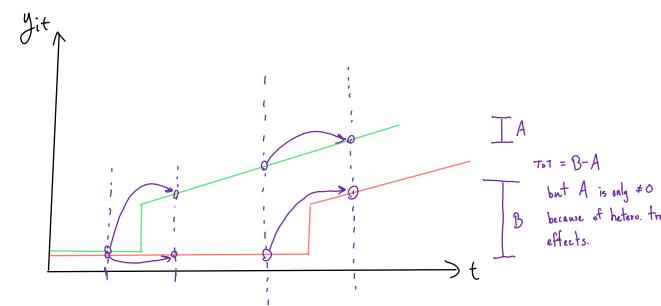
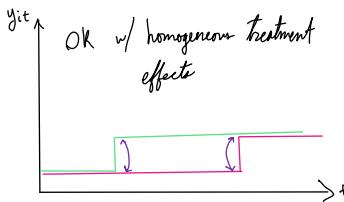


Diff in Diff issues

Goodman Bakon

→ issues w/ comparing Early treated
v Late treated groups



} will result in attenuation of treatment effects est.

$$y_{st} = \gamma_s + \delta_t + \beta D_{st} + \varepsilon_{st}$$

Partition reg:
that is equal to
TWFE reg

$$y_{st} = \alpha + \beta D_{st} + \varepsilon_{st}$$

$$\rightarrow y_{st} = \beta (\bar{D}_{st} - \bar{D}) + \varepsilon_{st}$$

$$\tilde{D}_{st} = D_{st} - \bar{D}_s - \bar{D}_t \quad (<0 \text{ for late periods for early treatment start group})$$

$$\text{resid from reg of } D_{st} \text{ on } s, t \text{ FE}$$

(<0 for late periods for early treatment start group)

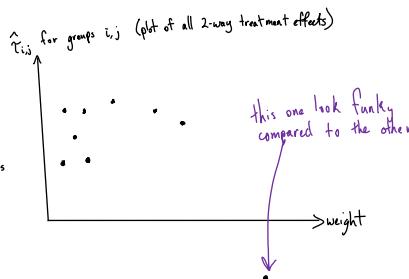
run reg. $y_{st} = \beta \tilde{D}_{st}$

→ we're effectively running a reg on
reg. treatment weight when running TWFE

Proposal

① Make a fig of weights:

look for 2-way group comparisons
that are outliers



② Think about effect heterogeneity

③ Think about dynamic treatment effects

→ probably should be est. event study instead

→ Sun & Abraham - Event Study - PS2b Fully-interacted Event Study

$$y_{st} = \sum_{j,c} \gamma_{ji} D_{jtc} + \gamma_i + \delta_t + \varepsilon_{st}$$

↳ Event time coef

↳ if state i was treated j periods ago as of time t
and is in treatment cohort c

j over treatment time cohort of obs
c = count time (t = cal time)

→ result: cohort-size-weighted γ for each t

ARE 213**Applied Econometrics****UC Berkeley Department of Agricultural and Resource Economics****SELECTION ON UNOBSERVABLES DESIGNS:****PART 3, CASE STUDIES WITH SYNTHETIC CONTROLS**

The Card (1990) and Card and Krueger (1994) papers are two examples of thoughtful diff-in-diffs studies, but they also highlight two central issues that affect much of the diff-in-diffs literature. First, in many cases there are multiple control units available for the researcher to choose from. For example, the choice of Eastern Pennsylvania as a control for Western New Jersey in the Card and Krueger (1994) study seems fairly obvious due to the direct geographic proximity of the two regions. The choice of Atlanta, Houston, Los Angeles, and Tampa as control cities for Miami in the Card (1990) study, however, seems somewhat more arbitrary. Though Card chose these cities because “they had relatively large populations of blacks and Hispanics and because they exhibited a pattern of economic growth similar to that in Miami over the late 1970s and early 1980s,” a different researcher might have chosen an entirely different set of control cities using a different (but still reasonable) algorithm. Second, as we have noted for both studies, the reported standard errors are not necessarily robust to the possibility of state-by-time (or city-by-time) specific shocks. For instance, perhaps New Jersey simply experienced some positive economic shock in late-1992 (the post-treatment period) – with only two observations on New Jersey, it’s impossible for us to even estimate what the variance of New Jersey’s statewide economic shocks might be.

1 Adadie et al. Synthetic Control Procedure

Abadie et al. (2010) present an estimation strategy – “synthetic controls” – that addresses both of these issues. At its heart, the synthetic controls strategy is basically a combination of diff-in-diffs and matching – control units are chosen based on how closely they resemble the

treated unit in the pre-treatment periods.¹ Furthermore, the strategy turns the large number of available control units into an advantage when estimating standard errors, because it's possible to study the variance of our estimator by constructing "placebo" estimates for units that were never treated.

Suppose that we observe a single treated unit and J control units over T time periods. Assume the policy intervention occurs at the end of period T_0 , so that periods $1, \dots, T_0$ are pre-intervention, and periods T_0+1, \dots, T are post-intervention. Let Y_{jt} represent the outcome of interest for unit j in period t , let D_{jt} represent the treatment, and let X_{jt} represent a set of K observed covariates. The question at hand is how to construct a synthetic control group for the treated unit out of the J potential control units.

Let $Z_1 = [\bar{X}_1, Y_{11}, Y_{1T_0/2}, Y_{1T_0}]'$ be a $K + 3 \times 1$ column vector of covariates and pre-intervention outcomes for the treated unit, where \bar{X}_1 contains the K observed covariates, each averaged over the pre-intervention time period. Let Z_0 be a $K + 3 \times J$ matrix that contains the same averaged covariates and pre-intervention outcomes for all J potential control units (each column corresponds to one of the control units). Our goal is to choose W^* , a $J \times 1$ column vector of weights. We will use these weights to combine all J control units into a single synthetic control unit against which to compare the treated unit. You could call Z_1 and Z_0 the "matching matrices" in the sense that they contain the variables that we are going to use to try to find the combination of control units that best matches the treated unit.

Choose W^* such that it minimizes the distance between Z_1 and Z_0W subject to $w_j^* \geq 0$ and $\sum_j w_j^* = 1$, where distance is defined as $\sqrt{(Z_1 - Z_0W)'V(Z_1 - Z_0W)}$ for some symmetric, positive semidefinite $(K + 3) \times (K + 3)$ matrix V .² In other words, choose the weights

¹Technically, some of the reanalyses of the LaLonde data are also combining diffs-in-diffs and matching. However, as I argued in an earlier footnote, I wouldn't necessarily consider the LaLonde paper to represent an archetype of a diffs-in-diffs paper, because it involves selection into treatment at the individual level rather than selection on the aggregate level.

²The weights need to be nonnegative and sum to one because, if they were not, it would be possible to perfectly fit Z_1 and Z_0W whenever there were more potential control units than elements in Z_1 . This also prevents extrapolation outside of the "support" (i.e., the convex hull) of Z_0 .

to minimize the distance between the treated unit's covariates (including pre-intervention outcomes) and the synthetic control unit's covariates (including pre-intervention outcomes). Note that this problem looks similar to estimating a regression of Z_1 on Z_0 with coefficients W ; we return to this interpretation in Section 2. For now, consider several likely candidates for the matrix V :

- (1) Set V equal to the inverse of a diagonal matrix in which the diagonal element in row i equals the variance of the covariate/pre-intervention outcome in row i of the Z matrices. This will minimize the normalized Euclidean distance.
- (2) Set V equal to the inverse of the variance-covariance matrix of the covariates/pre-intervention outcomes in Z . This will minimize Mahalanobis distance.
- (3) Set V to minimize the mean squared prediction error of the outcome variable during the pre-intervention period. Formally, let Y_1^p be a $T_0 \times 1$ column vector of pre-intervention outcomes for the treated unit, and let $\hat{Y}_0^p = Y_0^p W(V)$ be a $T_0 \times 1$ vector of pre-intervention outcomes for the synthetic control unit. Note that \hat{Y}_0^p is a function of W , which is itself a function of V . We will choose V such that the choice of V minimizes $(Y_1^p - \hat{Y}_0^p)'(Y_1^p - \hat{Y}_0^p)$. In other words, we choose V such that the resulting synthetic control provides the best fit for the pre-intervention outcome trajectory in the treated unit. In practice this means that we pick a value of V , solve for the W implied by that V , and see how closely the synthetic control unit produced by that W matches the pre-intervention outcomes of the treated unit. We keep doing this until we find the V that provides the best fit of the pre-intervention outcomes.³

The last algorithm is not trivial to implement; fortunately Abadie et al. provide code for implementing the synthetic control estimator in Stata (sort of), Matlab, and R on the web. Once you have estimated W and constructed the synthetic control unit $Y_0 W$, you could estimate a diffs-in-diffs model in which you compare the treated unit to the synthetic control

³This procedure seems somewhat convoluted in the sense that we could simply choose W to minimize $(Y_1^p - \hat{Y}_0^p)'(Y_1^p - \hat{Y}_0^p)$ to begin with. My guess is that the reason we don't do this is because Y_{jt} always contains some error, so leveraging the predictive power of the covariates can potentially give us a better out-of-sample forecast than just matching on pre-intervention outcomes and throwing away all the covariates.

unit. In practice, however, it is generally more informative to simply graph the outcomes for the treated unit against those for the synthetic control unit over all time periods, $1, \dots, T$.

Abadie et al. present a case study of California's Prop 99, passed at the end of 1988, which increased the cigarette tax and implemented other anti-tobacco measures. They apply the synthetic control group estimator to a data set that contains 30 years of data (18 year pre-intervention, 12 years post-intervention) and 50 states. After discarding from the control state "donor pool" other states that implemented tobacco control programs or raised their cigarette taxes by more than 50 cents, they retain 38 states. Estimating a synthetic control unit for California using 3 years of pre-intervention data (1975, 1980, and 1988) and a variety of covariates (e.g., retail cigarette price, per capita income, per capita beer consumption), Abadie et al. find positive weights for five states: Colorado, Connecticut, Montana, Nevada, and Utah. All other states receive a weight of zero.⁴

Figure 1 shows the evolution of cigarette sales in California versus Synthetic California. There is a clear break at the passage of Prop 99, but this break is somewhat deceptive. We constructed the synthetic control unit so that it tracked the treated unit closely in the pre-intervention period, so of course the two units will diverge more in the post-intervention period than they do in the pre-intervention period, even if the treatment has no effect. Fortunately, the plethora of untreated control units gives us a sample with which to conduct statistical inference and determine whether the post-intervention divergence is significant or not.

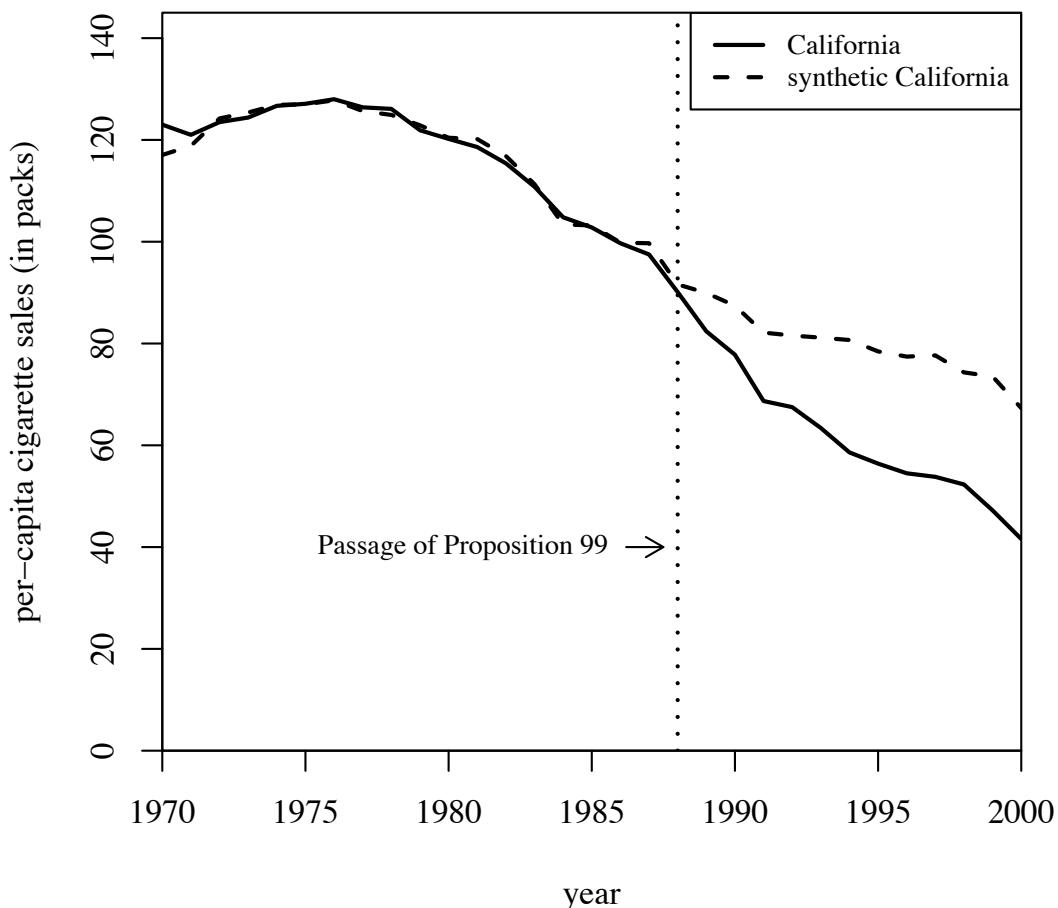
Abadie et al. suggest a variant on the exact permutation test (which we will discuss in greater detail when we talk about standard errors). Suppose that we randomly chose to implement Prop 99 in California. In this scenario, there is no bias in our estimate of the effect of Prop 99 – our only question is whether the divergence we observe between California and Synthetic California represents a real treatment effect or is simply due to chance. The key is that the variation in our results can be conceptualized as arising from the variation in the treatment assignment. In our world, we happened to assign Prop 99 to California,

⁴This may seem surprising until you consider that weights are always constrained to be zero or greater.

but in alternative worlds we could have assigned it to Delaware or Texas or Wyoming. If we assume that Prop 99 had no effect (the null hypothesis), we can pretend to assign it to these other states that were, in actuality, untreated. In doing so, we can map out the distribution of the estimator under the null hypothesis. This is exactly what Abadie et al. do in Figure 2.

Figure 2 graphs the difference between the “treated” state and its synthetic control for all 38 states in the study. The black line represents California (the actual treated state), while the gray lines represent all of the control states. These states received a “placebo” treatment in the sense that we pretended that they were treated, and estimated the synthetic control estimator as if they were, but in actuality they received no treatment. The fact that

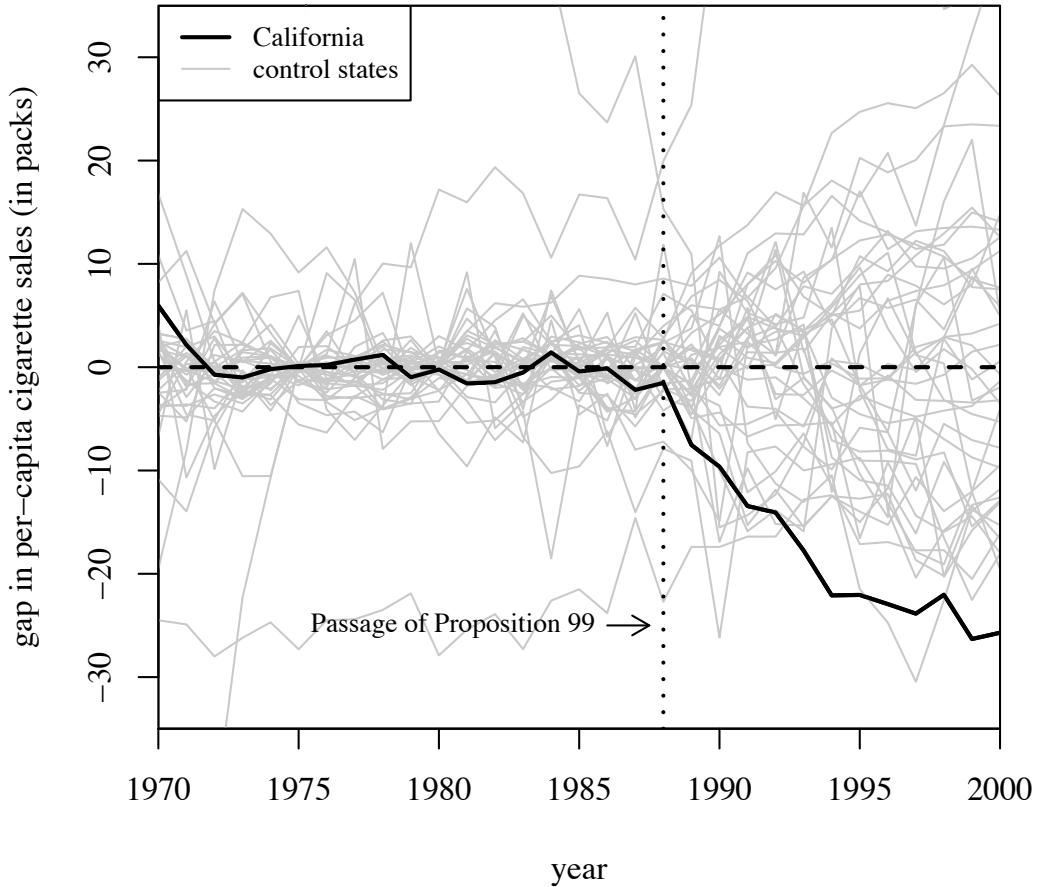
Figure 1: Per-Capita Cigarette Sales in CA vs. Synthetic CA. Source: Abadie et al. (2010)



the California line is at or near the bottom during the post-intervention period strongly suggests that these results are not simply due to chance. If we had randomly picked an untreated state and implemented the same procedure, it is unlikely that we would have found a post-intervention deviation of this magnitude.

To conduct a formal test, Abadie et al. compute the mean squared prediction error (i.e., the mean squared gap between the “treated” state and its synthetic control) for the pre- and post-intervention periods for all 38 states plus California. They find that the ratio of the post-intervention mean squared prediction error and the pre-intervention mean squared prediction error is higher for California than for any other state. Thus they reject the null hypothesis of no treatment effect at $p = 0.026$.

Figure 2: Per-Capita Cigarette Sales Gap: (Actual - Synthetic Control). Source: Abadie et al. (2010)



2 Alternative Procedures

One issue with the Abadie et al. estimator is that the synthetic control unit needs to match the treated unit on both levels and trends. Thus, if the treated unit happens to be a unit with an exceptionally high (or low) level of Y , the only control units that can receive positive weight are those with very high (or low) levels of Y . This is undesirable in that some control units may be better at matching the trends in the treated unit, but they can't receive positive weight unless they also have high levels of Y . From a diffs-in-diffs perspective, we may be discarding units that are good matches for satisfying the parallel trends assumption simply because they aren't good matches in terms of their baseline levels of Y .

Doudchenko and Imbens (2017) develop a general framework that nests the Abadie et al. synthetic control estimator. Recall that the goal with synthetic controls is to find a set of weights W that minimizes the distance between Z_1 and Z_0W , where Z is a set of pre-intervention measures. One way to achieve this fit would be to run a regression of Z_1 on Z_0 — the regression coefficients would give you the weights. But this regression fails the rank condition if the number of potential control units exceeds the number of variables in Z (as is often the case); in that case Z_0W can perfectly fit Z_1 . This overfitting problem suggests either placing restrictions on W , which is what Abadie et al. do (e.g. weights must be non-negative and sum to one), or regularization (i.e. penalizing overfitting, as in machine learning). Doudchenko and Imbens propose the latter solution.

Instead of regressing Z_1 on Z_0 using least squares, Doudchenko and Imbens suggest regressing Z_1 on Z_0 using an elastic net regression. This estimator is similar to the LASSO that we discussed previously, but its regularization term includes both the absolute values of the coefficient estimates (like LASSO) and the squares of the coefficient estimates. Letting μ be a column vector of identical constants, elastic net minimizes wrt W :

$$(Z_1 - \mu - Z_0W)'(Z_1 - \mu - Z_0W) + \lambda \cdot \sum_{i=1}^J (\frac{1-\alpha}{2} W_i^2 + \alpha |W_i|)$$

With regularization it is no longer necessary to impose restrictions on W , though in

practice one may still do so.

In the regression framework the following potential restrictions seem reasonable:

1. No intercept ($\mu = 0$)
2. Coefficients (weights) must sum to one.
3. Coefficients (weights) must be non-negative.
4. Pre-intervention trends in the outcome should exactly match (up to a constant) between treatment and synthetic control.

Abadie et al. impose the first three restrictions (and weight towards the fourth), while Doudchenko and Imbens propose regularization in combination with some subset of the restrictions. Allowing a non-zero intercept may seem appealing for addressing our earlier criticism that synthetic controls requires the composite control unit to match the treated unit on both levels and trends. Indeed, when matching on pre-intervention outcomes, one should generally allow for a non-zero intercept (as noted in Restriction 4 above). But it's less clear that one should allow for a non-zero intercept when matching across all covariates — what would it mean for CA to be equal to $2 + 0.5 \cdot NV + 0.5 \cdot CT$?

Doudchenko and Imbens do not give any hard rules on which restrictions to impose; they merely recommend choosing “on subjective grounds, a subset of the five restrictions.”⁵ Like the digital conveyer and kernel density estimation, it is seemingly more an art than a science. Even ignoring restrictions, however, one has to choose values for the elastic net tuning parameters, α and λ . Doudchenko and Imbens suggest a cross-validation procedure in which each control unit is coded as the “treated” unit. For each unit, run an elastic net regression, with tuning parameters α and λ , to predict the (placebo) treated unit using a synthetic control, with weights given by the regression. Record how well the prediction does in the final period — keep in mind that this “treated” unit is not actually treated, so the

⁵The fifth restriction they propose is that weights are constant across units. Presumably if you impose this restriction, plus the restriction that weights sum to one, your job is done.

prediction should be relatively accurate. Repeat this procedure for each control unit, a finite range of α values (0.1, 0.2, ..., 0.9), and a wide range of λ values. Select the values of α and λ that minimize squared prediction error in the final period across all control units. Then use these values of α and λ in an elastic net regression that estimates the weights for the actual treated unit's synthetic control.

3 Conclusion

To recap, the synthetic control estimator has two nice properties that can augment many diff-in-diffs designs. First, it provides a more rigorous, less ad-hoc way of selecting control units from a large pool of potential controls. Second, it leverages the large pool of potential controls to conduct permutation-based inference in a manner that is robust to the possibility of unit-by-time period specific shocks. In other words, it accounts for the fact that, even if we observed the entire population for each unit (e.g., state, city, etc.), there would still be some deviation between the treated unit and its synthetic control because there are aggregate (i.e., unit-level) shocks that occur at the unit-by-time level.

4 Additional References

Doudchenko, Nikolay, and Guido Imbens. "Balancing, Regression, Difference-in-differences and Synthetic Control Methods: A Synthesis." NBER Working Paper, 2017.