Functional Carbon Emissions Analysis

Sam Lee, Everett Andrew
Brigham Young University

February 2025

1 Introduction

Understanding global carbon emissions is essential for evaluating climate change mitigation efforts and forecasting future environmental impacts.

For our project, we will take a closer look at the per-capita carbon emissions of each country across time. We will use functional data analysis to capture smooth temporal trends, quantify the dynamics of emission distributions, and integrate confounding covariates into our models. We will first discuss the scientific questions we propose to answer and their significance. Then, we will describe the data we collected to answer these questions. Finally, we provide an in depth exploratory data analysis (EDA) analyzing the functional nature of this data.

2 Data

Our analysis relies on a comprehensive dataset constructed from three primary sources. The Emissions Database for Global Atmospheric Research (EDGAR) provides annual country-level CO_2 emissions per capita from 1970 onward, serving as our primary response variable.

To evaluate the role of renewable energy in shaping emissions trajectories, we incorporate data from *Our World in Data* (OWID) on renewable energy shares. This dataset records the proportion of total energy consumption derived from renewable sources, including wind, solar, hydro, and biomass, for each country.

Additionally, we integrate macroeconomic and governance indicators from the *World Bank*, which provides key economic measures such as GDP per capita, inflation, unemployment rates, and real interest rates. These variables will serve as controls for understanding emissions patterns in relation to economic development. We also incorporate the World Governance Indicators, which quantify aspects of political stability, corruption levels, regulatory quality, and government effectiveness.

Although this dataset spans over five decades and covers a wide range of countries, it is not without limitations. Earlier years exhibit data sparsity, particularly for developing nations, and certain variables contain missing observations due to inconsistent reporting practices.

We summarize the summary statistics for the comprehensive data set in Table 1.

3 Scientific Questions

Our study seeks to investigate three key questions related to the dynamics of carbon emissions.

A. How does a country's share of renewable energy (over time) affect its carbon emissions, holding other macroeconomic (GDP per capita, Inflation, Unemployment, Real Interest) and political covariates (World Governance Indicators) constant?

B. How can we create an ordinal classification to group countries in terms of carbon emissions to indicate how much and at what rate different groups of countries emit carbon?

C. How well can we predict a country's carbon emissions based on GDPPC, HDI, world governance indicators, real interest rate, and unemployment?

Addressing these scientific questions is crucial for understanding the economic, political, and technological forces that shape global emissions. The role of renewable energy in mitigating climate change has been widely discussed, yet its effectiveness in reducing carbon emissions remains conditional on economic structures, policy enforcement, and governance stability. By disentangling these effects, we will provide empirical insights into whether renewable energy adoption alone is sufficient to drive emissions reductions or whether complementary policies are necessary to achieve meaningful decarbonization.

A classification system based on emissions trajectories offers a more nuanced perspective than static comparisons of carbon output. By grouping countries based on emissions patterns rather than absolute levels, we can identify shared economic and policy characteristics among nations with similar trends.

Furthermore, if emissions trajectories can be reliably predicted, policymakers can take preemptive action to mitigate future carbon growth, while financial institutions can develop investment strategies that account for long-term sustainability risks.

4 Exploratory Data Analysis

We perform an exploratory functional data analysis to identify key trends, assess data quality, and evaluate the distributional properties of carbon emissions over time. All code for data preprocessing and figure creation is documented on this project's GitHub repository. For the purposes of this proposal, although we will augment several macroeconomic indicators, we examine the functional nature of both carbon emissions over time and the share of renewable energy over time across countries, as this covariate is the primary effect of interest as it pertains to our scientific questions.

4.1 Carbon Emissions

We begin this EDA by acknowledging the extreme variance in the data curves, particularly in the late 1900s (see Figure 7). Although we include a functional boxplot (see Figure 8) to show that there are many 'outliers', we propose that this extreme variation may explained by additional covariate information.

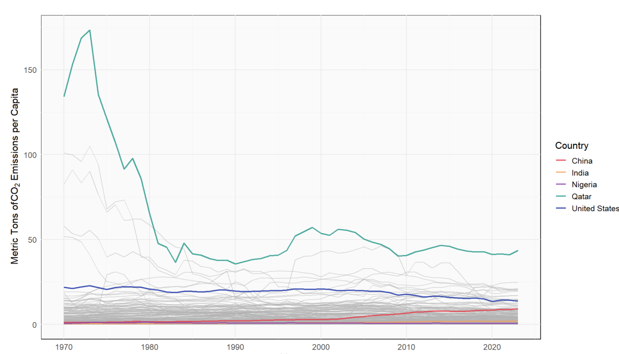

Figure 1: This is a plot of the CO2 emissions per capita for each country. This plot displays a modest trend downwards in emissions from 1970, and it shows a heavy skew towards 0. This was after removing the most extreme outlier, Palau.

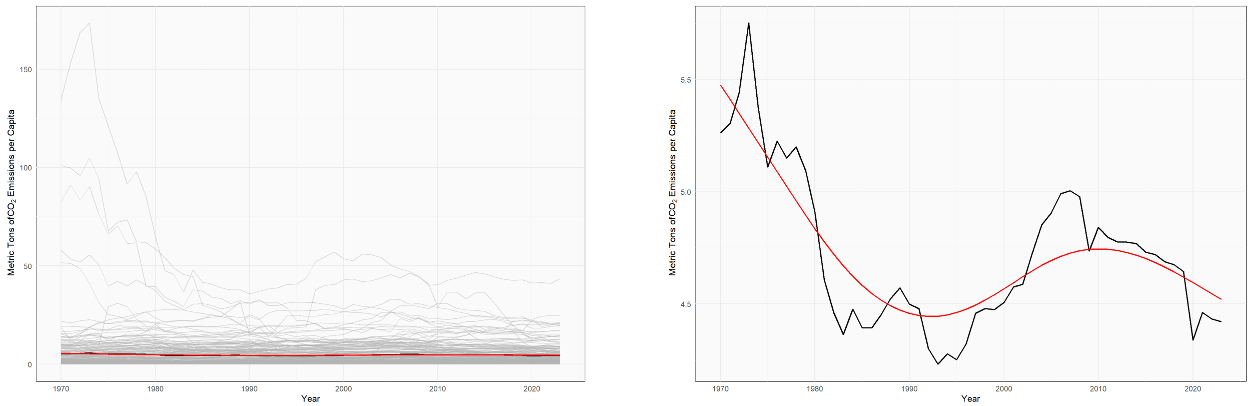

Figure 2: These are our mean estimation plots. We plotted the cross-sectional means (black) and penalized smooth splines (red) against the data first, but it was difficult to see the trends due to the intense skew. We then isolated the cross-sectional mean line and penalized smooth spline curve on their own plot to get a better idea of what's happening. The right figure was estimated using penalized regression splines, using GCV to select the penalty parameter.

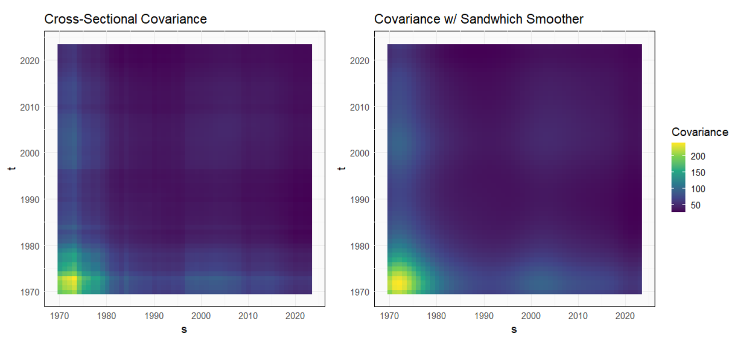

Figure 3: These plots show the cross sectional covariance and sandwich-smoothed covariance of our carbon emissions across years. We used `knots = 10` to get a more smoothed estimate.

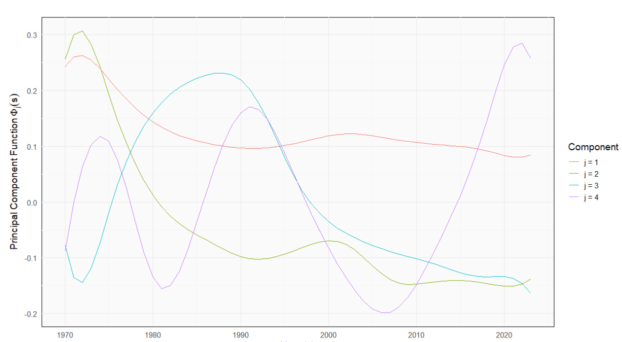

Figure 4: First four principal component functions, showing the 4 most dominant patterns of variation in the data. The total variance explained by these principal components is 0.987.

4.2 Shares of Renewable Energy

In analyzing this data, we found that the data frequency in the renewable energy shares across country was considerably sparse for certain years. Figure 10 shows how many curves (countries) we have data for by year. As we can see, we have much less data available prior to 2000, which will affect our analysis. As a result, we proceed with this section of the EDA using sparse functional data techniques.

Figure 5: Mean Estimation: Cross Sectional (black) vs Penalized Smooth Splines (red). Penalized regression splines were fit using GCV.

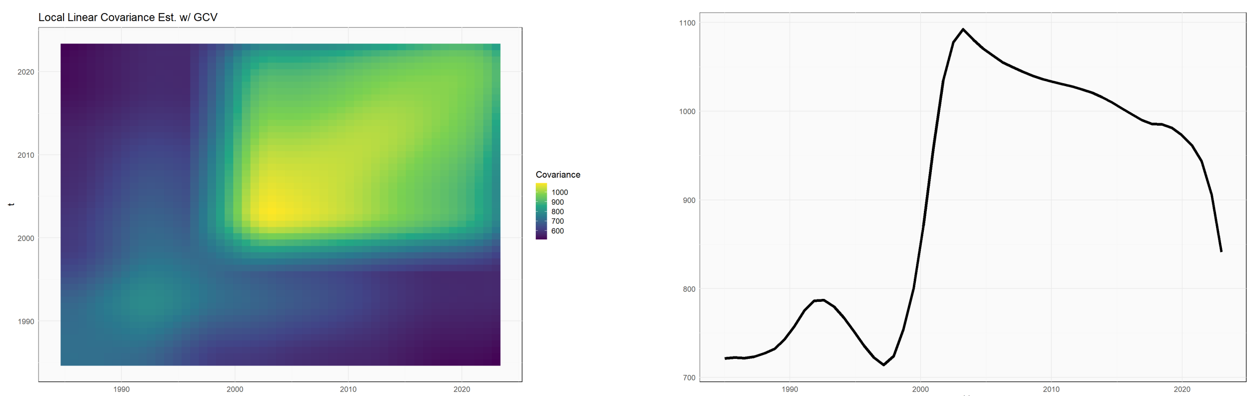

Figure 6: Local linear regression was fit to estimate the smoothed covariance. Epanechnikov kernel was used using GCV to choose bin width.

A Tables

Table 1: Summary Statistics of the Carbon Emissions Dataset

Variable	**Min**	**1st Quartile**	**Median**	**Mean**	**Max**
Key Variables					
CO_2 Emissions (Metric Tons per Capita)	0.0022	0.6509	2.3879	5.1771	178.2595
GDP per Capita	0.4942	4.4055	13.1340	21.4102	179.2007
Renewable Share (%)	0.00	2.34	17.37	30.85	100.00
Human Development Index (HDI)	0.2120	0.5410	0.6960	0.6703	0.9670
Inflation (%)	−31.57	1.82	4.59	35.46	26 762.02
Interest Rate (%)	−97.69	2.13	5.88	5.97	139.96
Unemployment Rate (%)	0.04	4.05	6.66	7.98	38.80
Governance Indicators					
Corruption Index	−1.85	−0.81	−0.26	−0.03	2.46
Government Effectiveness	−2.44	−0.76	−0.15	−0.01	2.47
Political Stability	−3.31	−0.67	0.04	−0.04	1.96
Rule of Law	−2.59	−0.81	−0.19	−0.04	2.12
Regulation Quality	−2.55	−0.72	−0.14	−0.01	2.31
Voice and Accountability	−2.31	−0.87	−0.02	−0.05	1.80

Note: Some variables have missing values, which are not displayed in this table.

B EDA Figures

B.1 Additional EDA Figures for the Carbon Emissions Data

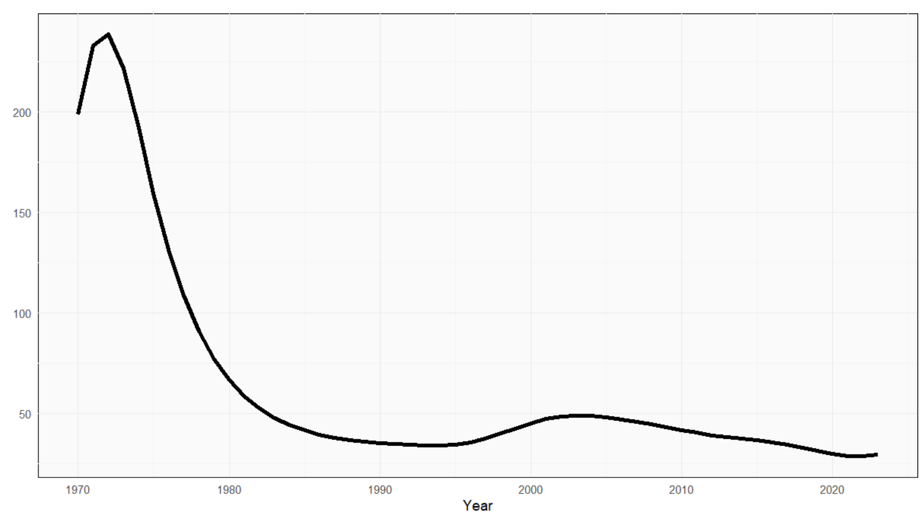

Figure 7: Carbon Emissions Variance Over Time

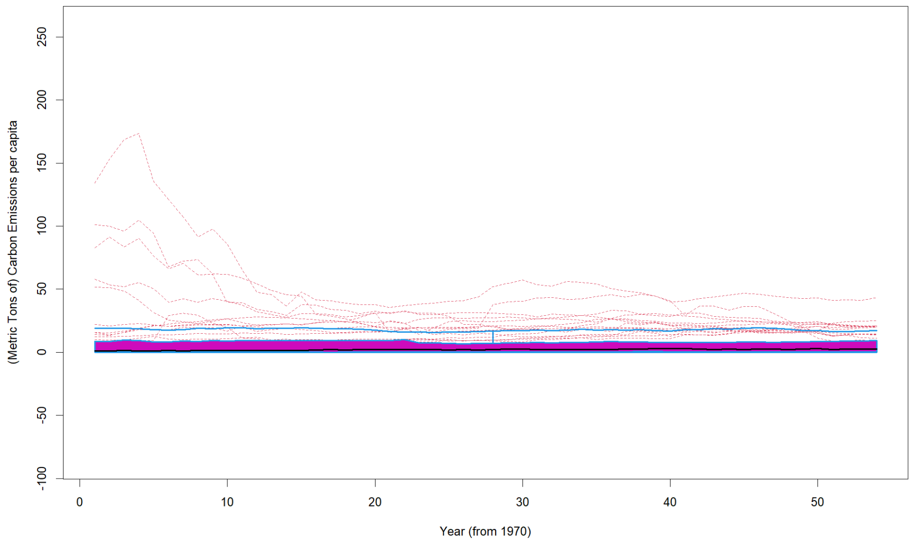

Figure 8: Carbon Emissions Functional Boxplot

B.2 Additional Figures for the Renewable Energy Shares Data

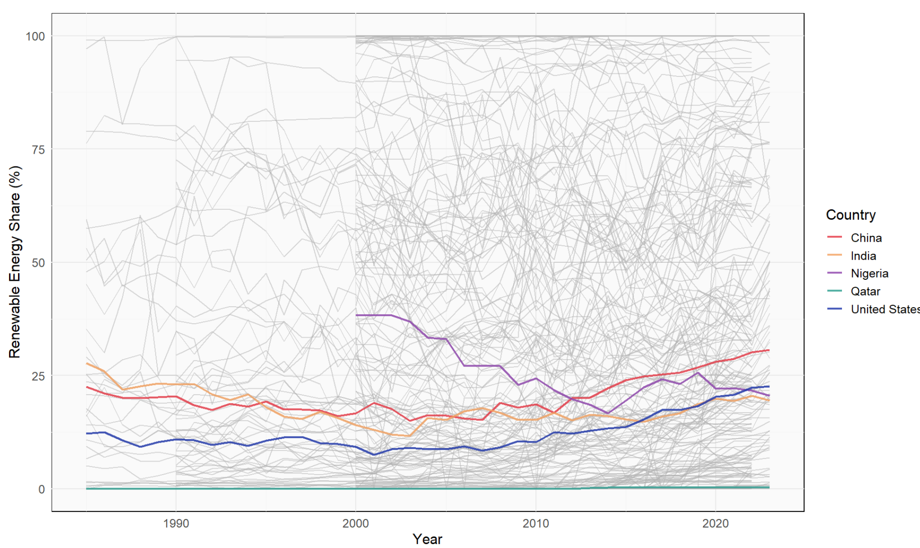

Figure 9: Renewable Energy Shares across countries over time. Specific countries (USA, China, India, Nigeria, and Qatar) are emphasized.

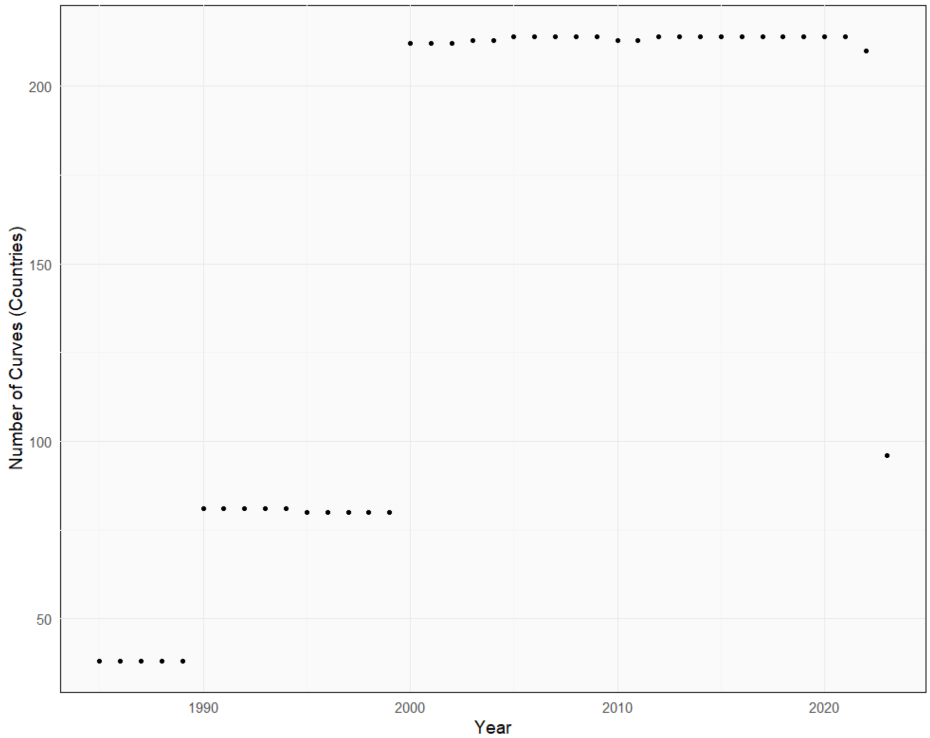

Figure 10: Data curves (number of countries) available per year.

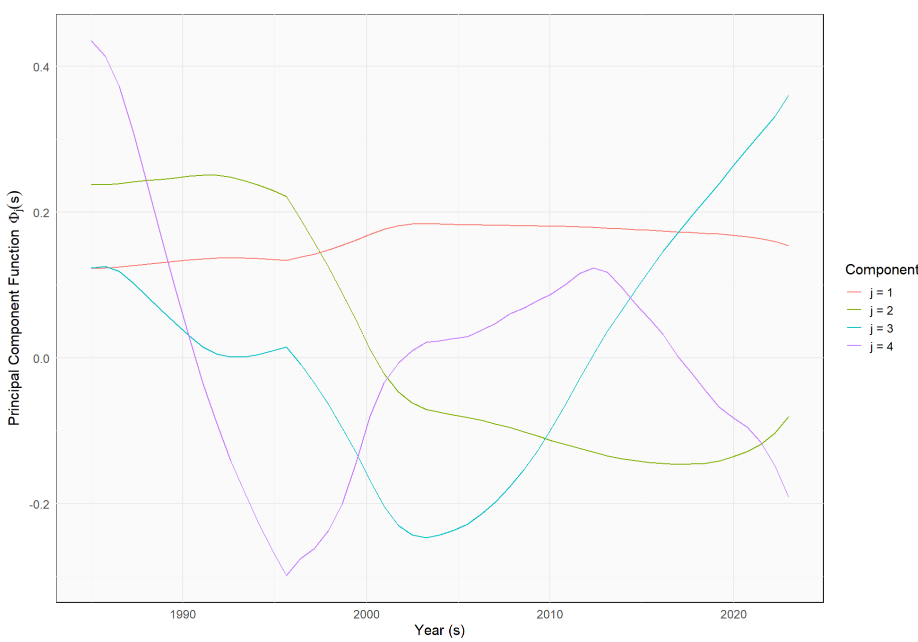

Figure 11: A plot of the first four principal component eigen functions are used extracted from the smoothed covariance estimate computed in Figure 6.