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

2025-10-24

Contents


```
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)
```

```r
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)

```r
# 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

```r
# 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)

```r
# 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`)

```r
# 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: $\log\left(\frac{\text{populati}+1}{\text{violent}+1}\right)$

```r
# Create outcome variable

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

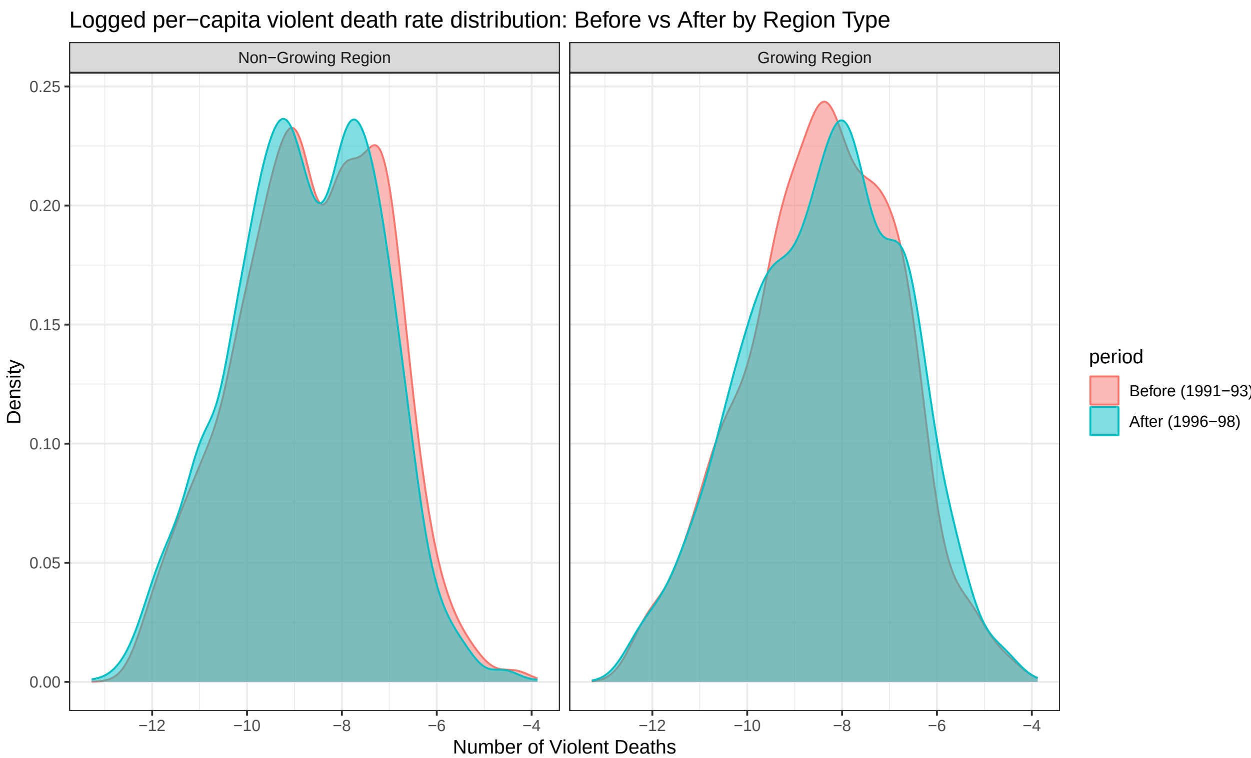
---

2 Q2. Visualizing Violence Before and After

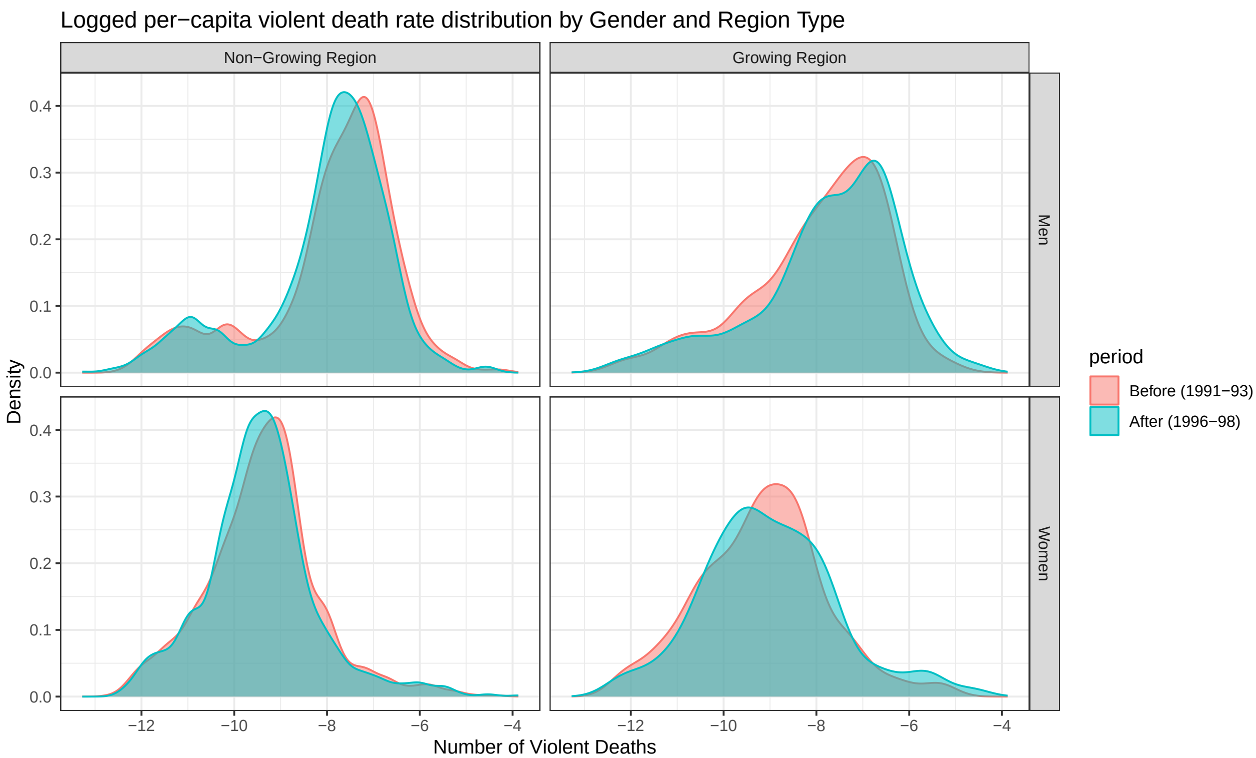
Tasks:

1. Create density plots for non-growing vs. growing regions, before vs. after
2. Extend to 2×2 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.

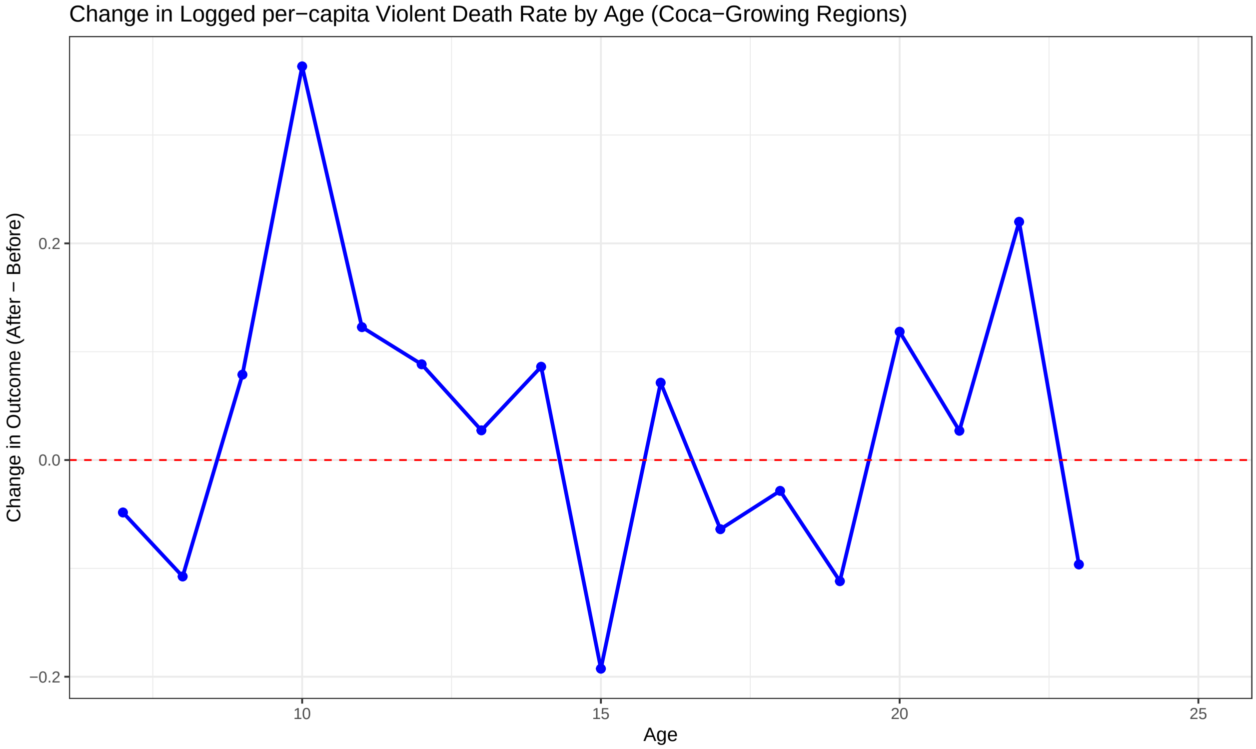
---

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) +
```

```r
  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 heavy 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.

---

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

```r
# Subset to pre-treatment years (1990-1993)
dta_pretreatment <- dta %>%
```

```r
  filter(year %in% c(1990, 1991, 1992, 1993)) %>%
  mutate(grow = ifelse(dep_ocu %in% department_list, 1, 0),
         outcome = log((violent + 1) / (populati + 1)))
```

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

summary(year_linear)
```

```
##
## Call:
## lm(formula = outcome ~ year + grow + year:grow, data = dta_pretreatment)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.972 -1.057  0.029  1.214  4.494
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.469e+01  4.921e+01  -0.299    0.765
## year         3.043e-03  2.471e-02   0.123    0.902
## grow         1.469e+00  9.748e+01   0.015    0.988
## year:grow   -6.358e-04  4.895e-02  -0.013    0.990
##
## Residual standard error: 1.552 on 4221 degrees of freedom
##   (238 observations deleted due to missingness)
## Multiple R-squared:  0.003242,   Adjusted R-squared:  0.002534
## F-statistic: 4.577 on 3 and 4221 DF,  p-value: 0.003328
```

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

summary(year_factor)
```

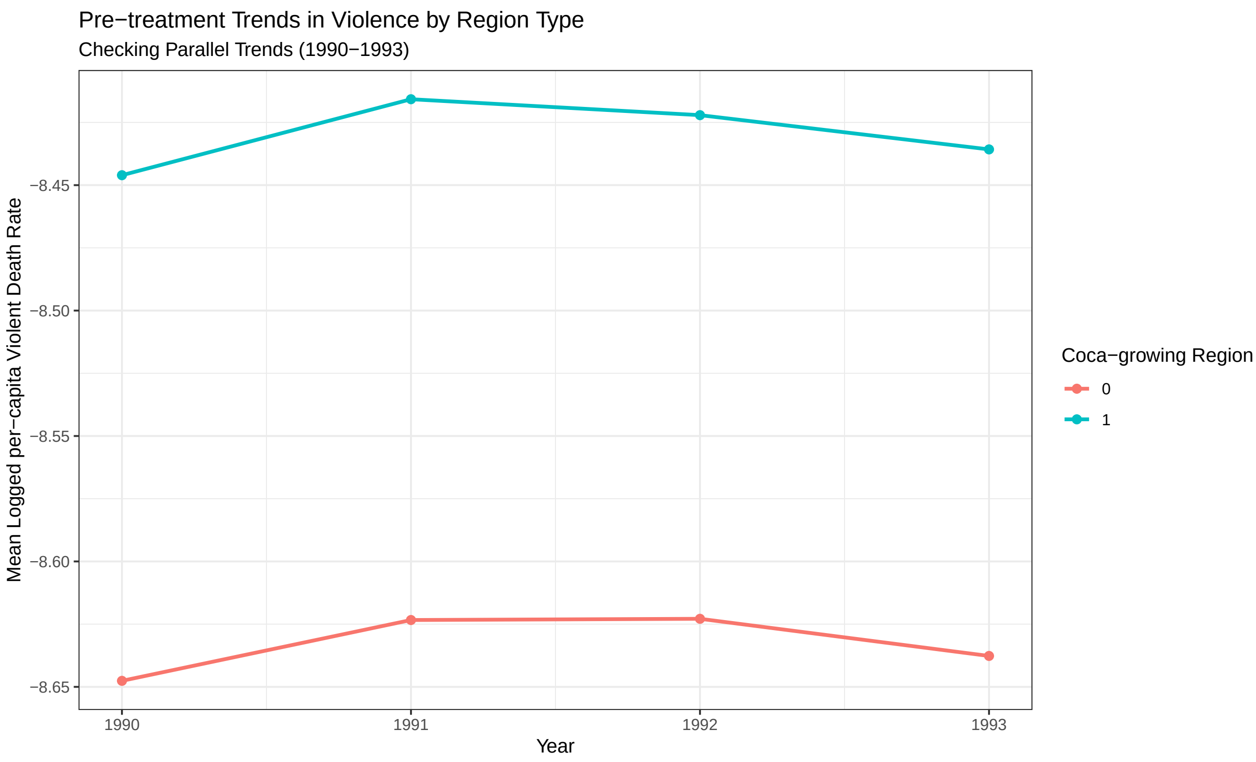
```
##
## Call:
## lm(formula = outcome ~ factor(year) + grow + factor(year):grow,
##     data = dta_pretreatment)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.9625 -1.0596  0.0296  1.2174  4.5045
##
## Coefficients:
##                         Estimate Std. Error  t value Pr(>|t|)
## (Intercept)           -8.6476041  0.0551588 -156.776   <2e-16 ***
## factor(year)1991       0.0242450  0.0780805    0.311    0.756
## factor(year)1992       0.0247314  0.0782302    0.316    0.752
## factor(year)1993       0.0099279  0.0781551    0.127    0.899
## grow                   0.2015622  0.1100062    1.832    0.067 .
## factor(year)1991:grow  0.0060478  0.1560531    0.039    0.969
## factor(year)1992:grow -0.0007938  0.1549368   -0.005    0.996
## factor(year)1993:grow  0.0003595  0.1547948    0.002    0.998
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 1.552 on 4217 degrees of freedom
##   (238 observations deleted due to missingness)
## Multiple R-squared:  0.003286,   Adjusted R-squared:  0.001631
## F-statistic: 1.986 on 7 and 4217 DF,  p-value: 0.05322
```

```
# Test if grow×year interactions are jointly zero
linearHypothesis(year_factor, matchCoefs(year_factor, ":grow"))
```

```
##
## Linear hypothesis test:
## factor(year)1991:grow = 0
## factor(year)1992:grow = 0
## factor(year)1993:grow = 0
##
## Model 1: restricted model
## Model 2: outcome ~ factor(year) + grow + factor(year):grow
##
##   Res.Df   RSS Df Sum of Sq      F Pr(>F)
## 1   4220 10162
## 2   4217 10162  3 0.0058328 8e-04      1
```

```
# Graph: Average outcome by year and group
dta_pretreatment %>%
  group_by(year, grow) %>%
  summarise(mean_outcome = mean(outcome, na.rm = TRUE), .groups = "drop") %>%
  ggplot(aes(x = year, y = mean_outcome, color = factor(grow))) +
  geom_line(size = 1) +
  geom_point(size = 2) +
  labs(
    title = "Pre-treatment Trends in Violence by Region Type",
    subtitle = "Checking Parallel Trends (1990-1993)",
    x = "Year",
    y = "Mean Logged per-capita Violent Death Rate",
    color = "Coca-growing Region"
  ) +
  theme_bw()
```



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

```
# Subset and create placebo variables
dta_placebo <- 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)),
    placebo_after = ifelse(year %in% c(1992, 1993), 1, 0)
  )
```

```
# Estimate placebo DiD model
placebo_did <- lm(outcome ~ placebo_after + grow + placebo_after:grow, data = dta_placebo)
summary(placebo_did)
```

```
##
## Call:
## lm(formula = outcome ~ placebo_after + grow + placebo_after:grow,
##     data = dta_placebo)
##
```

```
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -3.9699 -1.0574  0.0288  1.2145  4.4924 
## 
## Coefficients:
##                     Estimate Std. Error  t value Pr(>|t|)    
## (Intercept)        -8.635505   0.039023 -221.295  < 2e-16 ***
## placebo_after       0.005216   0.055292    0.094  0.92485    
## grow                0.204494   0.077990    2.622  0.00877 ** 
## placebo_after:grow -0.003148   0.109627   -0.029  0.97710    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 1.552 on 4221 degrees of freedom
##   (238 observations deleted due to missingness)
## Multiple R-squared:  0.00324,    Adjusted R-squared:  0.002532 
## F-statistic: 4.574 on 3 and 4221 DF,  p-value: 0.003342
```

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
dta_balance <- dta_subset %>%
  filter(after == 0)

balance_table <- dta_balance %>%
  group_by(grow) %>%
  summarise(
    mean_age = mean(age, na.rm = TRUE),
    mean_sex = mean(sex, na.rm = TRUE),
    mean_pop = mean(populati, na.rm = TRUE),
    .groups = "drop"
  )

balance_table
```

```
## # A tibble: 2 x 4
##    grow mean_age mean_sex mean_pop
##   <dbl>    <dbl>    <dbl>    <dbl>
## 1     0     15.6     1.51   78301.
## 2     1     15.5     1.59   40541.
```

```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.

```r
# Mean difference (after - before) for grow=1 and grow=0
did_table <- dta_subset %>%
  group_by(grow, after) %>%
  summarise(mean_outcome = mean(outcome, na.rm = TRUE), .groups = "drop") %>%
  pivot_wider(names_from = after, values_from = mean_outcome, names_prefix = "time") %>%
  mutate(diff = time1 - time0)

# DiD estimate: subtract the two
did_estimate <- diff(did_table$diff)
did_estimate
```

```
## [1] 0.1970235
```

Interpretation:

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

---

10 Q10. Regression Form of DiD

Tasks:

1. Estimate: 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

```r
# DiD regression model
did <- lm(outcome ~ after + grow + after:grow, data = dta_subset)
summary(did)
```

```
##
## Call:
## lm(formula = outcome ~ after + grow + after:grow, data = dta_subset)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.5075 -1.0675  0.0344  1.2000  4.8920
##
## Coefficients:
##             Estimate Std. Error  t value Pr(>|t|)
## (Intercept) -8.62797    0.03193 -270.247  < 2e-16 ***
## after       -0.14132    0.04485   -3.151  0.00163 **
## grow         0.20330    0.06317    3.218  0.00130 **
## after:grow   0.19702    0.08820    2.234  0.02554 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.55 on 6445 degrees of freedom
##   (367 observations deleted due to missingness)
## Multiple R-squared:  0.008903,   Adjusted R-squared:  0.008441
## F-statistic:  19.3 on 3 and 6445 DF,  p-value: 1.877e-12
```

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.

```r
# Model 1: outcome ~ grow + after + growafter
cov_did1 <- lm(outcome ~ after + grow + after:grow, data = dta_subset)
```

```r
# Model 2: Add age and sex
cov_did2 <- lm(outcome ~ after + grow + after:grow + age + sex, data = dta_subset)
```

```r
# Model 3: Add age, sex, and populati
cov_did3 <- lm(outcome ~ after + grow + after:grow + age + sex + populati, data = dta_subset)
```

```r
# Compare models
stargazer(cov_did1, cov_did2, cov_did3,      # List your models
          type = "text",                    # Output type: "text", "html", or "latex"
          title = "Regression Results: The Effect of Treatment on Outcome",
          align = TRUE,                     # Aligns numbers on decimal points
          dep.var.labels = "Outcome",       # A clean name for the dependent variable
          column.labels = c("Base Model", "Adds Demographics", "Full Model"),
          covariate.labels = c("After Treatment", "Treatment Group (Grow)", "Age",
                               "Sex", "Population", "Interaction: After x Grow"),
          notes = "Standard errors are in parentheses.",
          notes.align = "l")
```

```
##
```

```
## Regression Results: The Effect of Treatment on Outcome
## ==========================================================================================
##                                                        Dependent variable:
##                             -------------------------------------------------------------
##                                                              Outcome
##                                    Base Model          Adds Demographics            Full Model
##                                       (1)                     (2)                      (3)
## ------------------------------------------------------------------------------------------
## After Treatment                    -0.141***               -0.123***                -0.107***
##                                     (0.045)                 (0.036)                  (0.036)
##
## Treatment Group (Grow)              0.203***                0.304***                 0.231***
##                                     (0.063)                 (0.051)                  (0.051)
##
## Age                                                         0.142***                 0.123***
##                                                             (0.003)                  (0.004)
##
## Sex                                                        -1.057***                -1.057***
##                                                             (0.028)                  (0.028)
##
## Population                                                                         -0.00000***
##                                                                                      (0.00000)
##
## Interaction: After x Grow           0.197**                 0.131*                   0.123*
##                                     (0.088)                 (0.071)                  (0.071)
##
## Constant                           -8.628***               -9.162***                -8.731***
##                                     (0.032)                 (0.070)                  (0.079)
##
## ------------------------------------------------------------------------------------------
## Observations                         6,449                   6,449                    6,449
## R2                                   0.009                   0.350                    0.363
## Adjusted R2                          0.008                   0.350                    0.362
## Residual Std. Error            1.550 (df = 6445)        1.256 (df = 6443)        1.244 (df = 6442)
## F Statistic                 19.298*** (df = 3; 6445) 693.978*** (df = 5; 6443) 610.532*** (df = 6; 644
## ==========================================================================================
## Note:                       *p<0.1; **p<0.05; ***p<0.01
##                             Standard errors are in parentheses.
```

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]