AEA Papers and Proceedings 2024, 114: 614–617
https://doi.org/10.1257/pandp.20241050

Temporal Aggregation for the Synthetic Control Method†

By Liyang Sun, Eli Ben-Michael, and Avi Feller*

Empirical researchers often use the synthetic control method (SCM) to estimate the impact of a treatment on a single unit in panel data settings (Abadie, Diamond, and Hainmueller 2010). The synthetic control is a weighted average of control units that balances the treated unit's pretreatment outcomes as closely as possible.

Two challenges arise when using SCM with higher-frequency data, such as when the outcome is measured every month versus every year. First, because there are more pretreatment outcomes to balance, achieving excellent pretreatment fit is typically more challenging. Second, even when excellent pretreatment fit is possible, higher-frequency observations raise the possibility of bias due to overfitting to noise. A recent review by Abadie and Vives-i-Bastida (2022) explicitly cautions about such bias from using disaggregated outcomes in SCM. Instead, researchers can first aggregate the outcome series into lower-frequency (e.g., annual) observations and then estimate SCM weights that minimize the imbalance in these aggregated pretreatment outcomes. Doing so mechanically improves pretreatment fit as well.

In this paper, we propose a framework for temporal aggregation for SCM. Adapting recent results from Sun, Ben-Michael, and Feller (2023), we first derive finite-sample bounds on the bias for SCM under a linear factor model when using temporally disaggregated versus aggregated outcome series. With these bounds, we then show that temporal aggregation only reduces bias to the extent that doing so reduces noise without also overly reducing signal for the underlying factors. The optimal trade-off between the two approaches depends on unknown parameters. We argue, however, that finding synthetic control weights that jointly balance both the disaggregated and aggregate series is a promising compromise approach for practice.

Our setup builds on an expansive literature on SCM, especially recent papers that modify SCM to mitigate bias both due to imperfect pretreatment balance (e.g., Ferman and Pinto 2021; Ben-Michael, Feller, and Rothstein 2021) and bias due to overfitting to noise (e.g., Kellogg et al. 2021). Most directly relevant, Sun, Ben-Michael, and Feller (2023) discuss using multiple outcomes to mitigate both sources of bias.

I. Setup

For each unit $i = 1, \ldots, N$ and at each lower-frequency time interval $t = 1, \ldots, T$, we observe K higher-frequency observations of the outcome, denoted as Y_{itk}, where $k = 1, \ldots, K$. For instance, in a monthly series, $t = 1, \ldots, T$ could represent years and $k = 1, 2, \ldots, 12$ the months within each year. The choice of time interval t reflects how we compute temporal aggregates, such as aggregating monthly data into yearly averages. We maintain fixed values for N and K throughout the analysis.

We denote the exposure to a binary treatment by $W_i \in \{0, 1\}$. We restrict our attention to the case where a single unit receives treatment and follow the convention that this is the first one, $W_1 = 1$. The remaining $N_0 \equiv N - 1$ units are possible controls, often referred to as donor units. To simplify notation, we limit to one posttreatment period, $T = T_0 + 1$, though our results easily extend to larger T.

We denote the potential outcome under treatment w with $Y_{itk}(w)$. We are interested

* Sun: Department of Economics, University College London and CEMFI (email: liyang.sun@ucl.ac.uk); Ben-Michael: Department of Statistics and Data Science and Heinz College of Information Systems and Public Policy, Carnegie Mellon University (email: ebenmich@andrew.cmu.edu); Feller: Goldman School of Public Policy and Department of Statistics, University of California, Berkeley (email: afeller@berkeley.edu). We thank Elizabeth Stuart for sharing data for the application. Feller and Sun gratefully acknowledge support from the Institute of Education Sciences, US Department of Education, through Grant R305D200010. Sun also acknowledges support from Grant PID2022-143184NA-I00 funded by MCIN/AEI. See the associated augsynth library for R.

† Go to https://doi.org/10.1257/pandp.20241050 to visit the article page for additional materials and author disclosure statement(s).

in the treatment effects for the treated unit during the K higher-frequency observations during posttreatment period T: $\tau_{Tk} = Y_{1Tk}(1) - Y_{1Tk}(0)$ for $k = 1, \ldots, K$. Since we directly observe $Y_{1Tk}(1) = Y_{1Tk}$ for the treated unit, we focus on imputing the missing counterfactual outcome under control, $Y_{1Tk}(0)$.

Throughout, we will focus on de-meaned or intercept-shifted weighting estimators (Doudchenko and Imbens 2017; Ferman and Pinto 2021). We denote $\bar{Y}_{i\cdot\cdot} \equiv (1/T_0 K)\sum_{t=1}^{T_0}\sum_{j=1}^{K} Y_{itj}$ as the pretreatment average for the outcome for unit i and $\dot{Y}_{itk} = Y_{itk} - \bar{Y}_{i\cdot\cdot}$ as the corresponding de-meaned outcome. We consider estimators of the form $\hat{Y}_{1Tk}(0) \equiv \bar{Y}_{1\cdot\cdot} + \sum_{i=2}^{N} \gamma_i \dot{Y}_{iTk}$, where $\gamma \in \mathcal{C} \subset \mathbb{R}^{N-1}$ is a set of weights. Our paper centers on how to choose the weights γ from a set $\mathcal{C} = \left\{\gamma \in \mathbb{R}^{N-1} \middle| \|\gamma\|_1 \leq C, \sum_i \gamma_i = 1\right\}$ for a known C.

The first approach we consider is finding a synthetic control that has the best pretreatment fit on the (de-meaned) disaggregated high-frequency outcomes $q^{dis}(\cdot)$. We refer to this set of weights as the disaggregated weights $\hat{\gamma}^{dis}$:

$$\min_{\gamma \in \mathcal{C}} \frac{1}{T_0} \frac{1}{K} \sum_{k=1}^{K} \sum_{t=1}^{T_0} \left(\dot{Y}_{1tk} - \sum_{W_i = 0} \gamma_i \dot{Y}_{itk} \right)^2.$$

An alternative choice is to optimize the aggregated objective $q^{agg}(\cdot)$, the pretreatment fit for the temporally aggregated outcomes via averaging. We refer to the set of weights that minimize this objective as the aggregated weights $\hat{\gamma}^{agg}$:

$$\min_{\gamma \in \mathcal{C}} \frac{1}{T_0} \sum_{t=1}^{T_0} \left(\frac{1}{K} \sum_{k=1}^{K} \dot{Y}_{1tk} - \sum_{W_i = 0} \gamma_i \dot{Y}_{itk} \right)^2.$$

II. Bias Bounds

To derive finite sample bias bounds for $Y_{1Tk}(0) - \hat{Y}_{1Tk}(0)$ for each $k = 1, \ldots, K$, we assume that the outcomes under control are generated as $Y_{itk}(0) = \alpha_i + \beta_{tk} + L_{itk} + \varepsilon_{itk}$, with $\sum_{t=1}^{T} \beta_{tk} = 0$ for all k; see, for example, Athey et al. (2021). After incorporating the additive two-way fixed effects, the model component retains a term L_{itk} with $\sum_{i=1}^{N} L_{itk} = 0$ for all t, k and $\sum_{t=1}^{T} L_{itk} = 0$ for all i, k. We assume the idiosyncratic errors ε_{itk} are mean zero sub-Gaussian random variables with scale parameter σ, independent of the treatment status W_i. We also assume independence across units and time, which is plausible if the model components capture comovement in the outcome.

Let the matrix $\mathbf{L} \in \mathbb{R}^{N \times (TK)}$ contain L_{itk} for the treated unit and the remaining rows correspond to control units. Sun, Ben-Michael, and Feller (2023) show that a low-rank condition $rank(\mathbf{L}) < N - 1$ is necessary for there to exist oracle weights γ^* over donor units that yield an unbiased estimate for the control potential outcome: $E_{\varepsilon_{Tk}}[Y_{1Tk}(0) - \hat{Y}_{1Tk}(0)] = L_{1Tk} - \sum_{i=2}^{N} \gamma_i^* L_{iTk} = 0$. Here, the expectation is taken over the idiosyncratic errors in the respective posttreatment periods.

The deterministic model component can be written as a linear factor model, $L_{itk} = \phi_i \cdot \boldsymbol{\mu}_{tk}$, with $r = rank(\mathbf{L})$, where $\boldsymbol{\mu}_{tk} \in \mathbb{R}^r$ are latent time factors and each unit has a vector of time-invariant factor loadings $\phi_i \in \mathbb{R}^r$. Two important quantities for our discussion are $\underline{\xi}^{dis}$ and $\underline{\xi}^{agg}$, the smallest singular values of the variance-covariance matrix of, respectively, the time factors μ_{tk} and the averaged time factors $\bar{\mu}_t = (1/K)\sum_{k=1}^{K} \mu_{tk}$. Following previous literature, we assume that $\underline{\xi}^{dis} > 0$, which avoids issues of weak identification (Abadie, Diamond, and Hainmueller 2010).

For estimated weights $\hat{\gamma}$ based on pretreatment fit, the bias for the effect in period Tk is due to inadequate balance in the model components: $Bias(\hat{\gamma}) = L_{1Tk} - \sum_{i=2}^{N} \hat{\gamma}_i L_{iTk}$. We can decompose the bias from estimated weights $Bias(\hat{\gamma})$ into two terms using the linear factor model:

$$\sum_{t=1}^{T_0} \sum_{j=1}^{K} \omega_{tj} \left(\dot{Y}_{1tj} - \sum_{W_i=0} \hat{\gamma}_i \dot{Y}_{itj} \right) \quad \text{(imbalance)}$$

$$-\sum_{t=1}^{T_0} \sum_{j=1}^{K} \omega_{tj} \left(\dot{\varepsilon}_{1tj} - \sum_{W_i=0} \hat{\gamma}_i \dot{\varepsilon}_{itj} \right) \quad \text{(overfitting)},$$

where ω_{tj} are transformations of the factor values that depend on the estimator.

The first term is bias due to imperfect pretreatment fit (or imbalance) in the pretreatment outcomes, $\dot{Y}_{itj}$. The second term is bias due to overfitting to noise, also known as the approximation error. This term arises because the optimization problems minimize imbalance in observed pretreatment outcomes—noisy realizations of latent factors—rather than minimizing imbalance in the latent factors themselves.

Theorem 1 in the online Appendix formally states high-probability bounds on the bias terms, which we obtain using results from Sun, Ben-Michael, and Feller (2023). These bounds hold in finite samples and account for imperfect pretreatment fit. The leading terms in the bias due to imbalance and overfitting are

$$\left|Bias\left(\hat{\gamma}^{dis}\right)\right| = O\left(\frac{1}{\underline{\xi}^{dis}}\right) + O\left(\frac{1}{\underline{\xi}^{dis}\sqrt{T_0 K}}\right),$$

$$\left|Bias\left(\hat{\gamma}^{agg}\right)\right| = O\left(\frac{1}{\underline{\xi}^{agg}\sqrt{K}}\right) + O\left(\frac{1}{\underline{\xi}^{agg}\sqrt{T_0 K}}\right),$$

where $\underline{\xi}^{dis}$ and $\underline{\xi}^{agg}$ are the relevant smallest singular values for the disaggregated and aggregated series.

The first term, bias due to imbalance, does not vanish for either approach. Consistent with Ferman and Pinto (2021), bias remains because the SCM objective does not converge to the objective minimized by the oracle weights due to noise in the outcomes. Aggregation, however, reduces the level of noise in the objective and therefore reduces the bias by a factor of $1/\sqrt{K}$.

The second term in the bias is the contribution of overfitting to noise. This decreases in the total number of pretreatment observations $T_0 K$ for both the disaggregated and aggregated estimators. For the disaggregated weights, this follows prior results (e.g., Abadie, Diamond, and Hainmueller 2010). Aggregation reduces the noise in the objective, but this is counteracted by a commensurate reduction in the number of pretreatment observations, netting out to the same risk of bias due to overfitting as with the disaggregated weights.

In practice, the most important consideration in temporal aggregation is whether doing so eliminates the signal for the underlying factors. In particular, if (ignoring constant terms) $\sqrt{K}\underline{\xi}^{agg} > \underline{\xi}^{dis}$, aggregation will lead to tighter bias bounds. There are many scenarios where we expect long-run variation to persist, especially after seasonally adjusting the outcomes, and therefore believe this condition might hold. However, if there is not substantial long-run variation and $\underline{\xi}^{agg}$ is very small, aggregating can leave little behind to learn about the latent factor loadings and can possibly inflate the bias bound. Similar challenges arise in time series model estimation, where aggregation can lead to biased estimates for the true time series models (see Marcet 1991, among others).

Rather than choose between these two extremes, we propose finding SCM weights that control a linear combination of the two objectives, with weight ν on the aggregated objective and weight $(1-\nu)$ on the disaggregated objective. This creates an imbalance frontier, similar to the approach in Ben-Michael, Feller, and Rothstein (2022), where $\nu = 0$ corresponds to the disaggregated objective and $\nu = 1$ to the aggregated objective. As formalized in Lemma 1 in the online Appendix, if the SCM weights yield excellent pretreatment fit on both the disaggregated and aggregated outcomes, these weights will also achieve the minimum of the two bounds. In general, finding the optimal ν^* involves model-derived parameters and is infeasible. As a heuristic, we propose assigning equal weight to the disaggregated and aggregated fits, $\nu = 0.5$, and then assessing sensitivity to this choice.

There are many directions for future work that incorporate recent innovations in panel data methods, including first denoising (e.g., Amjad, Shah, and Shen 2018) or seasonally adjusting the disaggregated outcome series. We could also explore choosing an optimal level of temporal aggregation for a single SCM objective. Finally, questions about temporal aggregation also arise in event study and other panel data models, suggesting further avenues for fruitful research.

III. Texas 2021 Abortion Restrictions

We revisit the Bell, Stuart, and Gemmill (2023) study on SB8, a 2021 Texas law restricting abortion. Using monthly state-level live birth counts from 2016 to 2022, the authors construct a synthetic Texas to estimate SB8's impact on monthly births.

The original analysis uses SCM on monthly data ($\nu = 0$). We explore aggregating to yearly averages ($\nu = 1$) and intermediate values of ν as illustrated in Figure 1. Combining yearly and monthly births equally ($\nu = 0.5$) achieves substantial balance on both fronts, reducing potential bias (highlighted by "Yearly + monthly" in Figure 1). Online Appendix Figure 1 indicates a slightly larger estimated effect for posttreatment monthly births using the combined approach ($\nu = 0.5$) than the original analysis. Online Appendix Figure 2 demonstrates stable estimates across a wide range of ν.

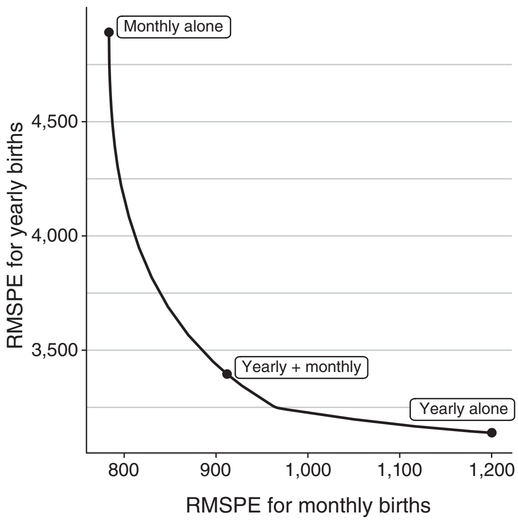

Figure 1. Imbalance Frontier

REFERENCES

Abadie, Alberto, and Jaume Vives-i-Bastida. 2022. "Synthetic Controls in Action." arXiv: 2203.06279.

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 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.

Amjad, Muhammad, Devavrat Shah, and Dennis Shen. 2018. "Robust Synthetic Control." *Journal of Machine Learning Research* 19 (22): 1–51.

Athey, Susan, Mohsen Bayati, Nikolay Doudchenko, Guido Imbens, and Khashayar Khosravi. 2021. "Matrix Completion Methods for Causal Panel Data Models." *Journal of the American Statistical Association* 116 (536): 1716–30.

Bell, Suzanne O., Elizabeth A. Stuart, and Alison Gemmill. 2023. "Texas' 2021 Ban on Abortion in Early Pregnancy and Changes in Live Births." *JAMA* 330 (3): 281–82.

Ben-Michael, Eli, Avi Feller, and Jesse Rothstein. 2021. "The Augmented Synthetic Control Method." *Journal of the American Statistical Association* 116 (536): 1789–803.

Ben-Michael, Eli, Avi Feller, and Jesse Rothstein. 2022. "Synthetic Controls with Staggered Adoption." *Journal of the Royal Statistical Society Series B: Statistical Methodology* 84 (2): 351–81.

Doudchenko, Nikolay, and Guido W. Imbens. 2017. "Balancing, Regression, Difference-in-Differences and Synthetic Control Methods: A Synthesis." arXiv: 1610.07748.

Ferman, Bruno, and Cristine Pinto. 2021. "Synthetic Controls with Imperfect Pretreatment Fit." *Quantitative Economics* 12 (4): 1197–221.

Kellogg, Maxwell, Magne Mogstad, Guillaume A. Pouliot, and Alexander Torgovitsky. 2021. "Combining Matching and Synthetic Control to Tradeoff Biases from Extrapolation and Interpolation." *Journal of the American Statistical Association* 116 (536): 1804–16.

Marcet, Albert. 1991. "Temporal Aggregation of Economic Time Series." In *Rational Expectations Econometrics*, edited by Lars Peter Hansen and Thomas J. Sargent, 237–82. Boulder, CO: Westview Press.

Sun, Liyang, Eli Ben-Michael, and Avi Feller. 2023. "Using Multiple Outcomes to Improve the Synthetic Control Method." arXiv: 2311.16260.