International Journal of Statistics and Applied Mathematics 2022; 7(3): 85-93

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2022; 7(3): 85-93

www.mathsjournal.com
Received: 05-03-2022
Accepted: 07-04-2022

Bello A Rasheed
Department of Mathematics and Statistics, Federal Polytechnic Bauchi, Bauchi State, Nigeria

Robiah Adnan
Department of Mathematics, Faculty of Science, UTM, Skudai, Johor, Malaysia

Seyed E Saffari
Centre of Education, Sabzevar University of Medical Sciences, Sabzevar, Iran

Dauda M Atiyaye
Department of Mathematics and Statistics, Federal Polytechnic Bauchi, Bauchi State, Nigeria

Corresponding Author:
Bello A Rasheed
Department of Mathematics and Statistics, Federal Polytechnic Bauchi, Bauchi State, Nigeria

Robust PC with wild bootstrap estimation of linear model in the presence of outliers, multicollinearity and heteroscedasticity error variance

Bello A Rasheed, Robiah Adnan, Seyed E Saffari and Dauda M Atiyaye

DOI: https://doi.org/10.22271/maths.2022.v7.i3b.826

Abstract
The regression model estimator is considered efficient if it is robust and resistant to the presence of heteroscedasticity variance, multicollinearity or unusual observations called outliers. However, in regard to these problems, the wild bootstrap and robust wild bootstrap are no longer efficient since they could not produce the smallest variance. Hence this research investigates the use of robust PC with wild bootstrap techniques on regression model as an estimator for real and simulation data in a situation where multicollinearity, heteroscedasticity and multiple outliers are present. This paper proposed a robust procedure based on the weighted residuals which combined the Tukey bisquare weighted function, principal component analysis (PCA) to remedy the multicollinearity problems, least trimmed squares (LTS) estimator, robust location and scale, and the wild bootstrap sampling procedure of Wu and Liu that remedy the heteroscedasticity error variance. RPCW Boot Wu and RPCW Boot Liu were obtained through a modified version of R Boot Wu and R Boot Liu. Finally, based on the real data and simulation study, the performance of the RPCW Boot Wu and RPCW Boot Liu is compared with the existing R Boot Wu, R Boot Liu and also with Boot Wu, Boot Liu using the biased, RMSE and standard error. The numerical example and simulation study shows that the RPCW Boot Wu and RPCW Boot Liu techniques have proven to be a good alternative estimator for regression model with lower standard error values.

Keywords: Wild bootstrap, heteroscedasticity, multicollinearity and multiple outliers

1. Introduction
In regression analysis, the ordinary least square is widely used to estimate the model parameters mostly because of tradition for optimal properties and ease of computation. Unfortunately, the mathematical elegance that makes the estimator so popular relied on a number of fairly strong and often unrealistic assumptions. Regression coefficients that involve tests of significance and confidence intervals are available in different popular statistical packages that researchers use regularly. But the results of tests statistics and the coverage probability of confidence intervals become valid largely depending on the extent to which these model's assumptions are met. However, if these assumptions are violated, the ordinary least square will no longer produce the best variance, resulting to the inefficiency in the parameter of the model.

One of these assumptions is the assumption of constant variance. The assumption of constant variance is one of the basic requirements of regression model. Researchers encounter a situation in which the variance of the response variable is relate to the value of one or more regressor variables resulting in heteroscedasticity. A common reason for the violation of this assumption is for the response variable to follow a probability distribution in which the variance is functionally related to the mean. Heteroscedasticity is said to be present if this assumption is violated. In the presence of heteroscedasticity, the OLS estimator will still remain unbiased. However, the most harmful consequence of heteroscedasticity would be the parameter covariance matrix. The elements in the diagonal matrix that are used to estimate the standard error becomes biased and unreliable [1]. On the other hand, the assumption of multicollinearity exists if there is no exact linear relationship between the explanatory variables. In the presence of multicollinearity, the OLS estimator will result in producing infinite variance that will lead to misleading interpretation in the test statistics.

In practice, the situation become worse when there are outliers in the data. Presence of outliers in the data will desterilize the parameter estimation in the model by inflating the test statistics which results in given wrong conclusions. However, most of the statistical data usually do not completely satisfy assumptions often made by the researchers which result in a dramatic effect on the quality of statistical analysis.

A heteroscedasticity bootstrap technique was firstly introduced by [2, 3]. They proposed the wild bootstrap technique which gives a better performance for the parameter estimates of the regression coefficients when the model exhibits both homoscedasticity and heteroscedasticity models. This type of weighted bootstrap is called the wild bootstrap in the literature. Wild bootstrap is a resampling procedure that is usually used to estimate bias, standard error and to construct the value of confidence interval of an estimator. The estimate of standard error and sampling distribution of the robust regression model can be evaluated from the drawn samples. [2, 3] described the wild bootstrap as procedures for treating sample data from the population at which the repeated sample is being drawn. In regression analysis, wild bootstrap method is suitable because it relaxes the assumption about the error terms which stated that the error distribution must follow a normal distribution [4]. To handle the multicollinearity problems, Principal component regression was introduced by [5] which is more precise than the OLS method in multicollinearity situation. The robust estimation is mainly used to overcome the problem of outliers by using a suitable weighted function of Tukey bisquare function to down weight the effect of outliers. The robust estimator used in this research is least trimmed squares (LTS) estimator that was introduced by [6]. We choose this weighted function also to improve the asymptotic relative efficiency of our LTS estimator.

Several attempts have been made to use the procedure of [2, 3] wild bootstrap techniques to remedy the problem of heteroscedasticity error variance [7] proposed a promising robust wild bootstrap estimator based on brain morphology to detect association between brain structure and covariates in order to diagnose severity of disease, such as age, IQ and genotype. A similarly modified wild bootstrap for quantile regression estimators was proposed and a simulation study was conducted based on median regression to relate with a number of bootstrap methods. Using a simple finite correction, the result indicates that the wild bootstrap can account for general forms of heteroscedasticity in regression model with fixed design point [8].

Most recently, a modified weighted bootstrap estimation method based on LTS to handle outliers and heteroscedasticity was proposed. This method will identify the exact number of outliers in the data and form two groups of observation, where the bootstrap sample is performed on these groups. The Alarmgir redescending M-estimator (ALARM) weighted procedure is used to estimate the regression model of each bootstrap sample [9]. The idea of this bootstrap method is to protect against excessive number of outliers and ensures efficient results [10] proposed the robust wild bootstrap based on [2, 3]. They disclosed that the problem of classical bootstrap is that the proportion of outliers involved in the bootstrap sample might be greater than that of the original data. Hence, the entire inferential procedure of bootstrap would be erroneous in the presence of outliers. They introduced robust wild bootstrap estimation based on MM-estimator introduced by [11]. This wild bootstrap procedure was to handle the problems of outlying observation and heteroscedasticity in the model. This study proposed alternative techniques that can handle problems of multicollinearity, heteroscedasticity and outliers in the model. We use a suitable combination of robust principal component regression with wild bootstrap techniques of [2, 3]. We proposed a slightly modification of robust wild bootstrap of MM-estimation, which is a combination of wild bootstrap and robust method. This study would examine the performance of the proposed method as an alternative to the existing methods for handling the multiple problems of multicollinearity, heteroscedasticity and outliers.

However, from the literature there is not much work devoted to this aspect of wild bootstrap method in a situation when multicollinearity, heteroscedasticity and outliers occur together. The vital role of wild bootstrap is to handle the problems of heteroscedasticity but it is not resistant to multicollinearity and outliers. We discussed the methodology of this research in section 2. In section 3, we introduced the newly proposed method and its performance was presented. Section 4 will contain the detailed conclusion of the study.

2. Methodology

A simulation study was designed to assess the performance of wild bootstrap [2, 3] and the robust wild bootstrap of [2, 3] with their proposed robust PC with wild bootstrap of [2, 3]. We generate the covariance of $x_1,\ x_2\ and\ x_3$ using the multiple linear regression model based on a combination of different regression conditions. Here, we follow a similar procedure used by [10]. The considered design for this experiment involved a regression model with intercept and covariance values. Suppose we consider the following linear model

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \sigma_i\, \varepsilon_t \qquad (1)$$

Where $i = 1, 2, ..., n$, the covariance values $x_1,\ x_2\ and\ x_3$ were generated using the following equation

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

where $i = 1, 2, ..., n, \quad j = 1, 2, 3$ and the parameter z_{ij} are the standard normal random numbers generated by the normal distribution and residuals are drawn from normal distribution with mean zero and variance 1, when no outliers were considered and for all I under heteroscedasticity $\sigma = 1$. The data is generated using $\beta_0 = \beta_1 = \beta_2 = \beta_3 = 1$. Next, we start contamination of the data. Randomly we replace some good observations of *i.i.d.* normal errors ε_t's. Now our main interest is to obtain a regression design that includes multicollinearity, heteroscedasticity and outliers in the model. We study the performance of each estimator according to severity of multicollinearity by using different degree of correlation ρ between the regressor variables. At the same time, the performance of the estimators was observed by increase percentage of outliers and the considered percentages of outliers are 0%, 10%, and 20% respectively. We form the heteroscedasticity generating procedure following [10, 12, 13] effort, where

$$\sigma_i^2 = \exp(2.6 x_{1i}) \qquad (3)$$

is used to generate the heteroscedasticity. Now the regression model of contaminated heteroscedastic is given as

$$y_i = \beta_0 + \beta_1 x1 + \beta_2\, x2 + \beta_3\, x3 + \sigma i \varepsilon_{t(Cont.)} \quad (4)$$

We first considered the sample size of n = 20 observations and applied the principal component analysis to estimate the component that contains all the information of the original data. In this design these components are then replicated five times to generate samples of n = 100, respectively. Here, we followed a similar procedure proposed by [10] who utilized the replication of covariate values to create large samples. For each simulated data set, with different sample size we fit the Boot Wu, Boot Liu, R Boot Wu, R Boot Liu, RPC Boot Wu and RPC Boot Liu linear regression model.

2.1 Wild Bootstrap Based on Wu's

This bootstrap procedure has been suggested by [2] for the situation when the additional assumption of $\mathbf{E(\varepsilon_i \mid X_i) = 0}$ is appropriate. The bootstrapping procedure of classical OLS bootstrap is slightly modified to estimate t* value. This is performed by drawing a random sample with replacement from an auxiliary distribution that has mean zero and variance one and attached with the fitted values of the model to obtain a fixed X-bootstrap of 4. As another alternative for the Wu's bootstrap procedure, the value of t* can be obtained with replacement using the following procedures;

Step 1: Fit an OLS regression model to the original sample of observations to get $\hat{\beta}$ the fitted values of

$$\hat{y}_i = f(x_i, \hat{\beta}) \quad (5)$$

Step 2: Use the fitted values to compute the residuals of $\varepsilon_i = y_i - \hat{y}_i$ of the fitted model.

Step 3: Generate the random sample of t* with replacement from a_i^R observations where

$$a_i = \frac{\hat{\varepsilon}_i - \bar{\hat{\varepsilon}}_i}{\sqrt{n^{-1}\sum_{i=1}^{n}(\hat{\varepsilon}_i - \bar{\hat{\varepsilon}}_i)^2}}$$, i = 1, 2, 3… n. and $\bar{\hat{\varepsilon}}_i = n^{-1}\sum_{i=1}^{n}\hat{\varepsilon}_i$.

The regression model that has intercept term $\bar{\hat{\varepsilon}}_i$ is usually approximately equal to zero [1].

Step 4: Obtain a fixed=x bootstrap sample of y_i^{*b} where

$$y_t^{*b} = f(x_i, \hat{\beta}_{ols}) + t_i^* \hat{\varepsilon}_i (1 - h_{ii)})^{-1} \quad (6)$$

and $h_{ii} = \mathbf{x_i^T (x^T x)^{-1} x_i}$ is the i-th leverage, y_t^{*b} is the new bootstrap response variable that can be used to obtain the first wild bootstrap coefficients and $\hat{\beta}^*$ is the least squares estimate based on the resample, $\hat{\boldsymbol{\beta}}^* = \mathbf{(x^T x)^{-1} x^T y^*}$

Step 5: Regress the obtained bootstrapped values of y_i^{*b} on the fixed x to obtain $\boldsymbol{\beta}^*$.

Step 6: Repeat Step 3 and Step 4 for k times to get $\hat{\boldsymbol{\beta}}^{*1}, \ldots, \hat{\boldsymbol{\beta}}^{*k}$ where K is the number of bootstrap replicates. This procedure is a nonparametric application of Wu's bootstrap sampling scheme, since the resampling is performed from the empirical distribution function of the normalized residuals. This method is referred to as Wu's bootstrap sample and denoted by Boot Wu.

Following the idea of wild bootstrap of [2, 3, 10] suggested a slight modification of the procedure of generating the t* value. The t* is randomly selected from auxiliary distribution that has third central moment equal to one, with zero mean and unit variance. It has been shown that when this is the case, Wu's share the usual second order asymptotic properties of the classical bootstrap. Put differently, a restriction is added that the third central moment be equal to one and such a selection is used to correct the skewness term in the edge worth expansion of the sampling distribution of $\mathbf{I}'\hat{\boldsymbol{\beta}}$, where I is an n-vector of ones. The procedure of Liu bootstrap can be performed by drawing a random sample of t* in the following ways. To generate the bootstrap sample, here we considered three construction of t* for the bootstrap regression model. If one assumes that t* puts mass only on two point distribution, then:

Step 1: $t^* = S_i - E(S_i), i = 1, 2, \ldots$, n, and $S_1, S_2, \ldots, S_n$ are independently and identically distributed normal distribution having density of $g_z(x) = [\alpha^{\beta} / (\beta - 1)!] x^{\beta - 1} e^{-ax} I_{(x \geq 0)}$ and α =2 and β = 4.

Step 2: $t_i^* = N_i M_i - E(N_i) E(M_i)$ $i = 1, 2, \ldots, n$ where $N_1, N_2, \ldots, N_n$ are independently and identically distributed normal distribution with mean $(1/2)(\sqrt{17/6}) + \sqrt{1/6}$ and variance $1/2$. $M_1, M_2, \ldots, M_n$ are also i.i.d. normally distributed with mean $(1/2)(\sqrt{17/6}) - \sqrt{1/6}$ and has variance $1/2$. N_i's and M_i's are independent.

Step 3: $t^* = (\delta_1 + V_{i,1}/\sqrt{2})(\delta_2 + V_{i,2}/\sqrt{2}) - \delta_1\delta_2$ where $V_{i,j}$'s are independent N(0,1)- distributed variables and where $\delta_1 = (3/4 + \sqrt{17/12})^{1/2}$ and $\delta_2 = (3/4 - \sqrt{17/12})^{1/2}$ respectively.

However, the three bootstrap procedures will generate the random sample of t_i^* Liu's. Both procedures will produce third central moments equal to one [10] suggested the most popular choice for the distribution of t_i^* is the second procedure as it always gives better results than the remaining one. Following [10], this research will make use of the second method for generating its bootstrap sample of t_i^*. The bootstrap procedure is called Liu bootstrap or Boot Liu.

2.2 Robust Wild Bootstrap MM-Estimator

Considering the idea of the classical bootstrap procedure based on [2, 3]. Another alternative of modified wild bootstrap techniques which is more robust was introduced to remedy the problem of heteroscedasticity and outliers [10]. This bootstrap method was based on MM- estimator procedure. The quantity of t* is obtained from equation (6) that is a robust normalized residual based on median and normalized median absolute deviation instead of mean and standard deviation. The bootstrap procedure of MM-estimator is summarized as follows:

Step 1: Obtain the fitted model of $y_i = x_i\beta + \varepsilon_i$ using MM-estimator of the sample data to estimate the robust parameter coefficient of $\hat{\boldsymbol{\beta}}_{\mathbf{MM}}$.

Step 2: Estimate the residuals of the MM-estimator of $\varepsilon_i^{MM} = y_i - \hat{y}_i$. Assign the estimated weight to each MM-residuals, ε_i^{MM} where the weight will equal to

$$w_{ii} = \begin{cases} 1 & \text{if } \left|\varepsilon_i^{MM}\right| / \sigma_{MM} \le c \\ c / (\left|\varepsilon_i^{MM}\right| / \sigma_{MM}) & \text{if } \left|\varepsilon_i^{MM}\right| / \sigma_{MM} > c \end{cases} \quad (7)$$

where c is the turning point. c is an arbitrary constant which is usually chosen between 2 and 3

Step 3: The estimate of the final weighted residuals for the robust MM-estimate of ε_i^{MM} is obtained by multiplying the weight with the residuals of MM-estimator of step 2.

Step 4: Obtain the bootstrap sample of (y_i^*, X), and

$$y_i^* = x\hat{\beta}_{MM} + \frac{t_i^* \varepsilon_i^{WMM}}{(1-h_{ii})} \quad (8)$$

where the estimate of t* is the required random sample obtained from step 4.

Step 5: Apply the OLS estimation procedure on the bootstrap sample of (y_i^*, X). This estimate is denoted by

$$\mathbf{R}_{\hat{\boldsymbol{\beta}}^*} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}^* \quad (9)$$

Step 6: Repeat step 4 and 5 for B times, where B is the required number of bootstrap replicates. The bootstrap procedure obtained from these techniques is called Robust Wild Bootstrap MM-estimator based on Wu and Liu i.e. (R Boot Wu) and (R Boot Liu).

2.3 Robust PC with Wild Bootstrap

We have discussed the robust wild bootstrap and classical wild bootstrap procedures. It is now evident that the OLS suffers a huge setback in the presence of outliers and multi collinearity. On the other hand, the MM-estimator will be affected in the presence of multicollinearity. Since both the robust wild bootstrap and classical wild bootstrap are not resistant to multicollinearity, in this study we proposed to use the high breakdown using a suitable weighted function that will provide high efficient robust LTS estimator [6] to obtain the robust residuals. This estimator has a breakdown point of $[((n-p)/2+1)/n]$. The coefficient $\hat{\boldsymbol{\beta}}_{\mathbf{LTS}}$ is to be estimated by minimizing $\sum_{i=1}^{h} \varepsilon_i^2$ where $\varepsilon_{(1)}^2 \le \varepsilon_{(2)}^2 \le \ldots \le \varepsilon_{(n)}^2$ are the i-th ordered residuals from the smallest to the largest bootstrap. This procedure allows us to trim a certain number of standardized residuals identified as outliers in each bootstrap sample $h = [n/2] + [(p+1)/2]$ where p is the number of parameters and n is the sample size. The procedure for estimating the Wu's bootstrap sample based on the robust wild bootstrap least trimmed squares is summarized as follows:

Step 1: We first centre and scale our observation using the following equation:

$$X^* = \frac{x_i - \bar{x}_i}{\sqrt{\sum (x_i - \bar{x}_i)^2}} \quad (10)$$

Step 2: Compute the estimate of the correlation matrix for centre and scale observation: $X^{*'}X^*$

Step 3: Compute the eigenvalues λ_i and eigenvecror, V using the correlation matrix obtained in step two.

Step 4: Compute the components Z, where $Z = X^*V$

Step 5: Compute the eigenvalues for the component to check the total amount of information captured by each component. The component associated with the smallest eigenvalues should not be deleted.

Step 6: Fit the regression model $y_i = f(x_i, \beta_{LTS})$ using the LTS method to the original sample of observation to get $\hat{\beta}_{LTS}$ and hence the fitted model now becomes

$$y_i = f(x_i, \beta_{LTS})$$

Step 7: Compute the residuals of the fitted function $\hat{\varepsilon}_{iLTS} = y_i - \hat{y}_i$ and estimate the standardized residuals.

Step 8. Estimate the initial weight for all the cases by using the inverse of this absolute fitted value obtained in step 3 and denotes it as w_{1i} which now becomes our initial weight.

Step 9: Estimate the scaled residuals r_i using the robust median and robust normalized median absolute deviation. The scaled residuals are defined as

$$r_i = \frac{\varepsilon_i}{MAD} \text{ and } MAD = \frac{1}{.6745} median\left\{\left|\varepsilon_i - median\{\varepsilon_i\}\right|\right\} \quad (11)$$

The constant value is 0.6745 and it is called the turning constant. It provides an unbiased estimate of σ for independent observations from a normal distribution.

Step 10: Estimate of the final weighted can be acquired from a suitable robust weighted function procedure of M-estimator. In this paper, we will use the Tukey bisquares weighted function which is defined as

$$w_{2i} = \begin{cases} \left[1 - \left(\frac{r_i}{c}\right)^2\right]^2, & |r_i| \le 4.685 \\ 0 & |r_i| > 4.685 \end{cases}, \quad i = 1, 2, ..., 51 \qquad (12)$$

where r_i is defined as the scaled residual or standardized residuals of the LTS obtained from step 2 and c is the tuning constant which is equal to 4.685.

Step 11: We obtained the estimate of the weighted residuals of LTS denoted as w_i^{LTS} by multiplying w_{1i} obtained from step 3 with the Tukey bisquares weighted function of w_{2i}. However, because of the presence of heteroscedasticity in the data, we have now modified the bootstrap schemes that will produce an efficient estimate of the regression parameter. This modified bootstrap method can also be used to obtain the standard error which is asymptotically corrected under heteroscedasticity of unknown form.

Step 12: Construct a bootstrap sample $(\mathbf{y_i^*}, \mathbf{x})$ where for each i, draw a value $\mathbf{t^*}$ with replacement from a distribution with zero mean and unit variance attached to $\hat{y}_i$. The fixed-x- bootstrap values $\mathbf{y_i^{*b}}$ can be obtained by using the following equation:

$$\mathbf{y_i^{*b}} = f(x_i, \hat{\beta}_{LTS}) + t^* \hat{\varepsilon}_i / \sqrt{1-h_{ii}} \qquad (13)$$

where $h_{ii} = x'(x'x)^{-1}x_i$ is the i- th leverage. The value of i-th leverage is used to reduce the influence of cases that are large leverage. We modified the Wu's procedure by computing the robust normalized residuals based on the median and normalized median absolute deviations (NMAD) to replace the mean and standard deviation which are not robust. Then we have, $\mathbf{t^*} = {}^R\mathbf{a}_i, {}^R\mathbf{a}_1, ..., {}^R\mathbf{a_n}$ and

$${}^R a_i = \frac{\hat{\varepsilon}_i^{WLTS} - median(\hat{\varepsilon}_i^{WLTS})}{NMAD_{norm}(\hat{\varepsilon}_i^{WLTS})} \qquad (14)$$

where the estimate of the normalized median absolute deviation of the weighted residuals is obtained from equation (16)

$$NMAD = \text{median}\left\{\left|\hat{\varepsilon}_i^{WLTS} - median(\hat{\varepsilon}_i^{WLTS})\right|\right\} \qquad (15)$$

Step 13: Fit the LTS to the bootstrapped values y_i^{*b} on the fixed -x- to obtain $\hat{\boldsymbol{\beta}}^{*b}{}_{LTS}$.

Step 14: Repeat Step 7 and Step 8 for k times to get $\hat{\boldsymbol{\beta}}^{*bi}{}_{LTS}, ..., \hat{\boldsymbol{\beta}}^{*bk}{}_{LTS}$ where k is the bootstrap replications. The bootstrap procedure obtained from these techniques is called Robust PC with Wild Bootstrap LTS estimator based on Wu or RPC Boot Wu. In this article we also want to modify the robust MM-estimator based on Liu's algorithm, as the significant difference between Wu and Liu implementation of wild bootstrap is the choice of the random sample t_i^*. In the proposed robust PC with wild bootstrap based on Liu wild bootstrap, for the choice of t_i^* we followed exactly the same procedures as the robust wild bootstrap of Liu. We called this wild bootstrap RPC Boot Wu.

3. Evaluation of the Bootstrap Method

To evaluate the performance of different robust wild bootstrap procedure used in this paper, we estimate the bias, RMSE and standard error. The best results from the estimate are the one that produced the smallest bias, RMSE and standard error. We estimate the bias, RMSE and standard error by employing the formulae of these estimates. The estimate of LTS is used as the initial estimate for the estimation of these regression models. The procedures continue to further perform the bootstrap estimate of the bias, RMSE and standard error of the RW Boot Wu LTS and RW Boot Liu estimate. The numerical calculation of Boot Wu, Boot Liu and the R Boot Wu and R Boot Liu is to perform the same procedure. The relevant formulas for the estimates are given as follows:

$$\hat{\beta}^*_{(bLTS)} = \frac{\sum_{b=1}^{k} \hat{\beta}^{*b}_{LTS}}{k} = \bar{\hat{\beta}}_{bLTS} \qquad (16)$$

The corresponding estimate of bias bootstrap is given by $\hat{\beta}^*_{bLTS} - \hat{\beta}_{bLTS}$ and the estimate of the bootstrap standard error is given as

$$\text{SE}(\hat{\beta}^*_{(bLTS)}) = \sqrt{\frac{\sum_{b=1}^{k} \hat{\beta}^{*b}_{(bLTS)} - (\hat{\beta}^*_{bLTS})^2}{k}} \qquad (17)$$

The estimate of the bootstrap variance is obtained using the following formula

$$s^2_{\hat{\beta}_{bLTS}} = \frac{1}{k-1} \Sigma (\hat{\beta}^{*b}_{LTS} - \hat{\beta}_{(bLTS)})^2 \qquad (18)$$

The covariance is given as the off diagonal values of the matrix and the estimate is given as

$$\text{cov}(\hat{\beta}_{bLTS}) = \frac{1}{k-1} \sum_{b=1}^{B} (\hat{\beta}^{*b}_{LTS} - \hat{\beta}_{(bLTS)})(\hat{\beta}^{*b}_{LTS} - \hat{\beta}_{(bLTS)})' \qquad (19)$$

The robust mean squared error is computed as

$$RMSE(\hat{\beta}_{bLTS}) = (bias)^2 + \text{var}(\hat{\beta}_{bLTS}) \qquad (20)$$

and the robust root mean squared error of the models is defined as

$$RRMSE = \sqrt{(bias)^2 + \text{var}(\hat{\beta}_{bLTS})} \qquad (21)$$

4. Example using Real Data Sets

This section will discuss the application of the RPC Boot Wu and RPC Boot Liu methods on real data by considering the numerical example that will show the advantages of the proposed method with respect to the other estimators, Boot Wu, Boot Liu, R Boot Wu and R Boot Liu estimator in the presence of outlier multicollinearity and heteroscedasticity error variance. The cigarette data is taken from [14]. The dataset contains measurements of weight and tar, nicotine, and carbon monoxide (CO) content for 25 brands of cigarettes. We checked whether the data set contained any outliers or not using the LTS residuals. It was discovered that five observations (about 20% of the sample of size 25) were identified as outliers.

We apply variance inflation factor (VIF) to test for the presence of multicollinearity in the data. The results disclosed that there is high correlation between the covariates. On the other hand, the modified robust Goldfeld-Quadl test is used for heteroscedasticity test and the null hypothesis is rejected which indicated that there is heteroscedasticity in the data. The wild bootstrap, robust wild bootstrap and robust PC with wild bootstrap methods were then applied to the data. The results obtained are based on 1000 bootstrap replicates are presented in Table 1 along with the standard errors, bias and RMSE of the parameter estimates from wild bootstrap, robust wild bootstrap and robust PC with wild bootstrap methods. Based on the results, it is interesting to observe that both the standard error bias and RMSE of the wild bootstrap method tend to be larger followed by robust wild bootstrap. The robust PC with wild bootstrap methods has the smallest standard errors.

Table 1: The Parameter estimate, Standard error, Bias and RMSE of non-robust wild bootstrap, robust wild bootstrap and robust PC with wild bootstrap of a collection of 25 cigarette data.

Par. Estm.	Boot Wu	Boot Liu	R Boot Wu	R Boot Liu	RPC Boot Wu	RPC Boot Liu
Estimate	11.87	11.874	12.952	12.952	3.1633	2.602
S.E	3.6877	3.5129	1.1587	0.8014	0.0077	0.0066
Bias	-4.1698	1.9157	-1.1885	-0.3058	-0.0045	-0.0013
RMSE	5.5665	4.0013	1.6598	0.8578	0.0089	0.0067
Estimate	-15.933	-15.974	-17.075	-16.923	0.7602	0.9581
S.E	0.2665	0.2469	0.1477	0.1290	0.0323	0.0322
Bias	0.1418	-0.2160	0.1847	0.0076	-0.0396	0.0040
RMSE	0.3019	0.3281	0.2365	0.1292	0.0511	0.0324
Estimate	-10.786	-10.809	-4.1841	-4.1112	0.6411	-2.3441
S.E	4.3248	4.0302	1.7351	1.3688	0.0447	0.0390
Bias	-0.7468	4.7028	-2.4218	0.2219	-0.0090	0.0170
RMSE	4.3888	6.1935	2.9792	1.3867	0.0456	0.0425
Estimate	11.7273	11.791	47.171	47.764	-0.8764	0.3854
S.E	4.1282	4.022	1.2220	0.8707	0.2360	0.2251
Bias	3.1917	-3.5044	1.4908	-0.0539	0.0297	-0.0366
RMSE	5.2181	5.3345	1.9277	0.8723	0.2379	0.2281

This cannot provide evidence of our final conclusion yet, which can only be done investigating the results obtained from real data. However, we can make a reasonable interpretation that the robust wild bootstrap and classical wild bootstrap are affected by multicollinearity and outliers.

5. Examples using Simulated Data Sets

The example of real data sets obtained in section 5 have shown that the RPC Boot Wiu and RPC Boot Liu coefficient estimates are generally found to be the most stable robust bootstrap estimates with the smallest RMSE, bias and standard error. This section will further investigate the robustness of our proposed RPC Boot Wu and RPC Boot Liu methods by performing a simulation using a multiple linear regression model of three regressor variables. Table 2 presents simulation results of the bias, RMSE and standard error of the parameter estimates obtained from different degrees of multicollinearity and percentage of outliers. As shown in the tables, the performance of Boot Wu and Boot Liu estimator is poor since the standard error is large when compared with the R Boot Wu, R Boot Liu, RPC Boot Wu and RPC Boot Liu at 10% level of contamination. The effect becomes very serious as the percentage of outliers increases to 20%. The R Boot Wu and R Boot Liu estimator without principal component techniques shows the worst performance since the standard error is larger than the proposed methods. On the other hand, incorporation of the principal component techniques reduces the standard error of values of the RPC Boot Wiu and R Boot Liu estimators. It is worth mentioning when the sample size, percentage of outliers and the degree of multicollinearity is increased to a sample size n = 100, both the Boot Wu and Boot Liu, R Boot Wu and R Boot Liu estimators show the worst performance since the standard error is very high when compared with the proposed methods. Results from the table also describe the estimate of the bias and RMSE for both methods. The proposed method seems to be the most resistant estimator towards the presence of 10% outliers and 0.50 level of multicollinearity by producing the smallest values of bias and RMSE as compared with the other methods. Furthermore, when the percentage of outliers increases to 20% and the degree of multicollinearity is 0.99, it is reported that the RPC Boot Wu and RPC Boot Liu estimators become superior by producing the lowest values of bias and RMSE. The performance of each method is described in Table 2- Table 4, in which each method is evaluated based on the lowest bias, RMSE and standard error values. Out of all methods, the RPC Boot Wu and RPC Boot Liu is the most robust and resistant to the presence of multicollinearity, heteroscedasticity and multiple outliers.

Table 2: Bias, RMSE and standard error of n = 20 and n = 100 (bold) for 0% level of contaminated data based on non-robust wild bootstrap, robust wild bootstrap and robust PC with wild bootstrap from normal distribution with 3 regressor variables.

Coef.	Method	$\rho = 0.2$			$\rho = 0.5$			$\rho = 0.99$		
		Bias	RMSE	SE	Bias	RMSE	SE	Bias	RMSE	SE
β_0	Boot Wu	-1.297	3.236	2.965	-1.703	1.741	0.364	2.235	3.856	3.143
		-1.113	1.646	1.213	-0.743	1.182	0.920	0.160	1.059	0.677
	Boot Liu	0.249	1.163	1.136	-0.004	0.214	0.214	0.277	0.788	0.737
		-0.293	1.221	1.185	-0.188	1.751	1.741	-0.048	0.695	1.058
	R Boot Wu	-0.003	0.445	0.445	0.001	0.476	0.476	-0.461	0.713	0.543
		0.276	0.350	0.216	0.037	0.165	0.161	0.286	0.379	0.250
	R Boot Liu	-0.368	0.460	0.275	-0.014	0.299	0.299	-0.186	0.224	0.124
		0.016	0.036	0.032	-0.026	0.060	0.055	-0.019	0.086	0.084
	RPC Boot Wu	0.037	0.051	0.035	0.010	0.037	0.035	-0.007	0.048	0.048
		-0.002	0.014	0.018	0.000	0.034	0.067	0.000	0.041	0.053
	RPC Boot Liu	0.004	0.016	0.015	0.009	0.018	0.016	-0.009	0.013	0.009
		-0.002	0.018	0.014	-0.009	0.067	0.034	-0.001	0.053	0.041
β_1	Boot Wu	-3.193	4.501	3.173	-4.744	4.804	0.758	-3.352	10.848	10.317
		-2.681	2.943	1.215	-1.580	1.791	0.844	1.318	9.955	3.271
	Boot Liu	2.354	2.583	1.062	0.339	0.402	0.216	-1.670	4.612	4.299
		0.074	1.344	1.342	0.591	1.997	1.907	-8.126	3.527	5.750
	R Boot Wu	-0.096	0.532	0.523	-0.170	0.577	0.551	-1.335	2.574	2.200
		0.370	0.413	0.185	0.023	0.111	0.108	-0.636	0.916	0.660
	R Boot Liu	-0.584	0.644	0.270	-0.105	0.232	0.207	-1.457	1.565	0.572
		0.002	0.030	0.030	-0.035	0.059	0.047	0.073	0.546	0.542
	RPC Boot Wu	-0.048	0.054	0.025	-0.002	0.030	0.030	0.002	0.011	0.011
		0.000	0.003	0.003	0.000	0.001	0.002	0.000	0.002	0.000
	RPC Boot Liu	-0.003	0.011	0.011	-0.011	0.015	0.011	-0.002	0.005	0.004
		-0.006	0.006	0.003	0.002	0.003	0.001	0.000	0.000	0.002
β_2	Boot Wu	5.602	6.269	2.815	9.237	9.258	0.628	12.603	25.104	21.712
		-1.496	1.771	0.948	-0.714	1.031	0.743	-0.969	12.371	3.188
	Boot Liu	-0.521	1.088	0.955	-0.210	0.291	0.201	9.688	10.327	3.577
		1.052	1.579	1.178	2.090	2.788	1.845	10.859	3.332	5.927
	R Boot Wu	0.021	0.206	0.205	-0.193	0.352	0.294	-1.602	3.429	3.031
		-0.016	0.194	0.193	-0.028	0.141	0.138	1.832	2.081	0.988
	R Boot Liu	-0.154	0.284	0.239	-0.184	0.263	0.188	0.873	1.012	0.512
		-0.011	0.016	0.012	0.000	0.023	0.023	0.419	0.502	0.277
	RPC Boot Wu	-0.017	0.036	0.032	0.003	0.027	0.027	0.008	0.328	0.328
		0.005	0.007	0.003	-0.005	0.007	0.008	0.000	0.006	0.006
	RPC Boot Liu	0.003	0.007	0.006	-0.004	0.007	0.006	0.026	0.042	0.033
		0.017	0.017	0.005	-0.002	0.008	0.004	-0.008	0.010	0.006
β_3	Boot Wu	-2.894	5.058	4.148	-3.234	3.252	0.338	-6.539	15.868	14.459
		-0.356	1.795	1.760	0.662	1.382	1.213	-1.017	6.315	4.539
	Boot Liu	2.423	2.811	1.425	0.561	0.641	0.310	-9.899	10.947	4.673
		-0.897	1.897	1.672	-0.811	2.622	2.493	-1.483	4.651	6.138
	R Boot Wu	-0.102	0.456	0.445	-0.690	0.878	0.543	5.595	7.128	4.416
		-0.219	0.257	0.135	0.041	0.142	0.136	-0.808	1.071	0.704
	R Boot Liu	-0.476	0.598	0.362	-0.154	0.305	0.263	0.287	0.709	0.649
		-0.032	0.036	0.016	0.013	0.022	0.018	-0.466	0.694	0.515
	RPC Boot Wu	-0.048	0.058	0.033	0.000	0.040	0.040	-0.009	0.341	0.341
		0.002	0.012	0.014	-0.009	0.011	0.006	0.000	0.008	0.006
	RPC Boot Liu	0.000	0.011	0.011	-0.002	0.015	0.015	0.081	0.127	0.098
		0.000	0.014	0.012	0.000	0.006	0.006	0.000	0.006	0.008

Table 3: Bias, RMSE and standard error of n = 20 and n = 100 (bold) for 10% level of contaminated data based on non- robust wild bootstrap, robust wild bootstrap and robust PC with wild bootstrap from normal distribution with 3 regressor variables

Coef.	Method	$\rho = 0.2$			$\rho = 0.5$			$\rho = 0.99$		
		Bias	RMSE	SE	Bias	RMSE	SE	Bias	RMSE	SE
β_0	Boot Wu	-0.108	6.443	6.442	0.005	7.391	7.391	0.371	3.829	3.811
		-0.918	3.726	3.611	-2.457	5.072	4.436	-0.165	3.218	3.182
	Boot Liu	-0.286	2.761	2.746	1.222	5.614	5.479	-0.021	2.306	2.306
		-0.010	3.180	3.180	-2.498	4.464	3.699	0.226	3.186	3.210
	R Boot Wu	0.129	0.920	0.911	0.360	0.734	0.640	-0.186	0.914	0.895
		0.185	0.278	0.207	-0.127	0.242	0.205	0.069	0.203	0.190
	R Boot Liu	-0.273	0.565	0.494	-0.135	0.432	0.411	-0.141	0.348	0.318
		0.044	0.064	0.046	-0.007	0.061	0.061	-0.041	0.110	0.102

	RPC Boot Wu	0.044	0.059	0.039	0.010	0.037	0.036	-0.139	0.179	0.113
		0.000	0.002	0.003	0.002	0.003	0.005	0.001	0.003	0.005
	RPC Boot Liu	0.004	0.017	0.016	0.009	0.018	0.016	0.033	0.045	0.031
		0.000	0.003	0.002	-0.001	0.005	0.002	0.004	0.006	0.003
β_1	Boot Wu	-10.610	13.734	8.721	-8.738	12.318	8.682	-11.054	24.169	21.494
		-4.905	6.033	3.514	-2.619	6.013	5.413	5.375	17.782	15.338
	Boot Liu	-2.575	3.674	2.621	-1.740	7.489	7.284	-6.185	14.616	13.243
		2.662	4.283	3.355	1.863	4.542	4.142	9.939	16.252	14.745
	R Boot Wu	-0.196	1.003	0.984	0.343	0.769	0.688	-1.648	4.432	4.114
		0.153	0.237	0.181	0.021	0.186	0.185	-0.482	0.941	0.807
	R Boot Liu	-0.967	1.109	0.543	-0.487	0.629	0.398	-1.575	2.403	1.816
		0.027	0.052	0.044	-0.028	0.067	0.061	-0.253	0.925	0.890
	RPC Boot Wu	-0.056	0.063	0.029	-0.002	0.030	0.030	-0.004	0.058	0.058
		0.000	0.002	0.002	0.000	0.007	0.000	0.000	0.007	0.005
	RPC Boot Liu	-0.003	0.012	0.011	-0.011	0.015	0.011	0.002	0.010	0.010
		0.000	0.002	0.002	0.000	0.000	0.007	0.000	0.005	0.007
β_2	Boot Wu	0.804	6.590	6.540	5.583	9.816	8.074	2.145	21.392	21.284
		-3.716	4.635	2.770	-2.846	3.990	2.796	-14.500	19.110	15.458
	Boot Liu	-0.521	2.254	2.193	-0.457	3.508	3.478	3.369	11.785	11.293
		2.064	3.522	2.854	3.027	5.094	4.097	11.814	21.195	15.021
	R Boot Wu	-0.086	0.305	0.293	-0.291	0.430	0.317	1.020	4.462	4.344
		0.004	0.230	0.230	0.170	0.288	0.233	-0.017	0.745	0.745
	R Boot Liu	-0.440	0.644	0.470	-0.596	0.715	0.395	0.810	1.563	1.337
		-0.018	0.023	0.014	-0.020	0.032	0.025	0.498	0.622	0.373
	RPC Boot Wu	-0.012	0.039	0.037	0.003	0.028	0.028	-0.026	0.650	0.649
		0.000	0.001	0.002	0.006	0.006	0.005	0.001	0.002	0.004
	RPC Boot Liu	0.003	0.008	0.007	-0.004	0.007	0.006	-0.103	0.259	0.237
		-0.002	0.003	0.001	-0.003	0.006	0.002	0.000	0.004	0.002
β_3	Boot Wu	-4.986	8.461	6.836	-3.198	9.098	8.518	8.968	27.145	25.621
		-4.601	7.046	5.337	-1.229	4.178	3.993	5.883	26.766	22.404
	Boot Liu	-1.131	3.614	3.432	-4.586	7.745	6.241	3.528	15.144	14.727
		0.386	4.177	4.159	-0.538	4.010	3.974	-16.405	23.164	21.149
	R Boot Wu	0.000	0.739	0.739	-0.441	0.819	0.690	1.243	5.579	5.439
		-0.296	0.332	0.150	-0.139	0.261	0.221	0.492	0.884	0.734
	R Boot Liu	-0.802	1.052	0.682	-0.213	0.579	0.538	0.422	2.077	2.033
		-0.038	0.043	0.019	-0.003	0.026	0.026	-0.286	0.900	0.854
	RPC Boot Wu	-0.056	0.069	0.040	0.000	0.040	0.040	0.011	0.718	0.718
		0.002	0.002	0.003	0.000	0.003	0.008	0.000	0.001	0.004
	RPC Boot Liu	0.000	0.013	0.013	-0.002	0.015	0.015	-0.103	0.342	0.327
		0.000	0.003	0.001	0.000	0.008	0.003	0.000	0.004	0.001

Table 4: Bias, RMSE and standard error of n = 20 and n = 100 (bold) for 20% level of contaminated data based on non- robust wild bootstrap, robust wild bootstrap and robust PC with wild bootstrap from normal distribution with 3 regressor variables.

Coef.	Method	$\rho = 0.2$			$\rho = 0.5$			$\rho = 0.99$		
		Bias	**RMSE**	**SE**	**Bias**	**RMSE**	**SE**	**Bias**	**RMSE**	**SE**
β_0	Boot Wu	6.873	12.088	10.646	7.708	13.144	9.944	11.350	14.310	8.715
		-3.712	6.466	5.294	-2.583	5.542	4.903	-2.463	5.526	11.440
	Boot Liu	-1.212	10.003	8.940	-0.823	8.978	9.929	-0.022	9.384	9.384
		-3.234	5.919	4.958	-1.216	4.482	4.314	-3.260	11.702	4.463
	R Boot Wu	0.034	5.011	4.191	-0.299	4.202	5.011	-0.118	6.214	6.213
		0.007	0.363	0.363	-0.206	0.565	0.527	-0.059	0.280	0.274
	R Boot Liu	-1.190	3.756	4.542	-1.453	4.769	3.562	-0.738	3.423	3.343
		0.045	0.133	0.125	0.013	0.070	0.069	-0.108	0.165	0.125
	RPC Boot Wu	0.176	0.196	0.060	-0.088	0.107	0.088	-0.001	0.014	0.014
		0.001	0.002	0.003	0.000	0.002	0.004	0.000	0.002	0.004
	RPC Boot Liu	0.006	0.010	0.019	0.011	0.021	0.008	-0.003	0.014	0.013
		-0.003	0.004	0.002	-0.001	0.004	0.002	0.004	0.006	0.002
β_1	Boot Wu	-13.294	17.449	11.602	-15.857	19.648	11.303	-63.437	81.502	51.169
		-8.529	11.146	7.176	-3.545	6.324	5.237	8.037	37.355	29.713
	Boot Liu	-18.232	20.557	8.913	-18.822	20.825	9.496	-90.290	105.200	53.986
		1.322	5.980	5.832	4.619	6.801	4.991	26.763	30.780	26.060
	R Boot Wu	-0.652	5.173	4.505	-0.832	4.582	5.132	-4.633	16.667	16.010
		-0.442	0.633	0.453	0.782	1.146	0.838	0.280	1.528	1.502
	R Boot Liu	-1.435	2.452	2.358	-1.249	2.668	1.989	-1.815	10.839	10.686
		0.026	0.113	0.110	-0.014	0.070	0.069	-0.742	1.306	1.075
	RPC Boot Wu	0.132	0.145	0.047	-0.092	0.104	0.060	-0.001	0.008	0.008

		0.000	0.001	0.001	0.000	0.004	0.000	0.000	0.003	0.003
	RPC Boot Liu	-0.004	0.006	0.011	-0.012	0.016	0.005	-0.002	0.008	0.008
		0.000	0.001	0.001	0.000	0.000	0.004	0.005	0.006	0.003
β_2	Boot Wu	16.322	19.145	10.006	20.950	24.006	10.007	84.542	96.692	46.925
		-6.297	7.316	3.723	-4.383	5.575	3.446	-12.568	24.859	27.418
	Boot Liu	9.136	11.869	7.577	9.071	11.738	7.577	75.268	86.609	42.845
		-3.857	6.659	5.429	-2.385	4.163	3.412	0.163	30.161	24.859
	R Boot Wu	-0.140	2.234	2.230	0.027	2.587	2.230	-1.918	10.357	10.178
		0.175	0.411	0.372	-0.277	0.716	0.660	0.006	1.135	1.135
	R Boot Liu	-0.497	1.359	1.265	-0.066	1.942	1.265	1.128	9.452	9.385
		-0.016	0.026	0.021	-0.038	0.046	0.026	0.432	0.739	0.599
	RPC Boot Wu	-0.150	0.167	0.073	0.049	0.075	0.073	0.011	0.082	0.081
		0.000	0.001	0.003	0.001	0.002	0.005	0.001	0.002	0.002
	RPC Boot Liu	-0.002	0.004	0.004	-0.026	0.030	0.004	0.006	0.092	0.092
		-0.002	0.003	0.001	0.000	0.005	0.001	-0.006	0.007	0.002
β_3	Boot Wu	-3.267	10.682	10.566	-4.082	11.327	10.170	-18.994	57.984	54.785
		-10.677	12.606	6.702	-3.867	6.510	5.237	3.108	38.096	23.178
	Boot Liu	-6.466	13.538	12.274	-6.853	14.057	11.893	12.952	55.163	53.621
		-6.868	8.456	4.932	-4.494	6.150	4.198	-25.217	23.385	28.556
	R Boot Wu	-1.443	3.624	4.150	-1.714	4.490	3.325	8.831	17.648	15.280
		-0.218	0.387	0.319	0.260	0.602	0.543	-0.204	1.096	1.077
	R Boot Liu	-1.064	2.307	2.467	-0.701	2.565	2.047	0.286	11.390	11.386
		-0.037	0.047	0.029	-0.021	0.040	0.034	0.262	1.128	1.097
	RPC Boot Wu	0.148	0.176	0.138	-0.105	0.173	0.095	0.014	0.175	0.175
		0.001	0.001	0.002	0.000	0.004	0.005	0.000	0.002	0.001
	RPC Boot Liu	-0.003	0.006	0.017	0.011	0.020	0.005	-0.040	0.085	0.075
		-0.004	0.004	0.001	-0.001	0.005	0.004	0.004	0.004	0.002

6. Conclusion

The presence of multicollinearity, outliers and heteroscedasticity error variance required a comprehensive and detailed investigation not only for usual regression analysis but also for principal component and wild bootstrap procedures. In the present paper, we have introduced a new wild bootstrap procedure based on Wu and Liu called RPC Boot Wu and RPC Boot Liu for regression analysis that will provide the enhancement protection against data with multiple problems in order to get numerically stable results. We present a numerical example and simulation studies to evaluate the performance of our proposed methods. The results obtained from the real data and simulated data disclosed that the RPC Boot Wu and RPC Boot Liu are a better choice when compared with the Boot Wu, Boot Liu, R Boot Wu and R Boot Liu, particularly when the data contain multicollinearity, heteroscedasticity and outliers. The performance of our proposed robust PC with wild bootstrap methods is a very robust alternative to other wild bootstrap and robust wild bootstrap procedures.

7. Acknowledgments

We acknowledge the financial support from Universiti Teknologi Malaysia for the research Universiti Grant (Q.J130000.2526.06H68) and the Ministry of Higher Education (MOHE) of Malaysia.

8. References

1. Habshah Midi, Md. Sohel Rana, Rahmatullah Imon AHM. The Performance of Robust Weighted Least Squares in the Presence of Outliers and Heteroscedastic Errors. WSEAS Transition of Mathematics. 2009;8(7):351-361.
2. Wu CFJ. Jackknife, bootstrap and other resampling methods in regression analysis. The Annals of Statistics. 1986;14(4):1261-1350.
3. Liu RY, Bootstrap procedures under some non-i.i.d. Models. The Annals of Statistics. 1988;16(4):1696-1708.
4. Meriam Zahari S, Norazan Mohamed Ramli, Balkiah Mokhtar. Bootstrap Parameter in Ridge Regression with Multicollinearity and Multiple Outliers. Journal of Environmental Biological Science. 2014;4(7S):150-156.
5. Massy WF. Principal components regression in exploratory statistical research. Journal of the American Statistical Association. 1965;60(309):234-256.
6. Rousseeuw PJ, Leroy A. Robust Regression and Outlier Detection. Wiley, New York; c1987.
7. Hongtu Z, Joseph GI, Bradley SP. A Statistical Analysis of Brain Morphology Using Wild Bootstrapping. IEEE Trans. med. imagine. 2007;26(7):954-966.
8. Xingdong He FX, Hu J. Wild Bootstrap for Quantile Regression. Biometrika. 2011;98(4):995-999.
9. Alamgir S, Ali A. Split Sample Bootstrap Method. World Applied Science Journal. 2013;21(7):983-993.
10. Rana S, Habshah Midi, Imon AHMR. Robust Wild Bootstrap for Stabilizing the Variance of Parameter Estimates in Heteroscedastic Regression Models in the Presence of Outliers; c2012, p14. Article ID 730328
11. Yohai VJ. High breakdown-point and high efficiency robust estimates for Regression, The Annals of Statistics. 1987;15(2):642-656.
12. Cribari-Neto F. Asymptotic inference under heteroscedasticity of unknown form, Computational Statistics & Data Analysis. 2004;45(2):215-233.
13. Rasheed BA, Robiah A, Ehsan S, Kafi D. Application of Robust Wild Bootstrap Estimation of Linear Model in Econometric. International Journal of Applied mathematics and Statistics. 2015;53(1):82-101.
14. Coultas DB, Stidley CA, Samet JM. Cigarette Yields of Tar and Nicotine and Markers of Exposure to Tobacco *Smoke*, American Review of Respiratory Disease. 1993;148(2):435-440.