Performance of Robust Wild Bootstrap Estimation of Linear Model in the Presence of Outliers and Heteroscedasticity Errors

R. Adnan[,1,a], B. A. Rasheed[1,b], S. E. Saffari[2,c] and K. D. Pati[1,d]*

[1]Department of mathematics, Faculty of Science, UTM, 81310 UTM Skudai, Johor, Malaysia
[2]Centre of Education, Sabzevar University of Medical Sciences, Sabzevar, Iran.
[a,*]*robiah@utm.my,* [b]*arasheedbello@yahoo.com,* [c]*ehsanreiki@yahoo.com,* [d]*kafi_dano@yahoo.com*

Abstract- *Bootstrap techniques are widely used today in many other fields such as economics, Business Administration, Physics, Engineering, Chemistry, Meteorological, Biological Sciences and Medicine. This paper is concerned with the estimation of linear regression model parameters in the presence of heteroscedasticity using wild bootstrap approaches of Wu and Liu. The empirical evidence has shown that these techniques are effective in the presence of heteroscedasticity. However, when there are outliers in the data, this method is no longer effective. To overcome this situation, this paper proposed robust wild bootstrap estimation methods where heteroscedasticity and outliers occur simultaneously. The proposed method is based on the Tukey-redesceding M-estimator which incorporate the LTS and LMS estimator, robust scale and location, and the wild bootstrap sampling procedures of Liu and Wu. Its performance is compared with other existing robust wild bootstrap estimator of MM-estimator using real data and simulation study. The results obtained from this study disclosed that the proposed methods offer a substantial improvement over the existing techniques and proved to be a good alternative estimator. Copyright © 2015 Penerbit Akademia Baru - All rights reserved.*

Keywords: Robust Estimation, Wild Bootstrap, Bias, Standard error and RMS.

1.0 INTRODUCTION:

Wild bootstrap method was first proposed by [1] which gives a better performance for homoscedastic and heteroscedastic models. However, a better alternative estimation method is introduced by [2-3] following the idea of [1] to estimate the regression model parameters. The most common bootstrap methods are the residuals bootstrap and the paired bootstrap which are defined in [4], and some of their asymptotic properties can be found in [5-7] among others. For bootstrap method, [8] proposed a bootstrap procedure based on random weight on the loss functions, [9] established a modified form of the residuals bootstrap, and [10] considered the validity of paired bootstrap techniques. [11] proposed a modified weighted bootstrap estimation method based on LTS. To account for heteroscedasticity [1-3] proposed the wild bootstrap techniques by randomly weighting the residuals. However, different attempts have been made to use the procedure of [1-2] wild bootstrap techniques to remedy the problem of heteroscedasticity error variance. Others including [12], [13-15] have considered the properties of wild bootstrap, but the existing theories of wild bootstrap are all based on ordinary least squares (OLS) method they can be seriously affected in the presence

of outliers. Recently [16] showed that the result produced from classical wild bootstrap is misleading in the presence of outliers. They proposed to replace the Wild bootstrap of OLS with a high-breakdown estimator and efficient robust estimator of MM-estimator to estimate the robust residuals.

[17-18] pointed out that the problem of Bootstrap MM-estimator is that the estimator tends to vary a lot when the sample size is too small or when the data are contaminated by a large number of measurement errors. As an alternative, a precise robust weighted bootstrap method with tolerance for high proportion of contaminated measurement errors can be used to maintain its robustness. To date, not much work has been devoted to wild bootstrap techniques in the presence of heteroscedasticity and outliers with high measurement errors in the data for both small and large sample sizes.

This paper considered the wild bootstrap of least trimmed squares (LTS) estimator [19] and the wild bootstrap of least median squares (LMS) estimator [20]. We have discovered a number of classical choices of weighted distribution in the wild bootstrap which are invalid for small sample size and are not tolerant to large measurement error. The numerical example and simulation study on proposed method is examined to demonstrate the relevance of our result in finite sample problems. We proposed a simple modification of robust wild bootstrap that is capable of handling small sample size and is tolerant to large measurement errors that give better performance which is asymptotically correct when the models are homoscedastic or heteroscedastic. In section 2, we discussed some existing robust wild bootstrap procedures found in the literature. Discussion on the two proposed methods RWBootWu and RWBootLiu will be done in Section 3. Results for the comparisons of the proposed methods with the existing methods (RBootWuMM, RBootLiuMM, RWBootWuLTS, RWBootLiuLTS, RWBootWuLMS, RWBootLiuLMS) through simulation and a real data set will be presented in Section 4 and 5. Discussion and conclusion are in Section 6.

2.0 VARIOUS WILD BOOTSTRAP PROCEDURES:

Consider the classical linear regression model in the following form

$$y = x\beta + \varepsilon \tag{1}$$

where $\mathbf{y}$ is a vector of observed values, $\mathbf{x}$ is a matrix of dimension (nxp) that contains the p vector of explanatory variables, β is a vector of regression parameter, ε is a vector of errors. The error assumes to satisfy a Gaussian distribution with mean zero, and unknown but constant variance. However, for many applications, the variance about the error terms are not homoscedastic and it is thus required to consider more consistent estimators for the variance of $\hat{\beta}$. Various wild bootstrap procedures are briefly discussed in the following section.

2.1 Robust Wild Bootstrap Techniques Based on Wu

We describe both robust wild bootstrap based on [1] and robust wild bootstrap based on [2] procedures here with a slight modification, that is, instead of MM-estimator for estimating $\hat{\beta}$, the LMS and LTS estimator using a Tukey weighting procedure will be used. This is because the existing method is greatly affected by large measurement errors and consequently the parameter produced from the estimate will be wrong.

First, we summarize the existing robust wild bootstrap scheme based on MM-estimate so that comparison can be made with the proposed methods. The steps involved are as follows:

Step 1: Fit model of $y_i = x_i\beta + \varepsilon_i$ using MM-estimator to estimate $\hat{\beta}_{MM}$. Step 2: Estimate the residuals of MM-estimator given as

$$\hat{\varepsilon}_i^{MM} = y_i - \hat{y}_i \tag{2}$$

Step 3: Assign weight to each MM-estimator residual $\hat{\varepsilon}_i^{MM}$ of equation 2. The weight will be equal to

$$w_{ii} = \begin{cases} 1 & if \ |\hat{\varepsilon}_i^{MM}| / \hat{\sigma}_{MM} \leq c \\ \dfrac{c}{(|\hat{\varepsilon}_i^{MM}| / \hat{\sigma}_{MM})} & if \ |\hat{\varepsilon}_i^{MM}| / \hat{\sigma}_{MM} > c \end{cases} \tag{3}$$

where $\hat{\sigma}_{MM}$ is the square root of the mean squares error of the residuals of MM-estimator Step 3. Compute the final weight of MM-estimator. This is obtained by multiplying the weight of Equation (3) with the estimate of the residuals of MM-estimator of equation (2). The weight that corresponds to good observation will have the final weighted residuals as: $\hat{\varepsilon}_i^{WMM} = 1 \times \hat{\varepsilon}_i^{WMM}$ and the weight that corresponds to bad observation will have the final weighted residuals as $\hat{\varepsilon}_i^{WMM} = c / (\hat{\varepsilon}_i^{MM} / \hat{\sigma}_{MM}) \times \hat{\varepsilon}_i^{MM}$

Step 4: Obtain the bootstrap sample of (y_i^*, X), where the estimate of y_i^* is given as

$$y_i^* = x\hat{\beta}_{MM} + \frac{t_i^* \hat{\varepsilon}_i^{WMM}}{(1 - h_{ii})} \tag{4}$$

where $h_{ii} = x'(x'x)^{-1}x$ is the i-th leverage; the value of ith leverage is used to reduce the influence of cases with large leverage point and for each i. The quantity t^* is drawn with replacement from a distribution with zero mean and unit variance or can be drawn from normalized residuals $a_1, a_2, ... a_n$, that is

$$^R a_i = \frac{\hat{\varepsilon}_i^{WMM} - \text{median}(\hat{\varepsilon}_i^{WMM})}{\text{NMAD}_{norm}(\hat{\varepsilon}_i^{WMM})} \tag{5}$$

where $MAD = \dfrac{1}{0.6745} median\{|\hat{\varepsilon}_i^{MM} - median(\hat{\varepsilon}_i^{MM})|\}$. The constant value 0.6745 is called the turning constant. It provides an unbiased estimate of σ^2_{MM} for independent observations from a normal distribution [16].

Step 5: Apply the OLS estimation procedure on the bootstrap sample of (y_i^*, X). This estimate is denoted by $^R \hat{\beta}^* = (X^1 X)^{-1} X^1 y_i^*$.

Step 6: Repeat Step 4 and 5 for k times, where k is the required number of bootstrap replicates. The bootstrap procedure is called RBootWuMM.

2.2 Robust Wild Bootstrap Techniques Based on Liu

The bootstrap MM-estimator based on [2] algorithm was also applied by [16] to estimate the parameter of the model. [2] suggested to modify the procedure of generating the $t*$ value. The $t*$ is randomly selected from auxiliary distribution that has a third central moment equal to one, in addition to the zero mean and unit variance. In this case, Wu shares the usual second order asymptotic properties of the classical bootstrap. Put differently, the addition of the restriction that the third central moment be equal to one and such kinds of selection are used to correct the skewness term in the edge worth expansion of the sampling distribution of $I'\hat{\beta}$, where I is an n-vector of ones.

Liu's bootstrap procedure can be applied as follows: As we want to discuss Step 4 of section 2.1, Steps 1 through 3 remain the same.

Step 4a: For each i, the quantity t^* is drawn with replacement from auxiliary distribution with zero mean and unit variance that has a third central moment equal to one, in addition to the zero mean and unit variance. They decided to estimate the value of $t*$ following the procedure of [21] which is described as follows:

$t_i^* = N_i M_i - E(N_i)E(M_i)$ 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 independently and identically distributed normally with mean $(1/2)(\sqrt{(\sqrt{17/6})} - \sqrt{1/6}$ and have variance of $1/2$. N_i's and M_i's are independent.

Step 4b: Form a bootstrap sample of (y_i^*, X), where the estimate of y_i^* is given as

$$y_i^* = x\hat{\beta}_{MM} + \frac{t_i^* \varepsilon_i^{WMM}}{(1 - h_{ii})} \tag{6}$$

Step 5: through 6 remain the same. The wild bootstrap obtained from such procedure is called RBootLiuMM.

3.0 PROPOSED METHODS: ROBUST WILD BOOTSTRAP LTS (RWBLTS) AND ROBUST WILD BOOTSTRAP LMS (RWBLMS)

The proposed robust wild bootstrap methods are based on [1] and [2], but the important difference lies in the choice of estimation method, weight distribution and estimation of standardized residuals. The existing robust wild method uses MM-estimator, but we use LTS and LMS estimators, because both methods have tolerance of large measurement errors because of their 50% breakdown point. Instead of estimating our standardized residuals from the square root of mean squares error of the residuals of the MM-estimator estimator, we estimate our standardized residuals based on median absolute deviation which is more robust.

The procedure for the proposed wild bootstrap algorithm based on the robust wild bootstrap least trimmed squares is summarized as follows:

Step 1: Fit the regression model $y_i = f(x_i, \beta_{LTS})$ using the LTS method to the original sample of observation to get $\hat{\beta}_{LTS}$, hence the fitted model becomes $y_i = f(x_i, \beta_{LTS})$

Step 2: Compute the residuals of the fitted model $\hat{\varepsilon}_i^{LTS} = y_i - \hat{y}_i$.

Step 3. Estimate the initial weight for all the cases by using the inverse of this absolute fitted value obtained in Step 1 and denote as w_{1i} where

$$w_{1i} = [X(X'X)^{-1}X'y]^{-1} \tag{7}$$

Step 4: Estimate the scaled residuals e_i, using the robust median absolute deviation given as:

$$e_i = \frac{|\hat{\varepsilon}_i^{LTS}|}{MAD} \text{ where } MAD = \frac{1}{0.6745} median\{|\hat{\varepsilon}_i^{LTS} - median(\hat{\varepsilon}_i^{LTS})|\} \tag{8}$$

It provides an unbiased estimate of σ_{LTS} for independent observations from a normal distribution [22].

Step 5: The final weight can be acquired from any robust weighted function procedure; however in this study, we used the Tukey bisquare weighted function, defined as:

$$\rho(e_i) = \begin{cases} \left\{ \left[\left(1 - (\frac{e_i}{c})^2\right)^2 \right] \right\} & for \ |e_i| \leq c \\ 0 & for \ |e_i| > c \end{cases} \tag{9}$$

where e_i is the standardize residual of LTS obtained from Step 4 and c = 4.685 is the turning constant which produces 95% efficiency relative to the sample mean for normal population [23].

Step 6: The final weighted residuals of the LTS estimate denoted as w_i^{LTS} are obtained by multiplying the weight w_{1i} with the weight w_{2i}. The weight that corresponds to good observation will have the final weight as $w_{2i} = \hat{\varepsilon}_i^{LTS} \times 1$ and the weight that corresponds to bad observation will also have the final weight as $w_{2i} = \hat{\varepsilon}_i^{LTS} \times 1.345 / e_i$

Step 7: The weighted residuals for the LTS estimate denoted as $\hat{\varepsilon}_i^{WLTS}$ are obtained by multiplying the weights in Step 6 with the residuals of the LTS estimate. Any observation that corresponds to good data point has the weighted residuals as $\hat{\varepsilon}_i^{WLTS} = 1 \times \hat{\varepsilon}_i^{LTS} \times [x'(x'x)^{-1}x'y]^{-1}$; otherwise the weighted residuals is $\hat{\varepsilon}_i^{WLTS} = \hat{\varepsilon}_i^{LTS} \times (1.345/|\hat{\varepsilon}_i|) \times [x'(x'x)^{-1}x'y]^{-1}$. However, because of the presence of heteroscedasticity in the data, the bootstrap schemes are modified to 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 8. Construct a bootstrap sample (y_i^*, x) where for each i, draw a value t^*, with replacement from a distribution with zero mean and unit variance attached to $\hat{y}_i$. For obtaining fixed-x-bootstrap values y_i^{*b}, where

$$y_i^{*b} = f(x_i, \hat{\beta}_{LTS}) + t^* \hat{\varepsilon}_i^{LTS} / \sqrt{1 - h_{ii}} \tag{10}$$

and $h_{ii} = x'(x'x)^{-1}x$ is the i-th leverage, the value of i-th leverage is used to reduce the influence of cases with large leverage point. We modified Wu's procedure by using 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. The following equation is then obtained as:

$$t^* = \frac{\hat{\varepsilon}_i^{WLTS} - \text{median}(\hat{\varepsilon}_i^{WLTS})}{\text{NMAD}_{\text{norm}}(\hat{\varepsilon}_i^{WLTS})} \quad for \ i=1,2,...,n \tag{11}$$

The estimate of normalized median absolute deviation of the weighted residuals is given as:

$$NMAD = \text{median}\left\{\left|\hat{\varepsilon}_i^{WLTS} - \text{median}(\hat{\varepsilon}_i^{WLTS})\right| / 0.6745\right\} \tag{12}$$

Step 9: Fit the LTS to the bootstrapped values y_i^{*b} on the fixed-x to obtain $\hat{\beta}_{LTS}^{*b}$.

Step 10: Repeat the procedures in Step 8 and 9 for k times to get $\hat{\beta}_{LTS}^{bi},...,\hat{\beta}_{LTS}^{*bk}$ where k is the number of bootstrap replications.

Step 11: Estimate the variance of the k vectors of estimated parameter obtained using the procedures in Steps 1 to 9.

The wild bootstrap obtained from such procedure is called RWBootWuLTS. Hence, in this study, we also applied the weighted bootstrap LTS based on Liu's algorithm to estimate the parameters of the model. It should be noted that the studies by [1] and [2] differ only in the choice of random sample of t*. In this research, the samples were randomly selected, following exactly the same procedure as proposed by [16]. The wild bootstrap obtained from such procedure is called RWBootLiuLTS. We also apply the same procedure as in LTS to LMS to estimate the parameter of the model. The wild bootstrap obtained from such procedure is called RWBootWuLMS and RWBootLiuLMS.

4.0 NUMERICAL EXAMPLE

To assess the performance of the proposed method, we consider body fat data. This data is used by many researchers such as [24] and [25]. It describes the percentage of body fat, age, weight, height, and ten body circumference measurements (e.g., abdomen) recorded from 252

men. It consists of the following components: y= PCTBF, x1= Density, x2= Age, x3= Weight, x4= Height, x5= Neck, x6= Chest, x7= Abdomen, x8= Hip, x9= Thigh, x10= Knee, x11= Ankle, x12= Biceps, x13= Forearm, x14= Wrist.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_{14} x_{14} + \varepsilon \tag{13}$$

Table 1: Wild Bootstrap Standard Error of the parameters for the Robustness of real data (Tukey)

Coef.	Robust MM-Estimation		Robust LMS-Estimation		Robust LTS-Estimation	
	RBootWu	RBootLui	RWBootWu	RWBootLui	RWBootWu	RWBootLui
β_0	0.8620	0.7302	0.1806	0.1054	0.1254	0.1377
β_1	0.6358	0.5465	0.1284	0.0741	0.0857	0.0972
β_2	0.0008	0.0006	0.0001	0.0001	0.0001	0.0001
β_3	0.0014	0.0012	0.0003	0.0002	0.0002	0.0003
β_4	0.0031	0.0026	0.0008	0.0005	0.0006	0.0007
β_5	0.0056	0.0048	0.0011	0.0007	0.0008	0.0009
β_6	0.0024	0.0020	0.0005	0.0003	0.0004	0.0004
β_7	0.0025	0.0021	0.0005	0.0003	0.0003	0.0004
β_8	0.0036	0.0030	0.0007	0.0004	0.0005	0.0005
β_9	0.0034	0.0028	0.0007	0.0004	0.0005	0.0005
β_{10}	0.0059	0.0050	0.0012	0.0007	0.0008	0.0009
β_{11}	0.0059	0.0053	0.0015	0.0009	0.0010	0.0011
β_{12}	0.0041	0.0035	0.0008	0.0005	0.0006	0.0006
β_{13}	0.0051	0.0044	0.0013	0.0008	0.0010	0.0010
β_{14}	0.0138	0.0110	0.0025	0.0016	0.0019	0.0020
AV.SE	0.1037	0.0883	0.0214	0.0125	0.0177	0.0163

We initially checked on the basis of whether this data contained any outliers or not by using standardized residuals of LTS and it is observed that it contained about 10% outliers. We applied robust Goldfeld-Quandt test on the data by testing the suspected regressor variable with the response variables without the points identified as outliers by the LTS procedures to test for the presence of heteroscedasticity. The MGQ test was performed on the basis that the people with high percent of body fat will be expected to have higher abdomen circumference. The result of modified Goldfeld-Qundl test of MGQ = 10.4591 is compared with the critical value for F-Statistics with $(n_1-c-2k)/2$ and $(n_2-c-2k)/2$ degree of freedom, which is F= 0.005 for the 5% significance level. As a result our alternative hypothesis that there is heteroscedasticity in the data was accepted. We applied three robust wild bootstrap techniques to the data and the results are presented in Table 1. This table exhibits the standard error of parameter estimate obtained from the robust methods. It is interesting to

observe that the robust wild bootstrap of MM-estimator have larger standard errors when compared with the robust wild bootstrap estimator for both LTS and LMS methods in the presence of outliers and heteroscedasticity in the data. This is not evidence for our final conclusion. Based on the results obtained (Table 1), so far, we can say that the MM-estimator is affected by the presence of outliers.

5. 0 SIMULATION RESULTS

Here, we carry out an extensive simulation study on finite sample to compare the performance of RBootWuMM, RBootWuMM, RWBootWuLMS, RWBootWuLMS, RWBootWuLTS and RWBootLiuLTS method. We considered data generating procedure similar to [16] and [26]. The design of this research involves a linear model of two covariates:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \sigma_i \varepsilon_i \qquad (14)$$

To generate the covariate values, we consider the sample sizes of n= 20, 60 and 100. For the case of sample size n=20, the covariate observations of x_{1i} and x_{2i} were generated from N (0, 1). We replicated these observations three and five times to generate the sample size n=60 and n=100 respectively. We performed the data generating procedures using $\beta_0 = \beta_1 = \beta_2 = 1$. For all i under homoscedasticity, $\sigma_i = 1$. Now we obtained heteroscedasticity generating procedure following [16] and [27]. The heteroscedasticity generation function is defined as $\sigma_i = \exp(1.5x_{1i} + 1.5x_{2i})$. The level of heteroscedasticity in this research is measured by $\phi = \max(\sigma_i^2) / \min(\sigma_i^2)$ and the degree of heteroscedasticity remains constant for different sample sizes. However, for each simulation run and for various sample sizes, $\varepsilon_i's$ were drawn from standard normal distribution with mean zero and variance one i.e. N (0, 1) for the case of data with no outliers. Our interest is to estimate a regression model that would involve outliers. We then start contaminating the data by randomly substituting some limited number of good observations with a certain percentage of outliers. In this respect, the points that we replaced will produce large residuals and thus identified as outliers in the data set. The heteroscedasticity linear model with outliers becomes

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \sigma_i \varepsilon_{i(conter\min ated)} \qquad (15)$$

and $\varepsilon_{i(conter\min ated)} = \alpha N(0,1) + (1 - \alpha) N(5,10)$ where α is selected based on the percentage of outliers. In this research we generate 5%, 10% and 15% outliers, the 95%, 90% and 85% of $\varepsilon_i's$ were drawn from N(0,1) and 5%, 10% and 15% outliers were generated from N(5,10). The simulation of each sample size was replicated 500 times with 1000 bootstrap sample replicates. We performed a similar simulation procedure following the design of [16] and [21].
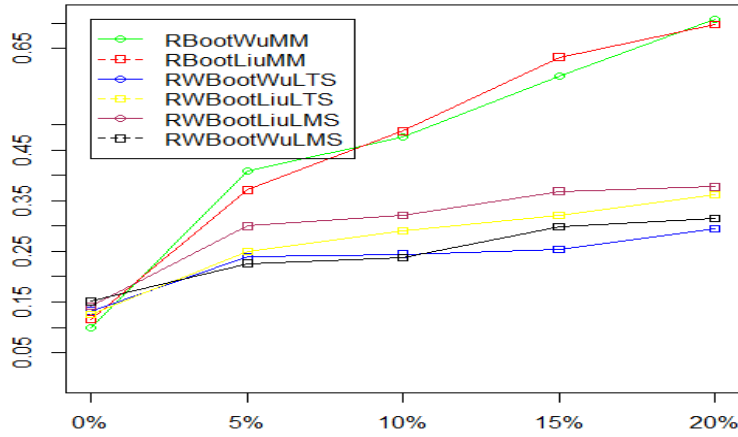
Figure 1: The average effect of outlier percentage on standard error for the sample size n=20
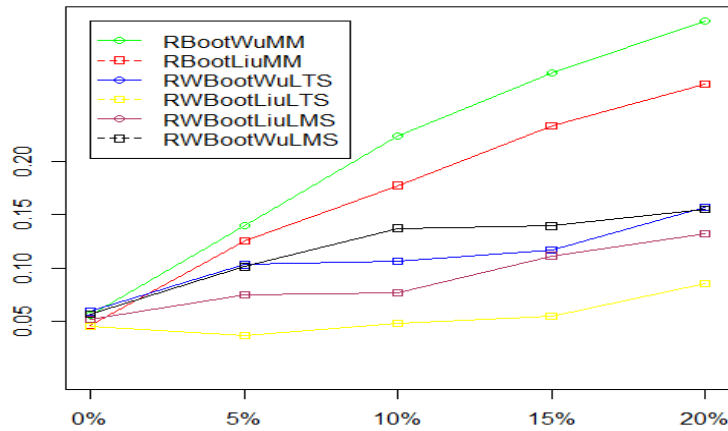

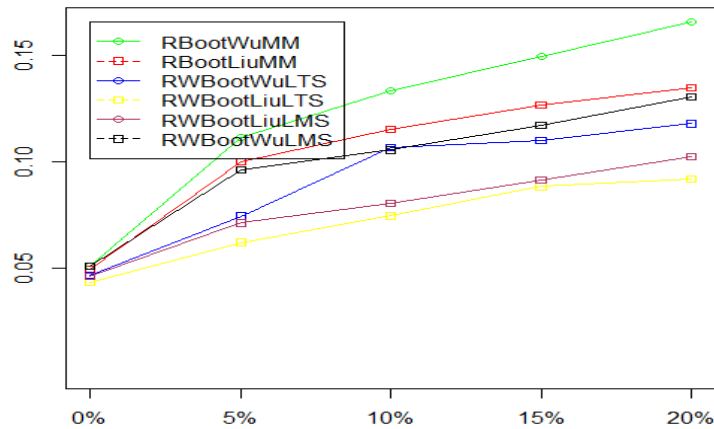**Figure 2:** The average effect of outlier percentage on standard error for the sample size n=60

Figure 3: The average effect of outlier percentage on standard error for the sample size n=100

Table 2: Root Mean Squares Error for Robust Wild Bootstrap estimation method data N= 20

Outliers	Coef.	MM-Estimation		LTS-Estimation		LMS-Estimation	
		RBootWu	RBootLui	RWBootWu	RWBootLui	RWBootWu	RWBootLui
0%	β_0	0.0864	0.1246	0.1263	0.1555	0.2603	0.1833
	β_1	0.0877	0.1265	0.1426	0.2338	0.1598	0.3424
	β_2	0.1398	0.1033	0.2385	0.1310	0.3706	0.4014
5%	β_0	0.2935	0.4569	0.2924	0.6146	0.3039	0.3299
	β_1	0.4442	0.3829	0.3005	0.3306	0.3857	0.2474
	β_2	0.3804	0.4317	0.2061	0.4520	0.2907	0.3394
10%	β_0	0.4599	0.4635	0.2679	0.4303	0.2726	0.6979
	β_1	0.4028	0.5495	0.2684	0.4673	0.4296	0.6826
	β_2	0.5032	0.5982	0.2176	0.3918	0.2501	0.6318
15%	β_0	0.5964	0.5492	0.1930	0.3641	0.3529	0.3246
	β_1	0.6692	0.8279	0.3364	0.3382	0.3275	0.4918
	β_2	0.5879	0.7197	0.2584	0.3484	0.2149	0.3466
20%	β_0	0.6048	0.6695	0.2945	0.3980	0.3304	0.3258
	β_1	0.8290	0.8791	0.3447	0.2983	0.3480	0.4759
	β_2	0.8002	0.9519	0.3145	0.3928	0.3270	0.3346

We estimate the bias, root mean squares errors and standard errors of various sample sizes with different percentage of outliers. The bootstrap bias of RBootWuMM and RBootLiuMM can be obtained by subtracting the difference between the true model and the estimated model as:

$$\text{Bias} = \hat{\beta}_{bMM} - \hat{\beta}_{MM} \tag{16}$$

Here $\hat{\beta}_{bMM} = (k)^{-1} \sum_{b=1}^{k} \hat{\beta}_{MM}^{*b}$ and the corresponding estimate of bootstrap standard error of RBootWuMM and RBootLiuMM can be acquired from the square root of the main diagonal of the covariance matrix and is given as

$$SE(\hat{\beta}_{bMM}) = \sqrt{\frac{\sum_{b1}^{k} (\beta_{(MM)}^{*b} - \hat{\beta}_{(bMM)})(\beta_{(MM)}^{*b} - \hat{\beta}_{(bMM)})'}{(k-1)}} \tag{17}$$

Table 3: Root Mean Squares Error for Robust Wild Bootstrap estimation data N= 60

Outliers	Coef.	MM-Estimation		LTS-Estimation		LMS-Estimation	
		RBootWu	RBootLui	RWBootWu	RWBootLui	RWBootWu	RWBootLui
0%	β_0	0.1079	0.0483	0.0957	0.0546	0.0675	0.0614
	β_1	0.0856	0.1583	0.0746	0.0404	0.0713	0.0539
	β_2	0.1487	0.0695	0.1098	0.0416	0.0532	0.0434
5%	β_0	0.1324	0.1740	0.1094	0.0395	0.1245	0.0992
	β_1	0.1648	0.4070	0.1130	0.0935	0.0993	0.1174
	β_2	0.1359	0.1302	0.0974	0.0454	0.1013	0.0907
10%	β_0	0.2449	0.1538	0.0815	0.0482	0.1171	0.0788
	β_1	0.2569	0.1767	0.1118	0.0549	0.1307	0.0868
	β_2	0.2785	0.2154	0.1270	0.0491	0.1789	0.0942
15%	β_0	0.3273	0.2398	0.1293	0.0555	0.1660	0.1469
	β_1	0.4496	0.2563	0.1893	0.0529	0.2953	0.0976
	β_2	0.2651	0.2783	0.1415	0.0726	0.1857	0.1466
20%	β_0	0.2442	0.2482	0.1665	0.1078	0.1720	0.2773
	β_1	0.4500	0.5492	0.1446	0.0805	0.1264	0.1084
	β_2	0.3735	0.2700	0.1651	0.0950	0.2222	0.2891

Table 4: Root Mean Squares Error for Robust Wild Bootstrap estimation data N= 100

Outliers	Coef.	MM-Estimation		LTS-Estimation		LMS-Estimation	
		RBootWu	RBootLui	RWBootWu	RWBootLui	RWBootWu	RWBootLui
0%	β_0	0.0668	0.0563	0.0247	0.0517	0.0578	0.0504
	β_1	0.0902	0.0635	0.0661	0.0404	0.0823	0.0571
	β_2	0.1622	0.0512	0.0537	0.0512	0.0795	0.0526
5%	β_0	0.1242	0.1014	0.0671	0.0685	0.1042	0.1058
	β_1	0.1603	0.2571	0.0816	0.0656	0.1060	0.1163
	β_2	0.1160	0.2512	0.0758	0.0623	0.1027	0.0767
10%	β_0	0.1297	0.1116	0.0930	0.0612	0.2144	0.0724
	β_1	0.1845	0.1119	0.1093	0.0775	0.2580	0.1461
	β_2	0.1404	0.1295	0.1180	0.0860	0.1130	0.0802
15%	β_0	0.1522	0.1248	0.1123	0.1072	0.1337	0.1268
	β_1	0.3773	0.2375	0.1221	0.0923	0.1132	0.1441
	β_2	0.1852	0.1330	0.1084	0.0917	0.1283	0.1652

	β_0	0.1575	0.1270	0.1143	0.0823	0.1296	0.1524
20%	β_1	0.2215	0.1512	0.1379	0.1053	0.1407	0.1415
	β_2	0.2000	0.1341	0.1413	0.1194	0.1209	0.1820

Table 5: Bias Measurement for Robust Wild Bootstrap estimation method data N=20

Outliers	Coef.	MM-Estimation		LTS-Estimation		LMS-Estimation	
		RBootWu	RBootLui	RWBootWu	RWBootLui	RWBootWu	RWBootLui
0%	β_0	0.0203	0.0186	-0.0020	-0.0986	0.2044	-0.0434
	β_1	0.0423	0.0287	0.0645	-0.1159	0.0711	-0.2747
	β_2	0.0199	0.0175	0.0473	-0.0662	0.2539	0.3666
5%	β_0	-0.1188	0.0092	-0.1615	-0.5769	0.1960	0.0348
	β_1	-0.2246	-0.2518	-0.1245	-0.0905	0.3523	-0.0679
	β_2	0.0579	-0.2252	0.0474	0.3961	-0.0316	-0.0614
10%	β_0	0.0598	-0.0089	-0.0274	-0.3217	0.1076	-0.6192
	β_1	0.1120	-0.1022	-0.0520	-0.3756	0.3419	-0.6075
	β_2	0.1965	-0.3833	-0.0772	0.2406	0.1474	0.5387
15%	β_0	0.0831	0.0190	0.0687	-0.0022	0.0090	0.1041
	β_1	-0.0262	0.5002	-0.0316	0.2198	-0.0318	0.0598
	β_2	-0.2529	0.1976	-0.0782	-0.0688	0.0965	-0.1533
20%	β_0	0.0530	0.1822	0.0280	-0.0478	0.0572	0.0516
	β_1	0.2412	0.7196	0.2053	0.0183	0.1896	0.0095
	β_2	-0.3332	0.1319	-0.0371	0.0036	-0.0070	0.0226

Table 6: Bias Measurement for Robust Wild Bootstrap estimation method data N=60

Outliers	Coef.	MM-Estimation		LTS-Estimation		LMS-Estimation	
		RBootWu	RBootLui	RWBootWu	RWBootLui	RWBootWu	RWBootLui
0%	β_0	0.0952	-0.0271	-0.0731	-0.0066	0.0394	-0.0157
	β_1	0.0564	-0.1496	0.0384	0.0023	0.0305	-0.0038
	β_2	-0.1395	-0.0518	-0.0966	0.0030	0.0123	-0.0019
5%	β_0	-0.0130	0.1213	-0.0341	-0.0155	-0.0202	-0.0246
	β_1	0.0482	-0.3849	0.0202	-0.0842	-0.0054	-0.0956
	β_2	-0.0412	0.0510	0.0249	0.0313	-0.0606	0.0678
10%	β_0	0.0217	0.0146	0.0152	-0.0098	-0.0352	-0.0236
	β_1	-0.1932	0.0270	0.0060	0.0141	-0.0168	0.0078
	β_2	-0.1019	-0.0702	-0.0105	-0.0206	0.0551	0.0639
15%	β_0	0.0187	0.0627	-0.0335	-0.0044	-0.1120	0.0620
	β_1	0.3685	-0.1247	0.1697	0.0053	0.2564	0.0170
	β_2	-0.0251	-0.1363	0.0105	0.0439	-0.1095	-0.1023

Outliers	Coef.						
20%	β_0	-0.0281	0.0077	0.0003	-0.0411	0.0363	-0.2281
	β_1	0.2113	0.4509	0.0333	0.0039	-0.0620	-0.0297
	β_2	-0.1228	0.0911	-0.0300	0.0559	-0.1207	0.2558

Table 7: Bias Measurement of Robust Wild Bootstrap estimation method data N=100

Outliers	Coef.	MM-Estimation		LTS-Estimation		LMS-Estimation	
		RBootWu	RBootLui	RWBootWu	RWBootLui	RWBootWu	RWBootLui
0%	β_0	-0.0406	0.0192	0.0028	0.0154	0.0392	0.0164
	β_1	-0.0802	-0.0365	0.0200	-0.0169	-0.0669	0.0410
	β_2	-0.1513	-0.0267	0.0123	0.0232	-0.0499	0.0127
5%	β_0	-0.0438	-0.0045	-0.0064	0.0082	-0.0078	0.0703
	β_1	0.1215	0.2388	0.0047	-0.0164	-0.0401	-0.1005
	β_2	-0.0219	0.2291	0.0118	0.0301	0.0558	-0.0108
10%	β_0	-0.0022	-0.0087	-0.0049	-0.0023	-0.1821	0.0061
	β_1	0.1223	-0.0159	0.0119	-0.0112	-0.2366	0.1157
	β_2	0.0471	0.0375	0.0075	0.0080	0.0514	-0.0004
15%	β_0	-0.0435	0.0541	-0.0045	-0.0248	0.0395	-0.0921
	β_1	-0.3435	0.1822	0.0535	-0.0377	0.0243	0.1060
	β_2	0.1134	-0.0659	0.0064	0.0505	-0.0609	0.1390
20%	β_0	0.0170	-0.0022	-0.0466	-0.0141	-0.0010	-0.1188
	β_1	0.1308	0.0420	0.0271	-0.0507	-0.0032	-0.0953
	β_2	0.1176	-0.0221	0.0832	-0.0607	-0.0045	-0.1473

Table 8: Standard Error for the Robust Wild Bootstrap estimation method data N= 20

Outliers	Coef.	MM-Estimation		LTS-Estimation		LMS -Estimation	
		RBootWu	RBootLui	RWBootWu	RWBootLui	RWBootWu	RWBootLui
0%	β_0	0.084	0.1232	0.1263	0.1203	0.1611	0.1781
	β_1	0.0768	0.1232	0.1272	0.2031	0.1431	0.2044
	β_2	0.1384	0.1018	0.2338	0.113	0.2699	0.1636
AV.SE		0.0997	0.1161	0.1624	0.1455	0.1914	0.1820
5%	β_0	0.2684	0.4568	0.2438	0.2119	0.2322	0.3281
	β_1	0.3832	0.2885	0.2735	0.318	0.157	0.2379
	β_2	0.376	0.3683	0.2006	0.2178	0.289	0.3338
AV.SE		0.3425	0.3712	0.2393	0.2492	0.2261	0.2999
10%	β_0	0.456	0.4634	0.2665	0.2858	0.2505	0.322
	β_1	0.3869	0.5399	0.2633	0.2781	0.2601	0.3113
	β_2	0.4632	0.4592	0.2034	0.3092	0.2021	0.3302
AV.SE		0.4754	0.4874	0.2444	0.2910	0.2376	0.3212
	β_0	0.5906	0.5489	0.1804	0.3641	0.3528	0.3075

15%	β_1	0.6687	0.6597	0.3349	0.257	0.326	0.4882
	β_2	0.5307	0.692	0.2463	0.3415	0.192	0.3109
AV.SE		0.5967	0.6335	0.2539	0.3209	0.2983	0.3689
	β_0	0.6025	0.6442	0.2932	0.3951	0.3254	0.3217
20%	β_1	0.7931	0.5049	0.2769	0.2977	0.2918	0.4758
	β_2	0.7275	0.9427	0.3123	0.3928	0.3269	0.3338
AV.SE		0.7077	0.6973	0.2941	0.3619	0.3147	0.3771

Table 9: Standard Error for the Robust Wild Bootstrap estimation method data N= 60

		MM-Estimation		LTS-Estimation		LMS-Estimation	
Outliers	Coef.	RBootWu	RBootLui	RWBootWu	RWBootLui	RWBootWu	RWBootLui
	β_0	0.0507	0.0400	0.0617	0.0542	0.0548	0.0594
0%	β_1	0.0644	0.0519	0.0639	0.0403	0.0645	0.0538
	β_2	0.0514	0.0464	0.0523	0.0415	0.0518	0.0434
AV.SE		0.0555	0.0461	0.0593	0.0453	0.0570	0.0522
	β_0	0.1318	0.1247	0.1039	0.0363	0.1228	0.0961
5%	β_1	0.1576	0.1323	0.1112	0.0406	0.0992	0.0681
	β_2	0.1295	0.1198	0.0942	0.0329	0.0812	0.0603
AV.SE		0.1396	0.1256	0.1031	0.0366	0.1011	0.0748
	β_0	0.2439	0.1531	0.0801	0.0472	0.1117	0.0752
10%	β_1	0.1693	0.1746	0.1116	0.0531	0.1296	0.0864
	β_2	0.2592	0.2036	0.1266	0.0446	0.1702	0.0692
AV.SE		0.2236	0.1771	0.1061	0.0483	0.1372	0.0769
	β_0	0.3268	0.2315	0.1249	0.0553	0.1225	0.1332
15%	β_1	0.2575	0.2239	0.0839	0.0526	0.1465	0.0961
	β_2	0.2639	0.2426	0.1411	0.0578	0.1500	0.105
AV.SE		0.2827	0.2327	0.1166	0.0552	0.1393	0.1114
	β_0	0.2426	0.2481	0.1665	0.0997	0.1681	0.1576
20%	β_1	0.3973	0.3135	0.1407	0.0804	0.1102	0.1042
	β_2	0.3527	0.2542	0.1624	0.0768	0.1865	0.1347
AV.SE		0.3309	0.2719	0.1565	0.0856	0.1549	0.1322

Table 10: Standard Error for the Robust Wild Bootstrap estimation method data N= 100

		MM-Estimation		LTS-Estimation		LMS-Estimation	
Outliers	Coef.	RBootWu	RBootLui	RWBootWu	RWBootLui	RWBootWu	RWBootLui
	β_0	0.0530	0.0529	0.0245	0.0494	0.0425	0.0477
0%	β_1	0.0412	0.0519	0.0630	0.0367	0.0480	0.0397

	β_2	0.0584	0.0437	0.0523	0.0456	0.0619	0.0510
AV.SE		0.0509	0.0495	0.0466	0.04329	0.0508	0.0461
	β_0	0.1162	0.1013	0.0668	0.0680	0.1039	0.079
5%	β_1	0.1046	0.0952	0.0815	0.0635	0.0981	0.0586
	β_2	0.1139	0.1031	0.0749	0.0545	0.0862	0.0759
AV.SE		0.1116	0.0999	0.0744	0.0620	0.0961	0.0712
	β_0	0.1297	0.1113	0.0929	0.0612	0.1132	0.0721
10%	β_1	0.1381	0.1108	0.1087	0.0767	0.1028	0.0892
	β_2	0.1323	0.1239	0.1178	0.0856	0.1006	0.0802
AV.SE		0.1334	0.1153	0.1065	0.0745	0.1055	0.0805
	β_0	0.1458	0.1125	0.1122	0.1043	0.1277	0.0872
15%	β_1	0.1562	0.1523	0.1098	0.0842	0.1106	0.0976
	β_2	0.1464	0.1155	0.1082	0.0765	0.1129	0.0892
AV.SE		0.1495	0.1268	0.1101	0.0883	0.1171	0.0913
	β_0	0.1566	0.1270	0.1044	0.0811	0.1296	0.0954
20%	β_1	0.1787	0.1453	0.1352	0.0923	0.1407	0.1046
	β_2	0.1618	0.1323	0.1142	0.1028	0.1208	0.1069
AV.SE		0.1657	0.1349	0.1179	0.0920	0.1304	0.1023

6.0 RESULTS AND DISCUSSION

Figures 1-3 described the influence of outliers on the standard errors of various robust wild bootstrap methods. The average standard errors of regression parameter are plotted at different percentage level of outliers for different wild bootstrap. The results disclosed that the performances of different robust wild bootstrap methods are fairly close to each other at 0% level of contamination. It appeared that the average standard errors of RBootWuMM, RBootLiuMM, and the proposed RWBootWuLTS, RWBootLiuLTS, RWBootWuLMS and RWBootLiuLMS are closer to each other. However, when the percentage of outliers is introduced to 5%, 10%, 15% and 20% levels of contaminated data, the RBootWuMM and RBootLiuMM estimate becomes large. It is curious to observe that not much outlier influence is feasible for the modified RWBootLiuLTS, RWBootWuLMS and RWBootLiuLMS methods at different percentage levels of outlier. In contrast, it is observed that at n=20 sample size, the average standard errors of RWBootLiuLTS and RWBootWuLMS are smaller than the RWBootLiuLMS and RWBootLiuLTS. When the sample size is increased to n=60, and n= 100 the RWBootLiuLMS and RWBootLiuLTS have the least average standard errors.

The simulation results of different robust wild bootstrap techniques are described in Table 2-10. Our focus is to investigate the effect of outliers and heteroscedasticity on the coefficient of regression model by looking at the bias and the standard errors as well as their root mean squares errors. Different interesting points can be seen in these tables. The results presented

in Table 2-10 described the bias and root mean squares error for small and large sample size of the proposed method with the existing methods. It was observed that for different sample size, the robust wild bootstrap method of RBootWuMM and RBootLiuMM does not perform well when compared with the proposed method of RWBootWuLTS, RWBootLiuLTS, RWBootWuLMS and RWBootLiuLMS in case of contaminated data. The RBootWu and RBootLiu performed well but the performance is not up to expectations when compared with our proposed method for all sample sizes. On the other hand, it is also observed that the RWBootWuLTS, RWBootLiuLTS, RWBootWuLMSm and RWBootLiuLMS provide the least standard error among all and reduces further with increased sample size. As the percentage of outliers increase to 10%, one can notice that the increase in bias and standard errors for RBootWuMM and RBootLiuMM is much larger than the increase in bias and standard errors for RWBootWuLTS, RWBootLiuLTS, RWBootWuLMS and RWBootLiuLMS. The proposed method appeared to be more resistant to outliers among the three estimation methods with respect to their bias, root mean squares errors and standard errors. The error measures of RWBootWuLTS, RWBootLiuLTS, RWBootWuLMS and RWBootLiuLLMS were consistently smaller for all sample sizes and in different percentage of outliers. It was observed that as the percentage of outlier's increases both the RBootWuMM and RBootLiuMM perform badly when compared with the proposed method as their bias, root mean square errors and standard errors would increase significantly. Finally, the results of this analysis clearly show that our new wild bootstrap method is very robust and is proven to be more resistant to outliers in the data since it will not easily be influenced by the presence of even large numbers of outliers. As shown, it produced the least standard errors, bias and root mean squares errors in both simulation and numerical examples. Hence, it becomes a robust wild bootstrap alternative to existing wild bootstrap techniques.

7.0 CONCLUSION

The presence of heteroscedasticity and outliers in the data required a comprehensive investigation of both regression and bootstrap methods. In this paper, we examined the finite sample behaviour of new bootstrap procedures namely RBootWuMM, RBootWuMM, RWBootWuLMS, RWBootWuLMS, RWBootWuLTS and RWBootLiuLTS in linear regression model in the presence of outliers and heteroscedasticity using numerical examples and simulation study. The results obtained from both numerical example and simulation study have revealed several important results. In our simulation study, this estimator performed as well as its robust competitors when there are no outliers. In the presence of outliers and heteroscedasticity, the results shows that robust bootstrap methods of RWBootWuLTS, RWBootLiuLTS, RWBootWuLMS and RWBootLiuLMS techniques have consistently outperformed both the robust bootstrap of RBootWu and RBootLiu when the percentage of outliers increases. It appeared that the Tukey bisquares weighted function and wild bootstrap procedures applied on RWBootWu, RWBootLiu, RWBootWuLMS and RWBootLiuLMS have improved the performance of robust wild bootstrap method. Finally, the RWBootWuLTS, RWBootLiuLTS, RWBootWuLMS and RWBootLiuLMS estimators of regression model have proven to be good alternative to other robust wild bootstrap procedures particularly when the data contain high percentage of outliers.

REFERENCES

[1]   C.F.J. Wu. Jackknife, Bootstrap and other resampling methods in regression analysis, The Annals of Statistics 14 (1986) 1261-1350.

[2]   R.Y. Liu, Bootstrap procedures under some non- i.i.d. Models, The Annals of Statistics 16 (1988) 1696-1708.

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

[4]   B. Efron, R. Tibshirani, Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy, Statistical Science 1 (1986) 54-77.

[5]   J. Shao, D. Tu, Jackknife and bootstrap, New York, Springer, (1986).

[6]   E. Ammen, When does bootstrap work? Asymptotic results and Simulation, New York, Springer, (1991).

[7]   E. Flachaire, A better way to bootstrap pairs, Economics Letters 64 (1999) 257-262.

[8]   C.R. Rao, L.C. Zhao, Approximation to the distribution of M-estimates in linear models by randomly weighted bootstrap, Samkhya 54 (1992) 323- 331.

[9]   S.N. Lahiri, Bootstraping M-estimators of a multiple linear regression parameter, Annals of Statistics 20 (1992) 1548-1570.

[10]   K. Knight, Asymptotic fot L_1 estimators of regression under heteroscedasticity, Canadian Journal of Statistics 27 (1999) 497-507.

[11]   S. Alamgir, A. Ali, Split Sample Bootstrap Method World Applied Science Journal 21 (2013) 983-993.

[12]   Z. Hongtu, G.I. Joseph, S.P. Bradley, A Statistical Analysis of Brain Morphology Using Wild Bootstrapping, IEEE Transactions on Medical Imaging 26 (2007) 954-966.

[13]   A.C. Davidson, E. Falchaire, The wild bootstrap tame at last, Journal of Econometrics 146 (2008) 162- 169.

[14]   M.R. Norazan, M. Habshah, A.H.M.R. Imon, Estimating Regression Coefficient using Weighted Bootstrap with Probability, WSEAS Transactions on Mathematics 8 (2009) 362-371.

[15]   F. Xingdong, X. He, J. Hu, Wild Bootstrap for Quantile Regression, Biometrika 98 (2011) 995-999.

[16]   R. Sohel, M. Habshah, A.H.M.R. Imon, Robust wild bootstrap for stabilizing the variance of parameter estimates in heteroscedastic regression models in the presence of outliers, Mathematical Problems in Engineering 2012 (2012) 1-14.

[17]   C.L. Cheng, J. Riu, On estimating linear relationships when both variables are subject to heteroscedastic measurement errors, Technometrics 48 (2006) 511.

[18] A.R. Bello, A. Robiah, S. Ehsan, D. Kafi, Application of robust wild bootstrap estimation of linear model, Econometric International Journal of Applied mathematics and Statistics 53 (2015) 82-101.

[19] J.P. Rousseeuw, A.M. Leroy, Robust Regression and Outlie Detection, Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics, John Wiley & Sons, New York, NY, USA (1987).

[20] P.J. Rousseeuw, Least median of squares regression, Journal of the American Statistical Association 79 (1984) 871-880.

[21] F. Cribari-Nato, S.G. Zarkos, Bootstrap methods for heteroscedastic regression models: evidence on estimation and testing, Econometric Reviews 18 (1999) 211-228.

[22] M. Habshah, R. Sohel, A.H.M. Rahmatullah Imon, The performance of robust weighted least squares in the presence of outliers and heteroscedastic errors, WSEAS Transactions on Mathematics 8 (2009) 351- 361.

[23] A.R. Bello, A. Robiah, S. Ehsan, D. Kafi, Robust weighted least squares estimation of regression parameter in the presence of outliers and heteroscedastic errors, Journal of technology 71 (2014) 11–18.

[23] K.W. Penrose, A.G. Nelson, A.G. Fisher, Generalized body composition prediction equation for men using simple measurement techniques, Medicine and Science in Sports and Exercise 17 (1985) 189.

[25] R.W. Johnson, Fitting percentage of body fat to simple body measurements, Journal of Statistics Education 4 (1996).

[26] H. White, J.G. Mackinnon, Some heteroscedasticity-consistent covariance matrix estimators with improved finite sample properties, Journal of Econometrics 29 (1985) 305-325.

[27] F. Cribari-Neto, Asymptotic inference under heteroscedasticity of unknown form, Computational Statistics & Data Analysis 45 (2004) 215-233.