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

Yicheng Shen, Yanjiao Yang, Huiyin Lin

25 April, 2023

1 Introduction

Panel data analysis is a widely used statistical tool in social sciences, econometrics and ecology to study the relationships between variables over time. However, this type of analysis is susceptible to confounding variables, which can bias the estimated coefficients and distort the interpretation of the results. Fixed and random effects models frequently emerge in panel data analysis to account for confounding variables that are time-invariant (Gunasekara et al. 2014) or time-varying (Li, Chen, and Gao 2011; Ahn, Lee, and Schmidt 2013). A common belief among econometricians is that fixed effects (FE) or random effects (RE) models can absorb unmeasured confounding variables (Angrist and Pischke 2009), but the mechanism behind this claim is mysterious and not well-understood. In this research paper, we aim to explore the relationship between fixed/random effects and time-invariant unmeasured confounding in panel data analysis and provide insights into whether and in what sense these models can address this issue.

Our simulation suggests that fixed/random effects can remedy the issue of unmeasured confounding to a certain extent, but there are still systematic bias resulting from the relationship between unmeasured confounding and the treatment assignment. [**more stuff here**]

2 Background

In typical observational studies, failing to capture significant unmeasured confounding gives rise to biased estimates of treatment effects, which compels practitioners to develop methods of assessing and handling uncontrolled confounding (VanderWeele and Arah 2011). The question of whether fixed/random effects models can account for unmeasured confounding in panel data analysis has been the subject of much debate particuarly in the econometrics literature. A number of studies have explored this issue from different angles and with varying degrees of empirical evidence.

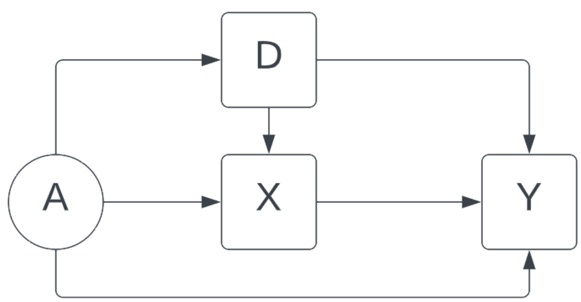
One line of research has focused on theoretical arguments for why fixed/random effects models might be effective at absorbing unmeasured confounding. For example, Angrist and Pischke (2009) discuss the strategies that use data with a time or cohort dimension to control for unobserved-but-fixed omitted variables. Hausman and Taylor (1981) argue that fixed effects models can control for the time-invariant confounders by essentially differencing them out, while random effects models can account for time-varying confounders that are uncorrelated with the fixed effects. More recently, Wooldridge (2010) has suggested that fixed effects models can be viewed as a form of quasi-experimental design that mimics a randomized controlled trial, and thus can address the unobserved component to the extent that such designs do.

Other scholars have challenged the notion that fixed/random effects models can fully absorbing unmeasured confounding. For example, Mundlak (1978) once argued that random effects models are biased when unobserved heterogeneity is correlated with observed variables, and that fixed effects models are limited by the fact that they cannot estimate time-invariant covariates. Hazlett and Wainstein (2022) point out that random effect estimates in multilevel models are equivalent to fixed effects estimates that have been shrunken through a regularization process. When the source of unmeasured confounding is at the group-level, the FE approach could unbiasedly estimates treatment effects, but with poor estimates of standard errors. Bias takes place in random effects models because their variables are not "allowed" to adjust for confounding as intended and thus fails to remove the unmeasured confounding. Furthermore, Bell and Jones (2015) note that while fixed effects models can provide reasonable estimates of treatment effects, they may still suffer from omitted variable bias if the unobserved confounding variable is correlated with the time-varying variables.

Therefore, the effectiveness of fixed/random effects models in accounting for unmeasured confounding and obtaining unbiased estimates of treatment effects in panel data analysis remains a topic of ongoing research and debate. As such, our projects seeks to explore the trends of potential bias when applying fixed/random effects models in the face of unmeasured confounding.

3 Method

We presume that specifying an intercept for every individual in the study is an appropriate method to address unmeasured confounding, A.



Generate data

3.1 Continuous Case

Our data generation process (DGP) uses the following model specifications:

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

In our DGP of artificial panel data, there are 100 individuals and 10 time points. Covariate, X, for each individual is simulated from $N(0, 1)$. The unmeasured counfounding, A, is either from $N(0, 1)$ or Uniform(0,1).

We designed two ways that the unmeasured confounding, A, could affect the treatment assignment: 1. As $A \sim N(0, 1)$ is above or below a certain cut-off point, the probability of receiving the treatment changes to a respectively value. 2. $A \sim \mathrm{Uniform}(0, 1)$ directly serves as the probability of receiving the treatment.

The outcome variable, $Y_i t$, is continuous and generated from a linear model including:

- The true coefficient for covariate, γ, is chosen arbitrarily as 5.
- The true treatment effect, i.e. the coefficient of treatment status, ρ, could var from to [ ] to [ ].
- The true coefficient of unmeasured confounding, γ, could vary from to [ ] to [ ], with zero meaning that there is no direct effect of unmeasured confounding on the outcome.
- The true coefficient of the interaction between the treatment and the unmeasured confounding, γ, could var from to [ ] to [ ], with zero meaning that there is no interaction effect.
- The noise term, ϵ, which follows $N(0,1)$.

Estimator

We have several types of methods below that provide estimations of the treatment effects.

Difference-in-difference (DID) estimator

The Difference-in-Differences (DID) method is a quasi-experimental approach used to estimate the causal effect of a treatment in panel data.

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

OLS model

very likely to be heavily biased

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

Fixed effects model

Probably good performance, but would be biased if there is interaction effect missing

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

Random effects model

Our random effects model is essentially a random intercept (RI) model. The argument that RI could account for unmeasured confounding is theoretically problematic due to its modeling specification. In particular, bias emerges when the random effects are correlated with the treatment. Because the group-specific intercepts in a RI model are regularized, they do not achieve the values that would "fully absorb" group-specific confounding, leaving components unexplained that can instead be captured by FE (Hazlett and Wainstein 2022).

We use `lmer` from the `lme4` package to fit the random intercept model.

$$Y_{it} = \alpha_i + \lambda_t + \rho \mathrm{D}_{it} + \delta \mathrm{X}_{it} + \epsilon_{it}$$
$$\text{where } \alpha_i | D, X \overset{i.i.d}{\sim} N(0, \sigma^2)$$

3.2 Binary Case

fit `glm`

RE: `glmer` from the `lme4`

4 Result

In Figure 1, we show that

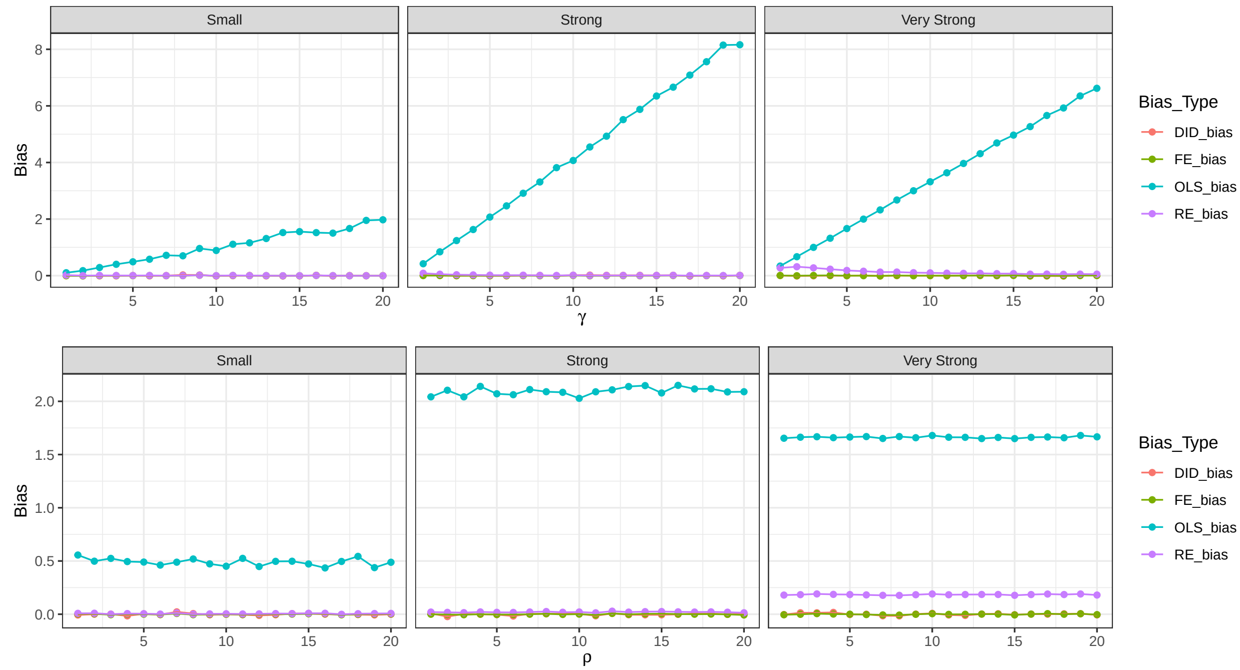


Figure 1: Average bias of estimated treatment effects from four types of estimation methods.

In Figure 2, we show that

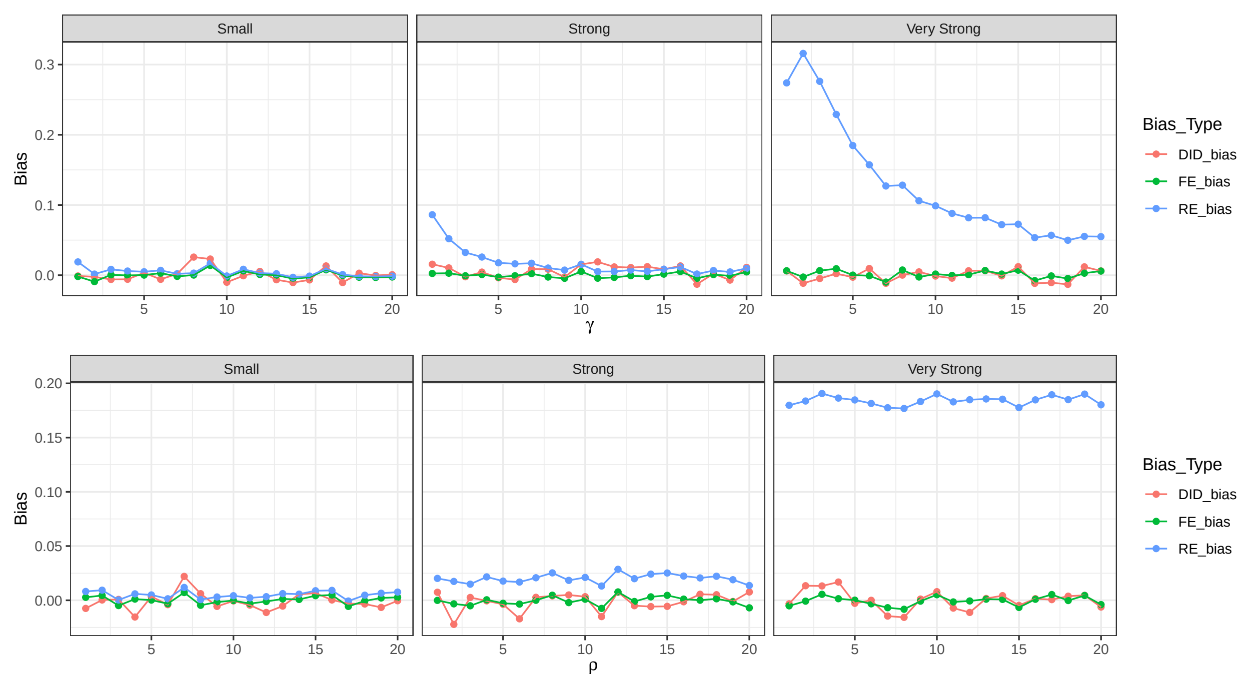


Figure 2: Compareing average bias of estimated treatment effects from DID, FE, and RE models.

5 Discussion

Limitation and further direction:

Time-variant confounding?

More time points?

Estimated coefficients of latent variable?

Overall, the literature suggests that while fixed/random effects models may be useful in controlling for unmeasured confounding in panel data analysis, they are not a panacea. Other methods, such as instrumental variables or regression discontinuity designs, may be more necessary in certain cases to fully address this issue.

6 Reference

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.

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.

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.