Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
`http://siba-ese.unisalento.it/index.php/ejasa/index`
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v11n1p21

On using the distribution of Cook-Weisberg statistic and identification of influential observations

By Jayakumar, Sulthan, Samuel

Published: 26 April 2018

Electronic Journal of Applied Statistical Analysis
Vol. 11, Issue 01, April 2018, 21-41
DOI: 10.1285/i20705948v11n1p21

On using the distribution of Cook-Weisberg statistic and identification of influential observations

Dr. G.S. David Sam Jayakumar*, A. Sulthan, and W. Samuel

Jamal Institute of Management
Tiruchirappalli 620 020, India

Published: 26 April 2018

This paper proposed the distribution of Cook-Weisberg statistic used to evaluate the influential observations in linear multiple regression analysis. The authors explored the relationship between the CW statistic and COV ratio (Co-variance ratio) in terms of two independent F-ratio's and they show the derived density function of the measure in a series expression form. Moreover, the first two moments of the distribution are derived in terms of Beta, di-gamma, poly-gamma functions, upper control limit of CW-statistic is also established and the authors computed the critical points of CW-statistic at 5% and 1% significance level for different sample sizes and varying no.of predictors. Finally, the numerical example shows the identification of the influential observations and the results extracted from the proposed approaches are more scientific, systematic and it's exactness outperforms the traditional approach.

keywords: CW-statistic, influential observations, COV-Ratio, F-ratio,Beta function

1 Introduction and Related work

The Studentized residuals and the plot of the residuals were considered the most appropriate statistical devices to detect potentially critical observations in the literature before the third quarter of the 20th century. Behnken and Draper (1972) have clarified that the estimated variance of the residuals, include pertinent information beyond

*Corresponding author: samjaya77@gmail.com

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

that provided by plots of residuals or studentized residuals. Similarly, they discussed the variances of residuals in several more complicated designs. Hoaglin and Welsch (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) were the first to establish a simple measure, Di that incorporates information from the 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 with reference to parametric regression models by the pioneers namely Cook (1977), Cook and Weisberg (1982), Belsley et al. (2005) and Chatterjee and Hadi (2009) 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 in local polynomial regression. The phrase 'influence measures' has glimpsed a great surge of research interests. The developments of different measures are investigated to identify the influential observation from the early criteria of Cook's to the present and a definition about influence, which appears most suitable, is given by Belsley et al. (2005). 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 of residual error, and it is monotonically increasing in the leverage values and in the squared residuals.Recently, Dıaz-Garcıa and González-Farıas (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. Considering the above reviews, the authors proposed the exact distribution of Cook-Weisberg statistic which in need to exactly identify the influential data points and it is discussed in the subsequent sections.

2 Relationship between Cook-Weisberg statistic and Covariance Ratio

The multiple linear regression model with random error is given by

$$Y = X\beta + e \tag{1}$$

where $Y_{(n\times1)}$ is the vector of values of the dependent variable, $X_{(n\times(p+1))}$is an full column rank matrix of predictors, $\beta_{(k\times1)}$ is the vector of beta co-efficients or partial regression co-efficients and $e_{(n\times1)}$is the vector of residual followed normal distribution N $(0,\sigma_e^2 I_n)$ respectively. From (1), statisticians concentrate and give importance to the error diagnostics such as outlier detection, identification of leverage points and evaluation of influential observations etc. Several error diagnostics techniques exist in the literature proposed by statisticians, but Cook-Weisberg statistic which is having an abbreviation $(CW)$is the interesting technique to evaluate the influential observations proposed by Cook and Weisberg(1980).It is the logarithm of the ratio of the volume of the $(1-\alpha)\,100\%$confidence ellipsoids with and without the *ith* observation as a measure of influence. The original version of the CW-statistic was reduced and it is given as

$$CW_i = -\frac{1}{2}\log\left(1-h_{ii}\right) + \frac{p}{2}\log\left(\frac{(n-p-2)\,F_{(\alpha;p+1,n-p-1)}}{\left(n-p-t_i^2\right)F_{(\alpha;p+1,n-p-2)}}\right) \tag{2}$$

From (2), CW_i is visualized in terms of the squared external studentized residual (t_i) follows t-distribution with *n-p-2* degrees of freedom,h_{ii} are the hat-values or the diagonal elements of the hat matrix (H) and$F_{(\alpha;p+1,n-p-1)}$, $F_{(\alpha;p+1,n-p-2)}$ are the upper α points of the F-distribution with the appropriate degrees of freedom. Further(2)can be modified and exhibit in a more convenient form as

$$CW_i = -\frac{1}{2}\log\left(CVR_i\right) + \frac{p}{2}\log\left(\frac{F_{(\alpha;p+1,n-p-1)}}{F_{(\alpha;p+1,n-p-2)}}\right) \tag{3}$$

From (3), Cook and Weisberg explored the relationship between the CW-statistic and the (CVR) stands for Covariance-ratio of the *ith* observation. They argue, if this quantity CW is large and positive, then deletion of the *ith* observation will result in a substantial decrease in volume and if it is large and negative, that will result in a substantial increase in volume. Hence, if the volume of the ellipsoid will increase or decrease or if the quantum of increase or decrease is very large, then the particular observation is said to be influential. Though the CW-statistic is scientific, will it helps the statisticians to exactly identify the influential observations in a sample? The answer will be no? Because Cook and Weisberg failed to give an exact calibration point which in need to accurately identify the influential observation. For this, the authors made an attempt to propose the exact distribution of CW-statistic in a finite sample and make this approach more scientific by fixing meaning full criterion as a calibration point. To identify the exact influential observations in a finite sample, the authors proposed the exact distribution for CW-statistic by utilizing its relationship with Co-variance ratio (CVR) and it discussed in the next section.

3 Relationship between Covariance Ratio and F-ratios

Kuh-Welsch ratio or COVRATIO is the mnemonic abbreviation of the term 'Covariance ratio' is the interesting technique which is also based on the volume of confidence ellipsoids. It is a simple fact, $(CVR_i)$is a measure of the influence of the *ith* observation on the variance of estimated regression co-efficients $|[|\,]\widehat{\Sigma}_{\hat{\beta}}$ can be measured by comparing the ratio of the two determinants $|[|]\widehat{\Sigma}_{\hat{\beta}_{(i)}}$ and $|[|]\widehat{\Sigma}_{\hat{\beta}}$,where $\left(\widehat{\Sigma}_{\hat{\beta}_{(i)}}\right)$ is the variance of estimated regression co-efficients without the *ith* observation. The general form of the CVR_i of the *ith* observation is given by

$$CVR_i = \frac{|[|]\widehat{\Sigma}_{\hat{\beta}_{(i)}}}{|[|]\widehat{\Sigma}_{\hat{\beta}}} = \frac{\left|\overset{\wedge}{\sigma}{}^2_{e_{(i)}}\left(X^T_{(i)}X_{(i)}\right)^{-1}\right|}{\left|\overset{\wedge}{\sigma}{}^2_e\left(X^TX\right)^{-1}\right|} \tag{4}$$

Where $\overset{\wedge}{\sigma}{}^2_e$,$\overset{\wedge}{\sigma}{}^2_{e_{(i)}}$ are the estimated variance of residuals, $\left(X^TX\right)$,$\left(X^T_{(i)}X_{(i)}\right)$ are the design matrix of predictors with and without the *ith* observation respectively. Kuh-Welsch suggested, absolute value of the covariance ratio minus one $|CVR_i - 1|$ for observations exceeds the rough calibration point of $3\,(p+1)\,/n$ which are treated as influential observations. Covariance ratio can also be written in an alternative form in terms of the hat values (h_{ii}) and it is given as

$$CVR_i = \left(\frac{\overset{\wedge}{\sigma}{}^2_{e(i)}}{\overset{\wedge}{\sigma}{}^2_e}\right)^p\left(\frac{1}{1-h_{ii}}\right) \tag{5}$$

It is known the estimate of the true error variance without the *ith* observation is $\overset{\wedge}{\sigma}{}^2_{e(i)} = \left((n-p-1)\,\overset{\wedge}{\sigma}{}^2_e - \hat{e}{}^2_i\,/(1-h_{ii})\right)/n{-}p{-}2$ and substitute $\overset{\wedge}{\sigma}{}^2_{e(i)}\,/\,\overset{\wedge}{\sigma}{}^2_e = \left((n-p-1)-r_i^2\right)/(n-p-2)$in (5) to get

$$CVR_i = \left(\frac{(n-p-1)-r_i^2}{n-p-2}\right)^p\left(\frac{1}{1-h_{ii}}\right) \tag{6}$$

From (6), it is visualized in terms of the internally studentized residual (r_i)which is equal to $\overset{\wedge}{e_i}\,/s\sqrt{1-h_{ii}}$, where S is the standard deviation of the estimated residuals. To propose the exact distribution for CW-statistic ,the authors utilizing the relationship among the (CVR_i), internally studentized residual(r_i) and hat elements(h_{ii}).The terms r_i and h_{ii} are independent, because the computation of (r_i) involves the error term $e_i \sim N(0,\sigma_e^2)$ and h_{ii} values involves the set of predictors $(H = X(X'X)^{-1}X')$.Therefore, from the property of least squares $E(eX) = 0$, so r_i and h_{ii} are also uncorrelated and independent. Using this assumption, we first determine the distribution of r_i based on the relationship given by Weisberg (1980) as

$$t_i = r_i\sqrt{\frac{n-p-2}{(n-p-1)-r_i^2}} \sim t_{(n-p-2)} \tag{7}$$

From (7)it follows student's t- distribution with $(n-p-2)$ degrees of freedom and it can be written in terms of the F-ratio as

$$r_i^2 = \frac{(n-p-1)t_i^2}{(n-p-2)+t_i^2}$$

$$r_i^2 = \frac{(n-p-1)F_{i(1,n-p-2)}}{(n-p-2)+F_{i(1,n-p-2)}} \tag{8}$$

From (8),if $t_i$follows student's t- distribution with $(n-p-2)$ degrees of freedom, then t_i^2 follows $F_{(1,n-p-2)}$ distribution with $(1, n-p-2)$ degrees of freedom. Similarly, authors identify the distribution of h_{ii} based on the relationship proposed by Belsley et al (1980) and they show when the set of predictors in a linear regression model followed multivariate normal distribution with(μ_X, Σ_X), then

$$\frac{(n-p)\left(h_{ii}-(1/n)\right)}{(p-1)(1-h_{ii})} \sim F_{(p-1,n-p)} \tag{9}$$

From (9) it follows F-distribution with $(p-1, n-p)$ degrees of freedom and it can be written in an alternative form as

$$h_{ii} = \frac{\left(((p-1)/(n-p))F_{i(p-1,n-p)}\right)+1/n}{1+((p-1)/(n-p))F_{i(p-1,n-p)}} \tag{10}$$

In order to derive the exact distribution of (CVR_i),substitute (8) and (10) in (6), it gives the CVR_i in terms of the two independent F-ratios with $(1, n-p-2)$ and $(p-1, n-p)$ degrees of freedom respectively and the relationship is given as

$$CVR_i = \frac{n}{n-1}\left(\frac{n-p-1}{n-p-2}\right)^p \left(\frac{1}{1+\frac{1}{n-p-2}F_{i(1,n-p-2)}}\right)^p \left(1+\frac{p-1}{n-p}F_{i(p-1,n-p)}\right) \tag{11}$$

$$CVR_i = \frac{n}{n-1}\left(\frac{n-p-1}{n-p-2}\right)^p \frac{\left(1/\left(1+\frac{1}{n-p-2}F_{i(1,n-p-2)}\right)\right)^p}{1/\left(1+\frac{p-1}{n-p}F_{i(p-1,n-p)}\right)} \tag{12}$$

From (12), it can be further simplified and (CVR_i) is expressed in terms of two independent beta variables namely θ_{1i} and θ_{2i} of the first kind by using the following facts

$$\frac{1}{1+\frac{1}{n-p-2}F_{i(1,n-p-2)}} = \theta_{1i} \sim \beta_1\left(\frac{n-p-2}{2}, \frac{1}{2}\right) \tag{13}$$

$$\frac{1}{1+\frac{p-1}{n-p}F_{i(p-1,n-p)}} = \theta_{2i} \sim \beta_1\left(\frac{n-p}{2}, \frac{p-1}{2}\right) \tag{14}$$

Then, without loss of generality (12) can be written as

$$CVR_i = \lambda(p,n)\frac{\theta_{1i}^p}{\theta_{2i}} \tag{15}$$

From (15), the authors shown the (CVR_i) measure in terms of $\theta_{1i} \sim \beta_1\left(\frac{n-p}{2}, \frac{p-1}{2}\right)$ and $\theta_{2i} \sim \beta_1\left(\frac{n-p-2}{2}, \frac{1}{2}\right)$ which followed beta distribution of first kind and $\lambda(p,n) = \frac{n}{n-1}\left(\frac{n-p-1}{n-p-2}\right)^p$ is the normalizing function with two shape parameters p and n respectively. To avoid complexity, further the relationship from (15) modified as

$$\frac{CVR_i}{\lambda(p,n)} = \frac{\theta_{1i}^p}{\theta_{2i}} = \psi_i \tag{16}$$

Based on the identified relationship from (16), the authors derived the exact distribution of the CVR measure and it is discussed in the next section.

4 Exact Distribution of Covariance ratio

Using the technique of two-dimensional Jacobian of transformation, the joint probability density function of the two beta variables of Kind-1 namelyθ_{1i},θ_{2i} were transformed into density function of ψ_i and it is given as

$$f(\psi_i, u_i) = f(\theta_{1i}, \theta_{2i})\,|J| \tag{17}$$

From (17), it is known $\theta_{1i}$and θ_{2i} are independent then rewrite (17) as

$$f(\psi_i, u_i) = f(\theta_{1i})\,f(\theta_{2i})\,|J| \tag{18}$$

Using the change of variable technique, substitute $\theta_{2i} = u_i$ in (16) it gives

$$\theta_{1i} = (\psi_i u_i)^{1/p} \tag{19}$$

Then partially differentiate (19), and compute the Jacobian determinant in (18) as

$$f(\psi_i, u_i) = f(\theta_{1i})\,f(\theta_{2i})\left|\frac{\partial(\theta_{1i}, \theta_{2i})}{\partial(\psi_i, u_i)}\right| \tag{20}$$

$$f(\psi_i, u_i) = f(\theta_{1i})\,f(\theta_{2i})\begin{vmatrix} \frac{\partial\theta_{1i}}{\partial\psi_i} & \frac{\partial\theta_{1i}}{\partial u_i} \\ \frac{\partial\theta_{2i}}{\partial\psi_i} & \frac{\partial\theta_{2i}}{\partial u_i} \end{vmatrix} \tag{21}$$

From (18), it is known θ_{1i} and $\theta_{2i}$are independent, then the density function of the joint distribution of θ_{1i} and θ_{2i} is given as

$$f(\theta_{1i}, \theta_{2i}) = \frac{1}{B\left(\frac{n-p-2}{2}, \frac{1}{2}\right)}\theta_{1i}^{\frac{n-p-2}{2}-1}(1-\theta_{1i})^{\frac{1}{2}-1} \times \frac{1}{B\left(\frac{n-p}{2}, \frac{p-1}{2}\right)}\theta_{2i}^{\frac{n-p}{2}-1}(1-\theta_{2i})^{\frac{p-1}{2}-1} \tag{22}$$

where $0 \le \theta_{1i}, \theta_{2i} \le 1$,$n, p > 0$ and

$$\frac{\partial(\theta_{1i}, \theta_{2i})}{\partial(\psi_i, u_i)} = \begin{vmatrix} u_i^{1/p}\psi_i^{(1/p)-1}/p & \psi_i^{1/p}u_i^{(1/p)-1}/p \\ 0 & 1 \end{vmatrix} = u_i^{1/p}\psi_i^{(1/p)-1}/p \tag{23}$$

Then substitute (22) and (23) in (21) in terms of the substitution of u_i, to get the joint distribution of ψ_i and u_i as

$$f(\psi_i, u_i) = \frac{1}{B\left(\frac{n-p-2}{2}, \frac{1}{2}\right)} \left((u_i\psi_i)^{1/p}\right)^{\frac{n-p-2}{2}-1} \left(1-(u_i\psi_i)^{1/p}\right)^{\frac{1}{2}-1} \times \frac{1}{B\left(\frac{n-p}{2}, \frac{p-1}{2}\right)} u_i^{\frac{n-p}{2}-1} (1-u_i)^{\frac{p-1}{2}-1} \times |J| \quad (24)$$

where $0 \le \psi_i \le 1, 0 \le u_i \le 1$ and $|J| = u_i^{1/p}\psi_i^{(1/p)-1}/p$.

Rearrange (24) using binomial expansion and integrate with respect tou_i, to get the marginal distribution of ψ_i *as*

$$f(\psi_i; p, n) = \frac{(1/\lambda(p,n))^{\frac{n-p-2}{2p}-1}}{pB\left(\frac{n-p-2}{2}, \frac{1}{2}\right) B\left(\frac{n-p}{2}, \frac{p-1}{2}\right)} (\lambda(p,n)\psi_i)^{\frac{n-p-2}{2p}-1} \sum_{r=0}^{\infty} \begin{pmatrix} -1/2 \\ r \end{pmatrix} \psi_i^{r/p} \int_0^1 u_i^{\frac{(n-p)(p+1)+2(r-1)}{2p}-1} (1-u_i)^{\frac{p-1}{2}-1} du_i \quad (25)$$

where $0 \le \psi_i \le 1$,$n, p > 0$, $n > p$ it is known, from (25)

$$\int_0^1 u_i^{\frac{(n-p)(p+1)+2(r-1)}{2p}-1} (1-u_i)^{\frac{p-1}{2}-1} du_i = B\left(\frac{(n-p)(p+1)+2(r-1)}{2p}, \frac{p-1}{2}\right) \quad (26)$$

Then substitute (26) in (25) and arrange the terms, to get the density function of ψ_i in the series expression form as

$$f(\psi_i; p, n) = \frac{(\lambda(p,n))^2/p}{B\left(\frac{n-p-2}{2}, \frac{1}{2}\right) B\left(\frac{n-p}{2}, \frac{p-1}{2}\right)} \psi_i^{\frac{n-p-2}{2p}-1} \sum_{r=0}^{\infty} \begin{pmatrix} -1/2 \\ r \end{pmatrix} \psi_i^{r/p} B\left(\frac{(n-p)(p+1)+2(r-1)}{2p}, \frac{p-1}{2}\right) \quad (27)$$

where $0 \le \psi_i \le 1$,$n, p > 0$, $n > p$.

In order to derive the density function of CVR_i measure, the authors again utilize the relationship between ψ_i and CVR_i and it is known $0 \le \psi_i \le 1$, then $0 \le CVR_i \le \lambda(p,n)$.Hence, from (16) using one dimensional Jacobian of transformation, the density function of CVR_i can be written as

$$f(CVR_i) = f(\psi_i)|J| \quad (28)$$

$$f(CVR_i) = f(\psi_i)\left|\frac{d\psi_i}{d(CVR_i)}\right| \quad (29)$$

Then substitute$\psi_i = \frac{CVR_i}{\lambda(p,n)}, \frac{d\psi_i}{d(CVR_i)} = \frac{1}{\lambda(p,n)}$ and (27) in (29), to get the final form of the density function of CVR_i as

$$f(CVR_i; p, n) = \frac{\lambda(p,n)/p}{B\left(\frac{n-p-2}{2}, \frac{1}{2}\right) B\left(\frac{n-p}{2}, \frac{p-1}{2}\right)} \left(\frac{CVR_i}{\lambda(p,n)}\right)^{\frac{n-p-2}{2p}-1} \sum_{r=0}^{\infty} \begin{pmatrix} -1/2 \\ r \end{pmatrix} \left(\frac{CVR_i}{\lambda(p,n)}\right)^{r/p} B\left(\frac{(n-p)(p+1)+2(r-1)}{2p}, \frac{p-1}{2}\right) \quad (30)$$

where$0 \le CVR_i \le \lambda(p,n)$,$n, p > 0$, $n > p$ and $\lambda(p,n) = \frac{n}{n-1}\left(\frac{n-p-1}{n-p-2}\right)^p$.

From (30), it is the density function of CVR_i measure which always lies between 0 and $\lambda(p,n)$,if $n \to \infty$ then $0 \le CVR_i \le 1$.This shows, if the sample size is very large, then the upper limit of the CVR_i will converge to 1.Moreover,(30) involves the normalizing constants in terms of Beta functions namely $B(\frac{n-p-2}{2}, \frac{1}{2})$,$B(\frac{n-p}{2}, \frac{p-1}{2})$ and $\lambda(p,n)$ with two shape parameters (p,n),where n is the sample size and p is the no. of predictors used in a multiple linear regression model. The derivation of the distribution of CW-statistic by using (30) will be discussed in the next section.

5 Distribution of Cook-Weisberg statistic

From (3) and (30), using one dimensional Jacobian of transformation, the density function of CW_i statistic can be written as

$$f(CW_i) = f(CVR_i)\,|J| \tag{31}$$

$$f(CW_i) = f(CVR_i)\left|\frac{d(CVR_i)}{d(CW_i)}\right| \tag{32}$$

Then substitute $CVR_i = \exp\left(-2\left(CW_i - \varphi\left(p, F_{(\alpha;p+1,n-p-1)}, F_{(\alpha;p+1,n-p-2)}\right)\right)\right)$ from (3), then $\frac{d(CVR_i)}{d(CW_i)} = -2e\left(-2\left(CW_i - \varphi\left(p, F_{(\alpha;p+1,n-p-1)}, F_{(\alpha;p+1,n-p-2)}\right)\right)\right)$ and (30) in (32), to get the final form of the density function of CW_i-statistic as

$$f(CW_i;p,n) = \omega(p,n)\,\psi(CW_i;p,n)\left(\frac{\psi(CW_i;p,n)}{\lambda(p,n)}\right)^{\frac{n-p-2}{2p}-1}\sum_{r=0}^{\infty}\begin{pmatrix}-1/2\\ r\end{pmatrix}\left(\frac{\psi(CW_i;p,n)}{\lambda(p,n)}\right)^{r/p} B\left(\frac{(n-p)(p+1)+2(r-1)}{2p}, \frac{p-1}{2}\right) \tag{33}$$

where,

$$-\frac{1}{2}\log(\lambda(p,n)) + \varphi\left(p, F_{(\alpha;p+1,n-p-1)}, F_{(\alpha;p+1,n-p-2)}\right) \le CW_i < \infty, n,p > 0, n > p$$

$$\omega(p,n) = \frac{2(\lambda(p,n)/p)}{B\left(\frac{n-p-2}{2}, \frac{1}{2}\right)B\left(\frac{n-p}{2}, \frac{p-1}{2}\right)},$$

$$\lambda(p,n) = \frac{n}{n-1}\left(\frac{n-p-1}{n-p-2}\right)^p \quad \varphi\left(p, F_{(\alpha;p+1,n-p-1)}, F_{(\alpha;p+1,n-p-2)}\right) = \frac{p}{2}\log\left(\frac{F_{(\alpha;p+1,n-p-1)}}{F_{(\alpha;p+1,n-p-2)}}\right),$$

and $\psi(CW_i;p,n) = \exp\left(-2\left(CW_i - \varphi\left(p, F_{(\alpha;p+1,n-p-1)}, F_{(\alpha;p+1,n-p-2)}\right)\right)\right)$

From (33), it is the density function of CW_i statistic which always greater than or equal to $CW_i \ge -\frac{1}{2}\log(\lambda(p,n)) + \varphi\left(p, F_{(\alpha;p+1,n-p-1)}, F_{(\alpha;p+1,n-p-2)}\right)$ and if $n \to \infty$ then $0 \le CW_i < \infty$.This shows, if the sample size is very large, then the lower limit of the CW_i will converge to 0.Moreover,(33) involves the normalizing constants $\omega(p,n)$(in terms of Beta functions namely $B(\frac{n-p-2}{2}, \frac{1}{2})$,$B(\frac{n-p}{2}, \frac{p-1}{2})$),$\alpha(p,n)$ and $\varphi\left(p, F_{(\alpha;p+1,n-p-1)}, F_{(\alpha;p+1,n-p-2)}\right)$ with two shape parameters (p,n),where n is the sample size and p is the no. of predictors used in a multiple linear regression model. Similarly, $\psi(CW_i;p,n)$ is the auxiliary function of CW_i statistic and the authors derived the first two moments of CW_i in terms of mean, variance in order to know the location and dispersion of CW_i statistic and it is shown as follows. Using (15) and substitute in (3), to get

$$CW_i = -\frac{1}{2}\log\left(\lambda\left(p,n\right)\right) - \frac{p}{2}\left(\log\left(\theta_{1i}\right)\right) + \frac{1}{2}\left(\log\left(\theta_{2i}\right)\right) + \varphi\left(p, F_{(\alpha;p+1,n-p-1)}, F_{(\alpha;p+1,n-p-2)}\right) \tag{34}$$

Then take expectation for (34) and substitute the log moments of two independent beta variables θ_{1i} and$\theta_{2i}$of kind-1, to get the first moment of CW_i as

$$E\left(CW_i\right) = -\frac{1}{2}\log\left(\lambda\left(p,n\right)\right) - \frac{p}{2}E\left(\log\left(\theta_{1i}\right)\right) + \frac{1}{2}E\left(\log\left(\theta_{2i}\right)\right) + \varphi\left(p, F_{(\alpha;p+1,n-p-1)}, F_{(\alpha;p+1,n-p-2)}\right)$$

$$E\left(CW_i\right) = -\frac{1}{2}\log\left(\lambda\left(p,n\right)\right) - \Omega\left(p,n\right) + \varphi\left(p, F_{(\alpha;p+1,n-p-1)}, F_{(\alpha;p+1,n-p-2)}\right) \tag{35}$$

where $\Omega\left(p,n\right) = \frac{p}{2}\left(\Psi\left(\frac{n-p}{2}\right) - \Psi\left(\frac{n-p+1}{2}\right) - \frac{2}{(n-p-1)(n-p-2)}\right) - \frac{1}{2}\left(\Psi\left(\frac{n-p}{2}\right) - \Psi\left(\frac{n+1}{2}\right) + \frac{2}{n-1}\right)$ and $\psi\left(\right)$ is the di-gamma function respectively.

Subtract (35) from (34), square both sides, then take expectation to get

$$V\left(CW_i\right) = \frac{1}{4}\left(p^2 V\left(\log\left(\theta_{1i}\right)\right) + V\left(\log\left(\theta_{2i}\right)\right)\right) \tag{36}$$

Then, substitute the variance of log of two independent beta variables θ_{1i} and$\theta_{2i}$of kind-1, to get the variance of CW_i as

$$V\left(CW_i\right) = \frac{p^2}{4}\left(\Phi_1\left(p,n\right) + \Phi_2\left(p,n\right)\right) \tag{37}$$

where
$\Phi_1\left(p,n\right) = \left(\Psi\left(1, \frac{n-p}{2}\right) - \Psi\left(1, \frac{n-p+1}{2}\right) + \frac{8(n-p)-12}{(n-p-1)^2(n-p-2)^2}\right)$
$\Phi_2\left(p,n\right) = \frac{1}{p^2}\left(\Psi\left(1, \frac{n-p}{2}\right) - \Psi\left(1, \frac{n+1}{2}\right) - \frac{4}{(n-1)^2}\right)$ and $\psi\left(1,\right)$ is the poly-gamma function respectively.

By using the mean and variance of CW_i-statistic from (35) and (37), the authors established the upper control limit of CW_i for different combination of (p,n).Therefore

$$UCL\left(CW_i\right) = E\left(CW_i\right) + \sqrt{V\left(CW_i\right)} \tag{38}$$

$$UCL\left(CW_i\right) = \varphi\left(p, F_{(\alpha;p+1,n-p-1)}, F_{(\alpha;p+1,n-p-2)}\right) - \left(\frac{1}{2}\log\left(\lambda\left(p,n\right)\right) + \Omega\left(p,n\right)\right) + \frac{p}{2}\sqrt{\Phi_1\left(p,n\right) + \Phi_2\left(p,n\right)} \tag{39}$$

where $n > p$

By using (39), as a first approach, the authors utilize the upper control limit as a cut-off when the φ function is equal to $\varphi\left(p, F_{(0.05;p+1,n-p-1)}, F_{(0.05;p+1,n-p-2)}\right)$, $\varphi\left(p, F_{(0.01;p+1,n-p-1)}, F_{(0.01;p+1,n-p-2)}\right)$ to identify the influential observation in a linear multiple regression model. The computed CW_i of any observation exceeds the upper control limit, then the observation is said to be influential. As a second approach, the authors adopted the test of significance approach of evaluating and identifying the influential observations in a sample. The approach is to derive the critical points of the CW_i statistic by using the following relationship by substitute (11) in (3) as

$$CW_{i(\alpha;p,n)} = -\frac{1}{2}\log\left(\frac{\frac{n}{n-1}\left(\frac{n-p-1}{n-p-2}\right)^{p}\left(1+\frac{p-1}{n-p}F_{i(\alpha;p-1,n-p)}\right)}{\left(1+\frac{1}{n-p-2}F_{i(\alpha;1,n-p-2)}\right)^{p}\left(\frac{F_{(\alpha;p+1,n-p-1)}}{F_{(\alpha;p+1,n-p-2)}}\right)^{p}}\right) \tag{40}$$

From (40), for different combination of values of (p, n) and the significance probability $p\left(CW_i > CW_{i(\alpha;p,n)}\right) = \alpha$, The authors computed the critical points of CW-statistic. By using the critical points, we can test the significance of the non-influential observation computed from a multiple linear regression model. The following table-1 visualizes the upper control limit of the CW computed from (39) and tables 2,3 exhibits the significant percentage points of the distribution of CW-statistic for varying sample size *(n)* and no.of predictors *(p)* at 5% and 1% significance (α).

Table 1: Upper control limit of CW-statistic for combinations of (p,n) when$\varphi\left(p, F_{(\alpha=0.05;p+1,n-p-1)}, F_{(\alpha=0.05;p+1,n-p-2)}\right)$

n	P									
	1	2	3	4	5	6	7	8	9	10
4	-0.47148	-	-	-	-	-	-	-	-	-
5	0.08499	-0.8058	-	-	-	-	-	-	-	-
6	0.15166	0.3185	-1.2197	-	-	-	-	-	-	-
7	0.15379	0.44449	0.5178	-1.6785	-	-	-	-	-	-
8	0.14372	0.43699	0.7249	0.6927	-2.1664	-	-	-	-	-
9	0.13166	0.40528	0.7196	0.9938	0.8505	-2.6734	-	-	-	-
10	0.12059	0.37091	0.67283	0.9978	1.2523	0.9951	-3.1947	-	-	-
11	0.11064	0.3398	0.6197	0.9406	1.2713	1.5028	1.1296	-3.7269	-	-
12	0.10191	0.31191	0.57059	0.8718	1.2065	1.5395	1.7463	1.2559	-4.2672	-
13	0.0943	0.28817	0.52646	0.80674	1.1246	1.4701	1.8038	1.9845	1.3764	-4.814
14	0.088	0.26708	0.48788	0.74696	1.045	1.3766	1.7313	2.064	2.2178	1.4912
15	0.081955	0.24889	0.4534	0.69502	0.9714	1.2842	1.628	1.9903	2.3209	2.447
16	0.077145	0.23265	0.4238	0.64758	0.9061	1.1972	1.5232	1.8776	2.2469	2.5745
17	0.07243	0.21891	0.39706	0.60712	0.84715	1.1201	1.4245	1.7624	2.1267	2.5015
18	0.068539	0.206	0.3739	0.57008	0.7954	1.0493	1.3349	1.6512	2.0003	2.3738
19	0.064867	0.19507	0.35273	0.5381	0.7484	0.9877	1.254	1.5512	1.8789	2.2374
20	0.061782	0.18414	0.3346	0.50806	0.70738	0.9308	1.1814	1.459	1.7678	2.1058
21	0.058577	0.17572	0.31698	0.48252	0.66993	0.88163	1.1161	1.3782	1.6661	1.9846
22	0.056222	0.16687	0.30194	0.45774	0.6362	0.83511	1.0579	1.3027	1.5749	1.8727
23	0.053344	0.15978	0.28791	0.43724	0.60532	0.79486	1.0041	1.2373	1.4917	1.773
24	0.051415	0.15246	0.27538	0.4168	0.57797	0.75673	0.9562	1.1753	1.4164	1.6809
25	0.049077	0.14704	0.26357	0.39961	0.55164	0.72357	0.91204	1.121	1.3481	1.5988
26	0.047495	0.14036	0.25294	0.38236	0.5291	0.69121	0.87262	1.0694	1.2868	1.5223
27	0.045407	0.1353	0.24268	0.3679	0.50714	0.66424	0.83502	1.0239	1.2291	1.4546
28	0.044141	0.12966	0.23433	0.35304	0.48745	0.6367	0.80166	0.9802	1.1779	1.3909
29	0.041979	0.12567	0.22521	0.34128	0.4687	0.61285	0.76908	0.94306	1.1292	1.3331
30	0.041102	0.12048	0.21755	0.32747	0.45309	0.58968	0.74089	0.9056	1.0857	1.279
40	0.030577	0.089109	0.161	0.24068	0.33198	0.42949	0.53784	0.65288	0.77927	0.91198
60	0.020983	0.058719	0.10587	0.15675	0.21658	0.2769	0.34841	0.41903	0.49873	0.57942
80	0.015896	0.042773	0.080443	0.11493	0.16125	0.20513	0.25857	0.30787	0.3672	0.42207
100	0.012078	0.035352	0.066507	0.090978	0.12751	0.16265	0.20436	0.24405	0.29045	0.33439
120	0.010112	0.02946	0.052423	0.077616	0.10524	0.13655	0.16821	0.20192	0.23854	0.27772

p-no.of predictors n-Sample Size

Table 2: Upper control limit of CW-statistic for combinations of (p, n) when$\varphi\left(p, F_{(\alpha=0.01;p+1,n-p-1)}, F_{(\alpha=0.01;p+1,n-p-2)}\right)$

n	***P***									
	1	2	3	4	5	6	7	8	9	10
4	-1.25678	-	-	-	-	-	-	-	-	-
5	-0.15471	-2.3829	-	-	-	-	-	-	-	-
6	0.04226	-0.1698	-3.5902	-	-	-	-	-	-	-
7	0.09269	0.21839	-0.222	-4.8432	-	-	-	-	-	-
8	0.10522	0.30939	0.3794	-0.2999	-6.1254	-	-	-	-	-
9	0.10536	0.32418	0.5232	0.5273	-0.3956	-7.4271	-	-	-	-
10	0.10149	0.31511	0.54723	0.7313	0.6642	-0.5049	-8.7431	-	-	-
11	0.09614	0.2991	0.5329	0.7694	0.9339	0.7926	-0.6247	-10.0702	-	-
12	0.09061	0.28111	0.50709	0.7531	0.9891	1.1307	0.9136	-0.7527	-11.4054	-
13	0.0852	0.26407	0.47806	0.71964	0.9732	1.2059	1.3232	1.029	-0.8867	-12.7471
14	0.0805	0.24758	0.44988	0.68046	0.9337	1.1922	1.42	1.5115	1.1394	-1.0265
15	0.075755	0.23299	0.4228	0.64272	0.8862	1.1484	1.4103	1.6317	1.6961	1.2455
16	0.071845	0.21925	0.3986	0.60528	0.8389	1.0931	1.3626	1.6264	1.8407	1.8775
17	0.06793	0.20751	0.37596	0.57222	0.79275	1.0378	1.3012	1.5767	1.8418	2.0476
18	0.064639	0.1962	0.356	0.54088	0.7506	0.9826	1.2373	1.5084	1.7894	2.055
19	0.061467	0.18667	0.33733	0.5132	0.7108	0.9326	1.1747	1.438	1.7165	2.0011
20	0.058782	0.17674	0.3213	0.48666	0.67528	0.8845	1.1158	1.367	1.6389	1.9237
21	0.055877	0.16922	0.30528	0.46392	0.64233	0.84213	1.061	1.302	1.5612	1.8398
22	0.053822	0.16107	0.29154	0.44144	0.6121	0.80101	1.0109	1.2386	1.488	1.755
23	0.051144	0.15458	0.27871	0.42284	0.58422	0.76526	0.9635	1.1825	1.4185	1.6753
24	0.049515	0.14776	0.26708	0.404	0.55927	0.73063	0.9208	1.128	1.3538	1.5985
25	0.047277	0.14274	0.25617	0.38811	0.53504	0.70047	0.88094	1.0798	1.294	1.5283
26	0.045895	0.13656	0.24624	0.37206	0.5142	0.67061	0.84502	1.0332	1.2396	1.4613
27	0.043907	0.1318	0.23658	0.3585	0.49364	0.64574	0.81042	0.9917	1.1876	1.4014
28	0.042741	0.12636	0.22873	0.34454	0.47525	0.62	0.77956	0.9515	1.141	1.344
29	0.040679	0.12267	0.22001	0.33348	0.4576	0.59775	0.74918	0.91716	1.0963	1.2914
30	0.040002	0.11778	0.21285	0.32027	0.44289	0.57598	0.72279	0.8823	1.0561	1.2418
40	0.029977	0.087709	0.1586	0.23708	0.32708	0.42299	0.52954	0.64248	0.76647	0.89638
60	0.020783	0.058119	0.10487	0.15535	0.21468	0.2745	0.34541	0.41523	0.49423	0.57402
80	0.015796	0.042473	0.079943	0.11423	0.16025	0.20393	0.25697	0.30597	0.3649	0.41937
100	0.012078	0.035152	0.066207	0.090478	0.12691	0.16185	0.20346	0.24285	0.28905	0.33279
120	0.010012	0.02936	0.052223	0.077316	0.10484	0.13605	0.16751	0.20112	0.23764	0.27662

p-no.of predictors n-Sample Size

Table 3: Significant two-tail percentage points of

CW**-statistic at** $p\left(CW_i > CW_{i(\alpha=0.05;p,n)}\right) = 0.05$

n	***P***									
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
4	.87908	-	-	-	-	-	-	-	-	-
5	.50581	1.12670	-	-	-	-	-	-	-	-
6	.34364	.56871	.37788	-	-	-	-	-	-	-
7	.25714	.34852	.02435	1.80253	-	-	-	-	-	-
8	.20431	.24046	.01102	.55788	3.19207	-	-	-	-	-
9	.16898	.17896	.01825	.32373	1.06396	4.56135	-	-	-	-
10	.14382	.14026	.02414	.23870	.61415	1.55365	5.91724	-	-	-
11	.12504	.11409	.02758	.19510	.44056	.89093	2.03224	7.26356	-	-
12	.11052	.09543	.02931	.16781	.35020	.63083	1.15829	2.50273	8.60266	-
13	.09897	.08158	.02997	.14858	.29429	.49531	.81299	1.41868	2.96701	9.93608
14	.08957	.07095	.02999	.13400	.25577	.41203	.63333	.98907	1.67366	3.42636
15	.08178	.06258	.02962	.12239	.22728	.35522	.52351	.76599	1.16038	1.92429
16	.07522	.05584	.02903	.11285	.20515	.31364	.44909	.63023	.89441	1.32782
17	.06962	.05032	.02831	.10482	.18734	.28166	.39498	.53868	.73314	1.01937
18	.06479	.04572	.02753	.09793	.17263	.25615	.35363	.47246	.62484	.83293
19	.06059	.04185	.02672	.09195	.16022	.23524	.32083	.42208	.54682	.70817
20	.05689	.03854	.02591	.08668	.14959	.21774	.29407	.38228	.48768	.61860
21	.05361	.03568	.02511	.08202	.14036	.20282	.27176	.34994	.44113	.55092
22	.05069	.03320	.02434	.07784	.13226	.18994	.25282	.32305	.40340	.49779
23	.04806	.03102	.02359	.07408	.12508	.17868	.23651	.30028	.37210	.45483
24	.04570	.02910	.02287	.07068	.11867	.16874	.22229	.28072	.34566	.41928
25	.04355	.02739	.02218	.06758	.11291	.15990	.20978	.26370	.32299	.38930
26	.04160	.02586	.02152	.06475	.10770	.15198	.19866	.24874	.30329	.36363
27	.03981	.02448	.02090	.06215	.10296	.14483	.18872	.23547	.28600	.34136
28	.03817	.02324	.02030	.05975	.09864	.13834	.17976	.22362	.27069	.32183
29	.03666	.02211	.01973	.05753	.09466	.13243	.17164	.21295	.25701	.30454
30	.03526	.02108	.01919	.05548	.09101	.12702	.16425	.20330	.24471	.28912
40	.02552	.01431	.01499	.04091	.06577	.09040	.11523	.14055	.16656	.19347
60	.01643	.00862	.01034	.02685	.04242	.05756	.07253	.08747	.10249	.11765
80	.01211	.00615	.00787	.01999	.03133	.04227	.05300	.06362	.07420	.08478
100	.00959	.00477	.00635	.01592	.02484	.03341	.04177	.05001	.05818	.06632
120	.00794	.00389	.00532	.01323	.02058	.02762	.03448	.04121	.04787	.05447
∞	0	0	0	0	0	0	0	0	0	0

p-no.of predictors n-Sample Size

Table 4: Significant two-tail percentage points of CW-statistic at $p\left(CW_i > CW_{i(\alpha=0.01;p,n)}\right) = 0.01$

n	*P*									
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
4	1.70223	-	-	-	-	-	-	-	-	-
5	1.06068	2.2468	-	-	-	-	-	-	-	-
6	.75388	1.2863	.06801	-	-	-	-	-	-	-
7	.57930	.8592	.42261	2.29349	-	-	-	-	-	-
8	.46834	.6298	.36082	.37231	4.48116	-	-	-	-	-
9	.39218	.4908	.28352	.08216	1.13691	6.64739	-	-	-	-
10	.33690	.3991	.22542	.01669	.50001	1.88384	8.79964	-	-	-
11	.29505	.3347	.18347	.00070	.29556	.90290	2.61899	10.94197	-	-
12	.26232	.2873	.15273	.00195	.20836	.56152	1.29566	3.34565	13.07684	-
13	.23604	.2512	.12964	.00081	.16337	.40472	.81863	1.68102	4.06584	15.20593
14	.21450	.2228	.11188	.00123	.13671	.31868	.59324	1.06919	2.06067	4.78092
15	.19653	.2000	.09790	.00325	.11917	.26528	.46699	.77593	1.31468	2.43579
16	.18132	.1812	.08668	.00501	.10669	.22904	.38754	.61005	.95406	1.55610
17	.16827	.1656	.07753	.00644	.09724	.20276	.33318	.50504	.74898	1.12849
18	.15697	.1524	.06995	.00758	.08976	.18274	.29362	.43298	.61878	.88455
19	.14708	.1411	.06359	.00846	.08364	.16689	.26345	.38050	.52934	.72945
20	.13835	.1313	.05819	.00914	.07848	.15396	.23959	.34049	.46420	.62287
21	.13060	.1227	.05356	.00965	.07404	.14316	.22018	.30890	.41459	.54529
22	.12367	.1152	.04956	.01003	.07017	.13397	.20402	.28324	.37548	.48628
23	.11743	.1085	.04606	.01031	.06675	.12603	.19033	.26194	.34377	.43980
24	.11179	.1026	.04299	.01050	.06368	.11908	.17854	.24392	.31748	.40219
25	.10667	.0972	.04027	.01062	.06092	.11293	.16826	.22845	.29529	.37106
26	.10199	.0924	.03785	.01069	.05841	.10744	.15920	.21500	.27627	.34481
27	.09771	.0880	.03569	.01071	.05612	.10251	.15115	.20317	.25976	.32235
28	.09377	.0840	.03374	.01070	.05402	.09804	.14393	.19267	.24527	.30288
29	.09013	.0804	.03198	.01066	.05207	.09397	.13742	.18329	.23243	.28581
30	.08677	.0770	.03039	.01061	.05027	.09025	.13152	.17484	.22097	.27071
40	.06316	.0543	.02006	.00951	.03750	.06501	.09262	.12069	.14951	.17933
60	.04089	.0341	.01172	.00729	.02497	.04201	.05875	.07538	.09204	.10882
80	.03023	.0248	.00821	.00580	.01874	.03110	.04314	.05499	.06674	.07847
100	.02398	.0195	.00630	.00480	.01500	.02470	.03410	.04332	.05241	.06143
120	.01987	.0161	.00510	.00408	.01250	.02049	.02821	.03574	.04316	.05050
∞	**0**	0	0	0	0	0	0	0	0	0

p-no.of predictors n-Sample Size

6 Numerical Results and Discussion

In this section, the authors shown a numerical study of evaluating the influential observations based on Cook-Weisberg statistic of the *ith* observation in a regression model. For this, the authors fitted Step-wise linear regression models with different set of predictors in a Brand equity study. The data in the study comprised of 18 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 five point likert scale from 1 to 5.After the data collection is over, only 275 completed questionnaires were used for analysis. The Step-wise regression results reveals 4 nested models were extracted from the regression procedure by using IBM SPSS version 22. For each model, the CW-statistic were computed and the result of the proposed approaches namely I and II are visualized in the following table.

Table 5: Identification of influential observations based on Proposed approach-I

Model	$\boldsymbol{p}$	$\varphi\left(p, F_{(0.05;p+1,n-p-1)}, F_{(0.05;p+1,n-p-2)}\right)$		$\varphi\left(p, F_{(0.01;p+1,n-p-1)}, F_{(0.01;p+1,n-p-2)}\right)$	
		UCL (CW_i)	($\boldsymbol{n}$) $CW_i >$ UCL (CW_i)	UCL (CW_i)	($\boldsymbol{n}$) $CW_i >$ UCL (CW_i)
1	1	0.004345	8	0.004345	8
2	2	0.012948	8	0.012848	8
3	3	0.022601	9	0.022601	9
4	4	0.033664	8	0.033564	8

Table 6: Identification of influential observations based on Proposed approach-II

Model	$\boldsymbol{p}$	$\varphi\left(p, F_{(0.05;p+1,n-p-1)}, F_{(0.05;p+1,n-p-2)}\right)$		$\varphi\left(p, F_{(0.01;p+1,n-p-1)}, F_{(0.01;p+1,n-p-2)}\right)$	
		UCL (CW_i)	($\boldsymbol{n}$) $CW_i >$ UCL (CW_i)	UCL (CW_i)	($\boldsymbol{n}$) $CW_i >$ UCL (CW_i)
1	1	0.00340	8	0.00853	8
2	2	0.00160	12	0.00678	10
3	3	0.00235	16	0.00205	17
4	4	0.00573	12	0.00188	13

Table 5 and 6 visualizes the results of the proposed approaches for evaluating the influential observations by using CW-statistic. At first four nested multiple regression models were fitted and the cut-off CW values for the proposed approaches are shown in the tables. Here, the authors couldn't disclose the results of the traditional approach because Cook-Weisberg failed to highlight the calibration point and they recommend if any observation which attained large negative or large positive CW-statistic, then the observation is said to be influential. But they don't mention how large the computed CW-statistic for the observation. Hence due to this incompetency of the traditional approach, the authors were unable to show the results of the traditional approach. Under the proposed approach-I, in model-1 the computed CW-statistic for 8 observations were above the UCL when the φ function is equal to $\varphi\left(p, F_{(0.05;p+1,n-p-1)}, F_{(0.05;p+1,n-p-2)}\right)$, $\varphi\left(p, F_{(0.01;p+1,n-p-1)}, F_{(0.01;p+1,n-p-2)}\right)$ respectively and hence these observations are said to be influential. Similarly, in model-2, model-3 and model-4 are concern, 8, 9 and 8 observations are finalized as influential respectively. Under the proposed approach-II, the authors adopted the test of significance approach to identify the influential observations. As far as model-1 is concern, the computed values of CW-statistic for 8 observations are exceeds the critical CW values at 5%, 1% significance level when the φ function is equal to$\varphi\left(p, F_{(0.05;p+1,n-p-1)}, F_{(0.05;p+1,n-p-2)}\right)$,$\varphi\left(p, F_{(0.01;p+1,n-p-1)}, F_{(0.01;p+1,n-p-2)}\right)$ respectively. Likewise, when the φ function is equal to $\varphi\left(p, F_{(0.05;p+1,n-p-1)}, F_{(0.05;p+1,n-p-2)}\right)$, then 12, 16, 12 observations are treated as influential at 1 % significance level in model-2, model-3 and model-4 respectively. Moreover, if the φ function is equal to $\varphi\left(p, F_{(0.01;p+1,n-p-1)}, F_{(0.01;p+1,n-p-2)}\right)$, then in model-2, model-3 and model-4, 10, 17 and 13 observations are considered as Influential at 1% level of significance respectively. Finally, among the two approaches, the proposed approach-II is systematic and scientific when compared to proposed approach-I, because the critical CW-values at different significance level is scientifically determined from the distribution of CW-statistic. Hence the authors observed, the proposed approach-II outperforms the traditional approach in identifying influential observations and the results emphasize the advantages of determined calibration point for CW-statistic and it is visualized through the graphical display from the following control charts.

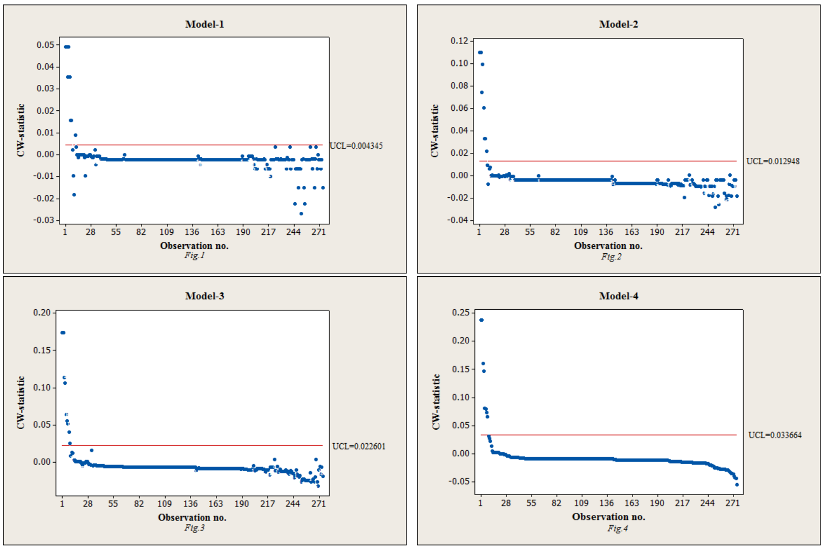

Figure 1: Control chart for fitted Models shows the Identification of Influential observation based on proposed Approach-I when $\varphi\left(p, F_{(0.05;p+1,n-p-1)}, F_{(0.05;p+1,n-p-2)}\right)$

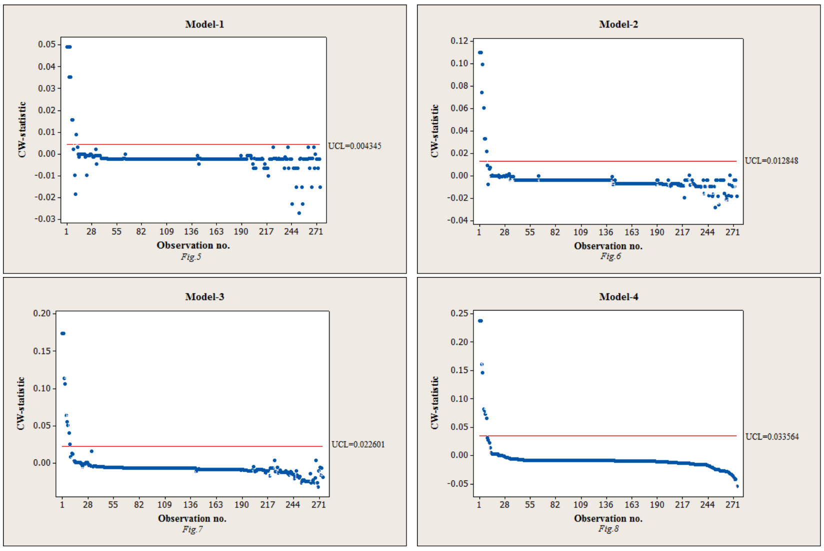

Figure 2: Control chart for fitted Models shows the Identification of Influential observation based on proposed Approach-I when $\varphi\left(p, F_{(0.01;p+1,n-p-1)}, F_{(0.01;p+1,n-p-2)}\right)$

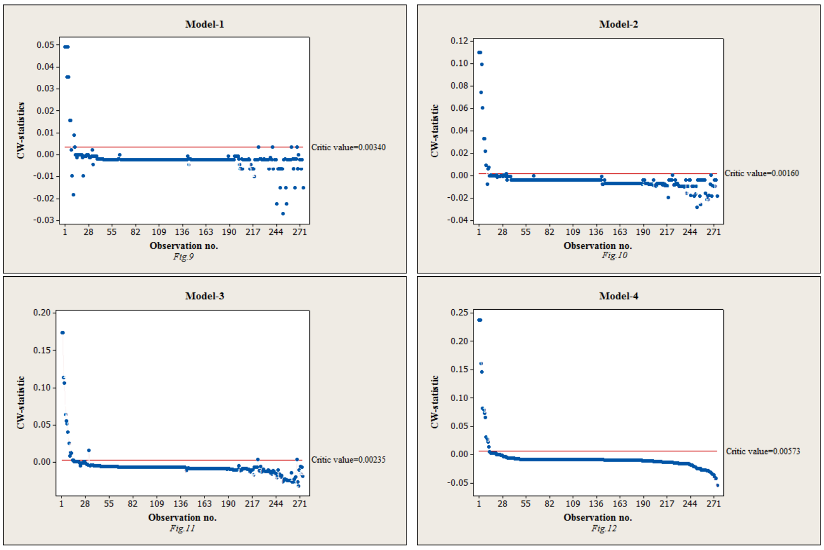

Figure 3: Control chart for fitted Models shows the Identification of Influential observation at 5% level based on proposed Approach-II when $\varphi\left(p, F_{(0.05;p+1,n-p-1)}, F_{(0.05;p+1,n-p-2)}\right)$

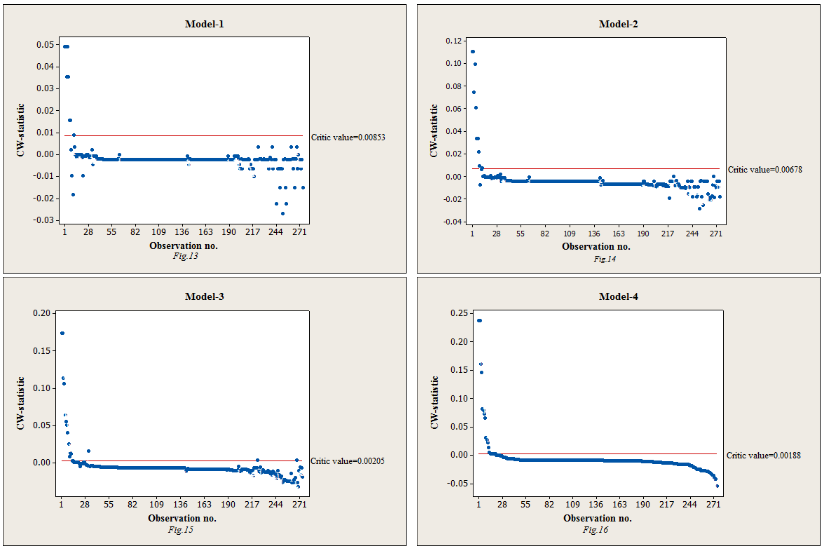

Figure 4: Control chart for fitted Models shows the Identification of Influential observation at 1% level based on proposed Approach-II when $\varphi\left(p, F_{(0.01;p+1,n-p-1)}, F_{(0.01;p+1,n-p-2)}\right)$

7 Conclusion

From the previous sections, the authors proposed a scientific approach which is based on the test of significance for Cook-Weisberg statistic to evaluate the influential observations in a multiple linear regression model. At first, the exact distribution of the CW-statistic was derived and the authors visualized the density function of CW-statistic in terms of series expression form with two shape parameters namely p and n.Moreover, the authors established the UCL of CW and also they computed the critical percentage points of CW at 5 %, 1% level of significance and it is utilized to evaluate the influential observations. Finally, the proposed approach II is more systematic and scientific because it is based on the test of significance and the results were superior when compared it with the traditional approach and proposed approach-I. So the authors conclude the proposed approaches over rides the use of traditional approach and believe that the proposed approaches took the process of identifying the influential observations based on the volume of confidence ellipsoids approaches to the next level which helps the statisticians to exactly identify the influential observations in multiple regression models.

References

Behnken, D. W. and Draper, N. R. (1972). Residuals and their variance patterns. *Technometrics*, 14(1):101–111.

Belsley, D. A., Kuh, E., and Welsch, R. E. (2005). *Regression diagnostics: Identifying influential data and sources of collinearity*, volume 571. John Wiley & Sons.

Chatterjee, S. and Hadi, A. S. (2009). *Sensitivity analysis in linear regression*, volume 327. John Wiley & Sons.

Cook, R. D. (1977). Detection of influential observation in linear regression. *Technometrics*, 19(1):15–18.

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

Dıaz-Garcıa, J. A. and González-Farıas, G. (2004). A note on the cook's distance. *Journal of statistical planning and inference*, 120(1-2):119–136.

Eubank, R. (1985). Diagnostics for smoothing splines. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 332–341.

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

Hoaglin, D. C. and Welsch, R. E. (1978). The hat matrix in regression and anova. *The American Statistician*, 32(1):17–22.

Kim, C. (1996). Cook's distance in spline smoothing. *Statistics & probability letters*, 31(2):139–144.

Kim, C. and Kim, W. (1998). Some diagnostic resultsin nonparametric density estimation. *Communications in Statistics-Theory and Methods*, 27(2):291–303.

Kim, C., Lee, Y., and Park, B. U. (2001). Cook's distance in local polynomial regression. *Statistics & probability letters*, 54(1):33–40.

Silverman, B. W. (1985). Some aspects of the spline smoothing approach to non-parametric regression curve fitting. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 1–52.

Thomas, W. (1991). Influence diagnostics for the cross-validated smoothing parameter in spline smoothing. *Journal of the American Statistical Association*, 86(415):693–698.