http://jaa.kbsaa.org

जर्नल ऑफ एकेडमिक एडवांसमेंट

JOURNAL OF ACADEMIC ADVANCEMENT

(Bi-Annual Peer Reviewed Refereed International Journal)

ISSN (Online): 2583-5203 | Volume: 5 | No.: 1 | June, 2026

https://doi.org/10.58574/jaa.2026.v5.i1.02

Publication Impact Factor (I2OR): 4.530 (2026)

RESEARCH ARTICLE

Global Persistence in CO_2 Emissions: Evidence from Panel Data Analysis Across Countries

Shreyasi Chatterjee

Research Scholar (Ph.D-JRF), Dept. of Statistics, Presidency University, Kolkata, India

Dr. Suman Guha

Assistant Professor, Dept. of Statistics, Presidency University, Kolkata, India

Dr. Surajit Sengupta

Assistant Professor, Dept. of Commerce, Bhairab Ganguly College, Kolkata, India

Corresponding Author: Shreyasi Chatterjee (shreyasi.rs@presiuniv.ac.in)

Received: April 27, 2026 | Revised: May 12, 2026 | Accepted: June 04, 2026

Index Terms: Temporal Dependence | Spatial Dependence | Generalised Least Squares

Abstract

This study examines different factors that affect emissions of carbon dioxide (CO_2) from the power sector. A panel dataset of countries has been used, ranging from 2016 to 2023. Given the global nature of emissions, both temporal and spatial dependence are assessed before model estimation. The evidence from autocorrelation analysis indicates strong temporal persistence in emissions. Accordingly, a pooled linear panel regression model that has a first-order autoregressive error structure has been selected. The model is estimated using Generalised Least Squares (GLS). The results show that economic growth, measured by GDP, has a non-linear relationship with CO_2 emissions. But population density does not exhibit a statistically significant association. The findings also suggest that emission dynamics are primarily driven by temporal processes. This highlights the importance of accounting for temporal dependence in global emission analyses.

1 INTRODUCTION

Carbon dioxide (CO_2) emissions from the power sector constitute a significant concern when it comes to global climate change. Although such emissions are often analysed at a global level, they are influenced by country-specific economic and demographic factors. Increasing global interconnectedness raises the possibility of spatial spillover effects, while persistence over time suggests the presence of temporal dependence. Many empirical studies either assume independence across observations or include spatial effects without formally testing their relevance. Such approaches may lead to biased inference and inappropriate model specification. The reason for such inaccuracies is that the underlying dependence structure is not properly identified. Thus, there is a clear need to distinguish between temporal dependence and spatial interaction for developing an appropriate analytical framework.

The present study examines the different factors that affect CO_2 emissions from the power sector. It evaluates both temporal and spatial dependence in a global panel dataset of countries that spans from 2016 to 2023. An appropriate model specification is adopted for empirical analysis keeping in mind the dependence structure that has been identified in this study.

2 LITERATURE REVIEW

Many studies are available that directly or indirectly highlight the relationship between economic activity and its effect on the environment. A key theoretical contribution to this field is provided by Gene M. Grossman and Alan B. Krueger. (1995). An inverted U-shaped relationship has been discussed by them. This framework demonstrates that as the economic growth increases, the degradation in environment also initially increases. As soon as the income reaches a definite threshold, the deterioration in the environment may start to decline. Thus, it indicates a relationship between economic advancement and different environmental outcomes, which is typically non-linear in nature.

Extensive empirical studies have examined the different drivers that results in CO_2 emissions using panel data approaches. For instance, Hassan Al-Mulali (2012) conducted such study for the Middle Eastern economies. He finds that economic growth, energy consumption, and foreign investments are important contributors to CO_2 emissions in that region. Similarly, S. M. Abbes and A. Mohammed (2015) report that there is comparable relationships within BRICS countries, emphasising the role of income and energy use. In contrast, Lijun Du, Chen Wei, and Sheng Cai (2012) show that in rapidly industrialising economies such as China, CO_2 emissions continue to increase along with economic expansion. This suggests that the EKC hypothesis does not hold universally.

An important characteristic of environmental panel data is that the current CO_2 emissions levels are influenced by past values. Thus, the data is said to have temporal dependence. Jeffrey M. Wooldridge (2010) and Badi H. Baltagi (2021) note that serial correlation is a common feature of panel datasets, and this should be properly addressed. If neglected, it may lead to an estimation that will be inefficient. As a result of that, any statistical inference obtained will be unreliable. From a theoretical standpoint, G.S. Maddala and Kajal Lahiri (1992), along with Damodar N. Gujarati et al. (2012), explain that such dependence may arise due to factors such as dynamic adjustment processes, omitted variables, or structural persistence. Emmanuel Alphonus Akpan and Imoh Udo Moffat (2018) illustrate the need to account for serial correlation in their work. They highlight that models can improve the reliability of empirical results by incorporating autoregressive error structures.

In some studies, the possibility of spatial interaction has also been widely discussed. Luc Anselin (1988), James P. LeSage and R. Kelley Pace (2009) provide the formal frameworks for modelling spatial dependence. Those frameworks are foundational in nature. However, empirical confirmation of spatial effects remains mixed. It indicates that such dependence is not universal. Thus, dependence should be evaluated in relation to the specific dataset and context.

Two important features of CO_2 emissions are highlighted in the literature in general. These are persistence over time and the potential for spatial interdependence. It can be seen that in many studies, models have been specified directly without clearly examining these forms of dependence. This underscores the need for a careful assessment of the underlying data structure before selecting an appropriate framework. This is especially needed while doing cross-country analyses of CO_2 emissions.

3 METHODOLOGY

Data Collection and Structure

The analysis is based on secondary data for different countries. It is drawn from the World Development Indicators (WDI) database. The study covers the time period from 2016 to 2023. This provides a perspective on emission dynamics across countries for several years. After data cleaning and filtering, the final dataset consists of 1809 observations spanning 189 countries over 8 years.

The resulting dataset is unbalanced. There are minor inconsistencies in data availability. However, the variation in observations across years is minimal, ensuring that the overall structure of the panel remains stable. The panel framework helps the analysis to capture both cross-country differences and temporal evolution in emissions. The response variable used here is emissions of CO_2 from the power sector, which is measured in Mt CO2eq. Although there are different categories of emissions available in the dataset, this sector contributes largely. This is an important source of global CO_2 emissions. It is also closely linked to economic activity.

The key explanatory variables include - 1) gross domestic product (GDP), expressed in constant 2015 US dollars, and 2) population density, measured as the number of people per square kilometre of land area. These are chosen to reflect the two most important aspects of this study, namely, economic activity and demographic pressures. These are closely linked to emission patterns. The objective was not to analyse different factors of emissions that are specific to certain sectors. Thus, the variables such as energy consumption, industrialisation, renewable energy use, and fossil fuel dependency were not included separately in the model. Here, GDP and population density were selected as the key explanatory variables. They represent the aggregate effect of different other indicators that are of interest in a broader context.

Data Preparation and Transformation

At first, the data was carefully examined for consistency. The variables changed gradually from one year to another. Thus, the last observation carried forward (LOCF) method was used as there was a small proportion of missing observations. Missing values within countries over time were addressed using the LOCF approach. This preserves the continuity of the time series, which is essential for capturing temporal dynamics. The percentage of observations that were imputed was considerably low. Now it is known that LOCF cannot impute missing observations that do not have preceding values. Thus, any remaining missing observations after this step were removed to ensure that the estimation is based on a consistent dataset. All the variables exhibited substantial right skewness. Logarithmic transformations were applied to all key variables. This reduces the skewness. This also allows the coefficients to be interpreted in elasticity terms. This is particularly meaningful in an environmental and economic context.

Descriptive Analysis

A sector-wise breakdown of CO_2 emissions is presented in Figure 1. This provides a descriptive overview of the relative contribution of each sector in the dataset.

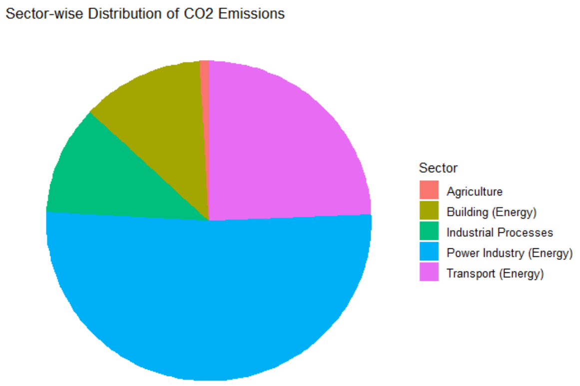

Figure 1: Sector-wise distribution of CO_2 Emissions

It can be clearly seen that the emissions from the power industry account for the largest share of total CO_2 emissions. Transport also contributes a considerable proportion. Emissions from industrial processes and buildings are relatively smaller. Agriculture can be seen to have a minor share in the overall distribution. This pattern highlights that the energy-related activities, particularly power generation, play a dominant role in global emissions. This gives us a strong ground to consider emissions from the power sector as the dependent variables. This is both relevant and empirically justified.

Examination of Dependence Patterns

In the analysis, a key step is to identify the dependence structure of the data. A better understanding of this directly influences the model selection. Two types of dependence were considered: temporal and spatial. Temporal dependence was assessed using residual diagnostics from an initial regression model. At first, country-specific first-lag autocorrelation coefficients were computed and summarised. A significant persistence in the error terms over time within each country was obtained. Then, the autocorrelation function (ACF) of the pooled residuals provides visual evidence of dependence across time. Thus, the assumption that the errors are independently distributed is violated. All of these highlight the need for a model that accounts for temporal dependence. The ACF of the pooled residuals from the model that was estimated using Ordinary least squares (OLS) is presented in the following Figure 2. The plot suggests strong autocorrelation at lower lags, but it decays gradually over subsequent lags.

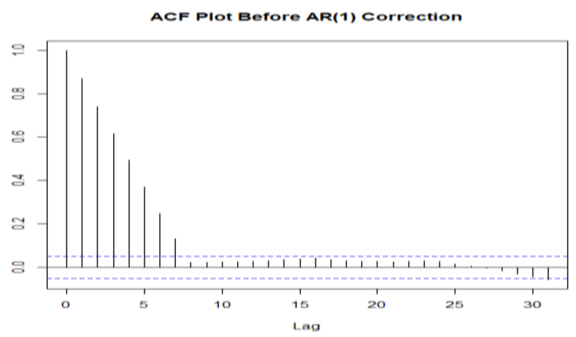

Figure 2: ACF of Pooled Residuals before AR (1) Corrections

The Breusch-Godfrey test was conducted to formally assess this. The results reject the null hypothesis of no serial correlation ($\chi^2 = 1139.3, df = 1, p < 0.001$). This confirms that temporal dependence is present significantly in the data. Moran's I statistic was used to examine the presence of spatial dependence among different countries. A distance-based weight matrix was constructed with the help of country coordinates. Countries that were located within 6000 km were considered neighbours. This threshold was selected so that countries can have sufficient neighbouring connections. Then the spatial matrix was row-standardised. After that, Moran's I was computed for each year.

The results showed very weak positive spatial autocorrelation. For a few years, Moran's I statistics were statistically significant (marginally) at the 5% level. But for the remaining years, the statistics were insignificant. The coefficients remained considerably low throughout the entire study period. After considering the above results, modelling temporal dependence was treated as the primary aim of the study. This insight plays a central role in guiding the selection of the final empirical model for this study.

Model Specification

A pooled linear panel regression model that has a first-order autoregressive error term is specified for the analysis. Different models were initially explored, including fixed effects (FE) model, random effects (RE) model, etc. The Hausman test indicated that the FE model should be used. However, this model with pooled specification was finally chosen as it aligns with the objective of this study. The primary aim was to study the overall global association rather than the effects that are unit-specific.

The model is selected based on the observed dependence structure. The first-order autoregressive specification allows current orders to depend on their immediate past values. There is no strong evidence in the data to support the consideration of higher-order autoregressive structures. The model is estimated using Generalised Least Squares (GLS). This method accounts for the correlation that is present in the error terms over time. Also, estimates obtained using this method are more efficient. The model is written as:

$y_{it} = \beta_0 + \beta_1 x_{1,it} + \beta_2 x_{1,it}^2 + \beta_3 x_{2,it} + \varepsilon_{it}$

...(i)

The error terms can be written as,

$\varepsilon_{it} = \rho \varepsilon_{i,t-1} + u_{it}$.

Here, the logarithm of CO_2 emissions from the power sector for country i in year t is represented as y_{it}. The logarithm of GDP and the logarithm of population density are denoted by the variables $x_{1,it}$ and $x_{2,it}$, respectively. The squared logarithm of GDP is represented by $x_{1,it}^2$. It examines the presence of non-linearity.

The term β_0 represents the intercept. β_1 , β_2 and β_3 These are the coefficients that measure how GDP, the square of GDP, and population density affect CO_2 emissions, respectively. These coefficients are to be interpreted in percentage terms as the logarithm of these variables is being taken. The error term, represented as ε_{it}, captures factors other than those already included in the model that can affect emissions. It follows an AR (1) structure, meaning that current errors depend on past errors within the same country. Here, the parameter ρ shows the strength of this time dependence, and u_{it} represents random shocks.

The empirical properties of the data are now well captured by this model. By accounting for temporal correlation, estimation using GLS yields more efficient estimates of the model coefficients. The estimates are also more reliable than those obtained by methods that assume independent errors.

Model Evaluation

The performance of the specified model was evaluated. The regression equation was estimated using both OLS as a benchmark and GLS to account for temporal dependence. The comparison between the two estimation approaches is based on coefficient estimates, standard errors, and model selection criteria that include the Akaike information criterion (AIC) and Bayesian information criterion (BIC). The Residual diagnostics are further used to assess model adequacy. Thus, the ACF of the pooled residuals were examined, and the Ljung-Box test was applied to evaluate the remaining autocorrelation.

Overall, the evaluation framework ensures that both the estimation method and model performance are assessed in a manner that is consistent with the underlying properties of the data.

4 RESULTS AND INTERPRETATION

The results are based on a linear panel regression model that has a first-order autoregressive error term. The estimation is done using GLS to account for temporal dependence.

Table 1. OLS and GLS Estimates

Model	OLS (β_0)	OLS (β_1)	OLS (β_2)	OLS (β_3)	GLS (β_0)	GLS (β_1)	GLS (β_2)	GLS (β_3)
Parameter	22.2708***	-2.3585***	0.0621 ***	-0.0214	27.2294***	-2.6555 ***	0.0660 ***	-0.0325
Standard Error	2.1107	0.1720	0.0035	0.0158	4.2498	0.3501	0.0072	0.0446

Table 2: Model Fit Statistics

	OLS	GLS
AIC	3897.221	-1570.083
BIC	3923.807	-1538.196

Source: Authors' Computation

Notes: *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

From Table 1 above, the estimated coefficient on GDP is -2.6555, which is statistically significant at the 1 per cent level. The coefficient for the squared GDP term is also statistically significant. It indicates a strong non-linear relationship between GDP and CO_2 emissions. Since the model is specified logarithmically, the estimated coefficients will be interpreted in elasticity terms. The population density has the coefficient of -0.0325. But it is not statistically significant. Thus, there is no meaningful association between CO_2 emissions and population density.

Evidence of temporal dependence is obtained by examining how the residuals are related over time. For each country, the correlation between the current residual and the residual from the previous year was calculated. These correlations are relatively high, with an average value of 0.5262 and a median of 0.6386, indicating that for most countries, the errors in one year are strongly related to the errors in the previous year. In some cases, this relationship is very strong. The maximum value obtained is very close to 1.

This implies that the unexplained part of CO_2 emissions does not change from year to year. It instead follows a persistent pattern over time. In other words, if

Overall, the results indicate that CO_2 emissions are non-linearly associated with economic activity and exhibit strong temporal persistence. But population density does not have a statistically significant effect. To

the model underestimates or overestimates emissions in one year, it may do the same in the following year. This pattern is further supported by the autoregressive parameter that was estimated using GLS. It was found to be 0.993. This value indicates a very high degree of persistence in the error term, which implies that the current errors are strongly influenced by past errors. Overall, these findings provide clear evidence that strong temporal dependence is present in the data. Now, the OLS estimation also yields a similar direction of results.

However, differences are observed in both coefficient magnitudes and standard errors. In particular, the standard errors under GLS are larger than those obtained from OLS, indicating that OLS underestimates variability when temporal dependence is ignored. By accounting for the correlation structure in the error term, the GLS estimation method provides more reliable statistical inference. Model comparison based on the AIC and BIC further supports the GLS specification. The AIC decreases substantially from 3897.221 under OLS to -1570.083 under GLS, with a similar improvement observed in BIC values (from 3923.807 to -1538.196), indicating a significantly better model fit when temporal dependence is taken into account. It was found that the final model with GLS specification was statistically significant overall (Wald Statistic = 524.9912, $p < 0.001$). evaluate the adequacy of the fitted model, diagnostic checks were performed on the residuals from the model estimated using GLS. The ACF of the pooled residuals (Figure 3) shows that although there is a spike at lag 0, the autocorrelation at other lags lies within the confidence bounds (except for a few). This implies that the temporal dependence has been properly addressed. It also indicates that no significant autocorrelation remains in the residuals.

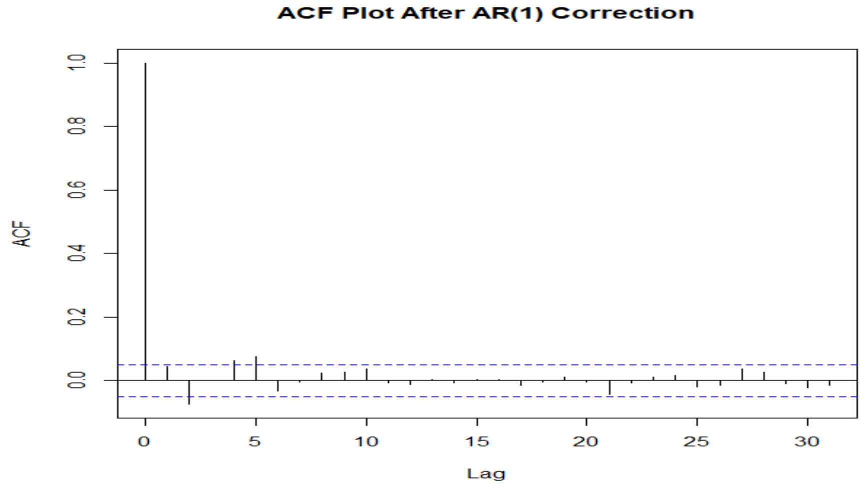

Figure 3: ACF of residuals after AR (1) Correction

The Ljung-Box test indicates that some remaining autocorrelation may still be present. This suggests that the AR (1) structure, though, captures the primary temporal dependence; still, minor correlation may persist.

Overall, the GLS model appears to capture the primary temporal dependence and the main relationships in the data, although some residual autocorrelation and deviations from normality remain.

CONCLUSION

The entire study provides empirical evidence on the factors influencing the emissions of CO_2 from the power sector. The analysis shows that the emissions exhibit strong persistence over time. It indicates that past emission levels have a very important role to play in shaping current outcomes. The findings suggest that variations in economic activity are more relevant in explaining changes in emissions than population density within the power sector.The incorporation of an appropriate error structure is necessary so that the analysis can provide more reliable estimates. It also demonstrates the importance of the estimation approach to align with the underlying characteristics of the data.

The study has its own limitations. The entire analysis is restricted to the power sector. It also includes a limited set of explanatory variables. This work may be extended in future by exploring other sectors. Thus, additional sectors may also be included. Alternative modelling approaches may also be explored to capture the broader dynamics of emissions in a better way.

REFERENCES

Abbes, S. M., & Mohammed, A. (2015). Factors affecting CO2 Emissions in the BRICS countries: A Panel Data Analysis. *Procedia Economics and Finance*, 26, 114-125.

Acemoglu, D., Aghion, P., Bursztyn, L., & Hemous, D. (2012). The environment and directed technical change. *American Economic Review*, 102 (1), 131–166.

Akpan, E. A., & Moffat, I. U. (2018). Modelling the autocorrelated errors in time series regression: A Generalised Least Squares Approach. *Journal of Advances in Mathematics and Computer Science*, 26 (4), 1-15. (p. 1)

Al-Mulali, U. (2012). Factors affecting CO2 emission in the Middle East: A Panel Data Analysis. *Energy*, 44 (1), 564-569.

Arellano, M., & Bond, S. (1991). Some tests of specification for panel data: Monte Carlo Evidence and an Application to Employment Equations. *Review of Economic Studies*, 58 (2), 277–297.

Breusch, T. S., & Pagan, A. R. (1980). The Lagrange Multiplier Test and Its Applications to Model Specification in Econometrics. *Review of Economic Studies*, 47 (1), 239–253.

Bozdogan, H. (1987). Model selection and Akaike's Information Criterion (AIC): The General Theory and its Analytical Extensions. *Psychometrika*, 52 (3), 345–370.

Chamberlain, G. (1982). Multivariate Regression Models for Panel Data. *Journal of Econometrics*, 18(1), 5-46.

Coondoo, D., & Dinda, S. (2008). Carbon Dioxide Emission and Income: A temporal analysis of cross-country distributional patterns. *Ecological Economics*, 65 (2), 375–385.

Dietz, T., & Rosa, E. A. (1997). Effects of population and affluence on CO2 emissions. *Proceedings of the National Academy of Sciences*, 94 (1), 175–179.

Du, L., Wei, C., & Cai, S. (2012). Economic Development and Carbon Dioxide Emissions in China: Provincial Panel Data Analysis. *China Economic Review*, 23 (2), 371-384.

Ehrlich, P. R., & Holdren, J. P. (1971). Impact of Population Growth. *Science*, 171 (3977), 1212 - 1217.

Gourieroux, C., Holly, A., & Monfort, A. (1982). Likelihood Ratio Test, Wald Test, and Kuhn-Tucker test in linear models with inequality constraints on the regression parameters. *Econometrica: Journal of the Econometric Society*, 63-80.

Grossman, G. M., & Krueger, A. B. (1995). Economic Growth and the Environment. *Quarterly Journal of Economics*, 110 (2), 353–377.

Kao, C. (1999). Spurious Regression and Residual-based Tests for Cointegration in Panel Data. *Journal of Econometrics*, 90 (1), 1–44.

Ljung, G. M., & Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65 (2), 297–303.

Müller-Fürstenberg, G., & Wagner, M. (2004). The Environmental Kuznets Curve: A Survey of Recent and Current Research. *Econometric Reviews*, 23 (4), 311-326.

Neath, A. A., & Cavanaugh, J. E. (2012). The Bayesian Information Criterion: Background, Derivation, and Applications. *WIREs Computational Statistics*, 4 (2), 199–203.

Sadorsky, P. (2009). Renewable Energy Consumption and Income in Emerging Economies. *Energy Policy*, 37 (10), 4021-4028.

Sengupta, S., Das, S., & Chatterjee, S. (2026). Influence of Employment and Investment in Circular Economy on the Economic Growth of EU Countries: A Dynamic Panel Data Regression Approach (2008-2024). *Social Science Journal for Advanced Research*, 6 (2), 48–53.

Ser, L. P. (2025). Shahbaz, M., Loganathan, N., Zeshan, M., & Zaman, K. (2015). Does Financial Development Stimulate Energy Consumption? *Renewable and Sustainable Energy Reviews*, 45, 1–10.

Stern, D. I. (2004). The Rise and Fall of the Environmental Kuznets Curve. *World Development*, 32(8), 1419 - 1439.

Ward, M. M., & Leigh, J. (1993). Pooled time series regression analysis in longitudinal studies. *Journal of Clinical Epidemiology*, 46(7), 645-659. https://doi.org/10.1016/0895-4356(93)90038-3

Watson, M. W., & Engle, R. F. (1985). Testing for Regression Coefficient Stability with a Stationary AR (1) Alternative. The Review of Economics and Statistics, 67(2), 341–346. https://doi.org/10.2307/1924737

York, R., Rosa, E. A., & Dietz, T. (2003). STIRPAT, IPAT and Impact: Analytic Tools for Unpacking the Driving Forces of Environmental Impacts. *Ecological Economics*, 46(3), 351–365.