ISSN 0971-0388

Aligarh Journal of Statistics
Vol. 41 (2021), 23-46

On Using the Multivariate Extension of Hadi's Influence Measure and Andrew-Pregibon Statistic in Diagnosing Multivariate Regression Residuals

G. S. David Sam Jayakumar[1], A. Sulthan[2] and W. Samuel[3]

[Received on September, 2019. Accepted on March, 2021]

ABSTRACT

This paper addressed the two frequently used error diagnostics in multiple linear regression analysis namely Hadi's influence measure and Andrew-Pregibon statistic which were logically extended and generalized to the multivariate form denoted as Multivariate Hadi's influence measure $\left({}_pH_i^2\right)$ and Multivariate Andrew-Pregibon statistic $\left({}_pAP_i\right)$. The Proposed multivariate measures were used to identify the Potential Outliers and Influential Observations in multivariate linear regression analysis. For this, 11 multivariate regression models were fitted and the multivariate $\left({}_pH_i^2\right)$, $\left({}_pAP_i\right)$ measures are utilized to scrutinize the residuals; results were exhibited along with the control charts.

1. Introduction and Related Work

Until the third quarter of the 20th century, to detect potentially critical observations Studentized residuals and the plot of the residuals were considered the most appropriate statistical methods. Behnken and Draper (1972) have explained that the estimated variance of the residuals includes pertinent information beyond that provided by plots of residuals or studentized residuals. They have also discussed the variances of residuals in several more complicated designs. Hoaglin and Welsh (1978) expressed, projection matrix known as the hat matrix contains this information and, together with the studentized residuals, provides a means of identifying exceptional data points. Cook (1977) has been the first to establish a simple measure, *Di* that incorporates information from the

---

✉ : G. S. David Sam Jayakumar
Email: samjaya77@gmail.com

Extended author information available after reference list of the article.

X-space and Y-space used for assessing the influential observations in regression models. The problem of outliers or influential data in the multiple or multivariate linear regression setting has been thoroughly discussed regarding parametric regression models by the pioneers namely Cook (1977), Cook and Weisberg (1982), Belsey *et al.* (1980) and Chatterjee and Hadi (1988) respectively. In non-parametric regression models, diagnostic results are quite rare. Among them, Eubank (1985), Silverman (1985), Thomas (1991), and Kim (1996) studied residuals, leverages, and several types of Cook’s distance in smoothing splines, and Kim and Kim (1998), Kim *et al.* (2001) proposed a type of Cook’s distance in kernel density estimation and local polynomial regression. Belsey *et al.* (1980) gave a suitable definition of influence after investigating the influential observations from early Cook’s measure to other various measures then. Cook’s statistical diagnostic measure is a simple, unifying and general approach for judging the local influence in statistical models. As far as the influence measures are concern in the literature, the procedures were designed to detect the influence of observations on a specific regression result. However, Hadi (1992), proposed a diagnostic measure called Hadi’s influence function to identify the overall potential influence which possesses several desirable properties that many of the frequently used diagnostics do not generally possess such as invariance to location and scale in the response variable, invariance to non-singular transformations of the explanatory variables, it is an additive function of measures of leverage and residual error, and it is monotonically increasing in the leverage values and the squared residuals. Diaz-Garcia and Gonalez-Farias (2004) modified the classical Cook’s distance with generalized Mahalanobis distance in the context of multivariate elliptical linear regression models and they also establish the exact distribution for identification of outlier data points. Using Mahalanobis distance, Jayakumar and Thomas (2013) proposed a procedure of clustering based on multivariate outlier detection. Using the relationship proposed by Weisberg (1980), Belsey *et al.* (1980); Jayakumar and Sulthan (2015; 2016; 2017) proposed an exact distribution of Cook’s distance used to evaluate the influential observations in multiple linear regression analysis. Further, they introduced a new method of regression clustering based on influential observations where the observations are treated as potential outliers. Lately, they established an exact distribution of Andre-Pregibon Statistic and evaluate influential observations in a multiple linear regression analysis, respectively. In this paper, the authors address two frequently used error diagnostics in multiple linear regression analysis namely Hadi's influence measure and Andrew-Pregibon statistic which were logically extended and generalized to the multivariate form and discussed in subsequent sections.

2. Some Preliminaries

The Multivariate linear regression model with random error is given by

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

where $\underset{(n\,X\,p)}{Y}$ $\underset{(n\,X\,q)}{X}$ are the matrix of dependent and independent variables, $\underset{(q\,X\,p)}{\beta}$ is the matrix of beta coefficients or partial regression coefficients and $\underset{(n\,X\,p)}{E}$ is the residual matrix followed multivariate normal distribution N(0, $\underset{(p\,X\,p)}{\Sigma_E}$) respectively, where rank $(X) = q$ and $p \leq q$. The best linear unbiased estimate of $\underset{(q\,X\,p)}{\beta}$ of the j^{th} regression is given as

$$\beta_j = \left(X^T X\right)^{-1} Y_j \tag{2.2}$$

where $j = 1, 2, 3\ldots, p$

From (2.2), the estimate of β which is the same as an equation by equation least squares estimation. Note that h_{ii} is the hat values of i^{th} observation or diagonal elements of the hat matrix $(H = X(X'X)^{-1}X')$ or prediction matrix play the same role as in multiple regression. Large values indicate that at least one component of the i^{th} observation may be an influential point in X-space. From (2.1), statisticians concentrate and give importance to the error diagnostics such as outlier detection, identification of leverage points and evaluation of influential observations. Several error diagnostics had been proposed in the past especially to scrutinize the residuals for multiple linear regression analysis and Hossain and Naik (1989) addressed the logical extension of the univariate diagnostics to the multivariate case. For a multivariate linear regression model, they generalized the internal studentized residual, external studentized residual in terms of Hoteling's T-square statistic, Cook's distance, Modified Cook's distance, Welsch-Kuh distance, Co-variance ratio and Likelihood displacement or Likelihood distance. Similarly, Diaz-Garcia *et al.* (2007) studied the exact distributions of multivariate classical, modified cook's distance to the multivariate elliptical linear regression model. Several error diagnostics techniques were extended to the multivariate form and Hadi's influence measure, Andrew Pregibon statistic so far not yet addressed and generalized to the multivariate linear regression. The Multivariate extension of Hadi's influence measure and Andrew Pregibon statistic is discussed in the following and subsequent sections.

3. Multivariate Extension of Hadi's Influence Measure

Hadi's $\left(H_i^2\right)$ influence measure is an interesting technique based on the fact that potentially influential observations in multiple linear regression are outliers in the X-space, the Y-space or both. The univariate form of the Hadi's influence measure of the *i*th observation is given by

$$H_i^2 = \frac{q\hat{e}_i^2}{\left(1-h_{ii}\right)\left(\hat{e}^T\hat{e}-\hat{e}_i^2\right)} + \frac{h_{ii}}{1-h_{ii}} \tag{3.1}$$

Where e_i^2 is the vector of squared estimated residuals, q is the dimensions of β, $\hat{e}^T\hat{e}$ is the sum of the squared estimated residuals and h_{ii} is the hat values of i^{th} observation or diagonal elements of the hat matrix $(H = X(X^{'}X)^{-1}X^{'})$. This diagnostic measure is the sum of two components each of which has an interpretation. A large value for the first term indicates that the model has a poor fit (a large prediction error) and a large value for the second term indicates the presence of an outlier in the X-space. Similarly, Hadi pointed these diagnostic measures possess several desirable properties and it also supplemented by a graphical display that shows the source of influence. He suggested, $\left(H_i^2\right)$ for observations more than a cut-off of $E\left(H_i^2\right)+c\sqrt{V\left(H_i^2\right)}$ which is treated as a potential outlier, where ***c*** is an appropriate constant. Now rewrite (2.2) in terms of the estimated sum of the square residual $\hat{e}^t\hat{e} = s^2\left(n-q\right)$ and the alternative form as

$$H_i^2 = \frac{q\left(\hat{e}_i^2 / s^2\left(n-q\right)\right)}{\left(1-h_{ii}\right)\left(1-\left(\hat{e}_i^2 / s^2\left(n-q\right)\right)\right)} + \frac{h_{ii}}{1-h_{ii}} \tag{3.2}$$

From (2.1), (2.2), (3.1) and the Prediction matrix H, they do the same role in multiple regression analysis except for the estimated residual part in (3.1).Hence the authors logically extended the Hadi's influence to *p*-variate residual and the multivariate Hadi's influence measure of i^{th} observation is denoted as ${}_pH_i^2$. Using quadratic forms of the estimated residual of *p*-variate and rewrite (3.2) as multivariate Hadi's influence measure which is given as

$$ {}_pH_i^2 = \frac{q\left(E_i^T S_p^{-1} E_i / (n-q)\right)}{(1-h_{ii})\left(1-E_i^T S_p^{-1} E_i / (n-q)\right)} + \frac{h_{ii}}{1-h_{ii}} \quad (3.3) $$

From (3.3), E_i is the $p\ X\ 1$ vector of squared estimated residuals, $S_p = E^T E/(n-q)$ is the variance-covariance matrix of estimated residuals and the quadratic form $E_i^T S_p^{-1} E_i = R_i^2(1-h_{ii})$ can be written in terms of the *p*-variate squared internal studentized residual (R_i^2) (see Hossain and Naik (2006) and the final form of Multivariate Hadi's influence measure is given as

$$ {}_pH_i^2 = \frac{q\left(R_i^2/(n-q)\right)}{1-\left((1-h_{ii})R_i^2/(n-q)\right)} + \frac{h_{ii}}{1-h_{ii}} \quad (3.4) $$

Since the Multivariate Hadi's influence measure from (3.4) can also be visualized in terms of the *p*-variate squared internal studentized residual (R_i^2) and if *p*=1, then the Multivariate Hadi's influence measure was reduced like (3.1) which is the univariate version of Hadi's influence measure. The Proposed measure enjoys all the properties like (3.1) and the authors suggested, $({}_pH_i^2)$ for observations more than a cut-off of $E({}_pH^2)+c\sqrt{V({}_pH^2)}$ which is treated as a potential outlier in multivariate regression analysis.

4. Multivariate Extension of Andrew-Pregibon Statistic

Andrew-Pregibon(AP_i) statistic is also an interesting technique based on the volume of confidence ellipsoids. It is a simple fact, (AP_i) statistic is a measure of the influence of the i^{th} observation on the estimated regression coefficients can be based on the change in volume of confidence ellipsoids with or without the i^{th}

observation. The general form of the (AP_i)-statistic of the i^{th} observation is given by

$$AP_i = 1 - h_{ii} - \left(\hat{e}_i^2 / \hat{e}^T \hat{e} \right) \tag{4.1}$$

Where e_i^2 is the vector of squared estimated residuals, $\hat{e}^T \hat{e}$ is the sum of the squared estimated residuals and h_{ii} is the hat values of i^{th} observation or diagonal elements of the hat matrix $(H=X(X'X)^{-1}X')$. Andrew-Pregibon suggested (AP_i) for observations less than a cut-off of $1-2q/n$ or if it is very small and close to zero, which are treated as influential observations. They do not distinguish between a high leverage point in the factor space and an outlier in the response factor space. By using the fact $\hat{e}^t \hat{e} = s^2(n-q)$, Andrew-Pregibon statistic can also be written in an alternative form as

$$AP_i = (1-h_{ii}) \left(1 - \frac{1}{n-q} \left(\frac{\hat{e}_i^2}{s^2 (1-h_{ii})} \right) \right) \tag{4.2}$$

From (2.1), (2.2), (4.1) and the hat values in the Prediction matrix H, they do the same role in multiple regression analysis except for the estimated residual part in (4.1).Hence the authors logically extended the Andrew-Pregibon statistic to *p*-variate residual and the multivariate Andrew-Pregibon statistic of i^{th} observation is denoted as ${}_pAP_i$. Now using quadratic forms of the estimated residual of *p*-variate and rewrite (4.2) as multivariate Andrew-Pregibon statistic which is written as

$${}_pAP_i = (1-h_{ii}) \left(1 - \frac{1}{n-q} \left(\frac{E_i^T S_p^{-1} E_i}{1-h_{ii}} \right) \right) \tag{4.3}$$

From (4.3), E_i is the $p \times 1$ vector of squared estimated residuals, $S_p = E^T E / (n-q)$ is the variance-covariance matrix of estimated residuals and the quadratic form $E_i^T S_p^{-1} E_i = R_i^2 (1-h_{ii})$ can be written in terms of the *p*-variate squared internal studentized residual $\left(R_i^2\right)$ (see Hossain and Naik (2006) and the final form of Multivariate Andrew Pregibon statistic is given as

$$ {}_pAP_i = \left(1-h_{ii}\right)\left(1-\frac{R_i^2}{n-q}\right) \tag{4.4}$$

Since the Multivariate Andrew Pregibon statistic from (4.3) can also be shown in terms of the *p*-variate squared internal studentized residual $\left(R_i^2\right)$ and if p=1, then the Multivariate Andrew-Pregibon statistic was reduced like (4.2) which is the univariate version of the statistic. The Proposed measure is having similar properties to (4.1) and the authors suggested adopting the same calibration point $\left({}_pAP_i\right)$ to identify the influential observations in multivariate regression analysis.

5. Numerical Results and Discussion

In this section, the authors have shown a numerical illustration of evaluating the potential outliers based on multivariate Hadi's influence measure and identifying the influential observations by using Multivariate Andrew-Pregibon statistic of the i^{th} observation in a Multivariate regression model. The multivariate functional data in this study comprised of 19 different attributes about a car brand and the data was collected from 275 car users. A well-structured questionnaire was prepared and distributed to 300 customers and the questions were anchored at a five-point Likert scale from 1 to 5. After the data collection is over, only 275 completed questionnaires were used for analysis. The authors fitted multivariate regression models with 4 response variables such as Top of the mind awareness (y_1), Brand Recall (y_2), Brand Recognition (y_3), Brand-familiarity (y_4) and 15 predictors namely Satisfaction (x_1), Commitment (x_2), Liking (x_3), Price-Premium (x_4), Best-in category (x_5), Popularity (x_6), Brand leader (x_7), Innovation (x_8), Esteem (x_9), Performance (x_{10}), Value Association (x_{11}), Organizational Association (x_{12}), Brand differentiation (x_{13}), Celebrity Association (x_{14}), Animal Association (x_{15}) were used in this study. 11 different Multivariate regression models were fitted by using Stata version 13 and for each model, the Multivariate Hadi's influence measures $\left({}_pH_i^2\right)$ and Multivariate Andrew-Pregibon statistic $\left({}_pAP_i\right)$ were computed and the results are visualized along with the controls charts in the following Table-1, 2, and 3. (See Appendix).

6. Discussion

Table 1 and 2 visualizes the results of Multivariate Hadi's influence measure $\left({}_{p}H_{i}\right)$ of evaluating the potential outliers. 11 nested multivariate regression models were evaluated and the cut-offs' for different c values are shown in the table-1.we discard models-3,5,6 because the insignificance of the Breusch-Pagan test of independence confirms that the estimated residuals of the above said models are independent and uncorrelated. This show the highlighted models are not multivariate regression models and these are equivalent to fitting separate regression models with the respective dependent variable and the predictors. Hence, for these models, univariate Hadi's influence measure and Andrew-Pregibon statistic can be used to identify the potential outliers and influential observations, which is not the scope and objective of the paper. As far as the fitted models-1, 2, 4, are concern, the computed Bivariate Hadi's influence measure for (30,14,6,4), (30,14,6,4), (39,24,6,0) observations are above the cut-off value for various values of c=1,2,3,4 respectively. Hence these observations are said to be potential outliers. Similarly, models-7,8,9,10 are concern, (42,15,5,3),(37,17,8,3),(37,14,7,4),(46,15,3,2) observations are finalized as potential outliers based on the calculated trivariate Hadi's influence measure and in the same manner, in model-11, the calculated Multivariate Hadi's influence measure for (49,13,7,1) observations were above the cut-off and hence these observations are said to be the potential outliers. Table 3 shows the results of the Multivariate Andrew-Pregibon statistic $\left({}_{p}A_{i}\right)$ of evaluating the influential Observations. As far as models-1, 2, 4 are concern, (37, 32, 34) observations are treated as influential because the calculated Bivariate Andrew-Pregibon statistics for these observations are below the recommended Cut-off. Similarly, for the fitted models-7, 8,9,10, the calculated tri-variate Andrew-Pregibon statistic for the set of Observations (38,43,41,38) are less than the recommended cut-off which is treated as influential. Finally, in the fitted model-11,44 observations are considered to be influential based on the calculated Multivariate Andrew-Pregibon statistic. Finally, in table-2, 3, the identity of potential outliers and influential observations are shown model-wise and the results are illustrated heuristically with the help of the following control charts.

Control chart for fitted Model-1 shows the potential outliers based on Bivariate Hadi's influence measure $\left({}_{2}H_i^2\right)$.

Fig. 1

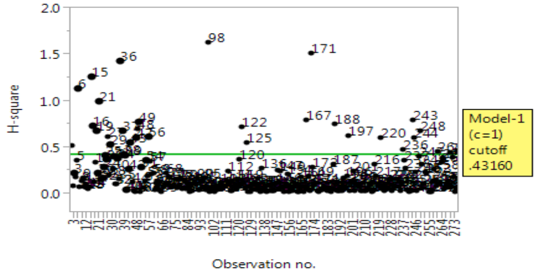


Fig. 2

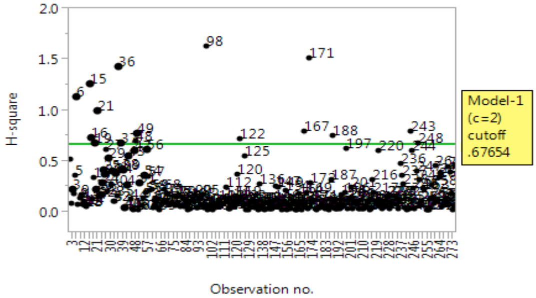


Fig. 3

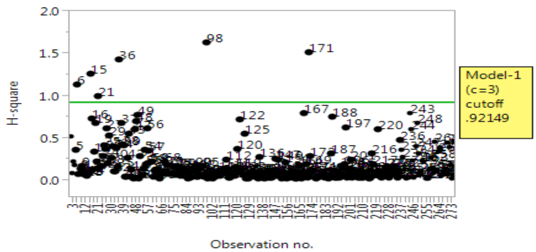
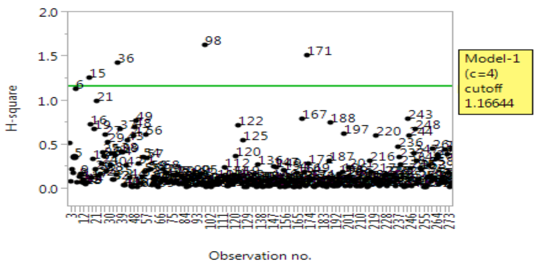


Fig. 4



Control chart for fitted Model-2 shows the potential outliers based on Bivariate Hadi's influence measure $\left({}_2H_i^2 \right)$.

Fig. 5

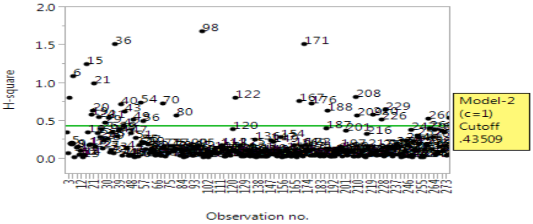


Fig. 6

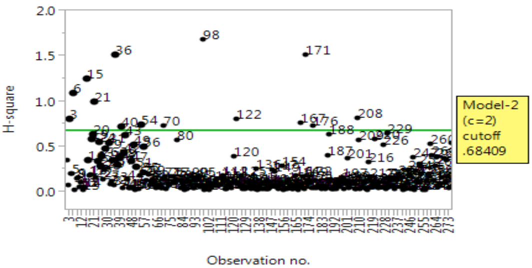


Fig. 7

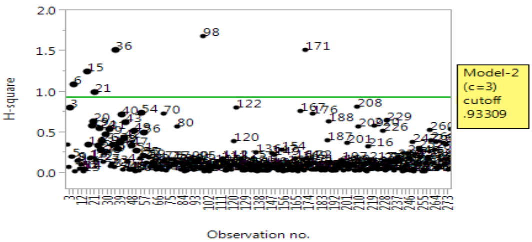
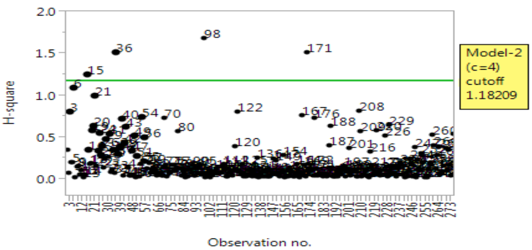


Fig. 8



Control chart for fitted Model-4 shows the potential outliers based on Bivariate Hadi's influence measure $\left({}_2H_i^2\right)$.

Fig. 9

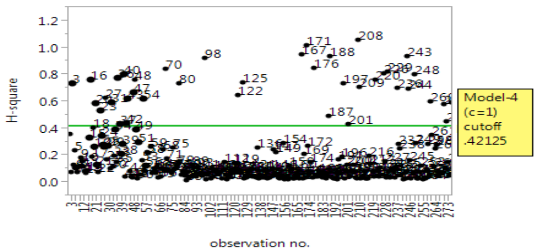


Fig. 10

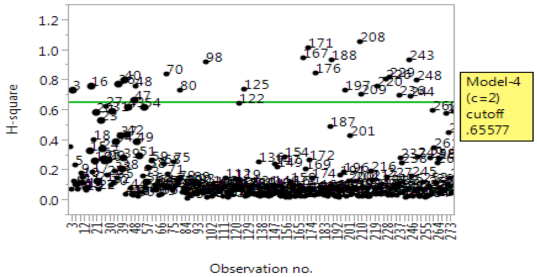


Fig. 11

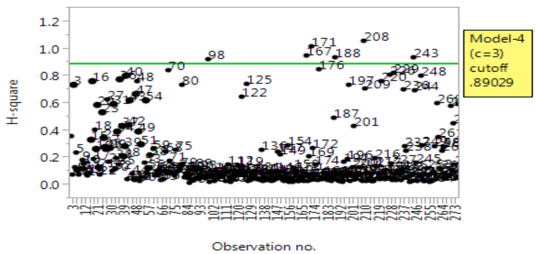
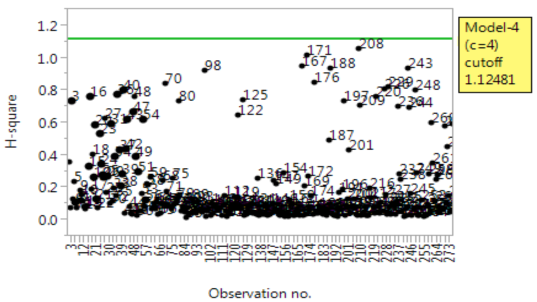


Fig. 12



Control chart for fitted Model-7 shows the potential outliers based on Tri-variate Hadi's influence measure $\left({}_{2}H_{i}^{2}\right)$

Fig. 13

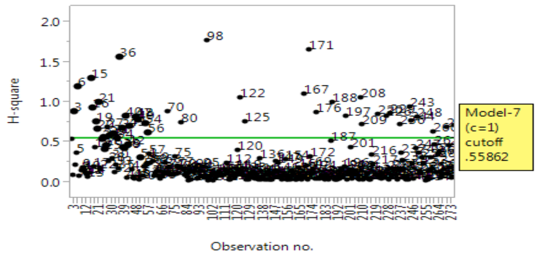


Fig. 14

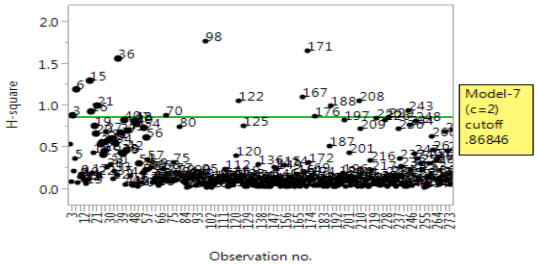


Fig. 15

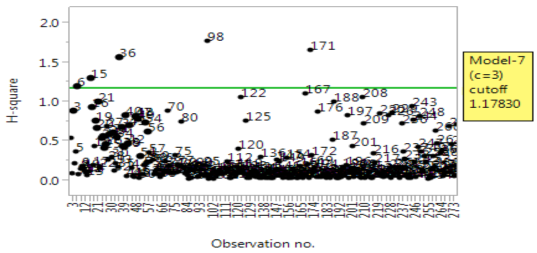
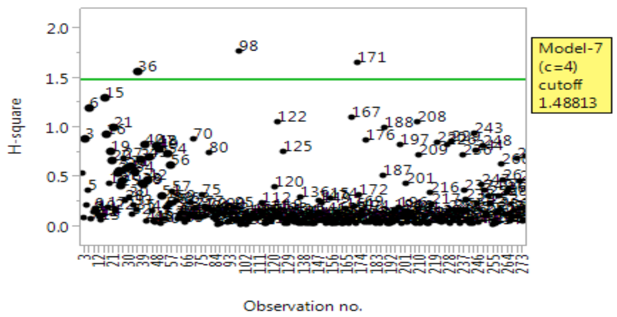


Fig. 16



Control chart for fitted Model-8 shows the potential outliers based on Tri-variate Hadi's influence measure $\left({}_{3}H_{i}^{2}\right)$

Fig. 17

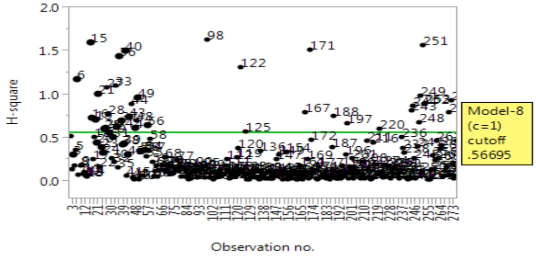


Fig. 18

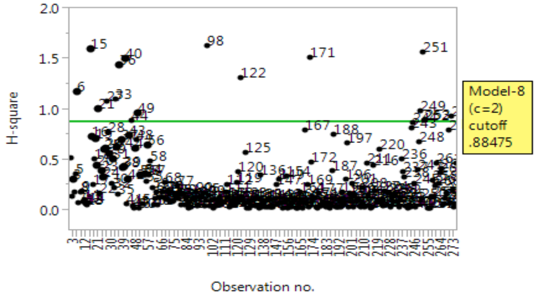


Fig. 19

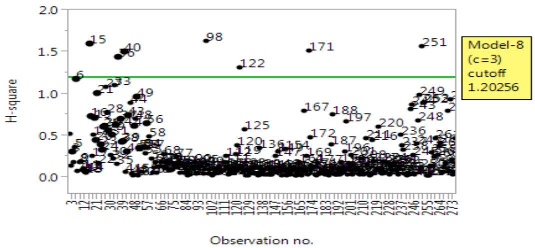
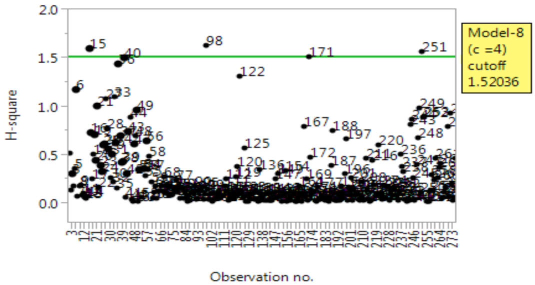


Fig. 20



Control chart for fitted Model-9 shows the potential outliers based on Tri-variate Hadi's influence measure $\left({}_3H_i^2\right)$

Fig. 21

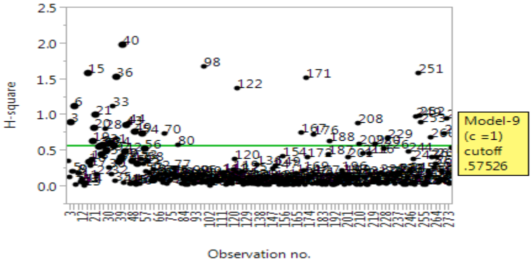


Fig. 22

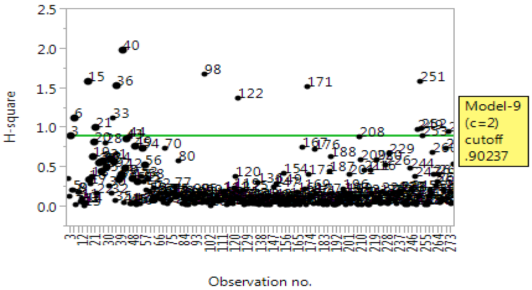


Fig. 23

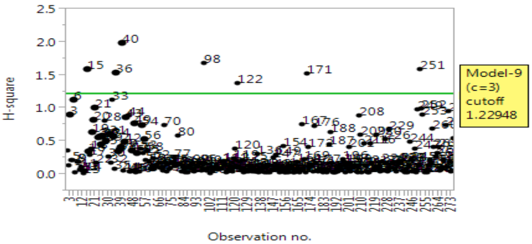
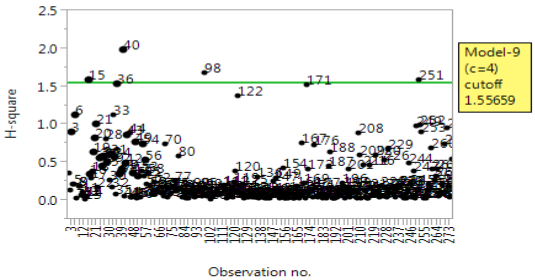


Fig. 24



Control chart for fitted Model-10 shows the potential outliers based on Tri-variate Hadi's influence measure $\left({}_3H_i^2\right)$

Fig. 25

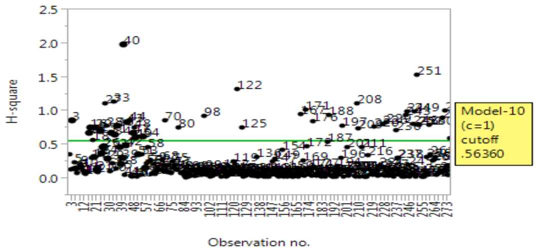


Fig. 26

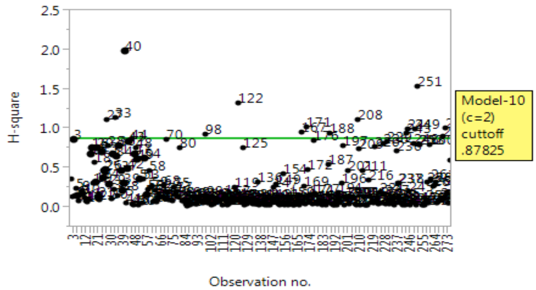


Fig. 27

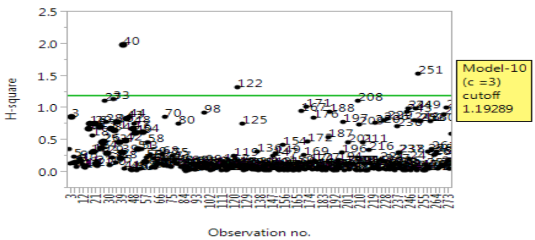
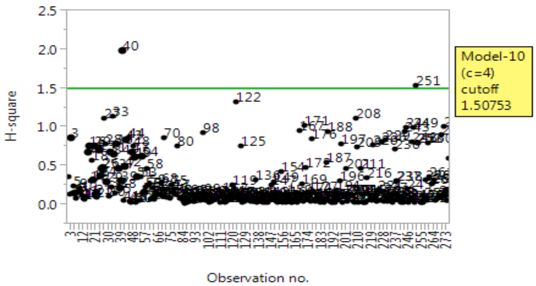


Fig. 28



Control chart for fitted Model-11 shows the potential outliers based on Multivariate Hadi's influence measure $\left({}_{4}H_i^2\right)$

Fig. 29

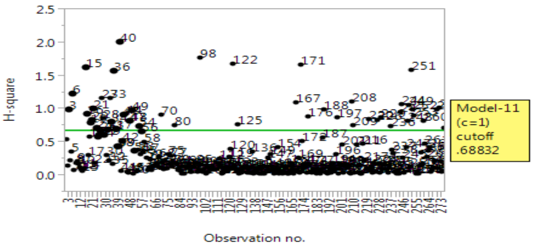


Fig. 30

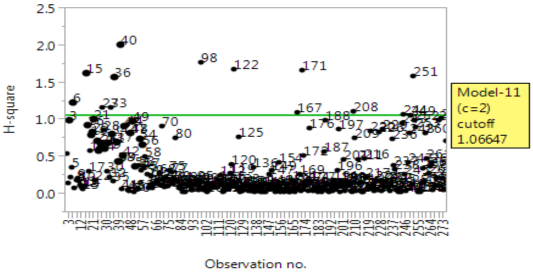


Fig. 31

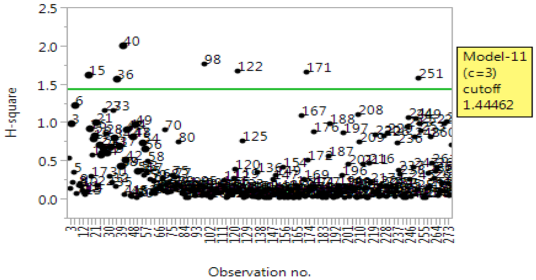
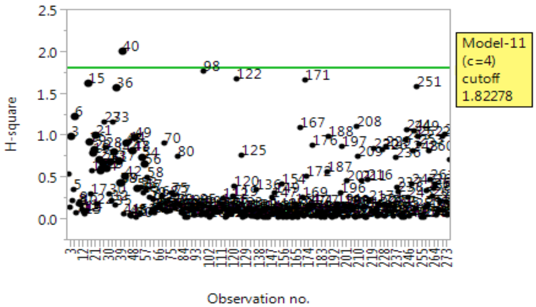


Fig. 32



Control chart for fitted Model shows the influential Observations based on Bivariate AP-statistic $\left({}_{2}AP_i\right)$

Fig. 33

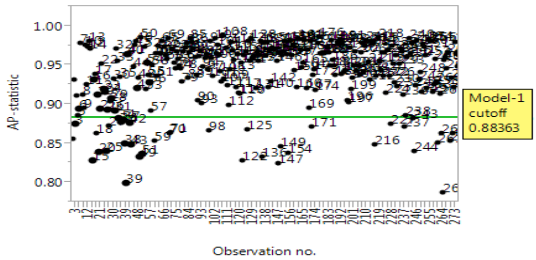


Fig. 34

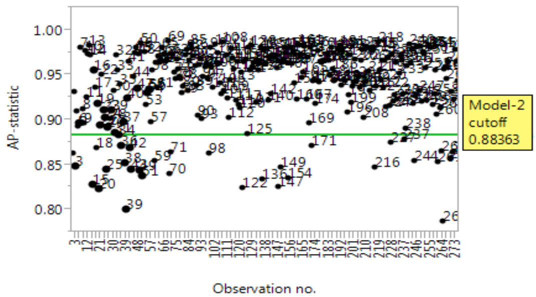
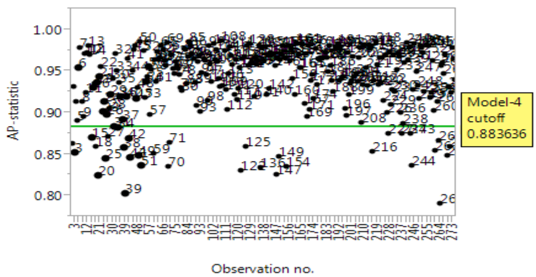


Fig. 35



Control chart for fitted Model shows the influential Observations based on Tri-variate AP-statistic $\left({}_3AP_i\right)$

Fig. 36

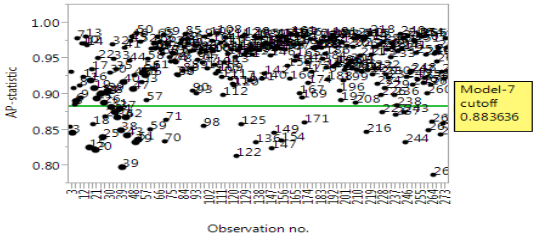


Fig. 37

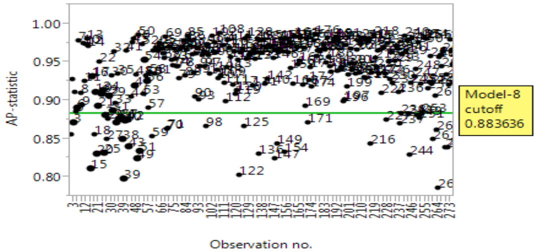


Fig. 38

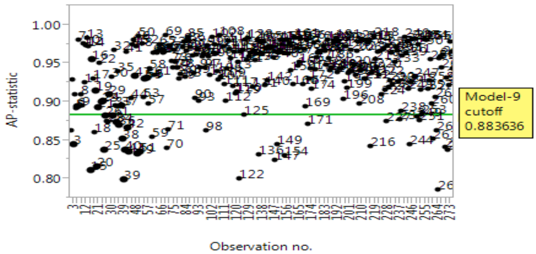
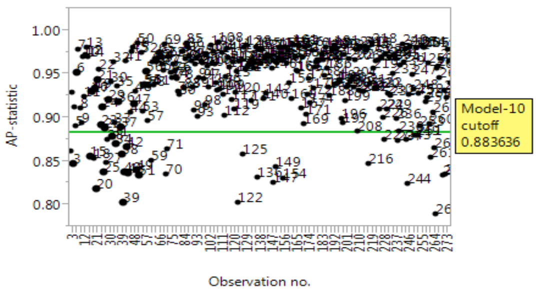
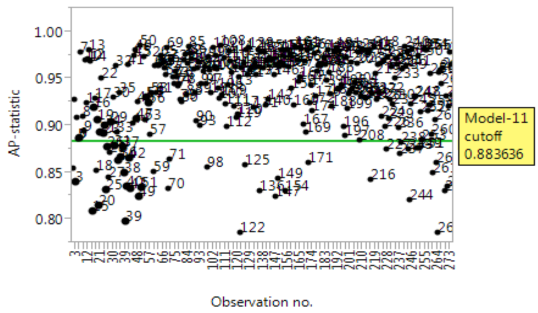


Fig. 39



Control chart for fitted Model shows the influential Observations based on Multivariate AP-statistic $\left({}_4AP_i\right)$

Fig. 40



7. Conclusion

From the previous sections, the authors proposed a multivariate extension of the univariate regression diagnostics namely Hadi's influence measure and Andrew-Pregibon statistic. The multivariate extension of these two frequently used diagnostic measure open the way to identify the potential outliers and influential observations in a linear multivariate regression model. To scrutinize the residuals in a fitted linear multivariate regression model, the authors recommended using these techniques along with the Breusch-Pagan test of independence will be more meaningful in identifying extreme observations. Finally, the authors suggested the exploration of the exact distribution of both measures will lead to a more scientific investigation of exact potential outliers and influential observations, which is left for future research.

References

Ali S. Hadi (1992): A new measure of overall potential influence in linear regression, *Computational Statistics & Data Analysis*, **14(1)**, 1-27

Behnken, D. W., and Draper, N. R. (1972): "Residuals and Their Variance Patterns," *Technometrics*, **14**, 101-111.

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

Cook, R.D. (1977): Detection of influential observation in linear regression. *Technometrics*, **19**, 15-18.

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

Cook, R. D., and Weisberg, S. (1982): *Residuals and influence in regression* (Vol. **5**). New York: Chapman and Hall.

Díaz-García, J. A., and González-Farías (2004): G. A note on the Cook's distance. *Journal of statistical planning and inference*, **120(1)**, 119-136.

Eubank, R.L. (1985): Diagnostics for smoothing splines. *J. Roy. Statist. Soc. Ser.* B **47**, 332–341.

Hoaglin, D.C., and Welsch, R.E. (1978): The Hat matrix in regression and ANOVA. *The Amer. Statist.*, **32**, 17-22.

Jayakumar, G. D. S., and Thomas, B. J. (2013): A new procedure of clustering based on multivariate outlier detection. *Journal of Data Science*, **11(1)**, 69-84.

Jayakumar, G. D. S., and Sulthan, A. (2015): Exact distribution of Cook's distance and identification of influential observations. *Hacettepe Journal of Mathematics and Statistics*, **44(1)**, 165-178.

Jayakumar, G. D. S., and Sulthan, A. (2016): Hadi's Influence Measure and Identification of Regression Clusters Based on Potential Outliers. *International Journal of Statistics & Economics™*, **17(2)**, 59-72.

Jayakumar, G. D. S., and Sulthan, A. (2017): On the Exact Distribution of Andrew-Pregibon (AP) Statistic and Identification of Influential Observations. *International Journal of Engineering and Future Technology™*, **14(3)**, 57-77.

Kim, C. (1996): Cook's distance in spline smoothing. *Statist. Probab. Lett.* **31**, 139–144.

Kim, C., Kim, W. (1998): Some diagnostics results in nonparametric density estimation. *Comm. Statist. Theory Methods* **27**, 291–303.

Kim, C., Lee, Y., Park, B.U. (2001): Cook's distance in local polynomial regression. *Statist. Probab. Lett.* **54**, 33–40.

Silverman, B.W. (1985): Some aspects of the spline smoothing approach to non-parametric regression curve 6tting (with discussion). *J. Roy. Statist. Soc. Ser.* **B** **47**, 1–52.

Thomas, W. (1991): Influence diagnostics for the cross-validated smoothing parameter in spline smoothing. *J. Amer. Statist. Assoc.* **86**, *693–698.*

Weisberg, S. (1980): *Applied linear regression*. New York: Wiley.

Authors and Affiliations

G. S. David Sam Jayakumar[1], A. Sulthan[2] and W. Samuel[3]

A. Sulthan
sulthan90@gmail.com

W. Samuel
wsamuel365@gmail.com

[2]KV Institute of Management and Information studies, Coimbatore-641107, India.
[1,3]Jamal Institute of Management, Tiruchirappalli – 620020, India.

Appendix

Table1: Multivariate Hadi's Influence measure $\left({}_pH_i^2\right)$ and Breusch-Pagan Test of Independence.

Model	Response variables (Y)	p	q	Predictors (X)	Multivariate Hadi's Influence measure $\left({}_pH_i^2\right)$								Breusch-Pagan Test of Independence	
					$c=1$	${}_pH_i^2$ >A	$c=2$	${}_pH_i^2$ >A	$c=3$	${}_pH_i^2$ > A	$c=4$	${}_pH_i^2$ >A	chi-square statistic	df
1	y_1, y_2	2	16	$x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}, x_{13}, x_{14}, x_{15}$	0.43160	30	0.676	14	0.921	6	1.166	4	*14.0	1
2	y_1, y_3				0.43509	30	0.684	14	0.933	6	1.182	4	*13.2	1
3	y_1, y_4				-	-	-	-	-	-	-	-	0.091	1
4	y_2, y_3				0.42125	39	0.655	24	0.890	6	1.124	0	*39.9	1
5	y_2, y_4				-	-	-	-	-	-	-	-	1.273	1
6	y_3, y_4				-	-	-	-	-	-	-	-	0.514	1
7	y_1, y_2, y_3	3			0.55862	42	0.868	15	1.178	5	1.488	3	*67.2	3
8	y_1, y_2, y_4				0.56695	37	0.884	17	1.202	8	1.520	3	*15.3	3
9	y_1, y_3, y_4				0.57526	37	0.902	14	1.229	7	1.556	4	*13.8	3
10	y_2, y_3, y_4				0.56360	46	0.878	15	1.192	3	1.507	2	*41.7	3
11	y_1, y_2, y_3, y_4	4			0.68832	49	1.066	13	1.444	7	1.822	1	*69.1	6

n=275 ***p**-no .of response variables* ***q**-dimensions of* β A -Cut-off$\left({}_pH_i^2\right)=\left(E\left({}_pH_i^2\right)+c\sqrt{V\left({}_pH_i^2\right)}\right)$ *p-value<0.01

Table 2: Model wise Identity of Potential Outliers.

Model	Potential Outliers no.							
	m	**c = 1**	***m***	**c = 2**	***m***	**c = 3**	***m***	**c = 4**
1	30	1,6,15,16,19,21,27,29,36,37,43,47,48,49,56,98,122, 125,167,171,188,197,220,236,243,244,248,261,269, 275	14	6,15,16,19,21,36,48,49,98, 122,167,171,188,243	6	6,15,21,36,98, 171	4	15,36,98, 171
2	30	3,6,15,19,20,21,24,29,31,36,40,43,49,54,56,70,80, 98,122,167,171,176,188,208,209,220,226,229,260, 275	14	3,6,15,21,36,40,54,70,98, 122,167,171,176,208	6	6,15,21,36,98, 171	4	15,36,98, 171
4	39	3,16,20,23,27,31,36,37,40,42,43,47,48,54,70,80,98, 122,125,167,171,176,187,188,197,201,208,209,220, 226,229,236,243,244,248,260,269,271,275	24	3,16,36,40,47,48,70,80,98, 125,167,171,176,188,197, 208,209,220,226,229,236, 243,244,248	6	98,167,171,188, 208,243	0	-
7	42	3,6,15,16,19,20,21,24,27,29,31,36,37,40,43,4748, 4954,56,70,80,98,122,125,167,171,176,188,197,208, 209,220,226,229,236,243,244,248,260,269,275	15	3,6,15,16,21,36,70,98,122, 167,171,176,188,208,243	5	6,15,36,98,171	3	36,98,171
8	37	6,15,16,19,21,25,26,27,28,33,34,36,37,40,43,44,47, 48,49,56,98,122,125,167,171,188,197,220,243,244, 248,249,251,252,253,269,271,	17	6,15,21,27,33,36,40,44,49, 98,122,171,249,251,252, 253,271	8	6,15,36,40,98, 122,171,251	3	15,98,251
9	37	3,6,15,19,20,21,24,26,28,31,33,34,36,40,43,44,49, 54,70,80,98,122,167,171,176,188,208,209,220,229, 249,251,252,253,260,269,271	14	3,6,15,21,33,36,40,98,122, 171,249,251,252,271	7	15,36,40,98,122, 171,251	4	15,40,98, 251
10	46	3,15,16,18,20,23,27,28,31,33,34,36,40,43,44,4748, 49,54,70,80,98,122,125,167,171,176,188,197,208, 209,220,226,229,236,243,244,248,249,251,252,253, 260,269,271,275	15	27,33,40,98,122,167,171, 188,208,243,244,249,251, 269,271	3	40,122,251,	2	40, 251
11	49	3,6,15,16,19,20,21,23,27,28,31,33,34,36,37,40,43,44, 47,48,49,54,70,80,98,122,125,167,171,176,188,197, 208,209,220,226,229,236,243,244,248,249,251,252, 253,260,269,271,275	13	6,15,27,33,36,40,98,122, 167,171,208,244,251	7	15,36,40,98,122, 171,251	1	40

m-no. of Potential outliers

Table 3: Multivariate Andrew-Pregibon statistic $\left({}_{p}AP_{i}\right)$, Breusch-Pagan Test of Independence and Model Wise Identification of Influential Observations.

Model	Response variables (Y)	p	q	Predictors (X)	Multivariate Andrew-Pregibon statistic $\left({}_{p}AP_{i}\right)$		Breusch-Pagan Test of Independence	
					$\left({}_{p}AP_{i}\right)$ <Cut-off	Influential Observation no.	chi-square statistic	*df*
1	y_1, y_2	2	16	$x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}, x_{13}, x_{14}, x_{15}$	37	1,3,15,18,20,25,27,34,36,37,38,39,42,43,49,51,59,70,71,98,122,125,136,147149,154,171,216,227,237,243,244,261,263,264,269,271	*14.033	1
2	y_1, y_3				32	1,3,15,18,20,25,36,38,39,42,43,49,51,59,70,71,98,122,136,147,149,154,171,216227,237,244,261,263,264,269,271	*13.279	1
3	y_1, y_4				-	-	0.091	1
4	y_2, y_3				34	1,3,15,18,20,25,27,31,34,38,39,42,43,49,51,59,70,71,122,125,136,147,149,154216,227,237,243,244,261,263,264,269,271	*39.945	1
5	y_2, y_4				-	-	1.273	1
6	y_3, y_4				-	-	0.514	1
7	y_1, y_2, y_3	3			38	1,3,15,18,20,25,27,31,34,36,37,38,39,42,43,49,51,59,70,71,98,122,125,136,147,149154,171,216,227 ,	*67.257	3
8	y_1, y_2, y_4				43	1,3,15,18,20,25,26,27,28,34,36,37,38,39,40,42,43,49,51,59,70,71,98,122,125,136,147149,154,171,216,227,237,238,243,244,249,251,261,263,264,269	*15.396	3
9	y_1, y_3, y_4				41	1,3,15,18,20,25,26,27,28,31,34,36,38,39,40,42,43,49,51,59,70,71,98,122,125,136,147,149,154,171,216,227,237,244,249,251,261,263,264,269,271	*13.884	3
10	y_2, y_3, y_4				38	1,3,15,18,20,25,27,28,31,34,38,39,40,42,43,49,51,59,70,71,122,125,136,147,149,154216,227,237,243,244,249,251,261,263,264,269,271	*41.732	3
11	y_1, y_2, y_3, y_4	4			44	1,3,15,18,20,25,26,27,28,31,34,36,37,38,39,40,42,43,49,51,59,70,71,98,122,125,136147,149,154,171,216,227,237,238,243,244,249,251,261,263,264,269,271	*69.135	6

n =275 ***p**-no.of response variables* ***q**-dimensions of* β Cut-off $({}_{p}AP_{i}) = 1 - 2q/n$ = 0.883636 **p*-value<0.01