OUTLIER RESISTANT REWEIGHTED REGRESSION ESTIMATOR FOR MULTI-COLLINEARITY PROBLEMS

Ekele, A., Michael, A. O., and Ekele, S. O.
Department of Mathematics and Statistics, Federal Polytechnic Idah, PMB 1037, Idah, Kogi State, Nigeria
Correspondence: ekelson2002@yahoo.com

ABSTRACT

Estimation problem may occur in regression methodology when the predictors are collinear. This situation becomes worse when data are contaminated with extreme observations called outliers because outliers themselves cause multi-collinearity. This problem may lead to a large variance, inconsistent parameter estimates and confidence intervals which tend to be much wider, especially when the ordinary least squares method is employed. The ridge regression has always been the common solution to multi-collinearity. However, the choice of the ridge parameter (k), is susceptible to the least squares problem and hence, the performance of the ridge regression is impaired. This article proposed alternative parameter estimation method through a reweighted scheme that is capable of reducing the collinear structures among the predictors, while resisting the influence of outlier data points. The performance of the proposed estimator is evaluated using the mean squared error criterion which compared with other existing estimators. Results obtained showed that the proposed estimator produces better reliable estimates.

Keywords: *Multi-collinearity, Least squares, Outlier, Reweighted Scheme*

.

INTRODUCTION

Due to the convenient properties of the classical regression in modeling, it enjoys a wide spread applicability in many disciplines such as Engineering, Physical, Social and Applied Sciences. This wide spread applicability of the regression models is pertinent to its ease of computation and close form solution during parameter estimation. In the course of parameter estimation, several assumptions are made based on the Ordinary Least Squares (*OLS*). One of such is that, "No predictor is a perfect linear function of another," while the other one is "Normality of error distribution". The breakdown of these assumptions, leads to the problem of multi-collinearity and outlier data respectively. Whereas multi-collinearity is defined as the existence of a perfect or exact linear relationship among some or all predictor variables of a regression model, outliers on the other hand, are data points that depart in trend from the bulk of the observations in correct metrics.

While outliers may be due to variability in the measurement or, it may indicate experimental error. Bagheri and Habshah (2009), states that there are several causes of multi-collinearity; a few list includes:

Data Sample Size Problem: The size of a sampled data, if too small, can cause high correlation among predictors. For instance, sampling over a limited range of data points taken by the predictors in the population, leads to multi-collinearity.

Variable Inclusion Problem: Study designs are sometimes constrained on the model or in the population being sampled to include a specified predictor variable that may be related to an already included predictor. For instance, Alih and Ong (2015), reported that in the regression of electricity consumption, y on income, x_1 and house size, x_2 there is a physical constrain in the population such that families with higher income generally have larger homes than families with lower incomes and hence, x_1 and x_2 are collinear predictor exhibiting multi-collinearity.

Model Specification Problem: Model specification problem causes multi-collinearity more often when, for example, polynomial terms are added to a regression model as predictors, especially when the range of the predictor variable is small.

An over Determined Model Problem: This occurs when the model has more predictor variables than the number of observations. This could happen in

medical research where there may be a small number of patients about whom information is collected on a large number of variables.

Under the breakdown of the Gaussian assumption such as dependency of predictor variables and non-normality of error distribution, the ordinary least squares (OLS) estimation method, produces large variability in the estimation of parameters. Most often, it causes the parameter estimates to be different from the true values by orders of magnitude or an incorrect sign. It may also inflate the variance of the estimations. Consequently, estimation and prediction inferences are impaired and validity of parameter estimates becomes questionable (Midi et al, 2010).

Two kinds of multi-collinearity exist in literature namely: (1) The exact or perfect multi-collinearity and (2) The near perfect or less than perfect multi-collinearity. As an illustration, consider the model:

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_p x_p + e_i \quad (1)$$

With the OLS estimator given as:

$$\hat{\beta}_j = (X'X)^{-1} X'Y, j = 0, \ldots, p \qquad (2)$$

If there exists perfect linear relationship among the explanatory variables, then it is treated as exact multi-collinearity, that is, an exact multi-collinearity occurs when $|X'X| = 0$ and $\hat{\beta}_j$ in Equation (2) becomes inestimable since the inverse $(X'X)^{-1}$ in Equation (2) will not exist. Furthermore, if the predictor variables are significantly correlated but not perfectly, then the scenario is referred to as near or less than perfect multi-collinearity. In this scenario, $|X'X| \neq 0$ and the inverse of $(X'X)^{-1}$ in Equation (2) exists but with related large diagonal element.

Several efforts have been made to improve the *OLS* estimation procedure. In general, there are two techniques. One technique is based on finding (biased) estimators with smaller mean squared error (*MSE*) than the OLS estimators. Ridge regression and many shrinkage type of estimators are examples (Stein, 1960; Scolve, 1968). This technique does not directly address itself to the issue of multi-collinearity, even though multi-collinearity is often the situation where the aforementioned estimators are used. In contrast, the second technique is concerned with the dependency nature of the predictor variables. The principal components regression as well as the latent root regression and the factor analysis methods, are few examples.

Hoerl and Kennard (1970a, b), propose the ridge estimator as an alternative to the *OLS* estimator when the predictors are collinear. The ridge estimator is given by:

$$\hat{\boldsymbol{\beta}}_R = (X'X + k\boldsymbol{I})^{-1} X'Y \qquad (3)$$

where $\boldsymbol{I}$ denote an identity matrix and *k*, is a positive number known as ridge parameter. The corresponding *MSE* is given by:

$$[MSE\left(\hat{\boldsymbol{\beta}}\right]_R) = \sigma^2 \sum_{i=1}^{p} \frac{\lambda_i}{[(\lambda]_i + k)^2} + k^2 \boldsymbol{\beta}'(X'X + k\boldsymbol{I})^{-2}\boldsymbol{\beta} \qquad (4)$$

Although the ridge regression estimator is bias, it yields minimum *MSE* compared to the *OLS* estimator for a certain value of *k* (Hoerl and Kennard, 1970a). However, the $[MSE\left(\hat{\boldsymbol{\beta}}\right]_R)$ depends on the unknown parameters *k*, $\boldsymbol{\beta}$ and σ^2, which can't be calculated in practice, but *k* has to be estimated from the data instead.

There are several methods for estimating the value of k in literature. A few list includes: Hoerl and Kennard (1970a), Hoerl et al., (1975), McDonald and Galarneau (1975); Lawless and Wang (1976), Hocking (1976), Wichern and Churchill (1978), Nordberg (1982), Singh and Tracy (1999), Wencheko (2000), Kibria (2003), Khalaf and Shukur (2005), Alkhamisi et al., (2006) and Alkhamisi and Shukur (2007).

Alkhamisi and Shukur (2007), suggest a new approach to estimate the ridge parameter. They also proposed some new estimators by adding $\frac{1}{\lambda_{max}}$ to some well-known estimators, where λ_{max} is the largest eigenvalue of $X'X$. They used Monte Carlo experiments and the *MSE* criterion to compare the proposed estimators with some well-known estimators. However, these estimators are biased and their efficiency depends on some values of *k*.

This article proposes a reweighted regression method with initial estimator, based on the g-inverse for a severe case of multi-collinearity and subsequent iterative reweighted scheme leading to an estimator that is unbiased with minimum variance at convergence. The proposed estimator

will be referred to as reweighted regression method and denoted as *Re-R*.

The Proposed Procedure

Consider the standard linear regression model given by:

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \qquad (5)$$

where $\boldsymbol{Y}$ is an n vector of responses, $\boldsymbol{X}$ is an $n \times p$ matrix of p predictor variables such that $p < n$, $\boldsymbol{\beta}$ is the unknown vector of regression parameter to be estimated, and $\boldsymbol{\varepsilon}$, is the n vector of residual (error) term whose conditional mean and variance are given by: $E(\boldsymbol{\varepsilon}|\boldsymbol{X}) = \boldsymbol{0}$ and $var(\boldsymbol{\varepsilon}|\boldsymbol{X}) = \boldsymbol{\sigma^2 I}_n$, where $\boldsymbol{\sigma^2}$ is an unknown parameter and $\boldsymbol{I}_n$ is an identity matrix of order n. The *OLS* estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\sigma^2}$ are given in Equation (2) and the residual mean square, $\hat{\sigma}^2 = \frac{RSS}{n-p}$ such that $\boldsymbol{e} = (\boldsymbol{I}_n - P)\boldsymbol{Y}$ is a vector of ordinary residuals, $RSS = \boldsymbol{e}^T\boldsymbol{e}$ is the residual sum of squares, and $P = X(X^TX)^{-1}X^T$ is the projection (hat) matrix with the i^{th} diagonal element denoted as p_{ii}. The adjusted residual suggested by Hadi and Simonoff (1993), is given by

$$r_{ai} = \frac{e_i}{\sqrt{1 - p_{ii}}} \qquad (6) \qquad (6)$$

According to Hadi and Son (1990), the square of the adjusted residual is the reduction in the residual sum of squares due to the deletion of the i^{th} observation. This is defined as:

$$r_a^2 = RSS - RSS_{(i)} \qquad (7)$$

where $RSS_{(i)}$ the residual sum of squares is when the subset indexed i is deleted.

The general approach to outlier identification is to form a "clean" subset of the sample that is presumably free of outliers, and then, tests the outlyingness of the remaining points relative to the clean subset. Let C be the set of indexes of the observations in the clean subset, and let X_C and Y_C be the subsets of observations indexed by C. Let $\widetilde{\beta}_C$, be the estimated regression coefficients obtained from fitting a OLS regression model to the set C, and let RSS_C and $\tilde{\sigma}_C^2$ be the corresponding residual sum of squares and residual mean square respectively. One way of finding the subset C, is to determine the subset of size V - outliers, the deletion of which produces the largest reduction in the residual sum of squares. This is equivalent to finding the subset of size $n - V$ with the minimum residual sum of squares. We refer to this subset as the minimum residual sum of squared ($MRSS$) subset.

The problem with this approach is that, V is rarely known. Even if V is known, finding the $MRSS$ subset by fitting a regression model to each of the C_V^n possible subsets of size $n - V$, involves extensive computations and may not even be feasible, especially for large n. We instead propose a method for approximating the $MRSS$ subset through cluster analysis and tests for the no-outlier hypothesis and further use the subset C to compute a robust weight for the Re-weighted Least Squares (RLS) algorithm as presented below. For simplicity, we provide two algorithms; one for finding the "clean" subset C, and the main algorithm, which computes robust distance for outlier identification and estimation of the regression parameters. These two-phase algorithms are able to address the influence of outliers and estimate regression parameters when the predictors are collinear.

Algorithm1:

The "Clean" Subset Algorithm

Step 1: Compute a Minimum volume ellipsoid (*MVE*) of the predictor-response space denoted as $\boldsymbol{Z}_y$ and obtain $\bar{x}_d$ and S_d as a robust measure of mean vector and covariance matrix of $\boldsymbol{Z}_y$ respectively. Determine the $(2p+1)\times p$ pivot point data matrix PP, which includes the $\bar{x}_d$, S_d and the cut-off point of *MVE* given by $c_f \chi^2_{(p,\alpha)}$. The pair of the cut-off points is given by the expression $\bar{x}_d \pm \sqrt{\lambda_i c_f \chi^2_{(p,\alpha)}}\boldsymbol{e}$, such that $p = k+1$,

$$c_f(\textit{correction factor}) = 1 + \frac{p+1}{n+p} + \frac{1}{n-h-p},$$

$h = integer\ part\ of\ (n+p+1)/2$, λ_i and $\boldsymbol{e}$ are the i^{th} eigenvalues and eigenvector of s_d respectively.

Step 2: Determine the $n \times p$ data matrix $\boldsymbol{B}$ with the i^{th} row of $\boldsymbol{B}$ denoted by a $(1 \times p)$ vector of $\boldsymbol{b}_i$ to be the estimator that results from an *OLS* regression of *PP* augmented by the i^{th} row of $\boldsymbol{Z}_y$. Obtain $S_d(\boldsymbol{B})$ and use it to compute an $n \times n$ similarity matrix $\boldsymbol{S}$ whose elements are defined by

$$S_{ij} = (\boldsymbol{b}_i - \boldsymbol{b}_j)^T \left(S_d(\boldsymbol{B})\right)^{-1} (\boldsymbol{b}_i - \boldsymbol{b}_j).$$

The elements of $\boldsymbol{S}$, serves as a distance metrics upon which an agglomerative hierarchical cluster analysis of a complete linkage is performed on the data.

Step 3: Order clusters from the most to the least extreme by the order of joining; the later a cluster joins, the more extreme it is. At each joining, identify the cases in the smaller cluster as more outlying. (For example, a cluster of size 1, is a single potential outlier; a cluster of size 2, is a pair of potential outliers).

Step 4: When the number of (most extreme) identified cases in Step 3 reaches $n-h$, the h remaining cases constitute the "clean" subset $\boldsymbol{C}$. Go to Step **2** of the main algorithm:

Algorithm 2: **The Main Algorithm**

Step 1: Find a "clean" subset $\boldsymbol{C}$, of size h from algorithm 1.

Step 2: Compute the distance

$$d_i = \frac{y_i - x_i^T \hat{\boldsymbol{\beta}}_C}{\hat{\sigma}_C \sqrt{1 - x_i^T \left(\boldsymbol{X}_C^T \boldsymbol{X}_C\right)^{-1} x_i}} \quad \text{if } i \in \boldsymbol{C},$$

$$d_i = \frac{y_i - x_i^T \hat{\boldsymbol{\beta}}_C}{\hat{\sigma}_C \sqrt{1 + x_i^T \left(\boldsymbol{X}_C^T \boldsymbol{X}_C\right)^{-1} x_i}} \quad \text{if } i \notin \boldsymbol{C}. \tag{8}$$

(Note that when $i \in \boldsymbol{C}$, d_i is internally studentized residuals and when $i \notin \boldsymbol{C}$, d_i is the scaled prediction error based on the clean subset $\boldsymbol{C}$).

Step 3: Arrange the observation in ascending order according to $|d_i|$ and let $d_{(h'+1)}$ be the $(h'+1)^{th}$ order statistic of $|d_i|$, where h' is the current size of $\boldsymbol{C}$.

i. If $d_{(h'+1)} \geq t_{(\alpha/2(h'+1),(h'-V))}$, then declare all the observation corresponding to

$|d_i| \geq t_{(\alpha/2(h'+1),(h'-V))}$ as outliers and go to step 4.

ii. Otherwise, update $\boldsymbol{C}$ by adding the first $(h'+1)^{th}$ ordered observations. If $n = h'+1$, then declare no outlier in the data and go to step 4; otherwise, go to step 2.

Step 4: Let d_i be the normalized distance obtained from step 2 of this algorithm, Define m_d to be the median of d_i and re-define d_i to be $d_i = 1/max(d_i, m_d)$. Then compute the squared normalized version of the latest d_i as:

$$sd_i = \frac{d_i^2}{\sum_{i=1}^{n} d_i^2} \tag{9}$$

Let $\hat{\boldsymbol{\beta}}^{\boldsymbol{0}}$ be the WLS estimate of the regression coefficient with weights defined by sd_i in (5).

Step 5: For $j = 1, 2, 3, ...,$ until convergence, let

$$\boldsymbol{e}^{j-1} = \boldsymbol{y} - \hat{\boldsymbol{y}}^{j-1} = \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}^{j-1}$$

be the residuals of the last fit. Compute the normalized square residual given by

$$sr_i^{j-1} = \left(e_i^{j-1}\right)^2 / \sum_{i=1}^{n} \left(e_i^{j-1}\right)^2$$

Let $p_i = (1 - sd_i) / \max\left(sr_i^{j-1}, m_{sr}\right)$

where m_{sr} is the median of $sr_i^{j-1}, i = 1, ..., n$. The new weight for the *Re-R* is given by

$$w_i^j = p_i^2 / \sum_{i=1}^{n} p_i^2$$

such that $\hat{\boldsymbol{\beta}}^{\boldsymbol{j}}$ is the reweighted estimate of the regression parameters when using w_i^j as a weight for the i^{th} observation.

Monte Carlo Simulation Study

Monte Carlo simulation experiment is conducted to study and compare the performance of the proposed *Re-R* with methods such as the ordinary least squares (*OLS*), the ridge regression (*RR*), and the generalized regression method

(*GRM*). Data were simulated based on the regression model defined as

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_1 x_{i2} + \varepsilon_i \qquad (14)$$

Let $\delta_i \sim iid(0,1)$ and $m_i \sim U[10,15]$ be thresholds and τ be the level of outlier contamination with standard values usually between $0 - 50\%$. Then, the inliers and outliers are planted in the simulated data set using Equation (15):

$$\varepsilon_i = \begin{cases} \delta_i \\ \delta_i + m_i \end{cases} \qquad (15)$$

Such that the target parameters are set at:

$$\beta_j = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}^t = \begin{pmatrix} 0.30 \\ 0.55 \\ 0.15 \end{pmatrix} \qquad (16)$$

Finally, the collinear explanatory variables are simulated as:

$$x_{ij} = (1 - \rho^2) z_{ij} + \rho z_{ij} \qquad (17)$$

such that $z_{ij} \sim iid(0,1)$ with values of $\rho = 0.00,\ 0.25,\ 0.95$ to mean no multi-collinearity, low multi-collinearity and severe multi-collinearity respectively.

The data are generated for sample sizes $n = 20$, $n = 30$, and $n = 100$ at $r = 10{,}000$ simulation runs and for each data-set. The coefficients are computed using the proposed method denoted as *Re-R*, and other existing methods such as the *OLS*, *RR*, and *GRM*. Thereafter, the mean squared error of the estimate is computed.

Note, that the mean squared error of a univariate component T, is given by:

$$MSE(T) = n \times \underset{r}{avr} (T^r - \theta)^2 \qquad (18)$$

where θ, is the true value of the parameter. Consequently, the MSE of the coefficients computed from the simulated datasets are defined as:

$$MSE(\beta_j) = n \times \underset{r}{avr} (\beta_j - \hat{\beta}_j)^2 \qquad (19)$$

The results obtained based on Equation (19), are presented in Tables 1-4.

Table 1: Simulation Results for Regression Estimators for $n = 20$

			$MSE(\hat{\beta}_j)$		
n	Estimators	Coefficients	$\rho = 0.00$	$\rho = 0.30$	$\rho = 0.95$
20	OLS	$\hat{\beta}_0$	0.0161	7.3340	7.5521
		$\hat{\beta}_1$	0.0344	7.7159	7.5482
		$\hat{\beta}_2$	0.0251	8.0041	7.6848
	Re-R	$\hat{\beta}_0$	0.0401	0.0421	0.0421
		$\hat{\beta}_1$	0.0512	0.0670	0.0677
		$\hat{\beta}_2$	0.0804	0.0818	0.1004
	RR	$\hat{\beta}_0$	0.5507	4.8840	4.6431
		$\hat{\beta}_1$	0.4514	3.1952	4.9411
		$\hat{\beta}_2$	0.4801	2.0856	3.0024
	GRM	$\hat{\beta}_0$	0.1105	3.0212	2.7891
		$\hat{\beta}_1$	0.1418	3.1121	3.1388
		$\hat{\beta}_2$	0.1200	2.9945	3.2835

Table 2: Simulation Results for Regression Estimators for $n = 30$

			$MSE(\hat{\beta}_j)$		
n	Estimators	Coefficients	$\rho = 0.00$	$\rho = 0.30$	$\rho = 0.95$
30	OLS	$\hat{\beta}_0$	0.0148	6.7473	6.9479
		$\hat{\beta}_1$	0.0316	7.0986	6.9443
		$\hat{\beta}_2$	0.0231	7.3638	7.0700
	Re-R	$\hat{\beta}_0$	0.0369	0.0387	0.0387
		$\hat{\beta}_1$	0.0471	0.0616	0.0623
		$\hat{\beta}_2$	0.0740	0.0753	0.0924
	RR	$\hat{\beta}_0$	0.5066	4.4933	4.2717
		$\hat{\beta}_1$	0.4153	2.9396	4.5458
		$\hat{\beta}_2$	0.4417	1.9188	2.7622
	GRM	$\hat{\beta}_0$	0.1017	2.7795	2.5660
		$\hat{\beta}_1$	0.1305	2.8631	2.8877
		$\hat{\beta}_2$	0.1104	2.7549	3.0208

Table 3: Simulation Results for Regression Estimators for $n = 50$

n	Estimators	Coefficients	$MSE(\hat{\beta}_j)$ $\rho = 0.00$	$\rho = 0.30$	$\rho = 0.95$
50	OLS	$\hat{\beta}_0$	0.0140	6.3806	6.5703
		$\hat{\beta}_1$	0.0299	6.7128	6.5669
		$\hat{\beta}_2$	0.0218	6.9636	6.6858
	Re-R	$\hat{\beta}_0$	0.0349	0.0366	0.0366
		$\hat{\beta}_1$	0.0445	0.0583	0.0589
		$\hat{\beta}_2$	0.0699	0.0712	0.0873
	RR	$\hat{\beta}_0$	0.4791	4.2491	4.0395
		$\hat{\beta}_1$	0.3927	2.7798	4.2988
		$\hat{\beta}_2$	0.4177	1.8145	2.6121
	GRM	$\hat{\beta}_0$	0.0961	2.6284	2.4265
		$\hat{\beta}_1$	0.1234	2.7075	2.7308
		$\hat{\beta}_2$	0.1044	2.6052	2.8566

Table 4: Simulation Results for Regression Estimators for $n = 100$

n	Estimators	Coefficients	$MSE(\hat{\beta}_j)$ $\rho = 0.00$	$\rho = 0.30$	$\rho = 0.95$
100	OLS	$\hat{\beta}_0$	0.0127	5.6472	5.7396
		$\hat{\beta}_1$	0.0272	5.9412	5.7366
		$\hat{\beta}_2$	0.0198	6.1632	5.8404
	Re-R	$\hat{\beta}_0$	0.0317	0.0324	0.0320
		$\hat{\beta}_1$	0.0404	0.0516	0.0515
		$\hat{\beta}_2$	0.0635	0.0630	0.0763
	RR	$\hat{\beta}_0$	0.4351	3.7607	3.5288
		$\hat{\beta}_1$	0.3566	2.4603	3.7552
		$\hat{\beta}_2$	0.3793	1.6059	2.2818
	GRM	$\hat{\beta}_0$	0.0873	2.3263	2.1197
		$\hat{\beta}_1$	0.1120	2.3963	2.3855
		$\hat{\beta}_2$	0.0948	2.3058	2.4955

SUMMARY, CONCLUSION AND RECOMMENDATION

Summary

Tables 1-4, show that the proposed estimator, *Re-R* perform significantly better in terms of mean squared error (*MSE*) at both the low (i.e $\rho = 0.30$) and high (i.e. $\rho = 0.95$) levels of multi-collinearity. However, it was out-performed by the OLS estimators when there is no multi-collinearity in data set (i.e. $\rho = 0$). The *RR*, and *GRM*, are similar in their *MSE* values and out-performed the *OLS* at both the low (i.e $\rho = 0.30$) and high (i.e. $\rho = 0.95$) levels of multi-collinearity. However, the OLS performed better than all the three estimators when the data set, has no multi-collinearity ($\rho = 0.00$). In all, the proposed *Re-R* is consistent in the *MSE* values.

Summary of Findings

1. The *MSE* decreases for all estimators as the sample sizes, n increases.
2. The *OLS* has the least *MSE* when data are not collinear and hence, it becomes a better estimator under this scenario.
3. The proposed *Re-R* method has the least *MSE* when data are collinear and hence, it becomes a better estimator under this scenario.

CONCLUSION/RECOMMENDATION

The real life application of multiple linear regression models requires among others, the simultaneous monitoring of collinear variables alongside spurious observations. Real life data from experimental outcomes often contain spurious observations whose causes can be traced to an assignable variation. These spurious variables often go unnoticed if proper statistical techniques are not employed prior to regression estimation. A multiple regression method known as the reweighted regression method, *Re-R* is proposed to simultaneously handle multi-collinearity and screen dataset for likely outlier structure in order to mimic the data trend prior to regression model estimation.

No single regression algorithm seems to be outstanding, and for any given estimator, it is easier to find a multi-collinearity level and outlier configurations in datasets where the estimator fails and hence, we state that the performance of the proposed *Re-R* method, can be quite poor when the level of datasets are not collinear and outliers are not present in datasets. Under this scenario, the *OLS* becomes a better alternative.

REFERENCES

Alih, E., and Ong, H. C. (2015), An outlier-resistant test for heteroscedasticity in linear models, *Journal of Applied Statistics*, **42**(8), 1617-1634,

Alkhamisi, M., Khalaf, G., and Shukur, G. (2006). Some modifications for choosing ridge parameters. *Communications in Statistics-Theory and Methods*, ***35***(11), 2005-2020.

Alkhamisi, M. A., and Shukur, G. (2007). A Monte Carlo study of recent ridge parameters. *Communications in Statistics-Simulation and Computation*, ***36***(3), 535-547.

Bagheri, A., and Habshah, M. (2009). Robust estimations as a remedy for multi-collinearity caused by multiple high leverage points. *Journal of Mathematics and Statistics*, **5**(4), 311- 321

Bagheri, A., Habshah, M, and Imon, A. H. M. R. (2010). The effect of collinearity-influential observations on collinear data set: A Monte Carlo simulation study. *Journal of Applied Sciences (Faisalabad)*, ***10***(18), 2086-2093.

Ehsanes Saleh, A. M., and Golam Kibria, B. M. (1993). Performance of some new preliminary test ridge regression estimators and their properties.*Communications in Statistics-Theory and Methods*, ***22***(10), 2747-2764.

Hadi, A. S., and Simonoff, J. S. (1993). Procedures for the identification of multiple outliers in linear models. *Journal of the American Statistical Association*, ***88***(424), 1264-1272.

Hadi, A. S., and Son, M. S. (1990). Some properties of, and relationships among several uncorrelated and homoscedastic residual vectors. *Communications in Statistics-Theory and Methods*, ***19***(7), 2625-2642.

Hocking, R.R. (1976). The analysis and selection of variables in linear regression. *Biometrics **32***, 1–49.

Hoerl, A. E., and Kennard, R. W. (1970a). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, ***12***(1), 55-67.

Hoerl, A. E., and Kennard, R. W. (1970b). Ridge regression: Applications to nonorthogonal problems. *Technometrics*, ***12***(1), 69-82.

Hoerl, A. E., Kannard, R. W., and Baldwin, K. F. (1975). Ridge regression: Some simulations. *Communications in Statistics-Theory and Methods*, ***4***(2), 105-123.

Khalaf, G., and Shukur, G. (2005). Choosing ridge parameter for regression Problems. *Communications in Statistics - Theory and Methods, **34**(5)* 1177-1182.

Kibria, B. G. (2003). Performance of some new ridge regression estimators. *Communications in Statistics-Simulation and Computation*, ***32***(2), 419-435.

Lawless, J. F. and Wang, P. (1976). A simulation study of ridge and other regression estimators. *Communications in Statistics-Theory and Methods*, ***5***(4), 307-323

McDonald, G. C., and Galarneau, D. I. (1975). A Monte Carlo evaluation of some ridge-type estimators. *Journal of the American Statistical Association*, ***70***(350), 407-416.

Midi, H., Bagheri, A., and Imon, A. H. M. R. (2010). The application of robust multi-collinearity diagnostic method based on robust coefficient determination to a non-collinear data. *Journal of Applied Sciences*, **10**(8), 611-619.

Midi, H., Arezoo, B., and Imon .A.H.M. R. (2011). *A Monte Carlo simulation study on high leverage collinearity-enhancing observation and its effect on multi-collinearity pattern. Sains Malaysiana*, ***40***(12), 1437-1447.

Nordberg, L. (1982). A procedure for determination of a good ridge parameter in linear regression. *Communications in Statistics-Simulation and Computation*, ***11***(3), 285-309.

Scolve, S.L. (1968). Improved estimators for coefficients in linear regression. *Journal of American Statistical Association **63***, 597–606.

Singh, S., and Tracy, D.S. (1999). Ridge-regression using scrambled responses. *Metrika*, 147–157.

Stein, C. M. (1960). Multiple Regression. In Essays in Honor of Harold Hotelling, Stanford: CA: Stanford University Press, chap. 37.

Wencheko, E. (2000). Estimation of the Signal-to-Noise in the Linear Regression Model. *Statistical Papers*, ***41***, 327-343.

Wichern, D. W., and Churchill, G. A. (1978). A comparison of ridge estimators. *Technometrics*, ***20***(3), 301-311.