Differnces-in-Differences with Spatial Spillover

Kyle Butts

September 30, 2020

Spatial Spillovers

Effects of treatments at states, counties, etc. can spill over boundaries.

- e.g. large employer opening/closing in a county have positive or negative employment effects on nearby counties

Differences-in-differences that compare outcomes among treated units with "control" units are biased from the spillover of effects.

Research Question

What are the sources of bias when researchers do not account for sptial spillover of treatment effects?

- In what contexts are these biases particularly large?
- Do current simple adjustments work?

Preview of Results

There are two sources of bias:

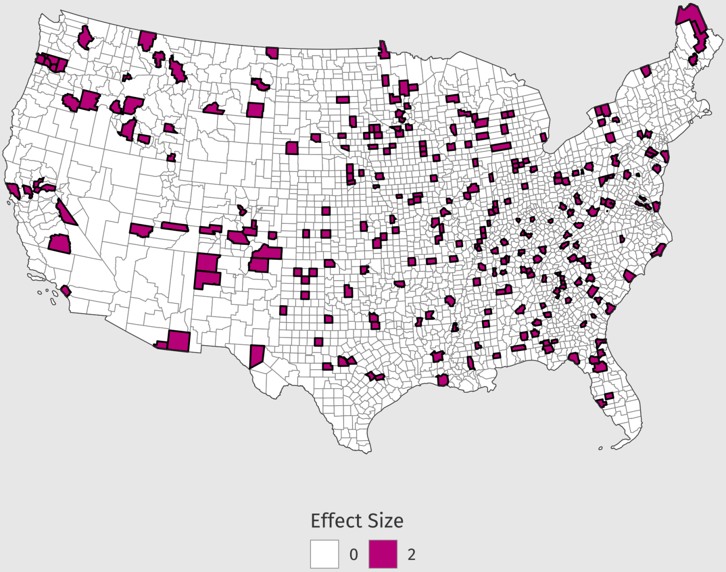
1. Treatment effects spillover onto control units
2. Treatment effects spillover onto also treated units

Solutions:

- Fixing bias 1 is relatively easy and works better when you have relatively few treated units
- Bias 2 worsens as treatment locations bunch more (spatial correlation of treatment increases)
- However, spatial correlation of treatment decreases bias 1

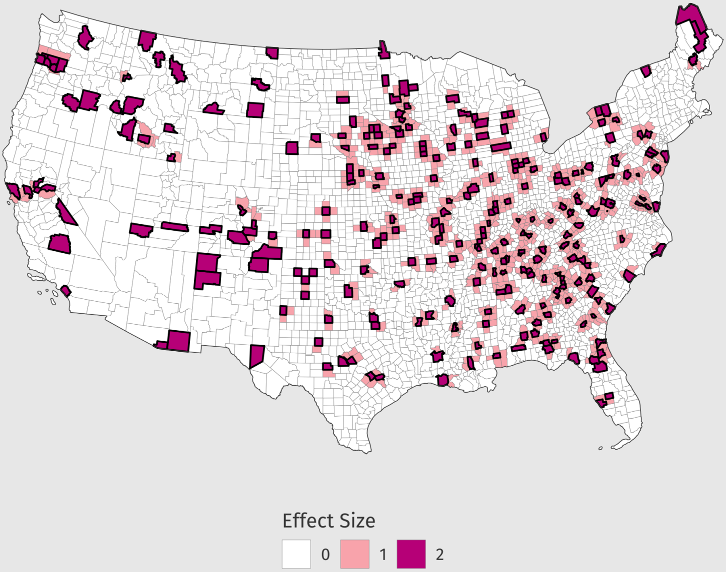
Direct Effect of Treatment

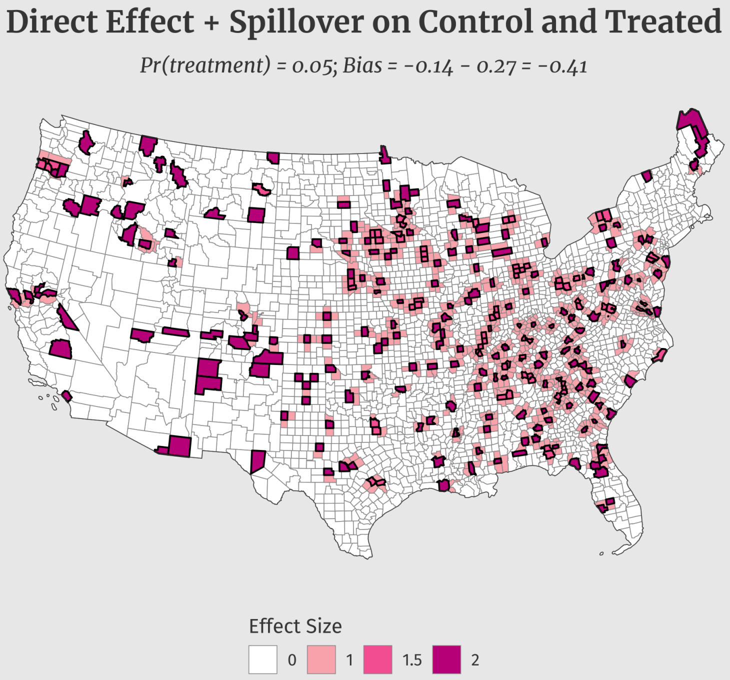
Pr(treatment) = 5%; Bias = 0



Direct Effect + Spillover on Control

Pr(treatment) = 5%; Bias = −0.27




Direct Effect + Spillover on Control and Treated
Pr(treatment) = 0.05; Bias = -0.14 - 0.27 = -0.41
Effect Size
0
1
1.5
2

Potential Outcomes Framework

For exposition, I will label units as counties. Assume all treatment occurs at the same time (2-periods or pre-post averages).[1]

- Let $Y_{it}(D_i, h_i(\vec{D}))$ be the potential outcome of county $i \in \{1, \dots, N\}$ at time t with treatment status $D_i \in \{0, 1\}$.
- $\vec{D} \in \{0, 1\}^N$ is the vector of all units treatments.
- The function $h_i(\vec{D})$ maps the entire treatment vector into a scalar that completely determines the spillover.
- In the world with no treatment, $\vec{0}$, we say $h_i(\vec{0}) = 0$.

[1] This form adapts the framework of Vazquez-Bare (2020) for the diff-in-diff setting.

Examples of $h_i(\vec{D})$

Examples of $h_i(\vec{D})$:

- **Treatment within x miles:**
  $h_i(\vec{D}) = 1$ if there is a treated unit within x-miles of unit i and $= 0$ otherwise
- **Average treatment among k-nearest neighbors:**
  $h_i(\vec{D}) = \sum_{j=1}^{k} D_{n(i,j)}/k$ where $n(i,j)$ represents the j-th closest neighbor to unit i.

Estimand of Interest

Estimand of Interest:

$$\tau_{\text{direct}} \equiv \mathbb{E}\left[Y_{i,1}(1,0) - Y_{i,1}(0,0) \mid D_i = 1 \right]$$

This is the direct effect in the absense of spillovers.

Parallel Trends

- I assume a modified version of the parallel trends assumption:

$$\mathbb{E}\left[Y_{i,1}(0,0) - Y_{i,0}(0,0) \mid D_i = 1, \vec{D} = 0\right]$$

$$= \mathbb{E}\left[Y_{i,1}(0,0) - Y_{i,0}(0,0) \mid D_i = 0, \vec{D} = 0\right],$$

- In the complete absence of treatment (not just the absence of individual i's treatment), the change in potential outcomes from period 0 to 1 would not depend on treatment status

What does Diff-in-Diff identify?

- With that, I decompose the diff-in-diff estimate as follows:

$$
\begin{aligned}
&\mathbb{E}_i\left[Y_{i1} - Y_{i0} \mid D_i = 1\right] - \mathbb{E}_i\left[Y_{i1} - Y_{i0} \mid D_i = 0\right] \\
&\quad = \mathbb{E}_i\left[Y_{i1}(1,0) - Y_{i1}(0,0) \mid D_i = 1\right] \\
&\qquad + \mathbb{E}_i\left[Y_{i1}(1, h_i(\vec{D})) - Y_{i1}(1,0) \mid D_i = 1\right] \\
&\qquad - \mathbb{E}_i\left[Y_{i1}(0, h_i(\vec{D})) - Y_{i1}(0,0) \mid D_i = 0\right] \\
&\quad \equiv \tau_{\text{direct}} + \tau_{\text{spillover, treated}} - \tau_{\text{spillover, control}}
\end{aligned}
$$

Signing the Bias

$$\tau_{\text{direct}} + \tau_{\text{spillover, treated}} - \tau_{\text{spillover, control}}$$

Table 1: Bias from Spillovers

	+ Spillover	- Spillover
onto Treated	+ Bias	- Bias
onto Control	- Bias	+ Bias

Example of Bias Simulation

Treated Spillover:

$\mathbb{E}_i \left[Y_{i1}(1, h_i(\vec{D})) - Y_{i1}(1, 0) \mid D_i = 1 \right] = -0.5$ if county i is within 40 miles of a treated county and 0 otherwise.

Control Spillover:

$\mathbb{E}_i \left[Y_{i1}(0, h_i(\vec{D})) - Y_{i1}(0, 0) \mid D_i = 0 \right] = 1$ if county i is within 40 miles of a treated county and 0 otherwise.

- Then, the bias term is

$$-0.5 * \% \text{ of treated within 40 miles of a treated county}$$
$$-1 * \% \text{ of control within 40 miles of a treated county}$$

Monte Carlo Simulations

- Simulated the following Data Generating Process

$$y_{it} = -2 + \mu_t + \mu_i + \tau_{\text{direct}} D_{it} + \tau_{\text{spillover, control}} (1 - D_{it}) \text{Near}_{it} \\ + \tau_{\text{spillover, treated}} D_{it} \text{Near}_{it} + \varepsilon_{it}$$

- Observation i is a U.S. county, year $t \in \{2000, \ldots, 2019\}$, treatment D_{it} turns on in 2010 and is assigned randomly or through a spatial process.
- Near_{it} is an indicator for having your county centroid be within 40 miles of a treated unit.
- $\tau_{\text{direct}} = 2$, $\tau_{\text{spillover, control}} = 1$ and $\tau_{\text{spillover, treat}} = -0.5$.

Source of Bias 1: Control Units

Simulation 1:

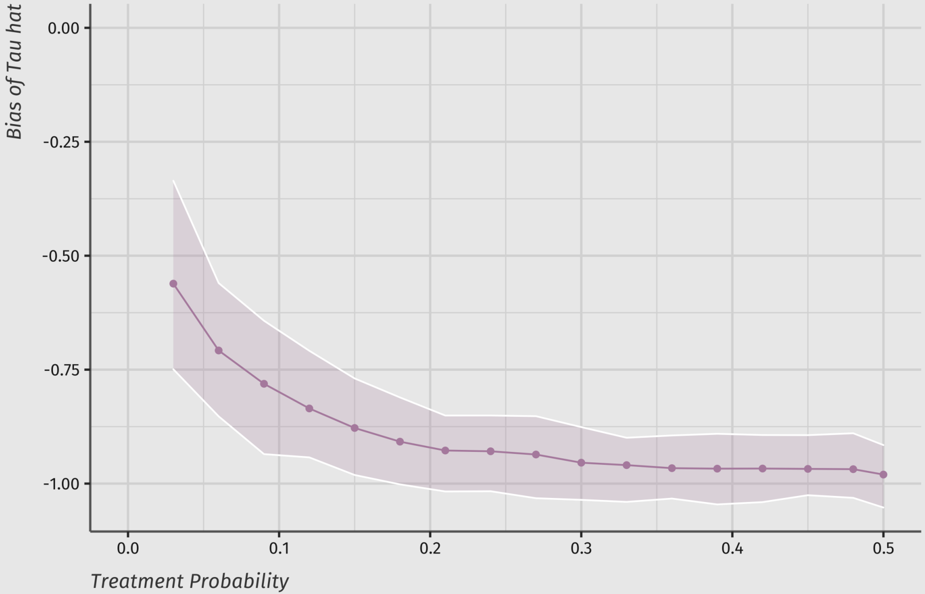
- The first simulation changes the probability of treatment, D_{it}.
- Assume, for now, there is only spillover on the control.
- Estimate the following equation:

$$y_{it} = \alpha + \mu_t + \mu_i + \tau D_{it} + \epsilon_{it}$$

- The bias of our estimate is $Bias = \hat{\tau} - 2$

As more units are treated, the bias increases

Monte Carlo simulations with 100 trials per simulation



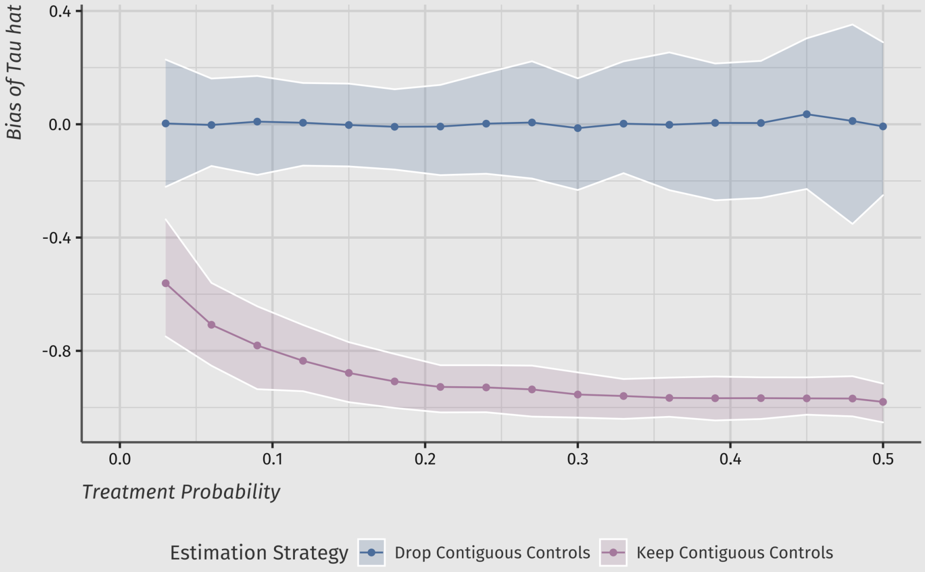
Solution: Removing "contaminated" controls

A common solution to the problem of spillover is to restimate without neighboring control units and see what happens to the bias.

This simulation assumes the correct spillover control units, i.e. control units with $h_i(\vec{D}) \neq 0$, which is not possible for researchers.

Dropping control units removes bias, but increases variance

Monte Carlo simulations with 100 trials per simulation



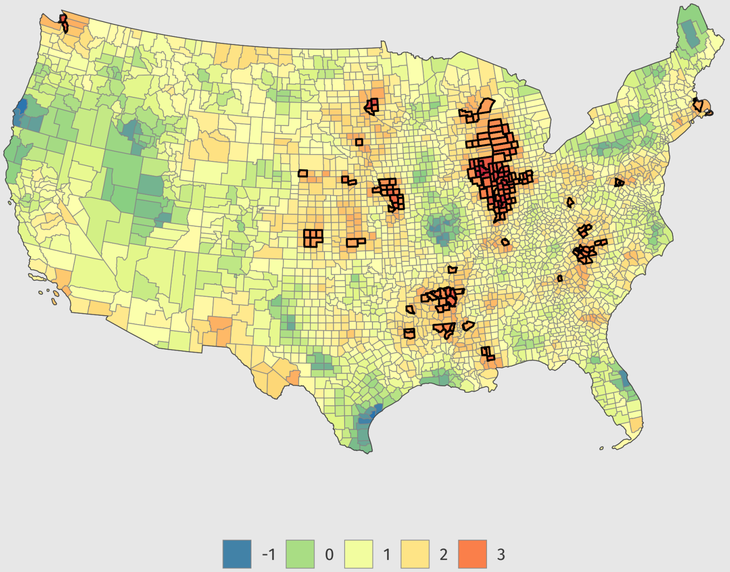
Source of Bias 2: Treated Units

Simulation 2:

- I now add in the bias from spillover of treated units on to other treated units.
- This source of bias only occurs when treated units are located next to eachother, so the magnitude of bias depends on spatial autocorrelation of D_{it}.
- Using a method from geosciences, I generate correlated "fields" across the US.
- Then, for the counties in the top 10% of values (hot spots in a field), I increase the probability of treatment.
- The unconditional probability remains fixed at 10%.

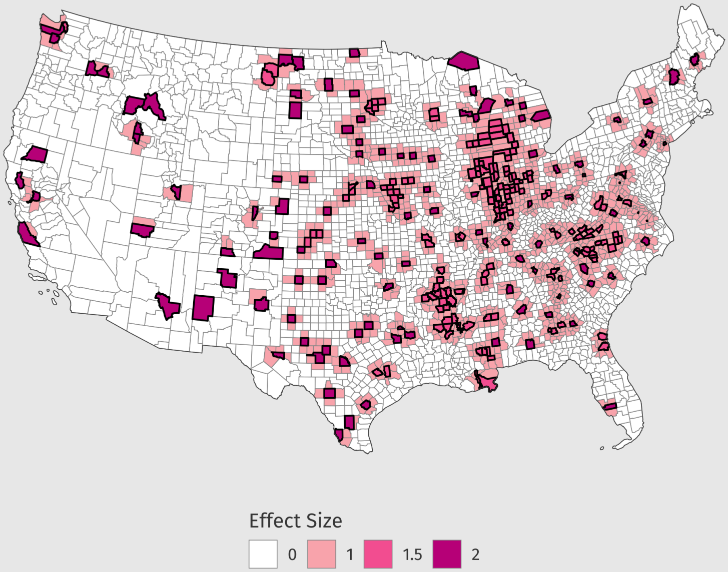
Kriging Example

Spatial Autocorrelation Measure = 1.4



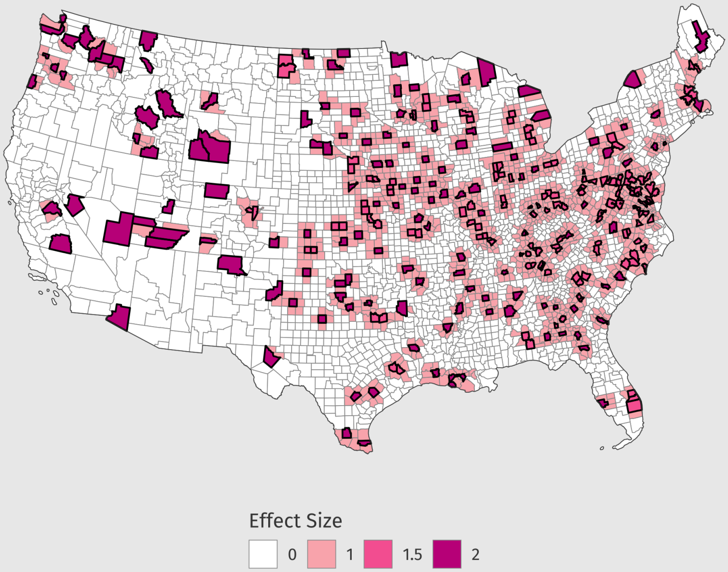
Large Spatial Autocorrelation

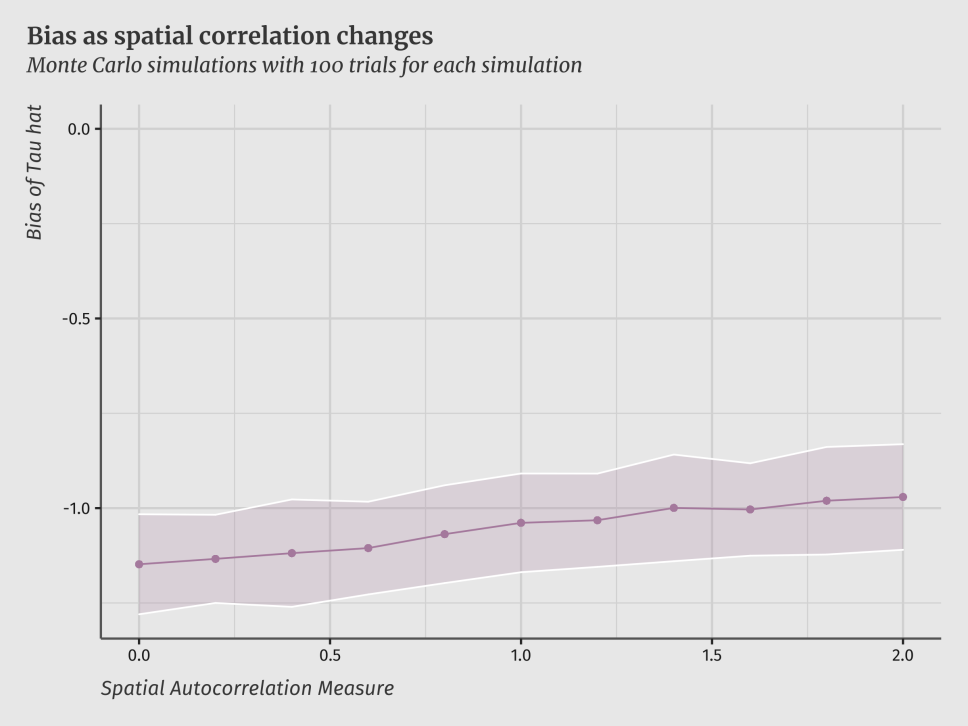
Pr(treatment) = 5%; Bias = -0.32 - 0.39 = -0.71; Spatial Autocorrelation Measure = 1.4

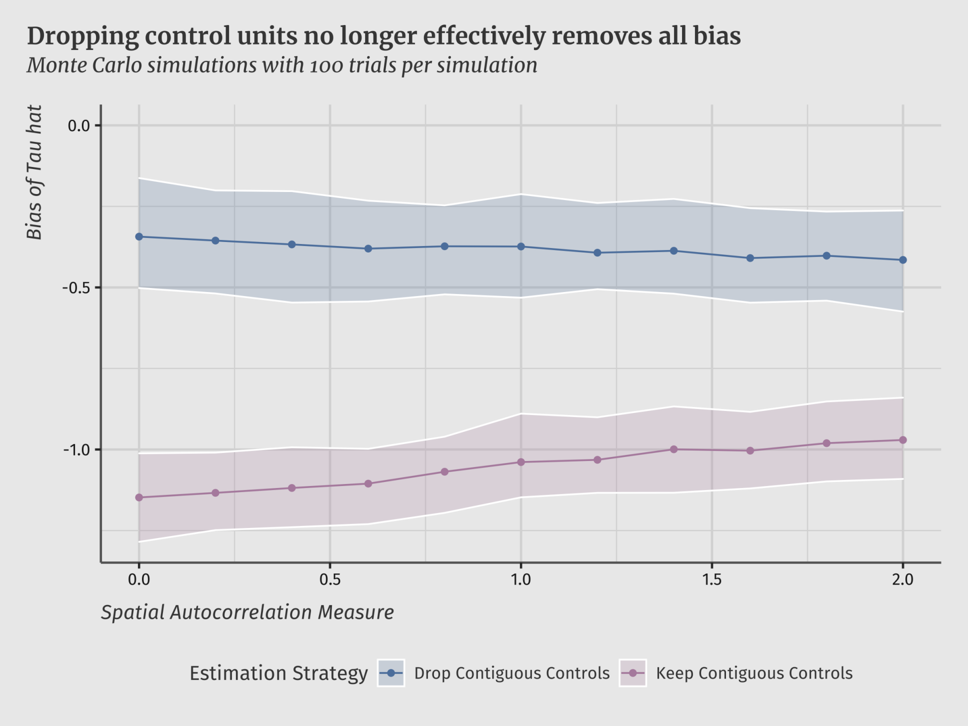


Small Spatial Autocorrelation

Pr(treatment) = 5%; Bias = −0.27 − 0.44 = −0.71; Spatial Autocorrelation Measure = 0.4




Bias as spatial correlation changes
Monte Carlo simulations with 100 trials for each simulation
Bias of Tau hat
0.0
-0.5
-1.0
0.0
0.5
1.0
1.5
2.0
Spatial Autocorrelation Measure



Dropping control units no longer effectively removes all bias
Monte Carlo simulations with 100 trials per simulation
Bias of Tau hat
0.0
-0.5
-1.0
0.0
0.5
1.0
1.5
2.0
Spatial Autocorrelation Measure
Estimation Strategy
Drop Contiguous Controls
Keep Contiguous Controls

Spillovers as estimand of interest

Until now, we assumed our estimand of interest is τ_{direct}.

However, the two other spillover effects are of interest as well:

- $\tau_{\text{spillover, control}}$: Do the benefits of a treated county come at a cost to neighbor counties?
- $\tau_{\text{spillover, treated}}$: Does the estimated effect change based on others treatment? (This is what you should consider if you are a policy maker)

To estimate the spillover effects, we have to parameterize $h_i(\vec{D})$ function and the potential outcomes function $Y_i(D_i, h_i(\vec{D}))$.

Next Steps

- Monte Carlo: estimating spillovers and the problem of misspecification
- Application: replicate paper(s) with a consideration of spillovers (any ideas?)