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

Kosuke Imai
Princeton University

2017 Joint Statistical Meetings

Joint work with In Song Kim (MIT)

August 1, 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

- Y_{it}: outcome variable
- X_{it}: binary treatment variable
- $\mathbf{U}_i$: unobserved time-invariant confounders

Assumption 1 (Linearity)

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

where $\alpha_i = h(\mathbf{U}_i)$ and $h(\cdot)$ is any function

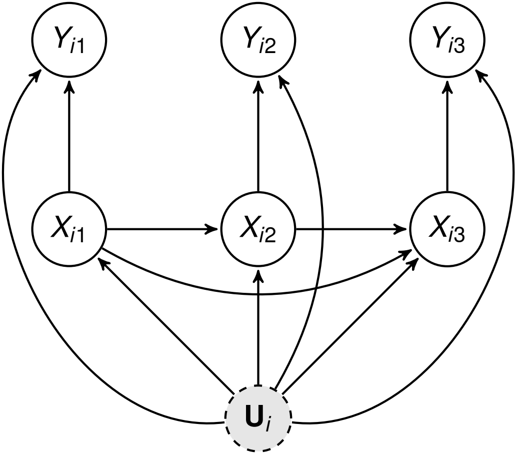
Assumption 2 (Strict Exogeneity)

$$\mathbb{E}(\epsilon_{it} \mid \mathbf{X}_i, \alpha_i) = 0$$

- What is the (nonparametric) identification assumption?

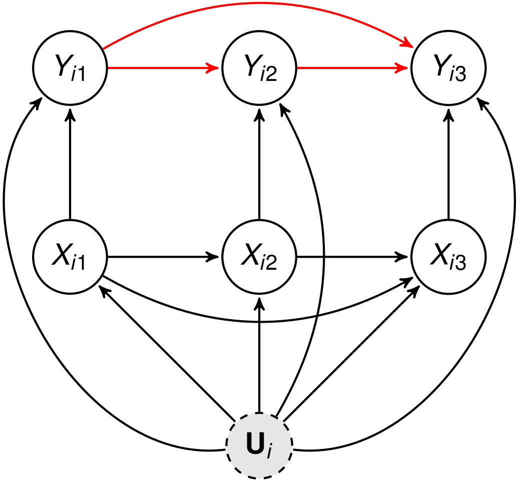
$$Y_{it} = g(X_{it}, \mathbf{U}_i, \epsilon_{it}) \quad \text{and} \quad \epsilon_{it} \perp\!\!\!\perp \{\mathbf{X}_i, \mathbf{U}_i\}$$

Directed Acyclic Graph (DAG)



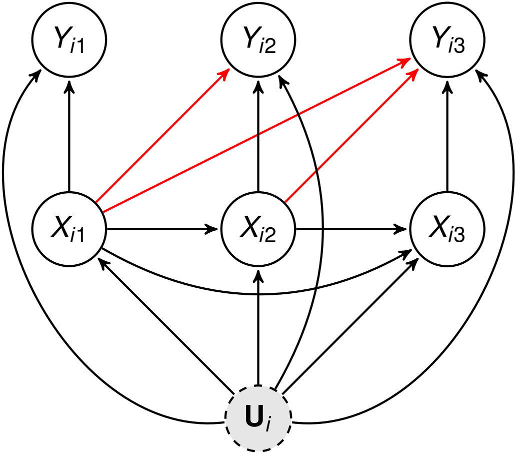
1. No unobserved time-varying confounders
2. Past outcomes do not directly affect current outcome
3. Past outcomes do not directly affect current treatment
4. Past treatments do not directly affect current outcome

Past Outcomes Don't Directly Affect Current Outcome



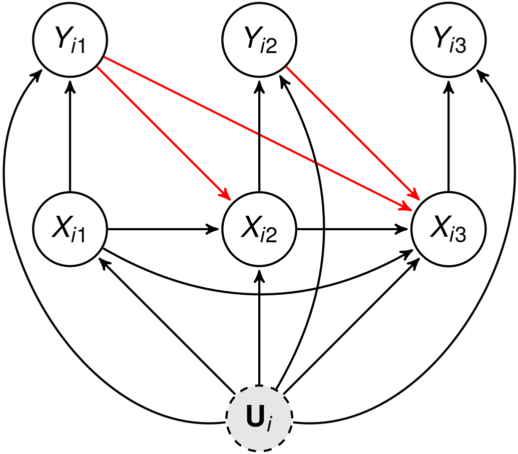
- Strict exogeneity still holds
- Past outcomes do not confound $X_{it} \longrightarrow Y_{it}$ given $\mathbf{U}_i$
- No need to adjust for past outcomes
- Cluster robust standard error
- The assumption can be relaxed

Past Treatments Don't Directly Affect Current Outcome



- 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
- The assumption can be partially relaxed

Past Outcomes Don't Directly Affect Current Treatment



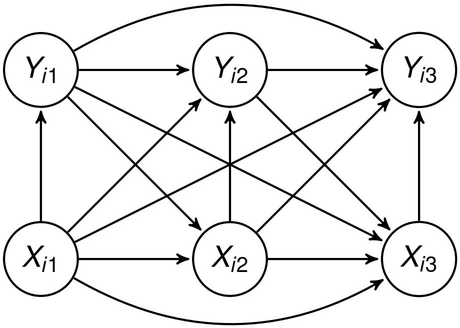
- Correlation between error term and future treatments
- Violation of strict exogeneity
- No adjustment is sufficient
- No dynamic causal relationships between treatment and outcome variables
- The assumption cannot be relaxed

What Randomized Experiment Satisfies Unit Fixed Effects Model?

- Experiment that satisfies the model assumptions:
  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

- Experiment that does not satisfy the model assumptions:
  1. randomize X_{i1}
  2. randomize X_{i2} given X_{i1} and Y_{i1}
  3. randomize X_{i3} given X_{i2}, X_{i1}, Y_{i1}, and Y_{i2}
  4. and so on

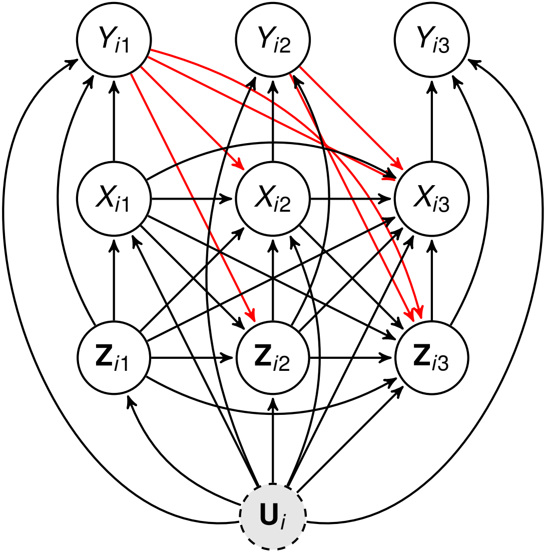
An Alternative Selection-on-Observables Approach



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

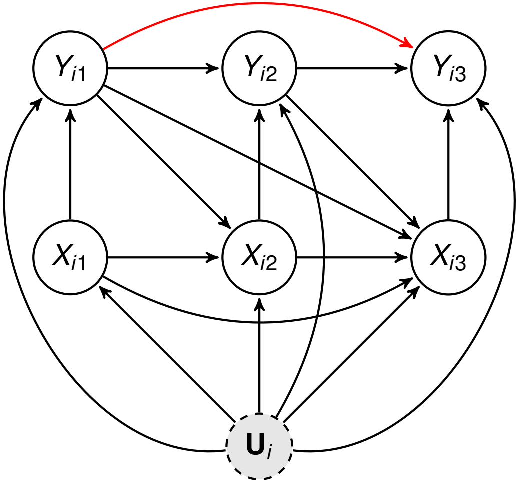
- Comparison across units within the same time rather than across different time periods within the same unit
- Marginal structural models $\rightsquigarrow$ can identify the average effect of an entire treatment sequence
- Trade-off $\rightsquigarrow$ no free lunch

Adjusting for Observed Time-varying Confounders



- $Y_{it} = \alpha_i + \beta X_{it} + \gamma^\top \mathbf{Z}_{it} + \epsilon_{it}$
- past outcomes cannot directly affect current treatment
- past outcomes cannot *indirectly* affect current treatment through $\mathbf{Z}_{it}$

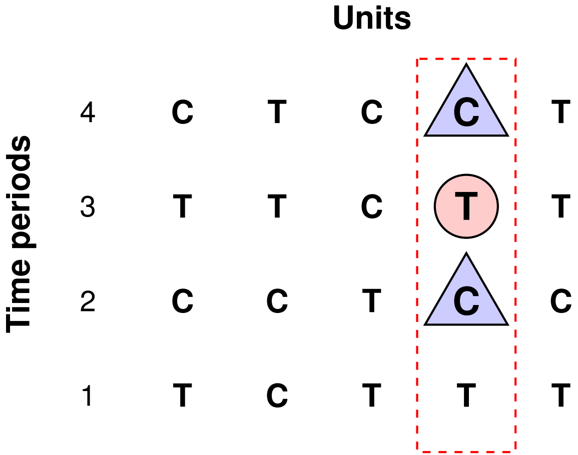
Instrumental Variables Approach



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

A Matching Framework for Fixed Effects Models

- Causal inference is all about the comparison of treatment and control observations
- FE models adjust for unit-specific unobservables through comparison across time periods within the same unit



The Within-Unit Matching Estimator

- Define: matched set $\mathcal{M}_{it}$ for observation (i, t)
- For example, one can match with all control observations:

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

- Or just match with the control observation in the previous period:

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

- A general matching estimator:

$$\hat{\tau} \quad = \quad \frac{1}{\sum_{i=1}^{N} \sum_{t=1}^{T} D_{it}} \sum_{i=1}^{N} \sum_{t=1}^{T} D_{it}(\widehat{Y_{it}(1)} - \widehat{Y_{it}(0)})$$

  where $D_{it} = \mathbf{1}\{\#\mathcal{M}_{it} > 0\}$ and

$$\widehat{Y_{it}(x)} \quad = \quad \begin{cases} Y_{it} & \text{if } X_{it} = x \\ \frac{1}{\#\mathcal{M}_{it}} \sum_{(i',t') \in \mathcal{M}_{it}} Y_{i't'} & \text{if } X_{it} = 1 - x \end{cases}$$

Matching as a Weighted Unit Fixed Effects Estimator

- *Any* within-unit matching estimator can be written as a weighted unit fixed effects estimator with different regression weights

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

where $\overline{X}_i^*$ and $\overline{Y}_i^*$ are unit-specific weighted averages

- Example: $\mathcal{M}_{it} = \{(i', t') : i' = i, X_{i't'} = 1 - X_{it}\}$ corresponds to

$$W_{it} = \begin{cases} \frac{T}{\sum_{t'=1}^{T} X_{it'}} & \text{if} \quad X_{it} = 1, \\ \frac{T}{\sum_{t'=1}^{T} (1 - X_{it'})} & \text{if} \quad X_{it} = 0. \end{cases}$$

- accommodates various identification strategies
- model-based standard errors, specification test

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., $\gamma_t = f(\mathbf{V}_t)$

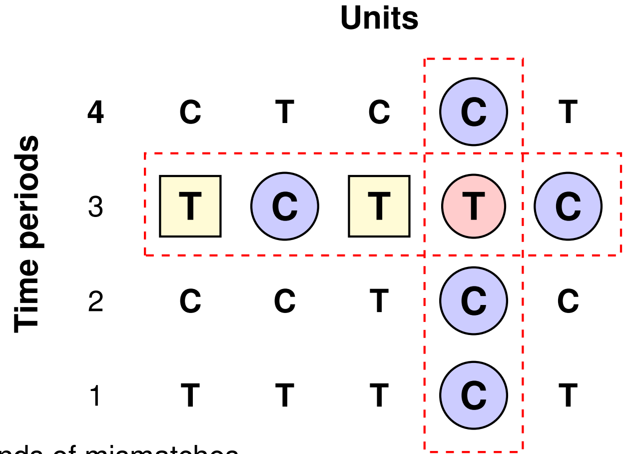
- Estimator:

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

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

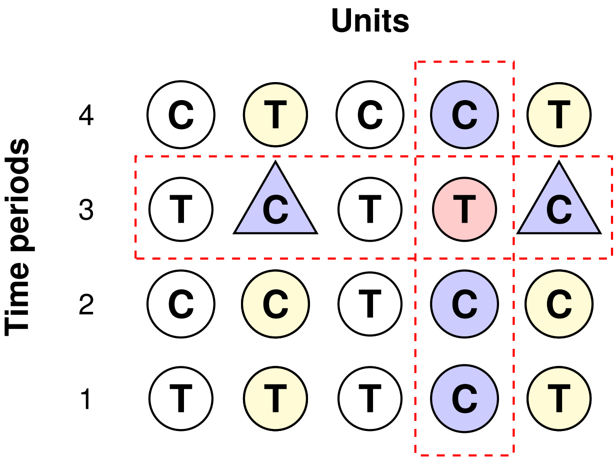
Matching and Two-way Fixed Effects Estimators

- Problem: No other unit shares the same unit and time



- Two kinds of mismatches
  1. Same treatment status
  2. Neither same unit nor same time

We Can Never Eliminate Mismatches

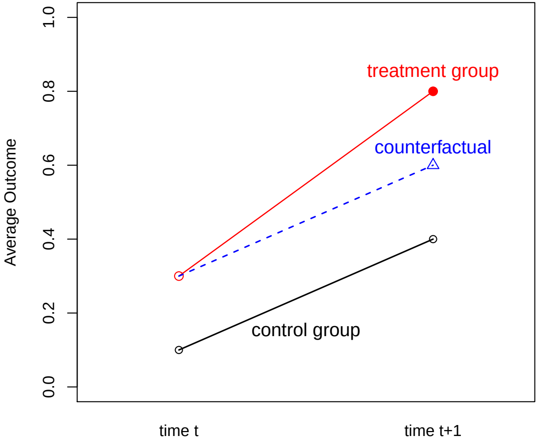


- To cancel time and unit effects, we must induce mismatches
- Solution: Difference-in-Differences

Difference-in-Differences Design

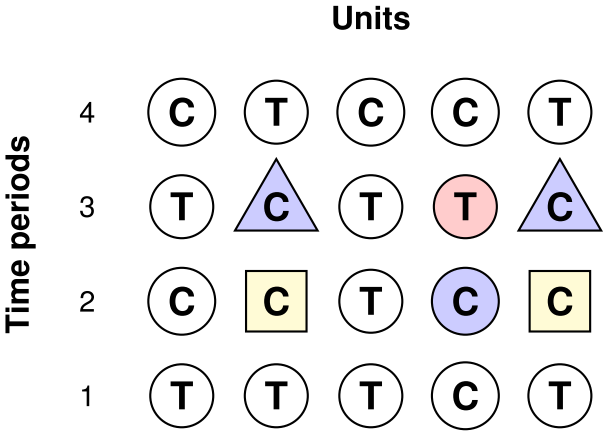
- Parallel trend assumption:

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



General DiD = Weighted Two-Way FE Effects

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



- We show the equivalence between the general DiD estimator and weighted two-way fixed effects estimator:

$$\underset{\beta}{\arg\min} \sum_{i=1}^{N} \sum_{t=1}^{T} W_{it} \{ (Y_{it} - \overline{Y}_i^* - \overline{Y}_t^* + \overline{Y}^*) - \beta (X_{it} - \overline{X}_i^* - \overline{X}_t^* + \overline{X}^*) \}^2$$

- Model-based 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

Concluding Remarks

- When should we use linear fixed effects models?
- A key (under-appreciated) causal assumption of fixed effects: past outcomes do not affect current treatment
- Tradeoff:
  1. unobserved time-invariant confounders $\rightsquigarrow$ fixed effects
  2. causal dynamics between treatment and outcome $\rightsquigarrow$ selection-on-observables
- A new matching framework:
  1. Equivalence between various matching estimators and (weighted) linear fixed effects regression estimators
  2. Model-based standard error, specification test
- R package **wfe** is available at CRAN