Lecture 4 - Identification: Fixed Effects

Topics in Econometrics

Vincent Bagilet

2025-09-24

Goal of the session

Outline of the course

1. Overview and fundamental hurdles
2. Simulations
3. Design: beyond identification
4. **Design: identification (fixed effects and related)**
5. Data visualization
6. Design: identification (IV and RDD)
7. Modelling
8. Analysis

Goal of the session

- **Fixed effects** are extremely common in applied economics
- What are they really doing?
- More generally, what are we really estimating in a specific model?
- What are we comparing to what?
- Where does the identifying variation come from?

Notes on Potential Outcomes

Potential outcomes framework

- Let's denote $D_i \in \{0, 1\}$, the treatment status, Y_i, the realized outcome, Y^0 and Y^1 the potential outcomes

Individual Treatment Effects (TEs)	$Y_i^1 - Y_i^0, \forall i$	*What we would ideally estimate*
Average Treatment Effects (ATE)	$\mathbb{E}[Y_i^1 - Y_i^0]$	*What we reasonably want to estimate*
Average Treatment Effects on the Treated (ATT)	$\mathbb{E}[Y_i^1 - Y_i^0 \mid D_i = 1]$	*What we reasonably want to estimate*
Difference in average observed outcomes	$\mathbb{E}[Y_i \mid D_i = 1] - \mathbb{E}[Y_i \mid D_i = 0]$	*What we can estimate*

SUTVA

- Stable unit treatment value assumption (SUTVA):
  - The potential outcome of one individual **does not depend on the treatment status of other individuals**
- Each unit has only 2 potential outcomes: Y_i^0 , Y_i^1
- Assumes no spillover effects
- Assumes no general equilibrium effects
- Often not realistic in economics

Selection bias

$$\underbrace{\mathbb{E}[Y_i|D_i = 1]}_{\text{Difference in average observed outcomes}} - \mathbb{E}[Y_i|D_i = 0] =$$

$$\underbrace{\mathbb{E}[Y_i^1 - Y_i^0|D_i = 1, X_i]}_{ATT} + \underbrace{\mathbb{E}[Y_i^0|D_i = 1, X_i] - \mathbb{E}[Y_i^0|D_i = 0, X_i]}_{\text{Selection Bias}}$$

- Goal: eliminate this selection bias to be able to say something about the quantity of interest (the ATT)
- **Selection bias**: average difference in Y_i^0 between the treated and untreated
- Assumptions regarding the assignment mechanisms can help eliminate it

Assumed assignment mechanisms

- **Random assignment** (*eg* experiments)
  - Treatment independent of potential outcomes $\Rightarrow$ no selection bias in expectation
  - It is the *Independence Assumption (IA)*: $(Y_i^0, Y_i^1) \perp D_i$
- **Selection on observables**
  - Random assignment conditional on some pre-treatment characteristic X
  - It is the *Conditional Independence Assumption (CIA)*: $(Y_0, Y_1) \perp D_i | X_i$
  - Compare outcomes within each stratum of X_i
- **Selection on unobservables**
  - Need other identification strategies to eliminate selection bias
  - Will still assume some other independence assumptions

Identifying assumptions

- Can recover an unbiased estimator of a causal effect iff an identifying/independence assumption holds:
  - IA: $(Y_i^0, Y_i^1) \perp D_i \Rightarrow$ can estimate the ATT
  - No IA but CIA: $(Y_i^0, Y_i^1) \perp D_i | X_i \Rightarrow$ can estimate the ATT in each stratum
  - No CIA but ∃ a relevant instrument Z_i that is an exogenous source of variation in D_i: $(Y_i^0, Y_i^1) \perp Z_i | X_i, \ \ Z_i \perp\!\!\!/ D_i | X_i \Rightarrow$ can estimate *a* LATE
- We always need an identification strategy that convinces us that an IA holds

Summary

- Goal: identifying **causal** effects
- *ie* a difference between two potential outcomes
- But, we cannot observe them
- We only see the differences in observed outcomes
- If (C)IA holds, we can estimate an unbiased ATT
  - Randomized Control Trial (RCT), the gold standard
- But (C)IA rarely holds $\Rightarrow$ need an **identification strategy** to elimate selection bias

Common identification methods

- **Randomized experiments (RCT)**
  - Randomization of treatment D
- **Difference-in-differences (DiD), event studies, synthetic control methods (SCM)**
  - Research designs that assume or construct parallel trends
- **Instrumental variables (IV)** or **regression discontinuity (RD)**
  - An instrument or discontinuity induces exogenous variation in treatment status
- **Matching estimators**:
  - Strategies solely based on matching are much less credible
  - But matching can complement natural or quasi-experimental design

Identification based on repeated observations

Adjusting for non-varying factors

- Repeated observations over some dimension allow **adjusting for all the unobserved characteristics that are constant across that dimension**
- Transform each variable into its **deviation from the group mean**
- Only keep **within variation** (discards the *between*)
- Two approaches to do that:
  - Manual demeaning
  - Including fixed effects
- Basically build a **counterfactual**

Event studies, DiD, and TWFEs

- *Objective*: estimate the impact of some treatment at a certain time
- Leverages repeated observations, typically **panel data**
- Builds a **counterfactual** that can be explicit or more implicit (*eg* TWFE):
  - Unit's outcome had the event not occurred

Event study

- All units are treated
- Assumed counterfactual: group's past value
- *Within* variation only
- + Flexible, allows looking at whether effects are **dynamic**
- − Difficult to rule out **other things changing at the same time**
  - *The rooster concluding the sun rises because of his crowing?*

$$Y_{it} = \sum_{t=-K}^{\tau-2} [\beta_t \mathbb{1}\{t\}] + \beta_\tau \mathbb{1}\{\tau\} + \sum_{t=\tau+1}^{L} [\beta_t \mathbb{1}\{t\}] + e_{it}$$

DiD, DiDiD, TWFE

- Some units never get treated
- Assumed counterfactual: **parallel trends** of treated and untreated are parallel
- *Within* and *between* variation
- + Pre-trends not a problem (unlike event studies) as long as trends of the groups are parallel
- − Issues when go beyond simple binary DiD (*we discuss that later*)

$$Y_{it} = \beta G_i P_t + \lambda_G + \lambda_P + e_{it}$$

Nuts and bolts of fixed effects

Interpreting fixed effects

- **Group FEs**: compare individuals within the group
- **Time FEs**: compare individuals within a time period
- **TWFEs**:
  - Average of TEs identified from variation within group **and** variation within period
  - ≠ variation within "that group that year" (this would be group-year FEs)
- Including **FEs changes the estimand**: we compare observation within a group or within a time period

Illustration of pooled estimate

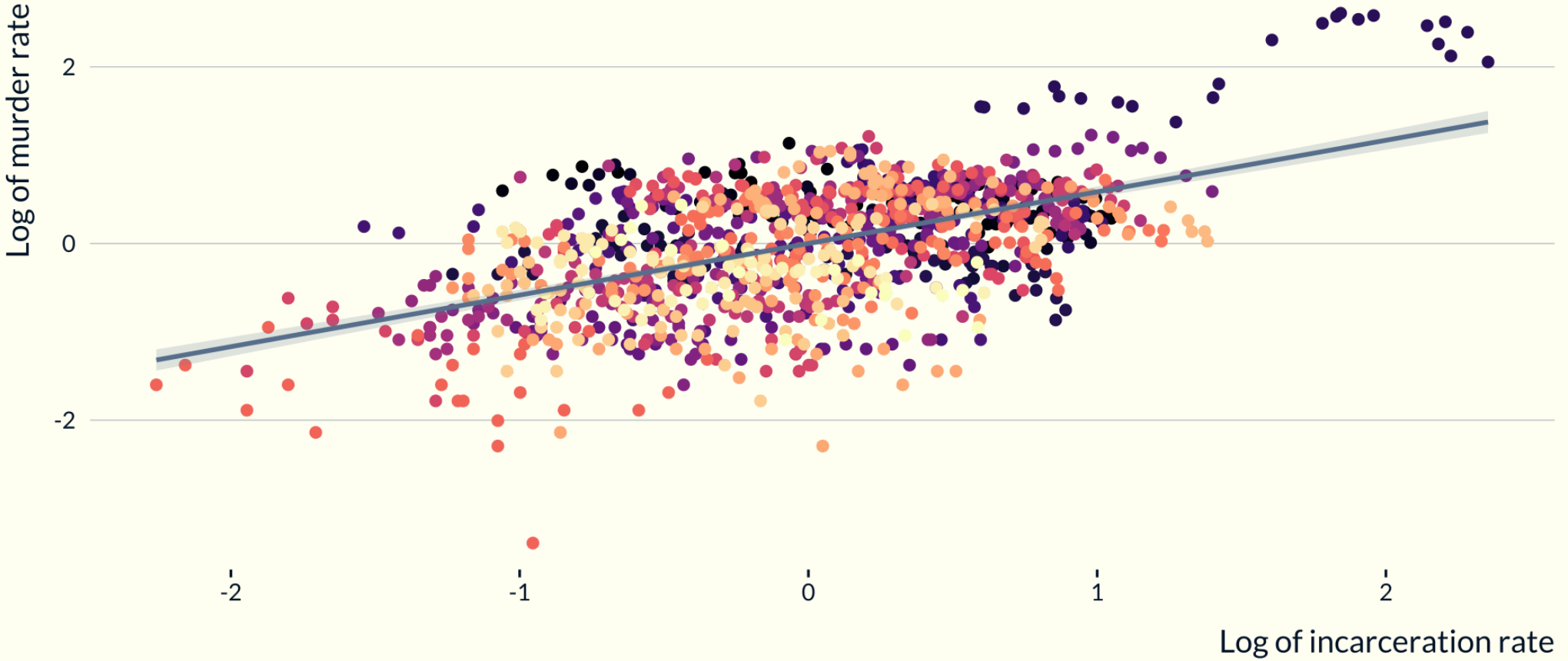
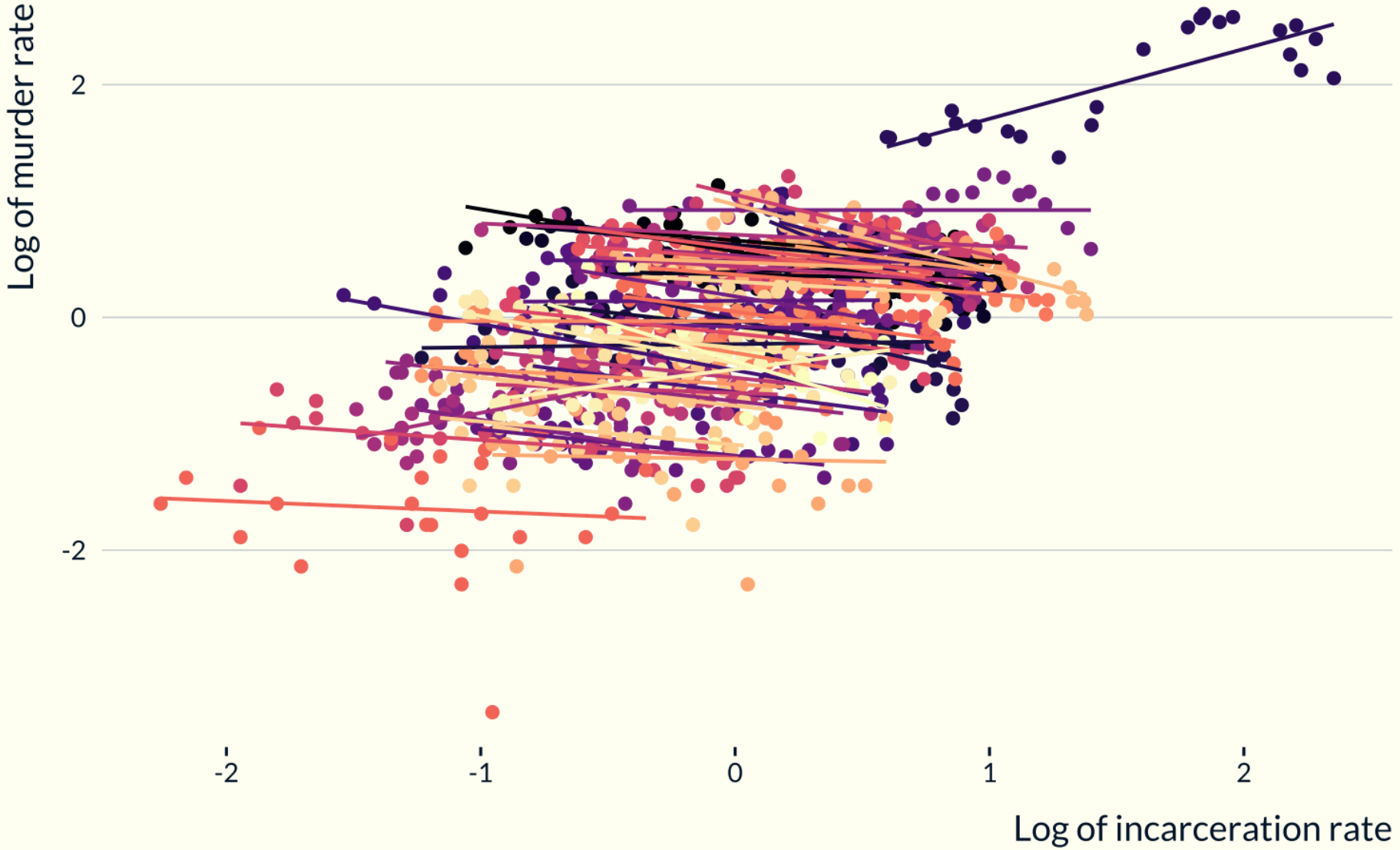
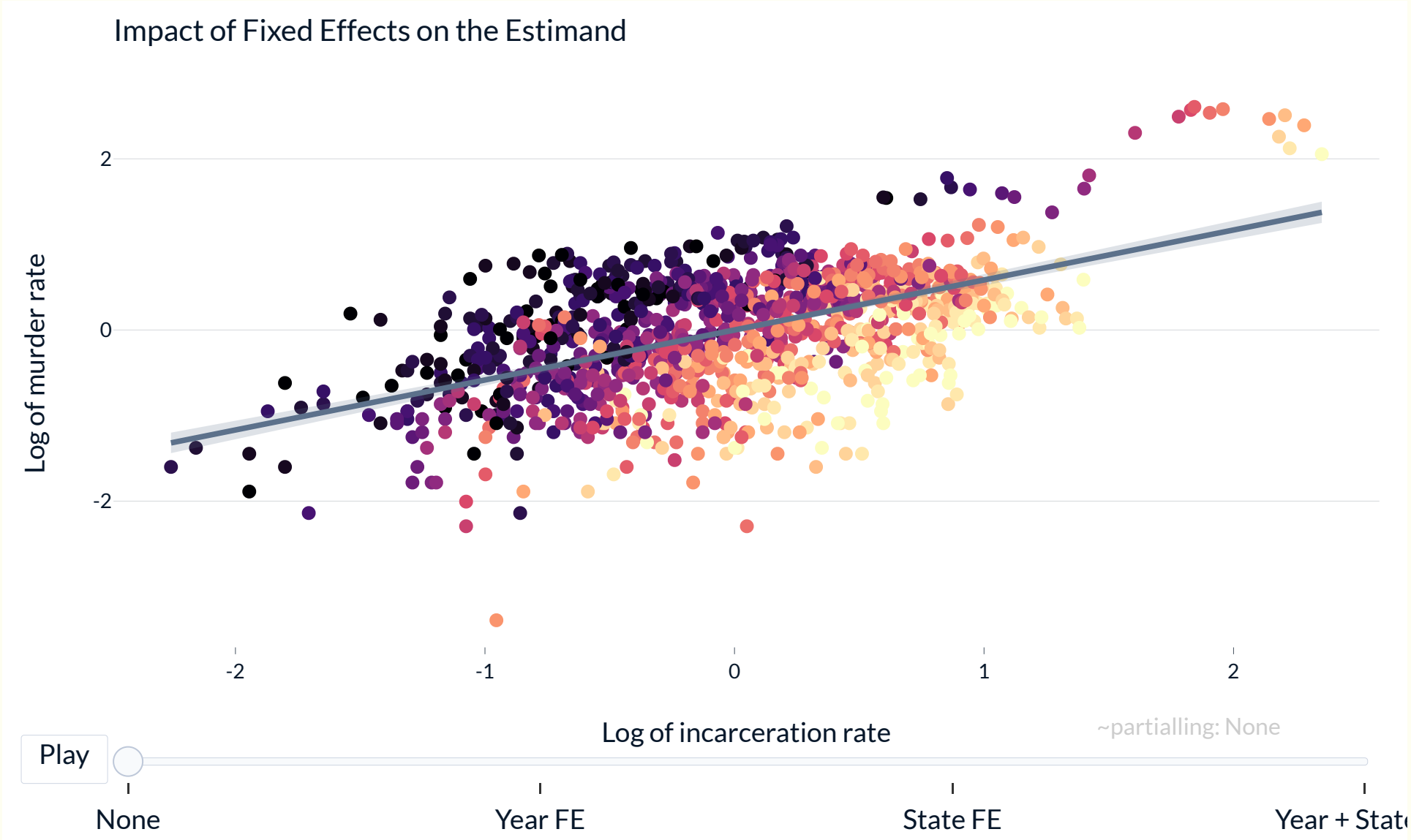


Illustration of within state relationship




Impact of Fixed Effects on the Estimand
Log of murder rate
2
0
-2
-2
-1
0
1
2
Log of incarceration rate
~partialling: None
Play
None
Year FE
State FE
Year + Stat

Regression as a projection

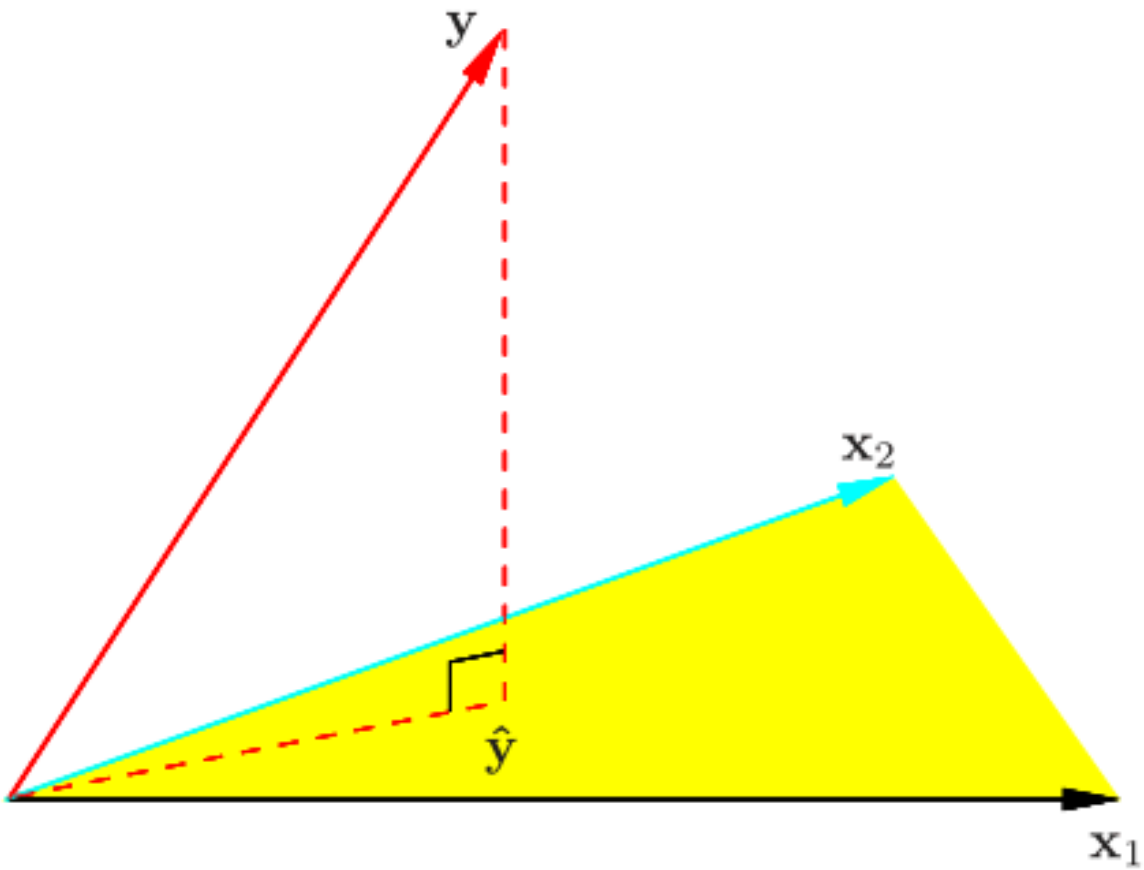


FIGURE 3.2. *The N-dimensional geometry of least squares regression with two predictors. The outcome vector* $\mathbf{y}$ *is orthogonally projected onto the hyperplane spanned by the input vectors* $\mathbf{x}_1$ *and* $\mathbf{x}_2$*. The projection* $\hat{\mathbf{y}}$ *represents the vector of the least squares predictions*

Frisch-Waugh-Lovell (FWL) Theorem

$$Y = X\beta + W\delta + U$$

- The estimate of β is the same as the estimate of $\tilde{\beta}$ in:

$$Y^{\perp W} = X^{\perp W}\tilde{\beta} + U^{\perp W}$$

- where $.^{\perp W}$ denotes each variable where W has been **residualized**
- *ie* its projection onto the **orthogonal space to W**
- Obtained using:
  - The **projection matrix** $P_W = W(W'W)^{-1}W'$
  - The **residual-maker matrix** $M_W = I - P_W$
- *eg* $X^{\perp W} = M_W X$
- Fixed effects regression = regression on variables after partialling out the fixed effects

In practice

- To compute the partialled out version of a regression:
  1. Compute the residualized version of y and x: regress them on controls/FE
  2. Regress the **residuals** on one another
- Exercise. Using the data bellow, run two regressions and compare the estimates obtained:
  1. Regress `l_murder` on `l_pris` with state fixed effects
  2. Regress their residualized versions on one another (partialling out state FEs)

```
library(AER)
data("Guns")

guns <- Guns |>
  as_tibble() |>
  mutate(
    l_pris = log(prisoners),
    l_murder = log(murder)
  )
```

Visualizing the raw data

Code | Graph levels | Graph logs

```
graph_levels <- guns |>
  ggplot(aes(x = prisoners, y = murder)) +
  geom_point() +
  labs(
    title = "Relationship between incarceration and murder rates",
    subtitle = "Variables in level: need to transform it",
    x = "Incarceration rate",
    y = "Murder rate"
  )

graph_log <- guns |>
  ggplot(aes(x = l_pris, y = l_murder)) +
  geom_point() +
  geom_smooth(method = "lm") +
  labs(
    title = "Relationship between incarceration and murder rates",
    subtitle = "Log are better suited",
    x = "Log of incarceration rate",
    y = "Log of murder rate"
  )
```

Equivalence residual vs manual demean

```r
#demeaning and showing that equal to residuals
sample_demean <- guns |>
  mutate(
    l_murder_res = feols(data = guns, fml = l_murder ~ 1 | state) |>
      residuals()
  ) |>
  group_by(state) |>
  mutate(mean_l_murder = mean(l_murder)) |>
  ungroup() |>
  mutate(
    l_murder_demean = l_murder - mean_l_murder
  ) |>
  select(l_murder_res, l_murder_demean) |>
  head(10)
```

l_murder_res	l_murder_demean
0.2824963	0.2824963
0.2170183	0.2170183
0.2094711	0.2094711
0.2094711	0.2094711
0.1057927	0.1057927
-0.0098917	-0.0098917
-0.1515422	-0.1515422
-0.1300360	-0.1300360
-0.0883633	-0.0883633
-0.0582103	-0.0582103

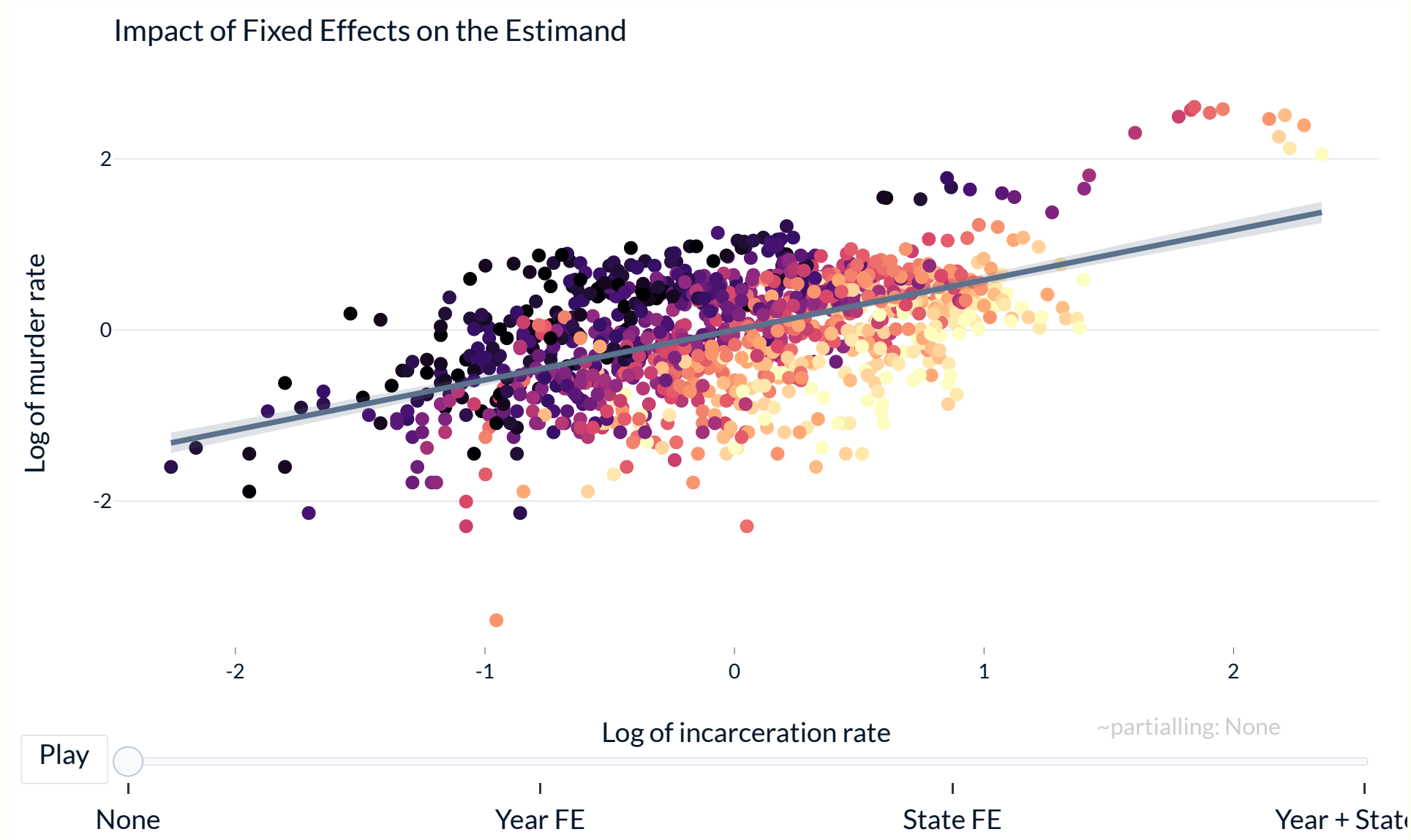
Illustration of the FWL theorem

```r
library(fixest)

#demeaning and showing that equal to residuals
guns_demean <- guns |>
  mutate(
    l_murder_res = feols(data = guns, fml = l_murder ~ 1 | state) |>
      residuals(),
    l_pris_res = feols(data = guns, fml = l_pris ~ 1 | state) |>
      residuals()
  )

reg_fe <- guns |>
  fixest::feols(fml = l_murder ~ l_pris | state)  |>
  broom::tidy() |>
  mutate(reg = "fixed_effects", .before = 1)

reg_res <- guns_demean |>
  feols(fml = l_murder_res ~ l_pris_res - 1, cluster = "state") |>
  broom::tidy() |>
  mutate(reg = "residualized", .before = 1)

rbind(reg_fe, reg_res) |>
  kable()
```

reg	term	estimate	std.error	statistic	p.value
fixed_effects	l_pris	-0.15834	0.0365294	-4.334587	7.05e-05
residualized	l_pris_res	-0.15834	0.0365138	-4.336438	7.01e-05

Identifying variation

- When adding FE (or controlling in general), we partial out or absorb some of the variation
- We **throw out variation**
- Good if throw out variation that:
  - Is endogenous
  - Explains some of the variance of y $\left(\text{since } \mathbb{V}_{\hat{\beta}} = \frac{\sigma_u^2}{n\sigma_x^2} \right)$
- Bad if throw out **identifying variation**, *ie* variation that allows you to identify the effect of interest


Impact of Fixed Effects on the Estimand
Log of murder rate
2
0
-2
-2
-1
0
1
2
Log of incarceration rate
~partialling: None
Play
None
Year FE
State FE
Year + Stat

ATE as a weighted average

- The estimate of the treatment coefficient is in fact **a weighted average of individual treatment effects**
  - See Aronow and Samii (2016) and Angrist and Pischke (2009) section 3.3.1)
- Weight: $w_i = (T_i - \mathbb{E}[T_i|X_i])^2$
- The weight represents:
  - How well the controls explain the treatment status
  - The conditional variance of the treatment, given X_i
- Actually **equivalent to leverage** in the residualized regression

Implications

- Observations whose treatment status is largely explained by covariates therefore **contribute little, if at all, to estimation**
- For FE: if for some groups there is little within variation, these groups do not contribute to identification
- Implications for **external validity** and **representativity**
- Implications for statistical power: the **effective sample** might be much smaller than the nominal sample

Effective sample vs nominal sample

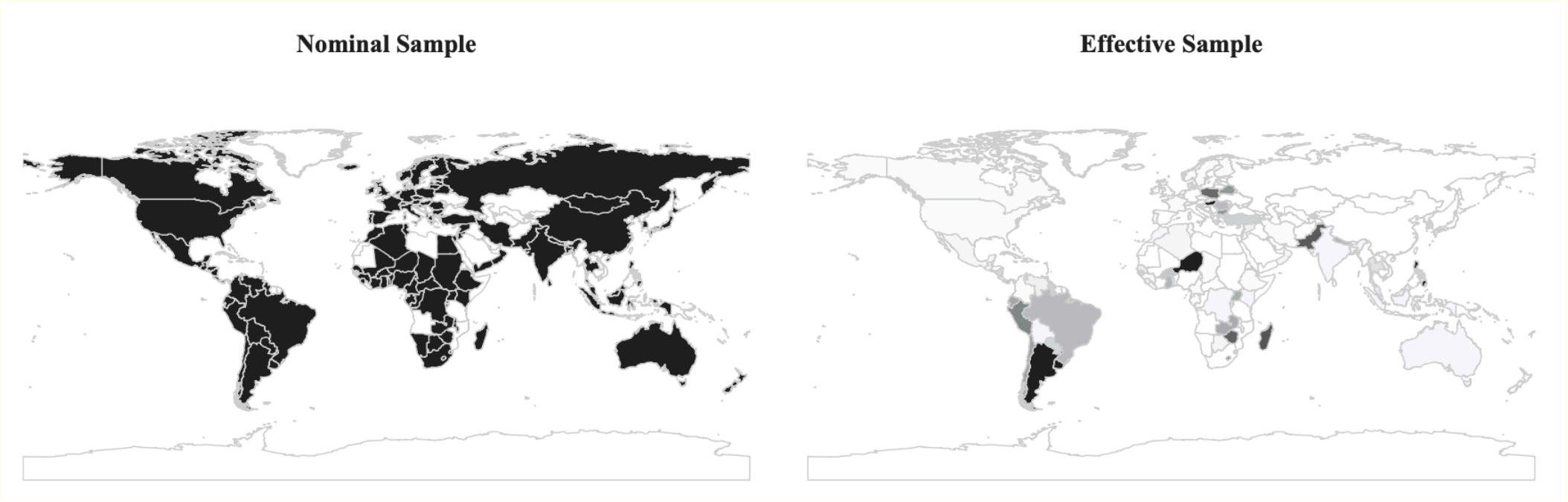


Figure from Aronow and Samii (2016)

Identifying contributing observations

- Let's run some R code together to identify contributing observations in a simple linear regression with fixed effects
- We will use the `gapminder` dataset and regress `lifeExp` on `log(gdpPercap)`
- Let's consider several regressions, with various sets of fixed effects
- I will share with you some code you a

Exercise

Summary

- Today we reviewed:
  - The basis of the **potential outcome framework**
  - Identification strategies based on repeated observations
  - How fixed effects work, under the hood
  - Issues with TWFE
- Hopefully you have a better understanding of:
  - Causal inference, from a bird's view
  - **How fixed effects really work**
  - Many details and **intuitions**

Take away messages

- The choice of FE is crucial and **affects the estimand**
- **FE can remove a lot of variation**:
  - Great if removes endogenous variation
  - Problematic if there is too little variation left