Tobler meets Robinson: Semiparametric methods for Covariate Adjustment using Spatial Data

Apoorva Lal, *Stanford*

Introduction

- Social scientists frequently attempt to study causal effects with geolocated data
  - Especially widespread in economic history, historical political economy
- Assignment mechanism unknown, but loose intuition that closer units are more comparable on unobservables – Tobler's law
  - Current practice: assume geolocations are just another covariate, proceed with standard unconfoundedness approaches
  - Or use spatial information to subset the data to 'close enough to a border', then fit outcome models : 'local' unconfoundedness
- **this project**: estimators that use geographic data to partial out effect of smooth spatial confounder using ideas from semiparametric estimation for causal effects

Setup

- Data:
  - Outcome: $Y_i(s) \in \mathbb{R}$
  - Treatment: $W_i(s) \in \{0, 1\}$
  - (Unobserved) Confounder: $U_i(s) \in \mathbb{R}$
- Each observation has spatial location $\mathbf{s}_i$
  - **point data**: $\mathbf{s}_i \in \mathbb{R}^2$: latitude, longitude
  - **areal data**: Data located on an *irregular lattice*

Robinson: Partial Linear Regression Formulation

- $\mathbf{X}_i = (1, W_i)$
- Since U is unobserved, 'long' regression is infeasible, but S_i available
- OVB: $\mathbb{E}[\tau - \widehat{\tau}] = \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{X}^\top \mathbb{E}[U|\mathbf{X}]$
- To adjust for S_i flexibly, a *partially linear* model is a reasonable starting point

$$Y_i = \tau W_i + \mu_{(0)}(\mathbf{S}_i) + \varepsilon_i$$

- In a seminal article, Robinson(1988) rewrites the above regression as

$$(Y_i - \mathbb{E}[Y_i|\mathbf{S}_i]) = \tau \cdot (W_i - \mathbb{E}[W_i|\mathbf{S}_i]) + \eta_i \quad ; \mathbb{E}[\eta, \mathbf{S}_i, W_i] = 0$$

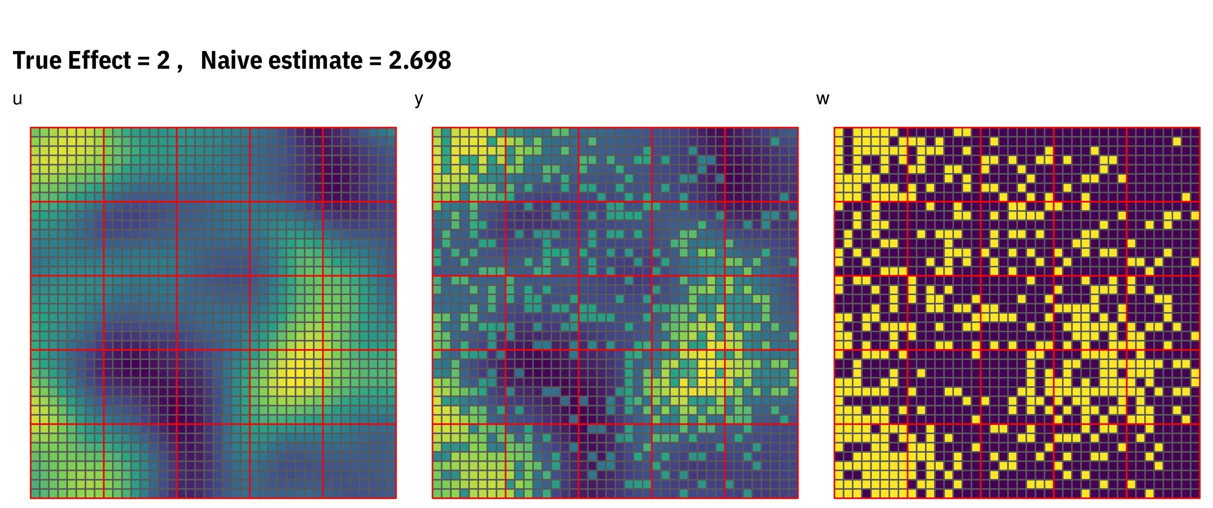
- where $\mathbb{E}[Y_i|\mathbf{S}_i] =: m(\mathbf{S}_i)$ and $\mathbb{E}[W_i|\mathbf{S}_i] =: e(\mathbf{S}_i)$ are nonparametric regressions ['nuisance' components]

Identification

1. **Consistency / SUTVA** : $Y_i = Y_i(w_i)$.
2. **Latent Unconfoundedness**: $Y_i(w) \perp\!\!\!\perp W|U$ *given unobserved confounder*
3. **Positivity** : $\mathbf{Pr}(W = 1|U) \in (0, 1)$
   - Treatment varies at a smaller scale than the spatial confounder
4. **Learnability of confounder** : $U = g(s)$ for a fixed, smooth (measurable) $g(\cdot)$
5. **Conditioning on S doesn't induce confounding**: $Y(w) \perp\!\!\!\perp W|U, S$

(4 + 5) reduces the problem to unconfoundedness given location: $Y_i(w) \perp\!\!\!\perp W|S$

This lets us identify the counterfactual mean $\mathbb{E}[Y(w)] = \mathbb{E}_{\mathcal{S}}[\mathbb{E}[Y|S, W = w]]$



Estimands

Estimands: $\tau^{\text{ATE}} = \mathbb{E}[Y^1 - Y^0], \quad \tau^{\text{ATT}} = \mathbb{E}[Y^1 - Y^0|W = 1], \quad \tau(\mathbf{x})^{\text{CATE}} = \mathbb{E}[Y^1 - Y^0|\mathbf{X} = \mathbf{x}]$

Conditional-variance weighted average of strata-specific effects

$$\tau_e = \frac{\mathbb{E}[\{e(\mathbf{S})(1 - e(\mathbf{S})\}\tau(\mathbf{S})]}{\mathbb{E}[e(\mathbf{S})(1 - e(\mathbf{S})]}$$

Estimators

Partially Linear Regression Coefficient is consistent for ATO

$$\widehat{\tau} = \sum_{i=1}^n (Y_i - \widehat{m}(\mathbf{S}_i))(W_i - \widehat{e}(\mathbf{S}_i)) / \sum_{i=1}^n (W_i - \widehat{e}(\boldsymbol{S}_i))^2$$

Augmented Inverse Propensity Weighting for ATE

$$\widehat{\tau}^{\text{ATE}}_{\text{AIPW}} = \frac{1}{n}\sum_{i=1}^n \left[\underbrace{\overbrace{\widehat{m}_1(\mathbf{S}_i)}^{\text{Regression}} + \overbrace{\frac{D_i(Y_i - \widehat{m}_1(\mathbf{S}_i))}{\widehat{e}(\mathbf{S}_i)}}^{\text{IPW of residuals}}}_{\text{estimator for } \mathbb{E}[Y_i(1)]}\right] - \frac{1}{n}\sum_{i=1}^n \underbrace{\left[\widehat{m}_0(\mathbf{S}_i) + \frac{(1 - D_i)(Y_i - \widehat{m}_0(\mathbf{S}_i))}{1 - \widehat{e}(\mathbf{S}_i)}\right]}_{\text{estimator for } \mathbb{E}[Y_i(0)]}$$

Tobler : Estimating $\widehat{m}, \widehat{e}$

- Weighted sum representation regressions $m(\mathbf{s}) = \mathbb{E}[Y|\mathbf{s} = \mathbf{s}]$ as $\widehat{m}(\mathbf{s}) = \sum_{i=1}^n \omega_i(\mathbf{s}_i) y_i$
  - **NN Regression, Kernel Regression, Random Forests**
- **Nearest Neighbours Differencing**: Compute $\widetilde{A}_i = \mathbf{W} A_i$ is the average of neighbours'values of A_i, where $\mathbf{W}$ is a (queen) weight matrix
  - Estimate regression with residuals

$$(Y_i - \widetilde{Y}_i) = \tau(W_i - \widetilde{W}_i) + \overbrace{U_i - \widetilde{U}_i}^{=0 \text{ by smoothness}} + \eta_i$$

  - With regular lattice data, Druckenmiller and Hsiang(2018) call this **Spatial First differences**
- **Markov Random Fields**: Unit random effects γ_j are assumed to be GMRF and $\mathcal{N}_j$ denotes the set of neighbours of unit j.

$$J(\boldsymbol{\gamma}) = \sum_{j=1}^n \sum_{i \in \mathcal{N}_j, i > j} (\gamma_j - \gamma_i)^2$$

Variances are computed using the jackknife or Bayesian Bootstrap.

Minimax Balancing

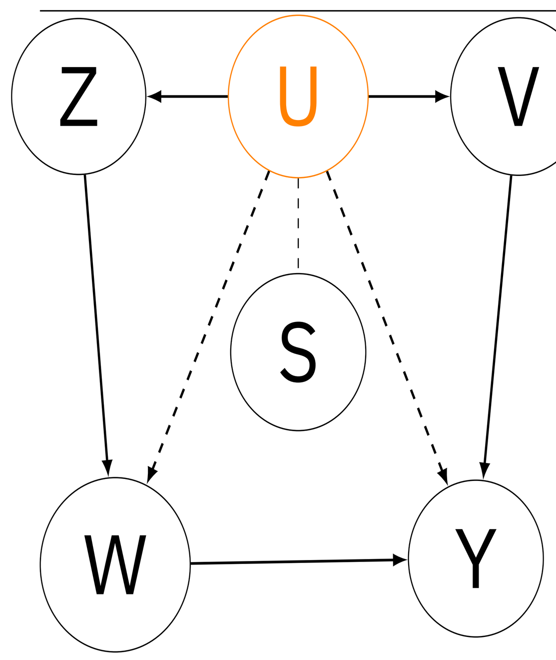
Balancing methods (Hirschberg and Wager 2021) seek to sidestep estimating a propensity score and instead directly fit an outcome surface and 'debias' it using weights minimise worst-case regression error for a convex function class (e.g. $\mathcal{F} : \{\|\nabla^2 \mu(\mathbf{x})\| \leq B)\}$

$$\mathcal{F}_{\mathcal{H}} = \left\{m : m(\mathbf{s}, w) = \mu(\mathbf{s}) + w\tau(\mathbf{s}), \|\mu\|^2_{\mathcal{H}} + \|\tau\|^2_{\mathcal{H}} \leq 1\right\}$$

$$\widehat{\gamma} = \underset{\gamma \in \mathbb{R}^n}{\text{argmin}} \sup_{\mu \in \mathcal{H}} \left(\frac{1}{n}\sum_{i=1}^n \gamma_i \mu(\mathbf{s}_i)\right) + \sup_{\tau \in \mathcal{H}} \left(\frac{1}{n}\sum_{i=1}^n (W_i \gamma_i - 1)\tau(\mathbf{s}_i)\right) + \frac{\sigma^2 \|\gamma\|_2^2}{n^2}$$

$$\widehat{\psi}^{\text{AML}} = \frac{1}{n}\sum_{i=1}^n (\widehat{\tau}(\mathbf{s}_i) - \widehat{\gamma}_i(\widehat{\mu}(\mathbf{s}_i) + W_i \widehat{\tau}(\mathbf{s}_i) - Y_i))$$

Proxy Outcome and Treatment Design



- **Bridge functions**: $\exists h(\cdot)$ that satisfies $\mathbb{E}[Y(w) - h(V, S, w)|U, S, W = w] = 0$ for $w \in \{0, 1\}$
  - h is a transformation of the pre-treatment outcome so that the effect of the unmeasured confounder on the transformation is the same as on the unobserved counterfactual outcome
- Then, counterfactual mean $\mathbb{E}[Y(w)] = \mathbb{E}[h(V, w, S)]$
- $h(\cdot)$ identified by moment condition

$$\mathbb{E}[Y - h(V, w, S; \boldsymbol{\gamma})|Z, W, S] = 0$$

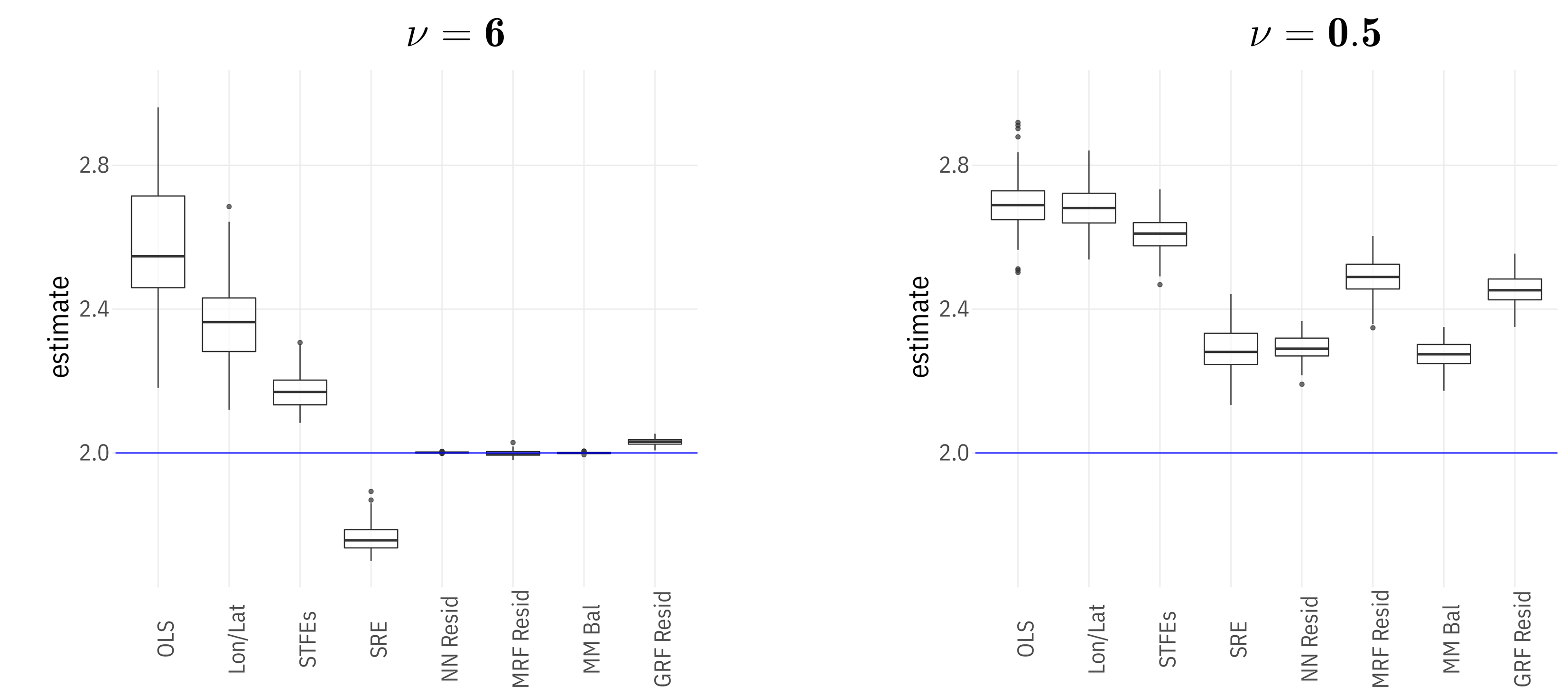
  - where $\boldsymbol{\gamma}$ is a finite-dimensional vector that characterises $h(\cdot)$
  - High-dimensional Generalised Method of Moments problem - need to handle non-uniqueness using regularisation (Imbens et al 2021)

Simulation Study

- Data on 40×40 cell grid
- U is a Gaussian Process with Matern covariance (stationary)

$$R(d; \theta, \nu) = \frac{1}{\Gamma(\nu)2^{\nu-1}} \left(\frac{2\sqrt{\nu}d}{\theta}\right)^\nu \mathcal{K}_\nu \left(\frac{2\sqrt{\nu}d}{\theta}\right)$$

  - K is a modified Bessel Fn
  - $\nu > 0$ is a smoothness parameter, θ is a range parameter
  - d is Euclidian distance between two locations
- $W_i \sim \text{Bernoulli}(\text{logit}(U + \eta_i)))$, η_i normal
- $Y_i = \tau W_i + U + \varepsilon_i$



References

[1] H. Druckenmiller and S. Hsiang. Accounting for unobservable heterogeneity in cross section using spatial first differences. 2018.

[2] B. Gilbert, A. Datta, J. A. Casey, and E. L. Ogburn. Approaches to spatial confounding in geostatistics, 2021.

[3] D. A. Hirshberg and S. Wager. Augmented minimax linear estimation. *The Annals of Statistics*, 49(6):3206–3227, 2021.

[4] W. Miao, X. Shi, and E. T. Tchetgen. A confounding bridge approach for double negative control inference on causal effects. *arXiv preprint arXiv:1808.04945*, 2018.

apoorvalal.github.io

apoorval@stanford.edu