Accounting for Unobservable Heterogeneity in Cross Section using Spatial First Differences (Druckenmiller & Hsiang, 2019)

Daniel C Posmik [1]

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

Key Ideas

- Global non-stationary trends often capture unobserved heterogeneity in many cross-sectional contexts
- If we restrict our between-unit comparison to two adjacent units only, we can difference this heterogeneity out because it constant across both units.
- The authors propose regressing the spatial first difference (SFD) of the outcome variable on treatment status - while omitting all other covariates since they are assumed to be constant in each pair of neighbors.

Why I chose this Paper

- Great revision (and extensions) of many key concepts in causal inference, e.g., the conditional independence assumption
- I am interested in causal inference for spatial treatments
- Note that this paper is **long** so I will be focusing on the estimator itself, rather than the applications. There are some sections of the paper that are ommitted in the interest of conciseness.
- Please do ask clarification questions during the paper review, but save the critical ones for the discussion!

Motivating the Idea

- Often, in the most general sense, we consider an outcome Y that is influenced by both (i) observed (short: x) and (ii) unobserved (short: c factors. Formally, we can write

$$Y_i = \beta x_i + \alpha c_i + \epsilon_i$$

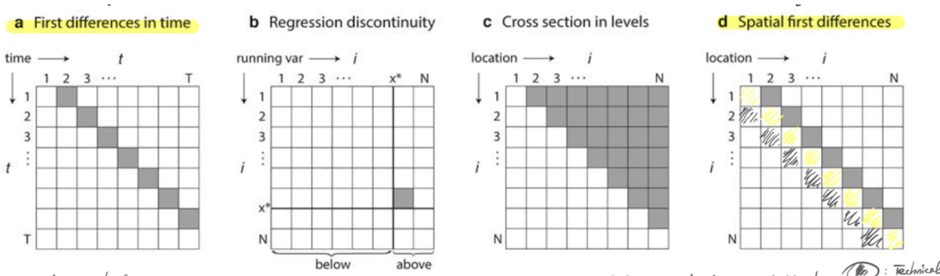
- Consider the coefficient β_L if c is omitted. β_L is a biased causal estimator because of omitted variable bias.

$$Y_i = \beta_L x_i + \epsilon_i \text{ where } \mathbb{E}(\beta_L - \beta) = \mathbb{E}[(xx')^{-1}(x'\alpha c)]$$

- Moreover, in many standard cross-sectional research designs, the conditional independence assumption (CIA) is levied on the entire sample. That means that all units are assumed to have their treatment assignment be independent of their potential outcomes, regardless of how far away they are from each other

SFD relaxes CIA drastically into the Local CIA. Units must only be conditionally independent within a neighbor pair, not across a pair.

Motivating the Idea (ctd.)



⚠ Important: Technically these should be symmetric but authors only shaded one side for simplicity!

: Technically also correct

: These are the same observations

Figure 1: **Comparison of pair-wise assumptions regarding the comparability of observations needed for identification in different research designs.** Graphical depiction of the various comparisons exploited to identify causal effects in (a) FD time-series models, (b) regression discontinuity designs with discontinuity at x^*, (c) the cross-sectional approach in levels, and (d) SFD. Each observation in a data set appears on both a row and column for a grid. Squares are grey if the observations for that row and column are assumed to be comparable (i.e. expected potential outcomes are conditionally equal) when using the associated research design.

The SFD Estimator

By exploiting Local CIA, and effectively purging the data of factors that are common to a comparison pair, the authors posit the following estimator:

$$\Delta Y_i = \beta_{SFD} \Delta X_i + \Delta\epsilon_i \text{ where } \hat{\beta}_{SFD} = (\Delta X' \Delta X)^{-1}(\Delta X' \Delta Y) \text{ via OLS estimation}$$

The authors argue that $\hat{\beta}_{SFD}$ is an unbiased estimator because:

- Case 1: If $\boldsymbol{c}$ is constant, it will be differenced out in SFD.
- Case 2: If $\boldsymbol{c}$ is not constant, $\hat{\beta}_{SFD}$ is still unbiased as long as the non-zero component (which remains from Case 1) of $\Delta\boldsymbol{c}$ is uncorrelated with $\Delta\boldsymbol{x}$ between neighbors.

Formally, the authors write the identifying assumption of SFD:

$$\mathbb{E}(\Delta X' \Delta C) = 0 \text{ (i.e. at least one of the components is zero} \Rightarrow \text{uncorrelated)}$$

Note that the SFD identifying assumption is even weaker than Local CIA.

Some commentary on asymptotic validity

The authors extend the asymptotic validity of the well-known time FD to its spatial analogue. Per Yatchew (1997),

$$\hat{\beta}_{SFD} \sim \mathcal{N}(\beta, \frac{1.5\sigma_\epsilon^2}{N} * \Omega_x^{-1})$$

Some interesting observations:

- As N increases, distance between spatial units decreases and Var($\hat{\beta}_{SFD}$) $\Rightarrow 0$
- Ω_x^{-1} describes the covariance between a treatment x and its spatial position (Remember: When $\Omega_x^{-1} \neq 0$, then "Global" CIA cannot be assumed)
- σ_ϵ^2 is the population variance of the error term which - under OLS estimation (Think: s_ϵ^2) - is autocorrelated (i.e. $\Delta\epsilon_i$ and $\Delta\epsilon_{i+1}$ both contain ϵ_i)
- This means Var($\hat{\beta}_{SFD}$) should be estimated using autocorrelation-robust approaches

How SFD works in Practice

Visually, SFD is intuitive in 1-D and 2-D grids:

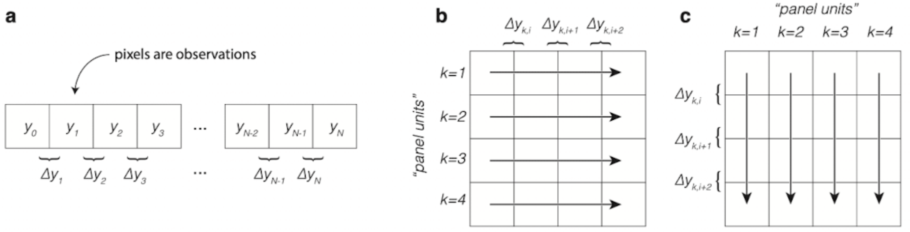


Figure 2: **Implementation of spatial first differences using gridded data.** Each square represents the location of an observation. (a) Implementation of SFD in one-dimensional space using a regular grid. The ith observation in the differenced data set contains the change in both the treatment $(\Delta \mathbf{x}_i)$ and the outcome (Δy_i) between immediately adjacent neighbors in positions i and $i-1$. Only Δy_i is shown. (b) Implementation of SFD in two-dimensional gridded data, where differences are computed in the West-East direction. Each row of observations (here indexed by k) is analogous to a panel unit in panel data. (c) Same, but differences are computed in the North-South direction.

Why SFD works in Practice

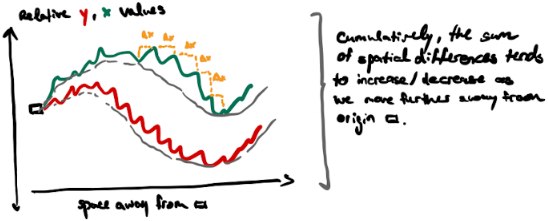
SFD allow us to omit not only unobservables (c), but also observable covariates (x) in the estimation procedure because:

- SFD filters out the influence of all factors that vary at low spatial granularity ($\Rightarrow$Global Bias eliminated)
- SFD differences out all common influences that only affect an idiosyncratic pair of adjacent neighbors ($\Rightarrow$Local Bias eliminated)

Moreover, SFD manages to to eliminate even complex correlation patterns.

- This is important in scenarios where treatment spills beyond just neighbors
- SFD eliminates these complex correlation patterns by subtracting the ”trend line” from an entire history of values as you move further away from the origin. The authors call this trend line the ”spatial history” (See next slide)

Why SFD works in Practice (ctd.)



What helped me understand this is thinking of SFD as a set of "mini regression discontinuity (RD)" designs between neighboring units. Regardless of how far you move away from the origin, you only ever consider immediately adjacent units. Suppose treatment x affects y's n steps away from the spatial origin, you will estimate the following:

$$\Delta Y_i = \beta_{SFD}\Delta X_i + \gamma_{SFD}\Delta X_{i+1} + ...\zeta_{SFD}\Delta X_{i+n} + \Delta\epsilon_i$$

The Applications

The authors consider two applications, one 1-D example and one 2-D example.

- Their 1-D example considers the effect of schooling on wages along long roads in Chicago and NYC

- Their results show that the ”levels” estimator is much higher than the SFD estimator. They conclude that the reason is the magnitude of omitted variable bias in the levels estimator.

Moreover,

- Their 2-D example considers the effect climate on crop yields

- They show that SFD in gridded data works both in horizontal and vertical alignment. Moreover, they use arbitrary rotations up to 360as a sanity-check.

- This extends well to non-gridded data as long as no unit is doubly counted (See next slide)

The Applications (ctd.)

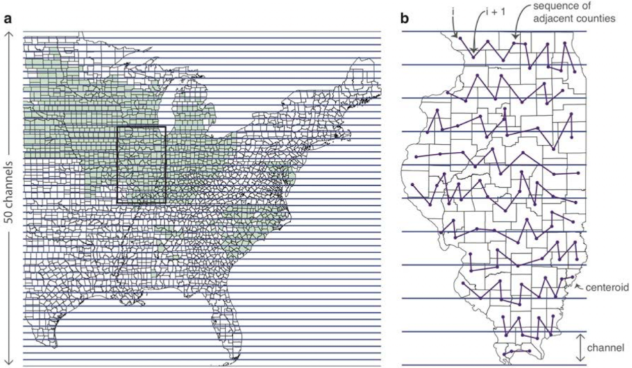


Figure 7: **Sampling procedure for spatial first differences with irregular (non-gridded) data.** (a) US counties east of the 100th meridian with sampling channels overlaid in blue. Counties included in the balanced panel are highlighted shaded green. (b) Detail of insert in a, depicting the algorithm used to generate sequences of adjacent counties to construct a “panel-like” data structure, analogous to Figure 2b (see text for description).

Thank You - Moving to the Discussion!

Discussion Ideas

- In general, what do you think are weaknesses of the SFD estimator?
- Do you think even the Local CIA is too strong? Can we just omit unobservables and especially (potentially idiosyncratic) covariates?
- Do you think an extension of SFD is possible in high-dimensional space, i.e. adding the time dimension? Rather than a 2-D plane, we would be working with a tensor.