Bounded-Leverage Weights for Robust Regression Estimators

by

Dovalee Dorsett and Richard F. Gunst

Technical Report No. SMU-DS-TR-171
Southern Methodist University

February 1983

Research sponsored by NASA Contract #NCC9-9

Department of Statistics
Southern Methodist University
Dallas, TX 75275

BOUNDED-LEVERAGE WEIGHTS FOR ROBUST REGRESSION ESTIMATORS

Dovalee Dorsett and Richard F. Gunst
Department of Statistics
Southern Methodist University
Dallas, Texas 75275

ABSTRACT

Both the least squares estimator and M-estimators of regression coefficients are susceptible to distortion when high leverage points occur among the predictor variables in a multiple linear regression model. In this article a weighting scheme which enables one to bound the leverage values of a weighted matrix of predictor variables is proposed. Bounded-leverage weighting of the predictor variables followed by M-estimation of the regression coefficients is shown to be effective in protecting against distortion due to extreme predictor-variable values, extreme response values, or outlier-induced multicollinearities. Bounded-leverage estimators can also protect against distortion by small groups of high leverage points.

Key Words: Influential Observations; Leverage Points; Regression Diagnostics

1. INTRODUCTION

Least squares estimators are well-known to be sensitive to a variety of model assumption violations. One aberration of the usual model assumptions which is difficult to detect and remedy is that of extreme values among the predictor variables. Recent interest in this problem (e.g., Denby and Larson 1977, Holland and Welsch 1977, and Krasker and Welsch 1982) is stimulated by the realization that extreme predictor-variable values can distort not only least squares estimators but also supposedly robust regression estimators. In this paper we examine a relatively straightforward procedure for weighting predictor variables in order to reduce the potential distortion of least squares or robust regression estimators.

Define a multiple linear regression model as

$$Y = X\beta + \varepsilon \ , \tag{1.1}$$

where Y is an n-dimensional vector of response variables, X is an $(n \times m)$ full-column-rank matrix of nonstochastic predictor variables, and ε is an unobservable n-dimensional vector of random error terms with $\varepsilon_i \sim NID(0,\sigma^2)$. The regression coefficient vector β is usually estimated by least squares; i.e., $\hat{\beta} = (\hat{\beta}_1,\ldots,\hat{\beta}_m)$ is chosen to minimize $\Sigma\rho(r_i)$, where $\rho(r_i) = r_i^2$, $r_i = y_i - x_i'\hat{\beta}$, and $x_i' = (x_{i1},\ldots,x_{im})$ is the ith row of X (note: if a constant term is included in model (1.1), $x_{i1} = 1$ for all i). Equivalently, $\hat{\beta}$ is the solution to the

system of equations

$$\Sigma x_{ij}\psi(r_i) = 0 \quad j=1,2,\ldots,m \tag{1.2}$$

where $\psi(t) = d\rho(t)/dt$. The resulting least squares estimator can be expressed in the familiar form

$$\hat{\beta} = (X'X)^{-1}X'Y. \tag{1.3}$$

Robust regression estimators seek to reduce the influence of aberrant response values while retaining the equivalence with the least squares estimator (1.3) when no wild response values occur. This is often accomplished by selecting alternative functions $\rho(\cdot)$ and $\psi(\cdot)$ which will leave "typical" residuals unchanged but which will lessen the influence of large residuals on the solution to eqns. (1.2). Huber (1964, 1973, 1981) popularized the use of a robust M-estimator which can be defined in terms of the following $\psi(\cdot)$ function:

$$\psi(r_i) = \max\{-c, \min(r_i,c)\}, \; c > 0.$$

The value of c used with this $\psi(\cdot)$ function is usually chosen to be a multiple of a robust estimator of σ.

Regardless of whether least squares or Huber's M-estimator is used to solve eqns. (1.2), extreme predictor-variable values can distort the solution since the respective $\psi(\cdot)$ functions winsorize the residuals but do not explicitly affect the x_{ij}. For illustration purposes, suppose eqn. (1.1) defines a single-variable, no-intercept model $y_i = \beta x_i + \varepsilon_i$. Let $x_k \to \infty$ while

holding x_i $i \neq k$ fixed. The solution to eqn. (1.2) becomes

$$\psi(y_k - \hat{\beta} x_k) = 0$$

or

$$\hat{\beta} = y_k / x_k \to 0.$$

Thus both least squares and Huber's M-estimator yield slope estimates which are zero regardless of the true value of β.

The failure of many robust estimation schemes to compensate for extreme predictor-variable values occurs because these estimators are specifically intended to cope with violations of error assumptions (including outliers in the errors) but not with aberrant predictor-variables values. In the remainder of this article we focus attention on a technique for weighting predictor variables in order to reduce the influence of a few data points on the estimation of regression coefficients.

2. BOUNDED-LEVERAGE ESTIMATORS

If one rewrites eqns. (1.2) as

$$\Sigma x_{ij} \phi(r_i) r_i = 0 \quad j=1,2,\ldots,m \tag{2.1}$$

where $\phi(r_i) = \psi(r_i)/r_i$, it is apparent that M-estimators weight each residual by a quantity $\phi(r_i)$ which approaches zero as $|r_i| \to \infty$. One can rewrite eqns. (2.1) in the form of a weighted least squares estimator,

$$\hat{\beta} = (X'\Phi X)^{-1} X'\Phi Y, \tag{2.2}$$

where $\Phi = \mathrm{diag}(\phi(r_i),\ldots,\phi(r_n))$. This equation can be solved iteratively to yield M-estimators of β.

The role of the weights $\phi(\cdot)$ in controlling the influence of residuals on the estimation of β suggests that a weighting of predictor variables could be beneficial for providing protection against distortion by extreme predictors. An alternative to the solution of eqns. (2.1) for estimating β is to weight the rows of X prior to inserting the x_i' in the estimating equations. The effect on the estimating equations is that one solves

$$\Sigma w_i x_{ij} \phi(r_i^*) r_i^* = 0 \quad j=1,2,\ldots,m \tag{2.3}$$

where $r_i^* = y_i - w_i x_i' \tilde{\beta}$.

Consider now the leverage values (e.g., Hoaglin and Welsch 1978) for a weighted predictor variable matrix WX. The leverage values $h_i(W)$ are the diagonal elements of the matrix $WX(X'W^2X)^{-1}X'W$:

$$\begin{aligned} h_i(W) &= w_i^2 x_i'(X'W^2X)^{-1}x_i \\ &= w_i^2 d^2(x_i) \ . \end{aligned} \tag{2.4}$$

One criterion for selecting weights w_i for use in the estimating equations (2.1) is to choose the w_i so that $\max\{h_i(W)\} \leq \eta^2$, where η^2 is a preselected value. The weights then satisfy

$$w_i = \min\{1, \eta/d(x_i)\} \ . \tag{2.5}$$

This is a deterministic weighting scheme (conditional on X) and can be viewed as a special case of a more general stochastic weighting procedure studied by Maronna (1976); see also Maronna, Buetos, and Yohai (1979) and Krasker and Welsch (1982).

Bounded-leverage regression estimates are computed as follows. For a prescribed value of η^2 iteratively determine the predictor-variable weights:

(1) initially set $w_i = 1$ for all i;

(2) calculate $h_i(W)$ and $d^2(x_i)$ from eqns. (2.4);

(3) calculate new weights $w_i = \min\{1, \eta/d(x_i)\}$;

(4) repeat steps (2) and (3) until convergence.

A unique solution to this algorithm is proven under very mild conditions in Maronna (1976). Like Maronna (1976), we are unable to prove convergence of the algorithm; nevertheless, convergence has been rapid on all examples studied thus far. For the two examples presented in the next section a maximum of 15 iterations on a C.D.C. 6600 were required to insure that two successive calculations of $h_i(W)$ differ by less than 10^{-4}. Once the predictor-variable weights are determined, $\tilde{\beta}$ can be obtained iteratively using eqns. (2.3) and algorithms such as those in Huber (1981, Section 7.8) or Dutter (1977).

3. APPLICATIONS AND EXAMPLES

Before illustrating the use of bounded-leverage estimators on two data sets, consider again the single-variable, no-intercept model which was discussed in Section 1. As $x_k \to \infty$, $w_k \to 0$ and the solution to eqn. (2.3) with bounded-leverage weights is the value of $\tilde{\beta}$ which satisfies

$$\sum_{i \neq k} w_i x_i \psi(y_i - w_i x_i \tilde{\beta}) = 0$$

i.e., $\tilde{\beta}$ is an M-estimator based on the (n-1) observations excluding (y_k, x_k).

Another interesting application of the bounded-leverage estimator occurs when extremely large values on two or more predictor variables for the same observation produce an outlier-induced multicollinearity. For example, if model (1.1) represents a two-variable, no-intercept model the estimating equations (1.2) are

$$\Sigma x_{ij} \psi(y_i - x_{i1}\tilde{\beta}_1 - x_{i2}\tilde{\beta}_2) = 0 \quad j=1,2. \tag{3.1}$$

Now let $x_{kj} \to \infty$, j=1,2, with $x_{k1}/x_{k2} = 1$. Equations (3.1) become

$$\psi(y_k - (\tilde{\beta}_1 + \tilde{\beta}_2) x_{k1}) = 0 \ , \tag{3.2}$$

which has a solution $\tilde{\beta}_1 + \tilde{\beta}_2 = y_k/x_{k1} \to 0$. The limiting solution to eqn. (3.2) forces $\tilde{\beta}_1 \approx -\tilde{\beta}_2$ but $\tilde{\beta}_1$ and $\tilde{\beta}_2$ can have arbitrary magnitudes - the same type of ambiguous solution which is characteristic of least squares estimation with multicollinear predictor variables. The bounded-leverage estimator yields $w_k \to 0$ as $x_{kj} \to \infty$ so the estimating equations (3.1) become

$$\sum_{i \neq k} w_i x_{ij} \psi(y_i - w_i x_{i1}\tilde{\beta}_1 - w_i x_{i2}\tilde{\beta}_2) = 0 \quad j=1,2$$

i.e., the bounded-leverage estimator eliminates the multicollinearity-inducing observation from the data set.

3.1 Single-Variable Example

Mickey, Dunn and Clark (1967) examine the use of a stepwise regression procedure for detecting outliers and illustrate the performance of their procedure on the data set in Table 1. Table 2 displays diagnostic statistics which are useful in assessing the overall impact of individual data points on the fit. Contained in Table 2 are leverage values h_i (Hoaglin and Welsch, 1978), Studentized (deleted) residuals t_i (e.g., Belsley, Kuh and Welsch 1980; Cook and Weisberg 1982), and Cook's (1977) distance measure D_i. We include in Table 2 leverage values which exceed Hoaglin and Welsch's suggested cutoff of $2m/n$, Studentized residuals which exceed 1.5 in magnitude, and D_i values which are greater than Cook's recommended comparison with a lower 10% point from an $F(m,n-m)$ distribution.

[Insert Tables 1 and 2]

It is clear from Table 2(a) that observation 18 has an extreme predictor-variable value ($h_i = .652$) and that it exerts a strong influence on the estimation of β ($D_i = .678$), but its relatively small Studentized residual ($t_i = -.845$) suggests that this observation can be fit reasonably well from the other twenty observations. As indicated by its small leverage value, observation 19 does not have an extreme predictor-variable value; nevertheless,

its Studentized residual reveals that it is fit very poorly by the least squares estimates from the other twenty observations.

Table 2(b) illustrates the effect that an extreme predictor-variable value can have if the data point is not consistent with the overall trend in the data. The only change in the analysis for Tables 2(a) and 2(b) is that the x_{10} value in Table 1 was changed from 20 to 50. This makes observation 10 extreme in X and not close to the trend of the other twenty observations. All the statistics in Table 2(b) are now large for observation 10 and because of its impact on the fit neither observation 18 nor observation 19 is fit well. Table 2(c) changes both x_6 and x_{10} to 50. Note that reinforcement by observation 6 greatly reduces the t_i and D_i values for observation 10 in Table 2(c) from the values in Table 2(b). This is the problem of masking observations: two or more influential observations are similar in magnitude and "mask" the deleterious impact of one another on the fit when statistics such as t_i and D_i are used to assess their effect. Other discussions of the masking problem are contained in Andrews and Pregibon (1978), Dempster and Gasko-Green (1981), and Draper and John (1981).

The second column of Table 3 lists the Huber weights $\phi(r_i)$ for these three situations using $c = 1.345\tilde{\sigma}$ in eqn. (1.2). Note in particular that the M-estimator does not weight observation 10 in Table 3(a), weights it in Table 3(b), and weights neither

observation 6 nor 10 in Table 3(c). The masking effect of observations 6 and 10 cause the M-estimator to weight observations 2 and 18, two observations which are consistent with the overall trend in the original data. These distortions are even more apparent in the coefficient estimates displayed in Table 4.

[Insert Tables 3 and 4]

Three bounded-leverage estimators were examined on this data set. The first estimator selects weights so that $h_i(W) < 2m/n$ and utilizes the estimating equations (2.3) with $c = 1.345\tilde{\sigma}$. The second estimator calculates leverage values for the matrix

$$H^* = X^*(X^{*\prime}X^*)^{-1}X^{*\prime}$$

where X^* is X without a column of ones. The leverage values are bounded by $h_i^*(W) \leq 2(m-1)/n$ and eqns. (2.3) are again used to estimate β. The third bounded-leverage estimator uses eqns. (2.3) to estimate β; however, prior to finding weights each column of X^* is centered by subtracting the median M_j of the observations in the column. Since centering by subtracting means (instead of medians) yields an average leverage of $\overline{h^*} = (m-1)/n$, we bound the median-centered leverage values by $h_i^*(W) \leq 2\overline{h^*}+1/n = (2m-1)/n$. Weighted values of the nonconstant predictor variables are then given by $x_{ij}^* = M_j + w_i(x_{ij} - M_j)$ and are used instead of $w_i x_{ij}$ in eqns. (2.3) and in the calculation of r_i^* (note: $x_{i1} = 1$ with no weight is still used for the constant term of the model).

Tables 3 and 4 show the weights and estimated coefficients, respectively, for the three bounded-leverage estimators. Note in particular the consistency of the latter two bounded-leverage estimators over all three data sets. Both of these estimators weight only the nonconstant predictor variables and use eqns. (2.3) to estimate the regression coefficients.

As a final comparison, Tables 3 and 4 contain the weights and coefficient estimates for Krasker and Welsch's (1982) bounded-influence estimator (with $\nu = 1.596\sqrt{m}$ and $c = \nu\tilde{\sigma}$). Bounded-influence estimators minimize (over all weighted M-estimators whose weights are only a function of $|y-x'\tilde{\beta}|$) the asymptotic variance of the estimator subject to a bound on the estimator's gross-error sensitivity. This estimator evidences distortion in the fit for both the latter two data sets; in particular, it behaves similar to Huber's M-estimator in the presence of extreme predictor-variable values and it does not effectively handle the masking problem. This example supports Huber's claim (1983, Section 1 and Rejoinder) that the bounded-influence estimator is not resistant to gross errors in the predictor variables.

3.2 Air Pollution and Mortality Example

Lave and Seskin (1970, 1977, 1979) conduct an extensive investigation on the effects of air pollution on mortality. Gibbons and McDonald (1980a, 1980b, 1982) discuss many of the

methodological problems surrounding the use of regression models for predicting mortality with the Lave and Seskin data, including the sensitivity of the results to individual observations in the data base. Using Lave and Seskin's 1960 cross-sectional data set on 117 Standard Metropolitan Statistical Areas (SMSAs) we now examine the least squares and robust fits to the eleven-variable prediction equation for Total Mortality Rate (deaths per 100,000 population) studied by Gibbons and McDonald (1982).

Table 5 lists the leverage values, Studentized residuals, and D_i values for SMSAs which exceed the suggested cutoff values which were mentioned in the previous section. In addition, two diagnostic statistics which were suggested by Belsley, Kuh, and Welsch (1980) are included. $DFFITS_i$ is a scaled measure of the difference between the predicted value of y_i from the full least squares fit and its predicted value from the fit which is obtained after deleting (y_i, x_i) from the data set. A size-adjusted cutoff of $2(p/n)^{1/2}=.64$ is used to highlight influential observations. The last column of Table 5 lists the largest $DFBETAS_{ij}$ value for each observation, using an absolute cutoff of 0.35. $DFBETAS_{ij}$ is a scaled measure of the change in the estimate of β_j when (y_i, x_i) is deleted from the data base. Each of these latter two statistics approximatley measure the number of estimated standard errors change in the predicted response or coefficient estimate, respectively, attributable to the deletion of observation (y_i, x_i).

[Insert Table 5]

Of particular interest in Table 5 is the extremely large leverage value for Jersey City. This SMSA possesses an unusually large population density (PM2) and is highly influential in the prediction of its own response and in the estimation of the coefficient for PM2 (see the values for $DFFITS_i$ and $DFBETAS_{ij}$, respectively). Also of interest are the large Studentized residuals and $DFFITS_i$ values for Tampa, Scranton, and Wilkes Barre, as well as the large $DFFITS_i$ value for Charleston. All four of these SMSAs exert a strong influence on the estimation of one or more of the regression coefficients, as indicated by the $DFBETAS_{ij}$.

M-estimator weights for Huber's M-estimator, the median-centered bounded-leverage estimator, and Krasker and Welsch's bounded-influence estimator are shown in Table 6. All three estimators place small weights on the residuals for Scranton and Wilkes Barre, SMSAs which have large residuals but small leverage values in Table 5. Among the more notable differences in the weights displayed in Table 6 are those assigned to Tampa, Charleston, and Jersey City. Although all three robust estimators assign small weights to Tampa, the bounded-leverage estimator weights both the predictor variables and the residuals. Huber's M-estimator fails to weight Charleston. Neither Huber's M-estimator nor the bounded-influence estimator weight Jersey City. Recall from Table 5 that all three of these SMSAs exert a substantial effect on the estimation of one or more of the regression coefficients and on the prediction of its own mortality rate.

[Insert Table 6]

The pattern in the weights assigned by the three M-estimators to Tampa, Charleston, and Jersey City reflects the ability of each of the three estimators to adjust for leverage points. The larger the leverage value, the less able are Huber's M-estimator and the bounded-influence estimator to compensate for the resulting distortion in the fit. This is the same tendency which was observed with the Mickey-Dunn-Clark data and which was demonstrated theoretically for a single-variable, no-intercept model as $x_k \to \infty$.

Table 7 displays the various coefficient estimates for this data set. Several of the robust coefficient estimates differ markedly from the corresponding least squares estimate, notably those for SMIN, SMEAN, and PMEAN. By far the largest difference among the robust coefficient estimates occurs for PM2. Both Huber's M-estimate and the bounded-influence estimate are similar to least squares. The bounded-leverage estimate for PM2 is much larger that the other but is similar in magnitude to that which would be obtained if Jersey City was deleted from the data set.

4. SUMMARY

Huber (1981) observed the need to achieve some type of predictor-variable weighting in order to protect against extreme predictor-variable values. In Section 2 of this article we introduced a weighting scheme which enables one to bound the leverage values of the weighted predictor variables. Coupled with M-esti-

mation as in eqn. (2.3), bounded-leverage estimators afford protection against extreme predictor variables, extreme response variables, and outlier-induced multicollinearities. Based on the above examples and others we've examined, we recommend that prior to computing the predictor-variable weights the constant term be removed from the matrix of predictor variables and the remaining variates be centered by subtracting the median of each column from all the observations in the column. Then the constant term ($x_{i1} = 1$) and the adjusted predictor-variable values $x^*_{ij} = M_j + w_i(x_{ij} - M_j)$ should be used in place of the $w_i x_{ij}$ in eqn. (2.3).

REFERENCES

ANDREWS, D. F. and PREGIBON, D. (1978), "Finding the Outliers that Matter," *Journal of the Royal Statistical Society, Series B*, 40, 85-93.

BELSLEY, D. A., KUH, E. and WELSCH, R. E. (1980), *Regression Diagnostics*, New York: John Wiley and Sons, Inc.

COOK, R. D. (1977), "Detection of Influential Observations in Linear Regression," *Technometrics*, 19, 15-18.

COOK, R. D. and WEISBERG, S. (1982). *Residuals and Influence in Regression*, New York: Chapman and Hall.

DEMPSTER, A. P. and GASKO-GREEN, M. (1981). "New Tools for Residual Analysis," *The Annals of Statistics*, 9, 945-959.

DENBY, L. and LARSON, W. A. (1977), "Robust Regression Estimators Compared Via Monte Carlo," *Communications in Statistics, A6* 355-362.

DRAPER, N. R. and JOHN, J. A. (1981), "Influential Observations and Outliers in Regression," *Technometrics*, 23, 21-26.

DUTTER, R. (1977), "Numerical Solution of Robust Regression Problems: Computational Aspects, A Comparison," *Journal of Statistical Computation and Simulation*, 5, 207-238.

HOAGLIN, D. C. and WELSCH, R. E. (1978), "The Hat Matrix in Regression and ANOVA," *American Statistician*, 32, 17-22.

HOLLAND, P. W. and WELSCH, R. E. (1977), "Robust Regression Using Iteratively Reweighted Least Squares," *Communications in Statistics, A6*, 813-827.

HUBER, P. (1964), "Robust Estimation of a Location Parameter," *The Annals of Mathematical Statistics*, 35, 73-101.

———(1973), "Robust Regression: Asymptotics, Conjectures and Monte Carlo," *The Annals of Statistics*, 1, 799-821.

———(1981), *Robust Statistics*, New York: John Wiley & Sons, Inc.

———(1983), "Minimax Aspects of Bounded-Influence Regression," *Technometrics*, 78, 66-80.

KRASKER, W. S. and WELSCH, R. E. (1982), "Efficient Bounded-Influence Regression Estimation," *Journal of the American Statistical Association*, 77, 595-604.

LAVE, L. B. and SESKIN, E. P. (1970), "Air Pollution and Human Health," *Science*, 169, 723-733.

LAVE, L. B. and SESKIN, E. P. (1977), *Air Pollution and Human Health*, The Johns Hopkins University Press.

LAVE, L. B. and SESKIN, E. P. (1979), "Epidemiology, Causality and Public Policy," *American Scientist*, 67, 178-186.

Marrona, R. A. (1976). "Robust M-Estimators of Multivariate Location and Scatter," *The Annals of Statistics*, 4, 51-67.

Marrona, R. A., Bustos, O., and Yohai, V. (1979). "Bias- and Efficiency-Robustness of General M-Estimators for Regression with Random Carriers," in *Smoothing Techniques for Curve Estimation* (Lecture Notes in Mathematics, No. 757), eds. T. Glasser and M. Rosenblatt, New York: Springer-Verlag.

Mickey, M. R., Dunn, O. J., and Clark, V. (1967). "Note on the Use of Stepwise Regression in Detecting Outliers," *Computers and Biomedical Research*, 1, 105-111.

TABLE TITLES

TABLE 1. Mickey-Dunn-Clark Data

Obn.	Age(Mo.)	Gesell Score	Obsn.	Age(Mo.)	Gesell Score
1	15	95	12	9	96
2	26	71	13	10	83
3	10	83	14	11	84
4	9	91	15	11	102
5	15	102	16	10	100
6	20	87	17	12	105
7	18	93	18	42	57
8	11	100	19	17	121
9	8	104	20	11	86
10	20	94	21	10	100
11	7	113			

TABLE 2. Diagnostic Statistics for Selected Observations, Mickey-Dunn-Clark Data

Obsn.	h_i	t_i	D_i
	(a) Original Data		
2			
6			
10			
18	.652		.678
19		3.607	.223
	(b) $x_{10} = 50$		
2			
6			
10	.523	2.585	2.824
18	.327	-2.277	1.031
19		2.555	.128
	(c) $x_6 = x_{10} = 50$		
2		-1.552	
6	.349		.163
10	.349	1.498	.563
18	.220	-2.522	.698
19		2.414	.116

TABLE 3. Comparison of M-Estimator Weights, Mickey-Dunn-Clark Data

		Bounded-Leverage							
	Huber	Constant Weighted		Constant Unweighted		Median-Centered		Bounded-Influence	
Obsn.	ϕ_i	w_i	ϕ_i	w_i	ϕ_i	w_i	ϕ_i	w_i	ϕ_i
				(a) Original Data					
2		.655		.734		.479		.471	.483
6				.954		.799		.675	
10				.954		.799		.675	
18		.311		.454	.569	.232	.555	.248	.247
19	.457		.483		.512		.512	.816	.430
				(b) $x_{10} = 50$					
2		.675		.734		.479		.513	.453
6				.954		.799		.714	
10	.800	.257		.382		.184		.223	.089
18	.898	.325		.454	.569	.232	.555	.276	.194
19	.619		.497		.512		.512	.840	.465
				(c) $x_6 = x_{10} = 50$					
2	.887	.721		.734		.479		.574	.345
6		.280		.382		.184		.259	
10		.280		.382		.184		.259	.357
18	.629	.354		.454	.569	.232	.555	.319	.117
19	.637		.523		.512		.512	.870	.570

TABLE 4. Comparison of Coefficient Estimates, Mickey-Dunn-Clark Data

	Intercept	Slope
(a) Original Data		
Least Squares	109.87	-1.13
Huber	109.74	-1.17
Bounded-Leverage		
Constant Weighted	105.32	-0.78
Nonconstant Predictors Weighted	108.83	-1.19
Median-Centered	108.60	-1.19
Bounded-Influence	106.78	-0.92
(b) $x_{10} = 50$		
Least Squares	102.76	-0.58
Huber	102.84	-0.62
Bounded-Leverage		
Constant Weighted	104.33	-0.72
Nonconstant Predictors Weighted	108.83	-1.19
Median-Centered	108.60	-1.19
Bounded-Influence	104.68	-0.76
(c) $x_6 = x_{10} = 50$		
Least Squares	101.47	-0.45
Huber	100.26	-0.38
Bounded-Leverage		
Constant Weighted	103.09	-0.60
Nonconstant Predictors Weighted	108.83	-1.19
Median-Centered	108.60	-1.19
Bounded-Influence	99.47	-0.27

TABLE 5. Diagnostic Statistics for Selected Observations, Air Pollution and Mortality Data

SMSA	h_i	t_i	D_i	$DFFITS_i$	$DFBETAS^*_{ij}$
Orlando FL		1.85			
Tampa FL	.278	-4.82	.596	-2.99	-2.17 (GE65)
Macon GA	.422			1.09	.95 (PMAX)
Savannah GA		2.00			
Terre Haute IN		1.85			
New Orleans LA		1.98			.35 (PMIN)
Albuquerque NM	.252				
Jersey City NJ	.894			-1.59	-1.47 (PM2)
Canton OH	.261	-1.98		-1.18	- .49 (SMIN)
Scranton PA		3.62		1.49	.66 (SMIN)
Wilkes Barre PA		3.73		1.07	.72 (POOR)
Sioux Falls SD		-1.70			
Austin TX		-2.24		-.65	.41 (PMIN)
Charleston WV	.557	-1.69		-1.90	.64 (SMIN)
Madison WI		-1.50			

*Only the largest $DFBETAS_{ij}$ is shown for each SMSA; the corresponding predictor variable is shown in parentheses.

TABLE 6. Comparison of M-Estimator Weights, Air Pollution and Mortality Data

SMSA	Huber	Bounded-Leverage		Bounded-Influence	
	ϕ_i	w_i	ϕ_i	w_i	ϕ_i
Bridgeport CT					.79
Miami FL	.63		.70		.63
Orlando FL	.55		.57		
Tampa FL	.25	.79	.49		.13
Macon GA		.54			.58
Savannah GA	.59		.60		.73
Terre Haute IN	.69		.77		.69
New Orleans LA	.60		.61		.64
Las Vegas NV	.64		.59		.60
Jersey City NJ		.14	.33		
New York NY		.50			
Canton OH	.74				.35
Scranton PA	.33		.35		.25
Wilkes Barre PA	.33		.35		.36
Sioux Falls SD	.64		.65		.59
Austin TX	.50		.50		.57
Waco TX			.77		
Charleston WV		.42	.79		.23
Madison WI	.76				

TABLE 7. Comparison of Coefficient Estimates, Air Pollution and Mortality Data

Predictor Variable*	Least Squares	Huber	Bounded-Influence	Bounded-Leverage
SMIN	.456	.314	.294	.245
SMEAN	.079	.248	.351	.279
SMAX	.056	-.017	-.004	.036
PMIN	.253	.166	.276	.107
PMEAN	.337	.149	.032	.110
PMAX	-.025	.018	.039	-.001
PM2	.089	.088	.079	.376
GE65	6.923	7.296	7.261	7.179
PNOW	.403	.425	.408	.355
POOR	.039	.062	.102	.191
LPOP	-.281	-.196	-.215	-.373
Constant	340.267	274.468	275.884	335.126

* SMIN: Smallest Biweekly Sulfate Reading ($\mu g/m^3 x10$)

SMEAN: Arithmetic Mean of Biweekly Sulfate Readings ($\mu g/m^3 x10$)

SMAX: Largest Biweekly Sulfate Reading ($\mu g/m^3 x10$)

PMIN: Smallest Biweekly Suspended Particulate Reading ($\mu g/m^3$)

PMEAN: Arithmetic Mean of Biweekly Suspended Particulate Readings ($\mu g/m^3 x10$)

PMAX: Largest Biweekly Suspended Particulate Reading ($\mu g/m^3 x10$)

PM2: SMSA Population Density (per square mile x .1)

GE65: Percent SMSA Population at least 65 years old (x 10)

PNOW: Percent of Nonwhites in SMSA Population (x 10)

POOR: Percent of SMSA Families with Income Below Poverty Level (x 10)

LPOP: Logarithm (Base 10) of SMSA Population (x 100)