International Journal of Multidisciplinary Research and Development
www.allsubjectjournal.com
Online ISSN: 2349-4182, Print ISSN: 2349-5979
Received: 22-08-2021, Accepted: 07-09-2021, Published: 22-09-2021
Volume 8, Issue 9, 2021, Page No. 161-167

Determining the best method to detect and remedy heterocedasticity of multiple linear regression by Appling in simulation data

Hessain Adam Hessain Iydam*, Afra Hashim Abdelateef Mohammed
Department of Statistics, College of Science, Sudan University of Science and Technology, Khartoum, Sudan

Abstract
Constructing an econometric model fit for forecasting necessarily requires that it be free from measurement problems. This research paper focused on the problem of Heterocedasticity by comparing common detection methods and remedies applied to simulated model was data is corresponding to the government expenditure. The descriptive and analytical approach used is statistical packages (SPSS V.20) and (E. Views V.9) and (Excel v.10), and the most important results after applied in Simulated data was that the best test led to the detection of Heterocedasticity is White's Test, based on the determination coefficient and the probability value, which proved its advantage in helping to detect the problem when applied in the simulated model and the remedies. The best remedy that led to the detection of the problem was the first remedy because it was proven that 6 out of the 8 detection methods led to the remedy, followed by the third and fifth (by using logarithm) Assumptions. It was proven that 5 out of the 8 detection methods led to the remedy. The paper recommended using the White's Test to detect the problem of Heterocedasticity and remedy by taking algorithms.

Keywords: heterocedasticity, autocorrelation, multicollinearity, lagrange multiplier (LM), akaike information criterion (AIC)

Introduction
Multiple linear regression is one of the branches of econometrics and it is concerned with studying and analyzing the effect of several independent quantitative variables on a quantitative dependent variable, where the multiple linear regression model is used as a means to predict future values by estimating the model parameters that are adopted in the estimated model for the purposes of forecasting and knowing the effect or relationship between the explanatory variables and the dependent variable (one variable) by estimating this relationship and building predictive model for it according to scientific methodology and statistical foundations. Building a multiple linear regression model and estimating its features without knowing or without verifying the conditions and assumptions that must be met when applying it leads to incorrect results and predictions. And it is necessary here to mention that the variables and data of any statistical study using the method of multiple linear regression often suffer from auto-correlation problems between random errors, Heterocedasticity and Multicollinearity. It has been found to have the problems above, and this requires building a problem-free model to work for a prediction. This research paper focused on the problem of Heterocedasticity, as the descriptive and analytical method was used in this research by describing and analyzing data using tests and criteria used in detection and remedying the problem of Heterocedasticity in the MLR model and by using the statistical packages (SPSS Version 20) and the program (E.Views 9) and (Excel v.10) in Simulated model.

Government Expenditures
Includes all government consumption, investment, and transfer payments [1]. In national income accounting, the acquisition by governments of goods and services for current use, to directly satisfy the individual or collective needs of the community, is classed as government final consumption expenditure. Government acquisition of goods and services intended to create future benefits, such as infrastructure investment or research spending, is classed as government investment (government gross capital formation). These two types of government spending; on final consumption and on gross capital formation, together constitute one of the major components of gross domestic product [2].

Simulation Concept
The great development in computers and the existence of simulation programs with a high degree of flexibility and ease of use made the use of simulation in solving industrial, economic, social, medical and environmental problems an easy matter to the extent that there are many scientists who reformulated a lot of applied sciences and verified their validity, depending on simulation methods. And simulation is one of the important means to solve problems, problem solving techniques, and it is the only and last way to solve any problem if it is difficult to solve it by analytical methods or numerical methods. The simulation depends on methods of resampling methods and the generation of numbers and random variables with specific characteristics [3].
Simulation models allow obtaining information, such as mean or median or confidence intervals, on variables that do not have an exact value, but for which we either know or assume a distribution. If some "result" variables depend on these "distributed" variables by the way of known or assumed formulae, then the "result" variables will also have a distribution. Simulation allows you to define the distributions, and then to obtain, through simulations, an empirical distribution of the input and output variables as well as the corresponding statistics.

Simulation models are used in many areas such as finance and insurance, medicine, oil and gas prospecting, accounting, or sales prediction [4].

1. Definition of Simulation

Simulation is an imitation or representation of the action of a real system over a specified period of time. Whether we run the simulation manually or using a computer, it includes the generation of an artificial history of the system in order to infer the operational properties of the real system [5].

It is also defined as a mathematical method for treating and implementing dilemmas in the computer, in which certain types of mathematical and logical relationships necessary to describe the behavior and form of a system for a complex real world and for long periods of time overlap, and the simulation process begins by building a model for the problem under study, then implementing experiments and solutions for the complex numerical model.

2. Advantages of Simulation

1. Simulation enables the study and experimentation of the internal interactions of any complex system or part of that system.
2. Economic, financial, social and environmental changes can be simulated and observed for this modification of the model's behavior.
3. By changing the simulation inputs and observing the resulting outputs, we can identify the important variables in the real system and the way in which they interact.
4. From the modeling and simulation process we obtain very useful information to improve the performance of the real system
5. Simulations are used to support many of the theoretical research findings.

Conceptual Framework

1. Linear Regression Model

Linear regression model, or linear model in statistics, is a statistical model used to interpret a variable y via another variable x or (some variables $x_1, x_2, ..., x_k$ according to a linear function [6].

A variable y is called the dependent variable and x_k variables are independent or explained variables, meaning that they statistically explain the change of the dependent variable.

It is divided into two types:

a. Simple linear regression as it consists of one dependent variable and one independent variable

a. b. Multiple linear regression consisting of dependent variable and several independent variables [7]

2. Multiple Linear Regression Model (MLR)

The multiple linear model consists of k independent variables $x_1, x_2, ..., x_k$. It takes the following form:

$$Y_i = \beta_o + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_k X_{ki} + U_i \quad (1)$$

Based on formula (1), there are $(k+1)$ parameters that are required to be estimated.

So it can be written briefly as follows:

So it can be written briefly as follows:

$$\underline{Y} = \underline{X}\ \underline{\beta} + \underline{U} \quad (2)$$

Where (3) is the General Linear Regression (GLM) [8]

3. Coefficient of determination

The coefficient of determination (goodness of fit) R^2 of a multi-linear model can be calculated from the correlation coefficient between measured and estimated values. It indicates how well the model equation fits the data.

However, the goodness of fit depends not only on the quality of fit but also on the number of observations and the number of variables. The goodness of fit can be deliberately brought towards 1.0 simply by including an increasing number of variables (descriptors) into the model equation [9].

So that is, the ratio of the contribution of the dependent variable to bringing about changes in the explanatory variables.

4. Problems of Regression Model

The problems facing the linear regression model are represented in several forms, including heterogeneity of variance, self-correlation, and multiple linear interference in three main problems:

a. Heterocedasticity

b. Autocorrelation between U_i's

a. c. Multicollinearity [10]

5. Heterocedasticity

One of the most important hypotheses of the model, as it is known, is that the error term u_i present in the regression function must be homogeneous, meaning that they all have the same value of variance. Violation of this hypothesis is known as the Heterocedasticity problem. It should be noted here that the phenomenon of Heterocedasticity of variance affects estimates of the variance of model estimators and that the tests used as the t test and the F test in this case become unrealistic and unreliable [11].

Failure to fulfill the assumption of variance homogeneity, results in:

a. The inapplicability of the formulas for estimators' variances $\hat{Y}_i$ ، $\hat{\beta}_1$ ، $\hat{\beta}_o$

b. If the error variance is homogeneous, then the least squares estimators will have the least variance even though they remain unbiased estimators.

c. Predictions in variable Y based on $\hat{\beta}$'s from the original data will have large variations [12].

6. Detection of Heterocedasticity

The Heterocedasticity of the variance is detected by several tests, including the following:

4.6.1. Breuch-Pagan Godfry Test

4.6.2. Harvey-Godfrey LM

4.6.3. Glejser test

4.6.4. Auto-Regressive Conditional Heterocedasticity (ARCH LM) Test

4.6.5. White Test

4.6.6. Park Test

4.6.7. Spearman's Rank Correlation Coefficient Test:

4.6.8. Gold-field Quandt Test

7. Remedies of Heterocedasticity

The remedy of Heterocedasticity via transformation of the original model is performed, and the form of the transformation of the original model depends on the Heterocedasticity pattern in the estimated original model.

The original model is assumed to be as follows: $Y_i = \beta_o + \beta_1 X_i + U_i$

There are several patterns (assumptions) for the Heterocedasticity, and the converted model or equation differs from one assumption to another.

7.1 First Assumption

$$\sigma_{ui}^2 = \sigma_u^2 X_i^2 \quad (3)$$

7.2 The second Assumption

$$\sigma_{ui}^2 = \sigma_u^2 X_i \quad (4)$$

7.3 Third Assumption

$$\sigma_{ui}^2 = \sigma_u^2 \left[E\left(\hat{Y}_i\right)\right]^2 \quad (5)$$

7.4 Fourth Assumption

$$\sigma_{ui}^2 = \sigma_u^2 \left|e_i\right| \quad (6)$$

7.5 The Fifth Assumption

It is a logarithmic transfer, as it is known that taking the logarithms of the values leads to the convergence of these values from each other, and this means a decrease in the variance of the values. In this assumption the logarithms of the values of the two variables are taken and we get the following regression model:

$$\log Y_i = \beta_o + \beta_1 \log X_i + U_i \quad (7)$$

This model is estimated by the OLS method to be devoid of Heterocedasticity [13].

7.6 Sixth Assumption (general case)

In all of the previous cases:

$$\sigma_{ui}^2 = \sigma_u^2 f\left(X_i\right) \quad (8)$$

Therefore, in order to eliminate Heterocedasticity of the random variable U, all the terms of the original regression model are divided by the square root of the coefficient σ_u^2, i.e. by $\sqrt{f(X_i)}$ [14].

Results and Discussion

First, a sufficiency test was applied for the model and a description of the variables, followed by estimating the equation of the standard model by including all the variables on it, and then testing the measurement problems and detecting them with all methods of detection and remedying them with all remedial methods and measuring the merits of each model to choose the best method of detection methods and the best remedy for each problem in the first model.

1. Description of study variables

This data was simulated by using Excel v.10; we suppose all the variables are following the normal distribution, the table (1) below showing the Descriptive Statistics of real data

Table 1: Descriptive Statistics of Real Data from 1977 to 2018

Variable	Minimum	Maximum	Mean	Std. Deviation
GDP	23.4	1176630.0	134478.599	255024.269
EXPE	.5	116090.0	15326.294	26183.53
INF	-1.04	181.475	40.512	41.90
EXCH_Of	.0004	24.3527	2.275288	4.009
MS	53.186	430786	30642.78	74942.59

Source: Researcher preparation, using SPSS V. 20, 2021

Where: EXPE: Government Expenditures (dependent variable) from 1977 to 2018

GDP: gross domestic product (First independent variable) from 1977 to 2018

INF: Inflation (Second independent variable) from 1977 to 2018

MS: money supply (Third independent variable represents) from 1977 to 2018

EXCH_Of: Exchange Rate (Fourth independent variable) from 1977 to 2018

According to the Mean and Std. Deviation for each variable in table (1) we Simulated data corresponding to the real data with sample size 42 observations (like real data sample size), (we add (_s) for all study variables to denotes for simulated variables (EXPE_s, GDP_s, INF_s, MS_s & EXCH_Of_s).

2. Sufficiency Test

Hypothesis of the sufficiency was tested by Kaiser-Meyer-Olkin (K M O) test, which values fall between 0 – 1. KMO measures the sampling adequacy which should be greater than 0.5 [15]

So by applying it to the study data, we get:

Table 2: Test value (K M O)

Test	value	d.f	Sig. value
K M O	0.516	10	0.000

Source: Researcher preparation by using SPSS V.20, 2021

From Table (2) we note that the test value is (0.516) and it falls within the range (0.5 - 1). Therefore, the data is sufficient for the test, and what confirms that it is the significance of the probability value which is equal to (0.000) which is less than 0.05.

3. Comparison between Detection Methods of Heterocedasticity for the Simulated Model

Table 3: Comparison between Detection Methods of Heterocedasticity for Simulated Model

Test	Test Value	S.E. of regression	AIC	R^2	Prob-Value	Decision
Breuch-Pagan Godfry	1.384	1650019	31.582	0.13	0.2427	Homogeneous
Harvey	14.65	1.372	3.58	0.61	0.0000	Heterocedasticity
Glejser	4.9046	554.3699	15.585	0.35	0.006	Heterocedasticity
ARCH LM Test	0.000722	1714020	31.59	0.0002	0.978	Homogeneous

White	1.316	1596561	31.677	0.41	0.254	Homogeneous
Park	2.477	1.958	4.293	0.21	0.1176	Heterocedasticity
					0.1715	
					0.0277	
					0.0368	
Spearman's Rank Correlation Coefficient	-2.344	-	-	-	-	Heterocedasticity
	-1.84					
	-0.166					
	-0.166					
Gold-Field Quandt	1.632	-	-	-	-	Homogeneous

Source: Researcher preparation by using Excel, EViews 9, and SPSS V.20, 2021.

4. Remedies of Heterocedasticity for the Simulated Model

First Assumption

After applied we got the model as follow:

$$\hat{EXPE}\_s01 = 0.0948 + 1573.858\hat{GDP}\_s01 + 2.004\hat{INF}\_s01 + 0.0281\hat{MS}\_s01 - 466.5705\hat{EXCH}\_s01$$

Whereas the model is Significance and the value of the coefficient of determination of the model is $R^2 = 0.998$

Detection of Heterocedasticity for the First Remedy of the Simulated Model

By applied for all detection methods like we did in the original data as before (see: 4.7), we get the table below:

Table 4: Comparison between Detection Methods of Heterocedasticity for First Remedy of the Simulated Model

Test	Test Value	S.E. of regression	AIC	R^2	Prob-Value	decision
Breuch-Pagan Godfry	0.333	0.000422	-12.59	0.035	0.8335	Homogeneous
Harvey	1.076	4.0138	5.729	0.104	0.3572	Homogeneous
Glejser	1.071	0.0094	-6.388	0.104	0.3596	Homogeneous
ARCH LM Test	0.560	0.0004	-12.688	0.014	0.4462	Homogeneous
White	5.0215	0.0003	-13.329	0.70	0.0058	Heterocedasticity
Park	1.556	3.924	5.6833	0.144	0.0348	Heterocedasticity
					0.2962	
					0.2822	
					0.4693	
Spearman's Rank Correlation Coefficient	0.024	-	-	-	-	Homogeneous
	-0.605					
	0.630					
	0.866					
Gold-Field Quandt	1.1723	-	-	-	-	Homogeneous

Source: Researcher preparation by using Excel, EViews 9, and SPSS V.20, 2021.

Second Assumption: $\sigma_{ui}^2 = \sigma_u^2 X_i$

The equation of the model after applied becomes as follows (see 3.7.2):

$$\left(\frac{\hat{EXPE}\_s}{\sqrt{GDP\_s}}\right) = -7.376 + 0.1094\left(\frac{\hat{GDP}\_s}{\sqrt{GDP\_s}}\right) + 15.33\left(\frac{\hat{INF}\_s}{\sqrt{GDP\_s}}\right) - 0.0528\left(\frac{\hat{MS}\_s}{\sqrt{GDP\_s}}\right) + 1223.9\left(\frac{\hat{EXCH}\_s}{\sqrt{GDP\_s}}\right)$$

Whereas the model is Significance and the value of the coefficient of determination of the model is $R^2 = 0.93$

Detection of Heterocedasticity for the Second Remedy of the Simulated Model

By applied for all detection methods like we did in the original data we get the table below:

Table 5: Comparison between Detection Methods of Heterocedasticity for Second Remedy of the Simulated Model

Test	Test Value	S.E. of regression	AIC	R^2	Prob-Value	decision
Breuch-Pagan Godfry	4.8385	43.0102	10.472	0.34	0.0061	Heterocedasticity
Harvey	8.4139	2.5053	4.786	0.48	0.0005	Heterocedasticity
Glejser	8.2065	2.9803	5.133	0.47	0.0006	Heterocedasticity
ARCH LM Test	1.306	50.8358	10.742	0.03	0.2491	Homogeneous
White	2.702	40.0984	10.493	0.58	0.013	Heterocedasticity
Park	2.5208	3.069	5.192	0.21	0.0073	Heterocedasticity
					0.3638	
					0.8465	
					0.7680	
Spearman's Rank Correlation Coefficient	-4.264	-	-	-	-	Heterocedasticity
	1.685					
	0.798					
	1.011					
Gold-Field Quandt	8.712	-	-	-	-	Heterocedasticity

Source: Researcher preparation by using Excel, EViews 9, and SPSS V.20, 2021.

Third Assumption: $\sigma_{ui}^2 = \sigma_u^2 \left[E\left(\hat{Y}_i \right) \right]^2$

Divided original data by $\hat{y}$ (See 4.7.3)

So the equation of the model after applied becomes as follows (see 4.7.3):

$$\left(\frac{EX\hat{P}E\_s}{\hat{y}}\right) = 4.439 - 0.365\left(\frac{GDP\_s}{\hat{y}}\right) - 4.61\left(\frac{INF\_s}{\hat{y}}\right) - 0.145\left(\frac{MS\_s}{\hat{y}}\right) + 2227.229\left(\frac{EXCH\_s}{\hat{y}}\right)$$

Whereas the model is Significance and the value of the coefficient of determination of the model is $R^2 = 0.971$

Detection of Heterocedasticity for the Third Remedy Model of the Simulated Model

By applied for all detection methods like we did in the original data as before (see: 4.7), we get the table below:

Table 6: Comparison between Detection Methods of Heterocedasticity for the third Remedy of the Simulated Model:

Test	Test Value	S.E. of regression	AIC	R^2	Prob-Value	decision
Breuch-Pagan Godfry	0.4594	0.037	-3.667	0.05	0.7650	Homogeneous
Harvey	2.514	1.975	4.310	0.21	0.058	Homogeneous
Glejser	1.254	0.087	-1.939	0.12	0.305	Homogeneous
ARCH LM Test	0.468	0.036	-3.750	0.01	0.498	Homogeneous
White	4.941	0.023	-4.413	0.72	0.0002	Heterocedasticity
Park	2.865	1.946	4.281	0.24	0.003	Heterocedasticity
					0.229	
					0.924	
					0.743	
Spearman's Rank Correlation Coefficient	2.798	-	-	-	-	Heterocedasticity
	0.012					
	0.915					
	1.190					
Gold-Field Quandt	2.941	-	-	-	-	Homogeneous

Source: Researcher preparation by using Excel, E Views 9, and SPSS V.20, 2021.

Fourth Assumption: $\sigma_{ui}^2 = \sigma_u^2 |e_i|$

Divided original data by $\sqrt{|e_i|}$

So the equation of the model after applied becomes as follows:

$$\left(\frac{\text{EX}\hat{\text{P}}\text{E\_s}}{\sqrt{|e_i|}}\right) = 5.5441 + 0.104\left(\frac{\text{GDP\_s}}{\sqrt{|e_i|}}\right) + 5.136\left(\frac{\text{INF\_s}}{\sqrt{|e_i|}}\right) + 0.031\left(\frac{\text{MS\_s}}{\sqrt{|e_i|}}\right) - 436.837\left(\frac{\text{EXCH\_s}}{\sqrt{|e_i|}}\right)$$

Whereas the model is Significance and the value of the coefficient of determination of the model is $R^2 = 0.999$

Detection of Heterocedasticity for the Fourth Remedy Model of the simulated model

By applied for all detection methods like we did in the original data as before, we get the table below:

Table 7: Comparison between Detection Methods of Heterocedasticity for the fourth Remedy for the simulated model

Test	Test Value	S.E. of regression	AIC	R^2	Prob-Value	decision
Breuch-Pagan Godfry	2.446	909.916	16.576	0.21	0.0634	Homogeneous
Harvey	20.623	1.002	2.953	0.69	0.0000	Heterocedasticity
Glejser	5.970	12.317	7.971	0.39	0.0008	Heterocedasticity
ARCH LM Test	0.052	992.175	16.685	0.001	0.8213	Homogeneous
White	1.315	923.548	16.767	0.40	0.2617	Homogeneous
Park	8.008	1.318	3.502	0.46	0.0047	Heterocedasticity
					0.0039	
					0.9448	
					0.8322	
Spearman's Rank Correlation Coefficient	-4.286	-	-	-	-	Heterocedasticity
	-4.622					
	-2.702					
	-2.545					
Gold-Field Quandt	3.026	-	-	-	-	Homogeneous

Source: Researcher preparation by using Excel, E Views 9, and SPSS V.20, 2021.

The Fifth Assumption: $\log Y_i = \beta_o + \beta_1 \log X_i + U_i$

By taking the logarithm for all variables in original data

So the equation of the model after applied becomes as follows (see 4.7.5):

$$\text{Log}\left(\text{EX}\hat{\text{P}}\text{E\_s}\right) = 2.942 + 0.715\left(\text{Log}\left(\text{GDP\_s}\right)\right) - 0.028\left(\text{Log}\left(\text{INF\_s}\right)\right) - 0.163\left(\text{Log}\left(\text{MS\_s}\right)\right) + 0.139\left(\text{Log}\left(\text{EXCH\_s}\right)\right)$$

Whereas the model is Significance and the value of the coefficient of determination of the model is $R^2 = 0.89$

Detection of Heterocedasticity for the Fifth Remedy Model of the simulated model

By applied for all detection methods like we did in the original data as before, we get the table below:

Table 8: Comparison between Detection Methods of Heterocedasticity for the fifth Remedy of the simulated model:

Test	Test Value	S.E. of regression	AIC	R^2	Prob-Value	decision
Breuch-Pagan Godfry	6.180	0.249	0.167	0.40	0.0007	Heterocedasticity

Harvey	0.338	2.367	4.672	0.03	0.8504	Homogeneous
Glejser	2.463	0.235	0.058	0.21	0.0620	Homogeneous
ARCH LM Test	0.211	0.312	0.554	0.01	0.6488	Homogeneous
White	20.401	0.110	-1.294	0.91	0.0000	Heterocedasticity
Park	0.420	2.446	4.765	0.06	0.8333	Homogeneous
					0.4176	
					0.7882	
					0.5807	
Spearman's Rank Correlation Coefficient	0.298	-	-	-	-	Homogeneous
	0.012					
	0.648					
	0.648					
Gold-Field Quandt	30.054	-	-	-	-	Heterocedasticity

Source: Researcher preparation by using Excel, EViews 9, and SPSS V.20, 2021.

Sixth Assumption (general case)

In all of the previous cases (see 4.7.6): $\sigma_{ui}^2 = \sigma_u^2 f\ (X_i)$

The Researcher suggested: $\sigma_{ui}^2 = \sigma_u^2 [E(\hat{y})]$

According to it we will divided all original variables by $\sqrt{\hat{y}}$

So the equation of the model after applied becomes as follows:

$$\left(\frac{EX\hat{P}E\_s}{\sqrt{\hat{y}}}\right) = -\ 0.046821 + 0.101\left(\frac{GDP\_s}{\sqrt{\hat{y}}}\right) + 20.316\left(\frac{INF\_s}{\sqrt{\hat{y}}}\right) + 0.071\left(\frac{MS\_s}{\sqrt{\hat{y}}}\right) - 1022.350\left(\frac{EXCH\_s}{\sqrt{\hat{y}}}\right)$$

Whereas the model is Significance and the value of the coefficient of determination of the model is $R^2 = 0.998$

Detection of Heterocedasticity for the Sixth Remedy for the simulated model

By applied for all detection methods like we did in the original data as before, we get the table below:

Table 9: Comparison between Detection Methods of Heterocedasticity for the sixth Remedy for the simulated model:

Test	Test Value	S.E. of regression	AIC	R^2	Prob-Value	decision
Breuch-Pagan Godfry	4.657	0.292	0.487	0.33	0.004	Heterocedasticity
Harvey	4.367	1.618	3.912	0.32	0.005	Heterocedasticity
Glejser	5.868	0.231	0.022	0.39	0.001	Heterocedasticity
ARCH LM Test	2.276	0.338	0.719	0.05	0.139	Homogeneous
White	3.827	0.243	0.277	0.66	0.001	Heterocedasticity
Park	4.869	1.589	3.876	0.34	0.5845	Heterocedasticity
					0.0906	
					0.0466	
					0.4970	
Spearman's Rank Correlation Coefficient	2.407	-	-	-	-	Heterocedasticity
	1.588					
	-3.520					
	-3.286					
Gold-Field Quandt	4.024	-	-	-	-	Heterocedasticity

Source: Researcher preparation by using Excel, EViews 9, and SPSS V.20, 2021.

Comparison of Detections Methods of Heterocedasticity for the simulated model:

Table 10: Comparisons of Detections Methods of Heterocedasticity for the simulated model:

Table Number in Comparisons Between Detection Methods	Maximum R^2	Minimum AIC	Minimum P.value
14	Harvey Test	Harvey Test	Harvey Test
17	White Test	Park Test	White Test
19	White Test	Harvey Test	Harvey Test
21	White Test	Glejser Test	White Test
23	Harvey Test	Harvey Test	Harvey Test
25	White Test	Glejser Test	White Test
27	White Test	Glejser Test	White Test & Glejser Test

Source: Researcher preparation 2021

We conclude from Table (10) that the best test that led to the detection of Heterocedasticity was White test, as it proved its best in helping to detect the problem when it was applied in the Simulated model and the remedies, According to AIC [16] the best tests is Glejser Test and Harvey Test.

Comparison of Remedies of Heterocedasticity:

Table 11: Comparisons of Remedies of Heterocedasticity for the simulated model:

Remedy	ratio of remedy to test
First	6: 8
Second	1: 8

Third	5: 8
Fourth	4: 8
Fifth	5: 8
Sixth	1: 8

Source: Researcher preparation 2021

From Table (11) we note that the best remedy led to remedy Heterocedasticity is First Remedy According to first assumption $\sigma_{ui}^2 = \sigma_u^2 X_i^2$, because it was proved that 6 out of the 8 detection methods led to the remedy, followed by the third and fifth Assumptions $\sigma_{ui}^2 = \sigma_u^2 \left[E\left(\hat{Y}_i\right)\right]^2$ and the logarithmic method $\log Y_i = \beta_o + \beta_1 \log X_i + U_i$ (By taking the logarithm) it where proven that 5 out of the 8 methods of detection led to the remedy of the problem by it.

Conclusions

From the obtained results we may conclude that, after estimating MLR for Simulated data, it was found that it suffers from the problem of Heterocedasticity, as 4 out of 8 tests helped to detect that; in addition to we noted:

1. The value of the coefficient of determination of the main Simulated estimated model was 0.997; meaning that 99.7% of the changes that occur to the dependent variable are caused by the independent variables.
2. All models estimated are significant.
3. The best test that led to the detection of Heterocedasticity was White's Test, based on the determination coefficient and the probability value, which proved its advantage in helping to detect the problem when applied in the simulated model and the remedies.
4. According to AIC the best tests are Glejser and Harvey Tests.
5. The best remedy led to remedy Heterocedasticity was First Remedy According to first assumption, because it was proved that 6 out of the 8 detection methods led to the remedy, followed by the third and fifth Assumptions i.e. And (By taking the logarithm in fifth assumption) it where proven that 5 out of the 8 methods of detection led to the remedy of the problem by it.

According to all these results for a simulated data, we recommend using White's Test to

References

1. Robert Barro, Vittorio Grilli. European Macroeconomics, Ch. 15–16. Macmillan, ISBN 0-333-57764-7, 1994.
2. https://en.wikipedia.org/wiki/Government_spending
3. Berri, Adnan Majid Abdul Rahman. “Modeling and simulation using Excel, SIMAN, Arena and General Purpose Simulation System (GPSS WORLD)”, King Saud University - KSA, 2002, 11.
4. Look Website link: https://help.xlstat.com/s/article/generating-many-distributions-in-a-simulation-model-by-copying?language=en_US
5. Ibid, 14.
6. Ismail, Muhammad Abd al-Rahman. “Linear Regression Analysis”, Institute of Public Administration, Kingdom of Saudi Arabia, 2006, 15-19.
7. Ronald E, Walpole. Probability & Statistics for Engineers & Scientists, Prentice Hall, Upper Saddle River, New Jersey, USA, 2007, 390-456.
8. Alrawi, Khashi Mahmoud. “Introduction to Regression Analysis”, Dar Al Kutub Directorate for Printing and Publishing, University of Mosul, Iraq, 1987.
9. http://www.statistics4u.info/fundstat_eng/ee_mlr_coeff_fit.html#:~:text=The%20coefficient%20of%20determination%20(goodness,equation%20fits%20to%20the%20data
10. Ibrahim, Bassam Younis, Haji, Anmar Amin, Younes, Adel Moussa. “*Econometrics*”, Azza Publishing House, Khartoum - Sudan, 2002, 151-279.
11. Samprit, Chatterjee, Ali S.Hadi. “Regression Analysis by Example”, Third Edition, John Wiley & Sons, 2006.
12. Ibrahim, op. cit, 205-206.
13. Ibid, 223.
14. Ibid, 223.
15. Look Website link: https://www.eduhk.hk/apfslt/v10_issue2/turkmen/turkmen3.htm
16. The Akaike information criterion (AIC) is an estimator of out-of-sample prediction error and thereby relative quality of statistical models for a given set of data, available at: https://en.wikipedia.org/wiki/Akaike_information_criterion