SYNTHETIC CONTROLS

Michal Kolesár*

April 29, 2024

1. BASICS OF THE SYNTHETIC CONTROL METHOD

To understand the strengths and weaknesses of any causal inference method, it is useful to break it down into two related pieces: (a) what is the precise estimand—the precise counterfactual that the method targets, and (b) how is the counterfactual constructed. Constructing the counterfactual is an extrapolation exercise, and it is useful consider the functional form restrictions, as well as the time periods or units that we are using for the extrapolation. For instance, we may be less concerned with linearity restrictions if we are only extrapolating over a short run.

Consider a simple difference-in-differences (DiD) setup with two time periods $t = 0, 1$ and two groups of units: control units that never get treated, and treated units that start treatment in period 1. We have seen than the DiD estimand corresponds to the average treatment effect for the treated (ATT), so that we are only constructing a counterfactual $Y_{i1}(0)$ for the treated units. Furthermore, the estimator the estimator can be decomposed as $\hat{\beta} = N_1^{-1}\sum_{i\colon D_{i1}=1}\hat{\tau}_i$, with $N_d = \sum_{i=1}^{n}\mathbb{1}\{D_{i1} = d\}$ denoting the number of units in each group, and

$$\hat{\tau}_i = Y_{i1} - \hat{Y}_{i1}(0), \quad \hat{Y}_{i1}(0) = \underbrace{\left(Y_{i0} - \frac{1}{N_0}\sum_{j\colon D_{j1}=0} Y_{j0}\right)}_{\hat{\mu}_i} + \frac{1}{N_0}\sum_{i\colon D_{j1}=0} Y_{j1}, \tag{1}$$

In other words, the DiD estimator puts equal weight $w_i = 1/N_0$ on each of the control units, and estimates the counterfactual outcome as $\hat{Y}_{it}(0) = \hat{\mu}_i + \sum_i w_i Y_{it}$, where $\hat{\mu}_i$ captures the differential level of outcome between the treated unit i and the control units. Equivalently, we're using the average change in the outcomes $\frac{1}{N_0}\sum_j (Y_{j1} - Y_{j0})$ to estimate $Y_{i1}(0) - Y_{i0}(0)$. If there are multiple post-treatment periods, and we use an event-study approach, the impact of the treatment in the later periods is estimated analogously, obtaining a dynamic treatment effect path.

*Email: mkolesar@princeton.edu.

Table 1: Difference-in-differences estimate on unemployment rates (African-American workers). Adapted from Card (1990).

	Year		
	1979	1981	Difference
Miami	8.3 (1.7)	9.6 (1.8)	1.3 (2.5)
Comparison Cities	10.3 (0.8)	12.6 (0.9)	2.3 (1.2)
Difference	−2.0 (1.9)	−3.0 (2.0)	−1.0 (2.8)

Imposing that each control unit has the same weight may be restrictive: after all, in cross-section methods for estimating the average treatment effect (ATE) or ATT under unconfoundedness, regression as well as propensity score weighting methods weight the control units by how similar their covariates are to the covariates of the treated units. Similarly, in regression discontinuity (RD), units farther away from the cutoff receive little or no weight in constructing the counterfactual.

The synthetic control (SC) estimator of Abadie and Gardeazabal (2003) and Abadie, Diamond, and Hainmueller (2010) provides a data-driven procedure to create a comparison unit in comparative case studies. The method is based on the observation that a weighted average of control units (a "synthetic control") often does a "better job" reproducing the characteristics of a treated unit than a simple average of control units (as DiD estimators do), and a better job than picking a single comparison unit (as nearest neighbor matching estimators do). As such, the ideas behind the SC method have a lot of potential, which is why Athey and Imbens (2017) call it "arguably the most important innovation in the policy evaluation literature in the last 15 years".

To motivate this method, consider the setup in Card (1990). Card is interested in the effect of the Mariel boatlift on the labor market outcomes of native workers in Miami. The Mariel boatlift was a mass emigration of Cubans, who traveled from Cuba's Mariel Harbor to the United States between April and October 1980, after Castro declared on April 20 that Cubans wishing to emigrate to the US were free to do so.

Using a difference-in-differences design, Card (1990) concludes that in spite of increasing the Miami labor force by 7%, the influx of immigrants did not have an effect on wages and unemployment rates of less-skilled native workers. In particular, Card uses four other cities in the south of the United States (Atlanta, Los Angeles, Houston, and Tampa-St. Petersburg) as a control group to approximate the change in native unemployment rates that would have been observed in Miami in the absence of the Mariel Boatlift. The summary of his analysis is in Table 1.

However, one may be concerned that the conclusions are sensitive to how the comparison units were chosen: is there a way of picking the comparison units in a less ad hoc fashion? This is the motivation for the reanalysis of the data in Peri and Yasenov (2019). Indeed, one can think of the SC method as trying to make the control group selection less ad hoc. If we can do that, then we can also hope to formalize the uncertainty about the ability of the control group to reproduce the counterfactual of interest.

1.1. Setup

Suppose there are $N_0 + 1$ units, $i = 0, \dots, N_0$, with unit 0 exposed to intervention in periods $T_0 + 1, \dots, T$. Units $1, \dots, N_0$ are the "donor pool" in the sense that they are potential controls. Our goal is to estimate $\tau_{0t} = Y_{0t}(1) - Y_{0t}(0)$ for $t > T_0$. Let X_i denote the vector of pre-intervention characteristics of unit i. Typically, $X_i = (Z_i, Y_{i1}, \dots, Y_{iT_0})$, where Z_i is a vector of fixed characteristics.

Similarly to the DiD case, we do not need to have a notion of potential outcomes $Y_{it}(1)$ for the control units. For example, Abadie, Diamond, and Hainmueller (2015) are interested in estimating the economic impact of German reunification on West Germany, using a donor pool of 16 OECD counties. It is hard to conceptualize what it would mean for these countries to "reunify".

The SC estimator constructs a single synthetic control unit as a weighted average of the units from the donor pool. These weights w^* minimize the discrepancy between the characteristics X_0 of the treated unit, and the characteristics of the synthetic control:

$$w^* = \underset{w}{\operatorname{argmin}} \|X_0 - \sum_i w_i X_i\| = \|X_0 - X'w\| \qquad \text{st} \quad \sum_{i=1}^{N_0} w_i = 1, \quad w_i \geq 0. \tag{2}$$

The synthetic control estimator is

$$\hat{\tau}_{0t} = Y_{0t} - \hat{Y}_{0t}(0), \qquad \hat{Y}_{0t}(0) = \sum_i w_i^* Y_{it}. \tag{3}$$

If there is more than one treated unit, we can estimate the treatment effect of each of them in the same way.

Remark 1. The restriction that the weights are positive plays three roles. First, it "precludes extrapolation" in the sense that the predicted counterfactual outcome $Y_{0t}(0)$ is always in the convex hull of the outcomes for the control units. However, allowing for extrapolation may be important if the treated unit is substantively different from the control units: in such cases, extrapolation may be necessary. Second, this restriction also allows the procedure to obtain unique weights even when the number of lagged outcomes is small relative to the number of control units, a common setting in applications, which aids interpretation. Finally, it ensures that, if X_0 doesn't belong to the convex null of the rows of X, then w^* is unique and sparse: thus, the restriction serves to regularize the estimator. In particular, if $w_i \geq 0$, then the synthetic control is the closest point to X_0 from the convex hull of $X_1, \dots, X_{N_0}$. But points on the surface of a convex hull are convex combinations of a few units: at most $\dim(X_i)$ if $X_1, \dots, X_{N_0}$ are in a "general position" (A set of points in a k-dimensional Euclidean space is in general position if no m of them lie in a $(m-2)$-dimensional hyperplane for $k = 2, \dots, p+1$). This sparsity makes the counterfactual outcome estimate very transparent, a big reason for the method's popularity.

Remark 2 (Comparison with DiD). Comparing eq. (1) with eq. (3), we see that there are two key differences between DiD and SC. First, SC doesn't allow for a non-zero intercept in the counterfactual prediction $\hat{Y}_{it}(0) = \mu + \sum_i w_i Y_{it}$: this is a critical feature of the DiD approach, allowing for permanently different levels for different units. On the other hand, DiD restricts the weights w_i to be equal: this is the key restriction that SC relaxes.

Doudchenko and Imbens (2017) explore this perspective further. Taking a machine learning perspective, they view the problem of constructing the counterfactual outcome $Y_{it}(0)$ as a prediction problem. Instead of imposing that the weights are non-negative, sum to one, and that there is no intercept, as the SC method does, for the case where X_i only consists of pre-treatment outcomes, they propose minimizing an elastic-net objective function $\|X_i - \mu - X'w\|_2^2 + \lambda((1-\alpha)\|w\|_2^2/\alpha + \alpha\|w\|_1)$.

To implement the SC method, one needs to pick the norm $\|\cdot\|$ in eq. (2). Abadie, Diamond, and Hainmueller (2015) use the weighted Euclidean norm $\|a\| = a'Va$, where V is a diagonal matrix with weights chosen by a split-sample method: divide pre-pretreatment period into a training and validation period. Then pick the weights to minimize the prediction error in the validation period. That is, let $t = 1, \ldots, T_0/2$ denote the training periods, and let $t = T_0/2 + 1, \ldots, T_0$ denote the validation period. Compute the mean squared prediction error (MSPE) $\sum_{t=T_0/2+1}^{T_0}(Y_{0t} - \sum_i w_i(V)Y_{it})^2$ over the validation period, using weights computed based on $X_i = (Z_i, Y_{i1}, \ldots, Y_{i,T_0/2})$ and minimizing eq. (2) over the training period. Minimize this over V, yielding V^*. Then use the resulting V^* along with the data $X_i = (Z_i, Y_{i,T_0/2}, \ldots, Y_{i,T_0})$ for the last $T_0/2$ periods before the intervention to estimate the $w^*(V^*)$.

Alternatively, Abadie and Gardeazabal (2003) propose choosing V to minimize the MSPE of the outcome variable for the pre-intervention periods, that is, they $\sum_{t=1}^{T_0}(Y_{0t} - \sum_i w_i^*(V)Y_{it})^2$ over V: if the only covariates in X_i are pre-intervention outcomes, this amounts to just using $V = I_{T_0}$.

Ultimately, the how V should be chosen depends on the relative importance of elements of X_0 as a predictor of the counterfactual post-intervention outcomes. Since the counterfactual outcomes are unobserved, the split-sample approach approximates them using the pre-intervention data.

We illustrate this approach using data from Abadie, Diamond, and Hainmueller (2015), who are interested in estimating the effect of the 1990 German reunification on the GDP of West Germany. The data spans 1960–2003, and the donor pool consists of 16 OECD countries. The covariates also include the inflation rate, trade openness (sum of exports plus imports as a percentage of GDP), industry share of value added (ten-year average), high school completion rate (five-year frequency), and ratio of domestic investment to GDP (five-year average). Years 1971–1980 are used to select the weights V. Of the 16 donors, only 5 receive weights greater than 0.01: USA 0.219, Austria 0.418, Netherlands 0.090, Switzerland 0.111 and Japan 0.155. Remarkably reasonable choices. Figure 1 plots the counterfactual estimates.

A note of caution when interpreting these types of figures: compared to DiD methods, the pre-treatment fit in these figures always looks remarkably good: but of course, this is because we choose the weights precisely to fit the pre-treatment outcome path!

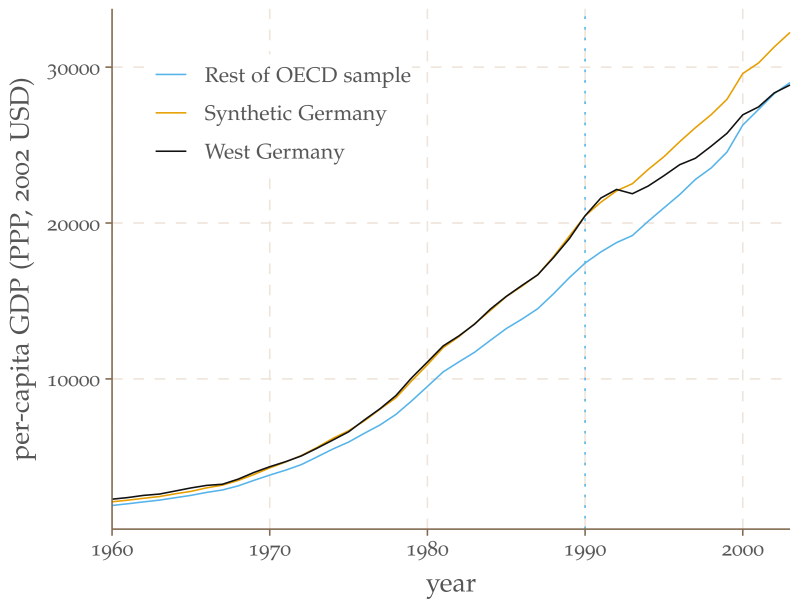

Figure 1: SC method estimates of the effect of German reunification.

This issue is magnified in this example because GDP is not stationary. For comparison, Figure 2 plots GDP growth rates, and we see that the fit is not quite as impressive. If we target GDP growth rates in constructing the synthetic weights, we get quite similar comparison group, with just Switzerland dropping out and replaced by Japan (USA 0.169; Austria 0.448; Netherlands 0.138; Japan 0.245). Figure 3 shows the counterfactual estimates. The fit does not improve, the root MSPE is about 0.5 percentage points for the period 1980–1989, and both ways of estimating the treatment effect yield root MSPE of about 1.5 percentage points post-intervention.

Remark 3. Similar to DiD studies, by saying that $Y_{it} = Y_{it}(0)$ if $t \leq T_0$ or if $i \geq 1$, we are assuming Stable unit treatment value assumption (SUTVA), which rules out spillover effects (a strong assumption for the German reunification example) and anticipation effects.

1.2. Falsification

- We can move the treatment date forward in time. For instance, we can suppose the reunification happened in 1980 and estimate the effect on 1981–1990 GDP

While the method has been increasingly popular, part of the limitation is that the assumptions underlying are not easily testable, for two reasons. First is that it's not entirely clear what the key underlying assumption is in the first place. In contrast, the parallel trends assumption in DiD is very clear. Second, there appear to be more researcher degrees of freedom, so it's a bit easier to adjust the estimator so we pass any

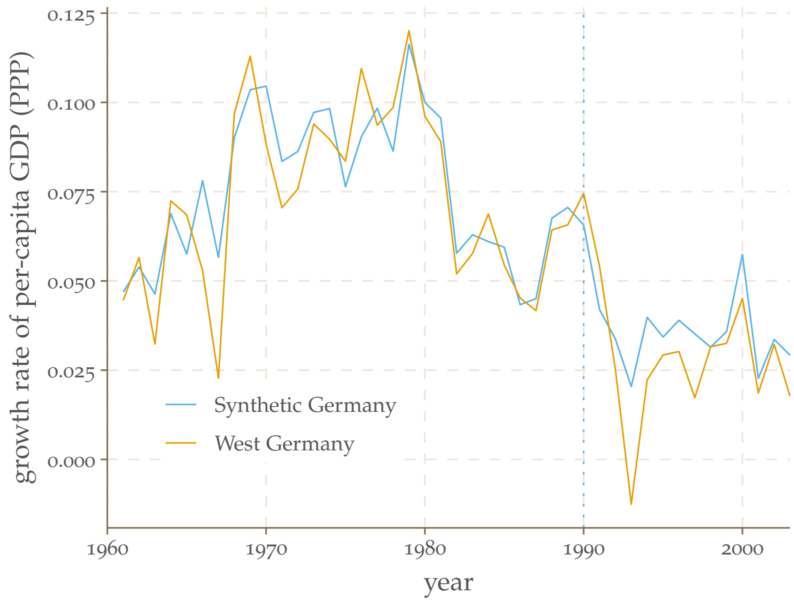

Figure 2: SC method estimates of the effect of German reunification on GDP growth rates. Method fitted on GDP.

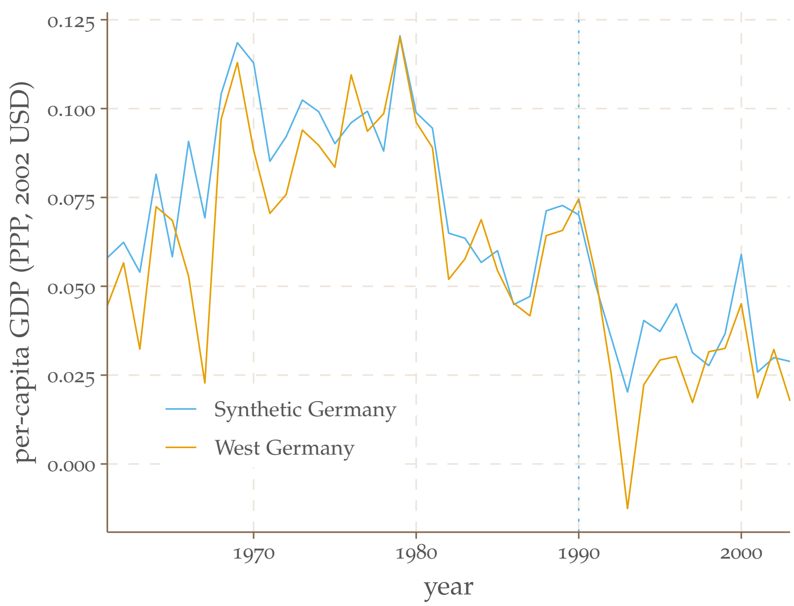

Figure 3: SC method estimates of the effect of German reunification. Method fitted on GDP growth rates.

placebo checks.

1.3. Formal properties

There is only one formal result available:

Proposition 4 (Abadie, Diamond, and Hainmueller 2010). Suppose $Y_{it}(0)$ is given by a factor model

$$Y_{it}(0) = \delta_t + Z_i'\theta_t + \lambda_t'\mu_i + \epsilon_{it},$$

where Z_i are observed covariates, λ_t are common factors, and μ_i are factor loadings. Suppose that the smallest eigenvalue of $\sum_{t=1}^{T_0} \lambda_t'\lambda_t / T_0$ is bounded away from zero, that $E|\epsilon_{it}|^p$ is bounded, and that that errors ϵ_{it} are independent of the factors, mean zero, and independent across time. Suppose also that λ_t has bounded support.

Suppose that the weights w satisfy $Z_0 = \sum_i w_i Z_i$ and $Y_{0t} = \sum_i Y_{it}$ for $t = 1, \ldots, T_0$. Then the prediction bias is bounded by

$$|E[Y_{0t}(0) - \sum_i w_i Y_{it}]| \leq K \frac{N_0^{1/p}}{T_0} \max\{(T_0 \overline{m}_p^{1/p}, (T_0 \overline{m}_2)^{1/2}\}.$$

where $K = C(p)^{1/p} \frac{\dim(\lambda_t) \max_{sj} |\lambda_{sj}|}{\min \operatorname{eig}((\lambda^{P\prime}\lambda^P / T_0))}$, $C(p)$ is a constant depending only on p given in Ibragimov and Sharakhmetov (2002), and $\overline{m}_p = \max_i T_0^{-1} \sum_{t=1}^{T_0} E|\epsilon_{it}|^p$.

Proof. Using the factor model, we can write.

$$Y_{0t}(0) - \sum_j w_j Y_{jt}(0) = (Z_0 - \sum_j w_j Z_j)'\theta_t + \lambda_t'(\mu_0 - \sum_j w_j \mu_j) + \sum_j w_j(\epsilon_{0t} - \epsilon_{jt}). \tag{4}$$

Let Y_j^P denote the vector of pre-intervention outcomes, and let λ^P denote the matrix of pre-intervention factors. Then under the assumption that we match Y_0^P and Z_0, we have

$$0 = \lambda^P(\mu_0 - \sum_j w_j \mu_j) + \sum_j w_j(\epsilon_0^P - \epsilon_j^P) = \lambda^P(\mu_0 - \sum_j w_j \mu_j) - \sum_j w_j \epsilon_j^P + \epsilon_0^P \tag{5}$$

Multiplying eq. (5) by $\lambda_t'(\lambda^{P\prime}\lambda^P)^{-1}\lambda^{P\prime}$, and subtracting it from eq. (4), we obtain

$$Y_{0t}(0) - \sum_j w_j Y_{jt}(0) = \sum_j w_j \left[\lambda_t'(\lambda^{P\prime}\lambda^P)^{-1}\lambda^{P\prime}\epsilon_j^P\right] + \sum_j w_j(\epsilon_{0t} - \epsilon_{jt}) - \lambda_t'(\lambda^{P\prime}\lambda^P)^{-1}\lambda^{P\prime}\epsilon_0^P.$$

If $t > T_0$, then the second term has mean zero. Furthermore, the third term always has mean zero since ϵ_{0t}^P is not correlated with the factors. Write

$$\sum_j w_j \left[\lambda_t'(\lambda^{P\prime}\lambda^P)^{-1}\lambda^{P\prime}\epsilon_j^P\right] = \sum_j w_j v_j,$$

where $v_j = \sum_{s=1}^{T_0} \kappa_{js}\epsilon_{js}$, with $\kappa_{ts} = \lambda_t'(\lambda^{P\prime}\lambda^P)^{-1}\lambda_s$. Now, by Hölder's inequality, since $\sum_j w_j^q \leq$

$\sum_j w_j = 1$, we have $|\sum_j w_j v_j| \le (\sum_j |v_j|^p)^{1/p}$. Therefore, by Jensen's inequality,

$$E|\sum_j w_j v_j| \le (\sum_j E|v_j|^p)^{1/p}.$$

Furthermore, by Cauchy-Schwarz inequality,

$$|\kappa_{ts}| \le \max_s \lambda_s (\lambda^{P\prime}\lambda^P)^{-1}\lambda_s \le \max \mathrm{eig}((\lambda^{P\prime}\lambda^P)^{-1}) \dim(\lambda_t) \max_{sj} |\lambda_{sj}| = \frac{\dim(\lambda_t) \max_{sj}|\lambda_{sj}|}{T_0 \min \mathrm{eig}((\lambda^{P\prime}\lambda^P/T_0))}$$

Now, by Rosenthal's inequality[1], letting $\overline{m}_p = \max_j T_0^{-1} \sum_{s=1}^{T_0} E[|\epsilon_{js}|^p]$

$$E|v_j|^p \le C(p) \max_{ts} |\kappa|_{ts}^p \max\{T_0 \overline{m}_p, (T_0 \overline{m}_2)^{p/2}\}.$$

Hence,

$$E|\sum_j w_j v_j| \le N_0^{1/p} C(p)^{1/p} (\max_{ts} |\kappa|_{ts}^p)^{1/p} \max\{T_0^{1/p} \overline{m}_p^{1/p}, (T_0 \overline{m}_2)^{1/2}\},$$

which yields the result. □

Some notes on this result:

1. The factor model is more flexible than assuming constant λ_t as DiD methods do. However, the result only gives a bound on the bias: it doesn't guarantee an unbiased estimate. Further, the constants in this bound are not known: we cannot use it directly for inference.
2. The result requires that we match X_0 exactly. If we do, and T_0 is large enough so that $T_0 \overline{m}_p < (T_0 \overline{m}_2)^{p/2}$, then the bias is of the order $N_0^{1/p} \overline{m}_2^{1/2} / T_0^{1/2}$. In other words, if N_0 is small relative to T_0, and we still manage to match X_0 exactly, then the bias is likely to be small. The intuition is that matching X_0 exactly means that we must be approximately matching the factor loadings μ_0.

   In other words, SC estimates are most credible if we ex ante reduce the donor pool to units comparable to unit 0, and have a sizable number of preintervention periods: otherwise tracking the treated unit's characteristics and outcomes over the pre-treatment period is not indicative of good post-treatment performance.
3. If we really believed this factor model, why not directly estimate it?

1.4. *Inference*

If treatment randomly assigned given X (a strong assumption!), can reassign the treatment to a random unit and recompute the treatment effect over $t = T_0 + 1, \ldots, T$. We can

1. Let Z_i be independent, mean zero, with $E[|Z_i|^t] < \infty$ for some $t > 2$. Then

$$E[|\sum_i Z_i|^t] \le C(t) \max\{\sum_i E[|Z_i|^t], (\sum_i E[Z_i^2])^{t/2}\}$$

where $C(t)$ is a constant. See Ibragimov and Sharakhmetov (2002).

plot these estimates to get a visual sense of how unusual the actual treated unit is (see Figure 4 in Abadie, Diamond, and Hainmueller (2010)). This method also delivers an exact p-value by comparing say the MSPE for the treated periods. As a test statistic, Abadie, Diamond, and Hainmueller (2010) propose using the ratio between pre-intervention and post-intervention MSPE,

$$r_j = \frac{\frac{1}{T-T_0}\sum_{t=T_0+1}^{T}(Y_{it} - \hat{Y}_{it}(0))^2}{\frac{1}{T_0}\sum_{t=1}^{T_0}(Y_{it} - \hat{Y}_{it}(0))^2}.$$

A related approach, suggested in Doudchenko and Imbens (2017) is to view the treatment unit as exchangeable with the control units in the absence of the treatment. So we can try to estimate $Y_{it}(0)$ for units $i \geq 1$ using the same estimator: drop unit 0 from the sample and let units other than i constitute the donor pool. Make a prediction $\hat{Y}_{it}(0)$. Do this for each unit i. Then we can estimate the variance of $\hat{Y}_{0t}(0)$ as $N_0^{-1}\sum_{i=1}^{N_0}(\hat{Y}_{it}(0) - Y_{it})^2$.

Because in most observational settings assignment to the intervention is not randomly assigned, the interpretation of these approaches in practice is not entirely clear. It is also unclear how one could construct a confidence interval for our predictions.

2. EXTENSIONS

In the SC prediction problem, there are three groups of control outcomes: pre-intervention outcomes for the treated unit, and pre- and post-intervention outcomes for the donor pool. The DiD method combines three groups of control units in a simple fashion, justified by a two-way fixed effects (2WFE) model, $Y_{it}(0) = \alpha_i + \lambda_t + \epsilon_{it}$. Arkhangelsky et al. (2021) augment the objective function to allow unit weights $\hat{w}_i$ and time weights $\hat{\lambda}_t$

$$\sum_{it}(Y_{it} - \mu - \alpha_i - \beta_t - D_{it}\tau)^2 \hat{w}_i \hat{\lambda}_t.$$

The unit weights are chosen to align pre-exposure trends of the treated and untreated, $\sum_{i=1}^{N_0}\hat{\omega}_i Y_{it} \approx \frac{1}{N_1}\sum_{i=N_0+1}^{n} Y_{it}$, while the time weights balance pre- and post-exposure time periods $\sum_{i=1}^{N_0}\sum_t \lambda_t Y_{it} \approx \frac{1}{T-T_0}\sum_i \sum_{t=T_0+1}^{T} Y_{it}$. In contrast, one can show that the SC estimator also minimizes the above objective if $\hat{\lambda}_t = 1$, subject to the constraint that $\alpha_i = 0$.

The justification for their procedure is provided by a factor model for $Y_{it}(0)$. They reason that because they are interested in treatment effects, rather than recovering the factor structure, they may achieve consistency for the treatment effects under weaker assumptions than those in Bai (2009).

Research Question. While the assumptions imposed in Bai (2009) are indeed stronger, is it the case that *consistency* of the factor-model estimators for treatment effects estimation does indeed require stronger assumptions? ⊠

- See JASA special issue Volume 116, Issue 536 that contains a number of articles building on the basic SC estimator.

REFERENCES

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, no. 490 (June): 493–505. https://doi.org/10.1198/jasa.2009.ap08746.

———. 2015. "Comparative Politics and the Synthetic Control Method." *American Journal of Political Science* 59, no. 2 (February): 495–510. https://doi.org/10.1111/ajps.12116.

Abadie, Alberto, and Javier Gardeazabal. 2003. "The Economic Costs of Conflict: A Case Study of the Basque Country." *American Economic Review* 93, no. 1 (March): 113–132. https://doi.org/10.1257/000282803321455188.

Arkhangelsky, Dmitry, Susan Athey, David Hirshberg, Guido Imbens, and Stefan Wager. 2021. "Synthetic Difference In Differences." *American Economic Review* 111, no. 112 (December): 4088–4118. https://doi.org/10.1257/aer.20190159.

Athey, Susan, and Guido W. Imbens. 2017. "The State of Applied Econometrics: Causality and Policy Evaluation." *Journal of Economic Perspectives* 31, no. 2 (May): 3–32. https://doi.org/10.1257/jep.31.2.3.

Bai, Jushan. 2009. "Panel Data Models With Interactive Fixed Effects." *Econometrica* 77, no. 4 (July): 1229–1279. https://doi.org/10.3982/ECTA6135.

Card, David. 1990. "The Impact of the Mariel Boatlift on the Miami Labor Market." *Industrial and Labor Relations Review* 43, no. 2 (January): 245–257. https://doi.org/10.2307/2523702.

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

Ibragimov, R., and Sh. Sharakhmetov. 2002. "The Exact Constant in the Rosenthal Inequality for Random Variables with Mean Zero." *Theory of Probability & Its Applications* 46, no. 1 (January): 127–132. https://doi.org/10.1137/S0040585X97978762.

Peri, Giovanni, and Vasil Yasenov. 2019. "The Labor Market Effects of a Refugee Wave: Synthetic Control Method Meets the Mariel Boatlift." *Journal of Human Resources* 54 (2): 267–309. https://doi.org/10.3368/jhr.54.2.0217.8561R1.