Journal of Korean
Data & Information Science Society
2005, Vol. 16, No. 3, pp. 665~682

Bootstrapping Regression Residuals

A.H.M. Rahmatullah Imon[1)] · M. Masoom Ali[2)]

Abstract

The sample reuse bootstrap technique has been successful to attract both applied and theoretical statisticians since its origination. In recent years a good deal of attention has been focused on the applications of bootstrap methods in regression analysis. It is easier but more accurate computation methods heavily depend on high-speed computers and warrant tough mathematical justification for their validity. It is now evident that the presence of multiple unusual observations could make a great deal of damage to the inferential procedure. We suspect that bootstrap methods may not be free from this problem. We at first present few examples in favour of our suspicion and propose a new method diagnostic-before-bootstrap method for regression purpose. The usefulness of our newly proposed method is investigated through few well-known examples and a Monte Carlo simulation under a variety of error and leverage structures.

Keywords : Diagnostics, Diagnostic-before-bootstrap, Fixed-X resampling, High leverage points, Jackknife-after-bootstrap, Monte Carlo simulation, Outliers, Random-X resampling

1. Introduction

Bootstrap technique proposed by Efron (1979) is such a procedure which creates a huge number of sub-samples from a pre-observed data set by a simple random sampling with replacement. These sub-samples could be later used to investigate the nature of the population without having any assumption about them. In recent years the application of bootstrap methods has become widespread (see Fox 1991,

1) First Author : Professor, Department of Statistics, University of Rajshahi, Rajshahi-6205, Bangladesh.
E-mail : imon_ru@yahoo.com

2) George and Frances Ball Distinguished Professor of Statistics, Department of Mathematical Sciences, Ball State University, Muncie, IN 47306-0490 USA.
E-msil : mali@bsu.edu.

Efron and Tibshirani 1993, Shao and Tu 1995, Davison and Hinkley 1997, Venables and Ripley 2000). This computer-based technique is used mainly for estimating standard error, bias, confidence interval and other statistical measures. In this paper we consider its application in regression analysis. For linear regression with normal random errors, the ordinary least squares (OLS) theory of regression estimation and inference are traditionally used mainly because of tradition and ease of computation. It is now evident (see Ryan 1997) that the OLS theory suffers a huge set back in the presence of unusual observations such as outliers. But in a real data the presence of 1-10% outliers are rather rule than exception (see Hampel et al. 1986). In linear regression, non-normality of errors often occurs in the presence of outliers in the data. We anticipate, however, that the bootstrap methods have the potential to provide more accurate analysis in a similar situation.

Bootstrap may be a very useful technique indeed but caution must be taken while considering this technique. To quote Efron (1992), *'it (Bootstrap) automatically produces accuracy estimates in almost any situation, including very complicated ones, without requiring much thought from the statistician. This is a considerable virtue, but a virtue that can be abused. The danger lies in the possibility that the bootstrap estimates of accuracy, so easily produced, might be accepted uncritically. This reinforces the truism the bootstrap data, like real data, deserve a thorough examination.'* In general, diagnostic methods are designed to find problems with assumptions in an analysis and these methods are being used extensively in all branches of regression analysis. But it is surprising that diagnostics in bootstrapping is not a much focused issue. However, Efron (1992) led the way when he introduced jackknife-after bootstrap and infinitesimal jackknife for estimating standard error and jackknife influence function for assessing the role of a single observation in the estimation of standard error for bootstrapping. The approaches he advocated can easily come under the broad heading diagnostics. A good number of bootstrap techniques are now available in the literature for estimating regression parameters (see Freedman 1981, Wu 1986, Ryan 1997). Imon and Das (2005) point out that neither of these techniques is entirely satisfactory when multiple outliers are present in the data. In Section 2 we introduce different commonly used bootstrap techniques suitable for regression analysis. We propose a new method diagnostic-before-bootstrap in Section 3 to handle situations when a group of outliers are present in the data. We apply this newly proposed technique to several well-known data sets to investigate its performance in Section 4. The usefulness of this technique is investigated and its performance is compared with different bootstrap techniques in Section 5 under a variety of situations.

2. Bootstrapping Regression Models and Residuals

Regression analysis is a statistical technique for investigating and modeling the relationship between variables. Application of regression is numerous and occurs in almost every branch of knowledge. In our study we are concerned with linear regression model

$$Y = X\beta + \epsilon \tag{2.1}$$

where $Y = (y_1, y_2, \cdots, y_n)^T$, $X = (x_1, x_2, \cdots, x_n)^T$, $x_i^T = (x_{i1}, x_{i2}, \cdots, x_{ip})$, and $\epsilon = (\epsilon_1, \epsilon_2, \cdots, \epsilon_n)^T$. In this equation β is a $k\times 1$ vector of unknown parameters to be estimated from the data, Y is an $n\times 1$ vector, X is an $n\times k$ data matrix of full rank $k\times n$ and ϵ is an $n\times 1$ vector of unobservable random errors with $E(\epsilon) = 0$ and $V(\epsilon) = \sigma^2 I$. For linear regression with normal random errors, the OLS theory of regression estimation and inference provides very satisfactory results. The least squares estimate of β is given by the formula $\hat{\beta} = (X^T X)^{-1} X^T Y$. Here we introduce few methods of bootstrapping regression models.

2. 1 Fixed-X resampling

The most popular way of bootstrapping in linear regression is to treat the fitted values from the model as giving the expectation of the response for the bootstrap samples. We could estimate a set of residuals by the OLS method and then can generate bootstrap residuals by random sub-sampling from that set. In other words, a distribution of bootstrap residuals can be obtained by bootstrapping OLS residuals. Attaching a random error to each $\hat{y}_i$ produces a fixed-X bootstrap samples, $y_r^* = \{y_{ri}^*\}$. The errors could be generated parametrically from a normal distribution with mean 0 and variance , if we are willing to assume that the errors are normally distributed or non-parametrically, by resampling residuals from the original regression. We would then regress the bootstrapped values y_r^* on the fixed X matrix to obtain bootstrap replications of regression coefficients. For a simulated set of errors $(\epsilon_1, \epsilon_2, \cdots, \epsilon_n)$, let $\varepsilon_r^* = (\varepsilon_{r1}^*, \varepsilon_{r2}^*, \cdots, \varepsilon_{rn}^*)$ be the rth bootstrap errors, where r = 1, 2,..., B, and B is the no. of bootstrap replications. Then the bootstrap responses $y_r^* = \{y_{ri}^*\}$ are generated by

$$y_{ri}^* = x_i^T \hat{\beta}^* + \varepsilon_{ri}^* \tag{2.2}$$

which is called model based resampling of linear regression model. This method also have some other names such as fixed X-resampling, bootstrapping the residuals of linear regression model or bootstrap 1 method of linear regression model. To obtain the bootstrap least squares (BLS) estimate we compute

$$\hat{\beta}^*_r = (X^T X)^{-1} X^T y^*_r \qquad r = 1, 2, \ldots, B \tag{2.3}$$

Thus the bootstrap 1 estimate of β is given by

$$\hat{\beta}_{(boot1)} = \frac{1}{B} \sum_{r=1}^{B} \hat{\beta}^*_r \ . \tag{2.4}$$

2.2 Random-X resampling

In random-X resampling or bootstrap 2 method, we follow a completely different approach. Assume that we want to fit a regression model with response variable Y and predictors $X_1, X_2, \cdots, X_p$. We have the pairs of observations $(y_i, x_{i1}, x_{i2}, \cdots, x_{ip})$, i = 1,2,..., n. We simply select B samples of pairs, fit the model and save the coefficients of each bootstrap sample. The resampling simulation therefore involves sampling pairs with replacement from $(x_1, y_1), \cdots, (x_n, y_n)$. This is equivalent to taking $\{x^*_{mi}, y^*_{mi}\} = \{x^*_m, y^*_m\}$, where $x^*_m = \{x^*_{mi}\}$, $y^*_{mi} = \{y^*_{mi}\}$; i = 1,2,..., n; m = 1,2,..., B. Let $\{x^*_m, y^*_m\}$ be independent, with uniformly distributed. Thus the $\hat{\beta}^*_m$ be the least squares estimate based on the resample

$$\hat{\beta}^*_m = \{\sum x^*_m x^{*T}_m\}^{-1} \sum x^*_m y^*_m \tag{2.5}$$

which is called bootstrapping pairs or case resampling or bootstrap 2 methods of linear regression estimate. Hence the bootstrap 2 estimate of β is given by

$$\hat{\beta}_{(Boot2)} = \frac{1}{B} \sum_{m=1}^{B} \hat{\beta}^*_m \ . \tag{2.6}$$

2.3 Jackknife

The generalized versions of well-known jackknife technique have also been proposed for estimating the distribution of estimate, as alternatives to the bootstrap. The jackknife predates the bootstrap and similarities to it. Here we would like to extend this idea to estimate regression coefficients and to estimate true errors. Let $\hat{\beta}$ be the usual OLS estimator of regression coefficients β

and $\hat{\beta}^{(-i)}$ be the corresponding estimate with i-th case deleted. Then the jackknife estimate of regression coefficients β is by definition

$$\hat{\beta}^{(Jack)} = n\hat{\beta} - \frac{n-1}{n}\sum_{i=1}^{n}\hat{\beta}^{(-i)}, \qquad i=1,\ 2,\ \cdots,\ n \tag{2.7}$$

2.4 Jackknife-after-bootstrap

Efron (1992) first considered the need of diagnostics in bootstrapping. He introduced the jackknife-after-bootstrap method to investigate the effect of a single observation in bootstrap. Let we have drawn B bootstrap samples and we want to estimate regression coefficients or residuals using the model (2.2). The jackknife-after-bootstrap method provides a way of estimating $\hat{\beta}_B$ using only paired bootstrap (Bootstrap 2) samples. Suppose we had a large computer and set out to calculate the jackknife estimate of $\hat{\beta}_B$. For i = 1,2,..., n, we should leave out data point i to recompute $\hat{\beta}_B$ and call the result $\hat{\beta}_{B(i)}$. Then the jackknife-after-bootstrap (Jack-Boot) residuals are defined as

$$\hat{\epsilon}_{B(i)} = y_i - x_i^T\hat{\beta}_{B(i)}, \qquad i = 1,2,\ldots,\ n \tag{2.8}$$

The main difficultly with this procedure is to compute $\hat{\beta}_{B(i)}$, which requires a completely new set of bootstrap samples for each i. Fortunately for each data point i, there are some bootstrap (Bootstrap 2) samples in which the data, does not appear, and we can use those samples to estimate $\hat{\beta}_{B(i)}$. In particular we estimate $\hat{\beta}_{B(i)}$ by the sample standard deviation $S(X^{*b}, y^{*b})$, b = 1, 2, ..., B, over the pairs of sample (X^{*b}, y^{*b}), that does not contain point i. Formally, if we let C_i denoted the indices of the bootstrap samples that does not contain point i and there are B_i such samples, then

$$\hat{\beta}_{B(i)} = \sum_{b\in c_i} S(X^{*b}, y^{*b})/B_i \tag{2.9}$$

Finally we have to consider how large should we take the number of bootstrap replications for regression analysis. There is no general agreement among the statisticians (see Efron 1987, Hall 1992, Efron and Tibshirani 1993, Booth and Sarker 1998) about the number of replications needed in bootstrap. Bootstrap replications may depend on the value of X and the complexity of the function we are dealing with. Even a small number of bootstrap replications, say B = 25, is usually informative. For estimating a standard error, the number B will ordinarily be in the range 25-250. But much bigger values of B (say 500-10000) are required for bootstrap confidence interval, regression analysis, cross-validation, randomization tests and permutation tests.

3. Diagnostics Before Bootstrap

In the previous section we introduce different bootstrap techniques for the estimation of regression errors. But we know that the classical inference procedure may often go in the wrong direction in the presence of unusual observations in the data set. We know that the traditional estimators, which form the basis of analysis could be extremely sensitive to outliers and to a greater extent when outliers are also the points of high leverages. Outliers are extreme observations that for one reason or another do not belong with the other observations in the data. In the framework of linear regression, we define an outlier to be an observation for which the fitted residual is large in magnitude compared to the other observations in the data set, that is, observations are judged as outliers on the basis of how unsuccessfully the fitted regression equation is in accommodating them and that is why observations corresponding to excessively large residuals are termed as outliers. High Leverage Points *'are those observations for which the input vector* x_i *is, in some sense, far from the rest of data'* (Hocking and Pendleton 1983). Observations that are isolated in the X space will have high leverages. If the OLS regression estimators are routinely applied to data, which contain a few wild observations, then the obtained estimates can be seriously misleading. It is therefore critically important to investigate the data for the presence of outliers and high leverage points whenever the OLS regression procedures are used. We also suspect that not all bootstrap techniques are equally efficient to handle this problem.

In this paper, we propose a new way of bootstrapping in linear regression where suspect outliers are identified and omitted from the analysis before performing bootstrap with the remaining set of observations. The bootstrap estimates of parameters will involve only good observations and for this reason they will not be affected by outliers. We could, however, fit the model for the entire data set and residuals for all observations could be estimated. We shall call this technique as diagnostic-before-bootstrap (Diag-Boot). An excellent review of identification of multiple outliers is available in Barnett and Lewis (1994), Ryan (1997), and Sengupta and Jammalmadaka (2003). For the identification of suspect outliers we would use the robust reweighted least squares (RLS) residuals proposed by Rousseeuw and Leroy (1987). To compute the RLS residuals, a regression line is fitted without the observations identified as outliers by the least median of squares (LMS) technique proposed by Rousseeuw (1984). The entire set of RLS residuals can be computed using the program PROGRESS developed by Rousseeuw and Leroy (1987) and that is also available in MINITAB and S-PLUS.

Let us denote a set of cases *remaining* in the analysis by R and a set of cases *deleted* by D. Let us also suppose that R contains $(n-d)$ cases after $d<(n-k)$ cases in D are deleted. Without loss of generality, assume that these observations are the last of d rows of X and Y so that they can be partitioned as

$$X=\begin{bmatrix} X_R \\ X_D \end{bmatrix}, \qquad Y=\begin{bmatrix} Y_R \\ Y_D \end{bmatrix}$$

Then the vector of estimated parameters after the deletion of d observations, denoted by $\hat{\beta}^{(-D)}$, is obtained as

$$\hat{\beta}^{(-D)}=(X_R^T X_R^T)^{-1} X_R^T Y_R \tag{3.1}$$

Thus an $n\times 1$ vector of deletion residuals can be defined as

$$\hat{\epsilon}^{(-D)} = Y - x\ \hat{\beta}^{(-D)} \tag{3.2}$$

From this the j-th deletion residual is defined by

$$\hat{\epsilon}_j^{(-D)} = Y_j - x_j^T \hat{\beta}^{(-D)} \qquad , j = 1,2,\dots,n \tag{3.3}$$

As Imon and Das (2005) point out that bootstrap 1 method performs best overall among all existing bootstrap techniques, we would like to consider it after the omission of suspect outliers. Let the deletion bootstrap responses $y_r^{*(-D)}=\{y_{ri}^{*(-D)}\}$ are generated by

$$y_{ri}^{*(-D)}= x_{Ri}^T \hat{\beta}^{*(-D)} + \varepsilon_{ri}^{*(-D)} \tag{3.4}$$

For each replication we compute

$$\hat{\beta}_r^{*(-D)}=(X_R^T X_R)^{-1} X_R^T Y_{Rr}^{*}\ , \qquad r = 1,2,\dots,B \tag{3.5}$$

Thus the diagnostic-before-bootstrap estimate of β is given by

$$\hat{\beta}^{*(-D)}=\frac{1}{B}\sum_{r=1}^{B}\hat{\beta}_r^{*(-D)}\ . \tag{3.6}$$

Then the Y's are fitted on the X's and the diag-boot residuals are computed by the equation

$$\hat{\epsilon}^{*(-D)} = Y - X\hat{\beta}^{*(-D)} \tag{3.7}$$

4. Examples

We now consider several well-known data sets that are frequently used in the study of the identification of outliers. We would like to compare the performance of our newly proposed diag-boot method with all other bootstrap methods together with the traditional OLS method. For both real life data and simulated data, we have used robust reweighted least squares (RLS) residuals proposed by Rousseeuw and Leroy (1987) for the identification of outliers before bootstrapping.

4.1 Hawkins-Bradu-Kass (1984) data

Hawkins, Bradu and Kass (1984) constructed an artificial three-predictor data set containing 75 observations with 10 outliers. Most of the traditional methods fail to focus on outliers and they produce a poor set of results. Although this is an artificial data we do not know what were the exact values of the true errors. All we know that the true errors corresponding to true errors were set nearly at 10. It has been reported by many authors (see Rousseeuw and Leroy 1987) and we observe from Table 1 that the robust RLS can produce a set of results, perhaps very close to the true ones. That is why we clearly identify all 10 outliers and we observe that their values are also very close to 10 each. But the unfortunate consequence of using OLS method is clearly visible here. All the 10 outliers get masked here, but three good observations (cases 11-13) are swamped in as outliers.

We now apply our newly proposed method for this data. It is worth mentioning that to compute results for bootstrap 1, bootstrap 2, jackknife-after-bootstrap and diagnostic-before-bootstrap we used three types of replications, i.e. B = 500, 1000 and 5000. All of these computations are done using S-PLUS 2000. We observe a little change in the results for different number of replications. For brevity we only present results of different bootstrap residuals based on 5000 replications and these results together with the OLS and jackknife residuals are presented in Table 1.

Table 1. Bootstrap residuals for Hawkins et al. (1984) data

	OLS	Jack	Boot1	Boot2	Jack-Boot	Diag-Boot	RLS
1	3.38039	3.39690	8.42045	3.22236	3.22011	9.7572	9.7386
2	3.99499	4.00813	8.81420	3.75519	3.75367	10.1601	10.1825
3	3.00259	3.01634	9.00975	2.94191	2.93874	10.3554	10.4053
4	2.56051	2.57028	8.20288	2.37546	2.37338	9.5595	9.6547
5	3.06102	3.07404	8.70995	2.89979	2.89744	10.0578	10.1071
6	3.43559	3.45553	8.72679	3.26571	3.26331	10.0558	9.9962
7	4.51295	4.53246	9.52722	4.27659	4.27475	10.8574	10.7955
8	3.83655	3.85042	9.01400	3.69731	3.69497	10.3577	10.3807
9	2.70905	2.71784	8.30094	2.56314	2.56086	9.6592	9.7668
10	3.03851	3.04468	8.59530	2.94538	2.94287	9.9593	10.1030
11	−7.83125	−7.81875	−1.49433	−8.06546	−8.06761	−0.1415	−0.0641
12	−9.37166	−9.36102	−1.70492	−9.36711	−9.37127	−0.3465	−0.2022
13	−6.11800	−6.09218	−0.56126	−6.57383	−6.57470	0.7582	0.6231
14	−3.80242	−3.76506	−1.12151	−5.11042	−5.10505	0.1721	−0.2147
15	−0.66050	−0.64647	−1.65679	−0.68080	−0.68140	−0.3372	−0.5034
16	0.86708	0.88290	−0.65040	0.81597	0.81571	0.6633	0.4559
17	0.64616	0.64218	−1.49082	0.54558	0.54659	−0.1277	−0.0730
18	−0.39398	−0.38918	−1.27765	−0.32397	−0.32487	0.0646	0.0328
19	0.65305	0.65277	−1.18408	0.51850	0.51947	0.1712	0.1823
20	0.34019	0.35322	−0.85797	0.26366	0.26353	0.4644	0.3062
21	0.67333	0.68155	−0.36935	0.68709	0.68649	0.9644	0.8793
22	0.93408	0.93119	−0.98938	0.77058	0.77186	0.3728	0.4184
23	−0.42803	−0.42008	−2.06704	−0.50127	−0.50106	−0.7332	−0.8327
24	1.35440	1.35897	−0.57305	1.24193	1.24263	0.7691	0.7087
25	−0.30497	−0.30699	−1.59075	−0.19760	−0.19834	−0.2330	−0.1803
...	....	....	....	....	....	....	....
71	0.01603	0.02135	0.00596	0.00596	0.00564	0.2660	0.2205
72	0.13812	0.13494	0.08367	0.08367	0.08414	−0.1294	−0.0707
73	0.44118	0.43352	0.44301	0.44301	0.44316	0.4710	0.6019
74	−0.38979	−0.39645	−0.48004	−0.48004	−0.47914	−0.8272	−0.7247
75	−0.34707	−0.35873	−0.19182	−0.19182	−0.19275	0.2692	0.4744
RSD	13.943	13.89	2.826	14.663	14.678	0.018	

We observe from this table that all 10 outliers are clearly identified by the newly proposed diagnostic-before-bootstrap method. These residuals are also very close to the RLS residuals. We also observe from this table that like the OLS residuals, bootstrap 2, jackknife and jackknife-after-bootstrap methods fail to focus on any of the genuine outliers but swamp in few good observations. As it is expected, bootstrap 1 method performs better than other bootstrap methods, but diag-boot residuals perform better than bootstrap 1.

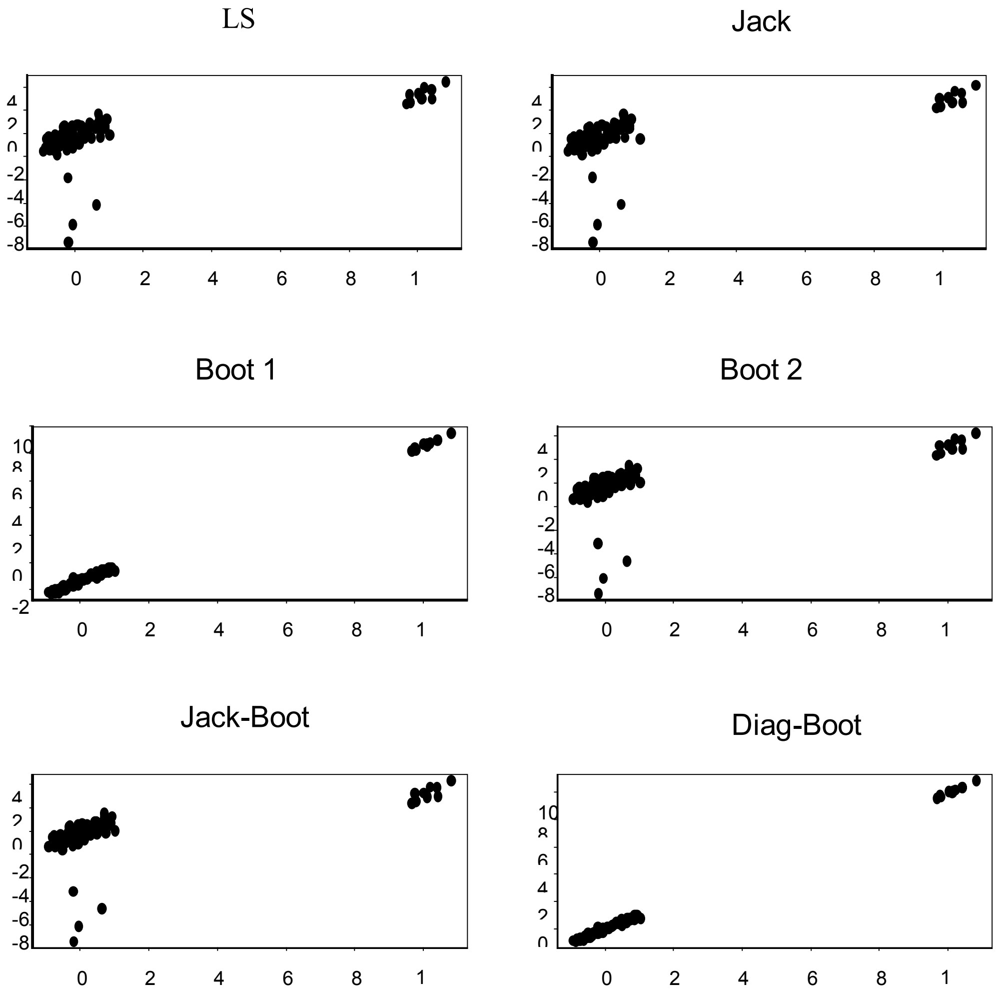

Figure 1. Residual plots against errors of different bootstrap techniques for Hawkins-Bradu-Kass (1984) data.

To measures which technique does best in estimating the true errors analytically we generally use the ratio of squared distances (RSD) proposed by Imon (2003). For any set of residuals $\epsilon_1^*, \epsilon_2^*, \cdots, \epsilon_n^*$, we define RSD as

$$\mathrm{RSD}(*) = \sum_{i=1}^{n} (\epsilon_i^* - \epsilon_i)^2 / k\sigma^2 \tag{4.1}$$

where we know the true errors. We would expect the RSD quantities to be close to 1 when the OLS residuals are used in it. As we do not know the exact ϵ here, we would use RLS residuals as their substitutes. We observe from the RSD values of different residuals that the performance of the OLS, jackknife, bootstrap 2 and jackknife-after-bootstrap methods are very poor as they yield very high RSD values. As it is expected, bootstrap 1 method performs better than other bootstrap methods, but the performance of the diagnostic-before-bootstrap method is quite outstanding. Its low RSD value clearly indicates that this method is able to estimate the entire set of residuals in a way that they become very close to the true errors.

Now we present few graphical displays to show the performances of different bootstrap techniques to estimate errors for Hawkins *et al.* (1984) data. In Figure 1, we have plotted the OLS, jackknife, bootstrap 1, bootstrap 2, jackknife-after-bootstrap and diagnostic-before-bootstrap residuals against the RLS residuals assuming the latter ones as true errors. For a good set of residuals we would expect a straight line from this scatter plot. We observe from this figure that the OLS residuals break down completely for this data. Similar remark may apply with the jackknife, bootstrap 2 and jackknife-after-bootstrap residuals. The performance of bootstrap 1 residuals are good but diagnostic-before-bootstrap residuals perform best over all and we observe almost a straight line when these residuals are plotted against the true errors.2

4.2 Belgian Telephone Data

In the Belgian Statistical Survey, we found a data set (see Rousseeuw and Leroy 1987) containing the total number (in tens of millions) of international phone calls made between the years 1950 and 1973. This time series data contains heavy contamination from 1964 to 1969. Upon inquiring, it turned out that during that period another recording system was used giving the total number of minutes of these calls. It has been reported by many authors that most of the commonly used diagnostic methods fail to identify 3 outliers (1964-1966) out of 7, but swamps in 3 good observations (1971-73) and thus produce a poor set of residuals as it is shown in Table 2. We also observe similar results when the jackknife, bootstrap 2 and jackknife-after-bootstrap techniques are used to estimate the true errors.

Table 2. Bootstrap residuals for Belgian telephone data

	OLS	Jack	Boot1	Boot2	Jack-Boot	Diag-Boot	RLS
1950	1.2385	1.246	−4.6086	1.3936	1.3935	−1.0089	0.2679
1951	0.7644	0.771	−4.5774	0.8999	0.9001	−0.9797	0.1675
1952	0.2602	0.266	−4.5762	0.3763	0.3762	−0.9804	0.0371
1953	−0.1239	−0.119	−4.455	−0.0274	−0.027	−0.8612	0.0267
1954	−0.5581	−0.554	−4.3838	−0.481	−0.4812	−0.792	−0.0337
1955	−0.9922	−0.989	−4.3126	−0.9347	−0.9349	−0.7227	−0.0942
1956	−1.4164	−1.4141	−4.2313	−1.3784	−1.3786	−0.6435	−0.1446
1957	−1.8505	−1.8491	−4.1601	−1.832	−1.8318	−0.5742	−0.205
1958	−2.1746	−2.1741	−3.9789	−2.1757	−2.1755	−0.395	−0.1554
1959	−2.5388	−2.5391	−3.8377	−2.5593	−2.5594	−0.2558	−0.1458
1960	−2.8929	−2.8941	−3.6865	−2.933	−2.9331	−0.1065	−0.1262
1961	−3.2571	−3.2591	−3.5453	−3.3167	−3.3168	0.0327	−0.1166
1962	−3.6412	−3.6441	−3.4241	−3.7203	−3.7205	0.1519	−0.127
1963	−3.6354	−3.6391	−2.9129	−3.734	−3.7337	0.6612	0.2526
1964	5.6405	5.6359	6.8683	5.5224	5.5224	10.4404	9.9022
1965	5.6363	5.6309	7.3695	5.4987	5.4988	10.9397	10.2718
1966	6.9322	6.9259	9.1707	6.7751	6.7754	12.7389	11.9414
1967	8.128	8.1209	10.872	7.9514	7.9511	14.4381	13.511
1968	9.9239	9.9158	13.1732	9.7277	9.7278	16.7374	15.6806
1969	12.4197	12.4108	16.1744	12.2041	12.2041	19.7366	18.5502
1970	−4.9844	−4.9942	−0.7244	−5.2196	−5.2195	2.8359	1.5198
1971	−7.3886	−7.3992	−2.6232	−7.6432	−7.6433	0.9351	−0.5106
1972	−7.5927	−7.6042	−2.322	−7.8669	−7.867	1.2343	−0.341
1973	−7.8969	−7.9092	−2.1208	−8.1906	−8.1908	1.4336	−0.2715
RSD	5.729	5.743	3.914	6.11	6.11	0.274	

When the newly proposed diagnostic-before-bootstrap method is applied to this data we observe from Table 2 that all the 6 outliers are clearly identified. The resulting residuals are also very close to the RLS residuals, which we assume as the true errors. We observe from the RSD values of different residuals that the performance of the OLS, jackknife, bootstrap 2 and jackknife-after-bootstrap methods are very poor as they yield very high RSD values. As it is expected, bootstrap 1 method performs better than other bootstrap methods, but the performance of the diagnostic-before-bootstrap method is quite outstanding. Its low RSD value clearly indicates that this method is able to estimate the entire set of residuals in a way that they become very close to the true errors.

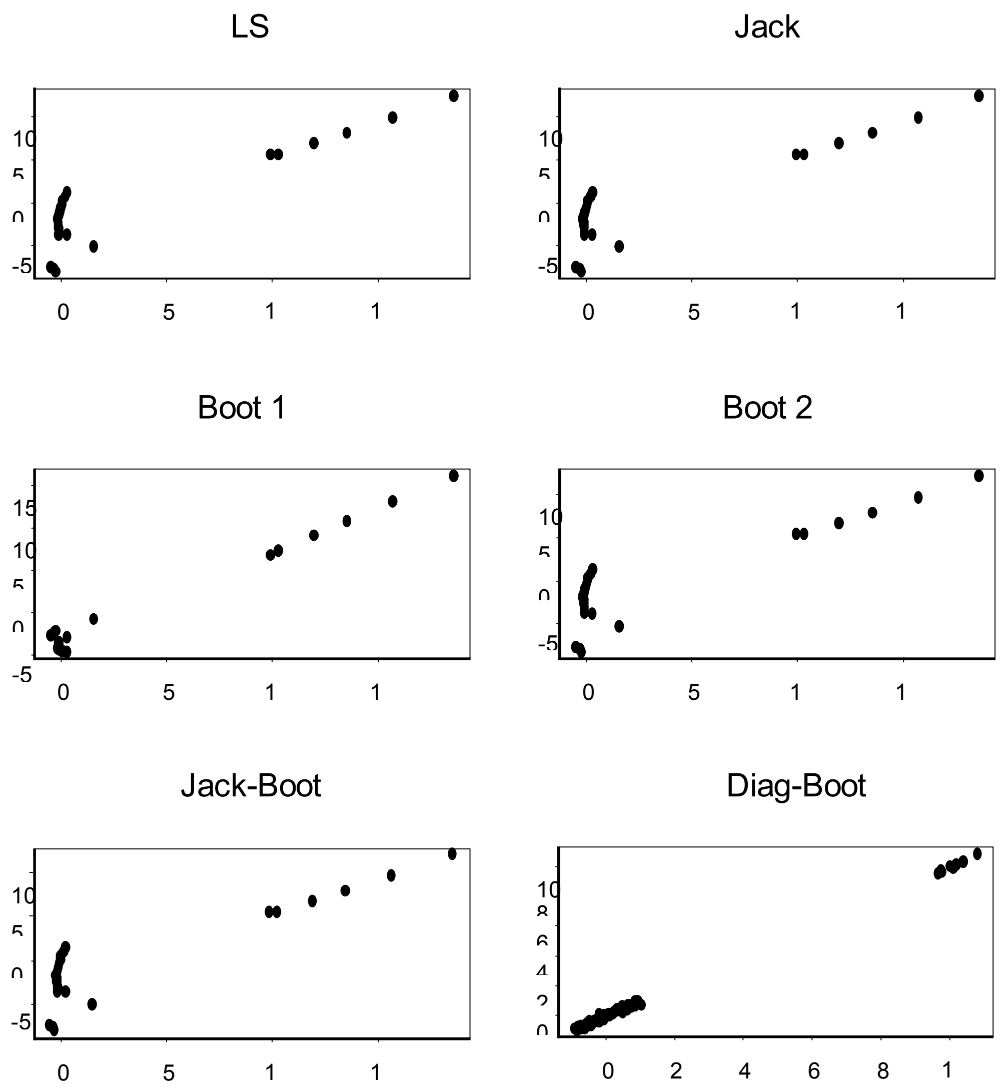

Figure 2. Residual plots against errors of different bootstrap techniques for Belgian telephone data.

Figure 2 shows the performances of different bootstrap techniques to estimate errors for Belgian telephone data. We observe from this figure that the OLS residuals break down completely for this data. Similar remark may apply with the jackknife, bootstrap 2 and jackknife-after-bootstrap residuals. The performance of bootstrap 1 residuals are good but diagnostic-before-bootstrap residuals perform best over all and we observe almost a straight line when these residuals are plotted against the true errors.

5. Simulation Results

So far we have considered examples to investigate the performance of our newly proposed diagnostic-before-bootstrap residuals as an improved set of estimates of the true errors. To understand the role of different estimation techniques more clearly we require simulation results with a good number of replications considering a variety of sample sizes. Although this type of data set is artificial in nature, it is very useful because here we know exactly the real situation; otherwise there is always uncertainty [see Cook and Hawkins (1990)] about which observations are unusual.

Now we carry out a simulation experiment to show how the RSD for different estimation techniques vary with the change in the sample sizes and the error and leverage structures. We have used five sets of designs in this experiment.

(i) **Normal errors**: Uniform (0,1) regressors with errors generated from Normal (0,4) distribution
(ii) **Single outlier**: Errors generated from Normal (0,4) distribution with the nth observation as outlier.
(iii) **Single high leverage outlier**: Uniform (0,1) regressor with the nth observation as high leverage point. Errors are generated from Normal (0,4) distribution but the nth observation is set as an outlier.
(iv) **Multiple (10%) outliers**: Errors generated from Normal (0,4) distribution with the last 10% observations as outliers.
(v) **Multiple (10%) high leverage** outliers: Uniform (0,1) regressor with the last 10% observations as points of high leverages. At the same time errors are generated from Normal (0,4) distribution with the last 10% observations as outliers.

For each of the design, Y is computed from equation

$$Y = 20 + 4.5X_1 - 1.5X_2 + 2.5X_3 + \epsilon \tag{5.1}$$

with n = 20, 30, 40, 50 and 100 and the results of RSD for different estimation techniques and different sample sizes based on 5000 simulations each are presented in Table 3.

Table 3. Simulation results of RSD for different bootstrap techniques

Sample size	Estimation technique	Normal	Single outlier	Single high leverage outlier	Multiple outliers	Multiple leverage outliers
n = 20	OLS	1.000	2.574	6.917	5.768	13.236
	Jack	1.001	2.497	6.702	5.736	13.233
	Boot 1	0.455	0.506	1.243	1.900	1.630
	Boot 2	1.108	2.493	6.219	4.745	12.079
	Jack-Boot	1.170	2.459	6.217	4.745	12.079
	Diag-Boot	0.457	0.460	0.495	0.528	0.514
n = 30	OLS	1.000	2.059	6.932	6.201	19.487
	Jack	1.000	1.998	6.717	6.188	19.482
	Boot 1	0.411	0.455	0.972	2.907	2.596
	Boot 2	1.067	1.995	6.232	5.599	17.784
	Jack-Boot	1.084	1.967	6.230	5.600	17.786
	Diag-Boot	0.412	0.413	0.439	0.536	0.520
n = 40	OLS	1.000	1.750	6.946	6.667	25.745
	Jack	1.000	1.698	6.730	6.653	25.739
	Boot 1	0.382	0.385	0.811	3.392	3.170
	Boot 2	1.016	1.695	6.245	6.325	23.495
	Jack-Boot	1.037	1.672	6.243	6.325	23.489
	Diag-Boot	0.383	0.383	0.403	0.533	0.521
n = 50	OLS	1.000	1.566	6.961	6.931	31.840
	Jack	1.000	1.527	6.607	6.916	31.839
	Boot 1	0.376	0.380	0.708	3.617	3.433
	Boot 2	1.000	1.645	6.332	6.716	31.503
	Jack-Boot	1.012	1.623	6.332	6.716	31.581
	Diag-Boot	0.377	0.377	0.393	0.538	0.529
n = 100	OLS	1.000	1.300	7.067	10.546	63.242
	Jack	1.000	1.271	7.067	10.523	63.242
	Boot 1	0.368	0.369	0.506	6.710	6.515
	Boot 2	0.996	1.303	6.248	10.336	63.199
	Jack-Boot	1.004	1.303	6.248	10.336	63.199
	Diag-Boot	0.368	0.368	0.375	0.685	0.675

We observe from the results given in Table 3 that for normal errors, as the standard theory tells, the OLS method produces RSD values very close to 1 for different sample sizes. The performances of OLS and jackknife are very much similar. But the performances of bootstrap 2 and jackknife-after-bootstrap methods are not very satisfactory because throughout the simulation they perform less than the OLS. However, their performances tend to improve with the increase in sample sizes. Diagnostic-before-bootstrap and bootstrap 1 method perform best throughout in the estimation of errors when they are normal as they produces much better result than the OLS. For a single outlier case, the OLS, jackknife, bootstrap 2 and jackknife-after-bootstrap produce much higher RSD values than diagnostic-before-bootstrap and bootstrap 1. But things become worse when outliers are also points of high leverage. The OLS and jackknife method break down completely for different sample sizes yielding very high RSD values. The performances of bootstrap 2 and jackknife-after-bootstrap often become poor. But it is interesting to note that performances of diagnostic-before-bootstrap and bootstrap 1 methods are quite satisfactory in this situation. But we face more serious consequences when 10% outliers are present in the data. We observe the worst set of results from the OLS followed by the jackknife, jackknife-after-bootstrap and bootstrap 2 methods. But it is interesting to note that, for the first time we observe that the so far successful bootstrap 1 method produces higher RSD values and these values tend to increase with the increase in sample sizes. Things become even worse when these outliers are associated with high leverage points. Here all of the estimation techniques except diagnostic-before-bootstrap breakdown and their corresponding RSD values tend to increase with the increase in sample size.

5. Conclusions

In this paper we propose a new bootstrap technique suitable for regression analysis keeping in mind the fact that like real data, bootstrap data also deserve a thorough examination especially when unusual observations are present in the data. We present few examples, which clearly show that all of the commonly used bootstrap techniques fail to produce accurate estimates of the regression parameters in the presence of a group of outliers. We also observe that things may become even worse when outliers are also points of high leverages. But the well-known examples and simulation experiments clearly show that our newly proposed method, where robust diagnostic is used before bootstrapping become very successful in a variety of situations. The diagnostic-before-bootstrap technique does extremely well in the presence of multiple outliers and high leverage outliers without harming a genuine normal case and performs in a robust way irrespective of error and leverage structure or sample sizes.

Acknowledgements

We thank the referees for their comments and suggestions which improved the presentation of the paper. We also thank Mr. Kallol Kumar Das for helping with computation.

References

1. Barnett, V., and Lewis, T. (1994). *Outliers in Statistical Data*. 3rd Ed., Wiley, New York.
2. Booth, J.G., and Sarkar, S. (1998). Monte Carlo approximation of bootstrap variances. *Amer Statist.*, 52, 354-357.
3. Cook, R.D., and Hawkins, S. (1990). Comment on paper by Rousseeuw, P.J., and van Zomeren, B.C., *J. Amer. Statist. Assoc.*, 85, 640-644.
4. Davison, A.C., and Hinkley, D.V. (1997). *Bootstrap Methods and Their Applications*. Cambridge University Press, London.
5. Efron, B. (1979). Bootstrap methods: Another look at the jackknife, *Ann. Statist.*, 7, 1-26.
6. Efron, B. (1987). Better bootstrap confidence intervals (with discussions), *J. Amer. Statist. Assoc.,*, 82, 171-200.
7. Efron, B. (1992). Jackknife-after-bootstrap standard errors and influence functions (with discussions), *J. Roy. Statist.*. Soc., Ser - B, 54, 83-127.
8. Efron, B., and Tibshirani, R. (1993). *An Introduction to the Bootstrap*. Chapman and Hall, New York.
9. Fox, J. (1991). *Regression Diagnostics*. Sage, Newbury Park.
10. Freedman, D.A. (1981). Bootstrapping regression models, *Ann. Statist.*, 9, 218-228.
11. Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. Springer, New York.
12. Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., and Stahel, W. (1986). *Robust Statistics: The Approach Based on Influence Function*. Wiley, New York.
13. 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.
14. Hocking, R.R. and Pendleton, O.J. (1983). The regression dilemma, *Commun. Statist.* - Theory Meth., 12, 497-527.
15. Imon, A.H.M.R. (2003) Residuals from deletion in added variable plots, *J. App. Stat.*, 30, 841-855.
16. Imon, A.H.M.R., and Das, K.K. (2005). A comparative study on the estimation of regression errors by bootstrap techniques, *Pak. J. Statist.*, 21, 109-23.
17. Rousseeuw, P.J. (1984). Least median of squares regression, *J. Amer. Statist.* Assoc.,, 79, 871-80.
18. Rousseeuw, P.J., and Leroy, A. (1987). *Robust Regression and Outlier Detection*. Wiley, New York.
19. Ryan, T.P. (1997). *Modern Regression Methods*. Wiley, New York.

20. Sengupta, D. and Jammalamadaka, S. (2003). *Linear Models: An Integrated Approach*. World Scientific, New Jersey.
21. Shao, J., and Tu, D. (1995). *The Jackknife and Bootstrap. Springer-Verlag*, New York.
22. Venables, W.N., and Ripley, B.D. (2000). *Modern Applied Statistics With S-Plus*. Vol. II, 3rd. Ed., Springer-Verlag, New York.
23. Wu. C.J.F. (1986). Jackknife, bootstrap and other resampling methods in regression analysis (with discussions), *Ann. Statist.*, 14, 1261-1350.

[received date : Jun. 2005, accepted date : Aug. 2005]