A New Statistic for Influence in Linear Regression

Daniel PEÑA

Department of Statistics
Universidad Carlos III de Madrid
28907 Getafe, Madrid, Spain
(*daniel.pena@uc3m.es*)

Since the seminal article by Cook, the usual way to measure the influence of an observation in a statistical model is to delete the observation from the sample and compute a convenient norm of the change in the parameters or in the vector of forecasts. In this article we define a new way to measure the influence of an observation based on how this observation is being influenced by the rest of the data. More precisely, the new statistic we propose is defined as the squared norm of the vector of changes of the forecast of one observation when each of the sample points are deleted one by one. We prove that this new statistic has asymptotically a normal distribution and is able to detect a group of high leverage similar outliers that will be undetected by Cook's statistic. We show in several examples that the proposed statistic is useful for detecting heterogeneity in regression models in large high-dimensional datasets.

KEY WORDS: Cook's distance; Influential observation; Large dataset; Leverage; Masking; Outlier.

1. INTRODUCTION

The seminal article by Cook (1977) had a strong influence on the study of outliers and model diagnostics. The books of Belsley, Kuh, and Welsch (1980), Cook and Weisberg (1982), Atkinson (1985), and Chatterjee and Hadi (1988) surveyed the field with applications to linear regression and other models. The study of influential observations has been extended to other statistical models using similar ideas to the ones developed in linear regression. (See Pregibon 1981 for logistic regression models, Williams 1987 for generalized linear models, and Peña 1990 for time series models.)

Influence is usually defined by modifying the weights attached to a point or group of points in the sample and looking at the standardized change of the parameter vector or the vector of forecasts. The point can be deleted, as proposed by Cook (1977) and Belsley at al. (1980), or its weight can be decreased, as in the local influence analysis introduced by Cook (1986). The local influence approach can also be used to introduce perturbations in specific directions of interest in a sample. (See Brown and Lawrence 2000 and Suárez Rancel and González Sierra 2001 for reviews of local influence in regression and many references, and Hartless, Booth, and Littell 2003 for recent results on this approach.) A related way to analyze influence by an extension of the influence curve methodology has been proposed by Critchley, Atkinson, Lu, and Biazi (2001).

In this article we introduce a new way to analyze influence. Instead of looking at how the deletion of a point or the introduction of same perturbation affects the parameters, the forecasts, or the likelihood function, we look at how each point is influenced by the others in the sample. That is, for each sample point we measure the forecasted change when each other point in the sample is deleted. In this way we measure the sensitivity of each case to changes in the entire sample. We show that this type of influence analysis complements the usual one and is able to indicate features in the data, such as clusters of high-leverage outliers, that are very difficult to detect by the usual influence statistics. For instance, it is well known that univariate influential statistics fail when we have a group of high-leverage outliers (see, e.g., Lawrance 1995 for a detailed analysis and Rousseeuw and Leroy 1987 for several examples). We show that the proposed statistic will indicate this type of situation. This statistic complements the usual influence analysis, and, in particular, a plot of a standard influence statistic, such as Cook's distance, and the proposed sensitivity statistic can be a useful diagnostic tool in linear regression. The proposed statistic can also be used together with any modification of Cook's distance, such as those proposed by Belsley et al. (1980), Atkinson (1981), and Welsch (1982), among others. (See Cook, Peña, and Weisberg 1988 for a comparison of some of these modifications.)

The objective of this article is not to propose a procedure for unmasking outliers or robust regression. Many procedures have been developed to solve these problems. (See, e.g., Rousseeuw 1984, Atkinson 1994, and Peña and Yohai 1999 and the references therein for outlier detection based on robust estimation, and Justel and Peña 2001 for a Bayesian approach to these problems.) We do not claim that the simple statistic that we propose will always be able to avoid masking; we do not suggest that this statistic can provide the same information as some of the computationally intensive methods. Our objective here is rather to propose a new statistic, very simple to compute and with an intuitive interpretation, that can be a useful tool in applied regression analysis. In particular, it can be very effective in large datasets in high dimension, where more sophisticated procedures are difficult to apply because of their high computational requirements.

The article is organized as follows. In the next section we present the notation and define our proposed statistic. In Section 3 we analyze some of its properties. In Section 4 we illustrate the application of this statistic in four examples, and in Section 5 we discuss the relationship between the proposed statistic and other procedures. Finally, in Section 6 we comment on the generalization of the proposed statistic to other statistical models.

© 2005 American Statistical Association and
the American Society for Quality
TECHNOMETRICS, FEBRUARY 2005, VOL. 47, NO. 1
DOI 10.1198/004017004000000662

2. A NEW STATISTIC FOR DIAGNOSTIC ANALYSIS IN REGRESSION

Consider the regression model with intercept

$$y_i = \mathbf{x}_i'\boldsymbol{\beta} + u_i,$$

where the u_i are independent random variables that follow a normal distribution with mean 0 and variance σ^2 and the $\mathbf{x}_i = (1, x_{2i}, \dots, x_{pi})$'s are numerical vectors in $\mathbb{R}^p$. We denote by $\mathbf{X}$ the $n \times p$ matrix of rank p whose ith row is $\mathbf{x}_i'$; by $\hat{\boldsymbol{\beta}}$ the least squares estimate (LSE), given by $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$; by $\hat{\mathbf{y}} = (\hat{y}_1, \dots, \hat{y}_n)'$ the vector of fitted values given by $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{H}\mathbf{y}$, where $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is the hat matrix; and by $\mathbf{e} = (e_1, \dots, e_n)'$ the vector of least squares residuals given by

$$\mathbf{e} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} = (\mathbf{I} - \mathbf{H})\mathbf{y}. \tag{1}$$

The study of influential observations is standard practice in statistical models. The general idea of influence analysis is to introduce small perturbations in the sample and see how these perturbations affect the fitted model. The most common approach is to delete one data point and see how this deletion affects the vector of parameters or the vector of forecasts. Of course, other types of perturbations are possible (see, e.g., Cook 1986). Let us call $\hat{\boldsymbol{\beta}}_{(i)}$ the LSE when the ith data point is deleted, and let $\hat{\mathbf{y}}_{(i)} = \mathbf{X}\hat{\boldsymbol{\beta}}_{(i)}$ be the corresponding vector of forecasts. Cook (1977) proposed measuring the influence of a point by the squared norm of the vector of forecast changes given by

$$D_i = \frac{1}{ps^2}\left\|\hat{\mathbf{y}} - \hat{\mathbf{y}}_{(i)}\right\|^2,$$

where $s^2 = \mathbf{e}'\mathbf{e}/(n-p)$. This statistic can also be written as

$$D_i = \frac{r_i^2 h_{ii}}{p(1-h_{ii})}, \tag{2}$$

where h_{ij} is the ijth element of $\mathbf{H}$ and

$$r_i^2 = \frac{e_i^2}{s^2(1-h_{ii})}$$

is the internally Studentized residual. The expected value of D_i can be approximated for large n by

$$E(D_i) \simeq \frac{h_{ii}}{p(1-h_{ii})}, \tag{3}$$

and it will be very different for observations with different leverages.

Instead of looking at the global effect on the vector of forecasts from the deletion of one observation, an alternative approach is to measure how the deletion of each sample point affects the forecast of a specific observation. In this way we measure how each sample point is being influenced by the rest of the data. In the regression model, this can be done by considering the vectors

$$\mathbf{s}_i = \left(\hat{y}_i - \hat{y}_{i(1)}, \dots, \hat{y}_i - \hat{y}_{i(n)}\right)'; \tag{4}$$

that is, we look at how sensitive the forecast of the ith observation is to the deletion of each observation in the sample. We define the new statistic at the ith observation, S_i, as the squared norm of the standardized vector $\mathbf{s}_i$, that is,

$$S_i = \frac{\mathbf{s}_i'\mathbf{s}_i}{p\widehat{\operatorname{var}}(\hat{y}_i)}, \tag{5}$$

and using the fact that

$$\hat{y}_i - \hat{y}_{i(j)} = \frac{h_{ji}e_j}{1-h_{jj}}$$

and $\widehat{\operatorname{var}}(\hat{y}_i) = s^2 h_{ii}$, this statistic can be written as

$$S_i = \frac{1}{ps^2 h_{ii}} \sum_{j=1}^{n} \frac{h_{ji}^2 e_j^2}{(1-h_{jj})^2}. \tag{6}$$

An alternative way to write S_i, is as a linear combination of the sample Cook's distances. From (2) and (6), we have

$$S_i = \sum_{j=1}^{n} \rho_{ji}^2 D_j, \tag{7}$$

where $\rho_{ij} = (h_{ij}^2/h_{ii}h_{jj})^{1/2} \le 1$ is the correlation between forecasts $\hat{y}_i$ and $\hat{y}_j$. Also, using the predictive residuals, $e_{j(j)} = y_j - \hat{\boldsymbol{\beta}}_{(j)}x_j = e_j/(1-h_{jj})$, we have that

$$S_i = \frac{1}{ps^2} \sum_{j=1}^{n} w_{ji} e_{j(j)}^2; \tag{8}$$

that is, S_i is a weighted combination of the predictive residuals.

3. PROPERTIES OF THE NEW STATISTIC

In this section we present three properties of the statistic S_i. The first property is that under the hypothesis of no outliers and when all of the h_{ii}'s are small, the expected value of the statistic is approximately equal to $1/p$. In other words, in a sample without outliers or high-leverage observations, all of the cases have the same expected sensitivity with respect to the entire sample. This is an important advantage over Cook's statistic, which has an expected value that depends heavily on the leverage of the case. The second property is that for large sample sizes with many predictors, the distribution of the S_i statistic will be approximately normal. This again is an important difference from Cook's distance, which has a complicated asymptotical distribution (see Muller and Mock 1997). This normal distribution allows one to compute cutoff values for finding outliers. Third, we prove that when the sample is contaminated by a group of similar outliers with high leverage, the sensitivity statistic will discriminate between the outliers and the good points.

Let us derive the first property. From (6),

$$E(S_i) = \frac{1}{ph_{ii}} \sum_{j=1}^{n} \frac{h_{ji}^2}{(1-h_{jj})} E(r_i^2), \tag{9}$$

and because $r_j^2/(n-p)$ is a beta variable with parameters $1/2$ and $(n-p-1)/2$ (see, e.g., Cook and Weisberg 1982, p. 19), $E(r_j^2) = 1$, and calling $h^* = \max_{1\le i\le n} h_{ii}$, we have that

$$E(S_i) = \frac{1}{ph_{ii}} \sum_{j=1}^{n} \frac{h_{ji}^2}{(1-h_{jj})} \le \frac{1}{p(1-h^*)} = \frac{1}{p} + \frac{h^*}{p(1-h^*)}.$$

In contrast, as $h_{jj} \geq n^{-1}$, we have

$$E(S_i) = \frac{1}{ph_{ii}} \sum_{j=1}^{n} \frac{h_{ji}^2}{(1-h_{jj})} \geq \frac{1}{p(1-n^{-1})}.$$

These results indicate that if h^* is small, then the expected influence of all of the sample points is approximately $1/p$. Thus we can look for discordant observations by analyzing those points that have values of the new statistic far from this value. It may seem that the hypothesis that all of the h_{ii}'s are small is very restrictive. However, when the sample contains a set of k similar high-leverage outliers, it can be proved (see Peña and Yohai 1995) that the maximum leverage of the outliers is $1/k$.

For the second property, we assume no outliers and that $h^* = \max_{1 \leq i \leq n} h_{ii} < c\overline{h}$, for some $c > 0$, where $\overline{h} = \sum_{i=1}^{n} h_{ii}/n$. Then, letting $n \to \infty$ and $p \to \infty$ but $p/n \to 0$, we show that the asymptotic distribution of S_i will be normal. This result comes from (6), writing

$$S_i = \sum_{j=1}^{n} w_{ij} \left(\frac{e_j^2}{s^2} \right),$$

where

$$w_{ij} = \frac{h_{ji}^2}{ph_{ii}(1-h_{jj})^2}$$

and, from (1), the residuals e_j are normal variables with covariance matrix $\sigma^2(\mathbf{I} - \mathbf{H})$. Thus when $n \to \infty$, $h_{ij} \to 0$, and the statistic S_i is a weighted combination of chi-squared independent variables with 1 degree of freedom. The coefficients w_{ij} are positive, and we now show that $w_{ij}/\sum w_{ij} \to 0$. Because

$$w_{ij} \leq \frac{h_{jj}}{p(1-h_{jj})^2} \simeq \frac{1}{p} h_{jj}(1+2h_{jj}),$$

we have

$$\frac{w_{ij}}{\sum w_{ij}} \leq \frac{h_{jj}(1+2h_{jj})}{p + 2\sum h_{jj}^2} \leq \frac{h_{jj}(1+2h_{jj})}{p},$$

and as $p \to \infty$, the relative weight of each chi-squared variable will go to 0. Thus the distribution of the statistic under these hypotheses will be asymptotically normal.

An implication of this property is that we can search for outliers by finding observations with values of the S_i statistic larger than $(S_i - E(S_i))/\mathrm{std}(S_i)$. Because the possible presence of outliers and high leverage points will affect the distribution of S_i, we propose using high-breakdown estimates for the parameters of the distribution. Using the median and MAD (median of the absolute deviations from the sample median), we propose considering as heterogeneous observations those that satisfy

$$|S_i - \mathrm{med}(S)| \geq 4.5MAD(S_i), \tag{10}$$

where $\mathrm{med}(S)$ is the median of the S_i values and $MAD(S_i) = \mathrm{median}\,|S_i - \mathrm{med}(S)|$. For normal data, $MAD(S_i)/.645$ is a robust estimate for the standard deviation, and the previous rule is roughly equivalent to taking three standard deviations in the normal case.

As an example, Figures 1(a) and 1(b) show the histograms of Cook's distance and S_i for a sample of 1,000 observations generated under the normal model with 20 predictors. It can be seen that whereas the distribution of Cook's distance is very skewed, the distribution of S_i is approximately normal. Figure 1(c) shows a plot of both statistics, which we call the C/S plot, and Figure 1(d) shows the individual S_i values and the cutoff limits defined in (10). It can be seen that these limits seem appropriate for noncontaminated data.

The third property that we prove is that when the data contain a group of high-leverage identical outliers, the sensitivity statistics will identify them. We show that the new statistic S_i is expected to be smaller for the outliers than for the good data points. We also show that Cook's statistic is unable to discriminate in this case. Suppose that we have a sample of n points $(y_1, \mathbf{x}_1'), \ldots, (y_n, \mathbf{x}_n')$ and let $\mathbf{X}_0' = [\mathbf{x}_1, \ldots, \mathbf{x}_n]$, $\mathbf{y}_0' = [y_1, \ldots, y_n]$, $\hat{\boldsymbol{\beta}}_0 = (\mathbf{X}_0'\mathbf{X}_0)^{-1}\mathbf{X}_0'\mathbf{y}_0$, and $u_i = y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}}_0$. Suppose now that this sample is contaminated by a group of k identical high-leverage outliers $(y_a, \mathbf{x}_a')$, and let $u_a = y_a - \mathbf{x}_a'\hat{\boldsymbol{\beta}}_0$ be the residual with respect to the true LSE and let $e_i = y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}}_T$ be the residual in the total regression with $n + k$ observations, where $\hat{\boldsymbol{\beta}}_T = (\mathbf{X}_T'\mathbf{X}_T)^{-1}\mathbf{X}_T'\mathbf{y}_T$ and $\mathbf{X}_T' = [\mathbf{X}_0'\mathbf{x}_a\mathbf{1}_k']$, where $\mathbf{1}_k$ is a vector of 1's of dimension $k \times 1$ and $\mathbf{y}_T' = [\mathbf{y}_0, y_a\mathbf{1}_k']$. Let $\mathbf{H} = \mathbf{X}_T(\mathbf{X}_T'\mathbf{X}_T)^{-1}\mathbf{X}_T'$ be the projection matrix with elements h_{ij} for the sample of $n + k$ data, and let $\mathbf{H}^0 = \mathbf{X}_0(\mathbf{X}_0'\mathbf{X}_0)^{-1}\mathbf{X}_0'$ be the projection matrix for the good data with elements h_{ij}^0. We partition the matrix $\mathbf{H}$ as

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{H}_{21} & \mathbf{H}_{22} \end{bmatrix},$$

where $\mathbf{H}_{11}$ has dimension $n \times n$ and $\mathbf{H}_{22}$ has dimension $k \times k$. We show in the Appendix that

$$\mathbf{H}_{11} = \mathbf{H}^0 - \frac{k}{kh_a^0 + 1}\mathbf{h}_{1a}^0(\mathbf{h}_{1a}^0)', \tag{11}$$

where $h_a^0 = \mathbf{x}_a'(\mathbf{X}_0'\mathbf{X}_0)^{-1}\mathbf{x}_a$ and $\mathbf{h}_{1a}^0 = \mathbf{X}_0(\mathbf{X}_0'\mathbf{X}_0)^{-1}\mathbf{x}_a$. Also,

$$\mathbf{H}_{12} = \frac{1}{kh_a^0 + 1}\mathbf{h}_{1a}^0\mathbf{1}_k{}' \tag{12}$$

and

$$\mathbf{H}_{22} = \frac{h_a^0}{kh_a^0 + 1}\mathbf{1}_k\mathbf{1}_k{}'. \tag{13}$$

The observed residuals, $e_i = y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}}_T$, are related to the true residuals, $u_i = y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}}_0$, by

$$e_i = u_i - kh_{ia}u_a, \qquad i = 1, \ldots, n, \tag{14}$$

and to the outlier points by

$$e_a = \frac{u_a}{1 + kh_a}. \tag{15}$$

Using (14), Cook's statistic for the good points is given by

$$D_i = \frac{(u_i - kh_{ia}u_a)^2 h_{ii}}{ps^2(1-h_{ii})^2},$$

where $s^2 = \sum e_i^2/(n + k - p)$. For the outlier points using (13), this statistic can be written as

$$D_a = \frac{u_a{}^2 h_a}{ps^2(1 + (k-1)h_a)^2(1 + kh_a)}. \tag{16}$$

Suppose now that we have high-leverage outliers, and let $h_a^0 \to \infty$. Then $\mathbf{H}_{12} \to \mathbf{0}$, which implies that $h_{ja} \to 0$ for

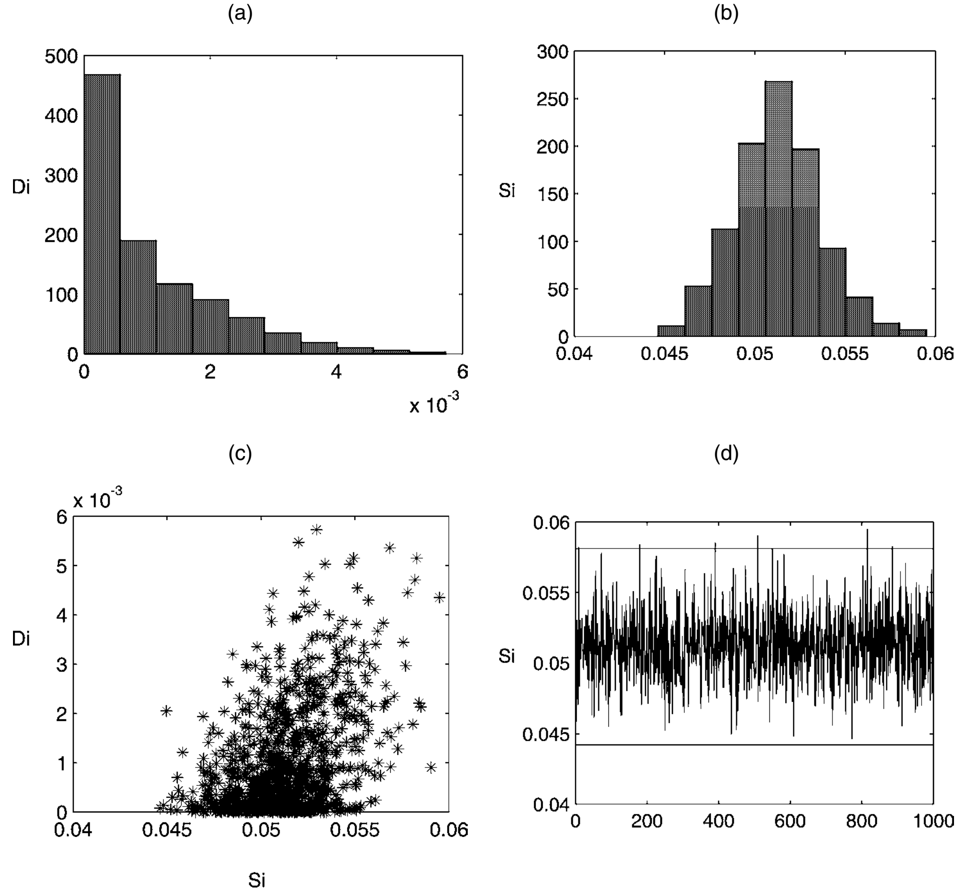

Figure 1. Influence Analysis of a Sample of 1,000 Observations With 20 Regressors and Linear Regression. (a) Histogram of Cook's distances. (b) Histogram of S_i. (c) C/S plot. (d) Plot of S_i versus case number.

$j = 1, \dots, n$, and $\mathbf{H}_{22} \to \frac{1}{k}\mathbf{1}_k\mathbf{1}_k'$, which implies that $h_a \to k^{-1}$, and

$$\rho_{ja}^2 = \frac{h_{ja}^2}{h_{jj}h_a}$$

will go to 0 for $j = 1, \dots, n$, and to 1 for $j = n+1, \dots, n+k$. Thus for the good observations, we have, from (7),

$$S_i = \sum_{j=1}^{n} \rho_{ji}^2 D_j, \qquad i = 1, \dots, n, \tag{17}$$

whereas for the outliers,

$$S_i = kD_a, \qquad i = n+1, \dots, n+k. \tag{18}$$

For the good points when $h_a^0 \to \infty$, $h_{ja} \to 0$ and, by (14), $e_i \to u_i$. Using the same argument as for computing $E(S_i)$, it is easy to show that at the good points, the expected value of S_i will be $1/p$. However, for the outliers from (15), when $h_a^0 \to \infty$, $e_a \to 0$, and $D_a \to 0$ and also $S_i \to 0$. Thus for high-leverage outliers, the new statistics will be close to 0 for the outliers and close to $1/p$ for the good observations. A similar result is obtained if we let $u_a \to \infty$ and $\|\mathbf{x}_a\| \to \infty$ but $u_a/\|\mathbf{x}_a\| \to c$.

The foregoing results indicate that this statistic can be very useful for identifying high-leverage outliers, which are usually considered the most difficult type of heterogeneity to detect in regression problems. Also, this statistic can be useful for identifying intermediate-leverage outliers that are not detected by Cook's distance. Suppose that we have a group of outliers with $h_a^0 \ge \max_{1 \le i \le n} h_{ii}$; that is, they have true leverage larger than the good points, but the true residual size, u_a, is such that the observed least squares residuals, e_a, given by (15), are not close to 0. Then the cross-leverage h_{ia} between the good points and the outliers for (18) will still be small, and thus ρ_{ia}^2 also will be small. Therefore, the new statistic for the outlier points will accumulate Cook's distances for all of them, and the value of the S_i statistic will be larger for the outliers than the value for the good points.

This statistic will not be useful in situations in which the outliers have low leverage. Suppose that case i in the sample corresponds to a single outlier due to some measurement error, so that $y_i = true(y_i) + c$. Suppose that the leverage at this point is small, that is, $h_{ii} < p/n$. Then if c is large, the point will appear as a clear outlier, due to its large residual, and also as influential, leading to a large value of D_i. However, because D_i will enter in the computation of the sensitivity statistic for all the observations, the value of S_i will not be very different from the others. But if the leverage of the point is large (close to 1), then, because the correlations ρ_{ij}^2 for $j = 1, \dots, n$, $j \neq i$ will be small, case i will have large values of both D_i and S_i and will be separated from the good points. This result generalizes in the same way for groups; with low-leverage outliers, the values of the statistic S_i at the outliers will not be much larger than for the rest of observations, whereas for intermediate outliers, it will larger. The group of low-leverage outliers will increase the variability

Table 1. Four Sets of Data With the Same Values of Observations 1–27 and in 28–30: S_a, No Outliers; S_b, Three High-Leverage Outliers; S_c, Three Intermediate-Leverage Outliers; S_d, Three Low-Leverage Outliers

	x	y	D_a	S_a	S_b	S_c	S_d
1	.3899	.0000	.0009	.4552	.5551	.5477	.8216
2	.0880	−.3179	.0069	.4893	.5601	.5352	.7536
3	−.6355	1.0950	.0364	.5327	.5628	.5318	.2308
4	−.5596	−1.8740	.1358	.5352	.5631	.5306	.2745
5	.4437	.4282	.0046	.4504	.5540	.5505	.8229
6	−.9499	.8956	.0345	.5139	.5603	.5391	.1311
7	.7812	.7310	.0293	.4291	.5449	.5702	.7961
8	.5690	.5779	.0123	.4408	.5510	.5575	.8184
9	−.8217	.0403	.0002	.5225	.5616	.5357	.1577
10	−.2656	.6771	.0095	.5286	.5630	.5290	.5059
11	−1.1878	.5689	.0171	.4984	.5571	.5460	.1152
12	−2.2023	−.2556	.0305	.4582	.5342	.5751	.1811
13	.9863	−.3775	.0309	.4216	.5379	.5827	.7678
14	−.5186	−.2959	.0055	.5360	.5632	.5300	.3014
15	.3274	−1.4751	.1179	.4613	.5563	.5447	.8169
16	.2341	−.2340	.0054	.4714	.5580	.5405	.8017
17	.0215	.1184	.0000	.4977	.5609	.5333	.7203
18	−1.0039	.3148	.0028	.5103	.5597	.5406	.1244
19	−.9471	1.4435	.0977	.5141	.5603	.5390	.1316
20	−.3744	−.3510	.0066	.5345	.5633	.5289	.4125
21	−1.1859	.6232	.0212	.4985	.5572	.5460	.1152
22	−1.0559	.7990	.0310	.5068	.5590	.5421	.1199
23	1.4725	.9409	.1325	.4129	.5168	.6089	.7027
24	.0557	−.9921	.0426	.4934	.5605	.5342	.7384
25	−1.2173	.2120	.0012	.4966	.5567	.5469	.1152
26	−.0412	.2379	.0004	.5056	.5615	.5318	.6823
27	−1.1283	−1.0078	.0834	.5021	.5580	.5442	.1163
28	($a = 1.02$) ($b, c, d = 20, 5, .5$)	($a = .72$) ($b, c, d = 5.0$)	.0384	.4207	.0160	.6567	.8220
29	($a = .75$) ($b, c, d = 20, 5, .5$)	($a = .\ 42$) ($b, c, d = 5.0$)	.0063	.4305	.0160	.6567	.8220
30	($a = -.44$) ($b, c, d = 20, 5, .5$)	($a = -.21$) ($b, c, d = 5.0$)	.0033	.5360	.0160	.6567	.8220

of the S_i values, but it will not separate the outliers from the good observations.

We illustrate the performance of the S_i statistic in the following way. A simulated sample of size 30 of two independent N(0, 1) random variables, x and y, is generated, and this is termed situation (a). Then three other datasets are built by modifying the three last cases of this sample by introducing three outliers of size $y = 5$ but with different leverages. Situation (b) corresponds to high leverage ($x = 20$), (c) corresponds to intermediate leverage ($x = 5$), and (d) corresponds to low leverage ($x = .5$). The data are given in Table 1. The four values assigned to cases 28, 29, and 30 for the different situations are given in parentheses for both x (four different leverage values considered) and y [the same outlier size, 5, for situations (b), (c), and (d)]. The table also provides the values of the S_i statistic in the four situations and the value of Cook's distance for the uncontaminated sample, situation (a). The values of Cook's distance, D_i, and the S_i statistic, are also represented in the four situations in the top row of Figure 2. The bottom row of the figure presents plots of S_i versus case number, including the reference values, $\mathrm{med}(S_i) + 4.5MAD(S_i)$ and $\max(0, \mathrm{med}(S_i) - 4.5MAD(S_i))$.

In (a), all of the values of the S_i statistic are close to its mean value of $1/2$. In (b), the three outliers have very high leverage, and therefore their residuals are close to 0. Then, as expected by the third property, the values of the S_i statistic for the outliers are close to 0, whereas for the good points they are close to the mean value, .5. In case (c), S_i is larger for the outliers than for the good points, and both groups are again well separated. Finally, in (d), the S_i statistic has a very large variance and is not informative, whereas Cook's distance takes larger values at the outliers than at the good points. This last situation is the most favorable one for Cook's distance, and a sequential deletion strategy using this distance will lead to the identification of the three outliers.

We can summarize the analysis as follows. In a good sample without outliers or high leverage points, the sensitivity of all the points, as measured by the statistic S_i, will be the same, $1/p$. If we have a group of high-leverage outliers, then the forecasts of these points will not change by deleting any point in the sample, and therefore, the sensitivity of these points will be very small. For a group of intermediate-leverage outliers, this means that the residuals at the outliers are not close to 0, the effect on the forecasts of the outliers after deleting an outlier point will be large, and therefore the sensitivity of the outliers will be larger than for the good points. Finally, if we have low-leverage outliers, then the forecasts of all of the points will be affected by deleting them, and the global effect is to increase the sensitivity of all the points in the sample. One could think of using the variance of S_i to identify the last situation, but this would not be very useful, because these low-leverage outliers are easily identified because of their large residuals.

4. EXAMPLES

We illustrate the performance of the proposed statistic with four examples. We have chosen two simple regression and two multiple regression examples and two real data examples and two simulated examples, so that we know the solution. Three of the four examples that we present have been extensively analyzed in the robust regression and diagnostic literature. In the

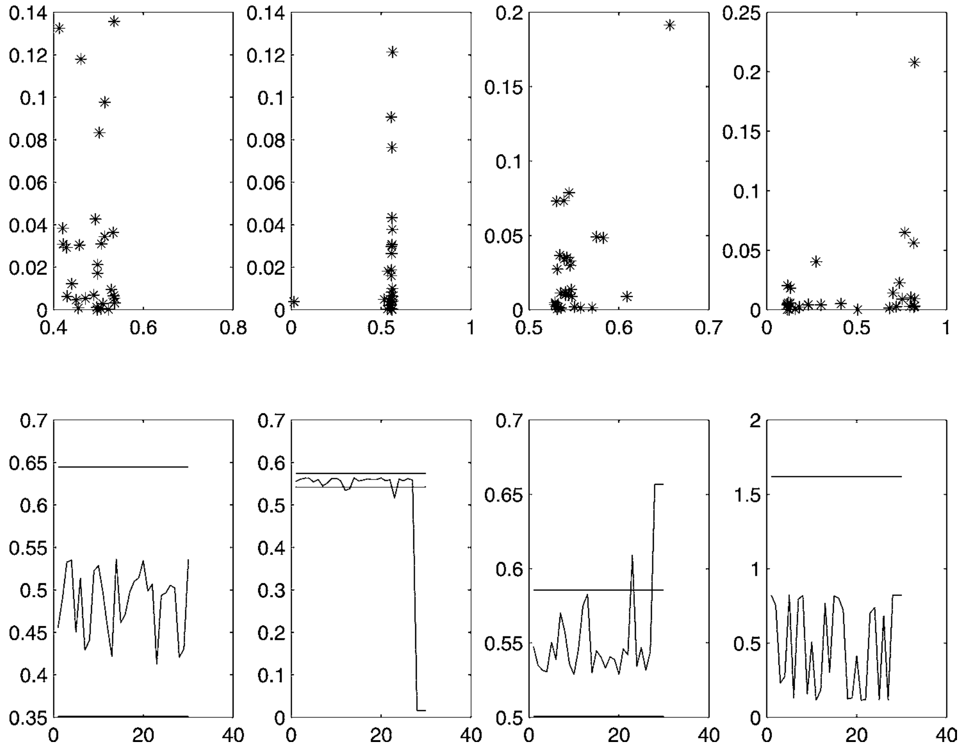

Figure 2. Plots of Cook's Distance versus the Proposed Statistic (top row) and Plots of S_i versus Case Four Situations (bottom row): (a) No Outliers, (b) Three High-Leverage Outliers; (c) Three Intermediate-Leverage Outliers; (d) Three Low-Leverage Outliers.

first example we present a situation in which we have a group of moderate outliers, and we illustrate it with the well-known Hertzsprung–Rusell diagram (HRD) data from Rousseeuw and Leroy (1987). In the second example we present a strong masking case using a simulated dataset proposed by Rousseeuw (1984). The third example illustrates the usefulness of the statistic in the well-known Boston Housing data from Belsley et al. (1980), which has been analyzed by many authors to compare robust and diagnostic methods. Finally, the fourth example is a simulated dataset with 2,000 cases and 20 variables and is presented to illustrate the advantages of the proposed statistic in routine analysis of high-dimensional datasets.

Example 1. Figure 3 shows the data for the HRD dataset. This data corresponds to the star cluster CYG OB1, which consists of 47 stars in the direction of Cygnus. The variable x is the logarithm of the effective temperature at the surface of the star, and y is the logarithm of its light intensity. These data were given by Rousseeuw and Leroy (1987) and have been analyzed by many authors as an interesting masking problem. In Figure 3 we observe that four data points (11, 20, 30, and 34) are clearly outliers and another two observations (7 and 14) seem to be far away from the main regression line. As we have $p = 2$, the approximate expected value for S_i is .5. Figure 4 shows an influence analysis of this dataset. The histogram of the Cook's distances will not indicate any observation as influential because of the masking effect. The histogram of the sensitivity statistic is more informative, because it shows a group of six observations separated from the others. The plot of Cook's distance versus S_i clearly indicates the six outliers. The good points have a value of S_i around .52, close to the expected value, whereas the six outliers have a value of S_i close to 1. The plot separates the two groups of data clearly. Finally, the comparison of the values of S_i with the cutoff defined in the previous section indicates that these six observations are outliers.

Example 2. Figure 5 shows a plot of the two groups' regression lines generated by Rousseeuw (1984). These data again have been analyzed by many authors and recently by Critchely et al. (2001), who presented them as a very challenging and difficult dataset. Figure 6 shows the influence analysis. Cook's distance does not show any indication of heterogeneity. The histogram of S_i clearly shows two groups of data. The larger group of 30 points has all of the values in the interval [.53, .59],

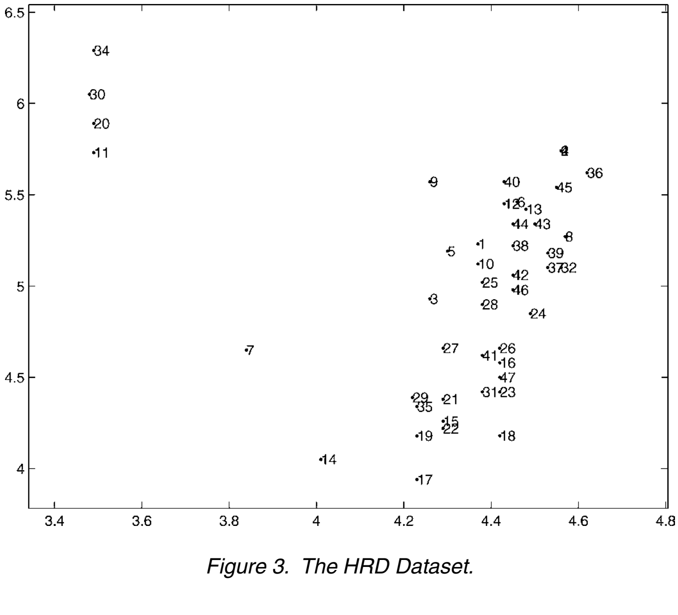

Figure 3. The HRD Dataset.

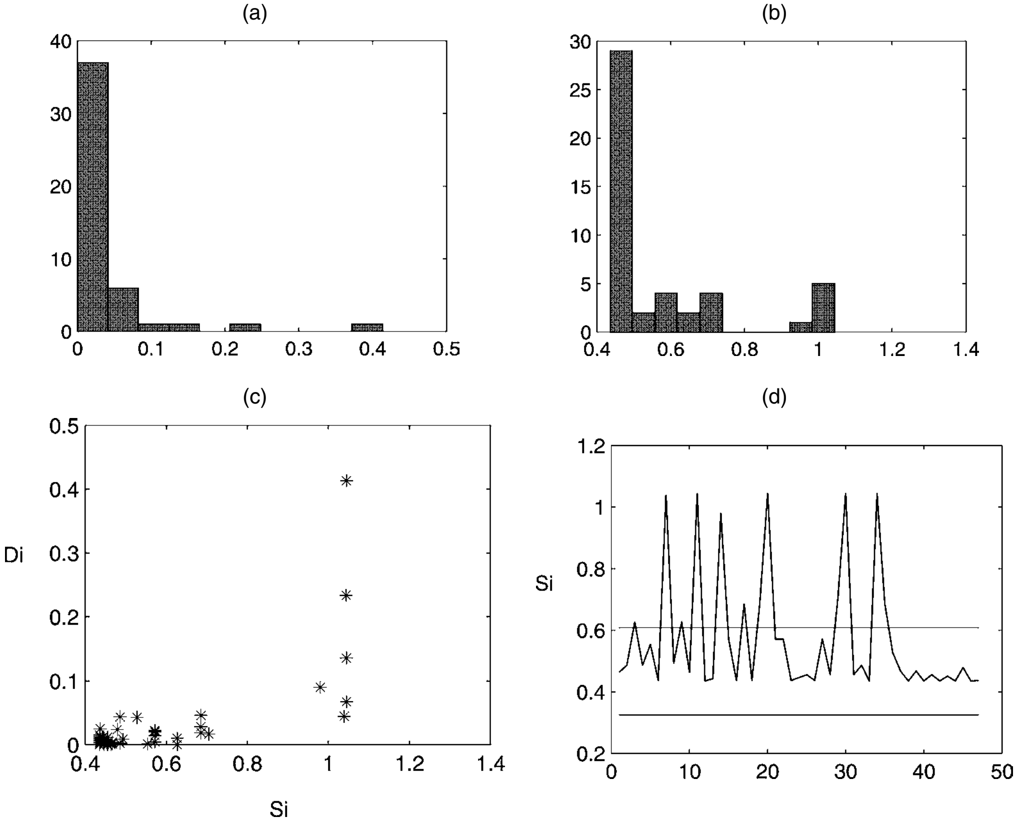

Figure 4. Influence Analysis of the HRD Data. (a) Histogram of Cook's distances. (b) Histogram of S_i. (c) C/S plot. (d) Plot of S_i versus case number.

whereas the group of 20 outliers has the values in the interval [.28, .30]. This is also shown in the C/S plot, where the group of outliers have large leverage and low influence, because the average value of the Cook statistic in the 20-point group is .0108, half of the value of this statistic in the 30-point group (.0218). Then, according to the analysis in Section 3, this group is expected to have a small value for S_i and will be separated from the good data. Finally, the comparison of the S_i statistic with the cutoff values again very clearly indicates the two groups.

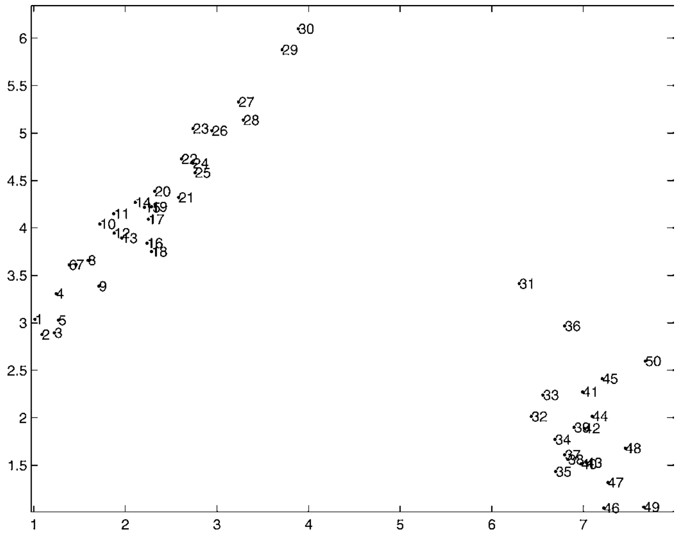

Figure 5. Data From Rousseeuw (1984).

Example 3. As a third example, we use the Boston Housing dataset, which consists of 506 observations on 14 variables and is available at *http://lib.stat.cmu.edu*. This dataset was given by Belsley et al. (1980) and has been considered by a number of authors for regression diagnostics and robust regression. Again, this is considered a difficult example (see Belsley et al. 1980). We have used the same regression as used by Belsley et al. (see also Alexander and Grimshaw 1996 for another analysis of these data), treating as dependent variables the logarithms of the median value of owner-occupied homes and as explanatory variables the 13 variables defined in Table 2.

Figure 7 shows the influence analysis of this dataset. In this example, neither the histograms nor the C/S plot are able to show the heterogeneity in the data. However, a comparison of the values of S_i to the cutoff values indicates 45 points as outliers. These 45 points correspond to observations in the range 366–480, as indicated in Figure 7(d). From Belsley et al. (1980), we obtain that cases 357–488 correspond to Boston, whereas the rest correspond to the suburbs. Also, the 45 points indicated by statistic S_i as outliers all correspond to some central districts of Boston, including Downtown, which suggests that the relation among the variables could be different in these districts than in the rest of the sample. To check this hypothesis, we fitted two regression lines, one to the sample of 461 points and the other to the 45 outliers. Deleting variables that are not significant, we obtain the two regression lines indicated in Table 3, which presents the regression coefficient estimates with the whole sample and the corresponding estimates when the

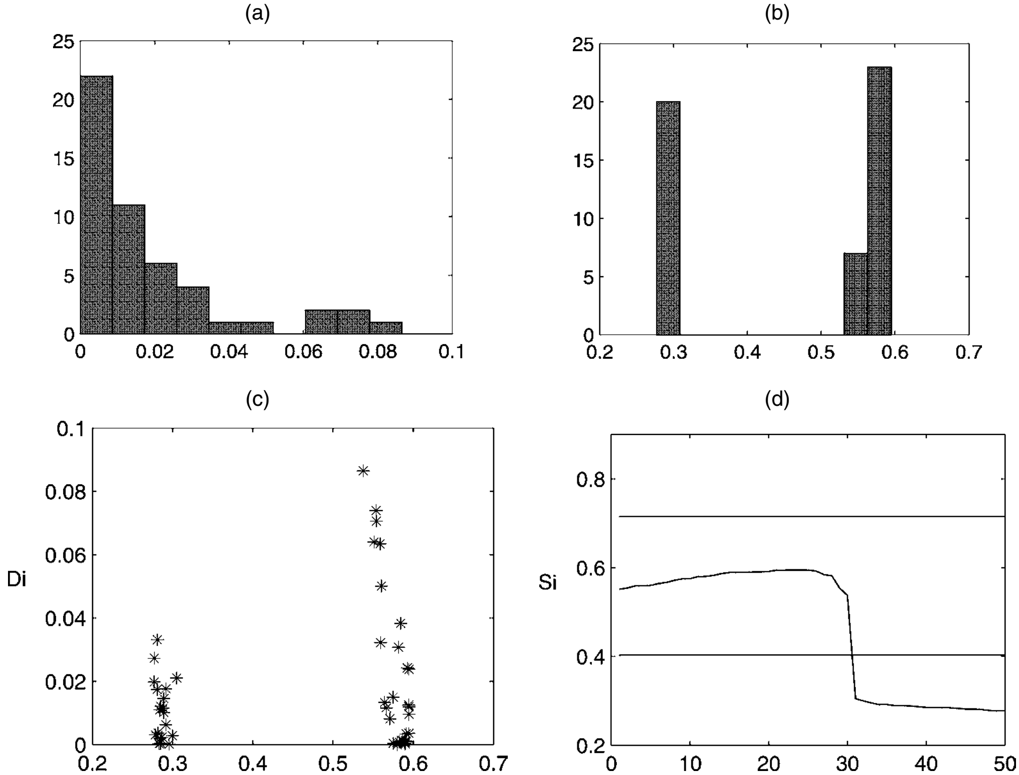

Figure 6. Influence Analysis of the Rousseeuw Two Regression Lines Data. (a) Histogram of Cook's distances. (b) Histogram of S_i. (c) C/S plot. (d) Plot of S_i versus case number.

model is fitted to each of the two groups. It can be seen that the effects of the variables are very different between the two groups of data. In fact, in the second group, only five variables are significant. Note the large reduction in residual sum of squares RSE when fitting different regression equations in the two groups.

Example 4. In this example we analyze the performance of the statistic in a relatively large dataset in high dimension. We consider a heterogeneous sample that is a mixture of two regressions with omitted categorical variable generated by the model

$$y = \beta_0 + \boldsymbol{\beta}_1'\mathbf{x} + \beta_2 z + u,$$

where the $\mathbf{x}$'s have dimension 20 and are independent random drawings from a uniform distribution and $u \sim \mathrm{N}(0, 1)$. The sample size is 2,000, and the first 1,600 cases are generated for the first regression with $z = 0$, and the last 400 cases are generated for the second regression with $z = 1$. The parameter values have been chosen so that the standard diagnosis of the regressing model does not show any evidence of heterogeneity. We have chosen $\beta_0 = 1$, $\boldsymbol{\beta}_1' = \mathbf{1}_{20}' = (1, \dots, 1)$, and $\beta_2 = -100$, and in the first regression the range of the explanatory variables is (0, 10), so that $\mathbf{x}|(z = 0) \sim [\mathrm{U}(0, 10)]^{20}$, whereas for the second the range is (9, 10), so that $\mathbf{x}|(z = 1) \sim [\mathrm{U}(9, 10)]^{20}$. This data has also been used in Peña, Rodriguez, and Tiao (2003).

Figure 8 shows the histogram of the residuals and the plots of residuals versus fitted values in a regression model fitted to the sample of 2,000 observations. No indication of heterogeneity is found. Figure 9 shows the influence analysis. Again, Cook's distance does not demonstrate any sign of heterogeneity, whereas the S_i statistic clearly indicates the two groups of data.

Table 2. Explanatory Variables for the Boston Housing Data

Name	*Description*
crim	Per capita crime rate by town
zn	Proportion of residential land zoned for lots over 25,000 sq. ft.
indus	Proportion of nonretail business acres per town
chas	Charles River dummy variable (= 1 if tract bounds river; 0 otherwise)
noxsq	Nitric oxide concentration (parts per 10 million) squared
rm	Average number of rooms per dwelling squared
age	Proportion of owner-occupied units built prior to 1940
dis	Log of weighted distances to five Boston employment centres
rad	Log of index of accessibility to radial highways
tax	Full-value property-tax rate per $10,000
ptratio	Pupil-teacher ratio by town
b	$(Bk - .63)^2$, where *Bk* is the proportion of blacks by town
lstat	Log of the proportion of lower status of the population

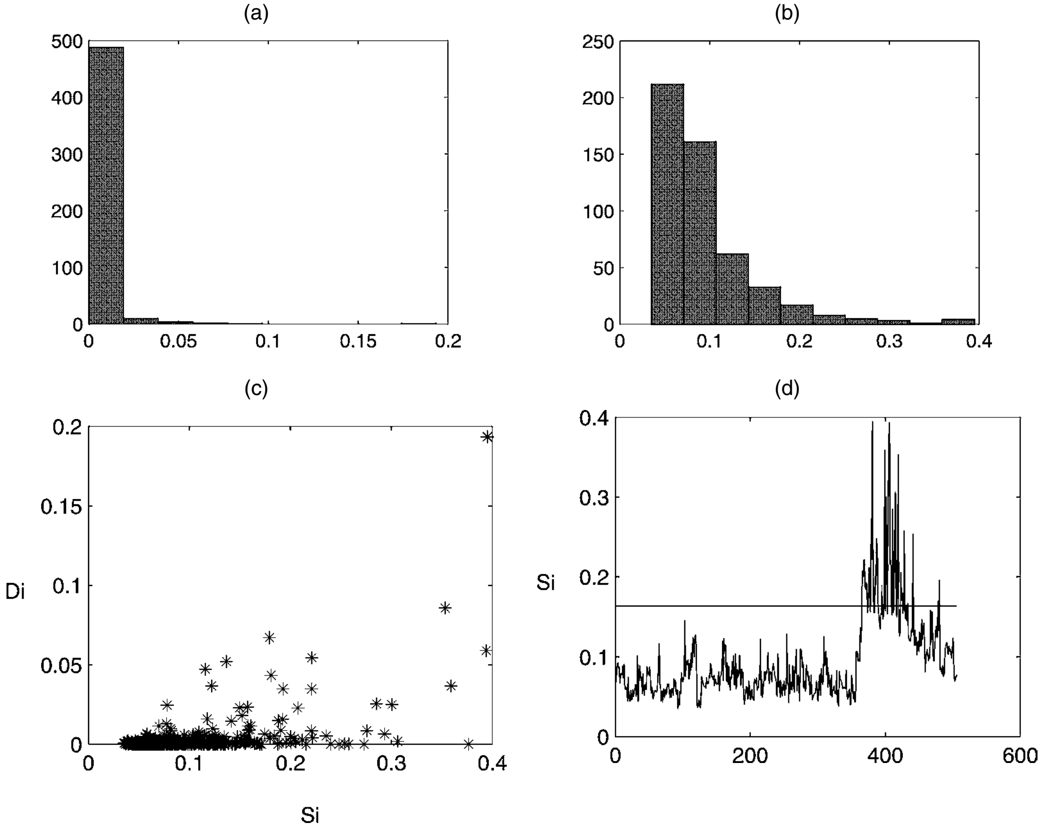

Figure 7. Influence Analysis of the Boston Housing Data. (a) Histogram of Cook's distances. (b) Histogram of S_i. (c) C/S plot. (d) Plot of S_i versus case number.

5. COMPARISON WITH OTHER APPROACHES

It is interesting to see the relationship of our statistic to other ways of looking at influence and dealing with masking. Cook (1986) proposed a procedure for the assessment of the influence on a vector of parameters $\boldsymbol{\theta}$ of a minor perturbation in a statistical model. This approach is very flexible and can be used to see the effect of small perturbations that normally would not be detected by the deletion of one observation. Cook suggested that one introduce an $n \times p$ vector $\mathbf{w}$ of case weights and use the likelihood displacement $(L(\hat{\boldsymbol{\theta}}) - L(\hat{\boldsymbol{\theta}}_w))$, where $\hat{\boldsymbol{\theta}}$ is the maximum likelihood estimator (MLE) of $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}_w$ is the MLE when the case-weight $\mathbf{w}$ is introduced. Then he showed that the directions of greatest local change in the likelihood displacement for the linear regression model are given by the eigenvectors linked to the largest eigenvalues of the curvature matrix $\mathbf{L} = \mathbf{EHE}$, where $\mathbf{E}$ is the vector of residuals. (See Hartless et al. 2003 for a recent contribution in this area proposing another eigenvalue analysis, and Suárez Rancel and González Sierra 2001 for a review of this approach in regression.) These eigenvalue analyses are related to the analysis of Peña and Yohai (1995), who showed that the global influence matrix that they introduced

Table 3. Regression Coefficients and RSEs in the Whole Sample and in the Two Groups Found by the SAR Procedure

	All sample		*Group 1*		*Group 2*	
Name	*Value*	*Std. error*	*Value*	*Std. error*	*Value*	*Std. error*
(intercept)	11.4655	.1544	10.4797	.1420	13.3855	.4535
crim	−.0119	.0012	−.0205	.0040	−.0088	.0024
zn	.0001	.0005				
indus	.0002	.0024				
chas	.0914	.0332	.0401	.0271		
noxsq	−.6380	.1131	−.2871	.0939	−2.7108	.8180
rm	.0063	.0013	.0149	.0013	−.0069	.0047
age	.0001	.0005	−.0012	.0004		
dis	−.1913	.0334	−.1273	.0253	−.7906	.2491
rad	.0957	.0191	.0894	.0153		
tax	−.0004	.0001	−.0004	.0001		
ptratio	−.0311	.0050	−.0288	.0038		
b	.3637	.1031	.7241	.1042		
lstat	−.3712	.0250	−.2020	.0237	−.6609	.0802
RSE	16.378		9.2210		2.8298	

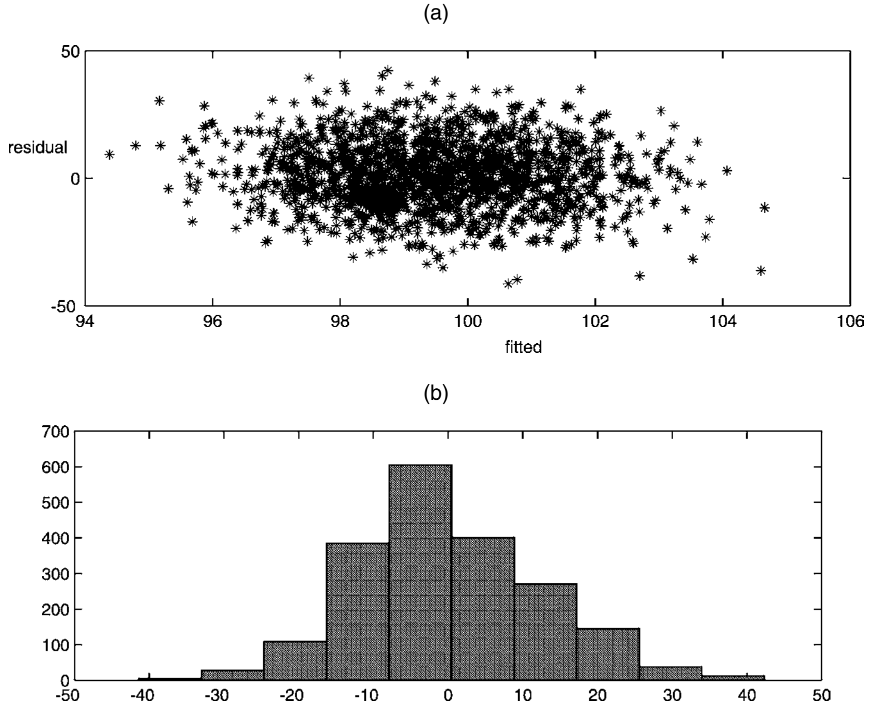

Figure 8. Plot of Residuals versus Fitted Values (a) and a Histogram of the Residuals (b) in the Two-Regression Simulated Example.

was a generalization of the **L** local influence matrix by using the standardized residuals instead of the matrix of least squares residuals. For low-leverage outliers, both matrices will be similar and will have similar eigenvalues, but for high-leverage out-

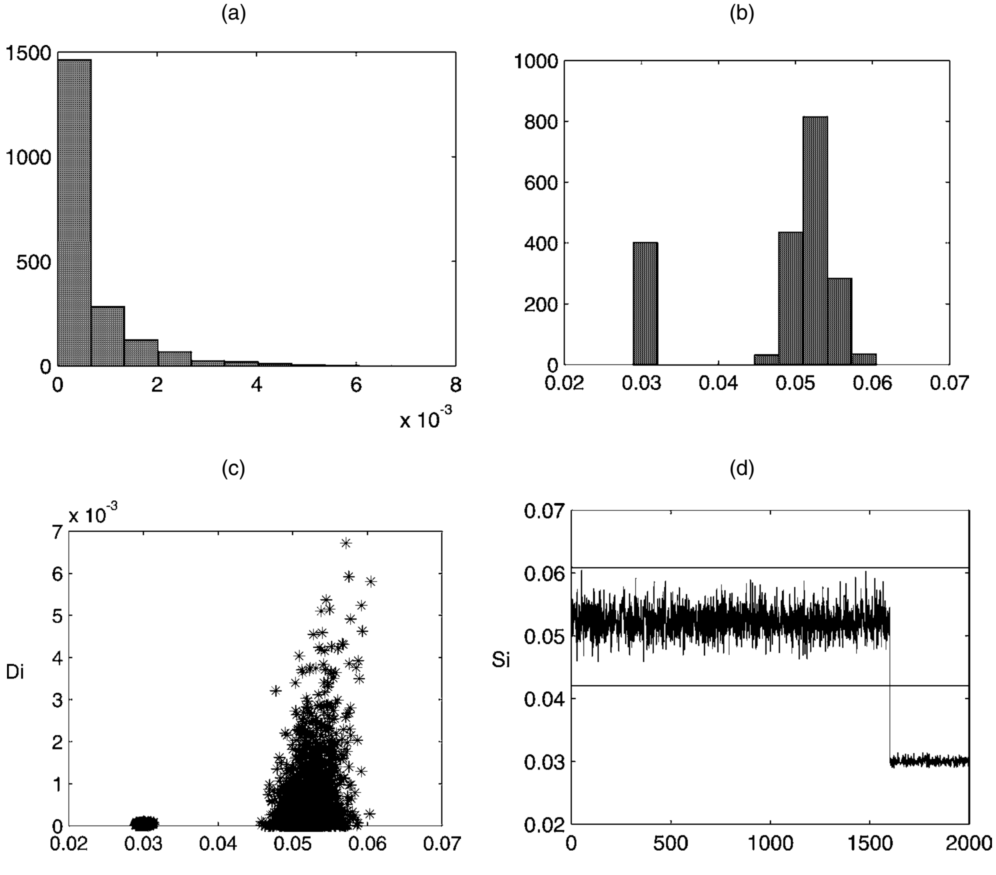

Figure 9. Influence Analysis of the Two Regression Simulated Data. (a) Histogram of Cook's distance. (b) Histogram of S_i. (c) C/S plot. (d) Plot of S_i versus case number.

liers the directions of local influence may be very different from those recommended by Peña and Yohai (1995) for outlier detection. The ideas presented in this article can be used to suggest new ways to apply the local influence approach by exploring the effect of perturbations involving all of the observations in the sample.

The statistic that we propose can also be used as a starting point to build robust estimation procedures for regression. These procedures find estimates defined by

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}\in A} S\big(e_1(\boldsymbol{\beta})\cdots e_n(\boldsymbol{\beta})\big), \tag{19}$$

where $A = \{\boldsymbol{\beta}^{(1)}, \dots, \boldsymbol{\beta}^{(N)}\}$ is a finite set. For instance, Rousseeuw (1984) proposed obtaining the elements of A by choosing at random N subsamples of p different data points, but this number increases exponentially with p, and thus the method based on random subsampling can be applied only when p is not very large. Atkinson (1994) proposed a fast method for detecting multiple outliers using a simple forward search from random starting points. Instead of drawing N basic subsamples, Atkinson suggested drawing $h < N$ random subsamples and using LSEs to fit subsets of size $p, p+1, \dots, n$, from each subsample. Then outliers are identified as the points having large residuals from the fit that minimizes the least median of squares criterion. This procedure requires that at least one of the h subsamples does not contain a high-leverage outlier, and will not be very effective when the number of variables p is large. Peña and Yohai (1999) proposed a procedure to build fast, powerful, robust estimates and identify outliers that uses an eigenvalue analysis of a sensitivity matrix built using ideas similar to the ones used here to build the statistic introduced in this article.

The main advantage of our proposed statistic is for routine analysis of large datasets in high dimension. In this situation, we have shown that a comparison of the S_i statistic with the cutoff values is able to identify groups of outliers in large high-dimensional datasets. This is a great advantage over alternative procedures based on graphical representations with no clear limits to identify outlying values, which will not be very useful for large datasets. Also, this is an advantage over robust estimation methods, which can be computationally very demanding and even unfeasible in some large datasets. As we have shown in Examples 3 and 4, the simple statistic that we propose will work very well in these situations with a trivial computational cost.

6. SENSITIVITY IN OTHER PROBLEMS

Our ideas can be easily generalized for more general models. Suppose that $y_1, \dots, y_n$ are independent random variables, where y_i has a probability density function $f_i(y, \boldsymbol{\theta}, \sigma)$, $\boldsymbol{\theta} \in \mathbb{R}^p$, and $\sigma \in \mathbb{R}$ is a nuisance parameter. This general setup includes linear and nonlinear regression and generalized linear models. For instance, in linear regression, f_i usually is a normal density with mean $\mathbf{x}_i'\boldsymbol{\theta}$ and variance σ^2, where $\mathbf{x}_i \in \mathbb{R}^p$. In nonlinear regression, f_i is a normal density with mean $g(\mathbf{x}_i', \boldsymbol{\theta})$ and variance σ^2. In generalized linear models,

$$f_i(y, \boldsymbol{\theta}, \sigma) = \exp\big\{(yh(\mathbf{x}_i'\boldsymbol{\theta}) - b(\mathbf{x}_i'\boldsymbol{\theta}))/a(\sigma) + c(y, \sigma)\big\};$$

that is, $f_i(y, \boldsymbol{\theta}, \sigma)$ belongs to an exponential family with parameters $h(\mathbf{x}_i'\boldsymbol{\theta})$ and σ.

Let $\hat{\boldsymbol{\theta}}$ and $\hat{\sigma}$ be the MLEs of $\boldsymbol{\theta}$ and σ, and let $\hat{\boldsymbol{\theta}}_{(i)}$ be the MLE of $\boldsymbol{\theta}$ when observation i is deleted. Let $\hat{y}_i$ be the forecast of y_i based on the minimization of some loss function, and let $\hat{y}_{i(j)}$ be the forecast based on the same loss function when observation j is deleted from the sample. The influence of the ith observation is measured by the standardized forecast change

$$D_i = \frac{(\hat{y}_i - \hat{y}_{i(i)})^2}{s^2(\hat{y}_i)},$$

where $s^2(\hat{y}_i)$ is an estimate of the variance of the forecast. The complementary S_i statistic,

$$S_i = \frac{\sum_{j=1}^n (\hat{y}_i - \hat{y}_{i(j)})^2}{s^2(\hat{y}_i)},$$

measures how the point is affected by each of the other sample points. Further research is needed on the properties of this generalization for the different models.

ACKNOWLEDGMENTS

The author thanks Victor Yohai for many fruitful discussions and joint work on this manuscript. Although he has decided not to be a coauthor, the main ideas in this article were developed when working together. This work was been supported by MCYT grant 2000-0167, and is dedicated to D. R. Cook whose work I admire very much.

APPENDIX: RELATION AMONG TRUE AND OBSERVED RESIDUALS IN THE CONTAMINATED SAMPLE

The projection matrix $\mathbf{H}$ is given by $\mathbf{H} = [\mathbf{X}_0', \mathbf{x}_a'\mathbf{1}_k]'(\mathbf{X}_0'\mathbf{X}_0 + k\mathbf{x}_a\mathbf{x}_a')^{-1}[\mathbf{X}_0', \mathbf{x}_a'\mathbf{1}_k]$ and using the Woodbury–Sherman–Morrison equation for the inverse of $\mathbf{X}_0'\mathbf{X}_0 + k\mathbf{x}_a\mathbf{x}_a'$ (see, e.g., Cook and Weisberg 1982, p. 136), we have that

$$\mathbf{H}_{11} = \mathbf{H}^0 - \mathbf{X}_0(\mathbf{X}_0'\mathbf{X}_0)^{-1}\mathbf{x}_a\mathbf{x}_a'(\mathbf{X}_0'\mathbf{X}_0)^{-1}\mathbf{X}_0\frac{k}{kh_a^0 + 1},$$

where $\mathbf{H}^0 = \mathbf{X}_0(\mathbf{X}_0'\mathbf{X}_0)^{-1}\mathbf{X}_0'$ and $h_a^0 = \mathbf{x}_a'(\mathbf{X}_0'\mathbf{X}_0)^{-1}\mathbf{x}_a$. Because $\mathbf{h}_{1a}^0 = \mathbf{X}_0(\mathbf{X}_0'\mathbf{X}_0)^{-1}\mathbf{x}_a$, we obtain (11). Also

$$\mathbf{H}_{12} = \mathbf{X}_0(\mathbf{X}_0'\mathbf{X}_0)^{-1}\mathbf{x}_a\mathbf{1}_k' - \mathbf{X}_0(\mathbf{X}_0'\mathbf{X}_0)^{-1}\mathbf{x}_a\mathbf{x}_a'(\mathbf{X}_0'\mathbf{X}_0)^{-1}\mathbf{x}_a\mathbf{1}_k'\frac{k}{kh_a^0 + 1},$$

and this leads to (12). In the same way, we have that

$$\mathbf{H}_{22} = \mathbf{1}_k\mathbf{x}_a'(\mathbf{X}_0'\mathbf{X}_0)^{-1}\mathbf{x}_a\mathbf{1}_k' - \mathbf{1}_k\mathbf{x}_a'(\mathbf{X}_0'\mathbf{X}_0)^{-1}\mathbf{x}_a\mathbf{x}_a'(\mathbf{X}_0'\mathbf{X}_0)^{-1}\mathbf{x}_a\mathbf{1}_k'\mathbf{1}_k\mathbf{1}_k'\frac{k}{kh_a^0 + 1},$$

and (13) is obtained. The parameters of both regressions are related by

$$\hat{\boldsymbol{\beta}}_T = \hat{\boldsymbol{\beta}}_0 + (\mathbf{X}_0'\mathbf{X}_0)^{-1}\mathbf{x}_a\frac{k}{kh_a^0 + 1}u_a,$$

and thus the observed residuals $e_i = y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}}_T$ are related to the true residuals $u_i = y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}}_0$ by

$$e_i = y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}}_0 - \mathbf{x}_i'(\mathbf{X}_0'\mathbf{X}_0)^{-1}\mathbf{x}_a \frac{k}{kh_a^0 + 1} u_a, \qquad \text{(A.1)}$$

which can be written as

$$e_i = u_i - \frac{h_{ia}^0 k}{kh_a^0 + 1} u_a, \qquad \text{(A.2)}$$

and, using (12), (14) and (15) are obtained.

[Received September 2002. Revised July 2004.]

REFERENCES

Alexander, W. P., and Grimshaw, S. D. (1996), “Treed Regression,” *Journal of Computational and Graphical Statistics*, 5, 156–175.

Atkinson, A. C. (1981), “Two Graphical Displays for Outlying and Influence Observations in Regression,” *Biometrika*, 68, 13–20.

——— (1985), *Plots, Transformations and Regression*, Oxford, U.K.: Clarendon Press.

——— (1994), “Fast Very Robust Methods for the Detection of Multiple Outliers,” *Journal of the American Statistical Association*, 89, 1329–1339.

Belsley, D. A., Kuh, E., and Welsch, R. E. (1980), *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, New York: Wiley.

Brown, G. C., and Lawrence, A. J. (2000), “Theory and Illustration of Regression Influence Diagnostics,” *Communications in Statistics, Part A—Theory and Methods*, 29, 2079–2107.

Chatterjee, S., and Hadi, A. S. (1988), *Sensitivity Analysis in Linear Regression*, New York: Wiley.

Cook, R. D. (1977), “Detection of Influential Observations in Linear Regression,” *Technometrics*, 19, 15–18.

——— (1986), “Assessment of Local Influence” (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 48, 133–169.

Cook, R. D., Peña, D., and Weisberg, S. (1988), “The Likelihood Displacement: A Unifying Principle for Influence,” *Communications in Statistics, Part A—Theory and Methods*, 17, 623–640.

Cook, R. D., and Weisberg, S. (1982), *Residuals and Influence in Regression*, New York: Chapman & Hall.

Critchely, F., Atkinson, R. A., Lu, G., and Biazi, E. (2001), “Influence Analysis Based on the Case Sensitivity Function,” *Journal of the Royal Statistical Society*, Ser. B, 63, 307–323.

Hartless, G., Booth, J. G., and Littell, R. C. (2003), “Local Influence of Predictors in Multiple Linear Regression,” *Technometrics*, 45, 326–332.

Hawkins, D. M., Bradu, D., and Kass, G. V. (1984), “Location of Several Outliers in Multiple Regression Data Using Elemental Sets,” *Technometrics*, 26, 197–208.

Justel, A., and Peña, D. (2001), “Bayesian Unmasking in Linear Models,” *Computational Statistics and Data Analysis*, 36, 69–94.

Lawrance, A. J. (1995), “Deletion Influence and Masking in Regression,” *Journal of Royal Statistical Society*, Ser. B, 57, 181–189.

Muller, E. K., and Mok, M. C. (1997), “The Distribution of Cook’s *D* Statistics,” *Communications in Statistics, Part A—Theory and Methods*, 26, 525–546.

Pregibon, D. (1981), “Logistic Regression Diagnostics,” *The Annals of Statistics*, 9, 705–724.

Peña, D. (1990), “Influential Observations in Time Series,” *Journal of Business & Economic Statistics*, 8, 235–241.

Peña, D., Rodriguez, J., and Tiao, G. C. (2003), “Identifying Mixtures of Regression Equations by the SAR Procedure,” in *Bayesian Statistics 7*, eds. J. M. Bernardo, et al., New York: Oxford University Press, pp. 327–347.

Peña, D., and Yohai, V. J. (1995), “The Detection of Influential Subsets in Linear Regression Using an Influence Matrix,” *Journal of the Royal Statistical Society*, Ser. B, 57, 145–156.

——— (1999), “A Fast Procedure for Robust Estimation and Diagnostics in Large Regression Problems,” *Journal of the American Statistical Association*, 94, 434–445.

Rousseeuw, P. J. (1984), “Least Median of Squares Regression,” *Journal of the American Statistical Association*, 9, 871–880.

Rousseeuw, P. J., and Leroy, A. M. (1987), *Robust Regression and Outlier Detection*, New York: Wiley.

Suárez Rancel, M., and González Sierra, M. A. (2001), “Regression Diagnostic Using Local Influence: A Review,” *Communication in Statistics, Part A—Theory and Methods*, 30, 799–813.

Welsch, R. E. (1982), “Influence Functions and Regression Diagnosis,” in *Modern Data Analysis*, eds. R. L. Launer and A. F. Siegel, New York: Academic Press.

Williams, D. A. (1987), “Generalized Linear Model Diagnostics Using the Deviance and Single Case Deletions,” *Applied Statistics*, 36, 181–191.