A Functional Carbon Emissions Analysis

Sam Lee, Everett Andrew
Brigham Young University

April 2025

Abstract

We investigate the temporal relationship between national renewable energy consumption and per-capita carbon emissions using a historical functional linear regression model. Using a compilation of economic, political, and environmental data across 178 countries from 1990 to 2023, we model carbon emissions as a smooth functional response influenced by past values of renewable energy consumption and macroeconomic covariates. We impose a historical constraint that ensures predictor values precede response values within a lag window. Our results suggest that increased renewable energy consumption between 1995 and 2010 is associated with reductions in emissions in subsequent years, after adjusting for governance and development indicators.

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 per-capita carbon emissions over time and the intertemporal consumption of renewable energy across countries. We will use functional data analysis to capture smooth temporal trends, quantify the dynamics of how carbon emissions data are affected by renewable energy consumption, and control for confounding covariates in our models.

The remainder of the paper is structured as follows. In this section, we state our research question (Section 1.1), describe the data sources and key variables, preprocessing steps (Section 1.2), and introduce the historical functional linear regression model (Section 1.3), 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 and figures are provided in the Appendix.

1.1 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?*

1.2 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 CO_2 emissions per capita from 1970 onward, serving as our primary response variable (see Figure 1) (Commission, 2023).

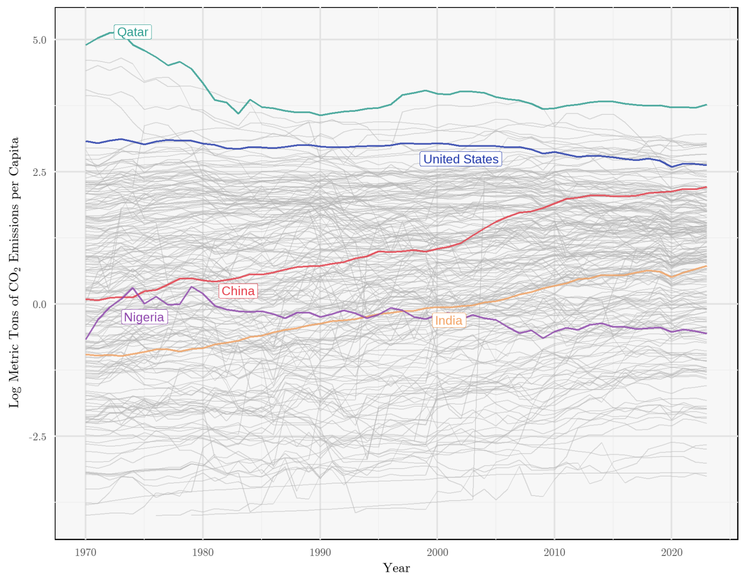


Figure 1: This is a plot of the CO2 emissions per capita for each country. This plot displays a modest trend downwards in emissions from 1970, and it shows a heavy skew towards 0.

To evaluate the role of renewable energy 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 2).

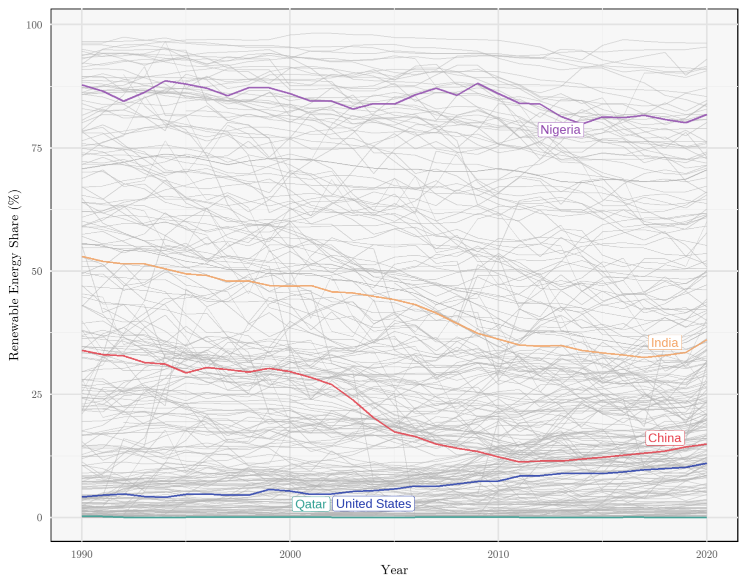


Figure 2: Renewable Energy Shares across countries over time. Specific countries (USA, China, India, Nigeria, and Qatar) are emphasized.

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, with covariates with a small fraction of missing data (e.g. %Missing < 10% as we define it), we use a functional mean imputation: 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[1] 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.

1.3 Historical Functional Linear Model

Traditional function-on-function models would not be an appropriate way to answer this question, since they would not account for significant differences in years between renewable energy share and carbon emissions. For example, the share of renewable energy of a country in 1992 would likely not have a very significant effect on its carbon emissions in 2020. In addition, 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 adds more noise to the model as opposed to any real signal[2].

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). 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 $x(s)$

[1] 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.

[2] 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 doesn't imply any significant increase in prediction power.

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: It takes into account these differences in the years. The historical functional linear model includes a constraint $s \leq t$ to ensure that the time points of the explanatory factors do not come after the time points of the responses. It also includes a lag parameter, δ, which we can determine 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.

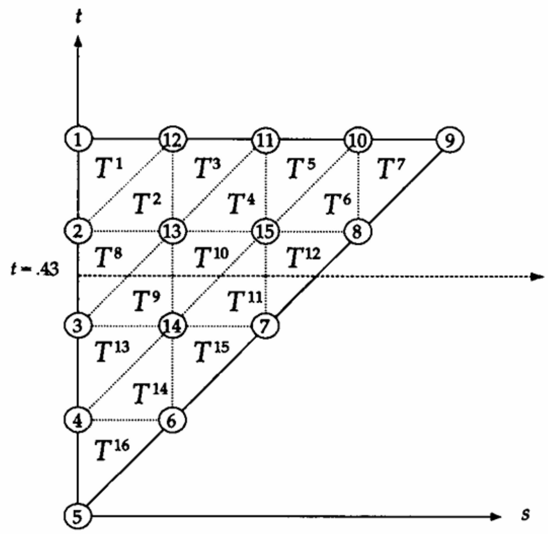


Figure 3: Triangular Support (Malfait and Ramsay, 2003)

2 Methods

To estimate the effect of renewable energy share and related covariates on carbon emissions over time, we adopt a 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 $s \geq t$.

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 (s) and response (t) 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:

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

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 $\mathcal{S}$. 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 14 and 4), we can still fit those "future" 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 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 a 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 applying our historical model:

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

#Functional data set compiled with scripts/cleaning/clean.R
emissions <- readRDS("data/clean/emissions.rds")

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

#Time lag
delta <- 20
#Integration limits for economic covariates
cov.limits <- function(s, t){
  s >= pmax(1990, t-delta)
```

```r
}
#Integration limits for WGI covariates
wgi.limits <- function(s, t){
  s >= pmax(1996, t-delta)
}
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"),

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

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:

```r
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")
```

To determine how well both models performed, we computed the functional R^2, which is given by:

$$R^2(t) = 1 - \frac{\sum_{i=1}^{n}\left(Y_i(t) - \hat{Y}_i(t)\right)^2}{\sum_{i=1}^{n}\left(Y_i(t) - \bar{Y}(t)\right)^2} \tag{3}$$

In R, we construct the following graphic in Figure 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)
colnames(yhat.default) <- co2.s
rownames(yhat.default) <- rownames(emissions$carbon)

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

#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)

ggplot(
  mapping=aes(x = co2.s),
) +
  geom_line(aes(y = fr2.default, color="Default"), linewidth=2)+
  geom_line(aes(y = fr2.historical, color="Historical"), linewidth=2)+
  ...
```

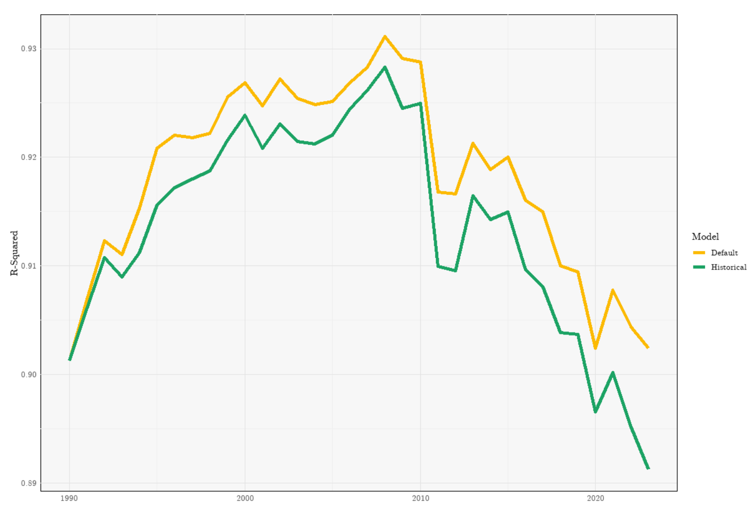


Figure 4: Functional R-squared comparison of the Function-on-function vs. Historical models

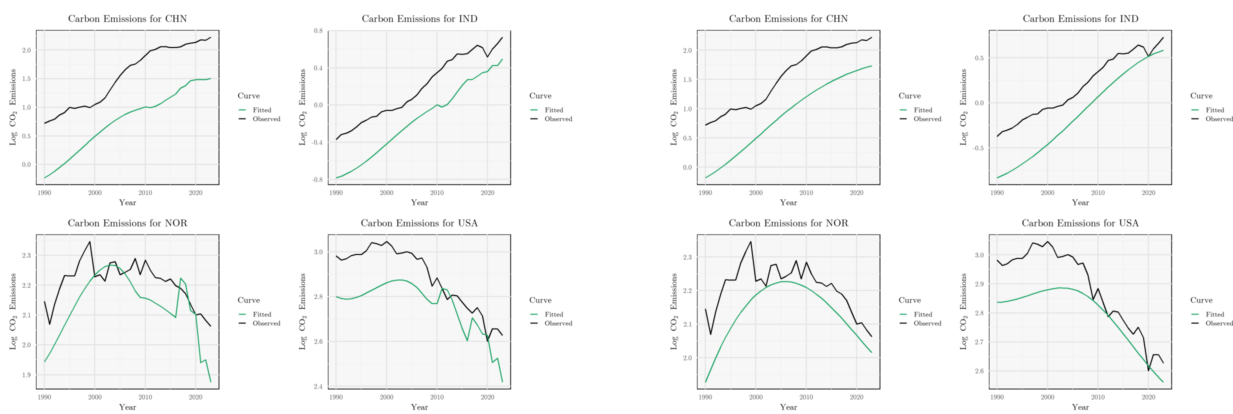
We can see that as the year increases, the default model appears to perform slightly better than the historical functional linear model. There is good reason for this—the function-on-function model erroneously uses renewable energy share data across all available years to predict CO_2 emissions. However, as we've discussed earlier, it's not appropriate to use data from the future to fit or predict a model on data in the past. The traditional function-on-function model is essentially fitting the noise that remains after the years up to time t have been accounted for. As a result, and as can be seen in the figure, the differences between the R^2 values balloon over time.

3.2 Choosing the Lag Parameter δ

Our process for selecting our lag parameter was rudimentary. Since we have only a short, discrete span of years we are working with, we chose it based on prior assumptions and trial and error. Given our range of years for CO_2 emissions and Renewable Energy Share (1990-2023 and 1990-2020, respectively), we had to choose δ between 30 and 3, and after testing a few values (3, 5, 10, 15, 20, 25, and 30), we settled on 20. We figured that the more prior years of renewable energy share data we included to fit a given CO_2 value, the more precise the fit would be. However, we also acknowledged that the effect of a prior year would also diminish the further back in time it is. After testing some fits on specific countries with a 20-year lag, we decided it was the most appropriate value.

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.



(a) Historical

(b) Traditional

Figure 5: 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 diverges post-2020, 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.

3.4 Effect of Renewable Energy Consumption

Figure 6 presents the estimated coefficient surface $\hat{\beta}_1(s, t)$ for the effect of renewable energy consumption on carbon emissions. Negative values of $\hat{\beta}_1(s, t)$ (in green) indicate that higher renewable energy consumption in past years is associated with lower present-day carbon emissions, as we hypothesized. This relationship is particularly pronounced for values of s between 1995 and 2010 and t between 2000 and 2020, suggesting that investments in renewable energy during this period have contributed meaningfully to emission reductions up to two decades later.

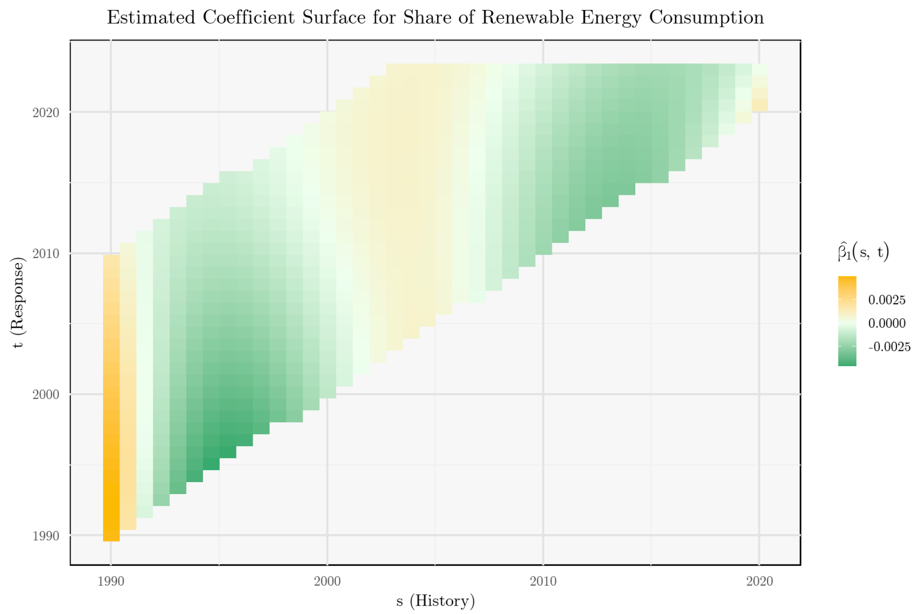


Figure 6: Estimated historical coefficient surface $\hat{\beta}_1(s,t)$ representing the effect of renewable energy share at time s on log carbon emissions at time t. The support is restricted to the triangular region defined by $s \leq t$ and $s \geq t - \delta$, with $\delta = 20$.

This figure was generated with the following R code:

```r
#Renewable Energy Consumption Coefficient
pffr.coefs <- coef(model)
energy.coef <- pffr.coefs$smterms[[2]]$coef
#For convenience, rename the history and response to the conventional s and t variables
colnames(energy.coef)[1:2] <- c("s", "t")

energy.coef %>%
  #Obtain the triangular support of interest
  dplyr::filter(s >= pmax(1990, t - delta), s <= t) %>%
  ggplot(aes(x = s, y = t, fill = value)) +
    geom_raster()
```

The surface also reveals a decay in influence as the temporal gap $t - s$ widens, consistent with the intuition that the marginal impact of earlier renewable energy use diminishes over time. Graphically, the darker tones along the diagonal suggest a moderately strong initial effect that peters out the further away we get from the initial consumption period. A small positive region in the early 1990s may reflect transitional inefficiencies or data sparsity—understanding that data before 1990 are not considered in this analysis. The other pockets of positive coefficient values may suggest the presence of startup costs for renewable energy during investment periods. Overall, the pattern supports the view that renewable energy share is a long-run negative determinant of emission trajectories.

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 better aligns with causal reasoning than 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 of our fitted intercept may have biased[3] our fitted results (as can be seen in the gaps between our fitted predictions and recorded log per capita carbon emissions in Figure 5). Our exclusion of sparse covariates such as unemployment and interest rates may have also hindered 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.

[3]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 10). Although we include a functional boxplot (see Figure 11) to show that log carbon emissions are well-behaved with respect to the variation of emissions across countries.

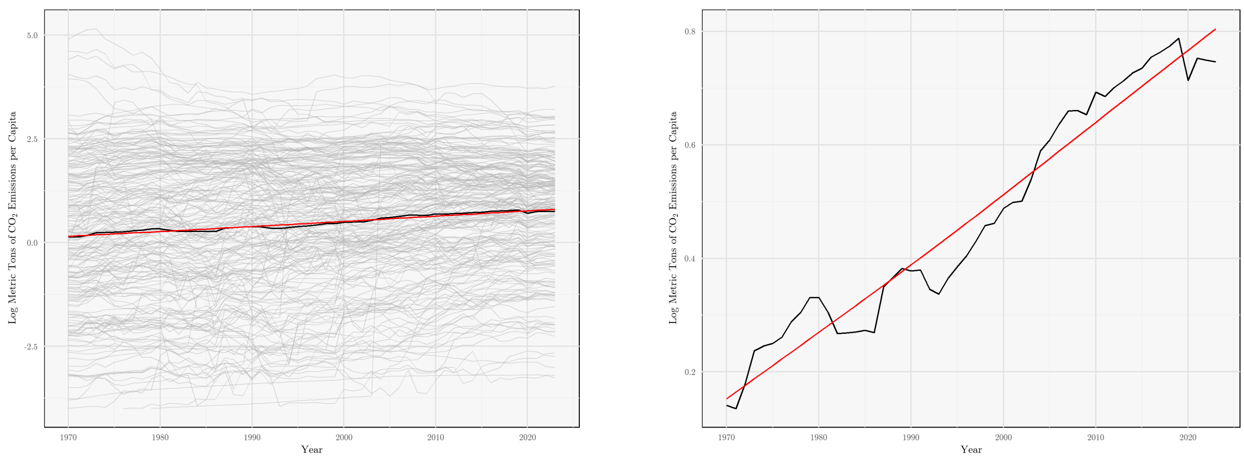


Figure 7: These are our mean estimation plots. We plotted the cross-sectional means (black) and penalized smooth splines (red) against the data first, but it was difficult to see the trends due to the intense skew. We then isolated the cross-sectional mean line and penalized smooth spline curve on their 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.

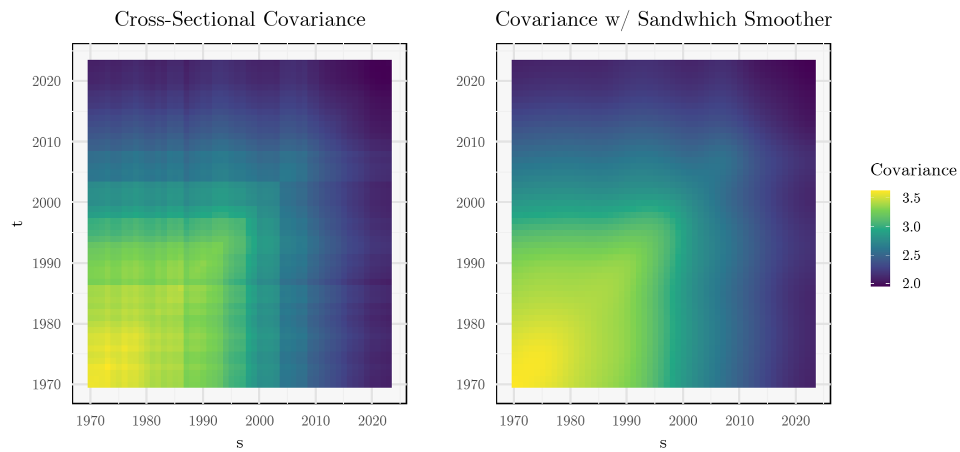


Figure 8: These plots show the cross sectional covariance and sandwich-smoothed covariance of our carbon emissions across years. We used `knots = 10` to get a more smoothed estimate.

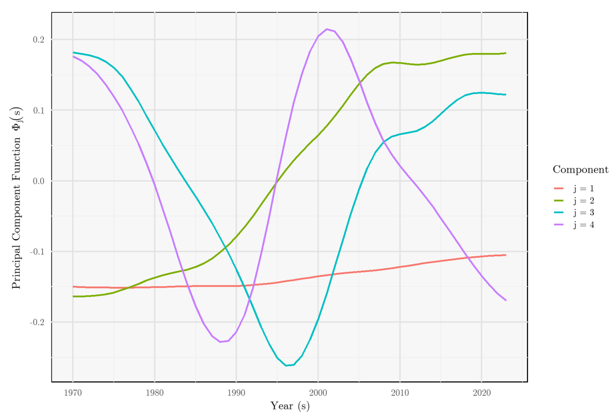


Figure 9: First four principal component functions, showing the 4 most dominant patterns of variation in the data. The total variance explained by these principal components is 0.987.

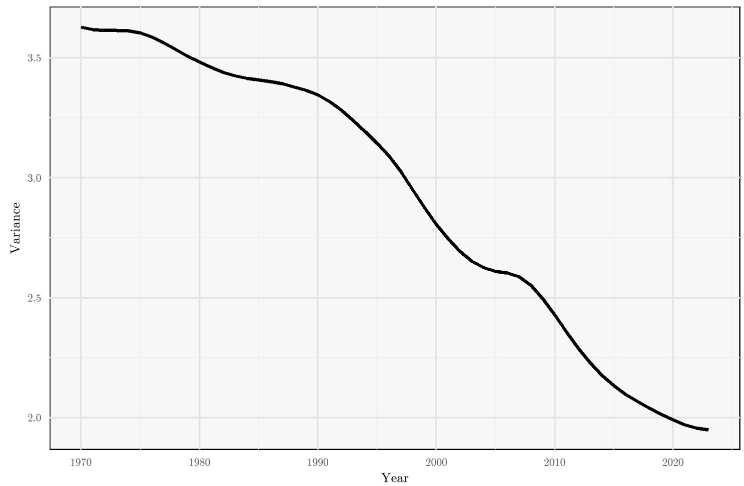


Figure 10: Carbon Emissions Variance Over Time

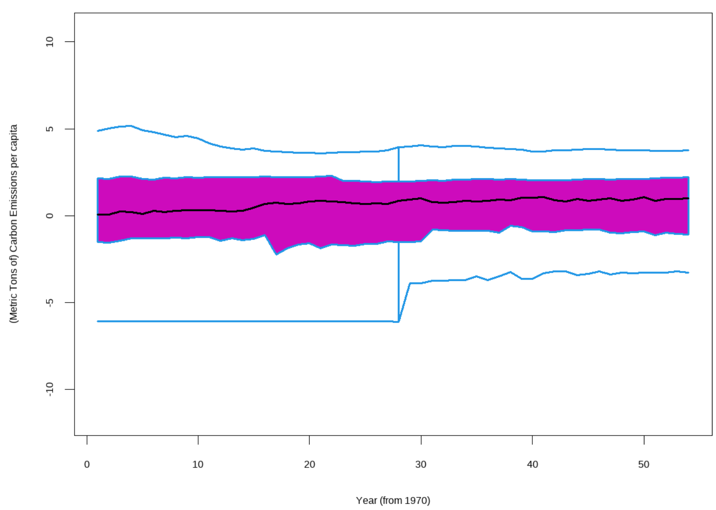


Figure 11: Carbon Emissions Functional Boxplot

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 14 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.

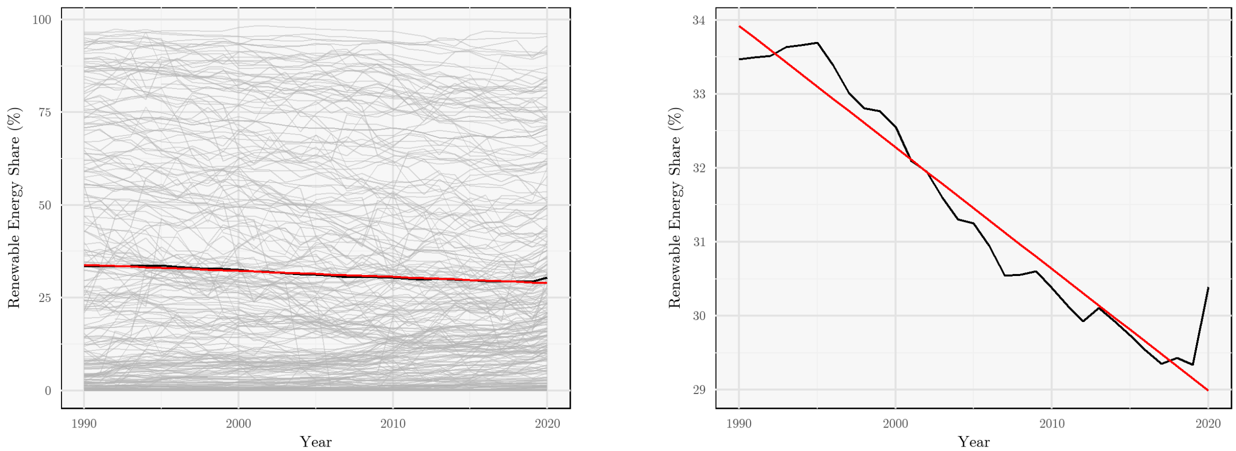


Figure 12: Mean Estimation: Cross Sectional (black) vs Penalized Smooth Splines (red). Penalized regression splines were fit using GCV.

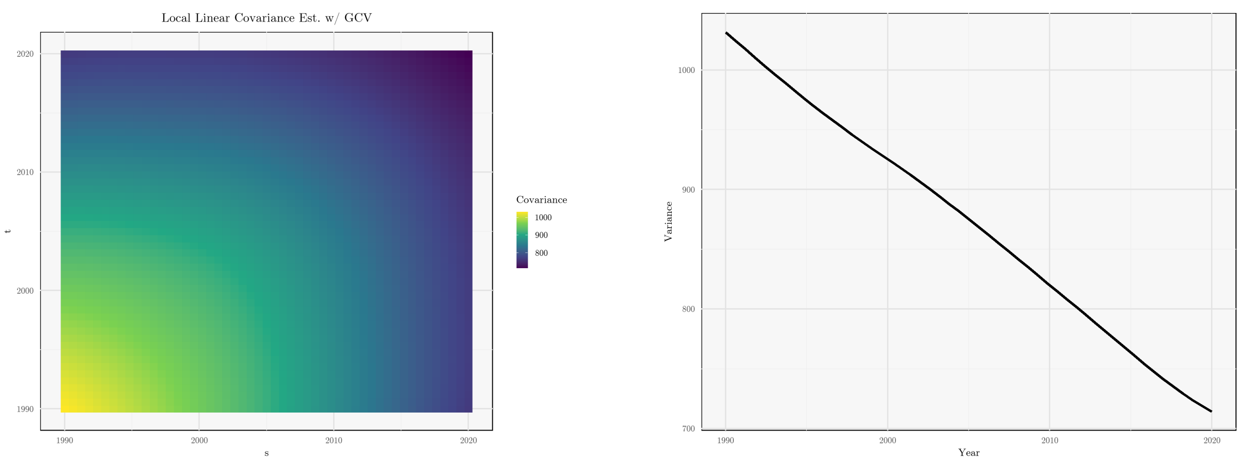


Figure 13: Local linear regression was fit to estimate the smoothed covariance. Epanechnikov kernel was used using GCV to choose bin width.

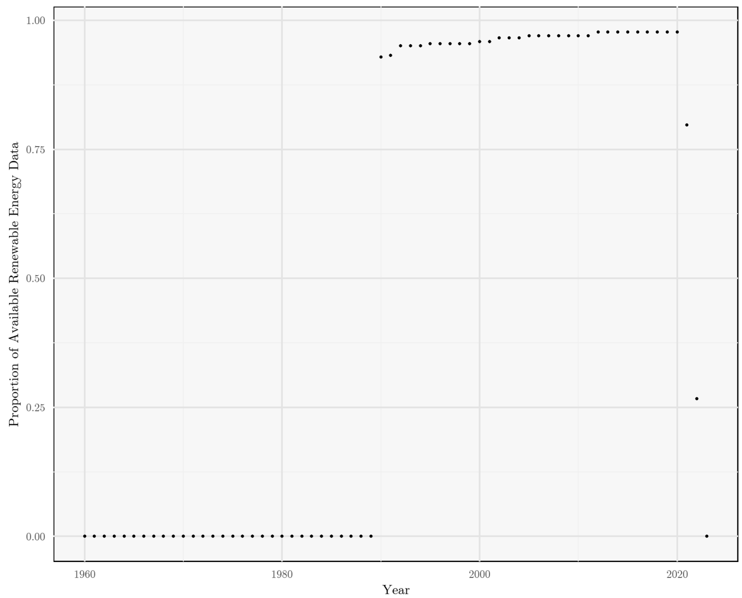


Figure 14: Proportion of data curves (as a fraction of the total number of countries in the data set) available per year.

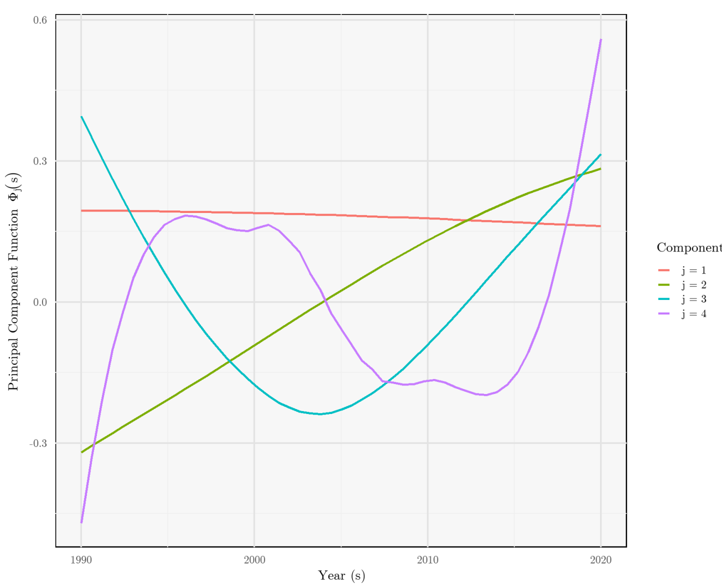


Figure 15: A plot of the first four principal component eigen functions are used extracted from the smoothed covariance estimate computed in Figure 13.

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:

```r
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 1 (Carbon Emissions data Visualization)

```r
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),
```

```r
             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 11 (functional boxplot for the Carbon Emissions data)

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

Code for Figure 7 (Mean estimation for Carbon Emissions data)

```r
#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 8 (Covariance estimates for Carbon Emissions data)

```r
#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,
```

```
                     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 10 (Variance of Carbon Emissions data)

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

Code for Figure 14 (Fraction of available curves for our Renewable Energy Consumption data)

```
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 2 (Renewable Energy Consumption data)

```
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 12 (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 13 (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 9 (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 15 (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 Figure 5 (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 6 (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.

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