

# A Brief Introduction to the Synthetic Control Method

Social Care Statistics Working Group

18/07/2023

Gintare Malisauskaite  
Personal Social Services Research Unit (PSSRU), University of Kent  
E-mail: [G.Malisauskaite@kent.ac.uk](mailto:G.Malisauskaite@kent.ac.uk)

# Introduction

- Proposed by Abadie and Gardeazabal (2003) and Abadie et al. (2010)
- Main principles of the Synthetic Control Method
- Application in Stata
- Examples
- Conclusions



# Main principles of the Synthetic Control Method

- Estimation of the treatment effect in econometric analysis is based on comparing the outcomes of the treated unit(s) with untreated units
  - However untreated units may be too different from the treated unit leaving the estimates biased
- This method offers constructing a synthetic control unit that mimics the treated unit prior to the treatment and predicts the outcome path post treatment assuming the unit is untreated
- Essentially, it's an estimation of what would have happened with the outcome variable if the treatment hadn't happened
- Synthetic control unit is constructed giving different weights between 0 and 1 (which also sum to 1) to untreated units to create a unit mimicking the treated unit based on a list of variables defining them
- Only most similar control units are given weights to construct synthetic unit



# Method

- Donor pool restrictions: units affected by the treatment or by events of a similar nature should be excluded from the donor pool
- Also units with large idiosyncratic shocks to the outcome of interest should also be excluded
- Requires sufficient data prior to treatment: credibility of a synthetic control depends upon how well it tracks the treated unit's characteristics and outcomes over an extended period of time prior to the treatment
- A sizeable number of post-intervention periods needed in cases when the effect of the intervention is gradual or changes over time
- Placebo tests
- Robustness checks: excluding units that received a positive weight



# Example I: Cigarette sales California and US, 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      |

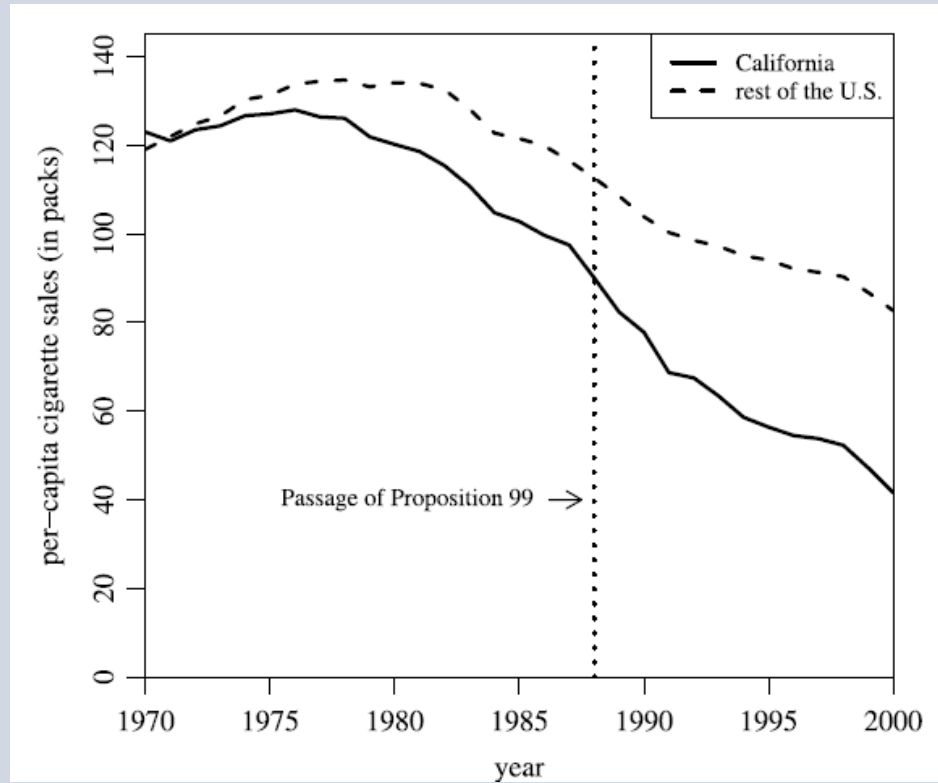
Table 1. Cigarette sales predictor means

| Variables                       | California |           | Average of<br>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                          |

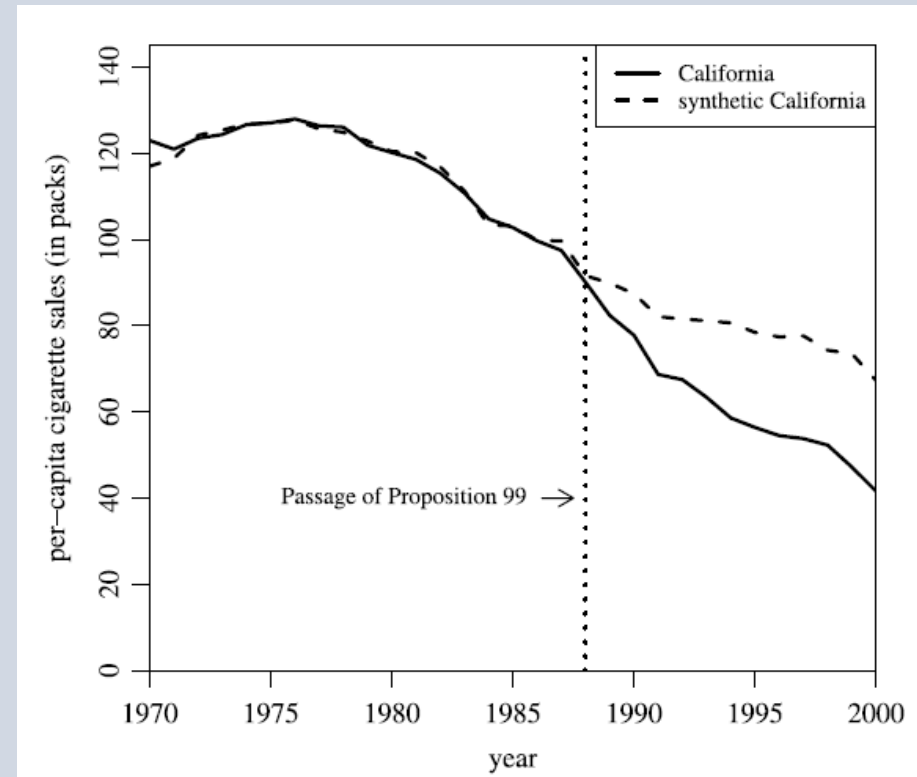
NOTE: All variables except lagged cigarette sales are averaged for the 1980–1988 period (beer consumption is averaged 1984–1988). GDP per capita is measured in 1997 dollars, retail prices are measured in cents, beer consumption is measured in gallons, and cigarette sales are measured in packs.



# Example I continued



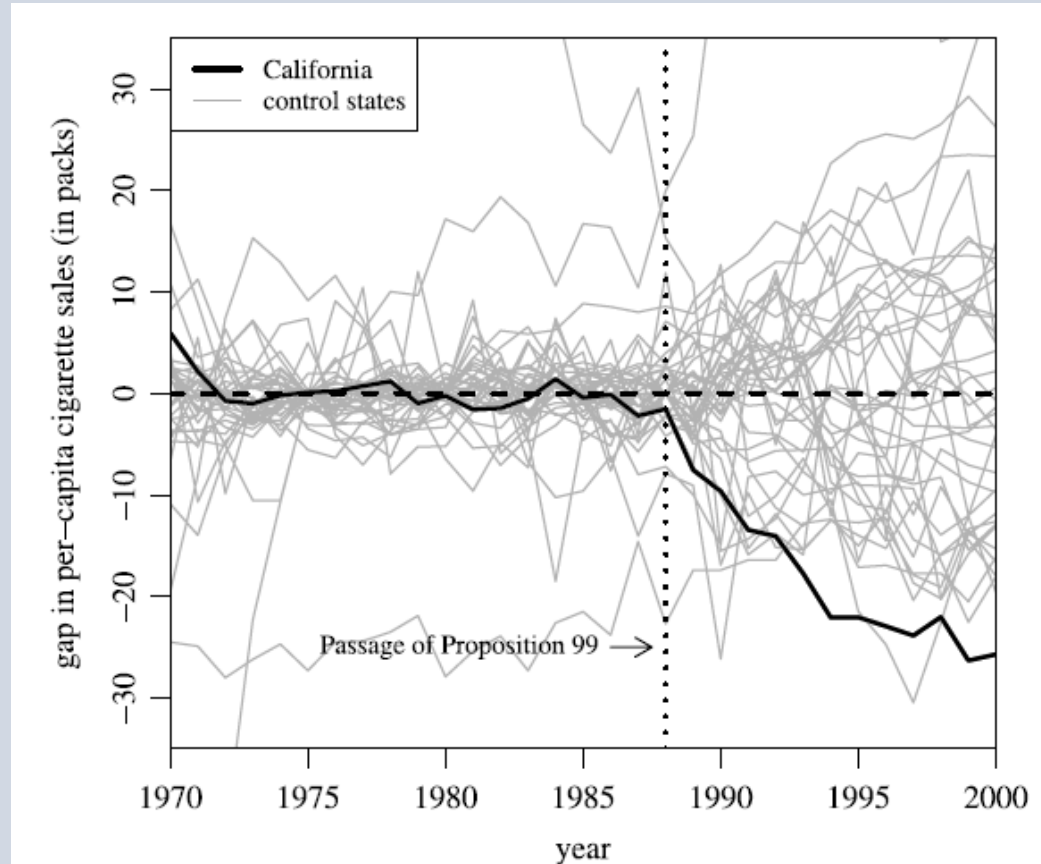
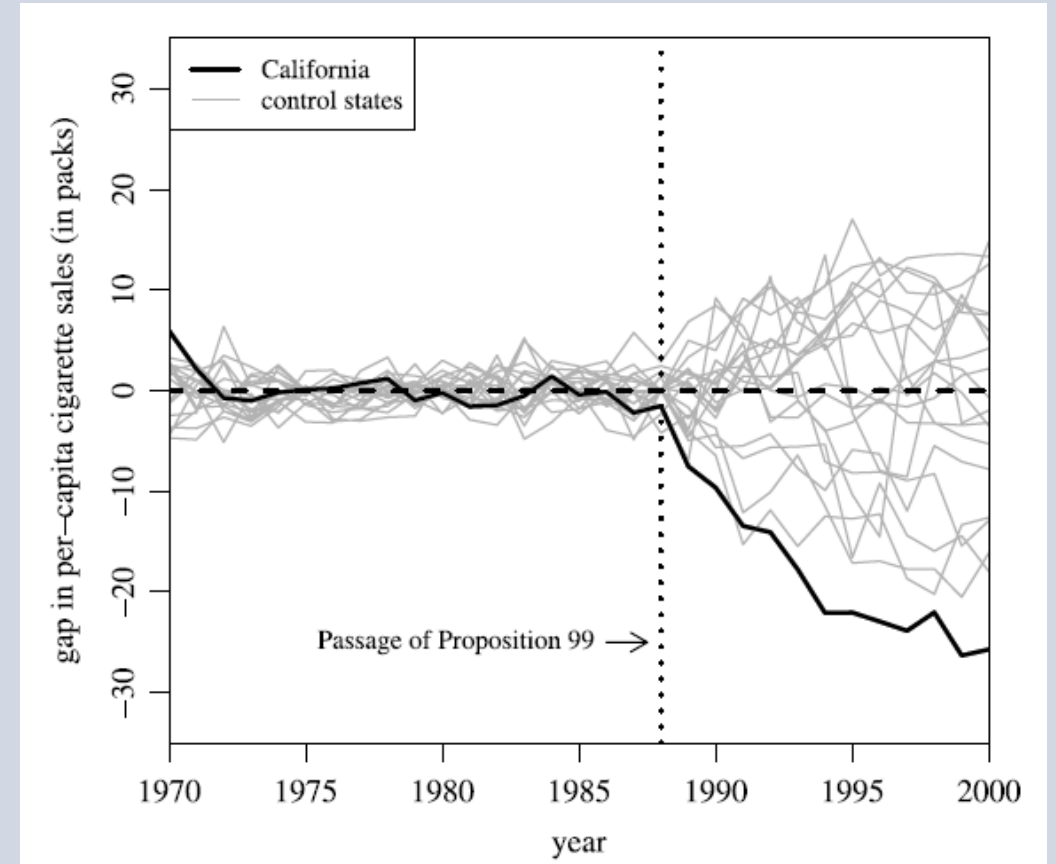
- Trends in per-capita cigarette sales: California vs. the rest of the United States.



- Trends in per-capita cigarette sales: California vs. Synthetic California.



# Example I continued: Placebo tests



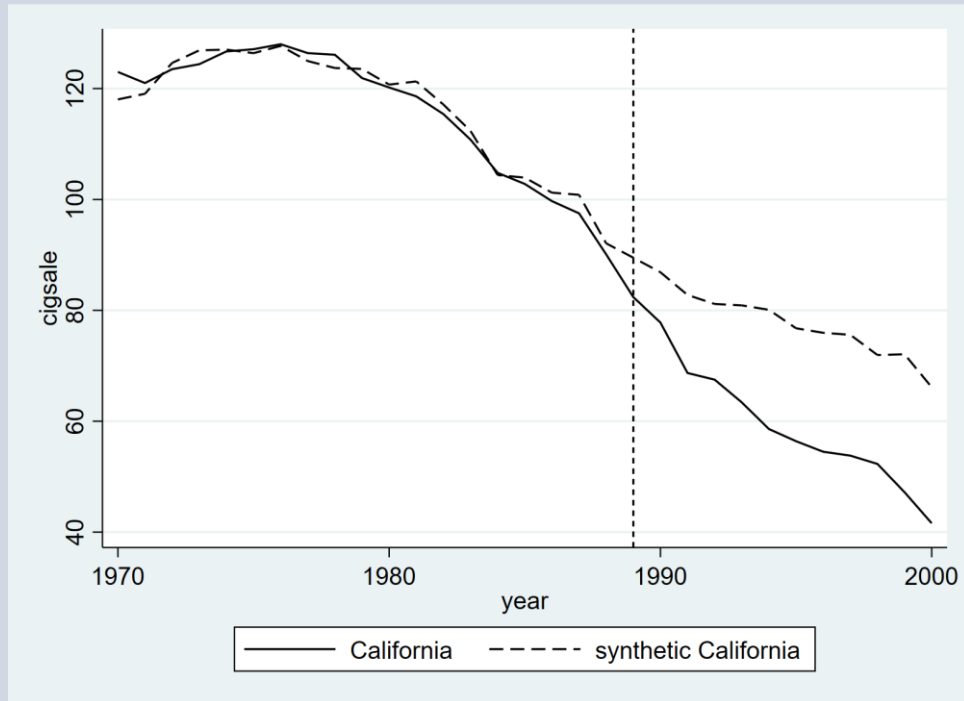
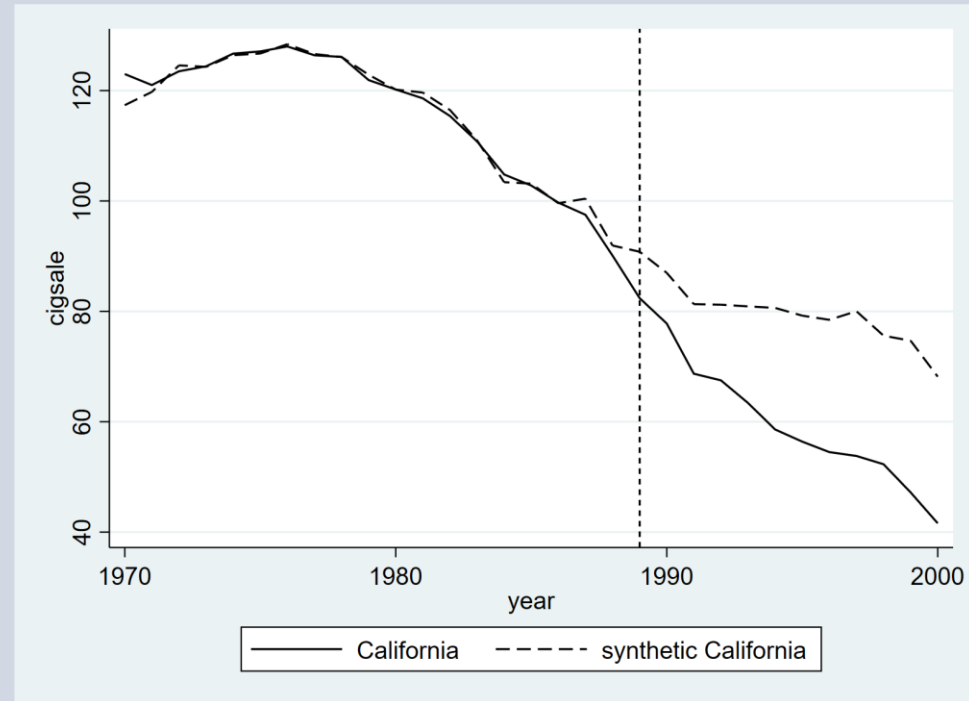
- 38 control states vs 19

# Stata help

- Command *synth* from Synth package developed by Jens Hainmueller available for Mat-lab, R, and Stata - Abadie et al. (2010)
- *synth\_runner*, builds on *synth* to help conduct multiple estimations, inference, diagnostics, help generate visualizations of results; designed to accompany *synth* but not supersede it, Stata help document - Galiani & Quistorff (2017)
- also *allsynth* – built on Synth package and offers bias-correction for inexact matching on predictor values, automated calculation of RMPSE p-values, expanded graphing functionality, uniqueness diagnostics – Wiltshire (2021)
- Things to note:
  - Control for the outcome variable prior to treatment
  - Output does not produce coefficients  $\beta$  or standard errors, only predicted outcome points in time



# Stata output



- `synth cigsale lncincome(1980(1)1988) age15to24(1980(1)1988) \\\`  
`retprice(1980(1)1988) beer(1984(1)1988) cigsale(1988) cigsale(1980) \\\`  
`cigsale(1975), trunit(3) trperiod(1989) fig`
- Abadie et. al. (2010)

- `synth cigsale cigsale(1988) cigsale(1987) cigsale(1986) \\\`  
`cigsale(1985) cigsale(1984) cigsale(1983) cigsale(1982) cigsale(1981) \\\`  
`cigsale(1980) cigsale(1979) cigsale(1978) cigsale(1977) cigsale(1976) \\\`  
`cigsale(1975) cigsale(1974) cigsale(1973) cigsale(1972) cigsale(1971) \\\`  
`cigsale(1970), trunit(3) trperiod(1989) fig`
- Lu (2021)

# Stata output: Galiani & Quistorff (2017)

```
. synth_runner cigsale beer(1984(1)1988) lnincome(1972(1)1988)
>   retprice age15to24 cigsale(1988) cigsale(1980) cigsale(1975),
>   trunit(3) trperiod(1989) gen_vars
Estimating the treatment effects
Estimating the possible placebo effects (one set for each of the 1 treatment
> periods)
|                                     | Total: 38
.....| 11.00s elapsed.

Conducting inference: 5 steps, and 38 placebo averages
Step 1... Finished
Step 2... Finished
Step 3... Finished
Step 4... Finished
Step 5... Finished

Post-treatment results: Effects, p-values, standardized p-values
```

|     | estimates | pvals    | pvals_std |
|-----|-----------|----------|-----------|
| c1  | -7.887098 | .1315789 | 0         |
| c2  | -9.693599 | .1842105 | 0         |
| c3  | -13.8027  | .2105263 | 0         |
| c4  | -13.344   | .1315789 | 0         |
| c5  | -17.0624  | .1052632 | 0         |
| c6  | -20.8943  | .0789474 | 0         |
| c7  | -19.8568  | .1315789 | .0263158  |
| c8  | -21.0405  | .1578947 | 0         |
| c9  | -21.4914  | .1052632 | .0263158  |
| c10 | -19.1642  | .1842105 | .0263158  |
| c11 | -24.554   | .1052632 | 0         |
| c12 | -24.2687  | .1052632 | .0263158  |



# Other examples: Economic impact of the 1990 German reunification on West Germany – Abadie et al (2015)

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

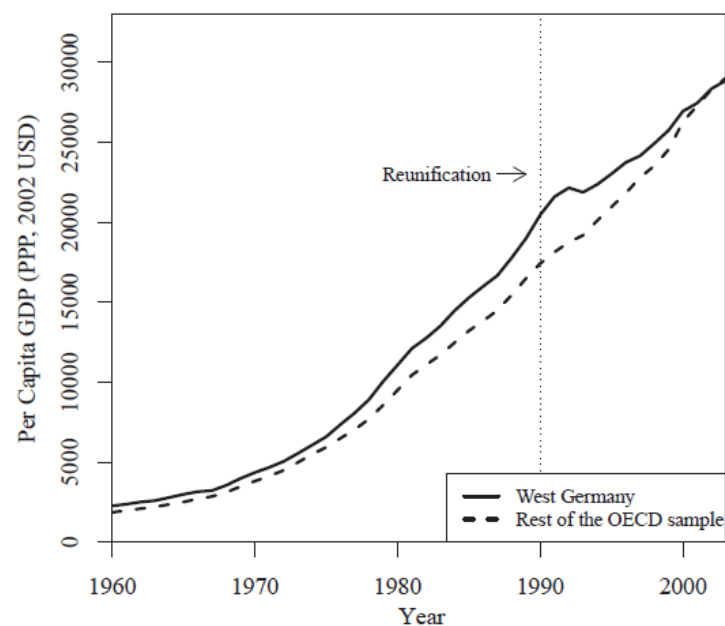


**FIGURE 2 Trends in per Capita GDP: West Germany versus Synthetic West Germany**

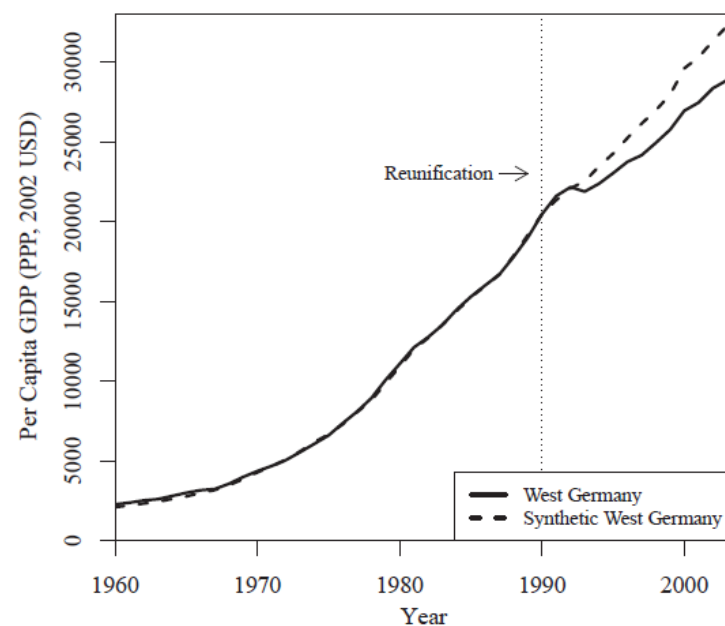
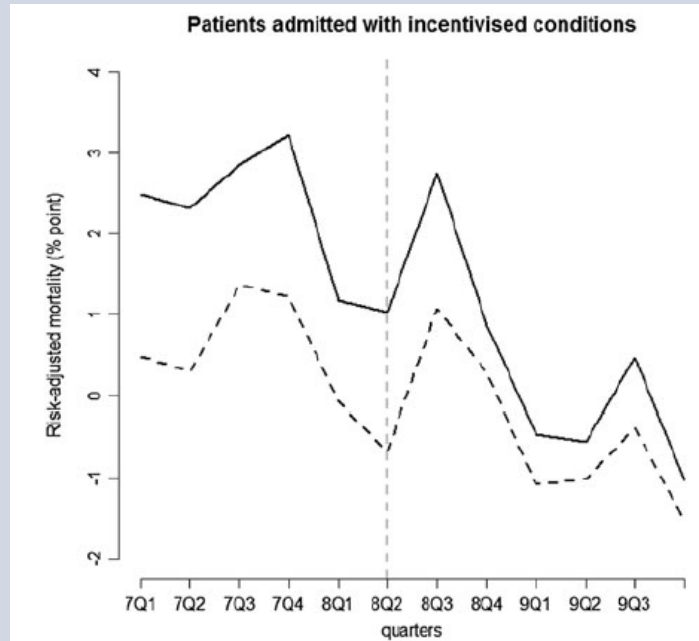
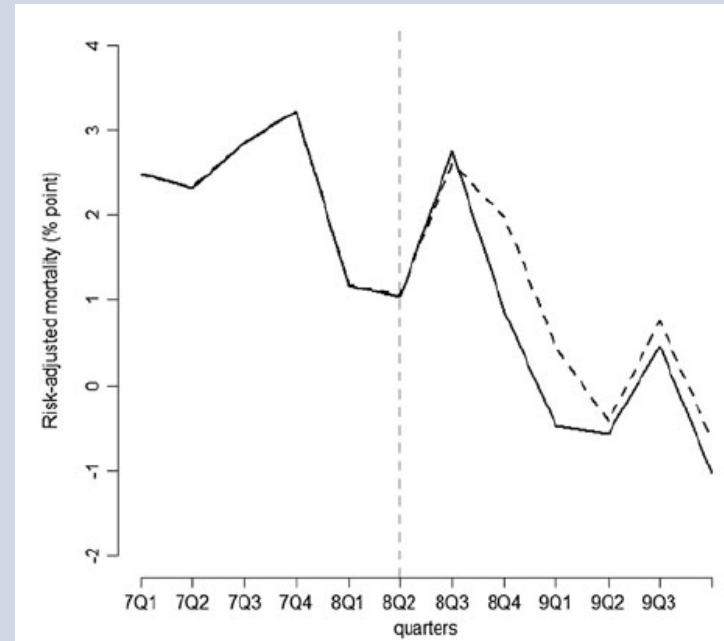


**TABLE 2 Economic Growth Predictor Means before German Reunification**

|                 | 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        |



# Other examples: Hospital pay-for-performance scheme in North West England and hospital mortality: multiple treated units – Kreif et al. (2015)



- Parallel trends assumption cannot be rejected

Table II. Estimated ATTs: DiD vs. synthetic control method (base case)

|                                 | Standard DiD ATT ( $p$ -value <sup>a</sup> ) | Synthetic control ATT ( $p$ -value <sup>b</sup> ) |
|---------------------------------|----------------------------------------------|---------------------------------------------------|
| All incentivised                | −0.88 (<0.01)                                | −0.45 (0.21)                                      |
| Pneumonia                       | −1.62 (<0.01)                                | −0.43 (0.45)                                      |
| Heart failure                   | −0.29 (0.52)                                 | 0.74 (0.18)                                       |
| Acute myocardial infarction     | −0.28 (0.44)                                 | −0.49 (0.15)                                      |
| All non-incentivised conditions | 0.39 (0.30)                                  | 0.90 (0.02)                                       |

After discarding 10 hospitals that did not contribute observations to each quarter, we apply both methods on a balanced panel of 24 treated and 122 potential control hospitals.

<sup>a</sup> $p$ -values based on standard errors from fixed effects model, clustered by hospital.

<sup>b</sup> $p$ -values calculated based on placebo tests.

# Other examples: BCF evaluation - Forder et al. (2018)

Figure 12. Synthetic control results – delayed transfers of care: delayed days per person 65+ (log) (Spec 1)

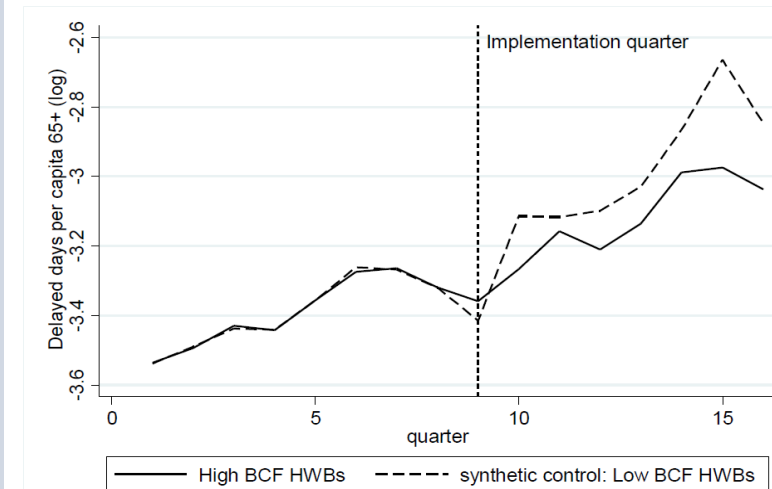


Figure 15. Synthetic control results – delayed transfers of care: delayed days per person 65+ (log) (Spec 2)

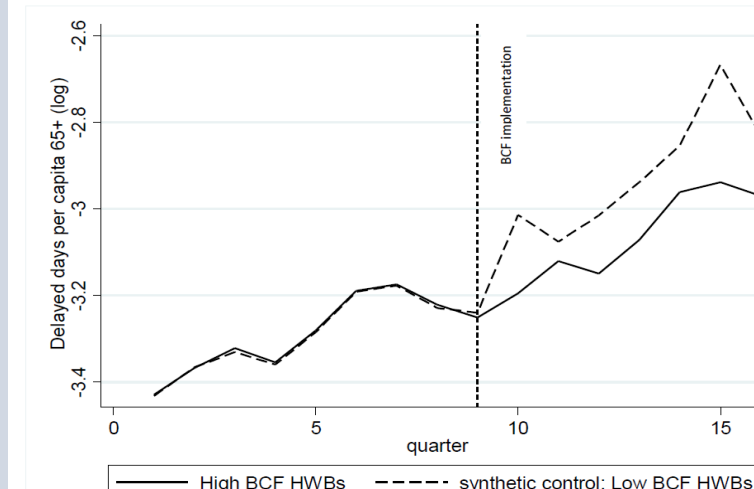


Figure 13. Synthetic control results – delayed transfers of care: delayed days per person 65+

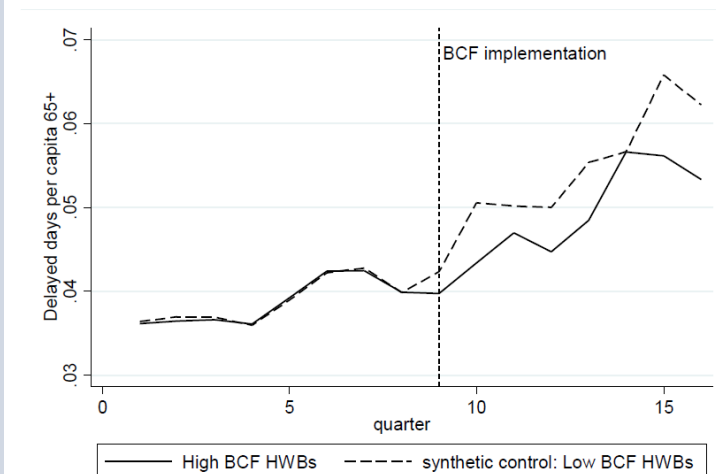
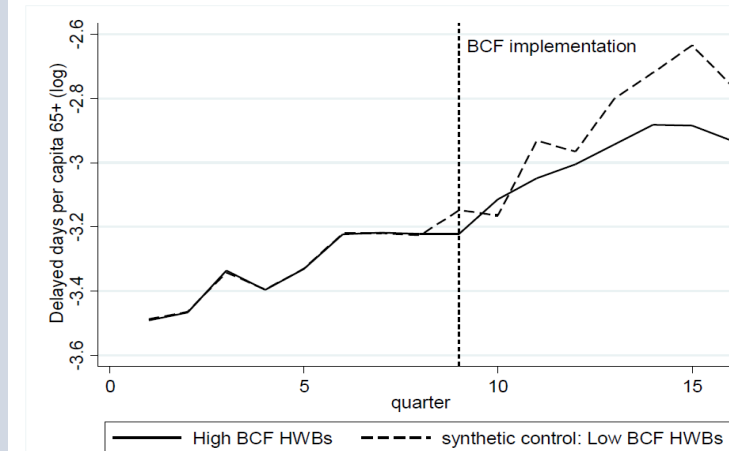


Figure 14. Synthetic control results – delayed transfers of care: delayed days per person 65+, defined at 75<sup>th</sup> percentile of BCF expenditure distribution



- Fig 12-13: Treated-above median BCF expenditure LAs; synth from lower than median half sample.
- Fig 14-15: Treated-top 25<sup>th</sup> percentile BCF expenditure LAs; synth from lower 75% of sample.

# Other examples: Urgent and Emergency Care Vanguard on DTOC - Malisauskaite et al. (2022)

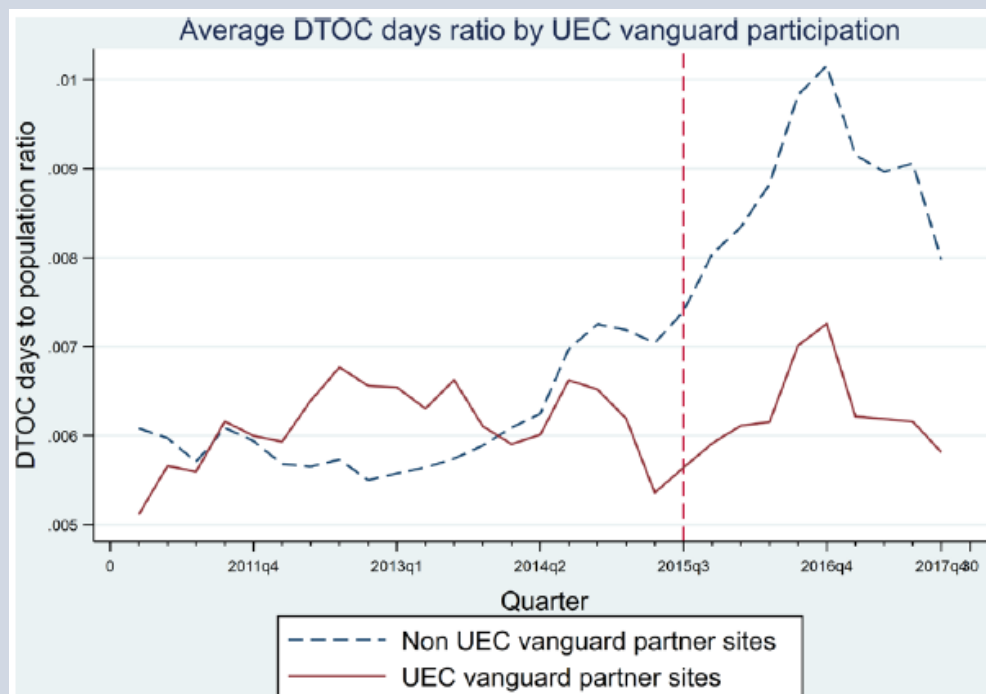


**Figure 1** Average DTOC days over time. Note: UEC vanguard partner sites include 29 LAs.<sup>14</sup>

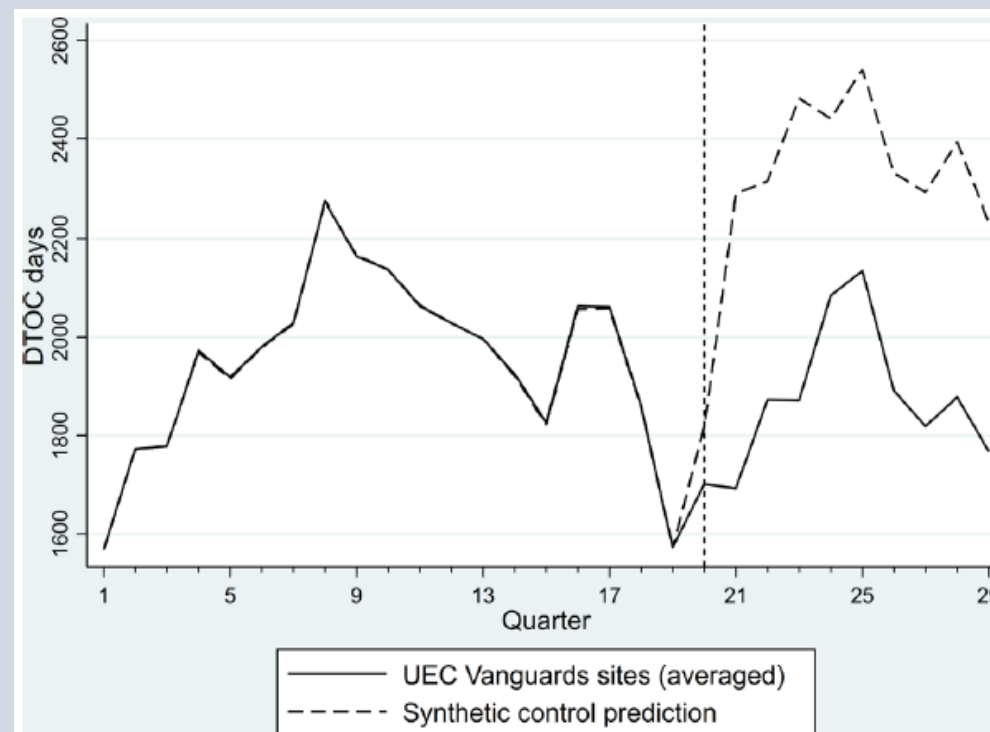


**Figure 2** Synthetic control estimation.



# Comparison with other methods

- In contrast to a difference-in-differences design, SCM does not give all untreated units the same weight in the comparison
- Regression estimator can be expressed as a weighted average of comparison units, and restricted for the weights to sum to one, but weights not restricted to be between 0 and 1
- In SCM contribution of each control unit is explicit and comparison made only with the most similar units
- Regression is based on extrapolation, while SCM avoids it
- DiD – parallel trends assumption, SCM useful when it doesn't hold



# Conclusions

- SCM useful when no single untreated unit provides a good comparison for the unit affected by the treatment or event of interest.
- Or when other methods are not appropriate
- An insight into what could have been without the treatment without extrapolation
- SCM is great for visual representation of the treatment effect
- Does not offer beta coefficients and their std. errors
- Becoming more popular, but literature still scarce



# References

- Abadie, A. (2021). "Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects." *Journal of Economic Literature*, 59(2), 391-425
- 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.
- Abadie, A., Diamond, A., & Hainmueller, J. (2015). "Comparative Politics and the Synthetic Control Method." *American Journal of Political Science*, 59(2), 495-510.
- Abadie, A., and Gardeazabal J. (2003). The economic costs of conflict: A case study of the Basque country. *American Economic Review* 93: 113–132.
- Chamberlain Seminar: Tutorial on Synthetic Control Methods (18/06/2021) <https://www.chamberlainseminar.org/past-seminars/spring-2021#h.pj7bcv69aoxc>
- Forder, J., Caiels, J., Harlock, J., Wistow, G., Malisauskaite, G., Peters, M., Marczak, J., D'Amico, F., Fernandez, J.L., Fitzpatrick, R., Jones, K. (2018). A system-level evaluation of the Better Care Fund: Final Report, [online], 1-140. Available at: <https://www.pssru.ac.uk/publications/pub-5424/>
- Galiani S, Quistorff B, (2017). The synth runner package: Utilities to automate synthetic control estimation using synth. *Stata Journal* 17(4):834-849.
- Hsiao, C., Ching, H. S., & Wan, S. K. (2012). "A Panel Data Approach for Program Evaluation: Measuring the Benefits of Political and Economic Integration of Hong Kong with Mainland China." *Journal of Applied Econometrics*, 27(5), 705-740.
- Kreif N, Grieve R, Hangartner D, Turner AJ, Nikolova S, Sutton M. (2015). Examination of the Synthetic Control Method for Evaluating Health Policies with Multiple Treated Units. *Health Economics* 25(12):1514-1528.
- Lu J. Synthetic Control Method (20/08/2021) [https://www.stata.com/meeting/china21-Uone-Tech/slides/China21\\_Jiaxuan.pdf](https://www.stata.com/meeting/china21-Uone-Tech/slides/China21_Jiaxuan.pdf)
- Malisauskaite, G., Jones, K., Allan, S., Roland, D., Birks, Y., Baxter, K., Gridley, K. (2022). How local partnerships to improve urgent and emergency care have impacted delayed transfers of care from hospitals in England: an analysis based on a synthetic control estimation method. *BMJ Open*, 12(2):e054568. DOI: [10.1136/bmjopen-2021-054568](https://doi.org/10.1136/bmjopen-2021-054568)
- Wiltshire J. C. allsynth: Synthetic Control Bias-Correction Utilities for Stata (05/08/2021) [https://www.stata.com/meeting/us21/slides/US21\\_Wiltshire.pdf](https://www.stata.com/meeting/us21/slides/US21_Wiltshire.pdf)



# Thank you!

- E-mail: [G.Malisauskaite@kent.ac.uk](mailto:G.Malisauskaite@kent.ac.uk)

