Adjusting for unmeasured spatial confounding with distance adjusted propensity score matching

Georgia Papadogeorgou

with Christine Choirat and Cory Zigler

OSU, June 4, 2019

Causal inference and unmeasured structured confounding

- Causal inference formalizes the notion of an *effect*, and provides identifiability assumptions
- One often invoked assumption is the no unmeasured confounding assumption (+ positivity = ignorability)
- It cannot be tested but sensitivity of results to violations of this assumption can be evaluated [Rosenbaum, 2002]

Causal inference and unmeasured structured confounding

- Causal inference formalizes the notion of an *effect*, and provides identifiability assumptions
- One often invoked assumption is the no unmeasured confounding assumption (+ positivity = ignorability)
- It cannot be tested but sensitivity of results to violations of this assumption can be evaluated [Rosenbaum, 2002]

- Can we use unmeasured confounders' *structure* to adjust for them?
  - Spatial structure: spatial variables vary continuously over space

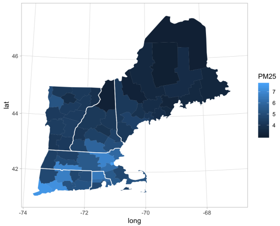
Spatial data and causal inference in air pollution research

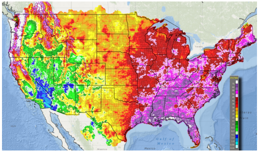
- The scientific questions are causal
  - Do emissions cause pollution?
  - What effect does an intervention on polluting sources have on air pollution concentrations?

https://kcstormfront.wordpress.com/2015/01/11/2014-in-review/

Spatial data and causal inference in air pollution research

- The scientific questions are causal
  - Do emissions cause pollution?
  - What effect does an intervention on polluting sources have on air pollution concentrations?
- The data are spatial
  - Spatially-indexed
  - Exposure, outcome, and covariates are spatially structured
  - Unmeasured confounders are spatial!

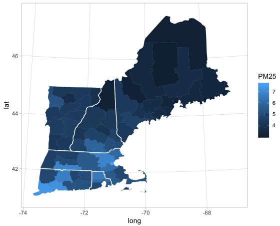


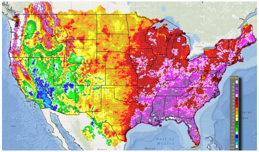


https://kcstormfront.wordpress.com/2015/01/11/2014-in-review/

Spatial data and causal inference in air pollution research

- The scientific questions are causal
  - Do emissions cause pollution?
  - What effect does an intervention on polluting sources have on air pollution concentrations?
- The data are spatial
  - Spatially-indexed
  - Exposure, outcome, and covariates are spatially structured
  - Unmeasured confounders are spatial!





Integration of spatial data and causal inference

https://kcstormfront.wordpress.com/2015/01/11/2014-in-review/

Air pollution regulations and their impact

- Regulations such as the Clear Air Act enforce stricter rules on emissions aiming to reduce ambient air pollution
  - Source-specific emissions like power plants and motor vehicles

---

NO_x: Nitric oxide and nitrogen dioxides, precursors of ozone, reacting with other compounds in the presence of sunlight to create ozone

SCR/SNCR: Selective Catalytic/Non-Catalytic Reduction technology

Air pollution regulations and their impact

- Regulations such as the Clear Air Act enforce stricter rules on emissions aiming to reduce ambient air pollution
  - Source-specific emissions like power plants and motor vehicles
- Power plants follow various strategies to comply to these regulations
- We focus on the installation of NO_x emission reduction control technologies
- Are SCR/SNCR more effective than alternative strategies in reducing ambient ozone concentrations?

NO_x: Nitric oxide and nitrogen dioxides, precursors of ozone, reacting with other compounds in the presence of sunlight to create ozone

SCR/SNCR: Selective Catalytic/Non-Catalytic Reduction technology

Comparative effectiveness of power plant NO_x emission reduction technologies

- SCR/SNCR systems are the most effective in reducing NO_x
- Ozone is a secondary pollutant
  - NO_x reacts with volatile organic compounds (VOCs) and carbon monoxide in the presence of sunlight to create ozone
- VOCs, sunlight might be spatial confounders and they are unmeasured

NO_x: Nitric oxide and nitrogen dioxides, precursors of ozone

Notation

- For unit i
  - Treatment $Z_i \in \{0, 1\}$
  - Potential outcomes $\{Y_i(0), Y_i(1)\}$
  - Observed outcome $Y_i = Y_i(Z_i)$
  - Covariates $\boldsymbol{X}_i = (X_{i1}, X_{i2}, \ldots, X_{ip})$
- Average treatment effect on the treated

$$\text{ATT} = E[Y(1) - Y(0) | Z = 1]$$

Identifiability and estimation of the ATT

- Common identifiability assumptions
  - Positivity: $p(Z = z|\boldsymbol{X}) > 0, z \in \mathcal{Z}$
  - No unmeasured confounding: $Y(z) \perp\!\!\!\perp Z|\boldsymbol{X}$
- Estimate the average potential outcome via propensity score methods, outcome regression, or combinations

Identifiability and estimation of the ATT

- Common identifiability assumptions
  - Positivity: $p(Z = z|\boldsymbol{X}) > 0, z \in \mathcal{Z}$
  - No unmeasured confounding: $Y(z) \perp\!\!\!\perp Z|\boldsymbol{X}$
- Estimate the average potential outcome via propensity score methods, outcome regression, or combinations
- Confounders $\boldsymbol{X} = (\boldsymbol{C}, \boldsymbol{U})$, $\boldsymbol{C}$ are observed, $\boldsymbol{U}$ are unobserved
- If $\boldsymbol{U}$ varies spatially, can we adjust for it?

Identifiability and estimation of the ATT

- Common identifiability assumptions
  - Positivity: $p(Z = z|\boldsymbol{X}) > 0, z \in \mathcal{Z}$
  - No unmeasured confounding: $Y(z) \perp\!\!\!\perp Z|\boldsymbol{X}$
- Estimate the average potential outcome via propensity score methods, outcome regression, or combinations
- Confounders $\boldsymbol{X} = (\boldsymbol{C}, \boldsymbol{U})$, $\boldsymbol{C}$ are observed, $\boldsymbol{U}$ are unobserved
- If $\boldsymbol{U}$ varies spatially, can we adjust for it?
- Matched pairs should be similar in terms of
  1. Observed covariates
  2. Unmeasured spatial covariates (small geographical distance)

Distance Adjusted Propensity Score Matching

- Propensity score model using measured variables $\boldsymbol{C}$:
  $P(Z_i = 1|\boldsymbol{C}_i) = \text{expit}\left(\boldsymbol{C}_i^T \beta\right)$

- For a treated unit i and a control unit j define

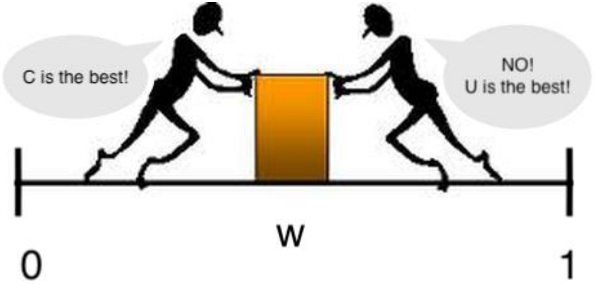
$$DAPS_{ij} = w|PS_i - PS_j| + (1-w) * Dist_{ij},\ w \in [0,1]$$

  where PS are propensity score estimates, and $Dist$ represents spatial proximity

- Small value of $DAPS_{ij}$ means:
  - Similar propensity scores
  - Points in close geographical distance (similar values of $\boldsymbol{U}$!)

- w: relative importance of the observed and unobserved confounders
  - High values of w - priority to observed covariates
  - Low values of w - priority to spatial proximity

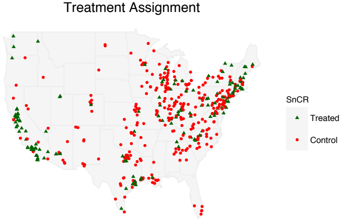
Choosing w



1. Match treated to control units for various values of w
2. Assess balance of the observed covariates
3. Choose the smallest value of w that achieves observed covariate balance

Data

- Coal and natural gas power plants during June-August 2004
- $Z = 1$ if at least half of facility heat input is used by units with installed SCR/SNCR technologies, $Z = 0$ otherwise
- 152 treated facilities, 321 controls

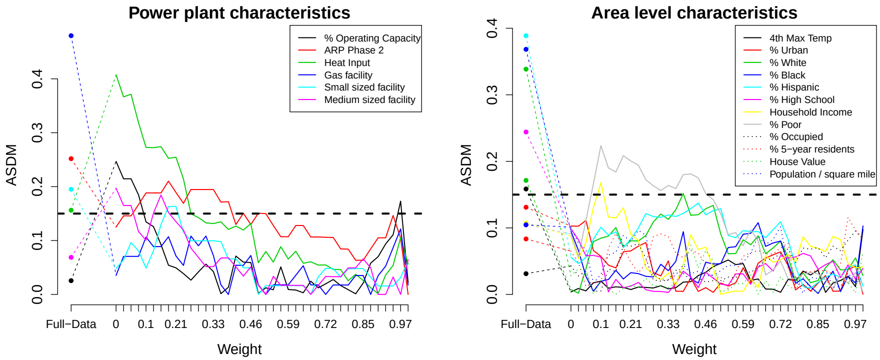


- Y: NO_x emissions / 4^{th} maximum ambient ozone concentration
- Covariates: Power plant characteristics, demographics, weather

Publicly available data sources: Air Markets Program Data, 2000 Census, EPA monitoring sites

Observed covariate balance as a function of w

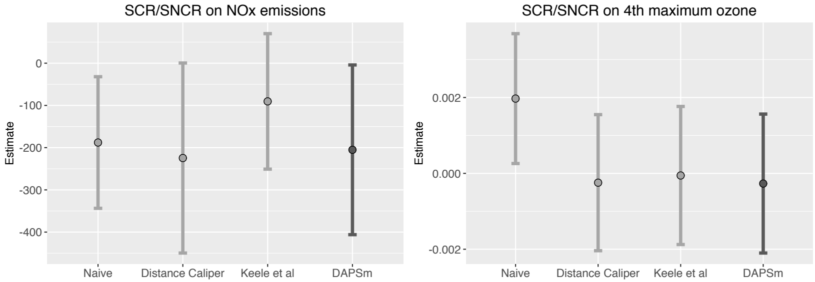
- Absolute standardized difference of means as a function of w



Matches

- Average distance of matched pairs
  - Naïve: 1066 miles
  - DAPSm: 141 miles

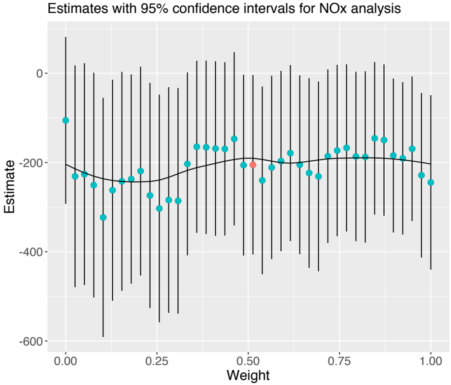
Results

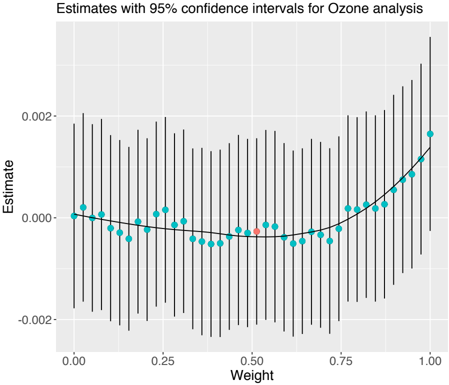


- Reduction by 205 tons of NO$_x$ emissions (95% CI: 4 – 406)
- -0.27 (95% CI: -2.1 to 1.56) parts per billion in ambient ozone

– The national ambient air quality standard for ozone is 70 parts per billion.
– Keele et al. [2015]

Evaluating the presence of unmeasured spatial confounding





Conclusions

- SCR/SNCR control technologies lead to
  - Reductions in NO_x emissions
  - Their effect on ozone is not significant
- When interference between units is accounted for, SCR/SNCR leads to reductions in ambient ozone concentrations [Papadogeorgou et al., 2019]

Papadogeorgou, Choirat, and Zigler [2018]

Conclusions

- SCR/SNCR control technologies lead to
  - Reductions in NO_x emissions
  - Their effect on ozone is not significant
- When interference between units is accounted for, SCR/SNCR leads to reductions in ambient ozone concentrations [Papadogeorgou et al., 2019]
- Approaches like DAPSm are not immediately compatible with spatial models
  - Bridging the two strands of literature (Patrick Schnell's talk yesterday)

Papadogeorgou, Choirat, and Zigler [2018]

Conclusions

- SCR/SNCR control technologies lead to
  - Reductions in NO_x emissions
  - Their effect on ozone is not significant
- When interference between units is accounted for, SCR/SNCR leads to reductions in ambient ozone concentrations [Papadogeorgou et al., 2019]
- Approaches like DAPSm are not immediately compatible with spatial models
  - Bridging the two strands of literature (Patrick Schnell's talk yesterday)
- Unmeasured confounding is one of the main criticisms of air pollution epidemiology
- We can address this using
  - Sensitivitiy analysis
  - Analysis mitigating bias by unmeasured structured confounders

Papadogeorgou, Choirat, and Zigler [2018]

References

https://github.com/gpapadog/DAPSm

Luke Keele, Rocio Titiunik, and José R Zubizarreta. Enhancing a geographic regression discontinuity design through matching to estimate the effect of ballot initiatives on voter turnout. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 178(1): 223–239, 2015.

Georgia Papadogeorgou, Christine Choirat, and Corwin M Zigler. Adjusting for unmeasured spatial confounding with distance adjusted propensity score matching. *Biostatistics*, 20(2): 256–272, 2018.

Georgia Papadogeorgou, Fabrizia Mealli, and Corwin M Zigler. Causal inference with interfering units for cluster and population level treatment allocation programs. *Biometrics*, 2019.

Paul R. Rosenbaum. *Sensitivity to Hidden Bias*, pages 105–170. Springer New York, New York, NY, 2002. ISBN 978-1-4757-3692-2.