Causal Inference with 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

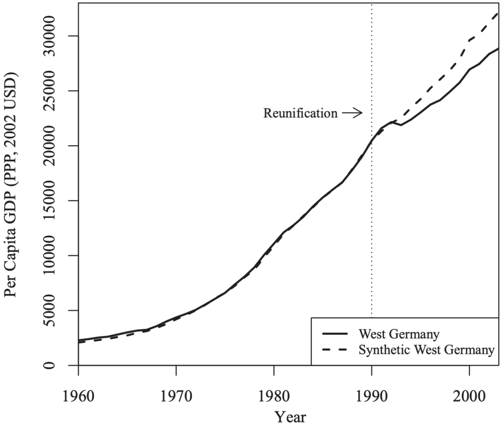
Motivation

Disclaimer

- This is a collection of preliminary, high-level ideas.
- The purpose of this presentation is to receive constructive criticism!
- My goal is to develop this project into a proper study in graduate school!

Synthetic Controls & DiD: A Review

- 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
- Causal estimator can be recovered in DiD, see Arkhangelsky et al, 2021
- **But**: This does not account for treatment effect heterogeneity - No insight into granular treatment responses due to spatial aggregation!



[Abadie et al., 2015]: German Reunification

Research Goals

I am interested in measuring the treatment effect $\tau_{it}(x)$ at location i and time t. Formally:

$\tau_{it}(x) = [Y_{it}^{(1)} - Y_{it}^{(0)} | X_{it} = x]$

where the $Y_{it}^{(1)}$ is the observed treated outcome, $Y_{it}^{(0)}$ denotes the unobserved untreated counterfactual. Let X_{it} denote a set of covariates and Z_{it} the treatment status, where $Z_{it} \in [0, 1]^d$.

My research goals are:

- Proposing a method for counterfactual prediction
- Matching comparison units to any treated unit $Y_{it}^{(1)}$
- Developing a difference-in-differences estimator

Counterfactual Prediction

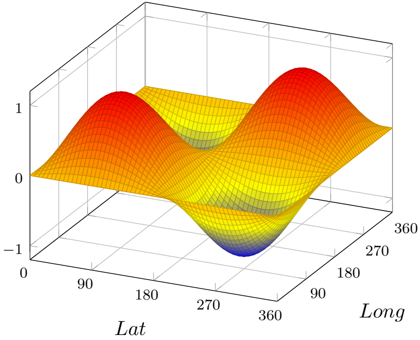
Finding a Prediction Algorithm

Need to find a prediction algorithm to generate synthetic (counterfactual) control data post-treatment.
This algorithm must be:

1. Excellent at modeling non-linear relationships
2. Robust to unobserved heterogeneity across space and time
3. Suitable for high-dimensional space

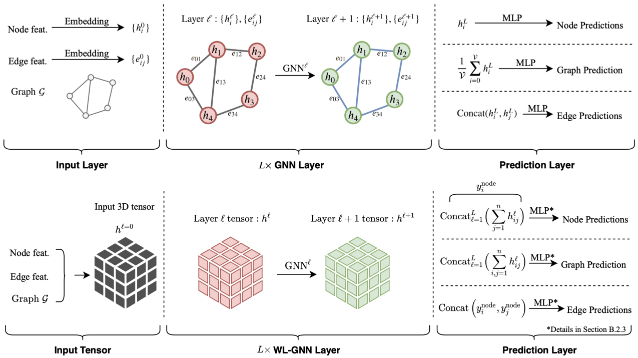
Note: Both discrete and continuous estimands are considered. The Application section elaborates.

How can we predict this shape?



A Geometric Learning Approach: Graph Neural Networks

1. Invariance to Permutation, Size, and Shape: Capable of handling complex data via graph representation
2. Reliable performance in non-Euclidian space: Important in high-dimensional setting
3. Features and similarity: Conveys both edge and node information

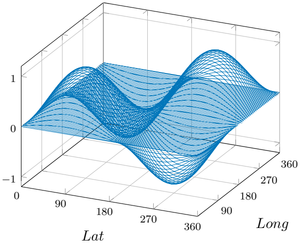


Summary of 2-D and 3-D Graph Neural Networks

Visualizing the End Goal

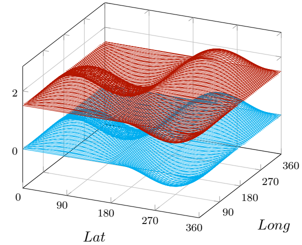
Consider treatment Z_{it} occuring at $t = t^*$ for all locations i. Hypothetically, a Graph Neural Network (GNN) can predict counterfactual untreated outcomes at various locations over time:

Some $t < t^*$: Untreated



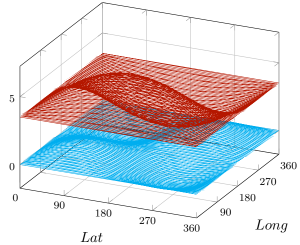
Pre-treatment, no treatment response, $Y_{it}(0)$

Some $t > t^*$: Treated, Counterfactual Untreated



Post-treatment, homogeneous response, $Y_{it}(1)$

Some $t >> t^*$: Treated, Counterfactual Untreated



Post-treatment, heterogeneous response, $Y_{it}(1)$

Choosing Comparison Units

Motivation

Let us decompose the treatment effect τ.

- $Y_{it}^{(1)}$: The set of treated outcomes is observed. Assuming the absence of measurement error, the data are accurate.
- $Y_{it}^{(0)}$: The set of untreated potential outcomes are not observed. Bias and inaccuracies likely stem from (i) prediction errors of the untreated data but also (ii) the choice of untreated units we choose to compare our observed treated unit with.

Therefore, even if we have generated reliable synthetic data for the untreated outcomes, that leaves us with one important choice:

Which control units $Y_{it}^{(0)}$ do we compare our treated unit $Y_{it}^{(1)}$ with to obtain τ?

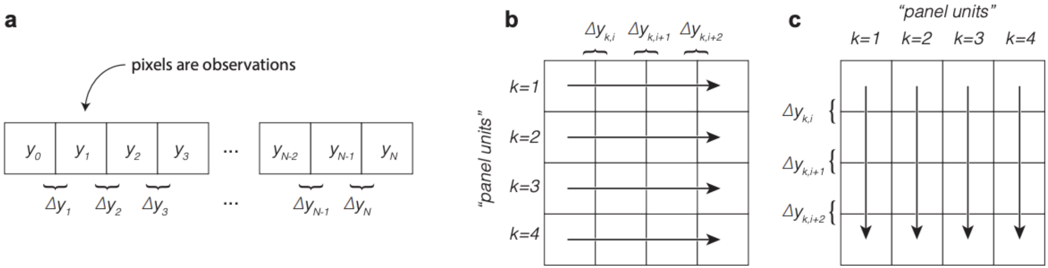
Context

Intuitively, this problem equates to not wanting to compare apples to oranges.

1. Using a set of covariates X_{it} to control for confounders (Pollmann WP)
   - Pros: It is easy to overlay spatial distributions of covariates in a map (Think: Each covariate is a separate heatmap)
   - Cons: Curse of dimensionality; does not capture unobserved heterogeneity
2. Using the k nearest units as comparison units
   - Pros: It is easy to implement and intuitive
   - Cons: Disregards more complex spatial relationships; breaks in higher dimensions; inherently arbitrary
3. Using spatial first differences (Druckenmiller & Hsiang, 2019)
   - Pros: It does account for unobserved heterogeneity (Think: Mini RDs between spatially adjacent units); No functional form assumptions and purely data-driven
   - Cons: Breaks in higher dimensions

The Spatial First Differences Idea

Spatial first differences do remove unobserved heterogeneity, great!

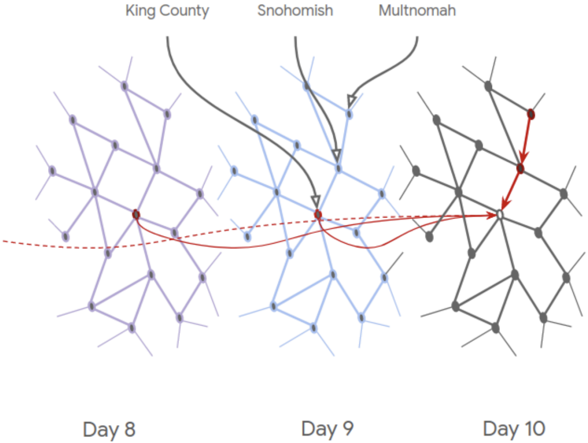


Druckenmiller & Hsiang, 2018

$\Rightarrow$ However, this approach breaks down when we want to account for a (non-pooled) time dimension. Can we borrow any ML tools to deal with this high-dimensional problem?

A Geometric Learning Approach

- Let us consider a graph where the nodes correspond to the units, and the edges between them represent the spatial differences.
- Then, the length of the edges can be a measure of the similarity of units under identical treatment status.
- When adding the time dimension, the graph consists of both spatial and temporal edges.

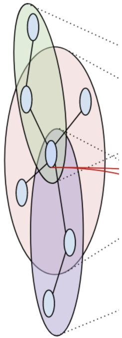


Kapoor, Ben et al., 2020: COVID-19 Infections

A Geometric Learning Approach (ctd.)

As long as we do not compare units across treatment status, we will obtain a three-dimensional graph. The edges represent a measure of similarity - a spatial or temporal first difference, respectively that is standardized across dimensions.

Formally, we can use a graph clustering algorithm to divide this 3-D graph into subsets. These k subsets will:



- Be between 1 < k < n, where 1 would include all units and n will construct a separate cluster for each unit in space and time
- Minimize the within-cluster average edge length: $\bar{E} = \frac{1}{k}\sum_{it}^{IT}(E)$; $\forall i \in I, t \in T$.
- Note: The spatial dimension shall be finely gridded. Coarse grids may omit valuable information.

Outcome and Practical Challenges

The desired outcome is a set of k clusters within which all units are comparable across space and time. Metaphorically, we are sorting a basket of apples and oranges into two separate baskets containing apples and oranges, respectively. Notably, there are considerable implementation challenges:

- As $k \to n$, $\bar{E}$ may monotonically decrease s.t. there is no k that minimizes $\bar{E}$ other than $k = n$
- Approaches to finding this optimal k have included the elbow and silhouette methods
- Optimization literature may give insight into appropriate constraints, e.g., continuity, non-separability, and overlap of clusters.

DiD Framework

Group-Time Average Treatment Effects

The **group-time average treatment effect**, $ATT(g,t)$, is a causal parameter in the context of DiD with multiple periods and multiple groups. The $ATT(g,t)$ has been approximated with the following estimators (τ):

- Outcome-regressions (Rubin, 1979)
  $\tau^{reg} = N^{-1} \sum_{i=1}^{N} \{\hat{\mu_1}(X_i) - \hat{\mu_0}(X_i)\}$

- Inverse Probability Weighting (Rosenbaum, 1987)
  $\tau^{ipw} = \frac{\sum_{i=1}^{N} Z_i Y_i / \hat{e}(X_i)}{\sum_{i=1}^{N} Z_i / \hat{e}(X_i)} - \frac{\sum_{i=1}^{N} (1-Z_i) Y_i / \{1-\hat{e}(X_i)\}}{\sum_{i=1}^{N} (1-Z_i) / \{1-\hat{e}(X_i)\}}$
  where $\hat{e}$ is the estimated propensity score.

- Doubly-Robust Methods
  $\tau^{dr} = \tau^{reg} + N^{-1} \sum_{i=1}^{N} \{ \frac{Z_i R_i}{\hat{e}(X_i)} - \frac{(1-Z_i) R_i}{1-\hat{e}(X_i)} \}$
  where $R_i = Y_i - \mu_{Z_i}(X_i)$ denotes the residual from outcome modeling.

A new estimator

... blah blah blah ...

Defining the g in ATT(g,t)

Frequently, units are clustered before inference. For example, the ATT(g,t) may be evaluated on the zip code level. However, this can bias the estimation of a treatment effect because:

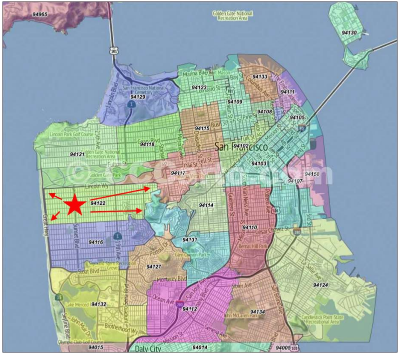
1. Pre-defined borders may not be applicable in treatment context
2. Biases of policymakers may be imposed on the study ex ante

We can overcome this issue by **inferring the clusters ex post**, i.e. after obtaining individual-level treatment effects. Formally, we ...

1. Measure idiosyncratic treatment effects over space and time, acknowledging that there will be missing treatment effects.
2. Interpolate missing treatment effects using the discrete subset we did measure.
3. Divide the continuous set of treatment effects into k distinct clusters which represent approximations of the ATT(g,t) that are robust to spatial heterogeneity.

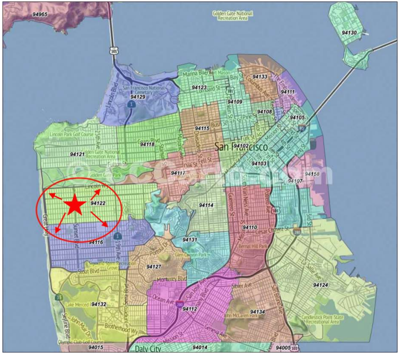
Defining the g in ATT(g,t): A Visualization

The existence of zip codes can be attributed to a variety of historical, political, and administrative factors, but these divisions do not necessarily correspond to a population that is homogeneous in their treatment response.



Ex Ante Comparison: Apples to Oranges

$\Rightarrow$



Ex Post Comparison: Apples to Apples

Control Strategy

Overlaying distributions of controls. This should also inform the clusters. More specifically: What if I can build the clusters based on all the covariates I consider and then represent them in a lower dimension in space???

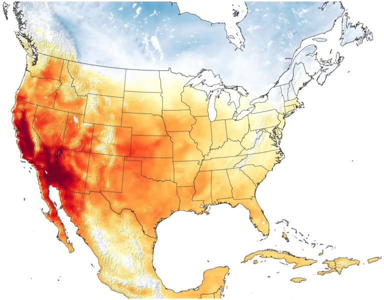
Limitations

- Vulnerable to structural differences in treatment effect density. Assumes the measurement of treatment effects across space and time distributed at random.

Heatwave Application

Motivation

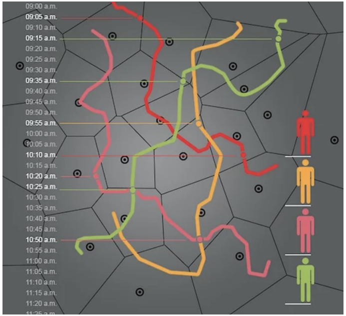
- Measuring the effect of climate disasters
  - Experimental approaches - if not unethical - are nearly impossible
  - Researchers are stuck with observational data
  - Recovering robust causal effects is critical policy input



My Research Question: What is the effect of heat waves on hospitalization rates?

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

Brainstorming

- Think about mapping hospitals etc. - But also public facilities that are AC controlled??? These could be identified with the help of and conversations with community members and stewards - Insight maybe: High-income - ppl just hunker down in their AC controlled places; low-income: Do people try to ”escape”/ ”flee” the heat to other places??? - classify/ define the concept of a ”HEAT SHELTER” in addition to hospitals - Define the median location of an individual as their home; also look into work location?

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.