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

Kosuke Imai

Department of Politics
Center for Statistics and Machine Learning
Princeton University

Joint work with In Song Kim (MIT)

Seminar at University of California, San Diego
May 6, 2016

Fixed Effects Regressions in Causal Inference

- Linear fixed effects regression models are the primary workhorse for causal inference with panel data
- Researchers use them to adjust for unobserved 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*)

Motivating Questions

1. What make it possible for fixed effects regression models to adjust for unobserved confounding?
2. Are there any trade-offs when compared to the selection-on-observables approaches such as matching?
3. What are the exact causal assumptions underlying fixed effects regression models?

Main Results of the Paper

- Identify causal assumptions of one-way fixed effects estimators:
  1. Treatments do not directly affect future outcomes
  2. Outcomes do not directly affect future treatments and future time-varying confounders

  $\rightsquigarrow$ can be relaxed under the selection-on-observables approach
- Develop within-unit matching estimators to relax the functional form assumptions of linear fixed effects regression estimators
- Identify the problem of two-way fixed effects regression models
  $\rightsquigarrow$ no other observations share the same unit and time
- Propose simple ways to improve fixed effects estimators using the new matching/weighted fixed effects regression framework
- Replace the assumptions with the design-based assumptions
  $\rightsquigarrow$ before-and-after and difference-in-differences designs

Linear Regression with Unit Fixed Effects

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

- Model:

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

- Estimator: "de-meaning"

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

where $\overline{X}_i$ and $\overline{Y}_i$ are unit-specific sample means

The Standard Assumption

Assumption 1 (Strict Exogeneity)

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

where $\mathbf{X}_i$ *is a* $T \times 1$ *vector of treatment variables for unit* i

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

Causal Assumption I

Assumption 2 (No carryover effect)

Treatments do not directly affect future outcomes

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

- Potential outcome model:

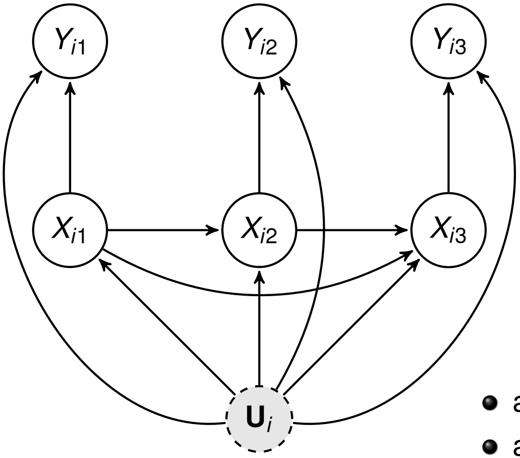
$$Y_{it}(x) \quad = \quad \alpha_i + \beta x + \epsilon_{it} \quad \text{for } x = 0, 1$$

- Average treatment effect:

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

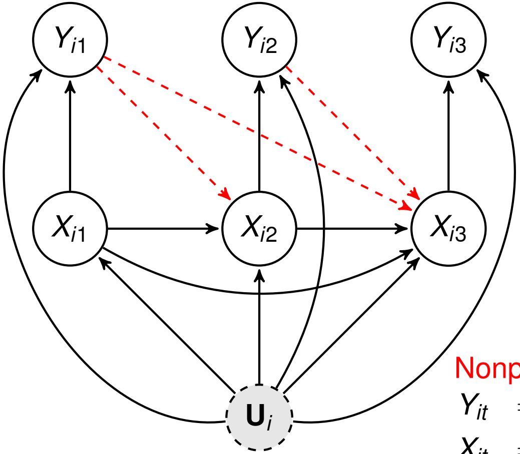
where $C_i = \mathbf{1}\{0 < \sum_{t=1}^{T} X_{it} < T\}$

Causal Directed Acyclic Graph (DAG)



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

Causal Directed Acyclic Graph (DAG)



Adding a red dashed arrow violates strict exogeneity

Nonparametric SEM (Pearl)

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

Causal Assumption II

- What randomized experiment satisfies strict exogeneity?

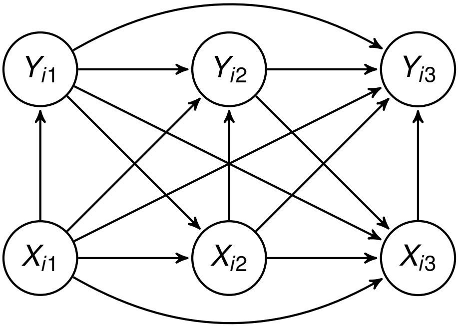
Assumption 3 (Sequential Ignorability with Unobservables)

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

- The "as-if random" assumption without conditioning on the previous outcomes
- Outcomes can *directly* affect future outcomes $\rightsquigarrow$ but no need to adjust for past outcomes
- Nonparametric identification result

An Alternative Selection-on-Observables Approach

- Marginal structural models in epidemiology (Robins)
- Risk set matching (Rosenbaum)
- Trade-off: unobserved time-invariant confounders vs. direct effect of outcome on future treatment



Within-Unit Matching Estimator

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

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

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

- The Within-unit matching estimator improves $\hat{\beta}_{\text{FE}}$ by relaxing the linearity assumption:

$$\hat{\tau}_{\text{match}} \quad = \quad \frac{1}{\sum_{i=1}^{N} C_i}\sum_{i=1}^{N} C_i\left(\frac{\sum_{t=1}^{T} X_{it}Y_{it}}{\sum_{t=1}^{T} X_{it}} - \frac{\sum_{t=1}^{T}(1-X_{it})Y_{it}}{\sum_{t=1}^{T}(1-X_{it})}\right)$$

Constructing a General Matching Estimator

- $\mathcal{M}_{it}$: matched set for observation (i, t)
- For the within-unit matching estimator,

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

- A general matching estimator just introduced:

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

where $D_{it} = \mathbf{1}\{\#\mathcal{M}(i,t) > 0\}$ and

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

Unit Fixed Effects Estimator as a Matching Estimator

- "de-meaning" $\rightsquigarrow$ match with all other observations within the same unit:

$$\mathcal{M}(i,t) \;=\; \{(i',t') : i' = i, t' \neq t\}$$

- mismatch: observations with the same treatment status
- Unit fixed effects estimator adjusts for mismatches:

$$\hat{\beta}_{\text{FE}} \;=\; \frac{1}{K}\left\{\frac{1}{\sum_{i=1}^{N}\sum_{t=1}^{T} D_{it}}\sum_{i=1}^{N}\sum_{t=1}^{T} D_{it}\left(\widehat{Y_{it}(1)} - \widehat{Y_{it}(0)}\right)\right\}$$

  where K is the proportion of proper matches

- The within-unit matching estimator eliminates all mismatches

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

- The proposed within-matching estimator:

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

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

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

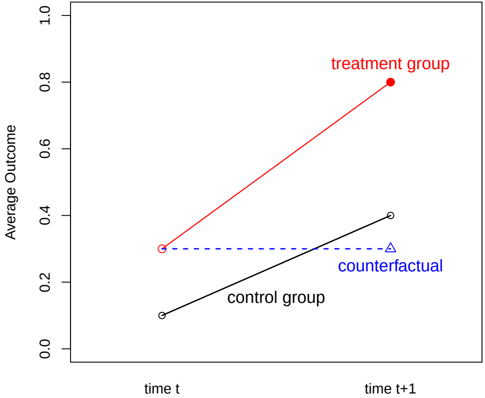
- We show how to construct regression weights for different matching estimators (i.e., different matched sets)
- Idea: count the number of times each observation is used for matching

- Benefits:
  - computational efficiency
  - model-based standard errors
  - double-robustness $\rightsquigarrow$ matching estimator is consistent even when linear fixed effects regression is the true model
  - specification test (White 1980) $\rightsquigarrow$ null hypothesis: linear fixed effects regression is the true model

Before-and-After Design

- The assumption that outcomes do not directly affect future treatments may not be credible
- Replace it with the design-based assumption:

$$\mathbb{E}(Y_{it}(x) \mid X_{it} = x') \quad = \quad \mathbb{E}(Y_{i,t-1}(x) \mid X_{i,t-1} = 1 - x')$$



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

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

- It is also the first differencing estimator:

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

- "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!

Remarks on Other Important Issues

1. Adjusting for observed time-varying confounding $\mathbf{Z}_{it}$
   - Proposes within-unit matching estimators that adjust for $\mathbf{Z}_{it}$
   - Key assumption: outcomes neither directly affect future treatments nor future time-varying confounders
2. Adjusting for past treatments
   - Impossible to adjust for all past treatments within the same unit
   - Researchers must decide the number of past treatments to adjust
3. Adjusting for past outcomes
   - No need to adjust for past outcomes if they do not directly affect future treatments
   - If they do, the strict exogeneity assumption will be violated
   - Past outcomes as instrumental variables (Arellano and Bond) $\rightsquigarrow$ often not credible

No free lunch: adjustment for unobservables comes with costs

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

- β_{FE}: bias due to time effects
- β_{FEtime}: bias due to unit effects
- β_{pool}: bias due to both time and unit effects

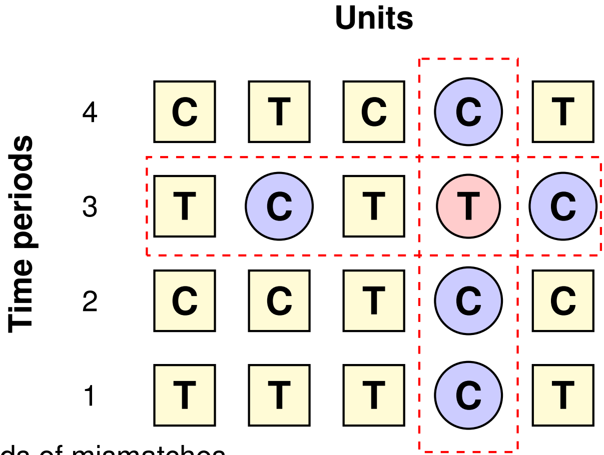
$$\hat{\beta}_{\text{FE2}} = \frac{\omega_{\text{FE}} \times \hat{\beta}_{\text{FE}} + \omega_{\text{FEtime}} \times \hat{\beta}_{\text{FEtime}} - \omega_{\text{pool}} \times \hat{\beta}_{\text{pool}}}{w_{\text{FE}} + w_{\text{FEtime}} - w_{\text{pool}}}$$

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

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

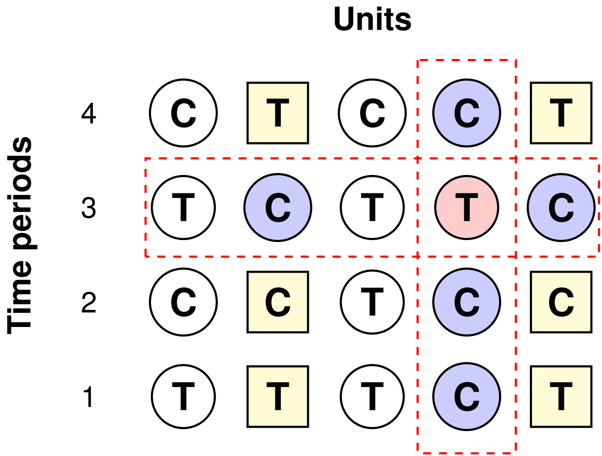
Matching and Two-way Fixed Effects Estimators

- 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

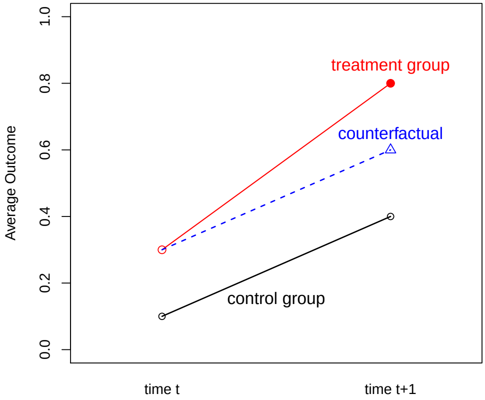


- To cancel time and unit effects, we must induce mismatches
- No weighted two-way fixed effects model eliminates mismatches

Difference-in-Differences Design

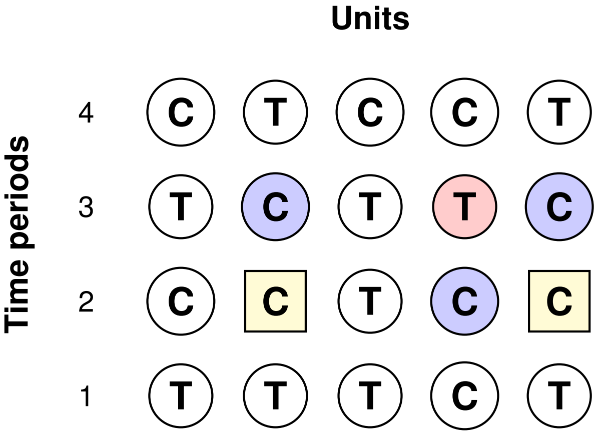
- Replace the model-based assumption with the design-based one
- Parallel trend assumption:

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

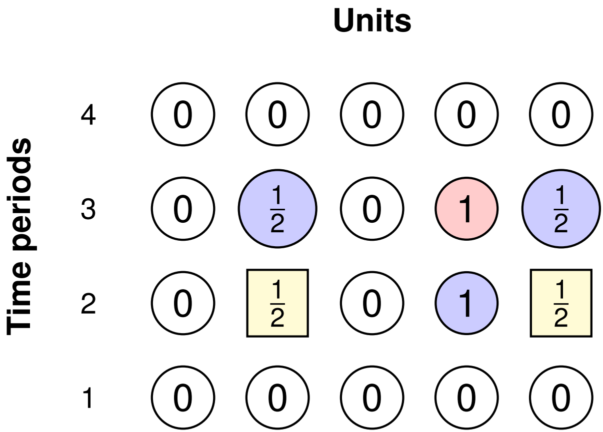


General DiD = Weighted Two-Way FE Effects

- $2 \times 2 \rightsquigarrow$ standard two-way fixed effects estimator works
- General setting: Multiple time periods, repeated treatments



- Regression weights:



- Weights can be negative $\Longrightarrow$ the method of moments estimator
- Fast computation is still available

Effects of GATT Membership on International Trade

1. Controversy
   - Rose (2004): No effect of GATT membership on trade
   - Tomz et al. (2007): Significant effect with non-member participants
2. The central role of fixed effects models:
   - Rose (2004): one-way (year) fixed effects for dyadic data
   - Tomz *et al.* (2007): two-way (year and dyad) fixed effects
   - Rose (2005): "I follow the profession in placing most confidence in the fixed effects estimators; I have no clear ranking between country-specific and country pair-specific effects."
   - Tomz *et al.* (2007): "We, too, prefer FE estimates over OLS on both theoretical and statistical ground"

Data and Methods

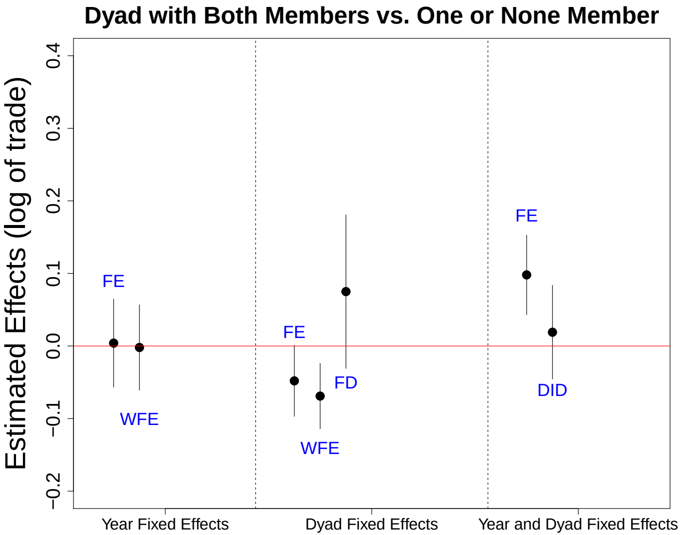
1. Data
   - Data set from Tomz et al. (2007)
   - Effect of GATT: 1948 – 1994
   - 162 countries, and 196,207 (dyad-year) observations
2. Year fixed effects model:

$$\ln Y_{it} = \alpha_t + \beta X_{it} + \delta^\top \mathbf{Z}_{it} + \epsilon_{it}$$

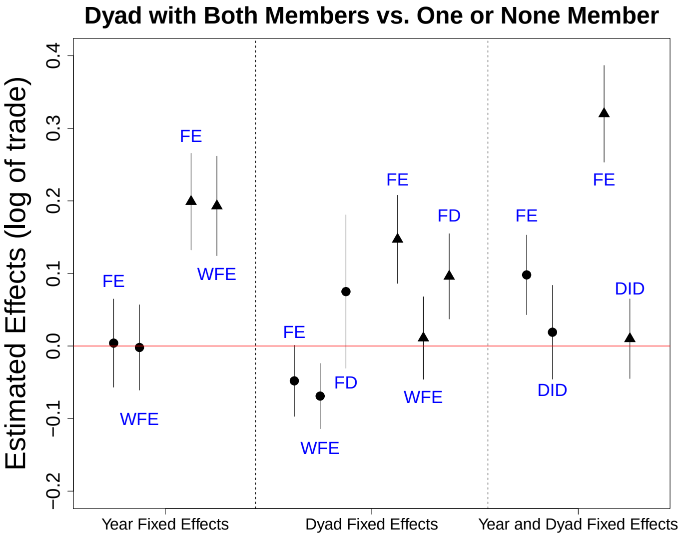
   - Y_{it}: trade volume
   - X_{it}: membership (formal/participants) Both vs. At most one
   - $\mathbf{Z}_{it}$: 15 dyad-varying covariates (e.g., log product GDP)
3. Weighted one-way fixed effects model:

$$\underset{(\alpha,\beta,\delta)}{\arg\min} \sum_{i=1}^{N} \sum_{t=1}^{T} W_{it}(\ln Y_{it} - \alpha_t - \beta X_{it} - \delta^\top Z_{it})^2$$

Empirical Results: Formal Membership



Empirical Results



Concluding Remarks

- Linear fixed effects models are attractive because they can adjust for unobserved confounders
- However, this advantage comes at costs
- Two key causal assumptions:
  1. treatments do not directly affect future outcomes
  2. outcomes do not directly affect future treatments and future time-varying covariates
- These assumptions can be relaxed under alternative selection-on-observables approaches
- Improve fixed effects estimators:
  1. Within-unit matching estimator $\rightsquigarrow$ no linearity assumption
  2. Design-based assumptions $\rightsquigarrow$ before-and-after, differnece-in-differences
  3. All of these can be written as weighted fixed effects regression
- R package **wfe** is available at CRAN

Send comments and suggestions to:

kimai@Princeton.Edu

More information about this and other research:

http://imai.princeton.edu