JOURNAL OF MULTIVARIATE ANALYSIS **61**, 144–158 (1997)
ARTICLE NO. MV971666

LAD Regression for Detecting Outliers in Response and Explanatory Variables

Yadolah Dodge

Statistics Group, University of Neuchâtel, Neuchâtel, Switzerland

Least absolute deviations regression resists outliers in the response variable but is relatively sensitive to outlying observations in the explanatory variables. In this paper a simple solution is proposed to overcome this problem. This is achieved by minimizing the absolute values of vertical and horizontal deviations in turn. Two algorithms are proposed: one for the simple and one for the multiple regression case. The methods presented have been tested on a variety of data and have proven to be quite effective.

1. INTRODUCTION

The classical linear model assumes a relation of the form

$$y_i = \sum_{j=0}^{p} \theta_j x_{ij} + e_i \qquad \text{for} \quad i = 1, ..., n, \tag{1}$$

where n is the number of observations and where x_{ij} is the jth explanatory variable, y_i the response variable, and e_i the error for the observation i. Let $\boldsymbol{\theta}$ be the column vector of parameters $(\theta_0, ..., \theta_p)'$ and let $\mathbf{x}_i$ be the row vector of explanatory variables $(x_{i0}, ..., x_{ip})$ for the observation i. A purpose of multiple regression is to estimate the vector of unknown parameters $\boldsymbol{\theta}$ by $\hat{\boldsymbol{\theta}} = (\hat{\theta}_0, ..., \hat{\theta}_p)'$ from observations $(x_{i0}, ..., x_{ip}, y_i)$, $i = 1, ..., n$.

To obtain the estimated values $\hat{y}_i$ we multiply $\mathbf{x}_i$ by $\hat{\boldsymbol{\theta}}$ which yields

$$\hat{y}_i = x_{i0}\hat{\theta}_0 + \cdots + x_{ip}\hat{\theta}_p = \mathbf{x}_i\hat{\boldsymbol{\theta}}.$$

The residual r_i of the observation i is the difference between what is actually observed and what is estimated:

$$r_i = y_i - \hat{y}_i.$$

Received March 6, 1996; revised October 16, 1996.

AMS 1990 subject classifications: 62J05, 62J20.

Key words and phrases: least absolute deviations regression, inverse least absolute deviations regression, robust regression, outliers, leverage points.

0047-259X/97 $25.00

The most popular method for estimating the parameters is the celebrated least squares (LS). The usual assumption for using this method which goes back to Legendre (1805) is that the errors e_i are distributed normally with mean 0 and variance σ^2. The least squares criterion is

$$\min_{\theta} \sum_{i=1}^{n} r_i^2.$$

In this setting, we usually considers the y_i as observed values and the $x_{i0}, ..., x_{ip}$ as fixed numbers. Hence, an observation $(x_{i0}, ..., x_{ip}, y_i)$ is called an outlier whenever y_i lies far away from the bulk of the observations. However, the explanatory variables $x_0, ..., x_p$ are also observed quantities subject to random variability particularly when the design has not been given in advance. Therefore, in fitting a regression line we have to confront outlying x_{ij} as well as outlying y_i. In this connection, a data point that has an extreme value for the response variable is called a *vertical outlier* and a data point that has an extreme value for one of the explanatory variables a *leverage point*.

While the least squares method enjoys well known properties within Gaussian parametric models, it is recognized that outliers, which arise from heavy-tailed distributions, have an unusually large influence on the resulting estimates. Outlier diagnostics have been developed to detect observations with a large influence on the least squares estimation. For excellent books related to such diagnotics the reader is refered to Cook and Weisberg (1982, 1994) and Chatterjee and Hadi (1988).

Parallel to diagnostics techniques, robust methods with varying degrees of robustness and computational complexity have been developed to modify the LS method so that the outliers have much less influence on the final estimates. Among others are the bounded influence estimators, the repeated median, the least median of squares method and the regression quantiles.

One of the simplest robust alternative to the LS is the least absolute deviations (LAD) method which was introduced in 1757 by Roger Joseph Boscovich. The least absolute deviations or L_1-norm criterion is

$$\min_{\theta} \sum_{i=1}^{n} |r_i|.$$

The LAD method is particularly well-suited to longer-tailed error distributions (e.g., Laplace or Cauchy). Other advantages of LAD method in robust regression are explained by Huber (1987). However, although robust to vertical outliers, this method is relatively sensitive to large leverage points. See, for example, Ellis and Morgenthaler (1992) and Gutenbrunner *et al.* (1993). In this paper we propose the following idea to overcome this problem.

At first outlying y_i are detected by regressing the response variable y on the explanatory ones $x_0, ..., x_p$ using the LAD method. Now suppose there are outlying x_{ij} for some j. Interestingly, these leverage points can be detected by regressing the jth explanatory variable x_j on $x_0, ..., x_{j-1}, y, x_{j+1}, ..., x_p$ using again the L_1-norm criterion. In effect as variables x_j and y have "switched" places, the leverage points turn to be vertical outliers. Since LAD regression protects efficiently against vertical outliers, outlying x_{ij} can be now identified.

It should be noted that for the estimation of the parameters, the fixed variables x_j should not be treated as random quantities or as a function of the response variable. Therefore in this setting it is inappropriate to determine a regression of x_j on y since such a pseudo regression would give spurious results. However, when the x_{ij} are themselves observed quantities, outliers can occur in these explanatory variables. In such a case with observed explanatory variables subject to outlier occurences, a regression of x_j on y is justifiable.

An algorithm is presented for the case of simple linear regression in Section 3 along with an example to clarify the algorithm. The multiple regression is treated in Section 4, and in Section 5 some remarks related to the proposed methods are discussed.

2. METHOD

A demonstration of the robustness of the LAD method to vertical outliers is given by the following theorem.

THEOREM 1. *Suppose* $\boldsymbol{\theta}^*$ *is a minimizer of* $F(\boldsymbol{\theta}) = \sum_{i=1}^n |y_i - \mathbf{x}_i\boldsymbol{\theta}|$. *Then* $\boldsymbol{\theta}^*$ *is also a minimizer of* $G(\boldsymbol{\theta}) = \sum_{i=1}^n |z_i - \mathbf{x}_i\boldsymbol{\theta}|$, *provided* $z_i \geqslant \mathbf{x}_i\boldsymbol{\theta}^*$ *whenever* $y_i > \mathbf{x}_i\boldsymbol{\theta}^*$ *and* $z_i \leqslant \mathbf{x}_i\boldsymbol{\theta}^*$ *whenever* $y_i < \mathbf{x}_i\boldsymbol{\theta}^*$.

Proof. F is a convex function and so $\boldsymbol{\theta}^*$ is a minimizer of F if and only if the zero vector is a subgradient of F at $\boldsymbol{\theta}^*$, that is, $\mathbf{0} \in \partial F(\boldsymbol{\theta}^*)$ (Osborne, 1985, Theorem 5.1, p. 23). It suffices to show that $\partial F(\boldsymbol{\theta}^*) \subseteq \partial G(\boldsymbol{\theta}^*)$. Write $F = \sum_{i=1}^n F_i$ and $G = \sum_{i=1}^n G_i$, where $F_i(\boldsymbol{\theta}) = |y_i - \mathbf{x}_i\boldsymbol{\theta}|$ and $G_i(\boldsymbol{\theta}) = |z_i - \mathbf{x}_i\boldsymbol{\theta}|$. By Lemma 4.7 on page 20 in Osborne (1985), $\partial F = \sum_{i=1}^n \partial F_i$ and $\partial G = \sum_{i=1}^n \partial G_i$, so it suffices to show $\partial F_i(\boldsymbol{\theta}^*) \subseteq \partial G_i(\boldsymbol{\theta}^*)$. From Lemma 1.1 on page 182 in Osborne (1985), we see that

$$\partial F_i(\boldsymbol{\theta}^*) = \begin{cases} \{-\mathbf{x}_i\} & \text{if} \quad y_i > \mathbf{x}_i\boldsymbol{\theta}^*, \\ [-\mathbf{x}_i, \mathbf{x}_i] & \text{if} \quad y_i = \mathbf{x}_i\boldsymbol{\theta}^*, \\ \{\mathbf{x}_i\} & \text{if} \quad y_i < \mathbf{x}_i\boldsymbol{\theta}^*, \end{cases}$$

where $[-\mathbf{x}_i, \mathbf{x}_i]$ denotes the line segment joining $-\mathbf{x}_i$ and $\mathbf{x}_i$. If $y_i > \mathbf{x}_i\boldsymbol{\theta}^*$ and $z_i > \mathbf{x}_i\boldsymbol{\theta}^*$, then $\partial F_i(\boldsymbol{\theta}^*) = \{-\mathbf{x}_i\} = \partial G_i(\boldsymbol{\theta}^*)$. If $y_i > \mathbf{x}_i\boldsymbol{\theta}^*$ and $z_i = \mathbf{x}_i\boldsymbol{\theta}^*$, then $\partial F_i(\boldsymbol{\theta}^*) = \{-\mathbf{x}_i\} \subseteq [-\mathbf{x}_i, \mathbf{x}_i] = \partial G_i(\boldsymbol{\theta}^*)$. The other cases can be verified similarly. Hence the proof.

The theorem says that the LAD fit is completely unaffected by any change in y_i whenever the sign of the residual remains the same. Thus the LAD regression method is robust to vertical outliers. Such robustness is independent of the number of independent variables for $p < n$.

We can therefore identify outlying y_i as points that lie far from the LAD fit, that is, the observations with large positive or negative residuals. The units of measurements must be taken into account, and so in order to decide if a residual r_i is large, we standarize the residuals by dividing them by an estimate of the standard deviation of the population of errors. Such an estimate $\hat{\sigma}$ must be robust itself and hence must depend only on good points and should not be affected by the outliers. One such estimate is

$$\hat{\sigma} = 1.4826 \text{ MAD}, \tag{2}$$

where MAD is the median of the absolute values of the nonzero residuals.

Sometimes MAD is computed using all the residuals (including the zeros). Hill and Holland (1977) and McKean and Schrader (1987) have found that estimates based on all the residuals tend to be too small. Otherwise if more than 50% of the points lie on the LAD regression line, $\hat{\sigma}$ can be computed by $0.4 \cdot \min\{|r_i|, r_i \neq 0\}$.

The multiplier 1.4826 ensures that, if a population of errors is assumed to have a normal distribution, then $\hat{\sigma}$ is a consistent estimate of σ.

We identify an observation i as an outlier if and only if $|r_i/\hat{\sigma}| \geqslant 2.5$. The value of 2.5 is of course arbitrary but it is chosen so that for Gaussian errors only about 1% of the observation will be qualified as outliers.

3. SIMPLE LINEAR REGRESSION CASE

We first consider the case of simple linear regression where n observations (x_i, y_i) are related by

$$y_i = \theta_0 + \theta_1 x_i + e_i. \tag{3}$$

This model is obtained from (1) by letting $p = 1$, $x_{i0} = 1$, and $x_{i1} = x_i$. According to Theorem 1, vertical outliers can be detected once the parameters θ_0 and θ_1 are estimated by $\hat{\theta}_0$ and $\hat{\theta}_1$ that minimize

$$\sum_{i=1}^{n} |y_i - \hat{\theta}_0 - \hat{\theta}_1 x_i|.$$

Now suppose there are leverage points in the data set. By switching in (3) the explanatory variable x and the response variable y, we obtain the new model

$$x_i = \eta_0 + \eta_1 y_i + f_i, \tag{4}$$

where f_i is the error for the observation i and where η_0 and η_1 are parameters estimated by $\hat{\eta}_0$ and $\hat{\eta}_1$ that minimize

$$\sum_{i=1}^{n} |x_i - \hat{\eta}_0 - \hat{\eta}_1 y_i|.$$

The leverage points in (3) turn to be vertical outlier in (4) and can be hence identified according to Theorem 1. The algorithm is thus the following.

ALGORITHM 1.

Step 1. Regress y on x using all observations by applying the LAD method. In addition to the parameters θ_0 and θ_1, the scale parameter σ has to be estimated by $\hat{\sigma}$ as in (2).

Step 2. For all observations i compute the standard residuals $r_i/\hat{\sigma}$, where r_i is the residual of the observation i with repect to the LAD fit. Remove all observations i for which $|r_i/\hat{\sigma}| \geqslant 2.5$.

Step 3. Regress the original explanatory variable x on the original response variable y using the remaining observations by applying the LAD method. Compute $\hat{\sigma}$ as in step 1.

Step 4. For each of the remaining observations i compute the standard residuals $r_i/\hat{\sigma}$ as in step 2. Remove all observations i for which $|r_i/\hat{\sigma}| \geqslant 2.5$.

Step 5. Regress y on x using the remaining observations by applying the LAD method.

The removed observations found in steps 2 and 4 are respectively vertical outliers and leverage points.

Remark 1. When we apply the Algorithm 1 to obtain the estimates of the regression coefficients and to identify the outliers in both directions we may find different results depending on which variable is chosen first to play the role of the dependent variable. With the proposed algorithm the vertical outliers are removed in Step 2. If it is desired to obtain an *invariant solution under ordering*, we can remove the vertical outliers found in Step 2 only after Step 4 so that the regression in Step 3 is computed with all the n observations. However, since the vertical outliers are more frequent than leverage points in a regression problem, it is perhaps natural

TABLE I

Data for the Hertzsprung–Russell Diagram of the Star Cluster CYG OB1

Index of Star (i)	$\log T_e$ (x_i)	$\log[L/L_0]$ (y_i)	Index of Star (i)	$\log T_e$ (x_i)	$\log[L/L_0]$ (y_i)
1	4.37	5.23	25	4.38	5.02
2	4.56	5.74	26	4.42	4.66
3	4.26	4.93	27	4.29	4.66
4	4.56	5.74	28	4.38	4.90
5	4.30	5.19	29	4.22	4.39
6	4.46	5.46	30	3.48	6.05
7	3.84	4.65	31	4.38	4.42
8	4.57	5.27	32	4.56	5.10
9	4.26	5.57	33	4.45	5.22
10	4.37	5.12	34	3.49	6.29
11	3.49	5.73	35	4.23	4.34
12	4.43	5.45	36	4.62	5.62
13	4.48	5.42	37	4.53	5.10
14	4.01	4.05	38	4.45	5.22
15	4.29	4.26	39	4.53	5.18
16	4.42	4.58	40	4.43	5.57
17	4.23	3.94	41	4.38	4.62
18	4.42	4.18	42	4.45	5.06
19	4.23	4.18	43	4.50	5.34
20	3.49	5.89	44	4.45	5.34
21	4.29	4.38	45	4.55	5.54
22	4.29	4.22	46	4.45	4.98
23	4.42	4.42	47	4.42	4.50
24	4.49	4.85			

to clean up the dependent variable in Step 2 from the contaminated data points.

EXAMPLE 1. This example comes from the field of astronomy. The data in Table I from the Hertzsprung–Russell diagram of the star cluster CYG OB1, which contains 47 observations (stars) in the direction of Cygnus (Rousseeuw and Leroy, 1987, p. 27).

Let us see how Algorithm 1 works with this data set. Steps 1 to 5 are illustrated by Figs. 1 and 2.

Step 1. By regressing y on x with all the 47 observations using the LAD method, we obtain as equation $\hat{y}_i = 8.149 - 0.693x_i$ with $\hat{\sigma} = 0.624$.

Step 2. All standard residuals fall inside the range $[-2.5, 2.5]$.

Step 3. By regressing x on y with all the 47 observations using the LAD method, we obtain as equation $\hat{x}_i = 3.687 + 0.143y_i$ with $\hat{\sigma} = 0.106$.

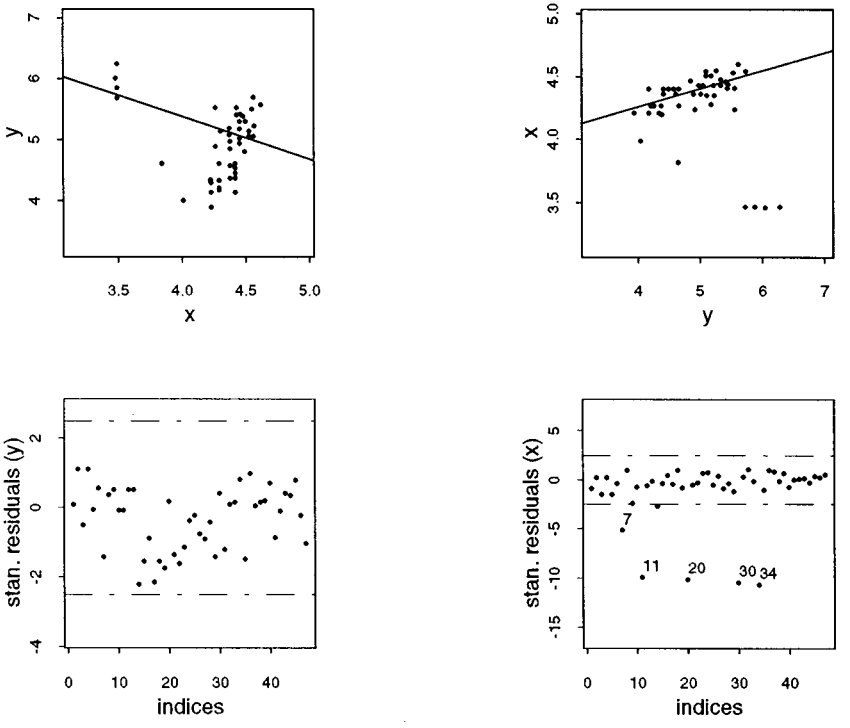

FIG. 1. Steps 1–4 for Example 1.

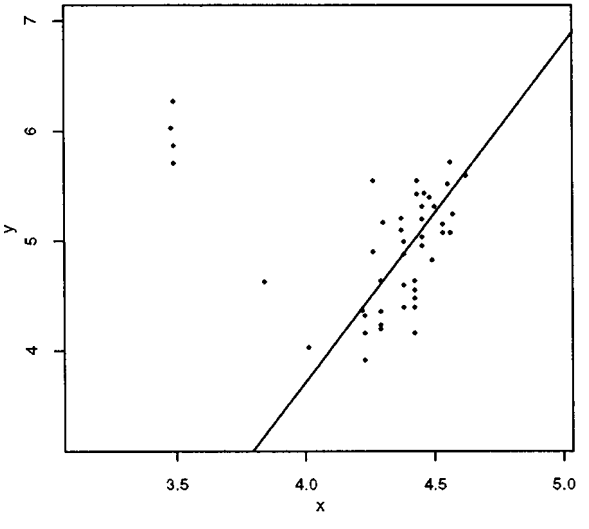

FIG. 2. Step 5 for Example 1 (final LAD fit without outliers).

Step 4. Standard residuals of observations 7, 11, 20, 30, and 34 are situated outside the range $[-2.5, 2.5]$. Thus we remove these five observations from the data set.

Step 5. By regressing y on x with the remaining 42 observations using the LAD method, we obtain as final equation $\hat{y}_i = -8.586 + 3.075x_i$ with $\hat{\sigma} = 0.425$.

Comparison. The least median of squares identifies observations 11, 20, 30, and 34 as leverage points with $\hat{y}_i = -12.298 + 3.898x_i$. The application of LS yields $\hat{y}_i = 6.793 - 0.413x_i$. The LS also identifies observations 11, 20, 30, and 34 as leverage points using the diagonal elements h_{ii} of the projection matrix $H = X(X'X)^{-1}X'$ (hat matrix). In fact the h_{ii} for these observations are larger than the cutoff value $2p/n = 0.085$. By applying the method of LS on the remaining 43 observations, we obtain the equation $\hat{y}_i = -4.057 + 2.048x_i$. At this stage the observations 7 and 14 are identified as outlying points. If we continue to apply the method of LS on the remaining 41 observations, we obtain the equation $\hat{y}_i = -8.208 + 2.984x_i$. Here the observation 9 is identified as outlying point. The final LS equation with the remaining 40 observations is $\hat{y}_i = -9.835 + 3.347x_i$.

4. MULTIPLE LINEAR REGRESSION CASE

We now consider the case of multiple linear regression, where n observations $(x_{i1}, ..., x_{ip}, y_i)$ are related by

$$y_i = \theta_0 + \theta_1 x_{i1} + \cdots + \theta_p x_{ip} + e_i. \tag{5}$$

This model is obtained from (1) by letting $x_{i0} = 1$. According to Theorem 1, vertical outliers can be detected once the vector of parameters $\boldsymbol{\theta} = (\theta_0, ..., \theta_p)'$ is estimated by $\hat{\boldsymbol{\theta}} = (\hat{\theta}_0, ..., \hat{\theta}_p)'$, that minimizes

$$\sum_{i=1}^{n} \left| y_i - \hat{\theta}_0 - \sum_{j=1}^{p} \hat{\theta}_j x_{ij} \right|.$$

Now suppose there are outlying x_{ij} in data set for some j. By switching in (5) the jth explanatory variable x_j and the response variable y, we obtain the new model

$$x_{ij} = \eta_0 + \eta_1 x_{i1} + \cdots + \eta_{j-1} x_{ij-1} + \eta_j y_i + \eta_{j+1} x_{ij+1} + \cdots + \eta_p x_{ip} + f_i, \tag{6}$$

where f_i is the error for the observation i and where $\boldsymbol{\eta} = (\eta_0, ..., \eta_p)'$ is the vector of parameters estimated by $\hat{\boldsymbol{\eta}} = (\hat{\eta}_0, ..., \hat{\eta}_p)'$ that minimizes

$$\sum_{i=1}^{p} \left| x_{ij} - \hat{\eta}_0 - \hat{\eta}_j y_i - \sum_{\substack{k=1 \\ k \neq j}}^{p} \hat{\eta}_k x_{ik} \right|.$$

As in the simple linear regression case, the leverage points in (5) turn to be vertical outliers in (6) and hence can be identified according to Theorem 1. Repeating this procedure for each explanatory variable x_j, we obtain the following algorithm.

ALGORITHM 2.

Step 1. Regress y on $x_1, ..., x_p$ using all observations by applying the LAD method. In addition to the vector of parameters $\boldsymbol{\theta}$, the scale parameter σ has to be estimated by $\hat{\sigma}$ as in (2).

Step 2. For all observations i compute the standard residuals $r_i/\hat{\sigma}$, where r_i is the residual of the observation i with respect to the LAD fit.

Step 3. Remove all observations i for which $|r_i/\hat{\sigma}| \geqslant 2.5$.

Step 4. For $j = 1, ..., p$, regress x_j on $x_1, ..., x_{j-1}, y, x_{j+1}, ..., x_p$, using the remaining observations by applying the LAD method. Compute $\hat{\sigma}_j$ as in Step 1.

Step 5. For $j = 1, ..., p$ and for each remaining observation i, compute the standard residuals $r_{ij}/\hat{\sigma}_j$ as in Step 2. Let D be the union of all observations i for which there is at least one j where $|r_{ij}/\hat{\sigma}_j| \geqslant 2.5$. Observe that order in which x_j becomes response variable does not affect the set D.

Step 6. Remove all observations found in D.

Step 7. Regress y on $x_1, ..., x_p$ using the remaining observations by applying the LAD method.

Remark 2. As in the simple linear regression case, results may differ depending on which variable is chosen in Step 1 to play the role of the dependent variable. An alternative algorithm for which the outcome is not affected by the ordering on which variable to regress first is to replace Steps 3 and 6 by Steps 3b and 6b defined as follows:

Step 3b. Let G be the set of observations for which $|r_i/\hat{\sigma}| \geqslant 2.5$.

Step 6b. Remove all observations found in $G \cup D$.

EXAMPLE 2. The data in Table II concern the incidences of fires in a residential area. We want to see how the incidence of fires is related to three characteristics of the area: the age of its houses, its incidence of theft,

TABLE II

AREA (i)	FIRE	log(FIRE) (y_i)	AGE (x_{i1})	THEFT (x_{i2})	INCOME (x_{i3})
1	6.2	1.825	0.604	29	11.744
2	9.5	2.251	0.765	44	9.323
3	10.5	2.351	0.735	36	9.948
4	7.7	2.041	0.669	37	10.656
5	8.6	2.152	0.814	53	9.730
6	34.1	3.529	0.526	68	8.231
7	11.0	2.398	0.426	75	21.480
8	6.9	1.932	0.785	18	11.104
9	7.3	1.988	0.901	31	10.694
10	15.1	2.715	0.898	25	9.631
11	29.1	3.371	0.827	34	7.995
12	2.2	0.788	0.402	14	13.722
13	5.7	1.740	0.279	11	16.250
14	2.0	0.693	0.077	11	13.686
15	2.5	0.916	0.638	22	12.405
16	3.0	1.099	0.512	17	12.198
17	5.4	1.686	0.851	27	11.600
18	2.2	0.788	0.444	9	12.765
19	7.2	1.974	0.842	29	11.084
20	15.1	2.715	0.898	30	10.510
21	16.5	2.803	0.727	40	9.784
22	18.4	2.912	0.729	32	7.342
23	36.2	3.589	0.631	41	6.565
24	39.7	3.681	0.830	147	7.459
25	18.5	2.918	0.783	22	8.014
26	23.3	3.148	0.790	29	8.177
27	12.2	2.501	0.480	46	8.212
28	5.6	1.723	0.715	23	11.230
29	21.8	3.082	0.731	4	8.330
30	21.6	3.073	0.650	31	5.583
31	9.0	2.197	0.754	39	8.564
32	3.6	1.281	0.208	15	12.102
33	5.0	1.609	0.618	32	11.876
34	28.6	3.353	0.781	27	9.742
35	17.4	2.856	0.686	32	7.520
36	11.3	2.425	0.734	34	7.388
37	3.4	1.224	0.020	17	13.842
38	11.9	2.477	0.570	46	11.040
39	10.5	2.351	0.559	42	10.332
40	10.7	2.370	0.675	43	10.908
41	10.8	2.380	0.580	34	11.156
42	4.8	1.569	0.152	19	13.323
43	10.4	2.342	0.408	25	12.960
44	15.6	2.747	0.578	28	11.260
45	7.0	1.946	0.114	3	10.080
46	7.1	1.960	0.492	23	11.428
47	4.9	1.589	0.466	27	13.731

and the income of its families. The data are for 47 predominantly residential areas in Chicago for the year 1975. The column labeled FIRE lists the number of fires per 1000 housing units in the area; the column labeled AGE list the proportion of housing units built before 1940; the column labeled THEFT list the number the thefts per 1000 residents; and the column labeled INCOME lists the median family income as a multiple of \$1000 (Andrews and Herzberg, 1985, p. 409).

Let us see how Algorithm 2 works with this data set. Steps 2 and 5 are illustrated by Fig. 3.

Step 1. By regressing y on x_1, x_2, x_3 with all the 47 observations using the LAD method, we obtain as equation $\hat{y}_i = 3.994 + 0.319x_{i1} + 0.007x_{i2} - 0.209x_{i3}$ with $\hat{\sigma} = 0.508$.

Step 2. Standard residual of observation 7 is 4.5 and hence is situated outside the range $[-2.5, 2.5]$.

Step 3. We remove observation 7 from the data set.

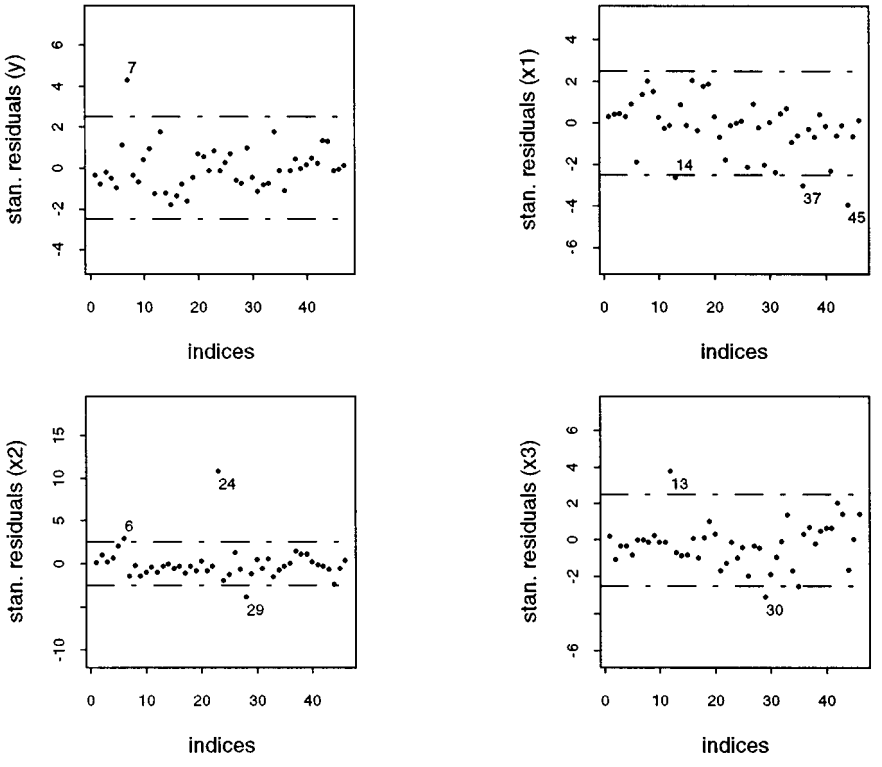

FIG. 3. Steps 2 and 5 for Example 2.

Step 4. By regressing x_1 on y, x_2, x_3 with the 46 remaining observations using the LAD method, we obtain the equation $\hat{x}_{i1} = 1.21 + 0.006y_i + 0.0002x_{i2} - 0.058x_{i3}$ with $\hat{\sigma}_1 = 0.1375$. By regressing x_2 on x_1, y, x_3 we obtain the equation $\hat{x}_{i2} = 18.48 + 10.57x_{i1} + 5.969y_i - 0.908x_{i3}$ with $\hat{\sigma}_2 = 9.385$, and by regressing x_3 on x_1, x_2, y we obtain the equation $\hat{x}_{i3} = 15.8 - 1.239x_{i1} + 0.0008x_{i2} - 2.02y_i$ with $\hat{\sigma}_3 = 1.098$.

Step 5. Standard residuals of observations $\{14, 37, 45\}$, $\{6, 24, 29\}$, and $\{13, 30\}$ are outside the range $[-2.5, 2.5]$ for $j = 1, 2$, and 3, respectively. Thus $D = \{6, 13, 14, 24, 29, 30, 37, 45\}$.

Step 6. We remove the eight observations found in D from the data set.

Step 7. By regressing y on x_1, x_2, x_3 with the 38 remaining observations using the LAD method, we obtain as final equation $\hat{y}_i = 4.205 - 0.021x_{i1} + 0.015x_{i2} - 0.238x_{i3}$ with $\hat{\sigma} = 0.485$.

Note that if we use the alternative algorithm which replaces Steps 3 and 6 by Steps 3b and 6b, we obtain the following results for the same example:

Step 1. By regressing y on x_1, x_2, x_3 with all the 47 observations using the LAD method, we obtain as equation $\hat{y}_i = 3.994 + 0.319x_{i1} + 0.007x_{i2} - 0.209x_{i3}$ with $\hat{\sigma} = 0.508$.

Step 2. Standard residual of observation 7 is 4.5 and hence is situated outside the range $[-2.5, 2.5]$.

Step 3b. Let $G = \{7\}$.

Step 4. By regressing x_1 on y, x_2, x_3 using the LAD method (with all the 47 observations!), we obtain the equation $\hat{x}_{i1} = 0.688 + 0.079y_i + 0.001x_{i2} - 0.025x_{i3}$ with $\hat{\sigma}_1 = 0.179$. By regressing x_2 on x_1, y, x_3 we obtain the equation $\hat{x}_{i2} = -3299 + 11.43x_{i1} + 14.15y_i + 2.344x_{i3}$ with $\hat{\sigma}_2 = 13.74$, and by regressing x_3 on x_1, x_2, y we obtain the equation $\hat{x}_{i3} = 16.11 - 2.201x_{i1} + 0.002x_{i2} - 1.952y_i$ with $\hat{\sigma}_3 = 1.161$.

Step 5. Standard residuals of observations $\{45\}$, $\{24, 29\}$, and $\{7, 13, 30\}$ are outside the range $[-2.5, 2.5]$ for $j = 1, 2$, and 3, respectively. Thus $D = \{7, 13, 24, 29, 30, 45\}$.

Step 6b. We remove six observations found in $G \cup D = \{7, 13, 24, 29, 30, 45\}$.

Step 7. By regressing y on x_1, x_2, x_3 with the 40 remaining observations using the LAD method, we obtain as final equation $\hat{y}_i = 3.75 + 0.182x_{i1} + 0.024x_{i2} - 0.237x_{i3}$ with $\hat{\sigma} = 0.408$.

Comparison. The least median of squares identifies observations 7, 13, 30, and 43 as outlying points with $\hat{y}_i = 6.373 - 0.024x_{i1} + 0.008x_{i2} - 0.422x_{i3}$. The application of LS yields $\hat{y}_i = 2.930 + 0.536x_{i1} + 0.013x_{i2} - 0.134x_{i3}$. The LS identifies observations 7 and 24 as leverage points using the diagonal elements of the hat matrix. After we delete these two observations from the data set and repeating the LS method on the 45 remaining observations, we obtain the equation $\hat{y}_i = 4.102 + 0.269x_{i1} + 0.010x_{i2} - 0.223x_{i3}$. At this stage observations 6, 13, and 45 are identified as outlying points. By repeating once again the LS method on the remaining 42 observations, we obtain the equation $\hat{y}_i = 4.360 + 0.397x_{i1} + 0.007x_{i2} - 0.251x_{i3}$. Here the observation 29 is detected as leverage point. The final LS equation with the remaining 41 observations is $\hat{y}_i = 4.027 + 0.345x_{i1} + 0.013x_{i2} - 0.233x_{i3}$.

5. FINAL REMARKS

An attempt has been made in this paper to provide a robust method of regression analysis that only uses the LAD criterion. This is achieved by taking into account the resistance of the LAD regression method to vertical outliers. That is, if there are leverage points, we have simply to switch the role of the response variable with the explanatory one. The effectiveness of the proposed method for finding a robust fit have been tested on a wide range of problems. It is to be noted, however, as seen in our examples, that the method of *repeated least squares* (after deleting outlying points), although not a common practice, produces results very comparable to robust methods such as the least median of squares or the LAD method introduced in this paper.

From the computational point of view, the LAD method is extremely simple and it requires only a routine to fit the LAD regression. There are several computer programs available for calculation of LAD estimates. See, for example, Sadovski (1974) and Farebrother (1988). Sadovski's code is based on Karst's (1958) algorithm. Farebrother (1988) gives a label-free Pascal translation of Sadovski's algorithm and a Pascal implementation of an improved variant of Sadovski's algorithm. For the case of multiple regression we can use, for example the modified simplex algorithm of Barrodale and Roberts (1973) approach that exist in the IMSL library under the name RLLAV. The LAD regression estimates are also obtainable from the function l1fit in the computer language S-Plus and from the robust regression package ROBSYS (Marazzi, 1993). Minimization of LAD can be formulated as a linear programming problem, as in Arthanari and Dodge (1993, Section 2–7). A recent discovery by Portnoy and Koenker (1997) on the computation of L_1 and other regression quantiles

in linear regression models, which involves the application of interior point linear programming methods, provides computational methods that are faster than least sequences for large n.

The statistical inference procedure to LAD estimation is quite similar to the classical analysis of variance. The Wald, likelihood ratio, and Lagrange multiplier tests were suggested by Koenker and Bassett (1982) for hypothesis testing in the context of LAD estimation. They showed that the three tests are asymptotically equivalent. For testing a general linear hypothesis the procedure is to replace the classical reduction in sum of squared residuals by the reduction in sum of absolute residuals. Using such direct analogy we can produce an LAD analysis of variance table which summarizes the LAD test of hypothesis. McKean and Schrader (1987) provided an estimate of the scale parameter that is used as a standardizing constant in the LAD analysis. This estimate is not only used for standardizing test statistic, but is also used in confidence intervals and multiple comparison procedures based on LAD estimates. Dielman and Rose (1995, 1996) compared the three approaches proposed by Koenker and Bassett (1982) for assessing the significance of coefficients in LAD. The results of their study suggest that the likelihood and Lagrange multiplier tests are preferable to the Wald test for multiple regression situation. For a complete treatment on LAD regression the reader is refered to Chapter 4 of Birkes and Dodge (1993).

ACKNOWLEDGMENTS

I thank David Birkes, Jean-Pierre Renfer, Valentin Rousson, Joe Whittaker, the referee, and the associate editor who contributed valuable comments to improve the article.

REFERENCES

Andrews, D. F., and Herzberg, A. M. (1985). *Data*. Springer-Verlag, New York.

Arthanari, T. S., and Dodge, Y. (1993). *Mathematical Programming in Statistics*, classical edition. Wiley, New York.

Barrodale, I., and Roberts, F. D. K. (1973). An improved algorithm for discrete L_1 linear approximation. *SIAM J. Numer. Anal.* **10** 839–848.

Birkers, D., and Dodge, Y. (1993). *Alternative Methods of Regression*. Wiley, New York.

Boscovich, R. J. (1757). De literraria expeditione per pontificiam synopsis amplioris operis, ac habentur plura eius ex exempla ria etiam sensorum impressa. *Bononiensi Scientiarum et Artium Inst. Atque Acad. Commen.* **4** 353–396.

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

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

Cook, R. D., and Weisberg, S. (1994). *An Introduction to Regression Graphics*. Wiley, New York.
Dielman, T. E., and Rose, L. E. (1995). A bootstrap approach to hypothesis testing in least absolute value regression. *Comput. Statist. Data Anal.* **20** 119–130.
Dielman, T. E., and Rose, L. E. (1996). A note on the hypothesis testing in LAV multiple regression: A small sample comparison. *Comput. Statist. Data Anal.* **21** 463–470.
Ellis, S. P., and Morgenthaler, S. (1992). Leverage and breakdown in L_1 regression. *J. Amer. Statist. Assoc.* **87** 143–148.
Farebrother, R. W. (1988). A simple recursive procedure for the L_1 norm fitting of a straight line. *Appl. Statist.* **37**, No. 3 457–489.
Gutenbrunner, G., Jureckova, J., Koenker, R., and Portnoy, S. (1993). Tests of linear hypotheses based on regression rank scores. *Nonparametric Statist.* **2** 307–331.
Hill, R. W., and Holland, P. W. (1977). Two robust alternatives to least squares regression. *J. Amer. Statist. Assoc.* **72** 828–833.
Huber, P. J. (1987). The place of the L_1-norm in robust estimation. In *Statistical Data Analysis—Based on L_1-norm and Related Methods* (Y. Dodge, Ed.), pp. 23–33. North Holland, Amsterdam.
Karst, O. J. (1958). Linear curve fitting using least deviations. *J. Amer. Statist. Assoc.* **53** 118–132.
Koenker, R., and Basset, G. W. (1982). Tests of hypotheses and L_1 estimation. *Econometrica* **50** 1577–1583.
Legendre, A. M. (1805). Nouvelles méthodes pour la détermination des orbites des comètes. In *Appendice sur la méthode des moindres carrés*, pp. 72–80. Courier, Paris.
Marazzi, A. (1993). *Algorithms, Routines, and S Functions for Robust Statistics*. Wadworth and Brook/Cole, California.
McKean, J. W., and Schrader, R. M. (1987). Least absolute errors analysis of variance. In *Statistical Data Analysis–Based on the L_1-norm and Related Methods* (Y. Dodge, Ed.), pp. 297–321. North-Holland, Amsterdam.
Osborne, M. R. (1985). *Finite Algorithms in Optimization and Data Analysis*. Wiley, New York.
Portnoy, S., and Koenker, R. (1997). The Gaussian hare and the Laplacian tortoise: Computation of squared-error vs. absolute-error Estimators. *Statistical Science*, forthcoming.
Rousseeuw, P. J., and Leroy, A. M. (1987). *Robust Regression and Outlier Detection*. Wiley, New York.
Sadowski, A. N. (1974). Algorithm AS74. L_1-norm fit of a straight line. *Appl. Statist.* **23** 244–248.

Printed in Belgium