JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2021, VOL. 116, NO. 536, 1713–1715: Theory and Methods Special Section on Synthetic Control Methods
https://doi.org/10.1080/01621459.2021.2002600

Introduction to the Special Section on Synthetic Control Methods

Alberto Abadie and Matias D. Cattaneo, Guest Editors

In June of 2010, *JASA* published the article "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program" (Abadie, Diamond, and Hainmueller 2010). This article proposed a way to estimate the effects of aggregate interventions (i.e., interventions that happen at the level of aggregate units, such as cities, states, or countries) using weighted averages of untreated units to approximate the counterfactual outcome that the treated units would have experienced in the absence of the intervention. The method had originally appeared in a study of the effect of terrorism on economic growth in the Basque Country in Spain (Abadie and Gardeazabal 2003). The specific characteristics of the empirical setting in Abadie and Gardeazabal (2003), where the units of analysis were entire regional entities, made alternative methods inappropriate and motivated the development of the synthetic control machinery. But it was not until after the publication of the 2010 *JASA* article that synthetic controls became a mainstay of empirical research in economics, the social sciences, and beyond. Today, synthetic controls are widely applied in the social sciences to study the effects of policy interventions and other treatments or events of interest. They have also been adopted as a research design in other disciplines, including the biomedical disciplines (especially in epidemiology) and engineering. Synthetic controls are also used by government agencies, multilateral organizations, NGOs, and business analytics units, and both the method and its applications have garnered the attention of the popular press and social media (see, e.g., Abadie 2021).

Further spurred by the work in Abadie, Diamond, and Hainmueller (2015), Doudchenko and Imbens (2017), Amjad, Shah, and Shen (2018), Arkhangelsky et al. (2018) and others, a new literature has emerged in statistics, econometrics, and machine learning to extend and improve the synthetic control methodology, to study the properties of synthetic control estimators, and to adapt the method to better handle particular data configurations. The articles in this special section on synthetic control methods cover many of the new research directions of the current methodological literature on synthetic control estimation and inference.

We briefly summarize below the articles that contribute to this *JASA*'s special section on synthetic controls, and describe their contributions in the context of the previous literature. To highlight the connections between the different pieces and facilitate the exposition, we group the articles into four broad categories: factor models/matrix completion methods, extensions/modifications/generalizations of the basic synthetic control estimator, time series analysis, and uncertainty quantification/inference.

Factor models/matrix completion methods: Agarwal et al. (2021), Athey et al. (2021), and Bai and Ng (2021)

A recent literature on synthetic controls and related methods recasts the synthetic control estimation problem as a matrix completion problem, and applies low rank approximation methods to impute the counterfactual outcomes that would have been observed for the treated units in the absence of the treatment. In this setup, potential outcomes without treatment form an $N \times T$ array, where N is the number of units in the sample and T is the number of time periods. Missing entries in this matrix represent the counterfactual outcomes that would have been observed for the treated units in the absence of the treatment.

For this setting, Athey et al. (2021) propose a matrix completion estimator that penalizes the complexity (nuclear norm) of the resulting matrix of potential outcomes. Their results allow for time series dependency in the patterns of missingness of potential outcomes. This is a crucial feature in the context of synthetic controls applications, where potential outcomes without treatment are missing from the time of treatment adoption. At a more general level, this article establishes interesting connections between matrix completion concepts and causal inference problems commonly studied in the program evaluation literature.

Agarwal et al. (2021) show that principal component regression is equivalent to performing linear regression after preprocessing the matrix of regressors via hard singular value thresholding. An implication of this result is that principal component regression is equivalent to the robust synthetic control estimator of Amjad, Shah, and Shen (2018). The authors use this equivalence result to establish novel properties of principal component regression and robust synthetic control estimators.

Bai and Ng (2021) investigate matrix completion methods for the case when both matrix dimensions are large. They demonstrate that, under the assumption of a strong factor structure, imputation of missing entries does not require iteration or regularization, and establish a variety of consistency and asymptotic normality results. When applied to synthetic control settings with missing potential outcomes, the article provides inferential tools for average and individual treatment effects.

CONTACT Alberto Abadie abadie@mit.edu

Time series analysis: Ferman (2021), Masini and Medeiros (2021)

Much of the analysis of the original synthetic control estimator has focused on small bias properties in finite samples, under the assumption that the researcher can always perfectly fit the observed characteristics of the treated unit using a synthetic control. Ferman (2021) studies the properties of synthetic control estimators under imperfect fit. He considers a setting with a large number of pre-intervention periods, a large number of untreated units, and a generative model with a linear factor structure for the outcome variable of interest. He shows that, when synthetic controls weights are spread among many untreated units, the resulting synthetic control unit is able to approximate the values of the factor loadings of the treated unit, which results in asymptotic unbiasedness.

Masini and Medeiros (2021) consider longitudinal settings where the data is high-dimensional and nonstationary, and propose a modified-LASSO estimator of counterfactual outcomes. In addition to establishing probability concentration results for the predicted counterfactuals, they propose asymptotically valid resampling-based inference methods for treatment effects.

Extensions/modifications/generalizations: Abadie and L'Hour (2021), Ben-Michael, Feller, and Rothstein (2021), Kellogg et al. (2021)

The presence of large discrepancies in the values of the predictors for the treated unit and the values of the predictors for the synthetic control unit may create biases in the synthetic control estimator. To address this problem, Ben-Michael, Feller, and Rothstein (2021) propose a bias-correction/make" procedure for synthetic control estimators that uses ridge regression to adjust for mismatches in the pretreatment outcomes between the treated units and the un-augmented synthetic control estimator. They show that their estimator can be interpreted as a synthetic control estimator that allows for negative weights (extrapolation), but penalizes the discrepancies between those weights and the weights of the un-augmented synthetic control estimator.

In essence, a synthetic control unit is a weighted average of untreated units chosen to match the values of predictors of the outcome variable for the treated unit. As such, synthetic control estimators could potentially incorporate substantial interpolation biases if the "donor pool" of untreated units that are allowed to contribute to a synthetic control includes units that are far from the treated unit in the space of the predictors of the outcome variable. Kellogg et al. (2021) propose combining synthetic control estimators and nearest-neighbor matching estimators in a weighted average to trade off interpolation and extrapolation biases. To select the weighting factors for their estimator, they adopt a cross-validation procedure based on a rolling window and one-step-ahead forecasts over the pre-intervention periods.

Abadie and L'Hour (2021) propose a synthetic control estimator that penalizes the discrepancies between the values of the predictors for the treated unit and the values of the predictors for each of the units that contribute to the synthetic control. They show that, under weak regularity conditions, their estimator is unique and sparse (in the sense that only a small number of untreated units contribute to the synthetic control), and propose a bias-correction technique similar to that in Ben-Michael, Feller, and Rothstein (2021). The properties of the synthetic control estimator in Abadie and L'Hour (2021) make it particularly suitable for empirical applications with many treated and many untreated units, a setting where un-penalized synthetic control estimators may often not be unique.

Uncertainty Quantification/Inference: Cattaneo, Feng, and Titiunik (2021), Chernozhukov, Wüthrich, and Zhu (2021), Shaikh and Toulis (2021)

Inference with synthetic control estimators is often carried out by permutation techniques after adopting a benchmark distribution for the assignment mechanism (see, e.g., Abadie, Diamond, and Hainmueller 2010; Firpo and Possebom 2018). Building on this mode of inference, Shaikh and Toulis (2021) consider settings where multiple units adopt a treatment at different times. They augment the synthetic control framework with a proportional hazard model for time of treatment adoption as a function of observed characteristics of the units. The authors leverage the model on treatment adoption to estimate the probabilities of becoming the first adopter of the treatment for each of the units in the sample. These probabilities provide an empirically estimated distribution for the assignment mechanism that can then be applied as the basis for a permutation test on the effect of the treatment.

Chernozhukov, Wüthrich, and Zhu (2021) propose a conformal inference method for testing sharp null hypotheses in synthetic control and similar settings with counterfactual outcomes. Their procedure starts with a model for the expected values of the potential outcomes without treatment for a treated unit. In the most basic version of the procedure, error terms representing the differences between the potential outcomes for the treated unit in the absence of the treatment and their expected values under the model are assumed to be exchangeable in time. Using this setup, the authors establish conditions for valid inference in tests based on permutations of the estimated model residuals.

In a related contribution, Cattaneo, Feng, and Titiunik (2021) propose prediction intervals to complement point estimates of treatment effects in synthetic control settings. Their framework incorporates two sources of uncertainty. The first source of uncertainty pertains to the in-sample estimation of the synthetic control weights. The second source of uncertainty comes from out-of-sample error components in treatment effects. The authors propose a simulation-based method to construct non-asymptotic probability bounds that account for both types of uncertainty. The method is valid for a large class of synthetic control estimators with both stationary and nonstationary data.

We conclude this brief introduction to the special section on synthetic controls expressing our appreciation to the many people who made it happen. We would first like to express our gratitude to the co-editors Regina Liu and Hongyu Zhao, who initially approached us to inquire about our willingness to guest edit a set of articles on synthetic controls for *JASA (T&M)*, and who provided us with much support and guidance during the entire editorial process. We also thank Jamie Hutchens and Eric Sampson for the superb editorial and logistical assistance in handling the submitted articles as well as the final publishing process. Finally, we thank to the contributing authors for their efforts putting together the articles for this issue, and for carefully and promptly addressing our editorial requests.

References

Abadie, A. "Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects," *Journal of Economic Literature*, 59, 391–425. [1713]

Abadie, A., Diamond, A., and 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, 493–505. [1713,1714]

——— . (2015), "Comparative Politics and the Synthetic Control Method," *American Journal of Political Science*, 59, 495–510. [1713]

Abadie, A., and Gardeazabal, J. (2003), "The Economic Costs of Conflict: A Case Study of the Basque Country," *American Economic Review*, 93, 113–132. [1713]

Abadie, A., and L'Hour, J. (2021), "A Penalized Synthetic Control Estimator for Disaggregated Data," *Journal of the American Statistical Association*, 116, 1817–1834, DOI: 10.1080/01621459.2021.1971535. [1714]

Agarwal, A., Shah, D., Shen, D., and Song, D. (2021), "On Robustness of Principal Component Regression," *Journal of the American Statistical Association*, 116, 1731–1745, DOI: 10.1080/01621459.2021.1928513. [1713]

Amjad, M., Shah, D., and Shen, D. (2018), "Robust Synthetic Control," *Journal of Machine Learning Research*, 19, 1–51. Available at: *http://jmlr.org/papers/v19/17-777.html* [1713]

Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W., and Wager, S. (2018), "Synthetic Difference in Differences," arXiv:1812.09970. [1713]

Athey, S., Bayati, M., Doudchenko, N., Imbens, G., and Khosravi, K. (2021), "Matrix Completion Methods for Causal Panel Data Models," *Journal of the American Statistical Association*, 116, 1716–1730, DOI: 10.1080/01621459.2021.1891924. [1713]

Bai, J., and Ng, S. (2021), "Matrix Completion, Counterfactuals, and Factor Analysis of Missing Data," *Journal of the American Statistical Association*, 116, 1746–1763, DOI: 10.1080/01621459.2021.1967163. [1713]

Ben-Michael, E., Feller, A., and Rothstein, J. (2021), "The Augmented Synthetic Control Method," *Journal of the American Statistical Association*, 116, 1789–1803, DOI: 10.1080/01621459.2021.1929245. [1714]

Cattaneo, M. D., Feng, Y., and Titiunik, R. (2021), "Prediction Intervals for Synthetic Control Methods," *Journal of the American Statistical Association*, 116, 1865–1880, DOI: 10.1080/01621459.2021.1979561. [1714]

Chernozhukov, V., Wüthrich, K., and Zhu, Y. (2021), "An Exact and Robust Conformal Inference Method for Counterfactual and Synthetic Controls," *Journal of the American Statistical Association*, 116, 1849–1864, DOI: 10.1080/01621459.2021.1920957. [1714]

Doudchenko, N., and Imbens, G. W. (2017), "Balancing, Regression, Difference-In-Differences and Synthetic Control Methods: A Synthesis," ArXiv:1610.07748. [1713]

Ferman, B. (2021), "On the Properties of the Synthetic Control Estimator With Many Periods and Many Controls," *Journal of the American Statistical Association*, 116, 1764–1772, DOI: 10.1080/01621459.2021.1965613. [1714]

Firpo, S., and Possebom, V. (2018), "Synthetic Control Method: Inference, Sensitivity Analysis and Confidence Sets," *Journal of Causal Inference*, 6, 20160026. [1714]

Kellogg, M., Mogstad, M., Pouliot, G., and Torgovitsky, A. (2021), "Combining Matching and Synthetic Controls to Trade off Biases From Extrapolation and Interpolation," *Journal of the American Statistical Association*, 116, 1804–1816, DOI: 10.1080/01621459.2021.1979562. [1714]

Masini, R., and Medeiros, M. C. (2021), "Counterfactual Analysis With Artificial Controls: Inference, High Dimensions and Nonstationarity," *Journal of the American Statistical Association*, 116, 1773–1788, DOI: 10.1080/01621459.2021.1964978. [1714]

Shaikh, A. M., and Toulis, P. (2021), "Randomization Tests in Observational Studies with Staggered Adoption of Treatment," *Journal of the American Statistical Association*, 116, 1835–1848, DOI: 10.1080/01621459.2021.1974458. [1714]