MALLOWS-TYPE BOUNDED INFLUENCE REGRESSION TRIMMED MEANS

by

P.J. DE JONGH
Institute for Maritime Technology
P.O. Box 181, Simon's Town 7995
Republic of South Africa

T. DE WET
Institute for Maritime Technology
P.O. Box 181, Simon's Town 7995
Republic of South Africa

A.H. WELSH
Department of Statistics
University of Chicago
5734 University Ave., Chicago, Il 60637

Keywords and phrases: trimmed means, regression, robust, bounded influence.

ABSTRACT

The influence functions of the regression trimmed mean estimators proposed by Koenker and Bassett (1978) and Welsh (1987) are bounded in the dependent variable space but not in the independent variable space. In this paper we follow the approach of Mallows (1973, 1975) and modify these estimators so that the resulting estimators have bounded influence functions. The large sample behaviour of these estimators is studied and it is shown that they have the same asymptotic distribution. The small sample behaviour of the ordinary and bounded influence regression trimmed means is then investigated by means of a Monte Carlo study and by applying the estimators to the water salinity data (see e.g. Ruppert and Carroll (1980)). Based on these results we conclude that potentially one can gain much by using bounded influence regression trimmed means over ordinary regression trimmed means. However, there does not seem to be a clear choice between the Koenker-Bassett and Welsh versions.

1. INTRODUCTION

In this paper we extend the regression trimmed means of Koenker and Bassett (1978) and Welsh (1987) by defining bounded influence regression trimmed mean estimators.

Consider the usual linear regression model

(1.1) $$Y_i = x_i' \beta + e_i, \quad i = 1,2,\ldots,n$$

with Y_i the i-th response (dependent variable), $x_i' = [1,x_{i2},\ldots,x_{ip}]$ the i-th row of the design matrix X (consisting of p-1 independent variables or carriers), β a vector of p unknown regression parameters and e_i the i-th error. We assume the errors to be independent and identically distributed according to some unknown distribution F. For simplicity, we will assume that F is symmetric about zero. We are concerned with constructing bounded influence regression trimmed mean estimators for β. A key concept in doing this is the influence function of an estimator introduced by Hampel (1968, 1974).

Let $0 < \alpha < 0.5$, then the influence function of the $100\alpha\%$ regression trimmed mean estimators of Koenker and Bassett (1978) and Welsh (1987) has the following form

(1.2) $$\Lambda(y,x) = Q^{-1}x\, g_\alpha(y-x'\beta),$$

where $\Lambda : R \times R^p \to R^p$ is the influence function, $Q = \lim_{n\to\infty}(X'X)/n$ and $g_\alpha(\cdot)$ is the influence function of the $100\alpha\%$ trimmed mean in the location case, i.e.

(1.3) $$g_\alpha(u) = \begin{cases} F^{-1}(\alpha)/(1-2\alpha) & u < F^{-1}(\alpha) \\ u/(1-2\alpha) & F^{-1}(\alpha) \le u \le F^{-1}(1-\alpha) \\ F^{-1}(1-\alpha)/(1-2\alpha) & u > F^{-1}(1-\alpha) \end{cases}.$$

Note that $g_\alpha(\cdot)$ limits the effect of the residuals and $\Lambda(y,x)$ is thus bounded in the dependent variable space. However, it is not bounded in the independent variable space.

Therefore one can conjecture that, in small samples, these ordinary robust estimators will be able to handle outliers in the Y-space but not outliers in the X-space. This is indeed so in the following situations (see de Jongh and de Wet (1987)):

Y_i is observed according to the model (1.1), but a gross error occurs in the corresponding design point x_i, e.g. a keypunch error.

Y_i is observed with a large error e_i and the corresponding x_i is an outlying design point (called a leverage point).

The above two cases are examples of so-called influential observations. See e.g. Denby and Larsen (1977), Hill (1982) and Chatterjee and Hadi (1986).

Methods of dealing with influential observations were proposed by e.g. Mallows (1973, 1975) and Krasker and Welsch (1982). The basic difference between their approaches is the way in which the influence function of an estimator is bounded. Mallows approach is to bound the influence separately over the design and residuals while this is done simultaneously by Krasker and Welsch. Recently, Giltinan et al. (1986) extended these methods to cover heteroscedastic regression models and they compared it over a number of cases. They found neither method clearly preferable in all cases. As advantages of the Mallows approach they note that it is more informative from a diagnostic viewpoint and that theoretically it should give more stable inference than the Krasker-Welsch approach since estimators in the Mallows class do have a bounded change-of-variance function (CVF). A disadvantage is that in general the Mallows approach will result in estimators that are less efficient at the normal model, since a point with high leverage is automatically downweighted, irrespective of the size of the residual. A nice rationale for applying Mallows' method is in the estimation of the parameters of an autoregressive process under an additive outlier model. See e.g. de Jongh and de Wet (1985).

In this paper we will follow the Mallows approach in defining bounded influence regression trimmed means based on the regression trimmed mean estimators of Koenker and Bassett (1978) and Welsh (1987). These estimators will be discussed in the next section. The asymptotic results are given in Section 3. In Section 4 we describe the Monte Carlo design and the results are given in Section 5. An example with real data is considered in Section 6.

2. THE ESTIMATORS

In this section we describe the bounded influence trimmed means.

2.1 THE BOUNDED INFLUENCE KOENKER-BASSETT ESTIMATOR

We define the $100\alpha\%$ bounded influence (Koenker-Bassett) trimmed mean as the solution $\hat{\beta}^{(w)}_{KB}(\alpha)$ to the minimization problem

$$\min_{\beta \in R^p} \sum_{i=1}^{n} d_i^{(w)}(\alpha)\ (Y_i - x_i'\beta)^2 \tag{2.1}$$

with

$$d_i^{(w)}(\alpha) = \begin{cases} w_i & x_i'\hat{\beta}^{(w)}_{RQ}(\alpha) \le Y_i \le x_i'\hat{\beta}^{(w)}_{RQ}(1-\alpha) \\ 0 & \text{otherwise} \end{cases} .$$

Here $\hat{\beta}^{(w)}_{RQ}(\theta)$ is the θ-th (bounded influence) sample regression quantile ($0 < \theta < 1$) which is obtained as any solution to the minimization problem

$$\min_{\beta \in R^p} \sum_{i=1}^{n} w_i\ \rho_\theta(Y_i - x_i'\beta) \tag{2.2}$$

with

$$\rho_\theta(u) = u\,\psi_\theta(u),\ \psi_\theta(u) = \theta - I(u < 0)$$

and $I(\cdot)$ the indicator function. This estimator is a weighted least squares estimator with weights $d_i^{(w)}(\alpha)$. Let $D_\alpha^{(w)} = \text{diag}(d_i^{(w)}(\alpha))$ then

$$(2.3) \qquad \hat{\beta}^{(w)}_{KB}(\alpha) = (X'D^{(w)}_{\alpha}X)^{-1}\, X'\, D^{(w)}_{\alpha}\, Y\ .$$

Essentially one can think of $\hat{\beta}^{(w)}_{KB}(\alpha)$ as the weighted least squares estimator (with weights w_i) of the observations remaining after discarding observations lying outside the α-th and $(1-\alpha)$-th bounded influence regression quantiles.

With $w_i=1$, $1 \leq i \leq n$, $\hat{\beta}^{(w)}_{KB}(\alpha) = \hat{\beta}_{KB}(\alpha)$, the original Koenker-Bassett estimator. Ruppert and Carroll (1980) showed that the asymptotic distribution of this estimator is given by

$$(2.4) \qquad \sqrt{n}(\hat{\beta}_{KB}(\alpha) - \beta) \overset{D}{\rightarrow} N(0, \sigma^2_{\alpha}\, Q^{-1})$$

$$\text{with } \sigma^2_{\alpha} = (1-2\alpha)^{-2} \left\{ \int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} e^2\, dF(e) + 2\alpha(F^{-1}(\alpha))^2 \right\}.$$

Here $\overset{D}{\rightarrow}$ denotes convergence in distribution and $N(\mu,\Sigma)$ a multivariate normal distribution with mean vector μ and covariance matrix Σ. Note that σ^2_{α}/n is the asymptotic variance of the trimmed mean in the location case.

2.2 THE BOUNDED INFLUENCE WELSH ESTIMATOR

Welsh (1987) based his estimator on the residuals obtained from a preliminary fit. Let $\hat{\beta}_{IN}$ be any initial estimator for β, denote the residuals from this fit by

$$(2.5) \qquad r_i = Y_i - x_i'\, \hat{\beta}_{IN}\ , \qquad i = 1,2,\ldots,n$$

and let $r_{(i)}$ be the i-th ordered value of the set of residuals $\{r_i\}$.

For $0 < \theta < 1$ and $s = [n\theta]$ define

$$(2.6) \qquad \xi_{\theta} = \begin{cases} r_{(s)} & s = n\theta \\ r_{(s+1)} & \text{otherwise} \end{cases}$$

with $[u]$ denoting the largest integer contained in u. We define the $100\alpha\%$ (Welsh) trimmed mean as

$$(2.7)\qquad \hat{\beta}^{(w)}_{WE}(\alpha) = (X'WB_{\alpha}X)^{-1}\, X'W(B_{\alpha}Y + h_{\alpha}(r))$$

where $W = \text{diag}(w_i)$ and $B_{\alpha} = \text{diag}(b_i(\alpha))$ with

$$(2.8)\qquad b_i(\alpha) = \begin{cases} 1 & \xi_{\alpha} \le r_i \le \xi_{1-\alpha} \\ 0 & \text{otherwise} \end{cases}$$

and $h_{\alpha}(r)$ a vector with i-th component $h_{\alpha}(r_i)$ where

$$(2.9)\qquad h_{\alpha}(u) = \begin{cases} \xi_{\alpha} - \alpha\,(\xi_{\alpha} + \xi_{1-\alpha}) & u < \xi_{\alpha} \\ -\alpha(\xi_{\alpha} + \xi_{1-\alpha}) & \xi_{\alpha} \le u \le \xi_{1-\alpha} \\ \xi_{1-\alpha} - \alpha(\xi_{\alpha} + \xi_{1-\alpha}) & u > \xi_{1-\alpha} \end{cases}.$$

Note the analogy between $B_{\alpha}Y + h_{\alpha}(r)$ and the influence function of the trimmed mean.

Again, with $w_i = 1$, $1 \le i \le n$, $\hat{\beta}^{(w)}_{WE}(\alpha) = \hat{\beta}_{WE}(\alpha)$, the original Welsh estimator. Welsh (1987, Theorem 1), showed that $\hat{\beta}_{WE}(\alpha)$ and $\hat{\beta}_{KB}(\alpha)$ are asymptotically equivalent. We will show in Section 3 that analogous results hold for $\hat{\beta}^{(w)}_{KB}(\alpha)$ and $\hat{\beta}^{(w)}_{WE}(\alpha)$.

It might seem that a natural approach to defining a $100\alpha\%$ trimmed mean would be to take the least squares estimator of the observations remaining after discarding observations whose residuals (from the preliminary fit) are less than ξ_{α} or greater than $\xi_{1-\alpha}$. This estimator, however, does not have asymptotic behaviour analogous to the trimmed mean. In fact the asymptotic behaviour depends heavily on the preliminary estimator (see Ruppert and Carroll (1980)) so that in general the estimator is neither robust nor efficient. However, Welsh (1987) showed that by "winsorizing" the preliminary residuals the desired asymptotic behaviour is obtained.

Remarks

(1) The above-defined estimators are scale and regression equivariant as well as equivariant wrt reparameterization of design. Note that for the well-known M-estimators (see e.g. Huber (1981)) this is only true if scale is estimated concurrently.

(2) Regression quantiles can be calculated very easily by formulating (2.2) as an L_1-problem and then using linear programming techniques. See de Jongh (1985) for details. Recently Narula and Wellington (1986) proposed a simple and efficient algorithm to find all regression quantiles associated with a data set.

(3) If least squares is used as the preliminary estimator the Welsh estimators have the advantage of computational simplicity over the Koenker-Bassett estimators. However, the Welsh estimators are sensitive to the choice of the initial estimator so that in practice one would prefer a robust estimator (e.g. L_1) over least squares.

(4) From the definitions of the estimators it follows that the Welsh estimator trims a fixed percentage ($200\alpha\%$) while the trimming proportion for the Koenker-Bassett estimator is variable. In practice the latter is typically slightly less than $200\alpha\%$.

3. ASYMPTOTIC RESULTS

The asymptotic behaviour of the bounded influence Koenker-Bassett trimmed mean and the bounded influence Welsh trimmed mean is given by Theorems 3A and 3B below. The proof of Theorem 3A follows very much along the lines of Ruppert and Carroll (1980) while the proof of Theorem 3B follows along the lines of Welsh (1987). Assuming symmetric trimming and a symmetric error distribution we give an outline of the proof of Theorem 3B. For the proof of Theorem 3A the interested reader is referred to de Jongh (1985). The results generalize in a straightforward way to the nonsymmetric case.

We shall proceed as follows. First we impose some conditions and then state the theorems. This will be followed by an outline of the proof of Theorem 3B.

3.1 CONDITIONS AND MAIN RESULTS

We impose the following conditions.

(i) F has continuous density f on the support of F and F is symmetric around 0.

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

with x_{ij} the (i,j)-th element of X.

(iii) For $x_i' = [x_{i1}, x_{i2}, \ldots, x_{ip}]$ the i-th row of X,

$$x_{i1} = 1 \qquad \forall\ i$$

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

Here $w = w(X)$ is a vector of weights that typically depends on the design X.

(iv) There exist positive definite matrices Q_w and B_w such that

$$\lim_{n\to\infty}(X'WX)/n = Q_w$$

$$\lim_{n\to\infty}(X'W^2X)/n = B_w$$

where $W = \text{diag}(w_i)$ and $W^2 = WW$.

(v) $\sqrt{n}(\hat{\beta}_{IN} - \beta)$ is bounded in probability.

We now state the theorems.

<u>THEOREM 3A</u>

Suppose conditions (i)-(iv) hold and let $0 < \alpha < 0.5$. Then the estimator $\hat{\beta}_{KB}^{(w)}(\alpha)$ satisfies

$$\sqrt{n}(\hat{\beta}_{KB}^{(w)}(\alpha) - \beta) = n^{-\frac{1}{2}} Q_w^{-1} \sum_{i=1}^{n} w_i x_i \, g_\alpha(e_i) + o_p(1)$$

and therefore

$$\sqrt{n}(\hat{\beta}_{KB}^{(w)}(\alpha) - \beta) \xrightarrow{D} N(0, \sigma_\alpha^2 \, Q_w^{-1} \, B_w \, Q_w^{-1})$$

with $g_\alpha(\cdot)$ defined in (1.3) and σ_α^2 defined following (2.4). Here $o_p(1)$ denotes a vector of which each component converges to zero in probability.

THEOREM 3B

Suppose that conditions (i)-(v) hold. Then, for $0 < \alpha < 0.5$. Then the estimator $\hat{\beta}_{WE}^{(w)}(\alpha)$ satisfies

$$\sqrt{n}(\hat{\beta}_{WE}^{(w)}(\alpha) - \beta) = n^{-\frac{1}{2}} Q_w^{-1} \sum_{i=1}^{n} w_i x_i \, g_\alpha(e_i) + o_p(1)$$

and therefore

$$\sqrt{n}(\hat{\beta}_{WE}^{(w)}(\alpha) - \beta) \xrightarrow{D} N(0, \sigma_\alpha^2 \, Q_w^{-1} \, B_w \, Q_w^{-1})$$

with $g_\alpha(\cdot)$ defined in (1.3) and σ_α^2 defined following (2.4).

3.2 PROOF OF THEOREM 3B

From Ruppert and Carroll (1980, Lemma A.4) it follows with

$D_{in} = n^{-\frac{1}{2}} w_i' x_i x_i'$ and $g(x) = 1$ that

$$n^{-1}(X' A_\alpha^{(w)} X) \xrightarrow{P} (1-2\alpha) Q_w$$

with $\xrightarrow{P}$ denoting convergence in probability and $A_\alpha^{(w)} = WB_\alpha$ as given in (2.7).

For $0 < u < 1$, put

$$T_{nu} = \hat{\beta}_{IN} - \beta + (\xi_u - F^{-1}(u), 0, \ldots, 0)'$$

with ξ_u given by (2.6) and define

$$J_i(T_{n\alpha}) = I(e_i < F^{-1}(\alpha) + x_i' \, T_{n\alpha})$$

$$K_i(T_{n\alpha}, T_{n(1-\alpha)}) = I(F^{-1}(\alpha) + x_i' \, T_{n\alpha} \le e_i \le F^{-1}(1-\alpha) + x_i' \, T_{n(1-\alpha)})$$

$$L_i(T_{n(1-\alpha)}) = I(e_i > F^{-1}(1-\alpha) + x_i' T_{n(1-\alpha)})$$

Then

$$|n^{-\frac{1}{2}}(X'A_\alpha^{(w)}X)(\hat{\beta}_{WE}^{(w)}(\alpha)-\beta) - (1-2\alpha) n^{-\frac{1}{2}} \sum_{i=1}^{n} w_i x_i g_\alpha(e_i)|$$

$$\leq \left| n^{-\frac{1}{2}} \sum_{i=1}^{n} w_i x_i \left[\xi_\alpha\{J_i(T_{n\alpha}) - \alpha\} - F^{-1}(\alpha)\{J_i(0) - \alpha + x_i' T_{n\alpha} f(F^{-1}(\alpha))\}\right]\right|$$

$$+ \left| n^{-\frac{1}{2}} \sum_{i=1}^{n} w_i x_i \left[e_i\{K_i(T_{n\alpha}, T_{n(1-\alpha)}) - K_i(0,0)\} + x_i'\{T_{n\alpha}F^{-1}(\alpha) f(F^{-1}(\alpha)) - T_{n(1-\alpha)}F^{-1}(1-\alpha)f(F^{-1}(1-\alpha))\}\right]\right|$$

$$+ \left| n^{-\frac{1}{2}} \sum_{i=1}^{n} w_i x_i \left[\xi_{1-\alpha}\{L_i(T_{n(1-\alpha)}) - \alpha\} + F^{-1}(1-\alpha)\{L_i(0) - \alpha + x_i' T_{n(1-\alpha)} f(F^{-1}(1-\alpha))\}\right]\right|$$

Now the first term on the right is bounded above by

$$|\xi_\alpha| \left| n^{-\frac{1}{2}} \sum_{i=1}^{n} w_i x_i [J_i(T_{n\alpha}) - J_i(0) - x_i' T_{n\alpha} f(F^{-1}(\alpha))]\right| +$$

$$|\xi_\alpha - F^{-1}(\alpha)| \left| n^{-\frac{1}{2}} \sum_{i=1}^{n} w_i x_i [J_i(0) - \alpha + x_i' T_{n\alpha} f(F^{-1}(\alpha))]\right|$$

which converges in probability to zero by Theorem 1 of Welsh (1986) and Lemma 4.1 of Bickel (1975). Similarly the third term converges in probability to zero by the argument of Theorem 1 of Ruppert and Carroll (1980). Thus the first result in Theorem 3B obtains. The asymptotic normality result follows in a straightforward way.

Remark

Typically the weights w_i depend on the design and for a particular choice $w_i = w(x_i)$ it follows from Theorems 3A and 3B that the influence function of both the bounded influence regression trimmed means is given by

$$\Lambda^{(w)}(y,x) = Q_w^{-1} w(x)x\, g_\alpha(y-x'\beta)$$

which remains bounded as long as $w(x)x$ remains bounded.

4. THE MONTE CARLO DESIGN

The purpose of the Monte Carlo study is to evaluate the small sample behaviour of the trimmed mean estimators described in Section 2 and of particular interest is the performance of these estimators in the presence of outliers and leverage points. In order to portray practice realistically we have to choose appropriate design matrices and error distributions in our Monte Carlo design. The estimators and the factors involved in the Monte Carlo study are now given.

Six estimators were considered, the least squares estimator (LS), the 10% Welsh trimmed mean based on LS-residuals (W-LS), the 10% Welsh trimmed mean based on residuals from an L_1-fit (W-L1), the 15% Koenker-Bassett trimmed mean (KB), the 15% bounded influence Koenker-Bassett trimmed mean (BI-KB) and the 10% bounded influence Welsh trimmed mean based on residuals from a weighted L_1-fit (BI-W-L1). Note that the trimming proportions of the Welsh and Koenker-Bassett estimators are not the same. This is due to the fact that the Koenker-Bassett estimators, as defined here, trim slightly fewer observations than the percentage required. For the factors considered here we found on average that the above-defined Koenker-Bassett estimators trimmed about 11% of the observations on either side. In order to obtain bounded influence estimators we use the following weight function introduced by Mallows (1973).

Denote the n observations of the (j-1)-th independent variable by $x_{1j},\ldots,x_{nj}$ for $j=2,3,\ldots,p$. Order the n observations $x_{(1)j},\ldots,x_{(n)j}$ and define $r_{1j},\ldots,r_{nj}$ as the vector of ranks of $x_{1j},\ldots,x_{nj}$. Let $L = [\tau n]+1$ and $U = n+1-L$. The weights associated with the (j-1)-th independent variable are now defined as

$$w_{ij} = \begin{cases} 1 & L \leq r_{ij} \leq U \\ (x_{(L)j} - x_{(U)j})/D_{ij} & r_{ij} < L \\ (x_{(U)j} - x_{(L)j})/D_{ij} & r_{ij} > U \end{cases}$$

where $D_{ij} = 2x_{ij} - x_{(U)j} - x_{(L)j}$ for $i=1,2,\ldots,n$. The Mallows weights are now defined as

$$w_i = \prod_{j=2}^{p} w_{ij}.$$

Note that the weights are chosen so that outliers in the independent variable space are given less weight according to their distance from the centre of the independent variable space. The basic idea in this case is to give weight to those Y_i having their corresponding x_i in the centre and decreasing weight if x_i is in the tails of the design space. We choose $\tau=0.15$, a choice we found to work well in practice. Recently Giltinan et al. (1986) introduced optimal weights which are chosen to bound the self-influence efficiently. Analogous weights for regression trimmed means may be defined in a similar way. We now give the factors involved in the study.

The sample size was taken as n=50 and the number of parameters as p=3. The distributions considered were the standard normal (N(0,1)), slash and three contaminated normal distributions, viz CN(0.1,9), CN(0.2,9) and CN(0.1,100). Here $CN(\epsilon,\sigma^2)$ denotes a standard normal contaminated with a $N(0,\sigma^2)$ with contamination probability ϵ.

Seven design matrices were considered. These design matrices, denoted by D1 to D7, were chosen to represent three cases viz that of a well-behaved design (D1), a design containing gross errors (D2,D3,D4 and D5) and a design containing leverage points but no errors (D6 and D7). Each of these designs consists of a column of ones and the remaining elements were obtained as follows

D1 : x_{ij} iid N(0,1) for i = 1,2,...,n and j = 2,3,...,p.

D2 : As D1 except one point is moved out 10 units in X space.

D3 : As D1 except two points are moved out 10 units in X space.

D4 : As D1 except one point is moved out 100 units in X space.

D5 : As D1 except two points are moved out 100 units in X space.

D6 : x_{ij} iid t_3 for i = 1,2,...,n and j = 2,3,...,p.

D7 : x_{ij} iid C for i = 1,2,...,n and j = 2,3,...,p.

Here t_3 is a t-distribution with 3 degrees of freedom and C the Cauchy distribution. Designs D1, D6 and D7 were orthogonalized so that X'X = nI. The points were moved out as follows (designs D2,D3,D4 and D5): The first (second) point to be moved out K units in design D1 was chosen as the point having largest (second largest) Euclidian distance from the centre of the design space. The resulting vector was then extended K units into X-space. Our motivation for doing this comes from experience with the simple regression case. There the most interesting behaviour of the estimators was observed when points on the edges of the design space were moved out.

Programming was done in FORTRAN VII on the WYLBUR system (IBM) of the University of North Carolina at Chapel Hill. Double precision arithmetic was used and 1000 Monte Carlo replicates were used to calculate the mean squared error of each estimator. Uniform random numbers were generated by using IMSL subroutine GGUBS, while normal random numbers were generated by IMSL subroutine GGNML. Regression quantiles were calculated using an adaptation of Barrodale and Roberts (1973) (see de Jongh (1985)).

5. MONTE CARLO RESULTS

The Monte Carlo results are presented and discussed in this section.

In Table I below we present these results in the form of efficiencies compared to that of the best estimator in a particular row, i.e.

$$\text{efficiency} = \frac{100 \text{ x Total mean square error of best estimator}}{\text{Total mean square error of estimator}}.$$

The total mean square error of an estimator (TMSE) was obtained as the sum of the mean square errors obtained for each parameter individually.

We now discuss the results obtained for design D1, then those for designs D2-D5 and lastly those for designs D6 and D7.

5.1 DESIGN D1

In this case we have a well-behaved design and the errors are distributed according to distributions having moderate to very heavy tails. This resembles the practical situation where we have no influential or leverage points, but the possibility of outliers in the response. The results are as expected and we give some remarks.

Remarks:

- The least squares estimator (LS) performs very badly for the heaviest tailed distributions. However, the Welsh estimator based on LS residuals (W-LS) performs surprisingly well, except when the errors come from the slash distribution.
- The ordinary robust estimators (W-L1 and KB) perform better than the bounded influence estimators (BI-W-L1 and BI-KB). Note that W-L1 is performing slightly better than KB for the distributions with moderate tails, but the situation is reversed for the heavier tailed distributions. We found that this is primarily due to the slightly higher percentage trimming obtained on the average by the Koenker-Bassett estimators.

TABLE I

Efficiencies of estimators compared to the best in the row.

DESIGNS	DISTRIBUTIONS	ESTIMATORS LS	W-LS	W-L1	KB	BI-KB	BI-W-L1
	N(0,1)	100	93	93	84	71	80
	CN(0.1,9)	76	100	99	93	79	86
D1	CN(0.2,9)	66	100	98	92	78	83
	CN(0.1,100)	16	95	100	100	87	86
	slash	0	0	91	100	90	81
	N(0,1)	75	70	66	58	90	100
	CN(0.1,9)	66	73	71	63	93	100
D2	CN(0.2,9)	69	82	78	73	95	100
	CN(0.1,100)	20	69	76	62	100	97
	slash	0	0	91	42	100	90
	N(0,1)	28	41	41	20	89	100
	CN(0.1,0)	32	41	41	24	94	100
D3	CN(0.2,9)	39	86	45	30	90	100
	CN(0.1,100)	18	42	43	26	100	92
	slash	0	0	68	29	100	87
	N(0,1)	33	30	30	30	89	100
	CN(0.1,9)	34	36	36	35	93	100
D4	CN(0.2,9)	27	44	44	43	95	100
	CN(0.1,100)	17	42	42	42	100	94
	slash	0	0	88	96	100	86
	N(0,1)	7	7	7	7	89	100
	CN(0.1,9)	8	8	8	8	94	100
D5	CN(0.2,9)	11	11	11	11	97	100
	CN(0.1,100)	9	11	10	10	100	84
	slash	0	0	4	4	100	72
	N(0,1)	100	93	92	82	70	78
	CN(0.1,9)	76	100	100	90	76	83
D6	CN(0.2,9)	67	100	99	91	80	83
	CN(0.1,100)	44	88	100	89	88	88
	slash	0	0	100	97	100	90
	N(0,1)	100	93	92	81	53	61
	CN(0.1,9)	81	100	98	90	60	67
D7	CN(0.2,9)	72	99	100	89	61	68
	CN(0.1,100)	19	49	100	71	77	76
	slash	0	0	100	64	99	88

5.2 DESIGNS D2, D3, D4 AND D5

Here, our data are observed exactly as in 5.1, but gross errors are introduced in design D1. This resembles the situation where we have possible outliers in the response and some gross errors in the design.

Remarks:

- The bounded influence estimators perform much better than the other estimators. The greater efficiency of these estimators is especially noticeable at designs D3 and D5.
- For the first three distributions (N(0,1), CN(0.1,9) and CN(0.2,9)) the bounded influence Welsh estimator (BI-W-L1) performs the best, while the bounded influence Koenker-Bassett estimator (BI-KB) does best for the two remaining distributions. Again this may be ascribed to the slightly higher percentage trimming on the average by the Koenker-Bassett estimator.
- The mean square error results for the individual parameters indicated that the direction of the error in X-space plays an important role in determining which of the parameters are influenced most. For example, in the above designs the first point that was moved out had largest error component with respect to the second explanatory variable and thus the estimates (using LS, W-LS, W-L1 and KB) for the second slope parameter (β_2) were severely biased. We moved points out in various directions and found the same phenomenon.

5.3 DESIGNS D6 and D7

In this case we have no errors in the design matrix but have some leverage points and outliers in the response. Influential observations may be present when a large error occurs at a leverage point.

Remarks:

- The bounded influence estimators have low efficiency compared with the ordinary robust estimators at the normal and moderately heavy tailed distributions. Note that at the normal distribution (especially for design D7) the efficiency of the bounded influence estimators is quite low. This is expected because a "good" leverage point is downweighted incorrectly. However, at the heavier tailed distributions the bounded influence estimators perform better. This may be attributed to the higher probability of obtaining an outlier at a leverage point.

- The ordinary Welsh estimator based on residuals from an L_1-fit (W-L1) seems to be the better estimator to use in this case.

We have repeated some of the above cases for n=100 and found qualitatively similar results. From the above discussion we have seen that a major advantage of the Mallows-type bounded influence trimmed means is when outliers occur both in the dependent and independent variable spaces.

6. AN EXAMPLE

We also applied the estimators of the previous section to the water salinity data (see Ruppert and Carroll (1980)). See also Carroll and Ruppert (1985). This is a data set of measurements of water salinity and river discharge taken in North Carolina's Pamlico Sound. Salinity was regressed against salinity lagged two weeks (LAGSAL), river discharge (FLOW) and a linear time trend (TREND). Estimates of the regression coefficients are listed in Table II for each estimator.

TABLE II

Results for the salinity data

ESTIMATOR	INTERCEPT	LAGSAL	TREND	FLOW	IQR	MAD	DELETED POINTS
LS	9.59	0.78	-0.03	-0.30	1.44	0.72	
W-LS	12.35	0.77	-0.09	-0.40	1.28	0.62	9,15,16,17
W-L1	13.07	0.76	-0.08	-0.43	1.25	0.56	9,15,16,17
KB	8.01	0.81	-0.01	-0.24	1.75	0.80	1,9,13,15,17
BI-KB	19.57	0.71	-0.18	-0.68	0.91	0.49	8,9,13,15,16
BI-W-L1	20.35	0.70	-0.22	-0.71	0.88	0.45	5,15,16,17

Since the true parameters are unknown, we can only measure performance by closeness of fit to the bulk of the observations. As such measures we use the median absolute value of the residuals (MAD) and the interquartile range of the residuals (IQR). These values as well as the points that are deleted by the trimmed mean estimators are also given in Table II.

Using MAD and IQR as our criteria the bounded influence estimators give the better results with BI-W-L1 having lowest IQR and MAD values. Carroll and Ruppert (1985) gave a detailed analysis of the data set under discussion. They identified points 3,5,9,15,16 and 17 as potential problems. Except for point 3, all these points show up in Table II. The absence of point 3 is expected since Carroll and Ruppert (1985) mentioned that graphical methods suggest point 3 conforms well with the bulk of the data.

From the results in the last two sections we conclude that potentially one can gain much by using bounded influence regression trimmed means over ordinary regression trimmed means. However, there does not seem to be a clear choice between the Koenker-Bassett and Welsh versions.

7 ACKNOWLEDGEMENT

The authors wish to thank Professor R.J. Carroll for comments that substantially improved the presentation of this article. The research of the first author was done while visiting the Department of Statistics at the University of North Carolina at Chapel Hill.

BIBLIOGRAPHY

BARRODALE, I. and F.D.K. ROBERTS (1973). An improved algorithm for discrete L_1 linear approximation, Siam Journal of Numerical Analysis, 10, 839-848.

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

CARROLL, R.J. and D. RUPPERT (1985). Transformations in regression: A robust analysis. Technometrics, 27(1), 1-12.

CHATTERJEE, S. and A.S. HADI (1986). Influential observations, high leverage points and outliers in linear regression, Statistical Science, 1(3), 379-416.

DE JONGH, P.J. (1985). The regression trimmed mean: An empirical study. Unpublished Ph.D. Thesis, The University of Cape Town, South Africa.

DE JONGH, P.J. and T. DE WET (1985). Trimmed mean and bounded influence estimators for the parameters of the AR(1) process. Comm. Stat. - Theor. Math. 14(6), 1361-1375.

DE JONGH, P.J. and T. DE WET (1987). Comment on paper by A.H. Welsh, The trimmed mean in the linear model, Ann. Statist., to appear.

DENBY, L. and W.A. LARSEN (1977). Robust regression estimators compared via Monte Carlo, Comm. Stat. - Theor. Math., 6(4), 335-362.

GILTINAN, D.M., CARROLL, R.J. and D. RUPPERT (1986). Some new estimation methods for weighted regression when there are possible outliers. Technometrics, 28(3), 219-230.

HAMPEL, F.R. (1968). Contributions to the theory of robust estimation, unpublished Ph.D. thesis, University of California, Berkeley, Dept. of Statistics.

HAMPEL, F.R. (1974). The influence curve and its role in robust estimation, J. Amer. Statist. Assoc., 69, 383-394.

HILL, R.W. (1982). Robust regression when there are outliers in the carriers: The univariate case, Comm. Stat. - Theor. Math., 11, 849-868.

HUBER, P.J. (1981). Robust Statistics. John Wiley and Sons, New York.

KOENKER, R.W. and G.W. BASSETT (1978). Regression quantiles, Econometrica, 46, 33-50.

KRASKER, W.S. and R.E. WELSCH (1982). Efficient bounded influence regression estimation, J. Amer. Statist. Assoc., 77, 595-604.

MALLOWS, C.L. (1973). Influence functions. Talk presented at conference on robust regression held at NBER, Cambridge Massachusetts.

MALLOWS, C.L. (1975). On some topics in robustness, unpublished memorandum, Bell Telephone Laboratories, Murray Hill New Jersey.

NARUCA, S.C. and J.F. WELLINGTON (1986). An algorithm to find all regression quantiles. J. Statist. Comput. Simul., 26, 269-281.

RUPPERT, D. and R.J. CARROLL (1980). Trimmed least squares estimates in the linear model, J. Amer. Statist. Assoc., 75, 828-838.

WELSH, A.H. (1986). Bahadur representations for robust scale estimators based on regression residuals, Ann. Statist., 14, 1246-1251.

WELSH, A.H. (1987). The trimmed mean in the linear model, Ann. Statist., to appear.