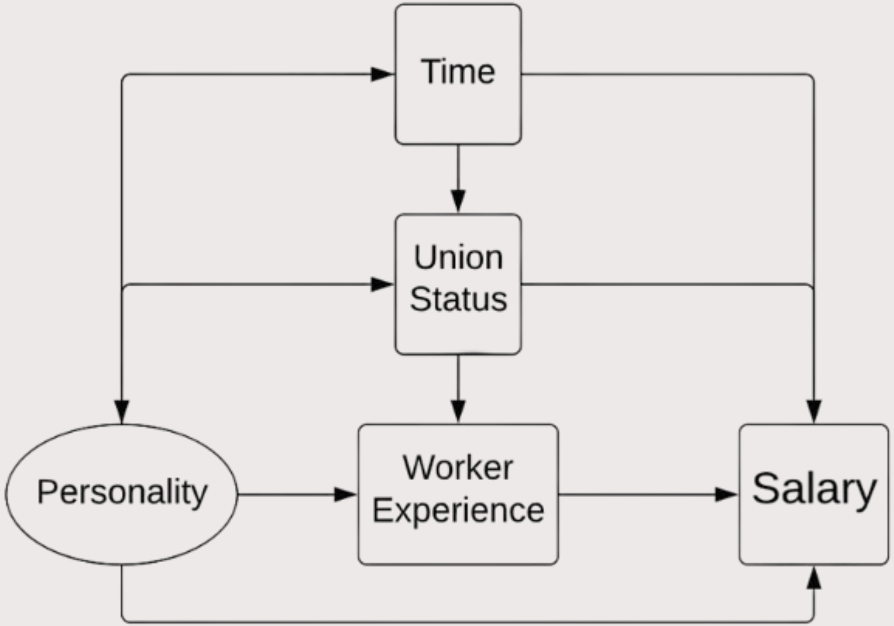
Demystifying the Relationship Between Fixed/Random Effects and Unmeasured Confounding in Panel Data Analysis

Yicheng Shen, Yanjiao Yang, Huiying Lin

April 2023

Panel Data & Unmeasured Confounder



Example inspired by Angrist, J. D., & Pischke, J. S. (2009)

Fixed (FE) & Random Effects (RE) Models

$$E(Y_{it}|A_i, X_{it}, t, D_{it}) = \alpha + \lambda_t + \rho D_{it} + A_i'\gamma + X_{it}\delta, \tag{5.1.2}$$

$$Y_{it} = \alpha_i + \lambda_t + \rho D_{it} + X_{it}\delta + \varepsilon_{it}. \tag{5.1.3}$$

where

$$\alpha_i \equiv \alpha + A_i'\gamma.$$

(Cautionary) Proponents:

With satisfied assumptions, fixed effects can absorb unmeasured confounding and provide consistent estimates.

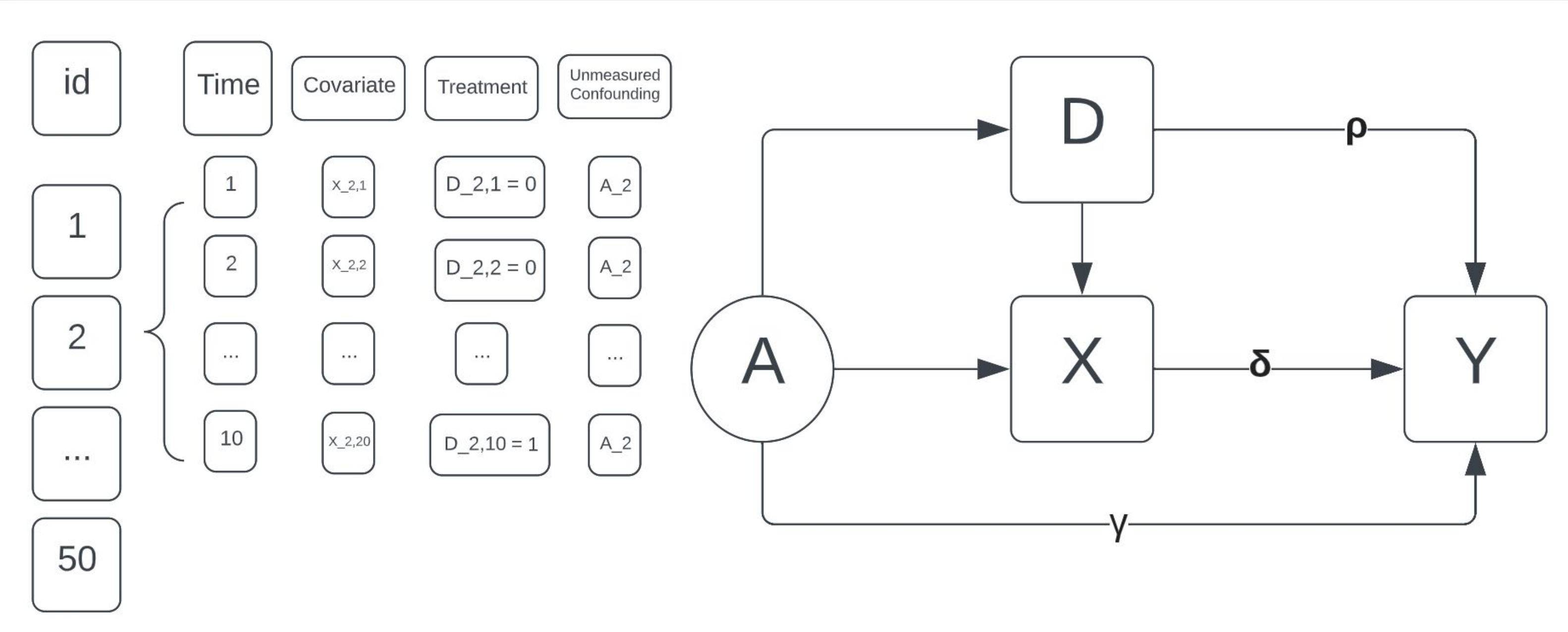
Hausman and Taylor (1981); Angrist and Pischke (2009); Wooldridge (2010), etc.

Opponents:

Assumptions are rarely met, both fixed and random effects would be biased due to interactions.

Mundlak (1978); Hazlett and Wainstein (2022); Bell and Jones (2015), etc.

Data Structure & Causal DAG



Simulation Settings (DGP)

$$Y_{it} = \alpha + \lambda_t + \delta \mathrm{X}_{it} + \rho \mathrm{D}_{it} + \gamma A_i + \phi \mathrm{D}_{it} A_i + \beta \lambda_t A_i + \epsilon_{it}$$

The outcome variable, Y_{it}, is continuous and generated from the linear model including:

- The true coefficient for covariate, δ, is chosen arbitrarily as 1 or 5.
- The true treatment/causal effect, i.e. the coefficient of treatment status, ρ, could vary from to 1 to 20.
- The true coefficient of unmeasured confounding, γ, could vary from to 0 to 20, with zero meaning that there is no direct effect of unmeasured confounding on the outcome.
- The true coefficient of the interaction, φ, between the treatment and the unmeasured confounding could vary from to 0 to 10, with zero meaning that there is no interaction of treatment effect.
- The true coefficient of the interaction, β, between the time and the unmeasured confounding could vary from to 0 to 10, with zero meaning that the effect of A on Y is time-invariant.
- The noise term, ε_{it}, is set to follow $\varepsilon_{it} \sim N(0,1)$.

Simulation Settings (Estimators)

- Difference-in-difference (DID) estimator:

$$\tau^{DID} = (\bar{Y}_{1,t+1} - \bar{Y}_{1,t}) - (\bar{Y}_{0,t+1} - \bar{Y}_{0,t}) = \hat{\tau}_1^{BA} - \hat{\tau}_0^{BA}$$

- OLS model:

$$Y_{it} = \alpha + \lambda_t + \rho \mathrm{D}_{it} + \delta \mathrm{X}_{it} + \epsilon_{it}$$

- Fixed effects (FE) model:

$$\alpha_i \equiv \alpha + \gamma A_i$$

$$Y_{it} = \alpha_i + \lambda_t + \rho \mathrm{D}_{it} + \delta \mathrm{X}_{it} + \epsilon_{it}$$

- Random effects (RE/RI) model:

$$Y_{it} = \alpha_i + \lambda_t + \rho \mathrm{D}_{it} + \delta \mathrm{X}_{it} + \epsilon_{it}$$

$$\text{where } \alpha_i | D, X \overset{i.i.d}{\sim} N(\mu, \sigma^2)$$

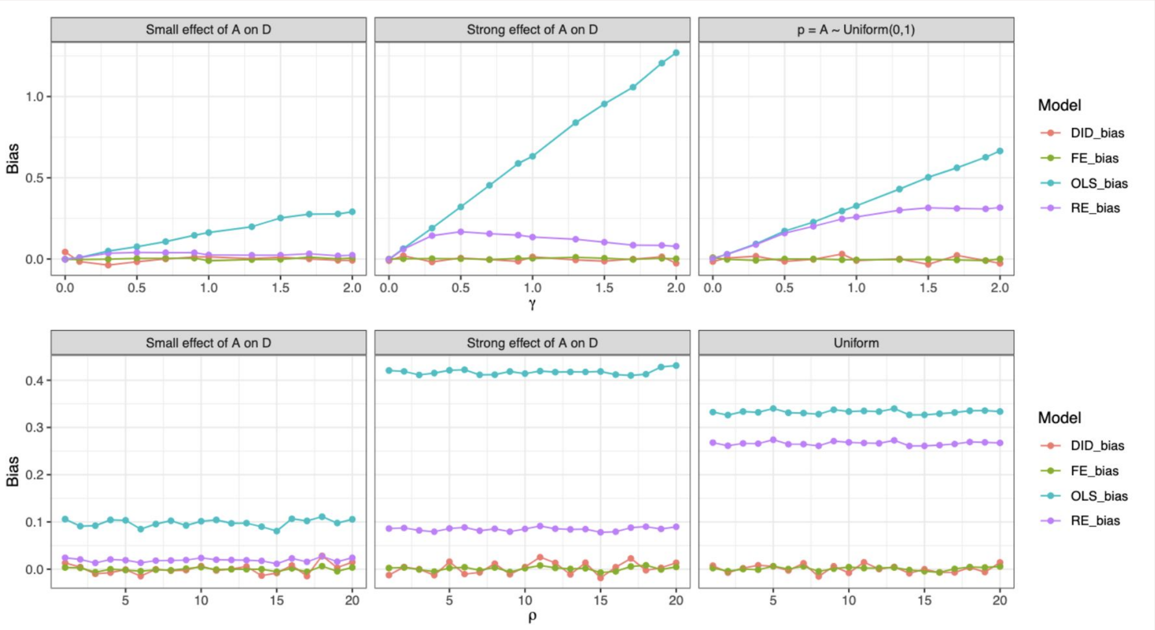
- Bias-corrected version of RE model:

$$Y_{it} = \alpha_i + \lambda_t + \rho \mathrm{D}_{it} + \rho' \bar{\mathrm{D}}_i + \delta \mathrm{X}_{it} + \epsilon_{it}$$

$$\text{where } \alpha_i | D, X \overset{i.i.d}{\sim} N(\mu, \sigma^2)$$

Results

No interaction and no time-varying unmeasured confounders

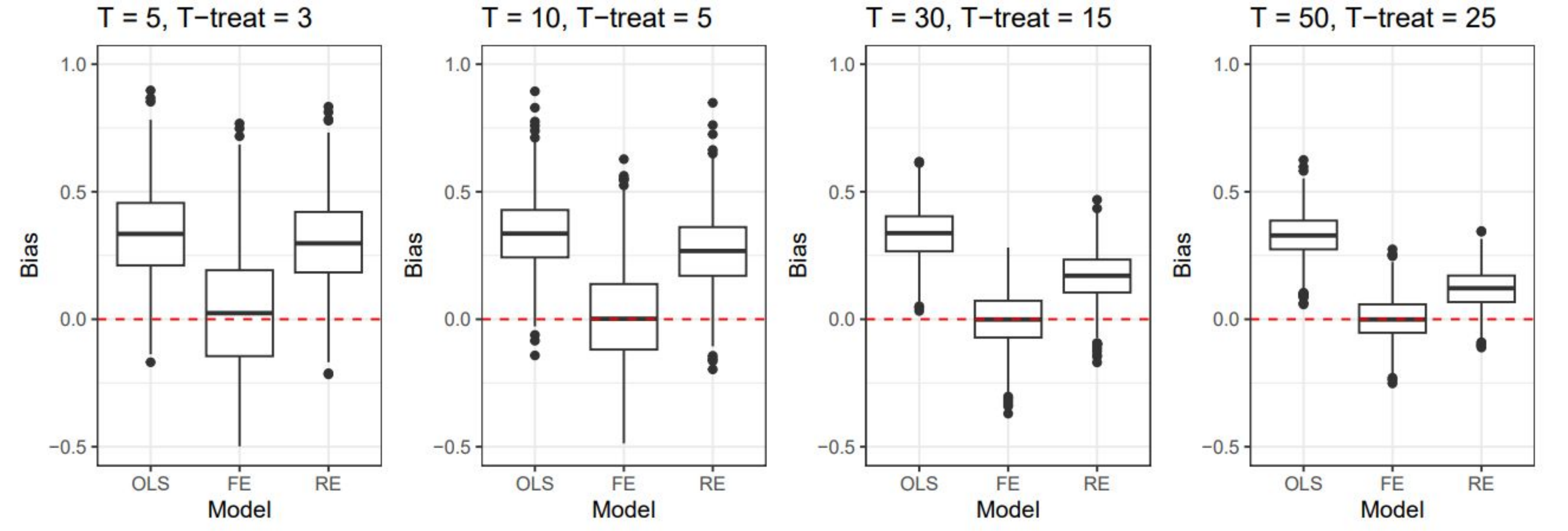


Average bias of estimated treatment effects from DID, OLS, FE, RE under N = 100, T = 10 and T-treat = 5

DID bias ≈ FE bias < RE bias < OLS bias

Results

No interaction and no time-varying unmeasured confounders

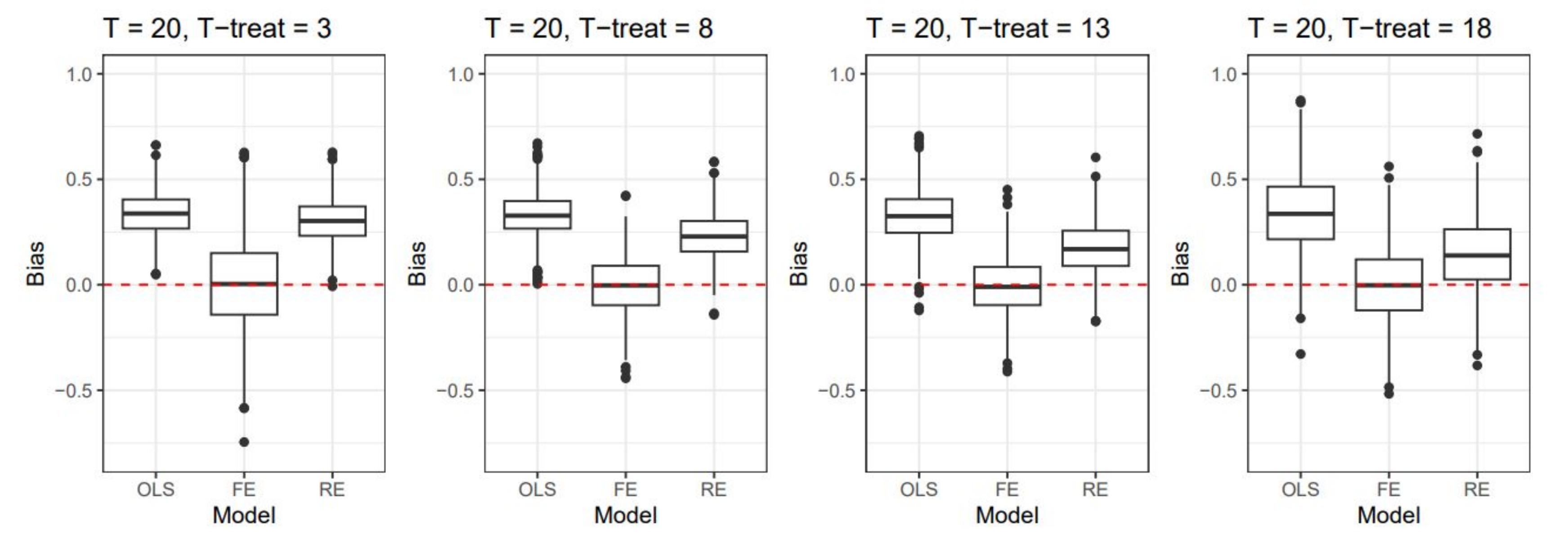


Distribution of bias across varying total time points

T ↗, RE bias ↘

Results

No interaction and no time-varying unmeasured confounders

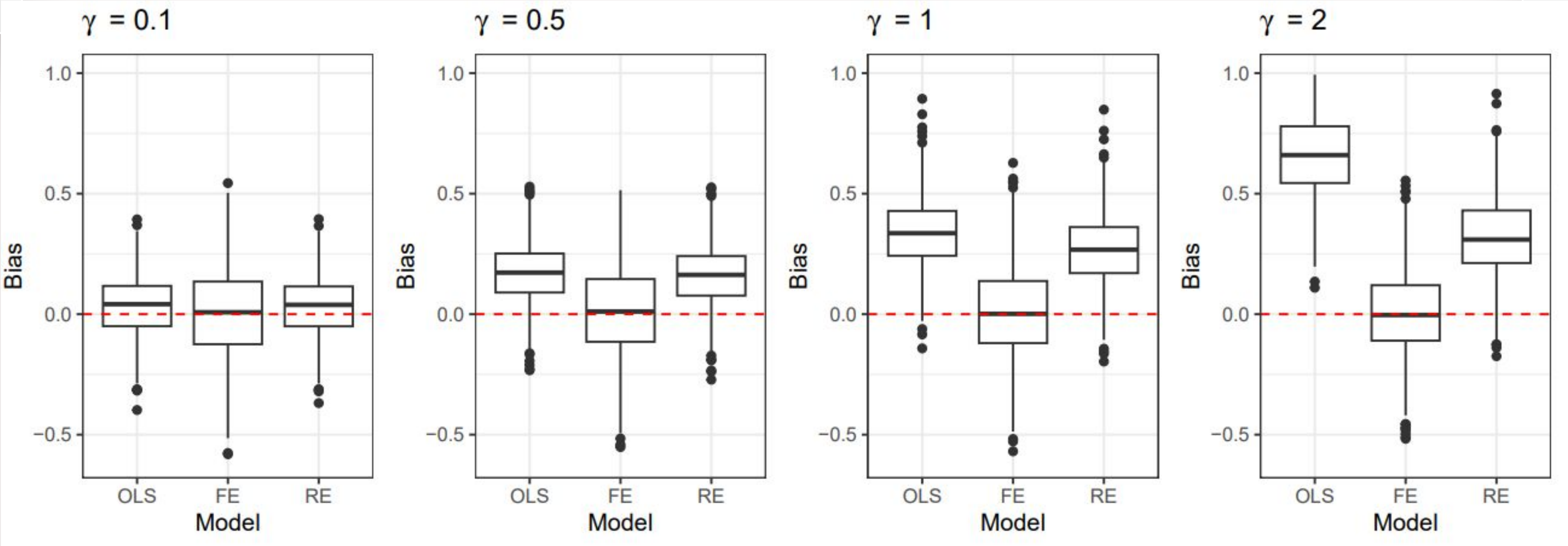


Distribution of bias across varying treatment time points

T−treat ↗, RE bias ↘

Results

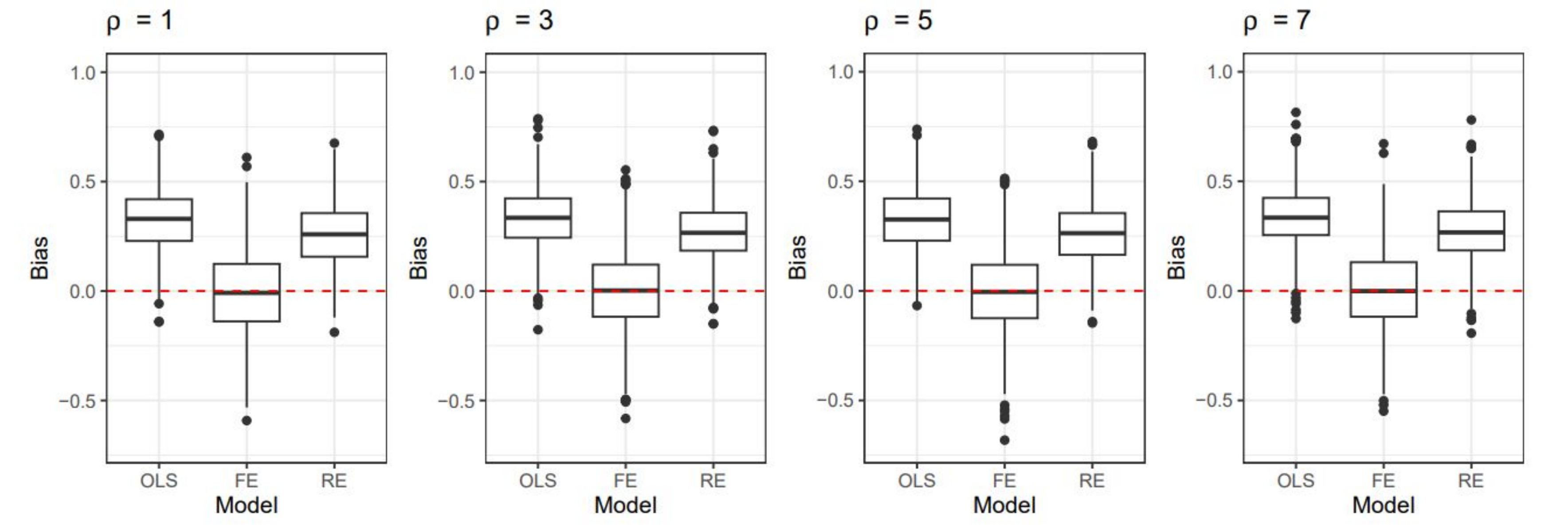
No interaction and no time-varying unmeasured confounders



Distribution of bias across varying unmeasured confounding coefficients
γ ↗, OLS bias ↗, RE bias ↗

Results

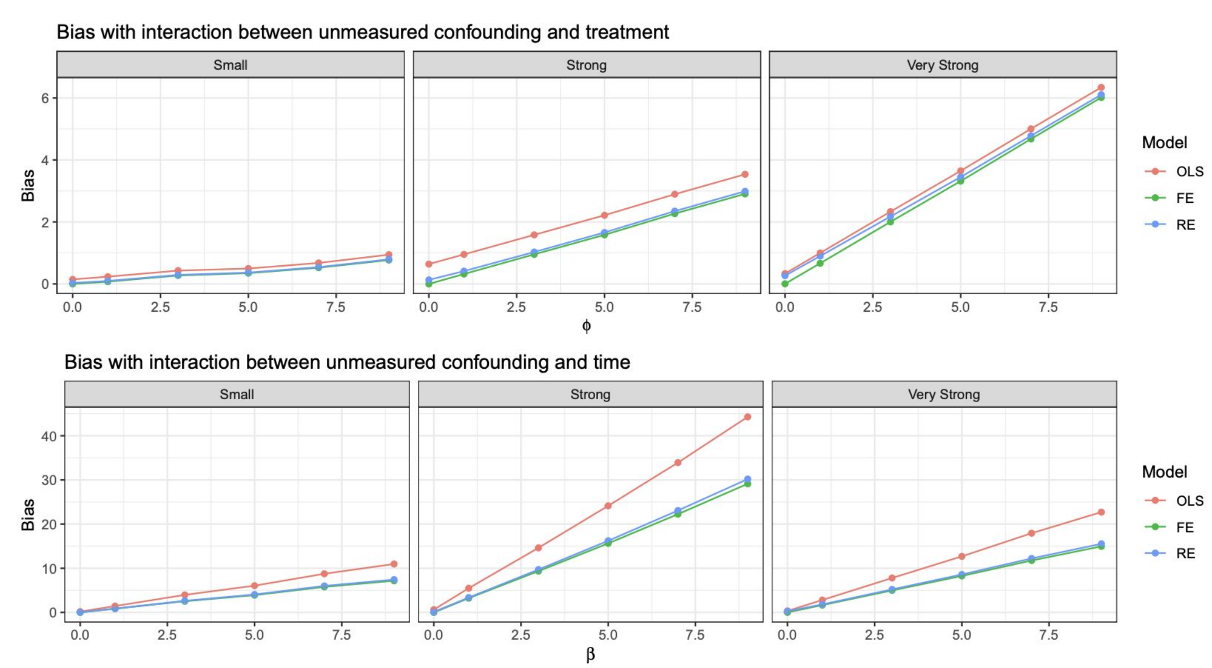
No interaction and no time-varying unmeasured confounders



Distribution of bias across varying true causal effects
ρ ↗, bias doesn't change

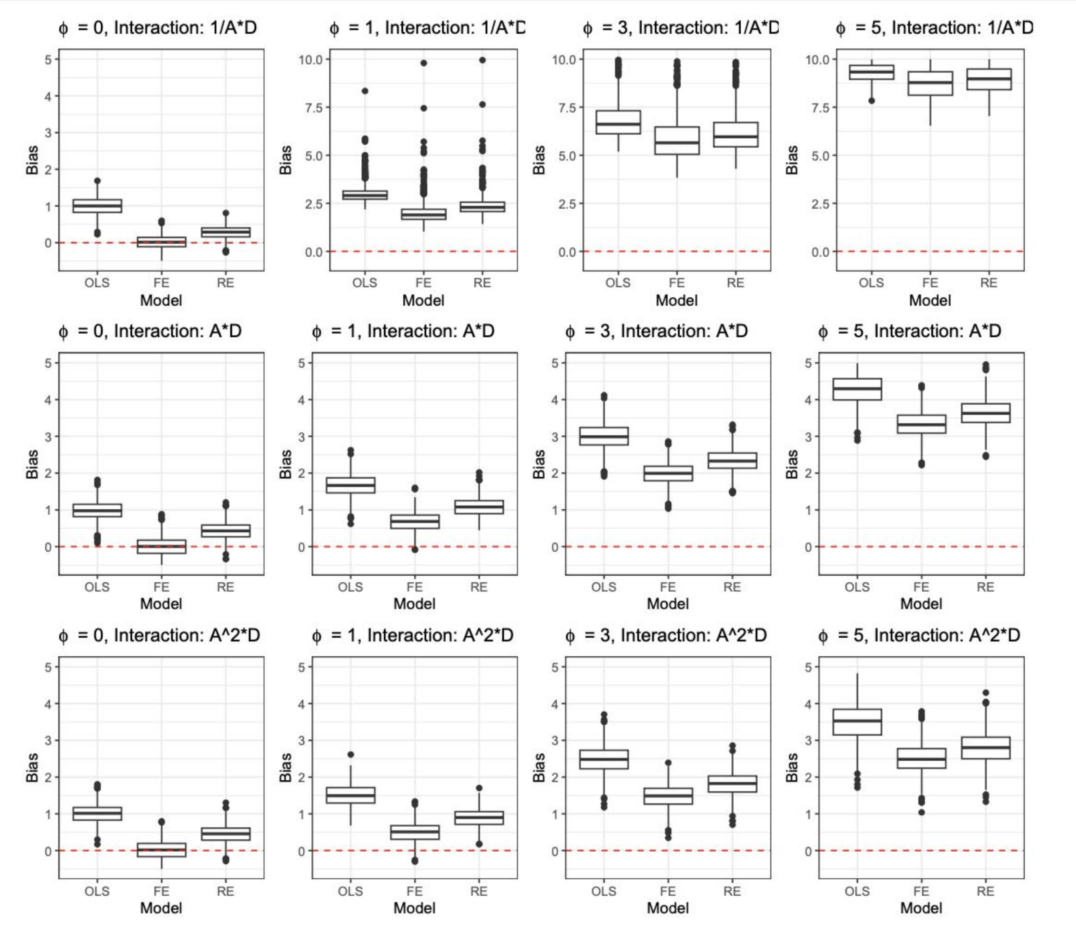
Results

With interaction and time-varying unmeasured confounders



- Even fixed effects models are not able to estimate the treatment effects consistently.

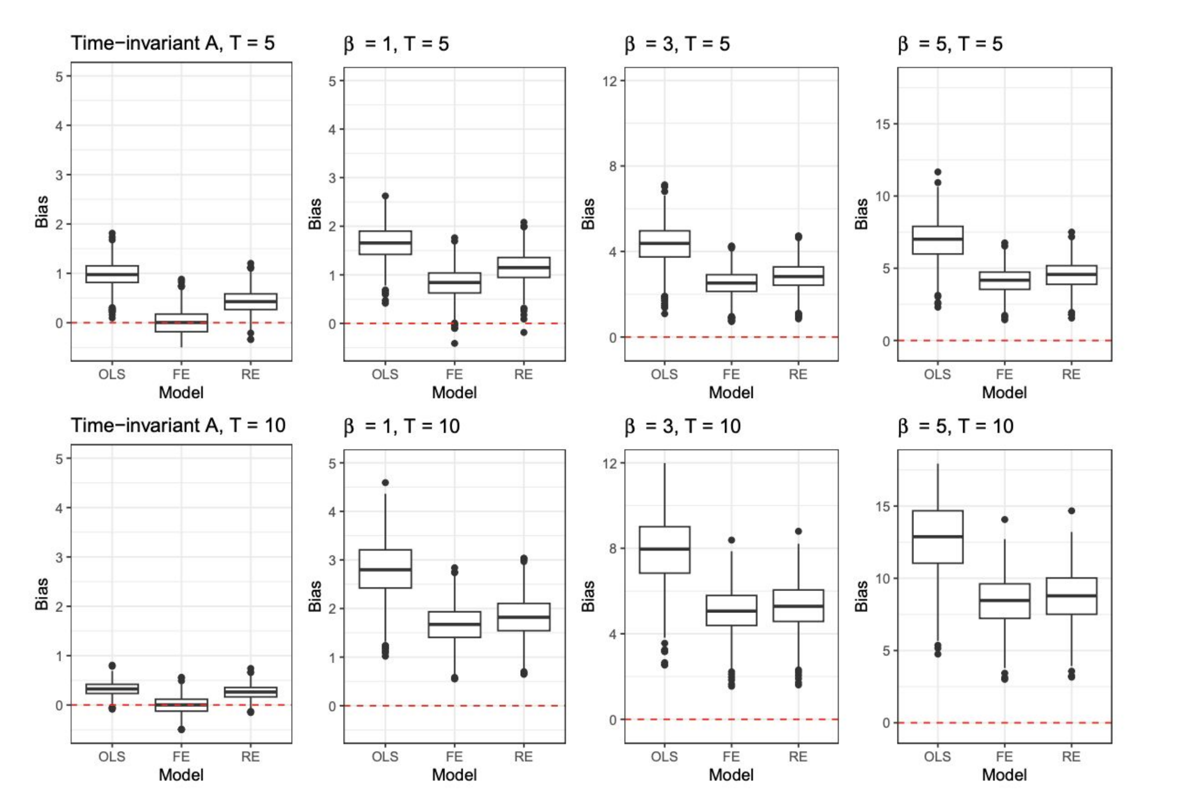
Results



- Φ ↗ (same interaction), bias↗
- FE bias < RE bias < OLS bias
- Fixed effects can no longer absorb the unmeasured confounders under large Φ

$$\alpha_i = \alpha + A_i(\gamma + \phi D_{it})$$

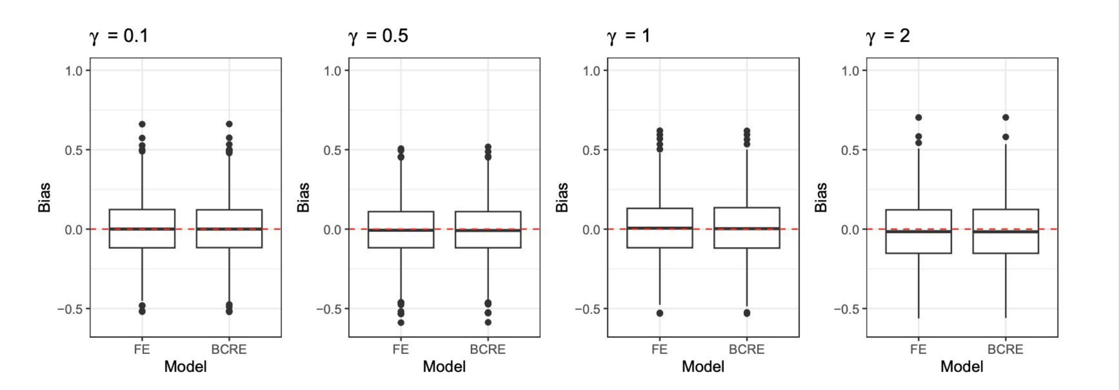
Results

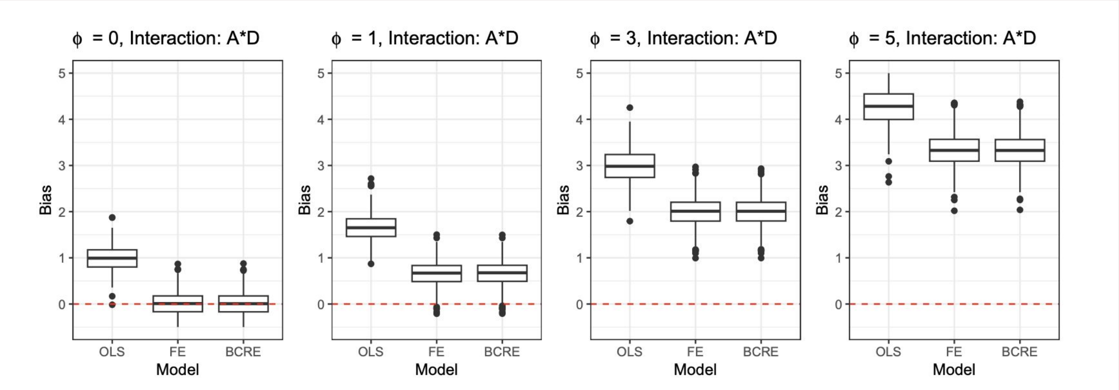


- β ↗ (same T), bias↗
- FE bias < RE bias < OLS bias
- Fixed effects can no longer absorb the unmeasured confounders under large β

$$\alpha_i = \alpha + A_i(\gamma + \beta\lambda_t)$$

Results Bias-corrected random effects model





- FE and bias-corrected RE models both give unbiased estimate under different values of γ.
- Φ ↗ , bias↗
- FE bias ≈ BCRE bias < OLS bias

Discussions

- Closed form of asymptotic bias
  Y. Li et al. (2020) : illustrates the closed form of bias in the presence of unmeasured within- and/or between-cluster confounders when the treatment is **continuous**

$$T_{ij} = a_{0i} + C'_{ij}\alpha_c + W'_{ij}\alpha_w + B'_i\alpha_b + \epsilon^t_{ij}$$

$$Y_{ij} = b_{0i} + \beta T_{ij} + C'_{ij}\beta_c + W'_{ij}\beta_w + B'_i\beta_b + \epsilon^y_{ij}$$

Under time-invariant unmeasured confounder
Expression of asymptotic bias for RE:

$$\frac{\alpha'_b V_b \beta_b}{(\sigma^2_a + \alpha'_b V_b \alpha_b) + \sigma^2_{\epsilon t}\frac{\sigma^2_{\chi e} + (n-1)\sigma^2_{de}}{\sigma^2_{\chi e}}}$$

Future work: the closed form of bias under binary treatment and will be helpful for obtaining insights of asymptotic behavior of the bias

Discussions

- Fixed/Random effects models do not estimate γA_i , other models such as instrumental variable (IV) can probably address the inadequacy

- Binary outcome: similar behavior to the simulation.
  Future work: conduct more simulations for binary confounders

- The time-varying confounders in this paper is set to change linearly with time

$$A_{it} = \lambda_t A_i$$

  Future work can specify more sophisticated time-varying confounders

- The coefficient of treatment:
  Future work can specify more complicated treatment effect which may change with time.

- Interactions between confounders and treatment: more variations besides

$$\phi A_i D_{it}, \phi A_i^2 D_{it}, \phi \frac{1}{A_i} D_{it}$$

Bibliography

Ahn, Seung C, Young H Lee, and Peter Schmidt. 2013. “Panel Data Models with Multiple Time-Varying Individual Effects.” Journal of Econometrics 174 (1): 1–14.

Angrist, Joshua D, and Jörn-Steffen Pischke. 2009. Mostly Harmless Econometrics: An Empiricist’s Companion, Chapter 5. Princeton university press.

Bell, Andrew, and Kelvyn Jones. 2015. “Explaining Fixed Effects: Random Effects Modeling of Time-Series Cross-Sectional and Panel Data.” Political Science Research and Methods 3 (1): 133–53.

Gunasekara, Fiona Imlach, Ken Richardson, Kristie Carter, and Tony Blakely. 2014. “Fixed Effects Analysis of Repeated Measures Data.” International Journal of Epidemiology 43 (1): 264–69.

Hausman, Jerry A, and William E Taylor. 1981. “Panel Data and Unobservable Individual Effects.” Econometrica: Journal of the Econometric Society, 1377–98.

Hazlett, Chad, and Leonard Wainstein. 2022. “Understanding, Choosing, and Unifying Multilevel and Fixed Effect Approaches.” Political Analysis 30 (1): 46–65.

Imbens, Guido, Nathan Kallus, and Xiaojie Mao. 2021. “Controlling for Unmeasured Confounding in Panel Data Using Minimal Bridge Functions: From Two-Way Fixed Effects to Factor Models.” arXiv Preprint arXiv:2108.03849.

Li, Degui, Jia Chen, and Jiti Gao. 2011. “Non-Parametric Time-Varying Coefficient Panel Data Models with Fixed Effects.” The Econometrics Journal 14 (3): 387–408.

Li, Yun, Yoonseok Lee, Friedrich K Port, and Bruce M Robinson. 2020. “The Impact of Unmeasured Within-and Between-Cluster Confounding on the Bias of Effect Estimatorsof a Continuous Exposure.” Statistical Methods in Medical Research 29 (8): 2119–39.

Mundlak, Yair. 1978. “On the Pooling of Time Series and Cross Section Data.” Econometrica: Journal of the Econometric Society, 69–85.

VanderWeele, Tyler J, and Onyebuchi A Arah. 2011. “Unmeasured Confounding for General Outcomes, Treatments, and Confounders: Bias Formulas for Sensitivity Analysis.” Epidemiology (Cambridge, Mass.) 22 (1): 42.

Wooldridge, Jeffrey M. 2010. Econometric Analysis of Cross Section and Panel Data. MIT press.

Thanks