Difference-in-Differences with Spatial Spillovers

Kyle Butts

July 27, 2021

Spatial Spillovers

Researchers aim to estimate the **average treatment effect on the treated**:

$$\tau \equiv \mathbb{E}\left[Y_{i1}(1) - Y_{i1}(0) \mid D_i = 1\right]$$

Estimation is complicated by **Spillover Effects**

Spillover effects occur when the effect of treatment extends over the treatment boundaries (states, counties, etc.). Example: Amazon Shipping Center

- A shipping center opening in county *c* has positive employment effects on **nearby control counties**
- Having nearby counties with shipping centers raises employment and therefore reduces the effect of **treated counties** through general equilibrium effects

This Paper

In this paper, I...

- Present a potential outcomes framework to formalize spillover effects and evaluate ad-hoc adjustments commonly found in the literature
- Propose an estimation strategy that improves on current practices by being more robust to spillovers
- Apply this framework to improve estimation of the local effect of place-based policies in Urban Economics

Bias from Spatial Spillovers

The canonical difference-in-differences estimate is:

$$\hat{\tau} = \underbrace{\hat{\mathbb{E}}\left[Y_{i1} - Y_{i0} \mid D_i = 1\right]}_{\text{Counterfactual Trend } + \tau} - \underbrace{\hat{\mathbb{E}}\left[Y_{i1} - Y_{i0} \mid D_i = 0\right]}_{\text{Counterfactual Trend}}$$

Two problems occur in the presence of spillover effects:

Bias from Spatial Spillovers

The canonical difference-in-differences estimate is:

$$\hat{\tau} = \underbrace{\hat{\mathbb{E}}\left[Y_{i1} - Y_{i0} \mid D_i = 1\right]}_{\text{Counterfactual Trend } + \tau} - \underbrace{\hat{\mathbb{E}}\left[Y_{i1} - Y_{i0} \mid D_i = 0\right]}_{\substack{\text{Counterfactual Trend} \\ + \text{ Spillover on Control}}}$$

Two problems occur in the presence of spillover effects:

- **Spillover onto Control Units:**
  Nearby "control" units fail to estimate counterfactual trends because they are affected by treatment

Bias from Spatial Spillovers

The canonical difference-in-differences estimate is:

$$\hat{\tau} = \underbrace{\hat{\mathbb{E}}\left[Y_{i1} - Y_{i0} \mid D_i = 1\right]}_{\substack{\text{Counterfactual Trend } + \tau \\ + \text{ Spillover on Treated}}} - \underbrace{\hat{\mathbb{E}}\left[Y_{i1} - Y_{i0} \mid D_i = 0\right]}_{\substack{\text{Counterfactual Trend} \\ + \text{ Spillover on Control}}}$$

Two problems occur in the presence of spillover effects:

- **Spillover onto Control Units:**
  Nearby "control" units fail to estimate counterfactual trends because they are affected by treatment
- **Spillover onto other Treated Units:**
  Treated units are also affected by nearby units and therefore combine "direct" effects

Remove Bias

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

$$\mathbb{E}[\hat{\tau}] = \tau + \text{Spillover on Control} + \text{Spillover on Treated}$$

Remove Bias

$$y_{it} = \mu_t + \mu_i + \tau D_{it} + \tau_{\text{spill,control}} \text{Within}_{it} * (1 - D_{it}) + \tau_{\text{spill,treat}} \text{Within}_{it} * D_{it} + +\varepsilon_{it}$$

$$\mathbb{E}[\hat{\tau}] = \tau,$$

so long as Within_{it} contains all the units with spillovers.

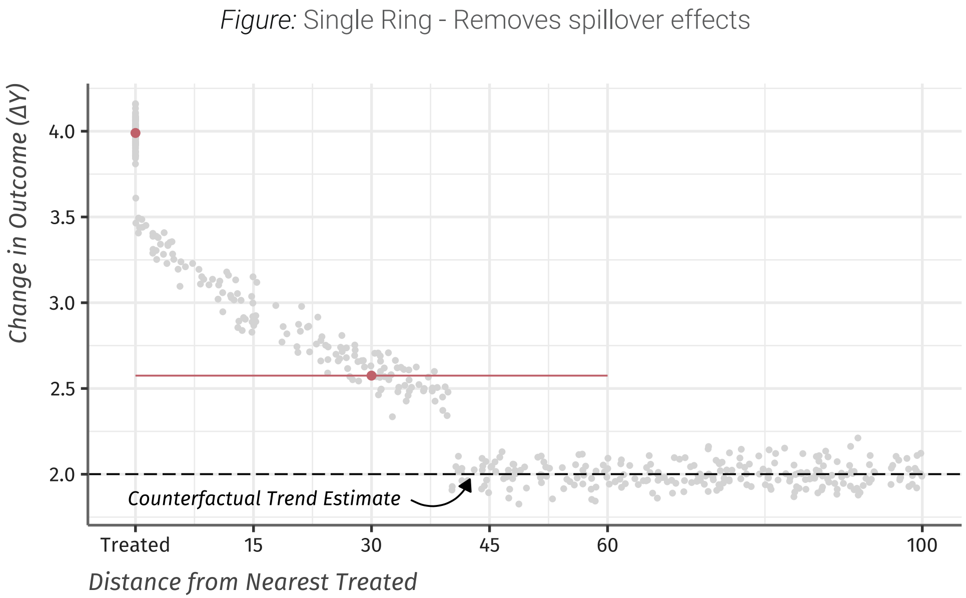
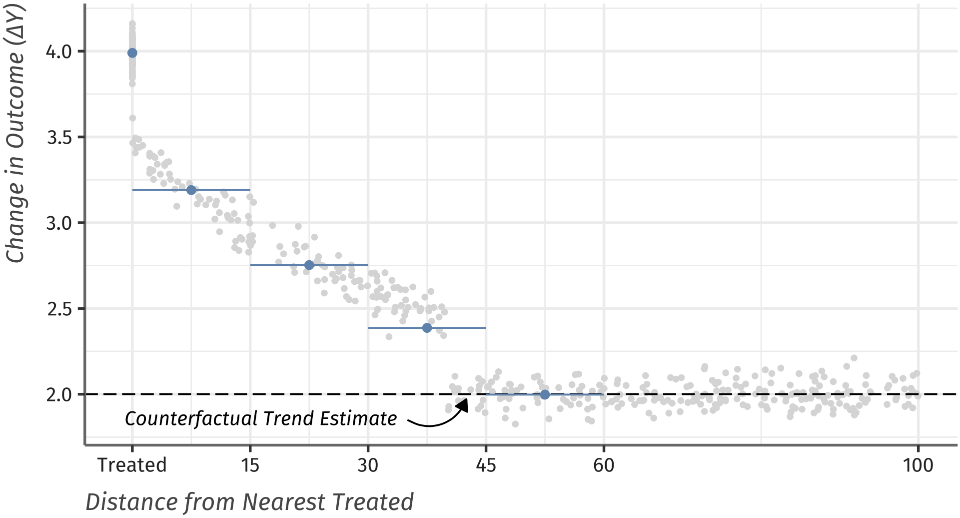


Figure: Single Ring - Removes spillover effects
Change in Outcome (ΔY)
4.0
3.5
3.0
2.5
2.0
Counterfactual Trend Estimate
Treated
15
30
45
60
100
Distance from Nearest Treated

Figure: Multiple Rings - Improves estimation of spillover effects



Outline

1– Formalize spillovers into a potential outcomes framework:

[Clarke (2017), Berg and Streitz (2019), and Verbitsky-Savitz and Raudenbush (2012)]

- I decompose the difference-in-differences estimator into three parts: Direct Effect of Treatment, Spillover onto Treated Units, Spillover onto Control Units
- Show that an indicator for being close to treated units remove *all bias* so long as the indicator contains all units affected by spillovers
- 'Rings' are able to estimate spillover effects while still removing all bias

Outline

2– Apply framework to Urban Economics

- Revisit Kline and Moretti (2014a) analysis of the Tennessee Valley Authority
  - The local effect estimate is contaminated by spillover effects to neighboring counties (Kline and Moretti, 2014b)
  - Large scale manufacturing investment creates an 'urban shadow' (Cuberes, Desmet, and Rappaport, 2021; Fujita, Krugman, and Venables, 2001)
- Discuss how framework can reconcile conflicting findings on effect of federal Empowerment Zones (Busso, Gregory, and Kline, 2013; Neumark and Kolko, 2010)
- Event Study estimates of Community Health Centers find highly localized efffects (Bailey and Goodman-Bacon, 2015)

Outline

Theory

Estimation of Spillovers

Application in Urban Economics

Event Study

Conclusion

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]

- $Y_{it}(D_i, h(\vec{D}, i))$ is 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 $h(\vec{D}, i)$ maps the entire treatment vector into an 'exposure mapping' which can be a scalar or a vector.
- No exposure is when $h(\vec{D}, i) = \vec{0}$.

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]

- $Y_{it}(D_i, h(\vec{D}, i))$ is 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 $h(\vec{D}, i)$ maps the entire treatment vector into an 'exposure mapping' which can be a scalar or a vector.
- No exposure is when $h(\vec{D}, i) = \vec{0}$.

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]

- $Y_{it}(D_i, h(\vec{D}, i))$ is the potential outcome of county $i \in \{1, \ldots, N\}$ 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 $h(\vec{D}, i)$ maps the entire treatment vector into an 'exposure mapping' which can be a scalar or a vector.
- No exposure is when $h(\vec{D}, i) = \vec{0}$.

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

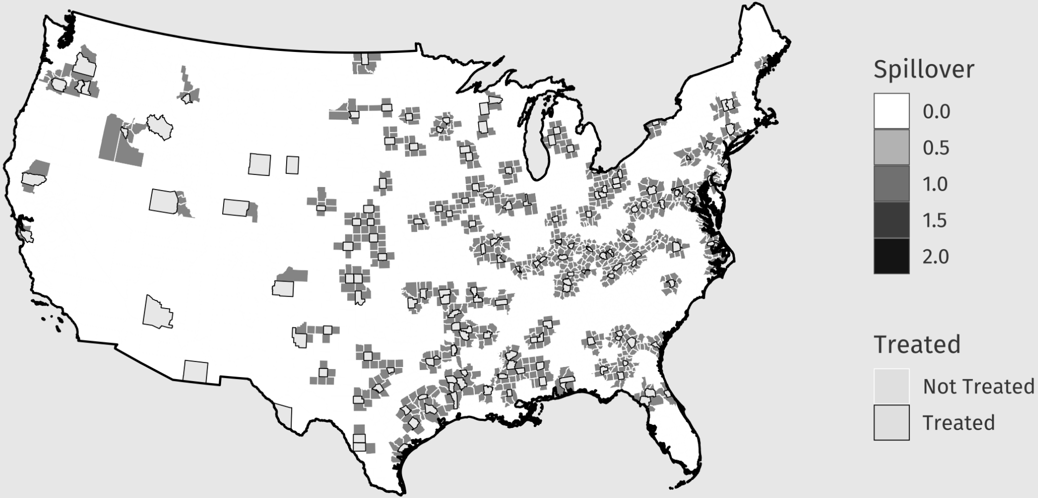
Examples of $h_i(\vec{D})$:

- **Treatment within x miles:**

  $h(\vec{D}, i) = max_j \, 1\left(d(i,j) \leq x\right)$ where $d(i,j)$ is the distance between counties i and j.

  - e.g. library access where x is the maximum distance people will travel
  - Spillovers are non-additive, i.e. spillover effects do not depend on number of nearby treated areas

Within 40mi.



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

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

- **Treatment within x miles:**

  $h(\vec{D}, i) = max_j \, 1(d(i,j) \leq x)$ where $d(i,j)$ is the distance between counties i and j.

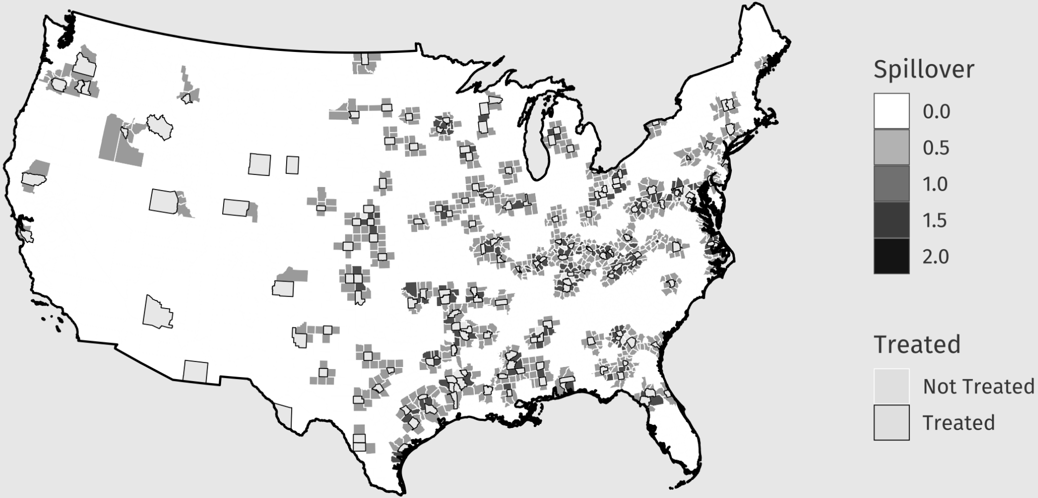
  - e.g. library access where x is the maximum distance people will travel
  - Spillovers are non-additive

- **Number of Treated within x miles:**

  $h(\vec{D}, i) = \sum_{j=1}^{k} 1(d(i,j) \leq x)$.

  - e.g. Amazon shipping center
  - Agglomeration economies suggest spillovers are additive

Within 40mi. (Additive)



Estimand of Interest

Estimand of Interest:

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

Spillover Effects:

$$\tau_{\text{spillover, treated}} \equiv \mathbb{E}\left[Y_{i1}(1, h_i(\vec{D})) - Y_{i1}(1,0) \mid D_i = 1\right]$$

$$\tau_{\text{spillover, control}} \equiv \mathbb{E}\left[Y_{i1}(0, h_i(\vec{D})) - Y_{i1}(0,0) \mid D_i = 0\right]$$

Estimand of Interest

Estimand of Interest:

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

Spillover Effects:

$$\tau_{\text{spillover, treated}} \equiv \mathbb{E}\left[Y_{i1}(1, h_i(\vec{D})) - Y_{i1}(1,0) \mid D_i = 1\right]$$

$$\tau_{\text{spillover, control}} = \mathbb{E}\left[Y_{i1}(0, h_i(\vec{D})) - Y_{i1}(0,0) \mid D_i = 0\right]$$

Parallel Trends

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

Assumption: *Parallel Counterfactual Trends*

$$\mathbb{E}\big[\underbrace{Y_{i,1}(0,\vec{0}) - Y_{i,0}(0,\vec{0})}_{\text{Counterfactual Trend}} \mid D_i = 1\big] = \mathbb{E}\big[\underbrace{Y_{i,1}(0,\vec{0}) - Y_{i,0}(0,\vec{0})}_{\text{Counterfactual Trend}} \mid D_i = 0\big]$$

In the *complete absence of treatment* (not just the absence of individual i's treatment): Changes in outcomes do not depend on treatment status

What does Difference-in-Differences identify?

With the parallel trends assumption and random assignment of D_i, I decompose the difference-in-differences estimate as follows:

$$\mathbb{E}\left[\hat{\tau}\right] = \underbrace{\mathbb{E}\left[Y_{i1} - Y_{i0} \mid D_i = 1\right] - \mathbb{E}\left[Y_{i1} - Y_{i0} \mid D_i = 0\right]}_{\text{Difference-in-Differences}}$$

$$= \mathbb{E}\left[Y_{i1}(1,0) - Y_{i1}(0,0) \mid D_i = 1\right]$$

$$+ \quad \mathbb{E}\left[Y_{i1}(1, h_i(\vec{D})) - Y_{i1}(1,0) \mid D_i = 1\right]$$

$$- \quad \mathbb{E}\left[Y_{i1}(0, h_i(\vec{D})) - Y_{i1}(0,0) \mid D_i = 0\right]$$

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

What does Difference-in-Differences identify?

With the parallel trends assumption and random assignment of D_i, I decompose the difference-in-differences estimate as follows:

$$\mathbb{E}[\hat{\tau}] = \underbrace{\mathbb{E}[Y_{i1} - Y_{i0} \mid D_i = 1] - \mathbb{E}[Y_{i1} - Y_{i0} \mid D_i = 0]}_{\text{Difference-in-Differences}}$$

$$\begin{aligned} &= \mathbb{E}[Y_{i1}(1,0) - Y_{i1}(0,0) \mid D_i = 1] \\ &\quad + \quad \mathbb{E}\left[Y_{i1}(1, h_i(\vec{D})) - Y_{i1}(1,0) \mid D_i = 1\right] \\ &\quad - \quad \mathbb{E}\left[Y_{i1}(0, h_i(\vec{D})) - Y_{i1}(0,0) \mid D_i = 0\right] \end{aligned}$$

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

Removing Bias

Assumption: *Spillovers are Local*

Let $d(i, j)$ be the distance between units i and j. There exists a distance $\bar{d}$ such that

(i) For all units i,

$$\min_{j:D_j=1} d(i, j) > \bar{d} \implies h(\vec{D}, i) = \vec{0}.$$

(ii) There are treated and control units such that $\min_{j:\ D_j=1} d(i, j) > \bar{d}$.

Removing Bias

With assumption that spillovers are local, define S_it to be an indicator equal to one in the post period for all units with $h(\vec{D}, i) \neq \vec{0}$ (and potentially some units with $= \vec{0}$).

Estimation of the following equation:

$$y_{it} = \mu_t + \mu_i + \tau D_{it} + \tau_{\text{spill,treat}} S_{it} * D_{it} + \tau_{\text{spill,control}} S_{it} * (1 - D_{it}) + \varepsilon_{it}$$

- $\hat{\tau}$ is consistent for τ_{direct}
- $\hat{\tau}_{\text{spill}}$'s are not consistent for average spillover effects (hence the need for rings).

Aside – Removing “contaminated” controls

A common solution to the problem of spillover is to re-estimate on a subsample with neighboring control units removed.

This is not recommended for two reasons:

- Removing control units from the sample decreases precision of the estimates
- Spillover effects on treated units will still remain in the estimand.

Outline

- Theory
- Estimation of Spillovers
- Application in Urban Economics
- Event Study
- Conclusion

Robustness to Misspecification

Generate data using the following data-generating process with different exposure mappings:

$$y = \mu_t + \mu_i + 2D_{it} + \beta_{\text{spill,control}} * (1 - D_{it}) h(\vec{D}, i) + \varepsilon_{it}$$

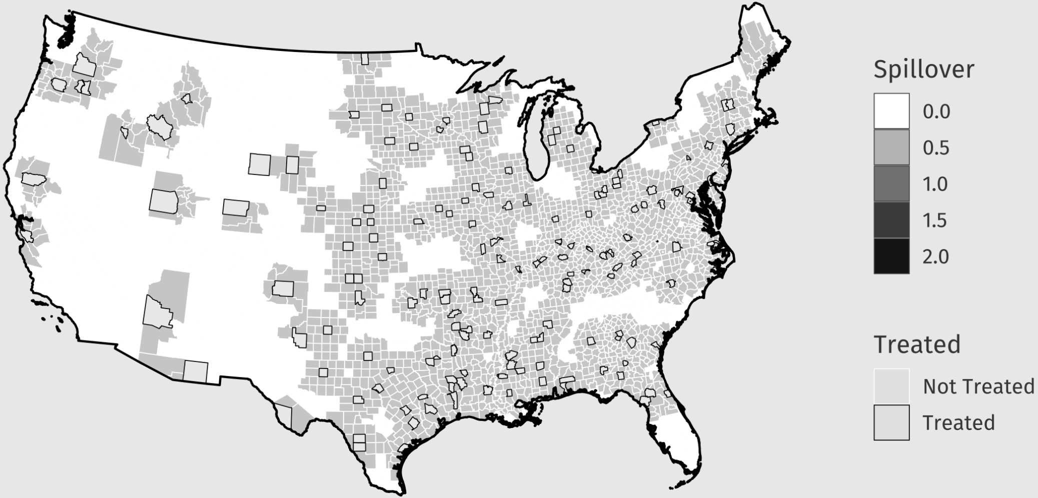
- Observation i is a U.S. county, year $t \in \{2000, \ldots, 2019\}$, treatment D_{it} turns on in 2010 and is assigned randomly.
- Details in the paper, but μ_t is normally distributed on a linear time trend, μ_i are normally distributed

Then, I estimate each data-generating process using (potentially) misspecified $\tilde{h}(\vec{D}, i)$ and report the average estimate bias.

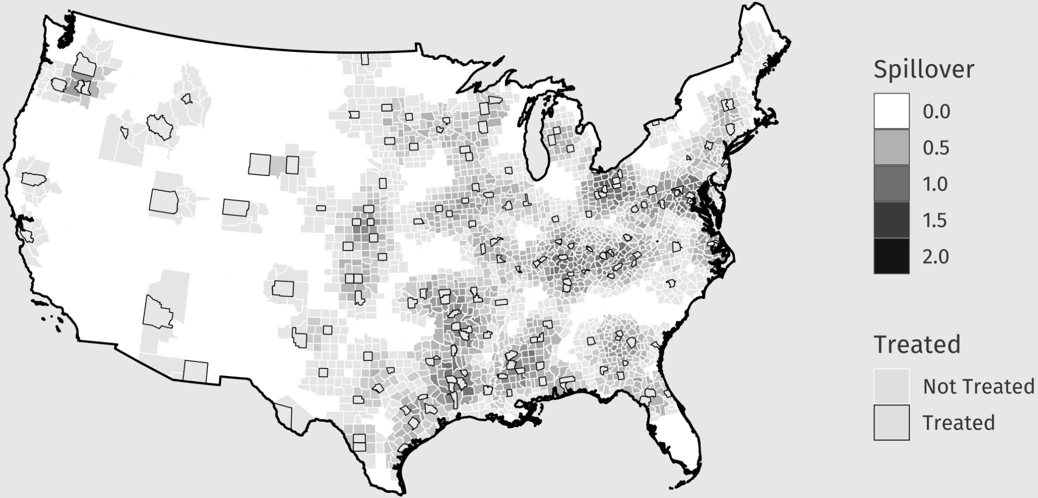
Specifications of $h(\vec{D}, i)$

- *Within* $40/80$*mi.*:
  - Indicator for nearest treated unit being within $40/80$ miles.
- *Within* $40/80$*mi. (Additive)*:
  - Number of treated units being within $40/80$ miles.

Within 80mi.



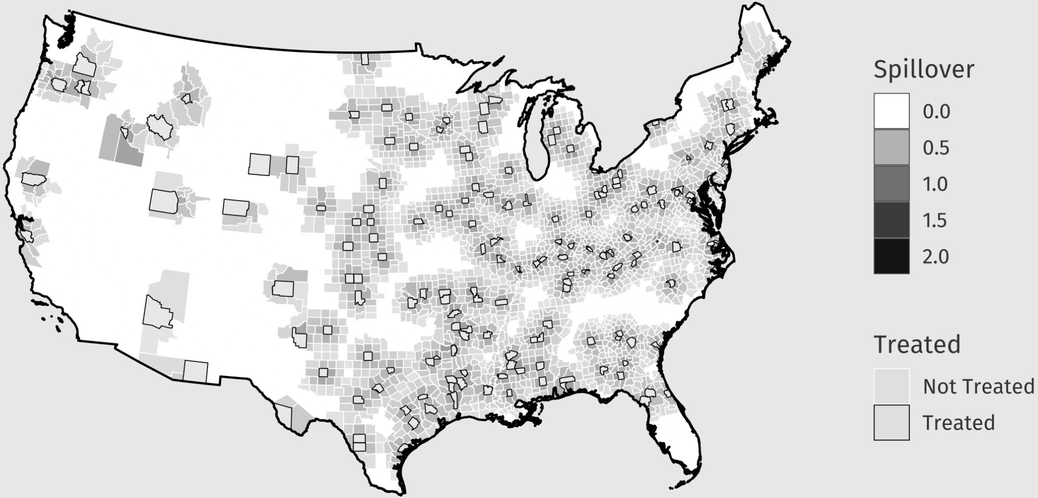
Within 80mi. (Additive)



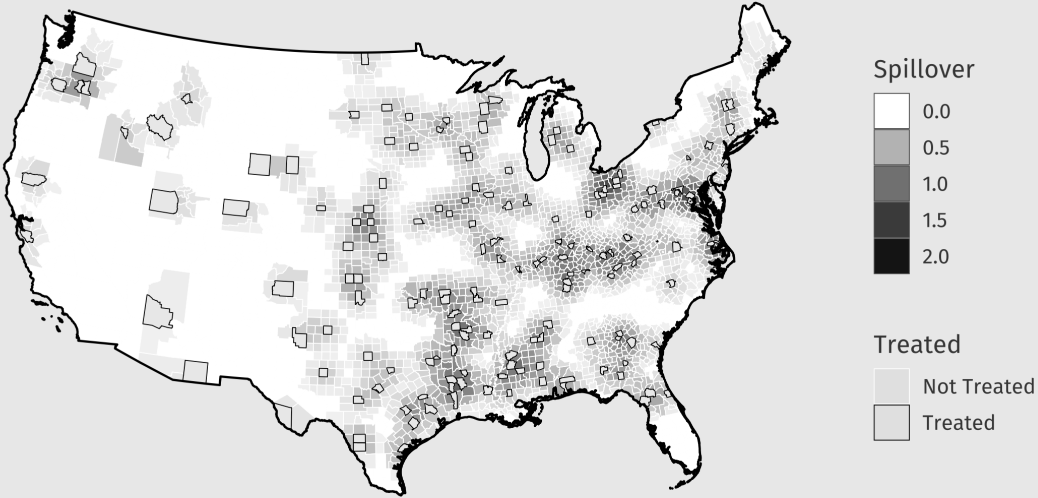
Specifications of $h(\vec{D}, i)$

- *Decay*:
  - $\max_j D_j * e^{-0.02d(i,j)} * 1(d(i,j) < 80)$
- *Decay (Additive)*:
  - $\sum_j D_j * e^{-0.02d(i,j)} * 1(d(i,j) < 80)$

Decay



Decay (Additive)



Specifications of $h(\vec{D}, i)$

- *Rings*:
  - Set of concentric rings. For each ring, indicator for nearest treated unit being within that distance bin
- *Rings (Additive)*:
  - Set of concentric rings. For each ring, number of treated units being within that distance bin

Rings (0-20, 20-30, 30-40, 40-60, 60-80)

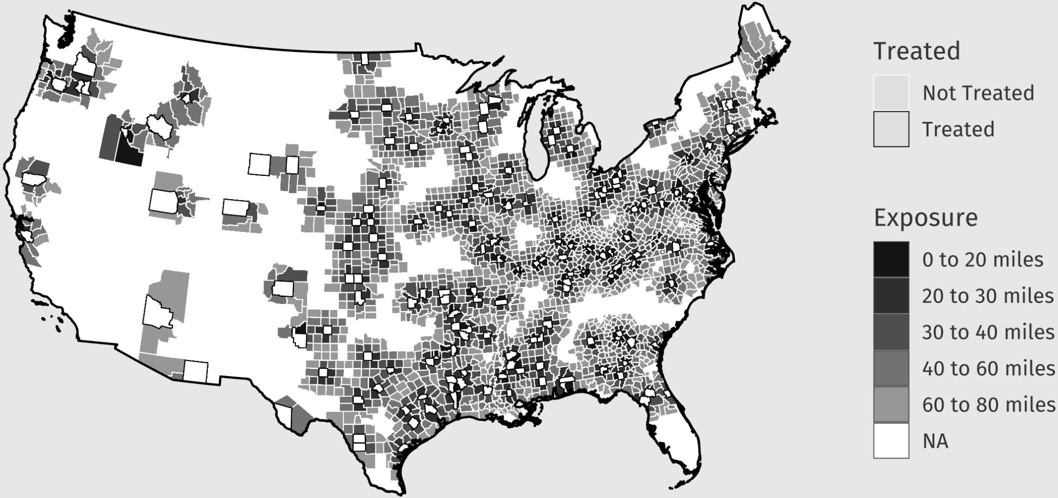


Table: Bias from Misspecification of Spillovers

	Data-Generating Process					
Specification	**Within 40mi.**	**Within 80mi.**	**Within 40mi. (Additive)**	**Within 80mi. (Additive)**	**Decay 80mi.**	**Decay 80mi. (Additive)**
TWFE (No Spillovers)	0.258	0.258	0.258	0.258	0.258	0.258
Within 40mi.	−0.005	0.213	−0.005	0.176	0.159	0.143
Within 80mi.	−0.009	−0.009	−0.009	−0.009	−0.009	−0.009
Within 40mi. (Additive)	0.043	0.221	−0.006	0.177	0.174	0.143
Within 80mi. (Additive)	0.034	0.134	−0.012	−0.009	0.099	−0.010
Decay 80mi.	−0.159	0.070	−0.174	0.014	−0.009	−0.033
Decay 80mi. (Additive)	−0.023	0.148	−0.084	0.019	0.088	−0.008
Rings (0-20, 20-30, 30-40)	−0.005	0.213	−0.005	0.176	0.159	0.143
Rings (0-20, 20-30, 30-40, 40-60, 60-80)	−0.009	−0.009	−0.009	−0.009	−0.009	−0.009
Rings (0-20, 20-30, 30-40, 40-60, 60-80) (Additive)	0.036	0.134	−0.008	−0.008	0.100	−0.009

Indicator (or set of rings) that captures *all* affected unit removes all bias

Table: Bias from Misspecification of Spillovers

	Data-Generating Process					
Specification	**Within 40mi.**	**Within 80mi.**	**Within 40mi. (Additive)**	**Within 80mi. (Additive)**	**Decay 80mi.**	**Decay 80mi. (Additive)**
TWFE (No Spillovers)	0.258	0.258	0.258	0.258	0.258	0.258
Within 40mi.	−0.005	0.213	−0.005	0.176	0.159	0.143
Within 80mi.	−0.009	−0.009	−0.009	−0.009	−0.009	−0.009
Within 40mi. (Additive)	0.043	0.221	−0.006	0.177	0.174	0.143
Within 80mi. (Additive)	0.034	0.134	−0.012	−0.009	0.099	−0.010
Decay 80mi.	−0.159	0.070	−0.174	0.014	−0.009	−0.033
Decay 80mi. (Additive)	−0.023	0.148	−0.084	0.019	0.088	−0.008
Rings (0-20, 20-30, 30-40)	−0.005	0.213	−0.005	0.176	0.159	0.143
Rings (0-20, 20-30, 30-40, 40-60, 60-80)	−0.009	−0.009	−0.009	−0.009	−0.009	−0.009
Rings (0-20, 20-30, 30-40, 40-60, 60-80) (Additive)	0.036	0.134	−0.008	−0.008	0.100	−0.009

Indicator (or set of rings) that captures *all* affected unit removes all bias

Table: Bias from Misspecification of Spillovers

Specification	Within 40mi.	Within 80mi.	Within 40mi. (Additive)	Within 80mi. (Additive)	Decay 80mi.	Decay 80mi. (Additive)
	Data-Generating Process					
TWFE (No Spillovers)	0.258	0.258	0.258	0.258	0.258	0.258
Within 40mi.	−0.005	0.213	−0.005	0.176	0.159	0.143
Within 80mi.	−0.009	−0.009	−0.009	−0.009	−0.009	−0.009
Within 40mi. (Additive)	0.043	0.221	−0.006	0.177	0.174	0.143
Within 80mi. (Additive)	0.034	0.134	−0.012	−0.009	0.099	−0.010
Decay 80mi.	−0.159	0.070	−0.174	0.014	−0.009	−0.033
Decay 80mi. (Additive)	−0.023	0.148	−0.084	0.019	0.088	−0.008
Rings (0-20, 20-30, 30-40)	−0.005	0.213	−0.005	0.176	0.159	0.143
Rings (0-20, 20-30, 30-40, 40-60, 60-80)	−0.009	−0.009	−0.009	−0.009	−0.009	−0.009
Rings (0-20, 20-30, 30-40, 40-60, 60-80) (Additive)	0.036	0.134	−0.008	−0.008	0.100	−0.009

Indicator (or set of rings) that captures *all* affected unit removes all bias

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 treatment of neighbors?

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

Estimation of Spillover Effects

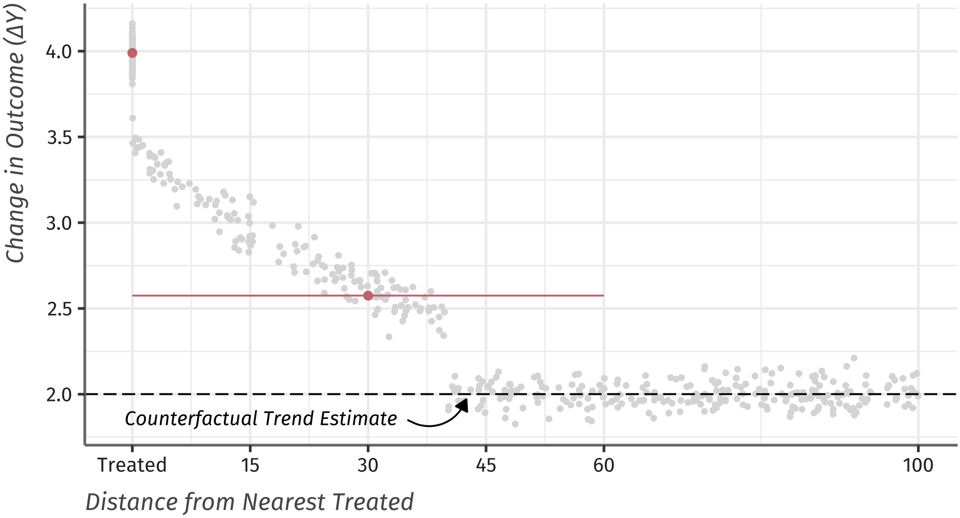
To see which specifications can predict spillover effects well, I estimate the spillover effects for each control unit, $\hat{\beta}_{\text{spill, control}} * \tilde{h}(\vec{D}, i)$.

Then calculate

$$1 - \frac{\overbrace{\sum_{i:D_i=0} (\beta_{\text{spill, control}} h(\vec{D}, i) - \hat{\beta}_{\text{spill, control}} \tilde{h}(\vec{D}, i))^2}^{\text{Mean Square Prediction Error}}}{\underbrace{\sum_{i:D_i=0} (\beta_{\text{spill, control}} h(\vec{D}, i))^2}_{\text{Normalization}}}$$

This gives the proportion of spillovers explained by $\tilde{h}(\vec{D}, i)$

Figure: Single Ring - Removes spillover effects



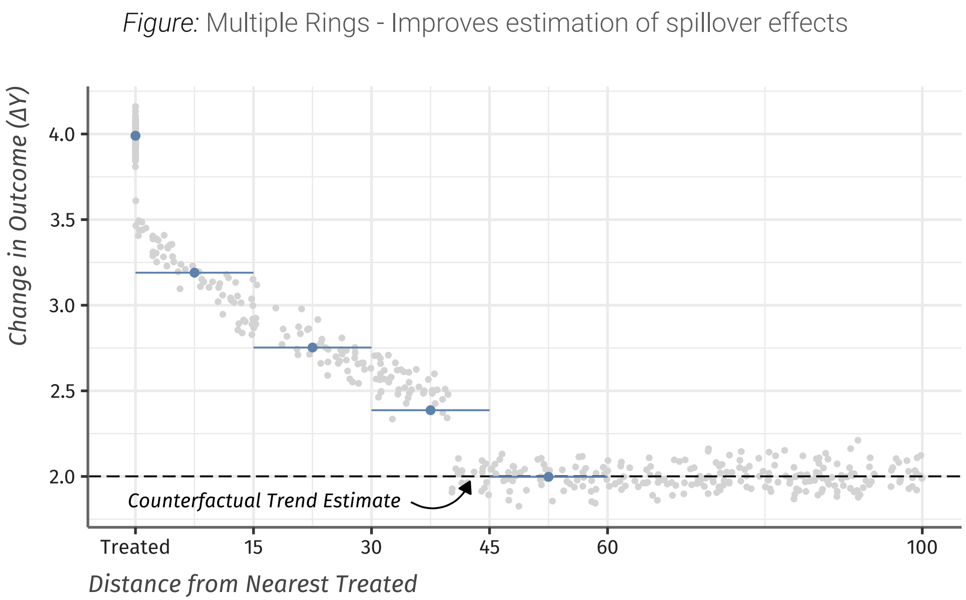


Figure: Multiple Rings - Improves estimation of spillover effects
Change in Outcome (ΔY)
4.0
3.5
3.0
2.5
2.0
Counterfactual Trend Estimate
Treated
15
30
45
60
100
Distance from Nearest Treated

Table: Percent of Spillovers Predicted by Specification

Specification	Within 40mi.	Within 80mi.	Within 40mi. (Additive)	Within 80mi. (Additive)	Decay 80mi.	Decay 80mi. (Additive)
	Data-Generating Process					
TWFE (No Spillovers)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Within 40mi.	99.4%	25.9%	85.6%	38.8%	59.5%	56.1%
Within 80mi.	39.8%	96.2%	34.3%	71.7%	85.6%	68.0%
Within 40mi. (Additive)	85.3%	21.2%	99.5%	40.6%	52.0%	60.7%
Within 80mi. (Additive)	45.8%	61.8%	47.2%	98.4%	71.0%	93.6%
Decay 80mi.	60.1%	82.5%	52.7%	75.8%	97.5%	82.2%
Decay 80mi. (Additive)	60.7%	56.9%	63.8%	93.5%	79.0%	98.7%
Rings (0-20, 20-30, 30-40)	98.4%	23.7%	85.9%	37.5%	58.9%	56.2%
Rings (0-20, 20-30, 30-40, 40-60, 60-80)	96.6%	91.7%	84.2%	72.7%	91.9%	78.4%
Rings (0-20, 20-30, 30-40, 40-60, 60-80) (Additive)	83.5%	57.4%	97.6%	95.0%	73.5%	94.9%

Donuts perform best at estimating spillover effects.

It is important to get Additive vs. Non-Additive correct.

Table: Percent of Spillovers Predicted by Specification

	Data-Generating Process					
Specification	**Within 40mi.**	**Within 80mi.**	**Within 40mi. (Additive)**	**Within 80mi. (Additive)**	**Decay 80mi.**	**Decay 80mi. (Additive)**
TWFE (No Spillovers)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Within 40mi.	99.4%	25.9%	85.6%	38.8%	59.5%	56.1%
Within 80mi.	39.8%	96.2%	34.3%	71.7%	85.6%	68.0%
Within 40mi. (Additive)	85.3%	21.2%	99.5%	40.6%	52.0%	60.7%
Within 80mi. (Additive)	45.8%	61.8%	47.2%	98.4%	71.0%	93.6%
Decay 80mi.	60.1%	82.5%	52.7%	75.8%	97.5%	82.2%
Decay 80mi. (Additive)	60.7%	56.9%	63.8%	93.5%	79.0%	98.7%
Rings (0-20, 20-30, 30-40)	98.4%	23.7%	85.9%	37.5%	58.9%	56.2%
Rings (0-20, 20-30, 30-40, 40-60, 60-80)	96.6%	91.7%	84.2%	72.7%	91.9%	78.4%
Rings (0-20, 20-30, 30-40, 40-60, 60-80) (Additive)	83.5%	57.4%	97.6%	95.0%	73.5%	94.9%

Donuts perform best at estimating spillover effects.

It is important to get Additive vs. Non-Additive correct.

Table: Percent of Spillovers Predicted by Specification

Specification	Within 40mi.	Within 80mi.	Within 40mi. (Additive)	Within 80mi. (Additive)	Decay 80mi.	Decay 80mi. (Additive)
	Data-Generating Process					
TWFE (No Spillovers)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Within 40mi.	99.4%	25.9%	85.6%	38.8%	59.5%	56.1%
Within 80mi.	39.8%	96.2%	34.3%	71.7%	85.6%	68.0%
Within 40mi. (Additive)	85.3%	21.2%	99.5%	40.6%	52.0%	60.7%
Within 80mi. (Additive)	45.8%	61.8%	47.2%	98.4%	71.0%	93.6%
Decay 80mi.	60.1%	82.5%	52.7%	75.8%	97.5%	82.2%
Decay 80mi. (Additive)	60.7%	56.9%	63.8%	93.5%	79.0%	98.7%
Rings (0-20, 20-30, 30-40)	98.4%	23.7%	85.9%	37.5%	58.9%	56.2%
Rings (0-20, 20-30, 30-40, 40-60, 60-80)	96.6%	91.7%	84.2%	72.7%	91.9%	78.4%
Rings (0-20, 20-30, 30-40, 40-60, 60-80) (Additive)	83.5%	57.4%	97.6%	95.0%	73.5%	94.9%

Donuts perform best at estimating spillover effects.

It is important to get Additive vs. Non-Additive correct.

Outline

Theory

Estimation of Spillovers

Application in Urban Economics

Event Study

Conclusion

Tennessee Valley Authority

Kline and Moretti (2014a) look at the long-run impacts of the Tennessee Valley Authority (TVA).

- The TVA was a large-scale federal investment started in 1934 that focused on improving manufacturing economy. (Hundreds of dollars spent anually per person)
- The program focused on large-scale dam construction that brought cheap wholesale electricity to the region

Research Questions:

- What are the local effects of TVA on manufacturing and agricultural economies?
- Do these effects come at the cost of other counties?

Tennessee Valley Authority

Kline and Moretti (2014a) look at the long-run impacts of the Tennessee Valley Authority (TVA).

- The TVA was a large-scale federal investment started in 1934 that focused on improving manufacturing economy. (Hundreds of dollars spent anually per person)
- The program focused on large-scale dam construction that brought cheap wholesale electricity to the region

Research Questions:

- What are the local effects of TVA on manufacturing and agricultural economies?
- Do these effects come at the cost of other counties?

Identification

Kline and Moretti (2014a) run the county-level difference-in-differences specification:

$$y_{c,2000} - y_{c,1940} = \alpha + \text{TVA}_c \tau + X_{c,1940}\beta + (\varepsilon_{c,2000} - \varepsilon_{c,1940}) \quad (1)$$

- y are outcomes for agricultural employment and manufacturing employment.
- TVA_c is the treatment variable
- $X_{c,1940}$ allow for different long-term trends based on covariates in 1940.

They trim the sample using a logit regression to predict treatment using $X_{c,1940}$ and then keep control units in the top 75% of predicted probability.

Spillovers in the TVA Context

In our context, there is reason to believe spillovers can occur to nearby counties

- **Agriculture:**
  - Employees might be drawn to hire wages for new manufacturing jobs in Tennessee Valley (negative spillover on control units)
- **Manufacturing:**
  - Cheap electricity might be available to nearby counties (positive spillover on control units)
  - Manufacturing jobs that would have been created in the control units in the absence of treatment might move to the Tennessee Valley (negative spillover on control units)

Specification including spillovers

$$\Delta y_c = \alpha + \text{TVA}_i \tau + \sum_{d \in \text{Dist}} \text{Ring}(d)\delta_d + X_{i,1940}\beta + \Delta\varepsilon_c \quad (2)$$

- Ring(d) is a set of indicators for being in the following distance bins (in miles) from the Tennessee Valley Authority:

$$d \in \{(0,50], (50,100], (100,150](150,200]\}$$

Effective Sample and Spillover Variables

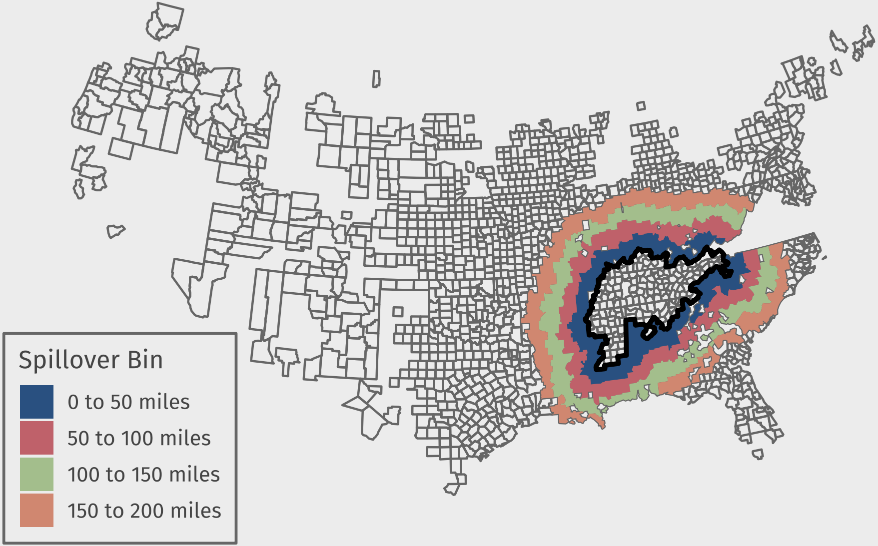


Table: Effects of Tennessee Valley Authority on Decadel Growth, 1940-2000

	Diff-in-Diff	**Diff-in-Diff with Spillovers**				
	TVA	TVA	TVA between 0-50 mi.	TVA between 50-100 mi.	TVA between 100-150 mi.	TVA between 150-200 mi.
Dependent Var.	(1)	(2)	(3)	(4)	(5)	(6)
Agricultural employment	-0.0514^{***}	-0.0739^{***}	-0.0371^{***}	-0.0164	-0.0298^{***}	-0.0157^{*}
	(0.0114)	(0.0142)	(0.0002)	(0.0114)	(0.0096)	(0.0088)
Manufacturing employment	0.0560^{***}	0.0350	-0.0203^{***}	-0.0245	-0.0331^{*}	-0.0296^{**}
	(0.0161)	(0.0218)	(0.0006)	(0.0282)	(0.0189)	(0.0142)

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table: Effects of Tennessee Valley Authority on Decadel Growth, 1940-2000

	Diff-in-Diff	**Diff-in-Diff with Spillovers**				
	TVA	TVA	TVA between 0-50 mi.	TVA between 50-100 mi.	TVA between 100-150 mi.	TVA between 150-200 mi.
Dependent Var.	(1)	(2)	(3)	(4)	(5)	(6)
Agricultural employment	-0.0514^{***}	-0.0739^{***}	-0.0371^{***}	-0.0164	-0.0298^{***}	-0.0157^{*}
	(0.0114)	(0.0142)	(0.0002)	(0.0114)	(0.0096)	(0.0088)
Manufacturing employment	0.0560^{***}	0.0350	-0.0203^{***}	-0.0245	-0.0331^{*}	-0.0296^{**}
	(0.0161)	(0.0218)	(0.0006)	(0.0282)	(0.0189)	(0.0142)

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Identification Strategies and Place-Based Policies

The literature on federal Enterprise Zones, place-based policy that gives tax breaks to businesses that locate within the boundary, has found conflicting results, suggesting positive or near-zero effects of the program (Neumark and Young, 2019).

Identification Strategies and Place-Based Policies

- Busso, Gregory, and Kline (2013) compare census tracts in Empowerment Zones to census tracts that qualified and were rejected from the program. They find significant large reduction of poverty.
- Neumark and Kolko (2010) compare census tracts in Empowerment Zones to census tracts within 1,000 feet of the Zone. They find near-zero effects on poverty.

My framework can explain both of these results. If census tracts just outside the Empowerment Zones also benefit from the policy, then the estimates of Neumark and Kolko (2010) are attenuated towards zero

Identification Strategies and Place-Based Policies

- Busso, Gregory, and Kline (2013) compare census tracts in Empowerment Zones to census tracts that qualified and were rejected from the program. They find significant large reduction of poverty.
- Neumark and Kolko (2010) compare census tracts in Empowerment Zones to census tracts within 1,000 feet of the Zone. They find near-zero effects on poverty.

My framework can explain both of these results. If census tracts just outside the Empowerment Zones also benefit from the policy, then the estimates of Neumark and Kolko (2010) are attenuated towards zero

Identification Strategies and Place-Based Policies

- Busso, Gregory, and Kline (2013) compare census tracts in Empowerment Zones to census tracts that qualified and were rejected from the program. They find significant large reduction of poverty.
- Neumark and Kolko (2010) compare census tracts in Empowerment Zones to census tracts within 1,000 feet of the Zone. They find near-zero effects on poverty.

My framework can explain both of these results. If census tracts just outside the Empowerment Zones also benefit from the policy, then the estimates of Neumark and Kolko (2010) are attenuated towards zero

Outline

Theory

Estimation of Spillovers

Application in Urban Economics

Event Study

Conclusion

Gardner (2021) Overview

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

The problem with estimating this by OLS is that the treatment variable becomes residualized $\tilde{D}_{it}$ and this leads to all sorts of problems... (see new diff-in-diff literature)

Gardner (2021) recommends a two-step approach:

1. Estimate μ_i and μ_t using never-treated/not-year-treated observations (D_{it} = 0). Then subtract off $\hat{\mu}_i$ and $\hat{\mu}_t$.
2. Then, regress $y_{it} - \hat{\mu}_i - \hat{\mu}_t \equiv \tilde{y}_{it}$ on τD_{it} (or event study leads/lags). This estimate is unbiased because D_{it} is non-residualized (standard errors require adjusting).

Gardner (2021) Overview

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

The problem with estimating this by OLS is that the treatment variable becomes residualized $\tilde{D}_{it}$ and this leads to all sorts of problems... (see new diff-in-diff literature)

Gardner (2021) recommends a two-step approach:

1. Estimate μ_i and μ_t using never-treated/not-year-treated observations (D_{it} = 0). Then subtract off $\hat{\mu}_i$ and $\hat{\mu}_t$.
2. Then, regress $y_{it} - \hat{\mu}_i - \hat{\mu}_t \equiv \tilde{y}_{it}$ on τD_{it} (or event study leads/lags). This estimate is unbiased because D_{it} is non-residualized (standard errors require adjusting).

Gardner (2021) Overview

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

The problem with estimating this by OLS is that the treatment variable becomes residualized $\tilde{D}_{it}$ and this leads to all sorts of problems... (see new diff-in-diff literature)

Gardner (2021) recommends a two-step approach:

1. Estimate μ_i and μ_t using never-treated/not-year-treated observations (D_{it} = 0). Then subtract off $\hat{\mu}_i$ and $\hat{\mu}_t$.
2. Then, regress $y_{it} - \hat{\mu}_i - \hat{\mu}_t \equiv \tilde{y}_{it}$ on τD_{it} (or event study leads/lags). This estimate is unbiased because D_{it} is non-residualized (standard errors require adjusting).

Controlling for Spillovers in Staggered Treatment Timing

$$y_{it} = \mu_i + \mu_t + \tau D_{it} + \tau_{\text{spill,control}} \text{Within}_{it} * (1 - D_{it}) + \tau_{\text{spill,treat}} \text{Within}_{it} * D_{it} + \varepsilon_{it}$$

Adjust two-step approach:

1. Estimate μ_i and μ_t using observations that aren't yet treated/affected by spillovers ($D_{it} = 0$ and $Within_{it} = 0$). Then subtract off $\hat{\mu}_i$ and $\hat{\mu}_t$.

2. Then, regress $\tilde{y}_{it}$ on $\tau D_{it} + \tau_{\text{spill,control}} \text{Within}_{it} * (1 - D_{it}) + \tau_{\text{spill,treat}} \text{Within}_{it} * D_{it}$ (or interacted event study leads/lags). This estimate is unbiased because D_{it} is non-residualized (standard errors require adjusting).

Controlling for Spillovers in Staggered Treatment Timing

$$y_{it} = \mu_i + \mu_t + \tau D_{it} + \tau_{\text{spill,control}} \text{Within}_{it} * (1 - D_{it}) + \tau_{\text{spill,treat}} \text{Within}_{it} * D_{it} + \varepsilon_{it}$$

Adjust two-step approach:

1. Estimate μ_i and μ_t using observations that aren't yet treated/affected by spillovers ($D_{it} = 0$ and $Within_{it} = 0$). Then subtract off $\hat{\mu}_i$ and $\hat{\mu}_t$.
2. Then, regress $\tilde{y}_{it}$ on $\tau D_{it} + \tau_{\text{spill,control}} \text{Within}_{it} * (1 - D_{it}) + \tau_{\text{spill,treat}} \text{Within}_{it} * D_{it}$ (or interacted event study leads/lags). This estimate is unbiased because D_{it} is non-residualized (standard errors require adjusting).

Community Health Centers

Bailey and Goodman-Bacon (2015) study the creation of federal community health centers between 1965 and 1974

Research Question: Do low-/no-cost health services lower the mortality rate of the treated counties? *New Question:* Do these effects spread to neighboring counties?

Figure: Effects of Establishment of Community Health Centers

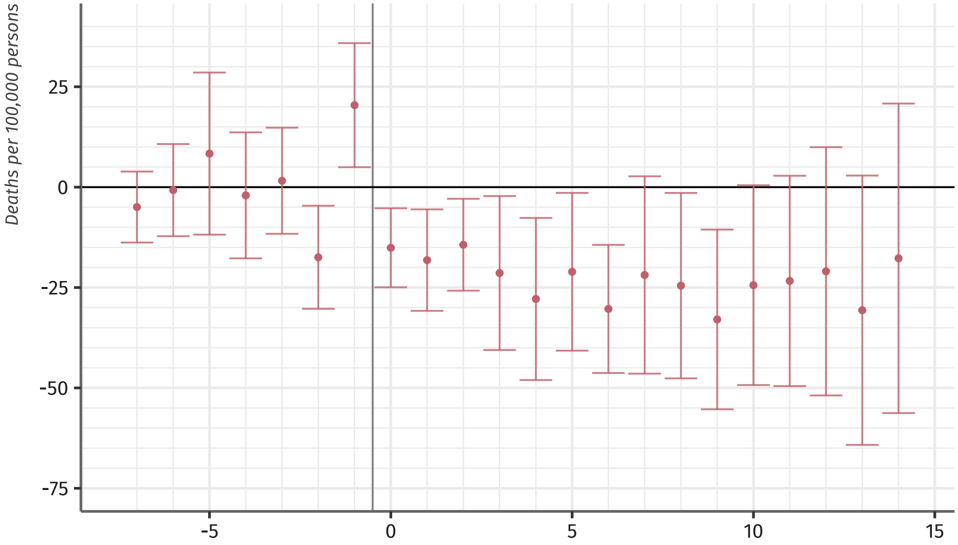
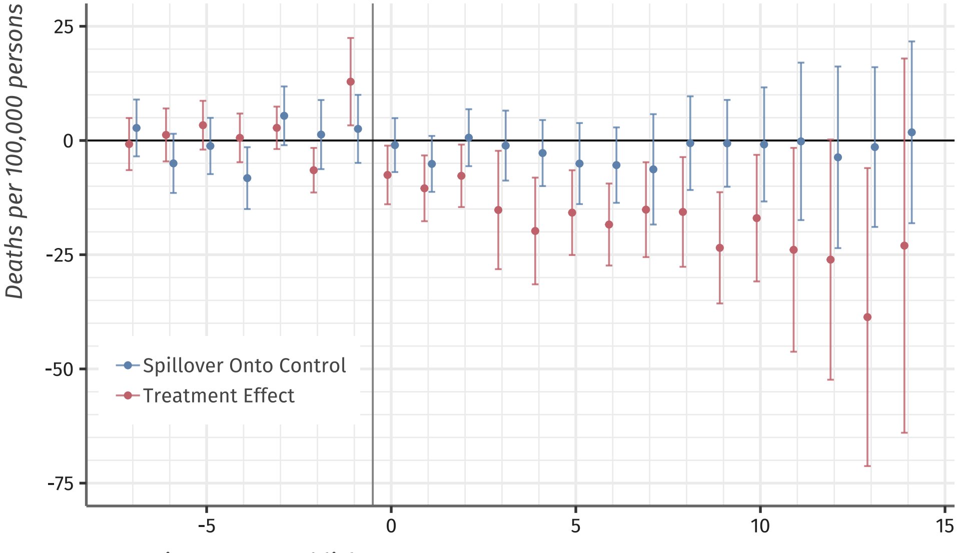


Figure: Direct and Spillover Effects of Community Health Centers



Outline

Theory

Estimation of Spillovers

Application in Urban Economics

Event Study

Conclusion

Conclusion

- I decomposed the TWFE estimate into the direct effect and two spillover terms
- I showed that a set of concentric rings allows for estimation of the direct effect of treatment and they are able to model spillovers well
- For place-based policies, I show the importance of considering spatial spillovers when estimating treatment effects
- More generally, identification strategies that use very close control units in order to minimize differences in unobservables should consider the problems with treatment effect spillovers.