Bima Journal of Science and Technology, Vol. 8(2B) July, 2024 ISSN: 2536-6041

DOI: 10.56892/bima.v8i2B.737

Robust Bootstrap for Handling Heteroscedasticity and Outliers in the Presence of High Leverage Point

M. Mijinyawa[1]*, B. A. Rasheed[2] and A. Abdulkadir[3]

[1]Mathematical Science Department, Faculty of Sciences, Abubakar Tafawa Balewa University Bauchi

[2]Mathematics Department, Faculty of Sciences, Gombe State University, Gombe

[3]Mathematical Science Department, Faculty of Sciences, Abubakar Tafawa Balewa University Bauchi

Corresponding Author: almubarack77@gmail.com

ABSTRACT

It's fascinating how researchers are constantly improving regression analysis methods to deal with issues like heteroscedasticity. The robust MM estimator seems like a smart choice to enhance the wild bootstrap process for more accurate results in regression analysis. Researchers are debating the best bootstrap technique for dealing with outliers and heteroscedasticity in linear regression. There is a push for a more efficient and accurate method, considering the drawbacks of the Minimum Volume Ellipsoid approach. The proposal to replace MVE with ISE in the modified method is a promising step towards better speed, accuracy, and efficiency in robust bootstrapping. The specific objective of this paper is to modify the existing robust bootstrap technique (WBootMM-GM6-Liu). The methodology understudied the existing models and compared four existing bootstrap techniques with the modified version of the WBootMM-GM6-Liu to ascertain the impact of the modification. The numerical test results revealed that the modified version of the technique has the least standard errors, bias, and root mean square errors (RSME) and therefore outperforms the existing models taking into account the presence of heteroscedasticity, outliers, and high leverage points (HLPs). In the case of further research, this model can possibly be improved upon based on assessing fixed and random effects with other variables apart from those considered in this paper.

Keywords: heteroscedasticity, outliers, high leverage points, minimum volume ellipsoid, index set equality, robust bootstrap

INTRODUCTION

In multiple regression, ordinary least squares (OLS) estimation is used if assumptions are met to obtain regression weights when analyzing data, OLS assumes that residual errors should be normally distributed, have equal variance at all levels of the explanatory variables, and be uncorrelated with both the independent variables and each other (Yan and Su, 2009). In practice, the assumption that residual errors should be normally distributed may not hold because of the possibility of skewness or the presence of outliers in data. In theory, when this assumption is not met, the OLS estimation for the regression coefficients β will be biased and/or non-efficient.

Homoscedasticity refers to the situation when the variance of the error terms is constant. Heteroscedasticity is a common problem in a linear regression model, which occurs when the variance of the error terms is not constant (Lukman et al., 2016). In this situation, the OLS estimator is no longer efficient. There are several methods to rectify the problem of heteroscedasticity (Habshah et al., 2011). A weighted bootstrap method proposed by Wu (1986) is one of the alternative methods to rectify this problem. Liu (1988) suggested a wild bootstrap approach that, under both homoscedastic and heteroscedastic models, is slightly different from the weighted bootstrap method and works better. Rana et al., (2012) suggested that there is evidence that the presence of outliers due to the use of ordinary least squares (OLS) in their algorithm causes such wild bootstrap estimators to suffer a huge setback. So, in the construction of the robust wild bootstrap process, they implemented the robust MM estimator. The MM estimator, however, does not have limited impact properties. Hence, in this study, we attempt to improve on the WBootMM-GM6 (Liu) introduced by Osama and Paul (2021) by incorporating the Index Set Equality (ISE) robust estimator in coming up with a Modified Wild Bootstrap.

The motivation to introduce the bootstrap method is the infeasibility of drawing many samples from the population to create a sampling distribution. The bootstrap procedure approximates the sampling distribution by repeatedly drawing samples and calculating the statistics from one original sample.

MATERIALS AND METHODS

Robust Regression

Robust regression is an alternative to Ordinary Least Squares OLS that can be appropriately used when there is evidence that the distribution of the error term is non-normal (heavy-tailed) and/or there are outliers that affect the regression equation (Ryan, 1993). A least squares method weights each observation equally in getting parameter estimates, whereas robust methods enable the observations to be weighted unequally (Draper and Smith, 1998). In matrix notation, the linear regression model is given by:

$$Y = X\beta + e$$

Where, for a sample of size n, y is the (n ×1) vector containing the values for the response variable, X is the (n × p) matrix containing the values for the P explanatory variables, and e is the (n ×1) vector containing the error terms. The (p ×1) vector β contains the unknown regression parameters. The vector of parameters estimated by OLS is then:

$$\hat{\beta}_{ols} = \arg\min_{\beta} \sum_{i=1}^{n} e_i\left(\hat{\beta}\right) \qquad i=1,2,\ \ldots,\ n$$

Where $e_i\left(\hat{\beta}\right)$ is the vector of residuals

$$e_i\left(\hat{\beta}\right) = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_0 x_{i2} - \ldots - \hat{\beta}_0 x_{ip}$$

Robust Wild Bootstrap Technique (WBootMM-GM6-Liu)

Wu (1986) noted that the objective of wild bootstrap is to estimate the standard errors of estimates that under heteroscedasticity are asymptotically correct. The drawback of the wild bootstrap is that the estimates of the standard errors become high in the presence of outliers. The wild bootstrap based on the MM estimator denoted as WBootMM-Liu is therefore adopted by Rana et al., (2012) further into a wild bootstrap algorithm. However, this estimator cannot adequately handle high leverage points (HLPS) because the MM estimator is robust to outliers in the y coordinate (Yohai, 1987). It is now evident that the GM6 is robust to high leverage points Ayinde et al., (2015). Osama and Dallatu (2021) incorporated the MM-GM6 estimator denoted as WBootMM-GM6-Liu in the wild bootstrap algorithm to downweight outliers in and directions. The algorithm of MM-GM6 wild bootstrap can be summarized as follows:

Step 1. Fit a model $y_i = x_i\beta + \varepsilon_i$ by using the MM estimator to the real data to obtain the robust MM parameters $\hat{\beta}_{MM}$ and then the fitted model is $\hat{y}_i = x_i\hat{\beta}_{MM}$

Step 2. The residuals of the MM estimate are obtained as $\hat{\varepsilon}_i^{MM} = y_i + \hat{y}_i$. Then, assign

DOI: 10.56892/bima.v8i2B.737

the weight of GM6 to each residual $\hat{\varepsilon}_i^{MM}$ to get new weighted residual $\min\left(1, \frac{x^2_{0.95,p}}{MVE}\right) \times \hat{\varepsilon}_i^{MM}$, where MVE is the minimum-volume ellipsoid.

Step 3. The MM estimate's final weighted residuals denoted as $\hat{\varepsilon}_i^{WMM}$ can be calculated by multiplying the new weight obtained in Step 2 with the value of t_i^* to get $\min\left(1, \frac{x^2_{0.95,p}}{MVE}\right) \times \hat{\varepsilon}_i^{MM} \times t_i^*$.

Step 4. A bootstrap sample (y_i^*, X) is then constructed, where $y_i^* = x_i \hat{\beta}_{MM} + \hat{\varepsilon}_i^{WMM}$ and t_i^* is randomly selected following Liu (1988) procedure.

Step 5. The MM method is then applied to the bootstrap sample (y_i^*, X) and the resulting estimate can be written as $\hat{\beta}^{*R} = \left(X^T X\right)^{-1} X^T y^*$.

Step 6. Steps 3 to 5 were repeated for R times, where R is the bootstrap replications.

Proposed Modified Robust Bootstrap Technique (MWBootMM-GM6-Liu)

In the proposed modified robust bootstrap technique, we begin by highlighting the major alteration in the existing robust bootstrap technique so as to ascertain the magnitude of the changes in comes with. It has been stated in so many literatures that the Minimum Volume Ellipsoid robust estimator has a number of setbacks which includes longer running time, inability to completely do away with outliers and so on. Due to the masking and swamping effects, the Diagnostic Robust Generalized Potential based on Index Set Equality, DRGP (ISE) takes off from Diagnostic Robust Generalized Potential based on Minimum Volume Ellipsoid, DRGP(MVE) and because the running time of ISE is much faster than MVE, Hock and Habshah (2016). Monte Carlo simulation study and numerical data indicate that DRGP (ISE) works excellently to detect the actual high leverage points and reduce masking and swamping effects in a linear model. The algorithm of the proposed modified MM-GM6 wild bootstrap is as follows:

Step 1. Fit a model $y_i = x_i\beta + \varepsilon_i$ by using the MM estimator to the real data to obtain the robust MM parameters $\hat{\beta}_{MM}$ and then the fitted model is $\hat{y}_i = x_i\hat{\beta}_{MM}$

Step 2. The residuals of the MM estimate are obtained as $\hat{\varepsilon}_i^{MM} = y_i + \hat{y}_i$. Then, assign the weight of GM6 to each residual $\hat{\varepsilon}_i^{MM}$ to get new weighted residual $\min\left(1, \frac{x^2_{0.95,p}}{ISE}\right) \times \hat{\varepsilon}_i^{MM}$, where ISE is the Index Set Equality.

Step 3. The MM estimate's final weighted residuals denoted as $\hat{\varepsilon}_i^{WMM}$ can be calculated by multiplying the new weight obtained in Step 2 with the value of t_i^* to get $\min\left(1, \frac{x^2_{0.95,p}}{ISE}\right) \times \hat{\varepsilon}_i^{MM} \times t_i^*$.

Step 4. A bootstrap sample (y_i^*, X) is then constructed, where $y_i^* = x_i \hat{\beta}_{MM} + \hat{\varepsilon}_i^{WMM}$ and t_i^* is randomly selected following Liu (1988) procedure.

Step 5. The MM method is then applied to the bootstrap sample (y_i^*, X) and the resulting estimate can be written as $\hat{\beta}^{*R} = \left(X^T X\right)^{-1} X^T y^*$.

DOI: 10.56892/bima.v8i2B.737

Step 6. Steps 3 to 5 were repeated for R times, where R is the bootstrap replications.

Minimum Volume Ellipsoid

The Diagnostic Robust Generalized Potential is a traditionally used measure for detecting high leverage points. The minimum volume ellipsoid (MVE), introduced by (Rousseeuw, 1985) was the first high-breakdown robust estimator of multivariate location and scatter that has come to be regularly used in practice. The MVE became popular thanks to its high resistance to outliers, which makes it a reliable tool for outlier detection, and the widely available, user-friendly implementations of its computational algorithm. However, the calculation of MVE involves a lot of computational effort. Due to this, the calculation of DRGP based on RMD-MVE takes too much computing time. Midi, Ramli and Imon (2009) proposed Diagnostic Robust Generalized Potential based on Minimum Volume Ellipsoid (DRGP(MVE)) for detecting HLPs.

Index Set Equality

The Index Set Equality (ISE) which is another new technique from fast MCD (Salleh 2013) is used as an alternative to MVE or MCD. ISEs' running time is very fast because the algorithm of ISE only takes into account a comparison of two index set.

The following steps illustrate the computation of ISE.

Step 1: Choose arbitrarily h observations from a dataset to be included in the subsample donated as H_{old} where $h = \frac{n+p+1}{2}$ and p is the number of independent variables (Rousseeuw andamp; Driessen, 1999).

Let $I_{old} = \left(\pi_1^{old}, \pi_2^{old}, \ldots, \pi_h^{old}\right)$ be the index set for H_{old}

Step 2: Compute the p-dimensional mean vector $\overline{T}\ H_{old}$ and the (pxp) covariance matrix of C_{old} from the subset H_{old}.

Step 3: Compute the squared Mahalanobis Distance for each observation, as

$$d_{old}^2(i) = \left(t_i - \overline{T_{H_{old}}}\right) C_{H_{old}}^{-1} \left(t_i - \overline{T_{H_{old}}}\right)$$
for 1,2,..,n.

Step 4: Arrange $d_{old}^2(i)$ in increasing order,

$$d_{old}^2(\pi(1)) \le d_{old}^2(\pi(2)) \le d_{old}^2(\pi(3)) \le \ldots \le d_{old}^2(\pi(n))$$

Where π is a permutation equal to $\{1,2,\ldots,n\}$.

Step 5: The first h items that correspond to the smallest $d_{old}^2(i)$ will be placed in set $H_{New} = \{t_{\pi(1)}, t_{\pi(2)}, t_{\pi(3)}, \ldots, t_{\pi(h)}\}$. Then list new index set, as

$$I_{New} = \left(\pi_{(1)}^{new}, \pi_{(2)}^{new}, \pi_{(3)}^{new}, \ldots, \pi_{(h)}^{new}\right)$$

The DRGP(ISE) consists of two steps, whereby in the first step:

Step I: The suspected HLPs are determined using RMD based on ISE.

Step II: The suspected HLPs will be placed in the 'D' set and the remaining in the 'R' set. The generalized potential (p_i) is employed in the second step to check all the suspected HLPs; those possessing a low leverage point will be put back to the 'R' group. This technique is continued until all points of the 'D' group have been checked to confirm whether they can be referred to as HLPs. The generalized potential is defined as follows:

$$\overline{p}_i = \begin{cases} h_i^{(-D)} & \\ \dfrac{h_i^{(-D)}}{1 - h_i^{(-D)}} & for\ i \in D \\ & for\ i \in R \end{cases}$$

The Cut-off point for DRGP is given by $cdi = median(\bar{p}_i) + 3Q_n(\bar{p}_i)$

Q_n, a pairwise order statistic for all distance proposed by (Rousseeuw and Croux, 1993), is employed to improve the accuracy of the identification of HLPs and is given by $Q_n = c\{|X_i - X_j|; < j(k)\}$, where $k = {}^hC_2 \approx {}^hC_2/4$ and $h = \lfloor n/2 \rfloor + 1$.

They make use of c = 2.2219, as this value will provide a consistent estimator for Gaussian data. If some identified $\bar{p}_i$ did not exceed cdi then the case with the least $\bar{p}_i$ will be returned to the estimation subset for re-computation of $\bar{p}_i$. The values of generalized potential based on the final 'D' set is the DRGP(ISE) represented by $\bar{p}_i$ and the 'D' points will be declared as HLPs.

Simulation Study

In this section, a simulation study is carried out based on the Monte Carlo procedure to investigate the performance of the proposed method denoted as MWBootMM-GM6-Liu in the presence of both heteroscedasticity and high leverage points. In this paper, we consider a multiple linear regression model with two explanatory variables and different sample sizes of 50, 100, and 150. According to Liu (1988), the design of a heteroscedastic model can be written as:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \delta_i \varepsilon_i$$

Where x_{1i} and x_{2i} are generated from U(0,1) for all the sample sizes. The parameters β_0, β_1, and β_2 are set equal to one as the true parameters of this model, and the generation function of heteroscedasticity is $\delta_i^2 = \exp(\theta_1 x_{1i} + \theta_1 x_{2i})$, where θ_1 is to be 0.4. In this paper, the heteroscedasticity's level $\xi = \frac{\max(\delta_i^2)}{\min(\delta_i^2)} = 4.\varepsilon_i's$ where the error term generated from (0,1) for the clean data. For 5% and 10% HLPS, the 95% and 90% of $\varepsilon_i's$ were generated from (0,1) and the 5% and 10% were generated from (0,20). The simulation for each sample size involves a total of 500000 replications with 1000 replications and 500 bootstrap samples each. This simulation was performed based on the procedure of Cribari-Neto and Zarkos (1999) and Furno (1997). The five estimation methods such as WBootOLS, WBootLiu, WBootMM-Liu, WBootMM-GM6-Liu and MWBootMM-GM6-Liu were then applied to the simulated data. The outcomes of simulation study are summarized in following tables.

Table 1: Standard Errors of the WBootOLS, WBootLiu, WBootMM-Liu, WBootMM-GM6-Liu, and MWBootMM-GM6-Liu Estimates

% outliers	Coeff	WBootOLS	WBootLiu	WBootMM-Liu	WBootMM-GM6-Liu	MWBootMM-GM6-Liu
Sample Size n= 50						
0%	β_0	1.7601	1.8706	2.2742	1.6005	1.2834
	β_1	2.1034	2.0324	2.2074	1.9803	1.9001
	β_2	2.4202	2.3059	2.1093	2.0989	1.9201
5%	β_0	7.3955	5.6802	6.1054	1.8219	1.2039
	β_1	8.4640	7.0022	7.0215	1.0934	0.9345
	β_2	6.9984	4.3024	5.0238	1.8309	1.03845
10%	β_0	8.1034	5.9031	5.2948	1.0294	1.9573
	β_1	9.4596	7.4946	6.9832	1.0586	0.9583
	β_2	9.4750	5.8920	5.4013	1.6747	0.9868
Sample Size n= 100						
0%	β_0	1.3840	1.4033	1.1984	0.9855	0.8792
	β_1	1.1938	1.3059	1.3895	0.9430	0.4976
	β_2	1.3019	1.5015	1.4903	1.0856	1.0096
5%	β_0	4.9585	5.0034	3.9850	1.4384	1.0348
	β_1	5.6982	5.9658	4.8591	1.8475	1.3049
	β_2	3.9985	4.3059	3.9485	1.4112	1.1012
10%	β_0	5.3295	5.5956	3.5736	1.2049	0.8874
	β_1	5.3958	5.7464	4.0193	1.9482	1.3048
	β_2	4.9358	5.3029	4.1298	1.7593	1.1093
Sample Size n= 150						
0%	β_0	0.5492	0.5829	0.4559	0.3304	0.2310
	β_1	0.7349	0.7294	0.6639	0.3989	0.3039

	β_2	0.5856	0.5832	0.4723	0.2934	0.2094
5%	β_0	3.5039	3.4948	3.6029	1.5938	0.9485
	β_1	3.7039	3.4923	3.3829	1.9485	0.8437
	β_2	2.9094	3.0011	3.2039	1.0293	0.9934
10%	β_0	3.0958	2.8450	2.9938	1.5730	1.0394
	β_1	4.0112	3.8576	3.4923	1.3928	1.0103
	β_2	3.2335	2.9859	2.7450	1.7394	1.4019

The following figures shows the effect of High Leverage Points HLPs on the standard errors of the parameter estimates. It is obvious from the plots that the standard errors of the parameter estimates of the proposed MWBootMM-GM6-Liu outperforms other methods across all percentages of HLPs because it has the smallest standard errors from the results obtained from the analysis of the simulated data.

Bima Journal of Science and Technology, Vol. 8(2B) July, 2024 ISSN: 2536-6041

DOI: 10.56892/bima.v8i2B.737

Table 2: Bias of the WBootOLS, WBootLiu, WBootMM-Liu, WBootMM-GM6-Liu, and MWBootMM-GM6-Liu Estimates

% outliers	Coeff	WBooOLS	WBootLiu	WBootMM-Liu	WBootMM-GM6-Liu	MWBootMM-GM6-Liu
Sample Size n= 50						
0%	β_0	0.2849	0.2578	0.1178	0.0957	0.0248
	β_1	0.2930	0.3067	0.1015	0.0829	0.0492
	β_2	0.3011	0.2856	0.0894	0.0583	0.0094
5%	β_0	1.2740	0.3560	0.1029	0.0830	0.0175
	β_1	0.1924	0.2032	0.1102	0.0928	0.0303
	β_2	0.2169	0.2265	0.0932	0.0753	0.0156
10%	β_0	3.1092	0.2767	0.1304	0.0793	0.0184
	β_1	0.3216	0.1796	0.1394	0.0674	0.0062
	β_2	0.3401	0.2340	0.1491	0.0923	0.0316
Sample Size n= 100						
0%	β_0	0.3012	0.2749	0.2019	0.0592	0.0193
	β_1	0.2948	0.3002	0.1649	0.0481	0.0201
	β_2	0.4293	0.4029	0.2049	0.0182	0.0095
5%	β_0	1.5859	0.6928	0.3012	0.0239	0.0102
	β_1	0.2940	0.2702	0.2015	0.0604	0.0203
	β_2	0.2954	0.2270	0.1938	0.0490	0.0192
10%	β_0	2.9650	0.8363	0.3029	0.0485	0.0194
	β_1	0.2948	0.3049	0.2384	0.0183	0.0059
	β_2	0.3954	0.3592	0.1928	0.0119	0.0045
Sample Size n= 150						

0%	β_0	0.4968	0.3950	0.3001	0.1039	0.0449
	β_1	0.2847	0.2592	0.1849	0.0939	0.0149
	β_2	0.3844	0.3982	0.2914	0.1039	0.0392
5%	β_0	1.1923	0.5920	0.2938	0.1495	0.0394
	β_1	0.3825	0.3602	0.2948	0.1384	0.0283
	β_2	0.2948	0.3003	0.1734	0.1283	0.0730
10%	β_0	3.4920	0.7393	0.4029	0.1938	0.0293
	β_1	0.2845	0.2843	0.2004	0.0945	0.0085
	β_2	0.4965	0.3849	0.2741	0.1380	0.0394

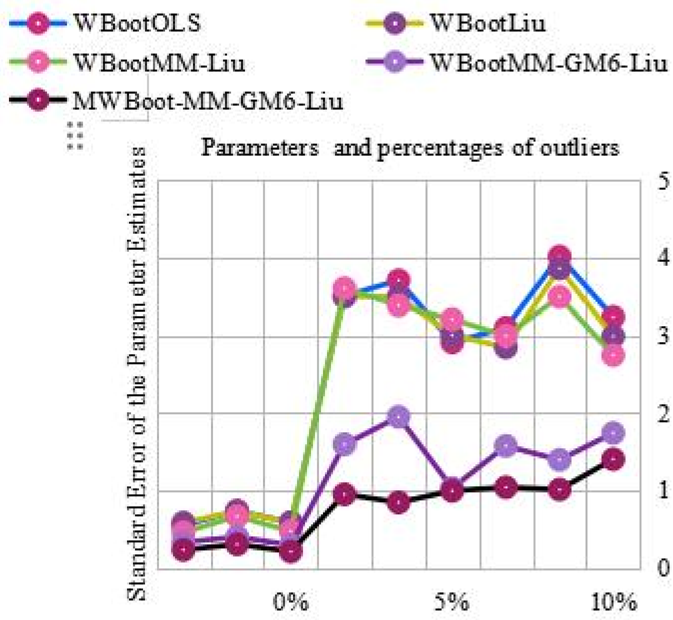

Figure 3: The Effects of 0%, 5%, 10% HLPs on the Standard Errors of the Parameter Estimates when Sample Size n = 150.

DOI: 10.56892/bima.v8i2B.737

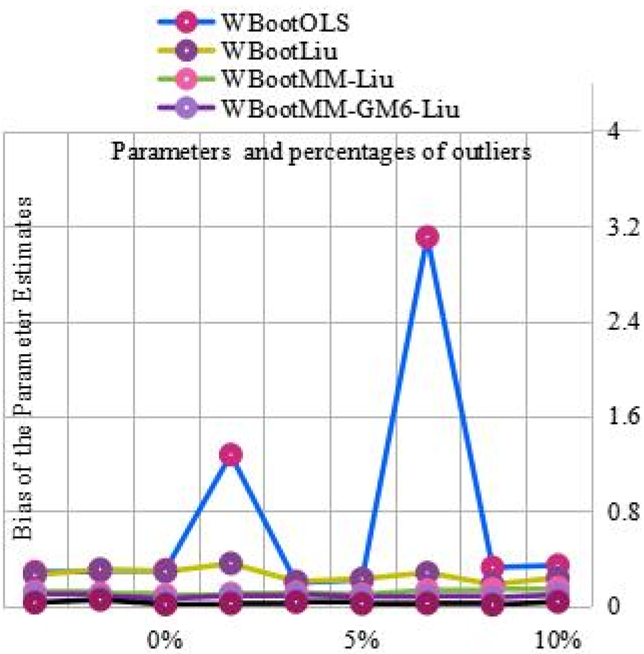

Figure 4: The Effects of 0%, 5%, 10% HLPs on the Bias of the Parameter Estimates when Sample Size n = 5.

The following figures shows the effect of High Leverage Points HLPs on the bias errors of the parameter estimates. It is obvious from the plots that the bias errors of the parameter estimates of the proposed MWBootMM-GM6-Liu outperforms other methods across all percentages of HLPs because it has the smallest bias errors from the results obtained from the analysis of the simulated data.

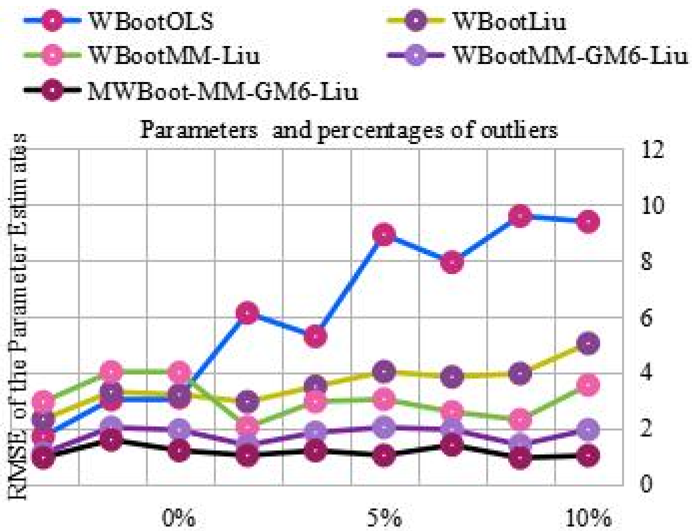

Figure 5: The Effects of 0%, 5%, 10% HLPs on the Bias of the Parameter Estimates when Sample Size n = 100.

DOI: 10.56892/bima.v8i2B.737

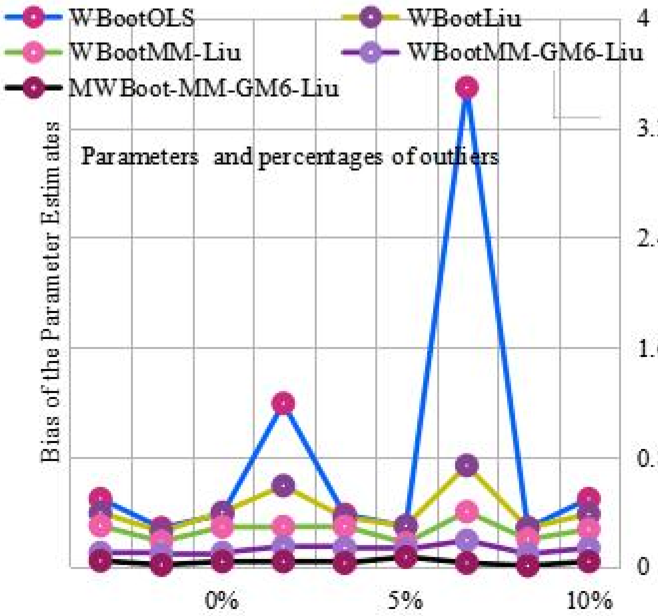

Figure 6: The Effects of 0%, 5%, 10% HLPs on the Bias of the Parameter Estimates when Sample Size n = 150.

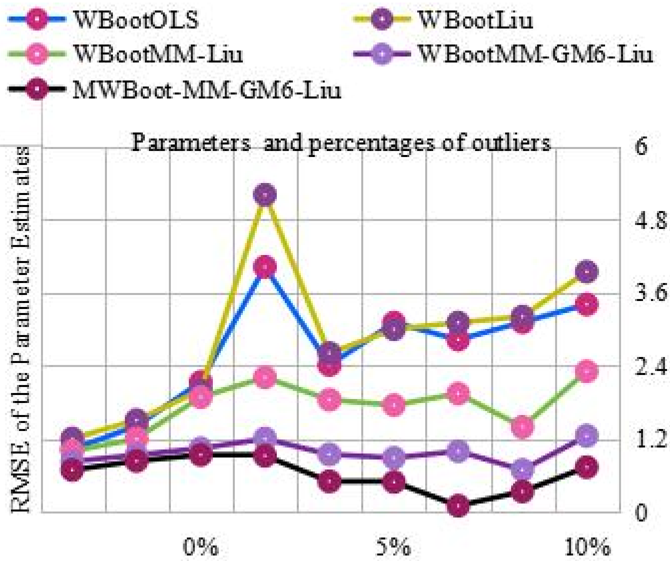

Figure 7: The Effects of 0%, 5%, 10% HLPs on the RMSE of the Parameter Estimates when Sample Size n = 50.

Table 3: RMSE of the WBootOLS, WBootLiu, WBootMM-Liu, WBootMM-GM6-Liu, and MWBootMM-GM6-Liu Estimates

% outliers	Coeff	WBooOLS	WBootLiu	WBootMM-Liu	WBootMM-GM6-Liu	MWBootMM-GM6-Liu
Sample Size n= 50						
0%	β_0	1.7032	2.3202	2.9475	1.1039	0.9487
	β_1	3.0295	3.3029	4.0196	2.0293	1.5853
	β_2	3.1002	3.2093	3.9965	1.9384	1.2049
5%	β_0	6.1293	2.9485	2.0395	1.3938	1.0291
	β_1	5.2934	3.4985	2.9548	1.8491	1.2029
	β_2	8.9234	4.0293	3.0395	2.0324	1.0293
10%	β_0	7.9284	3.8394	2.5850	1.9485	1.4023
	β_1	9.5876	3.9585	2.3049	1.4039	0.9384
	β_2	9.3845	5.0394	3.5494	1.9484	1.0239
Sample Size n= 100						
0%	β_0	1.0383	1.2029	1.0029	0.8329	0.6921
	β_1	1.4039	1.5029	1.2019	0.9384	0.8405
	β_2	2.130	1.9903	1.8802	1.0482	0.9406
5%	β_0	4.0193	5.2030	2.2039	1.2039	0.9248
	β_1	2.4094	2.6029	1.8370	0.9482	0.5056
	β_2	3.1039	2.9948	1.7491	0.8913	0.4958
10%	β_0	2.8239	3.1039	1.9384	1.0003	0.1049
	β_1	3.1103	3.2034	1.3928	0.6938	0.3394
	β_2	3.4029	3.9384	2.3029	1.2495	0.7390
Sample Size n= 150						

0%	β_0	0.9371	0.9659	0.7928	0.4029	0.1093
	β_1	0.9475	0.9530	0.8840	0.5019	0.2938
	β_2	0.8374	0.8723	0.7292	0.4102	0.1928
5%	β_0	2.5920	2.1038	1.2039	0.6039	0.2019
	β_1	3.0029	2.4029	1.6919	0.7039	0.2120
	β_2	1.9284	1.6390	1.0024	0.5948	0.1029
10%	β_0	1.7793	1.7392	1.1039	0.4029	0.0928
	β_1	2.3045	1.3921	1.0293	0.9001	0.3019
	β_2	4.0010	2.3032	1.4038	0.9034	0.3837

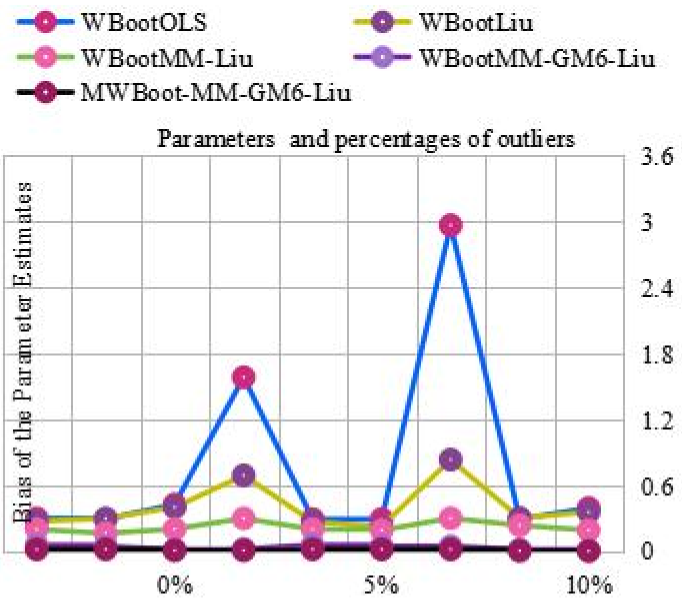

Figure 8: The Effects of 0%, 5%, 10% HLPs on the RMSE of the Parameter Estimates when Sample Size n = 100.

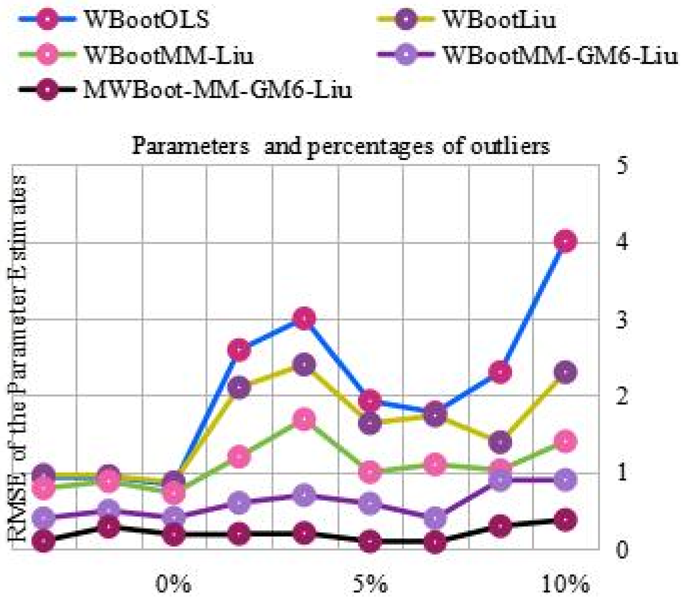

Figure 9: The Effects of 0%, 5%, 10% HLPs on the RMSE of the Parameter Estimates when Sample Size n = 150.

The following figures shows the effect of High Leverage Points HLPs on the Root Mean Square Errors RMSEs of the parameter estimates. It is obvious from the plots that the RMSEs of the parameter estimates of the proposed MWBootMM-GM6-Liu outperforms other methods across all percentages of HLPs because it has the smallest RMSEs from the results obtained from the analysis of the simulated data.

CONCLUSION

In this paper, we modified the Wild Robust Bootstrap technique model by replacing the existing estimator i.e Minimum Volume Ellipsoid with a faster and more efficient estimator i.e the Index Set Equality. The modified model was then compared with the existing Wild Robust Bootstrap model and other existing models and it was discovered that the modified Wild Robust Bootstrap technique model gave a more accurate estimates under different conditions as well as different sample sizes. The Wild Robust Bootstrap technique model was fit to handle linear regression problems in the presence of high leverage points, outliers and heteroscedeisca. Root Mean Square Error (RSME) level was studied across multiple datasets and models and it is found to be least in the case of the modified robust bootstrap model. Bias test also shows that the modified robust bootstrap technique is more accurate because it has less bias when tested on three different levels of outliers.

It is concluded in the results of the analysis that the modified wild robust bootstrap technique (MWBoot-MM-GM6-Liu) is the most accurate and has an added advantage of low running time.

REFERENCES

Aba, D. (2021). *Parametric bootstrapping in a generalized extreme value regression model for binary response*. Journal of Statistical Modeling, 3(1) 1–11.

Adnan, B., Rasheed, B.A., Saffar, S.E, and Pati, K.D. (2015). Performance of Robust Wild Boostrap Estimation of Linear Model in the Presence of Outliers and Heteroscedasticity Errors. Journal of advanced research in applied mechanics, 8(1), 13-31.

Aguinis, H., Gottfredson, R.K., and Joo, H. (2013). Best-practice recommendations for defining, identifying, and handling outliers. *Organizational Research Methods, 16*(2), 270-301.

Alfons, A., Ateş, N.Y., and Groenen, P.J.F. (2021). A Robust Bootstrap Test for Mediation Analysis. *Sage Journals, 21(3).*

Alguraibawi, M., Midi, H., and Imon A.H.M.R. (2015). *A New Robust Diagnostic Plot for Classifying Good and Bad High Leverage Points in a Multiple Linear Regression Model. Hindawi Publishing Corporation Mathematical Problems in Engineering Volume 2015, Article ID 279472, 12 pages.*

Amado, C., Bianco, A.M., Boente, G., and Pires, A.M. (2014). Robust bootstrap: An alternative to bootstrapping robust estimators. *Revstat Statistical Journal*, *12*(2), 169–197.

Bagheri, A., and Midi, H. (2012). *On the Performance of the Measure for Diagnosing Multiple High Leverage Collinearity-Reducing Observations*. Mathematical Problems in Engineering, Hindawi, vol. 2012, 1-16.

Cao, J., Zhang, X., Yang, G., and Zou, X. (2018). Robust control of pressure for LNG carrier cargo handling system via mirror-mapping approach. *Complexity*, Hindawi, vol. 2018, 1-11.

Croux, C., Flandre, C., and Haesbroeck, G. (2002). The breakdown behavior of the maximum likelihood estimator in the logistic regression model. *Statistics and Probability Letters*, 377-386.

Dalatu, P.I. (2020). *The implementation of wild bootstrap based on mm-gm6 estimator in the presence of heteroscedastic errors and high leverage points. Nigeria Journal of Engineering Science and Technology Research Vol. 6, No. 1, 2020(99-103).*

Efron, B. (2003). *Second Thoughts on the Bootstrap*. Journal of Statistical Science, 18 (2) 135 - 140.

Emmanuel, S.A., Kayode, A., and Lukman, A.F. (2016). *A comparative study of some robust ridge and liu estimators. Science World Journal, 11(4).*

Muthukrishnan, R., Radha, M. (2010). M-Estimators in Regression Models. Journal of Mathematics Research, 2(4). www.ccsenet.org/jmr.

Osama, A.A., and Dallatu, P.I. (2021). On the performance of wild bootstrap based on mm-gm6 estimator in the presence of heteroscedastic errors and high leverage point. Pakistan Journal of Statistics, 37(2), 121-134.