Controlling for spatial confounding and spatial interference in causal inference

Modeling insights and the spycause package

Tyler D. Hoffman[1,2,a] Peter Kedron[1,2,a]

[1]School of Geographical Sciences and Urban Planning, Arizona State University
[2]Spatial Analysis Research Center, Arizona State University

March 25, 2023

[a]This material is based upon work supported by the National Science Foundation through a Graduate Research Fellowship awarded to Hoffman under NSF Grant No. 026257-001 and Project Grant No. BCS-2049837 awarded to Kedron. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Introduction: Why cause?

- Want to deduce causal relationships through statistical models
- Space presents unique challenges: scale, confounding, and interference
- Need for meta-analysis of use cases and relative performance among existing spatial causal models

Introduction: Why cause?

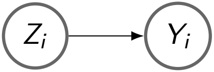
- Want to deduce causal relationships through statistical models
- Space presents unique challenges: scale, confounding, and interference
- Need for meta-analysis of use cases and relative performance among existing spatial causal models

Research objectives

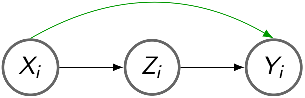
- Demonstrate that intuition from noncausal spatial modeling holds in causal spatial modeling
- Develop a standardized code base for spatial causal models

Inference setting

Ideal scenario

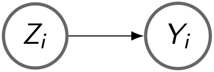


Nonspatial confounding

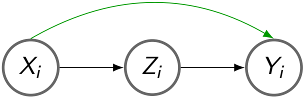


Inference setting

Ideal scenario



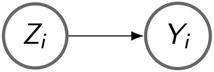
Nonspatial confounding



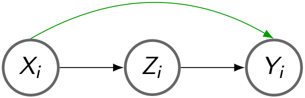
- **Study unit:** US states
- **Outcome:** drunk driving crashes
- **Treatment:** drinking age
- **Nonspatial confounder:** number of cars

Inference setting

Ideal scenario



Nonspatial confounding



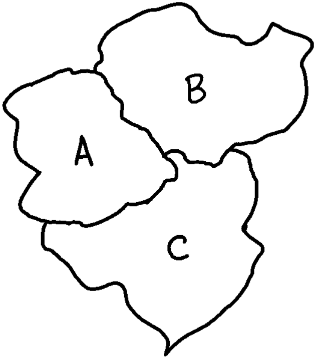
- **Study unit:** US states
- **Outcome:** drunk driving crashes
- **Treatment:** drinking age
- **Nonspatial confounder:** number of cars

Conditioning on X_i removes the green arrow, permitting inference on the treatment effect $Z_i \rightarrow Y_i$.

Spatial confounding

Challenge

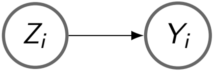
Non-treatment variables may contribute to the outcome through spatial relationships.



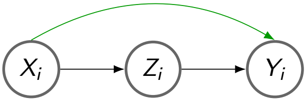
- Minute, unquantifiable, highly local qualities of places
- **Study unit:** US states
- **Outcome:** drunk driving crashes
- **Treatment:** drinking age
- **Spatial confounder:** location of bars

Spatial causal issues: Spatial confounding

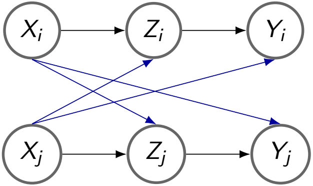
Ideal scenario



Nonspatial confounding



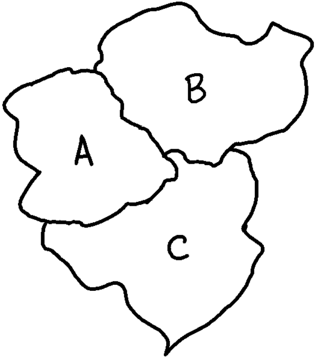
Spatial confounding



Spatial interference

Challenge

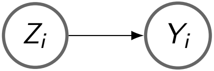
If units influence each others' responses to an intervention, then we cannot isolate the effect of the intervention.



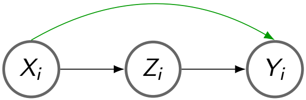
- Tobler's First Law: nearby things tend to be related
- **Study unit:** US states
- **Outcome:** drunk driving crashes
- **Treatment:** drinking age
- **Interference:** drinking age of neighboring states

Spatial causal issues: Spatial interference

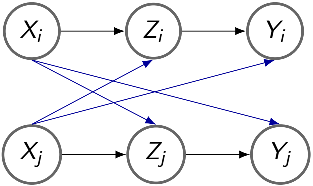
Ideal scenario



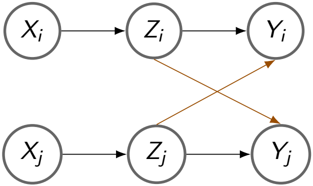
Nonspatial confounding



Spatial confounding



Spatial interference



Spatial causal models

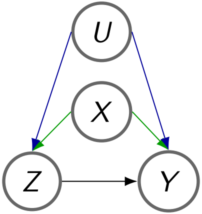
- Matching methods
- Regression adjustment
- Spatial instrumental variables
- Geographic regression discontinuity design
- Spatial difference-in-difference

(Herrera et al., 2014; Akbari et al., 2021; Reich et al., 2021)

Spatial confounding adjustments

Conditional autoregressive models

Let $U \sim N(0, \Sigma)$ where $\Sigma = \sigma_U^2(I - \rho_U W)^{-1}$ and W is a row-standardized weights matrix.

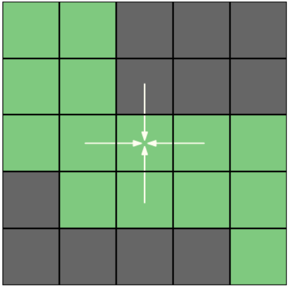


Strategy: incorporate U in models to control for unknown spatial confounding.

Spatial interference adjustments

Spatial lag adjustment

For a spatial weights matrix W, add a lag of the treatment variable WZ to the linear model.



Strategy: incorporate a spatial lag of the treatment variables to account for their affects on each other.

Simulation study: data

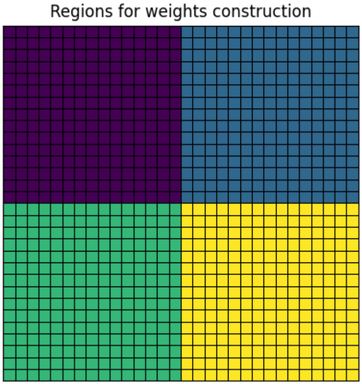
Parameters of interest:

- Structure of spatial confounding in data
- Structure of spatial interference in data

Weights matrices considered:

- None
- Binary (Queen contiguity)
- Distance-based
- Region-based

$4 \times 4 = 16$ total data scenarios



Simulation details

Research objectives

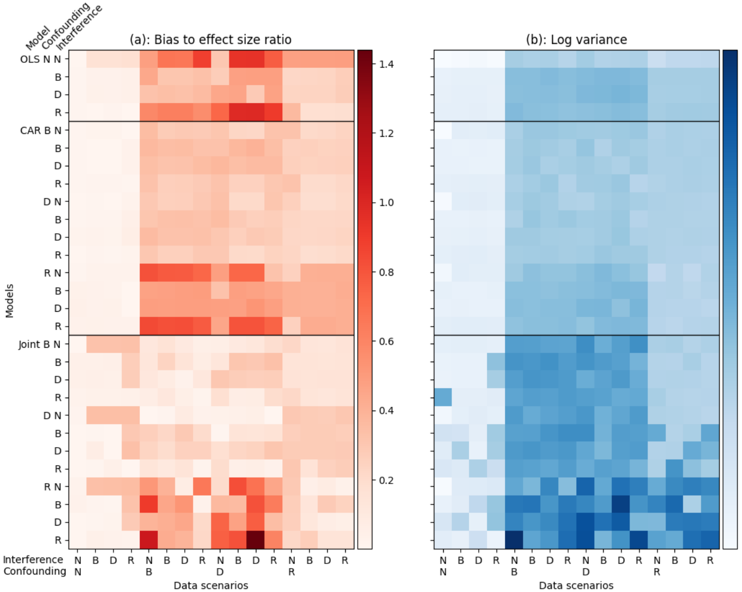
- Demonstrate that intuition from noncausal spatial modeling holds in causal spatial modeling
- Develop a standardized code base for spatial causal models

- **Designed a simulation experiment to begin analyzing relative performance of spatial causal models**
- 1 confounding adjustment for OLS, 3 each for CAR and Joint
- 4 interference adjustments (applicable for all models)
- 28 total models on 16 data scenarios = 448 combinations

Simulation results

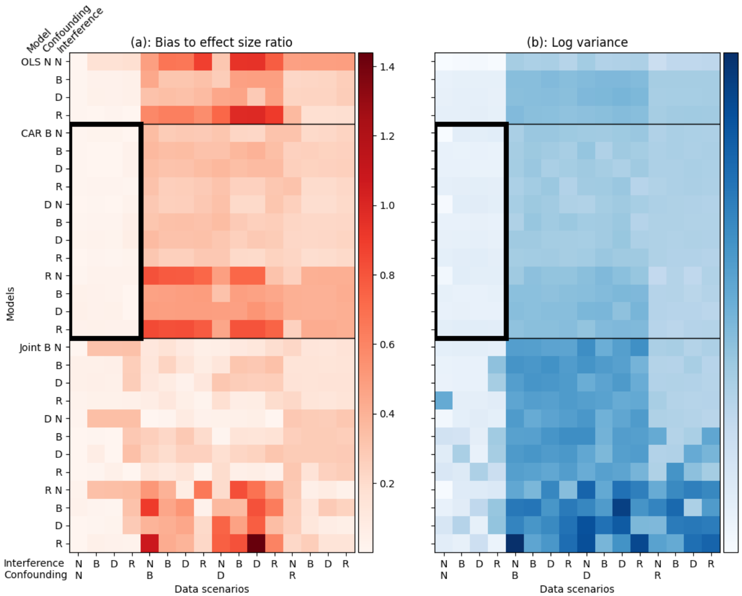
Result 1: Prefer less complex models

N: none
B: binary
D: distance
R: region



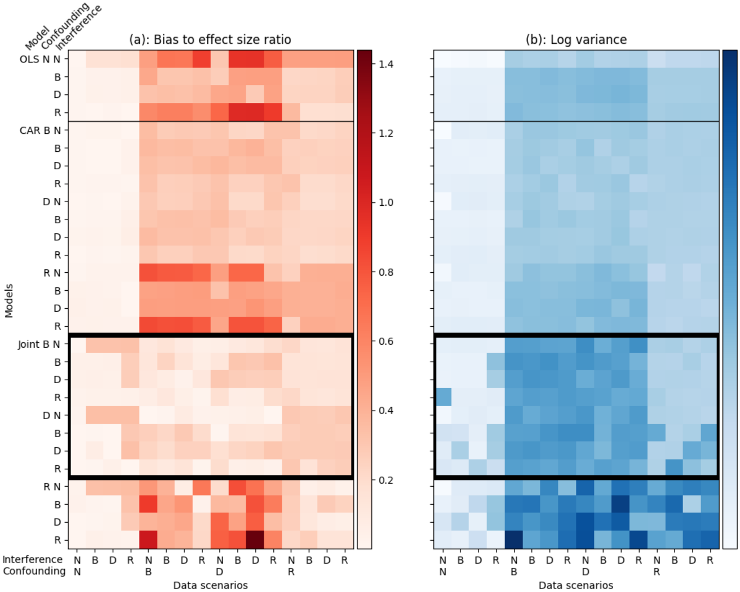
Result 2: Prefer CAR to OLS if spatial issues are possible

N: none
B: binary
D: distance
R: region



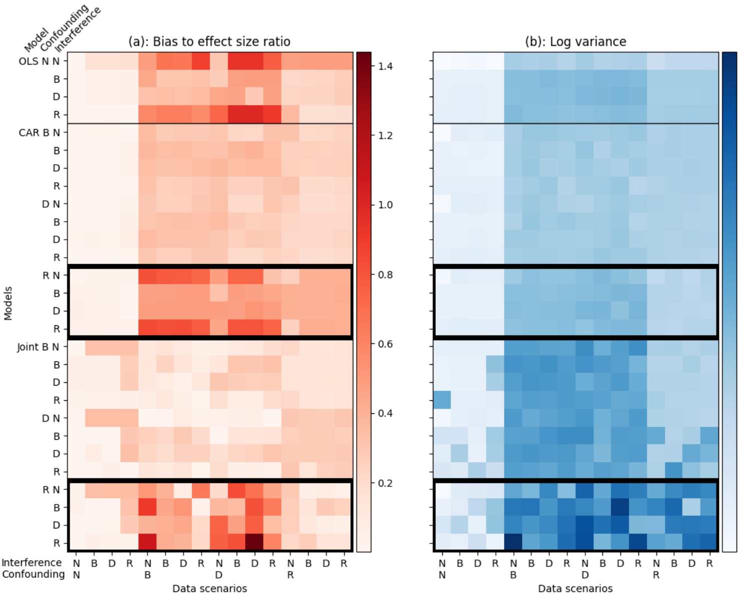
Result 3: Prefer Joint to CAR if spatial issues are likely

N: none
B: binary
D: distance
R: region

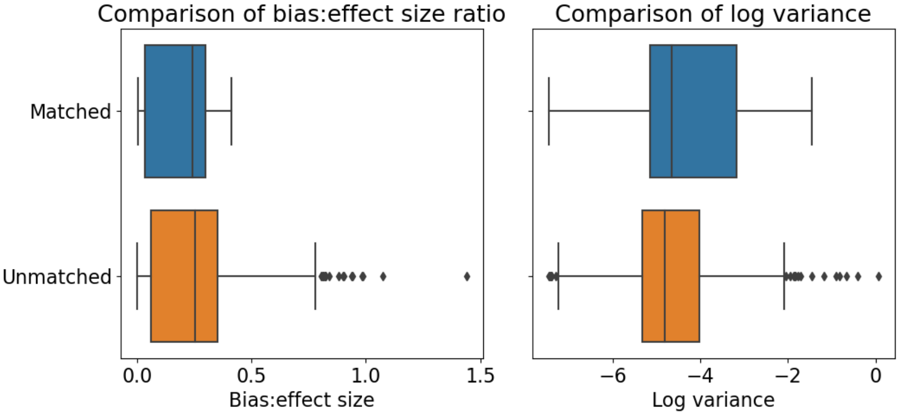


Result 4: Region-based weights have limited use cases

N: none
B: binary
D: distance
R: region



Result 5: *Post hoc* diagnostics are critical



Results: Summary of takeaways

1. Prefer less complex models
2. If there is a possibility of spatial issues, prefer CAR to OLS
3. If there is a strong possibility of spatial issues, prefer Joint (with binary confounding adjustment)
4. Be vigilant for region-based weight use cases
5. *Post hoc* diagnostics can illuminate issues in model structure

Key lessons

- Proliferation of new models in GIScience spurs need for meta-analytical research
- Value of working with domain experts on spatial problems
- Next steps include expanding the code base, developing tutorials, and documentation to enable widespread usage
- Python package and simulation data are available at `github.com/tdhoffman`

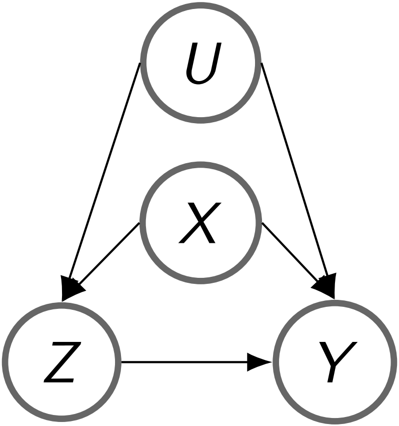
Acknowledgements: Thanks to Sarah Bardin, Drew Trgovac, Dylan Connor, Amy Frazier, and the Frazier-Connor-Kedron lab group for their constructive insights and feedback!

Simulation study: models

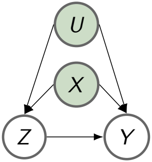
Name	Model
OLS	$y \sim N(X\beta + Z\tau, \sigma^2)$
CAR	$y \sim N(X\beta + Z\tau + U, \sigma^2)$ $U \sim \text{CAR}(\rho_U, \sigma_U^2)$
Joint	$y \sim N(X\beta + Z\tau + U, \sigma^2)$ $Z \sim \text{Bernoulli}(\pi)$ $\pi = \text{expit}(X\alpha + \phi U + V)$ $U \sim \text{CAR}(\rho_U, \sigma_U^2)$ $V \sim \text{CAR}(\rho_V, \sigma_V^2)$

- Interference adjustment: rewrite $\tilde{Z} = [Z, WZ]$ and $\tilde{\tau} = [\tau_1, \tau_2]$
- OLS cannot model confounding, while CAR and Joint must model confounding $\implies$ 28 total models

Data generating processes



Confounders

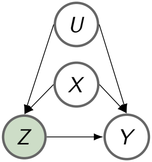


Confounders

Inputs: ρ_X, σ_X^2, ρ_U, σ_U^2

$$\varepsilon_X \sim \text{Unif}(-\sigma_X, \sigma_X)$$
$$X = (I - \rho_X W_C)^{-1} \varepsilon_X$$
$$U \sim \text{CAR}(\rho_U, \sigma_U^2)$$

Treatment

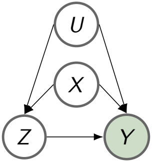


Treatment

$\rho_V \rightarrow$, $\sigma^2_V \rightarrow$, $\alpha \rightarrow$, $\phi \rightarrow$

$$V \sim \text{CAR}(\rho_V, \sigma^2_V)$$
$$\pi = \text{expit}(X\alpha + \phi U + V)$$
$$Z \sim \text{Bernoulli}(\pi)$$

Outcome



β, τ, $\tilde{\tau}$, σ_Y^2 →

Outcome

$$Y \sim N(X\beta + Z\tau + W_I Z\tilde{\tau} + U, \sigma_Y^2)$$