When Should We Use Linear Fixed Effects Regression Models for Causal Inference with Longitudinal Data?

Kosuke Imai
Princeton University

Asian Political Methodology Conference
University of Sydney

Joint work with In Song Kim (MIT)

January 10, 2017

Fixed Effects Regressions in Causal Inference

- Linear fixed effects regression models are the primary workhorse for causal inference with longitudinal/panel data
- Researchers use them to adjust for unobserved time-invariant confounders (omitted variables, endogeneity, selection bias, ...):
  - "Good instruments are hard to find ..., so we'd like to have other tools to deal with unobserved confounders. This chapter considers ... strategies that use data with a time or cohort dimension to control for unobserved but fixed omitted variables" (Angrist & Pischke, *Mostly Harmless Econometrics*)
  - "fixed effects regression can scarcely be faulted for being the bearer of bad tidings" (Green *et al.*, *Dirty Pool*)
- When should we use linear FE regression models for causal inference?

Linear Regression with Unit Fixed Effects

- Balanced panel data with N units and T time periods
- Y_{it}: outcome variable
- X_{it}: binary causal or treatment variable of interest

Assumption 1 (Linearity)

$$Y_{it} = \alpha_i + \beta X_{it} + \epsilon_{it}$$

- $\mathbf{U}_i$: a vector of unobserved time-invariant confounders
- $\alpha_i = h(\mathbf{U}_i)$ for *any* function $h(\cdot)$
- A flexible way to adjust for unobservables
- Average contemporaneous treatment effect:

$$\beta = \mathbb{E}(Y_{it}(1) - Y_{it}(0))$$

Strict Exogeneity and Least Squares Estimator

Assumption 2 (Strict Exogeneity)

$$\epsilon_{it} \perp\!\!\!\perp \{\mathbf{X}_i, \mathbf{U}_i\}$$

- Mean independence is sufficient: $\mathbb{E}(\epsilon_{it} \mid \mathbf{X}_i, \mathbf{U}_i) = \mathbb{E}(\epsilon_{it}) = 0$
- Least squares estimator based on de-meaning:

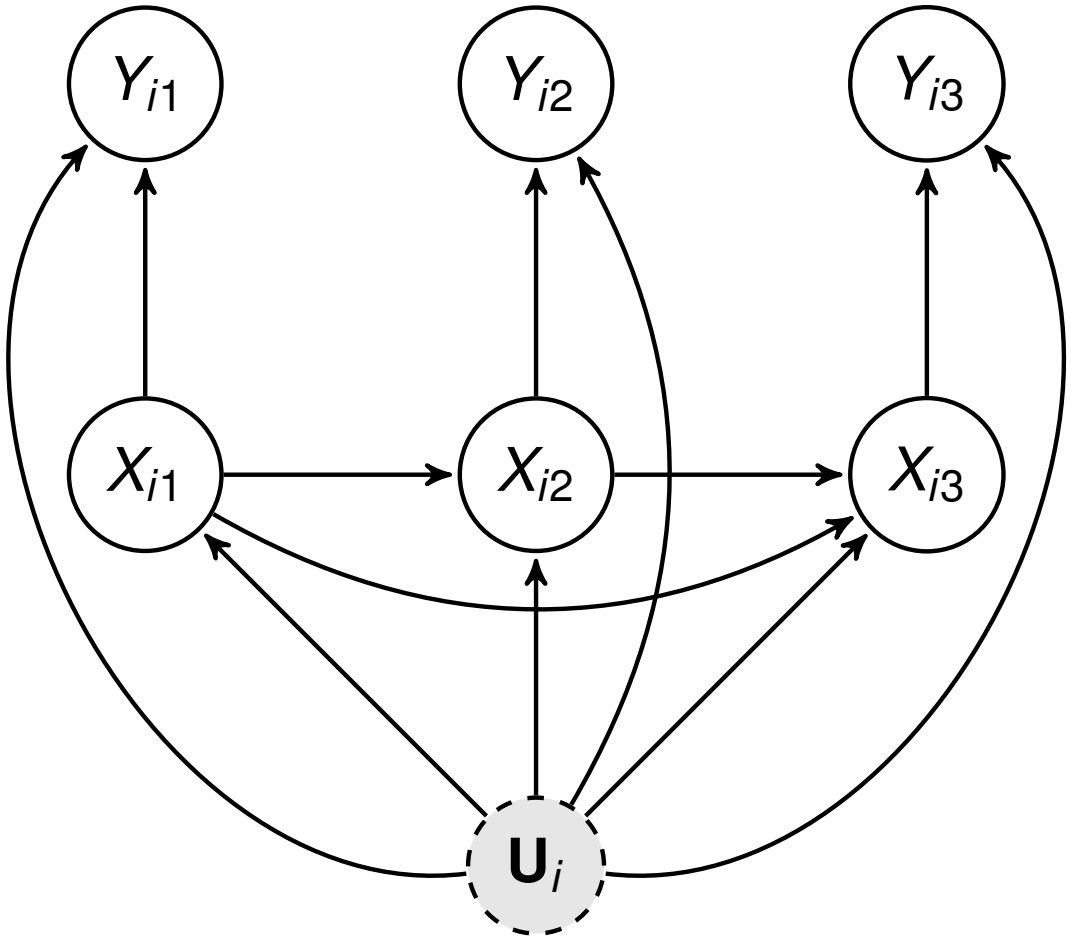
$$\hat{\beta}_{\text{FE}} = \arg\min_{\beta} \sum_{i=1}^{N} \sum_{t=1}^{T} \{(Y_{it} - \overline{Y}_i) - \beta(X_{it} - \overline{X}_i)\}^2$$

  where $\overline{X}_i$ and $\overline{Y}_i$ are unit-specific sample means
- ATE among those units with variation in treatment:

$$\tau = \mathbb{E}(Y_{it}(1) - Y_{it}(0) \mid C_{it} = 1)$$

  where $C_{it} = \mathbf{1}\{0 < \sum_{t=1}^{T} X_{it} < T\}$.

Causal Directed Acyclic Graph (DAG)



- arrow $=$ direct causal effect
- absence of arrows $\rightsquigarrow$ causal assumptions

Nonparametric Structural Equation Model (NPSEM)

- One-to-one correspondence with a DAG:

$$\begin{aligned} Y_{it} &= g_1(X_{it}, \mathbf{U}_i, \epsilon_{it}) \\ X_{it} &= g_2(X_{i1}, \ldots, X_{i,t-1}, \mathbf{U}_i, \eta_{it}) \end{aligned}$$

- Nonparametric generalization of linear unit fixed effects model:
  - Allows for nonlinear relationships, effect heterogeneity
  - Strict exogeneity holds ($\epsilon_{it} \rightarrow Y_{it} \leftarrow \{\mathbf{X}_i, \mathbf{U}_i\}$)
  - No arrows can be added without violating Assumptions 1 and 2

- Causal assumptions:
  1. No unobserved time-varying confounders
  2. Past outcomes do not directly affect current outcome
  3. Past outcomes do not directly affect current treatment
  4. Past treatments do not directly affect current outcome

Potential Outcomes Framework

- DAG $\rightsquigarrow$ causal structure
- Potential outcomes $\rightsquigarrow$ treatment assignment mechanism

Assumption 3 (No carryover effect)

Past treatments do not directly affect current outcome

$$Y_{it}(X_{i1}, X_{i2}, \ldots, X_{i,t-1}, X_{it}) = Y_{it}(X_{it})$$

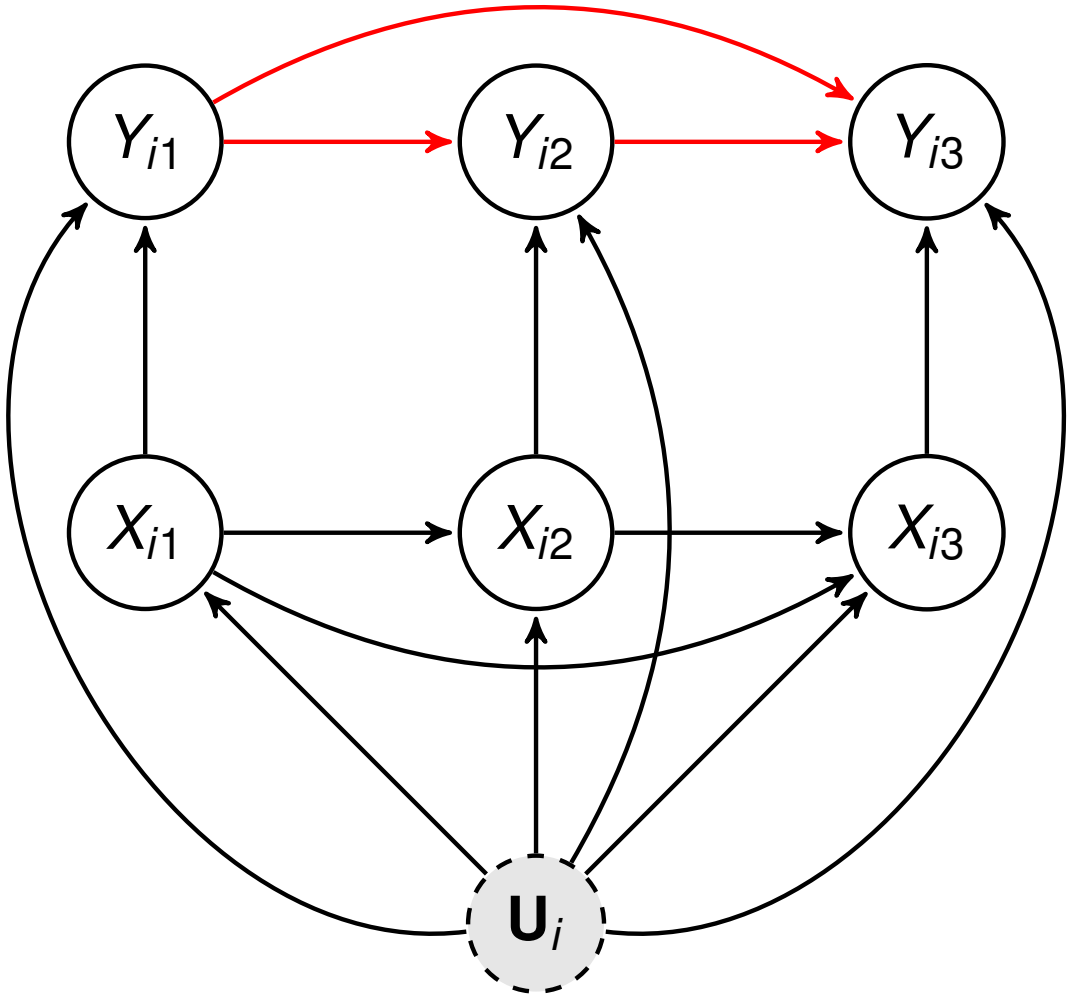
- What randomized experiment satisfies unit fixed effects model?
  1. randomize X_{i1} given $\mathbf{U}_i$
  2. randomize X_{i2} given X_{i1} and $\mathbf{U}_i$
  3. randomize X_{i3} given X_{i2}, X_{i1}, and $\mathbf{U}_i$
  4. and so on

Assumption 4 (Sequential Ignorability with Unobservables)

$$
\begin{aligned}
\{Y_{it}(1), Y_{it}(0)\}_{t=1}^{T} &\perp\!\!\!\perp X_{i1} \mid \mathbf{U}_i \\
&\vdots \\
\{Y_{it}(1), Y_{it}(0)\}_{t=1}^{T} &\perp\!\!\!\perp X_{it'} \mid X_{i1}, \ldots, X_{i,t'-1}, \mathbf{U}_i \\
&\vdots \\
\{Y_{it}(1), Y_{it}(0)\}_{t=1}^{T} &\perp\!\!\!\perp X_{iT} \mid X_{i1}, \ldots, X_{i,T-1}, \mathbf{U}_i
\end{aligned}
$$

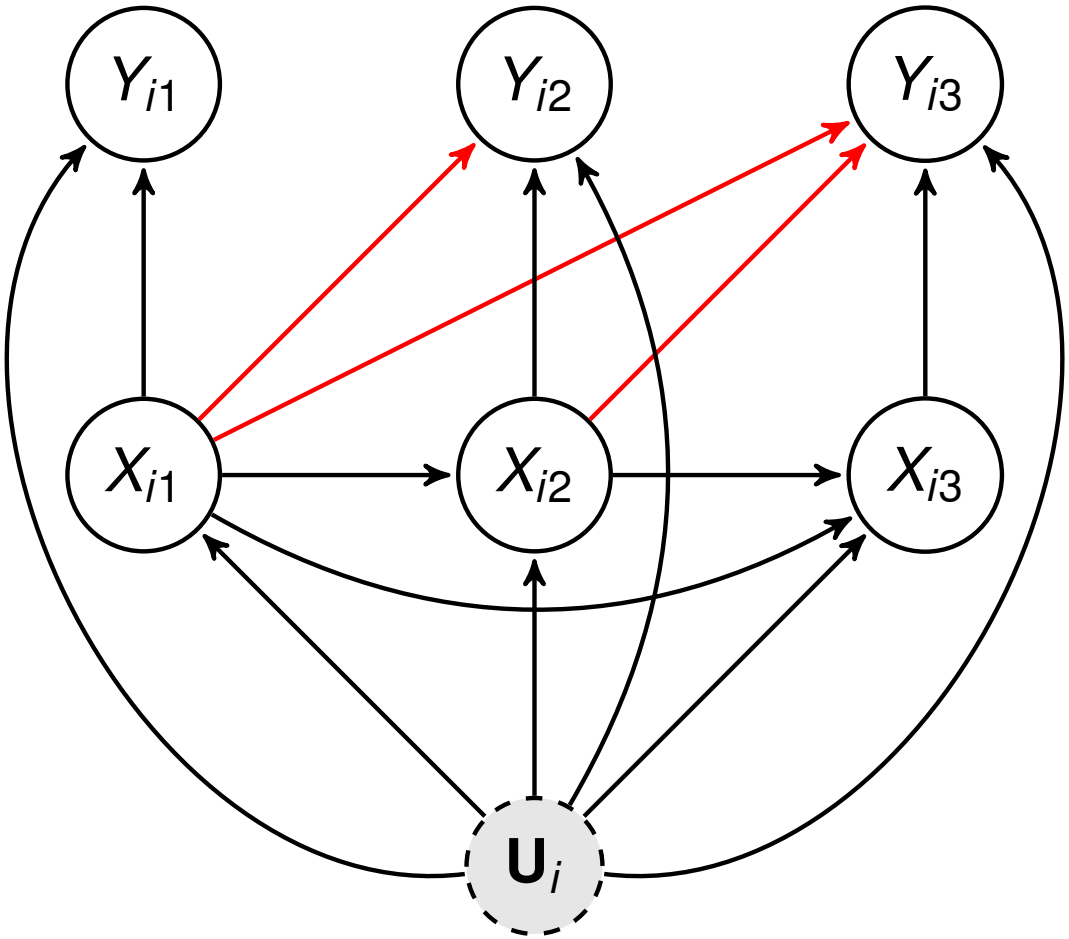
- "as-if random" assumption without conditioning on past outcomes
- Past outcomes cannot directly affect current treatment
- Says nothing about whether past outcomes can directly affect current outcome

Past Outcomes Directly Affect Current Outcome



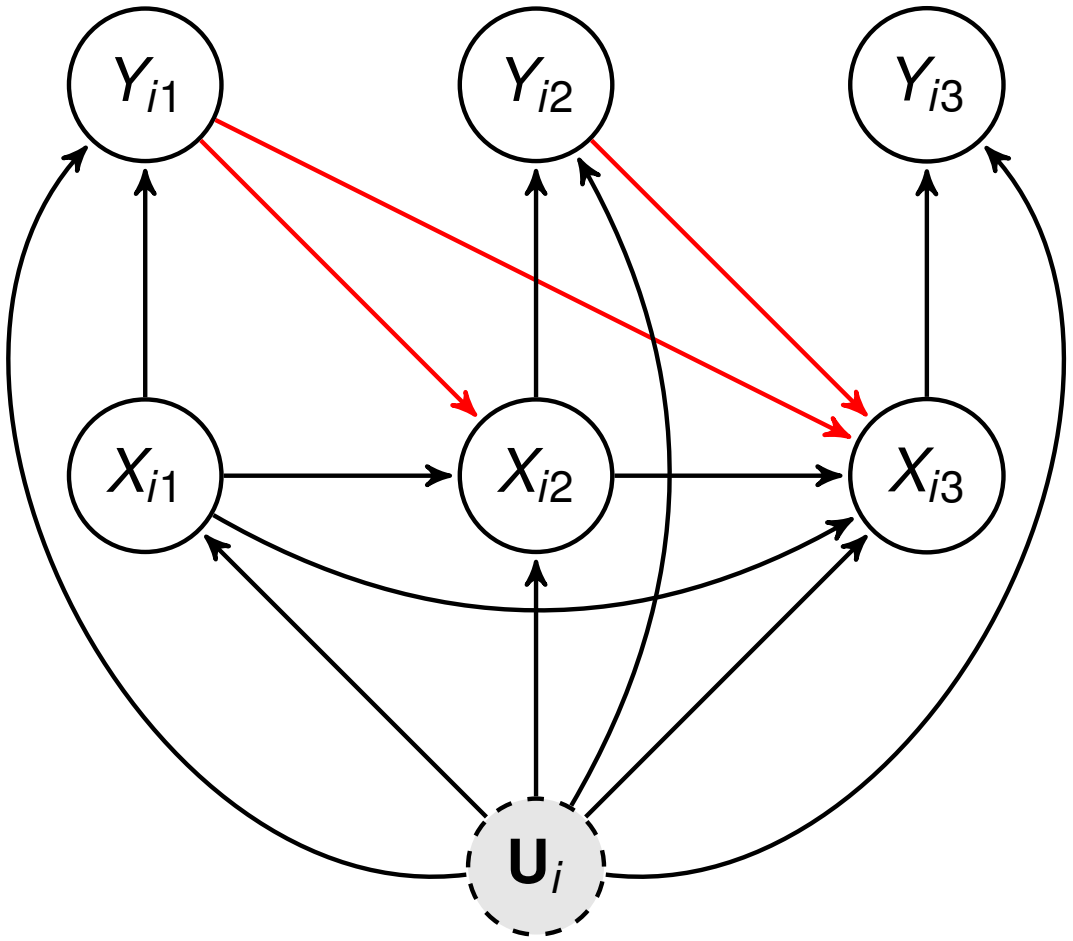
- Strict exogeneity still holds
- Past outcomes do not confound $X_{it} \longrightarrow Y_{it}$ given $\mathbf{U}_i$
- No need to adjust for past outcomes

Past Treatments Directly Affect Current Outcome



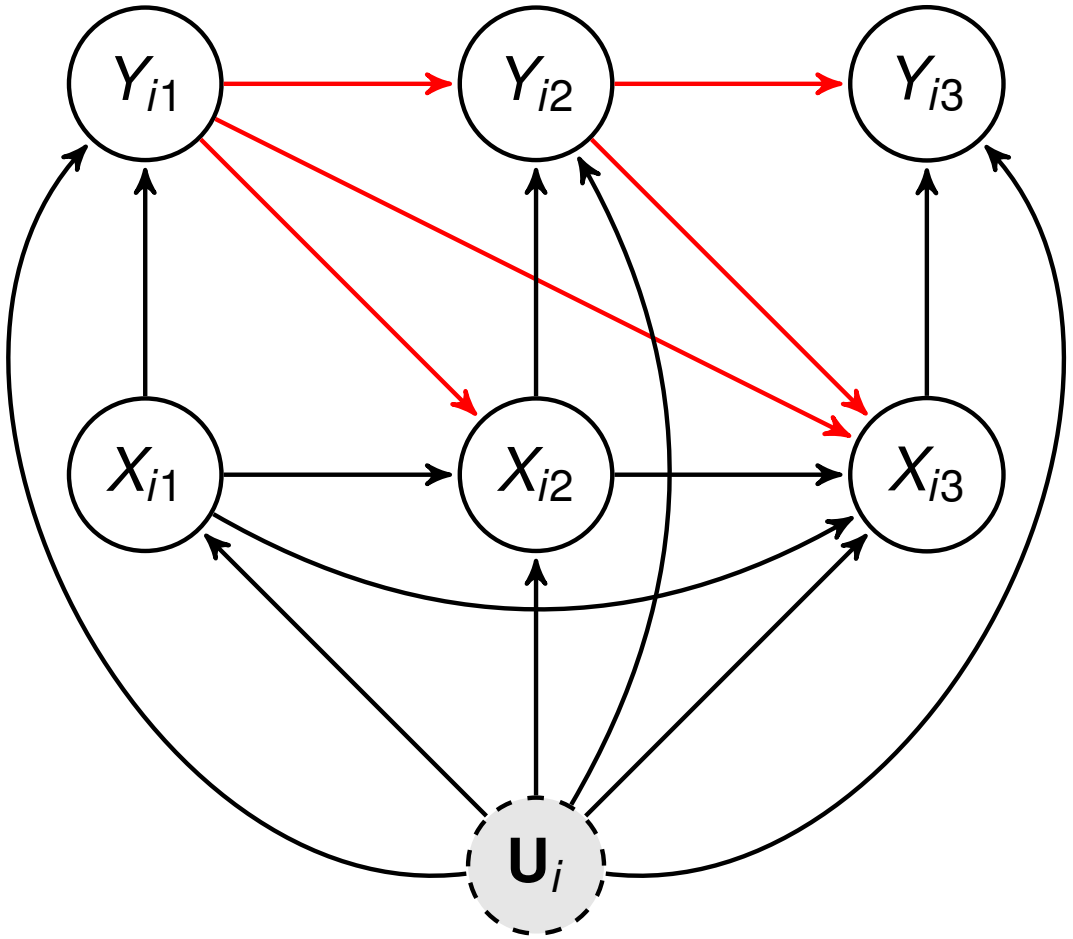
- Past treatments as confounders
- Need to adjust for past treatments
- Strict exogeneity holds given past treatments and $\mathbf{U}_i$
- Impossible to adjust for an entire treatment history and $\mathbf{U}_i$ at the same time
- Adjust for a small number of past treatments $\rightsquigarrow$ often arbitrary

Past Outcomes Directly Affect Current Treatment



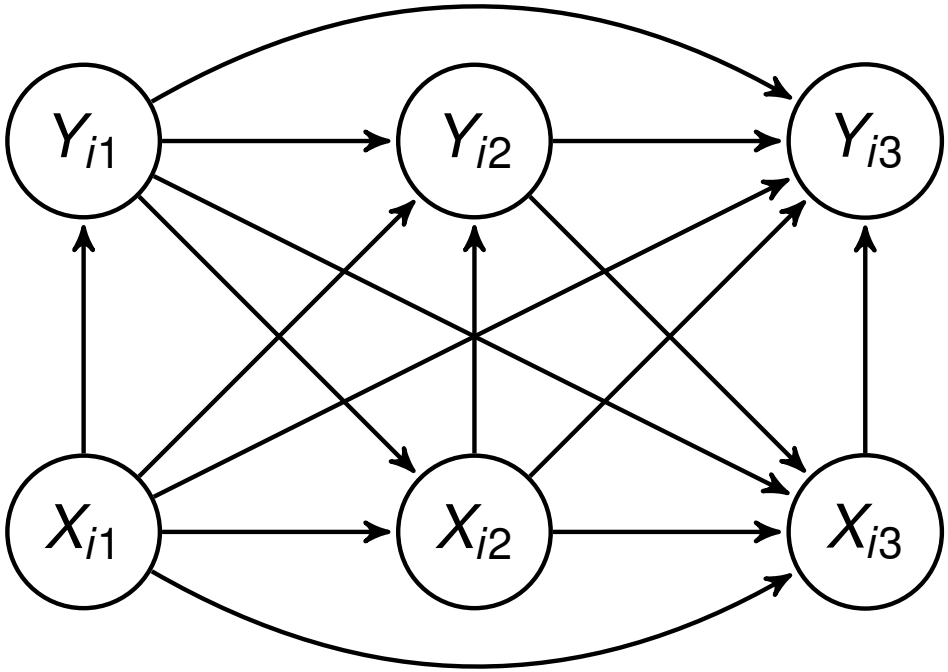
- Correlation between error term and future treatments
- Violation of strict exogeneity
- No adjustment is sufficient
- Together with the previous assumption $\rightsquigarrow$ no feedback effect over time

Instrumental Variables Approach



- Instruments: X_{i1}, X_{i2}, and Y_{i1}
- GMM: Arellano and Bond (1991)
- Exclusion restrictions
- Arbitrary choice of instruments
- Substantive justification rarely given

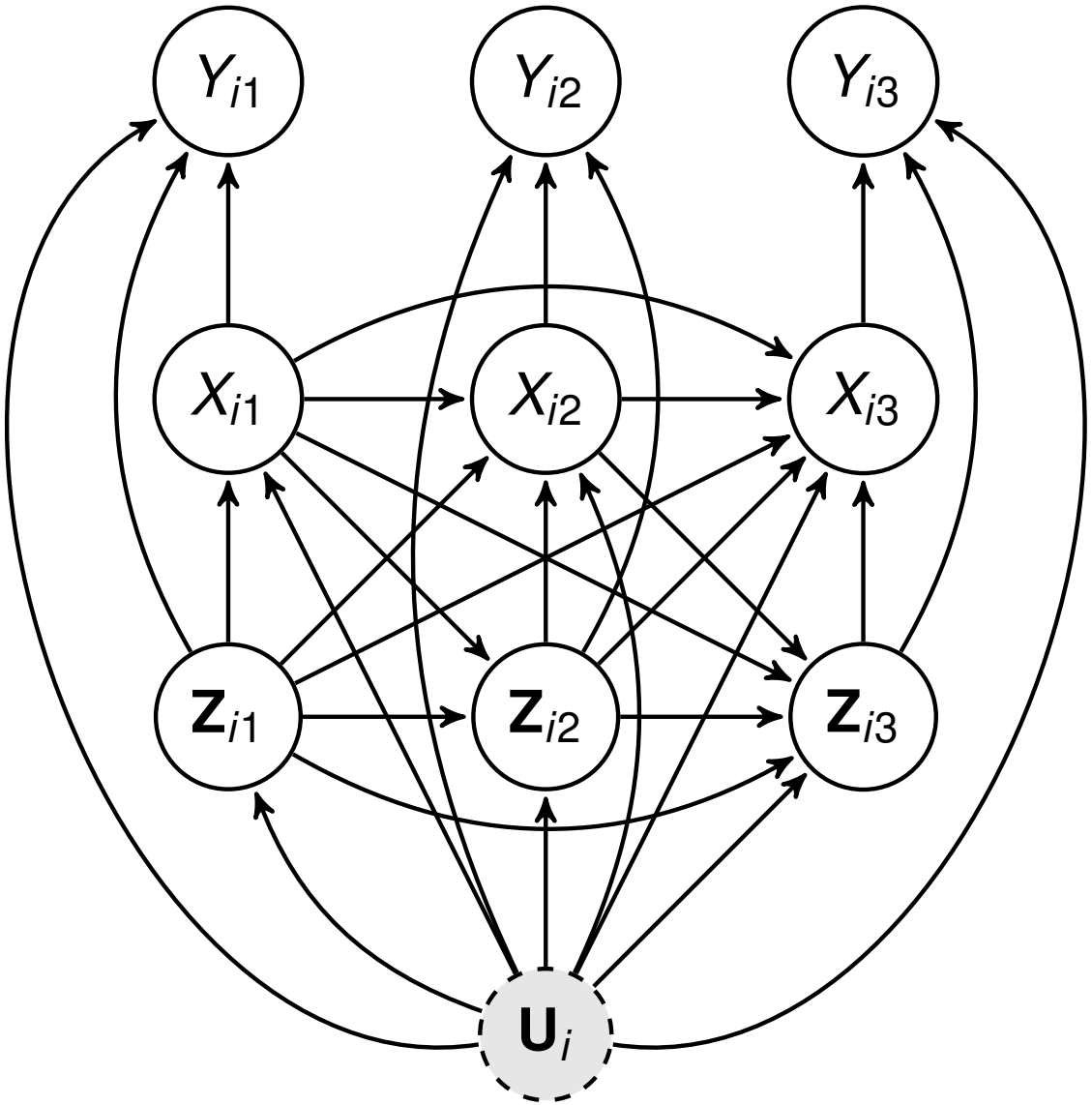
An Alternative Selection-on-Observables Approach



- Absence of unobserved time-invariant confounders $\mathbf{U}_i$
- past treatments can directly affect current outcome
- past outcomes can directly affect current treatment

- Comparison across units within the same time rather than across different time periods within the same unit
- Marginal structural models $\rightsquigarrow$ can identify the average effect of an entire treatment sequence
- Trade-off $\rightsquigarrow$ no free lunch

Adjusting for Observed Time-varying Confounders



- past treatments cannot directly affect current outcome
- past outcomes cannot directly affect current treatment
- adjusting for $\mathbf{Z}_{it}$ does not relax these assumptions
- past outcomes cannot *indirectly* affect current treatment through $\mathbf{Z}_{it}$

Linear Fixed Effects Estimator

- Even if these assumptions are satisfied, the the unit fixed effects estimator is inconsistent for the ATE:

$$\hat{\beta}_{\text{FE}} \xrightarrow{p} \frac{\mathbb{E}\left\{ C_i \left(\frac{\sum_{t=1}^{T} X_{it} Y_{it}}{\sum_{t=1}^{T} X_{it}} - \frac{\sum_{t=1}^{T} (1 - X_{it}) Y_{it}}{\sum_{t=1}^{T} 1 - X_{it}} \right) S_i^2 \right\}}{\mathbb{E}(C_i S_i^2)} \neq \tau$$

  where $S_i^2 = \sum_{t=1}^{T} (X_{it} - \overline{X}_i)^2 / (T - 1)$ is the unit-specific variance

- We show how to eliminate this bias using a general matching framework
- Equivalence between matching and weighted fixed effects estimators

Linear Regression with Unit and Time Fixed Effects

- Model:

$$Y_{it} = \alpha_i + \gamma_t + \beta X_{it} + \epsilon_{it}$$

where γ_t flexibly adjusts for a vector of unobserved unit-invariant time effects $\mathbf{V}_t$, i.e., $\gamma_t = f(\mathbf{V}_t)$

- Estimator:

$$\hat{\beta}_{\text{FE2}} = \arg\min_{\beta} \sum_{i=1}^{N} \sum_{t=1}^{T} \{(Y_{it} - \overline{Y}_i - \overline{Y}_t + \overline{Y}) - \beta(X_{it} - \overline{X}_i - \overline{X}_t + \overline{X})\}^2$$

where $\overline{Y}_t$ and $\overline{X}_t$ are time-specific means, and $\overline{Y}$ and $\overline{X}$ are overall means

Understanding the Two-way Fixed Effects Estimator

The two-way FE estimator combines three biased estimators!

1. β_{FE}: bias due to time effects
2. β_{FEtime}: bias due to unit effects
3. β_{pool}: bias due to both time and unit effects

$$\hat{\beta}_{\text{FE2}} = \frac{\omega_{\text{FE}} \times \hat{\beta}_{\text{FE}} + \omega_{\text{FEtime}} \times \hat{\beta}_{\text{FEtime}} - \omega_{\text{pool}} \times \hat{\beta}_{\text{pool}}}{w_{\text{FE}} + w_{\text{FEtime}} - w_{\text{pool}}}$$

with sufficiently large N and T, the weights are given by,

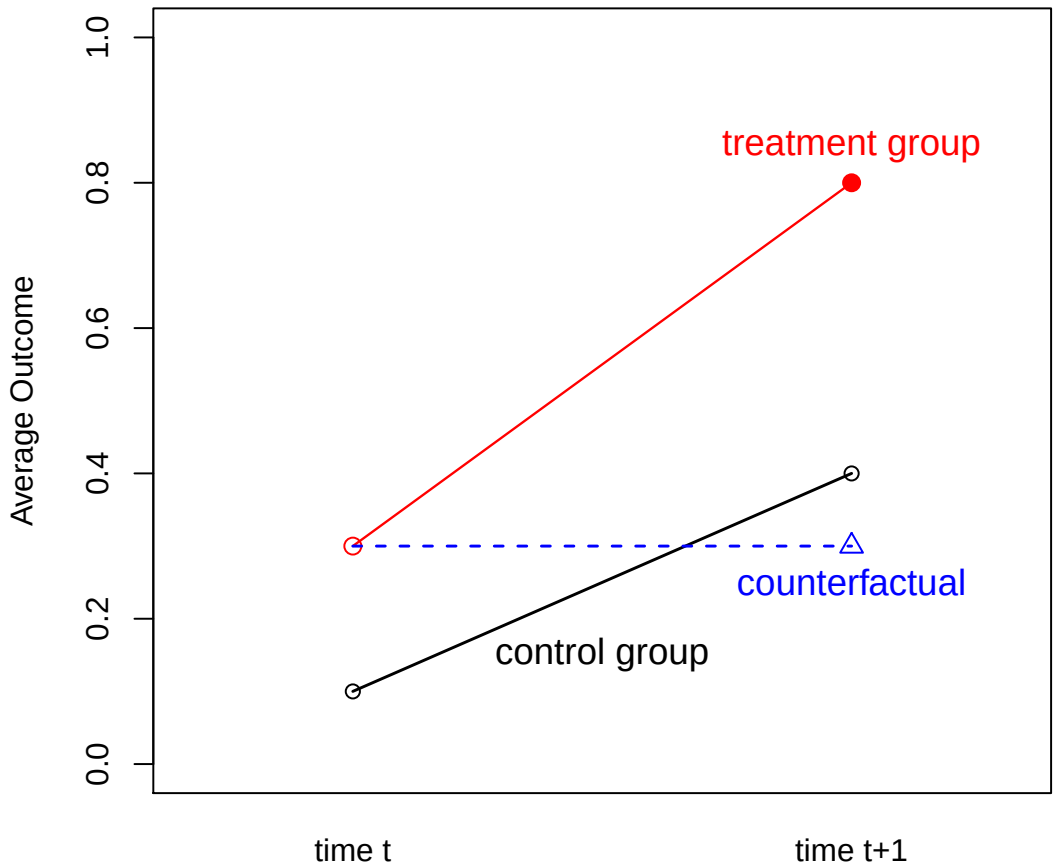
$$\begin{aligned} \omega_{\text{FE}} &\approx \mathbb{E}(S_i^2) = \text{average unit-specific variance} \\ \omega_{\text{FEtime}} &\approx \mathbb{E}(S_t^2) = \text{average time-specific variance} \\ \omega_{\text{pool}} &\approx S^2 = \text{overall variance} \end{aligned}$$

We consider various matching estimators including difference-in-differences and synthetic control

Before-and-After Design

- Accommodating various causal quantity of interest
- No time trend for the average potential outcomes:

$$\mathbb{E}(Y_{it}(x) - Y_{i,t-1}(x) \mid X_{it} \neq X_{i,t-1}) = 0 \quad \text{for} \quad x = 0, 1$$



- This is a matching estimator with the following matched set:

$$\mathcal{M}(i,t) \quad = \quad \{(i',t') : i' = i, t' \in \{t-1, t+1\}, X_{i't'} = 1 - X_{it}\}$$

- It is also the first differencing estimator:

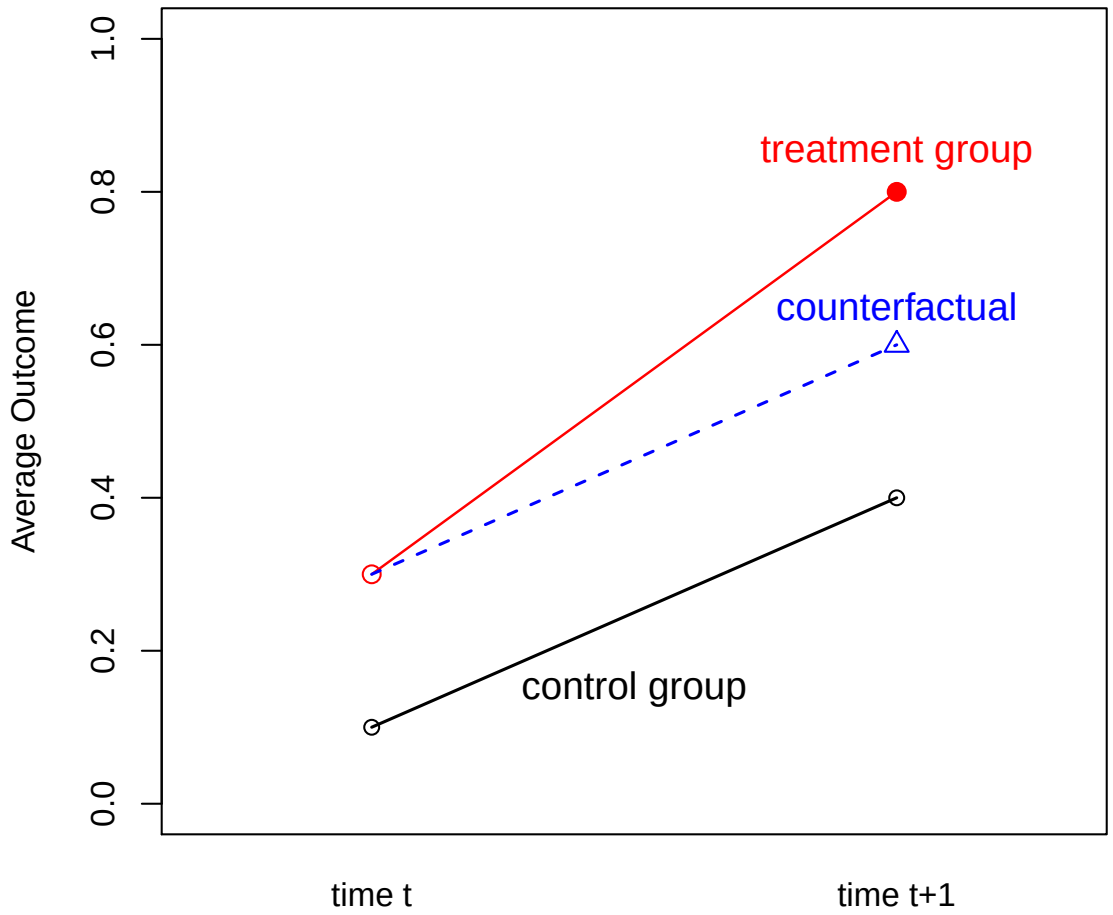
$$\hat{\beta}_{\text{FD}} \quad = \quad \arg\min_{\beta} \sum_{i=1}^{N} \sum_{t=2}^{T} \{(Y_{it} - Y_{i,t-1}) - \beta(X_{it} - X_{i,t-1})\}^2$$

- "We emphasize that the model and the interpretation of β are *exactly* as in [the linear fixed effects model]. What differs is our method for estimating β" (Wooldridge; italics original).
- The identification assumptions is very different
- But, still requires the assumption that past outcomes do not affect current treatment (Regression towards the mean)

Difference-in-Differences Design

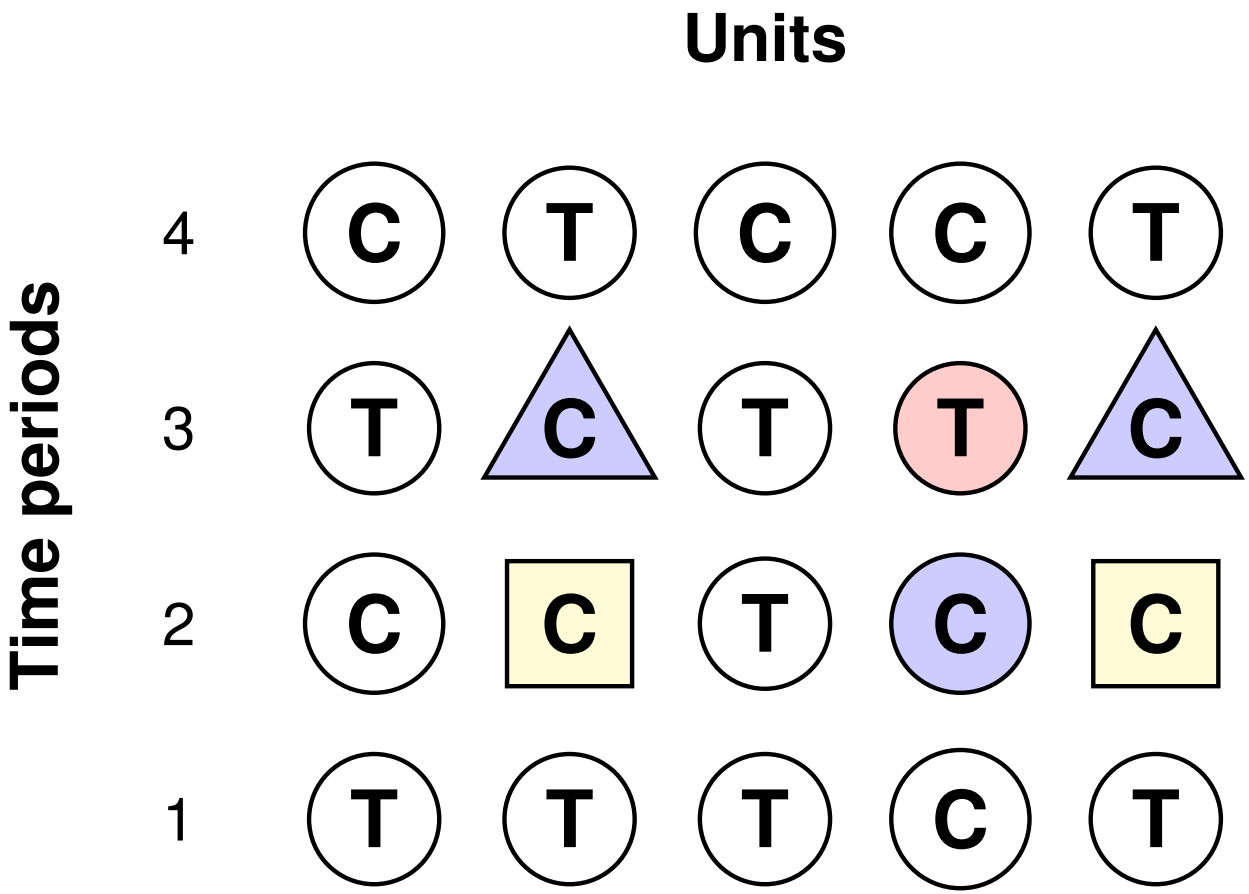
- Parallel trend assumption:

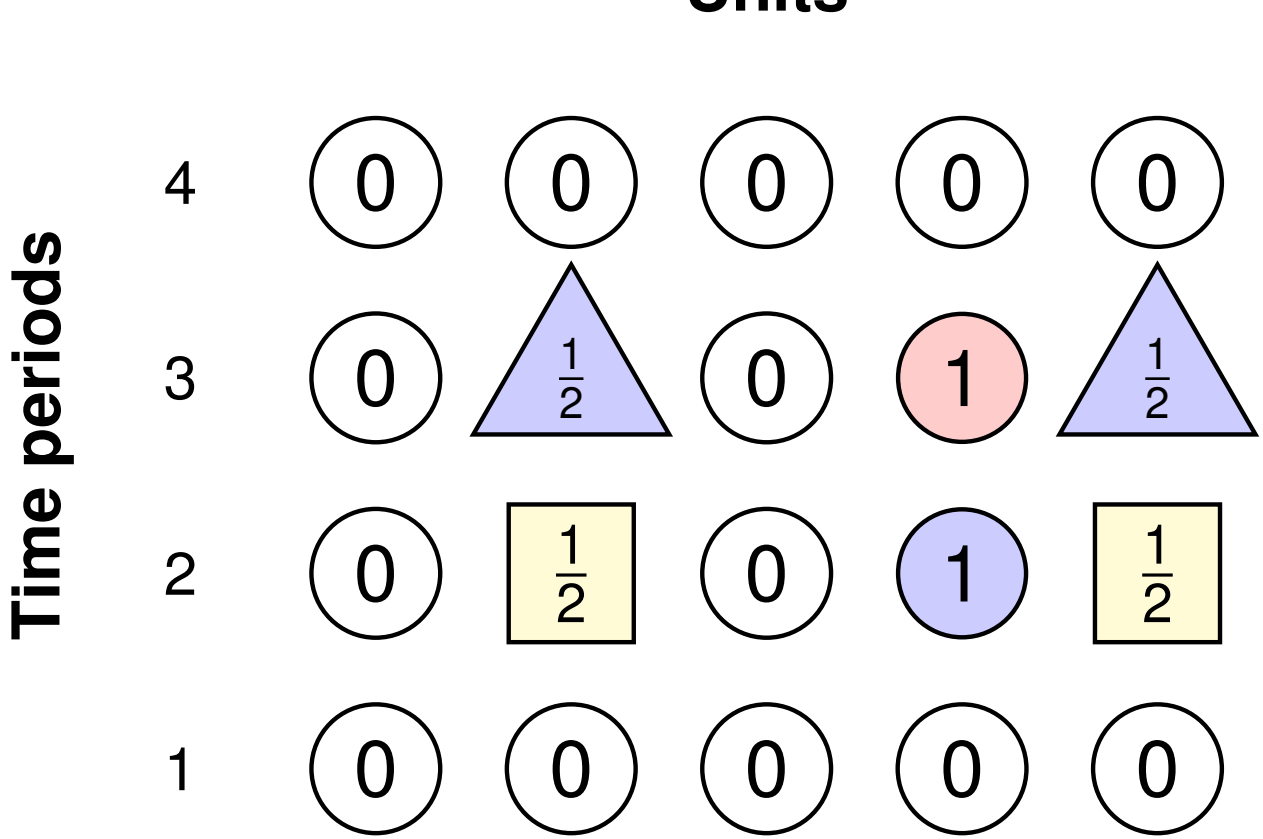
$$\begin{aligned} & \mathbb{E}(Y_{it}(0) - Y_{i,t-1}(0) \mid X_{it} = 1, X_{i,t-1} = 0) \\ = \; & \mathbb{E}(Y_{it}(0) - Y_{i,t-1}(0) \mid X_{it} = X_{i,t-1} = 0) \end{aligned}$$



General DiD = Weighted Two-Way FE Effects

- 2×2: equivalent to linear two-way fixed effects regression
- General setting: Multiple time periods, repeated treatments





- Fast computation, standard error, specification test
- Still assumes that past outcomes don't affect current treatment
- Baseline outcome difference $\rightsquigarrow$ caused by unobserved time-invariant confounders
- It should not reflect causal effect of baseline outcome on treatment assignment

Synthetic Control Method (Abadie et al. 2010)

- One treated unit i^* receiving the treatment at time T
- Quantity of interest: $Y_{i^*T} - Y_{i^*T}(0)$
- Create a synthetic control using past outcomes
- Weighted average: $\widehat{Y_{i^*T}(0)} = \sum_{i \neq i^*} \hat{w}_i Y_{iT}$
- Estimate weights to balance past outcomes and past time-varying covariates

- A motivating autoregressive model:

$$\begin{aligned} Y_{iT}(0) &= \rho_T Y_{i,T-1}(0) + \delta_T^\top \mathbf{Z}_{iT} + \epsilon_{iT} \\ \mathbf{Z}_{iT} &= \lambda_{T-1} Y_{i,T-1}(0) + \Delta_T \mathbf{Z}_{i,T-1} + \nu_{iT} \end{aligned}$$

- Past outcomes can affect current treatment
- No unobserved time-invariant confounders

Causal Effect of ETA's Terrorism

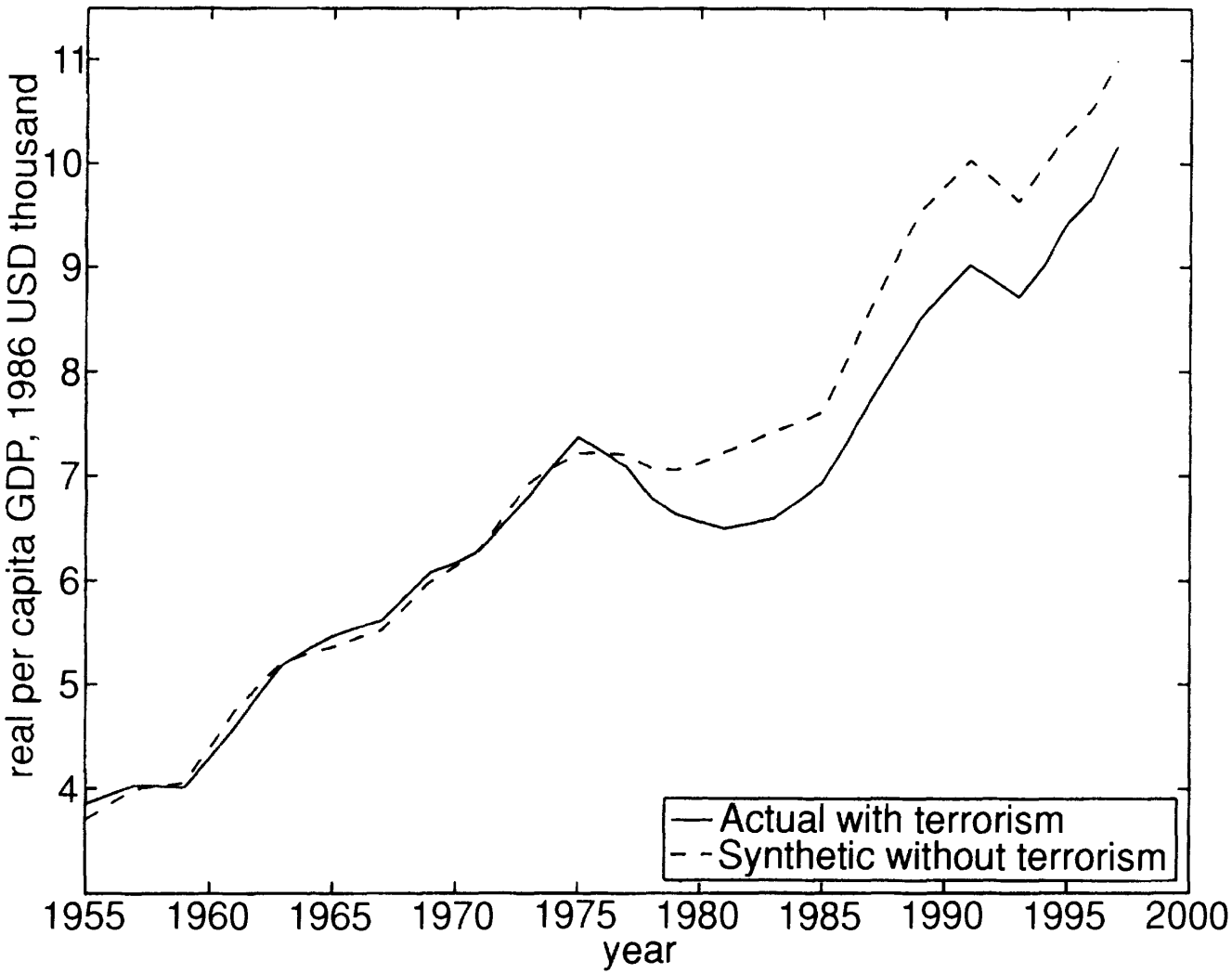


FIGURE 1. PER CAPITA GDP FOR THE BASQUE COUNTRY

Abadie and Gardeazabal (2003, AER)

- The main motivating model:

$$Y_{it}(0) \quad = \quad \gamma_t + \delta_t^\top \mathbf{Z}_{it} + \xi^\top \mathbf{U}_i + \epsilon_{it}$$

- A generalization of the linear two-way fixed effects model
- How is it possible to adjust for unobserved time-invariant confounders by adjusting for past outcomes?

- The key assumption: there exist weights such that

$$\sum_{i \neq i^*} w_i \mathbf{Z}_{it} = \mathbf{Z}_{i^* t} \text{ for all } t \leq T-1 \quad \text{and} \quad \sum_{i \neq i^*} w_i \mathbf{U}_i = \mathbf{U}_{i^*}$$

- In general, adjusting for observed confounders does not adjust for unobserved confounders

- The same tradeoff as before

Concluding Remarks

- When should we use linear fixed effects models?
- Key tradeoff:
  1. unobserved time-invariant confounders $\rightsquigarrow$ fixed effects
  2. causal dynamics between treatment and outcome $\rightsquigarrow$ selection-on-observables
- Two key (under-appreciated) causal assumptions of fixed effects:
  1. past treatments do not directly affect current outcome
  2. past outcomes do not directly affect current treatment
- A new matching estimator:
  1. Within-unit matching estimator $\rightsquigarrow$ no linearity assumption
  2. Various causal identification strategies can be incorporated including the before-and-after and difference-in-differences designs
  3. Equivalent representation as a weighted linear fixed effects regression estimator
- R package **wfe** is available at CRAN

Send comments and suggestions to:

kimai@Princeton.Edu
insong@mit.edu

More information about this and other research:

http://imai.princeton.edu
http://web.mit.edu/insong/www