Websites:
http://www.sciencepub.net/newyork
http://www.sciencepub.net

Emails:
newyorksci@gmail.com
editor@sciencepub.net

New York Science Journal

MARSLAND PRESS
Multidisciplinary Academic Journal Publisher

Monte Carlos study on Power Rates of Some Heteroscedasticity detection Methods in Linear Regression Model with multicollinearity problem

O.O. Alabi, Kayode Ayinde, O. E. Babalola, and H.A. Bello

Department of Statistics, Federal University of Technology, P.M.B. 704, Akure, Ondo State, Nigeria
Corresponding Author: O. O. Alabi, alabioo@futa.edu.ng

Abstract: This paper examined the power rate exhibit by some heteroscedasticity detection methods in a linear regression model with multicollinearity problem. Violation of unequal error variance assumption in any linear regression model leads to the problem of heteroscedasticity, while violation of the assumption of non linear dependency between the exogenous variables leads to multicollinearity problem. Whenever these two problems exist one would faced with estimation and hypothesis problem. in order to overcome these hurdles, one needs to determine the best method of heteroscedasticity detection in other to avoid taking a wrong decision under hypothesis testing. This then leads us to the way and manner to determine the best heteroscedasticity detection method in a linear regression model with multicollinearity problem via power rate. In practices, variance of error terms are unequal and unknown in nature, but there is need to determine the presence or absence of this problem that do exist in unknown error term as a preliminary diagnosis on the set of data we are to analyze or perform hypothesis testing on. Although, there are several forms of heteroscedasticity and several detection methods of heteroscedasticity, but for any researcher to arrive at a reasonable and correct decision, best and consistent performed methods of heteroscedasticity detection under any forms or structured of heteroscedasticity must be determined. This paper then consider seven (7) heteroscedasticity structures that were originally proposed by different authors for developing statistical tools for heteroscedasticity detection in linear regression model. Also, nine (9) heteroscedasticity detection methods were examine to determines some methods that are best to be used in determining the presence of heteroscedasticity in a linear regression model with multicollinearity problem consideration was placed on power rate exhibit by each of the heteroscedasticity detection method. In this work, Monte Carlo experiment was conducted one thousands (1000) times on a linear regression model with three predictor variable that exhibits some degree of multicollinearity $(\rho = 0.8,\ 0.9,\ 0.95,\ 0.99,\ 0.999)$, seven sample sizes $(n = 15, 20, 30, 50, 100\, and\ 250)$. The parameters of the model were specified to be $\beta_0 = 4$, $\beta_1 = 0.4$, $\beta_2 = 1.5$, $\beta_3 = 3.6$ and the various tests were examined at 0.1, 0.05 and 0.01 levels of significance. The Confident interval criterion (C.I) was used to determine the performances of the methods. The study concluded that when power rate of a test is considered to determine the preferred heteroscedasticity detection method (s) in a model with multicollinearity problem; at significance level of α = 0.1 and 0.05, BG test or GFQ test are the preferred methods at all levels of multicollinearity, while at α = 0.01, the preferred method to use in testing for the presence of heteroscedasticity is either BG or NVST at all levels of multicollinearity except when $\rho \succ 0.95,$ at these instances ST compete favourably well with BG and NVST.

[O.O. Alabi, Kayode Ayinde, O. E. Babalola, and H.A. Bello. **Monte Carlos study on Power Rates of Some Heteroscedasticity detection Methods in Linear Regression Model with multicollinearity problem.** *N Y Sci J* 2020;13(6):31-38]. ISSN 1554-0200 (print); ISSN 2375-723X (online). http://www.sciencepub.net/newyork. 5. doi:10.7537/marsnys130620.05.

Keywords: Regression model, heteroscedasticity, heteroscedasticity detection methods, heteroscedasticity structures, Multicollinearity, significance levels, Confidence I*nterval and* Power rates.

1. Introduction

Heteroscedasticity cause serious problems in econometrics data. The consequences of using Ordinary Least Square (OLS) estimator when there is heteroscedasticity also affects the population parameters that leads to unbiasedness but inefficient, biased variance estimates and invalid hypothesis. Given this fact, the detection of heteroscedasticity in a linear regression model needs to be identified. Alabi et al (2008), opined that the 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 preferred heteroscedasticity detection methods via power rate among some methods of detecting heteroscedasticity in linear regression model when there is no multicollinearity problem in the linear regression model while evaluating a computationally simple asymptotic test that was proposed by Spearman (1904), Rao (1948), Goldfeld and Quandt (1965), Breusch and Godfrey (1966), Park (1966), Glejser (1969), Breusch and Pagan (1979), Harrison and McCabe (1979) and White (1980). These tests, originally designed for structural form and methods of detecting heteroscedasticity with various sample sizes under appropriate assumptions. These were used to test for the presence or absence of heteroscedasticity in linear regression models without multicollinearity problem.

1.2. Preview On The Methods Of Detecting Heteroscedasticity

There are several existing methods for detecting the existence of heteroscedasticity in linear regression model. This paper considered nine existing methods of heteroscedasticity detection, the methods are:

2-1 Breush-Pagan test (BP)

Breusch and pagan (1979) 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. Bresuch-Pagan illustrates this test by considering the following:

In linear regresson model, multiple regressions assess relationship between one dependent variable and a set of independent variables. Ordinary Least Squares (OLS) Estimator is most popularly used to estimate the parameters of regression model. The estimator has some very attractive statistical properties which have made it one of the most powerful and popular estimators of regression model. A common violation in the assumption of classical linear regression model is the presence of heteroscedasticity. Heteroscedasticity is a situation that arises when the variance of the error term is not constant. The performance of OLS estimator is inefficient in the presence of heteroscedasticity even though it is still unbiased. It does not have minimum variance any longer (Gujarati, 2003). In literature, there are various methods existing in detecting heteroscedasticity. Among them is the Breush-Pagan test. Breusch and pagan (1979) developed a test used in examining the prescence of heteroskedasticity 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. Bresuch-pagan illustrates this test by considering the following regression model

$$Y = \beta_0 + \beta_1 X_1 + u \tag{3}$$

Suppose that we estimate the regression model and obtain from this fitted model a set of value for the residuals û, OLS constrains these so that their mean is 0 and give the assumption that their variance does depend on the independent variables, an estimate of this variance can be obtained from the average of the squared values of the residuals. If the assumption is not held to be true, a simple model might be that the variance is linearly related to independent variable. Such a model can be examined by regressing the squared residuals on the independent variable, using an auxiliary regression equation of the form. z

$$\hat{u}^2 = \gamma_0 + \gamma_1 X + v \tag{4}$$

Bruesch-pagan test is a chi-square test; the test statistic is distributed with $n\chi^2$ with k degrees of freedom. If the test statistic has a p-value below an appropriate threshold e.g $p < 0.05$ then the null hypothesis of homoscedasticity is rejected and heteroscedasticity assumed. The procedure under the classical assumptions, OLS is the best linear unbiased estimate (BLUE), i.e, it is unbiased and efficient. It remains unbiased under heteroscedasticity, but efficiency is lost. Before deciding upon an estimation method, one may conduct the Bruesch-pagan test to examine the presence of heteroscedasticity. The Breusch-pagan test is based on model of the type

$$\sigma^2 = h(z_i'\gamma) \tag{5}$$

For the variance of the observations where $z_i = (1, z_{2i}, \ldots z_{pi})$ explain the difference in the variances. The null hypothesis is equivalent to the (p-1) parameter restriction:

$$\gamma_2 = \ldots = \gamma_p \tag{6}$$

For Breusch-pagan test large range multiplier (LM) yields the test statistic.

$$LM = \left(\frac{\partial l}{\partial \theta}\right)^T \left(-E\left[\frac{\partial^2 l}{\partial \theta \partial \theta^1}\right]\right)^{-1} \left(\frac{\partial l}{\partial \theta}\right) \tag{7}$$

This test is analogous to follow the simple three-step procedure:

Step1: Apply OLS in the model

$$Y = X\beta + \varepsilon \tag{8}$$

And compute the regression residuals.

Step2: Perform the auxiliary regression

$$\varepsilon_i^2 = y_1 + y_2 z_{2i} + \ldots + y_p z_{pi} + \eta_i \tag{9}$$

Where z could be partly replaced by independent variable X

Step 3: The test statistic is the result of the coefficient of determination of the auxiliary regression in step 2 and sample size n with LM = nR^2. The test statistic is asymptotically distributed as χ_{p-1} under the null hypothesis of homoscedasticity. Finally, we assume to reject the null hypothesis and to highlight the presence of heteroscedasticity when LM-statistic is higher than the critical value.

2-2 Park Test (PT)

Park (1966) propose a LM test, the test assumes the proportionality between error variance and the square of the regressors. According to Gujarati, the Park LM test formulizes the graphical method by suggesting that σ^2 is a particular function of the explanatory variable X_i.

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 is likely to be heteroscedasticity in nature.

In order to obtain the error term $\hat{u}_i$, we run a regression equation

$$Y_i = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + u_i \quad (10)$$

$$Var(u_i) = \sigma_i^2$$

by running the auxiliary regression we obtain the model below

$$\sigma_i^2 = \sigma^2 X_i^{\beta} e^{v} \quad (11)$$

We need to find the log

$$\ln \sigma_i^2 = \ln \sigma^2 + \beta \ln X_i + v_i \quad (12)$$

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

$$\ln \mu_i^2 = \ln \sigma^2 + \beta \ln X_i + v_i$$
$$= \alpha + \beta \ln X_i + v_i \quad (13)$$

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 (1904) 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

$$r_{X,e} = 1 - \left[\frac{6\sum_i d_i^2}{n(n^2-1)} \right]; -1 \le r \le 1 \quad (14)$$

where d_i is the difference between the rank of X and the rank of ε in observations. i and n is the number of individual ranked. Under the assumption that the population correlation coefficient is 0, the rank correlation coefficient has a normal distribution with 0 mean and variance $1/(n-1)$ in large sample. The appropriate test statistic is $r_{x,\varepsilon}(\sqrt{n-1})$ and the null hypothesis of homoscedasticity will be rejected at the 5% level if its absolute value is greater than 1.96 and at 1% level if its absolute values are greater than 2.58, using two tailed tests. The test can be performing with any of the levels if the explanatory variable is more than one. The preceding rank correlation coefficient can be used for heteroscedasticity. We assume the following model

$$Y_i = \beta_1 + \beta_2 X_i + u_i;\ i = 1,2,\ldots,n \quad (15)$$

the following steps are involved in spearman's rank correlation test

Run the above regression and obtain the residual $\hat{u}_i$ and take their absolute values $|\hat{u}_i|$

Arrange $|\hat{u}_i|$ and X_i in either increasing order or decreasing order and run Spearman's Rank Correlation coefficient by using formula

$$r_{X,\varepsilon} = 1 - 6\left[\frac{\sum d_i^2}{n(n^2-1)} \right] \quad (16)$$

If there is a systematic relationship between u_i and X_i, the rank correlation coefficient between the two should be statistically significant in which heteroscedasticity can be suspected. Given the null hypothesis that the true population rank correlation coefficient is zero and that n >8, it can be shown that $t = \frac{r_s\sqrt{n-2}}{\sqrt{1-r_s^2}} \sim t_{n-2}$ follows student's t–distribution with (n-2) degree of freedom. Therefore, in an application the rank correlation coefficient is significant on the basis of the t-test, we do not reject the hypothesis that there is heteroscedasticity in the problem. We therefore reject the null hypothesis of heteroscedasticity whenever $t_o \ge t_{1-\alpha,(n-2)}$. If there are more than one explanatory variable, rank correlation can be computed between |ε_i| and each of the explanatory variable separately and be tested using t_o.

2-4 Glejser test (GLJ)

Glejser (1969) developed a test similar to the Park test, after obtaining the residual ($\hat{u}_i$) from the OLS regression. Glejser suggest 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.

Glejser illustrates this test by considering the following steps

Step1: Estimate original regression with OLS and find the sample residual ε_i

Step2: Regress the absolute value | ε_i | on the explanatory variable that is associated with heteroscedasticity

Step3: Select the equation with the highest R^2 and lowest standard errors to represent heteroscedasticity

Step4: Perform a t-test on the equation selected from step3 on Y_i. If Y_i is statistically significant, reject the null hypothesis of homoscedasticity.

2-5 Goldfeld-Quandt test (GFQ)

Goldfeld -Quandt (1965) 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:

$$F = \frac{RSS_1}{RSS_2} \tag{17}$$

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)

Breusch-Godfrey (1978) developed a LM test of the null hypothesis of no heteroscedasticity against heteroscedasticity of the form $\sigma_t^2 = \sigma^2 h(z_t'\alpha)$, 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)

White (1980) 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

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \tag{18}$$

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

$$\hat{u}_i^2 = b_1 + b_2 X_{2i} + b_3 X_{3i} + b_4 X_{2\,i}^2 + b_5 X_{3\,i}^2 + b_6 X_{2i} x_{3i} + v_i \tag{19}$$

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, $\text{var}(\varepsilon_i) = Var(Y_i) = \sigma^2$.

The alternative hypothesis is H_1, it aims at the fact that the variance of the residual is heteroscedasticity $\text{var}(\varepsilon_i) = Var(Y_i) = \sigma_i^2$ that is at least one of the b_i's is different from zero, the null hypothesis is rejected. The LM-statistic equal to nR^2, this follows a 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)

Harrison-McCabe (1979) 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 homoscedasticity hypothesis can be rejected with respect to the given level.

2-9 Non-Constant ariation Score test (NVST)

Rao (1948), Cox and Hinkley (1974) develop a test of null hypothesis $H_0 : E(\varepsilon^2 / X_{1,} X_{2,} \cdots X_k) = \sigma^2$ against an alternative (H_1) hypothesis with a general functional form.

We recall the central issue is whether $E(\varepsilon^2) = \sigma^2 w_i$ 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 X_i^2. Suppose that the relationship between ε^2 and X is linear

$$\varepsilon^2 = X\alpha + v \tag{20}$$

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 multiple linear regression model of the form:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_{21} X_{2t} + \beta_p X_{pt} + u_t \tag{21}$$

where, $u_t \sim N(0, \sigma_t^2)$; 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. on error variance containing heteroscedasticity structures considered;

$$\sigma_t^2 = \sigma^2 (X_{2t}^2)^2 \tag{22}$$

$$\sigma_t^2 = \sigma^2 (X_{2t}^2) \tag{23}$$

$$\sigma_t^2 = \sigma^2 (X_{2t}) \tag{24}$$

$$\sigma_t^2 = \sigma^2 \left[E(y_t)^2 \right] \tag{25}$$

$$\sigma_t^2 = \sigma^2 \left[E(y_t) \right] \tag{26}$$

$$\sigma_t^2 = \sigma^2 (1 + X_{2t}^2)^2 \tag{27}$$

$$\sigma_t^2 = \sigma^2 [\exp(\beta_0 + \delta\beta_1 X_{1t} + \delta\beta_2 X_{2t} + \delta\beta_3 X_{3t})] \tag{28}$$

where δ =0 and 0.2.

Statistical package R-Studio (5.0) software was used to run the simulation aspect of the structures of heteroscedasticity to be tested. The best structural forms was identified from the result of the simulation.

3-1Generation of error term

The error term ε_i was generated to be normally distributed with mean zero and variance σ^2, that is,

$$\varepsilon_i \sim N(0, \sigma^2). \tag{29}$$

The error term containing different explanatory variables, heteroscedasticity structure and dependent variable were generated.

3-2 Generation of explanatory variables

The procedure used by Mansson *et al* (2010), Lukman and Ayinde (2015) and Durogade (2016), was adopted to generate explanatory variables in this study. This is given as:

$$X_{ti} = (1 - \rho^2)^{0.5} * z_{ti} + \rho z_{tp} \tag{30}$$

where t =1, 2,...,n and $i = 1, 2, \cdots, p$, where z_{ti} is the independent standard normal distribution with mean zero and unit variance. Rho ($\rho = 0.8, 0.9, 0.95, 0.99$ *and* 0.999) is the correlation levels between any two explanatory variables in this study. i.e. the multicollinearity level, while p is the number of explanatory variables.

3-3 Generation of Dependant variables

The dependent variables was generated, equation (21) 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 n, p, 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.

3-4 Estimated significance levels to b e used to determine the power rate

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 power rate on the variance of each error terms. Intervals was then set for the significance level as follows; The interval set for α = 0.1 is (0.09 to 0.14), the interval set for α = 0.05 is (0.045 to 0.054), and the interval set for α = 0.01 is (0.009 to 0.014). These intervals was set to know the number of times each power rate level falls between the range set for the confidence interval of each method of detecting heteroscedasticity in order to reject the hypothesis or not.

3-5 Sample Sizes

The sample sizes used for this research work were varied from 15, 20, 30, 40, 50, 100 and 250.

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

3-6 Criterion For Comparison

At a particular α level a confidence interval was set for 10%, 5% and 1%, the number of times $\hat{\alpha}$ falls in between, the set confidence interval was counted over the sample size and heteroscedasticity structures.

The heteroscedasticity test with highest number of count is choosen to be the best test with interm of power of the test.

$$\hat{\alpha} = \frac{r}{R} \tag{31}$$

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.

The heteroscedasticity test with highest number of count is chosen to be the best.

3-7 Power Of Statistical Test

Statistical Power of a test can be define as the probability of rejecting null hypothesis given that null hypothesis is false. If we consider the heteroscedasticity structure in equation (13) to (18), in all the structures, the null hypothesis will results to equal error variances give homoscedasticity except equation (18), which will give unequal error variances when $\delta = 2$.

i.e

$$\sigma_t^2 = \sigma^2\left[\exp(\beta_0 + 2\beta\beta_1 X_{1t} + 2\beta_2 X_{2t} + 2\beta_3 X_{3t})\right]$$

, when $\delta = 2$. (32)

The null hypothesis (H_0) stated that there is homoscedasticity while the alternative hypothesis (H_1) stated that there is heteroscedasticity. under this study, α is the probability of rejecting null hypothesis when it is not correct. while power is the probability of rejecting the null hypothesis (H_0) when the alternative hypothesis (H_1) is correct. In considering the Null hypothesis (H_0), rejection of null hypothesis (H_0) given that the alternative hypothesis (H_1) is true indicates a correct decision which is the power of the test. we then use this to determine the performances of the chosen methods by consider their performances under different significance levels (10%, 5% and 1%).

3-8 Procedures For Determination Of The Preferred Heteroscedasticity Method

At a particular α level a confidence interval was set for 10%, 5% and 1%, Test was carried out based on earlier stated null and alternative hypothesis,

The particular significance confidence interval α falls was determine, the number of times α falls in between, the set confidence interval was counted over the sample size and heteroscedasticity structures.

$\hat{\alpha}$ was determined as;

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

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.

The heteroscedasticity test with highest number of count is chosen to be the best with respect to power rate of test.

4. Results And Discussion

Results obtained from the simulation experiment shows that the number of times the estimated $\hat{\alpha}$ which is the probability of taking correct decision (power of the test) of each methods fall in between the set confidence interval for α = 10%, 5% and 1% was counted over the sample sizes and heteroscedasticity structures for each heteroscedasticity method of detection to obtain the results in Table 1.1.

Table 1.1: The number of counts for each method of detecting heteroscedasticity structures that has the highest power rate value at all sample sizes when there is presence of multicollinearity in the model.

ρ	Method	Estimated Alpha ($\hat{\alpha}=0.1$)								Estimated Alpha ($\hat{\alpha}=0.05$)								Estimated Alpha ($\hat{\alpha}=0.05$)							
		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	0	1	1	0	0	1	1	4	0	0	1	0	0	1	1	3	0	0	0	0	0	0	1	1
	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	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
	NVST	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	3
	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	1	1	1	1	1	1	7	0	0	1	1	0	1	1	4	0	1	1	0	0	0	0	2
	BG	1	1	1	1	1	1	1	7	0	0	0	0	0	0	1	1	1	0	0	0	1	1	0	3
	HM	1	1	0	1	1	1	1	6	0	0	0	0	0	1	1	2	0	0	0	0	0	0	0	0
	WT	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
0.9	BPG	0	1	1	0	0	1	1	4	0	0	1	0	0	1	1	3	0	0	0	0	0	0	1	1
	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	1	1	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
	NVST	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	3
	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	1	1	1	1	1	1	7	0	0	1	1	0	1	1	4	0	1	1	0	0	0	0	2
	BG	1	1	1	1	1	1	1	7	0	0	0	0	0	0	1	1	1	0	0	0	1	1	0	3
	HM	1	1	0	1	1	1	1	6	0	0	0	0	0	1	1	2	0	0	1	1	0	0	0	2
	WT	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.95	BPG	0	1	1	0	0	1	1	4	0	0	1	0	0	1	1	3	0	0	0	0	0	0	1	1
	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	1	1	0	0	0	0	1	3	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
	NVST	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	3
	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	1	1	1	1	1	1	7	0	0	1	1	0	1	1	4	0	1	1	0	0	0	0	2
	BG	1	1	1	1	1	1	1	7	0	0	0	0	0	0	1	1	1	0	0	0	1	1	0	3
	HM	1	1	0	1	1	1	1	6	0	0	0	0	0	1	1	2	0	0	1	1	0	0	0	2
	WT	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.99	BPG	0	1	1	0	0	1	1	4	0	0	1	0	0	1	1	3	0	0	0	0	0	0	1	1
	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	1	1	1	1	1	1	6	0	0	0	0	0	0	0	0	0	1	0	1	1	0	1	4
	NVST	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	2	0	0	0	1	0	1	1	3
	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	1	1	1	1	1	1	7	0	0	1	1	0	1	1	4	0	1	1	0	0	0	0	2
	BG	1	1	1	1	1	1	1	7	0	0	0	0	0	0	1	1	1	0	0	0	1	1	0	3
	HM	1	1	0	1	1	1	1	6	0	0	0	0	0	1	1	2	0	0	1	1	0	0	0	2
	WT	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.999	BPG	0	1	1	0	0	1	1	4	0	0	1	0	0	1	1	3	0	0	0	0	0	0	1	1
	PT	0	1	1	1	1	1	0	5	0	0	1	0	0	0	0	1	0	0	0	0	1	1	0	2
	ST	0	1	1	1	1	1	1	6	0	0	0	1	0	0	1	2	0	0	0	1	1	0	1	3
	NVST	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	2	0	0	0	1	0	1	1	3
	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	1	1	1	1	1	1	7	0	0	1	1	0	1	1	4	0	1	1	0	0	0	0	2
	BG	1	1	1	1	1	1	1	7	0	0	0	0	0	0	1	1	1	0	0	0	1	1	0	3
	HM	1	1	0	1	1	1	1	6	0	0	0	0	0	1	1	2	1	1	0	0	0	0	0	2
	WT	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Source: Simulated data

5. Conclusion

This paper revealed the power rate exhibited by each of the nine (9) heteroscedaticity detection methods under consideration in a linear regression model considered on some heteroscedasticity structures and several sample sizes that was classified as small ($15 \leq n \leq 30$), medium ($40 \leq n \leq 50$) and large ($100 \leq n \leq 250$) through a Monte Carlo study. The results in this paper show that when power rate of a test was considered on each method of heteroscedasticity detection in the model, BG test or

GFQ test are consistently the most preferred methods of heteroscedasticity detection when there exist no problem of multicollinearity between the exogenous variables of the regression model.

The results shows that with low sample size, BG and GFQ are the best except that GFQ out performed BG when $\alpha = 0.05$,

It also reveals that with medium sample size, BG and GFQ are consistently best except that HM and NVST compete well with BG at $\alpha = 0.05$ and $\alpha = 0.01$.

And with large sample size, BG and GFQ are consistently best except that GFQ and HM compete with BG and BPG at $\alpha = 0.05$ and NVST compete well with BG at $\alpha = 0.01$.

Inconclusion, BG and GFQ methods of heteroscedasticity are consistently best tests for heteroscedasticity detection when there is no multicollinerity problem in the regression model. of in detecting no-heteroscedasticity when α = 0.1 in the presence of multicollinearity. Moreover, when α = 0.05, GFQ test is the best method of no-heteroscedasticity detection. The recommended best method to be used is either BG test or NVST when the estimated alpha equal 0.01. Thus, this study suggest that, either BG test or GFQ test are consistently the best method to detect no-heteroscedasticity when there is presence of multicollinearity in the model at each level of significance the methods performed.

References

1. Alabi, O. O., Kayode, Ayinde. and Oyejola, B. A. (2008). Empirical Investigation of Effect of Multicollinearity on Type I Error rates of the Ordinary Least Squares Estimators. Journal of Modem Mathematics and Statistics 2 (3): 120-122. ISSN 1994-5388.
2. Breusch, T. S. and Godfrey, L. G. (1978). Misspecification Tests and Their Uses in Econometrics. Journal of Statistical Planning and Inference. 49 (2): 241–260. doi:10.1016/0378-3758(95)00039-9.
3. Breusch, T. S. and Pagan, A. A (1979). A Simple Test for Heteroscedasticity and Random Coefficient Variation. Econometrica 47, 1287-1294.
4. Chatterjee, S. and Hadi, A. S. (2006). Regression analysis by example. Fourth edition. Hoboken: John Wiley & Sons.
5. Charles Spearman (1904) Spearman Rank Correletion Coefficient. International Encyclopedia of the Social Sciences. Copyright 2008.
6. Cox, D. R. and Hinkley, D. V. (1974) Theoretical Statistics. London: Chapman and Hall.
7. Durogade, A. V. (2016). New Ridge Parameters for Ridge Regression. Journal of the Association of Arab Universities for Basic and Applied Sciences, 1-6.
8. Glejser, H. (1969). A New Test for Heteroscedasticity. Journal of the American Statistical Association. 64(235). 315-323 doi:10.1080/01621459. 1969. 10500976. JSTOR 2283741.
9. Goldfield, S. M., Quandt, R. E. (1965). Some Test for Homoscedasticity. Journal of the American Statistical Association. Bo (310): 539-547.
10. Harrison, M. J. and P. McCabe (1979). A Test for Heteroscedasticity Based on Ordinary Least Square Residuals. Journal of the American Statistical Association 74, 494- 499.
11. Lukman, A. F. and Ayinde, K. (2015). “Review and Classifications of the Ridge Parameter Estimation Techniques,” Hacettepe Journal of Mathematics and Statistics, vol. 46(113): 1- 1. DOI: 10. 15672/HJMS.201815671.
12. Mansson, K., Shukur, G. and Kibria, B. M. G. (2010). On Some Ridge Regression Estimators: A Monte Carlo simulation study under different error variances. Journal of Statistics, 17(1), 1-22.
13. Park, R. E. (1966). Estimation with Heteroscedasticity Error Term. Econometrica 34(4): 888. JSTOR 1910108.
14. Rao, C. R (1948). Large sample test of Statistical Hypothesis concerning several parameters with application to problems of testing. Proc. Comb. Phil. Soc. 44, 50-7.
15. Vinod, H. D. and Ullah, A. (1981). Recent Advances in Regression Methods, Marcel Dekker Inc. Publication.
16. White, H. (1980). A Heteroscedasticity Consistent Covariance Matrix Estimation and a Direct Test of Heteroscedasticiy. Econometrica, vol. 48, pp. 817-818.

6/20/2020