Understanding and Improving Fixed Effects Regression Models for Causal Inference

Kosuke Imai **In Song Kim**

Department of Politics
Princeton University

Political Methodology Colloquium
September 30, 2011

Motivation

- Fixed effects models are the primary workhorse for causal inference in applied panel data analysis
- Researchers use them to adjust for unobservables (omitted variables, endogeneity, selection bias, confoundedness ...):
  - "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*)
- Fixed effects models are often said to be superior to matching estimators because the latter can only adjust for observables
- **Question:** What are the exact causal assumptions underlying fixed effects models?

Main Results

1. Standard (one-way and two-way) fixed effects estimators are equivalent to particular matching estimators
2. Common belief that fixed effects models adjust for unobservables but matching does not is wrong
3. Identify the information used implicitly to estimate counterfactual outcomes under fixed effects models
4. Point out potential sources of bias and inefficiency in fixed effects estimators
5. Propose simple ways to improve fixed effects estimators using weighted fixed effects regressions

Matching and Regression in Cross-Section Settings

Units	1	2	3	4	5
Treatment status	T	T	C	C	T
Outcome	Y_1	Y_2	Y_3	Y_4	Y_5

- Estimating the Average Treatment Effect via matching

$$
\begin{aligned}
Y_1 &- \frac{1}{2}(Y_3 + Y_4) \\
Y_2 &- \frac{1}{2}(Y_3 + Y_4) \\
\frac{1}{3}(Y_1 + Y_2 + Y_5) &- Y_3 \\
\frac{1}{3}(Y_1 + Y_2 + Y_5) &- Y_4 \\
Y_5 &- \frac{1}{2}(Y_3 + Y_4)
\end{aligned}
$$

Matching Representation of Simple Regression

- Cross-section simple linear regression model:

$$Y_i = \alpha + \beta X_i + \epsilon_i$$

- Binary treatment: $X_i \in \{0, 1\}$
- Equivalent matching estimator:

$$\hat{\beta} = \frac{1}{N} \sum_{i=1}^{N} \left(\widehat{Y_i(1)} - \widehat{Y_i(0)} \right)$$

where

$$\widehat{Y_i(1)} = \begin{cases} Y_i & \text{if } X_i = 1 \\ \frac{1}{\sum_{i'=1}^{N} X_{i'}} \sum_{i'=1}^{N} X_{i'} Y_{i'} & \text{if } X_i = 0 \end{cases}$$

$$\widehat{Y_i(0)} = \begin{cases} \frac{1}{\sum_{i'=1}^{N} (1 - X_{i'})} \sum_{i'=1}^{N} (1 - X_{i'}) Y_{i'} & \text{if } X_i = 1 \\ Y_i & \text{if } X_i = 0 \end{cases}$$

- Treated units matched with the average of non-treated units

Fixed Effects Regression

- Simple (one-way) fixed effects regression:

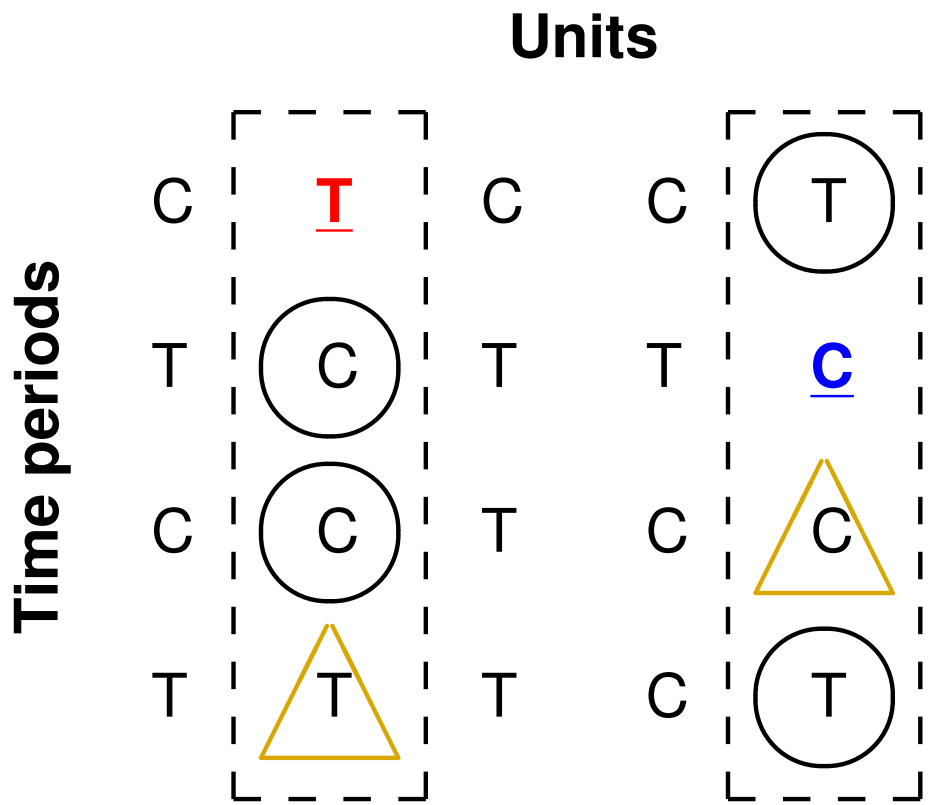
$$Y_{it} = \alpha_i + \beta X_{it} + \epsilon_{it}$$

- Binary treatment: $X_{it} \in \{0, 1\}$

- Unit fixed effects $\Longrightarrow$ within-unit comparison
- Estimates of all counterfactual outcomes come from other time periods within the same unit

- How is this done under the fixed effects model?

Mismatches in One-way Fixed Effects Model



- T: treated observations
- C: control observations
- **Circles**: Proper matches
- **Triangles**: "Mismatches" $\Longrightarrow$ attenuation bias

Matching Representation of Fixed Effects Regression

Proposition 1

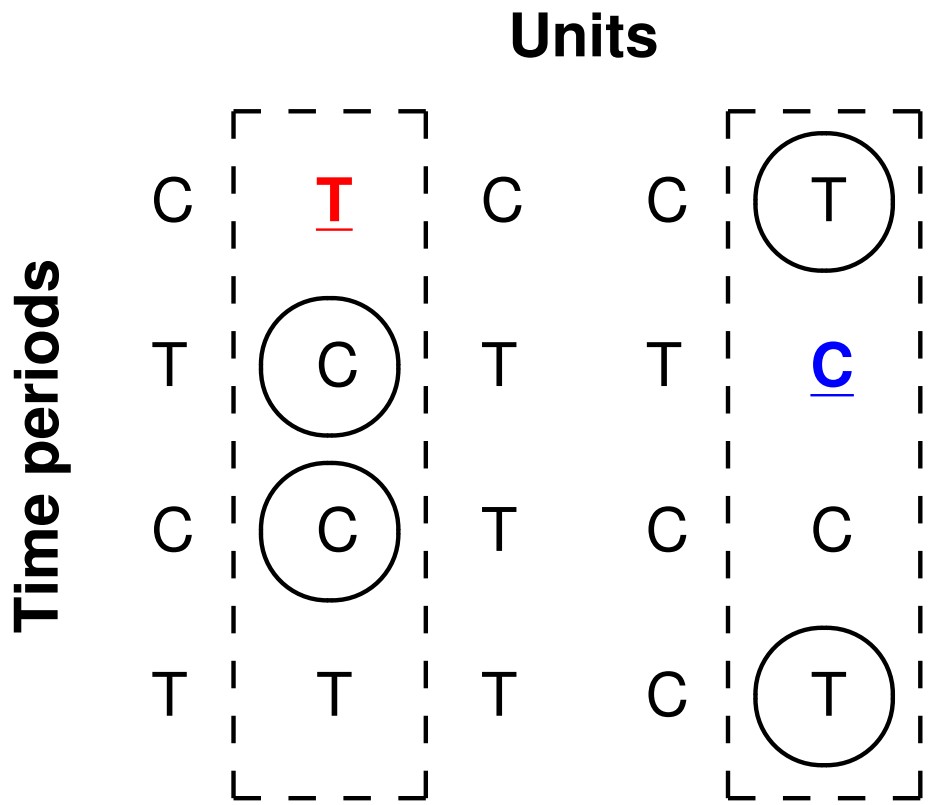
$$\hat{\beta}^{FE} = \frac{1}{K}\left\{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\widehat{Y_{it}(1)} - \widehat{Y_{it}(0)}\right)\right\},$$

$$\widehat{Y_{it}(x)} = \begin{cases} Y_{it} & \text{if } X_{it} = x \\ \frac{1}{T-1}\sum_{t' \neq t} Y_{it'} & \text{if } X_{it} = 1 - x \end{cases} \quad \text{for} \quad x = 0, 1$$

$$K = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{X_{it} \cdot \frac{1}{T-1}\sum_{t' \neq t}(1 - X_{it'}) + (1 - X_{it}) \cdot \frac{1}{T-1}\sum_{t' \neq t} X_{it'}\right\}.$$

- K: average proportion of proper matches across all observations
- More mismatches $\Longrightarrow$ larger adjustment
- Adjustment is required except very special cases
- "Fixes" attenuation bias
- Fixed effects estimator is a special case of matching estimators

Unadjusted Matching Estimator



- Only equal to fixed effects estimator if heterogeneity in either treatment assignment or treatment effect is non-existent

Unadjusted Matching as **Weighted** FE Estimator

Proposition 2

The unadjusted matching estimator

$$\hat{\beta}^M = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(\widehat{Y_{it}(1)} - \widehat{Y_{it}(0)} \right)$$

where

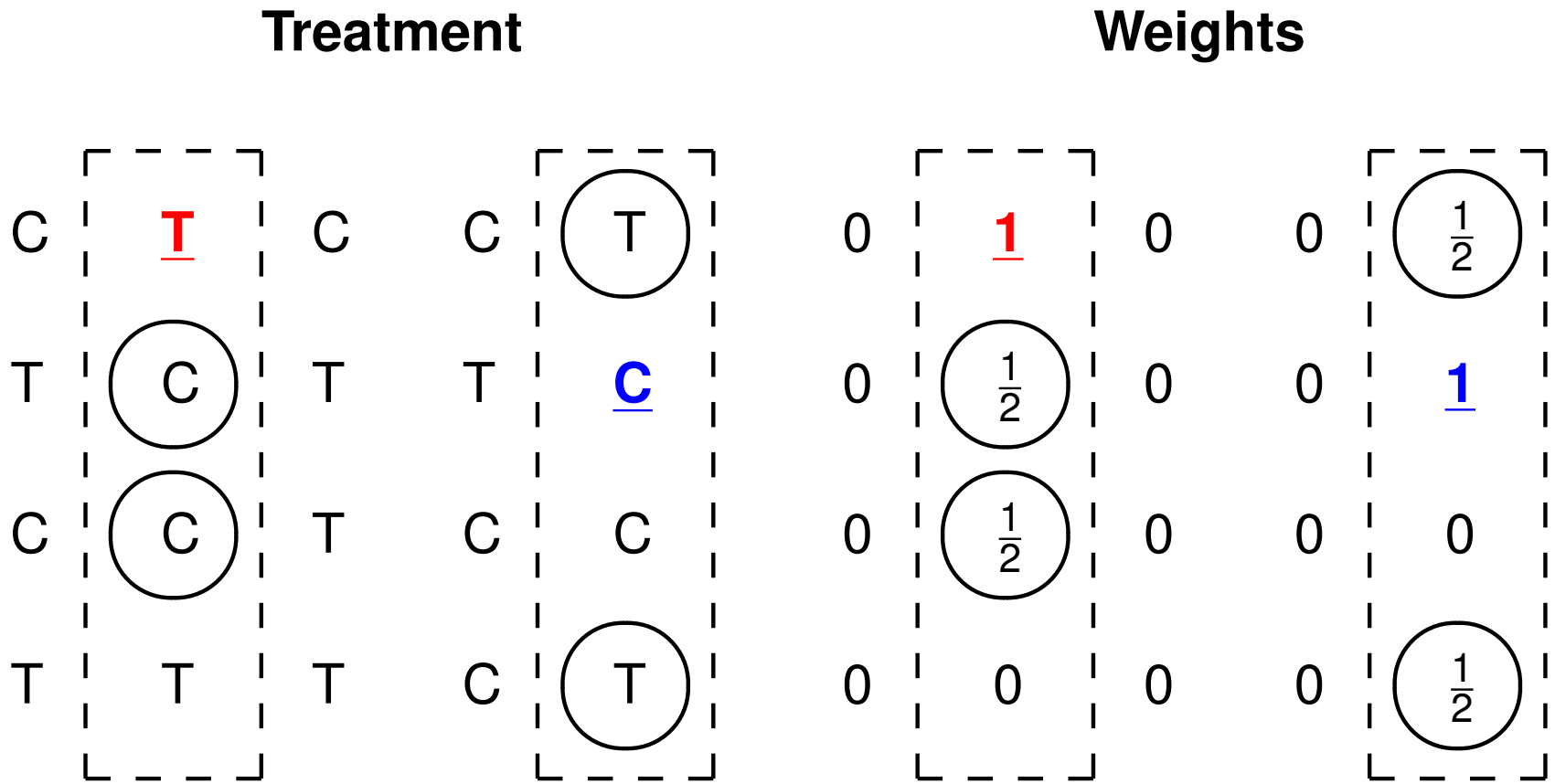
$$\widehat{Y_{it}(1)} = \begin{cases} Y_{it} & \text{if } X_{it} = 1 \\ \frac{\sum_{t'=1}^{T} X_{it'} Y_{it'}}{\sum_{t'=1}^{T} X_{it'}} & \text{if } X_{it} = 0 \end{cases} \quad \text{and} \quad \widehat{Y_{it}(0)} = \begin{cases} \frac{\sum_{t'=1}^{T} (1 - X_{it'}) Y_{it'}}{\sum_{t'=1}^{T} (1 - X_{it'})}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

is equivalent to the weighted fixed effects model

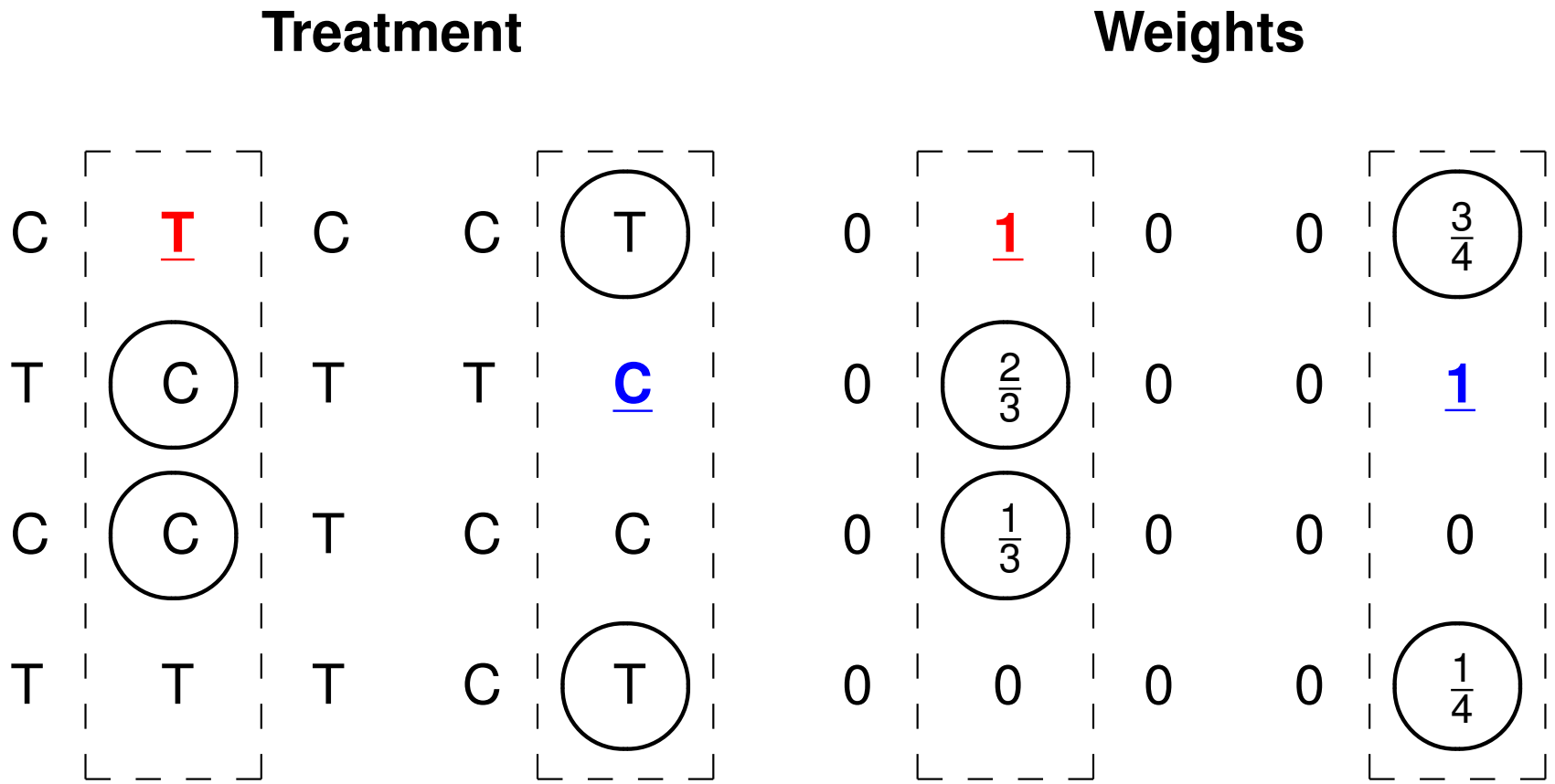
$$(\hat{\alpha}^M, \hat{\beta}^M) = \underset{(\alpha, \beta)}{\arg\min} \sum_{i=1}^{N} \sum_{t=1}^{T} W_{it} (Y_{it} - \alpha_i - \beta X_{it})^2$$

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

Equal Weights



Different Weights



Theorem: General Equivalence between Weighted Fixed Effects and Matching Estimators

General matching estimator

$$\tilde{\beta}^M = \frac{1}{\sum_{i=1}^N \sum_{t=1}^T C_{it}} \sum_{i=1}^N \sum_{t=1}^T C_{it} \left(\widehat{Y_{it}(1)} - \widehat{Y_{it}(0)} \right)$$

where $0 \leq C_{it} < \infty$, $\sum_{t=1}^T \sum_{i=1}^N C_{it} > 0$,

$$\widehat{Y_{it}(1)} = \begin{cases} Y_{it} & \text{if } X_{it} = 1 \\ \sum_{t'=1}^T v_{it}^{it'} X_{it'} Y_{it'} & \text{if } X_{it} = 0 \end{cases}$$

$$\widehat{Y_{it}(0)} = \begin{cases} \sum_{t'=1}^T v_{it}^{it'} (1 - X_{it'}) Y_{it'} & \text{if } X_{it} = 1 \\ Y_{it} & \text{if } X_{it} = 0 \end{cases}$$

$$\sum_{t'=1}^T v_{it}^{it'} X_{it'} = \sum_{t'=1}^T v_{it}^{it'} (1 - X_{it'}) = 1$$

is equivalent to the weighted one-way fixed effects estimator

$$W_{it} = \sum_{i'=1}^N \sum_{t'=1}^T w_{it}^{i't'} \quad \text{and} \quad w_{it}^{i't'} = \begin{cases} C_{it} & \text{if } (i,t) = (i',t') \\ v_{it}^{it'} C_{i't'} & \text{if } (i,t) \in \mathcal{M}_{i't'} \\ 0 & \text{otherwise.} \end{cases}$$

Adjusting for Time-varying Observed Confounders

- Confounders Z_{it} are correlated with treatment and outcome
- What do the above results (without such confounders) imply?

- Linear regression adjustment with:

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

- *Ex post* interpretation:

$$Y_{it} - \hat{\delta}^\top Z_{it} \quad = \quad \alpha_i + \beta X_{it} + \epsilon_{it}$$

- Inverse-propensity score weighting with normalized weights

$$\hat{\beta}^W \quad = \quad \frac{1}{N}\sum_{i=1}^{N}\left\{\sum_{t=1}^{T}\frac{X_{it}Y_{it}}{\hat{\pi}(Z_{it})}\Big/\sum_{t=1}^{T}\frac{X_{it}}{\hat{\pi}(Z_{it})} - \sum_{t=1}^{T}\frac{(1-X_{it})Y_{it}}{1-\hat{\pi}(Z_{it})}\Big/\sum_{t=1}^{T}\frac{(1-X_{it})}{1-\pi(Z_{it})}\right\}$$

  where $\pi(Z_{it}) = \Pr(X_{it} = 1 \mid Z_{it})$ is the propensity score
- within-unit weighting followed by across-units averaging

Propensity Score Weighting Estimator is Equivalent to Transformed Weighted FE Estimator

Proposition 3

$$(\hat{\alpha}^W, \hat{\beta}^W) = \underset{(\alpha,\beta)}{\arg\min} \sum_{i=1}^{N} \sum_{t=1}^{T} W_{it}(Y_{it}^* - \alpha_i - \beta X_{it})^2$$

where the transformed outcome Y_{it}^* is,

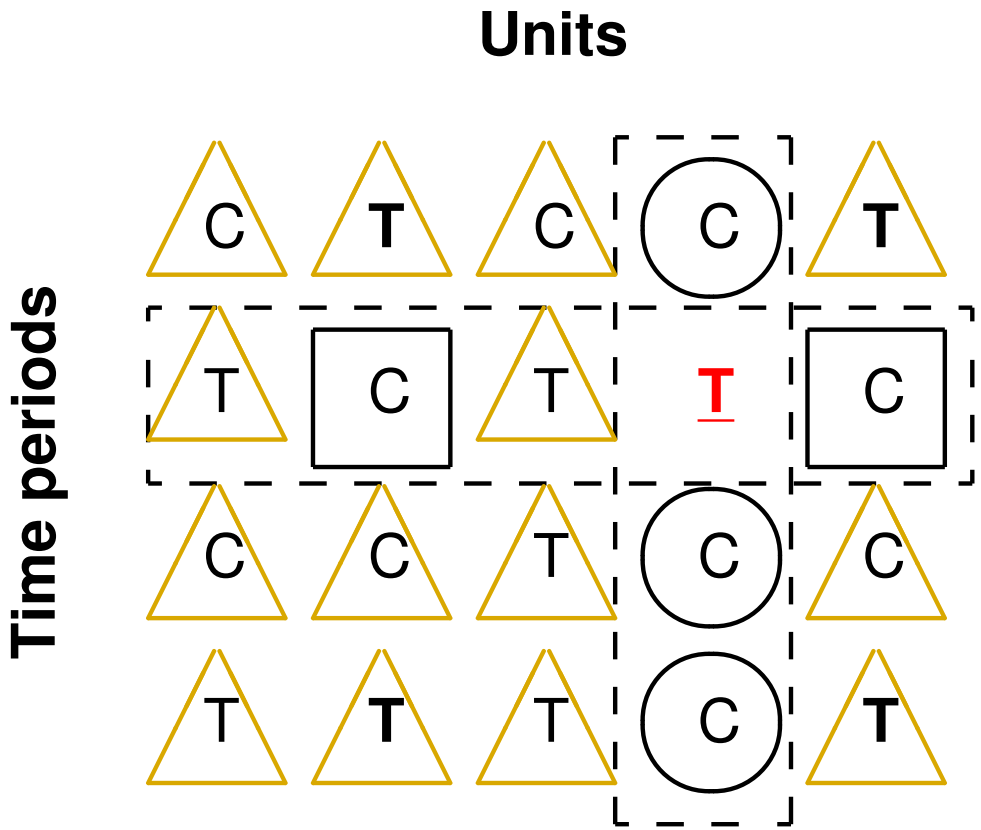
$$Y_{it}^* = \begin{cases} \frac{\left(\sum_{t'=1}^{T} X_{it'}\right) Y_{it}}{\hat{\pi}(Z_{it})} \Big/ \sum_{t'=1}^{T} \frac{X_{it'}}{\hat{\pi}(Z_{it'})} & \text{if} \quad X_{it} = 1 \\ \frac{\left\{\sum_{t'=1}^{T} (1 - X_{it'})\right\} Y_{it}}{1 - \hat{\pi}(Z_{it})} \Big/ \sum_{t'=1}^{T} \frac{(1 - X_{it'})}{1 - \pi(Z_{it'})} & \text{if} \quad X_{it} = 0 \end{cases}$$

and the weights are the same as before

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

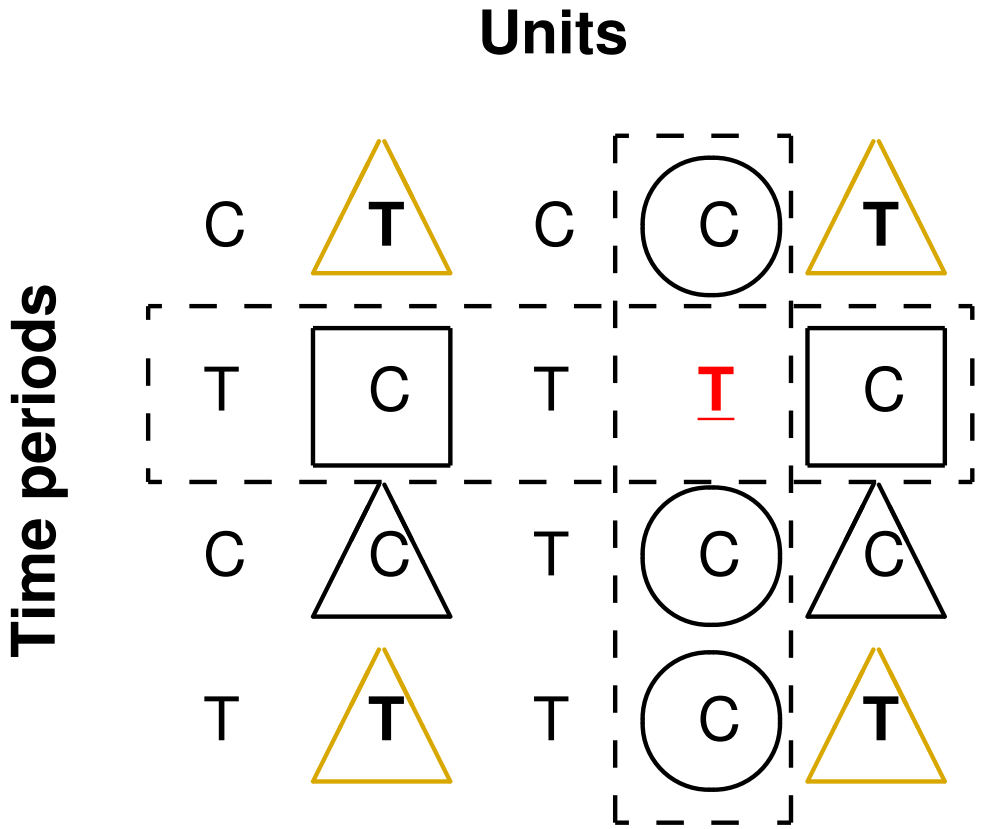
Mismatches in Two-way FE Model

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



- **Triangles**: Two kinds of mismatches
  - Same treatment status
  - Neither same unit nor same time

Mismatches in Weighted Two-way FE Model



- Some mismatches can be eliminated
- You can NEVER eliminate them all

Two-way **Weighted** FE Estimator

Proposition 4

The adjusted matching estimator

$$\hat{\beta}^{M^*} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{K_{it}}\left(\widehat{Y_{it}(1)} - \widehat{Y_{it}(0)}\right)$$

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

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

$$K_{it} = \frac{m_{it}n_{it}}{m_{it}n_{it} + a_{it}}$$

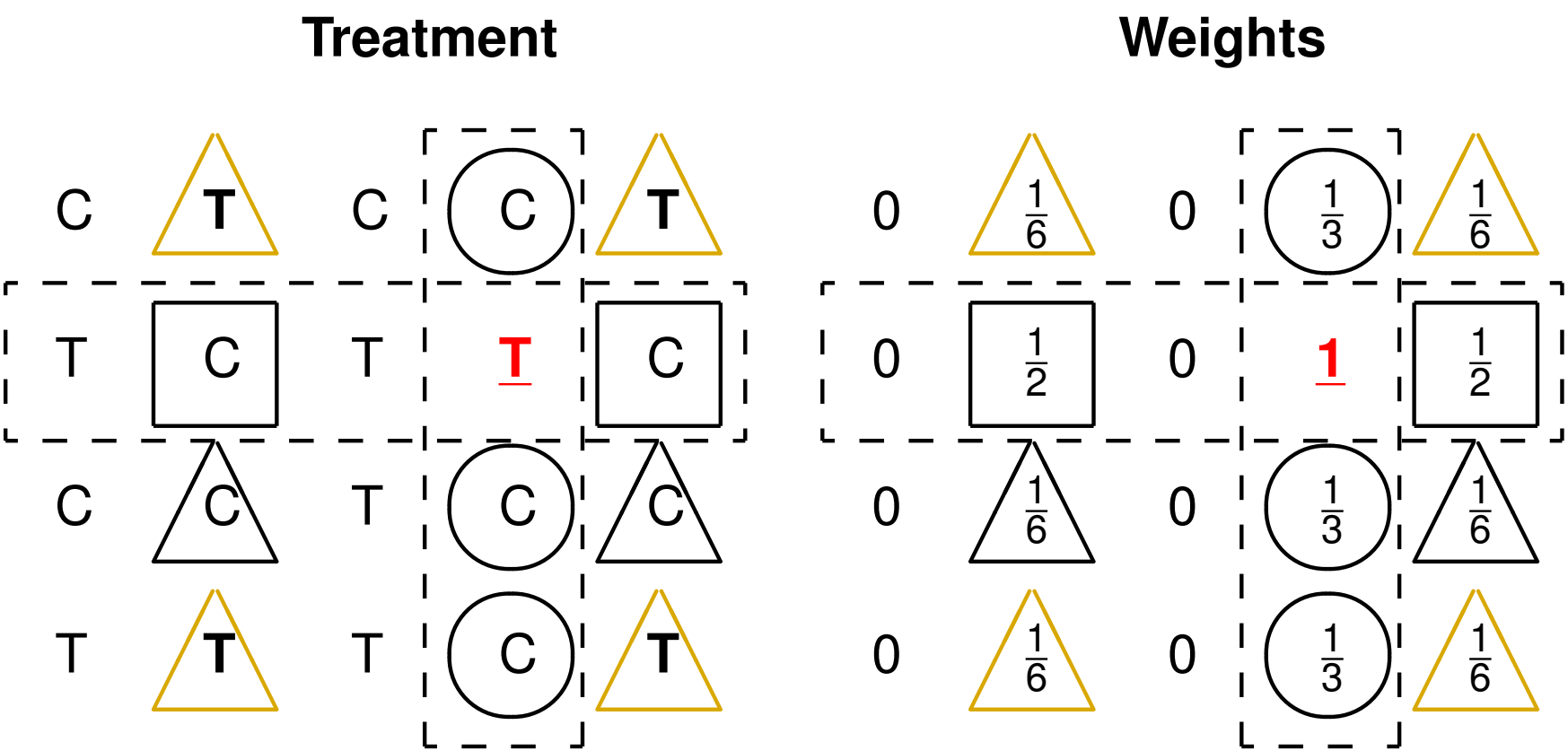
and $m_{it} = |\mathcal{M}_{it}|$, $n_{it} = |\mathcal{N}_{it}|$, and $a_{it} = |\mathcal{A}_{it} \bigcap \{(i',t') : X_{i't'} = X_{it}\}|$.

is equivalent to the following weighted two-way fixed effects estimator,

$$(\hat{\alpha}^{M^*}, \hat{\gamma}^{M^*}, \hat{\beta}^{M^*}) = \underset{(\alpha,\beta,\gamma)}{\arg\min}\sum_{i=1}^{N}\sum_{t=1}^{T} W_{it}(Y_{it} - \alpha_i - \gamma_t - \beta X_{it})^2$$

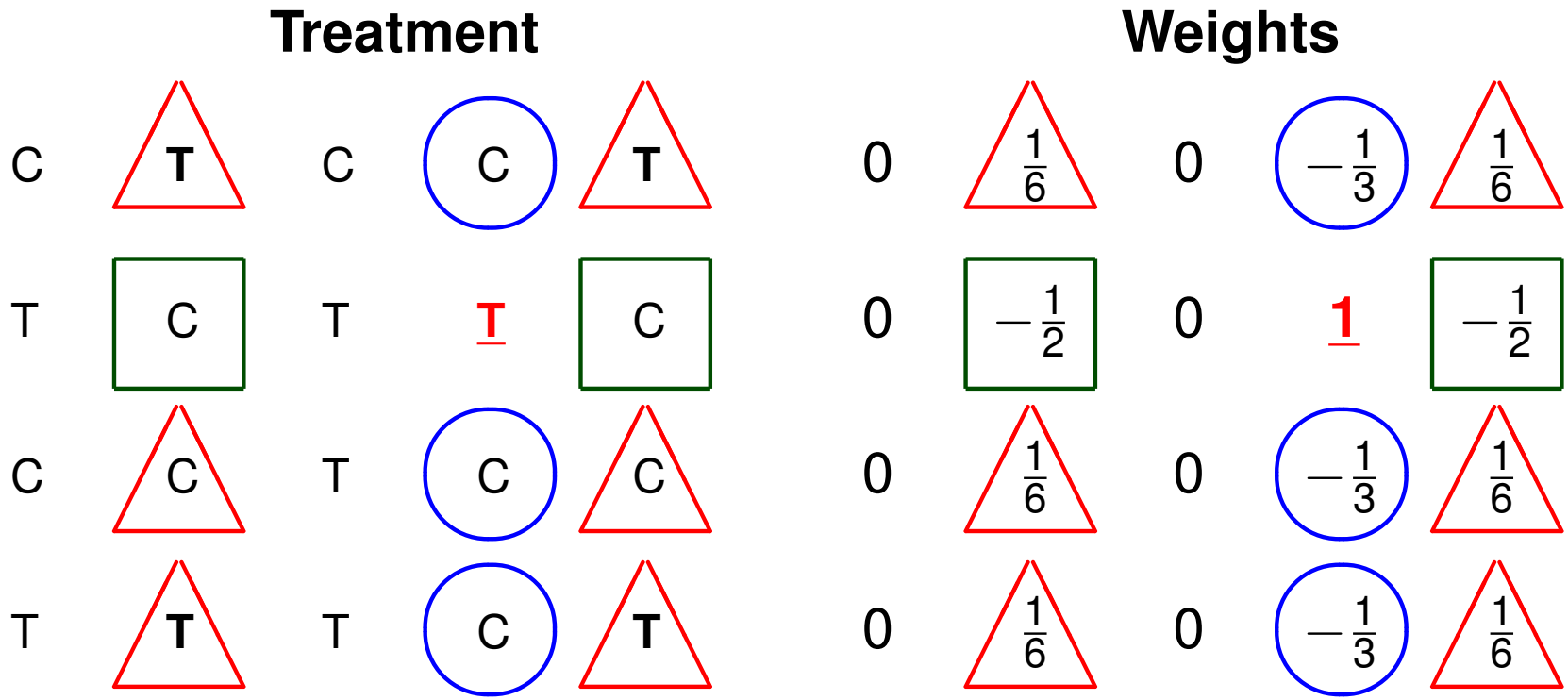
Weighted Two-way Fixed Effects Model

$$\hat{\beta}^{M*} = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{1}{K_{it}} \left(\widehat{Y_{it}(1)} - \widehat{Y_{it}(0)} \right)$$



Proof by Picture: $\sum_{i=1}^{N}\sum_{t=1}^{T} W_{it}(2X_{it}-1)\alpha_i^* = \sum_{i=1}^{N}\sum_{t=1}^{T} W_{it}(2X_{it}-1)\gamma_t^* = 0$

$$W_{it} = \sum_{i'=1}^{N}\sum_{t'=1}^{T} w_{it}^{i't'} \quad \text{and} \quad w_{it}^{i't'} = \begin{cases} \frac{m_{i't'}\, n_{i't'}}{m_{i't'}\, n_{i't'} + a_{i't'}} & \text{if } (i,t) = (i',t') \\ \frac{n_{i't'}}{m_{i't'}\, n_{i't'} + a_{i't'}} & \text{if } (i,t) \in \mathcal{M}_{i't'} \\ \frac{m_{i't'}}{m_{i't'}\, n_{i't'} + a_{i't'}} & \text{if } (i,t) \in \mathcal{N}_{i't'} \\ \frac{(2X_{it}-1)(2X_{i't'}-1)}{m_{i't'}\, n_{i't'} + a_{i't'}} & \text{if } (i,t) \in \mathcal{A}_{i't'} \\ 0 & \text{otherwise.} \end{cases}$$



Effects of GATT Membership on International Trade

1. Theory
   - Bagwell and Staiger (1999): Terms-of-trade incentives
   - Maggi and Rodrgues-Clare (2007): Domestic political incentives

2. Controversy
   - Rose (2004): No effect of GATT membership on trade
   - Tomz et al. (2007): Significant effect with non-member participants
   - Gowa and Kim (2005), Subramanian and Wei (2007): Asymmetrical effects

3. 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:

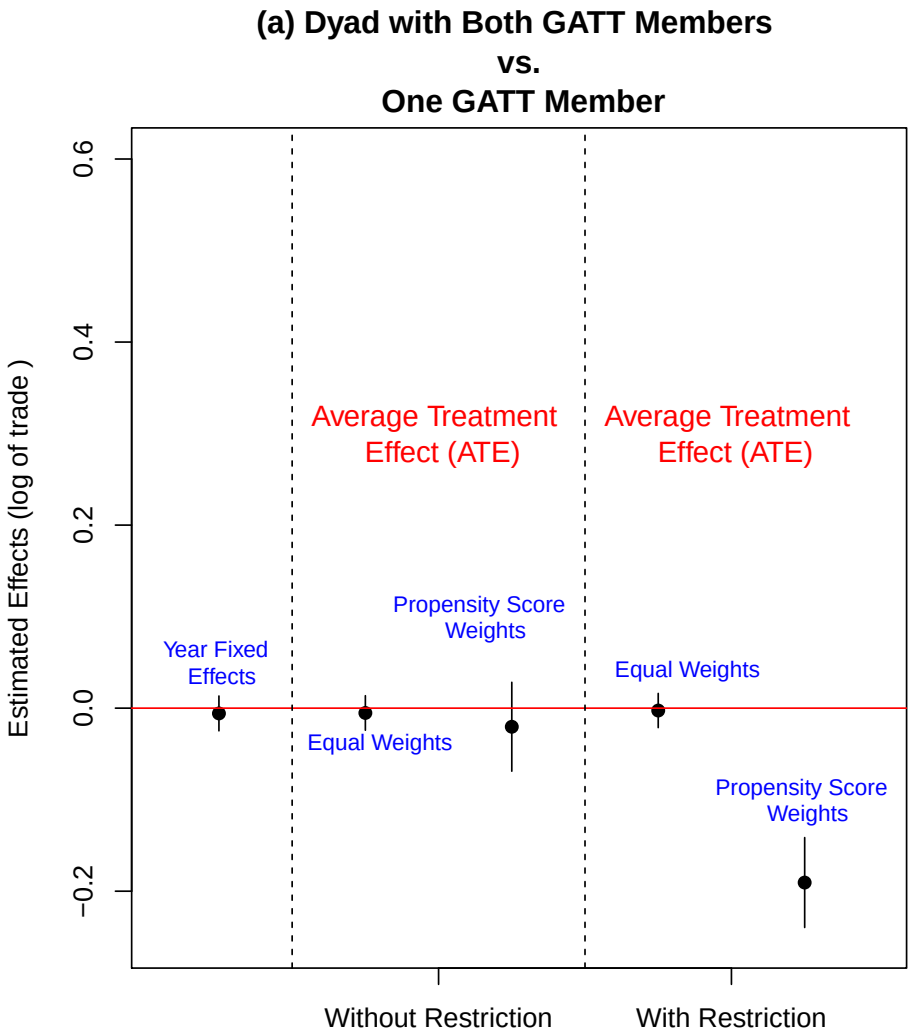
$$\ln Y_{it} = \alpha_t + \beta X_{it} + \delta^\top Z_{it} + \epsilon_{it}$$

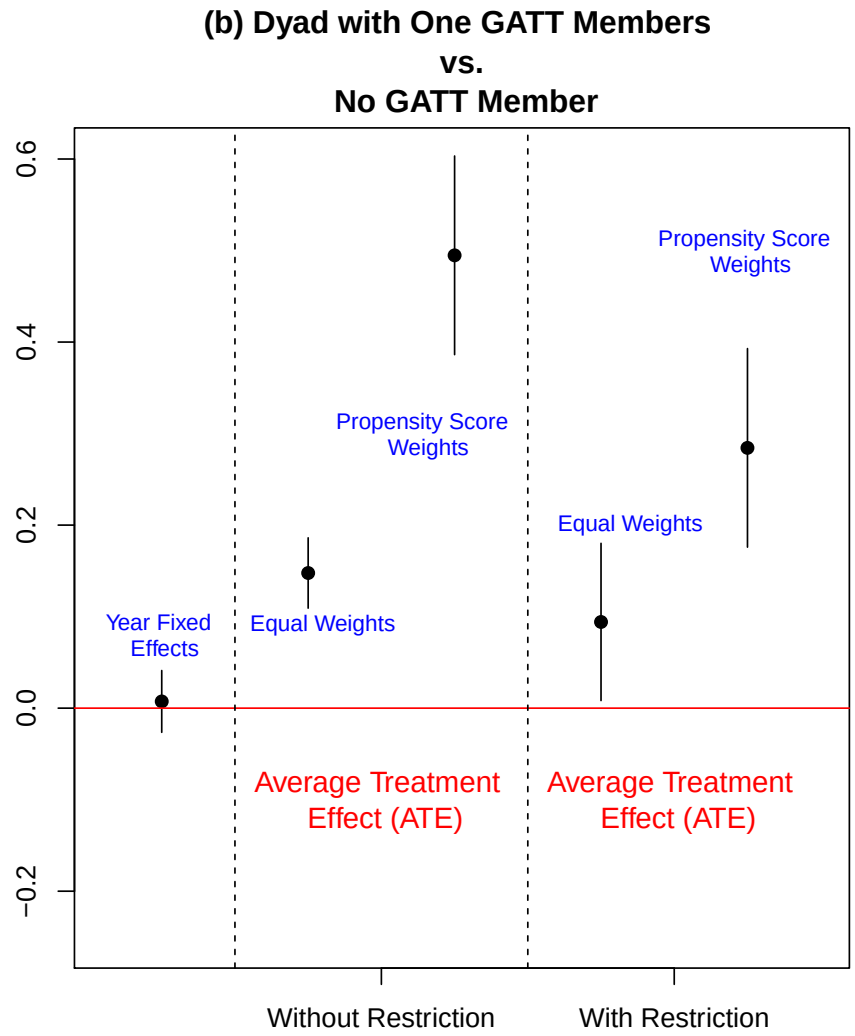
   - X_{it}: *Formal* membership (1) Both vs. One, (2) One vs. None
   - Z_{it}: 15 dyad-varying covariates (e.g., log product GDP)
3. Weighted one-way fixed effects model:

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

   - Equal weights
   - Inverse-propensity score weighting
   - With and without restriction (one country shared)

Empirical Results





Concluding Remarks and Practical Suggestions

- FE estimators are special cases of matching estimators
- FE models are not a magic bullet solution to endogeneity
- Key Question: "Where are the counterfactuals coming from?"
- Results can be sensitive to the underlying causal assumptions
- Standard (one-way) FE can be improved by the weighted FE regressions
- Time-varying covariates can be incorporated either through linear adjustment or propensity score weighting
- Use of two-way FE estimator is difficult to justify
- Second fixed effect can be incorporated into propensity score under the one-way FE framework

Future Research

1. Development of software for easy implementation
2. What about random effects estimators?
3. Simultaneous within-unit and within-time-period comparison
4. Bayesian modeling approach (joint work with Xun Pang)