EUROPEAN JOURNAL OF PURE AND APPLIED MATHEMATICS
Vol. 16, No. 1, 2023, 657-669
ISSN 1307-5543 – ejpam.com
Published by New York Business Global

Investigating the Association Between Carbon Dioxide (CO_2) Emissions and Energy Consumption on the Economic Growth of Clusterized Countries in the World

Laksmi Prita Wardhani[1], Alvida Mustika Rukmi[1,*], Thaliah Fauz Ardamayanti[1]

[1] *Mathematics, Faculty of Science and Data Analytics, Sepuluh Nopember Institute of Technology, Surabaya, East Java, Indonesia*

Abstract. This research aims to investigate the relationship between carbon dioxide (CO_2) emissions and energy consumption on the economic growth of countries in the world based on Gross Domestic Product (GDP). We use a large volume of data related to CO_2 emissions per capita and energy consumption per capita of each country in the world, so a clustering process is needed using the K-Means method, which divides the country into three cluster labels: high, medium, and low. Then, we performed multiple linear regression using the Generalized Least Square with Autoregressive (GLSAR) method for data that did not satisfy the Best Linear Unbiased Estimator (BLUE) criteria. The results are that countries included in the high cluster have a closer relationship between CO_2 emissions and energy consumption to GDP than the medium cluster. Likewise, for the medium cluster against the low cluster.

2020 Mathematics Subject Classifications: 62H30, 62J05, 62M10

Key Words and Phrases: Carbon dioxide (CO_2) emissions, Economic growth, Energy consumption, Generalized Least Square, K-Means

---

1. Introduction

Global warming is increasing the average temperature of the Earth's air, atmosphere, sea, and land, which is one of the crucial problems facing the world today. The contributing factors of this problem include carbon emissions, energy consumption, and the increasing number of population countries in the world [18]. Carbon emissions are gases released from the combustion of carbon-containing compounds, which are dominated by carbon dioxide (CO_2) gases [15]. While energy consumption is the use of power or energy in a system by utilizing a certain supply [19]. The increasing of carbon emissions and energy consumption is in line with the world's population, so it needs particular attention to maintain environmental stability.

*Corresponding author.
DOI: https://doi.org/10.29020/nybg.ejpam.v16i1.4614

Email addresses: laksmi_pw@matematika.its.ac.id (L. P. Wardhani), alvida@matematika.its.ac.id (A. M. Rukmi), thaliahfauz@gmail.com (T. F. Ardamayanti)

On the other hand, economic growth continues to be one of the most relevant and exciting sub-sectors of the economy [2]. Economic growth is an increase in the market value of goods and services adjusted for inflation over a certain period [10]. One of the indicators to measure the economic growth of a country is the Gross Domestic Product (GDP), which is the number of products in the form of goods and services produced by production units within the borders of a country (domestic) for one year [7]. The association between global warming and economic growth is one of the global challenges, especially to achieve net zero emissions, which countries have agreed upon in the Paris Agreement [9]. Countries in the world must try to optimize the use of global energy and reduce carbon dioxide (CO_2) emissions by implementing specific policies to improve their economy [14].

Certain countries, such as China, Poland, India, Philippines, Thailand, and Indonesia, have rapid economic and population growth, especially in the industrial sector, with high energy dependence. Their energy consumption and emissions have increased pollution significantly in recent decades [16]. Several studies have examined the relationship between CO_2 emissions and energy consumption on economic growth, using models based on econometric theory, panel data analysis, and statistical approaches. In general, the empirical results of the model show that there is a long-term causal relationship between these variables, which provides valuable information in terms of the implications of policies aimed at reducing or saving energy consumption to help reduce CO_2 emission [11][4][3][22][15][13]. However, these studies only applied in certain regions.

To answer these limitations, in this paper, we conduct a study to examine the relationship between CO_2 emissions and energy consumption on the economic growth of countries worldwide. Then, to handle a large amount of data to be more manageable, we grouped these countries into three groups, namely countries with high, medium, and low CO_2 emissions and energy consumption. This grouping aims to determine the country's characteristics in each group and to facilitate the analysis process. This similar grouping process has been carried out for 24 countries in Asia and obtained 6 clusters based on greenhouse gas emissions from 1990 to 2013 using the hierarchical clustering analysis method. The method gives good clustering results for those 24 countries [11]. However, there is another more straightforward clustering method for large amounts of data called the K-Means method [8].

On the other hand, no related studies have investigated the association between CO_2 emissions, energy consumption, and economic growth of the clusterized countries in the world. This paper contributes to this analysis. We use one of the machine learning models, namely multiple linear regression, which applies to each country that has been clusterized using K-Means. The aim is to investigate the simultaneous correlation between the observed variables. So that those countries can consider the right policies to increase economic growth by optimizing CO_2 emissions and energy consumption. Hence, this paper is organized into the following sections: Section 1 examines the background and problems solved in this paper, Section 2 discusses the literature review used, Section 3 shows the results of the research conducted, and Section 4 discusses the conclusion.

2. Clustering and Regression

2.1. K-Means Clustering

Clustering is a method of grouping data into several clusters or groups, so that data in one cluster has maximum similarity and data between clusters has minimal similarity. K-Means is one of the most widely used clustering methods. The similarity of characteristics in this method is based on the average (*mean*) of the object's distance.

The distance between objects is obtained by using the Euclidean Distance with Equation (1),

$$d(a,c) = \sqrt{\sum_{i=1}^{n} (a_i - c_i)^2} \tag{1}$$

where $d(a,c)$ is the Euclidean distance between a and c, a_i is the point present, c_i is the centroid, and n is the number of data.

The silhouette coefficient or silhouette score is a metric used to calculate the goodness of a clustering technique. The silhouette coefficient is denoted by the value $s(i)$ starting from $-1 \le s(i) \le 1$ which is calculated by Equation (2),

$$s(i) = \begin{cases} \frac{g(i)-f(i)}{\max(f(i),g(i))} & ; C_i > 1, \\ 0 & ; C_i = 1, \end{cases} \tag{2}$$

or it can be written as follows

$$s(i) = \begin{cases} 1 - \frac{f(i)}{g(i)} & ; f(i) < g(i), \\ 0 & ; f(i) = g(i), \\ \frac{g(i)}{f(i)} - 1 & ; f(i) > g(i) \end{cases} \tag{3}$$

where $s(i)$ is the silhouette coefficient, C_i is the number of cluster members i, $f(i)$ is the average distance between other objects in one cluster, and $g(i)$ is the minimum average distance to objects in other clusters. The silhouette coefficient close to 1 indicates that the grouping formed is improving. Conversely, if the silhouette coefficient is close to -1, then data grouping in one cluster worsens.

Before we do the clustering, we can normalize data with different ranges using the min-max scaler. Min-max scalers usually make it possible to transform data with varying scales so that no particular dimension dominates the statistics, and there is no need to make strong assumptions about the data distribution. The min-max scaler formula is as follows,

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{4}$$

where x_{scaled} is the normalized data, x is initial data, x_{min} is the minimum value of the attribute where data x resides, and x_{max} is the maximum attribute value.

2.2. Multiple Linear Regression

A multiple linear regression model describes the linear relationship between two or more independent variables $(X_1, X_2, X_3, ..., X_n)$ and one dependent variable (Y). Multiple linear regression analysis uses to predict the value of the dependent variable (Y) if the values of the independent variables are known and also to determine the direction of the relationship between the dependent variable and the independent variables. In this paper, we use two independent variables: X_1 is CO_2 emissions per capita, X_2 is energy consumption per capita, and the dependent variable Y is Gross Domestic Product (GDP) per capita. The two-variable multiple linear regression equation expresses as (5),

$$Y = b_0 + b_1X_1 + b_2X_2 + \varepsilon \tag{5}$$

where ε is a random residual, b_0, b_1, b_2 are parameters whose unknown values. The independent variables X_1 and X_2 are fixed and observed with negligible error. If we have a sample of size n, then the random residuals $\varepsilon_1, \varepsilon_2, ..., \varepsilon_n$ will be assumed to have a $mean = 0$, constant variance σ^2, independent or uncorrelated, and normally distributed. Furthermore, parameters $b_0, b_1,$ and b_2 in (5) are estimated using the Ordinary Least Square (OLS) with Equation (6),

$$\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \tag{6}$$

where $\mathbf{Y}$ is dependent variable matrix, $\mathbf{X}$ is the independent variable augmented matrix, and $\hat{\mathbf{b}}$ is the parameters estimator matrix for n samples as follows,

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x2p \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix}, \hat{\mathbf{b}} = \begin{bmatrix} \hat{b_1} \\ \hat{b_2} \\ \vdots \\ \hat{b_n} \end{bmatrix}.$$

After obtaining the regression estimator, we need to do the classic assumption test to examine whether the regression model satisfies the Best Linear Unbiased Estimator (BLUE) criteria. Regression models that satisfy the BLUE criteria can use as reliable estimators where the estimator is declared unbiased, consistent, normally distributed, and efficient. The classical assumption tests used to determine whether the regression model has satisfied the BLUE criteria consist of normality, multicollinearity, heteroscedasticity, and autocorrelation tests.

The normality test aims to examine whether the residuals obtained from the regression results were normally distributed or not. In this paper, we use the Jarque Bera (JB) test to examine the normality of the residuals with Equation (7),

$$JB = \frac{n}{6}\left(S^2 + \frac{(K-3)^2}{4}\right) \tag{7}$$

where

$$K = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4}{(\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2)^2} \tag{8}$$

$$S = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3}{(\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2)^{\frac{3}{2}}} \tag{9}$$

JB is Jarque Bera's coefficient value, n is the number of data samples, S is the expected skewness value, and K is the expected value of kurtosis. The null hypothesis declares that the residuals are normally distributed, and the critical statistical region to reject the hypothesis if $JB < \chi^2_{(2,\alpha)}$ with α is the significance level [23].

The multicollinearity test determines whether each independent variable is linearly correlated. Non-multicollinearity is one of the conditions that must be satisfied by multiple linear regression models. Multicollinearity can be detected by calculating the value of the Variance Inflation Factor (VIF) of each independent variable [6] with Equation (10),

$$VIF = \frac{1}{1 - R_j^2} \tag{10}$$

where R_j^2 is the model's coefficient of determination where one of the independent variables is used as the dependent variable on the other independent variables. If the $VIF \leq 10$, it can be concluded that there is no multicollinearity for that independent variable.

The heteroscedasticity test aims to test whether the regression model has a definite and constant residual variance. Heteroscedasticity occurs if the residual variance is not constant or different. Heteroscedasticity can be detected through the Glejser [12] test. Meanwhile, the autocorrelation test examines whether there is a correlation between the residuals in the current and previous observations. Autocorrelation often occurs in time series data. Autocorrelation can be detected by Durbin-Watson (DW) test [1] with Equation (11).

$$DW = \frac{\sum_{t=2}^{n}(\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^{n}\varepsilon_t^2} \tag{11}$$

where DW is the Durbin Watson coefficient, ε_t is the residual at the t^{th} observation, while n is the number of observed data. Decision-making is based on the following:

- If $0 < DW < dl$ or $4 - dl < DW < 4$, then it can be concluded that autocorrelation occurs.
- If $du < DW < 4 - du$, then it can be concluded that there is no autocorrelation.
- If $dl < DW < du$ or $4 - du < DW < 4 - dl$, then it can be concluded that there is no decision.

with the values of du and dl taken from the Durbin-Watson table.

The linear regression model with OLS must have no autocorrelation and heteroscedasticity. If these assumptions are not satisfied, then the OLS method is no longer appropriate to estimate the parameters of the multiple linear regression model, so that we can use the Generalized Least Square (GLS) method.

The first step in the GLS method is transforming the autocorrelation and heteroscedasticity data. Let the residual covariance matrix is of the form $\sigma^2\mathbf{V}$ where $\mathbf{V}$ is a non-singular matrix and positive definite so that there is a non-singular symmetric matrix $\mathbf{M}$ of size $n \times n$ with $\mathbf{M}'\mathbf{M} = \mathbf{M}\mathbf{M} = \mathbf{V}$. So that the estimator for the GLS is obtained as follows, [5]

$$\hat{\mathbf{b}}_{\mathbf{GLS}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y} \tag{12}$$

Estimation using the GLS method can be done by adding weights $\mathbf{V}$ to the parameters that have heteroscedasticity problems [21]. The weight used in this paper is a matrix $\mathbf{V}$ that follows the AR(1) model. The positive definite matrix $\mathbf{V}$ follows the first-order autoregressive or AR(1) model as in Equation (12). The GLS with AR(1) model is called the GLSAR method. We can use the *statsmodels* package in Python to determine the regression estimators using the GLSAR method.

$$\mathbf{V} = \frac{1}{1-\rho^2}\begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \cdots & 1 \end{bmatrix} \tag{13}$$

On the other hand, data transformation is also an effort to modify the measurement scale of the original data so that the data can meet the regression assumptions. Data transformation can fix the assumptions of normality, linearity, and homoscedasticity that must be satisfied to get the BLUE estimator [5]. In this paper, the transformation used is the natural logarithm transformation and the root transformation.

3. The Results

We use secondary data from the site Our World in Data regarding the data per capita on CO_2 emissions (tonnes), energy consumption (kilowatt per hour), and GDP (USD 2017) of 183 countries in the world for 1990–2020 [20].

3.1. Clustering of the Countries

The clustering process simplifies finding out the relationship between the use of CO_2 emissions, energy consumption, and GDP from 183 countries so we can efficiently conduct the association on each cluster. We use the K-Means method in Python and Google Colab to cluster the countries based on CO_2 emissions per capita and energy consumption per capita into three clusters, namely high (as cluster 1), medium (as cluster 2), and low (as a cluster 0). However, we first normalized the data using the min-max scaler function. This clustering results in the silhouette coefficient of 0.601, which is the highest of the other number of clusters. Figure 1 shows 121 countries included in cluster 0, marked in red; 13 countries in cluster 1, marked in green; and 49 countries in cluster 2, marked in blue. The distribution of countries by cluster is shown in Figure 2.

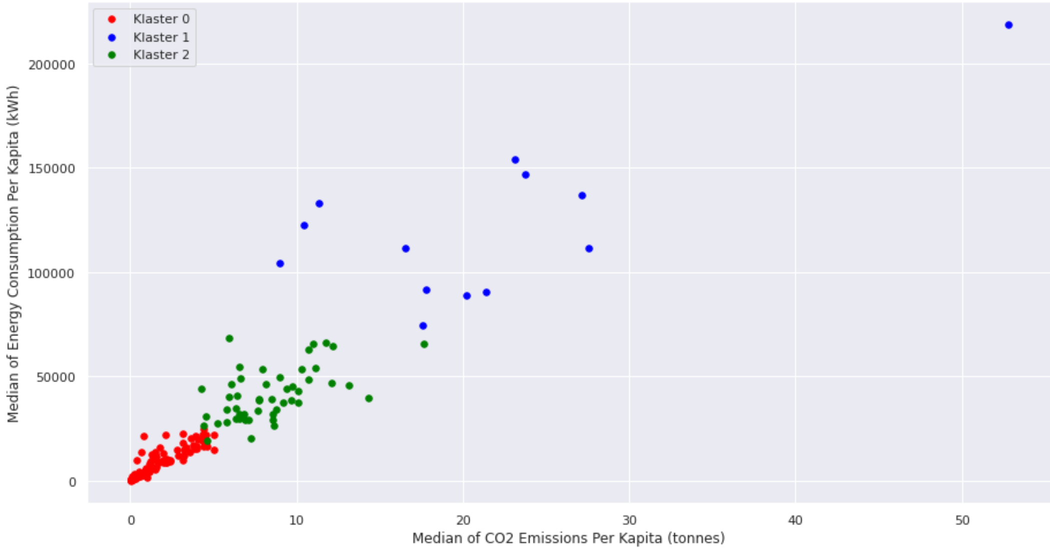

Figure 1: Clustering Result

3.2. The Regression Analysis

We process multiple linear regression analysis with Python. First, we estimated the regression parameters using the OLS method. Then we perform classical assumption tests: multicollinearity, heteroscedasticity, and autocorrelation tests using the methods previously described. Some examples of calculation results for several countries are presented in Table 1.

Table 1: OLS Results and Classical Assumption Tests for Singapore, Russia, and Gambia.

Entity		Singapore	Russia	Gambia
Cluster		High	Medium	Low
Estimator	b_0	-5,856.900	-2,064.000	2,539.511
	b_1	-844.643	-4,117.259	-2,472.734
	b_2	0.623	1.617	0.190
R^2		0.974	0.358	0.120
R^2_{Adj}		0.972	0.312	0.058
F-test		518.600	7.815	1.916
VIF		1.982	7.806	4.632
p-value	b_0	0.315	0.006	0.954
	b_1	0.262	0.003	0.727
	b_2	0.785	0.001	0.990
Durbin Watson		1.378	0.340	1.105
Conclusion		1. The model is significant. 2. No multicollinearity. 3. Homoskedasticity 4. Autocorrelation	1. The model is significant. 2. No multicollinearity. 3. Heteroskedasticity 4. Autocorrelation	1. The model is not significant. 2. No multicollinearity. 3. Homoskedasticity 4. Autocorrelation

Apart from these three countries, we find that all country data contain multicollinearity, heteroscedasticity, or autocorrelation problems. Therefore, we use the GLS method, specifically GLSAR, i.e., without and with transformation logarithmic and root, to solve

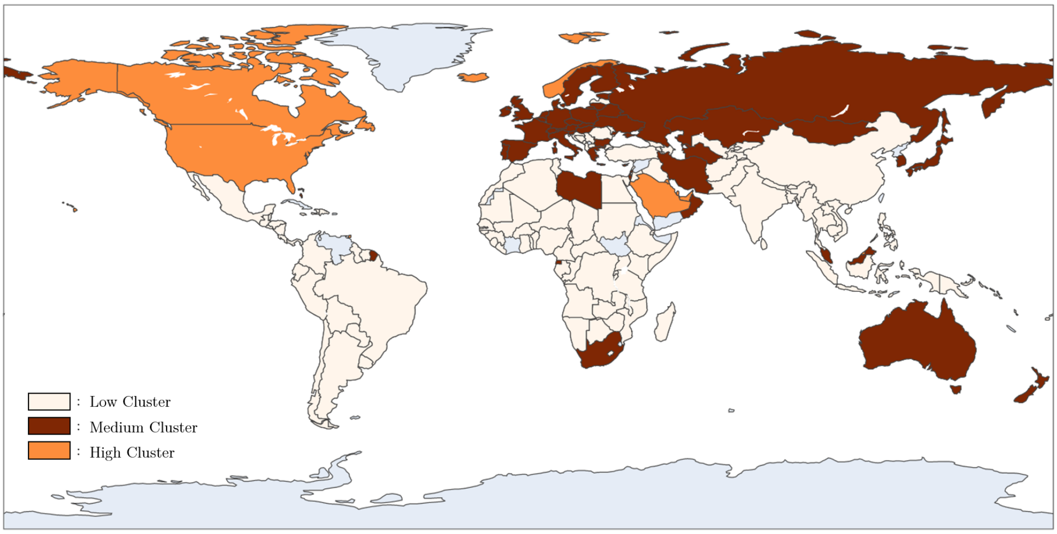

Figure 2: Distribution of Countries in the World Based on Their Clusters.

those problems. The methods result in the BLUE estimator for 45 of 183 countries: 3 countries in the high cluster, ten countries in the medium cluster, and 32 countries in the low cluster, shown in Table 2.

Table 2: Multiple Linear Regression Final Result

No	Country	Cluster	Transformation	r_{X_1Y}	r_{X_2Y}	$r_{X_1X_2}$	R^2_{Adj}
1	Qatar	High	Logarithm	-0,285	0,111	0,581	0,137
2	Singapore	High	Root	-0,649	0,968	-0,584	0,943
3	Trinidad and Tobago	High	No	0,736	0,789	0,809	0,624
4	Belgium	Medium	Logarithm	0,602	0,664	0,874	0,402
5	France	Medium	Logarithm	0,546	0,596	0,874	0,311
6	Hong Kong	Medium	Root	0,072	0,366	0,411	0,078
7	Hungary	Medium	No	-0,603	0,037	0,677	0,709
8	Italy	Medium	Logarithm	0,801	0,822	0,945	0,656
9	Libya	Medium	Logarithm	0,238	0,581	0,323	0,292
10	Malta	Medium	No	-0,203	0,757	0,047	0,603
11	Russia	Medium	No	0,817	0,787	0,689	0,745
12	Seychelles	Medium	No	0,392	-0,146	0,449	0,23
13	Ukraine	Medium	No	0,879	0,792	0,806	0,772
14	Argentina	Low	No	0,729	0,871	0,786	0,747
15	Armenia	Low	No	0,567	0,122	0,213	0,271
16	Bangladesh	Low	Logarithm	0,911	0,891	0,932	0,817
17	Bosnia and Herzegovina	Low	No	0,917	0,527	0,668	0,844
18	Botswana	Low	Root	0,366	0,066	0,055	0,067
19	Burkina Faso	Low	No	0,547	0,246	0,677	0,277

Table 2 : Multiple Linear Regression Final Result (continued)

No	Country	Cluster	Transformation	r_{X_1Y}	r_{X_2Y}	$r_{X_1X_2}$	R^2_{Adj}
20	Chile	Low	Root	0,479	0,575	0,528	0,326
21	Colombia	Low	No	0,363	0,71	0,443	0,47
22	Comoros	Low	Logarithm	0,431	0,434	0,874	0,185
23	Dominica	Low	No	0,842	0,877	0,944	0,754
24	Ecuador	Low	No	0,137	0,745	-0,126	0,577
25	Gabon	Low	Logarithm	0,381	-0,046	0,031	0,086
26	Gambia	Low	No	-0,054	-0,033	0,666	-0,071
27	Grenada	Low	No	0,222	0,262	0,654	0,004
28	Haiti	Low	Logarithm	0,758	0,667	0,722	0,576
29	Honduras	Low	No	0,506	0,656	0,525	0,427
30	Kiribati	Low	Logarithm	0,025	0,303	0,648	0,078
31	Kyrgyzstan	Low	No	0,693	-0,162	-0,105	0,45
32	Lesotho	Low	No	0,153	0,541	0,001	0,264
33	Madagascar	Low	Logarithm	0,475	0,623	0,559	0,368
34	Mali	Low	Logarithm	0,869	0,746	0,813	0,742
35	Nepal	Low	No	0,598	0,314	0,310	0,333
36	Niger	Low	Root	0,266	0,427	0,706	0,125
37	Paraguay	Low	No	0,538	0,159	0,129	0,245
38	Peru	Low	No	0,295	0,899	0,421	0,803
39	Saint Kitts Nevis	Low	No	0,360	0,444	0,548	0,158
40	Saint Vincent Grenadines	Low	Logarithm	0,188	0,018	0,468	-0,029
41	Samoa	Low	Logarithm	0,403	0,495	0,179	0,299
42	Sierra Leone	Low	No	0,695	0,303	0,247	0,465
43	Timor	Low	No	0,403	0,469	0,701	0,173
44	Tunisia	Low	No	0,911	0,652	0,652	0,96
45	Uganda	Low	Logarithm	0,457	-0,008	0,487	0,224

The partial correlation between CO_2 emissions and GDP shows that only five countries have a negative correlation, namely Singapore, Qatar, Hungary, Malta, and Gambia, as shown in Table 1. It means that when CO_2 emissions increase, GDP decreases, and vice versa. Singapore is one of the countries that succeeded in reducing CO_2 emissions and increasing GDP, as shown in Table 1 with a partial correlation r_{X_1Y} of -0.649 because it has implemented a carbon tax policy since 2019 [17]. This policy is an administrative fee with an ideal taxation system that covers all activities that generate gas CO_2 emissions in the agriculture, forestry, and industrial sectors to support carbon emission reductions which are currently a global problem. Other countries that have also succeeded in increasing economic growth and reducing CO_2 emissions are Malta and Gambia.

As for Qatar, CO_2 emissions tend to increase while its GDP decreases. According to the World Bank, Qatar is the largest emitter of greenhouse gases per capita, nearly three times more than the United States and nearly six times more than China. This Middle Eastern country uses about 60% of its electricity to cool the air. However, electricity in

Qatar comes from fossil fuels which release large amounts of CO_2 into the atmosphere and cause a climate emergency so that CO_2 emissions in Qatar tend to increase in the year of observation.

From the partial correlation between energy consumption and GDP, five countries have a non-unidirectional correlation, namely Gabon, Gambia, Kyrgyzstan, and Uganda from the low cluster and Seychelles from the medium cluster. Gabon, Gambia, Israel, Lithuania, and Kyrgyzstan are experiencing an increase in energy consumption and a decrease in GDP. Meanwhile, the countries whose energy consumption decreased and GDP increased are Seychelles and Uganda, although very weakly correlated. Meanwhile, Gabon, Gambia, and Kyrgyzstan experienced an increase in energy consumption followed by a decrease in their GDP. Since the correlation is very weak, it interprets that no meaningful policies can be used as examples from these countries.

As shown in Table 2, 36 other countries produce partial correlations that are positive or in the direction of GDP, meaning that if CO_2 emissions or energy consumption in that country increases, it will also follow by an increase in GDP. Vice versa, if CO_2 emissions or energy consumption decreases, it will be followed by a decrease in GDP.

Table 3: Statistics Calculation for Each Cluster.

Statistic	High Cluster	Medium Cluster	Low Cluster
mean	0,568	0,480	0,375
std	0,406	0,247	0,284
min	0,137	0,078	-0,071
max	0,943	0,772	0,960

From the R^2_{Adj} in Table 2, 24 of 45 or about 53% of these countries have a R^2_{Adj} value of less than 0.4. It is because while determining the regression equation, we only used two independent variables; meanwhile, many other variables that also explain GDP are not added. If we look at the statistical distribution of each cluster in Table 3, in general, clusters with a high amount of CO_2 emissions and energy consumption produce a higher mean of R^2_{Adj}. The low cluster resulted in the lowest mean of R^2_{Adj} of 0.375. The high cluster resulted in a higher mean of R^2_{Adj} than the medium and low clusters. These results conclude that the amount of CO_2 emissions and the energy consumption is directly proportional to each cluster's mean of R^2_{Adj}. Nevertheless, it is still necessary to observe in more detail each country in the three clusters because each country has different characteristics which affect the amount of CO_2 emissions, energy consumption, and GDP.

4. Conclusions

This research concludes that the number of good clusters for classifying countries in the world based on CO_2 emissions per capita and energy consumption per capita is three (3). In this research, the clusters of countries formed based on CO_2 emissions and energy consumption per capita are high, medium, and low.

The partial correlation test shows a correlation between CO_2 emissions and GDP, energy consumption and GDP, and CO_2 consumption and energy consumption in 183 countries in the world. Nine countries showed a non-unidirectional or negative correlation between CO_2 emissions or energy consumption on GDP, which means that if CO_2 or energy consumption increases, GDP will decrease, and vice versa. Conversely, the other 36 countries produce partial correlations in the same direction, which means that if CO_2 emissions or energy consumption increase, GDP will also increase, and vice versa. One country that can be an example is Singapore which has succeeded in reducing its annual CO_2 emissions in line with an increase in its GDP.

Based on the calculations, the high cluster resulted in the highest mean of R^2_{Adj} than the medium and low clusters. The medium cluster has a mean of R^2_{Adj} between the high and low clusters. Meanwhile, the low cluster has the lowest mean of R^2_{Adj}. So countries included in the high cluster have a closer association between CO_2 emissions and energy consumption to GDP than the medium cluster. Likewise, for the medium cluster against the low cluster.

Acknowledgements

The authors would like to thank the Sepuluh Nopember Institute of Technology for fully funding to support this research.

References

[1] Martha L. Abell, James P. Braselton, John Arthur Rafter, and John A. Rafter. *Statistics with Mathematica*, volume 1. Academic Press, 1999.

[2] Daron Acemoglu. Introduction to economic growth. *Journal of Economic Theory*, 147(2):545–550, 2012. Issue in honor of David Cass.

[3] Ching Chih Chang. A multivariate causality test of carbon dioxide emissions, energy consumption and economic growth in China. *Applied Energy*, 87(11):3533–3537, Nov 2010.

[4] Jaruwan Chontanawat. Relationship between energy consumption, CO_2 emission and economic growth in ASEAN: Cointegration and causality model. *Energy Reports*, 6:660–665, Feb 2020.

[5] Norman R Draper and Harry Smith. *Applied regression analysis*, volume 326. John Wiley & Sons, 1998.

[6] Rudolf J. Freund, Wilson J. Wilson, and Ping Sa. *Regression Analysis.* Regression Analysis. Academic Press, 2nd edition, 2006.

[7] International Monetary Fund. *Back to Basics: Economic Concepts Explained.* Finance & Development. International Monetary Fund, 2019.

[8] Paulene Govender and Venkataraman Sivakumar. Application of k-means and hierarchical clustering techniques for analysis of air pollution: A review (1980–2019). *Atmospheric Pollution Research*, 11(1):40–56, 2020.

[9] Niklas Höhne, Takeshi Kuramochi, Carsten Warnecke, Frauke Röser, Hanna Fekete, Markus Hagemann, Thomas Day, Ritika Tewari, Marie Kurdziel, Sebastian Sterl, and Sofia Gonzales. The paris agreement: resolving the inconsistency between global goals and national contributions. *Climate Policy*, 17(1):16–32, 2017.

[10] Alex Reuben Kira. The factors affecting gross domestic product (gdp) in developing countries: The case of tanzania. *European Journal of Business and Management*, 5(4):148–158, 2013.

[11] Yongbum Kwon, Hyeji Lee, and Heekwan Lee. Implication of the cluster analysis using greenhouse gas emissions of Asian countries to climate change mitigation. *Mitigation and Adaptation Strategies for Global Change*, 23(8):1225–1249, 2018.

[12] Weijun Li and Zhenyu Liu. A method of SVM with normalization in intrusion detection. *Procedia Environmental Sciences*, 11(A):256–262, Jan 2011.

[13] Mohammed Musah, Yusheng Kong, Isaac Adjei Mensah, Stephen Kwadwo Antwi, and Mary Donkor. The link between carbon emissions, renewable energy consumption, and economic growth: a heterogeneous panel evidence from West Africa. *Environmental Science and Pollution Research*, 27(23):28867–28889, 2020.

[14] Saiqa Naz, Rashida Sultan, Khalid Zaman, Abdullah Mohammed Aldakhil, Abdelmohsen A. Nassani, and Muhammad Moinuddin Qazi Abro. Moderating and mediating role of renewable energy consumption, FDI inflows, and economic growth on carbon dioxide emissions: evidence from robust least square estimator. *Environmental Science and Pollution Research*, 26(3):2806–2819, 2019.

[15] Ilhan Ozturk and Ali Acaravci. CO2 emissions, energy consumption and economic growth in Turkey. *Renewable and Sustainable Energy Reviews*, 14(9):3220–3225, Dec 2010.

[16] Dimitrios Pappas, Konstantinos J. Chalvatzis, Dabo Guan, and Alexis Ioannidis. Energy and carbon intensity: A study on the cross-country industrial shift from China to India and SE Asia. *Applied Energy*, 225:183–194, Sep 2018.

[17] Sébastien Postic and Marion Fetet. Global Carbon Accounts 2020. Technical report, International Atomic Energy Agency (IAEA), France, 2020.

[18] P. V. Vara Prasad, J. M.G. Thomas, and Sruthi Narayanan. Global Warming Effects. *Encyclopedia of Applied Plant Sciences*, 3:289–299, Jan 2017.

[19] Luis Pérez-Lombard, José Ortiz, and Christine Pout. A review on buildings energy consumption information. *Energy and Buildings*, 40(3):394–398, 2008.

[20] Hannah Ritchie, Max Roser, and Pablo Rosado. CO_2 and greenhouse gas emissions. *Our World in Data*, 2020. https://ourworldindata.org/co2-and-other-greenhouse-gas-emissions.

[21] Aditya R. Setyawan, Mustika Hadijati, and Ni Wayan Switrayni. Analisis masalah heteroskedastisitas menggunakan generalized least square dalam analisis regresi. *Eigen Mathematics Journal*, 1(2):61–72, 2019.

[22] Shaojian Wang, Qiuying Li, Chuanglin Fang, and Chunshan Zhou. The relationship between economic growth, energy consumption, and co2 emissions: Empirical evidence from china. *Science of The Total Environment*, 542:360–371, Jan 2016.

[23] Patrick K. Watson and Sonja S. Teelucksingh. *A Practical Introduction to Econometric Methods: Classical and modern.* University of West Indies Press, 2002.