

Il metodo del «controllo sintetico»

Samuele Poy



Quando si utilizza

- Le policies realizzate a livello aggregato (Stati, regioni, comuni, etc.)
- L'unità «trattata» è una (o sono poche); mentre, le potenziali unità di controllo sono molte.
→ Nella ricerca sociale, la selezione del termine di confronto è spesso ambigua
- I metodi tradizionali di regressione non sono utilizzabili



Il setting

- Si supponga di osservare $j+1$ unità, per un periodo temporale da $t=1...T$
- L'unità trattata è la numero «1». Essa inizia a essere trattata a partire da un certo momento T_0 , tale per cui $t=1...T_0 ...T$.
- L'interesse è sulla stima dell'effetto causale per l'unità trattata nei periodi $t > T_0$



Il setting

- Quindi:

$$\alpha_{1t} = Y_{1t} - Y_{1t}^N \text{ per ogni } t > T_0$$

- Dove:

Y_{1t} è il valore dell'outcome osservato dell'unità 1 (trattata)

Y_{1t}^N è il valore dell'outcome «controfattuale» (osservato nel caso in cui l'unità trattata non fosse stata trattata; valore non disponibile)



Il setting

- Si noti che Y_{it}^N è osservabile per tutte le unità, compresa quella «trattata», per $t < T_0$.

- Si definisca:

$$Y_{it}^N = \delta_t + \boldsymbol{\theta}_t \mathbf{Z}_i + \boldsymbol{\gamma}_t \boldsymbol{\mu}_i + \varepsilon_{it}$$

dove δ è la costante, \mathbf{Z} è un vettore di caratteristiche osservabili, $\boldsymbol{\mu}$ è un vettore di caratteristiche non osservabili, ε è il termine di errore.



Il setting

- Si consideri un vettore di pesi \mathbf{W} , che assegna a ciascuna unità (j) di controllo un valore.
- I pesi sono tali per cui:
 - $w_j \geq 0$ per ogni unità di controllo
 - La somma dei pesi è 1.
- Per ogni configurazione dei pesi assegnati a ciascuna unità di controllo (diversa combinazione di esse) si definisce un diverso «controllo sintetico» potenziale tale per cui:
 - $$\sum_{j=2}^{j+1} w_j Y_{jt} = \delta_t + \boldsymbol{\theta}_t \sum_{j=2}^{j+1} w_j \mathbf{Z}_{j+} + \boldsymbol{\gamma}_t \sum_{j=2}^{j+1} w_j \boldsymbol{\mu}_{j+} + \sum_{j=2}^{j+1} w_j \boldsymbol{\varepsilon}_{jt}$$



Il setting

- Abadie et al. (2003) suggeriscono di identificare una combinazione di pesi ottimale \mathbf{W}^* tale per cui:
 - $\sum_{j=2}^{j+1} w_j^* \mathbf{Z}_j = \mathbf{Z}_1$
 - $\sum_{j=2}^{j+1} w_j^* Y_{j1} = Y_{11} \dots \sum_{j=2}^{j+1} w_j^* Y_{jT_0} = Y_{1T_0}$
- \mathbf{W}^* sono pesi assegnati alle unità di controllo (il «controllo sintetico») tale per cui sia il valore delle diverse covariate \mathbf{Z} (pre-intervento) sia l'outcome Y in diversi periodi (pre-intervento) sono uguali tra unità trattata e controllo sintetico costituito.

Il setting

- E' dimostrato che, avendo a disposizione serie storiche relativamente lunghe pre-politica, W^* può comportare l'uguaglianza sia nell'ampio vettore di covariate (\mathbf{Z}) sia negli outcome Y (pre-policy) solo se anche i fattori non osservabili (μ) sono uguali tra unità trattata e controllo sintetico.
- Si noti che ad es. nei modelli Difference-in-Difference si tiene conto di fattori non osservabili, costanti nel tempo.
- Nel controllo sintetico si permette ai fattori non osservabili di variare nel tempo.



Il setting

- Ottenuta la configurazione di pesi ottimale (su dati pre-policy) si applica \mathbf{W}^* alle unità non trattate per Y nei vari periodi post policy, stimando così l'outcome Y controfattuale (controllo sintetico).
- L'effetto del trattamento è infine stimato per differenza tra valore reale del trattato e stima derivante dal controllo sintetico.

$$\begin{aligned}\alpha_{1t} &= Y_{1t} - Y_{1t}^N \text{ per ogni } t > T_0 \\ &= Y_{1t} - \sum_{j=2}^{j+1} w_j^* Y_{jt}\end{aligned}$$

•



Il setting

- Se il bilanciamento pre-policy riferito al controllo sintetico è povero (il controllo sintetico non mima Z e Y pre policy del trattato, le stime sono altrettanto poco credibili.



Alcuni esempi



L'effetto di una norma antifumo in California (Abadie et al., 2010)

- Nel 1988 la California approva una legge per il contrasto del consumo di sigarette:
 - Il prezzo aumenta di 0.25 al pacchetto
 - Destinazione delle entrate fiscali a politiche per la salute e contrasto al fumo
 - Campagne sui media
 - Altre misure di sensibilizzazione in tema ambiente e aria pulita, progetti anti-tobacco, etc.



L'effetto di una norma antifumo in California (Abadie et al., 2010)

- Gli altri stati americani sono il gruppo di potenziali controlli
- Sono esclusi dal «donor pool» (gruppo di potenziali controlli) gli Stati che negli anni seguenti hanno approvato misure analoghe



L'effetto di una norma antifumo in California (Abadie et al., 2010)

Table 1. Cigarette sales predictor means

Variables	California		Average of 38 control states
	Real	Synthetic	
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

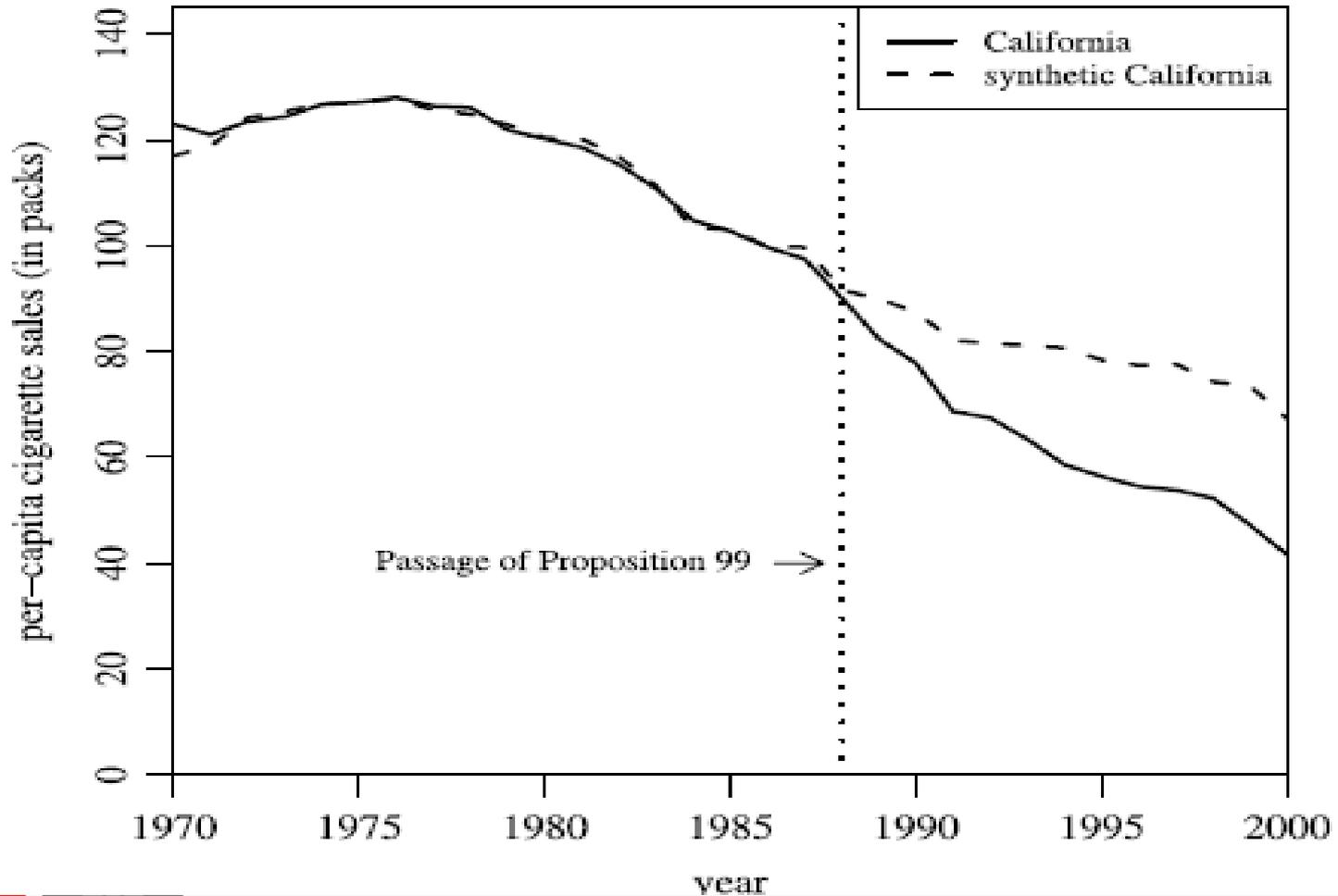
NOTE: All variables except lagged cigarette sales are averaged for the 1980–1988 period (beer consumption is averaged 1984–1988). GDP per capita is measured in 1997 dollars, retail prices are measured in cents, beer consumption is measured in gallons, and cigarette sales are measured in packs.

L'effetto di una norma antifumo in California (Abadie et al., 2010)

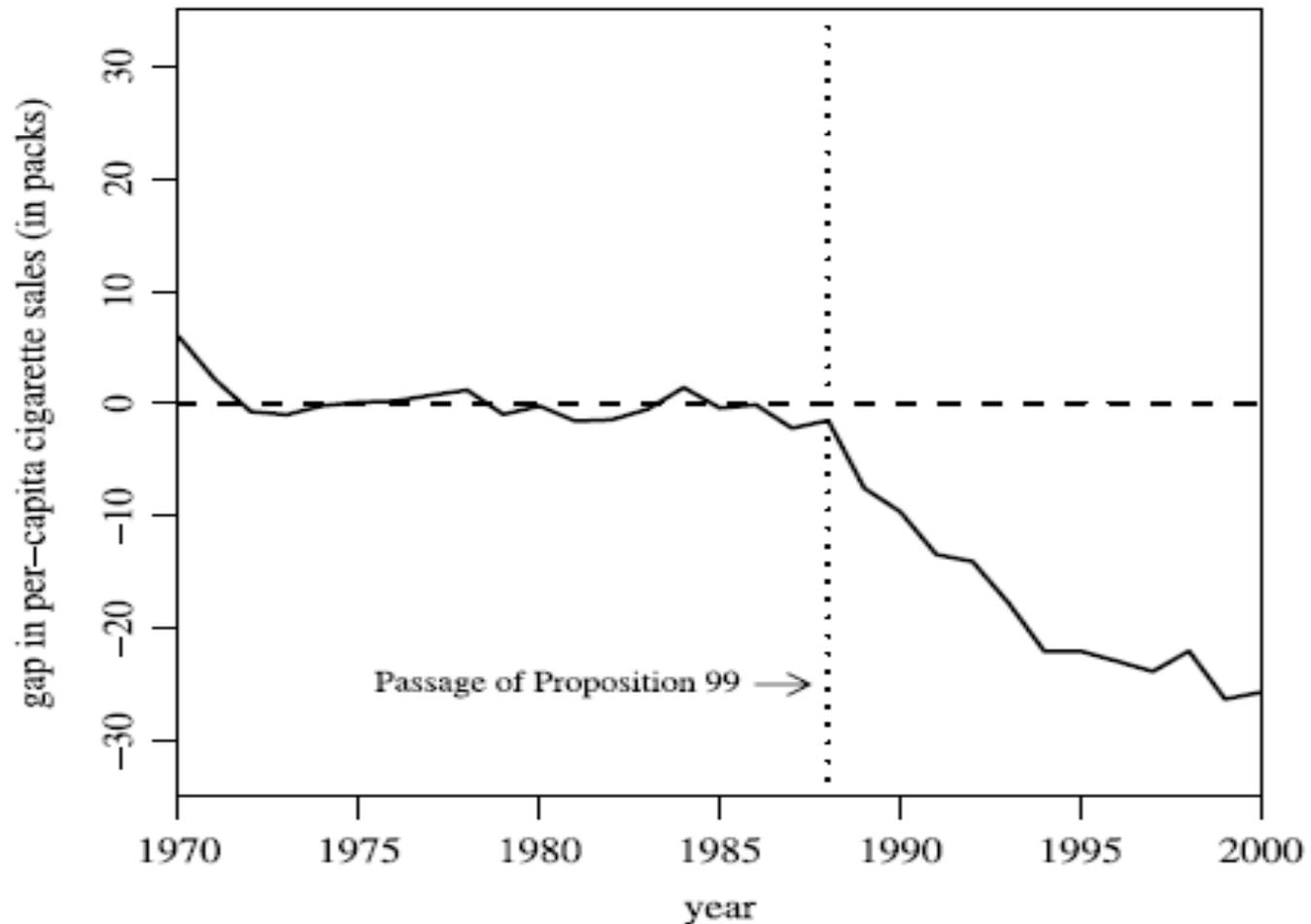
Table 2. State weights in the synthetic California

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

L'effetto di una norma antifumo in California (Abadie et al., 2010)



L'effetto di una norma antifumo in California (Abadie et al., 2010)

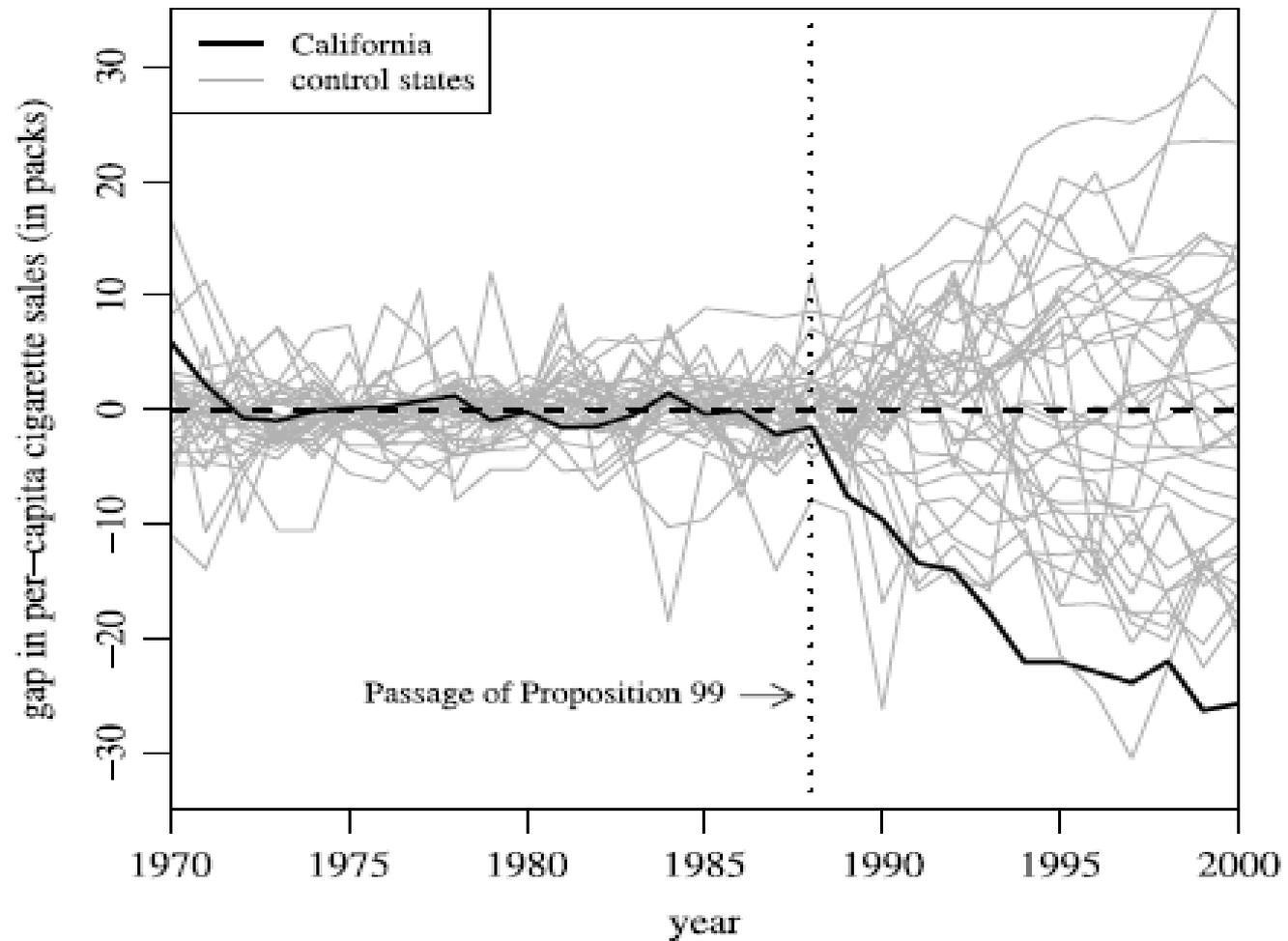


Inferenza

- Quanto è «dimensionalmente grande» l'effetto stimato per la California?
- Si usa fare test detti «placebo». Si esclude l'unità trattata davvero (California)
 - Si selezionano solo unità del gruppo di controllo, replicando iterativamente la procedura di controllo sintetico attribuendo a ciascuna di esse l'essere unità trattata («fake experiment»)
 - Si verifica quanto è grande l'effetto stima per la vera unità trattata (California) rispetto ai fake experiments.

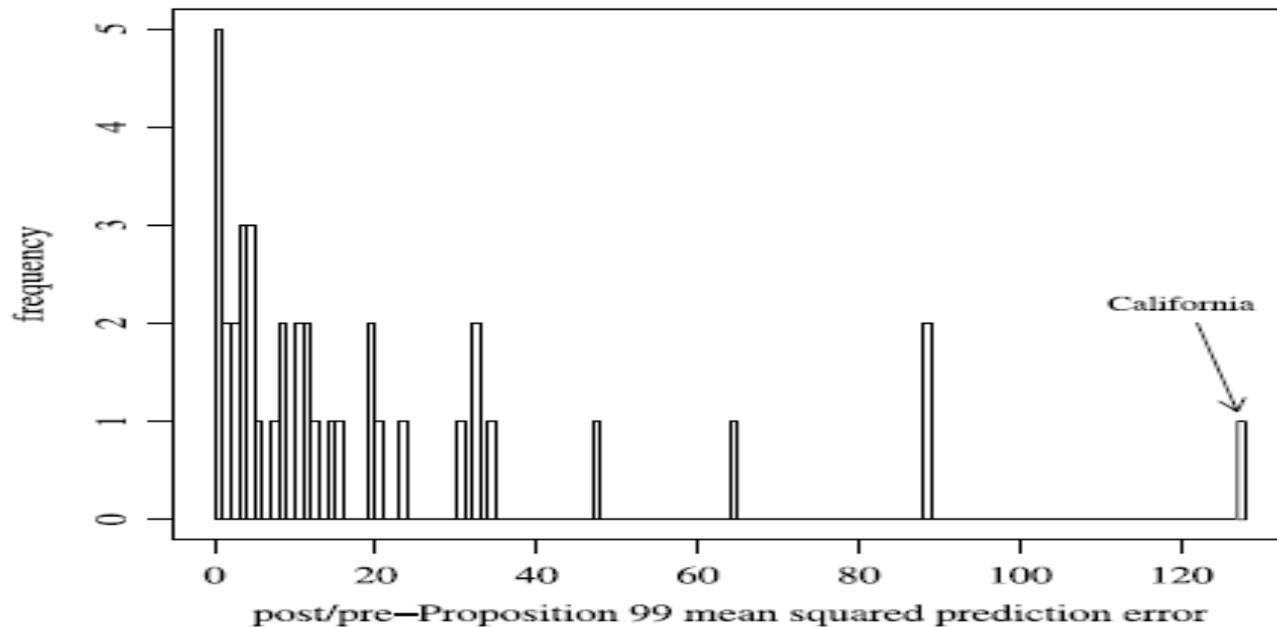


Inferenza

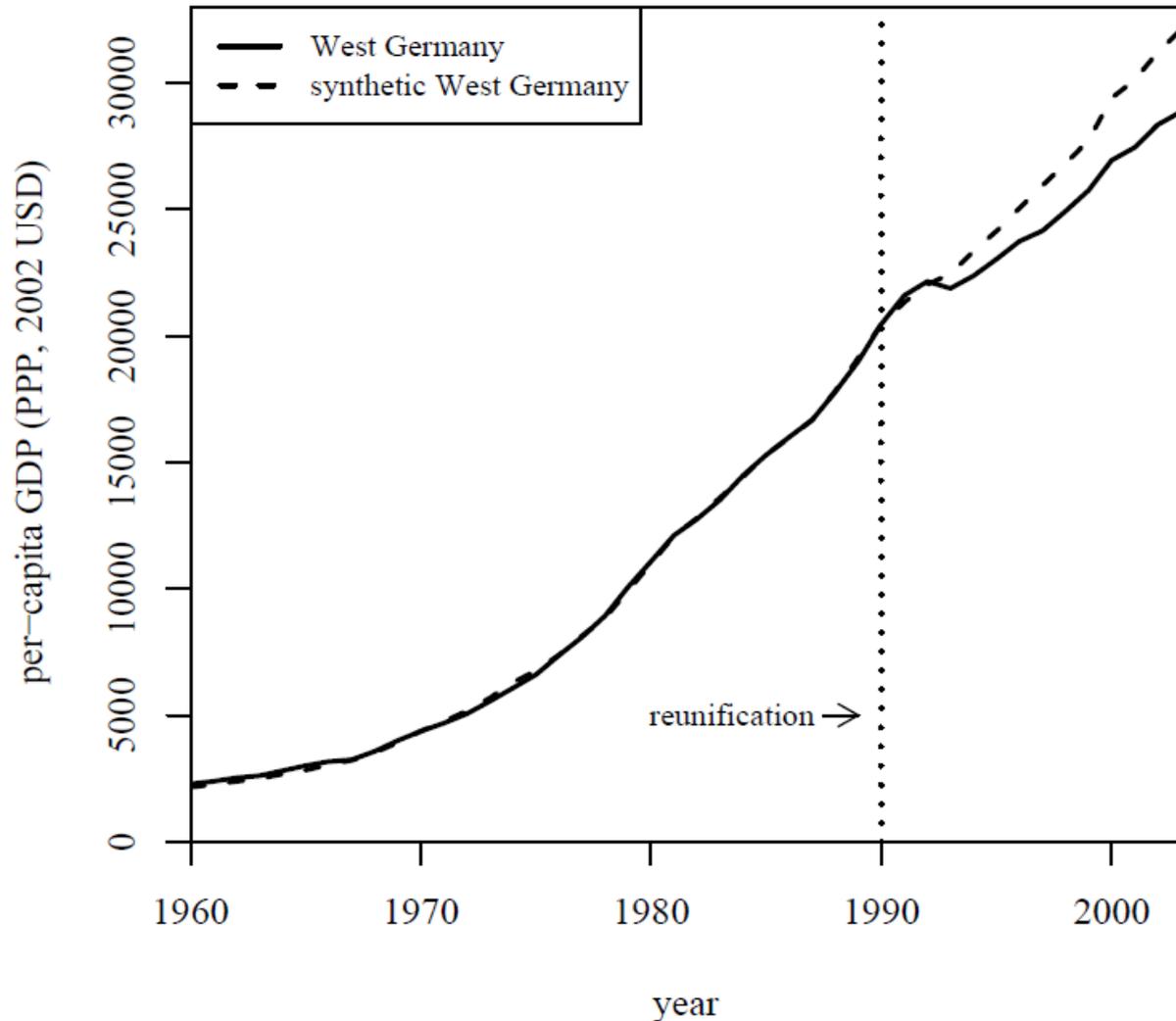


Inferenza

- Ci calcola l'errore quadratico medio di predizione (MSPE) sul periodo POST e PRE trattamento. Quindi, si calcola il rapporto POST rispetto a PRE di tale statistica per ogni placebo test.



L'effetto della reunificazione in Germania (Abadie et al., 2014)



L'effetto economico del terrorismo basco (Abadie et al., 2003)

THE AMERICAN ECONOMIC REVIEW

MARCO

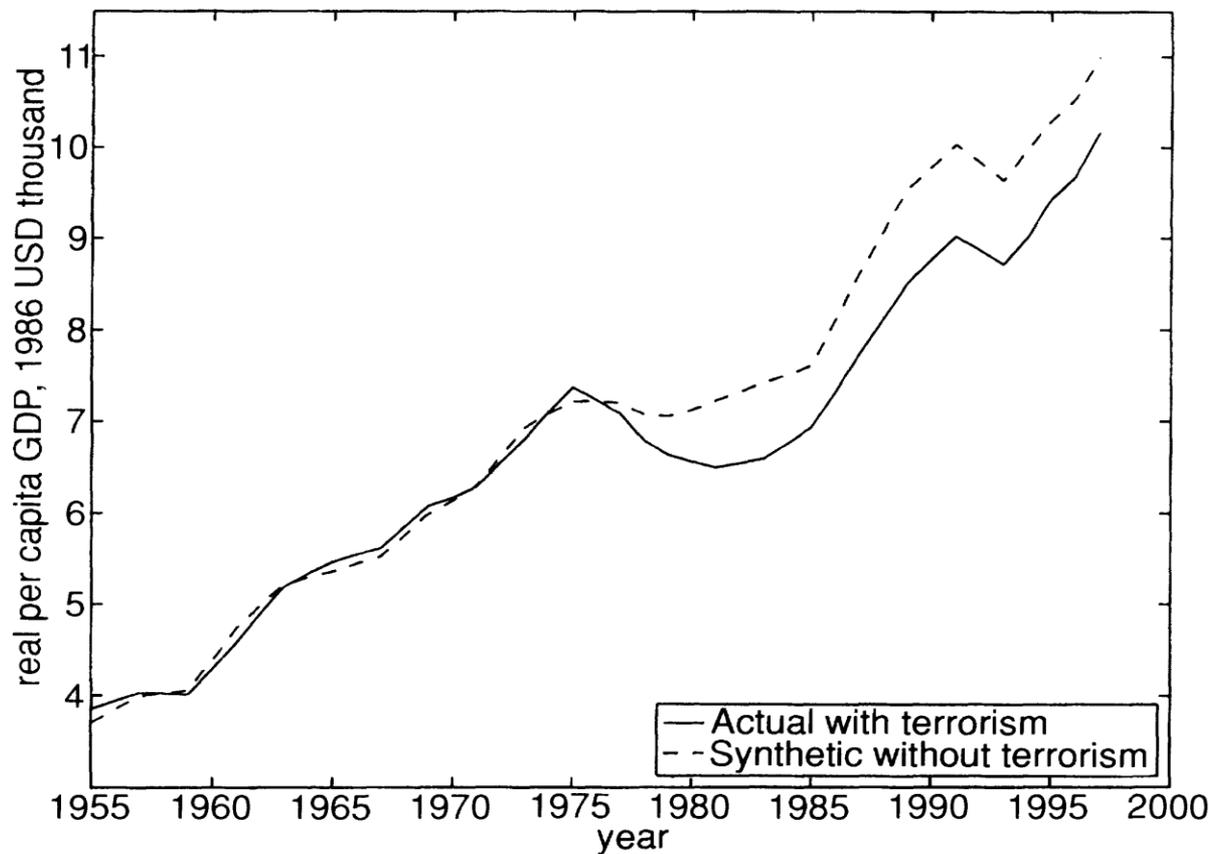


FIGURE 1. PER CAPITA GDP FOR THE BASQUE COUNTRY

L'effetto economico dell'indipendenza amministrativa locale (Dalmazzo et al., 2018)

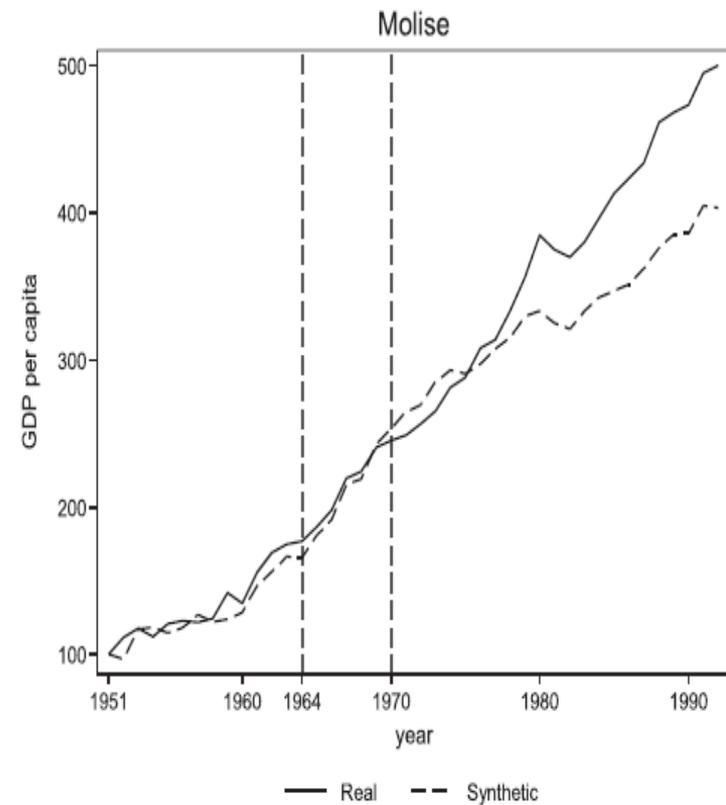
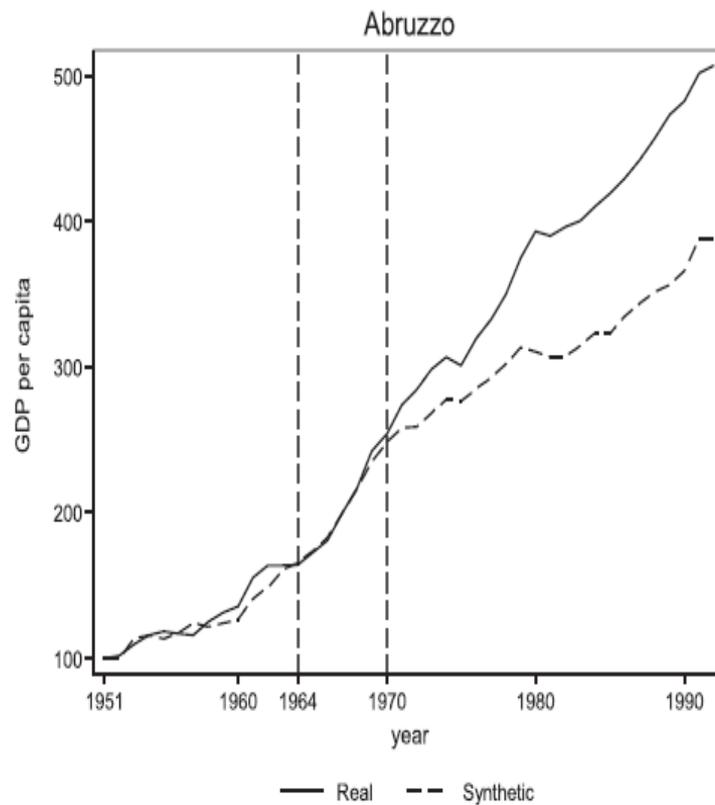
Table 1

Economic growth predictor means before secession.

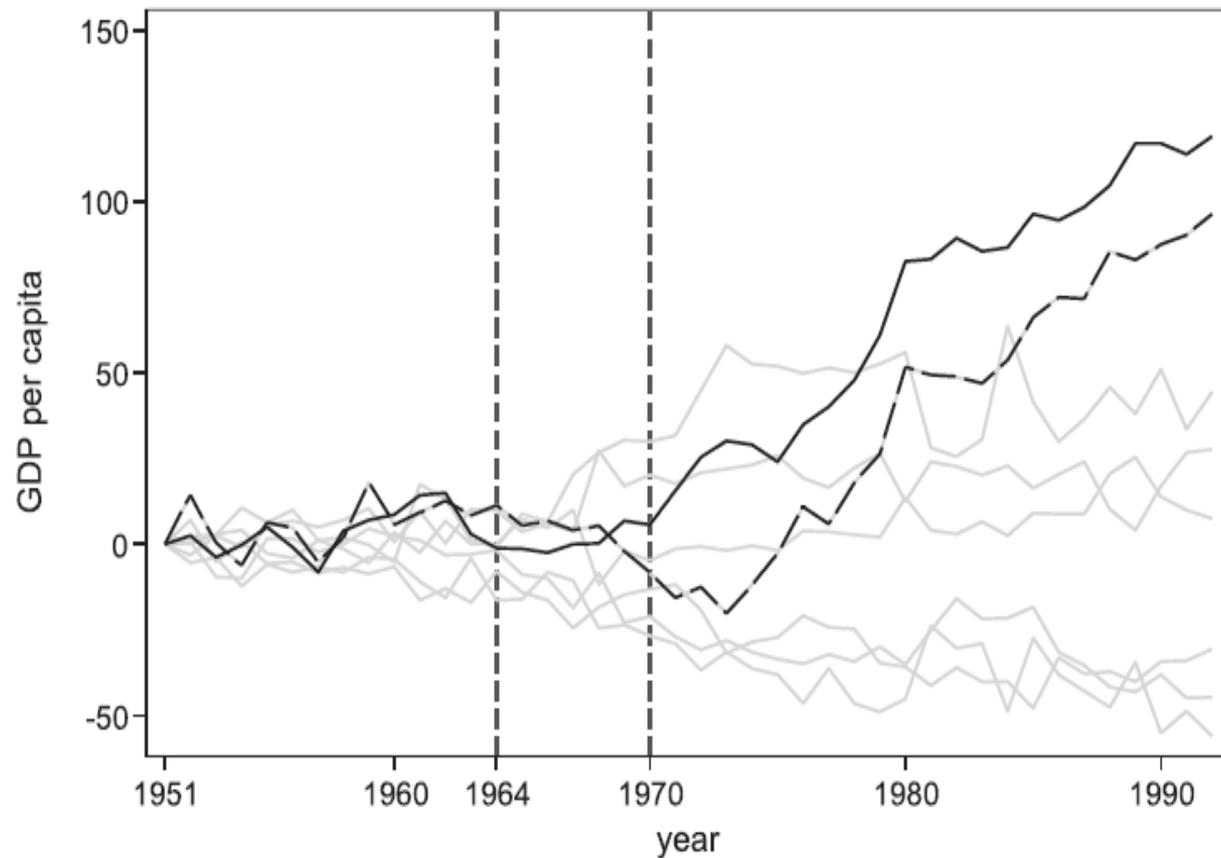
	Abruzzo		Molise		Average of other Mezzogiorno regions (not experiencing the secession)	Average of Central-Northern regions
	Real	Synthetic	Real	Synthetic		
GDP per capita	3315.8	3277.5	2920.4	2984.9	3126.1	5568.0
Annual GDP per capita growth rate	0.043	0.042	0.050	0.047	0.046	0.043
Investment-to-GDP ratio	0.255	0.323	0.214	0.367	0.313	0.271
Share of high educated	0.038	0.038	0.033	0.034	0.051	0.052
Population density	115.4	117.0	85.6	141.2	201.5	202.1
Net imports-to-GDP ratio	0.153	0.207	0.100	0.248	0.191	0.004
Minimum wage index	80.8	81.4	79.4	80.7	83.0	94.9
Agriculture share of VA	0.149	0.138	0.189	0.163	0.135	0.071
Industry share of VA	0.227	0.251	0.217	0.216	0.231	0.343
Market services share of VA	0.391	0.375	0.361	0.393	0.405	0.430

Notes: The weights used to build the synthetic controls are: Apulia (0.101), Sicily (0.397) and Sardinia (0.502) for the Abruzzo region; Apulia (0.589), Basilicata (0.233), Calabria (0.164) and Sicily (0.013) for the Molise region. The weights are chosen to minimize the distance between treated and synthetic control units in terms of variables described in [Appendix A3, Table A2](#). The last two columns represent population-weighted averages.

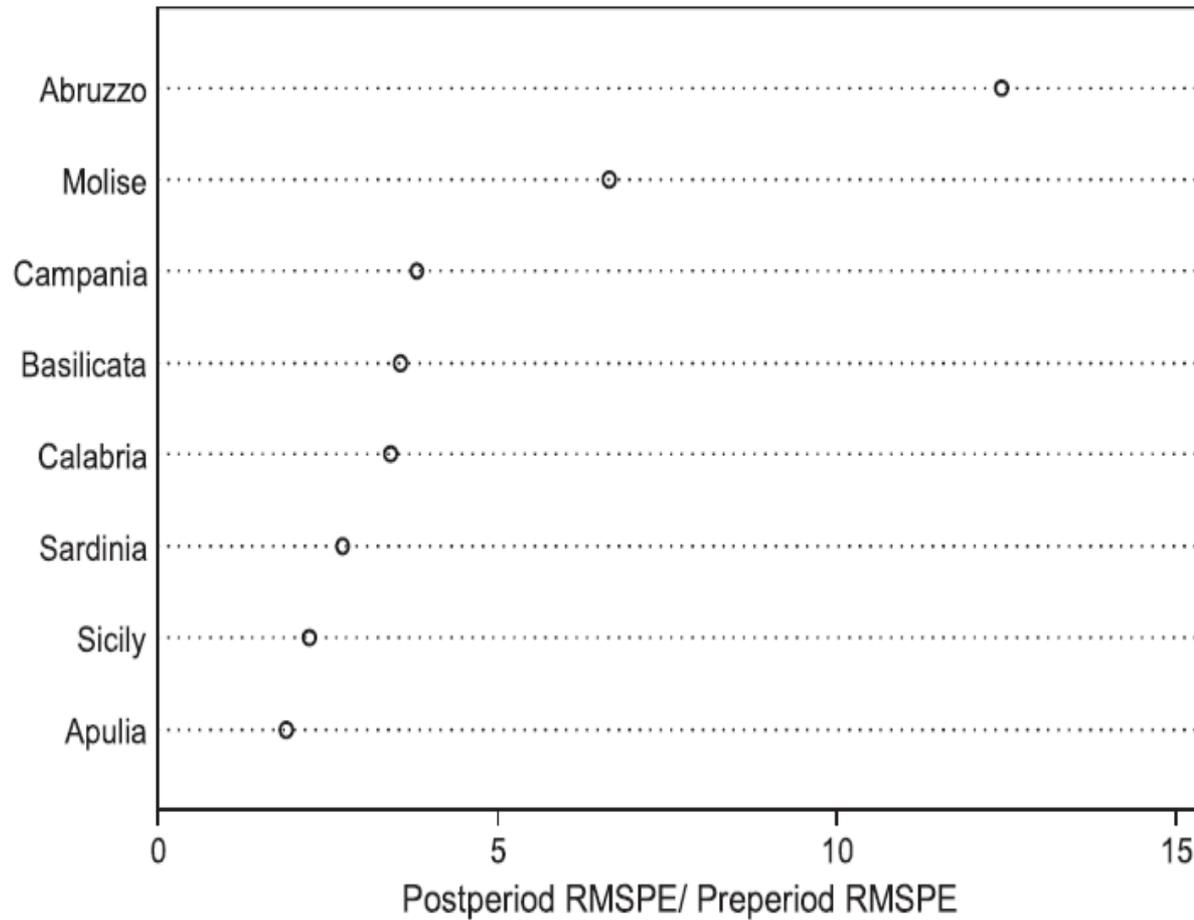
L'effetto economico della secessione locale (Dalmazzo et al., 2018)



L'effetto economico della secessione locale (Dalmazzo et al., 2018)



L'effetto economico della secessione locale (Dalmazzo et al., 2018)



Ulteriore materiale utile (e recenti avanzamenti)

Abadie, Alberto. 2021. "Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects." *Journal of Economic Literature*, 59 (2): 391-425.

Arkhangelsky, Dmitry, Susan Athey, David A. Hirshberg, Guido W. Imbens, and Stefan Wager (2020). Synthetic Difference in Differences, *American Economic Review*, forthcoming

Ben-Michael, Eli, Avi Feller, and Jesse Rothstein (2021). Synthetic Controls with Staggered Adoption. NBER Working Paper 28886

Abadie, Alberto, and Jérémy L'Hour (2021). A penalized synthetic control estimator for disaggregated data, *Journal of the American Statistical Association*

STATA (compreso utilities per stimare effetti con più unità trattate):

<https://journals.sagepub.com/doi/pdf/10.1177/1536867X1801700404>

