of Theorem 3.1. For $\underline{\Delta}_1$, $\underline{\Delta}_2$ in R^p and $\underline{\Delta} = (\underline{\Delta}_1, \underline{\Delta}_2)$, define

$$U(\underline{\Delta}) = n^{-1} \sum_{i=1}^{n} \underline{x}_i \underline{x}_i' I(\xi_1 + \underline{x}_i' \underline{\Delta}_1 n^{-\frac{1}{2}} \le Z_i \le \xi_2 + \underline{x}_i' \underline{\Delta}_2 n^{-\frac{1}{2}}\}$$

and

$$W(\underline{\Delta}) = n^{-\frac{1}{2}} \sum_{i=1}^{n} \underline{x}_i z_i I\{\xi_1 + \underline{x}_i' \underline{\Delta}_1 n^{-\frac{1}{2}} \le Z_i \le \xi_2 + \underline{x}_i' \underline{\Delta}_2 n^{-\frac{1}{2}}\} .$$

Using Lemma A.4, it is easy to show that for all $M > 0$,

$$\sup_{0 \le ||\underline{\Delta}|| \le M} |U(\underline{\Delta}) - (1-2\alpha)Q| = o_p(1) \tag{A8}$$

and

$$\sup_{0 \le ||\underline{\Delta}|| \le M} |W(\underline{\Delta}) - W(\underline{0}) - Q(\underline{\Delta}_2 \xi_2 f(\xi_2) - \underline{\Delta}_1 \xi_1 f(\xi_1))| = o_p(1) . \tag{A9}$$

Then using the fact that $\underline{x}_i' \underline{e} = 1$, we have

$$I\{r_{1n} \le r_i \le r_{2n}\} = I\{\xi_1 + \underline{x}_i'((\underline{\hat{\beta}}_0 - \underline{\beta}) + \underline{e}(r_{1n} - \xi_1)) \le Z_i$$
$$\le \xi_2 + \underline{x}_i'((\underline{\hat{\beta}}_0 - \underline{\beta}) + \underline{e}(r_{2n} - \xi_2))\} ,$$

and so replacing $\underline{\Delta}_\ell$ by $n^{\frac{1}{2}}((\underline{\hat{\beta}}_0 - \underline{\beta}) + \underline{e}(r_{\ell n} - \xi_\ell))$ for $\ell = 1,2$ in (A8) and (A9), we have

$$n^{-1}(X' AX) = (1-2\alpha)Q + o_p(1) \tag{A10}$$

and

$$n^{-\frac{1}{2}} X' A(\underline{y} - AX\underline{\beta}) = W(\underline{0}) + Q\{\xi_2 f(\xi_2) n^{\frac{1}{2}}(\underline{\hat{\beta}}_0 - \underline{\beta} + \underline{e}(r_{2n} - \xi_2))$$
$$- \xi_1 f(\xi_1) n^{\frac{1}{2}}(\underline{\hat{\beta}}_0 - \underline{\beta} + \underline{e}(r_{1n} - \xi_1))\} + o_p(1) . \tag{A11}$$

By (A10)

$$n^{\frac{1}{2}}(X' A(y - AX\underline{\beta})) = (1-2\alpha) n^{\frac{1}{2}} Q(\underline{\hat{\beta}}_{PE} - \underline{\beta}) + o_p(1) . \tag{A12}$$

By (A11), (A12) and Lemma 3.1

$$(1-2\alpha)n^{\frac{1}{2}} Q(\hat{\underline{\beta}}_{PE} - \underline{\beta}) = W(\underline{0}) + Q\{\xi_2 \underline{e}n^{-\frac{1}{2}} \sum_{i=1}^{n} \psi_{1-\alpha}(Z_i - \xi_2) - \xi_1 \underline{e}n^{-\frac{1}{2}} \sum_{i=1}^{n} \psi_\alpha(Z_i - \xi_1)$$

$$+ n^{\frac{1}{2}} a(I - \underline{e}\,\underline{e}')(\hat{\underline{\beta}}_0 - \underline{\beta})\} + o_p(1) .$$

Then (3.3) follows from the definition of $W(\underline{0})$. □

Proof of Theorem 5.3. Using (A4) and the method of Jurecková (1977, proof of Lemma 5.2) we can show that

$$n^{\frac{1}{2}}(\hat{\underline{\beta}}(\theta) - \underline{\beta}(\theta)) = O_p(1) .$$

Therefore, we can substitute $n^{\frac{1}{2}}(\hat{\underline{\beta}}(\underline{\theta}) - \underline{\beta}(\theta))$ for $\underline{\Delta}$ in (A4) and use (A2) to obtain

$$M(0) = f(\xi_\theta)Qn^{\frac{1}{2}}(\hat{\underline{\beta}}(\theta) - \underline{\beta}(\theta)) + o_p(1)$$

and Theorem 5.3 follows easily. □

Proof of Theorem 5.1. The proof is quite similar to the proof of Theorem 3.1 and can be found in Ruppert and Carroll (1978). □

Proof of Theorem 8.1. For $\underline{\Delta}_1, \underline{\Delta}_2, \underline{\Delta}_3$ in R^p define $\underline{\Delta} = (\underline{\Delta}_1, \underline{\Delta}_2, \underline{\Delta}_3)$ and

$$V(\underline{\Delta}) = n^{-1} \sum_{i=1}^{n} (Z_i - \underline{x}_i \underline{\Delta}_1 n^{-\frac{1}{2}} - \delta(\underline{\alpha}))^2 \, I(\xi_1 + \underline{x}_i' \underline{\Delta}_2 n^{-\frac{1}{2}} \le Z_i \le \xi_2 + \underline{x}_i' \underline{\Delta}_3 n^{-\frac{1}{2}}) .$$

We see that

$$S = nV(\sqrt{n}(\hat{\underline{\beta}}_{KB}(\underline{\alpha}) - (\underline{\beta} + \underline{\delta}(\alpha)) , \sqrt{n}(\hat{\underline{\beta}}(\alpha_1) - \underline{\beta}(\alpha_1)) , \sqrt{n}(\hat{\underline{\beta}}(\alpha_2) - \underline{\beta}(\alpha_2))) .$$

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

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

By a Taylor expansion of F and additional simple calculations

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

whence

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

Therefore by Corollary 5.1 and (5.2) we have

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

Now Var V(0) $\to$ 0, so by Chebyshev's inequality,

$$\begin{aligned} S &= EV(\underline{0}) + o_p(1) \\ &= E(Z_i - \delta(\underline{\alpha}))^2 \, I(\xi_1 \leq Z_i \leq \xi_2) + o_p(1) . \end{aligned}$$

Furthermore for $j = 1,2$

$$c_j = \xi_j - \delta(\underline{\alpha}) + o_p(1)$$

by Corollary 5.1 and (5.2), and Theorem 8.1 follows. □

<u>Proof of Theorem 8.2</u>. This follows in a straightforward manner from (5.2), Theorem 8.1, and Theorem 4.4 of Billingsley (1968). □

Table 1 - Variances of the asymptotic distribution of slope estimators or contaminated normal distributions. (The asymptotic covariance matrix is $\hat{Q}^{-1}$ multiplied by the displayed quantity.)

ε*	b**	Estimator											
					Trimmed Least Squares								
		Least Squares	Huber Proposal 2	Hampel One Step	$\hat{\underline{\beta}}_{PE}(LS)$ (Least Squares as Preliminary Estimate)			$\hat{\underline{\beta}}_{PE}(LAD)$ (Least Absolute Deviation as Preliminary Estimate)			$\hat{\underline{\beta}}_{PE}(RQ)$ (Average of αth and $1-\alpha$th Regression Quantiles as Preliminary Estimate)		
					α=.05	α=.10	α=.25	α=.05	α=.10	α=.25	α=.05	α=.10	α=.25
NORMAL		1.00	1.04	1.04	1.30	1.36	1.26	1.54	1.83	2.14	1.03	1.06	1.19
0.05	3.0	1.40	1.16	1.17	1.38	1.51	1.58	1.54	1.88	2.26	1.16	1.17	1.29
0.05	5.0	2.20	1.20	1.23	1.43	1.71	2.15	1.51	1.87	2.28	1.20	1.20	1.31
0.05	10.0	5.95	1.23	1.28	1.68	2.66	4.81	1.46	1.85	2.30	1.25	1.23	1.33
0.10	3.0	1.80	1.30	1.32	1.44	1.64	1.88	1.56	1.93	2.39	1.32	1.30	1.39
0.10	5.0	3.40	1.40	1.47	1.45	1.96	2.99	1.46	1.90	2.44	1.46	1.38	1.45
0.10	10.0	10.90	1.49	1.61	1.48	3.32	8.09	1.34	1.85	2.47	1.65	1.45	1.49
0.25	3.0	3.00	1.90	1.94	1.79	1.97	2.74	1.82	2.12	2.87	2.14	1.85	1.80
0.25	5.0	7.00	2.46	2.68	2.49	2.09	5.13	2.37	1.92	2.99	4.11	2.39	2.01
0.25	10.0	25.75	3.20	4.26	6.50	1.88	15.66	5.51	1.65	3.06	13.65	3.69	2.19

*Proportion of contamination

**Standard deviation of contamination

Table 2

The *water salinity* data set. The values are biweekly averages of SALINITY at time period i, SALLAG = salinity at time i-1, i = TREND = one of the six biweekly periods in March-May and H20FLOW = river discharge in time i.

OBS	SALINITY	SALLAG	TREND	H20FLOW	YEAR
1	7.6	8.2	4	23.005	72
2	7.7	7.6	5	23.873	
3	4.3	4.6	0	26.417	73
4	5.9	4.3	1	24.868	
5	5.0	5.9	2	29.895	
6	6.5	5.0	3	24.200	
7	8.3	6.5	4	23.215	
8	8.2	8.3	5	21.862	
9	13.2	10.1	0	22.274	74
10	12.6	13.2	1	23.830	
11	10.4	12.6	2	25.144	
12	10.8	10.4	3	22.430	
13	13.1	10.8	4	21.785	
14	12.3	13.1	5	22.380	
15	10.4	13.3	0	23.927	75
16	10.5	10.4	1	33.443	
17	7.7	10.5	2	24.859	
18	9.5	7.7	3	22.686	
19	12.0	10.0	0	21.789	76
20	12.6	12.0	1	22.041	
21	13.6	12.1	4	21.033	
22	14.1	13.6	5	21.005	
23	13.5	15.0	0	25.865	77
24	11.5	13.5	1	26.290	
25	12.0	11.5	2	22.932	
26	13.0	12.0	3	21.313	
27	14.1	13.0	4	20.769	
28	15.1	14.1	5	21.393	

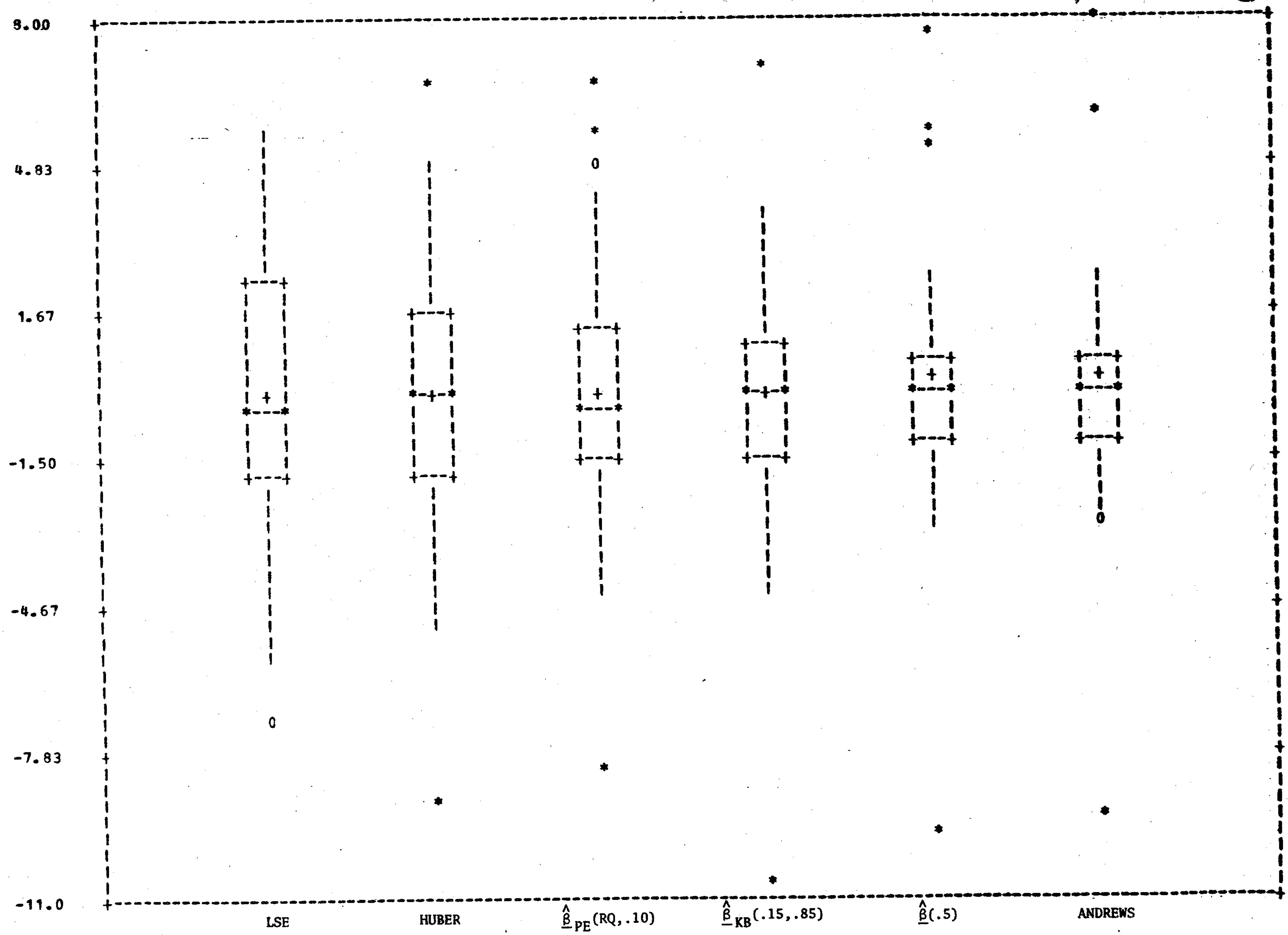
Table 3

Regression coefficients, MAD and IQR for the Stackloss data.

Code	Intercept	Air Flow	Temperature	Acid	MAD	IQR
LSE	39.92	-.72	-1.30	.15	1.92	3.12
$\hat{\underline{\beta}}(.50)$	39.69	-.83	-.57	.06	1.18	1.71
$\hat{\underline{\beta}}_{KB}(.15)$	42.83	-.93	-.63	.10	1.60	2.49
$\hat{\underline{\beta}}_{PE}(RQ,.10)$	40.37	-.72	-.96	.07	1.37	2.59
HUBER	41.00	-.83	-.91	.13	1.63	3.07
ANDREW	37.20	-.82	-.52	.07	.99	1.50

Water Salinity Data

Code	Intercept	SALLAG	TREND	H2OFLOW	MAD	IQR
LSE	9.59	.777	-.026	-.295	.72	1.38
$\hat{\underline{\beta}}(.50)$	14.21	.740	-.111	-.458	.50	.98
$\hat{\underline{\beta}}_{KB}(.15)$	9.69	.800	-.128	-.290	.67	1.36
$\hat{\underline{\beta}}_{PE}(RQ,.10)$	14.49	.774	-.160	-.488	.60	1.05
HUBER	13.36	.756	-.094	-.439	.56	1.02
ANDREW	17.22	.733	-.196	-.578	.47	.83



<u>Figure 1</u>. Boxplots of Residuals for the Stackloss Data.

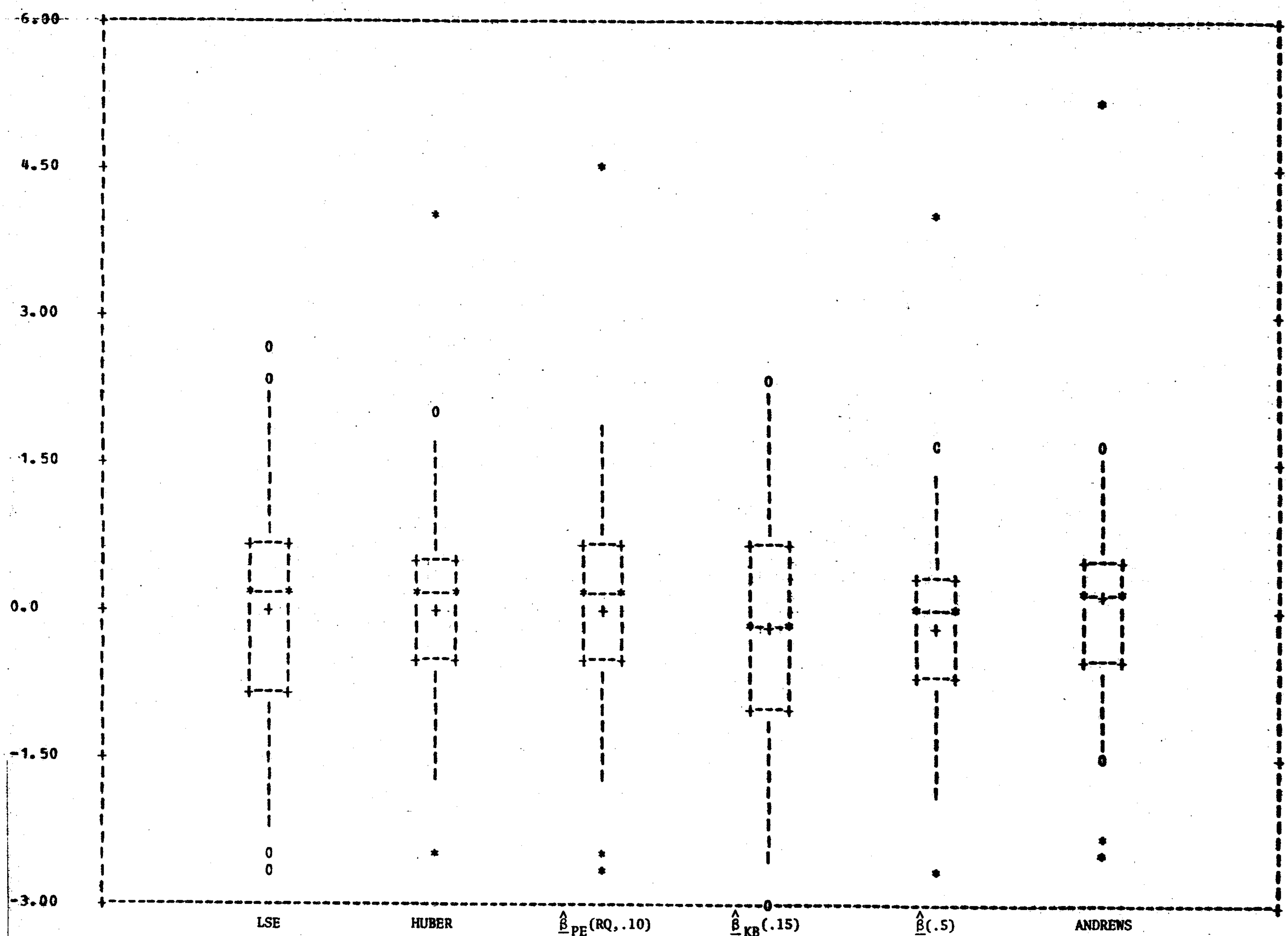


Figure 2. Boxplots of Residuals for the Water Salinity Data.

References

Andrews, D.F., et al. (1972). *Robust Estimates of Location: Survey and Advances*. Princeton, NJ: Princeton University Press.

Anscombe, F.J. (1961). Examination of residuals. In *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* (J. Neyman, ed.), pp. 1-36. Berkeley and Los Angeles, CA: University of California Press.

Bahadur, R.R. (1966). A note on quantiles in large samples. *Ann. Math. Statist. 37*, 577-580.

Bickel, Peter J. (1973). On some analogues to linear combinations of order statistics in the linear model. *Ann. Statist. 1*, 597-616.

Bickel, Peter J. (1975). One-step Huber estimates in the linear model. *J. Amer. Statist. Assoc. 70*, 428-433.

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

Carroll, Raymond J. (1979). On estimating variances of robust estimators when the errors are asymmetric. *J. Amer. Statist. Assoc. 74*, 674-679.

Carroll, Raymond J. and Ruppert, David (1979). Almost sure properties of robust regression estimates. *Institute of Statistics Mimeo Series #1240*, Department of Statistics, University of North Carolina at Chapel Hill.

deWet, T. and Venter, J.H. (1974). An asymptotic representation of trimmed means with applications. *South African Statistics Journal 8*, 127-134.

Ghosh, J.K. (1971). A new proof of the Bahadur representation of quantiles and an application. *Ann. Math. Statist. 42*, 1957-1961.

Gross, Alan M. (1973). A Monte-Carlo swindle for estimators of location. *App. Statist. 22*, 347-356.

Gross, Alan M. (1976). Confidence interval robustness with long-tailed symmetric distributions. *J. Amer. Statist. Assoc. 71*, 409-416.

Hogg, Robert V. (1974). Adaptive robust procedures: A partial review and some suggestions for future application and theory. *J. Amer. Statist. Assoc. 69*, 909-927.

Huber, Peter J. (1970). Studentizing robust estimates. In *Nonparametric Techniques in Statistical Inference* (M.L. Puri, ed.), pp. 453-463. Cambridge: Cambridge University Press.

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

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

Jurecková, Jana (1977). Asymptotic relations of M-estimates and R-estimates in linear regression model. *Ann. Statist. 5*, 464-472.

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

McKeown, P.G. and Rubin, D.S. (1977). A student oriented preprocessor for MPS/360. *Computers and Operations Research 4*, 227-229.

Ruppert, David and Carroll, Raymond J. (1978). Robust regression by trimmed least squares estimation. *Institute of Statistics Mimeo Series #1186*, Department of Statistics, University of North Carolina at Chapel Hill.

Stigler, Steven M. (1977). Do robust estimators work with real data? *Ann. Statist. 5*, 1055-1098.

Tukey, John W. (1960). A survey of sampling from contaminated distributions. In *Contributions to Probability and Statistics* (I. Olkin, ed.). Stanford, CA: Stanford University Press.

Tukey, John W. and McLaughlin, D.H. (1963). Less vulnerable confidence and significance procedures for location based on a single sample. *Sankhyā Ser. A 25*, 331-352.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle) Trimmed Least Squares Estimation in the Linear Model	5. TYPE OF REPORT & PERIOD COVERED TECHNICAL
	6. PERFORMING ORG. REPORT NUMBER Mimeo Series No. 1270
7. AUTHOR(s) David Ruppert and Raymond J. Carroll	8. CONTRACT OR GRANT NUMBER(s) NSF Grant MCS78-01240 AFOSR Contract AFOSR-75-2796 AFOSR Contract AFOSR-80-0080
9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research Bolling Air Force Base Washington, DC 20332	12. REPORT DATE February 1980
	13. NUMBER OF PAGES 40 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)

18. SUPPLEMENTARY NOTES

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

regression analogue, trimmed mean, regression quantile, preliminary estimator, linear model, trimmed least squares

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

We consider two methods of defining a regression analogue to a trimmed mean. The first was suggested by Koenker and Bassett and uses their concept of regression quantiles. Its asymptotic behavior is completely analogous to that of a trimmed mean. The second method uses residuals from a preliminary estimator. Its asymptotic behavior depends heavily on the preliminary estimate; it behaves, in general, quite differently than the estimator proposed by Koenker and Bassett, and it can be rather inefficient at the

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

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. normal model even if the percent trimming is small. However, if the preliminary estimator is the average of the two regression quantiles used with Koenker and Bassett's estimator, then the first and second methods are asymptotically equivalent for symmetric error distributions.