

Causal Impact

Ricky Truong

June 2025

Load libraries and data

```
# Delete everything in environment
rm(list = ls())

# Load libraries
library(tidyverse)
library(readxl)
library(readr)
library(CausalImpact)
library(patchwork)

# Load data
gdp <- read_delim("GDP.csv", delim = ";",
                  escape_double = FALSE, trim_ws = TRUE)
co2 <- read_excel("CO2.xlsx")
```

Wrangle CO2 data

```
# Delete unnecessary last four columns
co2 <- subset(co2, select = -c(...55 : ...58))

# Rename columns/variables using the fourth row
names(co2) <- as.character(unlist(co2[4,]))

# Delete unnecessary rows
co2 <- co2[-c(1:4, 57),]

# Select states and years only
co2 <- co2 %>%
  select(`1970`:`2022`, State)

# Convert data to long/tidy format
co2 <- co2 %>%
  pivot_longer(cols = -State, names_to = "Year", values_to = "CO2") %>%
```

```
mutate(Year = as.integer(Year),
       CO2 = as.numeric(CO2))
```

Wrangle GDP data

```
# Delete unnecessary rows
gdp <- gdp[-c(10, 53:60),]

# Select states and years only
gdp <- gdp %>%
  select(`1997`:`2023`, GeoName)

# Convert data to long/tidy format
gdp <- gdp %>%
  pivot_longer(cols = -GeoName, names_to = "Year", values_to = "GDP") %>%
  mutate(Year = as.integer(Year),
         GDP = as.numeric(GDP))

# Rename "GeoName" to "State" for consistency
gdp <- gdp %>%
  rename(State = GeoName) %>%
  mutate(State = case_when(State == "United States" ~ "Total of states",
                           State != "United States" ~ State))
```

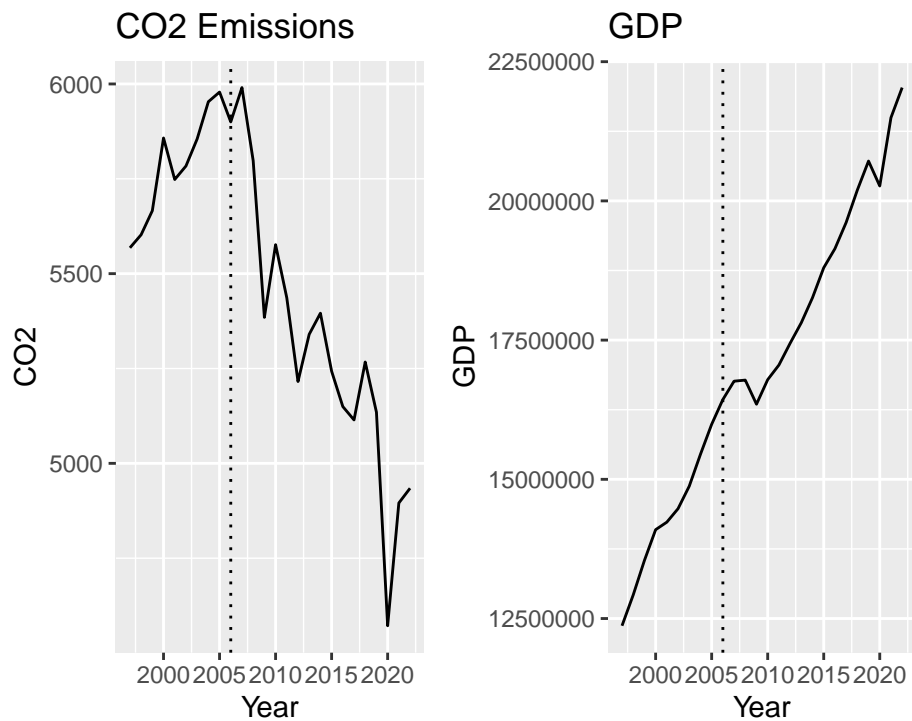
Create and visualize new combined data set

```
data <- full_join(co2, gdp, by = c("State", "Year")) %>%
  filter(Year >= 1997, Year <= 2022, State == "Total of states") %>%
  select(-State)

co2_plot <- data %>%
  ggplot(aes(x = Year, y = CO2)) +
  geom_line() +
  geom_vline(xintercept = 2006, linetype = "dotted") +
  labs(title = "CO2 Emissions")

gdp_plot <- data %>%
  ggplot(aes(x = Year, y = GDP)) +
  geom_line() +
  geom_vline(xintercept = 2006, linetype = "dotted") +
  labs(title = "GDP")

# Combine side by side
co2_plot + gdp_plot
```



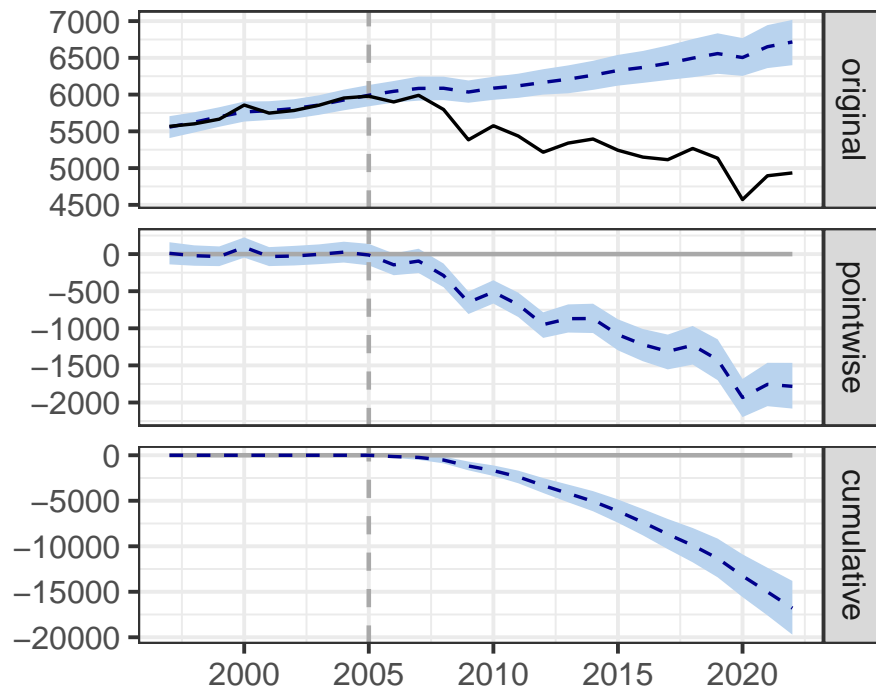
Ensure data set is a time series matrix

```
data <- data[order(data$Year), ]
ts_data <- ts(data[, c("CO2", "GDP")], start = data$Year[1])
```

Use CausalImpact to estimate causal effect

```
pre_period <- c(1997, 2005)
post_period <- c(2006, 2022)
impact <- CausalImpact(ts_data, pre_period, post_period)

plot(impact)
```



```
summary(impact)
```

```
## Posterior inference {CausalImpact}
##
##               Average          Cumulative
## Actual          5315           90348
## Prediction (s.d.) 6303 (86)      107145 (1461)
## 95% CI           [6127, 6474]    [104156, 110061]
##
## Absolute effect (s.d.) -988 (86)   -16798 (1461)
## 95% CI           [-1160, -812]    [-19713, -13809]
##
## Relative effect (s.d.) -16% (1.2%) -16% (1.2%)
## 95% CI           [-18%, -13%]    [-18%, -13%]
##
## Posterior tail-area probability p: 0.00102
## Posterior prob. of a causal effect: 99.89775%
##
## For more details, type: summary(impact, "report")
```