Hacettepe Journal of Mathematics and Statistics
Volume 38 (2) (2009), 185 – 198

RESEARCH INTO MULTIPLE OUTLIERS IN LINEAR REGRESSION ANALYSIS

Barış Aşıkgil*† and Aydın Erar*

Received 23 : 11 : 2006 : Accepted 08 : 06 : 2009

Abstract

Studying the observations in regression analysis it is seen that the output of regression is affected from outliers in the direction of the dependent and / or the independent variables. In this paper multiple outliers are examined in two real data sets. The results concerned with which method can determine multiple outliers better are examined with the help of some statistics and REC curve which can be used for determining efficiency. Also, the results are tried to support by using Monte Carlo Simulation.

Keywords: Regression, Multiple outliers, Forward search, Stalactite plot, REC curve.

2000 AMS Classification: 62 J 05, 62 J 20.

1. Introduction

The aim of regression analysis is to form a suitable model providing an explanation of the relationship between the dependent (Y) and independent (X_j) variables with the help of data. To form this model using the ordinary least squares (OLS) method some assumptions have to be satisfied. Specifically, the errors must have zero mean and equal variance, and be uncorrelated. If an inference is made, they must be normally distributed. Also, there must be no complete or approximate multicollinearity between the independent variables. The model of linear regression is given in matrix notation as follows:

$$\mathbf{Y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \tag{1}$$

The observation distances matrix (projection or hat matrix) is given by

$$H = X\left(X'X\right)^{-1}X', \tag{2}$$

*Department of Statistics, Faculty of Science and Letters, Mimar Sinan F. A. University, Çirağan Caddesi, Çiğdem Sokak No: 1, Besiktas 34349, Istanbul, Turkey.
E-mail: (B. Aşıkgil) basikgil@msgsu.edu.tr (A. Erar) aydinerar@msgsu.edu.tr
†Corresponding author

and the residual is defined as

$$\mathbf{e} = \mathbf{Y} - H\mathbf{Y} = \mathbf{Y} - \widehat{\mathbf{Y}}. \tag{3}$$

For given n and k, $k' = k+1$, $\mathbf{Y}$ is the $(n \times 1)$ vector of responses, X the $(n \times k')$ matrix of independent variables, $\boldsymbol{\beta}$ the $(k' \times 1)$ vector of unknown parameters, $\boldsymbol{\varepsilon}$ the $(n \times 1)$ vector of errors, H the $(n \times n)$ hat matrix and $\mathbf{e}$ the $(n \times 1)$ vector of residuals.

2. The concept of outlier and the study of a single outlier

Studying observations in regression analysis it is often seen that the output of regression is affected by observations far in the direction of the dependent variable (outlier) and / or observations far in the direction of the independent variables (leverage) [10].

A single outlier is studied with the help of ordinary graphs and statistics. The most used graphs in the study of a single outlier are the graphs of residual vs. each independent variable or the graph of residual vs. fitted value ($\widehat{\mathbf{Y}}$). The most often used statistics in the study of a single outlier are studentized or r-studentized residuals. Moreover, the diagonal elements of the hat matrix or the Mahalanobis Distance can be used to determine leverages. The effects of these observations on fitted parameters are studied with various statistics, such as DFFITS, DFBETAS, Cook Distance, COVRATIO, etc. [7].

3. The study of multiple outliers

Almost all techniques used for determining a single outlier are based on removing an observation from the data set. But, in some cases one outlier can affect another in various ways. These effects are known as masking and swamping effects. In the presence of the masking effect, outliers are hidden by other outliers and therefore, they cannot be determined. In other words, for a case with two outliers, leaving one of them from the data set can result in the other being determined as an outlier. In the presence of the swamping effect, outliers drag the fitted line towards themselves and some other observations can be determined as outliers because they are far from the fitted line. In other words, for a case with two outliers, as a result of leaving one of them from the data set the other can be determined as a good observation. In the presence of these effects statistics used for determining single outlier cannot give true results. Therefore, various methods insensitive to these effects are used for determining multiple outliers [4].

3.1. Forward search method. The forward search method (FSM) is a method which has been developed to overcome the masking and swamping effects created by multiple outliers in a data set. Two different approaches are available in the literature for the forward search method. These are the Hadi and Simonoff Approach [8] and the Atkinson and Riani Approach [3].

3.1.1. *Hadi and Simonoff's approach.* The forward search method can be used in multiple regression with more than two independent variables. Firstly, a regression model is formed for a data set with n observations and $|e_i|$ values are calculated. The basic subset is formed with the observations having the least $k+1$ (number of parameters in the model) $|e_i|$ values. If B is the basic subset, $\widehat{\boldsymbol{\beta}}_B$ is the column vector of the fitted parameter values obtained from observations in the basic subset, and X_B is the full-ranked matrix formed by observations in the basic subset. If X_B is not the full-ranked matrix, observations are added taking into consideration the $|e_i|$ values until the matrix has full rank. The subset M containing $h = \lfloor (n+k-1)/2 \rfloor$ observations and no outliers is determined using the steps given below, where $\lfloor \cdot \rfloor$ gives the integer part of the value [8]:

- A regression model is formed with the observations in B and

$$(4)\qquad \begin{cases} \dfrac{\left|y_i - \mathbf{x}_i'\widehat{\boldsymbol{\beta}}_B\right|}{\sqrt{1-\mathbf{x}_i'\left(X_B'X_B\right)^{-1}\mathbf{x}_i}}, & i \in B, \\ \dfrac{\left|y_i - \mathbf{x}_i'\widehat{\boldsymbol{\beta}}_B\right|}{\sqrt{1+\mathbf{x}_i'\left(X_B'X_B\right)^{-1}\mathbf{x}_i}}, & i \notin B. \end{cases}$$

  When calculating the values defined above the observations are given in ascending order.
- The size of the basic subset is denoted by s. If $s = h$, the subset M has the first h observations. If $s < h$, a new basic subset is formed with the ordered $s+1$ observations and the first two steps are applied again.

The forward search method continues with the subset M of size s having no outliers, by following the steps given below [8]:

- $$(5)\qquad u_i = \begin{cases} \dfrac{y_i - \mathbf{x}_i'\widehat{\boldsymbol{\beta}}_M}{\widehat{\sigma}_M\sqrt{1-\mathbf{x}_i'\left(X_M'X_M\right)^{-1}\mathbf{x}_i}}, & i \in M, \\ \dfrac{y_i - \mathbf{x}_i'\widehat{\boldsymbol{\beta}}_M}{\widehat{\sigma}_M\sqrt{1+\mathbf{x}_i'\left(X_M'X_M\right)^{-1}\mathbf{x}_i}}, & i \notin M. \end{cases}$$

  The values defined above are calculated, where $\widehat{\boldsymbol{\beta}}_M$ is the column vector of the fitted parameter values obtained from observations in the subset M, $\widehat{\sigma}_M$ is the standard deviation obtained from observations in the subset M, and X_M is the full-ranked matrix formed by the observations in this subset.
- Observations are put in ascending order of the values $|u_i|$, and $u_{(s+1)}$ is the value $|u_i|$ in the $(s+1)$-th place. If $u_{(s+1)} \geq t_{(\alpha/2(s+1),s-k)}$, all observations for which $|u_i| \geq t_{(\alpha/2(s+1),s-k)}$ are given as outliers and the forward search method is stopped. Otherwise, a new subset M is formed from the $(s+1)$ ordered observations and these steps are applied again. If $n = s+1$ the data set does not contain any outliers.

3.1.2. *Atkinson and Riani's approach.* Atkinson and Riani [3] propose that at the beginning of the forward search the subset size be taken as $m = k'$, and that many subsets be formed with m observations in each subset. The possible number of subsets is $\binom{n}{k'}$. If this number is very large, the number of subsets are usually taken as 1000. The beginning subset is chosen to definitely not contain an outlier and to have the least median squared residual. The $n \times (k'+1)$ matrix W is given by

$$(6)\qquad W = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} & y_1 \\ 1 & x_{12} & x_{22} & \cdots & x_{2k} & y_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} & y_n \end{bmatrix}$$

and

$$(7)\qquad S^{(m)}_{i_1,\ldots,i_m} \equiv \{w_{i_1},\ldots,w_{i_m}\}$$

are sets containing m different ordered observations, where $1 \leq i_1 \leq \cdots \leq i_m \leq n$, $i_1,\ldots,i_m$ is the i-th observation in these sets, and w_{i_1} is the i_1-th row of the matrix W. Also, $\iota' = [i_1,\ldots,i_m]$ and $e_{i,S^{(m)}_\iota}$ is the residual calculated for the i-th observation in

$S_\iota^{(m)}$. The beginning subset $S_*^{(k')}$ is chosen to satisfy

$$e^2_{[\text{Med.}],S_*^{(k')}} = enk_\iota \left[e^2_{[\text{Med.}],S_\iota^{(k')}} \right] \tag{8}$$

where $e^2_{[a],S_\iota^{(k')}}$ is the a-th squared residual chosen from $e^2_{i,S_\iota^{(k')}},\ i = 1, 2, \ldots, n$ and Med. is the integer part of the number $(n + k' + 1)/2$:

$$\text{Med.} = \left\lfloor (n + k' + 1)/2 \right\rfloor. \tag{9}$$

In this approach, by adding observations to the beginning subset $S_*^{(m)}$, it follows that all squared residuals $\left(e^2_{i,S_*^{(m)}},\ i = 1, 2, \ldots, n \right)$ calculated with the help of this subset are ordered and the observation with the least squared residual is added to the subset. Although one observation is usually added to the subset, two or more observations can sometimes be added. The forward search method continues until all observations are added to the subset. This method is a union of the robust and least squares fitting methods. Therefore, if the data set has q outliers these observations enter into the subset in the last q steps. Up to this time the residual graphs and fitted parameter values are approximately the same. But, the variance does not stay the same. If there are no outliers, $s^2_{S_*^{(m)}} < s^2_{S_*^{(n)}} = s^2$ for $m < n$. As a result of an increase in the observation number the values t decrease.

In 1994 Woodruff and Rocke stated that determining multiple outliers becomes difficult as a result of an increase in the size of the problem. Therefore, a few forward searches may be needed to determine the outliers. Because of this a stalactite graph is drawn to suggest some different beginning subsets. In the graph, rows show the size of the subset and columns show the observation number. Observations having big absolute residuals (bigger than 2 and 3) are marked with symbols in the graph. Multiple outliers can be seen from the graph [2] for all sizes of a subset.

Apart from the forward search method, some robust fitting methods can be used for determining multiple outliers.

3.2. The least median squares method. The least median squares (LMS) method, which has a big breakdown point (approximately 50%), is used for finding robust fitted values of parameters and determining multiple outliers by means of decreasing the effects of residuals. This method is defined below [11]:

$$\text{LMS} = \min_{\beta} [\underset{i}{\text{med}} (y_i - \mathbf{x}_i' \boldsymbol{\beta})^2]. \tag{10}$$

3.3. Huber's method. Multiple outliers can be determined by using some robust methods based on an alternative function and weights. In 1981 Huber defined the weight,

$$w_i = \begin{cases} 1, & |e_i/\hat{\sigma}| \le t \\ \dfrac{t}{|e_i/\hat{\sigma}|}, & |e_i/\hat{\sigma}| > t \end{cases} \tag{11}$$

where $\hat{\sigma}$ is the standard deviation of the residuals and t is a constant value which was proposed as $t = 2.0$ by Montgomery and Peck [9].

3.4. Regression error characteristic curves. Regression Error Characteristic (REC) curves are used for assessing quickly the relative merits of different regression functions. These curves are drawn with the squared residuals $\left(\mathbf{Y} - \widehat{\mathbf{Y}} \right)^2$ or the absolute deviations $\left| \mathbf{Y} - \widehat{\mathbf{Y}} \right|$ on the x-axis vs. the probability values on the y-axis.

Some general properties of REC curves are given below [5]:

- REC curves enable regression functions to be compared with one another and the model in the null hypothesis.
- The area of the REC curve gives the expected performance of the regression model. Also, the cumulative distribution function (CDF) can be calculated with the help of the curve. Therefore, the area over the curve (AOC) is a biased estimation of the expected error.
- Different REC curves can drawn with squared residuals or absolute deviations. In fact, there is no change in the relative condition of the curves.
- It can be easily determined with the help of REC curves whether different regression models are similar or quite different.
- As the size of the sample goes to infinity, the AOC converges to the expected value of the error. If e is based on absolute deviation, AOC converges to the mean absolute deviaton (MAD); if e is based on squared residuals, AOC converges to the mean squared error (MSE).
- The smaller the AOC, the better the regression model.

Aspects of REC curves for some special conditions are given below [5]:

- Because they are approximately preserved under monotonous transformations REC curves can be obtained for some logarithmic transformations, square root transformations, etc.
- If the data set has outliers, the upper part of the REC curve becomes smooth and does not reach 1 until the tolerance of error increases.
- REC curves sometimes increase rapidly in the middle and become smooth at the end. This condition shows a concave behaviour. A model having a REC curve like this is probably a biased model.

In order to draw the graph the algorithm given below can be used:

```
(1) Absolute or squared residuals are calculated and ordered (e).
(2) e0 = 0; sum = 0;
(3) for i = 1:n
(4) if e(i) > e0
(5) plot(e0, sum/n);
(6) e0 = e(i);
(7) end
(8) sum = sum + 1;
(9) end
(10) plot(e(n), sum/n);
```

REC curves can be used not only for comparing regression functions but also for determining multiple outliers with the help of AOC. In the application below, AOC is taken as a measurement of efficiency and the REC curves are used for determining multiple outliers.

4. Applications on two real data sets

The data set in Asikgil [1] contains information about houses to let in Kadikoy-Centre in Istanbul between August and November of the year 2005. The data set has 76 observations, 20 of which assumed to be free of outliers are chosen randomly and used for cross-validation analysis later. The variables used in the analysis are given below:

X1: The size of the house to let (m^2)
X2: The floor number of the house to let
X3: The deposit paid to the landlord (TL)

S: Type of heating of the house to let (stove, stove using natural gas, central heating, combined boiler)
M_B1: Kitchen and bathroom of the house to let are modern and well-kept or not
B1: The house to let is near to the sea or not
Y: Rent of the house (TL)

The data set was analyzed with SPSS11.5. It was seen that the assumptions are best satisfied for the logarithmic transformation of the variable Y.

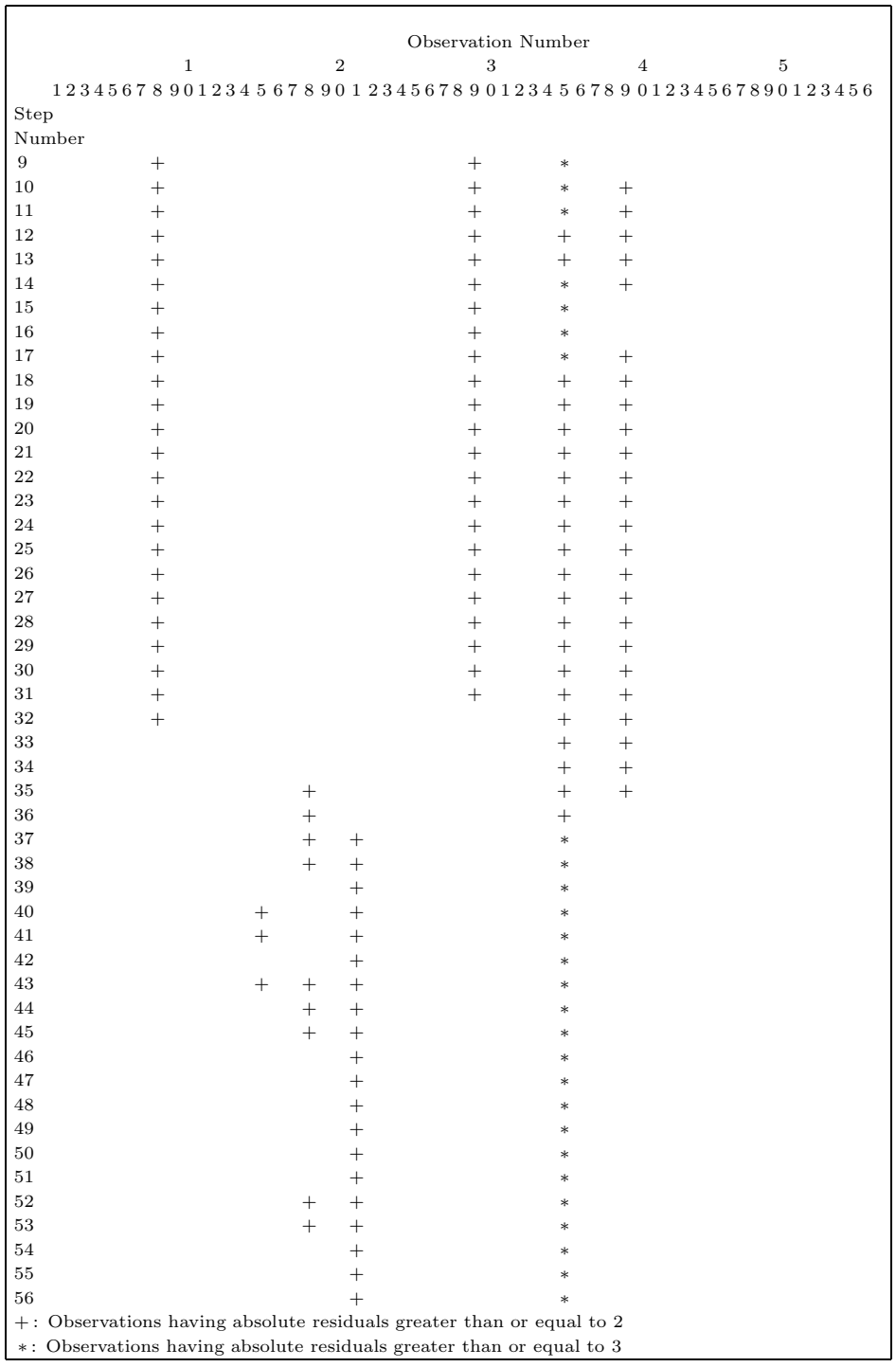
Some groups containing suspicious observations were examined with DFFITS, COVRATIO and the statistics of Tatlidil [12], but a definite result could not be obtained because of masking and swamping effects [1]. The "fwd" library of S-PLUS2000 was then used for determining multiple outliers in the data set. Using Atkinson and Riani's approach, the steps showing the entry of observations to the beginning subset were as given in Table 1.

Table 1. Entry of Observations to the Subset

Step Number (m)	Obs. Number	Step Number (m)	Obs. Number
10	50	34	32
11	42	35	5
12	51	36	7
13	10	37	49
14	4	38	54
15	37	39	22
16	45	40	9
17	55	41	16
18	1	42	27
19	34	43	20
20	53	44	33
21	26	45	24
22	56	46	52
23	25	47	29
24	48	48	46
25	40	49	17
26	19	50	39
27	23	51	11
28	28	52	8
29	38	53	15
30	47	54	18
31	12	55	21
32	31	56	35
33	14		

Using the output from S-PLUS2000 the stalactite plot based on the beginning subset is as given in Figure 1. It can be said that observations 8, 15, 18, 21, 29, 35 and 39 are outliers because of their large residuals compared with the other observations.

Figure 1. Stalactite Plot for the Forward Search Method



Combinations of suspicious observations obtained from LMS, FSM and those common to the two methods are described below:

$$\text{Research into Multiple Outliers} \begin{cases} \text{LMS} \begin{cases} (A)\ \text{Observations } 17, 18, 21, 35, 39 \end{cases} \\ \text{FSM} \begin{cases} (B)\ \text{Observations } 8, 21, 35, 39 \\ (C)\ \text{Observations } 8, 21, 29, 35, 39 \\ (D)\ \text{Observations } 8, 15, 18, 21, 29, 35, 39 \end{cases} \\ \text{Common} \begin{cases} (E)\ \text{Observations } 21, 35, 39 \end{cases} \end{cases}$$

By removing these combinations of observations from the data set the changes in some statistics were examined. The 20 observations chosen randomly at the beginning were used for calculating PRESS (Prediction Error Sum of Squares) and MATLAB7.1 was used for drawing the REC curves. All of the results are given in Table 2 and the graph of the REC curves is given in Figure 2.

Table 2. Examination of Multiple Outliers

MODELS	$\mathbf{R^2}$	$\widehat{\boldsymbol{\sigma}}$	**PRESS**	**AOC**
ENTIRE SET OLS	0.861	0.132	0.189	0.0134
A	0.916	0.094	0.165	0.0071
B	0.906	0.099	0.379	0.0077
C	0.892	0.099	0.206	0.0077
D	0.920	0.088	0.194	0.0060
E	0.896	0.103	0.183	0.0084
HUBER	0.886	0.119	0.180	0.0132

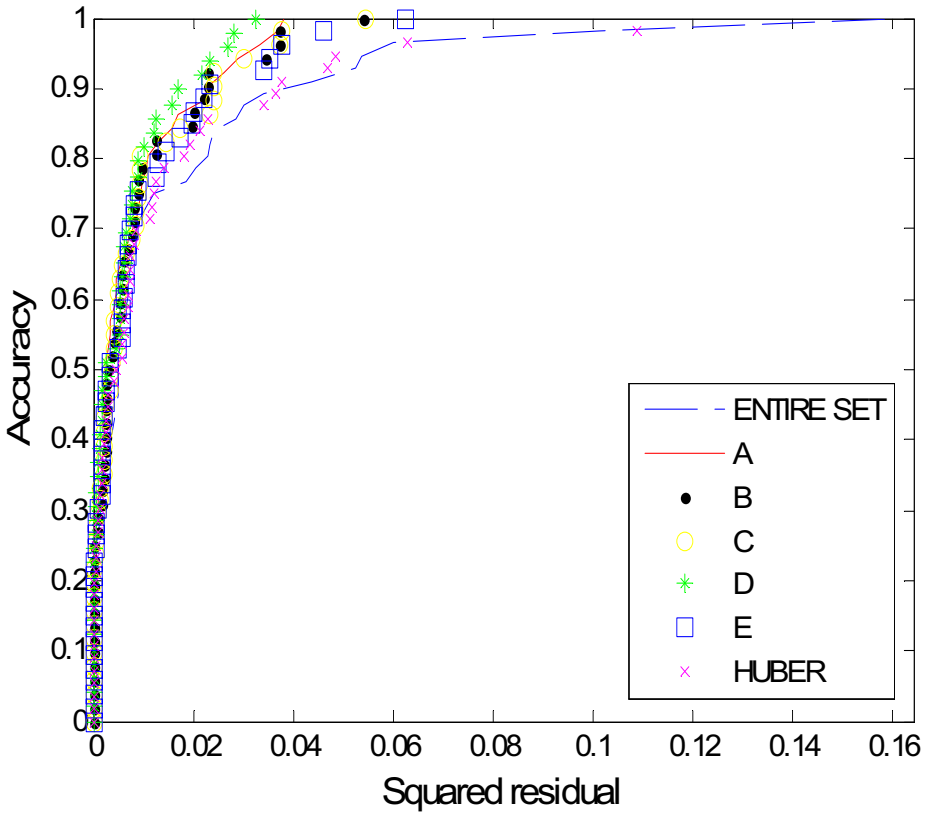
It can be seen from the REC curves in Figure 2 that if observations in combination (D) are removed from the data set, AOC becomes smaller than others. Finally, the following results are obtained regarding the examination:

(1) By taking into consideration the coefficient of determination, standart deviation and AOC in Table 2 it can be seen that the model D is the best. But, by taking into consideration PRESS obtained from cross-validation the model A becomes the best. To sum up, since the most three important statistics indicate the model D the observations in the combination (D) can be the true multiple outliers.

(2) The observations in the combination (A) are obtained from LMS and the observations in the combination (D) are obtained from the stalactite plot in FSM.

(3) Therefore, for this data set it can be said that the stalactite plot in FSM gives better results than robust methods in determining multiple outliers.

Figure 2. The Plot of REC Curves for Various Models



We now make a similar examination of the air pollution data set given in Candan [6]. The stalactite plot based on the beginning subset is given in Figure 3. The steps showing the entry of observations to the beginning subset are not given here.

Combinations of suspicious observations obtained from LMS, FSM and those common to the two methods are described below:

Research into Multiple Outliers

- LMS
  - (A) Observations $2, 8, 9, 24, 26, 29, 33$
- FSM
  - (B) Observations $26, 28, 29, 33, 34, 42$
  - (C) Observations $25, 26, 28, 29, 33, 34, 42$
  - (D) Observations $4, 24, 25, 26, 28, 29, 33, 34, 37, 42$
- Common
  - (E) Observations $26, 29, 33$

By removing the above combinations of observations from the data set the changes in some statistics can be examined in Table 3 and also in Figure 4.

Figure 3. Stalactite Plot for the Forward Search Method

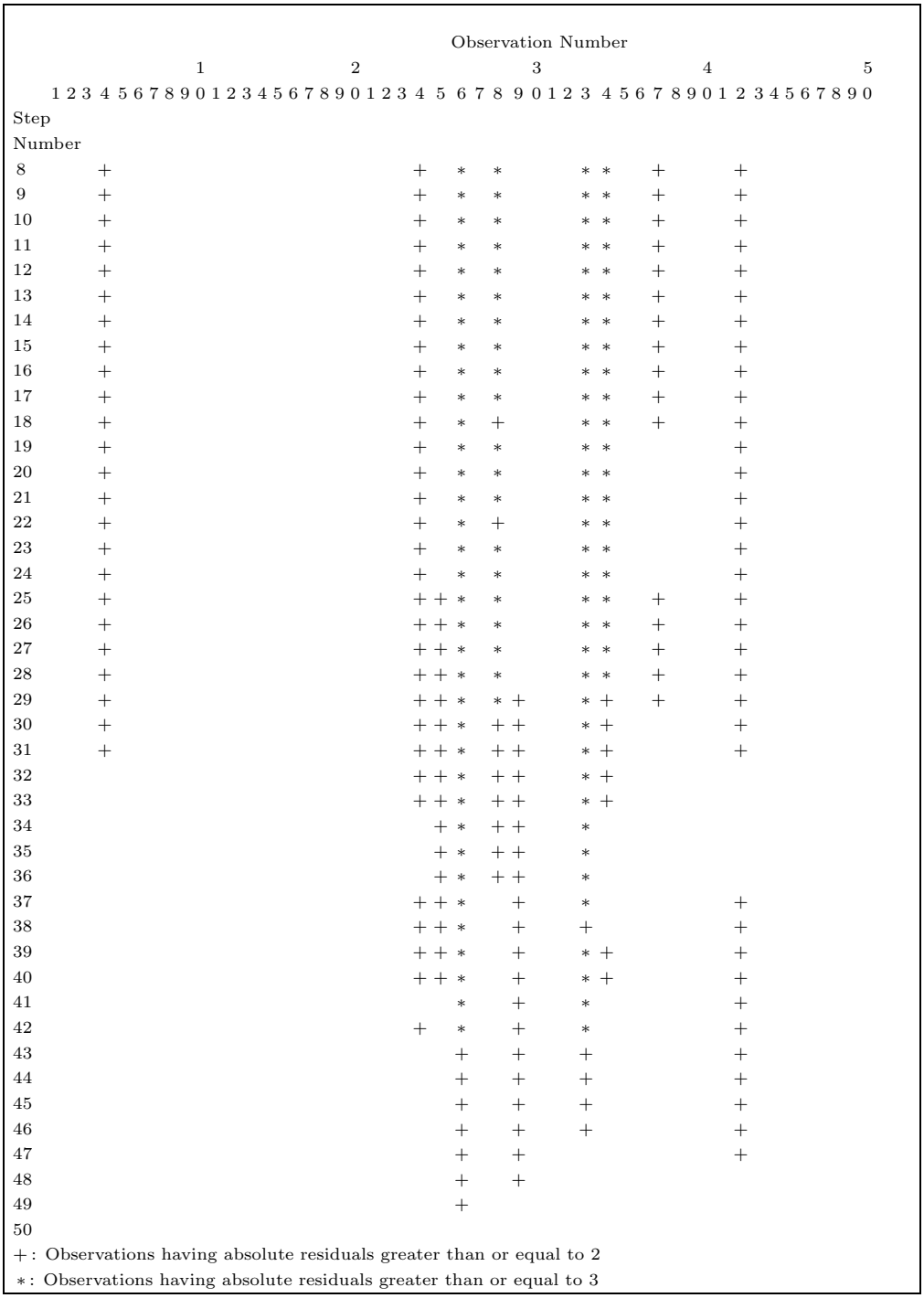
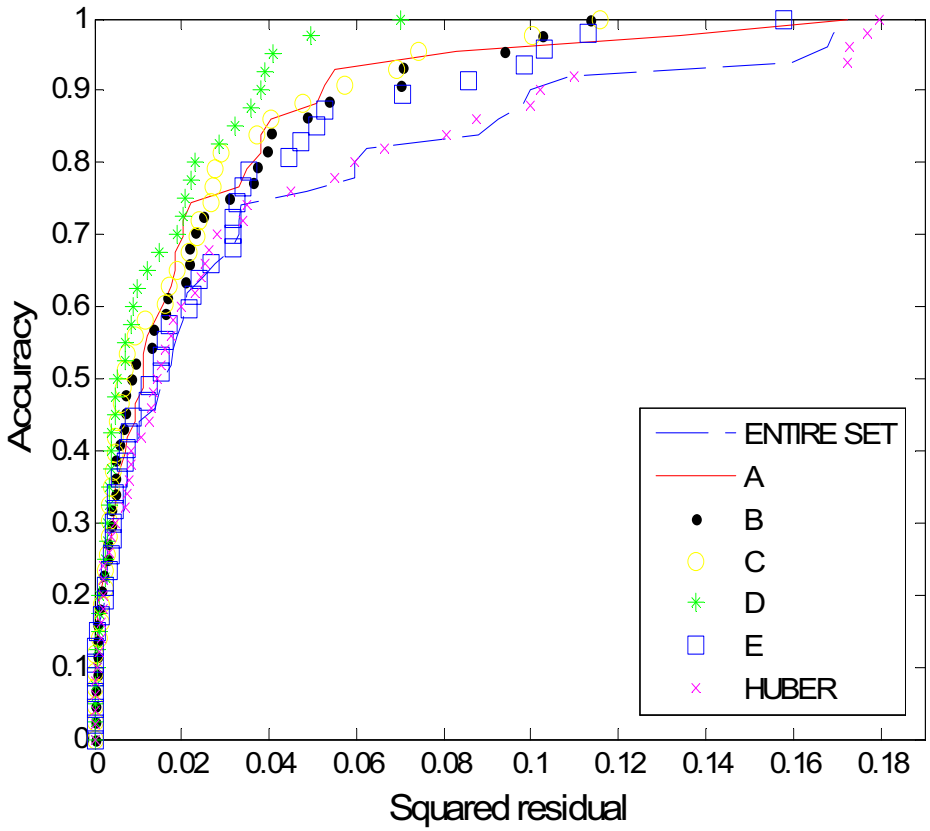


Table 3. Examination of Multiple Outliers

MODELS	R^2	$\widehat{\sigma}$	PRESS	AOC
ENTIRE SET OLS	0.410	0.207	0.456	0.0346
A	0.631	0.168	0.611	0.0211
B	0.596	0.165	0.456	0.0212
C	0.634	0.157	0.532	0.0188
D	0.719	0.132	0.753	0.0131
E	0.541	0.180	0.443	0.0255
HUBER	0.339	0.219	0.400	0.0346

Figure 4. The Plot of REC Curves for Various Models



It can be said that observations 4, 24, 25, 26, 28, 29, 33, 34, 37 and 42 are outliers by taking into consideration Table 3 and Figure 4. Finally, it can be seen that similar results concerned with which method can determine multiple outliers better are obtained for both of the data sets in [1] and [6].

5. Monte Carlo simulation

In this section a simulation study is considered in order to support our result on which method can determine multiple outliers better. The conditions under which the simulation was performed are:

(1) The multiple linear regression model is $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$,
(2) The first independent variable is generated from $N(3, 1)$ and the second independent variable is generated from $N(2, 1)$,
(3) The parameter vector is $\boldsymbol{\beta}' = [0.5 \quad 1.0 \quad 2.0]$,
(4) The error terms are independent and identically distributed from $N(0, 1)$,
(5) The sample size is determined as $n = 20$ and 50,
(6) The percentage of outliers for each sample size is determined as 10% and 20%,
(7) Specific observations are formed as outliers which lie approximately 2σ away from the data set for each sample size,
(8) The iteration number for each sample size is 1000.

Under these conditions, the simulation results are given in Table 4 and Table 5 for the sample size $n = 20$, and in Table 6 and Table 7 for the sample size $n = 50$. All observations are given in Table 4 to see the cases clearly. Only outliers are given in the other tables. All these tables present the percentages for determining different outliers by using Huber, LMS and FSM.

Table 4. Simulation result for $n = 20$ and 10% outlier

	Methods		
Observation Number	**Huber**	**LMS**	**FSM**
1	1	5	3
2	0	2	3
3	0	1	4
4	0	1	3
5	3	6	7
6	0	3	4
7	0	5	2
8	0	1	2
9	0	2	4
10	0	2	3
11	1	2	4
12	0	3	4
13	76	80	87
14	0	4	6
15	0	1	4
16	65	80	85
17	0	1	6
18	0	3	3
19	0	2	2
20	0	0	1

Table 5. Simulation result for $n = 20$ and 20% outlier

	Methods		
Observation Number	**Huber**	**LMS**	**FSM**
5	65	82	88
11	66	84	88
13	67	79	85
16	63	80	82

Table 6. Simulation result for $n = 50$ and 10% outlier

	Methods		
Observation Number	**Huber**	**LMS**	**FSM**
7	89	87	91
11	86	90	92
23	91	90	91
35	82	85	91
44	86	85	92

Table 7. Simulation result for $n = 50$ and 20% outlier

	Methods		
Observation Number	**Huber**	**LMS**	**FSM**
2	71	89	88
7	67	80	85
11	64	80	83
15	66	82	82
19	67	82	83
23	64	86	85
28	67	90	91
35	60	82	84
44	68	78	82
48	73	84	88

After the simulation study it can be said that:

(1) Taking Table 4 and Table 5 (not given in full) into consideration, we see that FSM tends to determine some non-outlier observations as outliers for small sample sizes such as $n = 20$.
(2) Taking Table 6 and Table 7 into consideration, we see that for big sample sizes such as $n = 50$ LMS tends to approach FSM in its ability to determining outliers,
(3) Generally, FSM has a better chance than the other methods in determining multiple outliers.

6. Conclusion

In this paper, we have tried to determine multiple outliers by using various methods in the presence of masking and swamping effects. We have discussed which of these methods can determine all multiple outliers, and it has been proposed that FSM gives better results than the other methods in determining multiple outliers.

To sum up, it can be said that the stalactite plot obtained from the output of FSM is especially useful in determining multiple outliers.

References

[1] Asikgil, B. *The Examination of Outliers and Influential Observations in Multiple Linear Regression Analysis and an Application* (Msc. Thesis, Mimar Sinan University, Istanbul, 2006).

[2] Atkinson, A. C. *Fast very robust methods for the detection of multiple outliers*, Journal of the American Statistical Association **89**, 1329–1339, 1994.

[3] Atkinson, A. C. and Riani, M. *Robust Diagnostic Regression Analysis* (Springer, New York, 2000).

[4] Barnett, V. and Lewis, T. *Outliers In Statistical Data* (Third Edition, John Wiley&Sons Ltd., Chichester, 1994).

[5] Bi, J. and Bennett, K. P. *Regression Error Characteristic Curves* (Proceedings of the Twentieth International Conference on Machine Learning (ICML), Washington, 2003).

[6] Candan, M. *Robust Estimators in Linear Regression Analysis* (Msc. Thesis, Hacettepe University, Ankara, 1995).

[7] Chatterjee, S. and Hadi, A. S. *Influential observations, high leverage points and outliers in linear regression*, Statistical Science **1**, 379–416, 1986.

[8] Hadi, A. S. and Simonoff, J. S. *Procedures for the identification of multiple outliers in linear models*, Journal of the American Statistical Association **88**, 1264–1272, 1993.

[9] Lawrence, D. K. and Arthur, J. L. *Robust Regression; Analysis and Applications* (Marcel Dekker, Inc., New York and Basel, 1990).

[10] Neter, J., Kutner, M. H., Nachtsheim, C. J. and Wasserman, W. *Applied Linear Statistical Models* (Fourth Edition, The McGraw-Hill Companies Inc., 1996).

[11] Rousseeuw, P. J. and Leroy, A. M. *Robust Regression and Outlier Detection* (John Wiley&Sons Inc., 2003).

[12] Tatlidil, H. *Test of Suspicious Observations in Linear Regression Analysis and Multivariate Data* (PhD. Thesis, Hacettepe University, Ankara, 1981).