LUND UNIVERSITY

Matching and synthetic controls

Nils Droste

2021 ClimBEco course

Causal Inference from observational data

Synopsis: Today, we will be looking into methods that help us find (aka *match*) or simulate (aka *synthesize*) a control group for inferring causal effects from observational data, and its recent developments

In particular, we will develop an understanding of

Causal Inference from observational data

Synopsis: Today, we will be looking into methods that help us find (aka *match*) or simulate (aka *synthesize*) a control group for inferring causal effects from observational data, and its recent developments

In particular, we will develop an understanding of

- matching approaches

Causal Inference from observational data

Synopsis: Today, we will be looking into methods that help us find (aka *match*) or simulate (aka *synthesize*) a control group for inferring causal effects from observational data, and its recent developments

In particular, we will develop an understanding of

- matching approaches
  - classical
  - machine-based learning

Causal Inference from observational data

Synopsis: Today, we will be looking into methods that help us find (aka *match*) or simulate (aka *synthesize*) a control group for inferring causal effects from observational data, and its recent developments

In particular, we will develop an understanding of

- matching approaches
  - classical
  - machine-based learning
- synthetic controls

Intuition

Consider a situation where the untreated are very different from the treated:

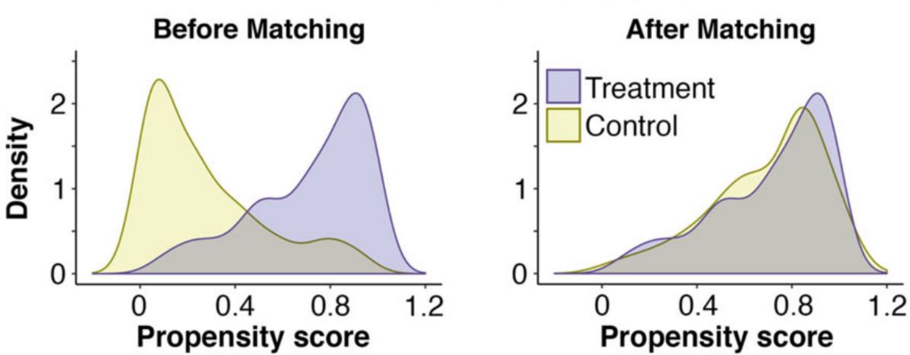


Image source: Schleicher et al. 2020

Intuition

Consider a situation where the untreated are very different from the treated:

> ***Matching**, def: any method that **strategically subsamples** dataset to **balance covariate distribution** in treated and control groups such that after matching both groups share an equal probability of treatment.*

Non-Random Treatment Assignment $\xrightarrow[\text{to Subsample}]{\text{Matching Methods}}$ ***Average Treatment Effect on the Treated + ~~Selection Bias~~***

Image source: Image source: Sizemore and Alkurdi 2019

Intuition

Consider a situation where the untreated are very different from the treated:

***Matching**, def: any method that **strategically subsamples** dataset to **balance covariate distribution** in treated and control groups such that after matching both groups share an equal probability of treatment.*

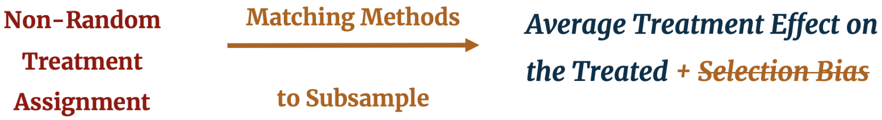
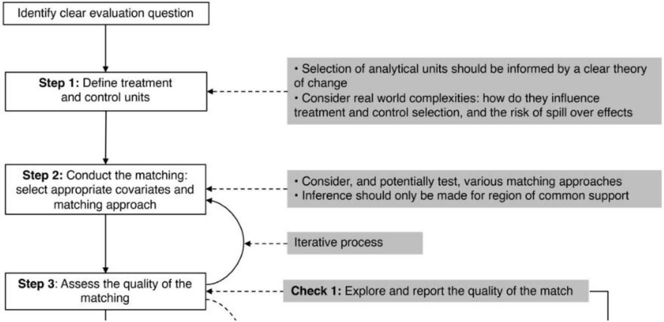


Image source: Image source: Sizemore and Alkurdi 2019

→ matching is a ***pre-analytical procedure***, allowing unbiased inference.

Procedure


Identify clear evaluation question
Step 1: Define treatment and control units
• Selection of analytical units should be informed by a clear theory of change
• Consider real world complexities: how do they influence treatment and control selection, and the risk of spill over effects
Step 2: Conduct the matching: select appropriate covariates and matching approach
• Consider, and potentially test, various matching approaches
• Inference should only be made for region of common support
Iterative process
Step 3: Assess the quality of the matching
Check 1: Explore and report the quality of the match

Procedure

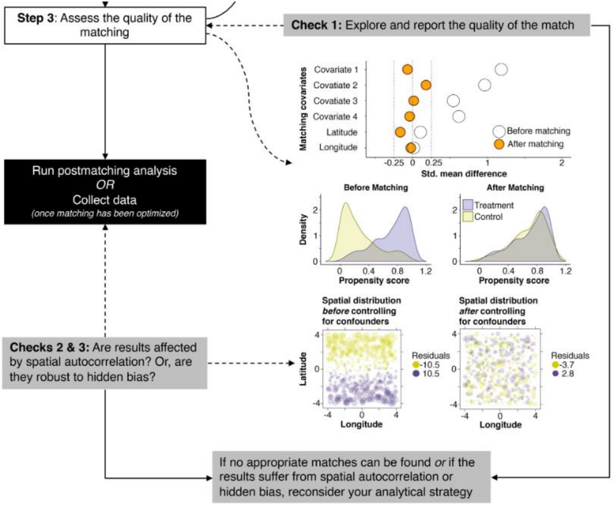


Image source: Schleicher et al. 2020

Basic conditions

The classical overarching conditions for robust causal inference:

- stable unit treatment value assumption (SUTVA)
  - treating one individual unit does not affect another's (potential) outcome
  - treatment is comparable [no (strong) variation in treatment]

Basic conditions

The classical overarching conditions for robust causal inference:

- stable unit treatment value assumption (SUTVA)
  - treating one individual unit does not affect another's (potential) outcome
  - treatment is comparable [no (strong) variation in treatment]
- unconfoundedness (strong ignorability)
  - $(Y(1), Y(0)) \perp T$: treatment assignment is independent of the outcomes
  - i.e. no omitted variable bias (recall the storch example)
  - or, at least, conditional unconfoundedness $(Y(1), Y(0)) \perp T|X$

Basic conditions

The classical overarching conditions for robust causal inference:

- stable unit treatment value assumption (SUTVA)
  - treating one individual unit does not affect another's (potential) outcome
  - treatment is comparable [no (strong) variation in treatment]
- unconfoundedness (strong ignorability)
  - $(Y(1), Y(0)) \perp T$: treatment assignment is independent of the outcomes
  - i.e. no omitted variable bias (recall the storch example)
  - or, at least, conditional unconfoundedness $(Y(1), Y(0)) \perp T|X$

$\rightarrow \pi(X_i) = Pr(D_i = 1|X_i)$ or *propensity score* can be used for matching

Basic conditions

The classical overarching conditions for robust causal inference:

- stable unit treatment value assumption (SUTVA)
  - treating one individual unit does not affect another's (potential) outcome
  - treatment is comparable [no (strong) variation in treatment]
- unconfoundedness (strong ignorability)
  - $(Y(1), Y(0)) \perp T$: treatment assignment is independent of the outcomes
  - i.e. no omitted variable bias (recall the storch example)
  - or, at least, conditional unconfoundedness $(Y(1), Y(0)) \perp T|X$

$\rightarrow \pi(X_i) = Pr(D_i = 1|X_i)$ or *propensity score* can be used for matching
$\rightarrow$ but should maybe not (King and R. Nielsen 2019), we will see alternatives

Overview

Here is a general overview of possible matching methods

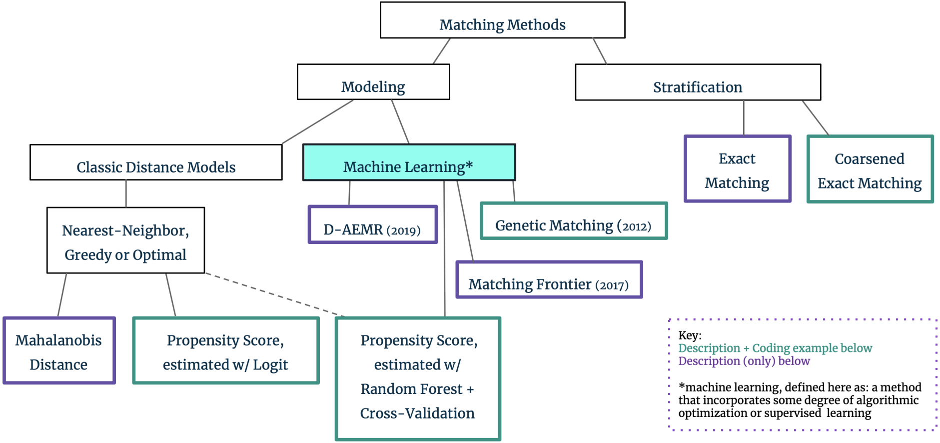


Image source: Sizemore and Alkurdi 2019

Notation

Consider that we aim to estimate *conditional average treatment effect* (CATE) (cf. Abrevaya, Hsu and Lieli 2015)

$$CATE = E(Y(1) - Y(0)|X = x) \tag{1}$$

Notation

Consider that we aim to estimate *conditional average treatment effect* (CATE) (cf. Abrevaya, Hsu and Lieli 2015)

$$CATE = E(Y(1) - Y(0)|X = x) \tag{1}$$

How to find the sufficiently similar subsamples?

Notation

Consider that we aim to estimate *conditional average treatment effect* (CATE) (cf. Abrevaya, Hsu and Lieli 2015)

$$CATE = E(Y(1) - Y(0)|X = x) \tag{1}$$

King and Nielsen (2019) formulate a general pruning (*matching*) function M:

$$X_\ell = M(X|A_\ell, T_i = 1, T_j = 0, \delta) \equiv M(X|A_\ell) \subseteq X \tag{2}$$

providing X_ℓ, subset of matched observation based on condition A_ℓ.

Notation

Consider that we aim to estimate *conditional average treatment effect* (CATE) (cf. Abrevaya, Hsu and Lieli 2015)

$$CATE = E(Y(1) - Y(0)|X = x) \tag{1}$$

King and Nielsen (2019) formulate a general pruning (*matching*) function M:

$$X_\ell = M(X|A_\ell, T_i = 1, T_j = 0, \delta) \equiv M(X|A_\ell) \subseteq X \tag{2}$$

providing X_ℓ, subset of matched observation based on condition A_ℓ.

$\rightarrow$ in what follows we will look at different pruning method ℓ
to produce the best matched subset δ.

Exact matching

For exact matching we find exactly equal pairs

$$X_{EM} = M(X|X_i = X_j) \tag{3}$$

Note: X can be a vector of covariates.

Coarsened Exact Matching (CEM)

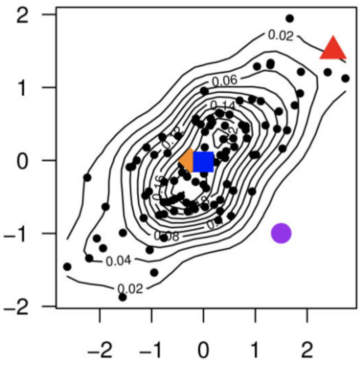
For coarsened exact matching we approximate

$$X_{CEM} = M(X|C_\delta(X_i) = C_\delta(X_i)) \tag{4}$$

where C_δ is a vector of same dimensions as X, but coarsened values, e.g. at ”*natural breakpoints*” such as years in one school type, levels of income, etc.

Mahalanobis Distance Method (MDM)

For multidimensional data, we can identify nearest neighbours in an n-dimensional space.



$$md(X_i, X_j) \quad = \{(X_i - X_j)^\top S^{-1}(X_i - X_j)\}^{\frac{1}{2}}$$

(Above) Mahalanobis distance measure, where S denotes the covariance matrix of X. [24]

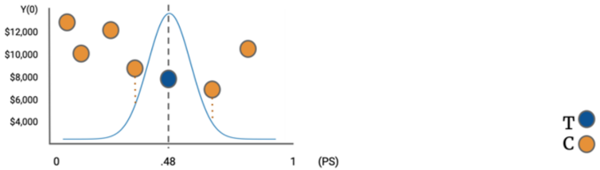
(Left) A contour plot is overlaid on a Mahalanobis distance scatter plot of 100 observations randomly drawn from a bivariate normal distribution. The centroid, in blue, is the reference point for distance between two points.

Image credit and description: Statistics How To: Mahalanobis Distance, Simple Definitions, Examples. Retrieved 10-08-2019 from: https://www.statisticshowto.datasciencecentral.com/mahalanobis-distance/

Image source: Sizemore and Alkurdi 2019

Propensity score matching (PSM)

Else, we can estimate probability of being treated, aka propensity score $\pi(X_i) = Pr(D_i = 1 | X_i)$ by logistic regression



Advantages	Disadvantages
solves matching problem for high dimensions	misspecification of PS model = bad matches
many available R packages for easy implementation	matched pairs may be dissimilar across X

Image source: Sizemore and Alkurdi 2019

example

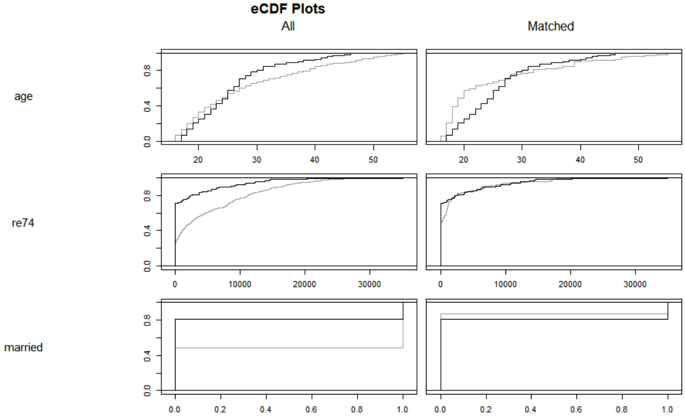

```
library(tidyverse)
library(MatchIt)

data("lalonde")
lalonde <- lalonde %>% as_tibble()

m.out <- matchit(treat ~ age + educ + race + married +
                 nodegree + re74 + re75, data = lalonde,
                 method = "full")
```

example


```
> m.out
A matchit object
 - method: Optimal full matching
 - distance: Propensity score
             - estimated with logistic regression
 - number of obs.: 614 (original), 614 (matched)
 - target estimand: ATT
 - covariates: age, educ, race, married, nodegree, re74, re75
```

example


```
plot(m.out, type = "ecdf", which.xs = c("age", "re74", "married")
```



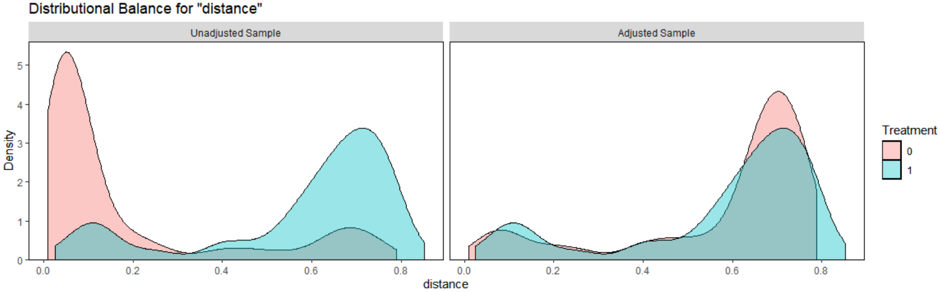
Code source: Greifer 2020

example


```
psFormula <- formula(treat ~ age + educ + race
                     + married + nodegree + re74 + re75)

lalonde$p.score <-
  glm(psFormula, data = lalonde,
      family = "binomial")$fitted.values

lalonde$att.weights <-
  with(lalonde, treat + (1-treat)*p.score/(1-p.score))
```

example


```
bal.plot(f.build("treat", covs0),
         data = lalonde, var.name = "p.score",
         weights = "att.weights", distance = "p.score",
         method = "weighting", which = "both")
```



Code source: Greifer 2020

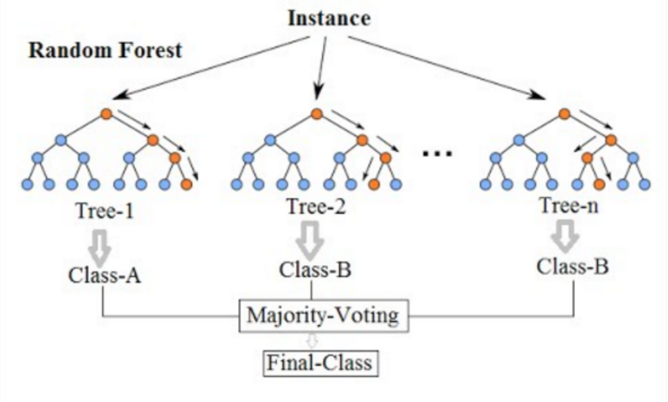
Intermediate discussion

There is a bit of critique on PSM

- King and Nielsen (2019)
  - *"PSM is ... uniquely blind to the often large portion of imbalance"*
  - *"easy to avoid by switching to one of the other popular methods of matching"*
  - i.e.: CEM and MDM
- Sizemore and Alkurdi (2019)
  - test PSM against machine learning based methods
  - logistic PSM $\succ$ random forest PSM $\succ$ genetic matching
  - CEM ???

Random forest (RF)

RF are multiple regression trees classifying the data by partitioning

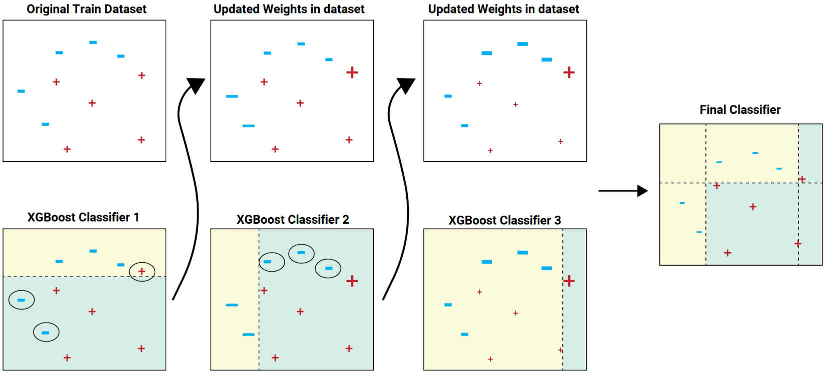


Code source: [Wikipedia]

We can use this to predict treatment (aka propensity scores)

eXtreme Gradient Boosting (XGBoost)

Machine learning such as XGBoost or even ensambles can also be used to



Code source: Quant Insti

→ predict treatment (aka propensity scores)

Genetic matching

Genetic Matching combines PSM and MDM

$$GMD(X_i, X_j, W) = \sqrt{(X_i)^T (S^{-\frac{1}{2}})^T W S^{-\frac{1}{2}} (X_i - X_j)} \tag{5}$$

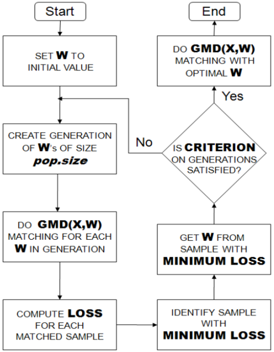
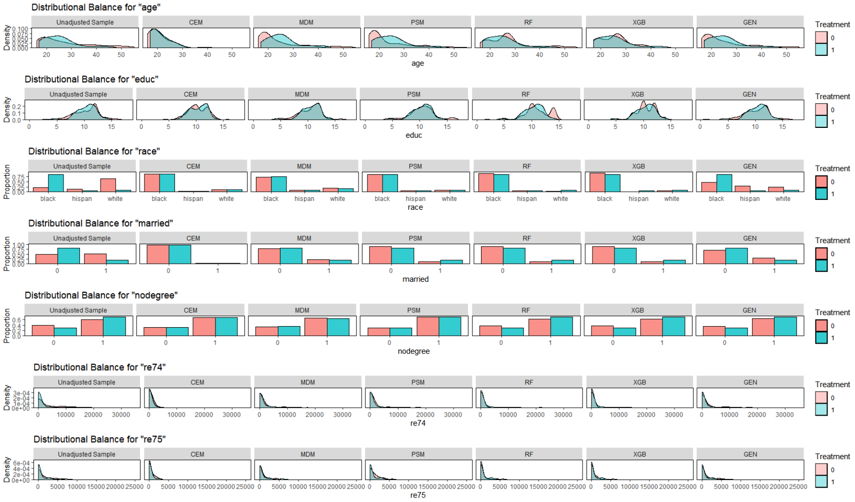


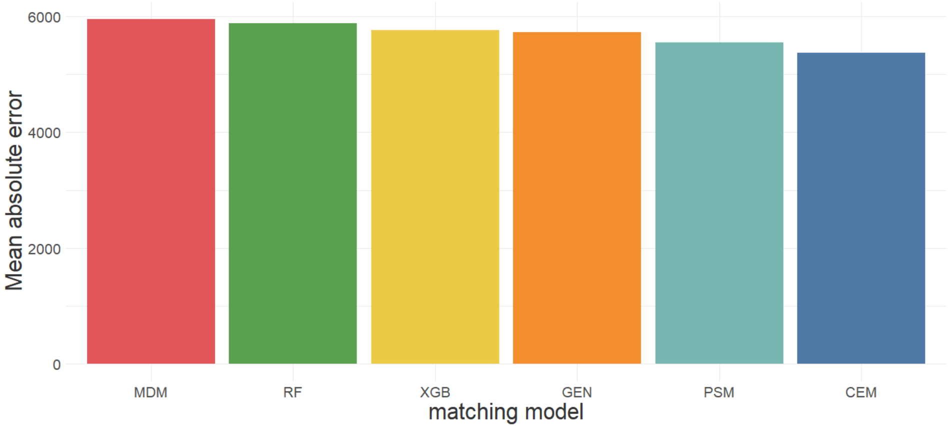
Image source: Sizemore and Alkurdi 2019

comparison - fitting distributions



plotting model comparisons for coviarates of the lalonde data set

comparison - mean absolute error



plotting model comparisons for the lalonde data set, cf. Colson et al. 2016

comparison - summary

- for the comparison above I used nearest neighbour matching, reducing sample size

comparison - summary

- for the comparison above I used nearest neighbour matching, reducing sample size
- maximizing post-match balance does not necessarily improve explanatory model power (Colson et al. 2016)

comparison - summary

- for the comparison above I used nearest neighbour matching, reducing sample size
- maximizing post-match balance does not necessarily improve explanatory model power (Colson et al. 2016)
- possibly both sample size and balance need to be taken into account (King, Lucas and R. A. Nielsen 2017)

comparison - summary

- for the comparison above I used nearest neighbour matching, reducing sample size
- maximizing post-match balance does not necessarily improve explanatory model power (Colson et al. 2016)
- possibly both sample size and balance need to be taken into account (King, Lucas and R. A. Nielsen 2017)
- latest approaches include almost exact matching (Dieng et al. 2018a; Dieng et al. 2018b), text matching (Roberts, Stewart and R. A. Nielsen 2020), generalized optimal matching (Kallus 2020)

comparison - summary

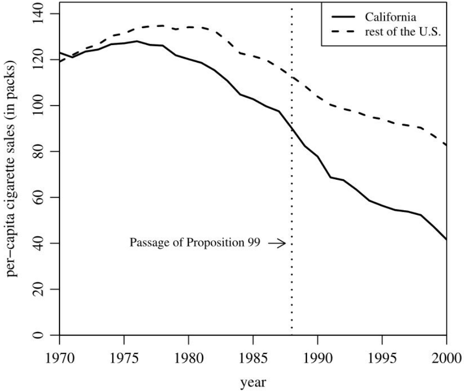
- for the comparison above I used nearest neighbour matching, reducing sample size
- maximizing post-match balance does not necessarily improve explanatory model power (Colson et al. 2016)
- possibly both sample size and balance need to be taken into account (King, Lucas and R. A. Nielsen 2017)
- latest approaches include almost exact matching (Dieng et al. 2018a; Dieng et al. 2018b), text matching (Roberts, Stewart and R. A. Nielsen 2020), generalized optimal matching (Kallus 2020)
- R packages include MatchIt, Matching, and PanelMatch

comparison - summary

- for the comparison above I used nearest neighbour matching, reducing sample size
- maximizing post-match balance does not necessarily improve explanatory model power (Colson et al. 2016)
- possibly both sample size and balance need to be taken into account (King, Lucas and R. A. Nielsen 2017)
- latest approaches include almost exact matching (Dieng et al. 2018a; Dieng et al. 2018b), text matching (Roberts, Stewart and R. A. Nielsen 2020), generalized optimal matching (Kallus 2020)
- R packages include MatchIt, Matching, and PanelMatch
- for the debate around propensity score matching (King and R. Nielsen 2019), see also Hünermund, (2019)

a case

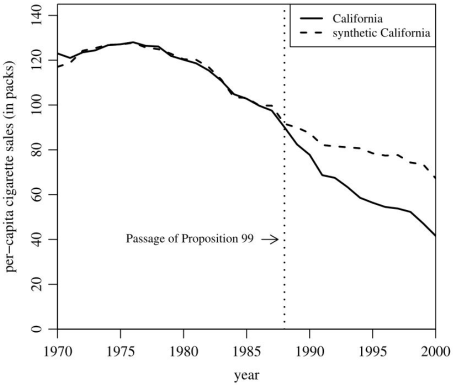
What if we do only have one treated unit?



California introduces tobacco control in 1988, cf. Abadie et al. 2010

and an idea

How about we compare to a weighted average of untreated?



California introduces tobacco control in 1988, cf. Abadie et al. 2010

and a notation

$$\hat{Y}_{t,post(0)} = \mu + \sum_{i=1}^{N} w_i Y_{i,T}^{obs} \tag{6}$$

”In other words, the imputed control outcome for the treated unit is a linear combination of the control units, with intercept μ and weights w_i for control unit i.” Doudchenko and Imbens 2016

the process

We compare the treated to the non-treated

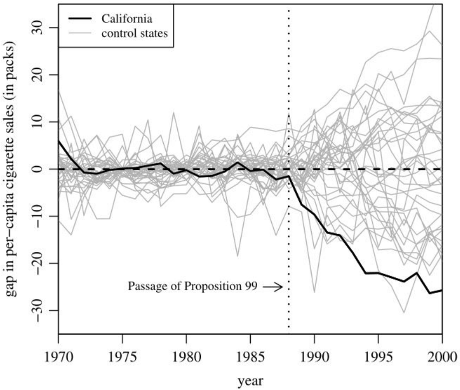


Figure 5. Per-capita cigarette sales gaps in California and placebo gaps in 34 control states (discards states with pre-Proposition 99 MSPE twenty times higher than California's).

A noisy control group, cf. Abadie et al. 2010

the process

And compute a synthetic control out of a weighted set of the untreated

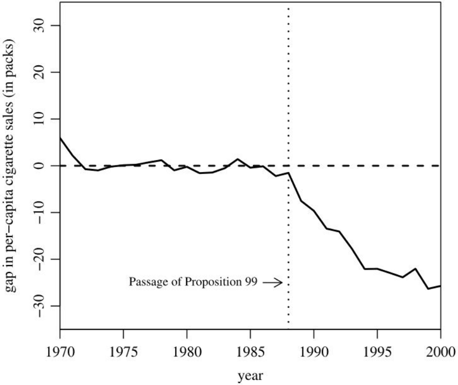


Figure 3. Per-capita cigarette sales gap between California and synthetic California.

California vs SynthCal, cf. Abadie et al. 2010

References I

Abadie, Alberto et al. (2010). 'Synthetic control methods for comparative case studies: Estimating the effect of California's Tobacco control program'. In: *Journal of the American Statistical Association* 105.490, pp. 493–505. ISSN: 01621459. DOI: 10.1198/jasa.2009.ap08746.

Abrevaya, Jason, Yu Chin Hsu and Robert P. Lieli (2015). 'Estimating Conditional Average Treatment Effects'. In: *Journal of Business and Economic Statistics* 33.4, pp. 485–505. ISSN: 15372707. DOI: 10.1080/07350015.2014.975555.

Colson, K. Ellicott et al. (2016). 'Optimizing matching and analysis combinations for estimating causal effects'. In: *Scientific Reports* 6.March, pp. 1–11. DOI: 10.1038/srep23222. URL: http://dx.doi.org/10.1038/srep23222.

Dieng, Awa et al. (2018a). 'Almost-Exact Matching with Replacement for Causal Inference'. In: *arXiv*, pp. 1–28. arXiv: 1806.06802. URL: http://arxiv.org/abs/1806.06802.

— (2018b). 'Collapsing-Fast-Large-Almost-Matching-Exactly: A Matching Method for Causal Inference'. In: *arXiv*, pp. 1–27. arXiv: 1806.06802. URL: https://arxiv.org/pdf/1806.06802.pdf.

Doudchenko, Nikolay and Guido W Imbens (2016). 'Balancing, Regression, Difference-In-Differences and Synthetic Control Methods: A Synthesis'. URL: http://www.nber.org/papers/w22791.

Kallus, Nathan (2020). 'Generalized optimal matching methods for causal inference'. In: *Journal of Machine Learning Research* 21, pp. 1–54. ISSN: 15337928. arXiv: 1612.08321.

References II

King, Gary, Christopher Lucas and Richard A. Nielsen (2017). 'The Balance-Sample Size Frontier in Matching Methods for Causal Inference'. In: *American Journal of Political Science* 61.2, pp. 473–489. DOI: 10.1111/ajps.12272.

King, Gary and Richard Nielsen (2019). 'Why Propensity Scores Should Not Be Used for Matching'. In: *Political Analysis* 27.4, pp. 435–454. ISSN: 14764989. DOI: 10.1017/pan.2019.11.

Roberts, Margaret E., Brandon M. Stewart and Richard A. Nielsen (2020). 'Adjusting for Confounding with Text Matching'. In: *American Journal of Political Science* 64.4, pp. 887–903. DOI: 10.1111/ajps.12526.

Schleicher, Judith et al. (2020). 'Statistical matching for conservation science'. In: *Conservation Biology* 34.3, pp. 538–549. ISSN: 15231739. DOI: 10.1111/cobi.13448.

Sizemore, Samantha and Raiber Alkurdi (2019). *Matching Methods for Causal Inference: A Machine Learning Update*. URL: https://humboldt-wi.github.io/blog/research/applied%7B%5C_%7Dpredictive%7B%5C_%7Dmodeling%7B%5C_%7D19/matching%7B%5C_%7Dmethods/ (visited on 01/05/2021).