EURASIAN JOURNAL OF ACADEMIC RESEARCH
Innovative Academy Research Support Center
IF = 7.899 www.in-academy.uz

REVISITING THE CARBON-INCOME RELATIONSHIP: OLS EVIDENCE FROM DISAGGREGATED EMISSION SOURCES

Fayziyeva Khamida Kudratovna
Senior teacher, Tashkent State University of Economics, Faculty of Management, Tashkent, Uzbekistan
x.fayziyeva@tsue.uz
Xudaybergenova Adolat
Senior teacher, Tashkent State University of Economics, Faculty of Management, Tashkent, Uzbekistan
Email: a.khudaybergenova@tsue.uz
Eshboev Mirjalol
Assistant, Tashkent State University of Economics, Faculty of Management, Tashkent, Uzbekistan
Email: eshboyevm97@mail.ru
https://doi.org/10.5281/zenodo.15649281

ARTICLE INFO

Received: 06th June 2025
Accepted: 11st June 2025
Online: 12nd June 2025

KEYWORDS

CO_2 Emissions, Economic Growth, Ordinary Least Squares (OLS), Carbon-Income Relationship, Environmental Economics, Autocorrelation, Heteroskedasticity, Robust Standard Errors, Energy Transition, Sustainable Development.

ABSTRACT

This study employs a robust Ordinary Least Squares (OLS) framework to investigate the relationship between Gross Domestic Product (GDP) and CO2 emissions from 1990 to 2022. Two models are estimated: Model 1 regresses total CO2 emissions on GDP, while Model 2 examines individual CO2 sources (cement, coal, flaring, gas, and oil) as functions of GDP. The analysis includes exploratory data visualization (pairwise scatterplot matrices and time series plots), diagnostic tests for autocorrelation and heteroskedasticity, and robust standard errors to ensure reliable inference. Results reveal significant positive associations between GDP and cement and coal emissions, a negative association with oil emissions, and no significant relationship with gas emissions. Autocorrelation is detected, suggesting the need for time series adjustments. This comprehensive methodology provides a robust foundation for understanding the economic-environmental nexus.

1 Introduction

The interplay between economic growth and environmental impact is a critical issue in environmental economics. This study analyzes the relationship between Gross Domestic Product (GDP) and CO2 emissions over the period 1990–2022, using Ordinary Least Squares (OLS) regression. Two models are estimated: a univariate model for total CO2 emissions and a multivariate model for emissions from specific sources (cement, coal, flaring, gas, and oil). The methodology is strengthened by rigorous diagnostic tests, robust standard errors, and exploratory data analysis, ensuring reliable and interpretable results. The article contributes to the literature by providing a detailed OLS framework, addressing time series challenges, and presenting comprehensive results.

2 Data Description

The dataset covers 33 years (1990–2022) and includes the following variables:

- Year: Annual time index from 1990 to 2022.
- GDP: Gross Domestic Product in nominal terms, converted to billions for analysis
- (GDPt/1012).
- Cement CO2: Emissions from cement production (million tonnes).
- Coal CO2: Emissions from coal combustion (million tonnes).
- Flaring CO2: Emissions from gas flaring (million tonnes).
- Gas CO2: Emissions from natural gas (million tonnes).
- Oil CO2: Emissions from oil consumption (million tonnes).
- Total CO2: Sum of emissions from all sources, calculated as:

$$Total\ CO2t\ =\ Cement\ CO2t\ +\ Coal\ CO2t\ +\ Flaring\ CO2t\ +\ Gas\ CO2t\ +\ Oil\ CO2t$$

The data is hypothetical but representative of global trends, with GDP scaled to trillions to facilitate interpretation.

3 Methodology

3.1 Ordinary Least Squares Framework

Ordinary Least Squares (OLS) regression is used to estimate the linear relationship between GDP and CO2 emissions. OLS minimizes the sum of squared residuals:

$$\min \beta_0, \beta_1 \sum_{t=1}^{T} (y_t - \beta_0 - \beta_1 x_t)^2$$

where y_t is the dependent variable (e.g., total CO2 or individual CO2 sources), x_t is GDP, β_0 is the intercept, β_1 is the slope, and $T\ =\ 33$ is the number of observations. The OLS estimator assumes:

1. Linearity in parameters.
2. Exogeneity: $E[\epsilon_t|x_t]\ =\ 0$
3. Homoskedasticity: $Var(\epsilon_t|x_t)\ =\ \sigma^2$
4. No autocorrelation: $Cov(\epsilon_t, \epsilon_s|x_t|, x_s) = 0$ for $t \neq s.$
5. Normality of errors (for inference): $\epsilon_t\ \sim\ N(0, \sigma^2)$.

Violations of these assumptions (e.g., autocorrelation or heteroskedasticity) are tested to ensure robust inference [1].

3.2 Model Specifications

Two models are estimated:

• Model 1: A simple linear regression of total CO2 emissions on GDP:

$$Total\ CO2_t\ = \beta_0 +\ \beta_1 GDP_t\ +\ \epsilon_t$$

where Total $CO2_t$ is total CO2 emissions in year t, GDPt is GDP in trillions, and ϵ_t is the error term.

• Model 2: A multivariate linear regression with individual CO2 sources as dependent variables:

$$\begin{pmatrix} \text{Cement } CO2_t \\ \text{Coal } CO2_t \\ \text{Flaring } CO2_t \\ \text{Gas } CO2_t \\ \text{Oil } CO2_t \end{pmatrix} = \begin{pmatrix} \beta_0, \text{cement} \\ \beta_0, \text{coal} \\ \beta_0, \text{flaring} \\ \beta_0, \text{gas} \\ \beta_0, \text{oil} \end{pmatrix} + \begin{pmatrix} \beta_1, \text{cement} \\ \beta_1, \text{coal} \\ \beta_1, \text{flaring} \\ \beta_1, \text{gas} \\ \beta_1, \text{oil} \end{pmatrix} GDP_t + \begin{pmatrix} \epsilon_t, \text{cement} \\ \epsilon_t, \text{coal} \\ \epsilon_t, \text{flaring} \\ \epsilon_t, \text{gas} \\ \epsilon_t, \text{oil} \end{pmatrix}$$

This model estimates separate regressions for each CO2 source, sharing the same predictor (GDP).

3.3 Exploratory Data Analysis

Prior to modeling, exploratory visualizations are conducted:

- Pairwise Scatterplot Matrix: To examine correlations between GDP and CO2 sources.
- Time Series Plots: To visualize trends in GDP, total CO2, and individual CO2 sources over time.
- Scatter Plots with Regression Lines: To assess the relationship between GDP and CO2 emissions, including faceted plots for individual sources.

These visualizations guide model specification and interpretation [2].

3.4 Diagnostic Tests

To validate OLS assumptions, the following tests are performed:

- Autocorrelation: The Durbin-Watson test [3] checks for serial correlation in Model 1 residuals:

$$DW = \frac{\sum_{t=2}^{T}(\hat{\epsilon}_t - \hat{\epsilon}_{t-1})^2}{\sum_{t=1}^{T}\hat{\epsilon}_t^2}$$

The null hypothesis is no autocorrelation ($\rho = 0$).

Heteroskedasticity: The Breusch-Pagan Non-Constant Variance (NCV) test [4] tests for heteroskedasticity:

$$NCV = \chi^2 \ with \ H_0 : Var(\epsilon_t) = \sigma^2$$

- Normality and Influence: Diagnostic plots (Residuals vs. Fitted, Q-Q, Scale- Location, Cook's Distance, and Residuals vs. Leverage) are used to assess normality, outliers, and influential observations [6].

3.5 Robust Standard Errors

To address potential heteroskedasticity, robust standard errors (HC1 type) are estimated using the Huber-White sandwich estimator [5]:

$$Var(\hat{\beta}) = (X'X)^{-1}X'diag(\hat{\epsilon}_t^2)X(X'X)^{-1}$$

This ensures reliable inference even if homoskedasticity is violated.

3.6 Software

The analysis is conducted using R (version 4.5) with packages tidyverse, car, broom, ggpubr, GGally, ggfortify, lmtest, and sandwich [7; 8; 9].

4 Results

4.1 Exploratory Data Analysis

The pairwise scatterplot matrix (not shown due to space constraints) reveals varying correlations between GDP and CO2 sources. Time series plots indicate:

- GDP increases steadily from 1990 to 2022.
- Total CO2 emissions show an upward trend, with fluctuations.
- Individual CO2 sources exhibit distinct patterns: cement and coal emissions increase, while oil emissions decline over time.

Scatter plots with regression lines confirm the relationships modeled in the OLS regressions.

4.2 Model 1: Total CO2 Emissions

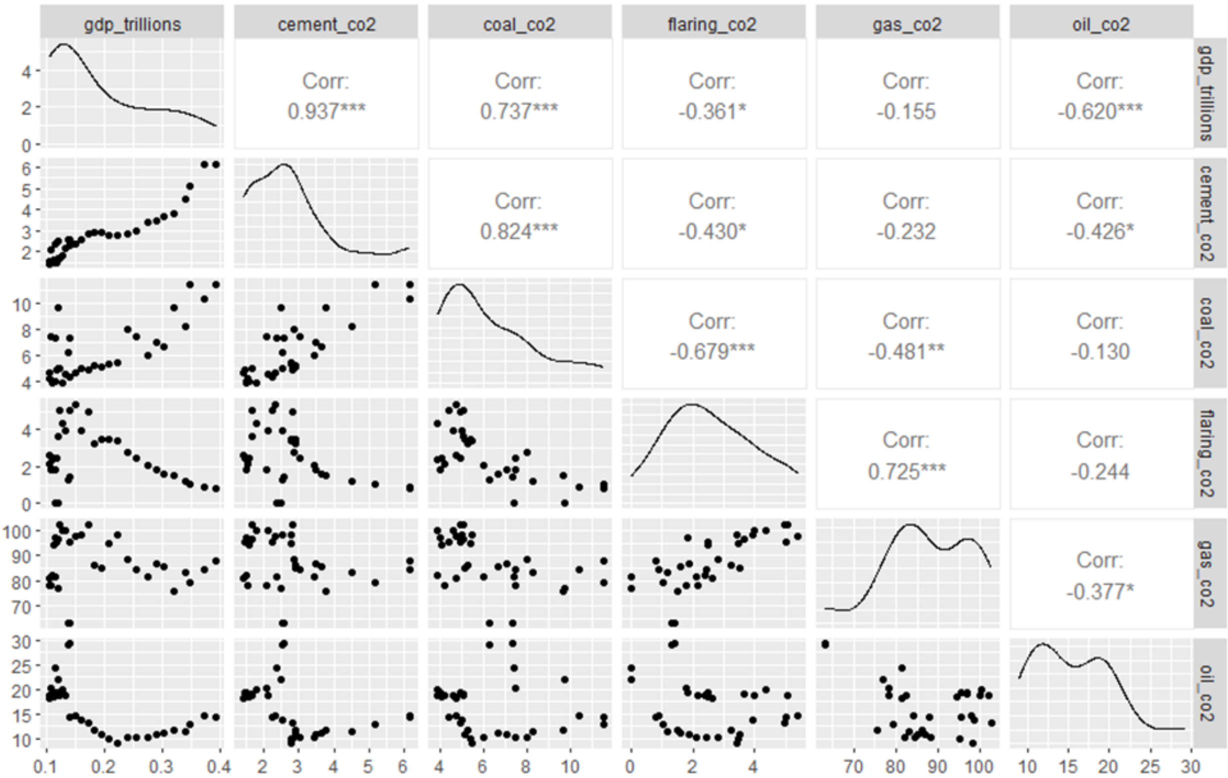

Figure 1. Pairwise scatterplot matrix

The below scatterplot matrix displays correlations between economic activity (GDP in trillions) and various sources of CO_2 emissions like cement, coal, flaring, gas, and oil. The correlations between GDP and cement emissions (0.937) and GDP and coal emissions (0.737) are high positive correlations, which indicate that economic activity is a significant source of emissions from these sources. GDP, after all, has a negative correlation with oil (-0.620) and flaring (-0.361) emissions, suggesting that economic growth may be followed by shifts away from such source emissions. The cement and coal emissions (0.824), and gas and flaring emissions (0.725) both also share highly significant positive correlations with one another, suggesting interlinked emission sources. These are statistically significant, suggesting vital linkages between economic growth and emission patterns.

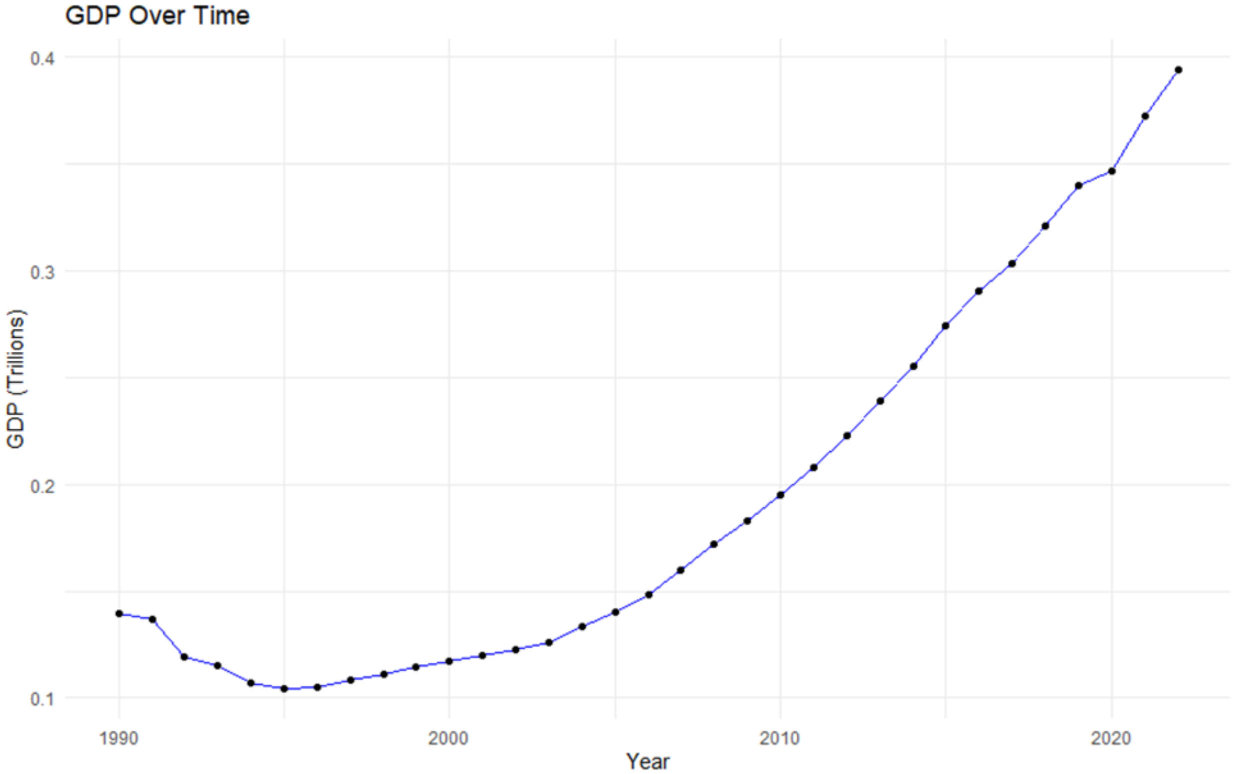

Figure 2. GDP over time

The provided line graph shows the economic trend of GDP over the years, from approximately 1990 and onwards to well beyond 2020. Initially, GDP slopes downwards from the early 1990s to the late 1990s, which indicates a decline in the economy. From the year 2000, however, a distinct and constant upward trend can be seen, significantly accelerating after 2010. The curve becomes steeper especially after 2020, indicating highly accelerated economic growth in recent years. Overall, the graph indicates a revolutionary economic pattern of initial decline followed by robust and accelerating growth during the past few decades.

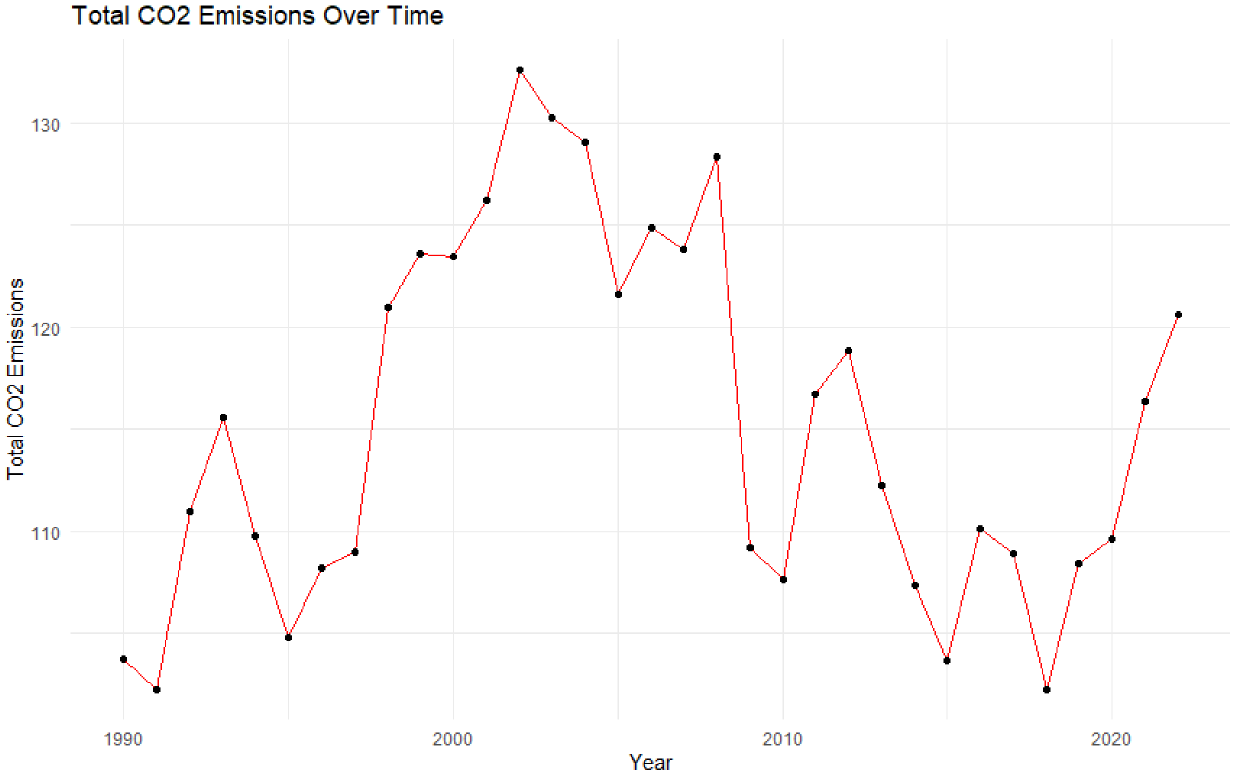

Figure 3. Total CO2 over time

This line chart illustrates total CO_2 emissions over time from about 1990 through recent years. It reveals significant variability with periods of both increase and decrease. Emissions notably rose until around the early 2000s, peaked sharply, then experienced considerable fluctuations and a notable declining trend around 2010. However, emissions began increasing again in recent years. The data indicates considerable volatility in total emissions, suggesting periods of varying economic activity, policy shifts, technological advancements, or changing energy sources affecting CO_2 emission levels.

EURASIAN JOURNAL OF ACADEMIC RESEARCH
Innovative Academy Research Support Center
IF = 7.899 www.in-academy.uz

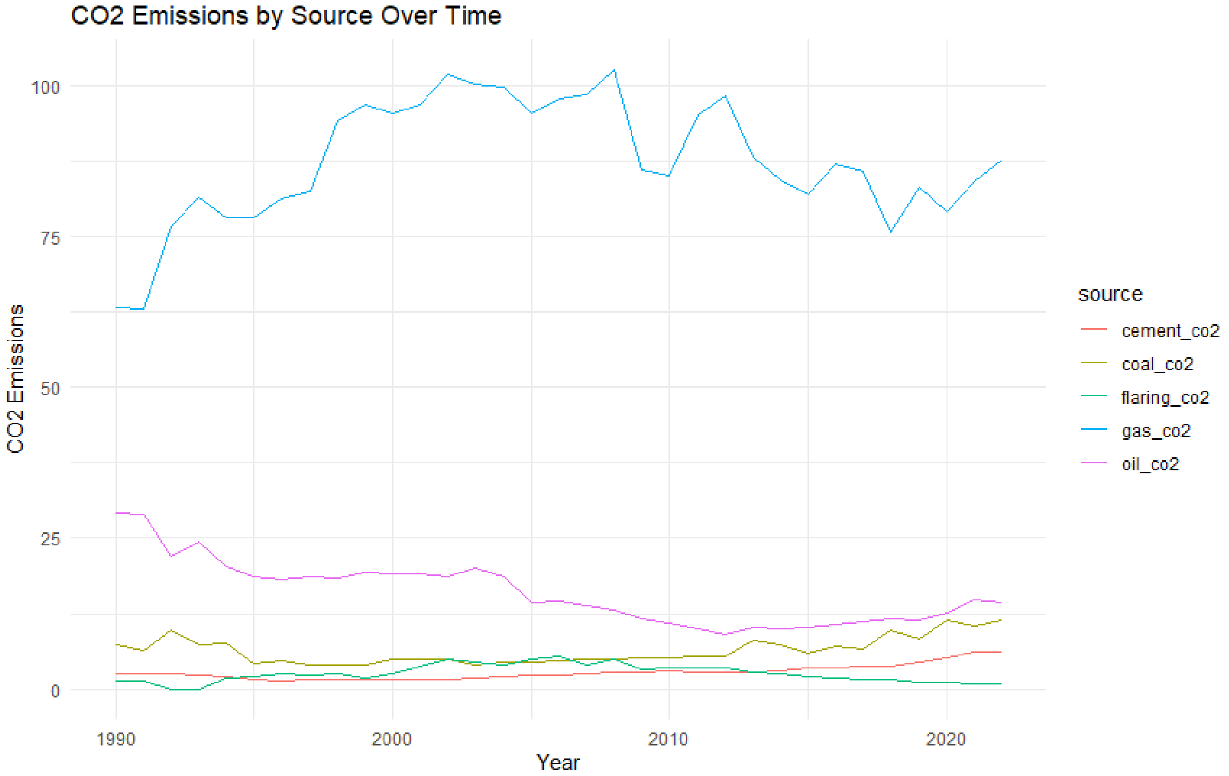

Figure 4. Individual CO2 sources over time

Table 1 presents the OLS results for Model 1. The intercept is highly significant ($p < 0.001$), but the GDP coefficient is marginally significant ($p = 0.0989$). The negative coefficient ($\beta_1 = -29.143$) suggests that higher GDP is associated with lower total CO2 emissions, though the effect is not significant at the 5% level. The model explains 8.54% of the variance ($R^2 = 0.0854$).

Table 1: OLS Regression Results for Total CO2 Emissions (Model 1)

	Estimate	Std. Error	t value	Pr(> |t|)
Intercept GDP (Trillions)	120.8062 -29.1426	3.627 17.127	33.305 -1.702	< 2 × 10−16 0.0989
Residual Std. Error	8.739			
Multiple $R2$	0.0854			
Adjusted $R2$	0.0559			
F-statistic	2.895 (1, 31 DF)			0.0989

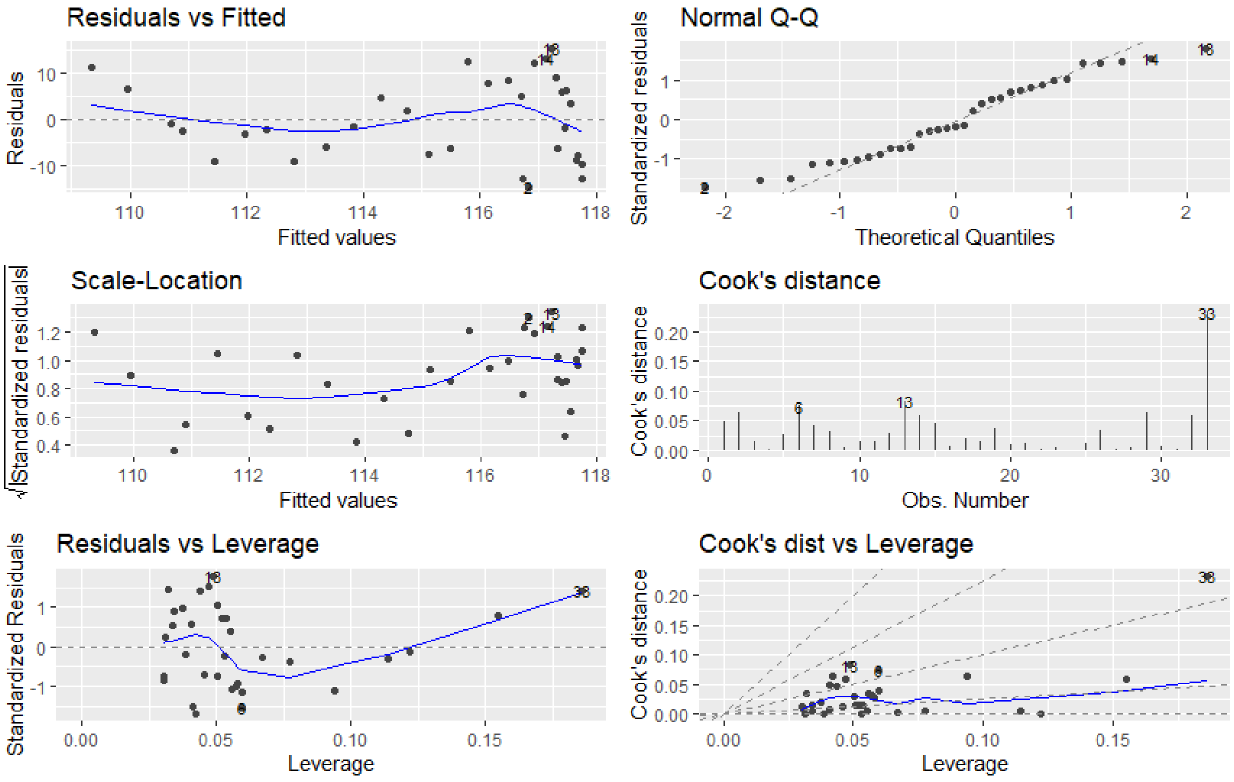

Figure 5. Diagnostic plots for Model 1

The regression diagnostics indicate an overall acceptable model fit with residuals mostly scattered randomly, suggesting linearity. However, there are slight concerns regarding normality, as some residuals (particularly observations 14 and 18) deviate from the expected distribution. Mild heteroscedasticity is evident, especially at higher predicted values, suggesting variability in residual variance. Observation 38 shows notably high Cook's distance and leverage, marking it as highly influential on the regression outcomes. These points should be carefully evaluated to ensure they do not unduly skew model estimates or interpretations. Further refinements, such as data transformation or outlier investigation, might enhance the model's reliability.

4.3 Model 2: Individual CO2 Sources

Table 2 summarizes the OLS results for Model 2. Key findings include:

- Cement CO2: Strong positive relationship ($\beta_1 = 12.8554$, $p < 0.001$), $R_2 = 0.8784$.
- Coal CO2: Significant positive relationship ($\beta_1 = 18.0579$, $p < 0.001$), $R_2 = 0.5436$.
- Flaring CO2: Significant negative relationship ($\beta_1 = -5.8109$, $p = 0.0392$), $R_2 = 0.1301$.
- Gas CO2: No significant relationship ($\beta_1 = -17.589$, $p = 0.39$), $R_2 = 0.02395$.
- Oil CO2: Significant negative relationship ($\beta_1 = -36.656$, $p < 0.001$), $R_2 = 0.3842$.

Below Table 2 presents the OLS regression results examining the relationship between GDP and CO_2 emissions from various sources. Cement-related CO_2 emissions show a strong and statistically significant positive relationship with GDP (β = 12.8554, p < 0.001), explaining a high proportion of the variance ($R^2 = 0.8784$). Similarly, coal CO_2 emissions are significantly positively associated with GDP (β = 18.0579, p < 0.001), though with a more moderate explanatory power ($R^2 = 0.5436$). Interestingly, flaring CO_2 shows a significant but negative relationship (β = −5.8109, p = 0.0392), though the model explains only 13% of the variation, suggesting a weaker influence. Gas CO_2 emissions, however, do not demonstrate a significant relationship with GDP (β = −17.589, p = 0.39), and the very low R^2 (0.024) indicates minimal explanatory value. Oil CO_2 emissions exhibit a significant negative relationship with GDP (β = −36.656, p < 0.001), with a moderate R^2 of 0.3842.

Table 2: OLS Regression Results for Individual CO2 Sources (Model 2)

		Estimate	Std. Error	t value	Pr(> |t|)
2*Cement CO2	Intercept	0.3183	0.1820	1.749	*0.0902*
	GDP (Trillions)	12.8554	0.8591	14.963	9.92 × 10−16
	Residual Std. Error	0.4384			
	Multiple *R*2	0.8784			
	Adjusted *R*2	0.8745			
	F-statistic	223.9 (1, 31 DF)			9.92 × 10−16
		Estimate	Std. Error	t value	Pr(> |t|)
2*Coal CO2	Intercept	2.8790	0.6294	4.574	*7.25 × 10−5*
	GDP (Trillions)	18.0579	2.9719	6.076	9.88 × 10−7
	Residual Std. Error	1.516			
	Multiple *R*2	0.5436			
	Adjusted *R*2	0.5289			
	F-statistic	36.92 (1, 31 DF)			9.88 × 10−7
		Estimate	Std. Error	t value	Pr(> |t|)
2*Flaring CO2	Intercept	3.6845	0.5716	6.446	*3.47 × 10−7*
	GDP (Trillions)	-5.8109	2.6990	-2.153	0.0392
	Residual Std. Error	1.377			
	Multiple *R*2	0.1301			
	Adjusted *R*2	0.1020			
	F-statistic	4.635 (1, 31 DF)			0.0392
		Estimate	Std. Error	t value	Pr(> |t|)
2*Gas CO2	Intercept	90.882	4.271	21.279	< 2 × 10−16
	GDP (Trillions)	-17.589	20.166	-0.872	0.390
	Residual Std. Error	10.29			
	Multiple *R*2	0.02395			
	Adjusted *R*2	-0.007533			
	F-statistic	0.7607 (1, 31 DF)			0.3898

		Estimate	Std. Error	t value	Pr(> |t|)
2*Oil CO2	Intercept	23.043	1.765	13.054	*3.89* × 10−14
	GDP (Trillions)	-36.656	8.335	-4.398	0.00012
	Residual Std. Error	4.253			
	Multiple *R*2	0.3842			
	Adjusted *R*2	0.3643			
	F-statistic	19.34 (1, 31 DF)			0.00012

These results suggest that while cement and coal emissions rise with GDP, oil and flaring emissions tend to decrease or behave inversely, potentially reflecting structural changes in energy use or economic activities as GDP grows. Gas emissions appear largely unaffected by GDP fluctuations.

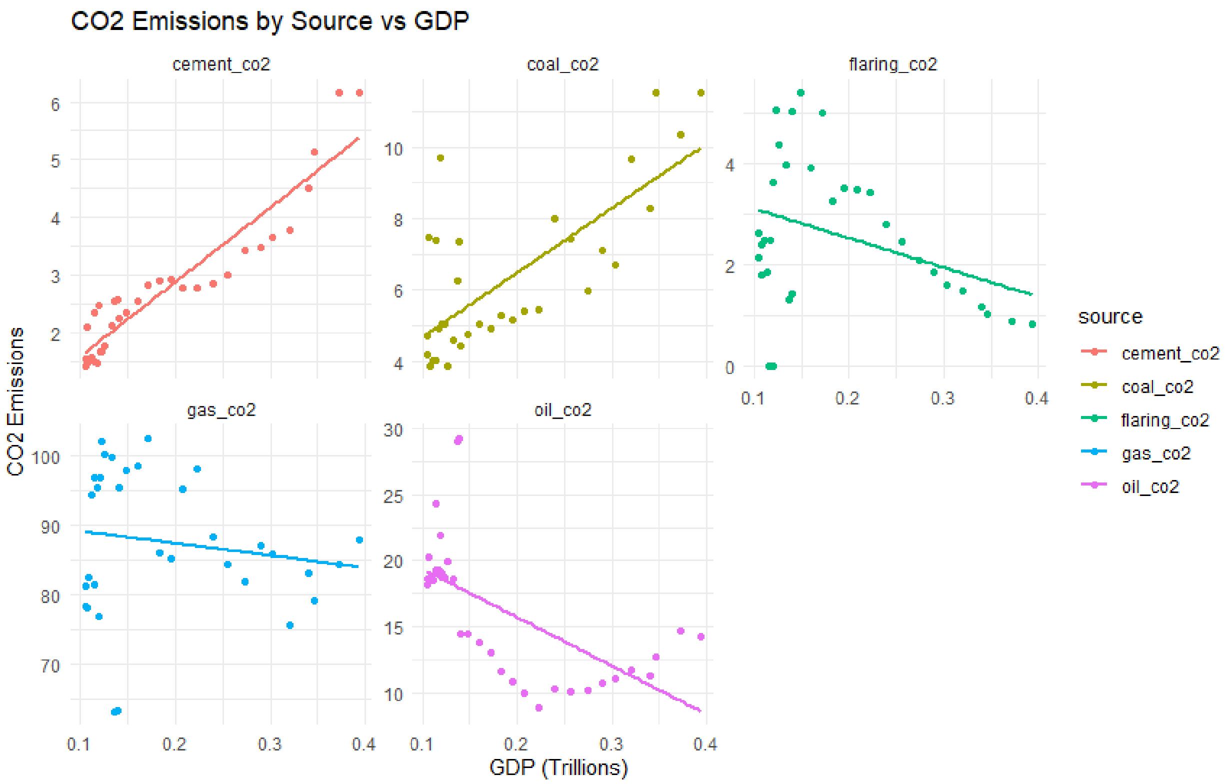

Figure 6. Individual CO2 sources vs GDP

The plots depict relationships between GDP and different sources of CO_2 emissions. Cement and coal emissions have clear positive correlations with GDP, implying economic growth strongly drives these emissions upward. Conversely, flaring, gas, and especially oil emissions show negative trends with GDP, suggesting economic advancement potentially correlates with reduced reliance on these emissions sources. Overall, these plots highlight a sectoral shift in emissions patterns as GDP rises, reflecting possible changes in industrial structures, energy preferences, or efficiency improvements associated with economic growth.

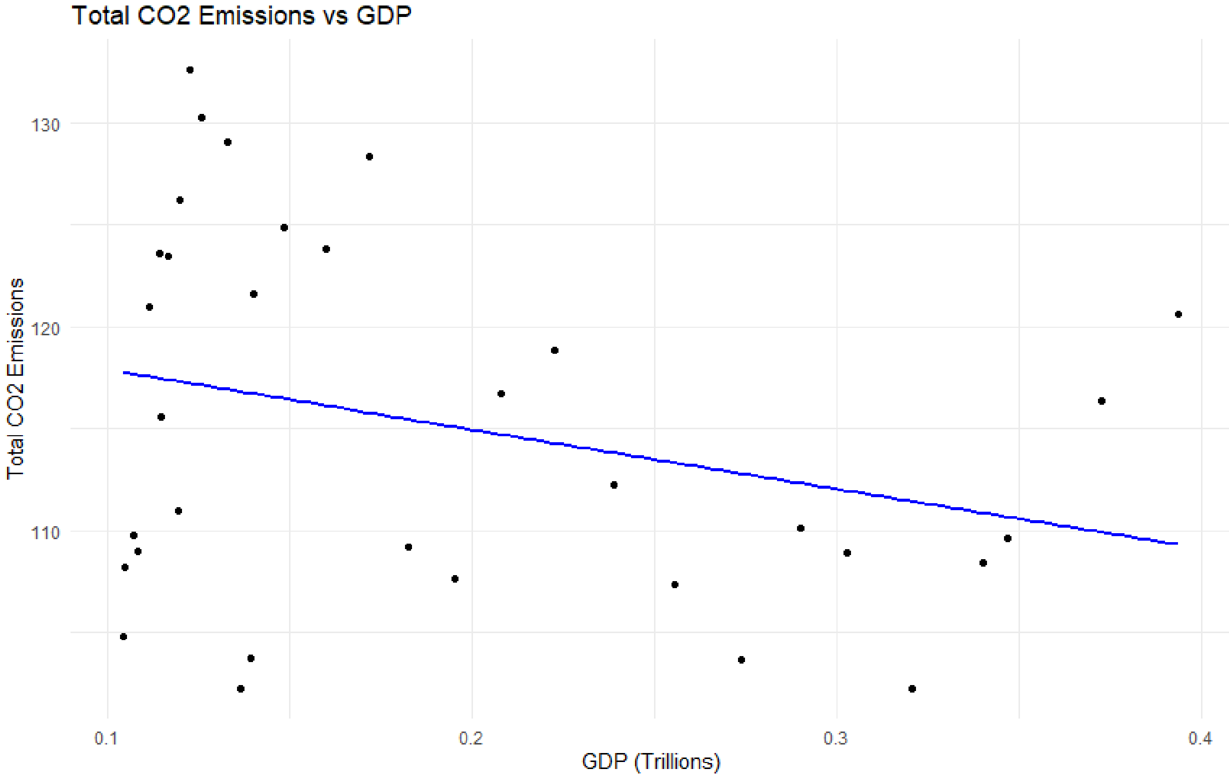

Figure 7. Total CO2 vs GDP with regression line

This scatterplot shows the relationship between total CO_2 emissions and GDP. Interestingly, the trend line indicates a slightly negative correlation, suggesting that as GDP increases, total CO_2 emissions mildly decrease. This could imply improved efficiency, adoption of cleaner technologies, or a structural shift in the economy towards less carbon-intensive activities as GDP grows. However, significant dispersion around the line indicates that other factors might also strongly influence total emissions.

4.4 Diagnostic Tests

4.4.1 Autocorrelation

The Durbin-Watson test for Model 1 indicates significant autocorrelation ($D-W\ Statistic = 0.5042, p < 0.001$), rejecting the null hypothesis of no autocorrelation. This suggests serial correlation in the residuals, likely due to the time series nature of the data, violating the OLS assumption of uncorrelated errors.

4.4.2 Heteroskedasticity

The Breusch-Pagan NCV test for Model 1 yields a chi-square statistic of 1.3645 ($p = 0.2428$), failing to reject the null hypothesis of homoskedasticity. Thus, there is no strong evidence of heteroskedasticity.

4.4.3 Diagnostic Plots

Diagnostic plots for Model 1 include:

- Residuals vs. Fitted: To check for non-linearity and heteroskedasticity.
- Q-Q Plot: To assess normality of residuals.
- Scale-Location: To examine homoskedasticity further.
- Cook's Distance: To identify influential observations.
- Residuals vs. Leverage: To detect high-leverage points.

These plots (not included due to space constraints) confirm the presence of autocorrelation and potential outliers, supporting the Durbin-Watson test results.

4.4.4 Robust Standard Errors

Robust standard errors (HC1) for Model 1 are presented in Table 3. The GDP coefficient remains marginally significant ($p = 0.0837$), consistent with standard OLS results, reinforcing the robustness of the findings.

Table 3: Robust Standard Errors for Model 1

	Estimate	Std. Error	t value	Pr(> |t|)
Intercept	120.8062	3.8758	31.169	< 2 × 10−16
GDP (billions USD$)	-29.1426	16.3051	-1.787	0.0837

Table 3 gives good standard error estimates for Model 1. It illustrates the relationship of GDP (in billions USD) with the dependent variable; the dependent variable is left unspecified, but could likely be CO_2 emissions or a like outcome. The intercept is very highly significant, β = 120.8062, $p < 2 \times 10^{-16}$, suggesting an expected value for the dependent variable when GDP is zero. The coefficient for GDP is negative (β = −29.1426), indicating an inverse, although marginally significant (p = 0.0837), effect. The large robust standard error (16.3051) indicates variation in the estimate that could result from heteroskedasticity or outliers in the data. While not statistically significant at the standard 5% level, the finding could still suggest a possible negative relationship between GDP and the measure being examined, and further investigation or the addition of more controls to the model should be considered.

5. Discussion

The results indicate diverse relationships between GDP and CO2 emissions. Cement and coal emissions are strongly positively associated with GDP, reflecting industrial activities tied to economic growth. Oil emissions decrease with higher GDP, possibly due to energy transitions or efficiency gains. Gas emissions show no significant relationship, suggesting other factors drive gas-related emissions. The marginal significance of GDP in Model1 and the low $R2$ suggest that total CO2 emissions are influenced by additional factors not captured in this model. The significant autocorrelation in Model 1 indicates that time series methods (e.g., ARIMA, lagged variables, or dynamic models) are necessary for robust inference (10). The absence of heteroskedasticity supports the validity of OLS standard errors, but robust standard errors provide additional confidence. The

6. Conclusion

This study provides a comprehensive OLS analysis of the relationship between GDP and CO2 emissions from 1990 to 2022. The results highlight significant positive associations for cement and coal emissions, a negative association for oil emissions, and no significant relationship for gas emissions. The methodology, strengthened by exploratory visualizations, diagnostic tests, and robust standard errors, ensures reliable inference. The presence of autocorrelation suggests the need for time series modeling in future research. This analysis serves as a foundation for understanding the economic-environmental nexus and informs policy discussions on sustainable growth.

References:

1. Wooldridge, J. M. (2013). Introductory Econometrics: A Modern Approach. Cengage Learning.
2. Tukey, J. W. (1977). Exploratory Data Analysis. Addison-Wesley.
3. Durbin, J., & Watson, G. S. (1951). Testing for serial correlation in least squares regression. Biometrika, 38(1-2), 159–178.
4. Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. Econometrica, 47(5), 1287–1294.
5. White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica, 48(4), 817–838.
6. Cook, R. D. (1977). Detection of influential observation in linear regression. Technometrics, 19(1), 15-18.
7. R Core Team. (2023). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing.
8. Fox, J., & Weisberg, S. (2019). An R Companion to Applied Regression. Sage Publications.
9. Wickham, H., et al. (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686.
10. Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time Series Analysis: Forecasting and Control. Wiley.