J. Educ. Soc. Stud. 4(2) 2023. 313-329

Available Online

Journal of Education and Social Studies

ISSN: 2789-8075 (Online), 2789-8067 (Print)

http://www.scienceimpactpub.com/jess

TESTING FOR HETEROSKEDASTICITY IN THE PRESENCE OF OUTLIERS

Muhammad Raza [1,*], Mumtaz Ahmed [2], Shahid Razzaque [3] and Hafsa Hina [3]

[1] *Pakistan Institute of Development of Economics (PIDE), Islamabad, Pakistan*
[2] *Department of Economics, COMSATS University Islamabad (CUI), Islamabad, Pakistan*
[3] *Department of Economics, Pakistan Institute of Development of Economics (PIDE), Islamabad, Pakistan*

ABSTRACT

Regression analysis is prone to the issue of heteroscedastic data in a variety of real-world cases, including macroeconomic data. Thus, it is crucial to test the data for possible heteroskedasticity. This is important because if the data is found to be heteroskedastic. Then, this may seriously impact the regression analysis's estimation and testing phase. It is emphasized that real data may contain one or more outliers; thus, it is important to use the appropriate test to test for the presence of heteroskedasticity when there is evidence of outliers present in the data. A commonly used test to test for heteroskedasticity is the Goldfeld-Quandt (GQ) test. However, its performance becomes questionable when the data contains one or more outliers. A modified version of GQ (MGQ) is available in the literature that considers the issue of outliers into account while testing for possible heteroskedasticity in the data. Though this is a good addition to an existing stream of heteroskedasticity tests, little attention is given to literature regarding its applicability. The present study takes the lead and makes the case that practitioners should use this newly proposed MGQ test. Various real-world cases using popular data sets are discussed, indicating the superiority of MGQ over the conventional GQ test when there are outliers in the data. The findings based on real-world data indicate that practitioners should use the MGQ test whenever the data contains outliers to avoid misleading conclusions.

Keywords: Real data; Outliers; Heteroscedasticity; Least trimmed squares robust.
** Email: muhammadraza_16@pide.edu.pk*

https://doi.org/10.52223/jess.2023.4209
Received: February 12, 2023; Revised: April 18, 2023; Accepted: April 26, 2023

INTRODUCTION

The concept of heteroskedasticity dates back to the early days of regression analysis, which was pioneered by Sir Francis Galton in the late 19th century. Galton (1883) used regression analysis to study the relationship between the height of fathers and their sons and found that the heights of sons tended to be closer to the average height of the population than their fathers' heights. However, Galton did not explicitly consider the issue of heteroskedasticity in his analysis. It wasn't until the mid-20th century that heteroskedasticity became recognized as an issue in regression analysis. The term "heteroskedasticity" was first coined by the economist Ragnar Frisch to describe situations where the variance of the dependent variable varied across observations (Frisch & Waugh, 1933). Heteroskedasticity is more common in cross-sectional data because cross-sectional data usually includes observations from different groups, regions, or categories that may have different levels of variability in the dependent variable (Wooldridge, 2002). This can result in varying levels of variance of the error term in the regression model, leading to heteroskedasticity. Heteroskedasticity refers to the situation where the variance of the dependent variable (or the error term) in a regression model varies across different groups or observations within a dataset (Greene, 2003).

Measurement error can also contribute to heteroskedasticity in cross-sectional data. If there is more measurement error in some observations compared to others, this can result in varying levels of variability in the dependent variable. Extreme values can also contribute to heteroskedasticity in cross-sectional data. If some observations have much larger or smaller values than others, this can result in varying levels of variability in the dependent variable. When the relationship between the independent and dependent variables is nonlinear, this can also result in heteroskedasticity in cross-sectional data (Wooldridge, 2002).

Heteroskedasticity can lead to incorrect statistical tests in regression analysis. It can lead to biased variance estimates in regression analysis. This is because the ordinary least squares (OLS) method assumes that the variance of the dependent variable is constant across all observations. When this assumption is violated, OLS estimator may produce biased covariance estimates (Gujarati, 2022; Ramsey, 1969). This means that the standard errors of the coefficient estimates may be larger than they should be, making it harder to detect statistically significant relationships. This can reduce the power of statistical tests and increase the probability of type II errors. This is because the OLS estimator places more weight on observations with higher variance, leading to a larger standard error and a smaller coefficient estimate. Therefore, it is important to test for heteroskedasticity and, if present, to use appropriate methods to correct for it (Wooldridge, 2002).

The detection of heteroskedasticity in regression analysis has a history, and researchers have developed many different tests and methods for detecting this issue. The detection of heteroskedasticity in regression analysis dates back to the mid-20th century. In the 1950s and 1960s, researchers developed several tests for detecting heteroskedasticity, including the Park test (Park, 1966), the Glejser test (Glejser, 1969), the Breusch-Pagan test (Breusch-Pagan, 1979), and the White and the Goldfeld-Quandt test (Goldfeld-Quandt, 1965).Park (1966) proposed a test for heteroskedasticity that involves regressing the absolute residuals from the original regression on the independent variables. If heteroskedasticity is present, the absolute residuals will be correlated with the independent variables, leading to a significant coefficient in the Park test. The Glejser test, proposed by Herbert Glejser (1969), is another test for heteroskedasticity that involves regressing the absolute residuals from the original regression on one or more independent variables that are thought to be related to the variance of the dependent variable. If heteroskedasticity is present, the absolute residuals will be correlated with the independent variables, leading to a significant coefficient in the Glejser test. The Breusch-Pagan test, proposed by Breusch and Pagan (1979), is one of the most commonly used tests for detecting heteroskedasticity. It is based on regressing the squared residuals from the original regression on the independent variables. If heteroskedasticity is present, the squared residuals will be correlated with the independent variables, leading to a significant coefficient in the Breusch-Pagan test. The White test, proposed by Halbert White (White, 1980), is another commonly used test for heteroskedasticity. It involves regressing the squared residuals from the original regression on the independent variables and their cross-products. If heteroskedasticity is present, the squared residuals will be correlated with the independent variables and their cross-products, leading to a significant coefficient in the White test. The Goldfeld-Quandt (GQ) test, proposed by David Goldfeld and Richard Quandt (Goldfeld & Quandt, 1965), is a test for heteroskedasticity that involves dividing the sample into two subgroups based on a particular independent variable and comparing the variances of the residuals in the two subgroups. If heteroskedasticity is present, the variances of the residuals in the two subgroups will be significantly different. The Harvey-Collier test, proposed by David Harvey and Paul Collier (Harvey & Collier, 1977), is a test for heteroskedasticity that involves regressing the absolute residuals from the original regression on a set of predetermined independent variables. If heteroskedasticity is present, the absolute residuals will be correlated with the predetermined independent variables, leading to a significant coefficient in the Harvey-Collier test. These tests have varying degrees of power and efficiency in detecting heteroskedasticity, and the choice of test often depends on the specific characteristics of the data and the research question at hand (Harvey & Phillips, 1974).

Researchers have also developed graphical methods for detecting heteroskedasticity, such as residual plots and scale-location plots (Rosopa et al., 2013). These methods involve plotting the residuals or the absolute residuals against the predicted values or the independent variables and visually examining whether there is a pattern of increasing or decreasing variance (Evans & King, 1988).

The GQ is an excellent tool for quickly determining if two or more samples have significantly different distributions. Furthermore, because the test is based on non-parametric assumptions and does not require the data to follow an exact distribution, it is well-suited for analyzing data sets of varying complexity. The Goldfeld-Quandt test (Goldfeld-Quandt, 1972) is particularly useful in cases where there is a suspected change in the variance of the residuals across different subsets of the data, such as with financial time-series data or cross-sectional data with a known grouping structure (Zaman, 1994). A comparison of six commonly used tests for the detection of heteroskedasticity shows that Harrison-McCabe test is the most powerful test. While the White test has the least power in above all mentioned tests (Uyanto, 2019). The literature highlights that heteroskedasticity is an econometric problem that impacts test procedure and estimation; hence, identifying the problem is crucial to resolve the issue (Abdul-Hameed & Matanmi, 2021). As the presence of an outlier in a data set and subsequent identification of heteroskedasticity lead to biased results, the study further suggested a modified version of the Breusch-pagan test for the identification of heteroskedasticity in the presence of outliers.

The Goldfeld-Quandt (GQ) test is a classical test used to detect heteroscedasticity in regression models. The test assumes that the variance of the errors in the regression model is a function of one or more of the independent variables, and it does not require any assumptions about the distribution of the errors or the functional form of the heteroscedasticity. The Modified Goldfeld-Quandt (MGQ) test, proposed by Rana et al. (2008), is an extension of the GQ test that allows for nonlinearity in heteroscedasticity.

The GQ test may fail in the presence of outliers or influential observations, which can bias the test results and lead to incorrect conclusions about the presence or absence of heteroscedasticity. The robust modification of the GQ test by Rana et al. (2008) addresses this issue by using a trimmed estimator to estimate the variances of the residuals replacing the OLS with the least trimmed squares (LTS) approach. The trimmed estimator is less sensitive to outliers and is designed to remove the influence of extreme observations on the test results. This approach provides a more robust approach to detecting heteroscedasticity and can improve the accuracy of regression analysis in practical applications. The use of robust statistical methods in detecting heteroscedasticity is particularly important in applications where the presence of outliers or influential observations is likely.

This study is to highlight the importance of MGQ in the presence of an outlier rather than the original GQ test. Taken several statistical examples of published articles to promote the MGQ test over the GQ test. So, MGQ will be used in case of outlier data in the future.

Historical Discussion of Heteroskedasticity Detection Tests

Heteroskedasticity tests are used to detect the presence of heteroskedasticity in a data set. Heteroskedasticity is an econometric term used to describe the presence of non-constant variance in the residuals of a regression. Heteroskedasticity can occur when the error terms of a regression are not independently and identically distributed (i.e., they are not homoskedastic). Heteroskedasticity is a form of non-constant variance in the residuals of a regression analysis. It is estimated that approximately 40% of all regression analyses suffer from heteroskedasticity. Graphical detection methods include plotting the residuals versus the independent variables and studying the pattern. If the points become more dispersed as the values of the independent variables increase, then heteroskedasticity may be present. Another graphical detection method is plotting the residuals versus the fitted values from the regression line. If the points become more dispersed as the fitted values increase, then heteroskedasticity may be present.

Statistical detection methods for heteroskedasticity include the Breusch-Pagan test, the White test, and the Goldfeld-Quandt test. The Breusch-Pagan test is a general test for heteroskedasticity and is used to test whether the variance of the residuals is related to the values of the independent variables. The White test is a more powerful version of the Breusch-Pagan test, which is used to test whether the variance of the residuals is related to the squared values of the independent variables. The Goldfeld-Quandt test is used to test whether the variance of the residuals is related to the fitted values from the regression line (Asteriou & Hall, 2015, 2017).

The Breusch-Pagan test is a test for heteroscedasticity in regression models (Breusch & Pagan, 1979). It is a form of the likelihood ratio test and is used to assess the null hypothesis that the variance of an error term is homogeneous across all observations. The test statistic is calculated by regressing the squared residuals on the independent variables and testing the significance of the resulting F-statistic. This test is used to detect heteroskedasticity in a linear regression model. It examines the correlation between the squared residuals and the independent variables. The Breusch-Pagan statistic is calculated by regressing the squared residuals on the independent variables and then testing the resulting F-statistic for significance. If the F-statistic is significant, then the model suffers from heteroskedasticity.

The Goldfeld-Quandt test is a test for heteroscedasticity in regression models (Goldfeld & Quandt, 1965, 1973). The test statistic is calculated by regressing the squared residuals on the independent variables and testing the significance of the resulting F-statistic. This test is used to detect heteroskedasticity in a linear regression model. It examines the correlation between the residuals and the independent variables. The Goldfeld-Quandt statistic is calculated by dividing the data into two groups based on the median value of the independent variables and then testing the difference in the variance of the residuals for the two groups for significance. If the variance in the two groups is significantly different, then the model suffers from heteroskedasticity.

The White test is a test for heteroscedasticity in regression models (White, 1980). Unlike the Breusch-Pagan test, the White test does not require the assumption of normally distributed errors. The test statistic is calculated by regressing the squared residuals on the independent variables and testing the significance of the resulting F-statistic. This test is used to detect heteroskedasticity in a linear regression model. It examines the correlation between the squared residuals and the independent variables. The White statistic is calculated by regressing the squared residuals on the independent variables and then testing the resulting t-statistic for significance. If the t-statistic is significant, then the model suffers from heteroskedasticity.

The Park test is a test for heteroscedasticity in regression models (Park, 1966). The test statistic is calculated by regressing the squared residuals on the independent variables and testing the significance of the resulting F-statistic. This test is used to detect heteroskedasticity in a linear regression model. It examines the correlation between the residuals and the independent variables. The Park statistic is calculated by dividing the data into two groups based on the median value of the independent variables and then testing the difference in the variance of the residuals for the two groups for significance. If the variance in the two groups is significantly different, then the model suffers from heteroskedasticity.

The Koenker-Bassett test is used to detect heteroskedasticity in a linear regression model. It examines the correlation between the absolute value of the residuals and the independent variables. The Koenker-Bassett statistic is calculated by regressing the absolute value of the residuals on the independent variables and then testing the resulting F-statistic for significance. If the F-statistic is significant, then the model suffers from heteroskedasticity.

The Weighted Least Squares (WLS) test is used to detect heteroskedasticity in a linear regression model. It examines the correlation between the residuals and the independent variables. The WLS statistic is calculated by weighting the residuals according to the size of the independent variables and then testing the resulting F-statistic for significance. If the F-statistic is significant, then the model suffers from heteroskedasticity.

The Generalized Least Squares (GLS) test is used to detect heteroskedasticity in a linear regression model. It examines the correlation between the residuals and the independent variables. The GLS statistic is calculated by weighting the residuals according to the size of the independent variables and then testing the resulting F-statistic for significance. If the F-statistic is significant, the model suffers from heteroskedasticity (Kariya & Kurata, 2004).

The Breusch-Godfrey test is a test for heteroscedasticity in regression models (Breusch & Godfrey, 1978). The test statistic is calculated by regressing the squared residuals on the independent variables and testing the significance of the resulting F-statistic. The Bartlett test is a test for heteroscedasticity in regression models (Bartlett, 1937). The test statistic is calculated by regressing the squared residuals on the independent variables and testing the significance of the resulting F-statistic. The Harvey-Collier test is a test for heteroscedasticity in regression models (Harvey & Collier, 1977). The test statistic is calculated by regressing the squared residuals on the independent variables and testing the significance of the resulting F-statistic.

The Glejser test is a test for heteroscedasticity in regression models (Glejser, 1969). The test statistic is calculated by regressing the squared residuals on the independent variables and testing the significance of the resulting F-statistic. The Engle-Granger test is a test for heteroscedasticity in regression models (Engle & Granger, 1987). The test statistic is calculated by regressing the squared residuals on the independent variables and testing the significance of the resulting F-statistic (Andrew, 1993; Andrew & Ploberger, 1994).

The Regression Model and Tests for Heteroskedasticity

Consider a multiple linear regression model:

$$Y = \alpha + X\beta + \mu \qquad (1)$$

Where, Y is a $N\times 1$ vector containing observations on the dependent variable, β is $K\times 1$ vector of unknown parameters, X is a N×K matrix of the regressors, and μ is an unobserved error term assumed to be independent with mean zero and non-constant variance, i.e., Var(ε_t)= $\sigma_n^2, n = 1,2,..,N$. Note that $\mu \sim N(0,\Sigma), where, \Sigma = diag(\sigma_n^2),\ n = 1,2,\ldots,N$.

The usual OLS estimate of true parameter β is: $\hat{\beta}_{OLS} = (X'X)^{-1}XY$ with covariance matrix, $Cov(\hat{\beta}_{OLS}) = \hat{\sigma}_n^2(X'X)^{-1}$ (since OLS assumes homoscedasticity), where, $\hat{\sigma}_n^2 = \frac{RSS}{N-K}$, , where RSS is the residuals sum of squares obtained from estimating regression in equation (1) above.

When the assumption of homoscedasticity gets violated, the OLS estimates though remain unbiased and consistent but no-longer efficient. In addition, the covariance matrix of OLS estimates, i.e., $Cov(\hat{\beta}_{OLS})$ becomes biased inconsistent and this leads to wrong n & F statistics and related confidence intervals thus, the significance of regressors also gets affected; specifically, a significant regressor may appear insignificant and vice versa. Thus, it is very important to test the regression errors for homoscedasticity by using available tests for heteroscedasticity. If null of homoscedasticity gets rejected then one can use heteroscedasticity consistent standard errors (HCSEs) while using OLS estimates to get valid inferences regarding the significance of regressors. The true covariance matrix of OLS estimates under heteroscedasticity is: $Cov(\hat{\beta}_{OLS}) = (X'X)^{-1}X'\Sigma(X'X)^{-1}$. Note that Σ contains unknown parameters, σ_n^2, n=1,2,...,N. so usually it is replaced with its estimate, $\hat{\Sigma}$ leading to estimated covariance matrix as: $Cov(\hat{\beta}_{OLS}) = (X'X)^{-1}X'\hat{\Sigma}(X'X)^{-1}$. Several variants of $\hat{\Sigma}$ are available in literature commonly known as heteroscedasticity consistent covariance matrix estimators (HCCMEs) (HC0 to HC5) (White, 1985; Hinkley, 1977; Horn et al., 1975; Mackinnon & White, 1985; Ahmed et al., 2017; Dutta & Zaman, 1989; Cook & Weisberg, 1983). The HCSEs are obtained by taking square root of diagonal entries of these HCCMEs.

An important challenge for the practitioner is to decide which test to use to test for the presence of heteroscedasticity. A large number of tests are available in the literature, such as Goldfeld and Quandt (1965), Glejser test (1969), Harrison–McCabe test (1979), Breusch–Pagan test (1969), White (1980), and Mackinnon and White (1985) among others. Our focus is on the GQ test and its several variants (LRGQ and MGQ). Before proceeding further, it is good to introduce first the conventional GQ test and then its variants. This is done in the following subsections:

The Goldfeld-Quandt Test

Goldfeld and Quandt (1965) proposed a test (commonly known as GQ test) to detect Heteroscedasticity in a linear regression model. This test is based on the idea of arranging the data in either ascending or descending order with respect to an identified variable with which residual variance is highly related. To set the stage for the GQ test and its variants, let's introduce a general framework which is used throughout in the discussion that follows from this point onward.

Consider a sample of N observations ranging from 1 to N ordered in such a way that variances are increasing. Divide the sample into two parts by choosing N_1 and N_2 in such a way that $1 < N_1 < N_2 < N$, and defining Y_1 and Y_2 to be $N_1 \times 1$ and $(N - N_2 + 1) \times 1$ vectors, with $T_1 \cong T/2$ and $T_2 \cong T_1 + 1$, as: $Y_1 = (y_1, y_2, \dots, y_{T_1})'$ and$Y_2 = (y_{T_2}, \dots, y_T)'$. Let X_1 and X_2 be $N_1 \times K$ and $(N - N_2 + 1) \times K$ matrices of corresponding values of the regressors and, μ_1& μ_2 be the corresponding $N_1 \times 1$ and $(N - N_2 + 1) \times 1$ error vectors assumed to follow a normal distribution with mean as zero vector and covariance matrices, $\sigma_1^2 I_{N_1}$ & $\sigma_2^2 I_{N-N_2+1}$ respectively. Consider the linear regression model separately for the two halves of the sample as:

$$Y_1 = X_1\beta_1 + \mu_1 \quad (1A)$$

$$Y_2 = X_2\beta_2 + \mu_2 \quad (1B)$$

Where, $\mu_1 = N(0, \sigma_1^2 I_{N_1})$ and $\mu_2 = N(0, \sigma_2^2 I_{N-N_2+1})$.

The focus of GQ is on testing the null hypothesis: $H_0: \gamma = 1$ against the alternative, $H_1: \gamma > 1$, where, $\gamma = \sigma_1^2/\sigma_2^2$, under the assumption that regression coefficients across the two halves are the same, i.e., $\beta_1 = \beta_2$.

Define $\hat{\beta}_1 = (X'_1X_1)^{-1}X'_1Y_1$ and $\hat{\beta}_2 = (X'_2X_2)^{-1}X'_2Y_2$ as OLS estimates of true parameters in each half of the sample and let $RSS_1^2 = \|Y_1 - X_1\hat{\beta}_1\|^2$ and $RSS_2^2 = \|y_2 - X_2\hat{\beta}_2\|^2$be the corresponding sum of OLS squared residuals. The original GQ test proposes omitting a few observations from the middle to increase the contrast between the variances in the first half of the sample and that of the second half of the sample. The GQ statistic is given by:

$$GQ = \hat{\sigma}_{2n}^2/\hat{\sigma}_{1n}^2 = \frac{RSS_2^2/df2}{RSS_1^2/df1} \quad (2)$$

The GQ is an exact test, and it follows an F-distribution under the null of homoscedasticity with $df_2 = N - (N_1 - K)$ and $df_1 = N_1 - K$ degrees of freedom and suggest rejecting the null when calculating the value of GQ-statistics (GQ_{cal}) is found to be greater than the critical value at 1%, 5%, and 10% significance levels (α). i.e.

$GQ_{cal} > F(\alpha, df_2, df_1)$

Following the brief description of a few common tests for the identification of heteroscedasticity. Yet, there is evidence that when outliers are present in the data, all of these tests suffer a significant setback. So, Rana et al. (2008) created a test that is not greatly impacted by outliers. Here, they suggest a brand test that modifies the Goldfeld-Quandt test. To replace the outlier-affected parts of the Goldfeld-Quandt test with reliable replacements.

Modified Goldfeld Quandt Test

The modified GQ (MGQ) test proposed by Rana et al. (2008) is a modification of the original Goldfeld-Quandt test for the case when there are outliers in the data. The key idea of the MGQ test is to identify the components of the GQ test that are affected by outliers, and then these are replaced by their robust alternatives to get better inferences under Heteroscedasticity.

This test works in parallel to the original GQ test, where one orders the observations with increasing variance and finds outliers via any robust Least Trimmed of Squares (LTS) method proposed by Rousseeuw and Leroy (1987) to estimate the regression models in [1A] and [1B] and then compute the deletion residuals (Imon, 2003) for the entire data set based on a fit without the points identified as outliers by the LTS fit. The modified GQ (MGQ) test is obtained as a ratio of the median of squared deletion residuals for the two halves of the entire sample, given as:

$$MGQ = \frac{MSDR_2^2/df_2}{MSDR_1^2/df_1} \qquad (3)$$

Where, $MSDR_1^2 = med\|y_1 - X_1\hat{\beta}_1\|^2$ and $MSDR_2^2 = med\|y_2 - X_2\hat{\beta}_2\|^2$ are the median of the squared deletion residuals (MSDR) for first and second half of the sample. The MGQ test follows an F distribution with $df_2 \equiv N - (N_1 - K)$ and $df_1 \equiv N_1 - K$ degrees of freedom under the null of homoscedasticity while the normality assumption holds true.

The Least Trimmed Squares (LTS) method is a robust regression technique used to fit a regression model in the presence of outliers or extreme observations (Rousseeuw & Driessen, 1999). Unlike traditional least squares regression, which is sensitive to outliers, the LTS method seeks to minimize the sum of the squares of a specified proportion of the smallest residuals.

The basic idea of the LTS method is to identify and exclude the extreme observations that have the largest impact on the traditional least squares regression fit. The LTS method computes the regression coefficients that minimize the sum of squares of a specified portion (trimming proportion) of the smallest residuals rather than the sum of squares of all residuals. The trimming proportion is typically set to a small value, such as 5% or 10%, in order to exclude a significant proportion of the outliers.

The LTS method provides a robust estimate of the regression coefficients in the presence of outliers and is particularly useful when the number of outliers is large or when the distribution of the residuals is unknown. However, the method is computationally intensive and may be more time-consuming than other robust regression techniques, such as the median absolute deviation method.

Real World Applications

This section provides several real-world cases, highlighting the superiority of the MGQ test over the conventional GQ approach to detect heteroskedasticity when there are outliers in the design matrix. The MGQ test is a statistical method used to detect heterogeneity in error variances, which can occur when the variance of errors in a regression model varies across different groups or subsets of data. By using this test, researchers can identify which groups have different variances and adjust their analysis accordingly. This can be particularly useful in fields such as finance, where heteroscedasticity (i.e., unequal variances) is common due to factors such as market volatility. The benefits of using the modified Goldfeld-Quandt test include the ability to improve the accuracy of statistical analysis and to identify one or more potential outliers or influential observations that may affect the results. Overall, this method provides a valuable tool for researchers and practitioners seeking to improve the quality and robustness of their statistical analyses.

Case 1: Brain—Body data

The first case is based on the dataset 'Animals', available in the robust base Package of R-Language. It contains data on the average body and brain weights of 62 terrestrial mammal species, including three additional species of dinosaurs at index 63—65 in Table 1. This data is used by several existing studies including Weisberg (1985) and Rousseeuw and Leroy (1987). The idea that a bigger brain is needed to regulate a bigger body makes sense. These three dinosaur species, which are represented in the data, have extraordinarily high brain-to-body ratios. Could it be that OLS's inability to recognize dinosaurs as exceptional has condemned the method to share their demise? Due of OLS's extraordinary sensitivity to "outliers," or data points that depart from the norm, this type of phenomenon regularly happens. In applied research, "special purpose" dummies are commonly required because they effectively remove the aberrant data from the observations. Sadly, finding poorly fitting observations typically requires some detective work. Examining the OLS residuals is a frequent method for this aim. The aberrant observation(s) may not, however, be revealed by the OLS residuals. The three dinosaur species' differences from the others can be seen by carefully analyzing this dataset and visually inspecting the data. When we have high-dimensional data, it is challenging to accomplish this.

Table 1. Body and brain data example.

Index	Animals	body (kg)	brain (g)	Index	Animals	body (kg)	brain (g)
1	Lesser short-tailed shrew	0.005	0.14	34	Water opossum	3.5	3.9
2	Little brown bat	0.01	0.25	35	Rock hyrax-b	3.6	21
3	Mouse	0.023	0.4	36	Yellow-bellied marmot	4.05	17
4	Big brown bat	0.023	0.3	37	Verbet	4.19	58
5	Musk shrew	0.048	0.33	38	Red fox	4.235	50.4
6	Star-nosed mole	0.06	1	39	Raccoon	4.288	39.2
7	E. American mole	0.075	1.2	40	Rhesus monkey	6.8	179
8	Ground squirrel	0.101	4	41	Potar monkey	10	115
9	Tree shrew	0.104	2.5	42	Baboon	10.55	179.5
10	Golden hamster	0.12	1	43	Roe deer	14.83	98.2
11	Mole	0.122	3	44	Goat	27.66	115
12	Galago	0.2	5	45	Kangaroo	35	56
13	Rat	0.28	1.9	46	Grey wolf	36.33	119.5
14	Chinchilla	0.425	6.4	47	Chimpanzee	52.16	440
15	Owl monkey	0.48	15.5	48	Sheep	55.5	175
16	Desert hedgehog	0.55	2.4	49	Giant armadillo	60	81
17	Rock hyrax-a	0.75	12.3	50	Human	62	1320
18	European hedgehog	0.785	3.5	51	Grey seal	85	325
19	Tenrec	0.9	2.6	52	Jaguar	100	157
20	Artic ground squirrel	0.92	5.7	53	Brazilian tapir	160	169
21	African giant pouched rat	1	6.6	54	Donkey	187.1	419
22	Guinea pig	1.04	5.5	55	Pig	192	180
23	Mountain beaver	1.35	8.1	56	Gorilla	207	406
24	Slow loris	1.4	12.5	57	Okapi	250	490
25	Genet	1.41	17.5	58	Cow	465	423
26	Phalanger	1.62	11.4	59	Horse	521	655
27	N.A. opossum	1.7	6.3	60	Giraffe	529	680
28	Tree hyrax	2	12.3	61	Asian elephant	2547	4603
29	Rabbit	2.5	12.1	62	African elephant	6654	5712
30	Echidna	3	25	63	Triceratops	9400	70
31	Cat	3.3	25.6	64	Dipliodocus	11700	50
32	Artic fox	3.385	44.5	65	Brachiosaurus	87000	154.5
33	Nine-banded armadillo	3.5	10.8	-	-	-	-

Source: Rousseeuw & Leroy (1987).

An LTS regression is run by iterated re-weighted least squares (IRLS), and residuals and weights are reported in Table 2.

Table 2. Bi-square weighting of body brain example.

Index	brain	Body	residual	weight	Index	brain	body	residual	weight
63	4.248	9.148	-4.694	0.000	9	0.916	-2.263	0.487	0.959
64	3.912	9.367	-5.194	0.000	61	8.434	7.843	0.466	0.963
65	5.040	11.374	-5.563	0.000	39	3.669	1.456	0.465	0.963
50	7.185	4.127	1.989	0.431	43	4.587	2.697	0.457	0.964
34	1.361	1.253	-1.691	0.565	22	1.705	0.039	-0.442	0.966
40	5.187	1.917	1.640	0.587	14	1.856	-0.856	0.377	0.975
42	5.190	2.356	1.315	0.722	51	5.784	4.443	0.352	0.979
15	2.741	-0.734	1.171	0.776	36	2.833	1.399	-0.328	0.981
19	0.956	-0.105	-1.084	0.806	20	1.740	-0.083	-0.315	0.983
47	6.087	3.954	1.019	0.828	29	2.493	0.916	-0.308	0.984
8	1.386	-2.293	0.979	0.840	59	6.485	6.256	-0.300	0.984
5	-1.109	-3.037	-0.961	0.846	30	3.219	1.099	0.282	0.986
41	4.745	2.303	0.909	0.861	60	6.522	6.271	-0.274	0.987
37	4.060	1.433	0.874	0.872	23	2.092	0.300	-0.250	0.989
55	5.193	5.257	-0.847	0.879	31	3.243	1.194	0.234	0.990
16	0.875	-0.598	-0.796	0.893	21	1.887	0.000	-0.231	0.991
49	4.394	4.094	-0.778	0.898	3	-0.916	-3.772	-0.220	0.992
53	5.130	5.075	-0.774	0.899	24	2.526	0.336	0.157	0.996
32	3.795	1.219	0.768	0.900	44	4.745	3.320	0.150	0.996
45	4.025	3.555	-0.745	0.906	1	-1.966	-5.298	-0.131	0.997
38	3.920	1.443	0.725	0.911	28	2.510	0.693	-0.125	0.997
12	1.609	-1.609	0.693	0.918	56	6.006	5.333	-0.090	0.999
18	1.253	-0.242	-0.684	0.920	2	-1.386	-4.605	-0.068	0.999
27	1.841	0.531	-0.673	0.923	48	5.165	4.016	0.051	1.000
33	2.380	1.253	-0.673	0.923	26	2.434	0.482	-0.044	1.000
58	6.047	6.142	-0.653	0.927	57	6.194	5.521	-0.043	1.000
17	2.510	-0.288	0.607	0.937	62	8.650	8.803	-0.035	1.000
11	1.099	-2.104	0.551	0.948	35	3.045	1.281	-0.029	1.000
10	0.000	-2.120	-0.536	0.951	6	0.000	-2.813	-0.019	1.000
13	0.642	-1.273	-0.526	0.952	54	6.038	5.232	0.017	1.000
4	-1.204	-3.772	-0.507	0.956	46	4.783	3.593	-0.015	1.000
52	5.056	4.605	-0.497	0.958	7	0.182	-2.590	-0.003	1.000
25	2.862	0.344	0.488	0.959	-	-	-	-	-

It can be seen that the weight increases as the absolute residuals decrease. To put it another way, cases with high absolute residuals are typically down-weighted. This result demonstrates that Triceratops, Dipliodocus and Brachiosaurus are heavily discounted with a weight of zero. Further, the findings of roubst regression will be close to OLS if most of the observations have weight equal to or close to 1. The bisquare weighting function zero the weight given to Triceratops, Dipliodocus and Brachiosaurus. In the cases where there is a significant difference in an OLS and a robust regression, then one should follow the findings of the robust regression as the high disparity between the two regression results indicates that outliers have a significant impact on the model's parameters. There are several weighting methods with merits and demerits. The bi-square weighting method may produce many solutions, while Huber weights may have trouble convergent or handling large outliers, so it's up to the practitioners which method to adopt in a particular situation. The results of GQ and the MGQ test are provided in Table 3, and the fitted OLS and LTS lines are shown in Figure 1.

Table 3. GQ and MGQ statistics—body and brain weight example.

Test	Statistic cal value	p-value
GQ	5.890	0.007
MGQ	1.047	0.474

It can be seen from Table 3 that the GQ test rejects the null of homoskedasticity (with a p-value of 0.007) while MGQ test does not reject the null of homoskedasticity indicating that OLS based GQ test findings are misled due to presence of outliers (see Figure 1) while the findings of MGQ test suggests that the data is homoscedastic. It is highlighted that the presence of outliers makes the GQ test to conclude that there is heteroskedasticity while the MGQ test takes care of outliers and thus provides a clearer picture and indicates that the data is homoscedastic. Thus, it is advocated that one should use MGQ rather than the GQ test when there are one or more outliers in the data.

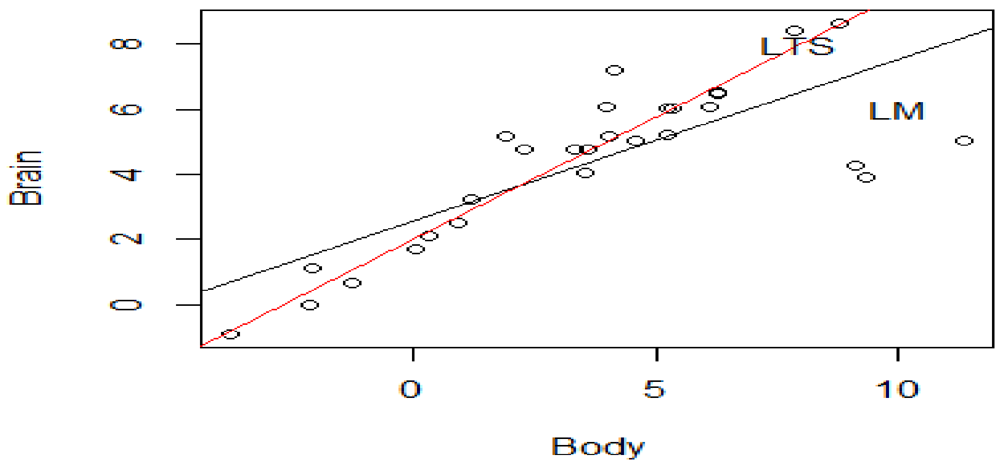

Figure 1. Case 1 (OLS and LTS).

This outlier can be recognized and ignored by the robust regression model, as shown in Figure 1, which can still identify the general trend in the remaining data. Because they can produce accurate answers that are unaffected by the outlier points, they are the perfect tool for studying datasets with outliers. The robust regression model is successful in bucking the outlier point's influence, as in Table 3, and detecting the trend in the remaining data.

Case 2: Hawkins, Bradu, Kass's data

Hawkins, Bradu, and Kass created an artificial data set (Hawkins et al., 1984), as in Table 4. Seventy-five observations in four dimensions make up the data collection (one response and three explanatory variables). That is a good illustration of how the masking effect works. The first 14 observations, divided into two groups of 1–10 and 11–14, are outliers. With robust distances calculated by, for example, MCD - covMcd, only observations 12, 13, and 14 show as outliers when using classical methods, but these are easily unmasked.

It is clear that the weight rises roughly in proportion to the fall in the absolute residual. Another way to describe it is that cases with significant residuals are often down-weighted. All observations that aren't represented above in Table 5 have a weight of 1. In an OLS regression, the weight assigned to each instance is 1. As a result, the conclusions of the OLS and robust regression are more similar as more cases in the robust regression have weights near to one. You should generally use the robust regression's results if there is a large difference between the outcomes of a regular OLS regression and the robust regression. Large discrepancies suggest that outliers have a considerable influence on the model's parameters in Figure 2. Many functions each have advantages and disadvantages.

Table 4. Hawkins, Bradu, Kass's data.

Index	X2	Y	Index	X2	Y
1	1.6	-0.2	38	0.5	-0.7
2	2.2	-0.9	39	0.1	-0.5
3	3.3	0.6	40	0.6	0.2
4	0	0.6	41	2.3	-0.8
5	3	-0.6	42	0.7	0.6
6	1.8	-0.9	43	0	-0.7
7	1.7	0.4	44	2.8	-0.4
8	0.4	0.2	45	2.5	0.2
9	1	-0.7	46	1.5	0.7
10	0.4	0.7	47	1.6	0
11	1.5	-0.9	48	3.4	-0.1
12	0	-0.7	49	2.2	0.9
13	3.2	0.3	50	3	-0.5
14	3.4	-0.3	51	0.1	-0.9
15	2.4	-0.4	52	2.2	0.6
16	2.4	0	53	1.4	0
17	2	-0.5	54	2.5	0.9
18	2	-0.2	55	3.4	0.4
19	2.9	0.1	56	1.6	0.1
20	0.1	0.6	57	2.4	0.3
21	3.3	-0.8	58	0.6	-0.5
22	0	-0.3	59	2.5	-0.7
23	2.2	0.7	60	2.9	-0.4
24	2.7	-0.3	61	1.6	-0.1
25	3	0.7	62	19.7	9.9
26	2.2	-1	63	20.5	10.1
27	0.9	0.3	64	20.7	9.6
28	0.7	-0.3	65	19.6	10.3
29	2.3	0.7	66	21.5	9.5
30	1.2	0.9	67	19.6	9.7
31	3.1	-0.6	68	21.1	10
32	2	-0.7	69	20.9	10.8
33	2.1	-0.3	70	20.2	10.3
34	0.7	0.7	71	20.4	10
35	3.2	0.1	72	34	0.1
36	0.8	0.3	73	24	-0.2
37	0.5	-0.4	74	26	0.7

Source: Hawkins et al. (1984).

Table 5. Huber weighting of Hawkins, Bradu, Kass's data.

Index	Y	X2	residual	weight	Index	Y	X2	residual	weight
72	-0.4	23	-11.601	0.000	56	0.1	3.2	-0.606	0.959
73	-0.2	24	-11.931	0.000	65	10.3	20.2	0.583	0.962
74	0.7	26	-12.091	0.000	30	-0.5	2	-0.570	0.964
75	0.1	34	-16.931	0.000	45	0.9	2.5	0.565	0.964
1	0.6	0	1.590	0.735	35	0.7	2.2	0.524	0.969

58	-0.8	3.3	-1.559	0.744	28	0.4	1.7	0.489	0.973
5	0.6	0.1	1.537	0.751	39	0.7	2.3	0.471	0.975
9	0.7	0.4	1.478	0.769	64	9.9	19.7	0.448	0.977
15	0.7	0.7	1.319	0.813	49	0.1	2.9	-0.447	0.978
20	0.9	1.2	1.254	0.830	6	-0.5	0.1	0.437	0.979
54	-0.6	3.1	-1.253	0.831	38	0.6	2.2	0.424	0.980
16	0.6	0.7	1.219	0.839	33	-0.3	2.1	-0.423	0.980
51	-0.6	3	-1.200	0.844	61	0.4	3.4	-0.412	0.981
36	-1	2.2	-1.176	0.850	55	0.3	3.2	-0.406	0.981
59	-0.3	3.4	-1.112	0.865	68	9.6	20.7	-0.382	0.984
53	-0.5	3	-1.100	0.868	10	-0.4	0.5	0.325	0.988
34	-0.9	2.2	-1.076	0.874	14	-0.3	0.7	0.319	0.989
46	-0.7	2.5	-1.035	0.883	63	9.7	19.6	0.301	0.990
40	-0.8	2.3	-1.029	0.884	2	-0.7	0	0.290	0.991
8	0.2	0.4	0.978	0.895	4	-0.7	0	0.290	0.991
50	-0.4	2.9	-0.947	0.901	42	0	2.4	-0.282	0.991
60	-0.1	3.4	-0.912	0.908	31	-0.2	2	-0.270	0.992
71	9.5	21.5	-0.906	0.910	21	0	1.4	0.248	0.993
62	10.3	19.6	0.901	0.910	26	0.1	1.6	0.242	0.993
23	0.7	1.5	0.895	0.912	19	-0.7	1	-0.240	0.994
48	-0.4	2.8	-0.894	0.912	67	10.1	20.5	0.224	0.994
12	0.2	0.6	0.872	0.916	70	10	21.1	-0.194	0.996
17	0.3	0.8	0.866	0.917	66	10	20.4	0.177	0.996
29	-0.9	1.8	-0.864	0.918	13	-0.5	0.6	0.172	0.997
18	0.3	0.9	0.813	0.927	57	0.6	3.3	-0.159	0.997
32	-0.7	2	-0.770	0.934	25	0	1.6	0.142	0.998
47	-0.3	2.7	-0.741	0.939	44	0.2	2.5	-0.135	0.998
37	0.9	2.2	0.724	0.942	52	0.7	3	0.100	0.999
69	10.8	20.9	0.712	0.944	24	-0.2	1.6	-0.058	1.000
22	-0.9	1.5	-0.705	0.945	27	-0.1	1.6	0.042	1.000
3	-0.3	0	0.690	0.947	7	-0.9	0.1	0.037	1.000
41	-0.4	2.4	-0.682	0.948	11	-0.7	0.5	0.025	1.000
-	-	-	-	-	43	0.3	2.4	0.018	1.000

The findings of GQ and the MGQ are presented in Table 6, which shows that the GQ test rejects the null of homoskedasticity (with a p-value of 0.000), whereas the MGQ test does not, indicating that the OLS-based GQ test findings are skewed due to the presence of outliers (see Figure 2), whereas the MGQ test provides a clearer picture and suggests that the data is homoscedastic. From this, we can conclude that the actual problem in the data is of outlier and not the heteroskedasticity. Once outliers have been taken care of, the data shows homoskedasticity. Thus, the conclusions based on MGQ provide a more accurate and truthful picture.

Table 6. GQ and MGQ statistics— Hawkins, Bradu, Kass's data example.

Test	Statistic cal value	p-value
GQ	47.062	0.000
MGQ	1.126	0.380

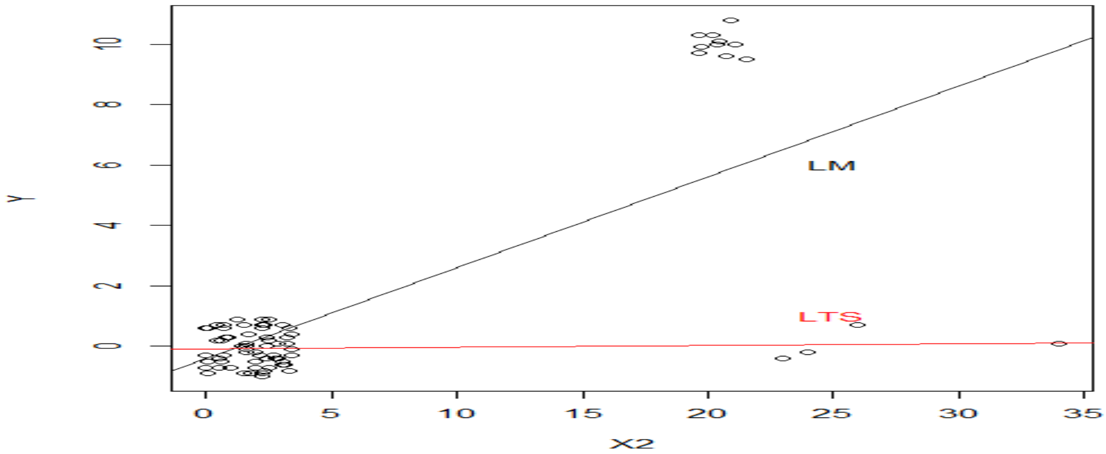

Figure 2. Case 2 (OLS and LTS).

Case 3: Boston housing data

The Boston Housing Dataset is derived from data gathered by the U.S. Census Agency about housing in the Boston Mass mentioned in Table 7. 506 observations of 14 attributes make up the dataset. The result or dependent variable in our model is the median dollar value of the cost of a home and the Average number of rooms per dwelling as an explanatory variable.

Table 7. Boston housing data.

Index	Cost (in $1000)	rooms	Index	Cost (in $1000)	Rooms
1	27.5	3.561	477	44	7.454
2	23.1	3.863	478	43.5	7.47
3	11.9	4.138	479	50	7.489
4	13.8	4.138	480	43.1	7.52
5	8.8	4.368	481	42.3	7.61
6	7	4.519	482	46	7.645
7	17.9	4.628	483	46.7	7.686
8	10.5	4.652	484	35.2	7.691
9	10.2	4.88	485	39.8	7.765
10	11.8	4.903	486	50	7.802
11	13.8	4.906	487	45.4	7.82
12	14.6	4.926	488	43.8	7.82
13	21.9	4.963	489	50	7.831
14	50	4.97	490	48.5	7.853
15	16.1	4.973	491	50	7.875
16	7.4	5	492	50	7.923
17	15.3	5.012	493	50	7.929
18	14.4	5.019	494	50	8.034
19	9.7	5.036	495	37.6	8.04
20	8.1	5.093	496	38.7	8.069
21	16.3	5.155	497	48.3	8.247
22	17.8	5.186	498	42.8	8.259
23	13.1	5.272	499	44.8	8.266
24	7.2	5.277	500	50	8.297
25	12	5.304	501	41.7	8.337
26	10.4	5.304	502	50	8.375

27	20	5.344	503	48.8	8.398
28	8.3	5.349	504	50	8.704
29	20.8	5.362	505	50	8.725
30	19.7	5.39	506	21.9	8.78

Source: Derived from U.S. Census service concerning housing in the area of Boston Mass.

Table 8 shows the residuals and weights obtained by running an LTS regression via IRLS.

Table 8. Bi-square weighting of Boston housing example.

Index	Median value of Owner-occupied home in $1000	Average number of rooms per dwelling	residual	Weight	Index	Median value of Owner-occupied home in $1000	Average number of rooms per dwelling	residual	Weight
469	15	7.313	-18.580	0.00	106	6.3	5.852	-11.824	0.39
334	7.2	6.434	-17.081	0.05	431	41.3	6.943	11.634	0.41
473	17.8	7.393	-16.627	0.07	345	13.1	6.471	-11.573	0.41
297	7.2	6.343	-16.119	0.09	319	12.5	6.405	-11.475	0.42
229	36.2	6.144	14.986	0.16	389	15.2	6.655	-11.419	0.42
344	9.6	6.461	-14.967	0.16	69	5	5.683	-11.337	0.43
413	13.3	6.794	-14.790	0.17	486	50	7.802	11.246	0.44
479	50	7.489	14.558	0.19	359	14.1	6.525	-11.144	0.44
362	10.9	6.545	-14.556	0.19	326	13	6.417	-11.102	0.45
419	14.1	6.833	-14.402	0.20	489	50	7.831	10.940	0.46
404	13.4	6.749	-14.214	0.21	252	10.9	6.202	-10.927	0.46
308	9.5	6.38	-14.210	0.21	56	26.4	5.604	10.899	0.46
171	5.6	5.987	-13.953	0.23	249	11	6.193	-10.732	0.48
4	13.8	4.138	13.808	0.24	395	16.4	6.701	-10.706	0.48
13	21.9	4.963	13.180	0.29	307	13.1	6.38	-10.610	0.49
381	13.4	6.629	-12.944	0.30	147	8.4	5.935	-10.602	0.49
343	11.8	6.459	-12.746	0.32	491	50	7.875	10.474	0.50
7	17.9	4.628	12.724	0.32	155	8.8	5.957	-10.435	0.50
388	13.9	6.649	-12.656	0.33	134	8.3	5.896	-10.290	0.51
232	8.7	6.152	-12.598	0.33	36	23.7	5.412	10.230	0.52
402	14.9	6.728	-12.492	0.34	253	11.7	6.208	-10.191	0.52
58	27.9	5.608	12.357	0.35	335	14.3	6.436	-10.003	0.53
3	11.9	4.138	11.908	0.38	492	50	7.923	9.966	0.54
316	12.1	6.404	-11.864	0.39	493	50	7.929	9.903	0.54
261	10.2	6.223	-11.849	0.39	273	12.6	6.251	-9.745	0.55

We see that the weight roughly rises as the absolute residue falls. To put it another way, down-weighting is often used for cases with substantial residuals. The total weight of all observations not shown in Table 8 above is 1. An OLS regression has a weight of 1 for each case. Because of this, the OLS and robust regression results are becoming more comparable as examples in the robust regression with weights near to one increase. You should probably use the results of the robust regression if there is a significant difference between the results of a regular OLS regression and those of the robust regression, as in Figure 3. Large discrepancies suggest that outliers significantly affect the model's input parameters. There are advantages and disadvantages to many functions.

Table 9. GQ and MGQ statistics— Boston housing example.

Test	Statistical value	p-value
GQ	1.318	0.025
MGQ	1.067	0.323

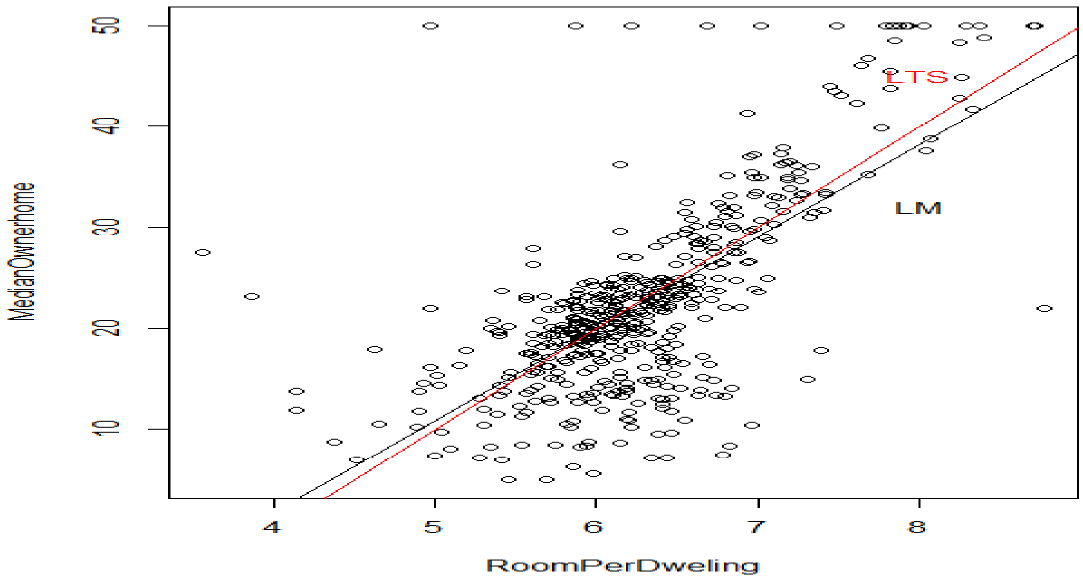

Figure 3. Case 3 (OLS and LTS).

Now that we have applied our suggested modified Goldfeld-Quandt test to the data sets from earlier investigations, the findings are shown in Table 9. From the three empirical examples presented above (Case 1—3), one can see that when there are outliers in the data sets, the Goldfeld Quandt test is unable to identify the heteroscedasticity issue, and in contrast, the MGQ provides a true picture and provides reliable inferences regarding the status of homoskedasticity in the data.

CONCLUSIONS

By detecting and addressing heterogeneity, researchers can ensure that their statistical models accurately reflect the data and avoid making incorrect conclusions. The commonly used test for heterogeneity testing is the one proposed by Goldfeld and Quandt (the GQ test). Recently, a modified version of GQ is proposed, which has the ability to work well when there are outliers in the data. The present study demonstrates that GQ test provides misleading results when the data contains one or more outliers. In contrast to GQ, the MGQ test works well under these situations and this study makes the case in favor of MGQ when outliers are present in the data. Several examples from real-world data are presented, showing the superiority of MGQ over GQ in the presence of outliers. The benefits of using the modified Goldfeld-Quandt test are numerous, as it allows researchers to identify potential sources of bias or confounding variables that may be affecting their results. By using the modified Goldfeld-Quandt test, researchers can account for these differences and ensure that their conclusions are robust and reliable. Overall, the modified Goldfeld-Quandt test is an essential tool for researchers working with complex datasets, as it allows them to detect and correct for heterogeneity in error variances and improve the accuracy of their statistical models in the presence of outliers.

REFERENCES

Abdul-Hameed, A. B., & Matanmi, O. G. (2021). A modified Breusch–Pagan test for detecting heteroskedasticity in the presence of outliers. Retrieved online from the website: http://article.pamathj.net/pdf/10.11648.j.pamj.20211006.13.pdf.

Ahmed, M., Haider, G., & Zaman, A. (2017). Detecting structural change with Heteroskedasticity. Communications in Statistics - Theory and Methods, 46(21), 10446-10455.

Andrews, D. W. (1993). Tests for parameter instability and structural change with unknown change point. Econometrica: Journal of the Econometric Society, 61, 821-856.

Andrews, D. W., & Ploberger, W. (1994). Optimal tests when a nuisance parameter is present only under the alternative. Econometrica: Journal of the Econometric Society, 62(2), 1383-1414.

Asteriou, D., & Hall, S. G. (2015). Applied econometrics: Macmillan International Higher Education, London.

Asteriou, D., & Hall, S. G. (2017). Applied econometrics (3rd Ed). Macmillan International Higher Education.

Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. Proceedings of the Royal Society of London. Series A-Mathematical and Physical Sciences, 160(901), 268-282.

Breusch, T., & Pagan, A. (1979). A simple test for heteroskedasticity and random coefficient variation. Econometrica, 47(5), 1287. https://doi.org/10.2307/1911963.

Cook, R. D., & S. Weisberg. (1983). Diagnostics for heteroskedasticity in regression. Biometrika 70 (1):1-10. doi:10.1093/biomet/70.1.1.

Dutta, J., & Zaman, A. (1989). What do tests for heteroskedasticity detect? (No. 248). University of Bonn, Germany.

Engle, R. F., & Granger, C. W. (1987). Co-integration and error correction: representation, estimation, and testing. Econometrica: Journal of the Econometric Society, 55(2), 251-276.

Evans, M., & King, M. (1988). A further class of tests for heteroskedasticity. Journal of Econometrics, 37(2), 265-276. https://doi.org/10.1016/0304-4076 (88)90006-1.

Frisch, R., & Waugh, F. V. (1933). Partial time regressions as compared with individual trends. Econometrica: Journal of the Econometric Society, 1(4), 387-401.

Galton, F. (1883). Inquiries into human faculty and its development. Macmillan, London.

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

Godfrey, L. (1978). Testing for multiplicative heteroskedasticity. Journal of Econometrics, 8(2), 227-236. https://doi.org/10.1016/0304-4076 (78)90031-3.

Goldfeld, S. M., & Quandat, R. E. (1972). Nonlinear methods in econometrics., Amsterdam Press, North Holland

Goldfeld, S. M., & Quandt, R. E. (1973). A Markov model for switching regressions. Journal of Econometrics, 1(1), 3-15.

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

Greene, W. H. (2003). Econometric analysis. Pearson Education, India.

Gujarati, D. (2002). Basic Econometrics (Terjemahan Sumarno Zain). Jakarta: Erlangga.

Harrison, M. J., & McCabe, B. P. (1979). A test for heteroskedasticity based on ordinary least squares residuals. Journal of the American Statistical Association, 74(366a), 494-499.

Harvey, A. C., & Collier, P. (1977). Testing for functional misspecification in regression analysis. Journal of Econometrics, 6(1), 103-119.

Harvey, A., & Phillips, G. (1974). A comparison of the power of some tests for Heteroskedasticity in the general linear model. Journal of Econometrics, 2(4), 307-316. https://doi.org/10.1016/0304-4076 (74)90016-5.

Hawkins, D. M., Bradu, D., & Kass, G. V. (1984). Location of several outliers in multiple-regression data using elemental sets. Technometrics, 26(3), 197-208.

Hinkley, D. V. (1977). Jackknifing in unbalanced situations. Technometrics, 19(3), 285-292.

Horn, S. D., Horn, R. A., & Duncan, D. B. (1975). Estimating heteroscedastic variances in linear models. Journal of the American Statistical Association, 70(350), 380-385.

Imon, A. R. (2003). Residuals from deletion in added variable plots. Journal of Applied Statistics, 30(7), 827-841.

Kariya, T., & Kurata, H. (2004). Generalized least squares. John Wiley & Sons: Hoboken, NJ, USA.

MacKinnon, J. G., & White, H. (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. Journal of Econometrics, 29(3), 305-325.

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

Ramsey, J. B. (1969). Tests for specification errors in classical linear least-squares regression analysis. Journal of the Royal Statistical Society: Series B (Methodological), 31(2), 350-371.

Rana, M. S., Midi, H., & Imon, A. R. (2008). A robust modification of the Goldfeld-Quandt test for the detection of heteroskedasticity in the presence of outliers. Journal of Mathematics and Statistics, 4(4), 277.

Rosopa, P. J., Schaffer, M. M., & Schroeder, A. N. (2013). Managing heteroskedasticity in general linear models. Psychological Methods, 18(3), 335.

Rousseeuw and Leroy (1987). Outlier detection and robust regression. John Wiley & Sons: Hoboken, NJ, USA.

Rousseeuw, P. J., & Driessen, K. V. (1999). A fast algorithm for the minimum covariance determinant estimator. Technometrics, 41(3), 212-223.

Uyanto, S. S. (2019). Monte Carlo power comparison of seven most commonly used heteroskedasticity tests. Communications in Statistics-Simulation and Computation, 51(4), 1-18.

Weisberg, S. (1985) Applied Linear Regression. 2nd edition. John Wiley & Sons: Hoboken, NJ, USA.

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

Wooldridge, J. M. (2002). Econometric analysis of cross section and panel data MIT Press, Cambridge.

Zaman, A. (1994). Most stringent and best invariant hypothesis tests for regression models. Department of Economics, Bilkent University, Turkey.