Title: Which Quasi-Experimental Estimator to Use Under What Conditions: Bias and Precision of ANCOVA, Difference-in-Difference, and Comparative Short Interrupted Time Series Estimators

Authors: Fatih Unlu, RAND Corporation (presenting author) and Cristopher Price, Abt Associates

Background/Context: Quasi-experimental (QE) designs typically rely on at least one and possibly multiple baseline or pre-treatment observations of the outcome measures (also called lagged outcome measures) to account for the non-random self-selection or assignment of units to the treatment and comparison conditions. For example, when estimating the impact of a program that started at time *t* on an outcome *m*, researchers seek to collect values of *m* for the treatment and comparison units at time *t-1*, *t-2*, and so on.
Researchers use two general types of QE estimators that differ by how they utilize the baseline measure(s). The first one is the difference-in-differences (DID) estimator in which pre- and post-treatment changes in the outcome levels of the treatment group is compared with that for the comparison group and the difference between the two is attributed to the treatment. When the pre-treatment data is available for multiple baseline time points, Comparative Short Interrupted Time Series (C-SITS) estimators that estimate overall or group-specific trends in the baseline period and extrapolate them into the posttreatment period could be implemented as an extension of the DID estimator. The second type of estimator entails controlling for the single or multiple baseline measures as independent variables along with other observable pretreatment covariates in multivariate regression models, which is also called the ANCOVA estimator.
DID/C-SITS and ANCOVA estimators utilize the baseline data in different ways and rely on different assumptions which have important implications for the bias and precision of the resulting estimates. The identifying assumption of the DID/C-SITS approach is that the treatment assignment is primarily related to time invariant attributes of program participants which are typically controlled for via fixed effects for the treatment group (or fixed effects for individuals) while the ANCOVA approach relies on a fundamentally different assumption that pertains to the ignorability of treatment assignment conditional on the lagged outcome measure(s) and other observables. Angrist and Pischke (2009) argue that using DID (ANCOVA) leads to over (under) estimates of positive treatment effects if in fact the assumption for the other approach holds. There are different views regarding the precision of the two estimators. Fitzmaurice, Laird, and Ware (2011) claim that the ANCOVA estimator tend to have more statistical power while Oakes and Feldman (2001) suggest conditions under which the DID estimator is more precise.

Purpose/Objective/Research Question: This paper examines the conditions under which the DID/C-SITS estimator is preferred over ANCOVA in terms of bias and precision. We also examine the extent to which combining each type of estimator with a matching process that

entails matching treatment individuals with the potential comparison individuals changes its bias and precision.

Research Design and Analysis: We conduct a series of simulation analyses to examine the bias and precision of DID/C-SITS and ANCOVA estimators. The simulations employ five data generation processes (DGPs) that primarily differ by the mechanism underlying the self-selection of individuals to the treatment and comparison conditions. Specifically, we consider the following scenarios under which treatment assignment is related to:

1. an fixed individual characteristic (e.g., more motivated students take up a college preparation intervention)
2. a shock experienced in the last baseline period (e.g., teachers who experienced health issues in the prior year and had lower attendance get more coaching in the following year)
3. the last lagged outcome measure (e.g., students selected for a reading intervention based on prior spring reading scores)
4. average of lagged outcome measures across multiple years (e.g., schools undergo a leadership change due to chronically low test scores)
5. the pretreatment time trend (e.g., schools with downward trends in test scores get new principals)

All simulated datasets include three pre-treatment periods and three post-treatment periods. Using these data, we implement two versions of the ANCOVA estimator and three versions of the DID/C-SITS estimator:

1. ANCOVA that controls for the last (closest to treatment) lagged outcome measure
2. ANCOVA that controls for all pre-treatment measures
3. DID that uses the last lagged and the first post-treatment outcome measure
4. C-SITS that uses all data to estimate a common linear trend for the treatment and comparison units
5. C-SITS that uses all data to estimate group-specific linear trends

We use each estimator with three different matching processes. The first one does not entail any matching, the second one matches treatment and comparison cases on the last pretest, and the third one uses the average of all pretest measures in the matching process.

Results: The overarching results of the simulation analysis is as follows:

- Thinking hard about the selection mechanism and using the an estimator that matches that mechanism (to the extent possible) is essential for minimizing bias. Specifically, when treatment assignment is related to:
    - **a fixed (time invariant) individual attribute:** All ANCOVA estimators (with or without explicit matching on pretest measures) yielded biased estimates. DID/C-SITS estimators with no matching produced unbiased estimates but combining them with matching introduced bias.

    - **a shock in the last baseline period:** All estimators yielded biased estimates. Only the estimator that directly compared the outcomes of treatment and comparison units yielded an unbiased effect.
    - **the last pretest:** DID/C-SITS estimators without matching yielded biased estimates. Combining them with matching reduced bias. The ANCOVA estimators (with or without matching) yielded unbiased estimates.
    - **the average pretest:** The ANCOVA estimator that controls for all lagged outcomes measures yielded unbiased estimates. Combining the other estimators with matching on the average pretest reduced bias substantially.
    - **a group-specific time trend:** Only the C-SITS model that explicitly controls for group-specific trends yielded unbiased effects.
- The C-SITS estimator that used all available data yielded the most precise estimates in the scenarios we considered. The ANCOVA estimator that used all data was the second best in terms of precision.
- Using additional pretreatment data when available in matching and/or estimation is generally helpful with bias and precision.
    - Examining pretest measures for trends is helpful especially if treatment selection may be based on the trend.

References

Angrist, J. D., & Pischke, J.-S. (2009). Mostly harmless econometrics: An empiricist's companion. Princeton: Princeton University Press.

Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2011). Applied longitudinal analysis (2nd ed.). Hoboken, N.J.: Wiley.

Oakes J.M., Feldman H.A. Statistical power for nonequivalent pretest-posttest designs the impact of change-score versus ancova models. Eval. Rev. 2001;25(1):3–28.