# Causal Inference with Synthetic Controls

Alexander J. Almeida
Stanford Graduate School of Business

July 20, 2023

# Roadmap of Talk

Synthetic control design: basics

Randomization inference: basics

Randomization inference for synthetic controls

Extensions and conclusion


MAR CANTÁBRICO
FRANCIA
SANTANDER
PAÍS VASCO
OVIEDO
ASTURIAS
SANTIAGO DE COMPOSTELA
CANTA-BRIA
VITORIA
NAVARRA
ANDORRA
GALICIA
PAMPLO-NA
LOGROÑO
LA RIOJA
CASTILLA Y LEÓN
ARAGÓN
CATALUÑA
VALLADOLID
ZARAGOZA
BARCELONA
P O R T U G A L
COM. MADRID
MADRID
OCÉANO ATLÁNTICO
TOLEDO
VALEN-CIA
PALMA DE MALLORCA
ISLAS BALEARES
EXTREMADURA
CASTILLA-LA MANCHA
COM. VALEN-CIANA
MÉRIDA
REG. DE MURCIA
MURCIA
MAR MEDITERRÁNEO
ANDALUCÍA
SEVILLA
CANARIAS
STA. CRUZ DE TENERIFE
LAS PALMAS DE GRAN CANARIA
CEUTA
MELILLA
ÁFRICA

- Euskadi Ta Askatasuna (ETA) formed to support Basque separatism
- Beginning terrorist activities in the late 1960s

TABLE 1—CHRONOLOGY OF ETA'S TERRORIST ACTIVITY

| Year | Killings | Kidnappings | Event |
|---|---|---|---|
| 1968 | 2 | 0 | First victim of ETA |
| 1969 | 1 | 0 | |
| 1970 | 0 | 1 | |
| 1971 | 0 | 0 | |
| 1972 | 1 | 1 | |
| 1973 | 6 | 1 | ETA kills Franco's Prime Minister Admiral Carrero-Blanco |
| 1974 | 19 | 0 | |
| 1975 | 16 | 0 | Dictator Franco dies |
| 1976 | 17 | 4 | |
| 1977 | 11 | 1 | First democratic elections in Spain after Franco's death |
| 1978 | 67 | 6 | Spanish Constitution approved in referendum |
| 1979 | 76 | 13 | Regional Autonomy Statute for the Basque Country approved |
| 1980 | 92 | 13 | |
| 1981 | 30 | 10 | Attempted military coup. Spain joins NATO |

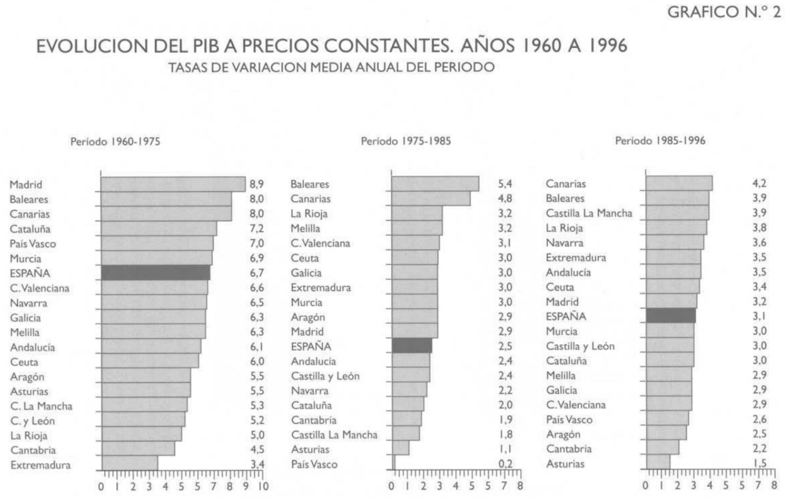
| | | | |
|---|---|---|---|
| 1982 | 37 | 8 | |
| 1983 | 32 | 5 | |
| 1984 | 32 | 0 | |
| 1985 | 37 | 3 | |
| 1986 | 41 | 3 | Spain joins European Community |
| 1987 | 52 | 1 | |
| 1988 | 19 | 1 | |
| 1989 | 19 | 1 | |
| 1990 | 25 | 0 | |
| 1991 | 46 | 0 | |
| 1992 | 26 | 0 | Barcelona hosts the Summer Olympic Games |
| 1993 | 14 | 1 | |
| 1994 | 13 | 0 | |
| 1995 | 15 | 1 | |
| 1996 | 5 | 2 | |
| 1997 | 13 | 1 | |
| 1998 | 6 | 0 | ETA declares indefinite cease-fire starting on September 18 |

- Question: What was the impact of terrorism on the people of the Basque Region?

- Question: What was the impact of terrorism on the people of the Basque Region?

- Less ambitious
    - What was the **economic** impact of terrorism on the people of the Basque Region?

    - Can we quantify this effect?




Source: BBV, 1999

How do we quantify the effect of terrorism?

How do we quantify the effect of terrorism?

- A naïve comparison of means before and after terrorism
  - ignores time-varying dynamics that are independent of terrorism
  - e.g. Business cycle variations in macroeconomic aggregates
- Comparing Basque country to other regions of Spain

How do we quantify the effect of terrorism?

- A naïve comparison of means before and after terrorism
  - ignores time-varying dynamics that are independent of terrorism
  - e.g. Business cycle variations in macroeconomic aggregates
- Comparing Basque country to other regions of Spain
  - Concentration of terrorism in Basque country supports this method

How do we quantify the effect of terrorism?

How do we quantify the effect of terrorism?

Compare the outcomes of Basque country to other regions that are similar.

Issue: we can't just compare it to the rest of Spain.

TABLE 3—PRE-TERRORISM CHARACTERISTICS, 1960'S

| | Basque Country (1) | Spain (2) |
|---|---|---|
| Real per capita GDP[a] | 5,285.46 | 3,633.25 |
| Investment ratio (percentage)[b] | 24.65 | 21.79 |
| Population density[c] | 246.89 | 66.34 |
| Sectoral shares (percentage)[d] | | |
| Agriculture, forestry, and fishing | 6.84 | 16.34 |
| Energy and water | 4.11 | 4.32 |
| Industry | 45.08 | 26.60 |
| Construction and engineering | 6.15 | 7.25 |
| Marketable services | 33.75 | 38.53 |
| Nonmarketable services | 4.07 | 6.97 |
| Human capital (percentage)[e] | | |
| Illiterates | 3.32 | 11.66 |
| Primary or without studies | 85.97 | 80.15 |
| High school | 7.46 | 5.49 |
| More than high school | 3.26 | 2.70 |

How do we quantify the effect of terrorism?

Compare the outcomes of Basque country to other regions that **are similar**.

How do we quantify the effect of terrorism?

Compare the outcomes of Basque country to other regions that **are similar**.

Any experimental control group of this form will be a weighted average of the options available.

- A difference in difference (DID) strategy would weight non-basque autonomous communities equally
- We can choose the weights more carefully: **synthetic control**

The synthetic control method as we currently know it is developed in Abadie and Gardeazabal, 2003.

# Empirical considerations

TABLE 3—PRE-TERRORISM CHARACTERISTICS, 1960'S

| | Basque Country (1) | Spain (2) | "Synthetic" Basque Country (3) |
|---|---|---|---|
| Real per capita GDP[a] | 5,285.46 | 3,633.25 | 5,270.80 |
| Investment ratio (percentage)[b] | 24.65 | 21.79 | 21.58 |
| Population density[c] | 246.89 | 66.34 | 196.28 |
| Sectoral shares (percentage)[d] | | | |
| Agriculture, forestry, and fishing | 6.84 | 16.34 | 6.18 |
| Energy and water | 4.11 | 4.32 | 2.76 |
| Industry | 45.08 | 26.60 | 37.64 |
| Construction and engineering | 6.15 | 7.25 | 6.96 |
| Marketable services | 33.75 | 38.53 | 41.10 |
| Nonmarketable services | 4.07 | 6.97 | 5.37 |
| Human capital (percentage)[e] | | | |
| Illiterates | 3.32 | 11.66 | 7.65 |
| Primary or without studies | 85.97 | 80.15 | 82.33 |
| High school | 7.46 | 5.49 | 6.92 |
| More than high school | 3.26 | 2.70 | 3.10 |

*Sources:* Authors' computations from Matilde Mas et al. (1998) and Fundación BBV (1999).

[a] 1986 USD, average for 1960–1969.

[b] Gross Total Investment/GDP, average for 1964–1969.

[c] Persons per square kilometer, 1969.

[d] Percentages over total production, 1961–1969.

[e] Percentages over working-age population, 1964–1969.

- Suppose there are $J$ potential control regions.

- Suppose there are $J$ potential control regions.
- Let $W = (w_1, \ldots, w_J)' \in \mathbb{R}^J$ be the weights. Each $W$ corresponds to a different synthetic control region. Usually we will want nonnegative weights that sum to 1.

- Suppose there are $J$ potential control regions.

- Let $W = (w_1, \ldots, w_J)' \in \mathbb{R}^J$ be the weights. Each $W$ corresponds to a different synthetic control region. Usually we will want nonnegative weights that sum to 1.

- Goal: Choose $W$ such that the synthetic control region resembles the actual region.

- Suppose there are $J$ potential control regions.

- Let $W = (w_1, \ldots, w_J)' \in \mathbb{R}^J$ be the weights. Each $W$ corresponds to a different synthetic control region. Usually we will want nonnegative weights that sum to 1.

- Goal: Choose $W$ such that the synthetic control region resembles the actual region.

    - Let $X_1 \in \mathbb{R}^K$ be the vector of pre-treatment observables.

    - Let $X_0 \in \mathbb{R}^{K \times J}$ be the matrix of pre-treatment observables for all regions.

    - We want to minimize $\|X_1 - X_0 W\|$

- We could minimize using the Euclidian norm

$$\|X_1 - X_0 W\| = \sqrt{\sum_{k=1}^{K} \left(X_1^k - X_0^k W\right)^2},$$

but we can incorporate greater flexibility by weighting the $k$th covariate by a value $V_{kk} \in \mathbb{R}$.

- We could minimize using the Euclidian norm

$$\|X_1 - X_0 W\| = \sqrt{\sum_{k=1}^{K} \left(X_1^k - X_0^k W\right)^2},$$

  but we can incorporate greater flexibility by weighting the $k$th covariate by a value $V_{kk} \in \mathbb{R}$.
- Operationalize this by letting $V \in \mathbb{R}^{K \times K}$ be diagonal (thus defining a semi-norm, see Abadie et al., 2010) where we collect the weights $V_{kk}$ along the diagonal.

$$\|X_1 - X_0 W\|_V := (X_1 - X_0 W)' V (X_1 - X_0 W)$$

- We could minimize using the Euclidian norm

$$\|X_1 - X_0 W\| = \sqrt{\sum_{k=1}^{K} \left(X_1^k - X_0^k W\right)^2},$$

  but we can incorporate greater flexibility by weighting the $k$th covariate by a value $V_{kk} \in \mathbb{R}$.
- Operationalize this by letting $V \in \mathbb{R}^{K \times K}$ be diagonal (thus defining a semi-norm, see Abadie et al., 2010) where we collect the weights $V_{kk}$ along the diagonal.

$$\|X_1 - X_0 W\|_V := (X_1 - X_0 W)' V (X_1 - X_0 W)$$

- What should we choose for $V$?

The weighting matrix $V$ could be chosen to weight relative importance of matching specific predictors (think 2-step GMM).

Instead, we will choose the matrix $V$ to match the treated unit's time series in the outcome variable.

Algorithm:

The weighting matrix $V$ could be chosen to weight relative importance of matching specific predictors (think 2-step GMM).

Instead, we will choose the matrix $V$ to match the treated unit's time series in the outcome variable.

Algorithm:

- Let $\mathcal{W}$ denote the set of nonnegative weights in $\mathbb{R}^K$ that sum to one. For each weighting matrix $V$, choose weights to minimize the weighted sum of squared errors,
$$W^*(V) = \text{argmin}_{W \in \mathcal{W}} (X_1 - X_0 W)' V (X_1 - X_0 W)$$

The weighting matrix $V$ could be chosen to weight relative importance of matching specific predictors (think 2-step GMM).

Instead, we will choose the matrix $V$ to match the treated unit's time series in the outcome variable.

Algorithm:

- Let $\mathcal{W}$ denote the set of nonnegative weights in $\mathbb{R}^K$ that sum to one. For each weighting matrix $V$, choose weights to minimize the weighted sum of squared errors,
$$W^*(V) = \text{argmin}_{W \in \mathcal{W}} (X_1 - X_0 W)' V (X_1 - X_0 W)$$
- For each balanced synthetic control $W^*(V)$, choose $V$ to minimize the difference between the time series of the treated unit and the synthetic control.

Performing this method tells us that

$$\texttt{Basque} = 0.8508(\texttt{Catalonia}) + 0.1492(\texttt{Madrid})$$

Performing this method tells us that

$$\texttt{Basque} = 0.8508(\texttt{Catalonia}) + 0.1492(\texttt{Madrid})$$

Based on the construction, we expect that

- pre-intervention observables of Basque region and synthetic Basque region are close, and

- pre-intervention per capita GDP time series for Basque and synthetic Basque regiones are close

TABLE 3—PRE-TERRORISM CHARACTERISTICS, 1960'S

| | Basque Country (1) | Spain (2) | "Synthetic" Basque Country (3) |
|---|---|---|---|
| Real per capita GDP[a] | 5,285.46 | 3,633.25 | 5,270.80 |
| Investment ratio (percentage)[b] | 24.65 | 21.79 | 21.58 |
| Population density[c] | 246.89 | 66.34 | 196.28 |
| Sectoral shares (percentage)[d] | | | |
| Agriculture, forestry, and fishing | 6.84 | 16.34 | 6.18 |
| Energy and water | 4.11 | 4.32 | 2.76 |
| Industry | 45.08 | 26.60 | 37.64 |
| Construction and engineering | 6.15 | 7.25 | 6.96 |
| Marketable services | 33.75 | 38.53 | 41.10 |
| Nonmarketable services | 4.07 | 6.97 | 5.37 |
| Human capital (percentage)[e] | | | |
| Illiterates | 3.32 | 11.66 | 7.65 |
| Primary or without studies | 85.97 | 80.15 | 82.33 |
| High school | 7.46 | 5.49 | 6.92 |
| More than high school | 3.26 | 2.70 | 3.10 |

*Sources:* Authors' computations from Matilde Mas et al. (1998) and Fundación BBV (1999).

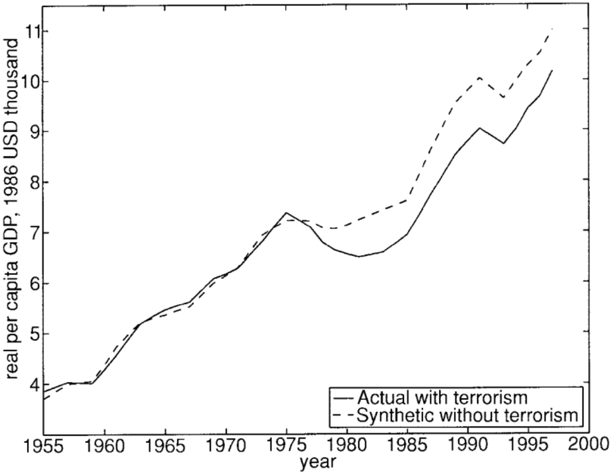
[a] 1986 USD, average for 1960–1969.

[b] Gross Total Investment/GDP, average for 1964–1969.

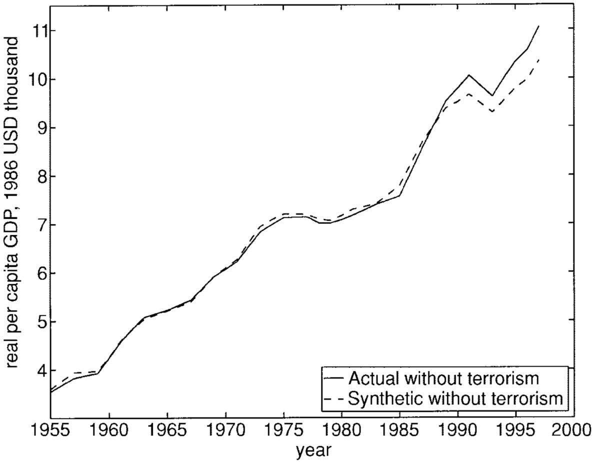
[c] Persons per square kilometer, 1969.

[d] Percentages over total production, 1961–1969.

[e] Percentages over working-age population, 1964–1969.


real per capita GDP, 1986 USD thousand
11
10
9
8
7
6
5
4
1955
1960
1965
1970
1975
1980
1985
1990
1995
2000
year
Actual with terrorism
Synthetic without terrorism

- No formal “inferences” are made in the paper: confidence bounds, standard errors
- Instead, the authors conduct a falsification exercise.
    - Suppose that in fact Catalonia experienced terrorism, and not the Basque Country.
    - If we conduct the same analysis on Catalonia and find a discrepancy between the synthetic Catalonia and the real Catalonia then we are in trouble.




The results of conducting a “placebo test” using the data from Abadie and Gardeazabal, 2003. A synthetic control is built for Catalonia, with the Basque Country excluded from the donor pool. Notice that the resulting time series reproduces the actual time series precisely until the late 1980s. Also relevant is the fact that the olympics took place in Barcelona in 1992.

- Outstanding question: is the result significant?
- The authors spend time offering arguments that it is, but we don’t have any formal statistical model at this point.
- Subsequent papers make strides in this direction leveraging randomization inference framework.

# Roadmap of Talk

Synthetic control design: basics

Randomization inference: basics

Randomization inference for synthetic controls

Extensions and conclusion

- The idea of randomization inference is often attributed to Ronald Fisher who introduced the method as an aside in his textbook *The Design of Experiments* (Fisher, 1935).
- In fact, the careful formulation of the idea is better attributed to Edwin Pitman (1937) and Bernard L. Welch (1937).

It’s argued that the idea’s attribution is an example of Stigler’s law of eponymy (Onghena, 2017):

“No scientific discovery is named after its original discoverer.”

- Stephen Stigler

Consider this example.

- Three students are given a study guide before an exam and three are not.
- The scores are $x_0 = (85, 60, 95, 55, 70, 85)$, where the first three elements in the vector are the treated students' scores.

The scores for the students that received the guide exceed the others by 10. Does the study guide work?

Consider this example.

- Three students are given a study guide before an exam and three are not.
- The scores are $x_0 = (85, 60, 95, 55, 70, 85)$, where the first three elements in the vector are the treated students' scores.

The scores for the students that received the guide exceed the others by 10. Does the study guide work?

Intuition: any three of these values are equally likely to be the treated units.
If we assume no treatment effect (i.e. $H_0 : \tau = 0$), then any permutation of the outcome vector, e.g. $(85, 60, 55, 95, 70, 85)$, equally likely to occur.

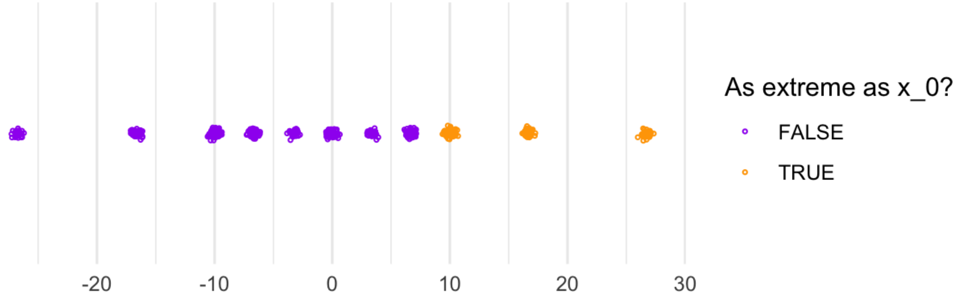
Consider this example.

- Three students are given a study guide before an exam and three are not.
- The scores are $x_0 = (85, 60, 95, 55, 70, 85)$, where the first three elements in the vector are the treated students' scores.

The scores for the students that received the guide exceed the others by 10. Does the study guide work?

Intuition: any three of these values are equally likely to be the treated units.
If we assume no treatment effect (i.e. $H_0 : \tau = 0$), then any permutation of the outcome vector, e.g. $(85, 60, 55, 95, 70, 85)$, equally likely to occur.

One solution: Enumerate all 6! permutations $x$ of $x_0$ & calculate the statistic for each one. This gives an exact distribution of the statistic $T$. Comparing $T(x_0)$ to this distribution allows us to make inferential statements about the treatment effect.




The empirical distribution of the difference in means statistic under the assumption that all permutations of the outcome vector are equally likely. The corresponding probability of observing a result at least as surprising as 10 under the null hypothesis is 30%.

An equivalent solution is to reason that ...

- The statistic $T$ depends only the partitioning of the students' scores into treated and untreated.
- There are $\binom{6}{3} = 20$ ways of choosing the three treated units.
- Enumerating all combinations will give us a p-value as well.

In this sense, the natural way to think about the experiment is with combinations, rather than permutations (see Imbens and Rubin, 2015, Chapter 5).

Nomenclature for randomization inference is occasionally confusing:

- Randomization tests are occasionally called permutation tests
- Randomization tests are occasionally more naturally identified with combinations rather than permutations of the data

Worse, Onghena relates how the *Encyclopedia of Statistical Sciences* in 1986 published conflicting entries which simultaneously stated that:

- Randomization test is special case of permutation test (Edgington).
- Permutation test is special case of randomization test (Gibbons).

Summary of randomization inference: by assuming that the treatment effect is zero we can calculate an exact null distribution of the statistic of interest by enumerating all of the possible assignments.

We can do this for any statistic:

- Transformations of the data. Consider taking $\log$ transformations for positive data.
- Robust statistics: the difference in medians or the difference in mean rank are more “robust” in the sense often attributed to Tukey.
- The $t$-statistic, but compare the statistic to the exact distribution rather than to the theoretical $t$-distribution.

A final note on randomization inference is that there is a close connection between randomization inference and bootstrapping.

- Bootstrapping consists of sampling with replacement from the data.
- Randomization consists of sampling without replacement from the data.

“The bootstrap distribution was originally called the ‘combination distribution.’ It was designed to extend the virtues of permutation testing to the great majority of statistical problems where there is nothing to permute.”

- Bradley Efron and Robert Tibshirani (1993, pp. 218).

# Roadmap of Talk

Synthetic control design: basics

Randomization inference: basics

Randomization inference for synthetic controls

Extensions and conclusion

Consider once again the synthetic controls set up in light of our randomization inference framework:

Consider once again the synthetic controls set up in light of our randomization inference framework:

- We want to compare the treated units to their synthetic controls via some statistic

Consider once again the synthetic controls set up in light of our randomization inference framework:

- We want to compare the treated units to their synthetic controls via some statistic
- Each placebo test calculates the same statistic (as in Abadie and Gardeazabal, 2003)

Consider once again the synthetic controls set up in light of our randomization inference framework:

- We want to compare the treated units to their synthetic controls via some statistic
- Each placebo test calculates the same statistic (as in Abadie and Gardeazabal, 2003)
- We care to determine if there is a significant effect associated with some intervention

Consider once again the synthetic controls set up in light of our randomization inference framework:

- We want to compare the treated units to their synthetic controls via some statistic
- Each placebo test calculates the same statistic (as in Abadie and Gardeazabal, 2003)
- We care to determine if there is a significant effect associated with some intervention

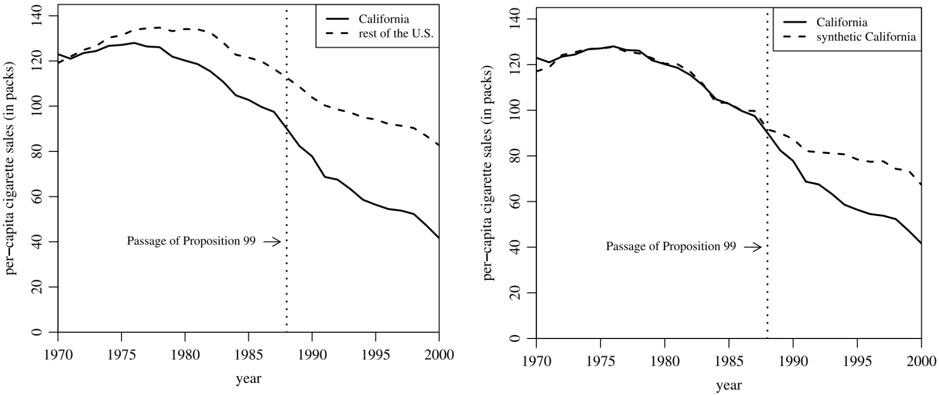
Without making any distributional assumptions about the statistic in question, we are able to calculate exact p-values using this methodology. This was noticed in Abadie et al., 2010.

- In 1988, California voters passed **1988 Proposition 99**
- The law included a 25-cent excise tax per carton of cigarettes
- What is an appropriate counterfactual?

- In 1988, California voters passed **1988 Proposition 99**
- The law included a 25-cent excise tax per carton of cigarettes
- What is an appropriate counterfactual? **Construct a synthetic California**

- In 1988, California voters passed **1988 Proposition 99**
- The law included a 25-cent excise tax per carton of cigarettes
- What is an appropriate counterfactual? **Construct a synthetic California**

$$\texttt{CA} = 0.164(\texttt{CO}) + 0.069(\texttt{CT}) + 0.199(\texttt{MT}) + 0.234(\texttt{NV}) + 0.334(\texttt{UT})$$




Headline: CA cigarette sales p.c. were **26 packs lower** because of Proposition 99.

Placebo studies are used to quantify the significance of the effect.

Placebo studies are used to quantify the significance of the effect.

Statistical inferences can be argued via randomization inference framework, and robustness can be strengthened by “in-time placebos.”

First, a useful construct is the **mean square prediction error** (MSPE) for unit *i*.

First, a useful construct is the **mean square prediction error** (MSPE) for unit $i$.

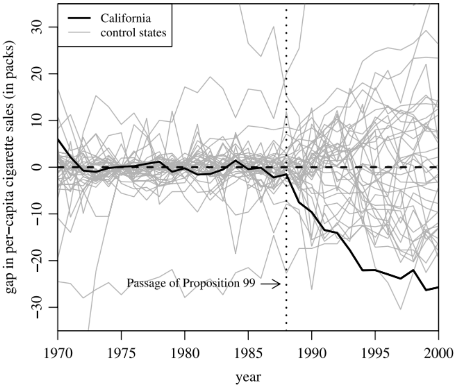
Suppose there are $N_c$ control periods and $N$ periods total. Let $\hat{Y}_{it}$ denote the synthetic control for unit $i$ at time $t$. Then MSPE over the pretreatment period is given by

$$\eta_{\text{pre}} := N_c^{-1} \sum_{t=1}^{N_c} (Y_{it} - \hat{Y}_{it})^2,$$

and the MSPE over the treatment period is given by

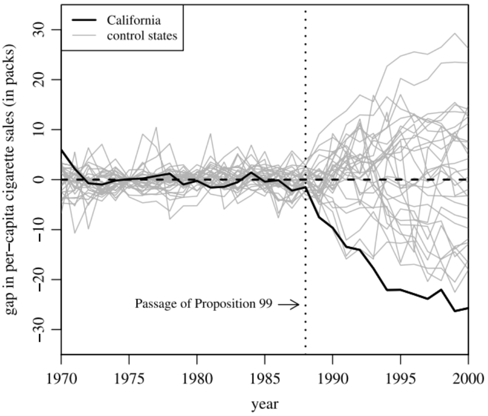
$$\eta_{\text{post}} := (N - N_c)^{-1} \sum_{t=N_c+1}^{N} (Y_{it} - \hat{Y}_{it})^2.$$

# Some of the synthetic controls perform quite poorly in the pretreatment period.




Time series showing the gap in per-capita cigarette sales using a synthetic control method. The black line shows the series for California and the grey lines show the result for all 38 other donor states. States were excluded if they enacted large-scale smoking reforms or big taxes between 1989 and 2000.

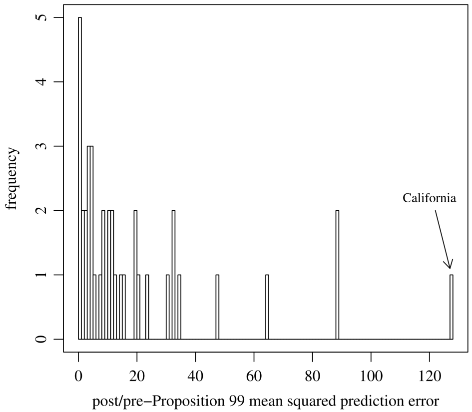
# After subsetting to well performing MSPE in the preintervention period we see a clearer picture.




Time series showing the gap in per-capita cigarette sales using a synthetic control method. The black line shows the series for California and the grey lines show the result for 29 donor states. States were excluded if they enacted large-scale smoking reforms or big taxes between 1989 and 2000 or if the mean square prediction error greater than five times that of California.

Another way to compare synthetic controls is based on the ratio of their posttreatment MSPE to their pretreatment MSPE: $\eta_{\text{post}} / \eta_{\text{pre}}$.

Another way to compare synthetic controls is based on the ratio of their posttreatment MSPE to their pretreatment MSPE: $\eta_{\text{post}} / \eta_{\text{pre}}$.




A bar chart where the frequency is tabulated over all ratios of posttreatment to pretreatment MSPE. Notice that California stands out drastically in this figure due to the fact that the pretreatment fit is tight, whereas the posttreatment series diverges from the synthetic control.

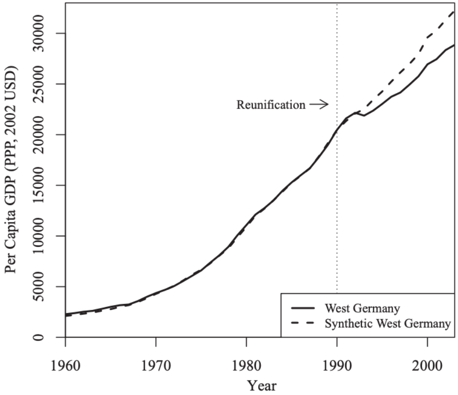
A further diagnostic is to report the results of creating a synthetic control using a different time period of treatment.

Such a diagnostic is considered an “in-time” placebo test, in contrast to an “in-place placebo.”

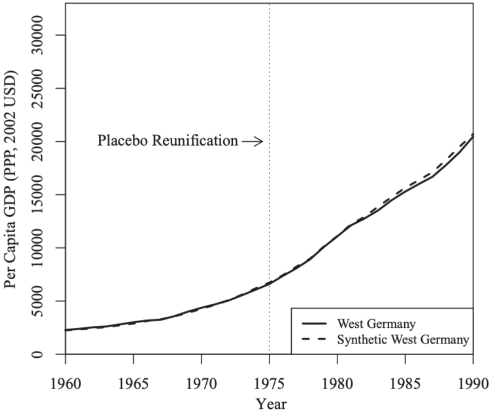
A further diagnostic is to report the results of creating a synthetic control using a different time period of treatment.

Such a diagnostic is considered an “in-time” placebo test, in contrast to an “in-place placebo.”

The in-place placebo is introduced in Abadie et al., 2015.




Time series for the evolution of West Germany and a Synthetic Control. Source: Abadie et al., 2015.




Time series for the evolution of West Germany and a Synthetic Control. The synthetic control is an in-time placebo where the year 1975 is taken to be the year of the intervention. Source: Abadie et al., 2015.

# Roadmap of Talk

Synthetic control design: basics

Randomization inference: basics

Randomization inference for synthetic controls

Extensions and conclusion

# Extensions

- Regression estimators for case studies also weight observations but might extrapolate outside of the convex hull of the data. Synthetic control corrects for this deficiency (Abadie et al., 2015).

- Diff-in-diff is nested in a broader synthetic control framework. We simply require that observations receive equal weight in DID (Arkhangelsky et al., 2021).

- Cross-validation and jackknife techniques for variance estimation that fall outside of the basic randomization inference setting that we have discussed.

# Conclusion

- Synthetic control is an influential technique for causal inference despite its relative novelty.
- Theory provides justification for statistical statements about the estimated treatment effects.
- Empirical applications abound in settings where the number of treated is relatively small.
- Research in both applied and theoretical settings is active and ongoing.

# References I


- Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. *Journal of the American Statistical Association*, *105*(490), 493–505. https://doi.org/10.1198/jasa.2009.ap08746
- Abadie, A., Diamond, A., & Hainmueller, J. (2015). Comparative Politics and the Synthetic Control Method. *American Journal of Political Science*, *59*(2), 495–510. https://doi.org/10.1111/ajps.12116
- Abadie, A., & Gardeazabal, J. (2003). The Economic Costs of Conflict: A Case Study of the Basque Country. *American Economic Review*, *93*(1), 113–132. https://doi.org/10.1257/000282803321455188
- Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W., & Wager, S. (2021). Synthetic Difference-in-Differences. *American Economic Review*, *111*(12), 4088–4118. https://doi.org/10.1257/aer.20190159
- BBV, F. (1999). *Renta Nacional de Espana y su Distribucioén Provincial* [OCLC: 907369585].

# References II


- Efron, B., & Tibshirani, R. (1993). *An introduction to the bootstrap*. Chapman & Hall.
- Fisher, R. A. (1935). *The design of experiments* (9. ed) [OCLC: 471778573]. Hafner Press.
- Imbens, G., & Rubin, D. B. (2015). *Causal inference for statistics, social, and biomedical sciences: An introduction*. Cambridge University Press.
- Onghena, P. (2017). Randomization Tests or Permutation Tests? A Historical and Terminological Clarification. In *Randomization, Masking, and Allocation Concealment*. Chapman; Hall/CRC.
- Pitman, E. J. G. (1937). Significance Tests Which May be Applied to Samples From any Populations. *Supplement to the Journal of the Royal Statistical Society*, *4*(1), 119. https://doi.org/10.2307/2984124
- Welch, B. L. (1937). On the z-Test in Randomized Blocks and Latin Squares. *Biometrika*, *29*(1/2), 21. https://doi.org/10.2307/2332405