Communication in Physical Sciences, 2026, 13(8):1319-1329

Monte Carlo Evaluation of Classical Heteroscedasticity Test in Linear Regression

Rita Nneka Nwaka* and Alaba Akinleye Obabire
Received: 18 April 2026/Accepted: 21 July 2026/Published: 28 July 2026
https://doi.org/10.4314/cps.v13i8.6

Abstract: *In order to assess and compare the finite-sample performance of five widely used tests for identifying heteroscedasticity in linear regression models—the Park, Breusch–Pagan, Goldfeld–Quandt, Glejser, and White tests—this study uses a Monte Carlo simulation framework. Both homoscedastic and heteroscedastic error structures are used to evaluate the tests' statistical power and empirical Type I error rates. Each experimental situation is duplicated 1,000 times at significance levels of 5% and 1%, and simulations are run for sample sizes ranging from 20 to 100 observations. Significant variations in the competing processes' performance are revealed by the simulation findings. In the majority of heteroscedasticity cases, the Glejser test typically exhibits the best statistical power, demonstrating strong sensitivity to deviations from constant error variance. Nevertheless, in homoscedastic conditions, this increased sensitivity is accompanied by rather high Type I error rates. On the other hand, as sample size and heteroscedasticity severity rise, the Breusch–Pagan and White tests show greater detection power and better control over the nominal Type I error rate. When the degree of heteroscedasticity is quite low, the Goldfeld–Quandt test typically records the lowest performance. Overall, the results indicate that no test predominates under all experimental circumstances. The Breusch-Pagan and White tests offer a better balance between Type I error control and statistical power than the Glejser test, which is extremely susceptible to heteroscedasticity. Therefore, the study offers helpful empirical data to help researchers choose the best heteroscedasticity diagnostic techniques for linear regression analysis.*

Keywords: *Heteroscedasticity; Monte Carlo simulation; Linear regression; Statistical power; Type I error; Diagnostic tests.*

Rita Nneka Nwaka*
Department of Mathematics and Statistics, University of Delta, Agbor, Delta State, Nigeria
Email: rita.nwaka@unidel.edu.ng
https://orcid.org/0009-0007-7280-576

Alaba Akinleye Obabire
Department of Statistics, The Federal Polytechnic, Orogun, Delta State, Nigeria
Email: obabire.akinleye@fepo.edu.ng

1.0 Introduction

One of the most used statistical methods for examining correlations between variables in the social sciences, engineering, economics, finance, and medicine is still linear regression. However, the degree to which the assumptions of the classical linear regression model (CLRM) are met determines the validity of statistical inference from a linear regression model. Among these presumptions, homoscedasticity—the requirement that the error terms have constant variance—is especially crucial since it ensures the validity of standard hypothesis tests and the effectiveness of the Ordinary Least Squares (OLS) estimator (Gujarati & Porter, 2009; Wooldridge, 2020).

The regression model is said to display heteroscedasticity when the assumption of constant variance is broken, suggesting that the variance of the error factors fluctuates systematically across observations.

Heteroscedasticity influences the estimators' efficiency and results in biased estimates of their standard errors, but it does not impact the OLS regression coefficients. Confidence intervals, t-tests, F-tests, and other inferential techniques may therefore become unreliable, causing researchers to draw erroneous conclusions about the importance of explanatory factors (Greene, 2020). Finding heteroscedasticity prior to doing statistical inference has become a crucial step in model validation as regression analysis is the foundation of empirical research in many fields. Over the past few decades, many diagnostic techniques have been developed to detect heteroscedasticity. The Breusch–Pagan test (Breusch & Pagan, 1979), the White test (White, 1980), the Goldfeld–Quandt test (Goldfeld & Quandt, 1965), the Park test (Park, 1966), and the Glejser test (Glejser, 1969) are some of the most popular. The theoretical presumptions, computational methods, and sensitivity to various types of heteroscedasticity vary among these tests.

For instance, the White test offers a more broad diagnostic that does not require definition of the functional form of heteroscedasticity, whereas the Park and Goldfeld–Quandt tests require previous assumptions regarding the relationship between the error variance and explanatory variables. In a similar vein, although the Breusch-Pagan test is computationally efficient, it is predicated on more robust assumptions about the underlying error structure. As a result, no single technique is always better in every experimental scenario. Thus, the statistical and economic literature has focused a great deal of attention on the relative performance of heteroscedasticity detection tests. The distribution of the disturbance term, the degree and functional form of heteroscedasticity, and sample size all affect how effective these tests are in finite samples (Onifade & Olanrewaju, 2020; Akewugberu et al., 2024). Certain tests identify heteroscedasticity more well at the expense of elevated Type I error rates, whereas others preserve the nominal significance level but have very low statistical power. For applied researchers looking for a suitable diagnostic process for a specific regression problem, these variations pose practical hurdles.

Because Monte Carlo simulation allows data to be generated under controlled conditions where the true data-generating process is understood, it has become one of the most popular approaches for assessing the finite-sample performance of statistical algorithms (Gentle, 2003).

Researchers can estimate empirical Type I error rates and statistical power by repeatedly simulating datasets under predetermined assumptions. This allows them to compare competing statistical tests objectively. The value of Monte Carlo techniques in evaluating heteroscedasticity detection techniques under different sample sizes and error variance structures is still being shown by recent research (Akewugberu et al., 2024; Farrar et al., 2025).

Nevertheless, comparative evidence involving the Park, Glejser, Breusch–Pagan, White, and Goldfeld–Quandt tests within a common simulation framework remains relatively limited, particularly across different levels of heteroscedasticity and finite sample sizes.

In order to close this gap, the Park, Glejser, Breusch–Pagan, White, and Goldfeld–Quandt tests—five popular heteroscedasticity identification tests—are thoroughly compared using Monte Carlo. Empirical Type I error rates and statistical power under both homoscedastic and heteroscedastic error structures across various sample sizes are used to assess these procedures' performance. Researchers in statistics, econometrics, and related fields might find the results useful in choosing suitable diagnostic techniques for identifying heteroscedasticity in regression analysis.

In addition, the study contributes to the growing body of simulation-based evidence on the finite-sample behaviour of classical

heteroscedasticity tests and offers recommendations for their application in empirical research.

2.0 Literature Review

2.1 Heteroscedasticity

The classical linear regression model assumes that the error terms have the same variance for every observation, a condition referred to as homoscedasticity (Gujarati & Porter, 2009; Wooldridge, 2020). In mathematical form, this assumption is written as $Var(\varepsilon_i) = \sigma^2$. A departure from this condition occurs when the error variance differs across observations, giving rise to heteroscedasticity, expressed as $Var(\varepsilon_i) = \sigma_i^2$, where $\sigma_i^2 \neq \sigma^2$. This problem is particularly common in cross-sectional data, where substantial differences may exist among observations in terms of income, expenditure, firm size, and other socioeconomic characteristics (Yang et al., 2019). Other possible sources include omitted variables, model misspecification, outliers, measurement errors, and the adoption of inappropriate functional forms (Greene, 2020).

Heteroscedasticity does not cause bias in the OLS coefficient estimates, but it reduces their efficiency and results in biased standard error estimates. Consequently, the reliability of hypothesis tests and confidence intervals can be affected (Gujarati & Porter, 2009). Identifying the presence of heteroscedasticity is therefore an essential part of regression diagnostics.

2.2 Statistical Power and Hypothesis Testing

Statistical power refers to the probability that a false null hypothesis will be correctly rejected. It is given by $1 - \beta$, where β represents the probability of a Type II error (Cohen, 1988). In the context of heteroscedasticity testing, power reflects the capacity of a diagnostic procedure to detect non-constant error variance when such a condition is actually present.

Hypothesis testing is also subject to Type I error, which occurs when a true null hypothesis is rejected. The probability associated with a Type I error is denoted by the significance level, α, which is commonly set at 0.05 (Montgomery et al., 2021). A Type II error occurs in the opposite situation, where a false null hypothesis is not rejected. Hence, a useful heteroscedasticity test should maintain its Type I error rate near the specified nominal level while achieving adequate statistical power.

2.3 Heteroscedasticity Detection Tests

A variety of statistical procedures are available for identifying heteroscedasticity. The Park test is based on a log-linear relationship between the error variance and an explanatory variable (Park, 1966). In the Glejser procedure, the absolute values of the residuals are regressed on the explanatory variables (Glejser, 1969). The Breusch–Pagan test examines whether systematic dependence exists between the error variance and the explanatory variables (Breusch & Pagan, 1979). The White test extends this approach by allowing the variance to be associated with the explanatory variables, their squared terms, and interaction terms (White, 1980). The Goldfeld–Quandt test, in contrast, assesses differences in residual variances between ordered subsets of the observations (Goldfeld & Quandt, 1965).

The performance of these heteroscedasticity detection procedures has also been examined in recent studies through Monte Carlo simulation (Onifade & Olanrewaju, 2020; Akewugberu et al., 2024; Farrar et al., 2025).

2.4 Monte Carlo Simulation

Monte Carlo simulation provides a controlled framework for assessing the finite-sample behaviour of statistical procedures. Its usefulness arises from the ability to generate data from a known data-generating process under predetermined conditions (Gentle, 2003). Through repeated simulations, researchers can obtain empirical estimates of Type I error rates and statistical power for different sample sizes and levels of heteroscedasticity.

Against this background, the present study employs a comprehensive Monte Carlo simulation to compare five commonly used

procedures for detecting heteroscedasticity: the Breusch–Pagan, White, Goldfeld–Quandt, Park, and Glejser tests. The simulation generates data under both homoscedastic and heteroscedastic error structures, with sample sizes and levels of heteroscedasticity varied across the experimental conditions. Performance is assessed using two principal measures, namely empirical Type I error rates and statistical power. This unified simulation framework enables an objective comparison of the strengths and limitations of each procedure under identical experimental conditions

3.0 Simulation Procedure

Consider a regression model with two variables:

$Y_i = \beta_1 + \beta_2 X_i + u_i$

Where σ_i^2 is positively related to X_i

$\sigma_i^2 = \sigma^2 X_i^2$

Where σ^2 is a constant.

Assume $\sigma_i^2 = \sigma^2 X_i^2$ reveals that σ_i^2 is proportional to the square of the X variable.

If $\sigma_i^2 = \sigma^2 X_i^2$ is suitable, it suggests that σ_i^2 increases with larger values of Xi. If this is the case, it implies that heteroscedasticity is probably present in the model.

Also, consider the datasets with sample sizes of (n = 20, 30, 40, 50, 60, 70, 80, 90,) and (100) were simulated under homoscedastic conditions and under low, mild, and high levels of heteroscedasticity. For each scenario, the Breusch–Pagan, White, Goldfeld–Quandt, Park, and Glejser tests were applied to the simulated regression residuals over 1,000 replications.

Under homoscedastic conditions, the empirical Type I error rate was obtained from the proportion of simulation runs that resulted in rejection of the null hypothesis of constant error variance. When heteroscedasticity was introduced, empirical power was measured by the proportion of simulations in which the null hypothesis was correctly rejected. The simulations were carried out at significance levels of α = 0.05 and α = 0.01.

The rejection frequencies were subsequently converted to percentages to facilitate comparison of the finite-sample performance of the five tests. A higher empirical power reflects a greater ability of a test to detect heteroscedasticity, whereas Type I error rates that remain close to the specified nominal significance level indicate better control of false rejections.

Table 1: Empirical Type I Error at an alpha level of 0.05

Sample size	Breusch – Pagan	White	Gold-field	Park	Glejser
20.0	4.20	3.10	4.30	5.30	7.10
30.0	5.10	5.70	5.70	4.60	8.80
40.0	4.60	5.50	5.20	5.70	8.50
50.0	6.10	4.70	4.70	4.80	10.00
60.0	4.40	4.60	4.80	4.90	8.60
70.0	4.00	5.20	3.80	4.90	9.60
80.0	4.40	4.20	5.30	3.50	8.70
90.0	5.50	5.40	5.40	4.50	9.80
100.0	5.40	4.30	4.80	5.90	10.50

Fig, 1: Empirical Type I Error at $\alpha = 0.05$

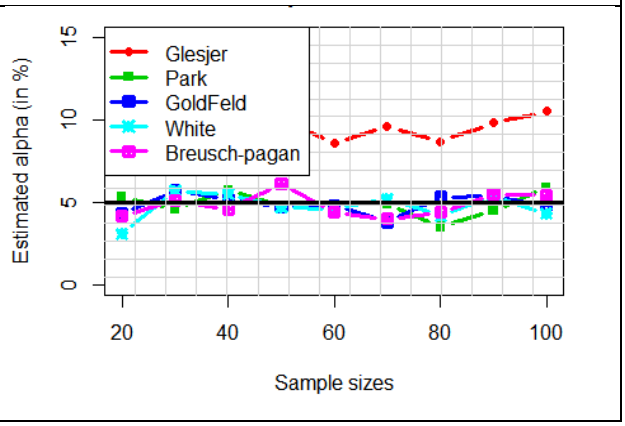


Table 1 and Fig. .1 present the empirical Type I error rates of the five heteroscedasticity tests under homoscedasticity at the 5% significance level. Each test was applied to the simulated regression residuals over 1,000 replications for each sample size. The Breusch–Pagan, White, Goldfeld–Quandt, and Park tests generally produce Type I error rates close to the nominal 5% level, indicating satisfactory control of false rejections. In contrast, the Glejser test consistently records higher-than-nominal Type I error rates, suggesting an increased tendency to falsely detect heteroscedasticity when the error variance is constant. Therefore, results from the Glejser test should be interpreted with caution.

Table 2: Empirical Type I Error at α =0.01

Sample size	Breusch –Pagan	White	Gold-field	Park	Glejser
20.0	0.50	0.90	0.70	0.70	1.50
30.0	1.20	0.50	1.20	2.00	2.70
40.0	0.60	0.60	1.20	1.00	1.80
50.0	1.20	0.70	0.60	0.70	2.60
60.0	0.70	0.40	1.00	1.30	1.80
70.0	0.80	0.50	1.60	0.90	1.60
80.0	1.20	1.30	1.40	1.00	3.90
90.0	0.50	1.20	0.60	1.00	1.80
100.0	0.80	0.80	0.60	0.40	1.50

Fig. 2: Empirical Type I Error at α = 0.01

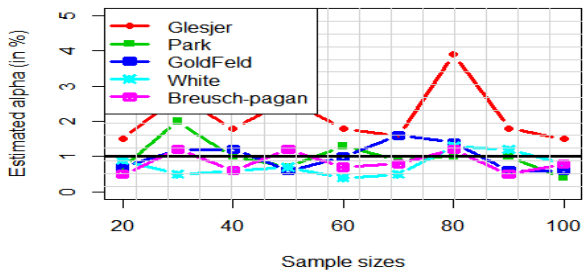


Table 2 and Fig. 2 show the empirical Type I error rates obtained for the selected tests at the 1% significance level under homoscedastic conditions. Here, the empirical Type I error rate refers to the proportion of simulation runs in which the true null hypothesis was rejected incorrectly.

For a test to be considered well calibrated, its rejection rate should remain reasonably close to the nominal 1% significance level. The results indicate that the Breusch–Pagan, White, Goldfeld–Quandt, and Park tests generally yield Type I error rates close to the specified level, suggesting satisfactory control of false rejections. The Glejser test, however, consistently produces higher rejection rates than the nominal level, pointing to an increased false-positive rate. Its results should therefore be interpreted with caution, particularly in applications where strict control of Type I error is required.

Table 3: Empirical Power at low heteroscedasticity level at α = 0.05

Sample size	Breusch –Pagan	White	Gold-field	Park	Glejser
20.0	22.40	14.20	5.70	19.50	28.90
30.0	37.80	26.50	5.90	30.10	49.60
40.0	50.10	37.20	5.80	40.70	62.20
50.0	61.90	51.20	6.10	47.60	75.50
60.0	71.90	63.10	7.20	58.20	85.40
70.0	78.30	73.20	5.60	65.90	91.70
80.0	85.00	82.00	6.70	71.70	95.70
90.0	89.90	88.20	6.60	80.00	96.90
100.0	93.60	93.10	4.50	83.40	98.60

Fig. 3: Empirical Power at low Heteroscedasticity level at α = 0.05

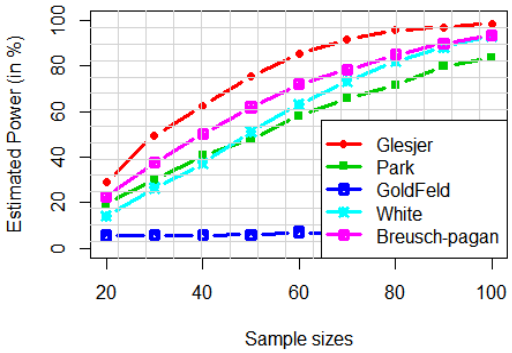


Table 3 and Fig. 3 report the empirical power of the selected heteroscedasticity tests under low heteroscedasticity. For each sample size, the tests were applied to the model residuals across 1,000 simulation replications.

With the exception of the Goldfeld–Quandt test, the empirical power of the procedures generally increases as the sample size becomes larger. This suggests that the ability to detect heteroscedasticity improves when more observations are available. Among the tests considered, the Glejser test produces the highest empirical power, indicating the greatest sensitivity to low levels of heteroscedasticity under the conditions examined in the simulation.

Table 4: Empirical Power at low heteroscedasticity level at α = 0.01

Sample size	Breusch –Pagan	White	Gold-field	Park	Glejser
20.0	5.60	2.90	1.30	7.80	10.00
30.0	11.30	6.60	1.70	12.30	19.00
40.0	22.30	13.20	1.40	17.50	31.90
50.0	33.10	20.80	1.70	26.20	46.80
60.0	44.70	29.90	0.90	32.00	57.70
70.0	53.00	38.60	1.30	39.70	70.50
80.0	59.00	46.90	1.70	47.20	75.50
90.0	71.30	63.00	1.50	54.90	85.40
100.0	75.90	68.40	1.50	62.30	89.00

Fig. 4: Empirical Power at low Heteroscedasticity level at α = 0.01

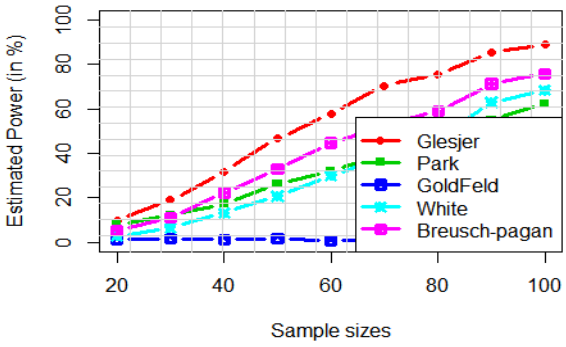


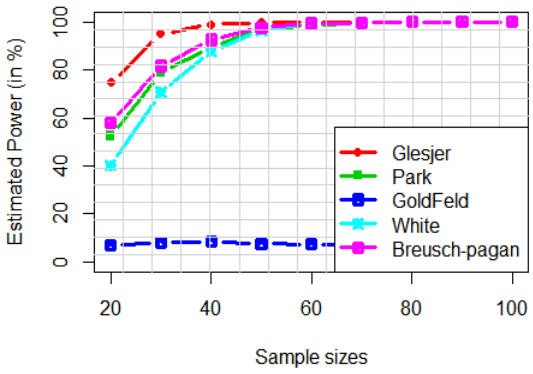
Table .4 and Fig. 4 report the empirical power of the selected tests under low heteroscedasticity at α = 0.01. For each sample size, the analysis was based on 1,000 simulation replications.

The Goldfeld–Quandt test displays an irregular pattern of empirical power and remains relatively low across the sample sizes considered. This suggests that the test has limited ability to identify mild departures from homoscedasticity under the simulated conditions. The other four procedures—Breusch–Pagan, White, Park, and Glejser—show a more consistent upward pattern in power as the sample size increases. Their performance therefore suggests greater reliability and sensitivity in detecting low levels of heteroscedasticity.

Table 5: Empirical Power at mild heteroscedasticity level at α = 0.05

Sample size	Breusch –Pagan	White	Gold-field	Park	Glejser
20.0	57.90	40.20	7.10	52.60	74.80
30.0	81.70	70.80	8.00	78.90	95.10
40.0	92.70	87.90	8.50	89.60	99.10
50.0	97.80	96.70	7.60	97.30	99.80
60.0	99.40	99.50	7.30	98.50	100.00
70.0	99.70	99.60	7.40	99.70	100.00
80.0	100.00	100.00	8.20	99.80	100.00
90.0	100.00	100.00	7.70	100.00	100.00
100.0	100.00	100.00	8.20	100.00	100.00

Fig. 5: Empirical Power at Mild Heteroscedasticity level at α = 0.05

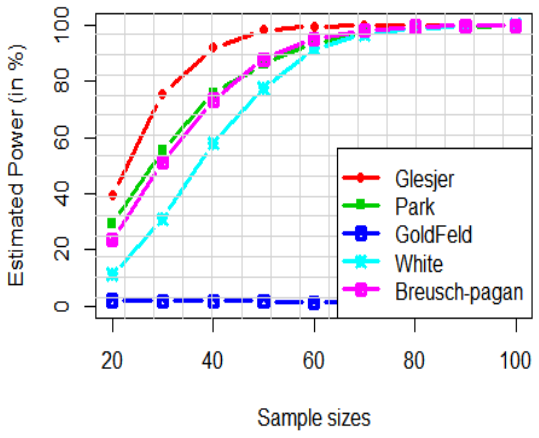


The empirical power of the chosen tests under mild heteroscedasticity is shown in Table 5 and Fig/ 5. For every sample size, 1,000 replications of each test were run on simulated model residuals. Every procedure achieves at least 80% power at n = 40, with the exception of the Goldfeld–Quandt test. Although the Glejser test yields the maximum power, care should be taken due to its propensity for exaggerated Type I error. The Breusch-Pagan test seems to have the best performance for identifying mild heteroscedasticity at n = 40 when power and error control are taken into account. The tests often show comparable power patterns for sample sizes larger than 40, but the Goldfeld-Quandt test is still relatively weaker.

Table 6: Empirical Power at mild heteroscedasticity level at α = 0.01

Sample size	Breusch –Pagan	White	Gold-field	Park	Glejser
20.0	23.60	11.40	2.00	29.20	39.50
30.0	51.30	31.10	1.80	55.30	75.40
40.0	73.40	57.90	1.90	75.70	92.00
50.0	87.90	77.60	1.90	86.10	98.30
60.0	95.20	91.70	1.10	93.70	99.50
70.0	98.00	96.70	1.90	96.80	99.80
80.0	99.40	99.00	1.70	99.40	100.00
90.0	99.80	99.60	1.50	99.60	100.00
100.0	99.80	100.00	2.40	99.80	100.00

Fig. .6: Empirical Power at Mild Heteroscedasticity level at α = 0.01



The empirical power of the chosen heteroscedasticity tests under moderate heteroscedasticity at α = 0.01 is displayed in Table 6 and Fig. .6. Power was calculated as the percentage of null hypothesis rejections, and each sample size was simulated 1,000 times. The methods attain sufficient power at roughly n = 60, with the exception of the Goldfeld–Quandt test, suggesting a positive correlation between sample size and statistical power. Under the simulated conditions, the Glejser test exhibits higher sensitivity to mild heteroscedasticity, consistently recording the highest power among the tests.

Table 7: Empirical Power at high heteroscedasticity level at α = 0.05

Sample size	Breusch – Pagan	White	Gold-field	Park	Glejser
20.0	93.60	80.80	9.70	90.60	99.30
30.0	99.70	97.50	9.30	98.50	100.00
40.0	100.00	99.60	9.40	99.60	100.00
50.0	100.00	99.90	12.10	99.90	100.00
60.0	100.00	100.00	9.90	100.00	100.00
70.0	100.00	100.00	10.60	100.00	100.00
80.0	100.00	100.00	12.60	100.00	100.00
90.0	100.00	100.00	11.20	100.00	100.00
100.0	100.00	100.00	13.70	100.00	100.00

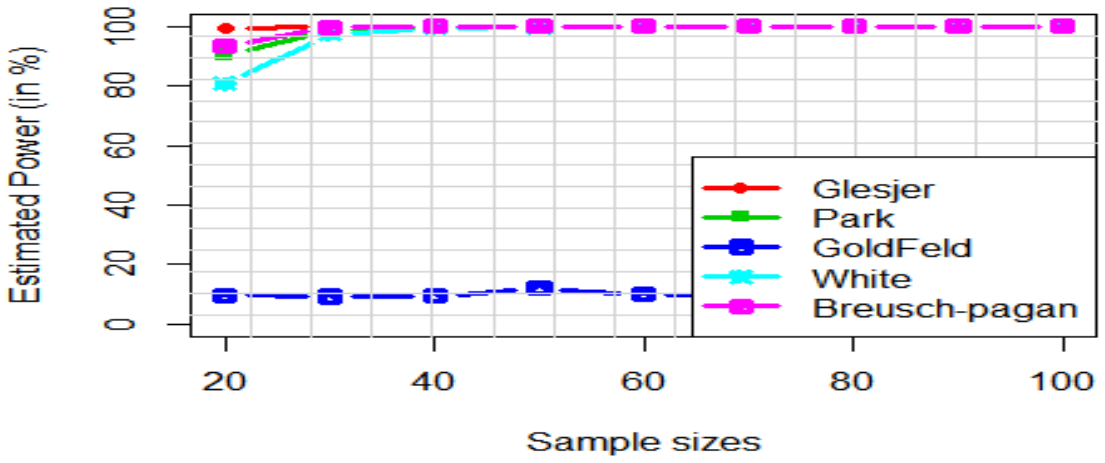


Fig. .7: Empirical Power at High Heteroscedasticity level at α = 0.05

The empirical power of the chosen heteroscedasticity tests at high heteroscedasticity at α = 0.5 is shown in Table 4.7 and Fig/ 7. Power was determined as the percentage of null hypothesis rejections, and each sample size was simulated 1,000 times. All methods, with the exception of the Goldfeld–Quandt test, attain power above 80%, suggesting a high likelihood of identifying strong heteroscedasticity. Among the tests, the Glejser test has the largest power, the strongest detection performance under simulated settings, and the highest sensitivity to significant departures from homoscedasticity.

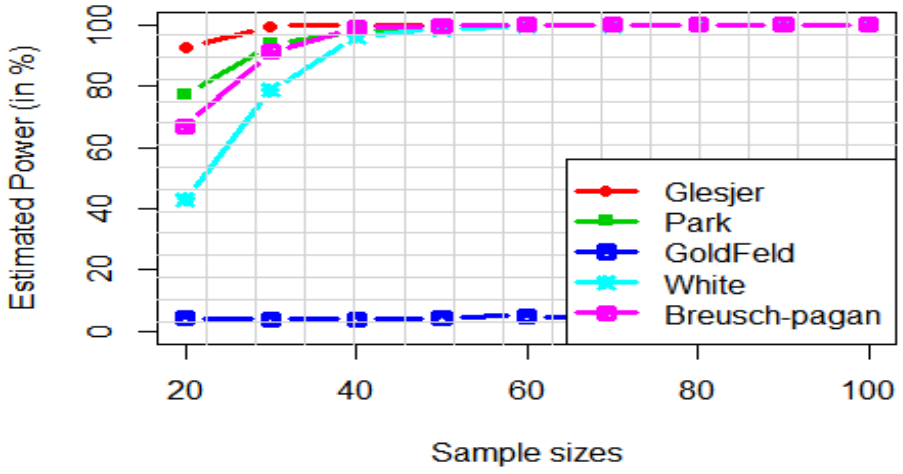


Fig8. : Empirical Power at High Heteroscedasticity level at α = 0.01

Table 8: Empirical Power at high heteroscedasticity level at α = 0.01

Sample size (n)	Breusch–Pagan (%)	White (%)	Goldfeld–Quandt (%)	Park (%)	Glejser (%)
20	66.90	43.00	4.20	77.30	92.50
30	91.20	78.70	3.90	93.80	99.60
40	99.10	96.20	4.00	98.00	100.00

50	99.70	98.70	4.40	99.60	100.00
60	100.00	99.80	5.10	100.00	100.00
70	100.00	99.90	3.80	100.00	100.00
80	100.00	100.00	4.30	100.00	100.00
90	100.00	100.00	5.30	100.00	100.00
100	100.00	100.00	4.20	100.00	100.00

The five heteroscedasticity tests' empirical power is shown in Table 8 and Fig. 8 under strong heteroscedasticity at $\alpha = 0.01$. Power was calculated as the percentage of rejections of the null hypothesis after 1,000 simulations of each sample size.
The findings demonstrate that, with the exception of the Goldfeld-Quandt test, the power of every test rises with sample size starting at $n = 20$. The Breusch-Pagan, White, Park, and Glejser tests demonstrate a good capacity to detect high heteroscedasticity, with sufficient power starting at about $n = 40$. On the other hand, for all sample sizes taken into consideration, the Goldfeld-Quandt test continues to be somewhat poor.

4.0 Summary of Results

The Breusch-Pagan, White, Goldfeld-Quandt, Park, and Glejser tests are compared across sample sizes, heteroscedasticity levels, and significance levels of 0.05 and 0.01 in Tables 4.1–4.8.
The Glejser test routinely displays inflated Type I error rates under homoscedasticity, while the Breusch–Pagan, White, Goldfeld–Quandt, and Park tests typically preserve Type I error rates near their nominal levels. Statistical power under heteroscedasticity often rises as sample size and heteroscedasticity severity do. The Park test performs mediocrely, whereas the Glejser test records the highest power in the majority of scenarios, followed by the Breusch-Pagan and White tests. When modest heteroscedasticity is present, the Goldfeld-Quandt test consistently yields the lowest power.
In terms of controlling Type I error, the Glejser test is the most sensitive overall, but it is less successful. Power and error-rate control are best balanced by the Breusch–Pagan and White tests, however the Goldfeld–Quandt test does not perform well. These results imply that for the practical identification of heteroscedasticity, the Breusch–Pagan and White tests would be more dependable options.

4.2 Discussion of Results

The Monte Carlo results demonstrate that the performance of heteroscedasticity tests varies with sample size and the severity of heteroscedasticity. Generally, statistical power increases as sample size increases, confirming the greater ability of larger samples to detect departures from homoscedasticity.
The Breusch–Pagan, White, Goldfeld–Quandt, and Park tests generally maintain Type I error rates close to their nominal levels, indicating satisfactory control of false rejections. In contrast, the Glejser test consistently records inflated Type I error rates, although it achieves the highest statistical power across most heteroscedasticity scenarios. Thus, its superior sensitivity is accompanied by a greater risk of false positives.
While retaining superior Type I error control, the Breusch–Pagan and White tests show robust and consistent power across various heteroscedasticity levels. The Park test performs moderately but steadily, especially as sample size grows. The Goldfeld–Quandt test, on the other hand, shows the lowest power in the majority of situations, indicating minimal efficacy under the conditions studied. Overall, the results show that the Breusch–Pagan and White tests offer the most balanced performance, combining adequate Type I error

management with strong detection power. When sensitivity is the main issue, the Glejser test is better, but the Goldfeld-Quandt test seems to be less accurate at identifying heteroscedasticity.

5.0 Conclusion

The Breusch–Pagan, White, Goldfeld–Quandt, Park, and Glejser tests are five commonly used heteroscedasticity tests. This work employed Monte Carlo simulations to examine the performance of these tests across various sample sizes, heteroscedasticity levels, and significance levels of 0.05 and 0.01. Empirical Type I error rates and statistical power were used to assess their performance.

The findings show significant variations between the tests. Strong sensitivity to heteroscedasticity is demonstrated by the Glejser test, which consistently yields the highest statistical power. Its comparatively high Type I error rates, however, point to a higher chance of false positives when homoscedasticity is present. As sample size and heteroscedasticity rise, the Breusch-Pagan and White tests offer a more balanced performance by combining moderate to high statistical power with adequate Type I error control. While the Goldfeld–Quandt test typically records the lowest power and seems less appropriate for general applications, the Park test shows moderate and reliable performance.

Overall, the results highlight how the reliability of regression diagnostics can be significantly impacted by the heteroscedasticity test selection. The Glejser test may be helpful when maximum sensitivity is required, provided that its high Type I error rate is taken into account. The Breusch–Pagan and White tests are advised when striking a balance between detection power and error management is crucial. Therefore, it is advised to use many diagnostic tests to improve the validity of results about heteroscedasticity.

Researchers and practitioners in economics, statistics, finance, and other domains that depend on regression analysis for empirical inference can benefit from the study's helpful empirical recommendations.

6.0 References

Akewugberu, H. O., Umar, S. M., Musa, U. M., Ishaaq, O. O., Ibrahim, A., Osi, A. A., & Ganiyat, A. F. (2024). Monte Carlo evaluation of White's test for detecting heteroscedasticity in generalized linear models. *FUDMA Journal of Sciences,* 8, 6, pp., 309–314. https://doi.org/10.33003/fjs-2024-0806-3040

Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica,* 47, 5, pp. 1287–1294. https://doi.org/10.2307/1911963

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Farrar, T., Blignaut, R., Luus, R., & Steel, S. (2025). A review and comparison of methods of testing for heteroskedasticity in the linear regression model. *Journal of Applied Statistics,* 52, 16, pp. 3121–3150. https://doi.org/10.1080/02664763.2025.2575038

Gentle, J. E. (2003). *Random number generation and Monte Carlo methods* (2nd ed.). Springer.

Glejser, H. (1969). A new test for heteroskedasticity. *Journal of the American Statistical Association,* 64, 325, pp. 316–323. https://doi.org/10.1080/01621459.1969.10500976

Goldfeld, S. M., & Quandt, R. E. (1965). Some tests for homoscedasticity. *Journal of the American Statistical Association,* 60, 310, pp. 539–547. https://doi.org/10.1080/01621459.1965.10480811

Greene, W. H. (2020). *Econometric analysis* (8th ed.). Pearson.

Gujarati, D. N., & Porter, D. C. (2009). *Basic econometrics* (5th ed.). McGraw-Hill Education.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to linear regression analysis* (6th ed.). John Wiley & Sons.

Onifade, O. C., & Olanrewaju, S. O. (2020). Investigating performances of some statistical tests for heteroscedasticity assumption in generalized linear model: A Monte Carlo simulations study. *Open Journal of Statistics,* 10, 3, pp. 453–493. https://doi.org/10.4236/ojs.2020.103029

Park, R. E. (1966). Estimation with heteroscedastic error terms. *Econometrica,* 34, 4, 888. https://doi.org/10.2307/1910108

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity. *Econometrica,* 48, 4, pp. 817–838. https://doi.org/10.2307/1912934

Wooldridge, J. M. (2020). *Introductory econometrics: A modern approach* (7th ed.). Cengage Learning.

Yang, K., Tu, J., Chen, T., & others. (2019). Homoscedasticity: An overlooked critical assumption for linear regression. *General Psychiatry, 32*(5), e100148, doi: 10.1136/gpsych-2019-100148

Declarations

Consent for Publication

Not applicable.

Availability of Data

The data used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Competing Interests

The authors declare that they have no competing interests and no conflict of interest regarding the publication of this manuscript.

Ethical Consideration

Not applicable.

Funding

The authors declared no source of external funding.

Authors' Contributions

Rita Nneka Nwaka contributed to conceptualization, methodology, Monte Carlo simulation, data analysis, interpretation, and manuscript preparation. Alaba Akinleye Obabire contributed to methodology, simulation validation, statistical analysis, literature review, interpretation, and critical manuscript revision. Both authors reviewed, revised, and approved the final manuscript and accepted responsibility for the integrity of the work.