Title: Synthetic Control and Weighted Event Study Models with Staggered Adoption

Authors:

Eli Ben-Michael, UC Berkeley, ebenmichael@berkeley.edu

Avi Feller, UC Berkeley, afeller@berkeley.edu
[presenting author]

Jesse Rothstein, UC Berkeley, rothstein@berkeley.edu

Problem:

Many important interventions and policy changes in education occur at an aggregate level, such as the level of the school, district, or state. In settings where randomization is infeasible or unethical, education researchers often turn to quasi-experimental research designs based on repeated observations of aggregate data. For example, a researcher might estimate the impact of a new reading program using school-level average test scores at multiple time points surrounding the introduction of the intervention (Jacob, Somers, Zhu, & Bloom, 2016). In our motivating example, at various times between 1964 and 1987, 33 states passed laws mandating that school districts bargain with teachers unions; the question is whether this increased teacher salaries and school spending.

Estimating effects in this *staggered policy adoption* setting, in which multiple units adopt treatment over time, remains a major statistical challenge. The workhorse model is the *event study* fixed effects estimator, a generalization of the difference-in-differences (DID) approach. Despite its popularity, this approach can often give misleading estimates (Goodman-Bacon, 2018), and requires a "parallel trends" assumption that can be violated in practice.

A recent alternative is the *synthetic control method* (Abadie, Diamond, & Hainmueller, 2010). Developed for settings with a single treated unit, SCM estimates the counterfactual untreated outcome via a weighted average of untreated units that seeks to match the treated unit's pre-treatment outcomes. This is attractive because it is often more plausible that some untreated units closely resemble the treated unit than that all units are on parallel trends, as assumed for DID. But there has been no natural way to apply SCM to the staggered adoption case, despite recent *ad hoc* approaches (Dube and Zipperer, 2015).

Prior methodological research:

There is an extensive methodological literature on estimating causal effects with panel data and adoption over time. We briefly highlight three relevant threads:

- **Outcome and time series modeling.** The most common estimation approach is to model the outcome (or time series) process directly. The workhorse method here is the Comparative Interrupted Time Series approach and the closely related fixed effects approach; see, for example, Hallberg, Williams, Swanlund, & Eno (2018). More recent general approaches include Xu (2017), who directly estimates a linear factor model and Athey, Bayati, Doudchenko, Imbens, & Khosravi (2017), who instead use "matrix completion" methods from the machine learning literature.
- **Weighting approaches.** Examples of weighting methods include weighting via multiple synthetic control weights (Arindrajit Dube, 2015) and inverse propensity score weighting (Callaway & Sant'Anna, 2018).
- **Combined estimators.** Finally, several recent papers propose to combine outcome and weighting methods for panel data, typically in settings with a single treatment time. Examples the Augmented SCM method of Ben-Michael, Feller, & Rothstein (2018) and

the Synthetic Difference-in-Differences approach of Arkhangelsky, Athey, Hirshberg, Imbens, & Wager (2018).

Method:
Our paper generalizes SCM to the staggered adoption setting. We show that the existing strategy of averaging separate SCM estimates for each treated unit amounts to solving a single joint optimization problem with a particular combination of unit-level balance criteria. The resulting weights, however, are not optimized for the Average Treatment Effect on the Treated (ATT). We pose a modified optimization problem, which we call *pooled SCM*, that generalizes SCM for this estimand. We then introduce a tuning parameter that allows the researcher to select the relative weight to put on balancing for the ATT versus balancing for individual units. We advocate for SCM with intermediate weights, which we call *hybrid SCM*, and argue that this approach inherits attractive properties from both the separate and pooled SCM approaches. We provide some interpretation of these estimators by showing that separate SCM weights are a form of inverse generalized propensity score weighting. Extending this, the hybrid and pooled SCM approaches correspond to partial and complete pooling, respectively, on the propensity score coefficients.

We then propose to combine SCM and event study outcome modeling, analogous to bias correction in matching or augmented IPW. This naturally extends the augmented SCM approach in Ben-Michael et al. (2018), which focuses on a single treated unit, to the staggered adoption setting. Specifically, the augmented estimator uses SCM to re-weight the residuals from a simple fixed effects model, as in Doudchenko & Imbens (2016) and Ferman & Pinto (2018) for the single treated unit case, further reducing covariate imbalance that remains even with hybrid SCM weights. This combined approach similarly reduces the event study model's reliance on strong homogeneity assumptions.

Setting:

We apply our approach to better understand the impact of mandatory teacher collective bargaining. Opponents have long argued that teachers unions are a key cause of poor educational performance in the United States, but the evidence regarding their first-order effects remains unresolved. The substantive question is timely in light of recent political and legal challenges to public sector unions. Methodologically, there is disagreement about whether the parallel trends assumptions needed to support difference-in-differences and event study analyses hold in this setting. In an important recent paper, Paglayan (2018) tries to adjudicate these disagreements by collecting rich historical data, which enables a more careful assessment of the underlying assumptions. Consistent with her results, we show that Paglayan (2018)'s newly available data do not support the assumptions needed for either event studies or separate SCM using all available states. We apply our proposed SCM estimators and weighted event study approach and show that they achieve better pre-treatment balance than either separate SCM or an event study estimator alone. We find no impact of teacher collective bargaining laws on either teacher salaries or student expenditures, similar to several recent papers (Frandsen, 2015) but counter to earlier claims (most notably Hoxby, 1996).

References

Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. *Journal of the American Statistical Association*, *105*(490), 493–505.

Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W., & Wager, S. (2018). Synthetic Difference in Differences. *arXiv*.

Athey, S., Bayati, M., Doudchenko, N., Imbens, G. W., & Khosravi, K. (2017). Matrix Completion Methods for Causal Panel Data Models. *arXiv*.

Ben-Michael, E., Feller, A., & Rothstein, J. (2018). The Augmented Synthetic Control Method. *arXiv*.

Callaway, B., & Sant'Anna, P. H. C. (2018). Difference-in-Differences with Multiple Time Periods and an Application on the Minimum Wage and Employment.

Doudchenko, N., & Imbens, G. W. (2016). Balancing, Regression, Difference-In-Differences and Synthetic Control Methods: A Synthesis. NBER Working Paper.

Dube, A., & Zipperer, B. (2015). Pooling Multiple Case Studies Using Synthetic Controls: An Application to Minimum Wage Policies, Working paper.

Ferman, B., & Pinto, C. (2018). Synthetic Controls with Imperfect Pre-Treatment Fit. Working paper.

Frandsen, B. R. (2015). The Effects of Collective Bargaining Rights on Public Employee Compensation. *ILR Review*, *69*(1), 84–112.

Goodman-Bacon, A. (2018). Difference-in-Differences with Variation in Treatment Timing. NBER Working Paper.

Hallberg, K., Williams, R., Swanlund, A., & Eno, J. (2018). Short Comparative Interrupted Time Series Using Aggregate School-Level Data in Education Research. *Educational Researcher*, *47*(5), 295–306.

Jacob, R., Somers, M.-A., Zhu, P., & Bloom, H. (2016). The Validity of the Comparative Interrupted Time Series Design for Evaluating the Effect of School-Level Interventions. *Evaluation Review*, *40*(3), 167–198.

Paglayan, A. S. (2018). Public-Sector Unions and the Size of Government. *American Journal of Political Science*, *63*(1), 21–36.

Xu, Y. (2017). Generalized synthetic control method: Causal inference with interactive fixed effects models. *Political Analysis*.