Scientific
Research
Publishing

Effects of Multicollinearity on Type I Error of Some Methods of Detecting Heteroscedasticity in Linear Regression Model

Olusegun Olatayo Alabi, Kayode Ayinde, Omowumi Esther Babalola, Hamidu Abimbola Bello, Edward Charles Okon

Department of Statistics, Federal University of Technology, Akure, Nigeria
Email: alabioo@futa.edu.ng, kayinde@futa.ng, habello@futa.ng, agunbiadewumi123@gmail.com, okoncharles2001@gmail.com

Abstract

Heteroscedasticity and multicollinearity are serious problems when they exist in econometrics data. These problems exist as a result of violating the assumptions of equal variance between the error terms and that of independence between the explanatory variables of the model. With these assumption violations, Ordinary Least Square Estimator (OLS) will not give best linear unbiased, efficient and consistent estimator. In practice, there are several structures of heteroscedasticity and several methods of heteroscedasticity detection. For better estimation result, best heteroscedasticity detection methods must be determined for any structure of heteroscedasticity in the presence of multicollinearity between the explanatory variables of the model. In this paper we examine the effects of multicollinearity on type I error rates of some methods of heteroscedasticity detection in linear regression model in other to determine the best method of heteroscedasticity detection to use when both problems exist in the model. Nine heteroscedasticity detection methods were considered with seven heteroscedasticity structures. Simulation study was done via a Monte Carlo experiment on a multiple linear regression model with 3 explanatory variables. This experiment was conducted 1000 times with linear model parameters of $\beta_0 = 4$, $\beta_1 = 0.4$, $\beta_2 = 1.5$ and $\beta_3 = 3.6$. Five (5) levels of mulicollinearity are with seven (7) different sample sizes. The method's performances were compared with the aids of set confidence interval (C.I.) criterion. Results showed that whenever multicollinearity exists in the model with any forms of heteroscedasticity structures, Breusch-Godfrey (BG) test is the best method to determine the existence of heteroscedasticity at all chosen levels of significance.

Keywords

Regression Model, Heteroscedasticity Methods, Heteroscedasticity Structures,

Multicollinearity, Monte Carlo Study, Significance Levels, Type I Error Rates

1. Introduction

The violation of the assumption of constant variance of error term in linear regression model results into heteroscedasticity problem. In practice, the nature of heteroscedasticity is usually unknown [1]. The consequences of OLS in the presence of heteroscedasticity are not BLUE, Inefficiency and Invalid Hypothesis testing. Given this fact, the detection of heteroscedasticity in a linear regression model needs to be identified. In reality, multicollinearity may co-exist with the problem of heteroscedasticity. The condition of severe non-orthogonality is referred to as a problem of multicollinearity. Multicollinearity exist when there is high linear relationships between two or more explanatory variables. According to [2] and [3], one should be very cautious about any conclusion with regression analysis when there is multicollinearity in the model, because [4], opined that effect of multicollinearity on type I error rates of the ordinary least square estimator is trivial in which the error rates exhibit no or little significance difference from the pre-selected level of significance. This paper attempts to determine the effects of multicollinearity on type I error rate of some heteroscedasticity detection methods in linear regression model. The heteroscedasticity detection methods chosen for this study are; Breusch Pagan test (BPG), Park test (PT), Spearman's Rank Correlation test (ST), Non-Constant Variation Score test (NVST), Glejser test (GLJ), Goldfeld-Quandt test (GFQ), Breusch-Godfrey test (BG), Harrison Mc Cabe test (HM) and White test (WT).

2. Background

Regression analysis is a statistical process for estimating the relationships among variables. It includes many techniques for modeling and analyzing several variables. Simple linear regression model postulated the relationship between dependent variable and one exogenous variable while multiple linear regression examine the relationship between dependent variable and a set of explanatory variables by fitting a linear equation to observed data. However, one of the assumption of classical linear regression model is that the variance of the error term is constant across observations (homoscedasticity). When homoscedasticity assumption is violated, it then leads to heteroscedasticity. Heteroscedasticity is a major concern in the application of regression analysis, which always occurs in cross sectional data, when the variances of the error terms are no longer constant. It is often investigated with the ideology of relationship between error terms and exogenous variables. According to [5] and [1], the consequences of using the Ordinary Least Square(OLS) estimator to obtain estimates of the population parameters when there is heteroscedasticity includes; inefficient parameter estimates and biased variance estimates which make standard hypothesis test

inappropriate. In practice, the nature of heteroscedasticity is usually unknown [1]. There are test procedures for establishing specific structures of heteroscedasticity. Brief literature reviews on some of these heteroscedasticity tests are as follows.

2.1. Breush-Pagan Test (BP)

[6] developed a test used in examining the presence of heteroscedasticity in a linear regression model. The variance of the error term was tested from a regression and is dependent on the value of the independent variables. [3] illustrates this test by considering the following:

Given the regression model

$$Y = \beta_0 + \beta_1 X + \mu \tag{1}$$

where Y is the dependent variables, X is the exogenous or explanatory variables, μ is the error term and βs are the regression coefficient.

[3] suggests that to determine the existence of heteroscedasticity in a given data the following procedures must followed;

Apply OLS in the model and compute the regression residuals.

Perform the auxiliary regression

$$\mu_i^2 = y_1 + y_2 z_{2i} + \cdots + y_p z_{pi} + \eta_i \tag{2}$$

where z could be partly replaced by independent variable X.

The test statistic is the result of the coefficient of determination of the auxiliary regression in (2) and sample size n with LM = nR^2. The test statistic is asymptotically distributed as χ^2_{p-1} under the null hypothesis of homoscedasticity.

2.2. Park Test (PT)

[5] proposes a LM test, the test assumes the proportionality between error variance and the square of the regressor. According to [1] and [5], LM test formulizes the graphical method by suggesting that σ^2 is a particular function of the explanatory variables. Park illustrates this test by regressing the natural log of squared residuals against the independent variable; if the independent variable has a significant coefficient, the data are likely to be heteroscedasticity in nature. Given the model below

$$\sigma^2 = \sigma^2 X_i^\beta e^v \tag{3}$$

We need to find the log

$$\ln \sigma_i^2 = \ln \sigma^2 + \beta \ln X_i + v_i \tag{4}$$

where v_i is the stochastic disturbance term, since σ_i^2 is not known, Park suggest using $\hat{u}_i^2$ as a proxy and run the following regression

$$\ln \mu_i^2 = \ln \sigma^2 + \beta \ln X_i + v_i = \alpha + \beta \ln X_i + v_i \tag{5}$$

If β turns out to be statistically significant, we there say that heteroscedasticity is present in the data and if it turns out to be insignificant, we may accept

the assumption of homoscedasticity.

2.3. Spearman's Rank Correlation Test (ST)

Spearman's Rank correlation [7] assumes that the variance of the disturbance term is either increasing or decreasing as X increases and there will be a correlation between the absolute size of the residuals and the size of X in an OLS regression. The data on X and the residuals are both ranked. The rank correlation coefficient is defined as

$$r_{X,e} = 1 - 6 \left[\frac{\sum_i d_i^2}{n(n^2 - 1)} \right]; -1 \leq r \leq 1 \tag{6}$$

where d_i is the difference between the rank of X and the rank of e in observations i and n is the number of individual ranked.

2.4. Glejser Test (GLJ)

[8] developed a test similar to the Park test, after obtaining the residual (u_i) (from the OLS regression. [8] suggests that regressing the absolute value of the estimated residuals on the explanatory variables that is thought to be closely associated with the heteroscedastic variance and attempts to determine whether as the independent variable increase in size, the variance of the observed dependent variable increases. This is done by regressing the error term of the predicted model against the independent variable. A high t-statistic (or low prob-value) for the estimate coefficient of the independent variable(s) would indicate the presence of heteroscedasticity.

2.5. Goldfeld-Quandt Test (GFQ)

[9] developed an alternative test to LM test, applying this test requires to perform a sequence of intermediate stages. First step involves to arrange the observations either is ascending or in descending order. Another step aims to divide the ordered sequence into two equal sub-sequences by omitting an arbitrary number P of the central observation. Consequently, the two equal sub-sequences will summarize each of them a number of $\frac{n-p}{2}$ observations. We then compute two different OLS regression the first one for the lowest values of X_i and the second for the highest values of X_i, in addition, obtain the residual sum of squares (RSS) for each regression equation, RSS$_1$ for the lowest values of X_i and RSS$_2$ for the highest values of X_i. An F-statistic is calculated based on the following formula:

$$F = \frac{RSS_1}{RSS_2} \tag{7}$$

The F-statistics is distributed with $\frac{N-P-2K}{2}$ degrees of freedom for both

numerator and denominator. Subsequently, compare the value obtained for the *F*-statistic with the tabulated values of *F*-critical for the specified number of degrees of freedom and a certain confidence level. If *F*-statistic is higher than *F*-critical, the null hypothesis of homoscedasticity is rejected and the presence of heteroscedasticity is confirmed.

2.6. Breusch-Godfrey Test (BG)

[10] developed a LM test of the null hypothesis of no heteroscedasticity against heteroscedasticity of the form $\sigma_t^2 = \sigma^2 h\left(z_t'\alpha\right)$, where z_t is a vector of independent variables. This vector contains the regressors from the original least square regression. The test is performed by completing an auxiliary regression of the squared residuals from the original equation on $(1, z_t)$. The test statistic follows a Chi-square distribution with degrees of freedom equal to the number of z under the null hypothesis of no heteroscedasticity.

2.7. White's Test (WT)

[11] proposed a statistical test that establishes whether the variance of the error in a regression model is constant. This test is generally, unrestricted and widely used for detecting heteroscedasticity in the residual from a least square regression. Particularly, White test is a test of heteroscedasticity in OLS residual. The null hypothesis is that there is no heteroscedasticity. The procedure for running the test is shows as follows:

Given the model

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \tag{8}$$

Estimate Equation (8) and obtained the residual $\hat{u}_i$ we then run the following auxiliary regression

$$\hat{u}_i^2 = b_1 + b_2 X_{2i} + b_3 X_{3i} + b_4 X_{2i}^2 + b_5 X_{3i}^2 + b_6 X_{2i} x_{3i} + v_i \tag{9}$$

The null hypothesis of homoscedasticity is $H_0 : b_1 = b_2 = \cdots = b_m = 0$ where H_0 highlights the fact that the variance of the residual is homoscedasticity *i.e.*, $var\left(\varepsilon_i\right) = Var\left(Y_i\right) = \sigma^2$. The alternative hypothesis is H_1 aims the fact that the variance of the residual is heteroscedasticity $var\left(\varepsilon_i\right) = Var\left(Y_i\right) = \sigma_i^2$ that is at least one of the b_i's is different from zero, the null hypothesis is rejected. The LM-statistic = nR^2 follows as χ^2 distribution characterized by $m - 1$, where n is the number of observation established to determine the auxiliary regression and R^2 is the coefficient of determination. Finally, we assume to reject the null hypothesis and to highlight the presence of heteroscedasticity when LM-statistic is higher than the critical value.

2.8. Harrison McCabe Test (HM)

[12] proposes a test to check the heteroscedasticity of the residuals. The breakpoint in the variances is set by default to the half of the sample. The *p*-value is estimated-using simulation. If the binary quality measure is false, then the ho-

moscedasticity hypothesis can be rejected with respect to the given level.

2.9. Non-Constant Variation Score Test (NVST)

[5] [13] and [14] develop a test of null hypothesis $H_0 : E\left(\varepsilon^2 / X_1, X_2, \cdots, X_k\right) = \sigma^2$ against an alternative (H_1) hypothesis with a general functional form. We recall the central issue is whether $E\left(\varepsilon^2\right) = \sigma^2 w_i$ is related to X and X_i. Then, a simple strategy is to use OLS residuals to estimate disturbance and check the relationship between ε_i^2 and X_i and that of X_i^2. Suppose that the relationship between ε_i^2 and X is linear

$$\varepsilon^2 = X\alpha + v \tag{10}$$

Then, we test $H_0 : \alpha = 0$ against $H_1 : \alpha \neq 0$ and base the test on how the squared OLS residual ε correlate with *X*.

3. Materials and Method

Consider the regression model of the form:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \cdots + \beta_p X_{pt} + u_t \tag{11}$$

$$u_t \sim N\left(0, \sigma_t^2\right);$$

where u_t is the error term and σ_t^2 is the heteroscedasticity variance that is considered. Y_t is the dependent variable, X_{pt} is the explanatory variables that contain multicollinearity and β_p is the regression coefficient of the model. A Monte Carlo Experiment was performed 1000 times, in generating the data for the simulation study. The error term containing different explanatory variables, heteroscedasticity structure and dependent variable were generated. The procedure used by [15] [16] and [17], was adopted to generate explanatory variables in this study. This is given as:

$$X_{ti} = \left(1 - \rho^2\right)^{\frac{1}{2}} * z_{ti} + \rho z_{tp} \tag{12}$$

$$t = 1, 2, 3, \cdots, n \quad \text{and} \quad i = 1, 2, 3, \cdots, p$$

where Z_{ti} is the independent standard normal distribution with mean zero and unit variance. Rho (ρ) is the correlation between any two explanatory variables and *p* is the number of explanatory variables. In this study, seven (7) error variance containing heteroscedasticity structures were considered, which are;

$$\sigma_t^2 = \sigma^2 \left(X_{2t}^2\right)^2 \tag{13}$$

$$\sigma_t^2 = \sigma^2 \left(X_{2t}^2\right) \tag{14}$$

$$\sigma_t^2 = \sigma^2 \left(X_{2t}\right) \tag{15}$$

$$\sigma_t^2 = \sigma^2 \left[E\left(y_t\right)^2\right] \tag{16}$$

$$\sigma_t^2 = \sigma^2 \left[E\left(y_t\right)\right] \tag{17}$$

$$\sigma_t^2 = \sigma^2 \left[\exp\left(\beta_0 + \delta\beta_1 X_{1t} + \delta\beta_2 X_{2t} + \delta\beta_3 X_{3t}\right)\right], \text{ where } \delta = 0 \text{ and } 0.2 \tag{18}$$

$$\sigma_t^2 = \sigma^2 \left(1 + X_{2t}^2\right)^2 \tag{19}$$

these tests were investigated and observed under type I error via the hypothesized values, to achieve this, Monte Carlo experiments is employed.

Moreover, in order to determine the dependent variables, Equation (1) was used in conducting the Monte Carlo experiments. The true values of the model parameters were fixed as follows; $\beta_0 = 4$, $\beta_1 = 0.4$, $\beta_2 = 1.5$, $\beta_3 = 3.6$. The sample sizes varied from 15, 20, 30, 40, 50, 100 and 250. At a specified value of sample size and multicollinearity level, the fixed X's are first generated; followed by the u_t and the values of Y_t were then determined. Then Y_t and X's were then treated as real life data set while the methods were applied.

The hypothesis about the methods of detecting heteroscedasticity under different forms of heteroscedasticity structures was tested at (10%, 5% and 1%) levels of significance to examine the (type I error rate) of each error terms. These intervals were referred to as the estimated significance level. The intervals was set to know the number of times each significance level falls between the range set for the confidence interval of each method of detecting heteroscedasticity in order to reject the hypothesis or not. At each level of significance; the interval set for $\alpha = 0.1$ is (0.09 to 0.14), the interval set for $\alpha = 0.05$ is (0.045 to 0.054), and the interval set for $\alpha = 0.01$ is (0.009 to 0.014).

Sample sizes were classified as small ($15 \leq n \leq 30$), medium ($40 \leq n \leq 50$) and large ($100 \leq n \leq 250$).

Multicollinearity levels were classified, the least value considered as low ($\rho = 0.8$), high ($\rho = 0.9$), very high ($\rho = 0.95$), Severe ($\rho = 0.99$) and very severe ($\rho = 0.999$).

At a particular α level a confidence interval was set for 10 percent, 5 percent and 1 percent, the number of times $\hat{\alpha}$ falls in between, the set confidence interval was counted over the sample size, multicollinearity and heteroscedasticity structures. The heteroscedasticity test with highest number of count is chosen to be the best.

$$\hat{\alpha} = \frac{r}{R} \tag{20}$$

where r is the number of times $\hat{\alpha}$ falls in between the confidence interval set at a particular significance level. While R is the number of times the experiment was carried out. At a given α, the number of time $\hat{\alpha}$ falls in between the set confidence interval at a particular sample sizes, multicollinearity levels and heteroscedasticity structures for each of the heteroscedasticity test was counted and the method with highest count is the best.

Procedure to Determine the Best Method of Detecting Heteroscedasticity When Multicollinearity Exist

1) α, which is the probability of committing type I error was chosen to be (10%, 5% and 1%).

2) Calculate $\hat{\alpha}$ where $\hat{\alpha} = \dfrac{r}{R}$, r is the number of times H_0 was rejected by a particular heteroscedasticity test in a particular sample sizes over a level of multicollinearity with a given heterosceasticity form. R is the number of replications.

3) Set confidence interval for each of the chosen level of the significance as follows; $\alpha = 0.1$ is (0.09 to 0.14), the interval set for $\alpha = 0.05$ is (0.045 to 0.054), and the interval set for $\alpha = 0.01$ is (0.009 to 0.014).

4) At a given α, count the number of time $\hat{\alpha}$ falls in between the set confidence interval at a particular sample sizes, multicollinearity levels and heteroscedasticity forms for each of the heteroscedasticity detection test.

5) The heteroscedasticity detection method with highest count in (4) is the best.

4. Results and Discussion

Results obtained from the simulation study show the number of times the estimated probability of type I error ($\hat{\alpha}$) fall in between the set confidence interval for $\alpha = 10\%$, 5% and 1% was counted over the sample sizes and heteroscedasticity structures for each heteroscedasticity detection method at different levels of multicollinearity as presented in Table 1.

Table 1. The number of time estimated probability of type I error $\hat{\alpha}$ fall in between the set confidence interval over the sample sizes, levels of multicollinearity and heteroscedasticity structures for various heteroscedasticity detection methods investigated.

Rho	Method	Estimated Alpha = 0.1								Estimated Alpha = 0.05								Estimated Alpha = 0.01							
		sample size (*n*)								sample size (*n*)								sample size (*n*)							
		15	20	30	40	50	100	250	Total	15	20	30	40	50	100	250	Total	15	20	30	40	50	100	250	Total
0.8	BPG	4	4	0	0	0	0	0	8	0	3	0	0	0	0	0	3	0	0	1	2	0	0	0	3
	PT	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	ST	4	2	0	0	0	0	1	7	1	0	0	0	0	0	0	1	2	1	1	0	0	0	1	5
	NVST	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	GLJ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	GFQ	0	0	0	0	3	3	0	6	1	0	0	1	0	0	0	2	0	0	1	4	0	0	0	5
	BG	3	3	6	4	4	6	3	**29**	2	1	0	2	1	1	3	**10**	1	3	3	2	3	1	3	**16**
	HM	0	2	1	0	3	3	0	9	0	0	0	1	0	0	0	1	0	2	0	3	0	0	0	5
	WT	0	0	0	4	2	3	0	9	0	0	0	0	0	2	0	2	0	0	0	2	0	1	0	3
0.9	BPG	3	2	0	0	0	0	0	5	0	2	0	0	0	0	0	2	0	0	0	0	0	0	0	0
	PT	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	ST	0	3	0	0	0	2	0	5	1	2	0	0	0	0	0	3	1	2	0	0	0	1	0	4
	NVST	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	GLJ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	GFQ	1	0	1	0	3	0	0	5	0	0	0	0	0	0	0	0	0	0	1	4	1	0	0	6
	BG	3	6	5	4	4	7	3	**32**	2	1	1	2	1	1	3	**11**	1	4	2	2	3	2	2	**16**

Continued

Level	Method																								
	HM	1	2	2	0	3	0	0	**8**	0	0	0	1	0	0	0	**1**	0	1	0	4	1	0	0	**6**
	WT	0	0	0	3	2	4	0	**9**	0	0	0	1	1	1	0	**3**	0	0	0	1	0	1	0	**2**
0.95	BPG	1	2	0	0	0	0	0	**3**	0	1	0	0	0	0	0	**1**	0	2	0	0	0	0	0	**2**
	PT	0	0	0	0	0	0	0	**0**	0	0	0	0	0	0	0	**0**	0	0	0	0	0	0	1	**1**
	ST	0	3	0	0	0	3	0	**6**	0	2	0	0	0	2	0	**4**	1	3	0	0	0	2	0	**6**
	NVST	1	0	0	0	0	0	0	**1**	0	0	0	0	0	0	0	**0**	0	0	0	0	0	0	0	**0**
	GLJ	0	0	0	0	0	0	0	**0**	0	0	0	0	0	0	0	**0**	0	0	0	0	0	0	0	**0**
	GFQ	2	0	1	0	3	0	0	**6**	2	0	1	0	1	0	0	**4**	2	0	0	5	1	0	0	**8**
	BG	3	5	4	4	5	7	3	**31**	1	2	2	2	2	0	2	**11**	1	5	2	3	4	3	3	**21**
	HM	4	2	1	0	2	0	0	**9**	0	0	1	1	0	0	0	**2**	2	1	2	6	1	0	0	**12**
	WT	0	0	0	4	3	4	0	**11**	0	0	0	2	0	1	0	**3**	0	0	0	1	0	2	0	**3**
0.99	BPG	1	1	0	0	0	0	0	**2**	0	1	0	0	0	0	0	**1**	0	2	0	0	1	0	0	**3**
	PT	0	1	0	0	0	0	0	**1**	0	0	0	0	0	0	0	**0**	1	0	0	0	0	0	1	**2**
	ST	0	2	0	0	0	1	0	**3**	0	1	0	0	0	1	0	**2**	0	2	0	0	0	1	0	**3**
	NVST	1	0	0	0	0	0	0	**1**	0	0	0	0	0	0	0	**0**	0	0	0	0	0	0	0	**0**
	GLJ	0	0	0	0	0	0	0	**0**	0	0	0	0	0	0	0	**0**	0	0	0	0	0	0	0	**0**
	GFQ	2	0	1	0	2	0	0	**5**	0	1	2	1	0	0	0	**4**	3	0	0	1	1	0	0	**5**
	BG	3	3	4	5	6	6	3	**30**	1	1	4	1	2	0	3	**12**	2	1	3	3	5	2	2	**18**
	HM	4	2	1	0	2	0	0	**9**	0	0	2	1	0	0	0	**3**	1	1	1	3	1	0	0	**7**
	WT	0	0	0	4	4	2	0	**10**	0	0	3	2	2	0	0	**7**	0	0	0	1	0	2	0	**3**
0.999	BPG	1	1	0	0	0	0	0	**2**	0	1	0	0	0	0	0	**1**	0	0	0	0	1	0	0	**1**
	PT	1	2	0	0	1	1	1	**6**	0	1	0	0	0	1	0	**2**	1	1	0	0	0	1	0	**3**
	ST	0	1	0	0	2	0	0	**3**	0	1	0	0	0	0	0	**1**	0	0	0	0	2	0	0	**2**
	NVST	1	0	0	0	0	0	0	**1**	0	0	0	0	0	0	0	**0**	0	0	0	0	0	0	0	**0**
	GLJ	0	0	0	0	0	0	0	**0**	0	0	0	0	0	0	0	**0**	0	0	0	0	0	0	0	**0**
	GFQ	2	1	1	3	2	0	0	**9**	0	0	1	1	0	0	0	**2**	1	0	0	2	0	0	0	**3**
	BG	3	3	5	5	7	6	3	**32**	1	0	3	1	4	0	3	**12**	1	0	3	2	6	2	3	**17**
	HM	4	2	1	2	3	0	0	**12**	0	0	0	3	1	0	0	**4**	2	1	0	3	0	0	0	**6**
	WT	0	0	0	4	3	3	0	**10**	0	0	0	1	1	2	0	**4**	0	0	0	0	1	0	0	**1**

Source: Simulated data.

From Table 1, the figures showing the performances of the heteroscedasticity detection method over the levels of multicollinearity were presented for $\hat{\alpha} = 0.1$, $\hat{\alpha} = 0.05$ and $\hat{\alpha} = 0.01$ in Figure 1, Figure 2 and Figure 3 respectively.

From Table 1 and Figure 1, when alpha = 0.1, it was generally observed that BG test is the best-performed method over all the structural forms of heteroscedasticity and sample sizes when there exist multicollinearity in the model.
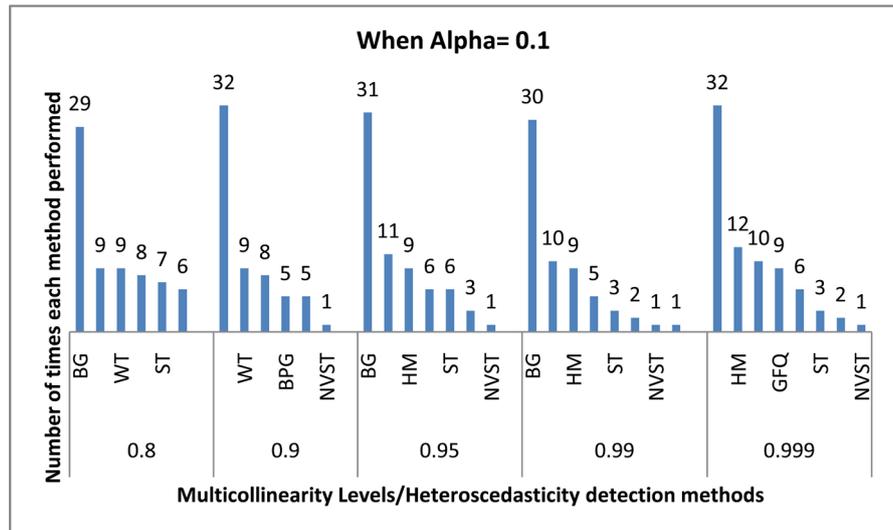
Figure 1. Figure showing the performances of the heteroscedasticity detection methods over the levels of multicollinearity when Alpha = 0.1.

Also, it was observed from Table 1 that;

1) When multicollinearity level is 0.8 and sample size is 15, BPG and ST methods of heteroscedasticity detection outperformed BG method.

When multicollinearity level is 0.8 and sample size is 20, BPG method of heteroscedasticity detection outperforms BG method.

When multicollinearity level is 0.8 and sample size is greater than 20 BG method's of heteroscedasticity detection outperformed all other methods.

2) When multicollinearity level is 0.9 BG method of heteroscedasticity detection performed best than all other methods except at sample size 15, at this instance, BPG method of heteroscedasticity detection outperformed equivalently well with BG method.

3) When multicollinearity level is grater or equal to 0.95, BG method of heteroscedasticity detection outperformed all other methods.

Hence, the performances of BG method of heteroscedasticity detection increase as the level of multicollinearity and sample sizes increases.

From Table 1 and Figure 2, when alpha = 0.05, it was generally observed that BG test is the best-performed method over all the structural forms of heteroscedasticity and sample sizes when there exist multicollinearity in the model.

Also, it was observed from Table 1 that;

1) When multicollinearity level is 0.8 and sample size is 15, BG method's of heteroscedasticity detection outperformed all other methods except at sample size 20 and sample size 100, at these instances, BPG and WT method of heteroscedasticity detection out performed BG method of heteroscedasticity detection respectively.

2) When multicollinearity level is 0.9, BG method of heteroscedasticity detection outperformed all other methods, except when sample size is 15, at this instance ST method of heteroscedasticity detection out performed BG method of
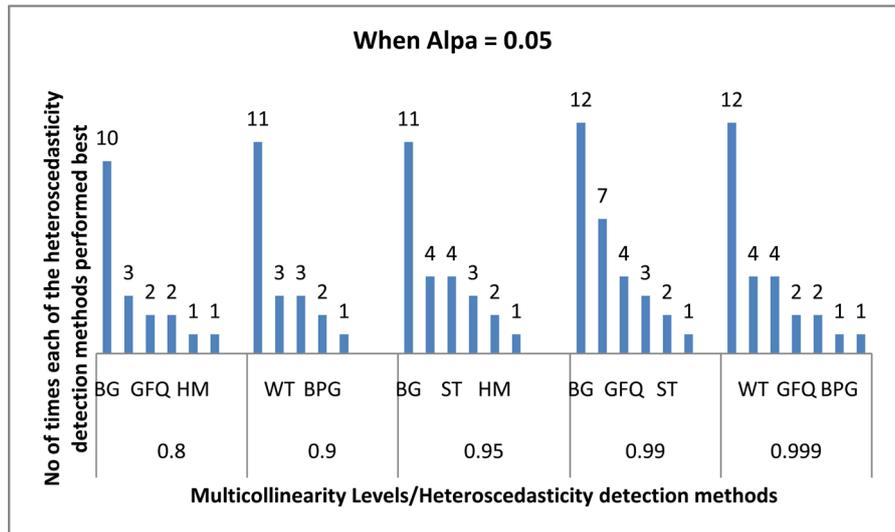
Figure 2. Figure showing the performances of the heteroscedasticity detection methods over the levels of multicollinearity when Alpha = 0.05.

heteroscedasticity detection.

3) When multicollinearity level is 0.95 BG method of heteroscedasticity detection outperformed all other methods except at sample size 40 and sample size 100, at these instance, WT method of heteroscedasticity detection out performed BG method of heteroscedasticity detection respectively.

4) When multicollinearity level is 0.99 BG method of heteroscedasticity detection performed best than all other methods except at sample size 10, at this instance, ST method of heteroscedasticity detection outperformed BG method.

5) When multicollinearity level is grater or equal to 0.999, BG method of heteroscedasticity detection outperformed all other methods except at sample size 15, at these instance, HM method of heteroscedasticity detection compete well with BG method of heteroscedasticity detection to outperform it.

Hence, the performances of BG method of heteroscedasticity detection increase as the level of multicollinearity and sample sizes increases.

From Table 1 and Figure 3, when $\hat{\alpha} = 0.01$, it was generally observed that BG test is the best-performed method of heteroscedasticity detection over all the structural forms of heteroscedasticity and sample sizes when there exist multicollinearity in the model.

Also, it was observed from Table 1 that;

1) When multicollinearity level is 0.8 and sample size is 15, BG method of heteroscedasticity detection outperformed all other methods except at sample size 15 and sample size 100, at these instances, GFQ and ST method of heteroscedasticity detection out performed BG method of heteroscedasticity detection respectively. Also, BPG method of heteroscedasticity detection compete favorably well with BG method at sample size 40.

2) When multicollinearity level is 0.9, BG method of heteroscedasticity detection outperformed all other methods, except when sample size is 40, at this instance
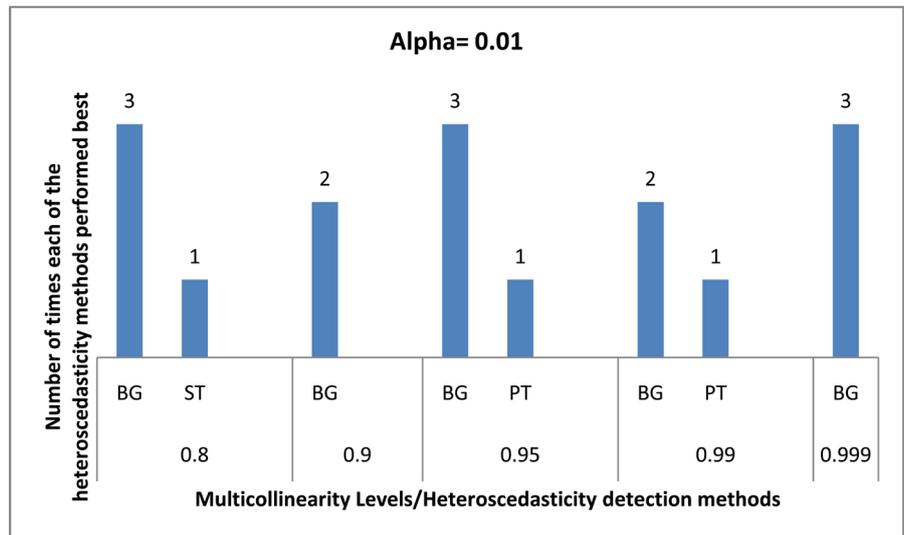
Figure 3. Figure showing the performances of the heteroscedasticity detection methods over the levels of multicollinearity when Alpha = 0.01.

GFQ and HM method of heteroscedasticity performed equivalently well to outperformed BG method of heteroscedasticity detection.

3) When multicollinearity level is 0.95, BG method's of heteroscedasticity detection outperformed all other methods except at sample size 15 and sample size 40, at these instances, GFQ and HM method of heteroscedasticity compete well to outperform BG method of heteroscedasticity detection respectively.

4) When multicollinearity level is 0.99, BG method of heteroscedasticity performed best than all other methods at all sample sizes except at sample size 15 and sample size 20; at these sample sizes, BPG and ST methods of heteroscedasticity detection performed well and outperformed BG method.

5) When multicollinearity level is grater or equal to 0.999, BG's method of heteroscedasticity detection outperformed all other methods at all sample sizes except at sample size 15 and sample size 20, at these instances, HM's method of heteroscedasticity detection compete well with BG's method of heteroscedasticity detection to outperform it.

Hence, the performances of BG's method of heteroscedasticity detection increase as the level of multicollinearity and sample sizes increases.

5. Conclusions

In spite of the level of multicollinearity, heteroscedasticity structures and sample size, we are able to conclude from the study on effects of multicollinearity on type I error rates of some methods of detecting heteroscedasticity when there exist multicollinearity in the model that;

The perfomances of BG's method of heteroscedasticity detection increases as the multicollinearity level increases at all the levels of significance.

The perfomances of BG method of heteroscedasticity detection increases as the sample size increases at all the levels of significance.

Whenever multicollinearity presents in the model with any heteroscedasticity structure, BG's test is the best method for heteroscedasticity detection in the model at different levels of significance in all sample size categories.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Gujarati, D.N. and Porter, D.C. (2009) Basic Econometrics. 5th Edition, McGraw-Hill Education, New York.

[2] Chatterjee, S. and Hadi, A.S. (2006) Regression Analysis by Example. 4th Edition, John Wiley & Sons, Hoboken. https://doi.org/10.1002/0470055464

[3] Chatterjee, S., Hadi, A.S. and Price, B. (2000) Regression Analysis by Example. 3rd Edition, John Wiley & Sons, Inc., Hoboken.

[4] Alabi, O.O., Ayinde, K. and Oyejola, B.A. (2008) Empirical Investigation of Effect of Multicollinearity on Type 1 Error Rates of the Ordinary Least Squares Estimators. *Journal of Modern Mathematics and Statistics*, **2**, 120-122.

[5] Park, R.E. (1966) Estimation with Heteroscedasticity Error Terms. *Econometrica*, **34**, 888. https://doi.org/10.2307/1910108

[6] Breusch, T.S. and Pagan, A.A. (1979) A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica*, **47**, 1287-1294. https://doi.org/10.2307/1911963

[7] Spearman, C. (1904) The Proof and Measurement of Association between Two Things. *America Journal of Psychology*, **15**, 72-101.

[8] Glejser, H. (1969) A Test for Heteroscedasticity. *Journal of the American Statistical Association*, **64**, 316-323. https://doi.org/10.1080/01621459.1969.10500976

[9] Goldfield, S.M. and Quandt, R.E. (1965) Some Tests for Homoscedasticity. *Journal of the American Statistical Association*, **60**, 539-547. https://doi.org/10.1080/01621459.1965.10480811

[10] Breusch, T.S. and Godfrey, L.G. (1978) Misspecification Tests and Their Uses in Econometrics. *Journal of Statistical Planning and Inference*, **49**, 241-260. https://doi.org/10.1016/0378-3758(95)00039-9

[11] White, H. (1980) A Heteroscedasticity-Consistent Covariance Matrix Estimation and a Direct Test of Heteroscedasticiy. *Econometrica*, **48**, 817-838. https://doi.org/10.2307/1912934

[12] Harrison, M.J. and McCabe, B.P.M. (1979) A Test for Heteroscedasticity Based on Ordinary Least Square Residuals. *Journal of the American Statistical Association*, **74**, 494-499. https://doi.org/10.1080/01621459.1979.10482544

[13] Rao, C.R. (1948) Large Sample Test of Statistical Hypothesis Concerning Several Parameters with Application to Problems of Estimation. *Mathematical Proceedings of the Cambridge Philosophical Society*, **44**, 50-57. https://doi.org/10.1017/S0305004100023987

[14] Cox, D.R. and Hinkley, D.V. (1974) Theoretical Statistics. Chapman and Hall, London.

[15] Mansson, K., Shukur, G. and Kibria, B.M.G. (2010) On Some Ridge Regression Es-

timators: A Monte Carlo Simulation Study under Different Error Variances. *Journal of Statistics*, **17**, 1-22.

[16] Lukman, A.F. and Ayinde, K. (2016) Review and Classifications of the Ridge Parameter Estimation Techniques. *Hacettepe Journal of Mathematics and Statistics*, **46**, 1-26.

[17] Durogade, A.V. (2013) New Ridge Parameters for Ridge Regression. *Journal of the Association of Arab Universities for Basic and Applied Sciences*, **15**, 1-6.