A Functional Carbon Emissions Analysis

Sam Lee, Everett Andrew
Brigham Young University

April 2025

Abstract

We investigate the temporal relationship between renewable energy consumption and per-capita carbon emissions using a historical functional linear regression model. Leveraging data from 178 countries between 1990 and 2023, we model carbon emissions as smooth functional responses influenced by past values of renewable energy share and a set of macroeconomic covariates. To enforce temporal causality, we impose a historical constraint that restricts the support of covariate effects. We find that recent increases in shares of renewable energy consumption are often associated with modest short-run increases in emissions, perhaps due to startup costs, transitional energy expansion, and infrastructure lags. However, renewable energy investments made further in the past show negative long-run effects on present-day emissions, on average.

1 Introduction

Understanding global carbon emissions is essential for evaluating climate change mitigation efforts and forecasting future environmental impacts. For our project, we take a closer look at per-capita carbon emissions over time and the intertemporal consumption of renewable energy across countries. We use functional data analysis to capture smooth temporal trends, and quantify the dynamics of how carbon emissions data are affected by renewable energy consumption.

1.1 Motivation

Figure 1 illustrates the global distribution of log CO_2 emissions per capita in 2023. Interestingly—and somewhat counter to popular narratives—it is the most economically developed countries that continue to exhibit the highest per-capita emissions. Notably, North America, Western Europe, and Australia appear among the top emitters, while emissions across much of sub-Saharan Africa and Southeast Asia remain significantly lower.

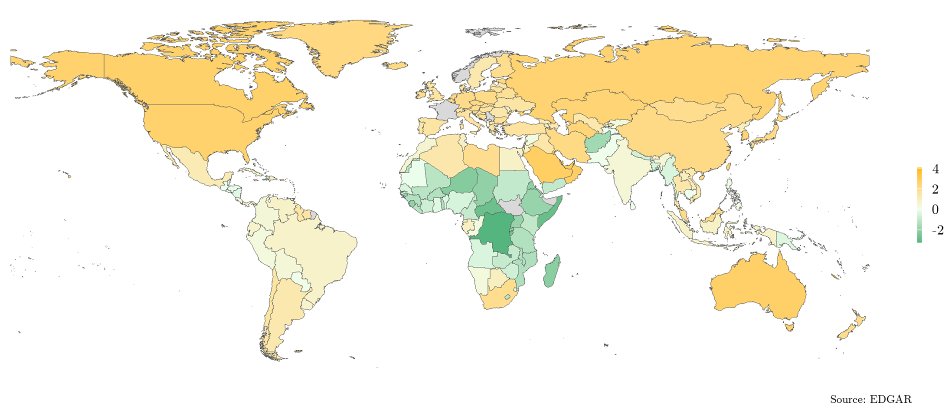


Figure 1: Log per capita metric tons of CO_2 emissions by country in 2023.

Nonetheless, we see a highly heterogeneous distribution of carbon emissions across countries. This has motivated pressing questions about what factors drive these disparities. How do they evolve over time? What types of energy and economic policies lead to long-run reductions in carbon emissions?

The remainder of the paper is structured as follows. In Section 1.2 we discuss a recent paper in the FDA literature that motivates this research. We state our research question in Section 1.3, describe the data sources and key variables, preprocessing steps in Section 1.4, and introduce the historical functional linear regression model in Section 1.5 where we give the theoretical motivation. Section 2 explains how we integrate the historical functional linear model into our analysis and details the estimation procedure. In Section 3, we present the empirical analysis, including model comparison, selection of the lag parameter, and interpretation of our main effect of interest. We discuss the substantive implications of our findings and conclude in Section 4. Additional exploratory analyses, figures, and additional R code are provided in the Appendix.

1.2 Related Literature

Recent work in applied statistics has leveraged functional data analysis to study the evolution of carbon emissions over time. Functional models are well-suited for this task, as they allow researchers to treat a country's emissions profile as a smooth curve rather than a collection of discrete observations.

To the best of our knowledge, the most recent published paper using FDA for the purpose of studying carbon emissions is Elayouty and Abou-Ali (2023), who employ a function-on-function linear model to estimate how electricity consumption and GDP predict log CO_2 emissions over time.

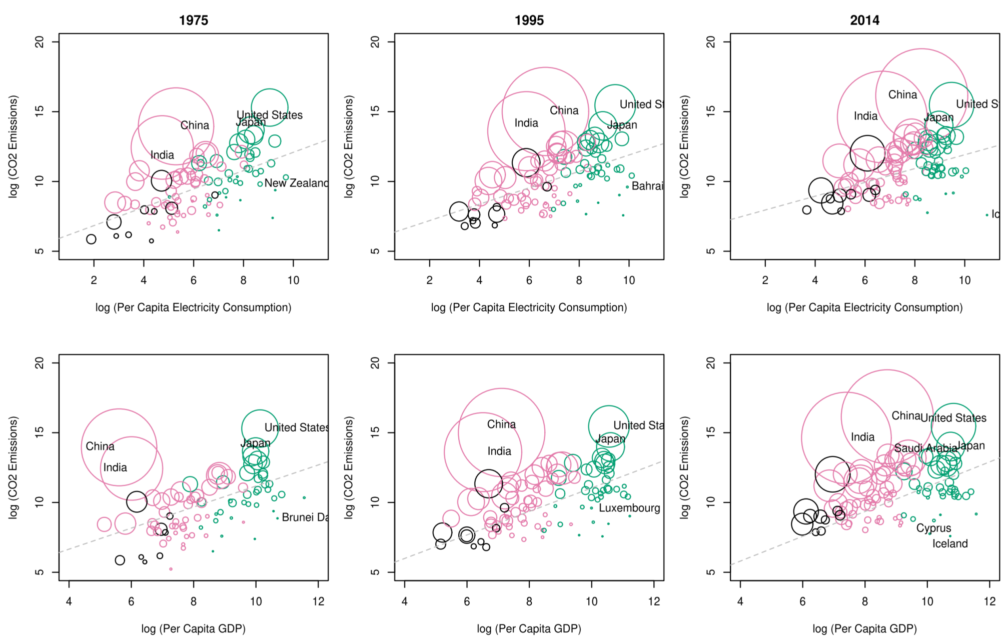


Figure 2: Functional regression of carbon emissions using electricity consumption and GDP. Source: Elayouty and Abou-Ali (2023).

These authors find that GDP and electricity consumption are dominant functional predictors of carbon emissions. However, their model does not impose a temporal ordering between predictor and response variables, implicitly allowing future values of electricity or GDP to influence past or present emissions. We give the critique that this lack of temporal structure makes interpretation difficult for policy evaluation and introduces unnecessary noise into the model fit. Our paper builds on this insight by introducing a historical functional linear model (HFLM) that explicitly constrains the support of covariates to precede the emission response in time.

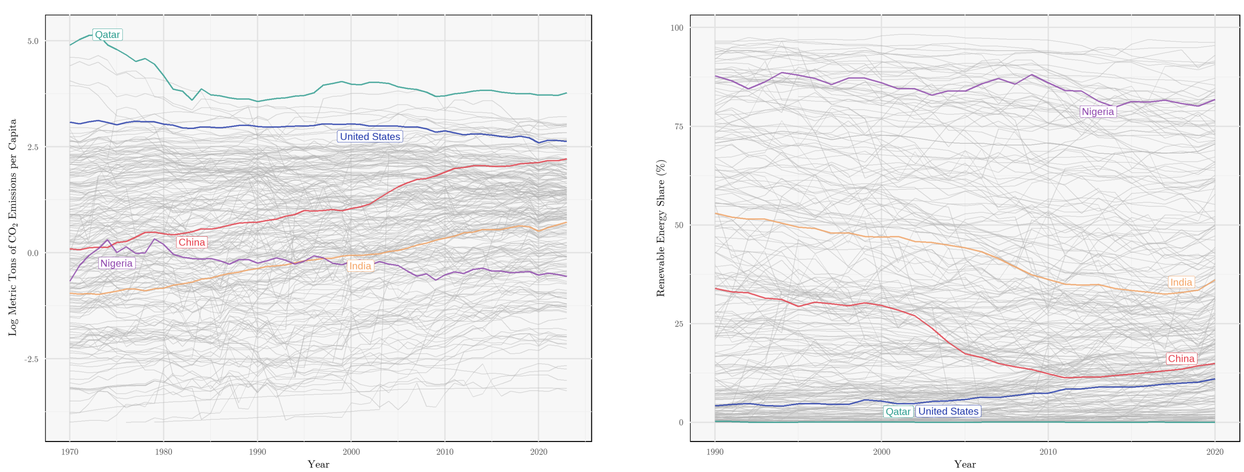
1.3 Research Question

A central challenge in environmental economics is understanding how national energy policies translate into long-run changes in carbon emissions. While many countries have increased the share of their renewable energy consumption, the timeline over which these efforts lead to measurable reductions in emissions is difficult to measure, since often we do not observe the immediate payoffs.

We seek to answer the following question: *To what extent does a country's historical share of renewable energy consumption influence its current carbon emissions, after adjusting for macroeconomic and political covariates, and how do these effects linger over time?*

1.4 Data

Our analysis relies on a comprehensive dataset constructed from two primary sources. The Emissions Database for Global Atmospheric Research (EDGAR) provides annual country-level metric tons of CO_2 emissions per capita from 1970 onward, serving as our primary response variable (see Figure 3a) (Commission, 2023).



(a) Log CO_2 emissions per capita (1970–2023).

(b) Renewable energy shares across countries.

Figure 3: Carbon emissions and renewable energy trends across countries.

To evaluate the role of renewable energy consumption in shaping emissions trajectories, we incorporate data from the *World Bank* on renewable energy shares (Bank, 2023). This dataset records the proportion of total energy consumption derived from renewable sources, including wind, solar, hydro, and biomass, for each country (see Figure 3b).

Additionally, we integrate macroeconomic and governance indicators from the *World Bank*, which provide key economic measures such as GDP per capita, inflation, unemployment rates, and real interest rates. These variables 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 (Bank, 2023).

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 countries, and certain variables contain a large amount of missing data. However, covariates with a small fraction of missing data (e.g.

%Missing $< 10\%$ as we define it) have missing values imputed using a functional mean imputer[1]: We use a cross-sectional mean of all the remaining curves at a given time point to fill in the missing data point at that given time point within a country. We summarize the summary statistics for the comprehensive data[2] set in Table 1. Our final analysis includes 178 countries.

Table 1: Summary Statistics of the Carbon Emissions Dataset

Variable	**Min**	**1st Quartile**	**Median**	**Mean**	**Max**	**%Missing**
Key Variables						
$\log CO_2$ Emissions (Metric Tons per Capita)	−6.0992	−0.6651	0.7331	0.4728	5.1559	0.00
GDP per Capita	0.4942	4.4055	13.1340	21.4102	179.2007	0.00
Renewable Share (%)	0.00	5.60	21.62	31.44	98.30	0.36
Human Development Index (HDI)	0.2120	0.5410	0.6960	0.6703	0.9670	7.27
Inflation (%)	−31.57	1.82	4.59	35.46	26 762.02	1.32
Interest Rate (%)	−97.69	2.13	5.88	5.97	139.96	39.07
Unemployment Rate (%)	0.04	4.05	6.66	7.98	38.80	41.70
Governance Indicators						
Corruption Index	−1.85	−0.81	−0.26	−0.03	2.46	0.76
Government Effectiveness	−2.44	−0.76	−0.15	−0.01	2.47	1.06
Political Stability	−3.31	−0.67	0.04	−0.04	1.96	0.36
Rule of Law	−2.59	−0.81	−0.19	−0.04	2.12	0.00
Regulation Quality	−2.55	−0.72	−0.14	−0.01	2.31	0.99
Voice and Accountability	−2.31	−0.87	−0.02	−0.05	1.80	0.00

Note: The final column reports the percentage of missing data for each variable. In practice, since GDP is heavily right-skewed, we log transform it.

1.5 Historical Functional Linear Model

Traditional function-on-function models would not be an appropriate way to answer our research question, since they would not account for significant differences in years between renewable energy share and carbon emissions. For example, a country's renewable energy share in 2020 could not possibly affect its carbon emissions in 1992. Including future years to predict a phenomenon in a current year simply adds more noise to the model as opposed to any real signal[3].

To answer this question, we will use a derivation of function-on-function analysis called the historical functional linear model, proposed by Nicole Malfait and James O. Ramsay (Malfait and Ramsay, 2003). We summarize their procedure as follows. Let 0 and T indicate initial and final times for a set of records $y_i(t)$, and let δ indicate a time lag, beyond which we conjecture that there is no feed-forward type influence of $x(s)$ on $y(t)$. That is, $y(t)$ is influenced by $x(s)$ for $s_o(t) \leq s \leq t$, with $s_o(t) = max(0, t - \delta)$. We will assume that $x(s)$ influences $y(t)$ linearly according to the following model that integrates this influence from $s_o(t)$ to t:

$$y_i(t) = \beta_0(t) + \int_{s_0(t)}^{t} x_i(s)\beta_1(s,t)ds + \varepsilon_i(t), \quad t \in [0, T] \tag{1}$$

This model is much like traditional function-on-function models, but it has one key difference. The historical functional linear model includes a constraint $s \leq t$ to ensure that the time points of the explanatory factors (s) do not come after the time points of the response (t). It also includes a lag parameter, δ, which we

[1]Alternatively, it may be more appropriate to integrate the principle components of these sparse covariates—we fail to do so here since FPCA does not integrate well with `pffr`.

[2]Since Unemployment and Interest Rate data are not well maintained on the global scale ($\sim 40\%$ sparsity), we omit these functional covariates from this analysis.

[3]As can be seen in our empirical analysis, the functional R-squared of a linear model that uses future covariates to predict a current response fits better. However, this improvement in R-squared does not imply any significant increase in prediction power.

estimate using context and cross-validation from the data. This parameter will be a threshold that defines how far ahead a country's carbon emissions can be explained by its past renewable energy consumption before the effect peters out. These constraints produce a triangular/trapezoidal support that allows us to phase out impossible effects between years.

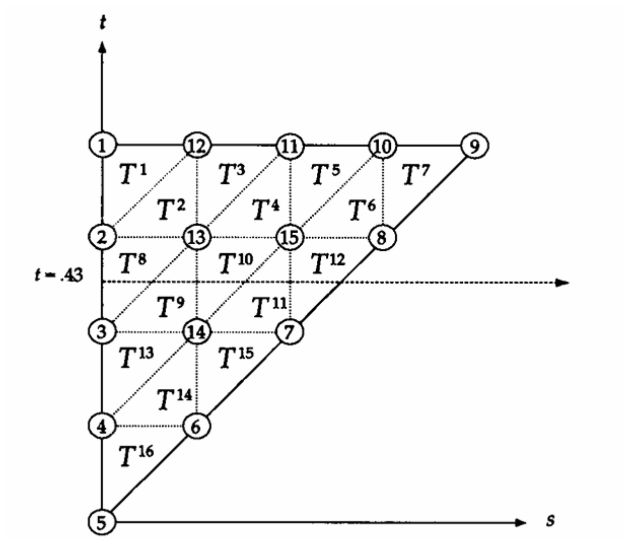


Figure 4: Triangular Support (Malfait and Ramsay, 2003)

2 Methods

To estimate the effect of renewable energy consumption and related covariates on carbon emissions over time, we adopt this historical functional linear model with a penalized spline basis function expansion. We model the time-varying relationship between a country's past renewable energy profile and its present carbon emissions, while accounting for lag constraints. Intuitively, macroeconomic activity that occurs in a country at year t can *only* affect carbon emissions in year s where $s \leq t$.

Specifically, we approximate each bivariate coefficient surface $\beta_k(s,t)$ using a tensor product of penalized B-spline basis functions (p-splines) in both the history and response directions. Letting $\{\phi_m(s,t)\}_{m=1}^{M}$ denote the bivariate basis functions defined over the triangular support $\{s \leq t\}$, we express the model as:

$$\begin{aligned} \log y_i(t) &= \beta_0(t) + \sum_{m=1}^{M} b_{1m} \int_{s_0(t)}^{2020} x_{1i}(s)\phi_m(s,t)\,ds \qquad (2) \\ &\quad + \sum_{k=2}^{10}\sum_{m=1}^{M} b_{km} \int_{s_0(t)}^{T_k} x_{ki}(s)\phi_m(s,t)\,ds + \varepsilon_i(t), \quad t \in [1990, 2023] \\ s_0(t) &:= \max(0, t-\delta) \qquad (3) \end{aligned}$$

Here, $y_i(t)$ denotes the (log-transformed) per capita carbon emissions for country i in year t. The functional predictors $x_{ki}(s)$ represent time-indexed covariates for each country, where $k=1$ corresponds to renewable energy share and $k=2,\ldots,10$ correspond to additional economic and governance controls such as HDI, inflation, and corruption.

Importantly, our functional response is defined over the space $\mathcal{S} = [1990, 2023]$. However, the functional space for which most of our covariates are defined over are a subset of $\mathcal{S}$, denoted as T_k for each kth covariate in the model. This is not a problem, however—although our estimates are less precise for the years 2021-2023 (since we do not have reliable renewable energy consumption data for those years; see Figures 15 and 5), we can still fit those "future" emission values using the past values from 2020 and before.

3 Data Analysis

In this section, we present our most important empirical results as well as a summary of the R code to replicate our results. While this section gives the appropriate code snippets, we recommend readers refer to the project's GitHub repository for more detailed documentation.

3.1 Function-on-function vs. Historical Functional Linear Regression Models

To estimate the model, we use the `pffr()` function from the `refund` package in `R`. For each functional covariate, we use penalized splines in both the s and t dimensions, with smoothness enforced by second-order difference penalties applied separately along the s and t directions.

The historical constraint $s \leq t$ is imposed using the `limits` argument within the `ff()` term, allowing us to restrict influence to a lagged time window $[t - \delta, \min(t, 2020)]$. The definite integrals in the model are approximated using a Riemann sum integration technique. This procedure yields smoothed coefficient surfaces $\hat{\beta}_k(s,t)$ for each covariate, allowing us to interpret how both renewable energy and macro-political covariates influence carbon emissions over time.

We present the R code for estimating our historical model:

```
#Time Supports for the response and covariates
co2.s <- 1990:2023
s <- 1990:2020
wgi.s <- c(1996, 1998, 2000, 2002:2023)

...

#Time lag
delta <- 20
#Integration limits for economic covariates
cov.limits <- function(s, t){
  s >= pmax(1990, t-delta) & s <= t
}
#Integration limits for WGI covariates
wgi.limits <- function(s, t){
  s >= pmax(1996, t-delta) & s <= t
}
model <- pffr(
  carbon ~
  #Functional predictors
    ff(energy, xind = s, yind = co2.s, limits=cov.limits) +
    ff(gdp, xind = s, yind = co2.s, limits=cov.limits) +
    ff(hdi, xind = s, yind = co2.s, limits=cov.limits) +
    ff(inflation, xind = s, yind = co2.s, limits=cov.limits) +
    ff(corruption, xind = wgi.s, yind = co2.s, limits=wgi.limits) +
    ff(Government, xind = wgi.s, yind = co2.s, limits=wgi.limits) +
    ff(Stability, xind = wgi.s, yind = co2.s, limits=wgi.limits) +
    ff(Law, xind = wgi.s, yind = co2.s, limits=wgi.limits) +
    ff(Regulation, xind = wgi.s, yind = co2.s, limits=wgi.limits) +
    ff(Voice, xind = wgi.s, yind = co2.s, limits=wgi.limits),
  #Support of carbon emissions
  yind = co2.s,
  data = emissions,

  #Thin plate basis splines for carbon emissions
  bs.yindex = list(bs = "tp"),
```

```r
  #Penalized splines for covariates
  bs.int = list(bs = "ps", k=4, m = c(2,2))
)
```

In addition, we include a general function-on-function model without any of the constraints on the support to compare how well it performs with our historical model (code given in the Appendix). To determine how well both models performed, we computed and compared the functional R^2 in Figure 5. This computation is defined by,

$$R^2(t) = 1 - \frac{\sum_{i=1}^{n}\left(Y_i(t) - \hat{Y}_i(t)\right)^2}{\sum_{i=1}^{n}\left(Y_i(t) - \bar{Y}(t)\right)^2} \tag{4}$$

```r
#Fitted values for the default model
yhat.default <- matrix(model.default$fitted.values, nrow=nrow(emissions$carbon),
               ncol=length(co2.s), byrow = T)

...

#Fitted values for the historical linear model
yhat <- matrix(model$fitted.values, nrow=nrow(emissions$carbon),
               ncol=length(co2.s), byrow = T)

...

#Computes a functional R-squared
fr2 <- function(y, y.hat, y.bar){
  n = nrow(y)
  ssr <- sapply(1:n, function(i){
    (y[i,]-y.hat[i,])^2
  }) %>% rowSums()
  sse <- sapply(1:n, function(i){
    (y[i,]-y.bar)^2
  }) %>% rowSums()
  #Returns a length(s) x 1 vector
  return(1-ssr/sse)
}

#Plot
fr2.default <- fr2(emissions$carbon, yhat.default, Y.bar)
fr2.historical <- fr2(emissions$carbon, yhat, Y.bar)

...
```

The default model achieves higher R^2 values $\forall t$ by construction, especially in the earlier years. However, this difference is attributable to the fact that the default model leverages covariate values from *future* years—violating temporal causality and thus introduces the risk of overfitting the model.

The historical model, in contrast, only draws on information available up to time t. As a result, its explanatory power is initially more limited. When t is close to the start of the time series, only one or two prior years of covariate history are available for integration. This explains the wider gap and greater dispersion in $R^2(t)$ values during the early 1990s. In our view, it is impressive that our HFLM achieves a similar fit: For any given time in which we observe carbon emissions for a given country, we only use on average less than half the available data through our temporal constraint; yet, the functional $R^2(t)$ achieves nearly[4] 90% (or higher) of the full FLM at any given t.

[4] In other words, if future information did, somehow, *actually* influence previous information (through what we might suppose as extraterrestrial means), we would expect that using less than half the data would give us half the predictive power.

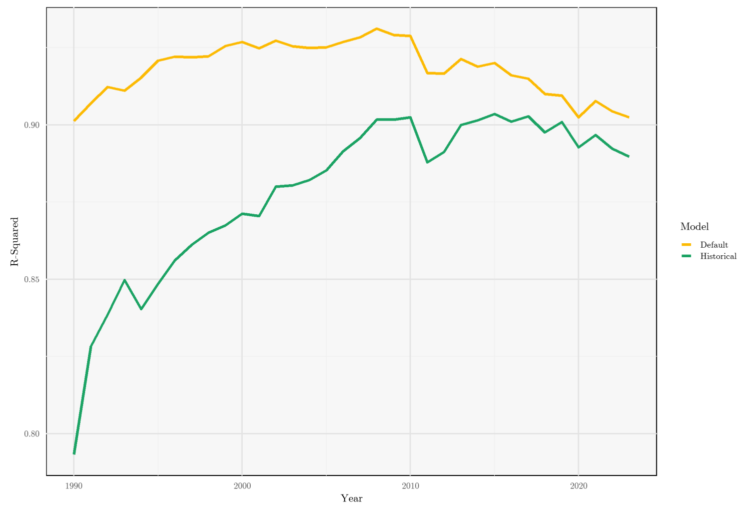


Figure 5: Functional R-squared comparison of the Function-on-function vs. Historical models

As time progresses, the historical model is able to accumulate more information from previously years, improving $R^2(t)$ and narrowing the gap relative to the unconstrained model. By the 2000s, both models achieve comparable predictive accuracy, but only the historical model does so within a causally interpretable framework. This supports our key argument: Temporal constraints yield more credible inferences about long-run energy policy effects without sacrificing much in-sample fit.

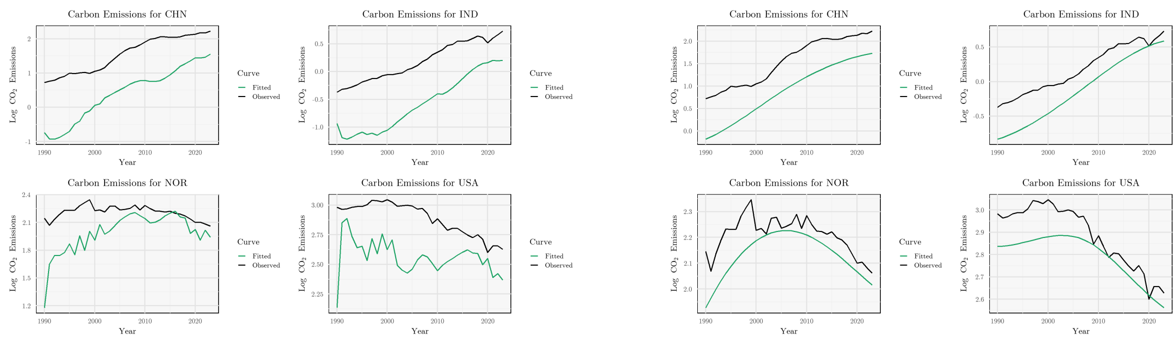
3.2 Choosing the Lag Parameter δ

Our selection of the lag parameter δ was based on heuristic judgment rather than formal cross-validation, due to computational constraints. Given the time spans of our response and covariates (1990-2023 for CO_2 and 1990-2020 for renewable energy consumption), we tested values of δ ranging from 3 to 30 and ultimately chose 20. Larger lag windows consistently improved in-sample fit, and 20 years struck a balance between model flexibility and plausibility when evaluated on specific country fits. While this approach is reasonable, future work should optimize δ out-of-sample—perhaps by minimizing a functional loss criterion such as $\int_{\mathcal{T}} R^2(t)dt$ across candidate values.[5]

3.3 Model Performance on Specific Countries

To evaluate the empirical accuracy of our historical model relative to the traditional function-on-function approach, we visualized the fitted and observed log CO_2 emission trajectories for four representative countries: China, India, Norway, and the United States.

[5]We omit this procedure here due to the computational burden of fitting ten separate historical models for each candidate δ_k.



(a) HFLM

(b) FLM

Figure 6: A comparison of our historical functional linear model and a traditional functional linear model.

Both China and India exhibit steep increases in emissions. Norway shows relatively stable emissions; the historical model captures the interior trend well but is quite noisy, while the unconstrained model offers a smoother trajectory. In the U.S., where emissions peaked and declined post-2007, both models reflect the downward trend, though the function-on-function model more closely tracks the steep drop.

We can see that the function-on-function model is generally smoother than our historical model. We conjecture that this may be because, by construction, we are limiting the amount of data used for estimating our coefficients at each (s, t) pair in our historical regression model—increasing the variance in the estimates. We might also suppose that our method of integration in Equation 2 may increase the variance—Riemann sum integration may perhaps introduce an exacerbated variance compared to the traditional trapezoidal integration used by `pffr`.

3.4 Effect of Renewable Energy Consumption

Figure 7 displays the estimated coefficient surface $\hat{\beta}_1(s, t)$ for the effect of renewable energy share at time s on log CO_2 emissions at time t, using a historical functional linear model with a 20-year lag constraint (the code for generating this plot is given in the appendix). Each coefficient $\hat{\beta}_1(s, t)$ represents the marginal effect of the share of renewable energy consumption in year s and emissions in year t, for $t - 20 \leq s \leq t$.

The surface reveals a temporally heterogeneous relationship. Early investments in renewable energy (notably from the 1990s) show a modest negative effect on emissions in later years. Along the diagonal, when we see the "immediate effect"of renewable energy consumption on emissions, we see negative or near-zero effects.

One interpretation of the pattern seen in the early to mid-2000s involves what we conjecture are implicit startup costs associated with renewable infrastructure. In the short run, investments in renewables often require significant capital inputs, including the expansion of grids, battery storage systems, and regulatory integration. These investments may temporarily increase overall energy demand or fossil fuel-based backup systems, leading to short-run increases in emissions. Importantly, the efficacy of renewables may not manifest until several years after deployment due to policy lags, as can be seen in the year 2000, for example, where the coefficient surface transitions from orange to a light green gradient at the end of 2020.

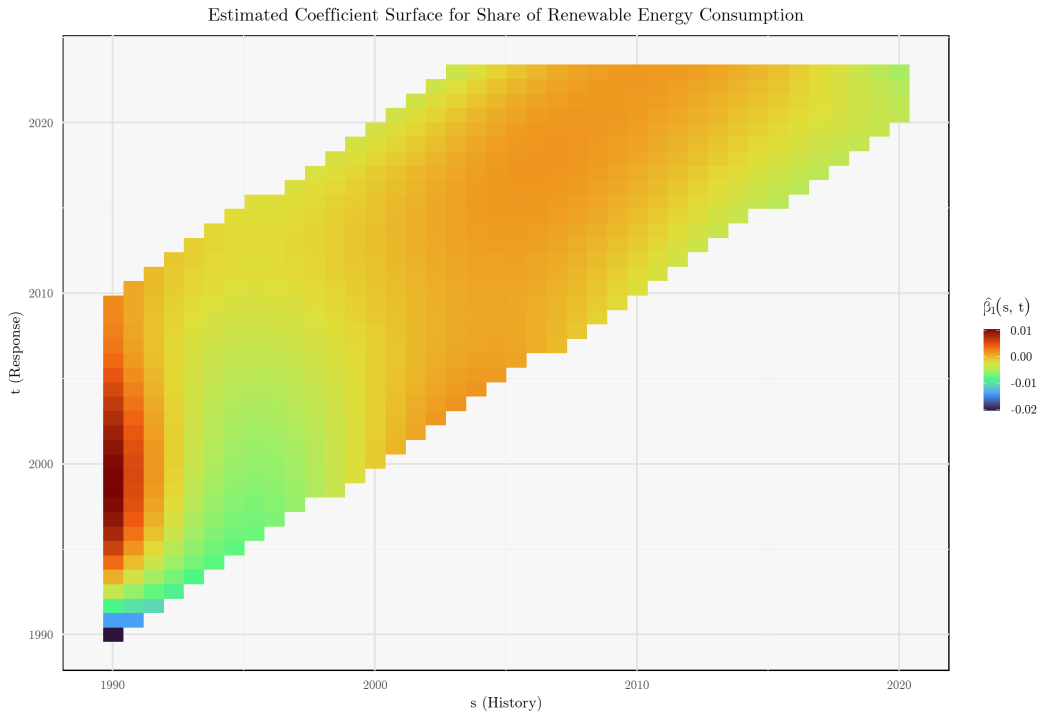


Figure 7: Estimated historical coefficient surface $\hat{\beta}_1(s,t)$ for the effect of renewable energy share. The triangular support enforces $s \leq t$ and $s \geq t - \delta$ with $\delta = 20$.

4 Conclusion & Limitations

Our project applies a historical functional linear model to evaluate how past renewable energy consumption influences present-day carbon emissions across countries. By restricting covariate influence to past values, our approach aligns with causal reasoning compared to conventional function-on-function models. We find that increases in renewable energy share are, on average, associated with long-run reductions in emissions, though the effect varies over time and diminishes as the temporal lag increases.

We discuss several limitations. Spatial correlation between countries may violate independence assumptions. We also acknowledge that potential misspecification may have biased[6] our fitted results (as can be seen in the gaps between our fitted predictions and recorded log per capita carbon emissions in Figure 6). Our exclusion of sparse covariates such as unemployment and interest rates may have also marginalized relevant explanatory variation. Moreover, using the share of renewable energy consumption within a country serves as an imperfect explanatory variable to measure how renewable energy projects affect carbon emissions within a country. For example, it is plausible for total energy consumption to increase within a country (and thus increasing carbon emissions) despite a stable renewable energy consumption *as a percent of total energy consumption* in that country. Future research will aim to address these shortcomings.

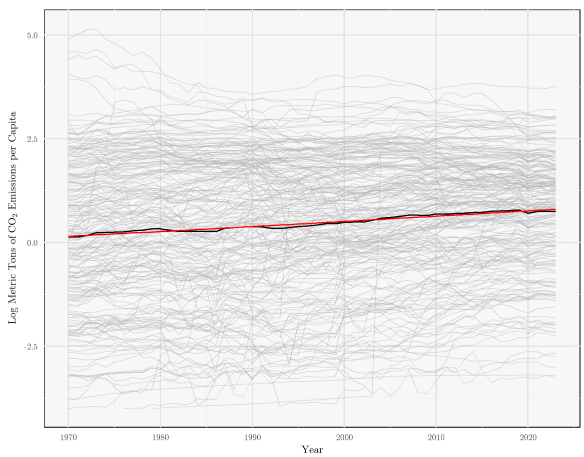
[6]We attempted to adjust for this bias by including an intercept without any penalization. After several attempts, we were not able to figure out the bias problem in our fitted values. While we leave this up to further research, it may be worth noting that `pffr` may not be the most effective tool for this model fit. We intend on exploring other methods in the future for estimating our historical functional linear model.

A Appendix

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.

A.1 Carbon Emissions

We begin this EDA by acknowledging the extreme variance in the data curves, particularly in the late 1900s (see Figure 11). Although we include a functional boxplot (see Figure 12) to show that log carbon emissions are well-behaved with respect to the variation of emissions across countries.



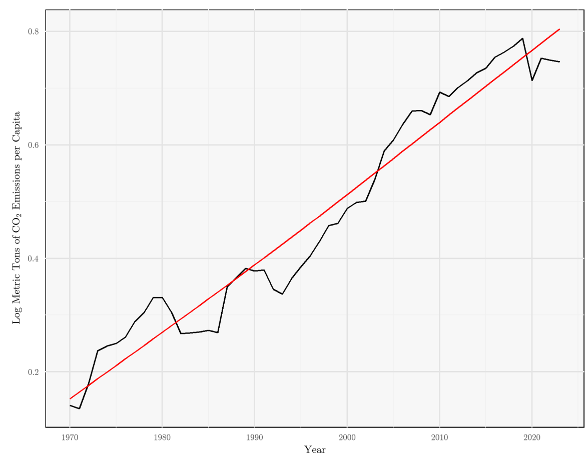


Figure 8: These are our mean estimation plots. We plotted the cross-sectional means (black) and penalized smooth splines (red) against the data first, but it was difficult to see the trends due to the intense skew. We then isolated the cross-sectional mean line and penalized smooth spline curve on their 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.

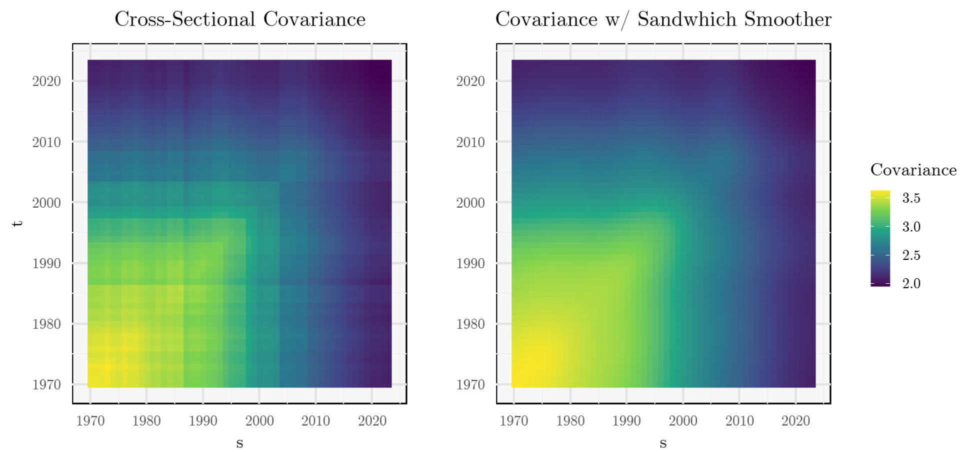


Figure 9: These plots show the cross sectional covariance and sandwich-smoothed covariance of our carbon emissions across years. We used `knots = 10` to get a more smoothed estimate.

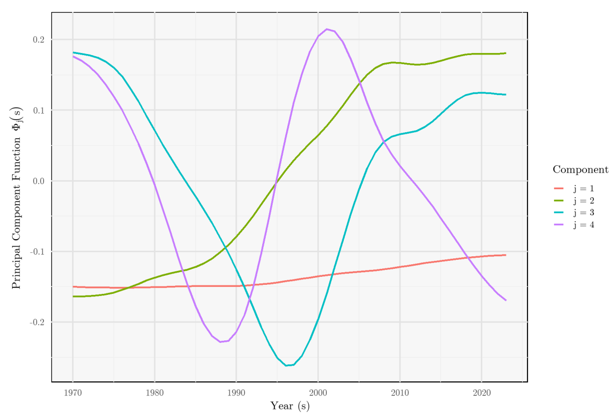


Figure 10: First four principal component functions, showing the 4 most dominant patterns of variation in the data. The total variance explained by these principal components is 0.987.

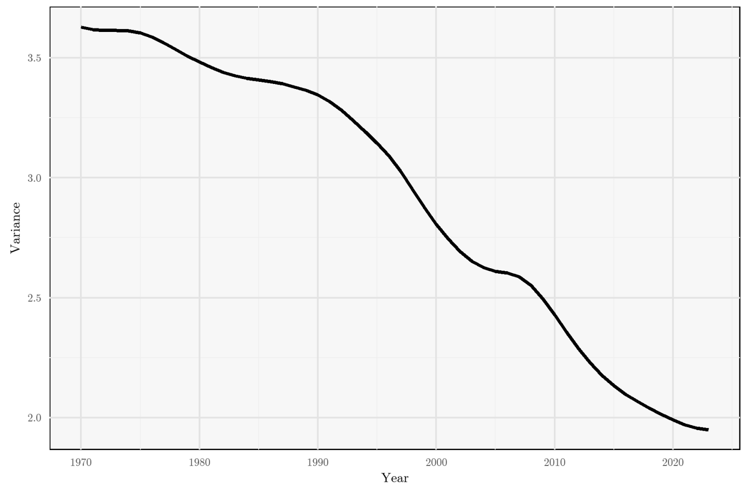


Figure 11: Carbon Emissions Variance Over Time

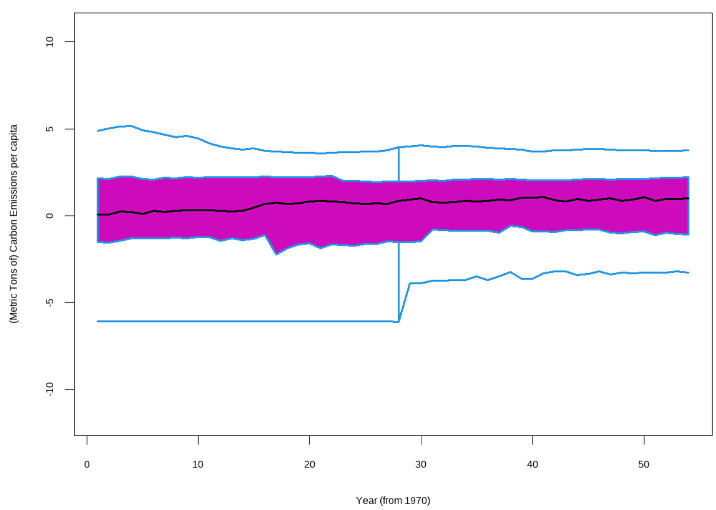


Figure 12: Carbon Emissions Functional Boxplot

A.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 15 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.

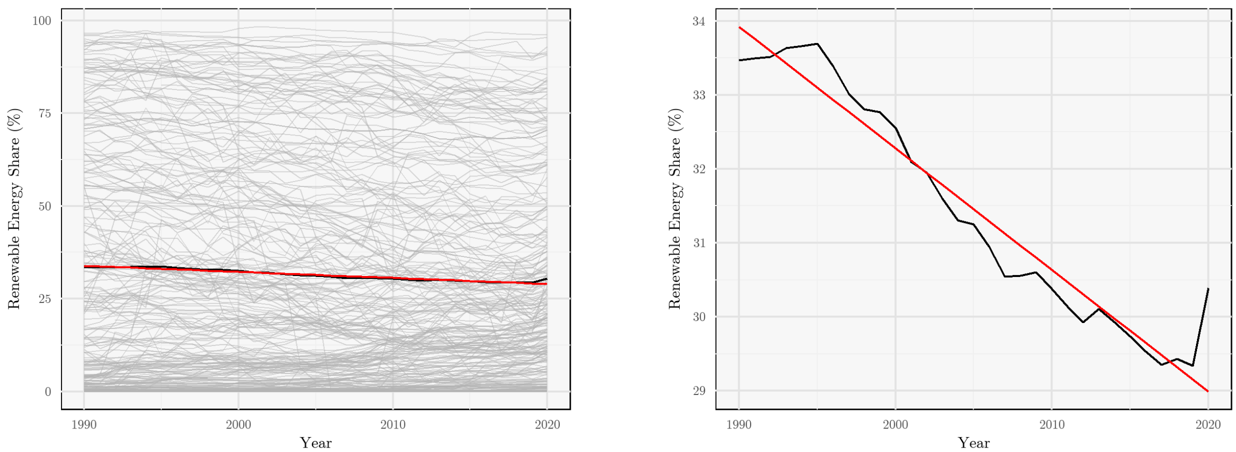


Figure 13: Mean Estimation: Cross Sectional (black) vs Penalized Smooth Splines (red). Penalized regression splines were fit using GCV.

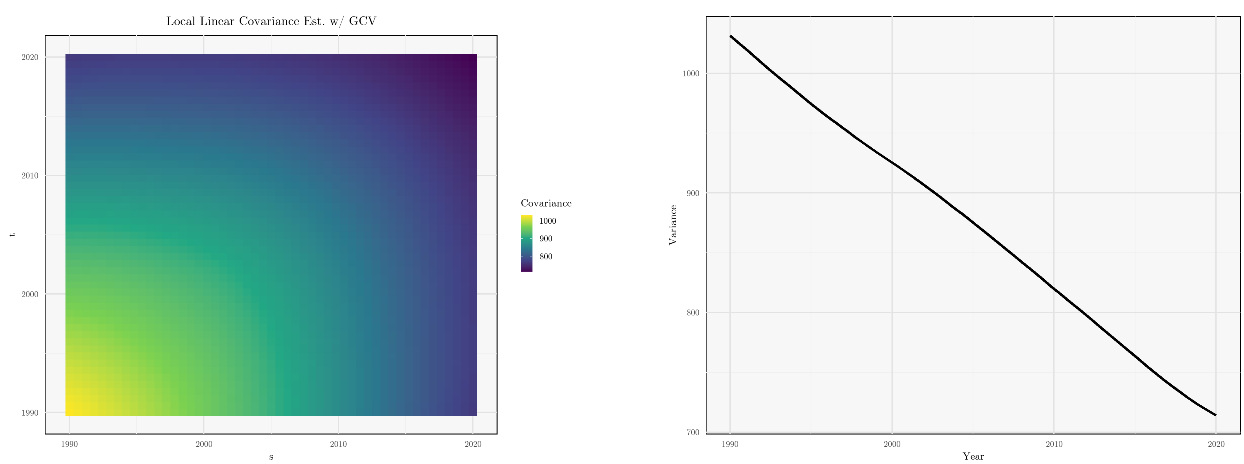


Figure 14: Local linear regression was fit to estimate the smoothed covariance. Epanechnikov kernel was used using GCV to choose bin width.

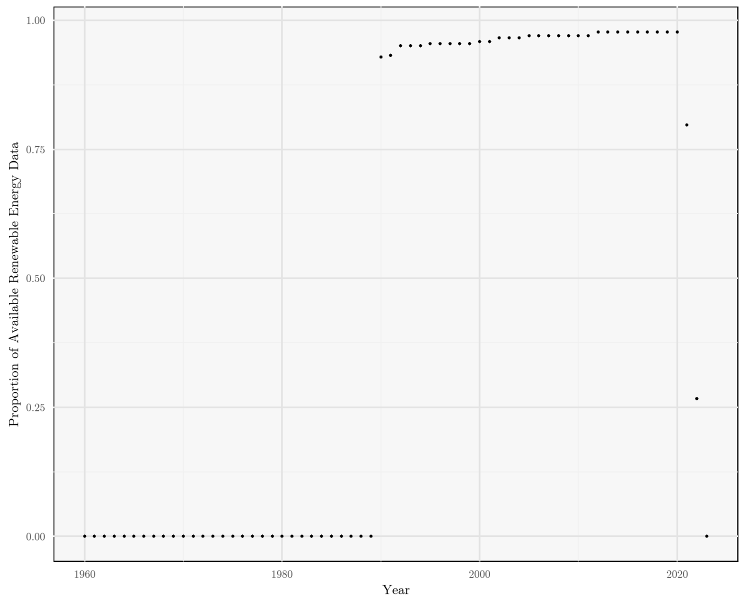


Figure 15: Proportion of data curves (as a fraction of the total number of countries in the data set) available per year.

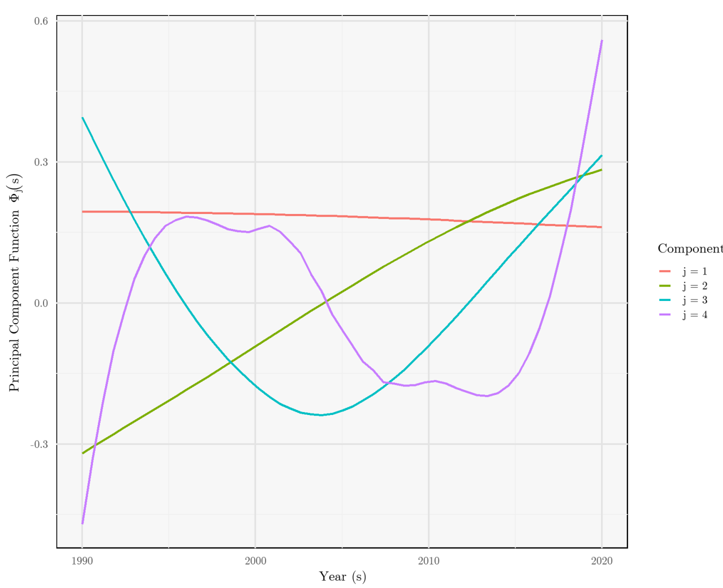


Figure 16: A plot of the first four principal component eigen functions are used extracted from the smoothed covariance estimate computed in Figure 14.

A.3 Code

While our GitHub provides a more organized view of our code (including data wrangling code), we provide the explicit code for all of our figures and analyses here.

A.4 EDA

We use this setup file to first load our wrangled data for visualization:

```
library(tidyverse)
library(npreg)
library(refund)

#Theme
library(sysfonts)

font_add("cm", regular="fonts/cmunrm.ttf")
showtext::showtext_auto()

theme <- theme_minimal(base_size = 24, base_family = "cm") +
  theme(
    #axis.text.x = element_text(angle = 75, hjust = 1),
    axis.title.x = element_text(face = "bold"),
    axis.title.y = element_text(face = "bold"),
    axis.text = element_text(face = "bold"),
    panel.background = element_rect(fill = "#F7F7F7"),
    panel.grid.major = element_line(color = "#E3E3E3"),
    panel.grid.minor = element_line(color = "#F0F0F0"),
    plot.title = element_text(hjust = 0.5, face = "bold")
  )

emissions <- read_csv("data/clean/emissions.csv")
carbon <- read_csv("data/clean/carbon.csv")

carbon.dat <- carbon %>% dplyr::select(Country, Year, log.CO2) %>%
  pivot_wider(
    names_from = Year,
    values_from = log.CO2
  )

W <- as.matrix(carbon.dat[,str_c(1970:2023)])
rownames(W) <- carbon.dat$Country
colnames(W) <- 1970:2023
```

Code for Figure 3a (Carbon Emissions data Visualization)

```
country_names <- c("QAT" = "Qatar", "USA" = "United States", "IND" = "India",
                   "CHN" = "China", "NGA" = "Nigeria")

country_colors <- c("China" = "#E63946",
                    "United States" = "#1D35AA",
                    "India" = "#F4A261",
                    "Qatar" = "#2A9D8F",
                    "Nigeria" = "#8E44AD")

carbon_filtered <- carbon %>%
  mutate(Country = recode(CC, !!!country_names)) %>%
  filter(log.CO2 > -4)

label_data <- carbon_filtered %>%
  filter(CC %in% names(country_names)) %>%
  group_by(Country) %>%
  slice_sample(n = 1)

ggplot() +
  geom_line(data = carbon_filtered,
            aes(x = Year, y = log.CO2, group = CC),
            color = "gray70", alpha = 0.5) +
  geom_line(data = carbon_filtered %>% filter(CC %in% names(country_names)),
            aes(x = Year, y = log.CO2, group = Country, color = Country),
```

```
            linewidth = 1, alpha=0.8, show.legend = F) +
  geom_label_repel(data = label_data,
                   aes(x = Year, y = log.CO2, label = Country, color = Country),
                   size = 8,
                   direction = "y",
                   segment.color = NA,
                   box.padding = 0.4,
                   show.legend = FALSE) +
  scale_color_manual(values = country_colors, name = "Country") +
  labs(
    x = "Year",
    y = expression("Log Metric Tons of"~CO[2]~"Emissions per Capita")
  )+
  theme
```

Code for Figure 12 (functional boxplot for the Carbon Emissions data)

```
fda::fbplot(t(W), xlab="Year (from 1970)", ylab="(Metric Tons of) Carbon Emissions per
    capita")
```

Code for Figure 8 (Mean estimation for Carbon Emissions data)

```
#Cross-sectional mean estimation
mu.hat <- colMeans(W)

#Using penalty chosen by GCV
mu.splines <- npreg::ss(x = carbon$Year, y = carbon$log.CO2,
                  method = "GCV")$y

mu.df <- data.frame(
  Year = as.integer(colnames(W)),
  MuHat = mu.hat,
  MuSplines = mu.splines
)

mu.df %>%
  ggplot() +
    geom_line(data = mu.df, aes(x = Year, y = MuHat), color = "black", linewidth=1)+
    geom_line(data = mu.df, aes(x = Year, y = MuSplines), color = "red", linewidth=1)+
  labs(
    x = "Year",
    y = expression("Log Metric Tons of"~CO[2]~"Emissions per Capita")
  )+
  theme

ggplot() +
  geom_line(data = carbon_filtered,
            aes(x = Year, y = log.CO2, group = CC),
            color = "gray70", alpha = 0.5) +  # Light gray for background countries
  geom_line(data = mu.df, aes(x = Year, y = MuHat), color = "black", linewidth=1)+
  geom_line(data = mu.df, aes(x = Year, y = MuSplines), color = "red", linewidth=1)+
  labs(
    x = "Year",
    y = expression("Log Metric Tons of"~CO[2]~"Emissions per Capita")
  )+
  theme
```

Code for Figure 9 (Covariance estimates for Carbon Emissions data)

```
#Sandwich smoother
W.tilde = scale(W, center = mu.splines, scale=F)
blt.fit <- refund::fpca.face(Y = W.tilde,
                     argvals = 1970:2023,
                     knots = 10,
                     npc = 10,
                     center = F,
```

```r
                     var = TRUE)

Khat.tilde <- blt.fit$efunctions %*% diag(blt.fit$evalues) %*%
  t(blt.fit$efunctions)

Covdf <- expand.grid(s = 1970:2023, t = 1970:2023)
Covdf$CS <- c(cov(W))
Covdf$Khat <- c(Khat.tilde)

CovPlot.CS <- ggplot(Covdf, aes(x = s, y = t)) +
  geom_raster(aes(fill = CS), show.legend = F) +
  scale_fill_viridis_c(
    values = scales::rescale(c(
      seq(0, 300, 50)
  ))) +
  xlab('s') + ylab('t') + theme_bw() +
  ggtitle('Cross-Sectional Covariance')+
  theme

CovPlot.Smooth <- ggplot(Covdf, aes(x = s, y = t)) +
  geom_raster(aes(fill = Khat)) +
  scale_fill_viridis_c(
    name = "Covariance",
    values = scales::rescale(c(
      seq(0, 300, 50))))+
  xlab('s') + ylab('') + theme_bw() +
  ggtitle('Covariance w/ Sandwhich Smoother') +
  theme

CovPlot.CS + CovPlot.Smooth
```

Code for Figure 11 (Variance of Carbon Emissions data)

```r
ggplot(mapping=aes(x = 1970:2023,
                   y = diag(Khat.tilde)))+
  geom_line(linewidth = 2)+
  labs(
    x = "Year",
    y = "Variance"
  )+
  theme
```

Code for Figure 15 (Fraction of available curves for our Renewable Energy Consumption data)

```r
avail.data <- renewable_shares %>% group_by(Year) %>%
  summarize(
    n = mean(!is.na(Renewable_Share))
  )

ggplot(avail.data, aes(x=Year, y=n))+
  geom_point()+
  labs(
    y = "Proportion of Available Renewable Energy Data"
  )+
  theme
```

Code for Figure 3b (Renewable Energy Consumption data)

```r
country_names <- c("QAT" = "Qatar", "USA" = "United States", "IND" = "India",
                   "CHN" = "China", "NGA" = "Nigeria")

label_data <- renewables_filtered %>%
  filter(CC %in% names(country_names)) %>%
  group_by(Country) %>%
  slice_sample(n = 1)

```

```r
country_colors <- c("China" = "#E63946",
                    "United States" = "#1D35AA",
                    "India" = "#F4A261",
                    "Qatar" = "#2A9D8F",
                    "Nigeria" = "#8E44AD")

renewables_filtered <- renewable_shares %>%
  filter(CC != "PLW" & Year %in% s) %>%
  na.omit() %>%
  mutate(Country = recode(CC, !!!country_names))

ggplot() +
  geom_line(data = renewables_filtered,
            aes(x = Year, y = Renewable_Share, group = CC),
            color = "gray70", alpha = 0.5) +
  geom_line(data = renewables_filtered %>% filter(CC %in% names(country_names)),
            aes(x = Year, y = Renewable_Share, group = Country, color = Country),
            linewidth = 1, alpha = 0.8, show.legend = F) +
  geom_label_repel(data = label_data,
                   aes(x = Year, y = Renewable_Share, label = Country, color = Country),
                   size = 8,
                   direction = "y",
                   segment.color = NA,
                   box.padding = 0.4,
                   show.legend = FALSE) +
  scale_color_manual(values = country_colors) +
  labs(
    x = "Year",
    y = "Renewable Energy Share (%)"
  )+
  theme
```

Code for Figure 13 (Renewable Energy Consumption mean estimation)

```r
#Cross-sectional mean estimation
mu.hat <- apply(W, 2, function(x){
  mean(x, na.rm=T)
})

#Using penalty chosen by GCV
mu.splines <- npreg::ss(x = renewables_filtered$Year, y =
    renewables_filtered$Renewable_Share,
                 method = "GCV")$y

mu.df <- data.frame(
  Year = as.integer(colnames(W)),
  MuHat = mu.hat,
  MuSplines = mu.splines
)

mu.df %>%
  ggplot() +
  geom_line(data = mu.df, aes(x = Year, y = MuHat), color = "black", linewidth=1)+
  geom_line(data = mu.df, aes(x = Year, y = MuSplines), color = "red", linewidth=1)+
  labs(
    x = "Year",
    y = expression("Renewable Energy Share (%)")
  )+
  theme

ggplot() +
  geom_line(data = renewables_filtered,
            aes(x = Year, y = Renewable_Share, group = CC),
            color = "gray70", alpha = 0.5) +
  geom_line(data = mu.df, aes(x = Year, y = MuHat), color = "black", linewidth=1)+
```

```
  geom_line(data = mu.df, aes(x = Year, y = MuSplines), color = "red", linewidth=1)+
  labs(
    x = "Year",
    y = expression("Renewable Energy Share (%)")
  )+
  theme
```

Code for Figure 14 (Renewable Energy Consumption covariance estimates)

```
W.tilde = scale(W, center = mu.splines, scale=F)

L <- fdapace::MakeFPCAInputs(IDs = renewables_filtered$CC,
                               renewables_filtered$Year,
                               renewables_filtered$Renewable_Share)

cov.sparse.gcv <- fdapace::FPCA(L$Ly, L$Lt,
                                optns = list(
                                  'kernel' = 'epan',
                                  'methodBwCov' = 'GCV',
                                  'error' = T,
                                  'useBinnedCov' = T,
                                  'dataType' = 'Sparse',
                                  'methodSelectK' = 4
                                ))

Covdf <- expand.grid(s = cov.sparse.gcv$workGrid, t = cov.sparse.gcv$workGrid)
Covdf$Khat <- c(cov.sparse.gcv$smoothedCov)

ggplot(Covdf, aes(x = s, y = t)) +
  geom_raster(aes(fill = Khat)) +
  scale_fill_viridis_c(
    name = "Covariance",
    values = scales::rescale(c(
      seq(0, 1000, 50))))+
  xlab('s') + theme_bw() +
  ggtitle('Local Linear Covariance Est. w/ GCV') +
  theme

#Variance Plot

ggplot(mapping=aes(x = cov.sparse.gcv$workGrid,
                   y = diag(cov.sparse.gcv$smoothedCov)))+
  geom_line(linewidth = 2)+
  labs(
    x = "Year",
    y = "Variance"
  )+
  theme
```

Code for Figure 10 (FPCA plot for Carbon Emissions)

```
PhiW = blt.fit$efunctions
ggplot(mapping = aes(x = rep(1970:2023, 4),
                     y = c(PhiW[,1:4]),
                     color = Component)) +
  geom_line(linewidth=1.2)+
  labs(
    x = "Year (s)",
    y = expression(
      "Principal Component Function " ~ Phi[j](s)),
    color = "Component"
  )+
  theme

sum(blt.fit$evalues[1:4])/sum(blt.fit$evalues)
```

Code for Figure 16 (FPCA Plot for Renewable Energy Consumption)

```
PhiW = cov.sparse.gcv$phi
Component <- rep(c("j = 1", "j = 2", "j = 3", "j = 4"),
                 each = length(cov.sparse.gcv$workGrid))
ggplot(mapping = aes(x = rep(cov.sparse.gcv$workGrid, 4),
                     y = c(PhiW[,1:4]),
                     color = Component)) +
  geom_line(linewidth=1.2)+
  labs(
    x = "Year (s)",
    y = expression(
      "Principal Component Function " ~ Phi[j](s)),
    color = "Component"
  )+
  theme
```

A.5 Model Fit

Code for Setup:

```
library(refund)

co2.s <- 1990:2023
s <- 1990:2020
wgi.s <- c(1996, 1998, 2000, 2002:2023)

emissions <- readRDS("data/clean/emissions.rds")

CCs <- emissions$carbon %>% rownames()
countries <- c("USA", "CHN", "IND", "NOR")
countries.index <- which(CCs %in% countries)
```

Code for Mean Imputations:

```
na.test <- function(var){
  nas <- sum(is.na(emissions[[var]]))
  n <- prod(dim(emissions[[var]]))
  return(nas/n)
}

for(v in names(emissions)){
  print(str_c(v, ": ", 100*round(na.test(v), 4), "%"))
}

emissions$energy <- apply(emissions$energy, 2, function(col) {
  col[is.na(col)] <- mean(col, na.rm = TRUE)
  return(col)
})

emissions$hdi <- apply(emissions$hdi, 2, function(col) {
  col[is.na(col)] <- mean(col, na.rm = TRUE)
  return(col)
})

emissions$inflation <- apply(emissions$inflation, 2, function(col) {
  col[is.na(col)] <- mean(col, na.rm = TRUE)
  return(col)
})

emissions$corruption <- apply(emissions$corruption, 2, function(col) {
  col[is.na(col)] <- mean(col, na.rm = TRUE)
  return(col)
})
```

```
emissions$Government <- apply(emissions$Government, 2, function(col) {
  col[is.na(col)] <- mean(col, na.rm = TRUE)
  return(col)
})

emissions$Stability <- apply(emissions$Stability, 2, function(col) {
  col[is.na(col)] <- mean(col, na.rm = TRUE)
  return(col)
})

emissions$Regulation <- apply(emissions$Regulation, 2, function(col) {
  col[is.na(col)] <- mean(col, na.rm = TRUE)
  return(col)
})

for(v in names(emissions)){
  print(str_c(v, ": ", 100*round(na.test(v), 4), "%"))
}
```

Code for Fitting the Default FLM Model:

```
model.default <- pffr(
  carbon ~ ff(energy, xind = s, yind = co2.s) +
           ff(gdp, xind = s, yind = co2.s) +
           ff(hdi, xind = s, yind = co2.s) +
           ff(inflation, xind = s, yind = co2.s) +
           ff(corruption, xind = wgi.s, yind = co2.s) +
           ff(Government, xind = wgi.s, yind = co2.s) +
           ff(Stability, xind = wgi.s, yind = co2.s) +
           ff(Law, xind = wgi.s, yind = co2.s) +
           ff(Regulation, xind = wgi.s, yind = co2.s) +
           ff(Voice, xind = wgi.s, yind = co2.s),
  yind = co2.s,
  data = emissions,

  #Thin plate basis splines for carbon emissions
  bs.yindex = list(bs = "tp"),

  #Penalized splines for covariates
  bs.int = list(bs = "ps", k=4, m = c(2,2))
)
saveRDS(model.default, "scripts/models/model_default.RDS")
```

Code for Figure 6 (Fitted vs. Observed Curves)

```
#Curve is the curve (row) index in the functional data
#This corresponds to how we formatted the fitted values
plot.fitted <- function(curve, yhat){
  plt <- ggplot(mapping=aes(x=co2.s))+
    geom_line(aes(y=emissions$carbon[curve,], color = "Observed"), linewidth=1)+
    geom_line(aes(y=yhat[curve,], color = "Fitted"), linewidth=1)+
    ...
  return(plt)
}

fitted.plots.default <- lapply(countries.index, function(cc){
  plot.fitted(cc, yhat.default)
})

#Using patchwork
wrap_plots(fitted.plots.default, ncol = 2)

#(The code is the same for the historical model fit, just with the appropriate arguments)
...
```

Code for Figure 7 (Coefficient surface for measuring the effect of renewable energy consumption on carbon emissions)

```
#Renewable Energy Consumption Coefficient
pffr.coefs <- coef(model)
energy.coef <- pffr.coefs$smterms[[2]]$coef
colnames(energy.coef)[1:2] <- c("s", "t")

energy.coef %>%
  dplyr::filter(s >= pmax(1990, t - delta), s <= t) %>%
  ggplot(aes(x = s, y = t, fill = value)) +
    geom_raster() +
    scale_fill_gradient2(
      low = "#1ca364",
      mid = "#EEFFEE",
      high = "#fcba03",
      midpoint = 0,
      name = expression(hat(beta[1])(s, t))
    )+
    ...
```

References

Bank, T. W. (2023). World development indicators. https://databank.worldbank.org/source/world-development-indicators.

Commission, E. (2023). Emissions database for global atmospheric research (edgar). https://edgar.jrc.ec.europa.eu/overview.php?v=booklet2023.

Elayouty, A. and Abou-Ali, H. (2023). Functional data analysis of the relationship between electricity consumption and climate change drivers. *Journal of Applied Statistics*, 50(10):2267–2285.

Malfait, N. and Ramsay, J. O. (2003). The historical functional linear model. *Canadian Journal of Statistics*, 31(1):115–128.