

# Detecting of Multicollinearity, Autocorrelation and Heteroscedasticity in Regression Analysis

**Abeer Mohamed Abd El Razek Youssef**

Faculty of Graduate Studies for Statistical Research, Cairo University, Giza, Egypt

**Email address:**

[abeer\\_yussef@yahoo.com](mailto:abeer_yussef@yahoo.com)

**To cite this article:**

Abeer Mohamed Abd El Razek Youssef. Detecting of Multicollinearity, Autocorrelation and Heteroscedasticity in Regression Analysis.

*Advances*. Vol. 3, No. 3, 2022, pp. 140-152. doi: 10.11648/j.advances.20220303.24

**Received:** August 15, 2022; **Accepted:** September 1, 2022; **Published:** September 29, 2022

---

**Abstract:** When we rely on the general linear regression model to represent the data, we use the ordinary least squares method to estimate the parameters of this model. This method, when applied, depends on the fulfillment of certain basic assumptions and conditions so that there is an accuracy in estimating the parameters of the regression model, and in many practical applications this hypothesis cannot be achieved, which makes the method of least squares ineffective in giving correct and accurate results, and this leads to falling into econometric problems. The estimated parameters lose the property of credibility, unbiased and make them not have the lowest possible variance and not expressive of the original theory. Most econometric models suffer from the problems of autocorrelation, multicollinearity, and heteroscedasticity. This paper presents a brief on these problems, their causes, how can be detected, tested, and minimized. The OLS method is based on several assumptions, and if these assumptions are fulfilled, we obtain unbiased, consistent, and efficient estimates (less variance compared to other methods). We discuss these problems as follows: First: the problem of multicollinearity Second: The problem of autocorrelation Third: Variation Heteroscedasticity. This article presents inference for many commonly used estimators - Variance Inflation Factors, Coefficient covariance matrix, Correlogram of Residuals, Normality Test for Residuals. Serial correlation LM test, Heteroskedasticity Test: Harvey, Actual and Estimated Residuals.

**Keywords:** Multicollinearity, Autocorrelation, Heteroscedasticity

---

## 1. Introduction

There is no doubt that the assumptions of the linear regression model may or may not be available. If available, the ordinary least squares method is valid for use in measuring the economic relations under study. But if it is not available, the method of ordinary squares does not become the appropriate method for estimating the parameters of economic relations, and these results in the emergence of some econometric problems that make this method an inappropriate method, and it is necessary to search in this case for other more appropriate standard methods.

I will present some standard problems encountered in the search:

1. The problem of autocorrelation
2. Variation homoskedasticity problem
3. The problem of lack of normal distribution
4. The problem of multicollinearity

## 2. Material and Methods

### 2.1. Autocorrelation and Detection Tests

#### *Causes of Autocorrelation*

- 1) Deleting some explanatory variables from the regression model results in the so-called deletion error, which in turn is reflected in the values of the random term.
- 2) Misidentification of the mathematical form of the model. For example, if the real relationship of a dependent variable is nonlinear, but the researcher has used a linear formula. Hence, without a doubt, the use of the linear formula instead of the non-linear one involves a certain type of error and is reflected in the random term [1].

Below are a few examples of some nonlinear formulas.

$$Y_i = A + BX_i^2 + U_i$$

$$Y_i = C + D(1/X) U_i$$

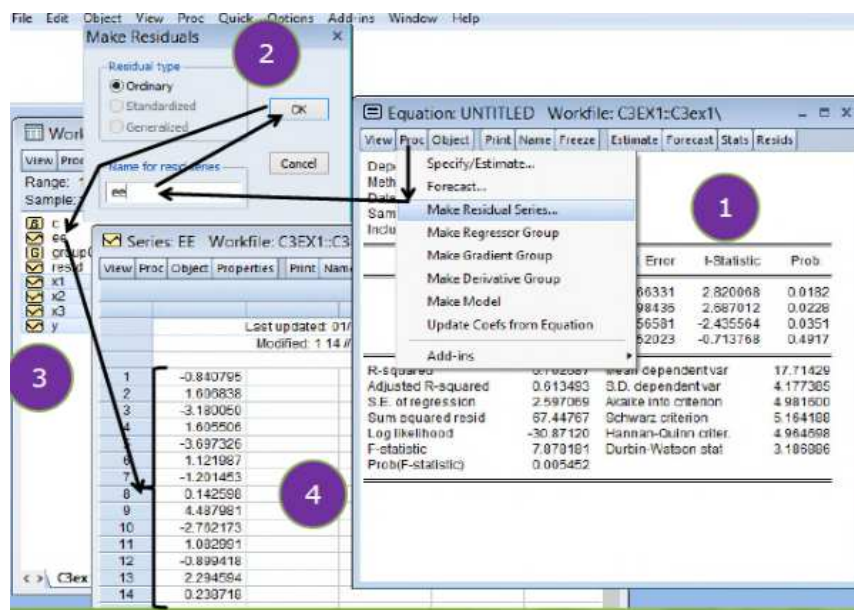
$$Y_i = F X_i^M U_i$$

As: A, B, C, D and F are constants whose value is estimated in the respective model. These formulas indicate that there is a nonlinear relationship between Y and the explanatory variable X in the three formulas. However, it is noted that redefining the variable X2 in the form numbered (1), as if we put  $X_2 = W$ , converts the original nonlinear relationship to a linear relationship:

$$Y = A + BW_i + U_i$$

The use of the mathematical logarithmic transformation transforms the relationship with the number (3) into a linear relationship as well:

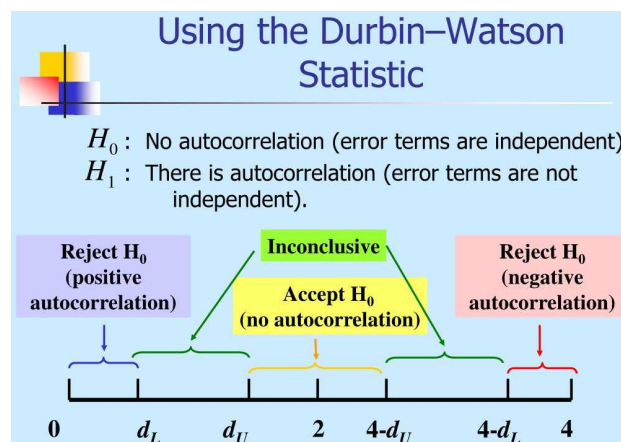
### 2.2.1. Steps of Residual Series



Source: Prepared by the researcher based on the statistical program EViews 10th Edition

Figure 1. Steps of Residual Series.

### 2.2.2. Using the Durban Watson Statistic



Source: Prepared by the researcher based on the statistical program EViews 10th Edition

Figure 2. Durban Watson.

$$\text{Log } Y_i = \text{Log } F + m \text{ Log } X + \text{Log } U_i$$

3-Data processing. In some cases, the published data may be monthly, and the researcher wants data on a quarterly basis, so he collects it and obtains an average of it. Perhaps it will provide fewer fluctuating data, which involves a kind of error that will be repeated from one observation to another because of the approximation process, which leads to the existence of autocorrelation [2].

### 2.2. Autocorrelation Tests

1. Durban Watson Test
2. Durban h test
3. Breusch-Godfrey serial correlation LM test

**Figure 4.** Steps serial correlation LM test.

But if we take the second case, the calculated  $\text{Obs} \cdot R$  squared statistic of 5.41 is greater than the tabular statistic which has a chi-square distribution, rejecting the null hypothesis and accepting the alternative hypothesis that is, the existence of autocorrelation [4].

#### 2.4. Inconsistency of Variance (Heterogeneity of Variance)

ARCH test: Test the variance hypothesis, using an autoregressive conditional variance test. We judge the results, whether the possibility of accepting the null hypothesis which states that the variance of the random error term in the estimated model is constant, or the rejection of the null hypothesis and the acceptance of the alternative hypothesis "absence of the homo of the variance." [5].

We conduct the ARCH Test to test the relationship between the square of residuals as a dependent variable and the square of the slowed residuals for one period to test the null hypothesis saying constant of variance. This test is based on either the classic Fisher test or the LaGrange multiple tests.

In the practical aspect of conducting this test, we follow the following steps [6]:

The first stage: calculating the error term  $t$  in the regression model.

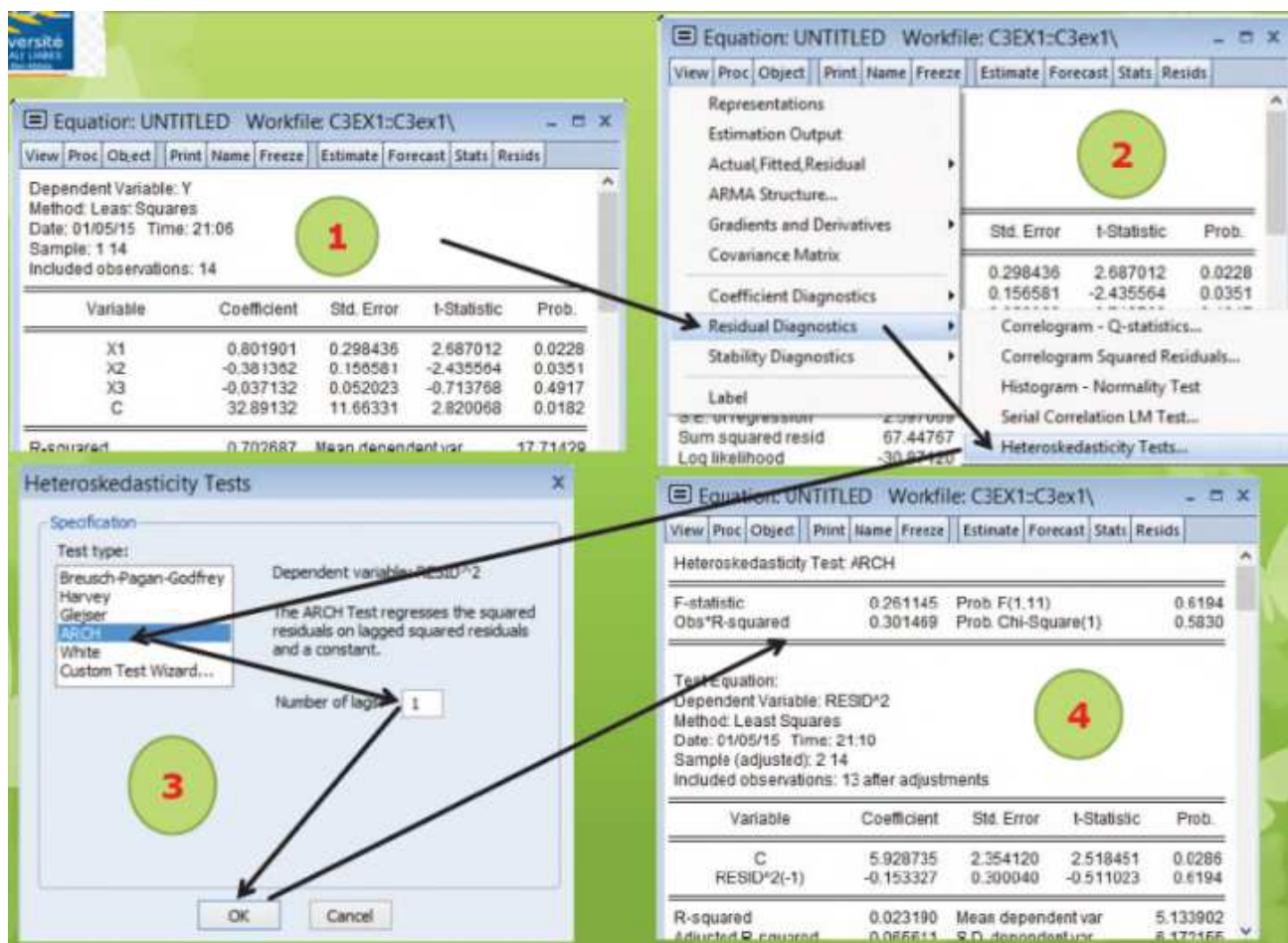
The second stage: calculating  $e_t^2$ .

The third stage: autoregression of residuals through  $p$  periods of deceleration. Significant delays are preserved.

Fourth stage: calculating the LaGrange multiplier statistic, where  $LM = N \cdot R^2$ , with  $N$  sample size and  $R^2$  representing the coefficient of determination [7].

### 3. Steps of Heteroscedasticity Test

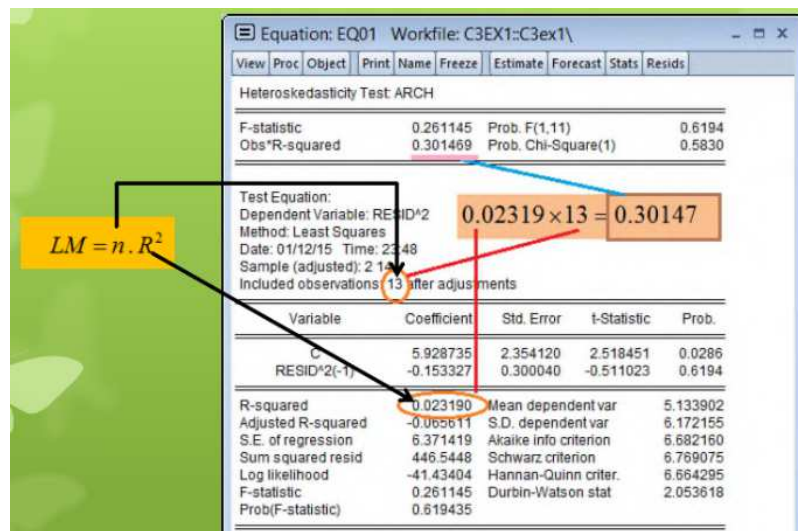
The result decides to accept the null hypothesis with the constant of the variance and the rejection of the inconstant of the variance for the error term series. This is the confirmation of the validity of the second hypothesis of the least squares method, which states constant of variance [8].



Source: Prepared by the researcher based on the statistical program EViews 10th Edition

Figure 5. Steps of heteroscedasticity test.

### 3.1. Steps of Heteroscedasticity Test



Source: Prepared by the researcher based on the statistical program EViews 10th Edition

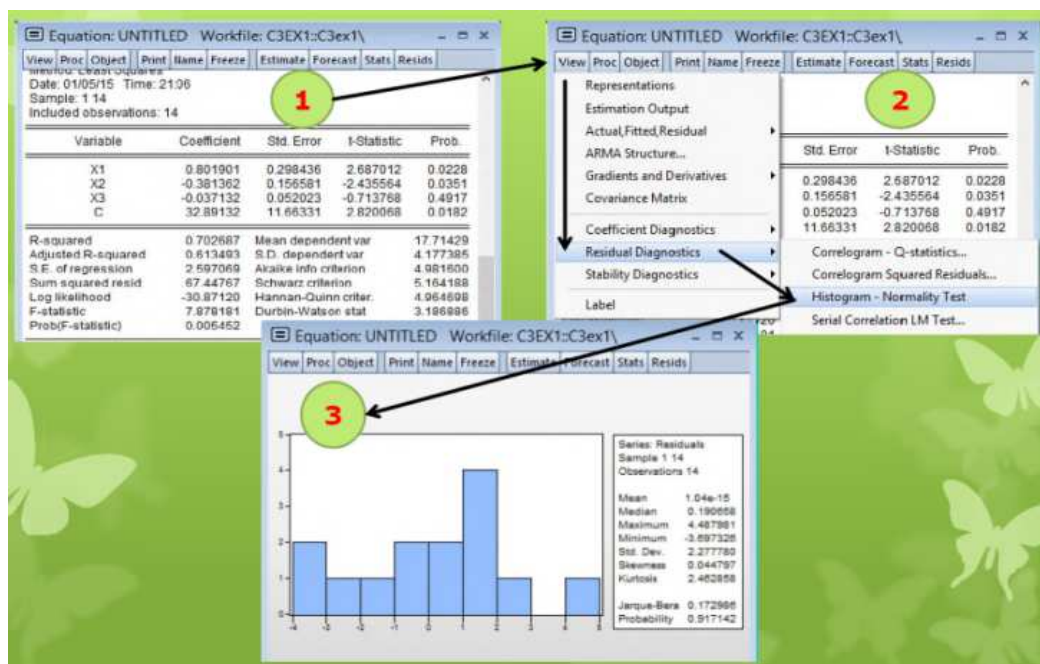
Figure 6. Calculation method of heteroscedasticity LM Test.

### 3.2. Test to Verify The Normalization of Residuals of the Regression Equation (Jarque-Bera Test)

For checking the normal distribution of residuals of the regression equation we use the Jarque-Bera test.

The null hypothesis that the residuals of the regression equation are normally distributed based on the statistic of this choice can be rejected or accepted [9].

We reject the null hypothesis if the JB statistic value is greater than the tabular value of the chi-square distribution [10].



Source: Prepared by the researcher based on the statistical program EViews 10th Edition

Figure 7. Steps of Normal Distribution Test.

## 4. Practical Application for Diagnostic Tests: Empirical Results

First hypothesis: There is a negative significant relationship with statistical significance between emissions of methane, nitrous and GDP.

Study variables: The variables used in estimating the model can be defined as follows:

*Table 1. Definition of Study variables.*

variable name	Definition	measruing unit	Variable type
LGDP	GDP	(current US\$)	dependent variable
LCO <sub>2</sub> _EMISSIONS	CO <sub>2</sub> emissions	(kt)	independent variable
LMETHANE_EMISSIONS	Agricultural methane emissions	(Thousand metric tons of CO <sub>2</sub> equivalent)	independent variable
LNITROUS_OXIDE_EMISSIONS	Agricultural nitrous oxide emissions	(Thousand metric tons of CO <sub>2</sub> equivalent)	independent variable

*Study population and sample:*

GDP was chosen as an indicator of production in Egypt and as a response variable (dependent); While were methane and nitrous included as an independent and explanatory variable, the study covers Egypt country with during the period 1990 to 2022, thus the number of observations used in the total sample is 33. The study was applied to Egypt [3].

**4.1. Application for Multiple Regression**

Overall Significance: Through the model, we find that the value of the F-statistic (0.000) is less than (0.05) indicating the overall Significance of the model, which is significant at the level of significance of 5%, meaning that the model is totally significant.

*Table 2. Estimation of Multiple Regression.*

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.590248	0.662695	3.908660	0.0005
LCO <sub>2</sub> _EMISSIONS	0.338262	0.076663	4.412349	0.0001
LNITROUS_OXIDE_EMISSIONS	-0.569029	0.142645	-3.989127	0.0004
LMANUFACTURING	0.920630	0.032070	28.70723	0.0000
LMETHANE_EMISSIONS	0.257899	0.123110	2.094861	0.0454
R-squared	0.997566	Mean dependent var		25.61279
Adjusted R-squared	0.997218	S. D. dependent var		0.753666
S. E. of regression	0.039749	Akaike info criterion		-3.473725
Sum squared resid	0.044240	Schwarz criterion		-3.246982
Log likelihood	62.31647	Hannan-Quinn criter.		-3.397433
F-statistic	2869.009	Durbin-Watson stat		0.832508
Prob (F-statistic)	0.000000			

Dependent Variable: LGDP

Method: Least Squares

Date: 07/26/22 Time: 04:26

Sample: 1 33

Included observations: 33

Source: Prepared by the researcher based on the statistical program EViews 10th Edition

As shown in the table significance of parameters this means that the model coefficients with statistical significance are represented the gross domestic production and emissions, where we find the probability of the fixed coefficient for each of them (0.0001), (0.0004), (0.0000), (0.0454) respectively, because they are less than the level of significance (0.05) [11].

It is clear from the Estimation of regression that the model suffers from problems to detect them, we follow the following tests:

**4.2. Variance Inflation Factors**

The variance inflation factor (VIF) is a measure of the amount of multiple linearity in a set of multiple regression variables. Mathematically, the VIF of a regression model variable is equal to the ratio of the total model variance to the model variance that includes only that single independent variable [12].

VIF values appear in the Centered VIF column (green values), VIF values show the variables that may be the cause of multicollinearity problem and whose value is higher than 10 (yellow values). Based on the highest value of VIF for an independent variable, it is the main cause of the multicollinearity [13].

*Table 3. Variance Inflation Factors.*

Variable	Coefficient	Uncentered	Centered
	Variance	VIF	VIF
C	0.439164	9172.407	NA
LCO <sub>2</sub> _EMISSIONS	0.005877	17576.45	17.57123
LMANUFACTURING	0.001028	12181.52	11.37272
LMETHANE_EMISSIONS	0.015156	29277.58	4.378833
LNITROUS_OXIDE_EMISSIONS	0.020348	38518.83	9.943755

Date: 07/26/22 Time: 09:02

Sample: 1 33

Included observations: 33

Source: Prepared by the researcher based on the statistical program EViews 10th Edition

**4.3. Coefficient Covariance Matrix**

The variance-covariance matrix forms the keystone artifact of regression models. The variance-covariance matrix of the regression model's errors is used to determine whether the model's error terms are homoscedastic (constant variance) and uncorrelated. The variance-covariance matrix of the fitted regression model's coefficients is used to derive the standard errors and confidence intervals of the fitted model's

coefficient estimates. Both matrices are used in forming the prediction intervals of the model's forecasts [14].

**Table 4.** Coefficient Covariance Matrix.

	C	LCO <sub>2</sub> _EMISSIONS	LMANUFACTURING	LMETHANE_EMISSIONS	LNITROUS_OXIDE_EMISSIONS
C	0.439164	0.020861	-0.004496	-0.039445	-0.021249
LCO <sub>2</sub> _EMISSIONS	0.020861	0.005877	-0.002078	0.001139	-0.005530
LMANUFACTURING	-0.004496	-0.002078	0.001028	-5.54E-05	0.000568
LMETHANE_EMISSIONS	-0.039445	0.001139	-5.54E-05	0.015156	-0.012460
LNITROUS_OXIDE_EMISSIONS	-0.021249	-0.005530	0.000568	-0.012460	0.020348

Source: Prepared by the researcher based on the statistical program EViews 10th Edition.

The variance-covariance matrix is a square matrix i.e., it has the same number of rows and columns. The elements of the matrix that lie along its main diagonal i.e., the one that goes from top-left to bottom-right contain the variances while all other elements contain the co-variances. Thus, the variance-covariance matrix of the fitted coefficients of a regression model contains the variances of the fitted model's coefficient estimates and the pairwise covariances between coefficient estimates.

#### 4.4. Correlation Matrix

A correlation matrix is a table showing correlation coefficients between sets of variables. Each random variable (Xi) in the table is correlated with each of the other values in the table (Xj). This allows you to see which pairs have the highest correlation [15].

**Table 5.** Correlation matrix.

	LGDP	LCO <sub>2</sub> _EMISSIONS	LMANUFACTURING	LMETHANE_EMISSIONS	LNITROUS_OXIDE_EMISSIONS
LGDP	1.000000	0.958840	0.997754	0.674685	0.817970
LCO <sub>2</sub> _EMISSIONS	0.958840	1.000000	0.953773	0.731382	0.889963
LMANUFACTURING	0.997754	0.953773	1.000000	0.675877	0.826621
LMETHANE_EMISSIONS	0.674685	0.731382	0.675877	1.000000	0.872754
LNITROUS_OXIDE_EMISSIONS	0.817970	0.889963	0.826621	0.872754	1.000000

Source: Prepared by the researcher based on the statistical program EViews 10th Edition

Table 5 shows at intersection of each cell, three values are given: Correlation, t-statistic, Prob.

The order from top to bottom (green values) and thus the degree of correlation of the independent variables X's with the dependent variable Y.

If <5% (Correlation) value P, we conclude that there is a significant association between the variable.

The function F and the rest of the independent variables, which means a successful selection of the variables.

There is an indication of a significant linear correlation between the values of the independent variables – before the application of the regression model (red values), which leads to the emergence of the problem of the linear correlation between Multicollinearity variables.

It is not preferable to have a significant association in the regression models between X's.

#### 4.5. Correlogram of Residuals

According to Table 6 the graphic representation of the functions of correlation and partial correlation, the series is not characterized by any format, initially (and graphically only) it can be said that the series is stable, where all parameters of the functions of correlation and partial correlation are within the confidence range, except for the first parameter of each statistical function [16].

**Table 6.** Correlogram of Residuals.

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
.  ***	.  ***	1	0.451	0.451	7.3469 0.007
.  *	.  .	2	0.152	-0.065	8.2029 0.017
.  .	.  .	3	0.015	-0.036	8.2113 0.042
.  *	.  *	4	-0.152	-0.171	9.1285 0.058
. **	.  *	5	-0.250	-0.139	11.703 0.039
. **	.  *	6	-0.316	-0.177	15.984 0.014
. **	.  *	7	-0.313	-0.134	20.324 0.005
. **	.  *	8	-0.295	-0.172	24.350 0.002
.  *	.  .	9	-0.149	-0.021	25.425 0.003
.  .	.  .	10	-0.007	-0.025	25.428 0.005
.  *	.  *	11	0.158	0.083	26.744 0.005
.  *	.  .	12	0.184	-0.054	28.598 0.005
. **	.  .	13	0.219	0.047	31.371 0.003
.  *	.  .	14	0.172	-0.052	33.178 0.003
.  *	.  *	15	0.162	0.082	34.870 0.003
.  .	.  *	16	0.015	-0.125	34.886 0.004

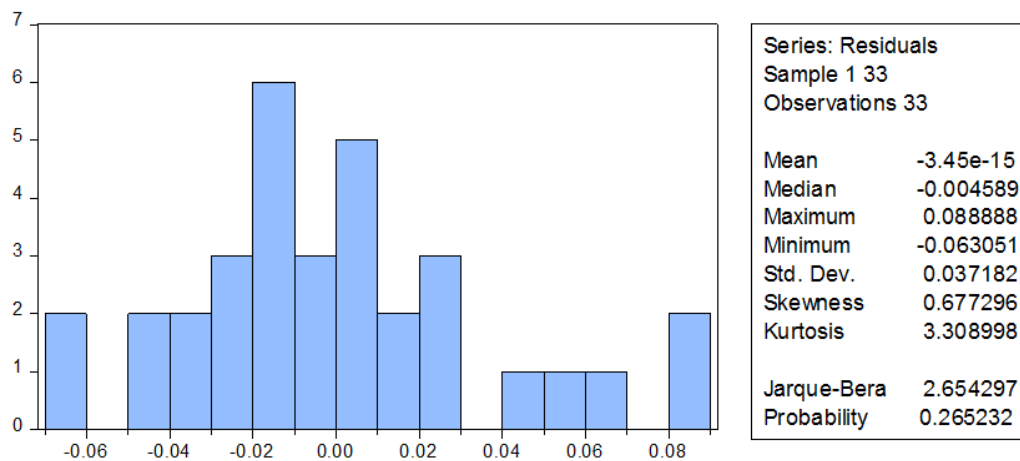
Date: 07/26/22 Time: 09:12

Sample: 1 33

Included observations: 33

#### 4.6. Normality Test for Residuals

Normality tests are used to determine if a data set is well-modeled by a normal distribution and measures a goodness of fit of a normal model to the data [17].



Source: Prepared by the researcher based on the statistical program EViews 10th Edition

**Figure 8.** Normality Test for Residuals.

From Figure 8, data of residuals are given. Normality of the above data was assessed. Result showed that data were not normally distributed as skewness (0.6772) and kurtosis (3.308) individually were within  $\pm 1$ . Jarque-Bera test ( $P = 2.654$ ) were statistically significant, that is, data were considered unnormal distributed.

Although both methods indicated that data were not normally distributed. As SD of residuals was less than half

mean value ( $0.037 < 3.000$ ), data were considered unnormally distributed.

#### 4.7. Autocorrelation Test

An identifiable relationship (positive or negative) exists between the values of the error in one period and the values of the error in another period.

**Table 7.** Serial correlation LM test.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.079918	0.585771	-0.136432	0.8925
LCO <sub>2</sub> _EMISSIONS	0.011606	0.070225	0.165269	0.8700
LMANUFACTURING	-0.013838	0.029776	-0.464722	0.6460
LMETHANE_EMISSIONS	-0.076020	0.111119	-0.684131	0.4999
LNITROUS_OXIDE_EMISSIONS	0.105422	0.130050	0.810629	0.4249
RESID (-1)	0.592287	0.204035	2.902872	0.0074
RESID (-2)	0.076091	0.245170	0.310360	0.7588
R-squared	0.278753	Mean dependent var		-3.45E-15
Adjusted R-squared	0.112312	S. D. dependent var		0.037182
S. E. of regression	0.035032	Akaike info criterion		-3.679287
Sum squared resid	0.031908	Schwarz criterion		-3.361846
Log likelihood	67.70824	Hannan-Quinn criter.		-3.572478
F-statistic	1.674782	Durbin-Watson stat		1.705451
Prob (F-statistic)	0.167141			

F-statistic: 5.024346; Prob. F (2,26): 0.0143

Obs\*R-squared: 9.198859; Prob. Chi-Square (2): 0.0101

Test Equation:

Dependent Variable: RESID

Method: Least Squares

Date: 07/26/22 Time: 09:17

Sample: 1 33

Included observations: 33

Presample missing value lagged residuals set to zero.

Source: Prepared by the researcher based on the statistical program EViews 10th Edition.

As shown in table Serial correlation LM test since Prob. Chi-Square (2) less than 5% is (0.0101) We reject  $H_0$  that there is autocorrelation. Obs\*R<sup>2</sup> = 9.198859 = NR<sup>2</sup>

$H_0: \rho = 0$

$H_1: \rho \neq 0$

P value (Coefficient) > 5% It means there is no correlation

between X 's and residuals.

#### 4.8. Heteroskedasticity and Serial Correlation

In statistics, heteroskedasticity (or heteroscedasticity) happens when the standard deviations of a predicted variable, monitored over different values of an independent variable or

as related to prior time periods, are non-constant. With the residual errors is that they will tend to fan out over time, heteroskedasticity, the tell-tale sign upon visual inspection of as depicted in the image below.

**Table 8.** Types of Heteroskedasticity.

Unconditional Heteroskedasticity	Unconditional heteroskedasticity occurs when the heteroskedasticity is uncorrelated with the values of the independent variables. Although this is a violation of the homoscedasticity assumption, it does not present major problems to statistical inference.
Conditional Heteroskedasticity	Conditional heteroskedasticity occurs when the error variance is related/conditional on the values of the independent variables. It poses significant problems for statistical inference. Fortunately, many statistical software packages can diagnose and correct this error.

**Table 9.** Heteroskedasticity Test: Harvey.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-27.44860	30.62924	-0.896157	0.3778
LCO <sub>2</sub> _EMISSIONS	-10.95344	3.543287	-3.091322	0.0045
LMANUFACTURING	5.328096	1.482235	3.594638	0.0012
LMETHANE_EMISSIONS	-1.402768	5.690064	-0.246529	0.8071
LNITROUS_OXIDE_EMISSIONS	3.894966	6.592940	0.590778	0.5594
R-squared	0.316628	Mean dependent var		-8.043523
Adjusted R-squared	0.219004	S. D. dependent var		2.078869
S. E. of regression	1.837179	Akaike info criterion		4.193067
Sum squared resid	94.50640	Schwarz criterion		4.419811
Log likelihood	-64.18561	Hannan-Quinn criter.		4.269360
F-statistic	3.243329	Durbin-Watson stat		2.457017
Prob (F-statistic)	0.026316			

F-statistic: 3.243329; Prob. F (4,28): 0.0263

Obs\*R-squared: 10.44874; Prob. Chi-Square (4): 0.0335

Scaled explained SS: 8.873285; Prob. Chi-Square (4): 0.0643

Test Equation:

Dependent Variable: LRESID2

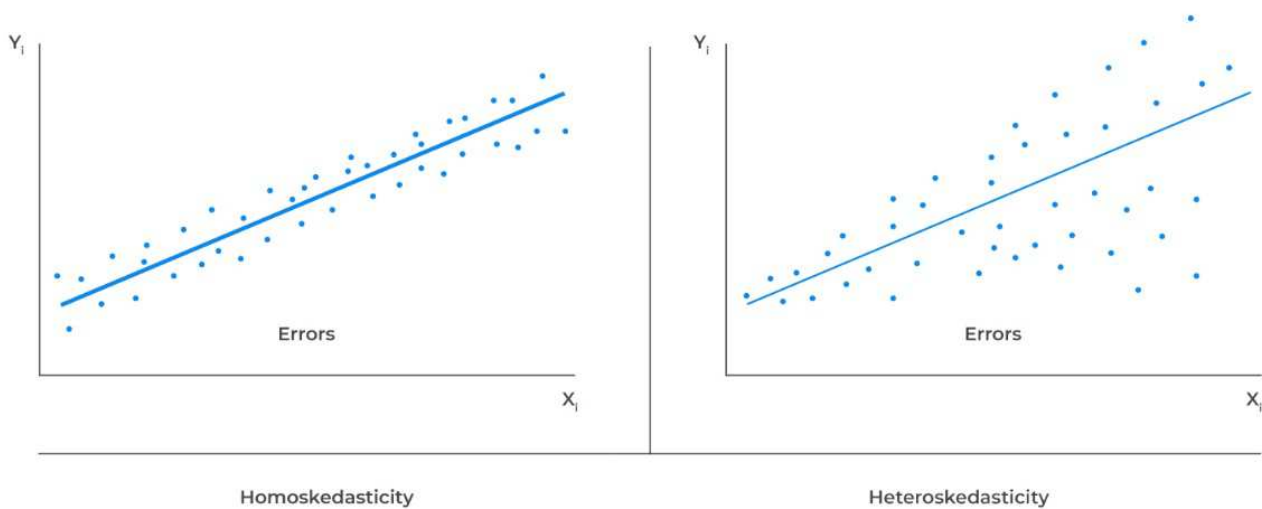
Method: Least Squares

Date: 07/26/22 Time: 09:21

Sample: 1 33

Included observations: 33

Source: Prepared by the researcher based on the statistical program EViews 10th Edition.



**Figure 9.** Heteroskedasticity and vs Homoskedasticity.

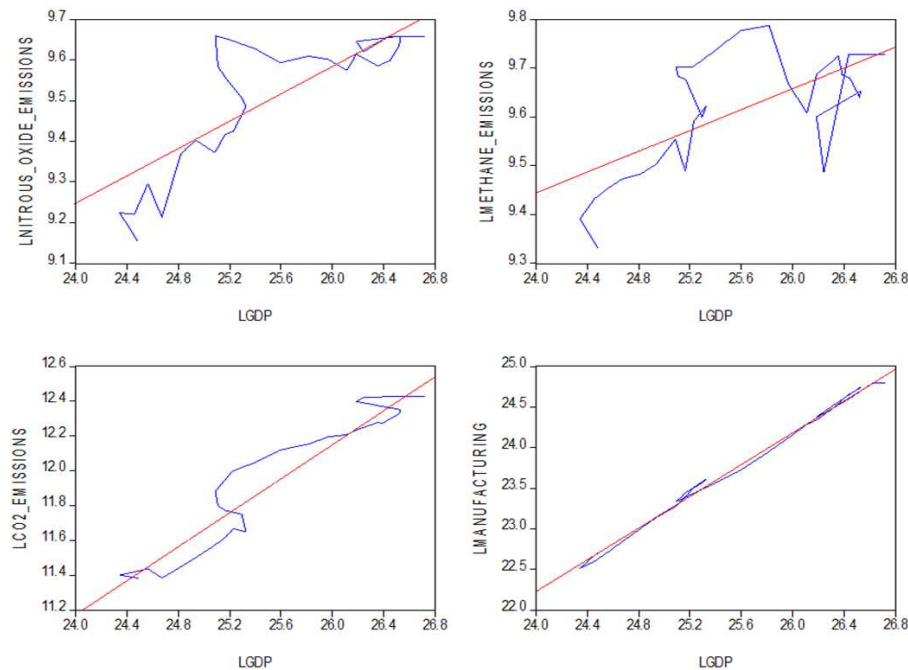
As shown in table Prob. Chi-Square (4) less than 5% is (0.0335) We reject  $H_0$  that there is Heteroskedasticity.  $Obs \cdot R^2 = 10.44874 = NR^2$

$H_0: \sigma_1^2 = \sigma_2^2 \dots \sigma_n^2 = 0$  The constant homogeneity of variances.

$H_1: \sigma_i^2 \neq 0$  at least heterogeneity of variances

## 5. Draw the Estimated Equation: Substituted Coefficients

$$LGDP = 2.59024769764 + 0.338262093151 \cdot LCO_2\_EMISSIONS + 0.920630344411 \cdot LMANUFACTURING + 0.257898931733 \cdot LMETHANE\_EMISSIONS - 0.56902863672 \cdot LNITROUS\_OXIDE\_EMISSIONS$$



Source: Prepared by the researcher based on the statistical program EViews 10th Edition

**Figure 10.** Simple regression.

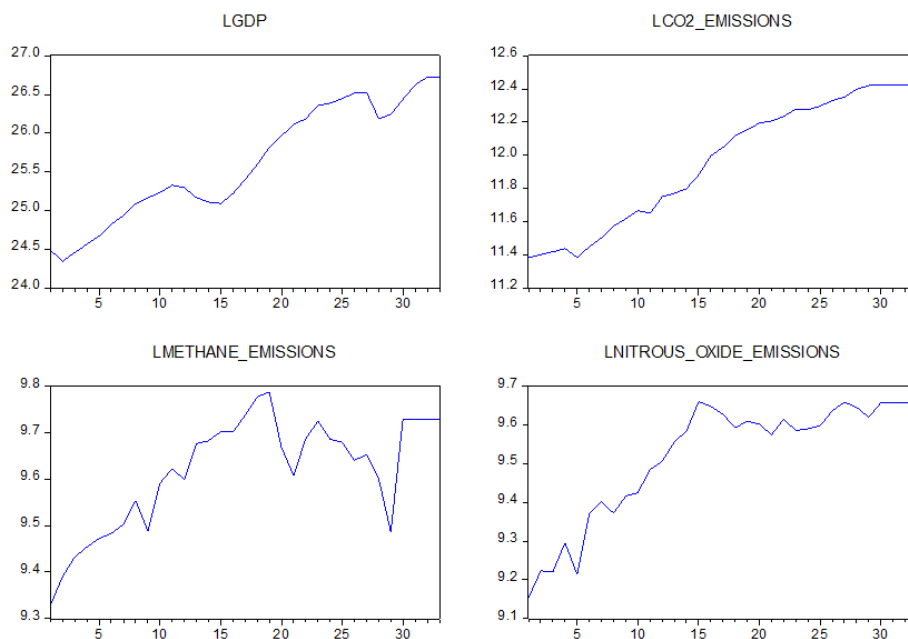
As appears in figure 10 diagram between independent variable and corresponding to the measurement of the LGDP (horizontal axis) and the LMETHANE\_EMISSIONS and LNITROUS\_OXIDE\_EMISSIONS (vertical axis). shows increasing positive relation among variables.

## 6. Trends of the Variables over Time

To avoid the problem of different scales for each variable, the log is chosen. The values of the variable are not affected

when the natural logarithm is taken for it. This method is used when the variable values are large, and it is intended to simplify them to reduce dispersion and variance between other variables. The log is characterized by not changing the shape of the distribution but changing the shape of the scale.

The figure 11 shows the variables over time. It is clear from the figure that the dependent and independent variables are in a continuous direction through time .and that there is an increasing direct relationship through time between the dependent and independent variable.



Source: Prepared by the researcher based on the statistical program EViews 10th Edition

**Figure 11.** logarithm of dependent and independent variables.

### 6.1. Actual and Estimated Residuals

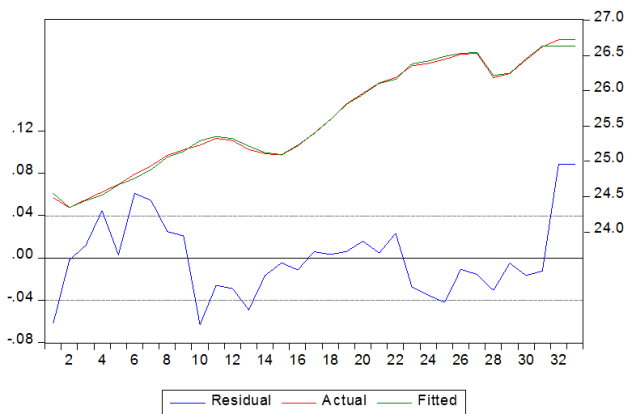


Figure 12. Actual and Estimated Residuals.

Source: Prepared by the researcher based on the statistical program EViews 10th Edition.

The curve in red represents the "true" actual values of the time series of the dependent variable in the figure 12.

The curve in green represents the estimated values of the dependent variable according to the estimated equation.

The blue curve represents the residuals of the regression equation "random error term perturbation"

From the Figure 12, we notice that there are "extreme" abnormal values in the random error term, and I think that the estimated model suffers from a disruption of variance.

There is no essential difference between the actual values, which are the values before including the explanatory variables, and the estimated values, which are the values after including the explanatory variables.

From the figure, we notice an almost similarity. As for the curve in blue, it indicates the behavior of residuals, which can theoretically be divided into three sections.

Where it is noticed that there is relative stability of the behavior of the residuals in the middle region of the chain, which in turn explained the absence of the imbalance between the actual values and the estimated values.

As for the sides of the series parties, we notice a fluctuation, which in turn affects the quality of the overall model. It is believed that the problem should be treated with the effect of outliers, as you can detect them through the mahaleb's test.

### 6.2. Logarithm of LGDP Residuals

To avoid the problem of different scales for each variable, the log is chosen. The values of the variable are not affected when the natural logarithm is taken for it. This method is used when the variable values are large, and it is intended to simplify them to reduce dispersion and variance between other variables. The log is characterized by not changing the shape of the distribution but changing the shape of the scale.

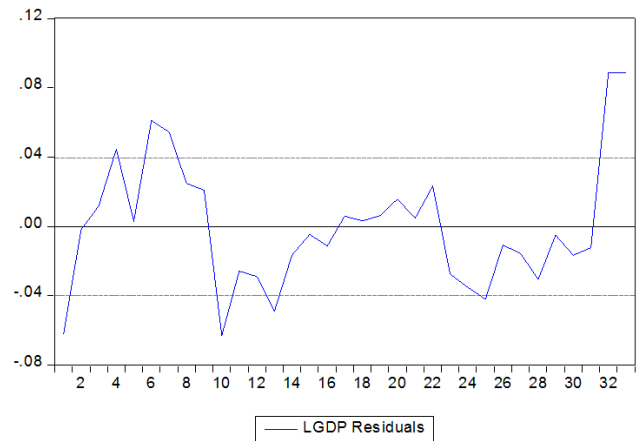


Figure 13. LGDP Residuals.

Source: Prepared by the researcher based on the statistical program EViews 10th Edition

The figure 13 shows the residual time series of the logarithm variable of GDP, which means that the residual series is unstable and often there is no co-integration.

## 7. Conclusion

From the above we conclude the following:

1. The model suffers from the problem of heterogeneity of variance, and this leads to that the predictions in the variable Y depending on the estimators  $B^{'}$ 's (the coefficients of the independent variables) from the original data will have large variances, and this means that the prediction will be inefficient and the reason for this is that the variance The predictions will include the U variance as well as the parameters variance.
2. The model suffers from a problem of autocorrelation, which means that  $Cov(u_j, u_i) \neq 0$ , and therefore the standard errors  $\sigma^2$  are rather large, which means that the accuracy in the model is low and therefore the confidence intervals and the model's significance will be unacceptable and unreliable in and inefficient.
3. The model suffers from the problem of linear interference. This means that the estimators' values are very large and biased, as well as the variances of these estimators and the covariances are very large, so the properties of estimators are not BLUE.

## Acknowledgements

I thank my father and mother for bearing hardships and stress on me during the exams, and for their service in difficult times, and for providing me the effort and time to allow me the opportunity to work hard and diligently day and night. I hope that God accepts their work and makes them worship the righteous and places them in the Paradise of Paradise, in the company of the prophets and friends.

I thank Dr. Hossam Elden M. Abdelkader, Associate

professor Economic Dep., Faculty of Administrative Sciences, Ain Shams University, Egypt .I thank him for teaching the monetary policy course, benefiting from his knowledge, and

teaching us advanced econometrics lectures in the preparatory year for the doctorate program, He is a distinguished young doctor, genius and a role model for the youth.

## Appendix of Study

*Table 10. Data of the Study Variables.*

Country	year	GDP	methane emissions	nitrous oxide emissions	CO <sub>2</sub> emissions	GDP growth (annual%)	Manufacturing	FDI
Egypt	1990	4.3E+10	11270	9450	87750	2.900791	7.3E+09	1.058425
Egypt	1991	3.74E+10	11980	10130	89370	3.973172	5.99E+09	2.420133
Egypt	1992	4.19E+10	12490	10100	90900	4.642459	6.54E+09	0.994028
Egypt	1993	4.66E+10	12760	10890	92660	4.988731	7.33E+09	0.940415
Egypt	1994	5.19E+10	13000	10030	87900	5.492355	8.31E+09	1.135376
Egypt	1995	6.02E+10	13140	11730	93720	5.575497	9.83E+09	1.268437
Egypt	1996	6.76E+10	13400	12120	98940	6.053439	1.12E+10	1.174393
Egypt	1997	7.84E+10	14100	11760	106060	6.370004	1.28E+10	1.236997
Egypt	1998	8.48E+10	13210	12290	110980	3.535252	1.44E+10	0.527385
Egypt	1999	9.07E+10	14640	12400	116540	2.390204	1.63E+10	0.759753
Egypt	2000	9.98E+10	15100	13170	114610	3.193455	1.8E+10	0.295684
Egypt	2001	9.67E+10	14760	13450	126700	4.092072	1.71E+10	1.590836
Egypt	2002	8.51E+10	15940	14140	129440	4.471744	1.53E+10	5.999509
Egypt	2003	8.03E+10	16040	14540	133020	6.843838	1.39E+10	9.348567
Egypt	2004	7.88E+10	16360	15680	144500	7.087827	1.36E+10	8.876336
Egypt	2005	8.96E+10	16350	15500	162220	7.156284	1.5E+10	5.831413
Egypt	2006	1.07E+11	16940	15180	170750	4.6736	1.72E+10	3.548351
Egypt	2007	1.3E+11	17630	14670	183400	5.147235	2E+10	2.916017
Egypt	2008	1.63E+11	17820	14910	189940	1.764572	2.53E+10	0.204543
Egypt	2009	1.89E+11	15810	14800	197660	2.2262	2.99E+10	1.002341
Egypt	2010	2.19E+11	14880	14400	200310	2.185466	3.53E+10	1.453434
Egypt	2011	2.36E+11	16110	14990	205770	2.915912	3.72E+10	1.50925
Egypt	2012	2.79E+11	16730	14560	215000	4.372019	4.51E+10	2.102581
Egypt	2013	2.88E+11	16090	14630	213860	4.346643	4.79E+10	2.438563
Egypt	2014	3.06E+11	15990	14740	219120	4.181221	5.13E+10	3.142826
Egypt	2015	3.29E+11	15380	15320	226280	5.314121	5.5E+10	3.260263
Egypt	2016	3.32E+11	15570	15670	231230	5.557684	5.6E+10	2.972837
Egypt	2017	2.36E+11	14770	15460	242230	3.569669	3.88E+10	1.602124
Egypt	2018	2.5E+11	13180	15070	247910	3.326742	4.04E+10	1.602124
Egypt	2019	3.03E+11	16800	15650	249370	3.326742	4.82E+10	1.602124
Egypt	2020	3.65E+11	16800	15650	249370	3.326742	5.88E+10	1.602124
Egypt	2021	4.04E+11	16800	15650	249370	3.326742	5.88E+10	1.602124
Egypt	2022	4.04E+11	16800	15650	249370	3.326742	5.88E+10	1.602124

Source: Data collected by researcher from world bank.

## References

- [1] Alin, A. (2010). Multicollinearity. Wiley interdisciplinary reviews: computational statistics, 2 (3), 370-374.
- [2] Mansfield, E. R., & Helms, B. P. (1982). Detecting multicollinearity. The American Statistician, 36 (3a), 158-160.
- [3] Daoud, J. I. (2017, December). Multicollinearity and regression analysis. In Journal of Physics: Conference Series (Vol. 949, No. 1, p. 012009). IOP Publishing.
- [4] Farrar, D. E., & Glauber, R. R. (1967). Multicollinearity in regression analysis: the problem revisited. The Review of Economic and Statistics, 92-107.
- [5] Kim, J. H. (2019). Multicollinearity and misleading statistical results. Korean journal of anesthesiology, 72 (6), 558-569.
- [6] Gunst, R. F., & Webster, J. T. (1975). Regression analysis and problems of multicollinearity. Communications in Statistics-Theory and Methods, 4 (3), 277-292.
- [7] Anderson, R. L. (1954). The problem of autocorrelation in regression analysis. Journal of the American Statistical Association, 49 (265), 113-129.
- [8] Kadiyala, K. R. (1968). A transformation used to circumvent the problem of autocorrelation. Econometrica: Journal of the Econometric Society, 93-96.
- [9] Griffith, D. A., Fischer, M. M., & LeSage, J. (2017). The spatial autocorrelation problem in spatial interaction modelling: a comparison of two common solutions. Letters in Spatial and Resource Sciences, 10 (1), 75-86.
- [10] Stimson, R. J., Mitchell, W., Rohde, D., & Shyy, P. (2011). Using functional economic regions to model endogenous regional performance in Australia: Implications for addressing the spatial autocorrelation problem. Regional Science Policy & Practice, 3 (3), 131-144.
- [11] Shrestha, N. (2020). Detecting multicollinearity in regression analysis. American Journal of Applied Mathematics and Statistics, 8 (2), 39-42.

- [12] Obite, C. P., Olewuezi, N. P., Ugwuanyim, G. U., & Bartholomew, D. C. (2020). Multicollinearity effect in regression analysis: A feed forward artificial neural network approach. *Asian journal of probability and statistics*, 6 (1), 22-33.
- [13] Zhang, T., Zhou, X. P., & Liu, X. F. (2020). Reliability analysis of slopes using the improved stochastic response surface methods with multicollinearity. *Engineering Geology*, 271, 105617.
- [14] Vörösmarty, G., & Dobos, I. (2020, October). Green purchasing frameworks considering firm size: a multicollinearity analysis using variance inflation factor. In *Supply Chain Forum: An International Journal* (Vol. 21, No. 4, pp. 290-301). Taylor & Francis.
- [15] Aslam, M., & Ahmad, S. (2020). The modified Liu-ridge-type estimator: a new class of biased estimators to address multicollinearity. *Communications in Statistics-Simulation and Computation*, 1-20.
- [16] Negret, P. J., Marco, M. D., Sonter, L. J., Rhodes, J., Possingham, H. P., & Maron, M. (2020). Effects of spatial autocorrelation and sampling design on estimates of protected area effectiveness. *Conservation Biology*, 34 (6), 1452-1462.
- [17] Stojkoski, V., Sandev, T., Kocarev, L., & Pal, A. (2022). Autocorrelation functions and ergodicity in diffusion with stochastic resetting. *Journal of Physics A: Mathematical and Theoretical*, 55 (10), 104003.