Volume Title: Evaluation of Econometric Models

Volume Author/Editor: Jan Kmenta and James B. Ramsey, eds.

Volume Publisher: Academic Press

Volume ISBN: 978-0-12-416550-2

Volume URL: http://www.nber.org/books/kmen80-1

Publication Date: 1980

Chapter Title: Regression Sensitivity Analysis and Bounded-Influence Estimation

Chapter Author: Roy E. Welsch

Chapter URL: http://www.nber.org/chapters/c11698

Chapter pages in book: (p. 153 - 167)

Regression Sensitivity Analysis and Bounded-Influence Estimation*

ROY E. WELSCH

SLOAN SCHOOL OF MANAGEMENT
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS

1. Introduction

Economists and others have been building and using econometric models for many years. A subset of these builders and users has always been concerned about model reliability, sensitivity, and validity. The energy crisis put certain aspects of modeling into the public and political spotlight. Many questions have been raised about the integrity of the modeling process, and in 1975 the National Science Foundation sponsored a conference at Vail, Colorado, on model formulation, validation, and improvement (Howrey, 1975). This conference caused a number of statisticians to pay more attention to the statistical questions raised in connection with model reliability, sensitivity, and validity. This paper briefly describes some of the progress that has been made by examining a particular model and set of data. What follows should in no way be construed as a complete analysis of the data or model.

We will denote the standard regression model by

$$y = X\beta + \varepsilon, \tag{1}$$

where X is $n \times p$. The least-squares (LS) estimates for β will be called b, the least-squares residuals, e, and the estimated standard error of the regression, s. The notation (i) will be used to indicate that the ith row or observation has been removed from a computation, and x_i will denote the ith row of the X matrix.

*Supported in part by National Science Foundation Grant 77-26902MCS to the M.I.T. Center for Computational Research in Economics and Management Science.

ISBN 0-12-416550-8

2. Data and Model

The data and model that we will discuss are taken from a recent paper by Harrison & Rubinfeld (1978). In this paper a two-step procedure is used to estimate the willingness to pay for reduced air pollution. The first step is the estimation of a hedonic housing price equation, and the second step is the estimation of a marginal willingness to pay function for households in an urban area. In what follows we will only examine in detail the first step.

The hedonic housing price model used by Harrison and Rubinfeld is

$$\begin{aligned}\text{LMV} = \beta_1 &+ \beta_2\text{CRIME} + \beta_3\text{ZONE} + \beta_4\text{INDUS} + \beta_5\text{CHAS}\\ &+ \beta_6\text{NOXSQ} + \beta_7\text{ROOM} + \beta_8\text{AGE} + \beta_9\text{DIST}\\ &+ \beta_{10}\text{HWAY} + \beta_{11}\text{TAX} + \beta_{12}\text{PTRATIO} + \beta_{13}\text{BLACK}\\ &+ \beta_{14}\text{STATUS} + \varepsilon. \end{aligned} \tag{2}$$

A brief description of each variable is given in Table 1. Further details may be found in the Harrison and Rubinfeld paper.

TABLE 1

DEFINITION OF MODEL VARIABLES

Variable	Definition
LMV	Logarithm of the median value of owner-occupied homes
CRIME	Per capita crime rate by town
ZONE	Proportion of a town's residential land zoned for lots greater than 25,000 square feet
INDUS	Proportion of nonretail business acres per town
CHAS	Charles River dummy variable with value 1 if tract bounds the Charles River
NOXSQ	Nitrogen oxide concentration (ppm) squared
ROOM	Average number of rooms squared
AGE	Proportion of owner units built prior to 1940
DIST	Logarithm of the weighted distances to five employment centers in the Boston region
HWAY	Logarithm of index of accessibility to radial highways
TAX	Full value property tax rate (per $10,000)
PTRATIO	Pupil–teacher ratio by town school district
BLACK	$(B-0.63)^2$ where B is the black proportion of the population
STATUS	Logarithm of the proportion of the population that is lower status

In order to obtain a measure of the willingness to pay for clean air W, the exponential of Eq. (2) is differentiated with respect to the pollution variable NOX. Assuming b_6 is negative this gives

$$\text{W}_i = -e^{\hat{y}_i}(2b_6\text{NOX}_i),$$

TABLE 2

Census Tracts

Observation	*Town*	*Observation*	*Town*
1	Nahant	275–279	Needham
2–3	Swampscott	280–283	Wellesley
4–6	Marblehead	284	Dover
7–13	Salem	285	Medfield
14–35	Lynn	286	Millis
36–39	Saugus	287	Norfolk
40–41	Lynnfield	288–290	Walpole
42–50	Peabody	291–293	Westwood
51–54	Danvers	294–298	Norwood
55	Middleton	299–301	Sharon
56	Topsfield	302–304	Canton
57	Hamilton	305–308	Milton
58	Wenham	309–320	Quincy
59–64	Beverly	321–328	Braintree
65	Manchester	329–331	Randolph
66–67	North Reading	332–333	Holbrook
68–70	Wilmington	334–341	Weymouth
71–74	Burlington	342	Cohasset
75–80	Woburn	343	Hull
81–84	Reading	344–345	Hingham
85–88	Wakefield	346–347	Rockland
89–92	Melrose	348	Hanover
93–95	Stoneham	349	Norwell
96–100	Winchester	350–351	Scituate
101–111	Medford	352–353	Marshfield
112–120	Malden	354	Duxbury
121–127	Everett	355–356	Pembroke
128–142	Somerville	357–488	Boston
143–172	Cambridge	357–364	Allston-Brighton
173–179	Arlington	365–370	Back Bay
180–187	Belmont	371–373	Beacon Hill
188–193	Lexington	374–375	North End
194–195	Bedford	376–382	Charlestown
196	Lincoln	383–393	East Boston
197–199	Concord	394–406	South Boston
200–201	Sudbury	407–414	Downtown (South Bay)
202–203	Wayland		
204–205	Weston	415–433	Roxbury
206–216	Waltham	434–456	Savin Hill
217–220	Watertown	457–467	Dorchester
221–238	Newton	468–473	Mattapan
239–244	Natick	474–480	Forest Hills
245–254	Framingham	481–484	West Roxbury
255–256	Ashland	484–488	Hyde Park
257	Sherborn	489–493	Chelsea
258–269	Brookline	494–501	Revere
270–274	Dedham	502–506	Winthrop

where $\hat{y}$ denotes the fitted values obtained for the model (2). It is easy to see that $\hat{y}$ and b_6 play especially important roles in the determination of W.

The basic data are from census tracts in the Boston Standard Metropolitan Statistical Area (SMSA) in 1970. With tracts containing no housing units or composed entirely of institutions excluded, the Boston sample contains 506 observations. To aid in understanding the results that follow, we provide a breakdown of the observations by town in Table 2.

In the process of developing the model (2), Harrison and Rubinfeld were careful to reduce collinearity as much as possible. To check this we computed the condition number (the square root of the ratio of the largest and smallest eigenvalues) of the scaled explanatory variable matrix and found it to be about 66, indicating that some collinearity remains [see Belsley, Kuh, & Welsch (1980)]. Harrison and Rubinfeld also discuss possible heteroscedasticity, but eventually settle on the model (2) with no heteroscedastic weighting.

TABLE 3

REGRESSION RESULTS[a]

	Estimated coefficient	Std error	t-statistic
INTER	9.756	0.15	65.23
CRIME	−0.0119	0.0012	−9.53
ZONE	0.0000804	0.0005	0.16
INDUS	0.000242	0.002	0.10
CHAS	0.0914	0.0332	2.75
NOXSQ	−63.807	11.315	−5.64
ROOM	0.00633	0.00131	4.82
AGE	0.0000897	0.0005	0.17
DIST	−0.191	0.033	−5.73
HWAY	0.0957	0.0191	5.00
TAX	−0.00042	0.00012	−3.42
PTRATIO	−0.0311	0.005	−6.21
BLACK	0.364	0.103	3.53
STATUS	−0.371	0.025	−14.83

[a] $R^2 = 0.81$, $s = 0.18$, $F(13{,}492) = 157$.

In order to summarize the data, we present some of the standard output from least-squares regression. Table 3 lists the estimated coefficients, standard errors, and t- statistics, along with F, R^2, and s.

Many econometric texts and discussions of econometric models fail to place enough emphasis on the verification of statistical assumptions. For example, Harrison and Rubinfeld do not mention or display a Gaussian

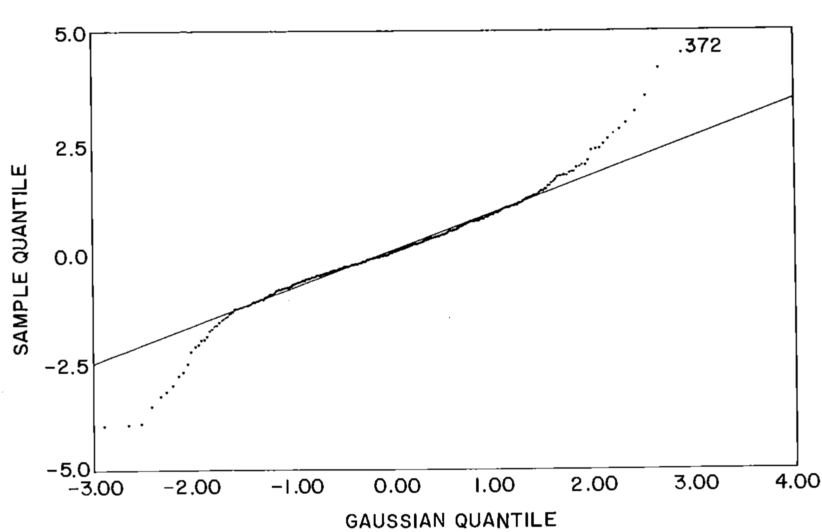

Fig. 1. Normal probability plot of studentized residuals. Resistant line is $Y = 0.85936X + 0.019074$.

(normal) probability plot of the residuals. Figure 1 is a probability plot of the studentized residuals (explained more fully in Section 4); we note the heavy tails indicating that the Gaussian error assumption may be suspect. The largest residual, 372, is denoted on the plot.

In general, we would also look at the residuals plotted against each explanatory variable. It is more informative to make partial-regression plots. This graphical device can be motivated as follows. Let $X[k]$ be the $n \times (p-1)$ matrix formed from the data matrix, X, by removing its kth column, X_k. Further let u_k and v_k, respectively, be the residuals that result from regressing y and X_k on $X[k]$. As is well known, the kth regression coefficient of a multiple regression of y on X can be determined from the simple two-variate regression of u_k on v_k. The partial-regression plot for b_k is a scatter plot of the u_k against the v_k along with their simple linear-regression line. The residuals from this regression line are, of course, just the residuals from the multiple regression of y on X, and the slope is b_k, the multiple regression estimate of β_k. Finally, the simple correlation between u_k and v_k is equal to the partial correlation between y and X_k in the multiple regression. The computational details for these plots are discussed by Mosteller & Tukey (1977).

The most interesting of these plots, the one for the crime variable, is shown in Fig. 2. Clearly, a few census tracts in Boston are rather influential in the determination of the crime coefficient.

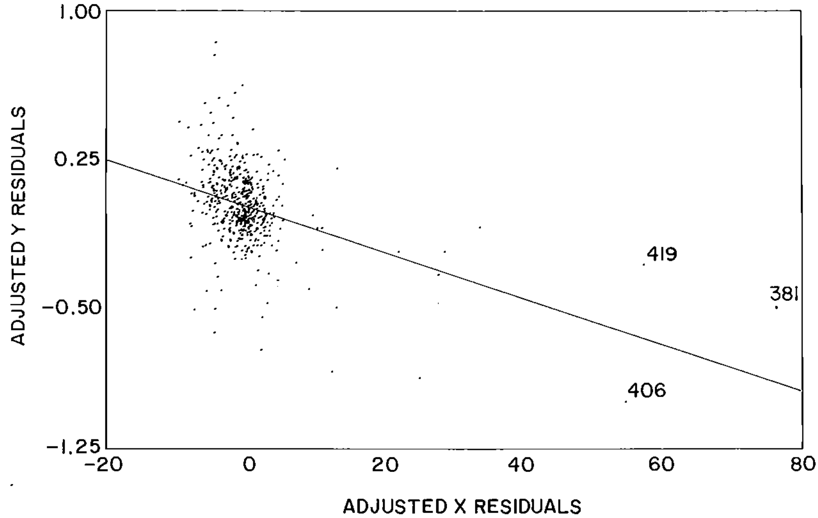

Fig. 2. Partial-regression plot for CRIME variable. Standard error = 0.001245. Regression line is $Y = -0.011866X$.

This preliminary look at the data (1 Gaussian probability plot and 14 partial-regression plots) has already caused us to question the Gaussian error assumption, to become conerned about influential observations and outliers, and to wonder if it was wise to put Boston and its suburbs together for calibrating this model. In the next section we try to assess the impact of the possible failure of the Gaussian error assumption.

3. Robust Estimation

If the error distribution for the Harrison and Rubinfeld model is not Gaussian, then we would like to use maximum likelihood estimates for the "correct" error model. Since this model is not known, a reasonable strategy is to explore models in a neighborhood of the Gaussian model to see how sensitive the estimated coefficients are to changes in the error model. For point estimation we would want a procedure that is reasonably efficient at the Gaussian model and at neighboring error models.

Huber (1973) has proposed such an estimator where the criterion function is given by

$$\rho_c(r) = \begin{cases} r^2/2, & |r| \leq c, \\ c|r| - c^2/2, & |r| > c, \end{cases} \tag{3}$$

and the parameters (including scale) are estimated by minimizing

$$\sum_{i=1}^{n} \sigma\rho[(y_i - x_i\beta)/\sigma] + d(n-p)\sigma. \tag{4}$$

Note that when $c = \infty$ this reduces to least-squares.

We chose $c = 1.345$ and $d = 0.3591$ which correspond to an estimator with 95% efficiency on Gaussian data. The results are presented in Table 4. There are a number of relatively large changes, including STATUS and ROOM (more than three LS standard errors). NOXSQ, of special interest to Harrison and Rubinfeld, changed by more than one standard error. My tentative conclusion based on these simple procedures is that the final results of the Harrison and Rubinfeld study should be stated using the LS and Huber estimation procedures in order to provide a range of values for consideration.

TABLE 4

ROBUST REGRESSION RESULTS

	LS	Huber
s	0.18	0.14
INTER	9.756	9.629
CRIME	−0.0119	−0.011
ZONE	0.0000804	0.0000368
INDUS	0.000242	0.001214
CHAS	0.0914	0.0768
NOXSQ	−63.807	−50.446
ROOM	0.00633	0.0115
AGE	0.0000897	−0.0006583
DIST	−0.191	−0.164
HWAY	0.0957	0.0704
TAX	−0.00042	−0.00036
PTRATIO	−0.0311	−0.0289
BLACK	0.364	0.551
STATUS	−0.371	−0.281

4. Regression Diagnostics

The partial-regression plots presented in Section 2 provide useful clues about influential data. There are a number of other diagnostic tools which can provide us with more precise information. The underlying philosophy is that by perturbing small portions of the data we will learn about observations that might be excessively influential in the determination of estimated coefficients, forecasts, and policy.

Two basic diagnostic quantities are the diagonal elements of the least-squares projection matrix $(X(X^TX)^{-1}X^T)$,

$$h_i = x_i(X^TX)^{-1}x_i^T, \tag{5}$$

and the studentized residuals,

$$e_i^* = e_i/\sqrt{1-h_i}\,s(i). \tag{6}$$

Both h_i and e_i^* are discussed extensively in Hoaglin & Welsch (1978) and Belsley, Kuh, & Welsch (1980). Briefly, when h_i is more than twice its average value, p/n, we say the ith observation is a *leverage* point and *may* be an influential observation. The point 381 in Fig. 2 is a leverage point, but the y data must be used before we can say it is an influential observation. The purpose of examining h_i is to detect multivariate outliers in the explanatory variable space that could not be detected via scatter plots. This measure is also an effective replacement for the tedious examination of all bivariate scatter plots. We prefer e_i^* to e_i since e_i^* has a Student's t-distribution with $n-p-1$ degrees of freedom when the Gaussian error assumptions hold. If we were to add a column to X consisting of all zeros except for a one in the ith row, then e_i^* is the t-statistic for testing the significance of the coefficient of this new δ column.

The two fundamental single-row diagnostic quantities are the scaled change in estimated coefficients (due to deleting the ith row),

$$\text{DFBETAS}_{ij} = \frac{b_j - b_j(i)}{s(i)\sqrt{(X^TX)^{-1}{}_{jj}}} = \frac{c_{ij}}{\sqrt{\sum_{k=1}^{n} c_{kj}^2}}\,\frac{e_i}{s(i)(1-h_i)}, \tag{7}$$

where c_{ij} is the appropriate element of $C^T = (X^TX)^{-1}X^T$, and the scaled change in fit,

$$\text{DFFITS}_i = \frac{x_i(b-b(i))}{s(i)\sqrt{h_i}} = \left(\frac{h_i}{1-h_i}\right)^{1/2} e_i^*. \tag{8}$$

At this point it is easy to see that neither h_i nor e_i^* alone will usually be sufficient to identify an influential observation (one with a large value of |DFBETAS| or |DFFITS|). DFFITS is essentially the product of these two quantities and if h_i is large, |DFFITS| can be large even if $|e_i^*|$ is small. The reverse is also true, and the examination of just the residuals (as is commonly done) can be misleading.

We now return to the Harrison and Rubinfeld data. The observations with $h_i \geq 2p/n = 0.055$, $|e_i^*| \geq 2.58$ $(\alpha = 0.01)$, and |DIFFITS| > 0.43 are listed in Table 5. The cutoff for DFFITS is based on the fact that when the X data is perfectly balanced, $h_i = p/n$ for all i and $\text{DFFITS}_i = (p/n - p)^{1/2}$ e_i^*. To provide some idea of the relative size of DFFITS we have plotted it against census tract in Fig. 3.

TABLE 5

REGRESSION DIAGNOSTICS

Observation	h_i	Observation	e_i^*	Observation	DEFITS
381	0.295	372	4.51	381	1.655
419	0.184	373	4.16	419	1.101
406	0.153	402	−3.99	373	0.986
411	0.112	401	−3.95	406	0.911
369	0.098	400	−3.94	369	0.879
365	0.089	490	−3.53	365	−0.859
156	0.084	413	3.52	490	−0.823
343	0.082	399	−3.30	413	0.788
366	0.077	398	−3.21	399	−0.724
163	0.077	410	3.16	215	0.721
153	0.074	506	−3.07	372	0.711
371	0.073	215	2.91	368	0.703
284	0.069	417	−2.85	401	−0.600
162	0.068	368	2.76	411	0.598
164	0.068	365	−2.75	506	−0.591
415	0.067	369	2.66	400	−0.581
143	0.067			417	−0.572
157	0.067			366	0.567
370	0.066			410	0.508
155	0.063			404	−0.477
368	0.061			491	−0.476
127	0.060			370	0.457
124	0.059			420	−0.454
215	0.058				
258	0.057				
160	0.056				

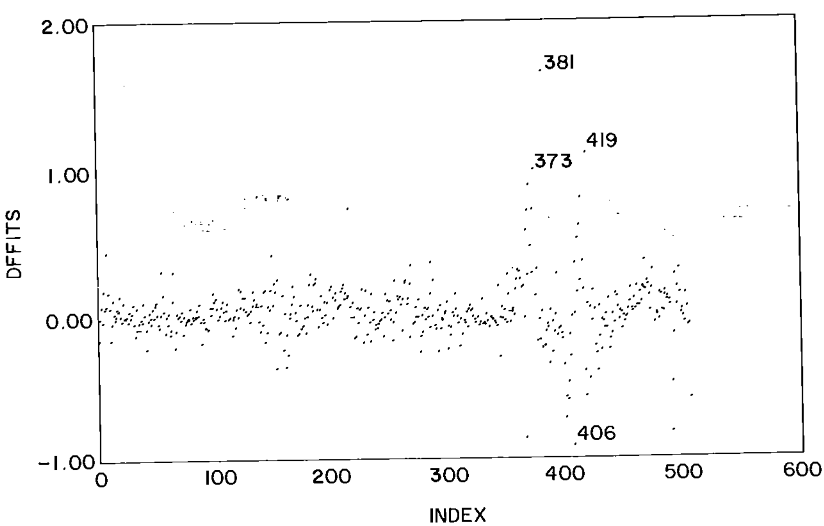

Fig. 3. Scatter plot of DFFITS versus Index.

Now we turn to the problem of deciding whether or not these points are causing special problems for the Harrison and Rubinfeld analysis. Using the rough rule that a value of $|\text{DFBETAS}| > 0.12\ (=2.58/\sqrt{n})$ may be troublesome, we found that for NOXSQ this value was exceeded in 22 cases with 0.38 (413) the largest $|\text{DFBETAS}|$. While 0.38 is well beyond the sample size adjusted cutoff of 0.12 it is rather small in the sense of statistical variability (0.38 standard deviations of the coefficient) and when compared to the values obtained for DFFITS and the DFBETAS for another variable, CRIME.

Figure 4 is a plot of DFBETAS against census tract for the CRIME variable. There are 14 $|\text{DFBETAS}|$ larger than the 0.12 cutoff and three points, 381(1.59), 419(1.00), and 406(0.87), which appear excessively influential. These same points surfaced in the DFFITS analysis and since the fit, $\hat{y}$, plays an important role in the determination of the willingness to pay, W, some attention to these points is warranted. We might wish to rerun the regression with these points deleted. This does not mean that these points should be forgotten, but they do appear to need special consideration.

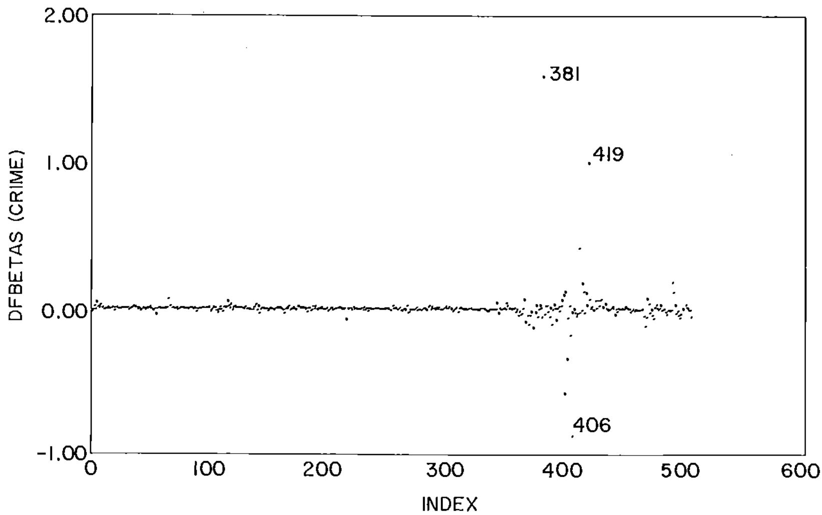

Fig. 4. Scatter plot of DFBETAS (CRIME) versus Index.

While it is tempting to use DFFITS as a summary measure, it can be wasteful if we are only interested in a few coefficients. If a point does not affect the coefficient we are interested in, then deleting it may increase the standard error of this coefficient needlessly. Influential data points may be the only points with certain kinds of information and our goal is to identify their influence and show that they need to be used wisely in model development.

5. Multiple-Row Diagnostics

If we look at each observation separately, the influence of one point may be masked by another, or the true impact and nature of a group of influential observations may not be fully diagnosed. Therefore, it is necessary to consider perturbing subsets of observations.

A variety of multiple-row methods are discussed in Belsley *et al.* (1980) Welsch & Peters (1978), and Andrews & Pregibon (1978). Many of these are quite costly for large data sets like the one under consideration. A cheaper stepwise approach has often proved to be quite effective for large data sets. Let D_m denote the index set of size m for a set of observations to be set aside. Thus $b(D_m)$ would be the least-squares estimates obtained without the use of the observations denoted by the row indices in D_m.

The stepwise procedure for each subset size, m, begins by finding a starting set, $D_m^{(0)}$, based on single row methods, for example, the indices of the m largest values of $|\text{DFFITS}|$. $D_m^{(1)}$ consists of the indices of the m largest values of

$$|x_i[b - b(D_m^{(0)})]|. \tag{9}$$

If $D_m^{(1)} = D_m^{(0)}$, stop. Otherwise form $D_m^{(2)}$ by considering the m largest values of

$$|x_i[b - b(D_m^{(1)})]|. \tag{10}$$

This process is continued until a set $D_m^{(*)}$ is found such that $D_m^{(*)} = D_m^{(*+1)}$ Generally, this procedure is performed for $m = 1, 2, 3$, etc., and this allows for a possible modification. Instead of going back to single row methods to start the process for each m, we just use the final set $D_m^{(*)}$ and find the $m + 1$ largest values of

$$|x_i[b - b(D_m^{(*)})]| \tag{11}$$

to start the process for $m + 1$.

The philosophy behind this procedure is that the largest changes in fit should occur for those points not used in the estimation of the coefficients. Different starting sets $D_m^{(0)}$ can lead to different final sets $D_m^{(*)}$. Rather than a drawback, we have found this to be an advantage and often use both of the starting procedures outlined above. After having obtained an ordered list of the values of (11) for each m, there is often a gap in each set. If there are more than m values larger than the location of the gap, we proceed to the set for $m + 1$ because more points probably should be set aside. If there are fewer than m points larger than the location of the gap, then some of the deleted fit values (11) are close to the nondeleted fit values and fewer points should be set aside. Thus, when there is a clear choice, m^* is chosen to be that value of m where the number of points beyond the gap also equals m. The set $D_{m^*}^{(*)}$ then denotes the potentially influential points that will require closer examination.

For the Harrison and Rubinfeld data both starting methods converged to 381, 419, 406, and 411, with 415 a possibility. The multiple row analysis has not revealed any masked points, except perhaps 415 if we had been looking only at DFFITS. However, 415 is in the h_i list in Table 5.

We have listed the results of deleting these five points in Table 6. There is little change in NOXSQ over the LS results, but a substantial change in the CRIME coefficient. This is not always the case and in some situations multiple-row techniques will affect a coefficient when single-row methods do not.

TABLE 6

DELETED REGRESSION RESULTS

	Deleted	LS
s	0.18	0.18
INTER	9.788	9.756
CRIME	−0.0191	−0.0119
ZONE	0.000312	0.0000804
INDUS	−0.000731	0.000242
CHAS	0.0873	0.0914
NOXSQ	−63.665	−63.807
ROOM	0.00626	0.00633
AGE	−0.0000808	0.0000897
DIST	−0.217	−0.191
HWAY	0.1106	0.0957
TAX	−0.00035	−0.00042
PTRATIO	−0.0305	−0.0311
BLACK	0.391	0.364
STATUS	−0.355	−0.371

6. Bounded-Influence Estimation

We have seen in the course of our analysis that there are some very influential observations, especially for certain coefficients. The Huber robust estimation procedure discussed in Section 3 is designed to maintain high levels of efficiency (low variance) when the error distribution is moderately heavy-tailed. Does it also insure that the perturbation of small subsets of the data will have moderate influence? If iteratively reweighted least-squares (Holland & Welsch, 1977) are used to solve (4), then the final Huber weights give five values of

$$\frac{x_i(b_w - b_w(i))}{s(i)\sqrt{h_i}} \tag{12}$$

greater than 0.43: 381(1.20), 419(0.74), 406(−0.66), 411(0.52), and 369(0.44). There has been a reduction in the number of potentially influential points as well as the size, but the influence of 381 is still pronounced. This is not a criticism of robust regression, but rather an effort to point out that another reasonable estimation strategy might be to bound the influence of small subsets of data. Our estimates will still be robust against long-tailed error, but less efficient because of the bound on the influence.

A measure of infinitesimal influence is to attach a weight, λ_i, to the ith observation and differentiate the weighted least-squares estimate $b(\lambda_i)$. This gives

$$\left.\frac{\partial b(\lambda_i)}{\partial \lambda_i}\right|_{\lambda_i=1} = (X^{\mathrm{T}}X)^{-1}x_i^{\mathrm{T}}e_i \tag{13}$$

and a corresponding infinitesimal influence for the fit of

$$x_i(X^{\mathrm{T}}X)^{-1}x_i^{\mathrm{T}}e_i = h_i e_i. \tag{14}$$

Note that (8), DFFITS, is closely related to (14).

A simple one-step bounded-influence estimate can be obtained by solving

$$\sum_{i=1}^{n} w_i x_i^{\mathrm{T}}(y_i - x_i\beta) = 0, \tag{15}$$

where

$$w_i = \begin{cases} 1 & \text{if } |\mathrm{DFFITS}_i| \le 0.34 \\ \dfrac{0.34}{|\mathrm{DFFITS}_i|} & \text{if } |\mathrm{DFFITS}_i| > 0.34 \end{cases}, \tag{16}$$

The cutoff of 0.34 is chosen for approximately 95% asymptotic efficiency. There are obviously other ways to choose w_i, and the procedure could also be iterated. Certain types of optimal weight functions have been developed by Krasker (1978) and Krasker & Welsch (1979). In effect, (16) is a way to start the iterative procedure for finding optimal weights. Since we do not have the space to describe the process fully, we will just examine the results of this starting step.

Using these weights in formula (12) gives 0.753 as the largest value of DFFITS (observation 414). We have done a better job of controlling the influence of individual observations. For example, the value of DFFITS for observation 381 is 0.13. (Iteration would allow a particular bound to be more nearly achieved.) Observation 414 appears in none of our other diagnostics. This "unmasking" is a by-product of bounded-influence estimation, which is acting on more than one observation and may therefore be considered a multiple-row technique. However, it acts in a smooth way (not $w_i = 1$ or 0) and therefore provides a different insight.

TABLE 7

BOUNDED-INFLUENCE COEFFICIENTS

	Bounded	LS	Huber
s	0.15	0.18	0.14
INTER	9.635	9.756	9.629
CRIME	−0.0141	−0.0119	−0.011
ZONE	−0.0000046	0.0000804	0.0000368
INDUS	0.000855	0.000242	0.001214
CHAS	0.083016	0.0914	0.0768
NOXSQ	−51.153	−63.807	−50.446
ROOM	0.01041	0.00633	0.0115
AGE	−0.000522	0.0000897	0.0006583
DIST	−0.167	−0.191	−0.164
HWAY	0.0763	0.0957	0.0704
TAX	−0.00033	−0.00042	−0.00036
PTRATIO	−0.0288	−0.0311	−0.0289
BLACK	0.502	0.364	0.551
STATUS	−0.299	−0.371	−0.281

Table 7 indicates that for NOXSQ the Huber robust and bounded influence approaches lead to similar estimated coefficients. (This is not unreasonable since we found no overly influential points for NOXSQ.) There is, however, quite a difference for CRIME as we might expect in view of the influential points we found. It is also of interest to look at the weights to see which observations or groups of observations were downweighted.

It is important to emphasize that bounded-influence estimation is not a substitute for a serious scrutiny of influential observations. This procedure will let the data as a whole provide a rough set of weights and an alternative fit. In the Harrison and Rubinfeld case, we should have been alerted to both long-tailed error (for NOXSQ) and influential data (for $\hat{y}$ and W). In general, we feel that alternative fits and coefficients should be given to users of a model so that they can get some idea about how sensitive the model is to the perturbation of data and of statistical assumptions.

REFERENCES

Andrews, D. F., & Pregibon, D. Finding the outliers that matter, *Journal of the Royal Statistical Society* Part C Applied Statistics, 1978, **27**, 85–93.

Belsley, D. A., Kuh, E., & Welsch, R. E. *Regression diagnostics: Identifying influential data and sources of collinearity*. New York: Wiley, 1980.

Harrison, D., Jr., & Rubinfeld, D. L. Hedonic housing prices and the demand for clean air. *Journal of Environmental Economics and Management*, 1978, **5**, 81–102.

Hoaglin, D. C., & Welsch, R. E. The hat matrix in regression and ANOVA. *The American Statistician*, 1978, **32**, 17–22 and Corrigenda, 1978, **32**, 146.

Holland, P. W., & Welsch, R. E. Robust regression using iteratively reweighted least-squares. *Communications in Statistics*, 1977, **A6**, 813–827.

Howrey, E. Philip. *Proceedings of the Vail conference on model formulation, validation and improvement*. Cambridge, Mass.: National Bureau of Economic Research, Inc., 1975.

Huber, P. J. Robust regression: Asymptotics, conjectures and Monte Carlo. *Annals of Statistics*, 1973, **1**, 799–821.

Krasker, W. S. Applications of robust estimation to econometric problems. Ph.D. Thesis, Department of Economics, Massachusetts Institute of Technology, Cambridge, Mass., 1978.

Krasker, W. S., & Welsch, R. E. Efficient bounded-influence regression estimation using alternative definitions of sensitivity. Technical Report #3, M.I.T. Center for Computational Research in Economics and Management Science, Massachusetts Institute of Technology, Cambridge, Massachusetts, 1979.

Mosteller, F. & Tukey, J. W. *Data analysis and regression*. Reading, Mass: Addison-Wesley, 1977.

Welsch, R. E., & Peters, S. C. Finding influential subsets of data in regression models. In A. R. Gallant and T. M. Gerig (Eds.), *Proceedings of the eleventh interface symposium on computer science and statistics*, Raleigh, N. C.: Institute of Statistics, North Carolina State University, 1978, Pp. 240–244.