Synthetic Control Method

Kosuke Imai

Harvard University

Spring 2021

- Setting: $N$ units and $T$ time periods
- One treated unit $i = N$ receiving the treatment at time $T$
- Quantity of interest: $Y_{NT}(1) - Y_{NT}(0) = Y_{NT} - Y_{NT}(0)$
- Create a synthetic control using past outcomes
- Weighted average:

$$
Y_{NT}(0) = \sum_{i=1}^{N-1} \hat{w}_i Y_{iT}
$$

where the weights balance past outcomes

$$
\hat{w} = \arg\min_w \sum_{t=1}^{T-1} \left( Y_{Nt} - \sum_{i=1}^{N-1} w_i Y_{it} \right)^2
$$

with $\sum_{i=1}^{N-1} \hat{w}_i = 1$ and $\hat{w}_i \geq 0$
Causal Effect of ETA’s Terrorism

FIGURE 1. PER CAPITA GDP FOR THE BASQUE COUNTRY

Model-based Justification

- The main motivating factor analytic model:

  \[ Y_{it}(0) = \gamma_t + \delta_t^T X_i + \xi_t^T U_i + \epsilon_{it} \]

- Generalization of the linear two-way fixed effects model
- Key assumption: there exist weights such that

  \[ \sum_{i=1}^{N-1} w_i X_i = X_N \quad \text{and} \quad \sum_{i=1}^{N-1} w_i U_i = U_N \]

- Another motivating autoregressive model with time-varying covariates:

  \[ Y_{it}(0) = \rho_t Y_{i,t-1}(0) + \delta_t^T X_{it} + \epsilon_{it} \]
  \[ X_{it} = \lambda_{t-1} Y_{i,t-1}(0) + \Delta_{t-1} X_{i,t-1} + \nu_{it} \]

- Past outcomes can affect current treatment
- No unobserved time-invariant confounders
We can generalize the synthetic control method

Pre-treatment covariates: \( \mathbf{Z}_i = (\mathbf{Y}_i^T, \mathbf{X}_i^T)^T \)

- lagged outcomes: \( \mathbf{Y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{i,T-1})^T \)
- lagged covariates \( \mathbf{X}_i = (X_{i1}^T, X_{i2}^T, \ldots, X_{i,T-1}^T)^T \)

Or some subsets or functions of these variables

Balance both the lagged outcomes and pre-treatment covariates

\[
\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} \left( \mathbf{Z}_N - \sum_{i=1}^{N-1} w_i \mathbf{Z}_i \right)^T \hat{\Sigma}^{-1} \left( \mathbf{Z}_N - \sum_{i=1}^{N-1} w_i \mathbf{Z}_i \right)
\]

subject to \( \sum_{i=1}^{N-1} w_i = 1 \), and \( w_i \geq 0 \) for all \( i = 1, \ldots, N - 1 \),

where \( \hat{\Sigma} \) is the covariance matrix of \( \mathbf{Z}_i \)
can do this for all control units and compare them with the treated unit.
Permutation Test

- Assumption: Errors are exchangeable

We can invert this test to obtain a point-wise confidence interval
Relationship with Regression

- The synthetic control can be seen as a constrained regression:

\[
(\hat{\alpha}, \mathbf{w}) = \arg \min_{\alpha, \mathbf{w}} \sum_{t=1}^{T-1} \left( Y_{Nt} - \alpha - \sum_{i=1}^{N-1} w_i Y_{it} \right)^2
\]

subject to the following constraints

1. zero intercept: \( \alpha = 0 \)
2. positive weights \( w_i \geq 0 \) for all \( i = 1, \ldots, N - 1 \)
3. sum-to-one constraint: \( \sum_{i=1}^{N-1} w_i = 1 \)

- No time-invariant difference between the treated and control units
- Treated unit in the convex hull of the control units
- Regularization required when \( N \) is large relative to \( T \)