Bol. Soc. Paran. Mat.
©SPM – E-ISSN-2175-1188
SPM: www.spm.uem.br/bspm

(3s.) **v. 2025 (43)** : 1–15.
ISSN-0037-8712
doi:10.5269/bspm.76456

Robust Analysis of Regression of Turkey's CO_2 Emission Dataset

Vijayabalan D*, Divya G, Muthumari G, Anand Gnana Selvam S, Kuppuswami G, Geethamalini S

ABSTRACT: This study uses multiple line regression and the generalised least squares (GLS) methods to investigate the factors influencing CO_2 emissions in Turkey in detail. It does this by demonstrating the differences in the impact of factors on CO_2 emissions at multiple levels and by proving the superiority of the multiple line regression method over the GLS, LS, and RR methods. This study looks at the impact of Turkey's industrialisation on CO_2 emissions from 1973 and 2021. The CO_2 emission dataset is analyzed using least squares (LS), generalized least squares (GLS), and robust regression (RR) techniques. The mean squared error (MSE) is used to compare these regression estimate methods. The paper also looks at important regression assumptions such multicollinearity, autocorrelation, heteroscedasticity, and normality. Model selection techniques are applied in the multicollinearity detection case to remove superfluous variables. To find the best model for the CO_2 emission dataset, all regression estimation methods are re-applied after variables have been eliminated. The study also examines and corrects for heteroscedasticity and autocorrelation, if they are found. This study intends to offer important insights into the environmental effects of economic development by thoroughly examining the connection between industrialization and CO_2 emissions in Turkey.

Key Words: CO_2 Emission, robust regression, generalized least squares, least squares.

Contents

1 Introduction

Significant improvements in human welfare, technological innovation, and product manufacturing were brought about by the Industrial Revolution. But in addition to these advantages, researchers are looking more closely at its drawbacks. These adverse effects are generally divided into three primary categories: contamination of the soil, water, and weather. In addition to directly endangering human health, atmospheric pollution fuels the greenhouse effect, which is one of the primary causes of global warming. According to scientific consensus, extensive glaciation might occur on Earth if the average global temperature were to drop by just 1.5 to 7.5 degrees Celsius. On the other hand, a temperature increase of 1.5 to 5.5 degrees Celsius might create a tropical climate on a worldwide scale. For more information, see Plass (1956). Thanks to easily accessible data from glaciers, a large portion of the scientific literature is devoted to the study of rising average temperatures. Important records of the chemical and physical makeup of the Earth's atmosphere can be found in ice sheets and caps. According

* Corresponding author.
2010 *Mathematics Subject Classification:* 62C10, 62C20.
Submitted April 19, 2025. Published September 22, 2025

to Thompson (2000), ice cores excavated from carefully chosen locations are frequently used to collect this data. One of the first to identify the possible link between greenhouse gases and global temperature was Arrhenius (1896). He postulated that the earth's temperature might be affected by the burning of fossil fuels. But because his calculations relied on the pace at which fuels burned at the time, he concluded that it would take around 3,000 years to double the amount of CO_2 in the atmosphere. In a similar vein, Ekholm (1901) added to this discussion by drawing attention to the effects of present coal use on global warming. His research highlighted the possibility that burning pit coal could raise the planet's average temperature. The foundation for future studies on the connection between human activity and climate change was established by these early discoveries.

Scientists started paying closer attention to this issue after industrialization had a major impact on rising CO_2 levels. In a groundbreaking study, Callendar (1938) estimated that the world temperature had increased by 0.3 degrees Celsius between 1880 and the 1930s. This study offered preliminary proof that industrial activity and temperature increases are related. Plass (1955) initiated additional research on the greenhouse effect and its influence on global climate change by delving deeper into the possible impacts of CO_2 in the atmosphere. His research served as a crucial basis for comprehending the mechanics underlying climate dynamics. Landsberg (1970) drew attention to how urbanization affects the climate worldwide. He underlined the difficulty of differentiating between human-induced changes and natural climatic variability, as well as the substantial impact that local changes have on the immediate environmental conditions. In addition, Landsberg expressed worries about the long-term effects of elements like heat rejection and CO_2 absorption, indicating that climate research and management continue to face difficulties. A significant study of temperature trends over a century, from 1880 to 1980, was conducted by Hansen and Lebedeff (1988) in order to demonstrate global temperature rise and the ensuing climate changes on a global scale. Their findings made a substantial contribution to our knowledge of long-term temperature trends and how they affect climate dynamics. More recently, Wang et al. (2023) examined global average temperatures from 1850 to 2021 using the Berkeley Earth Surface Temperature (BEST) dataset. Their analysis supported previous findings that the average global temperature has been rising noticeably over time. These repeated results highlight how urgent it is to solve climate change. In 2015, the United Nations held an important conference in Paris with the goal of reducing the increase in average world temperatures. With a more ambitious target of 1.5 degrees Celsius, the objective was to keep the increase in global temperature to less than 2 degrees Celsius. This emphasizes how climate change is acknowledged to have major detrimental effects on both natural systems and human populations. To lessen the adverse effects of climate change, immediate action is required. According to Crain et al. (2018), there is an urgent need to cut CO_2 emissions in a number of sectors. To reduce greenhouse gas emissions and fight climate change, policymakers are urged to enact stringent restrictions that target transportation, industry, and daily CO_2 usage. The integrity of the planet's ecosystems and the welfare of current and future generations depend on such actions. The analysis of CO_2 emission data and its consequences on Turkey is the specific subject of this article, which seeks to determine both the positive and negative effects. It expands on previous research that looks at CO_2 emissions in the context of Turkey. For instance, Halicioglu (2009) looked into the connection between Turkey's foreign commerce, energy consumption, income levels, and CO_2 emissions. The study examined patterns from 1960 to 2005 using variables such CO_2 emissions per capita, commercial energy consumption per capita, real income per capita, and openness ratio. Halicioglu sought to demonstrate the intricate connections between these variables and how they affect Turkey's CO_2 emissions through this analysis. The findings indicate a positive relationship between income levels, energy use, and CO_2 emissions, indicating that environmental deterioration has been exacerbated by Turkey's economic expansion. Halicioglu does, however, understand the complexity of this relationship and the possibility that attempts to slow economic growth could result in higher unemployment rates. Ozturk and Acaravci (2010) used the autoregressive distributed lag-bounds testing method of cointegration to examine the causal relationships and long-term trends between economic growth, carbon emissions, energy consumption, and employment ratio in Turkey between 1968 and 2005. Notably, energy consumption per capita tends to increase with income, although income growth is associated with a decrease in carbon emissions per capita. It's interesting to note that while employment has a short-term effect on GDP per capita, neither carbon emissions nor energy consumption directly affect it.

Aydin (2015) uses multiple linear regression and correlation analysis to investigate the factors impacting energy-related CO_2 emissions (ERCDE) in Turkey. The analysis covers five metrics from 1971 to 2010: GDP, combustible renewables and waste energy consumption (CRWC), alternative and nuclear energy consumption (ANEC), nation population (CP), and fossil fuel consumption (FFC). According to the results, CP appears to be the main factor affecting ERCDE, with GDP, ANEC, FFC, and CRWC following in order of relevance. Additionally, it is suggested that CP, GDP, and FFC can be used to accurately anticipate ERCDEs. From 1990 to 2013, Akbostancı et al. (2018) investigated CO_2 emissions in the Turkish economy from a variety of industries, including forestry, manufacturing, construction, transportation, residential areas, agriculture, and electricity production. They also emphasized how important it is to analyze emissions at the industry level in order to create effective industrial and environmental strategies, particularly in view of Turkey's energy sector challenges and conflicting energy policies. To determine CO_2 emissions in Turkey in 2015, Mangır and Şahin (2022) used ecologically extended input-output (EEIO) accounting of trade-emissions variables, consumption-based accounting, and production-based accounting. According to the findings, Turkey is a net importer of CO_2 throughout the year and imports greenhouse emissions equal to about 9.67 percent of the inventory of greenhouse gases and 7.7 percent for household use. Furthermore, Turkey is a net importer of CO_2 throughout the year, importing about 9.67% of total greenhouse gas emissions and 7.7% for domestic usage. Additionally, it is determined that the service sector has large emissions, particularly in sectors like transportation, financial intermediation, health and other services, education, and commercial operations. Telatar and Birinci (2022) examined the connection between CO_2 emissions, ecological footprint, and environmental levies. The researchers chose the years 1994–2019 in order to investigate this association. In conclusion, there was no sustained decrease in CO_2 emissions as a result of the implementation of ecological levies. Given these results, it was determined that environmental taxes have no effect on slowing or stopping ecological deterioration.

The literature identifies a variety of explanatory criteria for comprehending CO_2 emissions, stressing the significance of investigating their connections with important elements including GDP, oil and electrical energy consumption, and a nation's overall carbon footprint. For the purpose of directing policy decisions to solve environmental issues and advance sustainability, this investigation is essential. The study intends to thoroughly examine the connections between CO_2 emissions and different factors using this methodical approach, providing insightful information to stakeholders and policymakers entrusted with creating practical policies to reduce environmental impact and advance sustainable development.

1.1 Research objectives

The study's remaining sections are organized as follows: An outline of the variable elimination procedure and the LS, GLS, and RR approaches is given in Section 2. The analysis's CO_2 emission dataset is described in Section 3. Using LS, GLS, and RR techniques, Section 4 performs a thorough data analysis while looking for departures from underlying assumptions. This part also covers the variable elimination procedure used to improve the models. Lastly, the analysis's findings are shown in Section 5.

1.2 Contribution of the study

Examining the factors that influence CO_2 emissions the primary measure of environmental pollution that contributes to climate change is crucial. The investigation of variables that significantly affect Turkey's ability to reduce its CO_2 emissions served as the study's driving force. This article focuses on how Turkey's CO_2 emissions are affected by population growth, economic expansion, and the use of renewable energy.

This study seeks answers to the following research questions:

- Does Turkey's economic expansion impact CO_2 emissions?
- Does Turkey's population expansion have a statistically significant impact on CO_2 emissions?
- Does Turkey's shift to renewable energy sources effectively lower CO_2 emissions? The quantile regression model approach has not been used to analyse how Turkey's renewable energy use, population increase, and economic expansion affect CO_2 emissions. To address this research gap, the multiple line regression approach was used in this study to assess the correlations between the variables in question over the 1973–2021 period in Turkey. The following are some ways that this study adds to the literature:

(i) By investigating the connection between Turkey's use of renewable energy, population increase, economic expansion, and CO_2 emissions, this study adds to the body of literature.

(ii)The prediction performances of multiple line regression and other regression are compared in this study.

2 Method

One of the oldest techniques for modeling variables is regression analysis, which is applied extensively in many scientific fields. Numerous disciplines, including engineering, physics, chemistry, biology, economics, social sciences, and life sciences, can benefit from its adaptability. The associated regression techniques utilized in this work are briefly defined in the section that follows.

2.1 Multiple linear regressions

Assuming a linear relationship between y_i, x_i, the regression model can be represented as follows: Let $(x_1, y_1), \ldots, (x_n, y_n)$ be a given sample, where y_i is the dependent variable, $x_(i)$is the vector of independent variables, β is an unknown regression parameter vector, and ϵ_i is the random error.

$$y_i = x_i^{'}\beta + \epsilon_i \tag{2.1}$$

This regression model can also be expressed using matrix notation. Let X be a design matrix consisting of independent variables, y be the independent variable vector, β be an unknown regression parameter vector, and ϵ be the random error vector. The matrix notation of the regression model is:

$$y = X\beta + \epsilon. \tag{2.2}$$

2.2 LS Method

One popular method for determining the regression parameters in equation (2.1) is the least squares (LS) method. The error terms are minimized to yield the LS estimator of β, which is as follows:

$$\beta = (X'X)^{(-1)}X'y. \tag{2.3}$$

2.3 GLS Method

The Generalized Least Squares (GLS) method represents a refined approach to the LS method, especially advantageous in scenarios where the dataset demonstrates autocorrelation. The GLS estimator of β can be derived in the following manner:

$$\beta_G LS = \left(X'\Sigma^{(-1)}X\right)^{(-1)} X'\Sigma^{(-1)}y. \tag{2.4}$$

The matrix Σ in equation (2.4) is described as a positive and symmetric nn matrix whose $(i, j)^{th}$ element is equal to $Cov(e_i, e_j)$. Here, $Cov(e_i, e_j)$ represents the covariance of the residuals of β GLS. Furthermore, the GLS estimator can be obtained by transforming the LS method. Let S be the Cholesky decomposition of Σ. Then, the transformed model given in equation (2.1) becomes:

$$Y^* = X^*\beta + \epsilon^*, \tag{2.5}$$

Where $Y^* = S^{(-1)}y, X^* = S^{(-1)}x$ and $\epsilon^* = S^{(-1)}\epsilon$. Alternatively the GLS estimator can be obtained as:

$$\beta^* = (X^*X^*)^{(-1)}X^*Y^*. \tag{2.6}$$

When tackling potential challenges that may arise from datasets, including outliers, the Robust Regression (RR) method stands out as an important statistical tool. A substantial body of literature has been developed on RR methods, featuring contributions from Huber (1973), Andrews (1974), Rousseeuw and Yohai (1984), Hampel et al. (1986), Yohai (1987), and Rousseeuw and Leroy (1987).

The M-estimator, introduced by Huber in 1973, is widely recognised as one of the most frequently employed techniques in RR. Although M-estimators may not be strong in the presence of extreme values, they are commonly used with the understanding that outliers tend to appear more frequently in the dependent variable y. The aim of the method is to reduce the function of residuals, which escalates at a slower pace compared to the least squares function. The M-estimator of β can be derived through the following method:

$$min_\beta \sum_{i=2}^{n} \rho(\epsilon_i) = min_\beta \sum_{i=2}^{n} \rho(y_i - x_i\beta). \tag{2.7}$$

Where $x_i^{'}$ indicate ith row of X. The M-estimator is so named due to its use of the maximum likelihood method. The scale-invariant property of the M-estimator is given as:

$$min_\beta \sum_{i=2}^{n} \rho\left(\frac{\epsilon_i}{\sigma}\right) = min_\beta \sum_{i=2}^{n} \rho\left(\frac{y_i - x_i\beta}{\sigma}\right). \tag{2.8}$$

Where σ is a robust estimate of scale:

$$\sigma = median|\epsilon_i - median(\epsilon_i)|/(0.6745) \tag{2.9}$$

With the help of the following equation, β and σ parameters can be solved simultaneously:

$$Q(\beta,\sigma) = \frac{1}{n}\sum_{i=2}^{n}\left[\rho\left(\frac{y_i - x_i\beta}{\sigma} + \alpha\right)\right]\sigma \tag{2.10}$$

Various ρ functions exist in the literature, with one notable example being Tukey's bisquare function, proposed by Tukey (Beaton and Tukey, 1974), which is utilized in this study:

$$\rho_c = 1 - \left(1 - (x/c)^2\right)^3, |x| \le c \le 1, |x| > c. \tag{2.11}$$

3 Variable Elimination

A regression model that incorporates all explanatory variables is not always statistically significant. Consequently, in statistical analysis, choosing the appropriate variables is essential. The multicollinearity of variables in the model is a common issue that can occur. Removing some of the explanatory variables from the model is one strategy to address multicollinearity if it is discovered. It is advised to use variable elimination techniques in these situations. The literature has numerous documented variable removal techniques. To help with variable selection, Furnival and Wilson (1974) developed the leaps and bounds approach, which makes use of data derived by regression. Applying techniques like bootstrap resampling, predicted residual error sum of squares (PRESS), deviance or residual deviance, modified R-squared, or cross-validation can be challenging due to their computational cost, especially when dense criteria are involved. Consequently, these intricate techniques frequently necessitate modifications to computation, rendering them less practical for everyday application. Stepwise elimination is an alternate strategy. Stepwise approaches are helpful for recommending better alternatives, even though they might not always find the optimal model. Backward elimination, one of the stepwise techniques, is used in this study to eliminate variables. In order to increase overall statistical significance and decrease multicollinearity, this approach methodically eliminates the least significant variables from the model.

3.1 Backward elimination method

One popular variable removal strategy in the literature is the backward elimination method. Compared to other approaches, this one is typically less adversely impacted by correlations between regressors (Myers and Montgomery, 2002). In backward elimination, the contribution of each variable to lowering the sum of squared errors (SSE) determines whether it should be kept in the model or eliminated. A variable is eliminated from the model if its contribution to lowering SSE is deemed insignificant. The procedure is iterative: the model is reassessed to identify the subsequent least significant variable following each elimination. Until there are no more variables in the model or all of the remaining variables are statistically significant, this elimination process keeps going. Backward elimination's iterative process makes sure that only variables that significantly affect the model's performance are kept, improving the model's overall statistical robustness and interpretability.

4 Data

Carbon dioxide emissions are a major issue for the planet and its entire people. The literature looks at the following factors when analyzing the factors that directly affect CO_2 emissions: carbon footprint, GDP per capita, fuel and electricity consumption per capita, primary sector GDP percentage, secondary sector GDP percentage, and tertiary sector GDP percentage. Using data especially from Turkey for the years 1973–2021, we will examine CO_2 emissions and the factors influencing them in this study.

Using data from Worldmeters, WDB, and Eurostat, the CO_2 emissions per capita variable is computed as the ratio of total emissions in tonnes to the nation's population. WDB and CEIC data are used to calculate fuel consumption per capita, which is the total amount of fuel consumed by the nation's citizens in kilogram's. Using data from WDB and TURKSTAT (Turkish Statistical Institute), another measure, electricity consumption per capita, is calculated as the ratio of the nation's total electricity consumption in kWh to its population, including industrial units. The carbon footprint variable, which is derived from the Foot print network website on a per hectare basis, is the area of land needed to absorb all CO_2 emissions generated by human activities over the course of a person's lifetime (Pandey, 2011).

A figure known as the Gross Domestic Product (GDP) is computed to quantify the worth of the commodities and services that a nation produces during a given time period. The computed GDP number is divided by the country's population throughout the period to determine GDP per capita. The WDB's 2017 dollar exchange rate is used to standardize the GDP per capita variable in this analysis. The primary, secondary, and tertiary sectors are included in the GDP per capita variable. The primary sector is represented by the GDP's share of agricultural production, while the secondary sector is indicated by the GDP's share of industrial production. Lastly, the tertiary sector is represented by the percentage of income in the service sector, which is obtained from the WDB.

5 Analysis

This study involved an analysis of various factors that may impact CO_2 emissions, such as energy consumption, carbon footprint, oil consumption, GDP, and the distribution of GDP across the primary, secondary, and tertiary sectors. The data gathered for Turkey spanning from 1972 to 2021 are presented in Table 4.1. The CO_2 variable exhibits a mean of 3.2 with a standard deviation of ± 1.1. Energy consumption per capita shows a mean of 1551.1 and a standard deviation of ± 964.2. The carbon footprint presents a mean of 1.3 with a standard deviation of ± 0.4. Electricity consumption per capita has a mean of 1151.8 and a standard deviation of ± 391.1. Lastly, GDP per capita reveals a mean of 5084.2 with a standard deviation of ± 3924.3. The means for the primary (S1), secondary (S2), and tertiary (S3) sectors are 15.4, 27.5, and 50.2, with standard deviations of ± 8.9, ± 2.9, and ± 4.7, respectively. It is important to recognize that these variables are measured in different units, which requires standardization to address discrepancies prior to conducting the analysis.

To investigate the connection between CO_2 emissions and the independent variables, pair wise graphs are illustrated, as depicted in Figure 4.1. The graphs illustrate a linear correlation between CO_2 emissions per capita (CO_2-EPC) and various significant variables: CO_2-EPC, CO_2-CFP (carbon footprint per capita), CO_2-S1 (primary sector as a percentage of GDP), and CO_2-GDP (GDP per capita). Given the identified linear relationships, the linear regression method was deemed suitable for data analysis. Therefore, the regression model is as follows:

$$CO_2 = \beta_0 + \beta_1 EPC + \beta_2 CFP + \beta_3 S_1 + \beta_4 EUO + \beta_5 S2 + \beta_6 S_3 + \beta_7 GDP + \epsilon. \tag{5.1}$$

To test the appropriate hypotheses for the significance of the regression model, the following hypotheses are used:

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0, \quad H_1 : \beta_j \neq 0 \quad \text{for at least one } j. \tag{5.2}$$

Description	Variables	Source
CO2	Carbon Dioxide Emission Per Capita	Worldometers (1973-2014) and World Data Bank (2015-2021)
EPC	Energy Consumption Per Capita	World Data Bank (1973-2014) and TURKSTAT (2015-2021)
CFP	Carbon Foot Print Per Capita	Footprintnetwork (1973-2021)
S1	Sector 1	World Data Bank (1973-2021)
EUO	Oil Consumption Per Capita	World Data Bank (1973-2015) and CEICDATA (2016-2021)
S2	Sector 2	World Data Bank (1973-2021)
S3	Sector 3	World Data Bank (1973-2021)
GDP	Gross Domestic Products Per capita	World Data Bank (1973-2021)

Table 1: Defining Variables

Variables	Observation	Mean	Standard Deviation	Median	Mode	Minimum Value	Maximum Value	Standard Error
CO2	49	3.2	1.1	3.2	1.5	1.6	5.3	0.1
EPC	49	1551.1	964.2	1418.7	1285.4	292.9	3386	137.7
CFP	49	1.3	0.4	1.3	0.6	0.6	2.014	0.1
S1	49	15.4	8.9	14.5	9.5	5.5	36.002	1.3
EUO	49	1151.8	391.1	1109.9	460.1	640.4	1979.4	55.9
S2	49	27.5	2.9	26.8	3.4	22.5	32.9	0.4
S3	49	50.2	4.7	51.1	3.9	38.5	57.2	0.7
GDP	49	5084.2	3924.3	3100.4	2569.7	676.4	12507.8	560.61

Table 2: Descriptive statistics

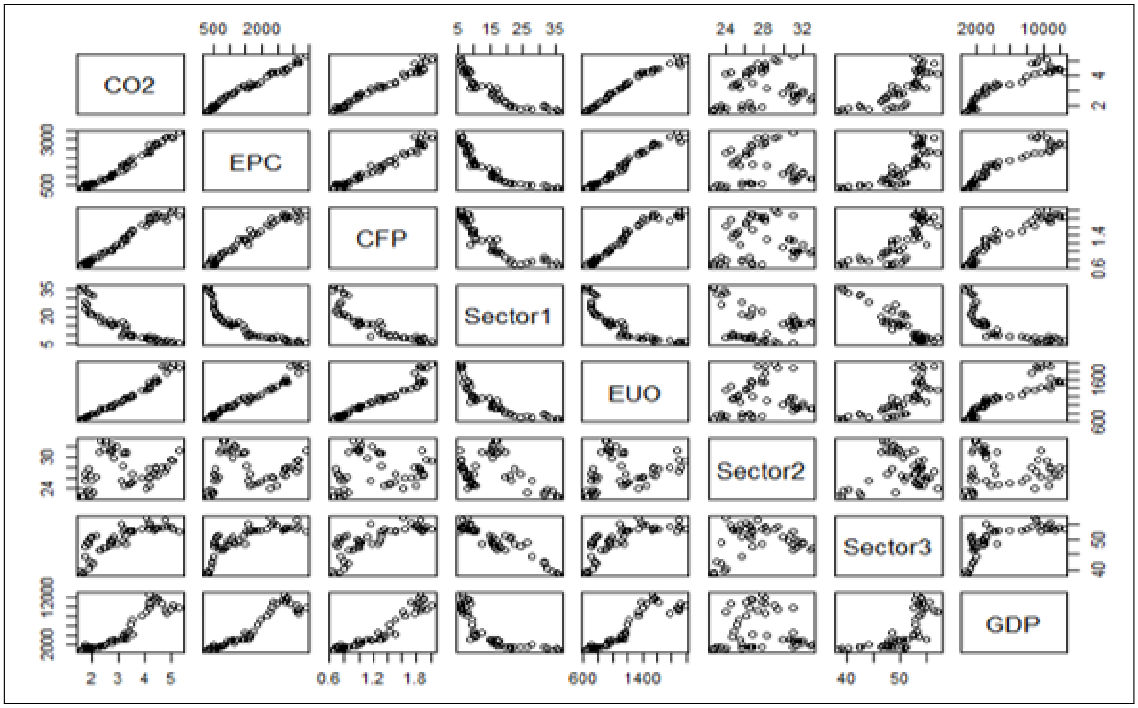

Table 3: Pair wise graphs

The estimation results for the LS, GLS, and RR methods are presented in Table 4.2. The table includes estimates of coefficients, standard errors, p values, and, Durbin-Watson test statistics p-values. While the coefficient estimates differ across the methods, the direction of the relationships remains consistent. The analysis show that EPC and CFP are significant at $\alpha = 0.001$ and positively correlated with CO_2, while S2 is significant at $\alpha = 0.05$ and also positively correlated. Other variables are not significant at $\alpha = 0.001$, 0.05, or 0.01 levels.

Additionally, unlike the other methods, the EUO variable is significant in the RR method. The RR method also has lower standard errors compared to LS and GLS. Furthermore, the RR method has the smallest mean squared error (MSE) value among the methods, indicating its superiority. Not all of the

LS				**GLS**			**RR (Bisquare)**		
Variables	**Coefficients**	**Standard Error**	**P**	**Coefficients**	**Standard Error**	**P**	**Coefficients**	**Standard Error**	**P**
Constant	-3.8056e-16	0.0160	1.000	0.0013006	0.01588027	0.9351	-0.0086863	0.01023	0.40077
EPC	0.6139731	0.1097	0.0001***	0.6363358	0.11303041	0.00001***	0.42975	0.1059	0.0001***
CFP	0.4981489	0.07042	0.0001***	0.4577831	0.06933529	0.00001***	0.52287	0.067961	0.0001***
S1	0.08152727	0.07439	0.279491	0.0761610	0.08288601	0.3635	0.024341	0.07179	0.7363
EUO	0.07090941	0.08078	0.385150	0.0969053	0.08508762	0.2614	0.1579	0.077956	0.049356**
S2	0.06622128	0.02086	0.002843**	0.0628082	0.02536862	0.0175*	0.048868	0.020129	0.019669**
S3	0.05755795	0.04477	0.205824	0.0499284	0.04875495	0.3118	0.036318	0.04321	0.4055
GDP	-0.1725638	0.04339	0.00028***	-0.1743816	0.05236960	0.0018***	-0.13868	0.041873	0.002**
Durbin-Watson P	0.0001***			0.038			0.019		
MSE	0.0742			0.07360			0.0716		
Significance levels 0.05** 0.001***									

Table 4: Estimation results for LS, GLS and RR methods

factors included are significant, even though the RR technique yielded better results. Furthermore, a regression model should satisfy all underlying assumptions. Based on Durbin-Watson test data, Table 4.2 clearly shows that autocorrelation is an issue with all approaches. For every approach, autocorrelation is detected at a significance level of 0.05. The additional assumptions will be evaluated later to handle possible departures from the model assumptions. The LS technique will be used to carry out these tests.

Figure 4.2 illustrates a range of diagnostic plots, including residuals plotted against predicted values, a Q-Q plot for the residuals, standardized residuals compared to predicted values, and Cook's distance utilized for identifying outliers. The plots reveal the existence of outlier observations within the dataset. Table 4.3 analyzes the Breusch-Pagan and Kolmogorov-Smirnov tests to assess the assumptions of ho-

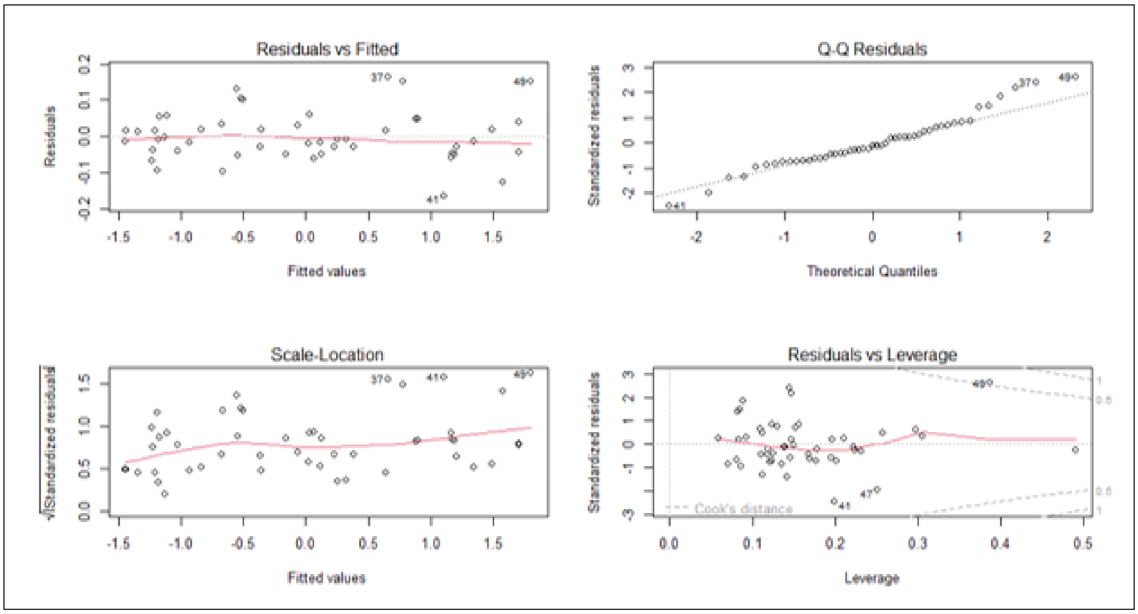

Table 5: Residual analysis

mogeneity and normal distribution of residuals. For the test of variance homogeneity, the hypotheses are stated as follows: H_0:Homoscedasticity is present (the residuals are distributed with equal variance) H_1:Heteroscedasticity is present (the residuals are not distributed with equal variance). The Breusch-Pagan test presented in Table 4.3 shows a p-value of 0.08595, indicating no significant heteroscedasticity. However, the Kolmogorov-Smirnov test, with a p-value of 6.014e-09, suggests that the residuals do not follow a normal distribution. The variance inflation factors (VIFs) for each independent variable are shown

Kolmogorov-Smirnov		**Breusch-Pagan**	
P	6.014e-09	**P**	6.014e-09

Table 6: Tests of homogeneity and normalization

in Table 4.4 in order to evaluate multicollinearity. The table shows that only the S2 variable satisfies the literature's VIF value requirement ($5 \leq VIF \leq 10$). However, based on the VIF values, the other factors show a very high association. To further evaluate multicollinearity, the variance decomposition ratios are shown in Table 4.5. As can be seen, there is a multicollinearity issue among the variables because only the eighth condition index (CI) is higher than thirty. Furthermore, the variance decomposition ratios in the table that are bolded indicate that the variance ratios of the EUO, CFP, and EPC variables, respectively, are greater than 0.5. It may be concluded that these variables are multicollinear, meaning

	EPC	CFP	S1	EUO	S2	S3	GDP
VIF	104.96	43.23	48.23	56.88	3.79	17.47	16.41

Table 7: Investigation of factors affecting variance

that the coefficients most impacted by multicollinearity are their variances. The backward elimination

Variance Ratios								
CI	β_0	**EPC**	**CFP**	**S1**	**EUO**	**S2**	**S3**	**GDP**
1.000	0.000	0.000	0.001	0.001	0.001	0.000	0.001	0.002
2.262	0.000	0.000	0.000	0.001	0.000	0.220	0.000	0.003
2.326	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3.688	0.000	0.001	0.002	0.007	0.005	0.042	0.080	0.011
8.297	0.000	0.008	0.001	0.001	0.078	0.058	0.001	0.438
12.437	0.000	0.001	0.217	0.163	0.045	0.108	0.283	0.200
19.551	0.000	0.127	**0.689**	0.383	0.006	0.157	0.425	0.043
30.807	0.000	**0.863**	0.090	0.445	**0.866**	0.416	0.210	0.303

Table 8: Variance Decomposition Ratios

method using Bayesian information criterion (Schwarz 1978) is used to eliminate variables in order to solve the multicollinearity problem. Table 4.6 presents the outcomes of this elimination technique. We see that the forward elimination approach produced comparable outcomes.

Equation 1			
	Sum of squares	**Residual sum of squares**	**Bayes information criteria**
EUO	0.004243	0.23000	-235.47
S1	0.006614	0.23237	-234.97
S3	0.009099	0.23486	-234.45
S2	0.055506	0.28126	-225.61
GDP	0.087097	0.31285	-220.40
EPC	0.172381	0.39814	-208.58
CFP	0.275530	0.50129	-197.29

Table 9: Backward elimination method results

The most important factors identified are GDP, S2, CFP, and EPC. The model selected through backward

Equation 2			
BayesiInformation criteria: -235.47			
	Sum of squares	**Residual sum of squares**	**Bayes information criteria**
S3	0.01390	0.24390	-236.49
S1	0.01625	0.24626	-236.02
S2	0.10687	0.33687	-220.66
GDP	0.12554	0.35554	-218.02
CFP	0.33189	0.5618	-195.59
EPC	0.83695	1.06695	-164.17

Table 10: Backward elimination method results

Equation 3			
Bayes information criteria: -236.49			
	Sum of squares	**Residual sum of squares**	**Bayes information criteria**
S1	0.00237	0.24628	-239.90
S2	0.10846	0.35237	-222.35
GDP	0.11892	0.36283	-220.92
CFP	0.32737	0.57127	-198.68
EPC	0.85398	1.09788	-166.66

Table 11: Backward elimination method results

elimination, shown in equation 5.3, is given by:

$$CO_2 = \beta_0 + \beta_1 GDP + \beta_2 S2 + \beta_3 CFP + \beta_4 EPC. \tag{5.3}$$

Table 11 displays the outcomes of applying the LS, GLS, and RR procedures to the model found via backward elimination provided in Equation (5.3). Coefficients, standard errors, Durbin-Watson p-values, p-values for significance, and mean squared error (MSE) values are provided for each technique in this table. The coefficients for the independent variables in Equation (5.3) are significant for all approaches, as can be seen.

All standard errors are also comparable. The RR and GLS approaches solve the autocorrelation issue. The RR approach has the minimum MSE, making it the superior model when comparing modeling performance in terms of MSE values.

Equation 4			
Bayes information criteria: -239.90			
	Sum of squares	Residual sum of squares	Bayes information criteria
GDP	0.11680	0.36308	-224.78
S2	0.11980	0.36608	-224.37
CFP	0.34396	0.59024	-200.97
EPC	0.86944	1.11572	-169.77

Table 12: Backward elimination method results

LS				GLS			RR (Bisquare)		
Variables	Coefficients	Standard Error	P	Coefficients	Standard Error	P	Coefficients	Standard Error	P
Constant	-7.52e-16	0.01069	1.000	3.20534	0.01765	0.0001***	3.2092	0.0179	0.0001***
EPC	0.6950	0.05576	0.0001***	0.78132	0.06902	0.0001***	0.7478	0.0699	0.0001***
CFP	0.4668	0.05955	0.0001***	0.50573	0.06751	0.0001***	0.4939	0.0684	0.0001***
S2	0.05448	0.01178	0.0001***	0.05175	0.01644	0.0029***	0.0565	0.0167	0.0014***
GDP	-0.1778	0.03891	0.0002***	-0.21463	0.05382	0.0002***	-0.1641	0.0545	0.0043***
Durbin-Watson P	0.0001			0.099			0.049		
MSE	0.0748			0.08076			0.058		
Significance levels 0.05** 0.001***									

Table 13: Comparison of backward elimination and robust regression methods

Conclusion and discussion

Effectively identifying and addressing the causes of environmental degradation is essential because it has an impact on ecosystems and human health. This study uses statistical models to investigate the relationship between CO_2 emissions, a significant contributor to environmental damage, and other influencing factors. According to the analysis, the RR approach performs better in terms of MSE than LS and GLS. It is noteworthy that whereas the LS and GLS approaches do not consider per capita oil consumption to be relevant, the RR method does. A few unimportant variables were identified by the LS analysis, indicating potential departures from the model's presumptions, including multicollinearity. The model was improved using a backward elimination technique in order to solve this problem. In terms of autocorrelation and MSE values, the RR approach outperformed the other two approaches. According to the report, Turkey's per capita GDP is negatively impacted by rising fuel and power use as well as industrial CO_2 emissions. This suggests that energy-driven industrial growth impedes economic advancement and leads to environmental deterioration. Strategies to reduce environmental harm should be given top priority by policymakers in order to assist Turkey's economic development. The same dataset will be used in subsequent analyses to further hone and validate these results using the RR elimination procedure.

Declaration:

1. **Availability of data and materials:** A a preprint of this manuscript has been publicly uploaded on ArXiv, accessible at the following link: https://arxiv.org/abs/2407.11063.

2. **Conflicts of interest:** The authors declare that they have no conflicts of interest and all the agree to publish this paper under academic ethics.

3. **Fundings:** No funding was received for this work.

4. **Author's contribution**: Author's contribution: All authors contributed significantly to this work. Vijayabalan D was responsible for conceptualization, methodology, and initial draft preparation, while Divya G and Muthumari G handled data curation, analysis, and visualization. Anand gnana selvam S and Kuppuswamy G contributed through investigation, validation, and supervision. Geethamalini S oversaw project administration, provided resources. All authors reviewed and approved the final manuscript.

References

1. Aydin, G. (2015). The development and validation of regression models to predict energy-related CO_2 emissions in Turkey. Energy Sources, Part B: Economics, Planning, and Policy, 10(2), 176-182.

2. Akbostancı, E., Tunç, G. İ., Türüt-Aşık, S. (2018). Drivers of fuel based carbon dioxide emissions: The case of Turkey. Renewable and Sustainable Energy Reviews, 81, 2599-2608.

3. Craig, M. T., Jaramillo, P., Hodge, B. M. 2018. Carbon dioxide emissions effects of grid-scale electricity storage in a decarbonizing power system. Environmental Research Letters, 13(1), 014004.

4. Callendar, G. S. (1938). The artificial production of carbon dioxide and its influence on temperature. Quarterly Journal of the Royal Meteorological Society, 64(275), 223-240.

5. Ekholm, ,N. "On the variations of the climate of the geological and historical past and their causes," Quarterly Journal of the Royal Meteorological Society 27, no. 117 (1901): 1–62.

6. Furnival, G. and Wilson, R. (1974). Regression by leaps and bounds. Technometrics, 16, 499–511.

7. Halicioglu, F. (2009). An econometric study of CO_2 emissions, energy consumption, income and foreign trade in Turkey. Energy policy, 37(3), 1156-1164.

8. Hansen, J., Lebedeff, S. (1988). Global surface air temperatures: Update through 1987. Geophysical Research Letters, 15(4), 323-326.

9. Hampel, F.R. 1978. Optimally bounding the gross-error-sensitivity and the influence of position in factor space. Proceedings of the ASA Statistical Computing Section, ASA, Washington, D.C., 59-64.

10. Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A. 1986. Robust Statistics: The Approach Based on Influence Functions, Wiley, 502 p., New York.

11. Huber, P.J. 1973. Robust regression: Asymptotics, conjectures and Monte Carlo. The Annals of Statistics, 1(5), 799-821.

12. Krasker, W.S. 1980. Estimation in linear regression models with disparate data points. Econometrica, 48(6), 1333-1346.

13. Krasker, W.S. and Welsch, R.E. 1982. Efficient bounded-influence regression.

14. Landsberg, H. E. (1970). Man-made climatic changes: man's activities have altered the climate of urbanized areas and may affect global climate in the future. Science, 170(3964), 1265-1274.

15. Mallows, C.L. 1975. On some topics in robustness. Technical memorandum, Bell Telephone Laboratories, Murray Hill, N.J.

16. Mangır, N., Şahin, Ü. A. (2022). An environmentally extended global multi-regional input–output analysis of consumption-based and embodied import-based carbon emissions of Turkey. Environmental Science and Pollution Research, 29(36), 54813-54826.

17. Manski, C. F. 1991. Regression. Journal of Economic Literature, 29(1), 34-50.

18. Myers, R.H., Montgomery, D.C., 2002. Response Surface Methodology-Process and Product Optimization using Designed Experiments. Wiley, New York.

19. Ozturk, I., Acaravci, A. (2010). CO_2 emissions, energy consumption and economic growth in Turkey. Renewable and Sustainable Energy Reviews, 14(9), 3220-3225.

20. Pandey, D., Agrawal, M., Pandey, J. S. 2011. Carbon footprint: current methods of estimation. Environmental monitoring and assessment, 178, 135-160

21. Plass, G. N. (1956). Carbon dioxide and the climate. American Scientist, 44(3), 302-316.

22. Plass, G. N. (1955). The carbon dioxide theory of climatic change. Tellus, 8(2), 140-154.

23. Rousseeuw, P.J. 1984. Least median of squares regression. Journal of the American Statistical Association, 79(388), 871-880.

24. Rousseeuw, P.J. and Yohai, V. 1984. Robust regression by means of S-estimators. Robust and Nonlinear Time Series Analysis, edited by J. Franke, W. Härdle, and R.D. Martin, 26, 256-274. Lecture Notes in Statistics, Springer, New York.

25. Schwarz, G. (1978). Estimating the dimension of a model. The annals of statistics, 461-464.

26. Svante Arrhenius, "On the influence of carbonic acid in the air upon the temperature of the ground," Philosophical Magazine and Journal of Science, 5th series, 41 (1896): 237–76.

27. Telatar, O. M., Birinci, N. (2022). The effects of environmental tax on Ecological Footprint and Carbon dioxide emissions: a nonlinear cointegration analysis on Turkey. Environmental Science and Pollution Research, 29(29), 44335-44347.

28. Thompson, L. G. (2000). Ice core evidence for climate change in the Tropics: implications for our future. Quaternary Science Reviews, 19(1-5), 19-35.

29. Wang, L., Wang, L., Li, Y., Wang, J. (2023). A century-long analysis of global warming and earth temperature using a random walk with drift approach. Decision Analytics Journal, 7, 100237.

30. Yohai, V.J. 1987. High breakdown-point and high efficiency robust estimates for regression. The Annals of Statistics, 15(2), 642.

Vijayabalan D,
Department of Mathematics
Veltech High-tech Dr. Rangarajan Dr. Sakunthala engineering college, Chennai,
India.
E-mail address: vijayabalantqb@gmail.com

and

Divya G,
Department of Mathematics,
S.A Engineering college, chennai, Tamilnadu,
India
E-mail address: mathsdivya3@gmail.com

and

Muthumari G,
Department of Mathematics,
S.A Engineering college, chennai, Tamilnadu,
India.
E-mail address: muthumarig@saec.ac.in

and

Anand Gnana Selvam S,
Department of Mathematics,
Periyar maniammai university of science and technology-thanjavur,
India.

E-mail address: anandjuslin@gmail.com

and

Kuppuswami G,
Department of Mathematics,
Panimalar Engineering college ,
India.
E-mail address: gkuppuswamiji@gmail.com

and

Geethamalini S ,
Department of Mathematics
Sathyabama Institute of Science and Technology, Chennai-600119, Tamilnadu,
India.
E-mail address: geetha malini@hotmail.com