Modeling Global Ecological and Socio-Economic Factors (2015-2023)

IDS702, Fall 2024

Meron Gedrago, Xiangyu Wang, Adil Gazder, Hongyi Duan

Abstract

This study explores the interconnected roles of education, income, and socioeconomic characteristics in shaping global carbon emissions. Using comprehensive global datasets, the research examines how variations in compulsory education duration impact carbon emissions across nations with differing income levels and population sizes. It also investigates the influence of life expectancy and labor force participation on a country's income group, highlighting indirect effects on environmental outcomes.The analysis seeks to uncover patterns and relationships that may provide insights into designing policies that simultaneously address climate change and socioeconomic disparities. By focusing on the interplay of education and income in mitigating carbon emissions, this study aims to contribute to a deeper understanding of sustainable development pathways and their implications for achieving global climate goals.

Introduction

Carbon emissions, primarily in the form of carbon dioxide (CO) released through the burning of fossil fuels, deforestation, and industrial activities, are the primary drivers of global warming. These emissions contribute to the accumulation of greenhouse gases in the atmosphere, trapping heat and causing a rise in global temperatures. The consequences of this warming include more frequent and severe weather events, rising sea levels, and widespread disruption to ecosystems and human livelihoods.

Global warming presents an urgent challenge as its effects are disproportionately felt by low-income nations, which often lack the resources to adapt to climate change. At the same time, high-income countries with historically higher carbon emissions bear significant responsibility for the crisis. This imbalance underscores the necessity for policies that address both environmental sustainability and economic equity. Education plays a crucial role in this effort, equipping individuals with the knowledge and skills to develop innovative solutions, adopt sustainable practices, and influence policy decisions.

Despite growing awareness of the climate crisis, global carbon emissions have continued to rise in recent years. This trajectory threatens progress toward international climate goals, such as those outlined in the Paris Agreement, and underscores the need for deeper exploration of the factors driving emissions and their interconnections with education and income. Addressing these issues

requires understanding how education can mitigate carbon emissions and how income inequality influences both education outcomes and environmental impacts.

Building on these concerns, this study aims to investigate the interplay between education, carbon emissions, and income on a global scale. Specifically, it seeks to answer the following questions:

1. How do National income and population size influence carbon emissions across countries for different compulsory education durations ?
2. How do life expectancy and labor force affect the the income group of a country, which could indirectly affect carbon emission?

Through an analysis of global datasets encompassing socioeconomic and environmental indicators, this study aims to provide insights into the dynamic relationship between compulsory education, carbon emissions, and socioeconomic characteristics of citizens in these countries.

Methods

Data Sourcing and Pre-processing

Our data was a curated dataset from the World Development Indicators (WDI) platform. WDI is the primary World Bank collection of development indicators, compiled from officially recognized international sources. It presents the most current and accurate global development data available, and includes national, regional and global estimates. The database contains 1,400 time series indicators for 217 economies and more than 40 country groups, with data for many indicators going back more than 50 years. We restricted our data to only our variables of interest and associated factors which may affect our research questions.

Variable Selection

The two main outcome variables we use is the total carbon emissions for a country for a given year and the Income group of a country for a given year. When we write about the variables we will use to estimate the two outcome variables, we have the following restriction on what we refer when we speak on our variables

- The Total CO2 Emissions variable is measured in terms of Mega-tonnes of CO2 equivalent (1 million metric tonnes) and is the total annual emissions of carbon dioxide equivalent units from human activities, excluding its use for Land Use, Land Use Change and Forestry (LULUCF). We only restrict the total CO2 emissions to focus only on emissions from direct anthropocentric activities and exclude natural or semi-natural processes.
- Income Group is defined as the classification of a country into one of the of four categories (Low Income (L), Lower Middle Income (LM), Upper Middle Income (UM) and High Income (H)), based on its Gross Net Income per capita (measured in US$) for that given year. Based on dynamic thresholds which vary per year, each country is classified into one of the following four categories.
- National income measures non inflation adjusted and is defined as the difference between Gross National Income and the consumption of fixed capital and natural resources depletion

- Compulsory Education Duration is the total number of years that children are legally obligated to attend school. For the purpose of this analysis, we have converted this variable into two. If the compulsory education years are below or equal to 10, it is classified as 'Low' and classified as 'High' otherwise.
- Labor force comprises people ages 15 and older who supply labor for the production of goods and services during a specified period. It includes people who are currently employed and people who are unemployed but seeking work as well as first-time job-seekers. Not everyone who works is included, however. Unpaid workers, family workers, and students are often omitted, and some countries do not count members of the armed forces.

Model fitting and Evaluation

We chose two different models for our research questions. We used a multiple linear regression to model the carbon emission for countries while we used a multinomial logistic regression model to predict the categories of income group for countries.

- Multiple Linear Regression: We conducted an anova test to measure the significance of the interaction term, and tested for linear regression assumptions through the residuals vs fitted and qq-plot.
- Ordinal Regression (Proportional Odds Logistic Regression): We considered modelling using ordinal regression and multinomial regression, due to improved accuracy modeled the relationship using an ordinal regression model.

Results

Research Question 1: Factors Affecting Carbon Emission

The data as shown below in 'Table 1' shows big differences between countries in carbon emissions, income, education, and population. A few countries produce a lot of carbon emissions, raising the global average, while most emit much less. Income is also uneven, with some very rich countries pulling the average far above what most others earn. Education levels are more even, with most countries requiring about 10 years of school, but improving the quality of education may help poorer nations. Bigger, richer countries tend to have larger populations and produce more carbon, but population size alone doesn't explain the differences. These patterns show that solving problems like climate change and poverty needs different plans for different countries—rich, high-emission countries need to lead the way in cutting emissions and sharing resources to help others grow in a fair and sustainable way.

Table 1: Characteristics of country carbon emission and demography

	Carbon Emission	Compulsory Education	National Income (billion USD)	Population (million)
Median	39.39	10	82.93	17.34
1st Quartile	7.14	9	13.22	4.39
3rd Quartile	447.70	11	944.3	126.7
Mean	1584.21	9.78	2651	391.3

	Carbon Emission	Compulsory Education	National Income (billion USD)	Population (million)

When we further explore the relationship of the four variables, we can see a strong positive association exists between Population and Carbon Emission, indicating that countries with larger populations tend to emit more carbon, likely due to increased industrial and energy demands. Similarly, National Income shows a positive trend with Carbon , suggesting that wealthier countries might contribute more to emissions. For countries with lower Compulsory Education, the relationship between National Income and Carbon Emission is stronger when compared to countries with higher Compulsory Education. Overall, all three variables look to be related to Carbon Emission from the scatterplot.

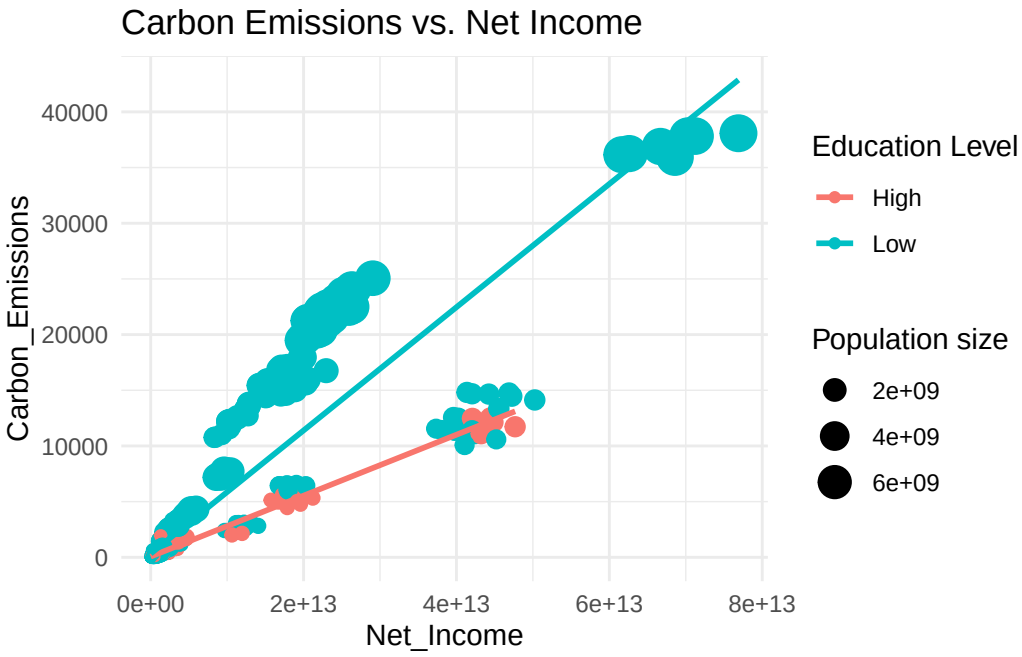
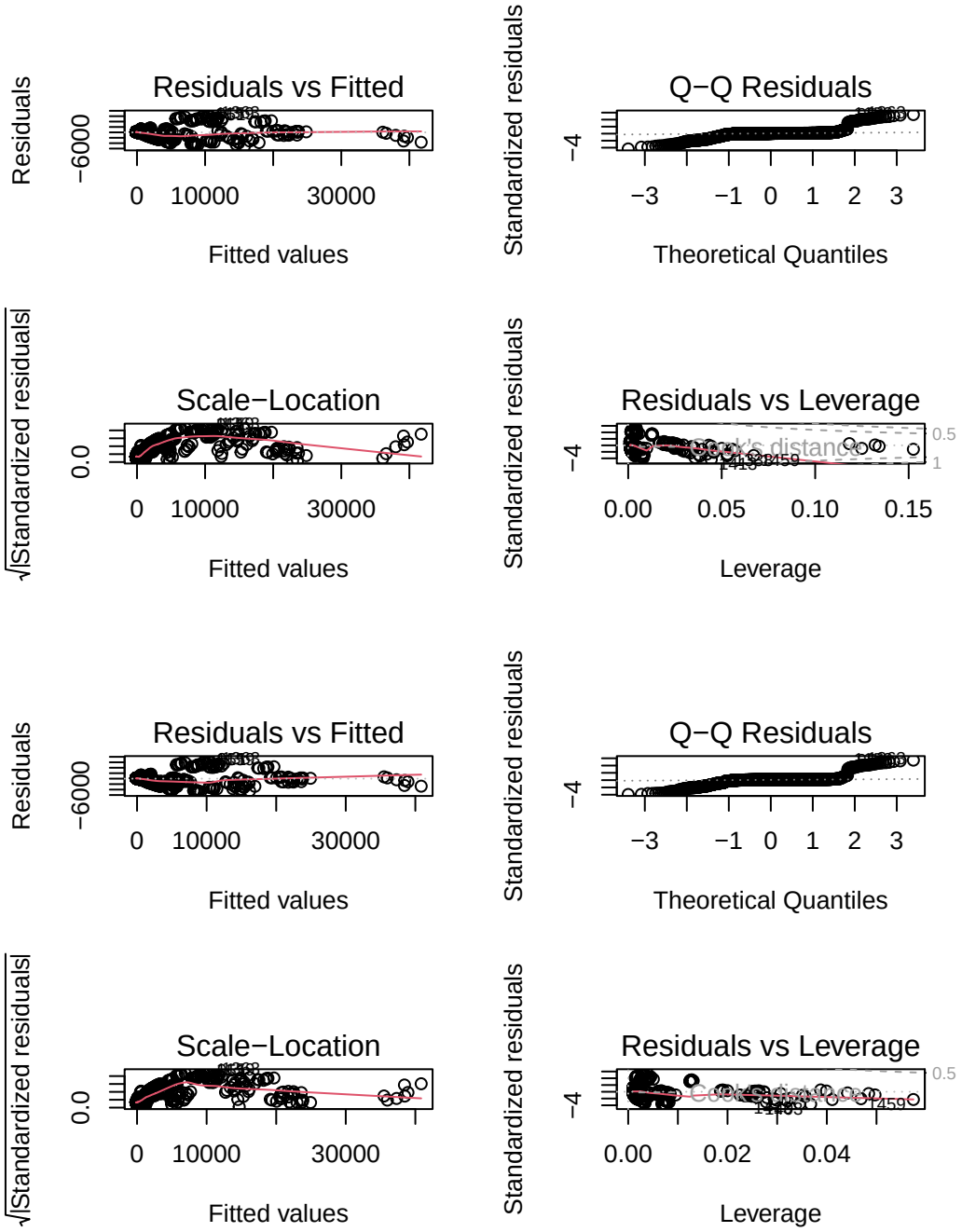


Table 2: Multiple Linear Regression Model Summary

Variable	Estimate	SE	t-value	p-value
Compulsory Education	-110.7	61.44	-1.504	0.133
National Income	2.040e-10	1.034e-11	19.728	<0.001
Population	2.471e-6	4.666e-08	52.966	<0.001
Compulsory Education * National Income	8.714e-11	1.168e-11	7.463	<0.001

The regression model shows a strong overall fit, indicating that 93.52% of the variation in carbon emissions is explained by Compulsory Education, National Income and Population.The regression model shows how factors like education, income, and population affect carbon emissions. It tells us that net income and population size are very important in predicting carbon emissions—higher income and more people lead to more carbon emissions. Education level alone doesn't seem to

have a big effect, but when combined with income, it changes the way income affects emissions. Specifically, countries with lower education seem to have a stronger link between income and carbon emissions. This means that while income and population are the biggest drivers, education can still influence how these factors interact with emissions.

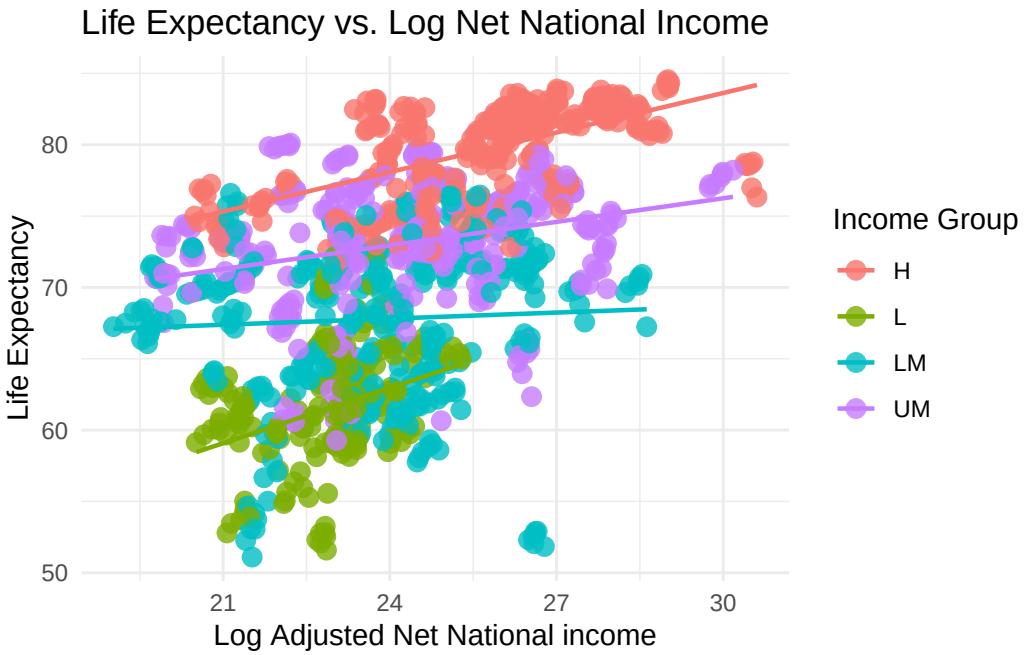


The factor model highlights that the relationship between education years and carbon emissions is nuanced, with certain education levels (like 8 and 8.5 years) being particularly impactful. This

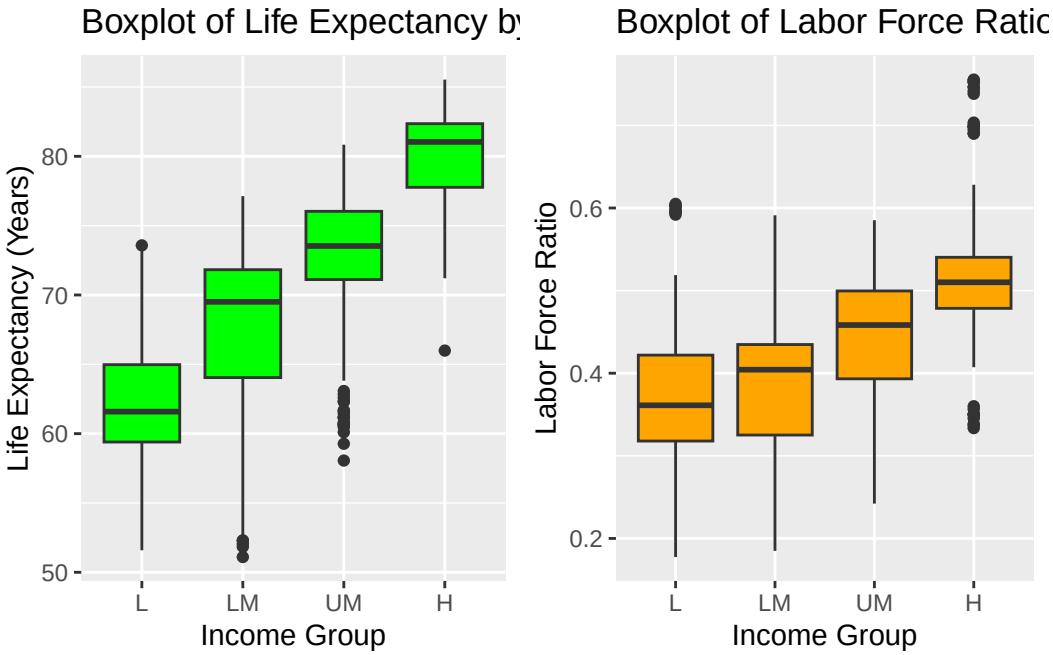
could indicate critical thresholds where education systems or societal behaviors influence emissions. For policy-making, focusing on specific education levels rather than increasing education years linearly might yield better outcomes for carbon reduction strategies.While both models perform well, the factor model provides deeper insights and better predictive power, making it more suitable for nuanced analyses and interventions.

Research Question 2: Factors affecting Income Group

We started with some exploratory data analysis to understand the impact of life expectancy as a function of adjusted net national income and notice that across various income groups, there is a positive correlation between the net national income of a country and the life expectancy of its people, illustrated in the figure below. This relationship highlights the significant role economic resources play in improving the quality of life and access to essential services, such as healthcare, nutrition, and education, which collectively contribute to increased life expectancy.



We further examine the relationship between the countries classified under various income groups and the corresponding life expectancy and labor force ratio for a given year. We again note that countries classified within higher income group brackets tend to have higher labor force ratios, undersocring the critical role of an active labor force in driving economic growth and maintaining a higher income classification.



We run the proportional odds model to try and model the income group classification of a country by splitting the data into training data (randomized 80% subset of the full data) and testing data (remaining 20% subset) to understand how accurately our model performs. The intercepts are detailed in the table below:

Term	Estimate	SE	t-value	p-value
Life Expectancy (Years)	2.547	0.113	22.45	<0.001
Labor Force Ratio	5.471	0.822	6.658	<0.001
L/LM	-1.330	0.365	-3.642	<0.001
LM/UM	1.819	0.361	5.042	<0.001
UM/H	4.313	0.387	11.154	<0.001

We notice the statistical significance of both the life expectancy and labor force ratio terms along with the intercepts for boundaries between the various levels of the income group are statistically significant with an overall accuracy of 66.2% (95% confidence interval of 60.5% - 71.5%) and a no information rate (NIR) of 32.1%. The confusion matrix from the testing data (as % of row totals) have been shown below. We do notice that there seems to be a slight difficulty in classification between Lower Middle (LM) and Upper Middle (UM) levels which can be improved with more extensive training data.

Prediction	L	LM	UM	H
L	72.0%	22.0%	6.0%	0.0%
LM	33.85%	41.54%	24.62%	0.0%
UM	2.30%	29.89%	56.32%	11.49%
H	0.0%	0.0%	11.34%	88.66%

We hence see that life expectancy and labor force ratio are critical indicators of a population's health and economic activity, both of which are strongly tied to a country's income group classification. Higher life expectancy and a larger labor force ratio are often associated with wealthier, more developed economies, while lower values indicate the challenges faced by poorer regions. The model highlights the importance of these factors in understanding socio-economic development, while also acknowledging that classifying income groups is complex and requires further refinement, especially when distinguishing between the middle income groups.

Conclusion

The project aimed to uncover the drivers of global CO2 emissions and explore the relationship between socioeconomic indicators and a country's income classification. Using the World Development Indicators (WDI) dataset, we analyzed variables such as population size, income levels, education years, and urbanization rates to address two key questions. This study offers valuable insights into the global dynamics of CO2 emissions and their link to socioeconomic indicators. By analyzing comprehensive data from the World Bank, we investigated the impact of variables like population size, income, and education on both emissions and income classification.

Through a systematic process of data sourcing, preprocessing, variable selection, and rigorous statistical analysis—including linear and non-linear regression models—we identified meaningful patterns. According to the statistical analysis , we found that income and population size are the most influential predictors of CO2 emissions, with wealthier and larger countries contributing disproportionately. Education, while less directly linked to emissions, plays a crucial role in how National Income impacts Carbon Emission. These results underscore the interplay between economic activity and environmental impact, highlighting the need for tailored strategies to address emissions across different income groups.

Future research should integrate industry-specific data, such as emissions from energy and transportation sectors, and leverage advanced modeling techniques to deepen understanding. This could guide policy interventions aimed at balancing economic development with environmental sustainability.

References

- World Development Indicators (2024). World Bank Group (Databank). Source Data. *(Link)*
- John Creamer (September 2024). United States Census Bureau. "Health Inclusive Poverty Measure in the United States: 2023". *(Link)*
- Our World in data. "CO2 emissions per capita vs GDP per capita, 2022". *(Link)*
- International Energy agency (February 2023). "The world's top 1% of emitters produce over 1000 times more CO2 than the bottom 1%". *(Link)*