Part 8: Program Evaluation (f): Synthetic Control Methods

Chris Conlon

April 9, 2021

Applied Econometrics

Motivation: Recap

Difference in Difference approaches have some drawbacks:

- We need to really believe parallel trends
    - Is ΔPA really a good counterfactual for ΔNJ?
    - Obvious question: why not pick ΔDE or ΔNY?
    - Measured effect shouldn't change if our assumption is valid (but it probably will!)
- With multiple treated individuals we can use 2WFE estimator

$$y_{it} = \beta X_{it} + \delta_i \cdot T_{it} + \gamma_i + \gamma_t + u_{it}$$

- It assumes additivity of FE $\gamma_i + \gamma_t$ are independent!
- States have different baseline levels γ_i but evolution over time is common γ_t.
- Turns out it doesn't really generalize the diff-in-diff with $I > 2$ (Imai and Kim 2020).

Motivation: Recap

Matching Estimators had drawbacks too:

- We need to believe in as if random-assignment conditional on X
- CIA: $\{Y_i(1), Y_i(0)\} \perp T_i | X_i$.

But we could use pretty flexible methods in constructing our matched controls:

- We could try to match on multiple dimensions of X.
- k–NN, kernels, etc.

What if we could use ideas from matching to better satisfy something like our parallel trends assumptions?

Example: Abadie, Diamond, Hainmueller (2010)

The Question

In 1988 California passes anti-smoking Prop 99

- increased excise tax by 25 cents per pack,
- earmarked the tax revenues to health and anti-smoking education budgets and funded anti-smoking ads
- led to indoor smoking bans in restaurants and bars city by city

What was the effect on per capita cigarette sales?

- Already a bunch of pre-existing trends.
- What is a good control for California?

Use state-level data from 1970-2000.

The Idea

Use a convex combination of other states to construct a synthetic counterfactual California.

- We observe Y_{it}, X_{it}, T_{it}.
- Assume only $i = 1$ and $t > T_0$ are treated.
- Construct a donor pool of potential controls subscripted by j.
- Choose some weights w_j for each entity (state) in donor pool. How?
    - Same $\mathbf{X_1} = \sum_j w_j \mathbf{X_j}$ as treated observations (like matching).
    - Same $\left(Y_{1,1}, \ldots, Y_{1,T_0}\right) = \sum_j w_j \cdot \left(Y_{j,1}, \ldots, Y_{j,T_0}\right)$ (like parallel trends).
    - Weights sum to one $\sum_j w_j = 1$ and maybe are non-negative (or not!)
- Idea is to match all of the X's and all of the Y_{it}'s in the pre-period

Table 1. Cigarette sales predictor means

Variables	California Real	California Synthetic	Average of 38 control states
Ln(GDP per capita)	10.08	9.86	9.86
Percent aged 15–24	17.40	17.40	17.29
Retail price	89.42	89.41	87.27
Beer consumption per capita	24.28	24.20	23.75
Cigarette sales per capita 1988	90.10	91.62	114.20
Cigarette sales per capita 1980	120.20	120.43	136.58
Cigarette sales per capita 1975	127.10	126.99	132.81

NOTE: All variables except lagged cigarette sales are averaged for the 1980–1988 period (beer consumption is averaged 1984–1988). GDP per capita is measured in 1997 dollars, retail prices are measured in cents, beer consumption is measured in gallons, and cigarette sales are measured in packs.

Donor Weights

Table 2. State weights in the synthetic California

State	Weight	State	Weight
Alabama	0	Montana	0.199
Alaska	–	Nebraska	0
Arizona	–	Nevada	0.234
Arkansas	0	New Hampshire	0
Colorado	0.164	New Jersey	–
Connecticut	0.069	New Mexico	0
Delaware	0	New York	–
District of Columbia	–	North Carolina	0
Florida	–	North Dakota	0
Georgia	0	Ohio	0
Hawaii	–	Oklahoma	0
Idaho	0	Oregon	–
Illinois	0	Pennsylvania	0
Indiana	0	Rhode Island	0
Iowa	0	South Carolina	0
Kansas	0	South Dakota	0
Kentucky	0	Tennessee	0
Louisiana	0	Texas	0
Maine	0	Utah	0.334
Maryland	–	Vermont	0
Massachusetts	–	Virginia	0
Michigan	–	Washington	–
Minnesota	0	West Virginia	0
Mississippi	0	Wisconsin	0
Missouri	0	Wyoming	0

Trend Check and Treatment Effects

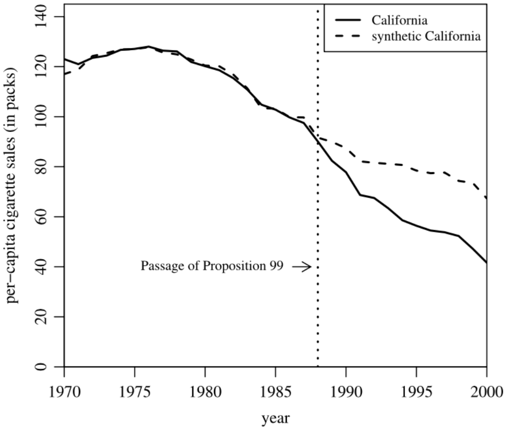
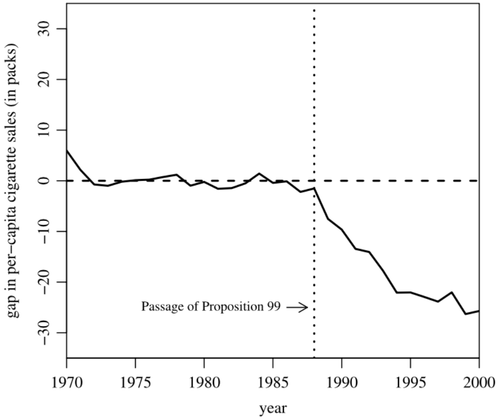


Figure 2. Trends in per-capita cigarette sales: California vs. synthetic California.



But still some issues

- How sensitive are weights estimates to different covariates?
  - "state-level measures of unemployment, income inequality, poverty, welfare transfers, crime rates, drug related arrest rates, cigarette taxes, population density, and numerous variables to capture the demographic, racial, and social structure of states".
- Can we run a placebo check? Do we detect effects where we know there is a null effect?
  - Put California in the donor pool.
  - Pick a state from the donor pool at pretend that receives the treatment after T_0
  - Choose w_j following the synthetic control procedure.
  - Compute the treatment effects in the same way.
  - Repeat for all states in donor pool.
  - Compare mean-square prediction error (MSPE) for $(Y_{1,1}, \ldots, Y_{1,T_0})$
  - This doubles as inference.

Placebo Test

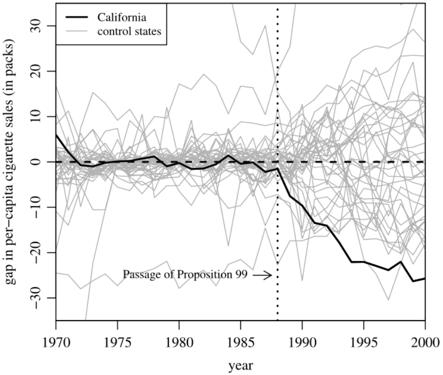


Figure 4. Per-capita cigarette sales gaps in California and placebo gaps in all 38 control states.

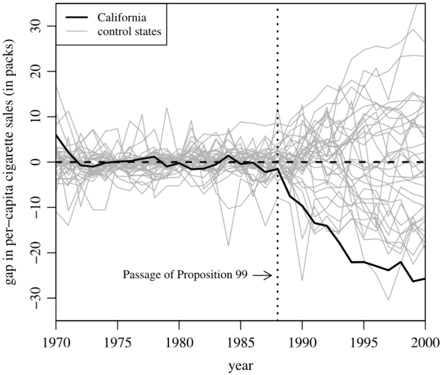


Figure 5. Per-capita cigarette sales gaps in California and placebo gaps in 34 control states (discards states with pre-Proposition 99 MSPE twenty times higher than California's).

More Placebo Tests: How unusual is the CA result?

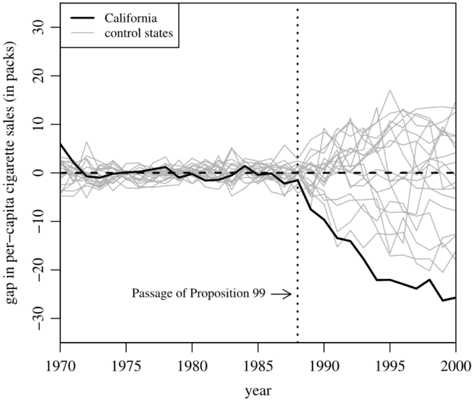


Figure 7. Per-capita cigarette sales gaps in California and placebo gaps in 19 control states (discards states with pre-Proposition 99 MSPE two times higher than California's).

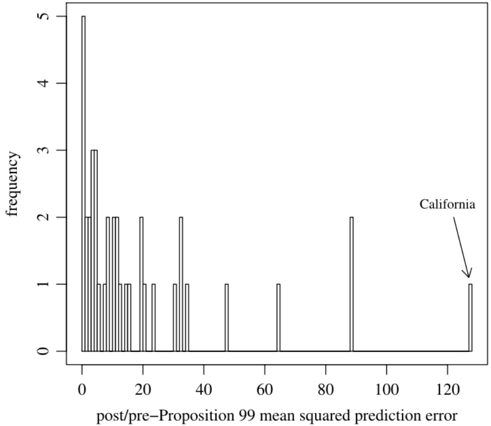


Figure 8. Ratio of post-Proposition 99 MSPE and pre-Proposition 99 MSPE: California and 38 control states.

More Formally

We're going to break this down into three parts:

1. How do we estimate w_j?
2. Conditional on $\hat{w}_j$'s, how do we estimate the treatment effect?
3. What kinds of models are compatible?

A Starting Point

Start with $T_{it} = 1$ IFF $i = 1$ and $t > T_0$:

$$Y_{it} = Y_{it}(0) + \alpha_{it} T_{it}$$

Estimate $\boldsymbol{\alpha_1} = (\alpha_{1,T_0+1}, \alpha_{1,T_0+2}, \ldots, \alpha_{1,T})$ (period by period treatment effect).
For $t > T_0$:

$$\alpha_{1t} = Y_{1t}(1) - Y_{1t}(0) = \underbrace{Y_{1t}}_{\text{observed}} - \underbrace{Y_{1t}(0)}_{\text{counterfactual}}$$

A Starting Point

Consider the model

$$Y_{it}(0) = \gamma_t + \theta_t \mathbf{Z}_i + \lambda_t \mathbf{F}_i + u_{it}$$

- δ_t is the time fixed effect
- Z_i are the usual covariates (fixed in i over time) with potentially time varying coefficients.
- F_i are i specific unobserved factors. Estimating these are complicated.
- λ_t are called factor loadings.

This is a generalization of the usual additive $\gamma_t + \gamma_i$ fixed effects model called interactive fixed effects (Bai 2009).

Some Algebra

So we define the synthetic observation as:

$$\sum_{j=2}^{J+1} w_j \underbrace{Y_{jt}(0)}_{\text{observed}} = \delta_t + \boldsymbol{\theta}_t \left(\sum_{j=2}^{J+1} w_j \mathbf{Z}_j \right) + \lambda_t \left(\sum_{j=2}^{J+1} w_j \boldsymbol{F}_j \right) + \sum_{j=2}^{J+1} w_j u_{jt}$$

And the difference in untreated observations:

$$Y_{1t}(0) - \sum_{j=2}^{J+1} w_j Y_{jt}(0)$$

$$= \underbrace{\boldsymbol{\theta}_t \left(\mathbf{Z}_1 - \sum_{j=2}^{J+1} w_j \mathbf{Z}_j \right)}_{\text{match this directly}} + \lambda_t \underbrace{\left(\boldsymbol{F}_1 - \sum_{j=2}^{J+1} w_j \boldsymbol{F}_j \right)}_{\text{match indirectly: } (Y_{i,1},\ldots,Y_{i,T_0})} + \sum_{j=2}^{J+1} w_j \underbrace{(u_{1t} - u_{jt})}_{\text{Usual} E[u_{it}|\mathbf{Z}_{it},\mathbf{F}_{it}]=0}$$

Some Algebra

The goal is to match:

$$\mathbf{Z}_1 = \sum_{j=2}^{J+1} w_j \mathbf{Z}_j, \quad \mathbf{F}_1 \approx \sum_{j=2}^{J+1} w_j \mathbf{F}_j$$

- The main result in the Abadie et. al (2010) paper proves that the above equation holds approximately if

$$(Y_{1,1}, \ldots, Y_{1,T_0}) = \sum_j w_j \cdot (Y_{j,1}, \ldots, Y_{j,T_0})$$

- This frame generalizes the 2WFE model since we can use time dimension to difference them out.
- We cannot difference out the $\lambda_t \mathbf{F}_i$.

The Challenge

Theory depends on matching:

$$\mathbf{Z}_1 = \sum_{j=2}^{J+1} w_j \mathbf{Z}_j, \quad (Y_{1,1}, \ldots, Y_{1,T_0}) = \sum_j w_j \cdot (Y_{j,1}, \ldots, Y_{j,T_0})$$

- One problem is the convex hull problem.
  - There may be no $w_2, \ldots, w_J$ that satisfy the equations
- Instead choose $\mathbf{w}$ in a minimum distance sense with some weighting matrix Ω
  - Can choose Ω to minimize variance of the resulting estimator.
- Hard to fit crazy out of bounds values: can't match CA population as linear combination of other states!

Other Pitfalls

Other things can go wrong

$$\mathbf{X}_i = (\mathbf{Z}_i, Y_{i,1}, \ldots, Y_{i,T_0})$$

$$\arg\min_{w_2,\ldots,w_J} (\mathbf{X}_i - \sum_j w_j \mathbf{X}_j)' \Omega (\mathbf{X}_i - \sum_j w_j \mathbf{X}_j) \quad \text{s.t. } \sum_j w_j = 1$$

- Can get negative weights: what do those mean?
- We can rule them out $w_j \geq 0$ constraints but the problem becomes more difficult to solve.
- This starts to look like a familiar LASSO problem: quadratic objective, L_1 penalty!
- We shouldn't be surprised by sparse models.

In practice

- The R package `synth` and accompanying examples are a good start.
- The R package `MSCMT` fixes some of the optimization issues in `synth`.
  - When there is no set of feasible weights and the non-negativity constraint binds – the problem is tricky to solve
  - `synth` can produce sub-optimal solutions (and slowly).

A more unified view

We can ask – what are the implied regression weights for linear regression? What is $E[Y_{1,t}(0)|\mathbf{X}]$?

- Get coefficients using just donor pool: $\widehat{\boldsymbol{B}} = \left(\overline{\boldsymbol{X}}_0 \overline{\boldsymbol{X}}_0'\right)^{-1} \overline{\boldsymbol{X}}_0 \boldsymbol{Y}_0'$
- Predict $Y_{1,t}(0) = \widehat{\boldsymbol{B}}' \overline{\boldsymbol{X}}_1$.
- Same as synthetic control but with weights: $\boldsymbol{W}^{reg} = \overline{\boldsymbol{X}}_0' \left(\overline{\boldsymbol{X}}_0 \overline{\boldsymbol{X}}_0'\right)^{-1} \overline{\boldsymbol{X}}_1$ (Projection matrix).

Why? Remember, everything is a kernel our prediction of $E[Y|X]$ is always some weighted average of other Y_j's.

Additonal Applications

- What is word of mouth effect of Superbowl ads? (Lovett, Peres, Xu QME 2019)
  - Construct synthetic versions of brands that don't buy Superbowl ads
- What is effect of soda tax on consumption in Berkeley, CA ? (Bollinger Sexton WP 2019).
  - Construct synthetic drugstores, supermarkets, and convenience stores using two donor polls: inside CA and outside CA.
- What are economic effects of German reunification?
  - Now we need a counterfactual "West Germany" (!)
    (Abadie, Diamond, and Hainmueller 2015)

Weighting

Table 1 Synthetic and Regression Weights for West Germany

Country	Synthetic Control Weight	Regression Weight	Country	Synthetic Control Weight	Regression Weight
Australia	0	0.12	Netherlands	0.09	0.14
Austria	0.42	0.26	New Zealand	0	0.12
Belgium	0	0	Norway	0	0.04
Denmark	0	0.08	Portugal	0	−0.08
France	0	0.04	Spain	0	−0.01
Greece	0	−0.09	Switzerland	0.11	0.05
Italy	0	−0.05	United Kingdom	0	0.06
Japan	0.16	0.19	United States	0.22	0.13

Notes: The synthetic weight is the country weight assigned by the synthetic control method. The regression weight is the weight assigned by linear regression. See text for details.

Balance Table

TABLE 2 Economic Growth Predictor Means before German Reunification

	West Germany	Synthetic West Germany	OECD Sample
GDP per capita	15808.9	15802.2	8021.1
Trade openness	56.8	56.9	31.9
Inflation rate	2.6	3.5	7.4
Industry share	34.5	34.4	34.2
Schooling	55.5	55.2	44.1
Investment rate	27.0	27.0	25.9

Notes: GDP per capita, inflation rate, trade openness, and industry share are averaged for the 1981–90 period. Investment rate and schooling are averaged for the 1980–85 period. The last column reports a population-weighted average for the 16 OECD countries in the donor pool.

Treatment Effects

FIGURE 1 Trends in per Capita GDP: West Germany versus Rest of the OECD Sample

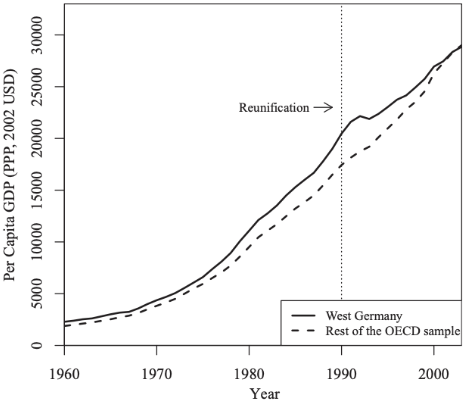
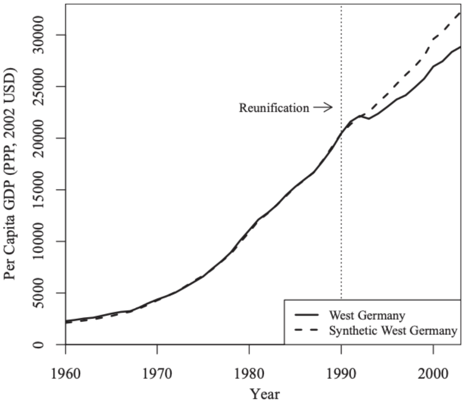


FIGURE 2 Trends in per Capita GDP: West Germany versus Synthetic West Germany



Treatment Effects and Placebo Timing

FIGURE 3 Per Capita GDP Gap between West Germany and Synthetic West Germany

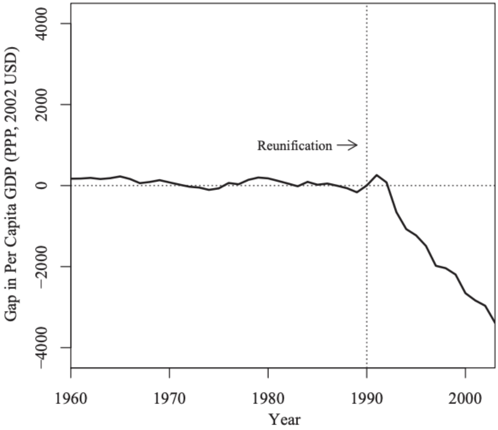
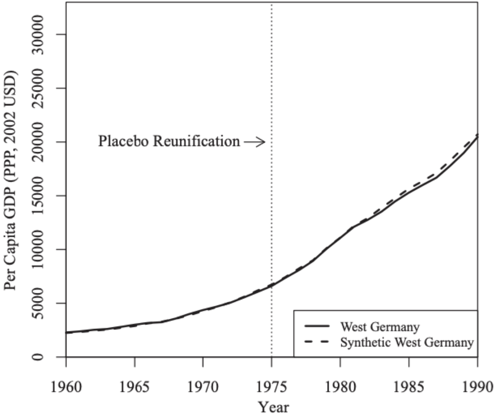
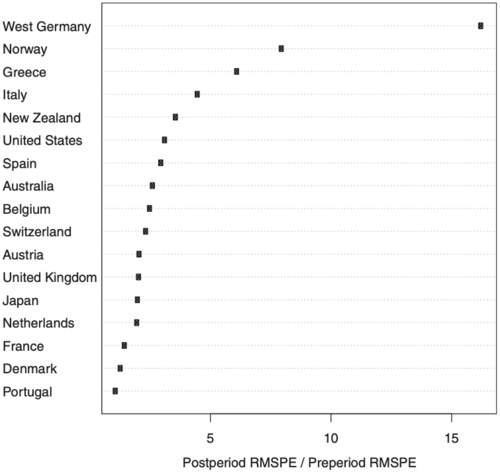


FIGURE 4 Placebo Reunification 1975–Trends in per Capita GDP: West Germany versus Synthetic West Germany



Treatment Effects and Placebo Timing

FIGURE 5 Ratio of Postreunification RMSPE to Prereunification RMSPE: West Germany and Control Countries



Thanks