Transactions of the Nigerian Association of Mathematical Physics
***Volume 2, (November, 2016*)**, **pp 143 – 150**

A New Class of Winsorized Shrinkage Estimators for Multiple Linear Regression

Iduseri A. and Osemwenkhae J.E

Department of Mathematics, Faculty of Physical Sciences, University of Benin, P.M.B. 1154, Benin City 300001, Edo State, Nigeria.

Abstract

In this paper, the simultaneous occurrence of multicollinearity and legitimate contaminant in Y-space due to non-normality of error variable is considered. To handle the problem of multicollinearity and legitimate contaminants in the data, a new class of modified Winsorized shrinkage estimators (MWSEs) is proposed and their performance is evaluated through estimated mean square error (EMSE) sense. A simulation studies reveal that the MWSEs show consistently minimum EMSE among the considered shrinkage estimators.

Keywords: Multicollinearity, legitimate contaminant, Winsorization, mean square error, multiple linear regression.

1.0 Introduction

The standard multiple linear regression model is expressed as

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

where y is an $n \times 1$ vector of response variable, X is a design matrix of order $n \times p$, β is a $p \times 1$ vector of regression coefficients and ε is a $n \times 1$ vector of random error, which is normally distributed with mean vector 0 and variance $\delta^2 I_n$. Here I_n is identity matrix of order n. The least square estimator (LSE) of β is a linear function of y and is defined as

$$\hat{\beta} = \left(X^I X\right)^{-I} X^I y \qquad (2)$$

and the covariance matrix of $\hat{\beta}$ is obtained as

$$Cov\left(\hat{\beta}\right) = \sigma^2 \left(X^I X\right)^{-I} \qquad (3)$$

According to the Gauss-Markov theorem, LSE of β yields the best linear unbiased estimate (BLUE) provided the following assumptions hold:

A1: $y = X\beta + \varepsilon$

This assumption states that there is a linear relationship between y and X.

A2: X is a $n \times p$ matrix of full rank.

This assumption states that there is no perfect multicollinearity. In other words, the columns of X are linearly independent. This assumption is known as the identification condition.

A3: $E\left[\varepsilon \middle| X\right] = 0$

This assumption (the zero conditional mean assumption) states that the disturbances average out to 0 for any value of X. Put differently, no observations of the independent variables convey any information about the expected value of the disturbance. The assumption implies that $E(y) = X\beta.$ This is important since it essentially says that we get the mean function right.

Corresponding author: Iduseri A., E-mail: augustine.iduseri@uniben.edu, Tel.: +2348036698860

A4: $E\left(\varepsilon\varepsilon'\middle|X\right)=\sigma^2 I$

This captures the familiar assumption of homoscedasticity and no autocorrelation.
A5: X may be fixed or random, but must be generated by a mechanism that is unrelated to ε
A6: $\varepsilon \mid X \sim N\left(0,\delta^2 I\right)$

It is well known that when the normality assumption holds with absence of multicollinearity, the ordinary least squares (OLS) estimator becomes a maximum likelihood estimator and the best linear unbiased estimator of the unknown regression parameters and has the smallest variance in the class of all linear unbiased estimators with stable coefficients. The traditional view that OLS estimator is robust to deviations from the assumptions of normality and moderate multicollinearity discourages users from applying other methods. In instances where the model adequacy diagnostics reveal a poor least squares fit due to outliers and multicollinearity, the practitioner is faced with the problem of specifying the correct model.
In the literature, the effect of violation of assumptions has been discussed by many authors [1-4]. Moderate multicollinearity may not be problematic. However, severe multicollinearity is a problem. Various techniques are available in the literature to deal with the problem of multicollinearity. Those of Hoerl and Kennard [5, 6], Hoerl et al. [7], Liu [8], and Liu [9] are praiseworthy. Similarly, the assumption of normally distributed errors is not required for multiple regression to provide regression coefficients that are unbiased and consistent, presuming that other assumptions are met [10]. Further, as the sample size grows larger, inferences about coefficients will usually become more and more trustworthy, even when the distribution of errors is not normal. This is due to the central limit theorem which implies that, even if errors are not normally distributed, the sampling distribution of the coefficients will approach a normal distribution as sample size grows larger, assuming some reasonably minimal preconditions. This is why it is plausible to say that regression is relatively robust to the assumption of normally distributed errors. However, when non-normality is caused by outlier rather than skewness, violation of this assumption has more serious consequences [11].
Another important problem that has received considerable attention is the presence of outliers in *Y*- space. Huber [12] and Rousseeuw and Leroy [13] as cited in [14], pointed out that the presence of outliers significantly affect the performance of the OLS estimator. Jadhav and Kashid [14] pointed out that outliers in *Y*- space are due to heavy tailed distribution of error variable. There are influential observations that occur as a result of either a faulty distributional assumption (i.e., when the data turns out with different structure than originally assumed) or as a result of a function of the inherent variability of the data (see [15, 16] for more details). These types of influential observations or outliers otherwise known as legitimate contaminants in this study are not as obvious or easily identified as ordinary outliers. Though this type of observation may have a legitimate place in the data set, but if non-randomly distributed can result in small but reasonable deviation from normal error distribution and may distort the least squares fit [12, 17]. Many robust parameter estimation methods available in the literature are proposed to handle the problem of outliers in the data. Another effective alternative is the use of Winsorization approach. Winsorization regression is an effective alternative to the least squares estimation method which reduces the effect of contamination on regression coefficient [14]. But how cases such as legitimate contaminants might be taken into consideration prior to building a multiple regression model has not been considered in the literature by statisticians or by methodologist. This is the main trust of this work.
In this paper, the simultaneous occurrence of multicollinearity, outliers and/or legitimate contaminants in *Y*-space resulting in non-normality of error variable is considered. To handle the problem of multicollinearity, outliers and/or legitimate contaminants in the data, a new class of Winsorized shrinkage estimators using a modified winsorization approach is introduced and their performance is evaluated through estimated mean square error (EMSE) criterion. A simulation study is conducted to evaluate the performance of the proposed estimators.

2.0 Some Estimators Used in Dealing with Outliers and Multicollinearity

The data in a linear regression model consist of n set of observations $(x_{1i}, x_{21}, \dots, x_{pi}, y_i)$ representing a random sample from a population. Traditionally, these observations are assumed to be independent and normally distributed with constant error variance, $X \sim N(0,\sigma I)$. The canonical form of the multiple regression model is given as

$$y = Z\alpha + \varepsilon \tag{4}$$

where $Z = XQ,\ \alpha = Q^I\beta$ and $Q = (q_1, q_2, \dots, q_p)$ is an orthogonal matrix of eigenvectors $q_1, q_2, \dots, q_p$ corresponding to the eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_p \geq 0$ of $\left(X^I X\right)$ matrix, respectively. Also, $Q^I Q = I$ (identity matrix) and $\left(X^I X\right)$ refers to $\Lambda = diag(\lambda_1, \lambda_2, \dots, \lambda_p)$ of $\left(X^I X\right)$. Given the regression model in (4), the aim is to find suitable estimate $\hat{\alpha}$ for the parameter α that minimizes the error sum of squares

$$e^I e = (y - Z\alpha)(y - Z\alpha)^I \quad (5)$$

Many methods are available in the literature for finding this estimate. Majority of the research on alternatives to least squares estimation in the presence of outliers and correlated regressors has addressed either the nonnormal issue or the collinearity issue but seldom addressed the combined problem and the possible contamination of the data set by legitimate contaminants. We give a brief review of some notable estimators.

2.1 Ordinary Least Squares (OLS) Estimator

When $\varepsilon \sim N(0, \sigma^2 I)$, then the optimal estimator of regression parameters is the OLS estimator [14]. Given the multiple linear regression in canonical form (4), the OLS method aims at finding the estimator $\hat{\alpha}$ that minimizes equation (5). It is denoted by

$$\hat{\alpha}_{OLS} = \Lambda^{-1} Z^I Y \quad (6)$$

where $\Lambda = diag(\lambda_1, \lambda_2, \ldots, \lambda_p)$.

The mean squares error of the OLS is given by

$$MSE\left(\hat{\alpha}_{OLS}\right) = \sigma^2 (Z^I Z)^{-I} \quad (7)$$

2.2 Ordinary Ridge Regression (ORR) Estimator

Among the several methods proposed to overcome the problem of multicollinearity, the ordinary ridge regression estimator (ORR) proposed by Hoerl and Kennard [5, 6] is one of the most popular biased estimators for regression parameters. By adding a biasing parameter of the ridge estimator, *k* to the diagonal elements of $(X^I X)$, the system then acts more like an orthogonal system. It is defined as

$$\hat{\alpha}_{ORR} = (\Lambda + kI)^{-I} \Lambda \hat{\alpha}_{OLS} \quad (8)$$

where $k > 0$ is a ridge parameter and I is an identity matrix of an order $p \times p$. Because, the ORR estimator is biased, the MSE of ORR estimator is defined as

$$MSE\left(\hat{\alpha}_{ORR}\right) = \sigma^2 \sum_{J=1}^{P} \frac{\lambda_j}{(\lambda_j + k)} + k^2 \sum_{j=1}^{p} \frac{\alpha_j^2}{(\lambda_j + k)^2} \quad (9)$$

where $\lambda_1, \lambda_2, \ldots, \lambda_p$ are the eigenvalues of $Z^I Z = Q^I X^I X Q = \Lambda = diag(\lambda_1, \lambda_2, \ldots, \lambda_p) = I_p$. The first term on the right hand side is the sum of the variances of $\hat{\alpha}_{(k)}$, and the second term is the square of the bias. If $k > 0$, note that the bias in $\hat{\alpha}_{(k)}$ increases with k .The ridge parameter *k* plays an important role in minimizing the MSE of the ORR estimator. Various choices for estimator of *k* are available in the literature, but the estimator proposed in [7] is widely used.

2.3 Liu Estimator

Liu [8] proposed another biased estimator of α known as LIU estimator and is given by

$$\hat{\alpha}_{LIU} = (\Lambda + I)^{-1} (\Lambda + dI) \hat{\alpha}_{OLS} \quad (10)$$

where $0 < d < 1$, is LIU parameter. The advantage of the LIU estimator is that $\hat{\alpha}_{LIU}$ is a linear function of d [14]. Consequently, it is easier to choose d in $\hat{\alpha}_{LIU}$ than to choose k in $\hat{\alpha}_{ORR}$.

The MSE of LIU estimator is defined as

$$MSE\left(\hat{\alpha}_{LIu}\right) = \sigma^2 \sum_{j=1}^{p} \frac{(\lambda_j + d)^2}{\lambda_j (\lambda_j + 1)^2} + (1 - d)^2 \sum_{j=1}^{p} \frac{\alpha_j^2}{(\lambda_j + 1)^2} \quad (11)$$

The unknown parameters α and σ^2 are replaced by their unbiased OLS estimates $\hat{\alpha}_{OLS}$ and $\hat{\sigma}_{OLS}$ respectively.

2.4 Winsorized Least Squares Estimator

Yale and Forsythe [18] introduced various methods of Winsorization. Winsorization is one of the robust techniques that aim to diminish the effect of outliers in the data. Their general Winsorization procedure proposed for simple linear regression can be easily generalized to the multiple regression. The general Winsorized least squares (WLS) procedure proposed in [18] involves, first obtaining LS estimates and the predicted values $\left(\hat{Y}_i\right)$ of Y $Y_i, i = 1, 2, \ldots, n$, as well as obtaining the residual values as $r_i = Y_i - \hat{Y}_i$ and ordering them such that $r_1 \leq r_2 \leq \ldots \leq r_n$ is the ordered LS residuals (see [14] for details). Further study in Winsorization can be found in [19, 20].

3.0 The Proposed New Class of Estimators

3.1 The Modified Winsorization Method

Generally, $\alpha\%$ Winsorization is achieved by replacing observations below the $\alpha\%$ value in a given dataset and values above $(100-\alpha)\%$ observation by the percentile value. Thus, we could still be substituting outlying observations for the Winsorized values and subsequently maintaining the contamination. However, the median, which is the 50th percentile of any given set of observations, is a robust estimator. Therefore, we propose a new method of Winsorization by the substitution of the median for the Winsorized observations which will result in a substantial reduction of outliers, as well as eliminate contamination proportion of legitimate contaminants in the dataset.

Algorithm for the Modified Winsorization Method (MWM)

Given a set of n observations $y_1, y_2, \ldots, y_n$, to achieve a $\alpha\%$ winsorization, the following algorithm is employed:

Step 1: Sort the observations in an increasing order such that $y_1 \leq y_2 \leq \ldots \leq y_n$ represent an ordered set of observations.

Step 2: Find the median of the ordered dataset in step 1.

Step 3: Compute the α percentile of the set of the ordered n observations.

Step 4: Replace observations corresponding to the computed percentile at both extremes of the data with the median obtained in step 2 to obtain the winsorized observations.

Example

For a sample of 15 observations (from x_1, the smallest, to x_{15}, the largest) the 10% winsorized data corresponds to:

$$Median = x_8 \; and \, required \;\; percentile = \frac{10}{100} \times 15 = 1.5$$

Hence, the Winsorized observation is: $x_8, x_8, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}, x_{13}, x_8, x_8$

Based on the modified Winsorization approach, a new class of estimators' is now proposed to handle the simultaneous occurrence of multicollinearity, outliers and legitimate contaminants in data. The proposed estimators are called "Winsorized Shrinkage Estimator". In the following section, some Winsorized Shrinkage Estimators (WSE) is introduced and their modified MSE expressions are obtained based on the technique suggested in [21].

3.1.1 Winsorized Ordinary Ridge Regression (WORR) Estimator

The Winsorized ordinary ridge regression (WORR) estimator of α based on the ORR estimator [5, 6] as defined in equation (8) is given as

$$\begin{aligned}\hat{\alpha}_{WORR} &= \left(\Lambda + kI\right)^{-I} Z^{I} y \\ &= \left(\Lambda + k_{WLS}\, I\right)^{-I} \Lambda \,\hat{\alpha}_{WLS}\end{aligned} \qquad (12)$$

k_{WLS} is the unknown ridge parameter estimated using

$$k = \frac{p\sigma^2_{WLS}}{\sigma^2_{WLS}} \qquad (13)$$

Where p denotes number of regressor variables, $\hat{\alpha}_{WLS}$ denotes the WLS estimate of α and

$$\hat{\sigma}^2_{WLS} = \frac{\left(y - Z\hat{\alpha}_{WLS}\right)^I\left(y - Z\hat{\alpha}_{WLS}\right)}{n-p} \tag{14}$$

is the estimator of variance σ^2 based on the WLS estimator.
The modified MSE is given by:

$$MSE\left(\hat{\alpha}_{WORR}\right) = \alpha^2\sum_{j=1}^{p}\frac{\lambda_j}{\left(\lambda_j + k_{WLS}\right)} + k_{WLS}^2\sum_{j=1}^{p}\frac{\alpha_j^2}{\left(\lambda_j + k_{WLS}\right)^2} \tag{15}$$

The unknown parameters σ^2 and α are replaced with their WLS estimators, $\hat{\sigma}^2_{WLS}$ and σ^2_{WLS} , respectively.

3.1.2 Winsorized Liu (WLIU) Estimator

The Winsorized Liu Estimator based on the LIU estimator [8] is defined as

$$\hat{\alpha}_{WLiu} = \left(\Lambda + 1\right)^{-I}\left(\Lambda + d_{WLS}\right)\hat{\alpha}_{WLS} \;,\quad 0 < d < 1 \tag{16}$$

where the LIU parameter d_{WLS} is obtained using

$$d_{WLS} = \sum_{j=1}^{p}\frac{\left(\alpha_j^2 - \sigma^2\right)}{\left(\alpha_j^2 + \frac{\sigma^2}{\lambda_j}\right)} \tag{17}$$

where the unknown parameters σ^2 and α are replaced by their WLS estimators. $\hat{\sigma}^2_{WLS}$ and $\hat{\alpha}_{WLS}$, respectively.
The modified MSE is given by:

$$MSE\left(\hat{\alpha}_{WLS}\right) = \sigma^2\sum_{j=1}^{p}\frac{\left(\lambda_j + d_{WLS}\right)^2}{\lambda_j\left(\lambda_j + 1\right)^2} + \left(d_{WLS} - 1\right)^2\sum_{j=1}^{p}\frac{\hat{\alpha}_j}{\left(\lambda_j + 1\right)^2} \tag{18}$$

where the unknown parameters are replaced by their corresponding estimates based on the WLS estimator.

3.2 Risk Functions for the New Class of Estimators

Let $\left(\hat{\theta} - \theta\right)^2$ be the quadratic loss function, then $E\left(\hat{\theta} - \theta\right)^2$ is termed as the risk function of the estimator, which in fact is the Mean Square Error (MSE) of estimator $\hat{\theta}$ of parameter θ. In this section, we present the risk functions of the OLS estimator, ORR estimator and the LIU estimator.

3.2.1 Risk Function of OLS Estimator

The risk function of OLS estimator is giving as,

$$\begin{aligned} Risk\left(\hat{\alpha}\right) &= \sigma^2 t_r\left[\left(Z^I Z\right)^{-1}\right] \\ Let \quad Z^I Z &= I_p \\ Risk\left(\hat{\alpha}\right) &= \sigma^2 t_r\left(I_p\right) \\ &= \sigma^2 I_p \end{aligned} \tag{19}$$

3.2.2 Risk Function of ORR Estimator

The risk function of ORR estimator is giving as,

$$Risk\left(\hat{\alpha}_k\right) = \frac{\sigma^2 p}{(1+k)^2} + \frac{k^2\Delta^2\sigma^2}{(1+k)^2} + \frac{\sigma^2}{(1+k)^2}\left[p + k^2\Delta^2\right], \; k > 0, \; \Delta^2 \geq 0 \quad (20)$$

where, Δ^2 is defined as the divergence parameter. It is the sum of squares of the normalized coefficients.

3.2.3 Risk Function of LIU Estimator

The risk function of Liu estimator is giving as,

$$Risk\left(\hat{\alpha}_d\right) = \frac{\sigma^2(1+d)^2 p}{4} + \frac{(d-1)^2\Delta^2\sigma^2}{4} = \frac{\sigma^2}{4}\left[(1+d)^2 p + (d-1)^2\Delta^2\right], \; d > 0, \; \Delta^2 \geq 0 \quad (21)$$

where Δ^2 is defined as the divergence parameter.

4.0 Simulation Study

A Simulation study was carried out to evaluate the performance of the proposed new class of estimators by comparing the performance of the new class of estimators (i.e., using the modified winsorized approach) with the existing estimators. Different ridge estimators corresponding to different values of ridge parameter k are considered under several degrees of multicollinearity (ρ) and contamination proportion (δ). The regressor variables are generated using a simulation design proposed in [22] as

$$X_{ij} = \left(1-\rho^2\right)^{\frac{1}{2}} Z_{ij} + \rho Z_{ip}\,; \; i = 1, 2, \ldots, n; \; j = 1, 2, \ldots, p. \quad (22)$$

where, Z_{ij} is independent standard normal pseudo-random numbers and ρ is linear correlation between any two explanatory variables. The following regression model is used to generate n observations on the response variables:

$$Y = 10 + 4X_1 + 6X_2 + 2X_3 + 8X_4 + \varepsilon$$

where the error variable ε is generated using the contaminated normal distribution. The $\delta\%$ contamination is done using the following mixture of normal distributions:

$$\varepsilon_i \sim f_{\varepsilon_i}(\cdot) = (1-\delta)\times N(0,1) + \delta \times N\left(0,10^2\right)$$

For δ = 0%, 10% , 20% and 30%, and n = 500, the different degrees of multicollinearity have been achieved by generating regressor variables using the model given in (22) for ρ = 0.5, 0.7 and 0.9. The 0.2n points are winsorized at each extreme to reduce the effect of outlier and legitimate contaminants observations. Hence, the 20% Winsorized estimators of OLS, ORR and LIU are obtained respectively. Also, the estimated MSE's (EMSE) of OLS, ORR, LIU, WLS, WORR and WLIU estimators are obtained by replacing the values of unknown parameters with their suitable estimates in their respective MSE expressions.

For the sample size (n), degree of multicollinearity (ρ) and low contamination proportion (δ), the above simulation experiment is repeated 1,000 times and the average EMSE (AEMSE) of these estimators are obtained and reported in the Tables 1- 4

Table 1: Winsorized and Un-Winsorized Estimators for Delta=0.3

Winsorized Estimators	n=500, delta=0.3			Un-Winsorized Estimators	n=500, delta=0.3		
	P=0.5	P=0.7	P=0.9		P=0.5	P=0.7	P=0.9
MSE. WLS	0.007667	0.010996	0.069997	MSE. LS	0.181477	0.638747	1.737099
MSE.WORR	0.005327	0.010992	0.069683	MSE. ORR	0.181476	0.638722	1.736531
MSE.WLIU	0.001072	0.000908	0.037177	MSE. LIU	0.005237	0.005237	0.005237

Table 2: Winsorized and Un-Winsorized Estimators for Delta=0.2

Winsorized Estimators	n=500, delta=0.2			Un-Winsorized Estimators	n=500, delta=0.2		
	P=0.5	P=0.7	P=0.9		P=0.5	P=0.7	P=0.9
MSE. WLS	0.007829	0.015706	0.019227	MSE. LS	0.319029	0.747888	0.031963
MSE.WORR	0.007828	0.015698	0.019196	MSE. ORR	0.319025	0.747853	0.031950
MSE.WLIU	0.002274	0.001136	0.002767	MSE. LIU	0.016764	0.034082	0.338747

Table 3: Winsorized and Un-Winsorized Estimators for Delta=0.1

Winsorized Estimators	n=500, delta=0.1			Un-Winsorized Estimators	n=500, delta=0.1		
	P=0.5	P=0.7	P=0.9		P=0.5	P=0.7	P=0.9
MSE. WLS	0.022316	0.010614	0.010539	MSE. LS	0.036300	0.495127	0.865632
MSE.WORR	0.022275	0.010610	0.010529	MSE. ORR	0.036283	0.049512	0.086586
MSE.WLIU	0.000346	0.001281	0.004050	MSE. LIU	0.000676	0.000723	0.000723

Table 4: Winsorized and Un-Winsorized Estimators for Delta=0.0

Winsorized Estimators	n=500, delta=0.0			Un-Winsorized Estimators	n=500, delta=0.0		
	P=0.5	P=0.7	P=0.9		P=0.5	P=0.7	P=0.9
MSE. WLS	0.008159	0.016198	0.025634	MSE. LS	0.028226	0.038572	0.991980
MSE.WORR	0.008157	0.016190	0.025581	MSE. ORR	0.028227	0.038554	0.991919
MSE.WLIU	0.002384	0.001172	0.002083	MSE. LIU	0.004026	0.000778	0.000723

The first section of Tables 1 to 3 shows the estimated MSE's for the Winsorized estimators that was obtained using our proposed modified Winsorization approach, for various degree of correlation (ρ) and proportion of outliers and/or legitimate contaminants (δ). Similarly, the second section of Tables 1 to 3 also shows the estimated MSE's for the existing or Un-Winsorized estimators. While the first and second sections of Table 4 show the estimated MSE's for the Winsorized and existing estimators with zero contamination proportion. A cursory look at Tables 1 to 4 show that the Winsorized estimators have a significant reduction in their MSE's compared to that of the existing or Un-Winsorized estimators even with as low as 10% contamination proportion (δ). Also in Tables 1 to 4, it was observe that the MSE's for the LIU estimators under both conditions are smaller when compared to that of OLS and ORR respectively.

5.0 Conclusion

A modified winsorized form of the OLS estimator, ORR estimator, LIU estimator is proposed. A cursory look at Tables 1 to 4 reveal that increase in the contamination proportion (δ) and the increase in degree of correlation (ρ) between the independent variables have a negative effect on the MSE, in the sense that it also increases. Also, the simulated study shows that the modified Winsorized biased estimators (i.e., WLS, WORR and WLIU) gave better performance in terms of their MSE values when compared to their corresponding OLS, ORR and LIU estimators. Specifically, the Winsorized LIU estimator with appropriate δ might be considered over the ridge regression counterpart, as observed from the simulated results.

6.0 References

[1] Birkes, D., & Dodge, Y. (1993).“Alternative methods of regression”. New York: John Wiley & Sons.

[2] Draper, N. R., & Smith, H. (1998). Applied Regression Analysis (3rd Ed.). New York: John Wiley & Sons

[3] Montgomery, D. C., Peck, E. A., & Vining, G. G. (2006). Introduction to Linear Regression Analysis (3rd ed.). New York: John Wiley & Sons.

[4] Chatterjee, S., & Hadi, A. S. (2012). Regression analysis by example (5th ed.). Hoboken, NJ: John Wiley & Sons.

[5] Hoerl, A. E., & Kennard, R. W. (1970a). “Ridge regression: Biased Estimation for Nonorthogonal Problems”. Technometrics, 12, 55-67.

[6] Hoerl, A. E., & Kennard, R. W. (1970b).“Ridge regression: Applications to Nonorthogonal Problems”. Technometrics, 12, 69-82.

[7] Hoerl, A.E., Kennard, R.W., Baldwin, K.F. (1975). “Ridge regression: some simulation”. Communications in Statistics 4: 105-123.

[8] Liu, K. (1993). A New Class of Biased Estimate in Linear Regression. Communications in Statistics – Theory and Methods, 22, 393-402.

[9] Liu, K. (2003). Using Liu-Type Estimator to Combat Collinearity. Communications in Statistics – Theory and Methods, 32(5), 1009-1020.

[10] Matt, N. W., Carlos, A. G. G., and Dason, K. (2013). Assumptions of Multiple Regression: Correcting Two Misconceptions. Practical Assessment, Research & Evaluation, Vol 18, No 11.

[11] Tabachnick, B. G. and Fidell, L.S. (2007). Using Multivariate Statistics (Fifth ed.). Pearson Education, Inc. USA

[12] Huber, P. J. (1973). Robust Regression: Asymptotics, Conjectures, and Monte Carlo. Annals of Statistics, 1, 799–821.

[13] Rousseeuw, P. J., & Leroy, A. M. (1987).“Robust Regression and Outlier Detection”. New York: John Wiley & Sons.
[14] Jadhav, Nileshkumar H. and Kashid, D N. (2014) "Robust Winsorized Shrinkage Estimators for Linear Regression Model," Journal ofModern Applied Statistical Methods: Vol. 13: Iss. 2, Article 7.
[15] Iglewicz, B. and Hoaglin, D. C. (1993). How to Detect and Handle Outliers. American Society for Quality Control, Milwaukee, WI.
[16] Osborne, J. and Amy, O. (2004). The Power of Outliers (and why researchers should always check for them). Practical Assessment Research & Evaluation, 9(6)
[17] Andrews, D. F. (1974). A Robust Method for Multiple Linear Regression. Technometrics, 16(4), 523-531.
[18] Yale, C., & Forsythe, A. B. (1976).“Winsorized Regression.Technometrics”, 18, 291-300.
[19] Tan, W. Y., & Tabatabai, M. A. (1988). A Modified Winsorized Regression Procedure for Linear Models. Journal of Statistical Computation and Simulation,30, 299-313.
[20] Chen, L. A., Welsh, A. H., & Chan, W. (2001). “Estimators for the Linear Regression Model based on Winsorized Observations”. Statistica Sinica, 11, 147-172.
[21] Kan, B., Alpu, O., &Yazici, B. (2013). “Robust ridge and robust Liu estimator for regression based on the LTS estimator”. Journal of Applied Statistics,40(3), 644-655.
[22] McDonald, G. C., & Galarneau, D. I. (1975). A Monte Carlo Evaluation of Some Ridge-type Estimators. Journal of the American Statistical Association,70(350), 407-416