Journal of Science Research and Reviews
PRINT ISSN: 1595-9074
E-ISSN: 1595-8329
DOI: https://doi.org/10.70882/josrar.2026.v3i6.99
Homepage: https://josrar.esrgngr.org
Original Research Article

Assessment of the Effects of Influential Observations on Heteroscedasticity Detection in Multiple Linear Regression

Muhammad Sani[1*], Aliyu Usman Goga[2] and Mustapha Isyaku[3]

[1, 2]Department of Statistics, Federal University Dutsin-Ma Katsina State, Nigeria.
[3]Department of Mathematics, Federal University Dutsin-Ma Katsina State, Nigeria.
*Corresponding Author's email: sanimksoro@gmail.com

KEYWORDS
Outlying observations,
Influential Observations,
Heteroscedasticity,
Linear Regression.

CITATION
Sani, M., Goga, A. U., & Isyaku, M. (2026). Assessment of the Effects of Influential Observations on Heteroscedasticity Detection in Multiple Linear Regression. *Journal of Science Research and Reviews*. https://doi.org/10.70882/josrar.2026.v3i6.99.

ABSTRACT
Influential observations can substantially affect regression diagnostics and lead to misleading conclusions about the presence of heteroscedasticity. This study investigated the effects of bad leverage points (BLPs) and vertical outliers (VOs) on heteroscedasticity detection in multiple linear regression using the classical White test (WT) and robust White test (RWT). Four numerical examples and Monte Carlo simulation experiments involving 2,000 replications were employed to evaluate the behaviour of both tests under different contamination conditions. The simulation experiments considered sample sizes ranging from 20 to 250 and contamination levels of 5% and 10%. Results indicated that introducing BLPs into initially homoscedastic datasets increased the tendency of the classical White test to reject the homoscedasticity hypothesis, while the robust White test generally retained the expected decision. Conversely, introducing VOs into heteroscedastic datasets reduced the ability of the classical White test to detect heteroscedasticity, whereas the robust test continued to reject the null hypothesis. These findings suggest that influential observations can either enhance or mask evidence of heteroscedasticity in conventional regression diagnostics. The study highlights the importance of robust diagnostic procedures when analysing contaminated regression datasets and recommends further evaluation of their empirical rejection rates and statistical properties under alternative contamination mechanisms.

INTRODUCTION

The presence of Influential observations (outlying observations) was identified as another source of heteroscedasticity in multiple linear regression (Alih and Choon, 2015; Rana et al., 2008). Imon and Khan (2003) as well as Habshah et al. (2009) reported that the presence of High Leverage Points (HLPs) in a data set causes the residual variances become heteroscedastic. Imon and Khan (2003) stated that in the presence of influential observations, the error variances become heteroscedastic. However, they did not discuss at length that influential observations may cause a heteroscedastic data set to become non-heteroscedastic and vice-versa. Therefore, it is very significant to investigate the effect of all types of outlying observations (good and bad leverage high points, vertical outlier) in both non-heteroscedastic and heteroscedastic data set as it may cause misleading conclusion about the fitting of a regression model. It is important to highlight that good leverage points do not affect heteroscedasticity and parameter estimation of a linear regression model rather they contribute to the precision of the estimate (Alguraibawi et al., 2015). We

This work is licensed under the Creative Commons Attribution 4.0 International License

define those observations that disrupt the heteroscedasticity pattern of a data set as Heteroscedasticity-Influential Observations (HIO). Those outliers or HLPs that are responsible for causing heteroscedasticity, we call them Heteroscedasticity-Enhancing Observations (HEO). On the other hand, HLPs or outliers that hide heteroscedasticity we call them Heteroscedasticity-Reducing Observations (HRO).

However, this source of heteroscedasticity is not similar to other sources of heteroscedasticity, but it can affect the detection of heteroscedasticity when the classical diagnostic's methods are used (Habshah et. al., 2009; Bagheri and Habshah, 2015). The most widely classical method of heteroscedasticity detection is the White's test (WT) developed by White (1980) which is independent of prior knowledge of the source of heteroscedasticity and normality. The major weakness of WT is that, in the presence of outlying observations it fails to detect heteroscedasticity. Recently, robust White test (RWT) was proposed by Sani et al., (2019) which is powerful and resistance to the effect of all types outlying observations. The RWT was constructed from the weighting method of Fast Improvised Influential Distance (FIID) developed by Habhash et al., (2021).

This study aimed to investigate the effects of bad leverage points and vertical outliers on heteroscedasticity detection in multiple linear regression by comparing the performance of the classical and robust White tests using empirical datasets and Monte Carlo simulations.

MATERIALS AND METHOD

White (1980) developed a test called White test (WT) which is a Lagrange Multiplier (LM) test to detect the presence of heteroscedasticity in linear regression. The test is based on the auxiliary regression analysis. The LM test statistic is the product of the R^2 value and sample size of the auxiliary regression given as,

$$LM = nR^2 \quad (1)$$

The LM test statistics follows a Chi squared distribution (White, 1980) having the degrees of freedom equal to *p*, where *p* is the number of regressors in the auxiliary regression. The null hypothesis (H_0) that the residual variances are constant is rejected if $nR^2 > \chi^2_{(p,0.05)}$ otherwise we fail to reject.

The algorithms for WT are as follows:

Step 1: Employ multiple linear regression model using OLS and obtain the residuals (r_i)

Step 2: Find auxiliary regression by regressing the squared residuals (r_i^2) obtained in Step 1 against the original regressors along with their squares and cross product terms using OLS.

Step 3: Compute the coefficient of determination R^2 from the auxiliary regression, given by:

$$R^2 = \frac{SSR}{SSE+SSR} \quad (2)$$

Where, SSR is the regression sum of squares and SSE is the errors sum of squares of the auxiliary regression.

Step 4: The null hypothesis (H_0) is rejected that the variances of the residual are constant if $nR^2 > \chi^2_{(p,0.05)}$, where *p* is the number of regressors in the auxiliary regression and *n* is the sample size.

Robust White Test (RWT)

Robust White test (RWT) developed by Muhammad et al., (2019) was constructed by replacing some components of the WT with robust alternative to form RWT. It has been observed that the WT needs two times minimization of sum of squares residual using OLS. The OLS was reported to be very sensitive and easily influenced by outlying observations. Therefore, MM-estimates developed by Yohai (1987) was employed which has 95% efficiency and 50% break down point relative to OLS under Gauss-Markov assumptions. In addition, the auxiliary regression here is based on weighted least squares (WLS) which uses the weight of GM-FIID developed by Muhammad at al., (2020). The algorithm of the RWT has been presented by Muhammad et al., (2019). Where the coefficient of determination R_R^2 from the auxiliary regression is given by:

$$R_R^2 = \frac{SSR_R}{SSE_R+SSR_R} \quad (3)$$

SSR_R is the sum of squares regression and SSE_R is the sum of squares errors of the auxiliary regression. Similarly, the null hypothesis (H_0) that the residual variances are constant is rejected if $nR_R^2 > \chi^2_{(p,0.05)}$, where *p* is the number of regressors in the auxiliary regression and *n* is the sample size.

Heteroscedasticity-Enhancing Observations (HEO)

Through a well-referred real homoscedastic data set and Monte Carlo simulation study, the conditions which cause a bad leverage point to become heteroscedasticity-enhancing observation are investigated in linear regression model. For an unusual observation Bad Leverage Points (BLP) to be declared as HEO it must satisfied the following two conditions for both WT and RWT:

Condition 1: In the absence of BLP, WT shows absence of heteroscedasticity, RWT shows absence of heteroscedasticity

Condition 2: In the presence of BLP, WT shows the presence of heteroscedasticity, RWT shows absence of heteroscedasticity

When these two conditions are satisfied, then the BLP is declared as HEO, meaning that the presence of heteroscedasticity is due to the presence of BLP. The classical WT shows absence of heteroscedasticity in the absence of BLP, but indicating heteroscedasticity in their presence. On the other hand, the RWT is not affected by the presence of BLPs. In order to verify that BLP is HEO, the following steps should be employed.

Steps to verify that BLP is HEO are:

Step 1: Identify BLP by plotting FIID classification with slight modification whereby,

An observation is declared as RO if;$|\mathrm{FGSR}| \leq \mathrm{CP}_{\mathrm{FGSR}}$ and $|\mathrm{FGLV}| \leq \mathrm{CP}_{\mathrm{FGLV}}$

An observation is declared GLP if;$|\mathrm{FGSR}| \leq \mathrm{CP}_{\mathrm{FGSR}}$ and $|\mathrm{FGLV}| > \mathrm{CP}_{\mathrm{FGLV}}$

An observation is declared as VO if;$|FGSR| > CP_{FGSR}$ and $|FGLV| \leq CP_{FGLV}$

An observation is declared as BLP if; $|FGSR| > CP_{FGSR}$ and $|FGLV| > CP_{FGLV}$

FGSR	FGLV	
	Vertical Outlier (VO)	Bad Leverage Point (BLP)
	Regular Observations (RO)	Good Leverage Point (GLP)
	Vertical Outlier (VO)	Bad Leverage Point (BLP)

Step 2: Compute WT and RWT in the absence of BLP

Step 3: Compute WT and RWT in the presence of BLP

Step 4: Declare an observation HEO when the two conditions are satisfied.

RESULTS AND DISCUSSION

This section presents numerical examples to investigate the effect of BLPs on a homoscedastic data set. The WT and RWT are applied to each example to evaluate the effect of BLPs. The HEO is then verified.

Education Expenditure Data

The data set contains 50 observations with three explanatory variables. Moreover, it was identified using Fast Improvised Influential Distance (FIID) of Habshah et al., (2021) that observation 49, i.e. Alaska (AK) is bad leverage points (BLP). We apply both classical and robust White tests to this data sets. The residuals versus fitted values of the data without AK and with AK observation are shown in Figure 1 and 2, respectively. We also modified the data to have additional BLP in order to evaluate the effect of multiple BLPs.

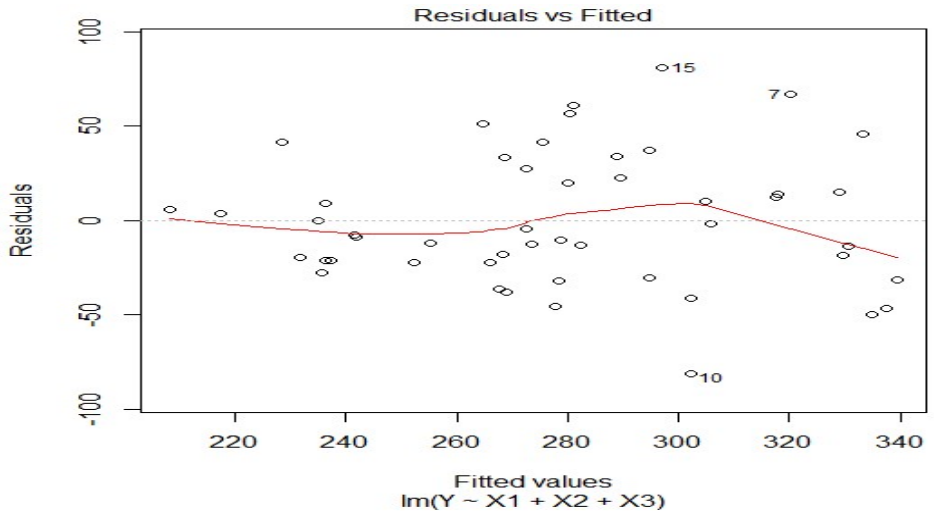


Figure 1: OLS residuals versus fitted values plot without AK observation

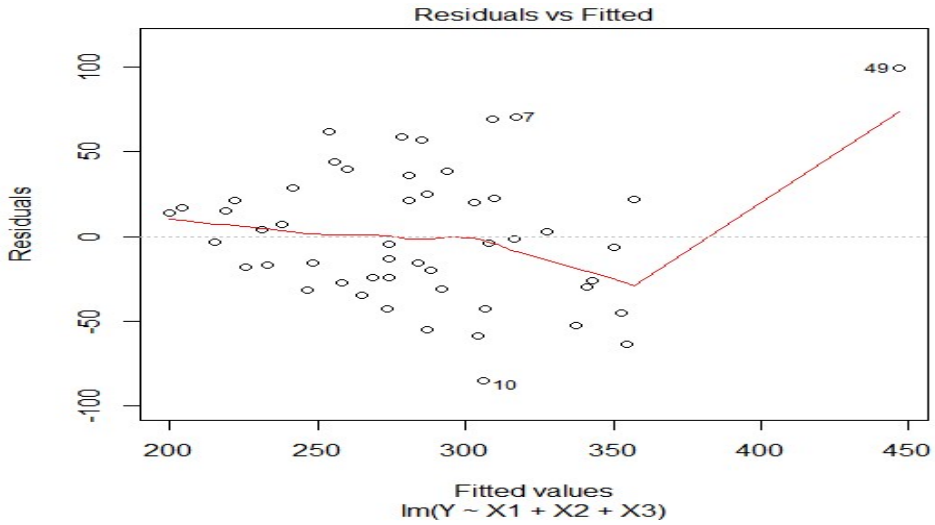


Figure 2: OLS residuals versus fitted values plot with AK observation

Table 1: Education Expenditure Data Heteroscedasticity Diagnostics

Tests	WT		RWT	
	LM = nR^2 (18.307)	***p*-values (0.05)**	**LM = nR^2 (18.307)**	***p*-values (0.05)**
Without AK(BLP)	5.7978	0.1219	5.2922	0.1516
With AK(BLP)	22.7817	4.48e-05	5.4339	0.1426
With AK and (2BLPs)	32.8399	3.48e-07	5.9064	0.1163

Table 1 indicates that introducing the Alaska observation changes the WT p-value from 0.1219 to 4.48 x 10^{-5}, whereas the RWT remains nonsignificant. This suggests that the classical test is sensitive to the identified leverage point. However, it does not independently establish that the underlying errors become heteroscedastic.

Housing Expenditure Data

The data was taken from Pindyck and Rubinfeld (1997). It possesses 20 observations for housing expenditure on four different income groups. As expected, people with higher income have relatively more variation in their expenditures on housing. The residual versus fitted plot of the data was shown in Figure 3 which clearly indicates presence of heteroscedasticity. Two BLPs (observation 16 and 17) were identified in this data as BLP. By removing these 2 BLPs as seen in Figure 4 and apply RWT and WT to detect the heteroscedasticity. We also modified the data to have additional 1 BLP in order to testify that BLP are HEO.

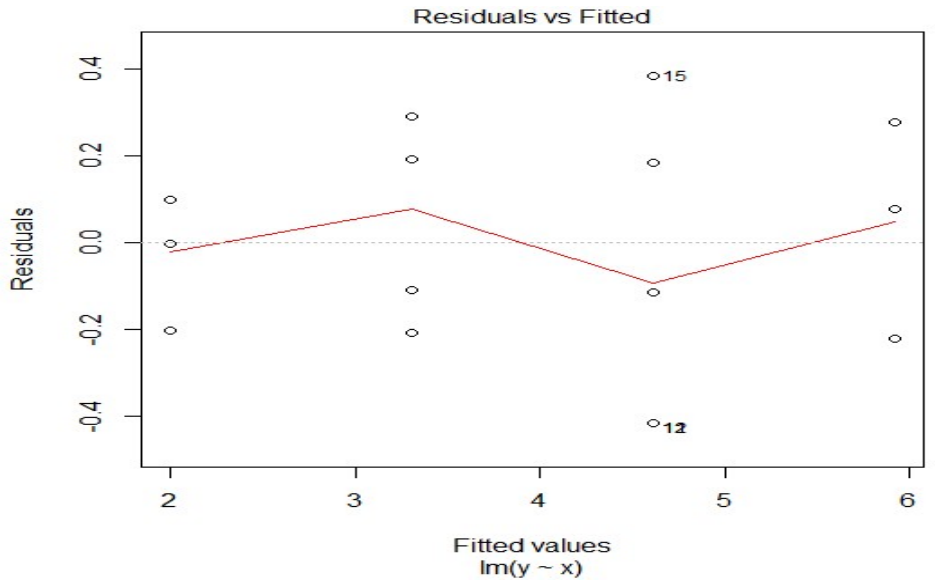


Figure 3: OLS residuals versus fitted values plot without two BLPs

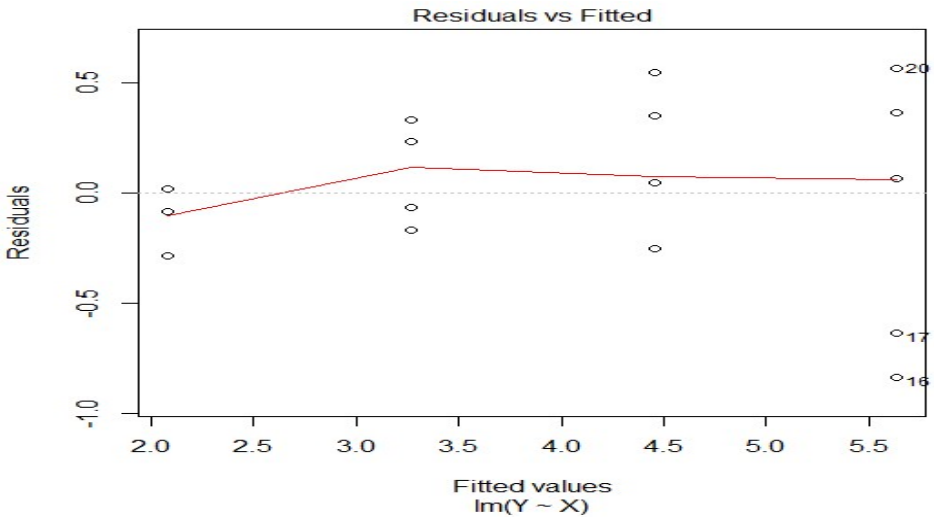


Figure 4: OLS residuals versus fitted values plot with two BLPs

Table 2: Heteroscedasticity Diagnostics for Housing Expenditure Data

Tests	WT		RWT	
	LM = nR^2 (3.841)	*p*-values (0.05)	LM = nR^2 (3.841)	*p*-values (0.05)
Without 2 BLPs	3.1454	0.0761	3.0705	0.0797
With 2 BLPs	7.1892	0.0073	3.7557	0.0664
With 2BLPs and 1 BLP	4.8080	0.0283	3.2617	0.0709

Table 2 similarly reports WT p-values of 0.0761 without BLPs and 0.0073 with BLPs, while RWT remains nonsignificant. From these two real examples, we have not come out with a final conclusion, but a reasonable interpretation up to this point is that BLPs can induce the variance of the errors to be heteroscedastic.

Monte Carlo Simulation

A Monte Carlo simulation experiment was conducted to investigate the effect of BLPs on a homoscedastic data set. In our simulation experiment, we used a simple linear model with three-variables as given in the equation below:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i \quad (4)$$

Where *y* is dependent variable, *X* is the (*n* x 3) vector of independent variables, all the values of *X* and the random errors (ε) were generated from normal distributions. The BLPs contamination was generated by random replacement of a certain percentage of regular observations with a values generated from *N*(10,1) for both *y* and X at 5% and 10% BLPs contamination. The simulation experiment was run for six different sample sizes *n* = 250, 200, 150, 100, 50 and 20. The effect of BLPs on a homoscedastic data was measured based on the *p*-value of the Lagrange Multiplier (LM) for RWT and WT with different contamination level and sample size. The level of significance was set at 0.05. In addition, the average of *p*-value and LM is recorded. The results were presented in Table 3 based on the average of 2,000 replications.

The probability (p-value) of obtaining a test statistic at least as extreme as the one observed from the sample, the p-value is obtained by R package using the following formula,

$$Pvalue = 1 - pchisq(LM, df) \quad (5)$$

Where *df* and LM are the degree of freedom and Lagrange Multiplier respectively.

Table 3: Average Lagrange Multiplier (LM) for WT and RWT in a Simulated Homoscedastic Data

Samples	Contamination Levels	WT		RWT	
		LM (18.307)	*p-value* (0.05)	LM (18.307)	*p-value* (0.05)
n = 20	Without BLPs	5.5423	0.0849	5.2900	0.1905
	5% BLPs	21.7264	0.0130	5.6196	0.1302
	10% BLPs	20.1148	0.0174	6.3921	0.1971
n = 50	Without BLPs	6.7743	0.06540	5.7681	0.1482
	5% BLPs	19.0929	0.0111	6.0550	0.2933
	10% BLPs	18.8878	0.0226	6.7342	0.1157
n = 100	Without BLPs	6.6951	0.1174	5.4209	0.1890
	5% BLPs	18.7821	0.0230	5.8611	0.1730

	10% BLPs	21.4894	0.0169	6.5379	0.1897
n =150	Without BLPs	6.8358	0.1012	5.4441	0.1886
	5% BLPs	23.7780	0.0216	5.9235	0.3409
	10% BLPs	25.4055	0.0079	6.4090	0.5163
n = 200	Without BLPs	6.9337	0.1735	5.4810	0.3282
	5% BLPs	26.6135	0.0011	5.7335	0.4605
	10% BLPs	24.0140	0.0312	6.3793	0.3496
n = 250	Without BLPs	6.9336	0.5517	5.4333	0.4147
	5% BLPs	31.0287	9.99e-16	8.9703	0.1955
	10% BLPs	27.2550	1.16e-07	6.2696	0.4573

Table 3 clearly shows that in the absence of BLPs the RWT and WT indicates the variances of the errors are homoscedastic since all the p-value greater than $\alpha = 0.05$. Looking at 5% and 10% BLP contamination for all sample sizes, the WT shows heteroscedasticity but not for RWT. This implies that, with 5% and 10% of BLPs contamination, the homoscedastic error variances completely turn to heteroscedastic error variances for all the sample sizes considered in this study. However, the RWT has not been affected by the BLPs and still indicates the absence of heteroscedasticity. Therefore, the classical White test becomes more likely to reject homoscedasticity under BLP contamination, whereas RWT remains comparatively stable.

Heteroscedasticity-Reducing Observations (HRO)

The heteroscedasticity-reducing observations (HRO) are those observations that hide or change a heteroscedastic data to a homoscedastic data. In this section, numerical examples and Monte Carlo simulation study are used to show how the vertical outliers (VO) affect or disrupt the heteroscedasticity in a linear regression model. For an unusual observation (VO) to be declared as HRO it must satisfied the following two conditions for both WT and RWT:

Condition 1: In the absence of VO; WT shows presence of heteroscedasticity, RWT shows presence of heteroscedasticity

Condition 2: In the presence of VO; WT show absence of heteroscedasticity, RWT shows presence of heteroscedasticity

When these conditions are satisfied then the VO is declared as HRO. This means the presence of VO hide the heteroscedasticity of a data set. The following steps should be employed to validate that VO reduces heteroscedastic data to homoscedastic.

Step 1*:* Identify VO by plotting FIID classification scheme with slight modification whereby,

A point is declared RO if;$|\mathrm{FGSR}| \leq \mathrm{CP}_{\mathrm{FGSR}}$ and $|\mathrm{FGLV}| \leq \mathrm{CP}_{\mathrm{FGLV}}$

A point is declared GLP if;$|\mathrm{FGSR}| \leq \mathrm{CP}_{\mathrm{FGSR}}$ and $|\mathrm{FGLV}| > \mathrm{CP}_{\mathrm{FGLV}}$

A point is declared VO if;$|FGSR| > CP_{FGSR}$ and $|FGLV| \leq CP_{FGLV}$

A point is declared BLP if; $|FGSR| > CP_{FGSR}$ and $|FGLV| > CP_{FGLV}$

FGSR	FGLV	
	Vertical Outlier(VO)	Bad Leverage Point(BLP)
	Regular Observations(RO)	Good Leverage Point(GLP)
	Vertical Outlier(VO)	Bad Leverage Point(BLP)

Step 2: Compute WT and RWT in the absence of VO

Step 3: Compute WT and RWT in the presence of VO

Step 4: Declare an observation HRO when the two conditions are met.

Numerical Examples

Two empirical examples were considered in this section to investigate the effect of vertical outlier (VO) on a heteroscedastic data set. The WT and RWT are applied to the data set in order to check for the presence of heteroscedasticity.

Hand Grip Strength Data

This data set was taken from Hossain et al. (2012). They studied the relationship between socioeconomic factors, anthropometric measurements and grip strength of 500 male and female healthy volunteers from the medical students, staff and visitors of the University of Malaya Medical Centre between January and April 2010. To determine the hand grip strength, they considered three variables body mass index (BMI), age and height. Sample of 108 women from this data based on their left-hand grip was considered. In this data, FIID plot reveals that two observations (21 and 45) were found to be VOs. Figure 5 indicates the existence of heteroscedasticity without observations 21 and 45. Figure 6 shows absence of

heteroscedasticity in the presence of observations 21 and 45.

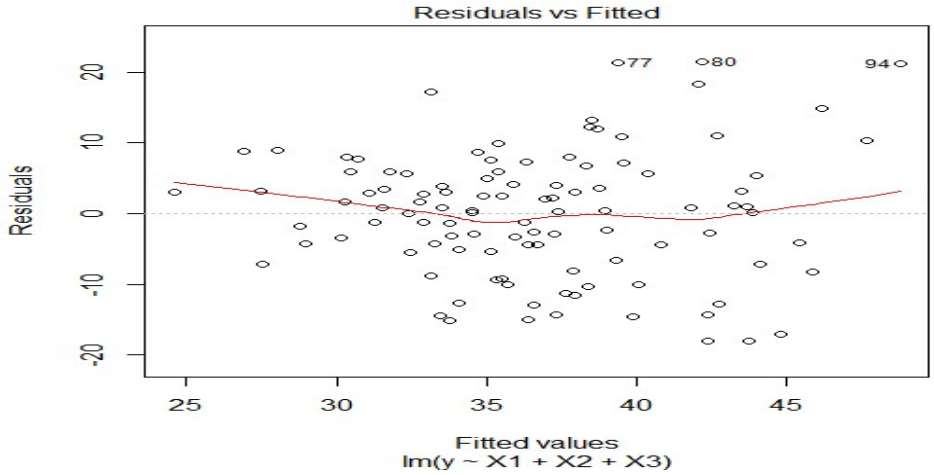


Figure 5: OLS residuals versus fitted values plot of hand grip data without 2 VO

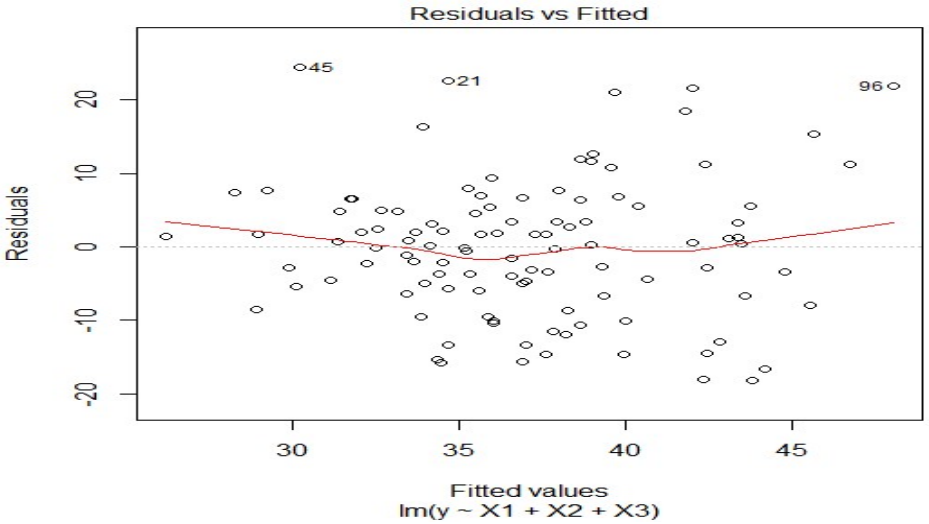


Figure 6: OLS residuals versus fitted values plot of hand grip data with 2 VO

Table 4: Heteroscedasticity Diagnostics for Hand Grip Strength Data

Tests	WT		RWT	
	LM = nR² (18.307)	*p*-values (0.05)	LM = nR² (18.307)	*p*-values (0.05)
Without 2 (VOs)	19.9996	0.0007	20.6590	0.0001
With 2 (VOs)	5.1458	0.1614	19.4424	0.0009

Table 4 shows that VOs are HRO. It can be observed that both WT and RWT tests shows heteroscedasticity in absence of 2VOs. On the other hand, in the presence of 2VOs the WT indicates absence of heteroscedasticity and while RWT still shows presence of heteroscedasticity.

Artificial heteroscedastic data

A heteroscedastic artificial data set was generated. The data has 3 predictor variables ($p = 3$) and 50 sample size that was replicated thrice to form sample size ($n = 150$). Each of the predicator variable was generated as $x \sim N(0,1)$. The error term follows $e_i \sim N(0, \sigma_i^2)$ with the skedastic function defined as $\sigma_i^2 = \exp\{c_1 x_{i1}\}$ (Lima et al., 2009) where the value of $c_1 = 0.35$. The true parameters were set at $\beta_0 = \beta_1 = \beta_2 = \beta_3 = 1$. The data is modified by adding some VOs in order to disrupt the heteroscedasticity pattern of data. We modified the data by incrementing observation 108 and 138 by 10 in the Y-direction in order to create VO. The White test and residual vs fitted plot were employed to check the existence of heteroscedasticity in both the original and modified data sets. Figure 7 indicates the evidence of heteroscedasticity due to the systematic pattern of the residuals and the funnel shape produce.

However, Figure 8 shows absence of heteroscedasticity when 2VOs are present.

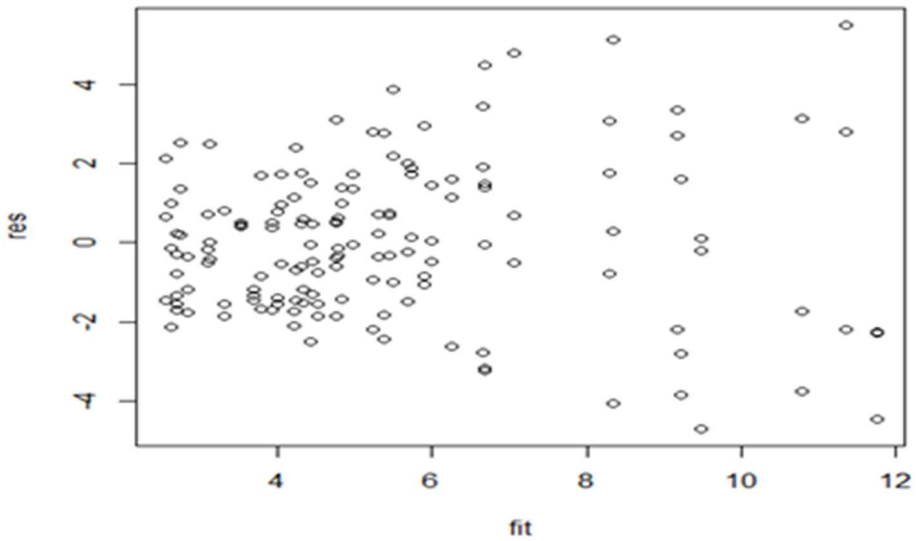


Figure 7: OLS residuals versus fitted values plot of artificial data (without 2VO)

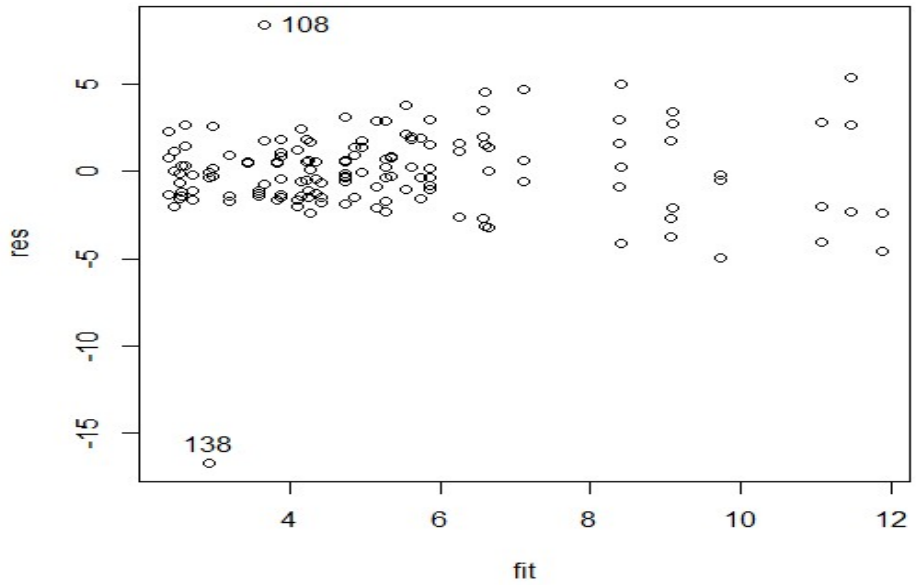


Figure 8: OLS residuals versus fitted values plot of modified artificial data (with 2VO)

Table 5: Heteroscedasticity Diagnostics for Artificial Heteroscedastic Data

Tests	**WT**		**RWT**	
	LM = nR² (18.307)	***p*-values (0.05)**	**LM = nR² (18.307)**	***p*-values (0.05)**
Without 2VOs	40.3217	9.11e-09	32.3815	4.35e-07
With 2VOs	0.5092	0.9169	33.2602	2.84e-07

The results of Table 5 clearly indicate that the two VOs are HRO since in the absence of VOs, both RWT and WT shows heteroscedastic errors but not for WT when VOs are present. Again, we are not pursuing the analysis of real data to a final conclusion. Nonetheless, up to this stage, we may say the VOs are the cause of reducing heteroscedasticity level on the WT results.

Monte Carlo Simulation

Employing Lima et al. (2009) simulation procedure, we consider a linear relation $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i,\ i = 1,2,\dots,n$. The three independent variables were generated from standard normal distribution, in which the true parameters were set at $\beta_0 = \beta_1 = \beta_2 =$

$\beta_3 = 1$, and $\varepsilon_i \sim N(0, \sigma_i^2)$. Six sample sizes n = 10, 25, 50, 75, 100 and 125 were replicated twice to form sample sizes of 20, 50, 100, 150, 200 and 250, respectively. The skedastic function is defined as $\sigma_i^2 = \exp\{c_1 x_{i1}\}$ (Lima et al., 2009). The contamination was done by replacement of some observation with vertical outliers (VOs) at different contamination levels (5% and 10%). The contamination was created by inflating the regular y-observation with $N(10, 1)$ for all the sample sizes. To assess the effect of VOs on a heteroscedastic data, we consider average p-value of the Lagrange Multiplier (LM) of both WT and RWT. To compute the p-value, the level of significance is 0.05.The average LM of WT and RWT was recorded based on the average of 2,000 replications as presented in Table 6.

Table 6 shows that the presence of VOs can turn a heteroscedastic data to a homoscedastic data for WT test. For both 5% and 10% VOs contamination, the simulated heteroscedastic data completely turn to homoscedastic. The RWT indicates the existence of heteroscedasticity irrespective of sample sizes and contamination levels. This clearly indicates the robustness of RWT against the effect of VOs.

Table 6: Average Lagrange Multiplier (LM) for WT and RWT in a simulated Heteroscedastic Data, λ = 5.08

Samples	Contamination Levels	WT		RWT	
		LM (18.307)	*p-value* (0.05)	LM (18.307)	*p-value* (0.05)
	Without VOs	20.8941	0.0117	21.4576	0.0078
n = 20	5% VOs	8.8275	0.3452	22.2118	0.0104
	10% VOs	6.9412	0.4655	25.1537	0.0036
n = 50	Without VOs	19.9551	0.0297	18.5188	0.0468
	5% VOs	7.5382	0.4738	21.1988	0.0197
	10% VOs	5.1930	0.5779	22.5867	0.0123
n = 100	Without VOs	19.3943	0.0355	22.4009	0.0131
	5% VOs	9.9364	0.3460	23.3325	0.0095
	10% VOs	6.9646	0.5287	24.4334	0.0065
n =150	Without VOs	21.4894	0.0179	23.5761	0.0088
	5% VOs	8.7827	0.3528	23.1182	0.0103
	10% VOs	5.4941	0.4558	26.5371	0.0030
n = 200	Without VOs	19.8585	0.0306	24.4418	0.0065
	5% VOs	9.7708	0.3432	24.9358	0.0054
	10% VOs	8.4058	0.3892	26.4904	0.0031
n = 250	Without VOs	22.9776	0.0108	23.8107	0.0081
	5% VOs	11.7905	0.2993	26.3359	0.0033
	10% VOs	9.8923	0.3452	26.4576	0.0031

The effects of VOs on heteroscedastic data sets have been investigated. The results of both numerical examples and Monte Carlo simulation show that the presence of VOs can change a heteroscedastic data to a homoscedastic data set. Similarly, when both VOs and BLPs simultaneously exist in either of homoscedastic or heteroscedastic data set, the VOs dominated the BLPs. As such in their presence, the variances of the errors become homoscedastic. Those VOs that disrupt a heteroscedastic data set are referred to as Heteroscedasticity-Reducing Observations (HRO).

CONCLUSION

This study investigated the effects of bad leverage points and vertical outliers on heteroscedasticity detection in multiple linear regression using the classical White test and robust White test. The empirical examples and Monte Carlo simulations indicated that bad leverage points can increase the tendency of the classical White test to reject homoscedasticity, whereas vertical outliers can mask evidence of heteroscedasticity in contaminated datasets. The robust White test generally demonstrated greater stability under the investigated contamination conditions. These findings support the importance of identifying influential observations before interpreting regression diagnostic results. Therefore, the results of the simulation study and numerical examples signifies that the Heteroscedasticity Influential Observations (HIOs) can be classified as Heteroscedasticity-Enhancing and Heteroscedasticity-Reducing Observations denoted by HEO and HRO respectively

REFERENCES

Alguraibawi, M., Midi, H., and Imon, A. H. M. (2015). A new robust diagnostic plot for classifying good and bad high leverage points in a multiple linear regression model. *Mathematical Problems in Engineering*, 1:1-12

Alih, E., and Choon, H. O. (2015). An outlier-resistant test for heteroscedasticity in linear models. *Journal of Applied Statistics*, 42(8):1617-1634.

Bagheri, A. and Habshah M. (2015). "Diagnostic plot for the identification of high leverage collinearity-influential observations." *SORT* 39(1): 51-70.

Habshah M., Norazan MR and Rahmatullah Imon AHM (2009). The performance of diagnostic-robust generalized potentials for the identification of multiple high leverage points in linear regression.*Journal of applied statistics* 36(5): 507-520.

Habshah Midi, Muhammad Sani Shelan Saied Ismaeel (2021) Fast Improvised Influential Distance for the Identification of Influential Observations in Multiple Linear Regression. *Sains Malaysiana*, 50 (7) (2021): 2085-2094 http://doi.org/10.17576/jsm-2021-5007-22

Hossain, M. G., Zyroul, R., Pereira, B. P., and Kamarul, T. (2012). Multiple regression analysis of factors influencing dominant hand grip strength in an adult Malaysian population. *Journal of Hand Surgery (European Volume)*, 37(1), 65-70.

Imon A.H.M.R. and Khan M.A.I. (2003) A solution to the problem of multicollinearity caused by the presence of multiple high leverage points, *Int. J. Stat. Sci*. 2:37–50

Lima, V.M.C., Souza, T, C., Cribari-Neto, F. and Fernandes, G.B. (2009). Heteroskedasticity- robust inference in linear regressions. *Communications in Statistics-Simulation and Computation* 39: 194-206

Pindyck, S.R and Rubinfeld, L.D. (1997). *Econometric Models and EconometricForecasts,* 4th Edition. New York: Irwin/McGraw-Hill.

Rana M.S., H. Midi, and A.H.M.R. Imon, (2008). A Robust Modification of the Goldfeld-Quandt Test for the Detection of Heteroscedasticity in the Presence of Outliers, *Journal of mathematics and Statistics,* 4(4), 2008, pp. 277-283

Sani, M., Habshah M. and Babangida I.B. (2019) Robust White's Test For Heteroscedasticity Detection In Linear Regression *FUDMA journal of science,* 3(2): 173- 178

Sani M. (2018) Robust diagnostic and parameter estimation for multiple linear and panel data regression models. Unpublished Ph.D. dissertation, Institute for Mathematical Research, Universiti Putra Malaysia

Sani, M., Midi, H. & Arasan, J. (2019) The Performance of Robust Heteroscedasticity Consistent Covariance Matrix Estimator. *Malaysian Journal of Mathematical Sciences,* 13(S): 71–88 (2019)

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48: 817-838

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