# Further selected topics

Christoph Hanck

Summer 2024

# Synthetic control

# Synthetic control

## Definition: Synthetic control

Synthetic control is a variant of **DiD with matching**. The same basic setup as for DiD is used with a policy that goes into effect at a particular time, but only for a particular group:

- The data from the pre-treatment period is used to adjust for differences between the treatment and control groups. We then check how they differ after treatment goes into effect.
- The post-treatment difference, adjusting for pre-treatment differences, is the effect.

# Synthetic control

## Synthetic Control vs. DiD

- The pre-treatment difference adjustments are not done with regression, but rather using matching.
- Unlike DiD with matching, the purpose of matching is to eliminate these prior differences, not to account for the propensity of treatment.
- Synthetic control relies on a long period of pre-treatment data.

# Synthetic control

## Synthetic Control vs. DiD

- After matching, the treated and control groups should have basically no pre-treatment differences. This is often accomplished by including the outcome variable as a matching variable.
- Statistical significance is generally not determined by figuring out the sampling distribution of the estimation method used beforehand, but rather by **randomization inference**:

  Randomization inference uses placebo tests to estimate a null distribution that can compared to the real estimate.

# Synthetic control

## Procedure

- Synthetic control starts with a treated group and a **donor set** of potential control groups.
- Using the pre-treatment data periods, it implements a matching algorithm that goes through all the control groups and assigns each of the potential controls a weight.
- These weights are designed such that the time trend of the outcome for the treated group should be almost exactly the same as the time trend of the outcome for the weighted average of the control group (the **synthetic control group**).

# Synthetic control

## Example: Effect of conflict on economic activity in Spain

- Abadie and Gardeazabal (2003) study the outbreak of violent conflict in the Basque region of Spain in the late 1960s to see what effect the conflict had on economic activity
- They compare growth in the Basque region to growth in a synthetic control group. They find that GDP in the Basque region was negatively impacted by the conflict.

# Synthetic control

## Example: Effect of conflict on economic activity in Spain

- Abadie and Gardeazabal match the Basque region to seventeen other nearby regions.
- The synthetic control algorithm creates weights for each of those seventeen regions in order to make prior trends equal, using information about those prior trends as well as other matching variables including population density, education, and investment levels.
- Cataluna and Madrid turn out to be the strongest matches and get the most weight, with the other regions getting very little weight. These weights are then applied across the entire time period.

# Synthetic control

## Example: Effect of conflict on economic activity in Spain

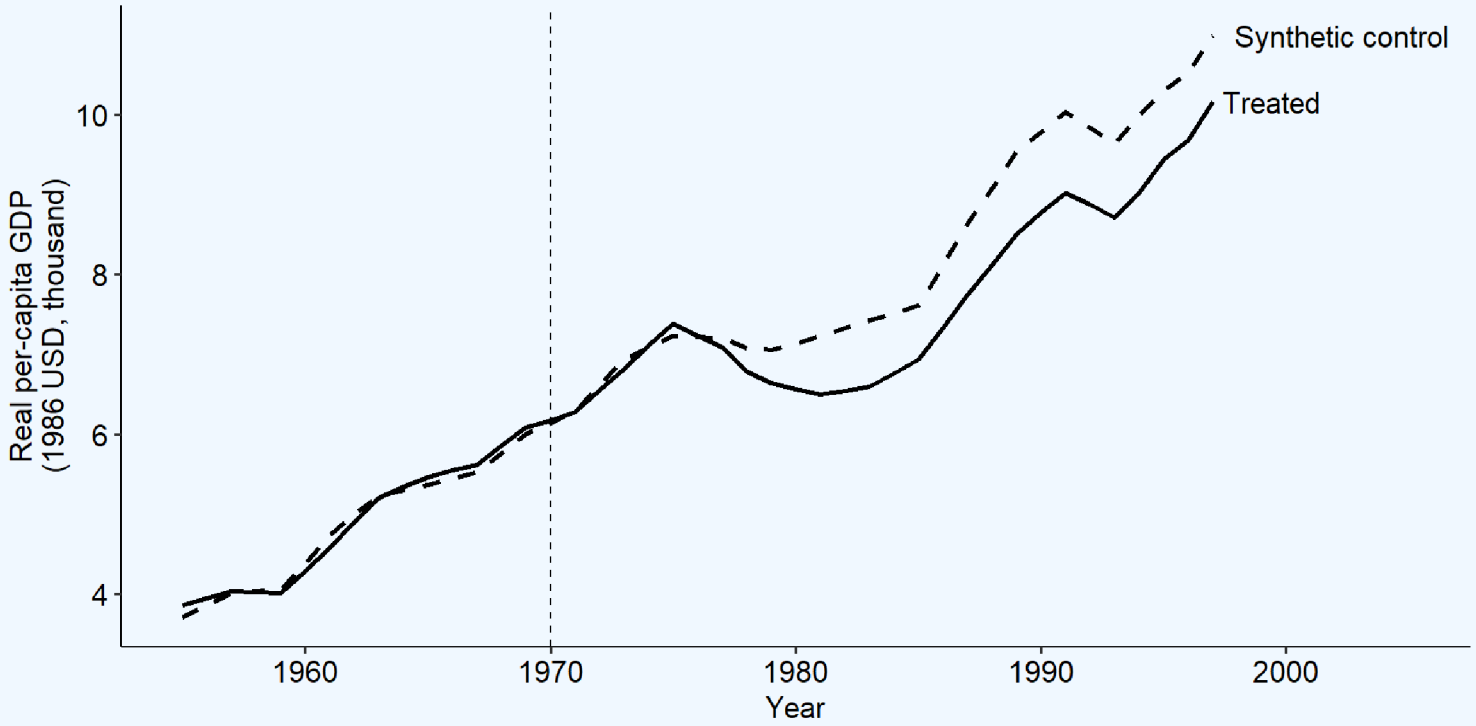


Figure 1: Synthetic control effect of conflict on Basque country economic growth

# Synthetic control

## Advantages

- Synthetic control does not need to rely on the parallel trends assumption as DiD does
- It makes the process of selecting a control group more disciplined
- Since synthetic control does not rely on regression, it is not as sensitive to functional form issues as DiD is
- It naturally shows dynamic effects without the need to estimate interaction terms

# Synthetic control

## Disadvantages

- Synthetic control is not as widely applicable as DiD: the matching quality will be poor if we do not have a lot of pre-treatment data
- It has a tendency to overfit noise in the outcome variable when matching
- Matching on the outcome variable seems problematic: using the outcome itself to close back doors feels a bit like cheating.
- That said, synthetic control only matches on pre-treatment outcomes, not the post-treatment outcomes where we actually estimate the effect