Heterogeneity-Robust Spatial Synthetic Controls

Daniel C Posmik [1]

[1]Becker Friedman Institute for Economic Research; University of Chicago
E-Mail: posmikdc@uchicago.edu

Table of Contents

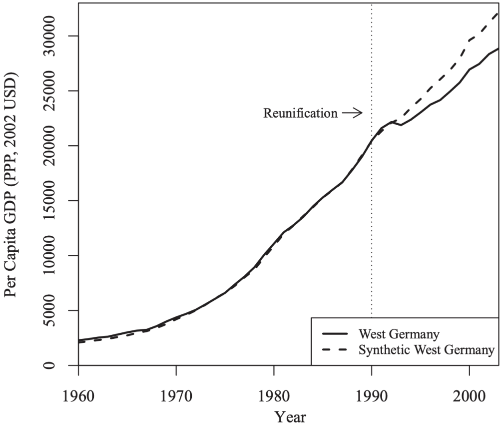
The Idea

Disclaimer

- This is a collection of preliminary ideas.
- The purpose of this presentation is to receive constructive criticism on many of these thoughts!
- My goal is to develop this project into a paper in graduate school!

Conventional Synthetic Controls

- Recover causal effects when no/ few untreated comparison units are available.
- Construct synthetic control units via pre-treatment/ exogenous characteristics
- Relies on parallel trends assumption



[Abadie et al., 2015]: German Reunification

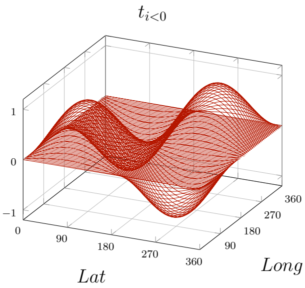
A Three-Dimensional Extension

- **Hypothesis**: The GDP effect of German reunification is not uniform across every state
  - Treatment effect heterogeneity?
  - Complex spatial relationships?
  - Information loss during data aggregation?

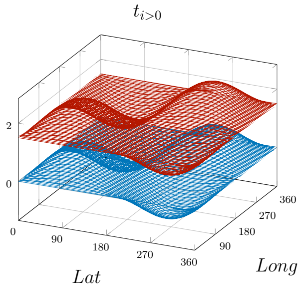
- Spatial Synthetic Controls (SSC) adds a spatial dimension to the *conventional* Synthetic Controls (SC) method
  - Disaggregation of information in the spatial dimension
  - Refine the average (ATE) to conditional (CATE) treatment effects
  - Causal inference in high-dimensional space

A Three-Dimensional Extension: Visualization

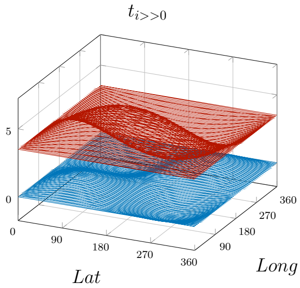
Treatment occurs at $t_{i=0}$. Let us consider three scenarios:



Pre-treatment, no treatment response



Post-treatment, homogeneous response



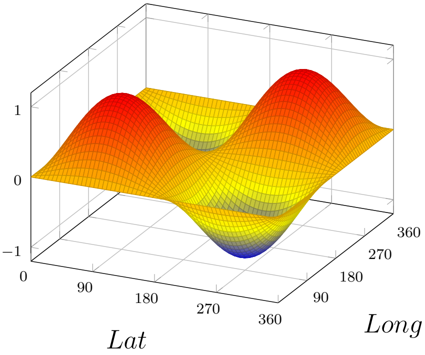
Post-treatment, heterogeneous response

Generating Synthetic Controls

Requirements for the Prediction Algorithm

1. Model non-linear relationships well
2. Robust to unobservables ("heterogeneity-robust") across spatial and temporal dimensions
3. Suitable for high-dimensional space

$t_{i<0}$, ..., $t_{i>0}$, ..., $t_{i>>0}$



Note: This problem is currently at the forefront of statistical research. If I had a definitive solution, I would be resting on my laurels right now.

Solutions?

Machine Learning	Non-Parametric	Dimensionality Reduction
Strong predictive performance in non-linear relationships	Robust to violations of functional form assumptions	Absorb temporal dimension into fixed effect.

Inspiration from Literature

- [Pouliot, 2022]
  - Krig-and-regress method for spatially misaligned data
  - Introduces Kriging (prediction) method into spatial econometrics
  - Relevant discussion on interpolation, aggregation, and spatial filtering techniques

- [Serra-Burriel et al., 2020]
  - Estimate the effects of wildfires on vegetation
  - An application of spatial synthetic controls, though treatment is discrete and units are static
  - Use of Nuclear Norm Matrix Completion Method (see, [Athey et al., 2021]

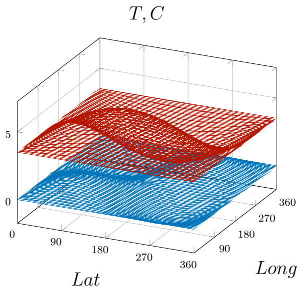
Heterogeneous Treatment Effects

A Set of Spatial Treatment Effects

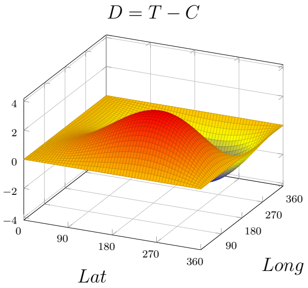
- Assume we have finite spatial grid which is divided in $i*j$ grids, where $i \in Lat$ and $j \in Long$.
- Then, we can obtain a set of treatment effects $D = \{\delta_{1,1}, \delta_{2,1}, \delta_{1,2}, ..., \delta_{i,j}\}$ by subtracting synthetic untreated outcomes in observed treated outcomes at position i, j.
- Commonly, it is assumed that all elements of D are constant over space. This assumption does not hold in spatially heterogeneous settings.

Reporting Heterogeneous Treatment Effects

The set of treatment effects D at time t_i may be spatially heterogeneous:



Treated (T), Control (C)



Treatment Effects (D)

Reporting Heterogeneous Treatment Effects

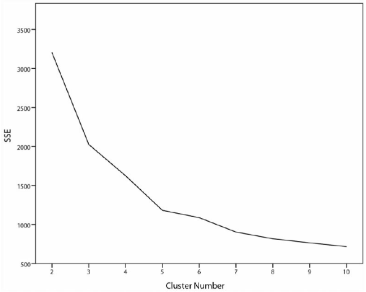
- How do we report these heterogeneous treatment effects?
  - Individual Treatment Effect: We report $\delta_{i,j}$ for $\forall i, j$.
  - Average Treatment Effect: We report $\frac{1}{n} * \sum(\delta_{i,j})\ \forall i, j$.
- The options are unrealistic and uninformative, respectively. Therefore, can we find a good middle ground?

Formally: We are looking to report the average of k subsets of D, where $1 < k < n$ in a plane with n grids.

The Clustering Problem

Rather than pre-define the reporting veracity (e.g., zip code), a clustering approach may allow the data to tell the whole story.

- How do we find the optimal k?
- Problem: As n increases, the error function monotonically decreases
- Formally: $\nexists k$ s.t. $argmin_k$(Error), outside of $k = n$.



Trade-Off: SSE vs. Cluster

A Solution?

Rather than pre-define the reporting veracity (e.g., zip code), a clustering approach may allow the data to tell the whole story.

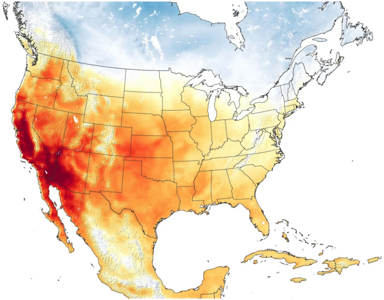
- Methods to find k: Elbow, Silhouette, etc.
- Problem: Curse of dimensionality!
- Uncertain: Is time really that important in short time horizons?

High-dimensional Clustering

Heatwave Application

Motivation

- Measuring the effect of climate disasters
  - Experimental approaches - if not unethical - are nearly impossible
  - Researchers are stuck with observational data
  - Recovering robust causal effects is critical policy input



Application: How do we measure the effect of heat waves on community health?

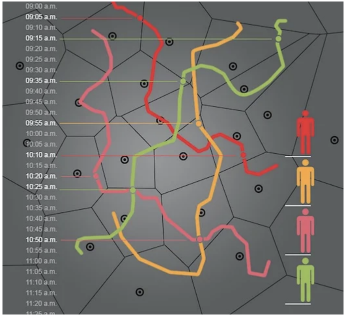
U.S. Cell Phone Ping Data

Access to cell phone ping data: Unique opportunity to analyze spatial settings where units are not bound to spatial location.

- Consider a densely populated area:

- Idea 1: Measure whether individuals visit hospitals via cell phone pings (Problem: Ping Irregularity)

- Idea 2: Measure the *dispersion effect* of heat waves



Conclusion and Discussion

Further Extensions

- Continuous treatment (consider [Callaway and Sant'Anna, 2021])
- Recent advances in spatial treatments (see, M. Pollmann)
- Digging into high-dimensional prediction and clustering

References I

Abadie, A., Diamond, A., and Hainmueller, J. (2015).
Comparative politics and the synthetic control method.
American Journal of Political Science, 59(2):495–510.

Athey, S., Bayati, M., Doudchenko, N., Imbens, G., and Khosravi, K. (2021).
Matrix completion methods for causal panel data models.
Journal of the American Statistical Association, 116(536):1716–1730.

References II

- Callaway, B. and Sant'Anna, P. H. (2021).
  Difference-in-differences with multiple time periods.
  Journal of Econometrics, 225(2):200–230.
- Pouliot, G. A. (2022).
  Spatial econometrics for misaligned data.
- Serra-Burriel, F., Delicado, P., Prata, A., and Cucchietti, F. (2020).
  Estimating heterogeneous wildfire effects using synthetic controls and satellite remote sensing.