Journal of Business and Social Sciences (ISSN: 2350-8876)
Research Department, Hetauda School of Management and Social Sciences (HSMSS), Hetauda

Autocorrelation and Heteroscedasticity in Regression Analysis

Dr. Nand Kishor Kumar

Lecturer, *Trichandra Campus, Tribhuvan University, Nepal*
nandkishorkumar2025@gmail.com

Cite this paper:

Kumar, N.K. (2023). Autocorrelation and heteroscedasticity in regression analysis. *Journal of Business and Social Sciences, 5*(1), 9 - 20.

https://doi.org/10.3126/jbss.v5i1.72442

Abstract

In the world of econometrics and time series analysis, autocorrelation and heteroscedasticity are two statistical issues that analysts often encounter when modeling data. Both phenomena can seriously undermine the reliability of regression results if left untreated, leading to biased conclusions and faulty predictions. Autocorrelation and heteroscedasticity are important considerations in any regression analysis, especially when explaining with time series analysis. Both phenomena undermine the accuracy of the model if left unaddressed. By applying the appropriate tests and corrective measures, such as Generalized Least Squares or robust standard errors, analysts can ensure that their models remain reliable, leading to more valid inferences and better decision-making. In practice, detecting and correcting for it should be a standard part of the model validation process to avoid incorrect conclusions and improve the robustness of the results.

Keywords: Autocorrelation, Heteroscedasticity, Regression analysis, Durbin- Waston test, Goldfeld-Quandt test, serial correlation coefficient

Introduction

A statistical technique known as an autocorrelation is used to determine if the variables in the prediction model have a correlation that varies with time. When the prediction model's autocorrelation assumption is met, the disturbance values are auto correlatedly paired rather than independent pairs. Checks for autocorrelation that may develop as a result of successive observations made over time and their relations.

The residuals' inability to remain independent across observations is the root source of this issue. The goal of the autocorrelation test is to display the connection between observation members arranged temporally or spatially. In addition, it tests whether the confounding error in a linear regression model are correlated. An autocorrelation problem arises when there is a correlation. The balance is not independent from one observation to the next, which leads to this issue.

It is common in time series data, where observations are often dependent on past values. For instance, a country's GDP in one year may be strongly related to its GDP in the previous year. If it exists, using Ordinary Least Squares (OLS) regression can lead to misleading statistical inferences.

As a result, there are several approaches for figuring out whether it exists. Durbin Watson (DW) test is a frequently used technique to identify autocorrelation in multiple linear regression. It is most frequently employed. This article explores these two concepts, their impact on regression models, detection methods, and solutions to correct them.

Literature Review

Kim (2019) explored that Multiple linear regression analysis findings are distorted by multicollinearity. The statistical insignificance of the regression coefficients and the widening of their confidence intervals are caused by the multicollinearity-induced inflation of their variances. If the variance inflation factor and condition number are greater than 5 to 10 and 10 to 30, respectively, then multicollinearity is found. Added that the creation of a trustworthy multiple linear regression model is made possible by the diagnosis of multicollinearity and the removal of multicollinear explanatory variables.

Baum and Lewbel (2019) explained that Heteroskedasticity can be used to construct instruments using estimators in the absence of external instruments. Astivia and Zumbo (2019) explored that high heteroskedasticity can be effectively addressed through the application of the bootstrap procedure. This technique can be used to calculate p-values and confidence intervals in instances where a researcher deals with small sample sizes.

Mukherjee and Laha (2019) described that Durbin-Watson test, which is commonly used for first-order autoregressive schemes, can be applied in small sample sizes to guarantee the validity of the collected information. Shrestha (2020) discussed that on identifying the characteristics of the visitors' destination satisfaction and how these elements affect their decision to return to Nepal. This study sent self-administered questionnaires to a sample of visitors, using the survey approach for data gathering. To identify important parameters associated with destination feature satisfaction, an exploratory factor analysis utilizing principal component analysis and varimax rotation was conducted. A probit model was examined using the maximum likelihood approach to examine the influence of visitors' satisfaction with their desire to return. Yin (2020) explained that estimation of autocorrelation is frequently challenging in instances where researchers deal with small sample sizes. The Durbin h test for autocorrelated error terms can be applied in instances where a researcher is dealing with a small sample size.

Youssef (2022) explored that majority of econometric models have issues with heteroscedasticity, multicollinearity, and autocorrelation. A summary of these issues, their causes, and methods for detection, testing, and minimization are provided in their work. The OLS approach is predicated on a number of norms, and if these presumptions hold true, we are left with estimates that are efficient, unbiased, and have less variation than other approaches. They talk about these issues as follows: First, the multicollinearity issue Second: The autocorrelation issue Variation Heteroscedasticity ranks third. The inference for several widely used estimators is presented in their paper, including the normality test for residuals, the correlogram of residuals, the coefficient covariance matrix, and variance inflation factors. The LM test for serial correlation, the Harvey heteroskedasticity test, and the estimated and actual residuals.

Literatures about this theory can be found in any standard text books, reference book and monographs of autocorrelation and heteroscedasticity, for instance we refer a few; Dobson & Basnett (2008), Kaufman (2014), Ye & Sun (2018), Fox (2008) etc. As this work explores these two concepts, their impact on regression models, detection methods, and solutions to correct them. These are two distinct concepts in regression analysis and time series analysis.

Method and Discussion

Let us consider the multiple regression equation get the form

$$Y_i = a + bX_i + e_i \quad \text{(i)}$$

Where a = Y_i- intercept
b = slope of regression line
X_i = independent variable
Y_i = dependent variable
e_i = errors or disturbance term

Estimated error value is the difference between observed (Y_i) and estimated ($\hat{Y}_i$) values of the dependent variable for a given value of X_i .

$$e_i = Y_i - \hat{Y}_i \quad \text{(ii)}$$

Here,

Y_i = observed value

$\hat{Y}_i$ = estimated value

The fundamental assumptions of the linear regression model regarding errors or disturbance are usually distributed with mean zero and variance σ^2 i.e. $e_i \sim N(0, \sigma^2)$ and the covariance between them are zero. So, the following are fundamental assumptions:

(i) e_i is normally distributed
(ii) $E(e_i) = 0$
(iii) $E(\sigma^2) = \sigma^2$
(iv) $E(e_i e_j) = 0$ for $i \neq j$ i.e. covariance is zero.

This suggests that the influence of the error term occurring at one period does not carry over to another period, observations are taken across time. Assume for the moment that we are researching the consumption function. In error term in time series missing data indicates that changes in consumption brought on by marriage ceremonies or other events are just momentary, affecting the current month's consumption. When we talk about cross-sectional statistics, we mean the model that says about the disruption of one family's consumption due to a relative's visit has no impact on the consumption of other households.

However, this presumption is false in actuality. When dealing with time series data, it is reasonable to assume that a significant number of random and independent components that are present during one period will partially or completely persist into subsequent periods. As a result, variables that were previously thought to be random begin acting permanently, which results in auto-correlation between subsequent disturbances. Stated differently, it may be said that if the least square estimator fails to satisfy or

$$E(e_i e_j) \neq 0 \text{ for } i \neq j \quad \text{(iii)}$$

The relationship between e_i and e_j, gives auto correlation. The series is auto correlated if

$E(e_i, e_j) \neq 0$ for i ≠j, otherwise, it is not correlated.

We have been discussing the independence of the errors, which is one of the fundamental assumptions of the regression model. Because a residual at any given point in time may tend to be comparable to residuals at adjacent points in time, this assumption is frequently broken when data are gathered over successive periods of time. As a result, it would be more probable for positive residuals to follow positive residuals and for negative residuals to follow negative residuals. We refer to this kind of residual pattern as auto-correlation or serial correlation. When a collection of data exhibits significant auto correlation, one may seriously question the validity of a fitted regression model.

Correlation between two time series data such as $e_1, e_2, \cdots\cdots e_{10}$ and as $e_2, e_3, \cdots\cdots e_{11}$, where the previous is the latter series lagged by one time period, is auto-correlation.

Causes of Auto-Correlation

There are several causes responsible for auto-correlation:

(i) **By removing explanatory variables**: An error term in one period may have a relationship with the error terms in other periods because in economics, one variable is affected by a large number of factors. If we have excluded the explanatory variables, the error term indicates the effect of omitted variables. As a result, absent explanatory factors lead to the auto correlation issue.

(ii) **By mis-lead of the mathematical model**: The second cause in autocorrelation is the mis-lead of the relationship between dependent variable and explanatory variables. So, care should be taken to specify it.

(iii) **By mis-lead of the true random:** The presence of measurement error in the disturbance term makes it potentially auto-correlated. The serial disturbances will auto-correlate if the explanatory variable is measured incorrectly because there is no way to fix it.

Effect of Auto-Correlation

When economic time series are used in many cases, the independence error term, may not hold. Let us consider the two variable regression model of the following form

$$Y_t = a + b_1 X_t + e_i \qquad \text{(iv)}$$

Where t is the time series data and u_t has five order regression as

$$E(e_t, e_{t+s}) \neq 0, \text{ and } s \neq 0 \qquad \text{(v)}$$

The error terms are generated as follows:

$$e_t = \rho\, e_{t-1} + \varepsilon_t, \ -1 < \rho < 1 \qquad \text{(vi)}$$

ρ is the coefficient of auto correlation and ε_t is the stochastic disturbance, satisfies the ordinary least square (OLS).

$E(e_t) = 0$

E ($\varepsilon_t^{\ 2}$) = var (ε_t) = σ^2

E (e_t, e_{t+s}) = cov. (e_t, e_{t+s}) $= 0$ for s $\neq 0$

When e_t and e_{t+s} are not independent, so define

$\rho = \frac{\text{cov.}(\ e_t, e_{t-1}\ )}{\text{var}\ (e_t)} = \frac{\sum_{t=2}^{n} e_t, e_{t-1}}{\sum_{t=1}^{n} e_t^{\ 2}}$ as the first order autocorrelation coefficient of the e-series. Similarly, for the variable linear regression model

$Y_t = \text{a} + b_1\ X_t + b_2 X_2\ +\ e_i$

$\rho^2 = \frac{\text{cov.}(\ e_t, e_{t-2}\ )}{\text{var}\ (e_t)} = \frac{\sum_{t=3}^{n} e_t, e_{t-2}}{\sum_{t=1}^{n} e_t^{\ 2}}$ is the second order auto correlation coefficient of the e-series.

$\rho^s = \frac{\text{cov.}(\ e_t, e_{t-s}\ )}{\text{var}\ (e_t)} = \frac{\sum_{t=s+1}^{n} e_t, e_{t-s}}{\sum_{t=1}^{n} e_t^{\ 2}}$ is the s^{th} order of the e-series. If s =0, then ρ = 1. The zero-order autocorrelation coefficient is constantly unity

Detection of Auto-correlation- Durbin-Watson Test

The Durbin Watson statistic is a test statistic that was created by statisticians Durbin and Watson to identify the existence of autocorrelation in the residuals. The correlation between each residual and the residual for the time period right before the one of interest is measured by this statistic. The Durbin Watson statistic is used to determine if the error terms are independents or serially correlated (auto correlated). The Durbin Watson statistic has the following definition:

$$d = \frac{\sum_{t=2}^{n} (e_t, e_{t-1}\ )^2}{\sum_{t=1}^{n} e_t^{\ 2}} \qquad \text{(vii)}$$

where e_t in time t.

To better know what the Durbin-Watson statistic is, we need to examine the composition of the 'd' statistic presented in equation (vii).

The numerator $\sum_{t=2}^{n} (e_t, e_{t-1}\ )^2$

represents the squared difference in two consecutive residuals. The denominator represents the sum of the squared residuals,

$$\sum_{t=1}^{n} e_t^{\ 2}$$

Since the value of d-statistic from e_t depends on the given X's, unlike the t, F or χ^2 tests, there is unique critical value of it, which will lead to the rejection of the acceptance of the null hypothesis that there is no first-order serial correlation in the disturbance e_t.

Durbin and Watson derived a lower bound d_L and an upper bound d_U . If the computed value of d-statistic lies beyond the critical values, then there is company of positive or negative serial correlation. The values of d_L and d_U depend only on the number of observations and the number of explanatory variables k.

Range of d

The limits of d are 0 and 4 that can be derived as:

The Durbin-Watson Test is

$$d = \frac{(e_t, e_{t-1})^2}{e_t^2} = \frac{\sum e_t^2 + \sum e_{t-1}^2 - 2\sum e_t e_{t-1}}{e_t^2} \qquad \text{(viii)}$$

Since, $\sum e_t^2$ and $\sum e_{t-1}^2$ differ in only one observation, they are approximately equal. So,

$\sum e_t^2 = \sum e_{t-1}^2$ in (viii), get

$$d = \frac{\sum e_t^2 + \sum e_t^2 - 2\sum e_t e_{t-1}}{e_t^2} = \frac{2\sum e_t^2 - 2\sum e_t e_{t-1}}{e_t^2} = 2\left[1 - \frac{\sum e_t \sum e_{t-1}}{\sum e_t^2}\right] = 2\{1-\rho],$$

where ρ is coefficient of auto correlation

$$\frac{\sum e_t \sum e_{t-1}}{\sum e_t^2}$$

Or, $d = 2[1-\rho]$ (x)

Where $-1 \leq \rho \leq 1$

Nature of First Order Auto correlation

(i) $\hat{\rho} = 0$ or d=2, then there is no first order auto correlation. So, errors are statistically independent.

(ii) $\hat{\rho} = +1$ i.e. d = 0, then there is perfect positive auto correlation in the residuals

(iii) If $\hat{\rho} = -1$ i.e. d = 4, then there is perfect negative correlation among successive residuals

(iv) If $\hat{\rho} > 0$, then d< 2 (i.e. closer is d to zero) in such a case there is greater evidence of positive correlation auto correlation.

(v) If $\hat{\rho} < 0$, then d > 2 (i.e. closer is d to 4), in such a case there is greater the evidence of negative auto correlation.

In order to determine whether the errors terms are statistically independent or auto correlated, the steps for testing auto correlation by Durbin-Watson test statistic are as follows.

Step 1: Setting the null hypothesis and alternative hypothesis

Null hypothesis Ho: $\rho = 0$. There is no first order positive (or negative) auto correlation between the error terms.

Alternative hypothesis H□: p > 0. There is first order positive auto correlation between error terms. In other words, they are dependent means that they are positively auto correlated.

Alternative hypothesis H□: p <0. There is first order negative auto correlation between the error terms.

Alternative hypothesis H□: p≠0.There is significant evidence of auto correlation.

Step 2: Compute Durbin Watson test statistic d under H is

$$d = \frac{\sum_{t=2}^{n} (e_t , e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}$$

Step 3: Finding out the critical d_L and d_U values at appropriate level of significance at particular sample size and variables.

Step. 4: Comparing the computed value of d with critical d_L and d_U values.

Choice for positive autocorrelation

(i) When d <4- d_L , d is significant and the null hypothesis H_0 is rejected and the alternative hypothesis H_1 that there is positive first order auto correlation.

(ii) When d > 4- d_U is not significant then the null hypothesis Ho is not first order positive auto correlation.

(iii) When d_L <d< d_U, test is inconclusive.

Choice for negative autocorrelation

(i) When d>4-d_1, d is significant and null hypothesis Ho is alternative hypothesis H

(ii) When d<4-du, d is not significant and null hypothesis Ho then there is no negative auto correlation.

(iii) When 4-du <d<4- d_L, then test is inconclusive, the result does not have a common application.

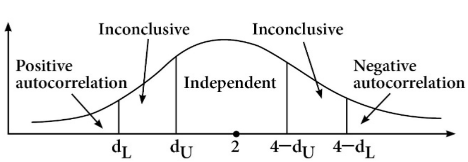
Choice for two-tailed case

(i) When 4- d_L<d<d_L , then d is significant and rejecting H_0 to accept H_1 gives auto correlation.

(ii) When d_U <d < 4- d_U, then d is not significant and accepting H_0 gives no auto correlation.

Figure 1 shows the distribution of 'd' statistics is approximately

centered around the value d = 2.



Distribution of d-statistics

Method of Removing Auto-correlation

If auto correlation exits, we may attempt to remove it by transforming the data (say X and Y) into new variables (say X' and Y') and then applying the method of least squares. We shall explain here the case where the it is removed. A method that is used to remove it is to transform the variables as follows.

Let the regression equation

$$Y = a + b_1 X + e \quad \text{(xi)}$$

and first order autocorrelation coefficient is

$$\rho = \frac{\sum_{t=2}^{n} e_t \cdot e_{t-1}}{\sum_{t=1}^{n} e_t^{2}} \quad \text{(xii)}$$

Let Durbin- Waston test shows there is presence of auto correlation in the series. Then, to remove auto correlation, data are transformed from given variables (X and Y) into the new variables (say X' and Y') as follows:

$X_2 - \rho X_1 = X'_2$ $\quad$ $Y_2 - \rho Y_1 = Y'_2$

$X_3 - \rho X_2 = X'_3$ and $\quad$ $Y_3 - \rho Y_2 = Y'_3$

$\vdots$ $\quad$ $\vdots$

$X_n - \rho X_{n-1} = X'_n$ $\quad$ $Y_n - \rho Y_{n-1} = Y'_n$

Then, apply the method of least square to y' and x'. After the regression of Y' on X' is found, if there is still auto-correlation remaining repeat this process.

Heteroscedasticity

The term heteroscedasticity describes a scenario in which the residuals' variance varies across all levels of the independent variables. The homoscedasticity assumption, which specifies that the residuals should have constant variance across observations, is broken.

The heteroscedasticity test is a method for evaluating regression models that establishes the variance inequality between the remaining data and other observations. Heteroskedasticity is typically described by the term "non-constant error variance" or by a regression model in which the residual variability varies as a function of an independent variable. Researchers use the ARCH test to see if this model has a heteroscedasticity problem. Because of the non-constant variance, the outcome estimations' variance rises.

The extent of the estimated variance will therefore affect the hypothesis test as the test is predicated on the amount of estimated variance. The theory will become untrustworthy as a result. In large samples, heteroscedasticity white is the consistent variance; the standard error may be used to modify the standard errors of the Ordinary Least Squares.

In the general linear model

$$Y_i = \beta X_i + u_i \quad \text{(xiii)}$$

We assume that E $(u^2 i) = \sigma^2 u$ is fixed for all values of all values of i. The assumption of same variance $((\sigma^2 u)$ of the residuals is known as homoscedasticity and violation of this is known as heteroscedasticity. we illustrate the problem with some examples.

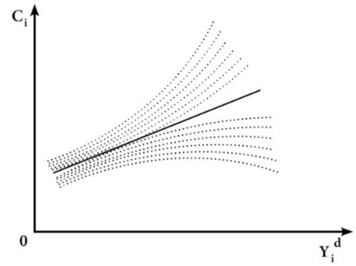
Examples of Heteroscedasticity

Estimation of the consumption function

$$C_i = C_a + cY^d i + u_i \qquad \text{(xiv)}$$

Where c is MPC, C_a is autonomous consumption, C_i and $Y^d i$ are consumption expenditure and disposable income of the household, u is the stochastic disturbance term. The scatter of C_i and $Y^d i$ generally looks like that given in the Figure 2.

Figure 2.



It is seen that at low levels of income. There is little variability in consumption, but as income increases the dots are more and more dispersed.

Consequences of Heteroscedasticity

Consider the two – variable model, assuming that is no problem other than heteroscedasticity.

$$Y_i = \beta_o + \beta_t X_i + u_i$$

E($u_i) = 0 E(u^2 i) = \sigma^2 i$

$= X_i \sigma_u^2$

function: $x^2 i\, \sigma^2 u$ or $\sqrt{X_i}\, \sigma^2 u$

E $(u_i u_i) = 0, \quad i \neq j$

The OLS estimator is

$$\widehat{\widehat{\beta_1}} = \frac{\Sigma x_1 y_1}{\Sigma x_i^2} = \beta + \frac{\Sigma x_i u_i}{\Sigma x_i^2}$$

So, E ($\widehat{\widehat{\beta_1}}$) = β, so $\widehat{\widehat{\beta_1}}$ is an unbiased of β_1 and Var ($\widehat{\widehat{\beta_1}}$)

$$= E\left[\frac{x_1 u_1 + x_2 u_2 + \cdots\cdots + x_n u_n}{\Sigma x_i^2}\right]^2$$

$$= \frac{1}{(\Sigma x_i^2)^2} E[\, x_1^2 u_1^2 + \cdots\cdots + 2\, x_1 x_2 u_1 u_2 + \cdots] = \frac{\Sigma x_i^2 \sigma_i^2}{(\Sigma x_i^2)^2}$$

Variance of the OLS estimator of β_1, when there is no heteroscedasticity is $\frac{\sigma^2 u}{\Sigma x_i^2}$

Now, suppose E $(u_i^2) = X_i \sigma_u^2$

$$\text{Var} (\widehat{\widehat{\beta_1}}) = \frac{\Sigma X_i x_i^2\ \sigma_u^2}{(\Sigma x_i^2)^2} = \frac{\sigma_u^2\ \Sigma (X_i - \bar{X})^2 X_i}{(\Sigma x_i^2)^2}$$

$$= \text{Var}(\widehat{\beta_1}) \quad \frac{\frac{n^2\sum X_i^2+(\sum X_i)^2-2n\sum X_i\sum X_i^2}{n^2}}{\sum x_i^2}$$

Where $\hat{\beta}_1$ is the OLS estimator of β_1

So, $$\frac{\text{Var}(\widehat{\beta_1})}{\text{Var}(\widehat{\widehat{\beta_1}})} = \frac{n^2\sum x_i^2}{n^2\sum X_i^2+(\sum X_i)^2-2n\sum X_i\sum X_i^2} \neq 1 \qquad \text{(xv)}$$

Hence, Var $(\widehat{\beta_1})$ $\neq$ Var ($\widehat{\widehat{\beta_1}}$).

Generally, Var $(\widehat{\beta_1})$ < Var ($\widehat{\widehat{\beta_1}}$)

Test for Nature of Heteroscedasticity

Let

$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_1 X_{2i} + \mu_i$

And the variance of μ_i is related to variable X_{1i} is given by

$\sigma_i^2 = f(X_{1i})$

Or, E (μ_i^2) = f (X_{1i}) where the exact functional form is not known, but if approximate f (X_{1i}), then easy to transform the variables multiplying by

$\frac{1}{\sqrt{f(X_{1i})}}$ and applying OLS to estimate parameters.

Comparison

Autocorrelation is general in time series data, where observations from one period are correlated with observations from previous periods. It can arise due to:

(i) Omitted variables.
(ii) Model misspecification.
(iii) Data with trends or seasonality.

Heteroscedasticity arise when there is unequal variability in the response variable.

(i) Changes in the scale or distribution of the dependent variable over time.
(ii) Presence of outliers or leverage points.
(iii) Non-linear relationships between independent and dependent variables.

Effect on Regression Models

The Ordinary Least Squares (OLS) estimators are impacted by autocorrelation, which renders them inefficient but yet impartial. When autocorrelation is present, standard errors may be overestimated, inflating t-statistics and perhaps leading to erroneous conclusions (such as Type I errors). In OLS regression, heteroscedasticity results in ineffective and perhaps biased parameter estimations. A mistake in the computed standard errors might result in erroneous confidence ranges and hypothesis testing.

Detection Methods

Autocorrelation can be detected from

The Durbin-Watson test for first-order autocorrelation.

The Breusch-Godfrey test for higher-order autocorrelation.

Autocorrelation Function (ACF) plots to visualize patterns over time.

Heteroscedasticity can be detected from

Breusch-Pagan test or White's test.

Residual plots, where plotting residuals against fitted values shows whether the spread of residuals is consistent.

The Goldfeld-Quandt test for testing heteroscedasticity in time series.

Solution

Autocorrelation can be addressed in following way:

Adding lagged variables.

Using Generalized Least Squares (GLS) or Newey-West standard errors to adjust for autocorrelation.

Differencing the data to remove trends.

Heteroscedasticity can be corrected in following way:

Using robust standard errors (e.g., White's robust standard errors).

Applying transformations to the dependent variable (e.g., log transformation).

Employing Weighted Least Squares (WLS), where observations with larger variances receive less weight.

Visual Representation

Autocorrelation can often be visualized using ACF or PACF plots, showing spikes that indicate correlation between residuals at various lags. Heteroscedasticity is typically visualized using a scatter plot of residuals versus fitted values, where a funnel shape suggests increasing variance.

Table1: Summary of Comparison

Feature	**Autocorrelation**	**Heteroscedasticity**
Definition	Residuals correlated with their own past values.	Residuals have non-constant variance.
Common in	Time series data	Cross-sectional or time series data
Impact on OLS	Inefficient estimators, inflated t-statistics.	Biased estimators, incorrect standard errors.
Detection	Durbin-Watson test, ACF plot, Breusch-Godfrey test.	Breusch-Pagan test, residual plots, White's test.
Correction Methods	GLS, Newey-West errors, differencing, lags.	Robust standard errors, WLS, transformations.

Conclusion

In conclusion, both autocorrelation and heteroscedasticity can severely impact the validity of regression models, but they do so in different ways and require different approaches for detection and correction.

References

Abeer Mohamed Abd El Razek Youssef. (2022). Detecting of Multicollinearity, Autocorrelation and Heteroscedasticity in Regression Analysis. *Advances, 3*(3), 140-152. https://doi.org/10.11648/j.advances.20220303.24

Astivia, O. L. O., & Zumbo, B. D. (2019). Heteroscedasticity multiple regression analysis: what it is, how to detect it and how to solve it with application in R and SPSS. Practical Assessment, *Research, and Evaluation, 24*(1), 1-17. Web.

Baum, C. F., & Lewbel, A. (2019). Advice on using heteroskedasticity-based identification. *Stata Journal, 19*(4), 757–767. Web.

Charpentier, A., Ka, N., Mussard, S., & Ndiaye, O. H. (2019). Gini coefficients and heteroscedasticity. *Econometrics, 7*(1), 1–16. Web.

Dobson, A. J., & Barnett, A. G., (2008). *An introduction to general linear models (3rd ed.*). Boca Raton, FL: CRC Press.

Kaufman, R.L. (2014). *Heteroskedasticity in Regression; Detection and Correction*. Sage Publication, Inc. Doi: https://doi.org/10.4135/9781452270128.

Kim, J. H. (2019). Multicollinearity and misleading statistical results. *Korean journal of anesthesiology, 72* (6), 558-569.

Kumar, N.K. (2024). F-Test and Analysis of Variance (ANOVA) in Economics. *Mikailalsys Journal of Mathematics and Statistics*, *2*(3), 102-113. https://doi.org/10.58578/mjms.v2i3.3449

Mukherjee, A. Kr., & Laha, M. (2019). The problem of autocorrelation in linear regression detection and remedies. *International Journal of Multidisciplinary Research and Modern Education, 5*(1), 105–110. Web.

Shrestha, N. (2020) Detecting Multicollinearity in Regression Analysis. *American Journal of Applied Mathematics and Statistics, 8*, 39-42. https://doi.org/10.12691/ajams-8-2-1

Ye, X. & Sun, Y. (2018). *Heteroscedasticity and Autocorrelation Robust F and T test in Stata*. San Diego: Department of Economics.

Yin, Y. (2020). Model-free tests for series correlation in multivariate linear regression. *Journal of Statistical Planning and Inference, 206,* 179–195. Web.