# Rural Windfall or a New Resource Curse? Coca, Income, and Civil Conflict in Colombia

Based on Angrist & Kugler (2008)

Maximilian Birkle Daniel Lehmann Henry Lucas

2025-10-25

## Contents

```r
knitr::opts_chunk$set(echo = TRUE, warning = FALSE, message = FALSE,
                      fig.width = 10, fig.height = 6)
# Setup
if (!require(haven)) install.packages("haven")
library (haven)
if (!require(dplyr)) install.packages("dplyr")
library(dplyr)
if (!require(foreign)) install.packages("foreign")
library(foreign)
if (!require(plm)) install.packages("plm")
library(plm)
if (!require(stargazer)) install.packages("stargazer")
library(stargazer)
if (!require(ggplot2)) install.packages("ggplot2")
library (ggplot2)
if (!require(sandwich)) install.packages("sandwich")
library (sandwich)
if (!require(lmtest)) install.packages("lmtest")
library (lmtest)
if (!require(tidyverse)) install.packages("tidyverse")
library (tidyverse)
if (!require(BART)) install.packages("BART")
library (BART)
if (!require(grf)) install.packages("grf")
library (grf)
if (!require(car)) install.packages("car")
library (car)


# Load Data and take a look at the dataset
dta <- read_delim("data00_AngristKugler.tab", delim = "\t")
```

---

# 1 Q1. Setup and Data Construction

**Tasks:**

1. Create `grow` variable (1 if `dep_ocu` $\in \{13, 18, 19, 50, 52, 86, 95, 97, 99\}$, 0 otherwise)

```
# Creating a new variable grow
department_list <- c(13, 18, 19, 50, 52, 86, 95, 97, 99)

dta <- dta %>%
  mutate(grow = ifelse(dep_ocu %in% department_list, 1, 0))
```

2. Subset data to years 1991, 1992, 1993 and 1996, 1997, 1998

```
# Subsetting the dataset to years 1991 - 1993 and 1996 - 1998
dta_subset <- dta %>%
  filter(year %in% c(1991, 1992, 1993, 1996, 1997, 1998))
```

3. Create `after` variable (1 if year $\in \{1996, 1997, 1998\}$, 0 otherwise)

```
# We create a variable called after with 1 for years 1996 - 1998
dta_subset <- dta_subset %>%
  mutate(after = ifelse(year %in% c(1996, 1997, 1998), 1, 0))
```

4. Create `growafter` variable (`grow` $\times$ `after`)

```
# Creating growafter variable (grow * after)
dta_subset <- dta_subset %>%
  mutate(growafter = grow * after)

dta_subset %>%
  count(grow, after, growafter)
```

```
## # A tibble: 4 x 4
##    grow after growafter     n
##   <dbl> <dbl>     <dbl> <int>
## 1     0     0         0  2491
## 2     0     1         0  2566
## 3     1     0         0   852
## 4     1     1         1   907
```

5. Create outcome variable: $\log\left(\frac{\text{populati+1}}{\text{violent+1}}\right)$

```
# Create outcome variable

dta_subset <- dta_subset %>%
  mutate(outcome = log((violent + 1) / (populati + 1)))
```

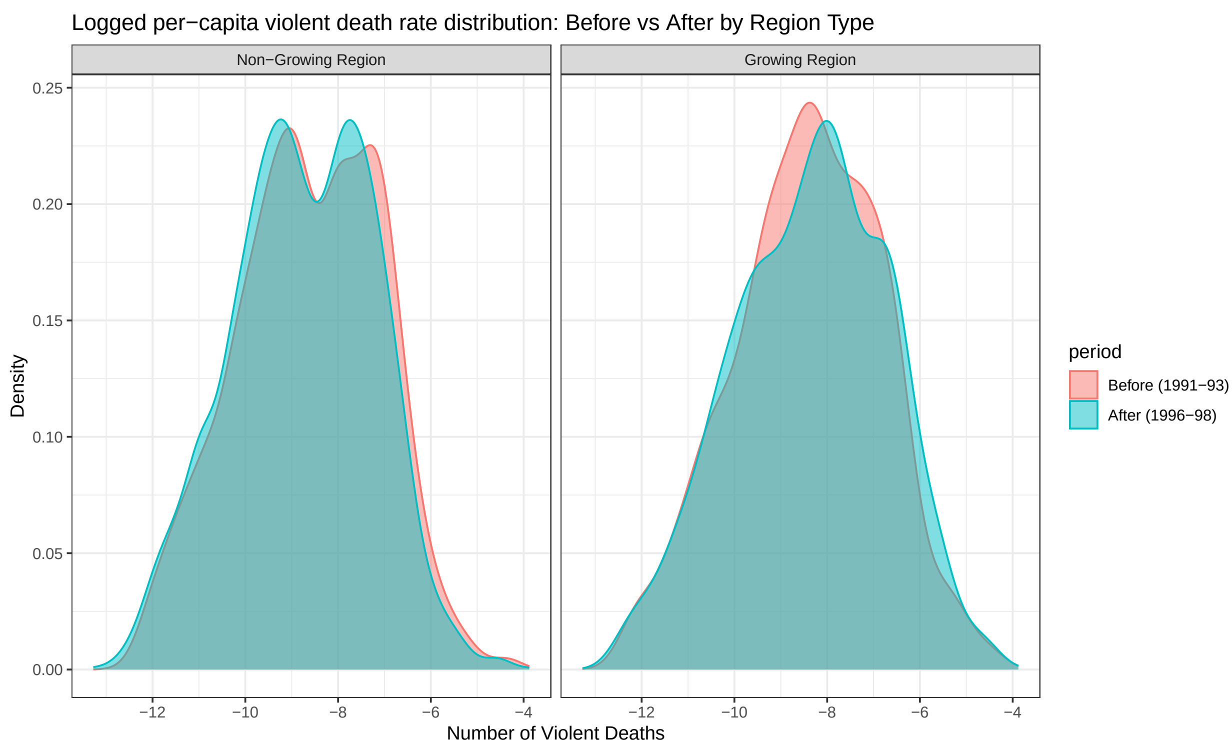
---

# 2 Q2. Visualizing Violence Before and After

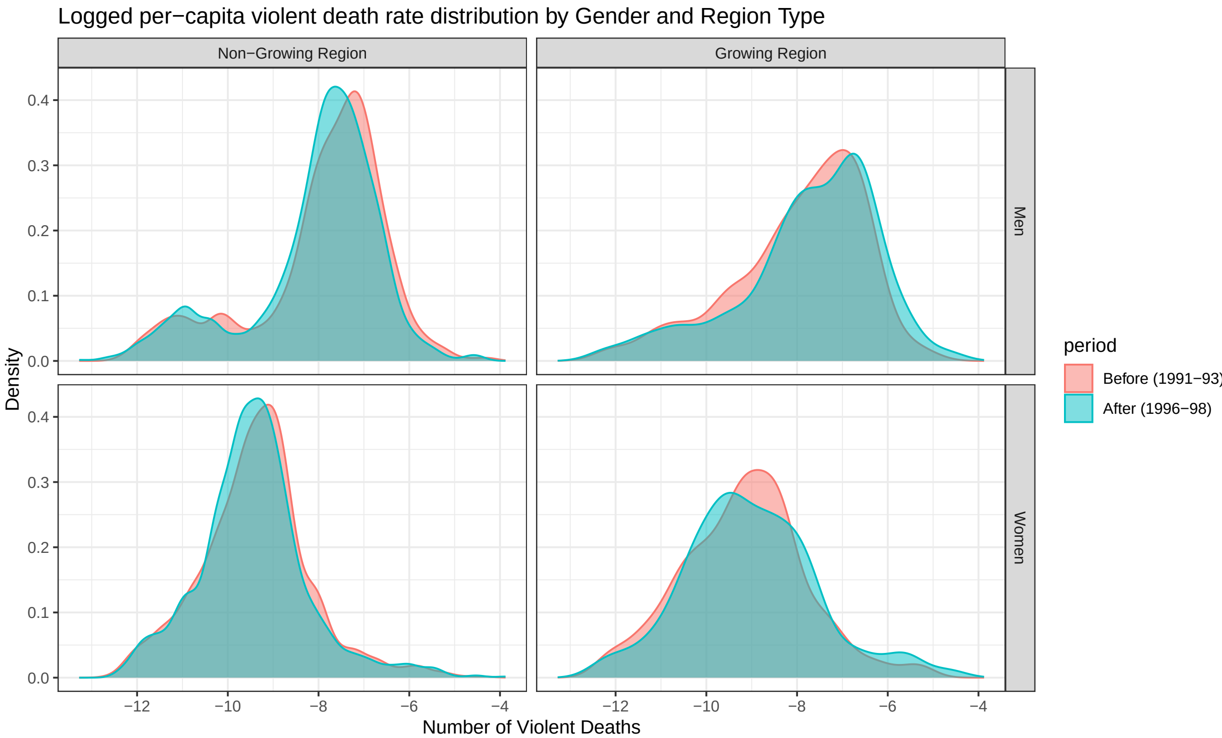
**Tasks:**

1. Create density plots for non-growing vs. growing regions, before vs. after
2. Extend to $2 \times 2$ grid by gender (men: `sex=1`, women: `sex=2`)
3. Interpret: Evidence of shifts in violence? Different by gender?

```r
# Density plots: Non-growing vs. Growing regions

density_plot <- dta_subset %>%
  mutate(
    period = factor(after, levels = c(0, 1), labels = c("Before (1991-93)", "After (1996-98)")),
    region_type = factor(grow, levels = c(0, 1), labels = c("Non-Growing Region", "Growing Region")),
    gender = factor(sex, levels = c(1, 2), labels = c("Men", "Women"))
  )


density_plot %>%
  ggplot(aes(x = outcome, fill = period, color = period)) +
  geom_density(alpha = 0.5) +
  facet_wrap(~ region_type) +
  labs(title = "Logged per-capita violent death rate distribution: Before vs After by Region Type",
       x = "Number of Violent Deaths",
       y = "Density") +
  theme_bw()
```




```r
# 2x2 grid: Top=men, Bottom=women; Left=non-growing, Right=growing

density_plot %>%
  filter(!is.na(gender)) %>%
  ggplot(aes(x = outcome, fill = period, color = period)) +
  geom_density(alpha = 0.5) +
  facet_grid(gender ~ region_type) +
  labs(title = "Logged per-capita violent death rate distribution by Gender and Region Type",
       x = "Number of Violent Deaths",
       y = "Density") +
  theme_bw()
```




**Interpretation:**

The evidence indicates a significant shift in violence following the air-bridge disruption, with the effect being highly specific to both region and gender. For the treatment group (**coca-growing regions**), there was a slight increase in the per-capita violent death rate among men. In contrast, the control group (**non-growing regions**) showed no meaningful change for either gender, which suggests that the increase in violence was not due to a nationwide trend. Furthermore, the pattern seems to be strongly gendered; the effect on women in the treatment group was minimal compared to a much larger effect on men. This suggests that the impact of the coca boom on violence was almost exclusively concentrated among the male population, who seem to have been the primary participants in the conflict.

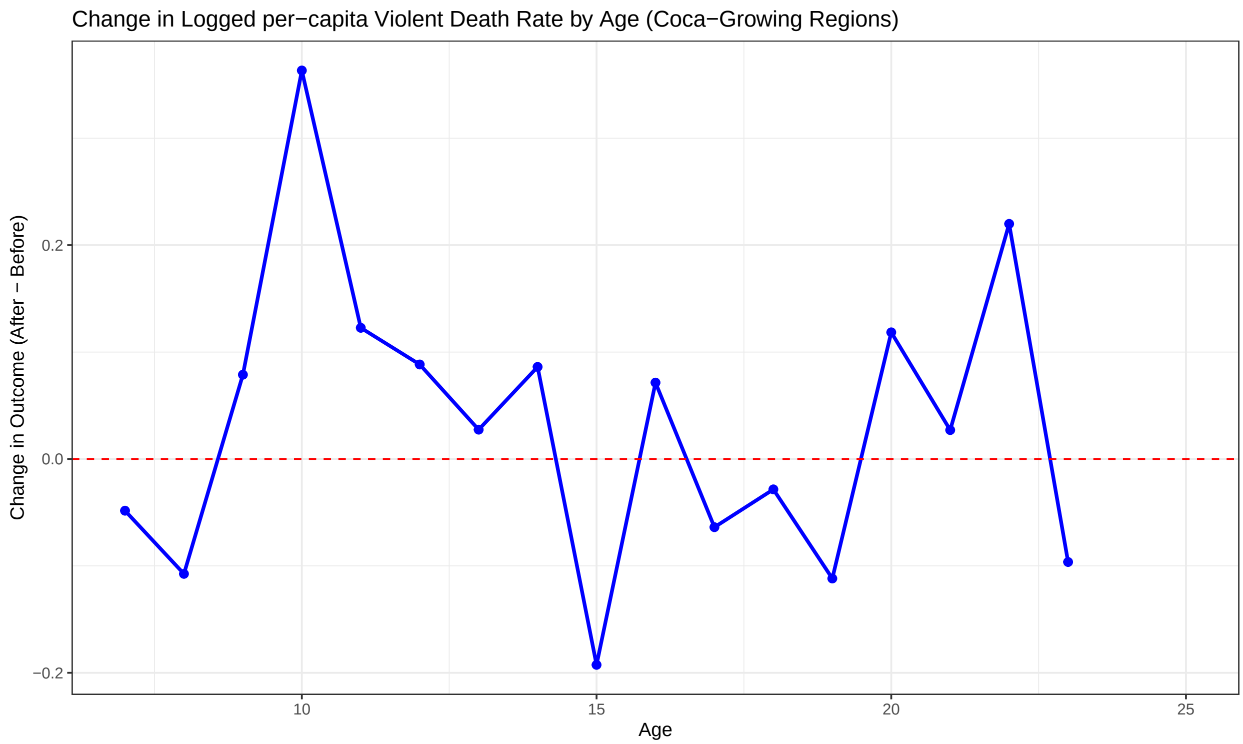
---

# 3 Q3. Age-Specific Effects

**Task:** For coca-growing regions only, plot the change in outcome (after - before) by age group.

```r
# Calculate mean difference by age for growing regions only

age_effects <- dta_subset %>%
  filter(grow == 1) %>%
  group_by(age, after) %>%
  summarise(mean_outcome = mean(outcome, na.rm = TRUE), .groups = "drop") %>%
  pivot_wider(names_from = after, values_from = mean_outcome, names_prefix = "period_") %>%
  mutate(change = period_1 - period_0)

# Plot age-specific effects
age_effects %>%
  ggplot(aes(x = age, y = change)) +
```

```
  geom_line(color = "blue", size = 1) +
  geom_point(color = "blue", size = 2) +
  geom_hline(yintercept = 0, linetype = "dashed", color = "red") +
  labs(title = "Change in Logged per-capita Violent Death Rate by Age (Coca-Growing Regions)",
       x = "Age",
       y = "Change in Outcome (After - Before)") +
  theme_bw()
```



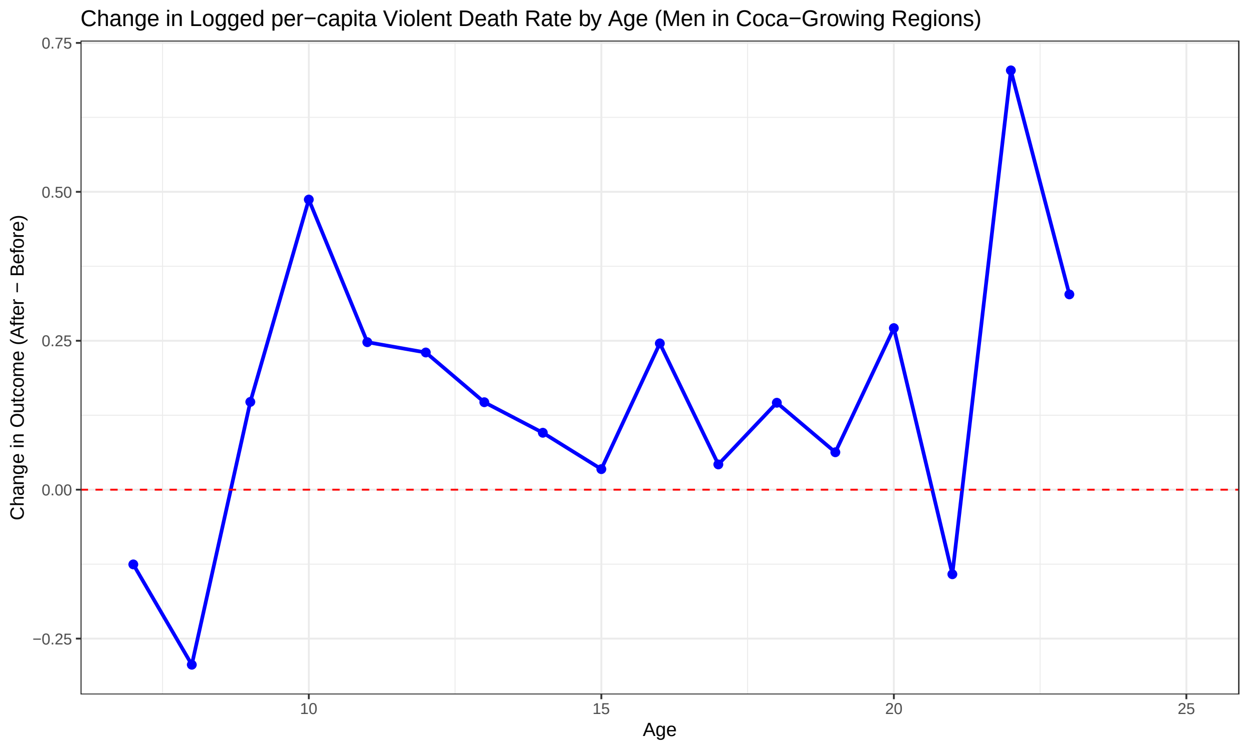

**Interpretation:**

This plot illustrates how the average violence rate changes for each age group by calculating the difference between the average violence rate for an age group after the air-bridge disruption and the average violence rate for that same group before the disruption. Clearly, the plot shows that the effect varies dramatically across different age groups, with a concentration in certain age brackets. The increase in conflict and violence seems to have disproportionately affected individuals of fighting age, as can be seen from the peaks in the graph between the ages of 17 and 22. Additionally, younger adolescents around the age of 10 seem to have been particularly involved in violent activities, resulting in a much higher death rate after the air disruption. However, it is important to note that, for a few age groups — especially those around 15 years old — the rate of violence decreased after the air disruption.

This seemingly volatile distribution across age groups reveals a key insight. If the increase in violence was solely due to increased criminal activity, we would most likely see it concentrated among cohorts that met the age of fighting and above. However, since we cleary see that also much younger age groups show an increase in violent deaths, we can conclude that this is mostly caused by conflict-related violence (e.g. civil war) after the 'air bridge' intervention, which corresponds to the paper's findings.

To explore this interesting trend, we created a graph for the per-capita violent death rate in coca-growing regions before and after for men only. What we can see here is that the increase in violence was not confined by the typical "fighting age" bracket but was high across the entire young male distribution, including children around 10. This pattern strongly suggests widespread, conflict-related violence. The coca boom fueled this by creating two distinct sets of victims: older cohorts were recruited as soldiers, while younger boys, drawn in as laborers, became collateral damage in the armed groups' fight to control the coca fields and labor force. The overall increased death rate of civillians may also have driven this trend.

```
age_effects_men <- dta_subset %>%
  filter(grow == 1, sex == 1) %>%
  group_by(age, after) %>%
  summarise(
    mean_outcome = mean(outcome, na.rm = TRUE),
    .groups = "drop"
  ) %>%
  pivot_wider(
    names_from = after,
    values_from = mean_outcome,
    names_prefix = "period_"
  ) %>%
  mutate(change = period_1 - period_0)

# Plot for age specific effects
age_effects_men %>%
  ggplot(aes(x = age, y = change)) +
  geom_line(color = "blue", size = 1) +
  geom_point(color = "blue", size = 2) +
  geom_hline(
    yintercept = 0,
    linetype = "dashed",
    color = "red"
  ) +
  labs(
    title = "Change in Logged per-capita Violent Death Rate by Age (Men in Coca-Growing Regions)",
    x = "Age",
    y = "Change in Outcome (After - Before)"
  ) +
  theme_bw()
```




---

# 4 Q4. Testing the Parallel Trends Assumption (Pre-Treatment)

**Tasks:**

1. Use pre-treatment years (1990-1993)
2. Estimate: outcome $= \alpha + \beta \cdot \text{year} + \gamma \cdot \text{grow} + \delta \cdot (\text{grow} \times \text{year}) + u$
3. Test if `grow` $\times$ `year` interactions are jointly zero (year as linear and categorical)
4. Create graph of average outcome by year and group

```
# Subset to pre-treatment years (1990-1993)
dta_pretreatment <- dta %>%
  filter(year %in% c(1990, 1991, 1992, 1993)) %>%
  mutate(grow = ifelse(dep_ocu %in% department_list, 1, 0),
         outcome = log((violent + 1) / (populati + 1)))
```

```
# Model with year as linear
year_linear <- lm(outcome ~ year + grow + year:grow, data = dta_pretreatment)

# Model with year as categorical (factor)
year_factor <- lm(outcome ~ factor(year) + grow + factor(year):grow, data = dta_pretreatment)

# Stargazer table for better visualization

# 1. Define the professional labels for your variables.
pro_labels <- c(
  "Coca-Growing Region",              # for 'grow'
  "Year (Linear) &times; Growing Region", # for 'year:grow' (use &times; for HTML)
  "1991 &times; Growing Region",          # for 'factor(year)1991:grow'
```