International Journal of Computational and Theoretical Statistics
ISSN (2210-1519)
Int. J. Comp. Theo. Stat. 2, No. 2 (Nov-2015)

Influential Observations and Cutoffs of Different Influence Measures in Multiple Linear Regression

Mintu Kr.Das[1] and Bipin Gogoi[2]

[1] *Research Scholar, Department of Statistics, Dibrugarh University Assam, India*
[2] *Professor, Department of Statistics, Dibrugarh University Assam, India*

Received 13 February 2015; Revised 17 Apri, 2015, Accepted 10 May 2015, Published 1 November 2015

Abstract: The analysis of data for outliers is a part of model building and data summarizing for model testing, parameter estimation, prediction and peculiarity investigations. Any influential point can disproportionately pull the ordinary least squares line and distort the predictions. Thus the detection of outlying observations is very essential in the course of model building in various disciplines, such as, medical research, economics, sociology, computer science, etc. A point is an influential one if it causes dramatic change in the model after its deletion. Each of the available test statistics has different cutoff values that indicate the amount of outlyingness. Sometimes only one statistic is sufficient to provide the information about influential points but often it is necessary to examine the cutoff of more than one influence measure. The reason behind is that all the cutoff values are either a function of the sample size or number of predictors or both. Also validity of the cutoff value is subjected to some additional conditions. In this paper we try to critically examine those conditions with the help of simulation study. We shall use a few combinations of (n, k), where n is the sample size and k is the no. of outliers for assessing the performances.

Keywords: Influential observations, Cooks distance, DFFITS, COVRATIO and Leverage.

1. INTRODUCTION

The analysis of data for outliers is a part of model building and data summarizing for model testing, parameter estimation, prediction and peculiarity investigations, however the perception of outlier is not at all simple [1]. Detection of outlying observations is very essential in the course of model building in various disciplines, viz., medical research, economics, sociology, computer science, etc. The ordinary least-squares (OLS) regression line passes through the centroid of the data and it assigns equal weights to all the observations. So the least square estimation is highly sensitive to outliers and influential observations. In the absence of outliers and with the fulfillment of the assumptions of zero mean, constant variance and uncorrelated errors the OLS provides us the BLUE of the regression parameters. But any anomalous point can disproportionately pull the line and distort the predictions. Detection of outlying observations is a very essential part of good regression analysis. The concept of influential point plays the central role in the detection process. In the linear regression framework, there are various ways to detect influential observations. An influential observation is one which either individually or together with several other observations, has a demonstrably larger impact on the calculated values of various estimates (coefficient, standard errors, t-values, etc.) than to the case for most of the other observation [21].The outlier is indeed an influential point, but the contrary is not always true [1]. An outlier (either in response space or in the space of the predictors) will not necessarily influential in affecting the regression equation. Tukey (1977) emphasized the importance of exploring the data, rather than just examining one of two summary statistics [15]. Thus two aspects of outlier detection are there which may be classified as ***case analysis*** and ***summary measure analysis.*** Another approach is the robust regression which is said to be insensitive to such wild points. But interestingly Hurber (1977) illustrated that in case of outliers in the predictor space robust approach may be inefficient, which indicates a need for ***case analysis*** in case of robust regression. The misleading consequence of the summary statistic was proved by Anscombe (1973) where same value of R^2 was obtained for three different data sets. [15]. Hocking (1983) found that the hat diagonals and "deleted" studentized residuals provide most of the evidence needed to track down maverick cases. Contrary to that, Welsch (1980) found that neither the hat diagonals nor studentized residuals alone will sufficient. He suggested $DFFITS_i$ [16].

E-mail: mintukrdas@gmail.com, bipin.gogoi@rediffmail.com

Since the work initiated by Mickey, Dunn and Clark (1967), several influence measures were developed to measure their influence on the forecast [2]. To be an influential one a point should cause dramatic change in the model after its deletion. Each of those test statistics has different cutoff values that indicate the amount of outlyingness. Sometimes only one statistic is sufficient to provide the information about influential points but often it is necessary to examine the cutoff of more than one statistic. The reason behind is that all the cutoff values are either a function of the sample size or number of predictors or both. Also validity of the cutoff value is subjected to some additional conditions [1]. The cut-off points should be used with caution. Diagnostic methods are not designed to be used as formal test of hypothesis (Hadi, 1992). Now-a-days the major statistical packages like BMDP, SPSS, SAS, MINITAB, R etc. avail the values of some outlier diagnostics which make the use of those diagnostics much wider. But the usage needs some proper guidelines regarding the benchmark value. In this paper we try to critically examine those conditions with the help of simulation study. We shall use a few combinations of (n, k) to assess the performances.

2. MEASURES OF INFLUENCE

The inclusion or exclusion of a point often make some change either in the estimated regression parameters or fitted values. Thus the influence of any point in a model is measured in terms of the change that occurred when that suspicious point is omitted from the model. Considering the linear regression model with intercept in matrix form

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \qquad (1)$$

where $\boldsymbol{y_{n\times 1}}$ is the response vector; $\boldsymbol{X} = \left(x_{ij}\right)_{n\times(p+1)}$ with $\boldsymbol{x_{i0}} = \boldsymbol{1}$ is the design matrix; $\boldsymbol{\beta}_{(p+1)\times 1}$ is the vector of parameters and $\boldsymbol{\varepsilon} \sim \boldsymbol{N}(\boldsymbol{0}, \sigma^2\boldsymbol{I})$ .With the fulfillment of the major assumptions concerning the model (1) the ordinary least-squares (OLS) regression yields the best linear unbiased estimate (BLUE) of the regression parameters. Those four assumptions are (i) Linearity at least approximately of the response with the predictors ; (ii) Independence nature of errors ; (iii) Equal error variance σ^2 (iv) Normality of the errors [3]. The least squares estimate (LSE) of $\boldsymbol{\beta}$ is given by $\boldsymbol{\beta} = (\boldsymbol{X^TX})^{-1}\boldsymbol{X^Ty}$ and the vector of fitted values as $\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{Hy}$ where $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X^TX})^{-1}\boldsymbol{X^T}$ is the ***hat*** (**weight** or ***projection***) ***matrix***. Here $\boldsymbol{H}$ maps the vector of observed values into a vector of fitted values [1]. Thus the vector of LS residuals becomes $\hat{\boldsymbol{e}} = \boldsymbol{y} - \hat{\boldsymbol{y}} = (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{y}$. The different types of measures with their usual cutoff are as follows.

2.1. Cook's Distance $\boldsymbol{D_i}$: Cook (1997) proposed a measure using the information from the studentized residuals and the variances of residuals and predicted values. Denoting the LSE of $\boldsymbol{\beta_i}$ without the i[th] observation as $\hat{\boldsymbol{\beta}}_{(-i)}$ the statistic is given by-

$$\boldsymbol{D_i} = \frac{\left(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(-i)}\right)^T \boldsymbol{X^TX}\left(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(-i)}\right)}{(p+1)s^2}$$

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

Where r_i is the studentized residuals defined by $r_i = e_i/\left[s(1-h_{ii})^{1/2}\right]$ and $s^2 = MS_{Res}$ is the estimate of the error variance. Here $\boldsymbol{D_i}$ combines residual magnitude and the location of the i[th] point in X-space to access influence. Since $\boldsymbol{X}\hat{\boldsymbol{\beta}}_{(-i)} - \boldsymbol{X}\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{y}}_{(-i)} - \hat{\boldsymbol{y}}$, we can write

$$\boldsymbol{D_i} = \frac{\left(\hat{\boldsymbol{y}}_{(-i)} - \hat{\boldsymbol{y}}\right)^T\left(\hat{\boldsymbol{y}}_{(-i)} - \hat{\boldsymbol{y}}\right)}{(p+1)s^2} \qquad (3)$$

Removal of the i[th] data point should keep the $\hat{\boldsymbol{\beta}}_{(-i)}$close to $\hat{\boldsymbol{\beta}}$ unless the point is an outlier. $\boldsymbol{D_i}$ as influence measure is based on confidence ellipsoid whose centre is at $\hat{\boldsymbol{\beta}}$. Thus $\boldsymbol{D_i}$ is the squared Euclidean distance that the vector of fitted response moves when the i[th] observation is deleted. Or in other words we can say that $\boldsymbol{D_i}$ examines the changes occurred in estimates for $\hat{\boldsymbol{\beta}}$ when some cases are deleted. This is the basic idea in influence analysis as introduced by Cook [3]. $\boldsymbol{D_i}$ is based on a scaled Mahalanobis type squared distance between $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}_{(-i)}$. Stevens (1984) suggested the use of $\boldsymbol{D_i}$ after having some signal of outlyingness from large hat diagonals (Mahalanobis distance) and studentized residuals [15]. Jensen and Ramirez (1998) established the effectiveness of Cook's statistic with the use of eigen values as well as associated p-values [12].

Cook and Weisberg (1994) suggested the use of the median of F-distribution with $(p+1)$and $(n-p-1)$degrees of freedom as a benchmark for identifying the influential subset. In another piece of work Muller and Mok (1997) showed that if we use the above benchmark then the test size will systematically and rapidly decrease with n. Ideally

cut off point should allow consistent interpretation across regression analysis. But the median or any other quantile of $F_{(p+1,n-p-1)}$, does not allow us such consistency [11].

2.2. COVRATIO$_i$,: This measure uses the concept based on the role of the ith observation on the precision of estimation and it is defined by

$$COVRATIO_i = \frac{\left|\left(\boldsymbol{X}_{(-i)}^T \boldsymbol{X}_{(-i)}\right)^{-1} \boldsymbol{S}_{(-i)}^2\right|}{|(\boldsymbol{X}^T\boldsymbol{X})^{-1} MS_{Res}|} \qquad (4)$$

$$COVRATIO_i = \frac{\left(\boldsymbol{S}_{(-i)}^2\right)^{(p+1)}}{MS_{Res}^{(p+1)}}\left(\frac{1}{1-h_{ii}}\right) \qquad (5)$$

Here $COVRATIO_i$ >1 indicates that ith observation improves the precision of estimation and $COVRATIO_i$ <1 indicates degradation. Belsley et al. (1980) recommended that the ith observation is influential if $\{1+3(p+1)/\mathrm{n}\} < COVRATIO_{\mathrm{i}} < \{1-3(p+1)/\mathrm{n}$ [1].

2.3. DFFITS$_i$, : It is a measure of influence introduced by Belsley, Kuh and Welsch (1980), which measures how the deletion of the ith observation influence the predicted or fitted values [1]. It is given as

$$DFFITS_i = \left(\hat{\boldsymbol{y}} - \hat{\boldsymbol{y}}_{(-i)}\right)/\sqrt{\boldsymbol{S}_{(-i)}^2 h_{ii}} \qquad (6)$$

$$DFFITS_i = \left(\frac{h_{ii}}{1-h_{ii}}\right)^{1/2} t_i \qquad (7)$$

The rule of thumb is that an observation for which $|DFFITS_i| > 2\sqrt{(p+1)/n}$ warrants attention. Vellman and Wesch(1981) suggested that $|DFFITS_i|$ values greater than 1 to 2 warrant special attention. $DFFITS_i$ is affected by both leverage and prediction error [1].

2.4. Hat matrix diagonals as a measure of influence:

The fitted response is a linear combination of the observed response values and elements of hat matrix. Thus $\hat{y}_i = h_{ii}y_i + \sum_{i\neq j} h_{ij}y_j$ indicates how heavily $y_i$contributes to$\hat{y}_i$ viah_{ii}. Cook and Weisberg (1982) say that "for any $h_{ii} > 0$, $\hat{y}_i$ will be dominated by $h_{ii}y_i$ if y_i is an outliers [4]. The hat matrix plays an important role for detecting influential observations as it determines the variance and covariance of residuals $\{Var(e_i) = \sigma^2(1-h_{ii})\}$ and that of fitted responses $\{Var(\hat{y}) = \sigma^2\boldsymbol{H}\}$**.** The elements h_{ij} may be regarded as the amount of leverage (impact/influence) exerted by the ith observation on the ith fitted value. The diagonal h_{ii} is the standardized measure of the distance of the ith observation from the centre (or centroid) of the X-space [6]. Generally $1/n \le h_{ii} \le 1$ with an average value $\overline{h_{ii}} = trace(H)/n = (p+1)/n$ and the data points with $h_{ii} > 2(p+1)/n$ may be regarded as outliers in X-space [3]. Larger the value of h_{ii}, smaller is the $Var(e_i)$. So large leverage will lead to the closer fit of $\hat{y}_i$ to y_i. Leverage reflects the position of an observation in the multidimensional space of the carriers or predictors [17]. In extreme case $h_{ii} = 1$, the error variance becomes zero. That's why a point which is located extremely in the X-space may not always influential unless it has an unusual value in Y-space. Another guideline is that $h_{ii} > 0.5$ indicate very high leverage, whereas $0.2 \le h_{ii} \le 0.5$ indicate moderate leverage. Additional evidence of an outlying case is the existence of a gap between the leverage values for most of the cases and the unusually high leverage value(s) [5].

2.5. Atkinson's A_i(Modified D_i):

To incorporate the deletion effect on variance estimate Atkinson (1985) modified $\boldsymbol{D_i}$ by replacing s^2 by $s_{(-i)}^2$ and scaling by a factor $(n-p-1)/(p+1)$ instead of $1/(p+1)$ and then taking the square root . As a result the modified Cooks measure becomes

$$\boldsymbol{A_i} = \frac{\left(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(-i)}\right)^T \boldsymbol{X}^T\boldsymbol{X}\left(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(-i)}\right)(n-p-1)}{(p+1)\, s_{(-i)}^2}$$

$$\boldsymbol{A_i} = \sqrt{\boldsymbol{D_i}(n-p-1)s^2/s_{(-i)}^2} \qquad (8)$$

$$\boldsymbol{A_i} = DFFITS_i\sqrt{(n-p-1)/n} \qquad (9)$$

The above measures Cook's $\boldsymbol{D_i}$, $DFFITS_i$ and Atkinson's $\boldsymbol{A_i}$ tells us parallel stories for most sets of data [4]. Draper and Smith suggested using these according to the user's preferences.

2.6. Potentials: Hadi (1992) found that if there is a high leverage point then the information matrix might have broken down and consequently the observations may not have the appropriate leverage. He introduced a single case deleted measure of leverage known as potentials defined as

$$p_{ii} = x_i^T\left(X_{(-i)}^T X_{(-i)}\right)^{-1}x_i$$

where $X_{(-i)}$ is the data matrix with i[th] row deleted. Its relationship with hat diagonals is given by $p_{ii} = h_{ii}/(1-h_{ii})$. Those observations with very large potentials are considered as high leverage points [16]. Hadi suggested using the cut off as
$Mean(p_{ii}) + c \times st.dev(p_{ii})$. Here c is a constant appropriately selected such as 2 and 3. Also realizing the fact that mean and s.d. are non-robust even for one extreme observation, Hadi suggests using median and median absolute deviation (MAD) respectively.

2.7. Peña's statistic: This statistic measures the influence of an observation by the rest of the data. Peña suggested that instead of examining the overall effect on the vector $\hat{\boldsymbol{y}}$due to deletion of one observation we can measure how the deletion of each sample point affects the forecast of a specific observation separately [10]. In regression model, it can be done by considering the vectors

$$\boldsymbol{a_i} = \left(\hat{y}_i - \hat{y}_{i(-1)}, \ldots, \hat{y}_i - \hat{y}_{i(-n)}\right)^T$$

The proposed statistic at the i[th] observation is defined as the squared norm of the standardized vector $\boldsymbol{a_i}$, that is,

$$\boldsymbol{S_i} = \frac{\boldsymbol{a_i^T a_i}}{(p+1)var(\hat{y}_i)} \qquad (10)$$

$$\boldsymbol{S_i} = \frac{1}{(p+1)s^2 h_{ii}}\sum_{l=1}^{n}\frac{h_{li}^2 \mathrm{e}_\mathrm{i}^2}{(1-h_{ll})^{\,2}} \qquad (11)$$

$$\boldsymbol{S_i} = \sum_{l=1}^{n}\rho_{li}^2 D_l \qquad (12)$$

where D_l is the Cook's distance and $\rho_{il} = \sqrt{(h_{il}^2/h_{ii}h_{ll})} \leq 1$ is the correlation between the forecasts $\hat{y}_i$ and $\hat{y}_l$. Thus $\boldsymbol{S_i}$ is a linear combination of Cook's distances and it has some important properties: (i) in case of outlier free data and small hat diagonals $E(\boldsymbol{S_i}) = 1/(p+1)$; (ii) $\boldsymbol{S_i}$ approximately follows normal distribution; (iii) in the presence of high-leverage identical outliers, the sensitivity statistics will identify them and $\boldsymbol{S_i}$ will be smaller for the outliers than for the good data points [10].

3. SIMULATION STUDY

Here we are comparing the above measures using Monte-Carlo simulation study. The linear model (1) can be written as follows as a way to model the outlier which is known as mean shift outlier model.

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\delta} + \boldsymbol{\varepsilon} \qquad (13)$$

where $\boldsymbol{\delta}$ is an $n \times 1$ vector consisting of $k(\ll n)$ unknown non-zero values, at k suspicious locations and having zeros elsewhere. Also for simplicity we consider $\boldsymbol{\delta}$ as the resultant contaminated vector having two components from response and predictor space outliers, i.e., $\delta = \delta_y + \delta_x$. For a hypothesized value of k there are $\binom{n}{k}$ partitions of the data and out of that $\binom{n-k}{k}$ will have clean data and $\binom{n}{k} - \binom{n-k}{k}$ will include the outliers. Without any loss of generality, the $\boldsymbol{\varepsilon}$'s for osbservations other than planted outliers are generated as $N(0,1)$ random variables. The outliers are introduced by adding known quantity to the first k of these variates. The explanatory variables are constructed by generating independent uniform (0,1) variates. The residual correlations resulting from such a set of explanatory variables are expected to be fairly small, and therefore the test procedure is expected to perform in optimal way. The

results of the study are based on 1,000 simulations. The power comparisons were made in such a way that for $n =$ sample size, $k =$number of outliers, c1= number of times the statistic detects outliers and c2= number of times the statistic exactly detects k outliers; the performance are being measured as the quantity c1/(Total number of repetitions × number of outliers). Also if the statistic is free from swamping and masking effect then its actual performances are being obtained by the quantity c1/(Total number of repetitions × number of outliers).

Table 1: Empirical performances of different influence measures for various combinations

(δ_y, δ_x)	n	k	$\mathbf{COVRATIO_i}$ $(1 \pm 3(p+1)/n)$	$\mathbf{DFFITS_i}$ $\left(2\sqrt{(p+1)/n}\right)$	$h_{ii}\left(\frac{2(p+1)}{n}\right)$	p_{ii} $(mean + 2sd)$	S_i $(mean + sd)$	$D_i(4/n)$
(5,5)	10	2	1.747	0.732	0.3015	0.162	0.972	0.0105
(5,5)	15	2	1.673	0.839	1.112	0.7665	1.4125	0.1355
(5,5)	20	3	1.2833	0.6273	1.0497	0.4997	1.2867	0.1877
(5,5)	25	4	1.1195	0.5115	0.9475	0.3883	1.2255	0.2013
(5,5)	30	4	1.2022	0.575	1.0965	0.635	1.4418	0.2765
(5,20)	10	2	1.7335	0.7505	0.2955	0.055	1.0415	0.015
(5,20)	15	2	1.6885	0.851	1.112	0.9315	1.541	0.1375
(5,20)	20	3	1.2977	0.631	1.085	0.4897	1.4647	0.186
(5,20)	25	4	1.1183	0.5175	1.0818	0.347	1.3923	0.2048
(5,20)	30	4	1.2148	0.579	1.0958	0.685	1.6753	0.2852

Table 2: Empirical performances of different influence measures detecting exact number of outliers for various situations

(δ_y, δ_x)	n	k	$\mathbf{COVRATIO_i}$ $(1 \pm 3(p+1)/n)$	$\mathbf{DFFITS_i}$ $\left(2\sqrt{(p+1)/n}\right)$	$h_{ii}\left(\frac{2(p+1)}{n}\right)$	p_{ii} $(mean + 2sd)$	S_i $(mean + sd)$	$D_i(4/n)$
(5,5)	10	2	0	0	0	0.0005	0.0005	0.0135
(5,5)	15	2	0	0	0.0005	0	0	0
(5,5)	20	3	0.0003	0.0003	0.0003	0	0	0
(5,5)	25	4	0.0003	0	0	0	0	0
(5,5)	30	4	0	0	0.0003	0	0	0
(5,20)	10	2	0	0	0.0015	0.0165	0.0005	0
(5,20)	15	2	0	0.0005	0.0005	0.0005	0	0.004
(5,20)	20	3	0	0.0003	0.0003	0.0003	0	0
(5,20)	25	4	0.0003	0	0.0003	0.0003	0	0
(5,20)	30	4	0	0	0.0003	0	0	0

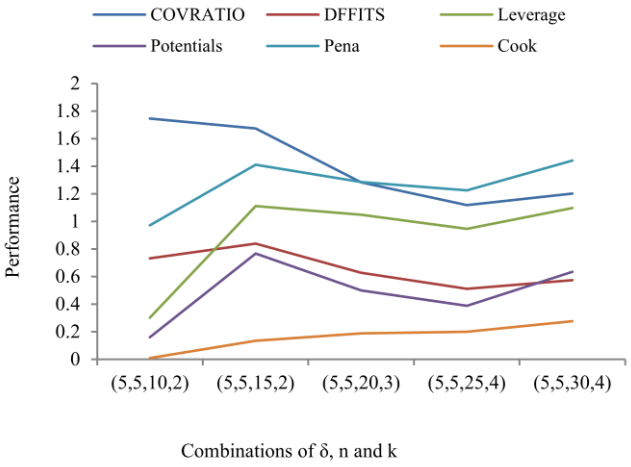

Fig.1(a): Detection performance of influence measures for low-leverage outliers

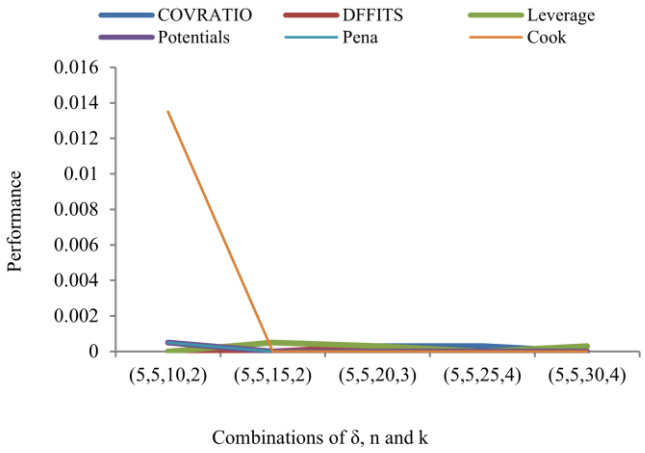

Fig.2(a): Detection performance of influence measures for exact number of low-leverage outliers

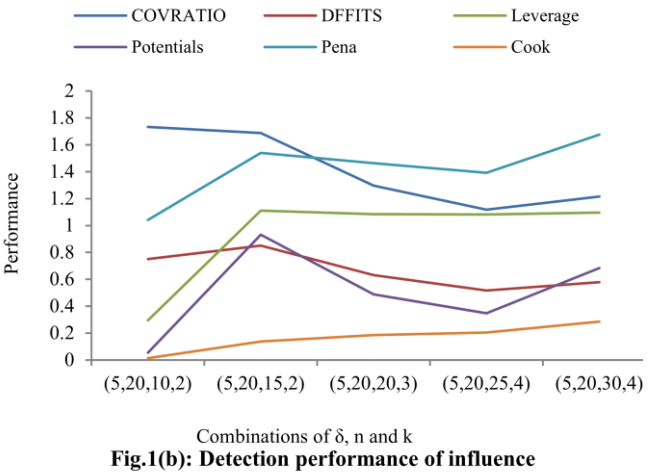

Fig.1(b): Detection performance of influence measures for high-leverage outliers

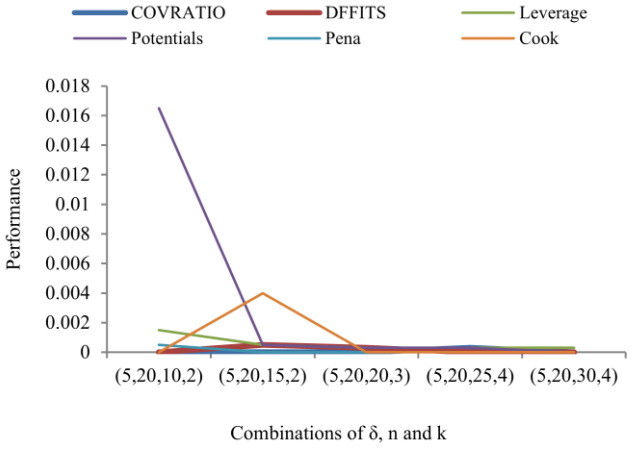

Fig.2(b): Detection performance of influence measures for exact number of high-leverage outliers

The resulting empirical powers are given the tables 1and 2. Here the first five rows corresponds to low leverage outliers and rest of the rows represents the high-leverage outliers. For the low-leverage outliers in both predictor and response space we see that COVRATIO and Pena's Statistic performs better than the others, especially better than DFFITS. Here we have not presented the performance Atkinson's $\boldsymbol{A_i}$ as it is the scaled version of DFFITS. As sample size and number of outlier increases Pena's measure can indentify more outlier than the others. Although Pena's statistic does not show any monotonic change either increase or decrease in its power as the combination of n, k, and δ changes. But the performance of COVRATIO falls down for large n and $k > 2$. Cooks distance gains the least power among the others, but its detection power increases as n and k increases. But if we consider the cutoff for Cook's distance as 1 then there were no detection of any observation as outlier. The detection performance is based on the cutoff$4/n$. Regarding the performance of hat diagonals we observe that its performance does not increase for the high-leverage outliers. Now if we take into account of the swamping and masking effects then we see that only two of the statistics are relatively better than others. Table-2 indicates that the hat diagonals and the Potentials may be considered to lessen the masking and swamping effects. While the Cooks and Pena's statistics are very poor in exact detection of outliers.

5. CONCLUSION

On the basis of the simulation results we conclude that for large sample with k>>1 COVRATIO and Pena's statistics seems to be suitable to track down outliers. While Pena's statistics is more reliable than COVRATIO to overcome masking and swamping effects. For high-leverage outliers DFFITS can be useful to detect outliers when sample size is moderately large. Also hat diagonal remains the only alternative to locate the outliers when outliers are in predictor space. Above all one noticeable result was that instead of using the ambiguous cutoff the best way locate the outliers is to examine the index plots of each statistic (or residuals) against the observation numbers. The plots like $COVRATIO_i$vs.i, $\boldsymbol{D_i}$ vs. i, etc. are useful to identify the influential observations.

ACKNOWLEDGMENT

The authors are thankful to the referee and the associate editor for their valuable suggestions. Also we acknowledge the support from the University Grant Commission, New Delhi, India for the financial assistance under the scheme UGC-BSR Fellowship in Sciences.

REFERENCES

[1] Montgomersy, D.C., Peck, E.A., Vining, G.G., "Introduction to Linear Regression", Analysis,3rd edn, Wiley series in probability and statistics, (2001).
[2] Marasinghe, M. G., "A Multistage Procedure for Detecting Several Outliers in Linear Regression", Technometrics, Vol. 27, No. 4, pp. 395-399, (1985).
[3] Su, X., Yan, X., Tsai, C.L., "Linear regression", WIREs Comp Stat, Vol.4, pp.275-294, (2012).
[4] Draper, N.R., Smith, H., "Applied Regression analysis", 3rd edn, Wiley series in probability and statistics. (2011).
[5] Kuntur, M.H., Nachtsheim, C.J., Neter, J., "Applied linear regression models", 4th edn, McGrow Hill, (2004).
[6] Seber, G.A.F., Lee, A. J., "Linear regression analysis", 2nd edn, Wiley series in probability and statistics, (2003).
[7] Gentleman, J. F, and Wilk, M. B. "Detecting Outliers. II. Supplementing the Direct Analysis of Residuals", Biometrics, Vol. 31, No. 2, pp. 387-410, (1975).
[8] Mickey, M. R., Dunn, 0. J., and Clark, V. "Note on the Use of Stepwise Regression in Detecting Outliers," Computers and Biomedical Research, No.1, pp.105-111, (1967).
[9] Cook, R. D. "Detection of Influential Observation in Linear Regression," Technometrics, vol.19, pp.15-18, (1977).
[10] Peña, D. A., "New Statistic for Influence in Linear Regression", Technometrics, Vol. 47, No. 1 (Feb., 2005), pp. 1-12, (2005).
[11] Muller, K. E. and Mok, M. C., "The distribution of cook's D-statistic", Communications in Statistics - Theory and Methods, 26:3, pp.525-546, (1997).
[12] Jensen, D. R. and Ramirez, D. E., "Detecting outliers with Cook's D-statistic", *Computing Science and Statistics* 29(1), 581-586, (1998b).
[13] Paul, S.R. and Fung, K.Y., "A Generalized Extreme Studentized Residual Multiple-Outlier-Detection Procedure in Linear Regression", Technometrics, Vol. 33, No. 3, pp. 339-348, (Aug., 1991).
[14] Cook, R.D. and Weisberg, S., "An introduction to regression graphics", Wiley series in probability and mathematical statistics. Probability and mathematical statistics, (1994).
[15] Stevens, J.P., "Outliers and Influential Data Points in Regression Analysis", Psychological Bulletin 1984, Vol. 95, No. 2, pp.334-344, (1984).
[16] Imon, A. H. M. R., "Identifying multiple influential observations in linear regression", Journal of Applied Statistics, 32:9, pp.929 — 946, (2005).
[17] Velleman, P.F., and Welsch, R.E., "Efficient Computing of Regression Diagnostics", The American Statistician,35:4, pp.234-242., (1981).
[18] McGill R., Tukey J. W. and Larsen W. A., The American Statistician, Vol. 32, No. 1, pp. 12-16, (Feb., 1978).
[19] Hadi. A.S., A new measure of overall potential influence in linear regression, Comput. Stat. Data Anal. 14 (1992), pp. 1–27
[20] Nurunnabi. A.M., Hadi. A.S. and Imon. A.H.M.R., Procedures for the identification of multiple influential observations in linear regression, Journal of Applied Statistics, 41:6, pp-1315-1331, (2014).
[21] Belsley, D.A., Kuh. E and Welsch, R.E., Regression Diagnostics: Identifying Influential Data and Sources of Collinearity, Wiley, New York, (1980).