Applied Microeconometrics II , Lecture 5

Ciprian Domnisoru
Aalto University

Outline

- Motivational look back at omitted variable bias
- Synthetic control method
- Summarizing DD robustness checks.

Omitted variable bias

- Omitted variable bias arises if an omitted variable is both: (i) a determinant of Y and (ii) correlated with at least one included regressor.
- When reviewing research:
  - identify a plausible candidate for the omitted variable
  - predict the direction of the bias, based on its expected correlation with X and Y
  - demonstrate the effect by controlling for the omitted variable or a proxy, and showing how results change
- two-way fixed effects/ difference in differences estimators do a great deal to alleviate omitted variable bias

OVB example

```
. reg fatalityrate beertax, robust

Linear regression                               Number of obs     =        336
                                                F(1, 334)         =      47.59
                                                Prob > F          =     0.0000
                                                R-squared         =     0.0934
                                                Root MSE          =     .54374

------------------------------------------------------------------------------
             |               Robust
fatalityrate |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     beertax |   .3646054   .0528524     6.90   0.000     .2606399     .468571
       _cons |   1.853308   .0471297    39.32   0.000     1.760599    1.946016
------------------------------------------------------------------------------
```

OVB example

```
. reg fatalityrate beertax year2-year7 state2-state48, cluster(state)

Linear regression                               Number of obs     =        336
                                                F(6, 47)          =          .
                                                Prob > F          =          .
                                                R-squared         =     0.9089
                                                Root MSE          =     .18788

                                  (Std. Err. adjusted for 48 clusters in state)
------------------------------------------------------------------------------
             |               Robust
fatalityrate |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     beertax |  -.6399799   .3857867    -1.66   0.104    -1.416083    .1361229
       year2 |  -.0799029   .0379069    -2.11   0.040    -.1561617    -.003644
       year3 |  -.0724206   .0474088    -1.53   0.133    -.1677948    .0229537
       year4 |  -.1239763   .0497587    -2.49   0.016    -.2240779   -.0238747
       year5 |  -.0378645   .0616479    -0.61   0.542    -.1618841    .0861552
       year6 |  -.0509021   .0687224    -0.74   0.463    -.1891536    .0873495
       year7 |  -.0518038   .0695801    -0.74   0.460    -.1917809    .0881732
      state2 |  -.5468622   .5064424    -1.08   0.286    -1.565693    .4719685
      state3 |  -.6385298   .3986016    -1.60   0.116    -1.440413    .1633531
      state4 |  -1.485192   .5892726    -2.52   0.015    -2.670655   -.2997283
      state5 |  -1.461534   .5521075    -2.65   0.011    -2.572231   -.3508375
```

OVB example

```
. reg fatalityrate beertax year2-year7 state2-state48, robust

Linear regression                               Number of obs     =        336
                                                F(54, 281)        =     128.89
                                                Prob > F          =     0.0000
                                                R-squared         =     0.9089
                                                Root MSE          =     .18788

------------------------------------------------------------------------------
             |               Robust
fatalityrate |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     beertax |  -.6399799   .2547149    -2.51   0.013    -1.141371   -.1385884
       year2 |  -.0799029   .0502708    -1.59   0.113    -.1788579    .0190522
       year3 |  -.0724206   .0452466    -1.60   0.111     -.161486    .0166448
       year4 |  -.1239763   .0460017    -2.70   0.007     -.214528   -.0334246
       year5 |  -.0378645   .0486527    -0.78   0.437    -.1336344    .0579055
       year6 |  -.0509021   .0516113    -0.99   0.325    -.1524958    .0506917
       year7 |  -.0518038     .05387    -0.96   0.337    -.1578438    .0542361
      state2 |  -.5468622   .3569504    -1.53   0.127    -1.249498     .155774
      state3 |  -.6385298    .292595    -2.18   0.030    -1.214486   -.0625735
      state4 |  -1.485192   .4121171    -3.60   0.000     -2.29642   -.6739631
```

Omitted Variable Bias

- The appropriate equation is a *long regression*

$$Y_i = \alpha^l + \beta^l P_i + \gamma^l X_i + e_i^l$$

but instead we estimate a *short regression* leaving out X_i

$$Y_i = \alpha^s + \beta^s P_i + e_i^s$$

- In the *short regression* we have bias since we don't account for selection
- The *omitted variable bias* (OVB) formula shows

$$\beta^s = \beta^l + \pi_1 \gamma^l$$

where π_1 is the coefficient of the regression

$$X_i = \pi_0 + \pi_1 P_i + u_i$$

- Essentially in the *short* regression β^s estimates the effect of P_i *including* effects that result from X_i.

Omitted Variable Bias

- The OLS estimator of β^s is

$$\beta^s = \frac{Cov(Y_i, P_i)}{Var(P_i)}$$

Substitutes in the "true" long regression for $\ln Y_i$

$$\begin{aligned}\beta^s &= \frac{Cov(\alpha^l + \beta^l P_i + \gamma^l X_i + e_i^l, P_i)}{Var(P_i)} \\ &= \frac{\beta^l Var(P_i) + \gamma^l Cov(X_i, P_i) + Cov(e_i^l, P_i)}{Var(P_i)} \\ &= \beta^l + \frac{Cov(X_i, P_i)}{Var(P_i)}\gamma^l \\ &= \beta^l + \pi_1 \gamma^l\end{aligned}$$

  - *Angrist and Pishcke: short equals long plus the effect of omitted times the regression of omitted on included*

Omitted Variable Bias in diff-in-diff context

- The appropriate equation is a *long regression*

$$Y_{it} = \alpha^l + \beta^l P_{it} + \gamma^l X_{it} + \delta_t + \lambda_i + e^l_{it}$$

but instead we estimate a *short regression* leaving out X_{it}

$$Y_{it} = \alpha^s + \beta^s P_{it} + \delta_t + \lambda_i + e^s_{it}$$

- In the difference in differences context, a variable correlated with the treatment effect as well as the outcome would be a failure of the paralell trends assumption- selection into treatment / bad controls.
- In two-way fixed effects models without quasi-exogeneous variation, correlation of the independent variable with time-varying covariates represent OVB bias.
- Other mechanical biases remain, particularly when panels are unbalanced and treatments heterogeneous over time within unit.

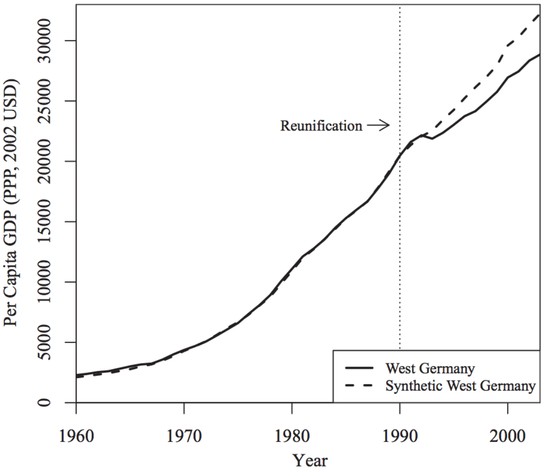
The Synthetic Control Method

Synthetic Control Method

- The synthetic control method provides a systematic way to choose comparison units in *small-sample* comparative case studies
- Distinctive feature of comparative studies: units of analysis are often aggregate entities – countries, regions, states – for which suitable single comparisons often do not exist
- a combination of comparison units – the "synthetic control" – often does a better job of reproducing the characteristics of a treated unit than any single comparison unit alone
- comparison unit is selected as the weighted average of all potential comparison units that best resembles the characteristics of the case of interest
- Example: Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2015. "Comparative Politics and the Synthetic Control Method." *American Journal of Political Science*, 59(2): 495-510.

Application: The Impact of the 1990 German Reunification

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



The Synthetic Control Method: Set-up

- Suppose that there is a sample of $J + 1$ units (e.g., countries) indexed by j
  - unit $j = 1$ is the case of interest, or the “treated unit”
    - the unit exposed to the event or intervention of interest
  - units $j = 2$ to $j = J + 1$ are potential comparisons
    - they constitute the “donor pool”, a reservoir of potential comparison units
    - comparison units are meant to approximate the counterfactual of the treated unit in the absence of intervention
    - thus, it is important to restrict the donor pool to units with outcomes that are thought to be driven by the same structural process as for the treated unit
  - in the application, we will investigate the effects of the 1990 German reunification on the economic prosperity in West Germany
    - and the set of potential comparisons is a sample of OECD countries

The Synthetic Control Method: Set-up

- We assume that the sample is a balanced panel
  - a longitudinal dataset where all units are observed at the same time periods, $t = 1, ..., T$
- We also assume that the sample includes
  - a positive number of preintervention periods, T_0
  - a positive number of postintervention periods, T_1
  - $T = T_0 + T_1$
- The goal of the study is to measure the effect of the intervention of interest on some postintervention outcome
  - unit 1 is exposed to the intervention ("treatment") during periods $T_0 + 1, ..., T$
  - and the intervention has no effect during the pretreatment period $1, ..., T_0$

The Synthetic Control Estimator

- We define a synthetic control as a weighted average of the units in the donor pool
  - $0 \leq w_j \leq 1$ for $j = 2, ...J$, and $w_2 + ... + w_{J+1} = 1$
  - a synthetic control can be represented by a $(J \times 1)$ vector of weights $W = (w_2, ..., w_{J+1})'$
- Choosing a particular value for W is equivalent to choosing a synthetic control
  - the method selects the value of W such that the characteristics of the treated unit are best resembled by the characteristics of the synthetic control

The Synthetic Control Estimator

- Let X_1 be a $(k \times 1)$ vector containing the values of the preintervention characteristics of the treated unit that we aim to match as closely as possible
- Some risk of specification search: Ferman, Bruno, Cristine Pinto, and Vitor Possebom. 2020. "Cherry Picking with Synthetic Controls." Journal of Policy Analysis and Management.
- Let X_0 be the $(k \times J)$ matrix collecting the values of the same variables for the units in the donor pool
  - the preintervention characteristics in X_1 and X_0 may include preintervention values of the outcome variable
- The difference between the preintervention characteristics of the treated unit and a synthetic control is given by the vector $X_1 - X_0 W$
  - the synthetic control, W^*, minimizes the size of this difference

Synthetic control: finding the weights

- Minimizing the quantity X1-WX_0

$$\|\boldsymbol{X}_1 - \boldsymbol{X}_0\boldsymbol{W}\| = \left(\sum_{h=1}^{k} v_h \left(X_{h1} - w_2 X_{h2} - \cdots - w_{J+1} X_{hJ+1}\right)^2\right)^{1/2}$$

- Notice an additional v weight: that is the weight we place on various characteristics X (as opposed to the weights placed on control units J).
- The weights v can be chosen based on evaluating how good of a predictor of Y_{1t} they are [subjective assessment of predictive power of X, regression, cross-validation].
- **synth** command in Stata/Matlab/R, see criticism: Kuosmanen et al. Design Flaw of the Synthetic Control Method. Working paper.

The Synthetic Control Estimator

- Let Y_{jt} be the outcome of unit j at time t
  - let Y_1 be a $(T_1 \times 1)$ vector collecting the postintervention outcome values for the treated unit: $Y_1 = (Y_{1,T_0+1}, ..., Y_{1,T})'$
  - similarly, let Y_0 be a $(T_1 \times J)$ matrix, where column j contains the postintervention outcomes values for unit $j+1$
- The synthetic control estimator of the effect of the treatment is given by the comparison $Y_1 - Y_0 W^*$
  - comparison between postintervention outcomes between the treated unit, which is exposed to the intervention, and the synthetic control, which is not exposed to the intervention
- Alternatively, for a postintervention period t (with $t \geq T_0$), the synthetic control estimator of the effect of the treatment is given by the comparison between the outcome for the treated unit and the outcome for the synthetic control at that period

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

The Synthetic Control Estimator

- The matching variables in X_0 and X_1 are meant to be predictors of postintervention outcomes
  - the may include preintervention values of the outcome variable
- Abadie, Diamond, and Hainmueller (2010) argue that if the number of preintervention periods in the data is large, matching on preintervention outcomes also controls for time-varying unobserved heterogeneity on the outcome
  - *intuition*: only units that are alike in both observed and unobserved determinants of the outcome variable should produce similar trajectories of the outcome variable over extended periods of time

Application: The Impact of the 1990 German Reunification

- After the fall of the Berlin Wall on Nov. 9, 1989, the German Democratic Republic and the Federal Republic of Germany ("West Germany") officially reunified on Oct. 3, 1990
  - Abadie, Diamond, and Hainmueller (2015) examines the impact on per capita GDP in West Germany
- Data: annual country-level panel data for the period 1960-2003
  - dataset allows for a preintervention period of 30 years
- The synthetic West Germany is constructed as a weighted average of potential control countries in the donor pool
  - donor pool includes a sample of 16 OECD member countries: Australia, Austria, Belgium, Denmark, France, Greece, Italy, Japan, the Netherlands, New Zealand, Norway, Portugal, Spain, Switzerland, the U.K., and the U.S.

Application: The Impact of the 1990 German Reunification

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.

Application: The Impact of the 1990 German Reunification

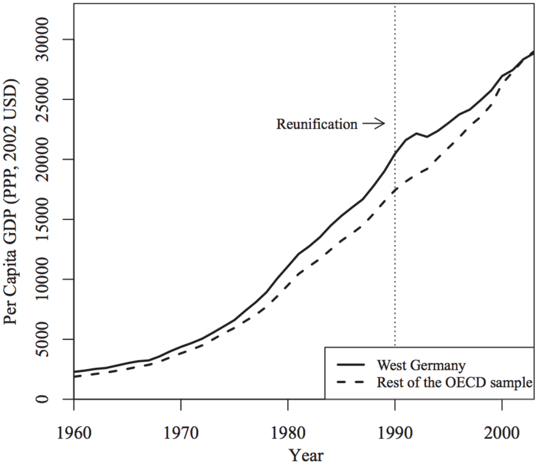
Table 1 Synthetic and Regression Weights for West Germany

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.

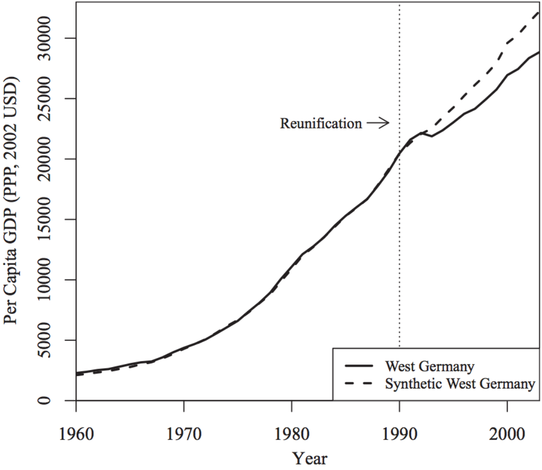
Application: The Impact of the 1990 German Reunification

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



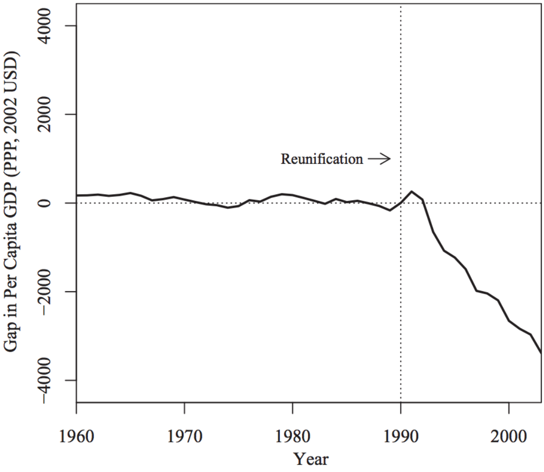
Application: The Impact of the 1990 German Reunification

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



Application: The Impact of the 1990 German Reunification

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

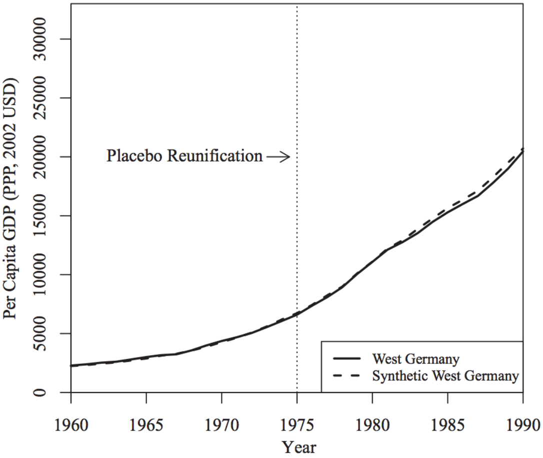


Application: Placebo Studies

- To evaluate the credibility of the results, researchers conduct placebo studies
  1. treatment of interest is reassigned to a year other than 1990
     - compare effect for West Germany to a placebo effect - a period before the reunification actually took place
     - a large placebo estimate would undermine confidence in the results (see Fig 4)
  2. treatment of interest is reassigned to different countries
     - compare the estimated effect on West Germany to the distribution of placebo effects obtained for other countries
     - deem the effect significant if the estimated effect for West Germany is unusually large relative to the distribution of placebo effects
     - or, if the ratio of postreunification RMSPE (root mean square prediction error) to prereunification RMSPE for W Germany is larger than for control countries (see figure 5)
     - RMSPE$=\sqrt{\frac{1}{T}\sum_{t=1}^{T}(Y_{1t}-\sum_{j=2}^{J+1}w_j^*Y_{jt})^2}$

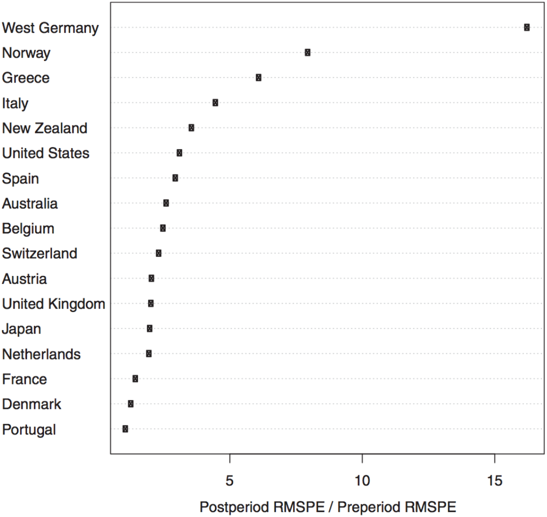
Application: Placebo Studies

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



Application: Placebo Studies

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

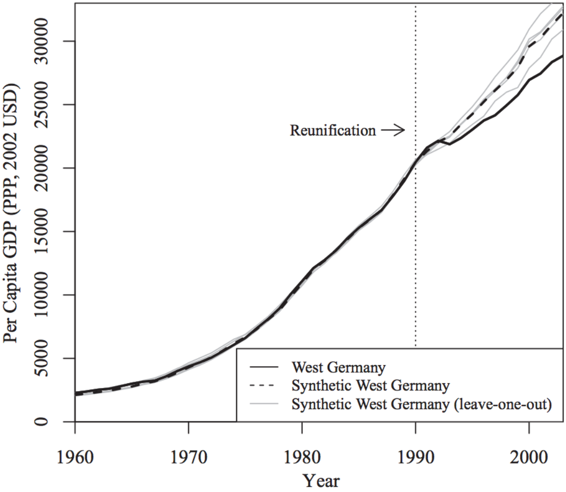


Application: Sensitivity Analysis

- To evaluate the credibility of the results, researchers run sensitivity analysis
  1. test the robustness of main results to changes in the country weights
     - iteratively reestimate the baseline model to construct a synthetic West Germany omitting in each iteration one of the countries that received a positive weight
     - this check allows us to evaluate the extent to which results are driven by any particular control country (see figure 6)
  2. test the robustness to reducing the number of units in the synthetic control
     - the original synthetic West Germany is a weighted average of five control countries: Austria, the U.S., Japan, Switzerland, and the Netherlands
     - now construct synthetic controls for West Germany allowing only combinations of four, three, two, and a single control country, respectively (see tables 3 and 4, and figure 7)

Application: Sensitivity Analysis

FIGURE 6 Leave-One-Out Distribution of the Synthetic Control for West Germany



Application: Sensitivity Analysis

Table 3 Synthetic Weights from Combinations of Control Countries

Synthetic Combination	Countries and W-Weights				
Five control countries	Austria	USA	Japan	Switzerland	Netherlands
	0.42	0.22	0.16	0.11	0.09
Four control countries	Austria	USA	Japan	Switzerland	
	0.56	0.22	0.12	0.10	
Three control countries	Austria	USA	Japan		
	0.59	0.26	0.15		
Two control countries	Austria	USA			
	0.76	0.24			
One control country	Austria				
	1				

Notes: Countries and W-weights for synthetic control are constructed from the best-fitting combination of five, four, three, and two countries, as well as one country. See text for details.

Application: Sensitivity Analysis

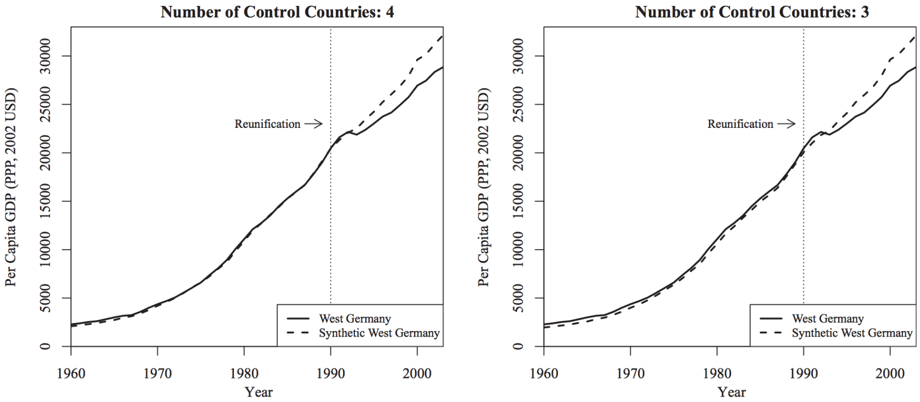
Table 4 Economic Growth Predictor Means before the German Reunification for Combinations of Control Countries

	West	Synthetic West Germany Number of countries in synthetic control					OECD
	Germany	**5**	**4**	**3**	**2**	**1**	**Sample**
GDP per capita	15808.9	15802.2	15800.9	15492.9	15580.9	14817.0	8021.1
Trade openness	56.8	56.9	55.9	52.5	61.5	74.6	31.9
Inflation rate	2.6	3.5	3.6	3.6	3.8	3.5	7.4
Industry share	34.5	34.4	34.6	34.8	34.3	35.5	34.2
Schooling	55.5	55.2	57.6	57.7	60.7	60.9	44.1
Investment rate	27.0	27.0	27.2	26.8	25.6	26.6	25.9

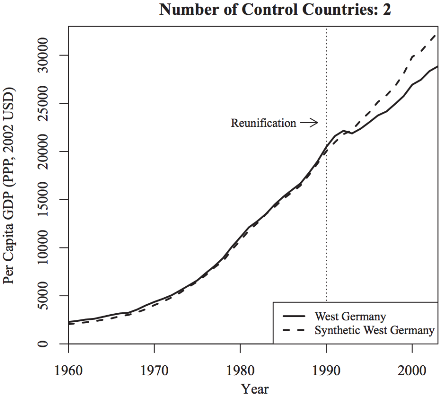
Notes: GDP per capita, inflation rate, and trade openness are averaged for the 1981–1990 period. Industry share is averaged for the 1981–1990 period. Investment rate and schooling are averaged for the 1980–1985 period.

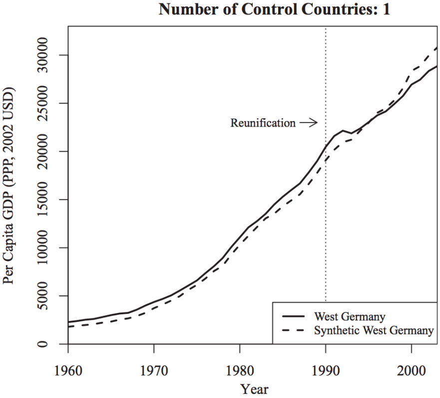
Application: Sensitivity Analysis

FIGURE 7 Per Capita GDP Gaps between West Germany and Sparse Synthetic Controls



Application: Sensitivity Analysis





Application: The Effect of CA's Tobacco Control Program

- Abadie et al. (2010) study the effects of Proposition 99
  - a large-scale tobacco control program that California implemented in 1988
- They use *synthetic control methods*
  - CA is quite different from other states in tobacco consumption
  - it is difficult to know how CA would have evolved in absence of the tobacco control program
- Findings
  - following Proposition 99, tobacco consumption fell markedly in CA relative to a comparable synthetic control region
  - by the year 2000, annual per-capita cigarette sales in CA were about 26 packs lower than what they would have been in the absence of Proposition 99

Application: The Effect of CA's Tobacco Control Program

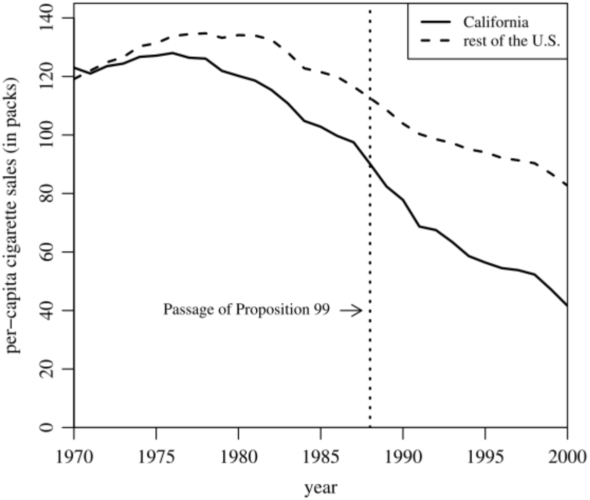


Figure 1. Trends in per-capita cigarette sales: California vs. the rest of the United States.

Application: The Effect of CA's Tobacco Control Program

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

Application: The Effect of CA's Tobacco Control Program

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.

Application: The Effect of CA's Tobacco Control Program

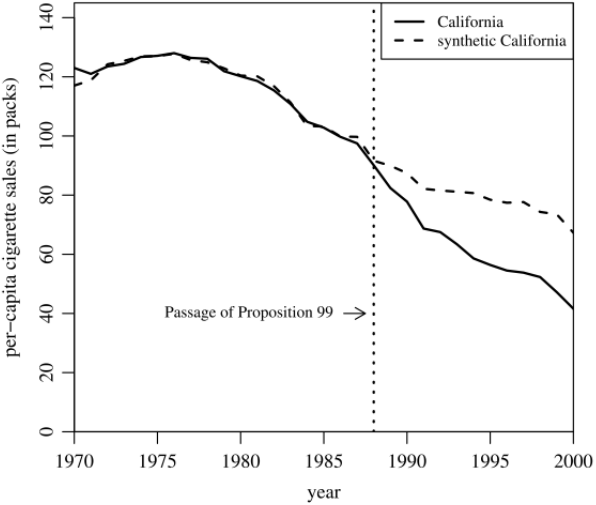
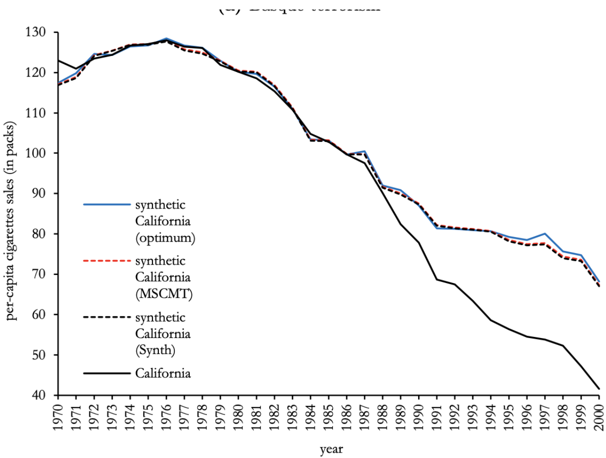


Figure 2. Trends in per-capita cigarette sales: California vs. synthetic California.

Kuosmanen et al. : criticism of **synth**



(b) California's tobacco control program

Fig. 1. The impact of suboptimal **w** weights on the evolution of synthetic controls.

Application: The Effect of CA's Tobacco Control Program

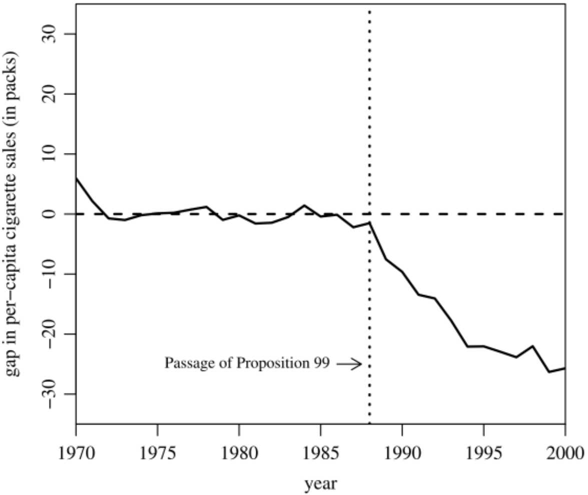


Figure 3. Per-capita cigarette sales gap between California and synthetic California.

Application: Placebo Studies

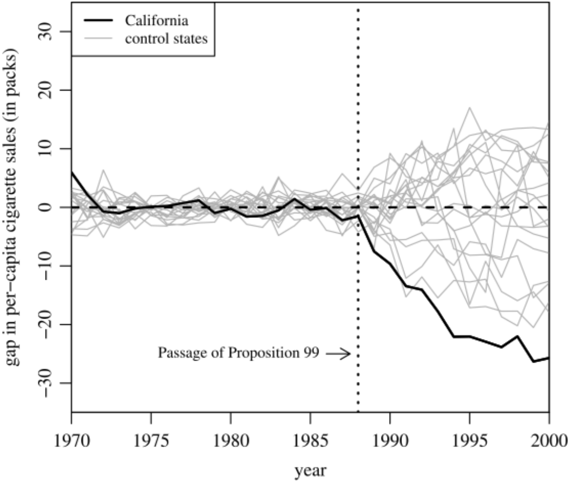


Figure 7. Per-capita cigarette sales gaps in California and placebo gaps in 19 control states (discards states with pre-Proposition 99 MSPE two times higher than California's).

Application: Placebo Studies

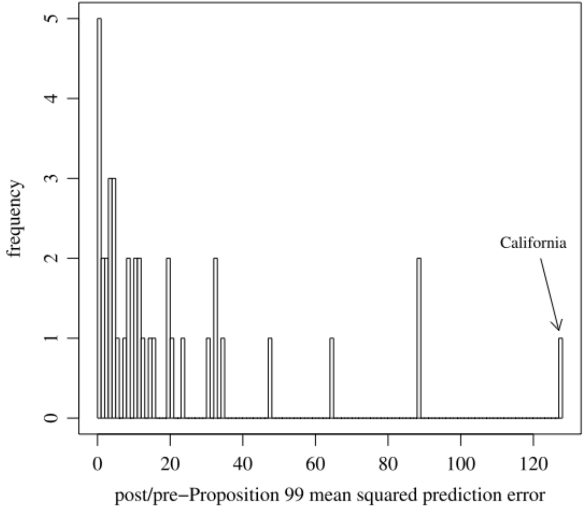


Figure 8. Ratio of post-Proposition 99 MSPE and pre-Proposition 99 MSPE: California and 38 control states.

Difference-in-differences checklist

Difference-in-differences checklist

- Good controls: check for covariates (balance tables), plot pre-trends.
- Did the policies happen when you think they happened?
- Watch out or try to control for packaged policies
- Spillovers to the control group? Think of precautionary behavior in cities that didn't have lockdowns during the COVID-19 pandemic.
- Quasi-exogeneity of the policy: anticipatory behavior?
- Same composition of treatment and control groups? Is migration a risk?
- Triple differences if you have a good second unaffected control group.
- Ideally use balanced panels, keep an eye out for corrections to unbalanced panels and/or treatment heterogeneity over time (Callaway and Sant'Anna, Abraham and Sun, more papers to follow).

Difference-in-differences checklist

- Conducting event studies: useful for picking up pre-trends and detecting reverse causality, mapping treatment intensity over time
- Clustering- use bootstrap if few clusters. If very few treatment units, consider regional variation (impact), RDDs, Fisher permutation test or...
- Synthetic control methods
- Placebo treatments
- In simple two way FE models, use instrumental variable for X_{it} you suspect is correlated with time-varying error term.
- You may just have to sign and discuss any remaining biases.