Proceedings of Science and Mathematics
Faculty of Science, Universiti Teknologi Malaysia
Vol. 1, 2020, page 75-86

# A New Heteroscedasticity Model in the Presence of Outliers and Multicollinearity using Hampel and Andrews Sin Psi Function

**Bello Abdulkadir Rasheed[a], Robiah Adnan[b]*, Salisu Lukunti[a]**
[a]Department of Mathematics and Statistics, Federal Polytechnic Bauchi, Bauchi State, Nigeria
[b]Department of Mathematical Science, Faculty of Science, UTM, 81310 UTM Skudai, Johor, Malaysia
*Corresponding author: robiaha@utm.my

**Abstract**
The presence of heteroscedasticity, multicollinearity and outliers are classical problems of data within the linear regression framework. This research is a proposal of new methods which can be a potential candidate for weighted robust wild bootstrap regression as well as the multicollinearity robust regression model with outliers' pattern based on Latin root. This proposal arises as a logical combination of principles used in the Latin root, wild bootstrap sampling procedure of Wu and Liu. The weighted robust GM-estimator of Krasker and Welsch (1982) with initial MM-estimator of Yohai (1987) and S-estimator of Rousseeuw and Yohai (1984) together with two different weighting procedures of Hampel's and Andrews sin weighted function are considered in the analysis. This paper investigates the nonresistance of weighted robust wild bootstrap (WRWBoot) regression and our proposed method for resistance to multicollinearity, outliers and heteroscedasticity error variance. The use of modified weighted robust wild bootstrap methods (WRWBoot) based on Latin root with multicollinearity and outlier diagnostic method yields more reliable trend estimations. From numerical example and simulation study, the resulting of the modified weighted robust wild bootstrap methods based on Latin root with multicollinearity and outlier diagnostic method (WRWBoot) is efficient than other estimators, using Standard Error (SE) and the Root Mean Squared Error criterion for numerical example and simulation study respectively for many combinations of error distribution and degree of multicollinearity.
**Keywords:** multicollinearity; outliers; Latin root; robust GM-estimator and wild bootstrap

## Introduction

In regression analysis, the ordinary least square is widely used to estimate the parameter of the models mostly because of tradition for optimal properties and ease of computation. Unfortunately, the mathematical elegance that makes the estimator so popular relayed on a number of fairly strong and many times 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 becomes valid depend largely to the extent in which these model's assumptions are met. However, if these assumptions are violated, the ordinary least square will no longer produced 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 remain unbiased. But 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. On the other hand, if there is no exact linear relationship between the explanatory variables, this is called assumption of multicollinearity. In the present of multicollinearity, the OLS estimator will results 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 desterilized the parameter estimation in the model by inflating the test statistics which resulted 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 Wu (1986) and Liu (1988). They proposed the wild bootstrap technique which gives a better performance for the parameter estimates of the regression coefficients when the model exhibits both the homoscedasticity and heteroscedasticity models. This type of weighted bootstraps 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. Wu (1986) and Liu (1988) 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 (Zahari et al., 2014).

To handle the multicollinearity problems, latent root regression was introduced 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 Hampel and Andrews sin psi function to down weight the effect of outliers. The robust estimator used in this research is GM-estimator of Krasker and Welsch (1982) with initial MM-estimator of Yohai (1987) and S-estimator of Rousseeuw and Yohai (1984) together with two different weighting procedures of Hampel's and Andrews sin weighted function are considered in the analysis. We choose this weighted function also to improve the asymptotic relative efficiency of our estimator the GM-estimator.

Several attempts have been made to use the procedure of Wu (1986) and Liu (1988) wild bootstrap techniques to remedy the problem of heteroscedasticity error variance. Zhu et al. (2007) 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 the. 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 Feng et al. (2011).

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, the idea of this bootstrap method is to protect against excessive number of outliers and ensures efficient results Alamgir and Ali (2013). Rana et al. (2012) proposed the robust wild bootstrap based on Wu (1986) and Liu (1988). They disclosed that the problem of classical bootstrap is that the proportion of outliers involve 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 Yohai (1987). 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 latent root regression with wild bootstrap techniques of Wu (1986) and Liu (1988). 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 not resistance to multicollinearity and outliers. We discussed the methodology of this research in section 2. In section 3, we introduce the newly proposed method and its performance are presented. Section 4 will contain the detailed of conclusion of the study.

**Methodology**

A simulation study was design to assess the performance of wild bootstrap Wu (1986) and Liu (1988) and the robust wild bootstrap of Wu (1986) and Liu (1988) with the proposed robust latent root with wild bootstrap of Wu (1986) and Liu (1988). We generate the covariance of $x_1, x_2 \; and \; x_3$ using the multiple linear regression model based on the combination of different regression condition. Here, we follow a similar procedure used by Rana et al. (2012), The considered design for this experiment involved a regression model with intercept and covariance values. Suppose we consider the following linear model. Where

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

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 is drawn from normal distribution with mean zero and variance 1.When no outliers was 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 $\varepsilon_t$'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 $\rho$ between the regressor variables. At the same time, the performance of the estimators were observed by increase percentage of outliers and the considered percentage of outliers are 0%, 10%, and 20% respectively. We form the heteroscedasticity generating procedure following Cribari-Neto (2004), Rana et al. (2012) and Rasheed et al. (2015) effort, where

$$\sigma_i^2 = \exp(2.6 x_{1i}) \qquad (4)$$

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

$$y_i = \beta_0 + \beta_1 x1 + \beta_2 \, x2 + \beta_3 \, x3 + \sigma i \varepsilon_{t(Cont.)} \qquad (5)$$

We first considered the sample size of n = 20 observations and apply the principal component analysis to estimate the component that contain 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 Rana et al. 2012 who utilized the replication of covariate values to create large samples. For each simulated data set, with different sample size we fit the linear regression model.

**The Latent Root Regression (LRR)**

The latent root regression utilizes the latent roots and latent vectors of the correlation matrix of the dependent and independent variables, denoted as A. The latent roots, $\lambda_j$ and latent vectors,j $\gamma_j$ of A'A are defined by:

$$|A^T A\_ - \lambda_j I| = 0 \; and \; (A^T A\_ - \lambda_j I)\gamma_j = 0$$
$$j = 0,1,...,K$$

Analysis of these latent roots and latent vectors enables one to:

- Identify near singularities in X
- Determine whether the near singularities have predictive value
- Obtain the modified least squares estimates of parameters which adjust for non-predictive near singularities

The OLS estimator in (2) can also be expressed in terms of these latent roots and latent vectors:

$$\hat{\beta} = \eta \sum_{j}^{k} \alpha_j \gamma_j^0$$ where $$\eta^2 = \sum_{i}^{k} (Y_i - \bar{Y})^2$$ and $$\alpha_j = \frac{\gamma_{oj} \lambda_j^{-1}}{\sum_{i}^{k} \gamma oi^2 \lambda_i^{-1}}$$ and $$\gamma_j^{oi} = (\gamma_{1j} \; , \gamma_{2j} \; , ..., \gamma_{kj}$$

and the residual sum of squares given by:

$$SSE = {}^2 \eta \left( \sum_{j=1}^{k} \frac{\gamma_{oj}^2}{\lambda_j} \right)^{-1}$$

Rana et al., 2012 suggested small latent roots and latent vectors in which $\lambda_j \leq 0.3$ and $|\gamma_{oj}| \leq 0.1$| which indicates the presence of non-predictive singularities. But later, they discovered that a tighter cut-off value of $\lambda_j \leq 0.2$ and $|\gamma_{oj}| \leq 0.1$ could improve the analysis.

Suppose now that the latent vectors $\gamma_o, \gamma_1 \quad , \gamma_{p-1} \leq 0.1$ correspond to non-predictive near singularities. The non-predictive multicollinearity is eliminated and only the predictive are retained. The above OLS estimator can be adjusted by setting $\alpha_0 = \alpha_1 = ... = \alpha_{p-1} = 0.$ Then the modified least squares coefficients are:

$$\hat{\beta}_{LRR} = \eta \sum_{j=1}^{k} \alpha_j \gamma_j^0$$ where $$\alpha_j = \frac{\gamma_{oj} \lambda_j^{-1}}{\sum_{i}^{k} \gamma oi^2 \lambda_i^{-1}} \quad j = p, p+1, ...k$$

with residual sum of squares,

$$SSE_{LRR} = \eta^2 \left( \sum_{j=1}^{k} \frac{\gamma_{oj}^2}{\lambda_j} \right)^{-1}$$

If all of the principal components for the correlation matrix of the dependent and independent variables are predictive, then none of the j a 's equal zero, the latent root estimator and the OLS estimator will be identical. It is well-known that the variance covariance matrix for the OLS estimator is given by $\sigma^2 (X'X)^{-1}$ and its trace (sum of diagonals) represents it unweight mean squared error:

$$MSE(\hat{\beta}) = \sigma^2 tr(X'X)^{-1}$$

or in terms of latent roots of X'X, $MSE(\hat{\beta}) = \sigma^2 \sum_{J=1}^{P} e_j^{-1}$

$e_j$ are the latent root of X'X and are ordered such that $e_1 \le e_2 \le \ldots \le e_k$. For a near multicollinearity situation, approaches 0 and (6) implies that $MSE(\hat{\beta})$ approaches infinity, that is ˆb is subjected to very large variance. This inflation causes the estimation becomes less accurate and less precise, thus unstable.

**Wild Bootstrap Based on Wu's**

This bootstrap procedure has been suggested by Wu (1986) and Beran (1986) 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). 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 $\hat{y}_i = f(x_i, \hat{\beta})$ (6)

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 equals to zero Wu (1986).

Step 4. Obtained a the random sample of t* from $a^R$ can be used to multiply it with $\hat{\varepsilon}_i (1 - h_{ii})^{-1}$ to obtained a $y_t^{*b}$ and $h_{ii} = \mathbf{x_i^T (x^T x)^{-1} x_i}$ is the i-th leverage where

$$y_t^{*b} = f(x_i, \hat{\beta}_{ols}) + t_i^* \hat{\varepsilon}_i (1 - h_{ii})^{-1} \quad (7)$$

$y_t^{*b}$ is the new bootstrap response variable that can be used to obtained 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 the procedures of Step 3 and Step 4 for *k* times to get $\hat{\boldsymbol{\beta}}^{*1},...,\hat{\boldsymbol{\beta}}^{*k}$ 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 referring as Wu's bootstrap sample and denote BootWu.
Following the idea of wild bootstrap of Wu (1986) and Beran (1986). Liu (1988) provide a suggestion by slightly modifying the procedure of generating the t* value. The t* is randomly selected from auxiliary distribution that has third central moment equal to one, In addition with zero mean and unit variance. She has shown that when this is the case, Wu.'s shares the usual second order asymptotic properties of the classical bootstrap. put differently, the addition of the restriction that the third central moment equal to one and Such kind of 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 (1988) 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* put mass only on two-point distribution, then

Step1. $t^{*} = S_i - E(S_i), i = 1, 2, ..., \text{n, and } S_1, S_2, ..., S_n$ 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 $\alpha$ =2 and $\beta$ = 4.

Step2. $t_i^{*} = N_i M_i - E(N_i)E(M_i)\ i = 1, 2, ..., n$ where $N_1, N_2, ..., N_n$ are independently and identically distributed normal distribution with mean $(1/2)\ (\sqrt{17/6)} + \sqrt{1/6}$ and variance $1/2$. $M_1, M_2, ..., M_n$ are also i.i.d. normally distribution 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 the $V_{i,j}$'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 (1988). Both procedures will produce third central moment equals to one. Rana and Midi (2012) 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 Rana Rano [13], this research will make used of the second method for generating its bootstrap sample of $t_i^{*}$. The bootstrap procedure is called liu bootstrap BootLiu.

**Robust Wild Bootstrap MM-Estimator**

They consider the idea of the classical bootstrap procedure based on Wu's (1986), Liu (1988) and Beran (1986). Another alternative of modified wild bootstrap technique which is more robust was introduced to remedy the problem of heteroscedasticity and outliers Rana and Midi (2012). This bootstrap method was based on MM- estimator procedure. The quantity of t* is obtained from equation (6) that is a robust normalized residuals 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 $\hat{\boldsymbol{\beta}}_{MM}$.

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, $\varepsilon_i^{MM}$ where the weight will equal to,

$$w_{ii} = \begin{cases} 1 & \text{if} \quad \left|\varepsilon_i^{MM}\right| / \sigma_{MM} \le c \\ c/(\left|\varepsilon_i^{MM}\right| / \sigma_{MM}) & \text{if} \quad \left|\varepsilon_i^{MM}\right| / \sigma_{MM} > c \end{cases} \qquad (8)$$

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 $\varepsilon_i^{MM}$ 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

$$y_i^* = x\hat{\beta}_{MM} + \frac{t_i^* \varepsilon_i^{WMM}}{(1-h_{ii})} \qquad (9)$$

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

Step5. Apply the OLS estimation procedure on the bootstrap sample of $(y_i^*, X)$. This estimate is denoted by

$$\mathbf{R}_{\hat{\boldsymbol{\beta}}^*} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}^* \qquad (10)$$

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's and Liu i.e (RBootWu) and (RBootLiu).

**Robust Latent with Wild Bootstrap**

Robust latent root regression incorporates resistance in the ordinary latent root regression. This is done by imposing weight to the correlation matrix of the dependent and independent variables, A'A. The pair wise Pearson correlation coefficient for the two variables is defined as:

$$r = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})(X_i - \bar{X})}{\sqrt{\sum_{i=1}^{n}(Y_i - \vec{Y})^2 \sum (X_i - \bar{X})^2}} \quad \text{where} \quad \bar{Y} = \frac{\sum_{i=1}^{n} Y_i}{n} \quad \text{and} \quad \bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

The correlation coefficient, r in given above is based on sample means $\bar{X}$ and $\bar{Y}$, respectively, which are known to be very sensitive to the presence of outliers. As an alternative, a robust location estimates which are less affected by outliers are proposed to replace $\bar{X}$ and $\bar{Y}$, in r. Following the idea of Mokhtar (1986), we propose using the weighted correlation coefficient between the dependent and the independent variables. We may use the weight from the final step of any robust estimators, but in this study the weight is confined to the final step of the robust M-estimation. The pair wise correlation coefficient in Eq. 9 is modified to obtain a weighted pair wise correlation coefficient, as follows.

$$r = \frac{\sum_{i=1}^{n} w_i (Y_i - \bar{Y}_W)(X_i - \bar{X}_W)}{\sqrt{\sum_{i=1}^{n} w_i (Y_i - \vec{Y}_W)^2 \sum w_i (X_i - \bar{X}_W)^2}}$$

where

$$\bar{Y}_W = \frac{\sum_{i=1}^{n} w_i Y_i}{\sum_{i=1}^{n} w_i} \quad \text{and} \quad \bar{X}_W = \frac{\sum_{i=1}^{n} w_i \bar{X}_i}{\sum_{i=1}^{n} w_i}$$

In this study, we have chosen the Hampel and Andrews sin psi function in the M estimation technique. The robust weighted correlation matrix for dependent and independent can be formulated. Based on this weighted correlation matrix, the latent roots and the latent vectors are computed, and the latent root routines are then incorporated to estimate the parameters of the model. We call this method the Latent Root- M based Regression (LRMB) because here we have employed the weight of the M-estimator in the weighted correlation matrix. We then obtained the estimate the weighted residuals of GM-estimator with MM-estimator as initial estimator of the GM-estimator using the Latin variables.

However, other procedure of Rboot for both Wu's and Liu will remain the same. In addition, 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 for a situation when outliers, multicollinearity and heteroscedasticity error variance are presence in the data. This modified bootstrap method can also be used to obtain the standard error which is asymptotically corrected under heteroscedasticity of unknown form. The bootstrap procedure obtained from Wu's procedure is called Hampel Robust Latent root with Wild Bootstrap GM -estimator based on Wu's or HRWLTBootWuGM and bootstrap procedure obtained from Liu's procedure is called Hampel Robust Latent root with Wild Bootstrap GM -estimator based on Liu's or HRWLRBootLiuGM. This procedure is applied Andrews Psi function and the procedure in this case is called Andrews Robust Latent root with Wild Bootstrap GM -estimator based on Wu's or ARWLTBootWuGM and bootstrap procedure obtained from Liu's procedure is called Andrews Robust Latent root with Wild Bootstrap GM -estimator based on Liu's or ARWLRBootLiuGM.

**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 GM-estimator 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 HRWLRBootWuGM, ARWLTBootWuGM, ARWLTBootLiuGM and HWLRBootLiuGM estimate. The numerical calculation of BootWu, BootLiu and the RBootWu and RBootLiu is to be Perform the same procedure.

**Example using Real Data Sets**

This section will discuss the application of the RWLRBootWu. and RLRBootLiuGM methods on real data by considering the numerical example that will show the advantages of the proposed method with respect to the other estimators, BootWu, BootLiu, RBootWu and RBootLiu estimator in the presence of outlier's multicollinearity and heteroscedasticity error variance. The cigarette data is taken from Coultas et al. (1993). 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 contain any outliers or not using the LTS residuals. It was discovered that five observations (about 20% of the sample of size 25) identifies 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 LRwith wild bootstrap methods were then applied to the data. The results obtain are based on 1000 bootstrap replicate and are presented in Table 1. The standard errors, bias and RMSE of the parameter estimates from wild bootstrap, robust wild bootstrap and robust latent root with wild bootstrap methods for both Wu's and Liu are presented in Table 1. Based on the results, it is interesting to observed both the standard error bias and RMSE of the wild bootstrap method tend to be larger followed by robust wild bootstrap. The HRWLRBootWu and HRWLRBootLiu methods has the smallest standard errors with HRWLRBootLiu as the best.

**Table 1:** The Parameter estimate, Standard error, Bias and RMSE of non-robust wild bootstrap, robust wild bootstrap and robust latent root with wild bootstrap of 25 collection of cigarette data

| Par. Estm.\ | BootWu | BootLiu | RBootWu | RBootLiu | HRWLR BootWu. | HRWLR BootLiu. | ARWLR BootWu. | ARWLR BootLiu. |
|---|---|---|---|---|---|---|---|---|
| Estimate | 11.87 | 11.874 | 12.952 | 12.952 | 3.1633 | 2.602 | 3.1863 | 2.625 |
| S.E | 3.6877 | 3.5129 | 1.1587 | 0.8014 | 0.0077 | 0.0066 | 0.0307 | 0.0296 |
| Bias | -4.1698 | 1.9157 | -1.1885 | -0.3058 | -0.0045 | -0.0013 | 0.0185 | 0.0217 |
| RMSE | 5.5665 | 4.0013 | 1.6598 | 0.8578 | 0.0089 | 0.0067 | 0.0319 | 0.0297 |
| Estimate | -15.933 | -15.97 | -17.075 | -16.923 | 0.7602 | 0.9581 | 0.7832 | 0.9811 |
| S.E | 0.2665 | 0.2469 | 0.1477 | 0.1290 | 0.0323 | 0.0322 | 0.0553 | 0.0552 |
| Bias | 0.1418 | -0.216 | 0.1847 | 0.0076 | -0.0396 | 0.0040 | -0.0166 | 0.027 |
| RMSE | 0.3019 | 0.3281 | 0.2365 | 0.1292 | 0.0511 | 0.0324 | 0.0741 | 0.0554 |
| Estimate | -10.786 | -10.81 | -4.1841 | -4.1112 | 0.6411 | -2.3441 | 0.6641 | -2.3211 |
| S.E | 4.3248 | 4.0302 | 1.7351 | 1.3688 | 0.0447 | 0.0390 | 0.0677 | 0.062 |
| Bias | -0.7468 | 4.7028 | -2.4218 | 0.2219 | -0.0090 | 0.0170 | 0.014 | 0.04 |
| RMSE | 4.3888 | 6.1935 | 2.9792 | 1.3867 | 0.0456 | 0.0425 | 0.0686 | 0.0655 |
| Estimate | 11.7273 | 11.791 | 47.171 | 47.764 | -0.8764 | 0.3854 | -0.8534 | 0.4084 |
| S.E | 4.1282 | 4.022 | 1.2220 | 0.8707 | 0.2360 | 0.2251 | 0.259 | 0.2481 |
| Bias | 3.1917 | -3.5044 | 1.4908 | -0.0539 | 0.0297 | -0.0366 | 0.0527 | -0.0136 |
| RMSE | 5.2181 | 5.3345 | 1.9277 | 0.8723 | 0.2379 | 0.2281 | 0.2609 | 0.2511 |

This cannot evidence up as our conclusion yet, only by investigating the results obtain from real data, but we can make a reasonable interpretation that the robust wild bootstrap and classical wild bootstrap are affected by multicollinearity and outliers.

**Examples using Simulated Data Sets**

The example of real data sets obtains in section 5 have shown that the RWLRBootWu. and RWLRBootLiu. 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 HRWLRBootWu, ARWLRBootWu, HRWLRBootLiu and ARWLRBootLiu methods by performing a simulation using a multiple linear regression model of three regressor variables. However, Table 2 presents simulation results of the bias, RMSE and standard error of the parameter estimates obtain from different degree of multicollinearity and percentage of outliers. As we can witness from the tables, the performance of BootWu and BootLiu estimator is poor since the standard error is large when compared with the BootWu, RBootLiu, RWLRBootWu. and

RWLRBootLiu at 10% level of contamination. The effect become very serious as the percentage of outliers is increases to 20%. The RBootWu and RBootLiu estimator without latent root techniques shows the worst performance since the standard error is larger than the proposed methods. On the other hand, incorporation of the Latent root techniques reduces the standard error of values of the HRWLRBootWu, ARWLRBootWu, HRWLRBootLiu and ARWLRBootLiu. estimators. It is worst to mention when the sample size, percentage of outliers and the degree of multicollinearity is inreases to a sample size n = 100, both the BootWu and BootLiu, RBootWu and RBootLiu estimator shows 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 propose methods 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 is increases to 20% and the degree of multicollinearity is 0.99, it is reported that the HRWLRBootWu,

**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 latent root 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** |
| $\beta_0$ | BootWu | -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** |
| | BootLiu | 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** |
| | RBootWu | -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** |
| | RBootLiu | -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** |
| | RWLRBootWu. | 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** |
| | RWLRBootLiu. | 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** |
| $\beta_1$ | BootWu | -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** |
| | BootLiu | 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** |
| | RBootWu | -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** |
| | RBootLiu | -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** |
| | HRWLRBootWu. | -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** |
| | HRWLRBootLi | -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** |
| | ARWLRBootWu. | 0.360 | 0.056 | 0.109 | 0.105 | 0.076 | 0.100 | 0.100 | 0.081 | 0.401 |
| | | **0.038** | **0.078** | **0.080** | **0.076** | **0.068** | **0.080** | **0.081** | **0.073** | **0.079** |
| | ARWLRBootLiu. | 0.065 | 0.076 | 0.080 | 0.079 | 0.069 | 0.080 | 0.079 | 0.099 | 0.115 |
| | | **0.038** | **0.090** | **0.090** | **0.078** | **0.071** | **0.081** | **0.077** | **0.065** | **0.083** |
| $\beta_2$ | BootWu | 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** |
| | BootLiu | -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** |
| | RBootWu | 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** |
| | RBootLiu | -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** |
| | HRWLRBootWu. | -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** |
| | HWLRBootLiu. | 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** |
| | ARWLRBootWu. | 0.051 | 0.087 | 0.254 | 0.274 | 0.138 | -0.010 | 0.185 | 0.166 | 0.077 |
| | | **0.039** | **0.077** | **0.079** | **0.080** | **0.081** | **0.078** | **0.080** | **0.082** | **0.078** |
| | ARWLRBootLiu | 0.051 | 0.086 | 0.084 | 0.088 | 0.097 | 0.089 | 0.099 | 0.086 | 0.075 |
| | | **0.043** | **0.075** | **0.075** | **0.082** | **0.080** | **0.077** | **0.082** | **0.080** | **0.082** |
| $\beta_3$ | BootWu | -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** |
| | BootLiu | 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** |
| | RBootWu | -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** |
| | RBootLiu | -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** |
| | HRWLRBootWu. | -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** |
| | HRWLRBootLi | 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** |
| | ARWLRBootWu. | 0.043 | 0.025 | 0.127 | 0.098 | 0.071 | 0.103 | 0.103 | 0.075 | 0.084 |
| | | **0.032** | **0.073** | **0.076** | **0.076** | **0.073** | **0.074** | **0.075** | **0.073** | **0.075** |
| | ARWLRBootL | 0.036 | 0.070 | 0.084 | 0.084 | 0.062 | 0.088 | 0.084 | 0.071 | 0.078 |
| | | **0.034** | **0.067** | **0.079** | **0.076** | **0.075** | **0.076** | **0.074** | **0.073** | **0.073** |

ARWLRBootWu, HRWLRBootLiu and ARWLRBootLiu estimators becomes most 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 HRWLRBootWu, ARWLRBootWu, HRWLRBootLiu and ARWLRBootLiu are the most robust and resistance toward the presence of multicollinearity and multiple outliers.

## Conclusion

The presence of multicollinearity, outliers and heteroscedasticity error variance required a comprehensive and details 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 HRWLRBootWu, ARWLRBootWu, HRWLRBootLiu and ARWLRBootLiu, for the 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 obtain from the real data and simulated data disclosed that the HRWLRBootWu, ARWLRBootWu, HRWLRBootLiu and ARWLRBootLiu are better choice when compared with the BootWu, BootLiu, RBootWu and RBootLiu particularly, when the data contain multicollinearity, heteroscedasticity and outliers. The performance of our proposed robust latent root with wild bootstrap methods using the weighted function of Hampel and Andrews function are robust alternative to other wild bootstrap and robust wild bootstrap procedures.

## Acknowledgement

We and Federal Polytechnic Bauchi of Nigeria wish to acknowledge the financial support from the TETFund for the sponsorship of our conference attendance.

## References

Alamgir, S., Ali, A. (2013). Split sample bootstrap method. *World Applied Sciences Journal*, *21*(7), 983-993.

Beran, R. (1988). Prepivoting test statistics: a bootstrap view of asymptotic refinements. *Journal of the American Statistical Association*, *83*(403), 687-697.

Coultas, D. B., Stidley, C. A., & Samet, J. M. (1993). Cigarette yields of tar and nicotine and markers of exposure to tobacco smoke. *American Journal of Respiratory and Critical Care Medicine*, *148*(2), 435-440.

Cribari-Neto, F. (2004). Asymptotic inference under heteroskedasticity of unknown form. *Computational Statistics & Data Analysis*, *45*(2), 215-233.

Feng, X., He, X., Hu, J. (2011). Wild bootstrap for quantile regression. *Biometrika*, *98*(4), 995-999.

Krasker, W. S., Welsch, R. E. (1982). Efficient bounded-influence regression estimation. *Journal of the American Statistical Association*, *77*(379), 595-604.

Liu, R. Y. (1988). Bootstrap procedures under some non-iid models. *The annals of statistics*, *16*(4), 1696-1708.

Rana, S., Midi, H., Imon, A. R. (2012). Robust wild bootstrap for stabilizing the variance of parameter estimates in heteroscedastic regression models in the presence of outliers. *Mathematical Problems in Engineering*, *2012*.

Rasheed, B. A., Adnan, R., Saffari, S. E., & Pati, K. D. (2015). Application of Robust Wild Bootstrap Estimation of Linear Model in Econometric. *International Journal of Applied mathematics and Statistics*, *53*(1), 82-101.

Rousseeuw, P., Yohai, V. (1984). Robust regression by means of S-estimators. In *Robust and nonlinear time series analysis* (pp. 256-272). Springer, New York, NY.

Wu, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *The Annals of Statistics*, *14*(4), 1261-1295.

Yohai, V. J. (1987). High breakdown-point and high efficiency robust estimates for regression. *The Annals of Statistics*, 642-656.

Zahari, S. M., Ramli, N. M., Mokhtar, B. (2014). Bootstrapped parameter estimation in ridge regression with multicollinearity and multiple outliers. *Journal of Applied Environmental and Biological Sciences*, *4*, 150-156.

Zhu, H., Ibrahim, J. G., Tang, N., Rowe, D. B., Hao, X., Bansal, R., Peterson, B. S. (2007). A statistical analysis of brain morphology using wild bootstrapping. *IEEE Transactions on Medical Imaging*, *26*(7), 954-966.