P-ISSN 1858-2192 | E-ISSN 2686-5467
BALANCE
JURNAL EKONOMI
PENERBIT:
PROGRAM STUDI ILMU EKONOMI DAN STUDI PEMBANGUNAN
FAKULTAS EKONOMI DAN BISNIS
UNIVERSITAS MUHAMMADIYAH MAKASSAR

A Theoretical Study of Multicollinearity and Linearity in Econometric Models for Economic Research

Muhammad Jiyad Naufal[1], Dicky Perwira Ompusunggu[2], Rika Angelina Sinaga[3], Marwindi Dola Anggia Sitohang[4], Teresia Novita Gunawan[5], Magdalena Simatupang[6], Nur Syifa Salsabila[7], Tesalonika Simanullang[8], Bobin Trianko Hutasoit[9]
Faculty of Economics and Business, University of Palangka Raya
email: jiyadnaufal27@gmail.com

Keywords:
multiple linear regression, multicollinearity, linearity, econometric model, classical assumptions, literature review

Abstract

Multiple linear regression is a central analytical tool in econometric research used to model the relationship between a dependent variable and multiple independent variables. However, the accuracy and validity of such models are highly dependent on classical assumptions, particularly multicollinearity and linearity. Multicollinearity, characterized by high correlations among predictor variables, can inflate standard errors and obscure the true effects of individual variables. Linearity, meanwhile, ensures that the relationships between variables follow a straight-line pattern, which is essential for valid estimation and inference. This theoretical study aims to deepen the understanding of both assumptions, explore their causes, impacts, and identify methodological approaches for detection and correction. Employing a descriptive literature review method, the study synthesizes insights from contemporary econometric research to provide a conceptual framework for handling these issues. Key findings highlight that multicollinearity often arises from overlapping variables, small samples, and measurement errors, and can be addressed through variable elimination, transformation, or penalized regression techniques such as ridge and lasso regression. Linearity violations, frequently resulting from model misspecification or temporal dependencies, may be mitigated using data transformations, polynomial regression, or robust regression approaches. The study concludes that proper diagnostic tools and corrective strategies are essential for improving model reliability and enhancing the credibility of econometric findings in economic research.

1. Introduction

Multiple linear regression is a widely used statistical technique to examine the relationship between a single dependent variable and multiple independent variables (Kasemset et al., 2014). This method is essential in various fields, including economics, social sciences, and health, for modeling and predicting complex phenomena. However, to ensure the validity and reliability of the regression model, researchers must adhere to the classical assumptions of regression analysis.

These classical assumptions include linearity (the relationship between variables should be linear), independence (observations are independent), homoscedasticity (constant variance of residuals), normality (normally distributed residuals), and the absence of multicollinearity (no high correlation among independent variables) (Morrissey & Ruxton, 2018). Violations of these assumptions can result in biased, inefficient estimates and incorrect inferences (Sebayang & Yuniarto, 2017).

Recent studies have emphasized the importance of diagnosing and correcting assumption violations in regression models. Hasanah et al. (2021) highlight the significance of detecting multicollinearity, while Nastiti et al. (2023) demonstrate advanced diagnostic tools and data transformation techniques to address assumption violations. Nwaigwe et al. (2004) underscore the role of proper variable selection in minimizing assumption violations and improving model accuracy.

Multicollinearity, one of the most common issues in multiple regression, occurs when independent variables are highly correlated. It can be identified through diagnostic tools such as the Variance Inflation Factor (VIF) and Tolerance (Widana & Muliani, 2020). Conversely, linearity refers to whether the relationship between the dependent and independent variables can be accurately represented by a straight line (Martaningtyas et al., 2024).

Given the critical role of these assumptions in regression modeling, this study aims to provide a theoretical exploration of multicollinearity and linearity. It identifies their causes, impacts, and provides practical approaches to address them. This paper also emphasizes the need for proper training in regression diagnostics to enhance the credibility and reliability of economic research outcomes.

2. Literature Review

2.1 Multicollinearity Test

Multicollinearity refers to a high correlation between two or more independent variables in a multiple linear regression model (Widarjono, 2010). This condition can distort the estimation of regression coefficients, inflate standard errors, and reduce the statistical power of the model. A good regression model assumes that all independent variables are mutually uncorrelated or orthogonal, meaning the correlation value between them approaches zero (Ghozali, 2016). When multicollinearity is present, it becomes difficult to determine the individual effect of each independent variable on the dependent variable, as their shared variance may overlap. For instance, in a regression model that uses motivation, leadership, and job satisfaction as predictors of performance, high correlation between any of these predictors may compromise the validity of the model (Marsuni, 2024). Although multicollinearity is a matter of degree rather than type, it can be detected using several statistical techniques. Common diagnostic tools include the Variance Inflation Factor (VIF), Tolerance, Pearson correlation matrix, and Condition Index (CI) (UT, 2021). A VIF value above 10 or a Tolerance value below 0.1 generally indicates a serious multicollinearity problem. Addressing this issue may involve variable elimination, data transformation, or using alternative modeling techniques such as ridge regression.

2.2 Linearity Test

Linearity is a fundamental assumption in regression analysis, where the relationship between independent and dependent variables is expected to follow a straight-line equation. Mathematically, a linear relationship can be expressed as $y = m x + b$ y=mx+b, where y y is the dependent variable, x x is the independent variable, m m is the slope, and b b is the intercept. This assumption is critical for the accuracy and interpretability of the model. Linear models are preferred for their simplicity, ease of computation, and intuitive interpretation. The property of superposition in linear relationships allows the use of multiple independent variables in a combined form to predict the dependent variable. However, not all relationships in real-world data are linear. In cases where variables exhibit curvilinear or complex interactions, applying a linear model may lead to misleading results. Techniques such as curve estimation, residual plots, or scatterplots are used to examine the linearity of the relationship (Martaningtyas et al., 2024). If residuals display non-random patterns, it suggests that the linearity assumption is violated, and a non-linear model may be more appropriate. Before fitting a regression model, researchers should perform a linearity test to ensure the assumptions are met. Failure to do so can result in biased estimates and poor predictive accuracy, ultimately weakening the credibility of the research findings (Supriyadi et al., 2017).

3 Research Methods

This study employs a descriptive qualitative approach using a systematic literature review to explore and analyze the concepts of multicollinearity and linearity within the framework of classical assumptions in multiple linear regression. The focus is to build a theoretical foundation that enables researchers to better understand how violations of these assumptions can compromise the reliability and accuracy of econometric models.

The literature review method was chosen to synthesize theoretical insights from a variety of scholarly sources, including peer-reviewed journal articles, academic books, and relevant empirical studies. The inclusion criteria for the literature were based on relevance to the topics of multicollinearity, linearity, and classical

P-ISSN 1858-2192 | E-ISSN 2686-5467
BALANCE
JURNAL EKONOMI
PENERBIT:
PROGRAM STUDI ILMU EKONOMI DAN STUDI PEMBANGUNAN
FAKULTAS EKONOMI DAN BISNIS
UNIVERSITAS MUHAMMADIYAH MAKASSAR

assumption testing in econometrics, with a preference for recent and high-impact publications. The stages of this research included:

1. Identifying and selecting relevant academic literature;
2. Extracting key concepts and findings related to regression assumptions;
3. Analyzing and synthesizing theoretical perspectives to propose appropriate diagnostic and corrective measures;
4. Compiling the results into a coherent discussion that addresses the implications of assumption violations for regression-based research.

As this is a theoretical study, no primary data collection or fieldwork was conducted. The methodology emphasizes conceptual analysis and critical evaluation of existing studies, aiming to produce a comprehensive reference for researchers concerned with ensuring the validity and reliability of regression models. This approach provides both a foundational and practical understanding of how to identify and manage multicollinearity and linearity issues in economic research.

4 Results and Discussion

4.1 Multicollinearity Test

4.1 Cause Symptom / Violation

Multicollinearity occurs when two or more independent variables in a regression model are highly correlated, leading to redundancy in the information they provide. This often stems from poor research design, natural relationships among variables (e.g., income and consumption), the inclusion of overlapping indicators (e.g., "years of education" and "education level"), or excessive use of dummy variables (Sriningsih et al., 2018; Arisandi et al., 2021). Time-related variables in time-series data may exhibit similar trends (e.g., inflation and GDP), increasing the risk of multicollinearity. Measurement error and small sample sizes further exacerbate this issue, making spurious correlations appear stronger than they truly are. The correlation between these variables can be measured by the Pearson correlation formula as follows:

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \; \sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

Description :

r : Pearson correlation coefficient that measures the relationship between variables X and Y

$\bar{X}$ dan $\bar{Y}$: means of variables X and Y

X_i dan Y_i : individual values of variables X and Y

n : number of data in the sample

This relationship can occur naturally without intervention from the researcher, but can cause the analysis results to be biased if not handled properly. Multicollinearity can also arise when the independent variables used in the model are too similar or substitute for each other. Under these conditions, difficulties arise in determining the individual effect of each variable on the dependent variable, which can result in unstable estimates and less clear interpretations (Yaldi et al., 2022).

The use of overlapping variables is often the cause of multicollinearity. This usually occurs when researchers include two or more variables that measure the same aspect of a phenomenon. For example, in an analysis involving education level, including variables such as "years of education" and "education level" in the same model can result in multicollinearity because the two variables are conceptually closely related (Sriningsih et al., 2018). Excessive or inappropriate dummy variables can also be a source of multicollinearity. This is often found in models that try to capture the effects of certain categories with dummy variables, especially when all categories are included in the model. This situation is known as the "dummy variable trap," where the number of dummy variables equals the number of categories, causing information redundancy.

The effect of time or trends in the data can also cause multicollinearity, especially in time series analysis. Variables that change over time, such as inflation rates, interest rates, and GDP growth, tend to have similar trend patterns. Poor research design can be the root of the multicollinearity problem. If researchers do not conduct an initial exploration of the data or do

P-ISSN 1858-2192 | E-ISSN 2686-5467
BALANCE
JURNAL EKONOMI
PENERBIT:
PROGRAM STUDI ILMU EKONOMI DAN STUDI PEMBANGUNAN
FAKULTAS EKONOMI DAN BISNIS
UNIVERSITAS MUHAMMADIYAH MAKASSAR

not pay attention to the relationship between independent variables before building the model, multicollinearity is likely to appear (Arisandi et al., 2021). Therefore, it is important for researchers to conduct preliminary analysis, such as evaluating the correlation matrix or using other exploratory methods, to ensure that the independent variables included in the model do not have too strong a linear relationship.

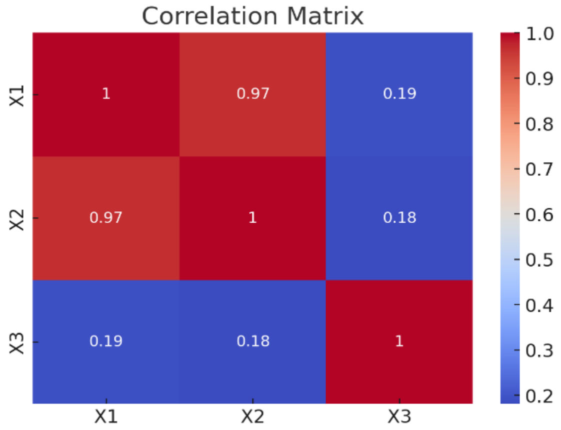

Figure 1 *Correlation Matrix*

Inaccurate measurement of variables can also trigger multicollinearity. When data is measured in an inconsistent way or there are measurement errors, the results can be inaccurate and lead to false relationships between the independent variables. For example, if researchers use data taken from different sources with different measurement methods, the variables may show relationships that do not reflect reality. In terms of sample size, small sample size can exacerbate the problem of multicollinearity. In statistical analysis, a limited amount of data often makes the relationship between variables more obvious. In other words, in a small sample, independent variables are more likely to appear strongly correlated even though the relationship is actually insignificant in a larger population.

4.2 Impact

Multicollinearity inflates the variance of coefficient estimates, making them unstable and statistically insignificant despite actual effects. This undermines interpretability and predictive power (Lin et al., 2011). Tools such as the Variance Inflation Factor (VIF) and Condition Index are commonly used to detect its presence. A VIF >10 or a Condition Index >30 typically signals serious multicollinearity (Aditiya et al., 2023).

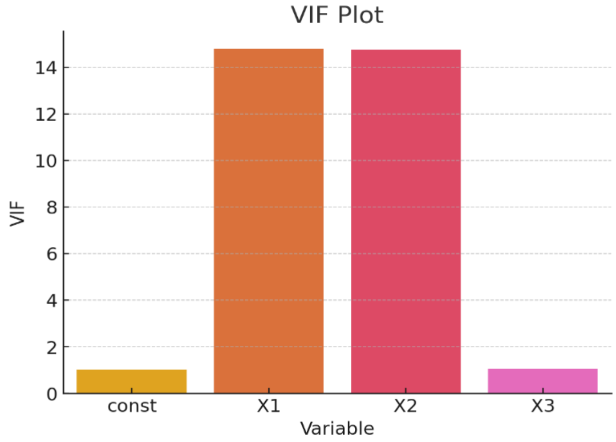

Figure 2 *VIF Plot*

In addition, multicollinearity can cause the variance value of parameter estimates to be very large. This high variance makes parameter estimation inefficient and tends to have large standard errors. In other words, the confidence intervals for the regression coefficients become wide, making it difficult to accurately test the significance of the independent variables. In such a situation, although the relationship between the independent and dependent variables is actually significant, statistical tests may show the opposite result. This can also be seen through the Condition Index Plot, which shows the severity of multicollinearity in the model. According to Aditiya et al., (2023) condition index values above 30 are often an indication that the model matrix is in an unstable condition, making the analysis results less reliable.

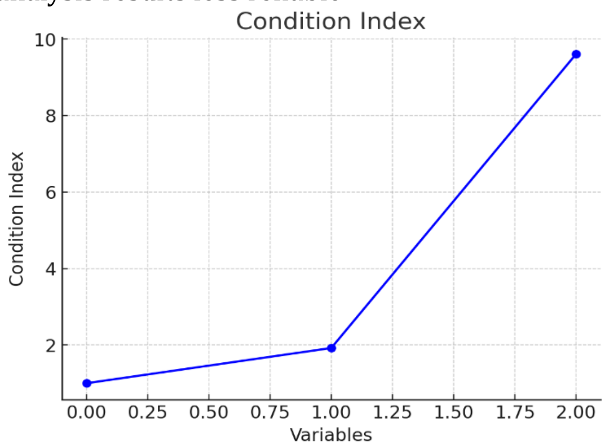

Figure 3 *Condition Index Plot*

Multicollinearity can also cloud the interpretation of the regression model. When there is a strong linear relationship between independent variables, it is difficult to determine whether the effect of a particular variable on the dependent variable is the direct result of that variable or the result of a relationship with other independent variables. As a result, conclusions drawn from the regression model may be misleading or irrelevant in the context of the study.

Another impact of multicollinearity is a decrease in the value of the t-statistic for a particular independent variable. This decrease is due to the increased standard errors associated with the regression coefficients. As a result, independent variables that should be significant in influencing the dependent variable may appear insignificant in the analysis results. This can lead to the removal of actually important variables from the model, which ultimately harms the interpretation and validity of the results.

Regression coefficients produced by models affected by multicollinearity also often have signs or values that do not make sense (Agustin & Astuti, 2020). For example, variables that theoretically have a positive relationship with the dependent variable may show a negative relationship in the regression results. This casts doubt on the validity of the model and makes it difficult to logically explain the phenomenon under study. In addition, multicollinearity can reduce the model's ability to accurately predict new data. Models that suffer from multicollinearity tend to be overfitting, where the model overfits itself to the sample data but fails to predict the same pattern in other data. In a business or policy context, this can result in incorrect decisions based on biased model predictions.

Finally, multicollinearity can decrease the efficiency of data utilisation. When it occurs, highly correlated independent variables do not actually add much new information to the model, but instead complicate the analysis (Widiastuti & Raharjo, 2020). As a result, the model becomes more complex without providing a significant improvement in the ability to explain the dependent variable. Therefore, identifying and addressing multicollinearity is essential to ensure valid, efficient and reliable analysis results.

4.3 Handling

Multicollinearity is a common problem that can arise in regression analysis, especially when the independent variables have a very strong linear relationship with each other. While multicollinearity does not necessarily violate the basic assumptions of regression, its presence can cloud the interpretation of results and reduce the validity of the model. One of the simplest ways to deal with multicollinearity is to remove one of the independent variables that is highly correlated with the other variables. This approach is effective if the deleted variable is not theoretically important or does not contribute significant information to the model. According to Suyono in Amelia & Putra, (2023), researchers can use correlation analysis or Variance Inflation Factor (VIF) to identify which variables should be removed from the model. VIF measures how much the variance of the regression coefficient increases due to multicollinearity. The formula for calculating VIF is as follows:

$$VIF = \frac{1}{1 - R^2}$$

where R2 is the coefficient of determination from regressing that independent variable on all other independent variables in the model. VIF values greater than 10 indicate serious multicollinearity and need to be addressed.

Another approach is to combine highly correlated independent variables into one new variable. This technique is often referred to as index formation or composite score. For example, if there are two variables that both measure a certain aspect of a phenomenon, they can be combined into one using methods such as averaging or factor analysis. This approach preserves the information that the variables have without causing multicollinearity.

Variable transformation can also be an option to overcome multicollinearity. One commonly used form of transformation is logarithmising or squaring the data. This transformation can reduce the correlation between independent variables, especially if the relationship between the variables is non-linear. In addition, this technique can also help correct abnormal data distribution, which is sometimes the root of multicollinearity problems. In addition, it can also be considered to use regression analysis techniques that are more robust to multicollinearity, such as ridge regression or lasso regression. Ridge regression adds a penalty to the estimated regression coefficients, which makes the parameters smaller and more stable, despite the presence of multicollinearity (Çankaya et al., 2019). The ridge regression model can be written with the following formula:

$$\hat{\beta}^{ridge} = \arg\min_{\beta}(\sum_{i=1}^{n}(Y_i - X_i\beta)^2 + \lambda\sum_{j=1}^{P}\beta_j^2)$$

di where λ is a penalty parameter that controls how much penalty is applied to the coefficients.

Lasso regression (Least Absolute Shrinkage and Selection Operator) is another method that can be used to overcome multicollinearity (Saleh, 2020). Lasso also adds a penalty, but uses the L1 norm to calculate the penalty, which can force the coefficients to be zero. This provides the added advantage of being able to select the most relevant variables for the model. The formula for lasso regression is:

$$\hat{\beta}^{lasso} = \arg\min_{\beta}(\sum_{i=1}^{n}(Y_i - X_i\beta)^2 + \lambda\sum_{j=1}^{P}|\beta_j|)$$

If multicollinearity comes from time series data or data with a certain trend, detrending can be an effective solution (Subagyo, 2018). Detrending is done by removing the trend effect from the independent variable, for example by calculating the difference between the current value and the previous value. This technique helps to reduce correlations caused by time patterns, making the model more stable.

Increasing the sample size can also help reduce multicollinearity. When the sample size is small, the relationship between independent variables tends to look stronger than it actually is. By increasing the amount of data, the variance of the parameter estimates will decrease, and the relationship between the independent variables will look closer to reality. However, this approach requires greater resources, such as time and money, so it is not always feasible.

It is important for researchers to conduct preliminary analyses of the data before building a regression model. Steps such as evaluating the correlation matrix, calculating VIF values, and performing other exploratory analyses can help identify potential multicollinearity early on. Through a better understanding of the relationship between independent variables, researchers can design more efficient models and minimise the risk of multicollinearity.

4.2 Linearity Test

a. Cause Symptom/Violation

Violation of the linearity assumption occurs when the relationship between the independent and dependent variables is not linear. In many studies, this relationship is assumed to be linear based on existing theory, but it is not always proven in practice. If the relationship is not linear, then the regression model constructed will give inaccurate and misleading results. Some types of violations that can occur include bias in model specification because some important variables are not included. This situation arises when the analysis does not consider variables that should be present, which can lead to inappropriate or misleading conclusions, as well as a potential misunderstanding of the relationship between the variables under study (Rizky et al., 2024).

In addition, the interconnectedness of time series data can lead to inertia, where there is often an interdependent relationship between successive values, such that a change in one variable does not affect another variable immediately, but takes some time to show its effect. Time lags can also affect the relationship between variables, where the impact of one variable on another does not occur immediately, but rather takes time before the effect is seen, making it difficult to determine an accurate cause-and-effect relationship. Improper data processing can also result in distortions in the analysis results; inappropriate data manipulation or processing actions can undermine the accuracy of the analysis results, including the removal of invalid data or the application of analysis methods that are not suitable for the characteristics of the data (Awalia & Sihombing, 2022).

The cobweb phenomenon can arise in the context of variable fluctuations, where a change in one variable triggers an unstable chain reaction in another variable, often occurring in an economic or market context. When producers or consumers respond to changes in price or demand, they may make inappropriate decisions based on incomplete information, which can lead to a continuous cycle of fluctuations that are difficult to predict (Rizky et al., 2024).

b. Impact

Violation of the linearity assumption in regression models can lead to various serious problems, one of which is estimation error. When the model does not fulfil this assumption, the resulting parameter estimates become inaccurate, which means that the parameter values do not reflect the values that actually exist in the population. As a result, the prediction of the value of the dependent variable becomes inappropriate. For example, if the model used to predict sales based on advertising expenditure does not consider the non-linear relationship between the two variables, the prediction results can be very different from reality, which has a negative impact on marketing strategies and decision-making in business (Nastiti et al., 2023).

In addition, violations of the linearity assumption can also lead to incorrect interpretations of the relationship between the variables being analysed. For example, if the true relationship is quadratic but the analysis is conducted by imposing a linear model, the resulting conclusions will be erroneous. This misinterpretation can lead to inappropriate decision-making, as the researcher or decision-maker may feel that they have correctly understood the relationship between the variables, when in fact this is not the case. The consequences of this misinterpretation can be very serious, especially in the context of scientific research or data analysis used for public policy. In addition, limitations in prediction are also an issue, where regression models that do not conform to the assumption of linearity will reduce the ability to make accurate predictions. This inaccuracy not only reduces the reliability of the analysis results, but can also result in poorly informed data-driven decisions. In a business context, decisions made based on inaccurate predictions can lead to significant financial losses, as companies may allocate resources inefficiently or fail to respond appropriately to market changes (Awalia & Sihombing, 2022).

c. Identification Method

One way to detect a linearity violation is to examine the residual pattern of the regression model. Residuals are the difference between the value predicted by the model and the actual value of the dependent variable. By plotting the residuals against the predicted value or independent variable, we can identify if there are certain patterns that indicate that the relationship between the variables is not linear. If the residuals show a systematic pattern, such as a curve or repeating pattern, this could be an indication that the model used is not appropriate and needs to be improved (Mardiatmoko, 2020).

In addition to visual analysis, we can also use statistical tests to determine whether the regression model fulfils the linearity assumption. Tests such as Ramsey RESET (Regression Specification Error Test) or Lagrange Multiplier can be used to detect model misspecification. These tests help in determining if there are missing variables or if the relationship between the variables does not conform to the linearity assumption required for valid regression analysis. By using this method, researchers can be more confident that the model they are using is an accurate representation of the data being analysed (Anam, 2020).

d. Handling

To handle violations of the linearity assumption in regression analysis, there are several approaches that can be taken. One of them is data transformation, where researchers can use the logarithm or square of the variables to try to change the relationship to be more linear (Awalia & Sihombing, 2022). If the transformation is unsuccessful, researchers can consider using non-linear models or polynomial regression that better fits the existing data pattern (Setya Budi et al., 2024). In addition, the application of robust methods can also be a solution, as these methods can help correct the influence of outliers and provide more stable estimates even if classical assumptions are violated (Anam, 2020). By applying these approaches, researchers can improve the accuracy and reliability of the analysis results.

5 Closing

5.1 Conclusion

Multicollinearity and violations of the linearity assumption are two critical challenges in regression analysis that can significantly distort model accuracy, interpretation, and predictive reliability. Multicollinearity arises when independent variables are highly correlated, leading to inflated variances and unstable coefficient estimates. Similarly, non-linearity between predictors and the dependent variable results in biased estimations and reduced model validity. Both issues often stem from inadequate research design, data-related problems, or model misspecification. Addressing these problems is essential to ensure the robustness of empirical findings.

5.2 Recommendation

Researchers are advised to perform comprehensive diagnostic tests before interpreting regression outputs. For multicollinearity, tools such as Variance Inflation Factor (VIF) should be routinely used, and corrective measures—like variable elimination, dimension reduction, or penalized regression methods (ridge, lasso)—should be considered when necessary. For linearity issues, residual plot analysis and tests like the Ramsey RESET should be conducted, followed by appropriate model adjustments, such as variable transformation or the application of non-linear or robust regression techniques. By implementing these methodological safeguards, future studies can produce more reliable and valid regression models that better inform theory and practice.

Bibliography

Aditiya, N. Y., Evani, E. S., & Maghfiroh, S. (2023). The concept of classical assumption test on multiple linear regression. 2(2), 102-110.

Amelia, S., & Putra, A. A. (2023). Principal Component Regression in Overcoming Multicollinearity in Factors Affecting Local Revenue in West Sumatra. 7, 10906-10914.

Anam, C. (2020). Types of statistical tests for analysis of research results. Study, 23(4), 115-117.

Ani, N. (2023). The Effect of Digital Marketing, Electronic Word of Mouth and Lifestyle on Purchasing Decisions at Tiktok Shop Indonesia. BISMA: Business and Management Journal, 1(04), 37-44. https://doi.org/10.59966/bisma.v1i04.398

Arisandi, R., Ruhiat, D., & Marlina, E. (2021). Implementation of ridge regression to overcome multicollinearity symptoms in rainfall modelling based on climatological time series data. JRMST| Journal of Research ..., 1(November), 1-11. https://ejournal.unibba.ac.id/index.php/jrmst/article/view/735%0Ahttps://ejournal.unibba.ac.id/index.php/jrmst/article/download/735/666

Awalia, S., & Sihombing, W. L. (2022). The Effect of Time Token Type Cooperative Learning Model on Students' Motivation to Learn Mathematics on Triangle Material in Class Vii Mts. Amin Darussalam Tembung in the 2021/2022 academic year. Indonesian Multi Disciplinary Scientific Journal, 1(9), 1278-1285.

Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer.

Budi, A. D. A. S., Septiana, L., & Mahendra, B. E. P. (2024). Understanding Classical

Assumptions in Statistical Analysis: An In-depth Study of Multicollinearity, Heteroscedasticity, and Autocorrelation in Research. Multidisciplinary Journal of Western Science, 3(01), 01-11.

Çankaya, S., Eker, S., & Hasan, S. (2019). Comparison of Least Squares, Ridge Regression and Principal Components Approaches in the Presence of Multicollinearity in Regression Analysis. 7(8), 1166-1172.

CME. (2001). ΝοΔιαγνωστικές εξετάσεις για τον καρκίνο του ήπατος Title. 2017, 4(2), 1-11. http://www.helpa-prometheus.gr/διαγνωστικές-εξετάσεις-για-τον-καρκί/

Draper, N. R., & Smith, H. (1998). Applied Regression Analysis. New York: John Wiley & Sons.

Fox, J. (2015). Applied Regression Analysis and Generalized Linear Models. Thousand Oaks, CA: Sage Publications.

Ghozali, I. (2005). Application of Multivariate Analysis with the SPSS Program. Semarang: Diponegoro University Publishing Agency.

Ghozali, I. (2016). Multivariate Data Modeling. Semarang: Diponegoro University Publishing Agency.

Ghozali, I. (2021). Multivariate Statistics for Economics and Business. Semarang: Diponegoro University Publishing Agency.

Gujarati, D. N. (2015). Basic Econometrics. McGraw-Hill Education

Hasanah, N., Mutiasari, & Hartati, S. (2021). Analysis of Factors Affecting the Performance of Employees of the Secretariat of the Regional People's Representative Council (Dprd) of Cilacap Regency. AmaNU: Journal of Management and Economics, 4(1), 53-67.

Jampachaisri, K., & Tinochai, K. (2019). Parameter estimation methods in multiple linear regression analysis with intraclass correlation and heavy-tailed distributed data. Journal of Applied Science, 18(2), 11-21. https://doi.org/10.14416/j.appsci.2019.07.002

Kasemset, C., Sae-Haew, N., & Sopadang, A. (2014). Multiple regression model for forecasting off-season litchi supply quantity. Chiang Mai University Journal of Natural Sciences, 13(3), 391-402. https://doi.org/10.12982/cmujns.2014.0044

Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). Applied Linear Statistical Models. New York: McGraw-Hill.

Lin, D., Foster, D. P., & Ungar, L. H. (n.d.). VIF Regression: A Fast Regression Algorithm for Big Data. 19104.

Mardiatmoko, G. (2020). The Importance of the Classical Assumption Test in Multiple Linear Regression Analysis. BAREKENG: Journal of Mathematical and Applied Sciences, 14(3), 333-342. https://doi.org/10.30598/barekengvol14iss3pp333-342

Martaningtyas, N. U., Septiyaningrum, E. A., & Maulana, Z. (2024). The Impact of Classical Assumption Violation on Inference Error in Econometric Analysis. SYNERGY Multidisciplinary Scientific Journal, 1(4), 255-265. https://e-journal.naureendigition.com/index.php/sjim

Montgomery, D. C., & Runger, G. C. (2010). Applied Statistics and Probability for Engineers. New York: John Wiley & Sons.

Morrissey, M. B., & Ruxton, G. D. (2018). Multiple Regression Is Not Multiple Regressions: The Meaning of Multiple Regression and the Non-Problem of Collinearity. Philosophy, Theory, and Practice in Biology, 10(20220112). https://doi.org/10.3998/ptpbio.16039257.0010.003

Nastiti, E., Damayanti, T., & Madina, S. A. (2023). The Impact of Classical Assumption Violations on Econometric Model Estimation. Journal of Pijar Management and Business Studies, 1(3), 566-577. https://e-journal.naureendigition.com/index.php/pmb

Nurcahya, W. A., Arisanti, N. P., & Hanandhika, A. N. (2024). Application of the Classical Assumption Test to Detect Errors in Data as an Effort to Avoid Violations of Classical Assumptions. Madani: Multidisciplinary Scientific Journal, 1(12).

Nwaigwe, C. C., Uche, P., & Onuoha, C. (2004). A test for the parameters of multiple linear regression models. Global Journal of Mathematical Sciences, 3(2).https://doi.org/10.4314/gjmas.v3i2.21364

Rizky, M., Saputra, H., Ramadhan Basuki, R., & Muhtadin, I. A. (2024). Regression analysis on classical assumption violations in linear regression. Multidisciplinary Scientific Journal, 307(1), 307-314. https://doi.org/10.5281/zenodo.10537197

Rousseeuw, P. J., & Leroy, A. M. (1987). Robust Regression and Outlier Detection. John Wiley & Sons.

Saleh, S. (2020). ROBUST VARIABLE SELECTION IN LINEAR REGRESSION MODELS. January 2015. https://doi.org/10.13140/RG.2.2.10348.39044

Sebayang, J. S., & Yuniarto, B. (2017). Comparison of Artificial Neural Network Estimation Model Optimization Genetic Algorithm and Multiple Linear Regression. STATISTICS MEDIA, 10(1), 13. https://doi.org/10.14710/medstat.10.1.13-23

Setya Budi, A. D. A., Septiana, L., & Panji Mahendra, B. E. (2024). Understanding Classical Assumptions in Statistical Analysis: An In-depth Study of Multicollinearity, Heteroscedasticity, and Autocorrelation in Research. West Science Multidisciplinary Journal, 3(01), 01-11. https://doi.org/10.58812/jmws.v3i01.878

Sholihah, S. M. A., Aditiya, N. Y., Evani, E. S., & Maghfiroh, S. (2023). The Concept of Classical Assumption Test in Multiple Linear Regression. Soedirman Journal of Accounting Research, 2(2), 102-110.

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

Sriningsih, M., Hatidja, D., & Prang, J. D. (2018). Multicollinearity Handling Using Principal Component Regression Analysis in the Case of Rice Imports in North Sulawesi Province. Scientific Journal of Science, 18(1), 18. https://doi.org/10.35799/jis.18.1.2018.19396

Subagyo, A. (2018). Economic statistical analysis techniques.

Supriyadi, E., Mariani, S., & Sugiman. (2017). Comparison of Partial Least Square (PLS) and Principal Component Regression (PCR) Methods to Overcome Multicollinearity in Multiple Linear Regression Models. Unnes Journal of Mathematics, 6(2), 117-128.

UT. (2021). Statistics and Research Methods.

Widana, W., & Muliani, P. L. (2020). Analysis Requirements Test Book. In Analysis of Minimum Service Standards at the Outpatient Installation at Semarang City Hospital.

Widarjono, A. (2010). Econometrics: Theory and Applications. Yogyakarta: UPP STIM YKPN.

Yaldi, E., Pasaribu, J. P. K., Suratno, E., Kadar, M., Gunardi, G., Naibaho, R., Hati, S. K., & Aryati, V. A. (2022). Application of Multicollinearity Test in Human Resource Management Research. Scientific Journal of Management and Entrepreneurship (JUMANAGE), 1(2), 94-102. https://doi.org/10.33998/jumanage.2022.1.2.89