Temporal Aggregation for the Synthetic Control Method

Liyang Sun[1], **Eli Ben-Michael**[2], **Avi Feller**[3],

[1]UCL and CEMFI, [2]CMU, [3]UC Berkeley

2024 ASSA

Motivation: Synthetic Control Method (SCM)

- Abadie, Diamond, and Hainmuller (2010) propose the SCM for common empirical settings:
  - aggregated panel data: one treated unit and several control
  - with many more controls than pre-treatment years
  - treatment is not random
- Effectively a matching estimator that estimates the counterfactual outcome for the treated unit
- Use pre-treatment outcome data to identify the weighted average of control units that most closely approximates the treated unit.

Temporal Aggregation in SCM

- Abadie, Diamond, and Hainmuller (2010) caution the SCM can be biased if the in-sample pre-treatment fit is poor
  - Achieving excellent pre-treatment fit is typically more challenging for higher frequency
- Examples of different frequency of measurements of the outcome
  - GDP analysis often annual, e.g. ADH (2010), Billmeier and Nannicini (2013), Pinotti (2015)
  - Housing data is available quarterly, e.g. Bohn et al (2014)
  - Employment analyzed monthly, e.g. Jardim et al (2022)
  - Firm trading behavior analyzed daily e.g. Acemoglu et al (2016)
- Should we try to achieve better fit by aggregating, e.g., from monthly to yearly averages?

This paper

- Formalize the intuition:
    - Aggregating outcome series into lower-frequency observations can potentially mitigate SCM bias under a linear factor model.
- But no free lunch:
    - Aggregation can help by reducing noise, but it may also eliminate valuable signals, possibly worsening bias.
- Practical recommendation:
    - Can jointly balance aggregated and disaggregated series to minimize bias.

Outline

Literature review

SCM Review

Bias of the SCM

Application

Literature review: SCM for single outcome

Huge literature on SCM for single outcome with fixed temporal aggregation

- Improvement from incorporating an outcome model (Doudchenko and Imbens 2017, Ben-Michael, Feller, and Rothstein 2021, etc.)
- Conformal inference method (Chernozhukov, Wüthrich, Zhu, 2021)
- All implemented in R package `augsynth`

Literature review: SCM for multiple outcomes

This paper also relates to the active research on extending SCM for multiple outcomes

- Amjad et al (2019) propose mRSC that extends RSC to a setting of multiple outcomes
- Sun, Ben-Michael, and Feller (2023, WP) recently explored SCM with multiple outcomes; provide theoretical conditions for when incorporating multiple outcomes can mitigate SCM bias
- Key technical formulation is common latent factor structure across multiple outcomes
- We directly apply their setup in the context of temporal aggregation

Outline

Aggregate Panel Data

- Consider a panel data setting of N units and T lower-frequency time intervals (Abadie, Diamond, and Hainmuller, 2010)
- For each unit $i = 1, \ldots, N$ and at each time period $t = 1, \ldots, T$, we observe K higher-frequency observations of the outcomes Y_{itk} where $k = 1, \ldots, K$
  - For example, we can represent a long quarterly series where $t = 1, \ldots, T$ indexes year, and $k = 1, 2, 3, 4$ indexes quarter within each year
- Denote the potential outcome under treatment w with $Y_{itk}(w)$
- A single unit receives treatment, and the convention 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 post-treatment period, $T = T_0 + 1$

Weighting estimators

- The estimands are treatment effects for the treated unit: $\tau_{Tk} = Y_{1Tk}(1) - Y_{1Tk}(0)$ for all $k = 1, \ldots, K$
- 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)
  - We denote $\bar{Y}_{i\cdot\cdot} \equiv \frac{1}{T_0 K} \sum_{t=1}^{T_0} \sum_{j=1}^{K} Y_{itj}$ as the pre-treatment 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: $\widehat{Y}_{1Tk}(0) \equiv \bar{Y}_{1\cdot\cdot} + \sum_{i=2}^{N} \gamma_i \dot{Y}_{iTk}$, where $\gamma \in \mathcal{C} = \{\gamma \in \mathbb{R}^{N-1} \mid |\gamma_i| \leq C, \sum_i \gamma_i = 1\}$
  - Abadie, Diamond, and Hainmuller (2010) argue the simplex constraint ensures that the weights will be sparse and provides regularization
  - We slightly relax the simplex constraint

Model: Assumption on Counterfactual Outcomes

- Under what assumptions is the SCM a good estimator?

Assumption (Fixed component)

The outcome under control is generated as

$$Y_{itk}(0) = \alpha_i + \beta_{tk} + L_{itk} + \varepsilon_{itk}$$

where L_{itk} is a deterministic model component, and the idiosyncratic errors ε_{itk} are mean zero, independent of the treatment status W_{it}, independent across units and time.

- These deterministic model components are equivalent to **linear factor model**, a common assumption in the SCM literature (Abadie, Diamond, and Hainmuller, 2010)

Connection to linear factor model

- Let the matrix $L \in \mathbb{R}^{N \times (TK)}$ contain L_{itk}. If $rank(L) = r > 0$, then the model component decomposes
$$L_{itk} = \phi_i \cdot \mu_{tk} \tag{1}$$
where $\mu_{tk} \in \mathbb{R}^r$ are the latent time-outcome factors and each unit has a vector of time-outcome-invariant factor loadings $\phi_i \in \mathbb{R}^r$
- Allows the unobserved factors to affect the treated unit differently, which would violate the parallel trends assumption that motivates TWFE
- TWFE assumes $L_{itk} = 0$ so that $Y_{itk}(0) = \alpha_i + \beta_{tk} + \varepsilon_{itk}$

Implied bias for SCM under factor model

- N and T_0 are usually small for directly estimating the factor model
- SCM instead only tries to recover model components for the treated units
- For any estimated weights $\hat{\gamma}$, the estimation error is a function of

$$Y_{1Tk}(0) - \widehat{Y}_{1Tk}(0) = \underbrace{\beta_{Tk}\left(1 - \sum_{W_i=0} \hat{\gamma}_i\right) + L_{1Tk} - \sum_{W_i=0} \hat{\gamma}_i L_{iTk}}_{\text{bias}} + \underbrace{\varepsilon_{1Tk} - \sum_{W_i=0} \hat{\gamma}_i \varepsilon_{iTk}}_{\text{noise}}$$

Model: Assumption on Oracle Weights

Assumption (Oracle Weights)

There exists $\gamma^* \in \mathcal{C}$ that solves the following system of TK equations

$$L_{1tj} = \sum_{W_i=0} \gamma_i^* L_{itj}, \; \forall t = 1, \ldots, T, \; j = 1, \ldots, K$$

- Sun, Ben-Michael and Feller (2023, WP) argues that a necessary condition is that L is low rank ($r < N - 1$)
- Intuitively, the less complicated is the factor structure, the more likely there is a solution (in fact, infinitely many solutions)
- The additional constraint on γ^* provides regularization

Outline

Measure of imbalance

- Applying the classical synthetic control directly to the disaggregated high-frequency outcomes gives the *disaggregated objective* $q^{\text{dis}}(\cdot)$:

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

- An alternative choice is the *aggregated objective* $q^{\text{agg}}(\cdot)$, the pre-treatment fit for the temporally aggregated outcomes:

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

Bias decomposition

- For any estimated weights $\hat{\gamma} \in \mathcal{C}$ that minimize pre-treatment imbalance, the bias term $L_{1Tk} - \sum_{W_i=0} \hat{\gamma}_i L_{iTk}$ can be further related to their objective function:

$$\sum_{t=1}^{T_0} \sum_{j=1}^{K} \omega_{tj} \left(\dot{Y}_{1tj} - \sum_{W_i=0} \hat{\gamma}_i \dot{Y}_{itj} \right) \quad (R_0 : \text{imbalance})$$

$$- \sum_{t=1}^{T_0} \sum_{j=1}^{K} \omega_{tj} \left(\dot{\varepsilon}_{1tj} - \sum_{W_i=0} \hat{\gamma}_i \dot{\varepsilon}_{itj} \right) \quad (R_1 : \text{overfitting bias})$$

- the weights are projected factor values that depend on the specific estimator

Finite-sample bias bounds from Sun, Ben-Michael and Feller (2023, WP)

Theorem

In addition to assumptions stated above, suppose the idiosyncratic errors are sub-Gaussian with scale parameter σ. Assume the time factors are bounded above by M. Then with high probability,

$$\left|Bias(\hat{\gamma}^{dis})\right| \leq \frac{rM^2}{\xi^{dis}} \left(4(1+C)\sigma + 2\delta + \frac{\tilde{\sigma}}{\sqrt{T_0 K}} \right),$$

$$|Bias(\hat{\gamma}^{agg})| \leq \frac{rM^2}{\xi^{agg}} \left(\frac{4(1+C)\sigma}{\sqrt{K}} + 2\delta + \frac{\tilde{\sigma}}{\sqrt{T_0 K}} \right).$$

- Key step: the minimized in-sample imbalance is bounded above by the in-sample imbalance obtained by the oracle weights

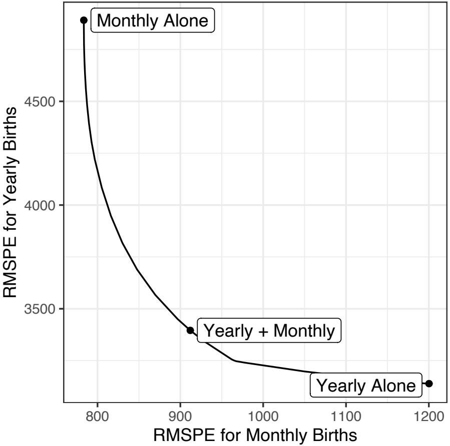
Bias from imperfect pre-treatment balance

- Leading terms in the bias due to imbalance are $O\left(1/\xi^{\text{dis}}\right)$ versus $O\left(1/(\xi^{\text{agg}}\sqrt{K})\right)$
- Consistent with Ferman and Pinto (2021), the SCM objective function does not converge to the objective function minimized by the oracle weights, and therefore remains biased
- However, the bias due to imbalance for the aggregate weights will decrease with the number of aggregation periods K
- This is because aggregating outcomes reduces the level of noise in the objective function

Overfitting

- Leading terms in the bias due to overfitting are $O\left(1/(\underline{\xi}^{\text{dis}}\sqrt{T_0 K})\right)$ versus $O\left(1/(\underline{\xi}^{\text{agg}}\sqrt{T_0 K})\right)$
- Overfitting bias cannot be reduced by aggregation
- Aggregation can potentially amplify the bias if $\underline{\xi}^{\text{agg}} \ll \underline{\xi}^{\text{dis}}$, which can happen if aggregation leaves little time variation behind to infer about the latent loadings
    - Here $\underline{\xi}^{\text{dis}}$ and $\underline{\xi}^{\text{agg}}$ are the lower bounds for $\sigma_{min}\left(\frac{1}{T_0 K}\sum_{tk}\mu_{tk}\mu_{tk}'\right)$ and $\sigma_{min}\left(\frac{1}{T_0}\sum_t (\bar{\mu}_t)(\bar{\mu}_t)'\right)$ where $\bar{\mu}_t = \frac{1}{K}\sum_{k=1}^{K}\mu_{tk}$ and $\sigma_{\text{min}}(A)$ denotes the smallest singular value of a matrix A
- Similar issue arises in time series (Marcet, 1991)

One Practical Solution



- Minimize $\nu q^{\text{dis}}(\cdot) + (1-\nu) q^{\text{agg}}(\cdot)$
- The optimal combination achieves a bias bound that is the minimum of the two bounds
- Reach the imbalance "frontier" (Ben-Michael, Feller and Rothstein, 2022)
- In the application, we consider equal combination

Outline

Literature review

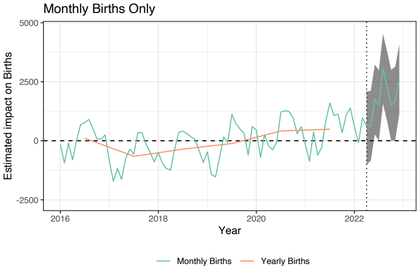
SCM Review

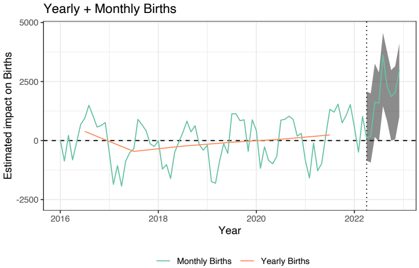
Bias of the SCM

Application

Texas Abortion Ban

- Bell, Stuart and Gemmill (2023) study the impact of the 2021 Texas abortion ban on birth outcomes
- They collected monthly counts of live births in all 50 states plus the District of Columbia for 2016 through 2022
- Births start in April 2022 are exposed to the ban





- The original paper applies SCM to monthly births, which has poor pre-treatment balance
- Aggregating to yearly averages improves pre-treatment balance, though we might be concerned about loss of signal
- Balancing in both monthly and yearly can mitigate such concern

Conclusion

- If outcomes are measured with high frequency, well-understood SCM might be overfitting to noise (Abadie and Vives-i-Bastida, 2022)
- We re-analyze the bias of SCM under a latent factor model for different levels of temporal aggregation
- However, there is a tradeoff
  - aggregation reduces noise and improves pre-treatment balance
  - aggregation can also reduce signal, amplifying the bias
- One practical solution is to jointly balance aggregated and disaggregated series to optimize such tradeoff
- Future research: de-noised SCM, extend insights to event study models, augmented methods and Synth DiD?

References

- Amjad, Muhammad, Devavrat Shah, and Dennis Shen, "Robust Synthetic Control," *Journal of Machine Learning Research*, 2018, 19 (22).
- Amjad, Muhammad, Vishal Misra, Devavrat Shah, and Dennis Shen, "mRSC: Multi-dimensional Robust Synthetic Control," *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, June 2019, 3 (2), 37:1-37:27.
- Ferman, Bruno and Cristine Pinto, "Synthetic Controls with Imperfect Pre-Treatment Fit," *Quantitative Economics*, 2021.
- Doudchenko, Nikolay, and Guido W. Imbens. 2017. "Difference-In-Differences and Synthetic Control Methods: A Synthesis." arXiv 1610.07748.
- 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): 351-381.
- Bell, Suzanne O., Elizabeth A. Stuart, and Alison Gemmill. 2023. "Texas' 2021 Ban on Abortion in Early Pregnancy and Changes in Live Births." *JAMA*.

References

- Abadie, Alberto, Alexis Diamond, and Jens Hainmueller, “Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California’s Tobacco Control Program,” Journal of the American Statistical Association, 2010, 105 (490), 493–505
- Ben-Michael, Eli, Avi Feller, and Jesse Rothstein, “The Augmented Synthetic Control Method,” Journal of the American Statistical Association, 2021, 0 (0), 1–27
- Sun, Liyang, Eli Ben-Michael, and Avi Feller. 2023.“Using Multiple Out- comes to Improve the Synthetic Control Method."
- Marcet, Albert. 1991. “Temporal aggregation of economic time series." In Rational expectations econometrics. , ed. Thomas J Sargent and Lars Peter Hansen, 237-282. Westview Press

References

- Pinotti, P. (2015). The Economic Costs of Organised Crime: Evidence from Southern Italy. *The Economic Journal, 125*, F203-F232.
- Bohn, S., Lofstrom, M., and Raphael, S. (2014). Did the 2007 Legal Arizona Workers Act Reduce the State's Unauthorized Immigrant Population? *The Review of Economics and Statistics, 96*, 258-269.
- Billmeier, A., and Nannicini, T. (2013). Assessing Economic Liberalization Episodes: A Synthetic Control Approach. *The Review of Economics and Statistics, 95*, 983-1001.
- Acemoglu, D., Johnson, S., Kermani, A., Kwak, J., and Mitton, T. (2016). The value of connections in turbulent times: Evidence from the United States. *Journal of Financial Economics, 121*, 368-391.