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

2025-10-19

Contents


```
knitr::opts_chunk$set(echo = TRUE, warning = FALSE, message = FALSE,
                      fig.width = 10, fig.height = 6)
# Setup
library(tidyverse)
library(haven)
library(knitr)
library(kableExtra)
library(broom)
library(lmtest)
library(car)
library(gridExtra)


# Load Data and take a look at the dataset
```

```r
data <- read_delim("data00_AngristKugler.tab", delim = "\t")
glimpse(data)
```

```
## Rows: 12,544
## Columns: 11
## $ dep_ocu  <dbl> 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5~
## $ year     <dbl> 1990, 1990, 1990, 1990, 1990, 1990, 1990, 1990, 1990, 1990, 1~
## $ sex      <dbl> 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1~
## $ age      <dbl> 7, 7, 8, 8, 9, 9, 10, 10, 11, 11, 12, 12, 13, 13, 14, 14, 15,~
## $ death    <dbl> 1440, 1110, 155, 105, 221, 106, 1696, 256, 2697, 342, 2329, 2~
## $ violent  <dbl> 10, 8, 11, 4, 81, 19, 1479, 138, 2345, 184, 1976, 123, 1317, ~
## $ disease  <dbl> 1335, 1041, 61, 65, 62, 58, 85, 84, 143, 113, 148, 119, 163, ~
## $ accident <dbl> 94, 61, 82, 36, 77, 29, 131, 34, 208, 45, 204, 25, 173, 20, 1~
## $ violent_ <dbl> 11, 8, 12, 4, 82, 19, 1480, 138, 2346, 184, 1977, 125, 1319, ~
## $ homicide <dbl> 10, 8, 10, 4, 78, 16, 1463, 120, 2296, 168, 1943, 117, 1289, ~
## $ populati <dbl> 555689, 555689, 520770, 520770, 492630, 492630, 469523, 46952~
```

---

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)

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

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

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

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

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

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

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

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

# Confirm that both are coded correctly
data_subset %>%
  filter(grow == 1) %>%
  distinct(dep_ocu) %>%
  arrange(dep_ocu)
```

```
## # A tibble: 9 x 1
##   dep_ocu
```

```
##     <dbl>
## 1      13
## 2      18
## 3      19
## 4      50
## 5      52
## 6      86
## 7      95
## 8      97
## 9      99
```

```r
data_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
```

```r
# First Table: Grow takes the value of 1 only for defined departments.
# Second Table: 'growafter' only takes the value of 1 if grow and after are 1.
```

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

```r
# Create outcome variable (it makes more sense if it is death over population)
data_subset <- data_subset %>%
  mutate(outcome = log((violent + 1) / (populati + 1)))
```

---

2 Q2. Visualizing Violence Before and After

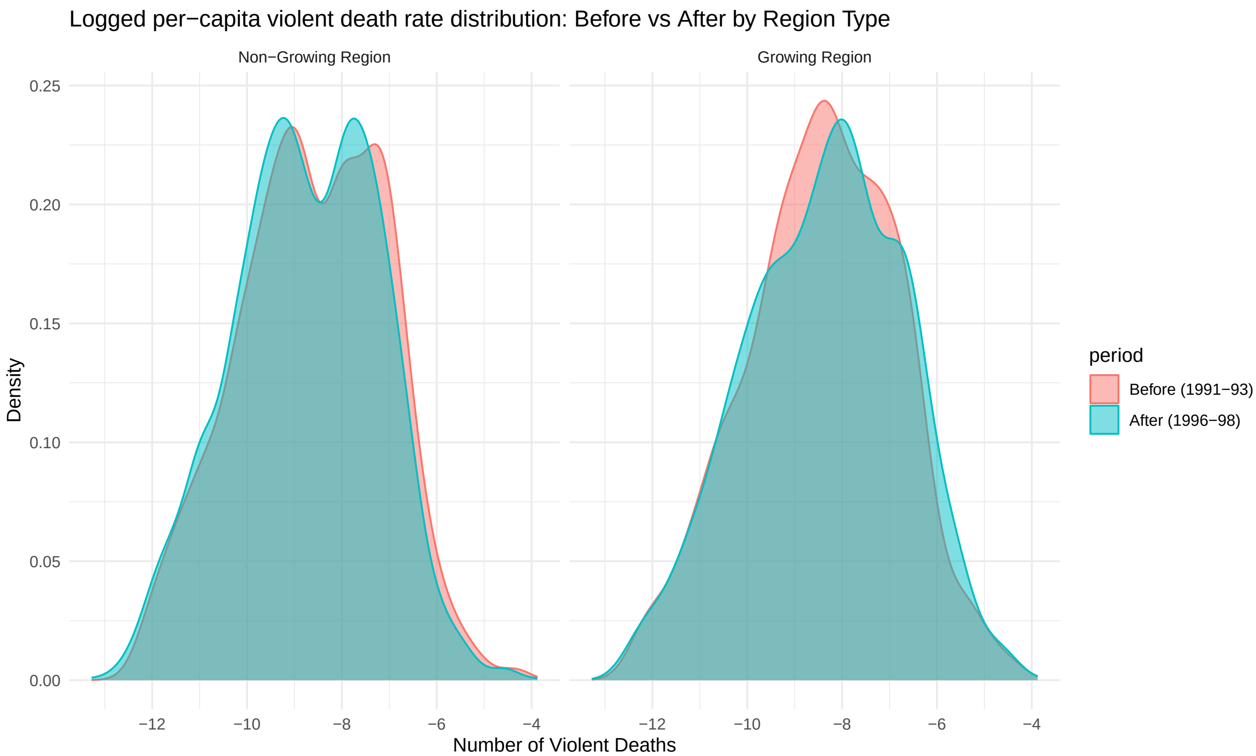
Tasks:

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

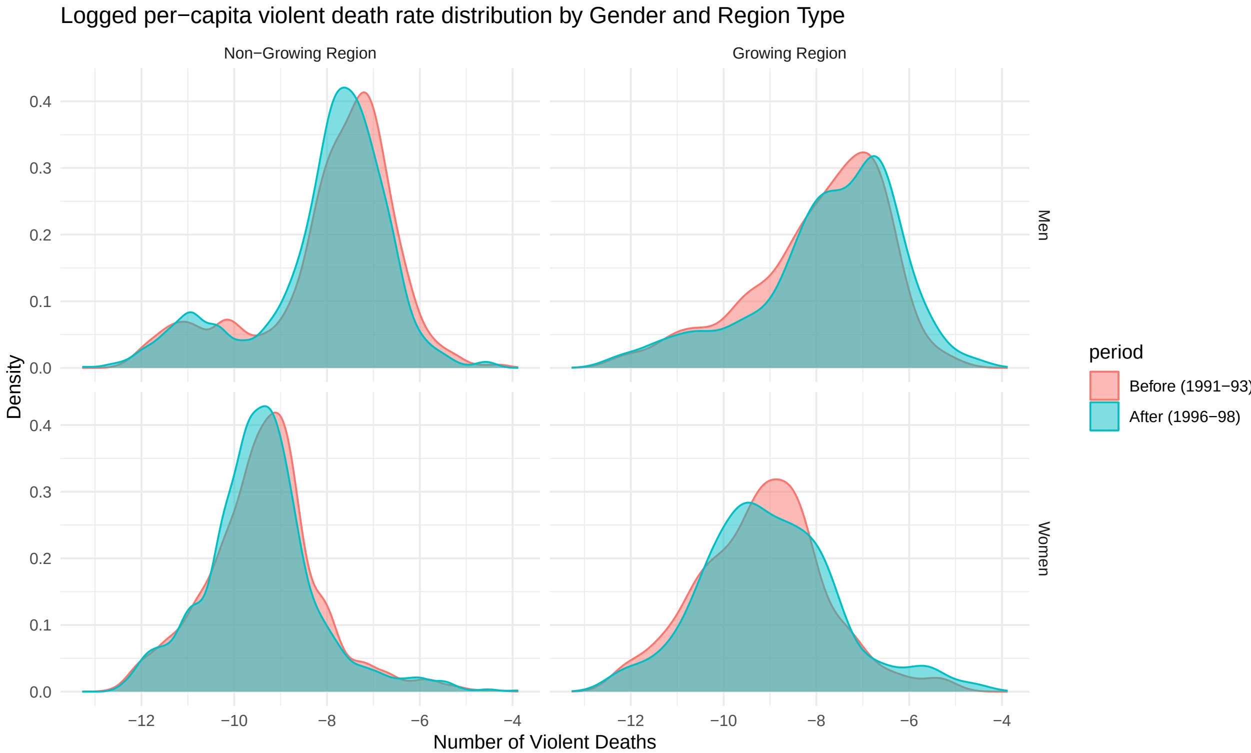
```r
# Setup
density_plot <- data_subset %>%
  mutate(
    period = factor(after, levels = c(0, 1), labels = c("Before (1991-93)", "After (1996-98)")),
    region_type = factor(grow, levels = c(0, 1), labels = c("Non-Growing Region", "Growing Region")),
    gender = factor(sex, levels = c(1, 2), labels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


# Density plots: Non-growing vs. Growing regions
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_minimal()
```




```
# 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_minimal()
```



Interpretation:

[Your interpretation here]

---

3 Q3. Age-Specific Effects

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

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


# Plot age-specific effects
```

Interpretation:

[Does the effect vary by age?]

---

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` × `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)
```

```r
# Model with year as linear
```

```r
# Model with year as categorical (factor)
```

```r
# Test if grow×year interactions are jointly zero
```

```r
# Graph: Average outcome by year and group
```

Interpretation:

[What do the p-values tell us about parallel trends?]

---

5 Q5. Placebo DiD Test

Tasks:

1. Create `placebo_after` (1 if year = 1992 or 1993, 0 if year = 1990 or 1991)
2. Estimate placebo DiD model
3. Interpret `placebo_after` × `grow` coefficient

```r
# Subset and create placebo variables
```

```r
# Estimate placebo DiD model
```

Interpretation:

[Should the placebo effect be significant? What would significance suggest?]

---

6 Q6. Covariate Balance at Time 0

Task: Compare treatment and control regions on `age`, `sex`, and `populati` using pre-treatment data.

```r
# Create balance table
```

```r
# Optional: Standardized difference plot
```

Discussion:

[Why is covariate balance critical? What would imbalance imply?]

---

7 Q7. Why Covariate Balance Matters

Discussion Questions:

1. If covariates are balanced at time 0, what does this imply about confounding?
2. What role do these variables play after assignment?
3. If violence trends already differ before treatment, how might this bias DiD?

[Your answers here]

---

8 Q8. Covariate Timing and Post-Treatment Bias

Discussion Questions:

1. Should we include covariates from time 0, time 1, or both?
2. What happens if you include a covariate measured after treatment?
3. When might adjusting for post-treatment variables be appropriate?

[Your answers here]

---

9 Q9. Computing the DiD Estimate

Task: Compute manual DiD estimate.

```
# Mean difference (after - before) for grow=1


# Mean difference (after - before) for grow=0


# DiD estimate: subtract the two
```

Interpretation:

[Why is DiD preferable to simple before-after comparison?]

---

10 Q10. Regression Form of DiD

Tasks:

1. Estimate: $\text{outcome} = \beta_0 + \beta_1 \cdot \text{after} + \beta_2 \cdot \text{grow} + \beta_3 \cdot (\text{after} \times \text{grow}) + u$
2. Report β_3 and p-value
3. Show analytically that β_3 equals the manual DiD estimate

```
# DiD regression model
```

Interpretation:

[What does β_3 tell us about the causal effect?]

Analytical proof:

[Show that $\beta_3 = (\bar{Y}_{1,1} - \bar{Y}_{1,0}) - (\bar{Y}_{0,1} - \bar{Y}_{0,0})$]

---

11 Q11. Adding Covariates

Task: Estimate three models and compare.

```
# Model 1: outcome ~ grow + after + growafter
```

```
# Model 2: Add age and sex
```

```
# Model 3: Add age, sex, and populati
```

```
# Compare models
```

Discussion:

[Does β_3 change? Do covariates matter? Which specification is most credible?]

---

12 Q12. Interpretation and Reflection

Summary:

1. Did violence increase or decrease after the air-bridge disruption?
2. Does the evidence support a "resource-curse" interpretation?
3. What are the remaining identification threats?

[Your final interpretation here]