Differences-in-Differences with Spatial Spillover

Kyle Butts

Univ. of Colorado, Boulder

November 20, 2020

ABSTRACT

Empirical work often uses treatment variables defined by geographic boundaries. When researchers ignore the common problem that the effects of treatment cross over borders, difference-in-differences produces biased estimates for the average treatment effect. In this paper, I decompose this bias in two parts. First, the control group no longer identifies the counterfactual trend because their outcomes are affected by treatment. Second, changes in treated units' outcomes reflect the effect of their own treatment status and the effect from the treatment status of "close" units. I use Monte Carlo simulations to demonstrate when the biases are particularly large. Further, I show that a common solution used in the literature of removing 'contaminated' control units only prevents one source of bias. Lastly, I propose improved estimation strategies that can remove both sources of bias.

1 — Introduction

Empirical work in economics often considers treatment assigned by geographic boundaries such as cities, counties, and states. The effect of these treatments do not typically stay within these boundaries whether it be from people crossing borders or from general equilibrium effects affecting neighboring areas. In the causal inference literature, the Stable Unit Treatment Value Assumption (SUTVA) is often assumed which says that treatment effects do not depend on the treatment status of any other unit (Rubin, 1980). This article considers a common violation of SUTVA in a process I label 'spatial spillovers' where 'close' units treatment assignment affects outcomes of nearby units.[1]

Despite the problem of spatial spillovers being common across many settings, Berg and Streitz (2019) document that little empirical analysis takes the problem seriously. In a survey of eight top Economics and Finance Journals in 2017, they find that only 21 articles out of 108 that run differencse-in-differences estimation discuss spillovers.[2] Of those 21, only eight include spillovers in their empirical specification to prevent bias in the estimation of treatment effects.

In the presence of these spatial spillovers, I identify two sources of bias that result when estimating treatment effects by standard difference-in-differences methodologies. First, untreated units that are 'close' to treated units experience effects of treatment and therefore these 'control' units fail to identify the counterfactual trend. When estimating by difference-in-differences, the spillover onto the 'close' control units is averaged into the untreated change in outcomes. In this case, the spillover is subtracted from the estimated treatment effect and biases the estimate in the opposite sign of the spillover effect. For example, if a factory opening in a given county benefits both that county and neighboring counties, the treatment effect estimate is negatively biased because the change in outcome in neighboring counties is higher than it would be absent treatment.

[1] Close can refer to many different things, e.g. geographic distance, node distance in a graph, or social relationships in schools or cities.

[2] The articles surveyed are American Economic Review, Econometrica, the Journal of Political Economy, the Quarterly Journal of Economics, the Review of Economic Studies, the Journal of Finance, the Journal of Financial Economics, and the Review of Financial Studies.

Second, changes in treated units' outcomes reflect the effect of their own treatment status and the effect from the treatment status of "close" units. The spillover is added to the treated units' change in outcomes. Therefore the bias of treatment effect is the same sign as the sign of the spillover. For example a factory opening in two neighboring counties might cause the benefit of each individual factory to decrease. The estimated treatment effect will be negatively biased because there is a negative spillover onto also treated units.

The magnitude of the bias depends on two factors: the size of the spillover and the number of units affected by spillovers. If spillovers are quite large in magnitude or spread far over distance, then the bias will generally be large. This paper provides an explicit form for this bias and enables researchers to give bounds for it under assumptions about the size of the spillover effect and on number of units affected. I use Monte Carlo simulations to quantify the magnitude of bias in various spillover settings.

Berg and Streitz (2019) find that six of the eight papers that include spillovers directly in their estimation strategy simply drop the control units they suspect to be affected by spillovers. Since this is a common strategy by researchers, I use Monte Carlo simulations to show that this method is effective in the case where there is *only* spillovers on to control units. Dropping control unit observations, however, makes estimates less precise, so parameterizing the spillover function may be a better strategy especially in the case where there is a high proportion of control units neighboring treated units.

Removing both kinds of bias requires a researcher to take a stance and parameterize a functional form for the spillover function. The particular functional form depends on the context of the application however there are a few practical considerations. First, the researcher should argue whether spillovers on control units and on treated units exist. Second, they should argue what the maximum distance these spillovers can occur. Last, they should consider if spillovers are additive on the number of treated close treated units.

1.1. Literature Review

There are two different models of SUTVA violations in the literature. Within-group spillovers are when units are in distinct groups and outcomes depend on the treatment

status of your group only. For example, SUTVA is violated if a person's likelihood of getting sick depends on the vaccine rate of his community. In this case, the community is the group within which spillovers can occur. Estimation in this case is done in a 'partial intereference' framework.[3] This setting compares "control" units in a partially treated group with control units in completely untreated groups to estimate spillover effects.[4] However, the setting in my paper does not feature distinct groups which allow for the researcher to seperate areas into distinct non-overlapping groups.

Between-group spillovers occur when groups are overlapping which allows for more general forms of SUTVA violations. For example, a counties' economic outcomes depend on the economic outcomes of nearby counties (even across state borders). Vazquez-Bare (2019) models within-group spillovers such that their exposure mapping is only a function of unit i's group's treatment vector. He develops a potential outcomes framework which assumes that potential outcomes are a function of both own treatment-status D_i and a function of the vector of treatment assignments $h_i(\vec{D})$. The function $h_i(\vec{D})$ is referred to as an exposure mapping and adding an exposure mapping to the potential outcomes framework allows to explicitly model SUTVA violations. Using this framework, Vazquez-Bare (2019) shows that a difference in means comparison between treated and control units is a biased estimated for the treatment effect. My paper uses this potential outcome framework in the case where treatment spillover is not limited to within-group spillovers and considers estimation of treatment effects by difference-in-differences.

Sävje, Aronow, and Hudgens (2019) consider a similar potential outcome framework as Vazquez-Bare (2019), but they instead include spillovers as a part of their definition of "treatment effect". Then, the authors describe how the ATE estimated by difference-in-differences averages over both individual heterogeneity and, central to their contribution, over heterogeneity in the exposure mapping, $h_i(\vec{D})$. This paper develops a strategy to allow researchers to seperately identify treatment effects and spillovers.

The paper closest to mine is Delgado and Florax (2015) where they consider spillover

[3] Halloran and Struchiner (1995) considers community-vaccine rates in epidemology; Sobel (2006) considers interference in the Moving to Opportunity Program; and Angrist (2014) studies the context of school peer effects.

[4] Angelucci and Di Maro (2016) provides a summary for estimation of within-group treatment effects.

only on to control units and identify a bias that can come from estimating a difference-in-differences model without explicitly controlling for spillovers. Using Monte-Carlo simulations, they find that there is an omitted variable bias problem by not including a measure of spillovers. This paper derives an explicit form for this bias in terms of potential outcomes and also includes the presence of spillovers on to also treated units.

Similarly, Clarke (2017) finds an explicit form for bias when estimating a difference-in-differences model with spillover onto control units. However, their model only allows spillovers onto control units and assumes spillovers are constant regardless on the number of nearby treated units. Lastly, Berg and Streitz (2019) and Verbitsky-Savitz and Raudenbush (2012) find results for a specific potential outcome function where $h_i(\vec{D})$ is the proportion of treated contiguous counties. They allow this spillover to enter the potential outcome additively and the coefficient is allowed to differ by own treatment status. In my framework, if I assume the particular functional forms for potential outcomes of Clarke (2017), Berg and Streitz (2019), and Verbitsky-Savitz and Raudenbush (2012), I arrive at the same bias equation as theirs.

The paper is structured as follows. Section 2 presents the potential outcomes framework, defines the estimand of interest, and shows the resulting bias from estimating a classical difference-in-differences model. Section 3 presents Monte Carlo simulations to illlustrate the bias result and to evaluate currently used solutions in the literature. Section 4 presents an application for evaluating the Tennessee Valley Authority program.

2 — Potential Outcomes Framework

Following the canonical difference-in-difference framework, there is a time t_0 where treatment turns on and remains on afterwards. As in Vazquez-Bare (2019), potential outcomes for unit i at time t, denoted $Y_{i,t}(D_i, h_i(\vec{D}))$, are a function of own treatment-status D_i and, departing from the canonical framework, of a function of the entire vector of treatment assignments $h_i(\vec{D})$ where $\vec{D} \in \{0,1\}^n$. The function $h_i(\vec{D})$ is referred to as an 'exposure mapping' and is a non-negative scalar.[5] The 'exposure mapping' measures

[5] The derivation of bias does not require this assumption.

the intensity at which unit i is affected by spatial spillovers. When unit i is sufficiently 'far' away, it has no exposure to spatial spillovers and $h_i(\vec{D}) = 0$. The exposure mapping formalizes the SUTVA violation in that potential outcomes are affected by other unit's treatment assignment and can be summarized by $h_i(\vec{D})$.

To help better understand the exposure mapping function, I give three examples of $h_i(\vec{D})$ that are commonly used in the literature. First, $h_i(\vec{D})$ can be a 0/1 variable that equals one only if there is a treated unit within $\bar{d}$-miles of unit i (similarly a dummy for counties that share borders is commonly used). Let $d(i, j)$ be a geographic distance measure which tells the distance unit i is from unit j. In this case

$$h_i(\vec{D}) = \max_{i \neq j} D_j * 1[d(i, j) < \bar{d}]$$

This exposure mapping is useful when it is assumed that spillovers do not decay over distance and the number of neighboring units treated do not impact the intensity of spillovers. For example, this mapping likely applies in the context of new library creation (Berkes and Nencka, 2020). In this case the distance $\bar{d}$ would be the distance people would likely travel to a nearby library. Access to a neighboring town library likely does not depend on whether you can access 1 or more nearby libraries, so the binary value is a good approximation to the level of spillovers.

Second, $h_i(\vec{D})$ can be a function that equals the proportion of the k-nearest neighbors that are treated, i.e.

$$h_i(\vec{D}) = 1/k \sum_{j \in k(i)} D_j,$$

where $k(i)$ is the index of unit i's k-nearest neighbors. This exposure mapping is no longer binary, so intensity of spillovers depend on the number of units treated. In the context of large store openings and agglomeration economies, this would imply that as the number of nearby counties receiving new factories increase, the effect on own-county outcomes increase as well (Basker, 2005).

Last, is the spatial decay function where exposure decreases with distance. This depends on the decay parameter α that is decided by the researcher. In this case, spillover intensity is the sum across all treated observations' decay term, i.e.

$$h_i(\vec{D}) = \sum_{j \neq i} D_j e^{-\alpha d(i,j)}.$$

This exposure mapping allows for the intensity of spillovers to depend on distance to treatment and also is additive in the number of nearby units treated.[6] In the literature on R&D investment, Keller (2002) uses a modified version of this exposure mapping where D_j is country j's R&D expenditure. This allows for the technology spillovers to decay over distance to an exponential degree.

The last thing a researcher must do is specify the functional form of the potential outcomes. Typically, the spillover function enters into the regression linearly and ocasionally the coefficient is allowed to differ by treatment status.

2.1. Spatial Spillovers

With the potential outcomes defined, I now formalize what is meant by 'spatial spillovers'. I define 'spillover onto control units' as:

$$Y_i(0, h_i(\vec{D})) - Y_i(0,0).$$

The spillover measures the difference in non-treated potential outcomes between being exposed and not being exposed at level $h_i(\vec{D})$. Then, the average spillover onto control units averages over potential heterogeneity in spillovers and over heterogeneity in exposure intensity $h_i(\vec{D})$:

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

To emphasize, the average spillover onto conrol units averages over each control unit's exposure mapping. For example, assume the potential outcomes is additively linear in D_i and $h_i(\vec{D})$ and that the coefficient $\beta_{\text{spill,control}}$ measures the effect of $h_i(\vec{D})$ on outcome Y among control units. Then an individual control unit's spillover effect is $\beta_{\text{spill,control}}\, h_i(\vec{D})$. The average spillover onto control unit would therefore be $\tau_{\text{spill,control}} = \beta_{\text{spill,control}} * \mathbb{E}_i\left[h_i(\vec{D})\right]$, i.e. the average over all control units exposure mapping.

Similarly, we define the average spillover onto also treated units as:

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

[6] However, this specification assumes that all units are affected by all other units. This creates problems with inference because it implies potential correlation between all units' error terms. For this reason, this function often is summed over only the k-nearest neighbors.

It is important to clarify what I am assuming is the estimand of interest researchers would like to estimate when using difference-in-differences. I assume that what the 'Average Treatment Effect' is trying to measure in this context is what I will call the 'direct effect of treatment':

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

which measures the effect of being treated in the absence of exposure to spillovers. This differs from Sävje, Aronow, and Hudgens (2019) where they define the Average Treatment Effect as

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

where the expectation is over individuals and their exposures. I prefer the former because it allows for seperate identification of the direct effect of treatment and the spillover effects themselves.[7]

2.2. Bias in Difference-in-Differences Estimation

In this section, I identify the two sources of bias in difference-in-differences estimation. For exposition, I will refer to one pre- and post-period, $t = 0$ and $t = 1$, but this can be replaced by averages across Y in the pre-period and the post-period respectively. To estimate the 'direct effect of treatment', researchers estimate the canonical two-way fixed effects model,

$$y_{it} = \tau D_{it} + \mu_i + \mu_t + \epsilon_{it}. \tag{1}$$

The estimand $\hat{\tau}$ is a biased estimate for τ_{direct}. To show this, I first present the equivalent to the parallel counterfactual trends assumption in the context of the new potential outcome framework.

Assumption 1 (Parallel Counterfactual Trends)

$$\mathbb{E}\left[Y_{i1}(0,0) - Y_{i0}(0,0) \mid D_i = 1\right] = \mathbb{E}\left[Y_{i1}(0,0) - Y_{i0}(0,0) \mid D_i = 0\right]$$

[7] The spillover effects themselves might be of interest to the researcher, so clearly seperating them is beneficial in estimation. Later they can be combined to estimate the 'net effects' that Sävje, Aronow, and Hudgens (2019) estimate.

This assumption states that in the absence of treatment and with zero exposure (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. This generalizes to the classic parallel counterfactual trends when SUTVA is satisfied because then every unit has zero exposure.

When researchers run the canonical difference-in-differences regression, the estimand $\hat{\tau}$ will be a biased estimate for τ_{direct}. Given that Assumption 1 holds, the estimate can be decomposed as the direct effect and the two sources of spillover bias. The proof is given in Appendix A.

Theorem 1 (Bias from Difference-in-Differences Estimation)
If Assumption 1 holds, the expectation of the estimand $\hat{\tau}$ from Equation 1 is

$$\begin{aligned}
\mathbb{E}[\hat{\tau}] &= \underbrace{\mathbb{E}\left[Y_{i1} - Y_{i0} \mid D_i = 1\right] - \mathbb{E}\left[Y_{i1} - Y_{i0} \mid D_i = 0\right]}_{\textit{Difference-in-Differences}} \\
&= \mathbb{E}\left[Y_{i1}(1,\vec{0}) - Y_{i1}(0,\vec{0}) \mid D_i = 1\right] + \mathbb{E}\left[Y_{i1}(1, h_i(\vec{D})) - Y_{i1}(1,\vec{0}) \mid D_i = 1\right] \\
&\quad - \mathbb{E}\left[Y_{i1}(0, h_i(\vec{D})) - Y_{i1}(0,\vec{0}) \mid D_i = 0\right] \\
&= \tau_{direct} + \tau_{spill,treated} - \tau_{spill,control}
\end{aligned}$$

The intuition behind the biases are as follows. First, the change in outcomes among treated units combines the direct effect and the spillover from nearby treated units. Therefore the first difference adds the average spillover onto the treated units, $\tau_{\text{spill,treated}}$. Second, the change in outcomes among control units combines the parallel counterfactual trend with the average spillover onto control units. Since $\hat{\tau}$ is found by subtracting this second difference, we subtract the average spillover onto the control, $\tau_{\text{spill,control}}$.

2.3. Bounding the Bias

The previous section finds an explicit form of the bias when estimating Equation 1. The explicit form of bias can be used by researchers to provide bounds on the bias in a manner simmilar to Rambachan and Roth (2020). To provide bounds on the bias, researchers must identify the set $\Delta = [\underline{\Delta}, \overline{\Delta}]$ such that it contains the bias

$$\tau_{\text{spill,treated}} - \tau_{\text{spill, control}} \in \Delta$$

Choosing this set requires considerations of the economic context. It is guided by choices of parameterization of the exposure mapping and potential outcomes. To exemplify this bias, I will show a simple example that allows for a maximum bias bound.

2.3.1. Example Parameterization of Bias

To exemplify the above result, I will derive the bias for a particular parameterization of the exposure mapping and the potential outcomes. For the exposure mapping, I will assume that spillovers occur for any county that is within $\bar{d}$ miles of a treated county, i.e.

$$\text{Near}_{it} \equiv h_i(\vec{D}) = \max_{i \neq j} D_j * 1[d(i,j) < \bar{d}]. \tag{2}$$

This specification is a simple one in that the spillover is not additive in the number of nearby treated units and does not decay over distance. The potential outcome is assumed to be additive in both treatment and in the exposure mapping, but the coefficient on Near_{it} is allowed to differ between treated and non-treated unitts:

$$\begin{aligned} y_{it} = \mu_t + \mu_i + \beta_{\text{direct}} D_{it} + \beta_{\text{spill,control}}(1 - D_{it})\text{Near}_{it} \\ + \beta_{\text{spill,treated}} D_{it}\text{Near}_{it} + \varepsilon_{it}, \end{aligned} \tag{3}$$

where μ_t and μ_i are unit and time fixed effects respectively. In this case, the spillover on an individual control unit is given by $\tau_{\text{spill,control}}$ and on an treated unit is $\tau_{\text{spill,treated}}$.

If a researcher estimates Equation 1, then the bias will be $\tau_{\text{spill,treated}} - \tau_{\text{spill,control}}$ where the two bias terms are

$$\tau_{\text{spill,control}} = \beta_{\text{spill,control}} \frac{\sum_{i:D_{it}=0} \text{Near}_{it}}{N_C}$$

and

$$\tau_{\text{spill,treated}} = \beta_{\text{spill,treated}} \frac{\sum_{i:D_{it}=1} \text{Near}_{it}}{N_T},$$

where N_C and N_T are the number of control and treated units respectively. Simply put, the bias will be the product of the proportion of treated units that receive spillover effects and the size of the effect on treated units minus the proportion of control units that receive spillover effects time the size of the effect on control units.

Then researchers wanting to create the bias set Δ, will assume minimum and maximum values of $\beta_{\text{spill,control}}$, $\beta_{\text{spill, treated}}$, and a maximum proportion of exposed treated and

control units. With these assumptions, the set Δ can be approximated and the partially identified set is given by $[\hat{\tau} + \underline{\Delta}, \hat{\tau} + \overline{\Delta}]$. Inference on this set can be done in the manner described in Rambachan and Roth (2020).

2.4. Improving Estimation

The previous section shows how to estimate the magnitude of potential bias in the estimation of Equation 1. However, parameterization of the potential outcomes and exposure mapping allows for unbiased estimates of τ_{direct} by controlling for the exposure mapping directly. This is recommended for two reasons.

First, controlling for exposure mapping allows for unbiased estimates of the direct effect of treatmnet if the potential outcomes are correctly specified. In the case where potential outcomes are 'somewhat' incorrectly specified, this can remove some of the bias in estimation. Simulations below highlight that since spillovers are likely to decay over distance, controlling for immediate neighbors removes a large portion of the average spillover effects.

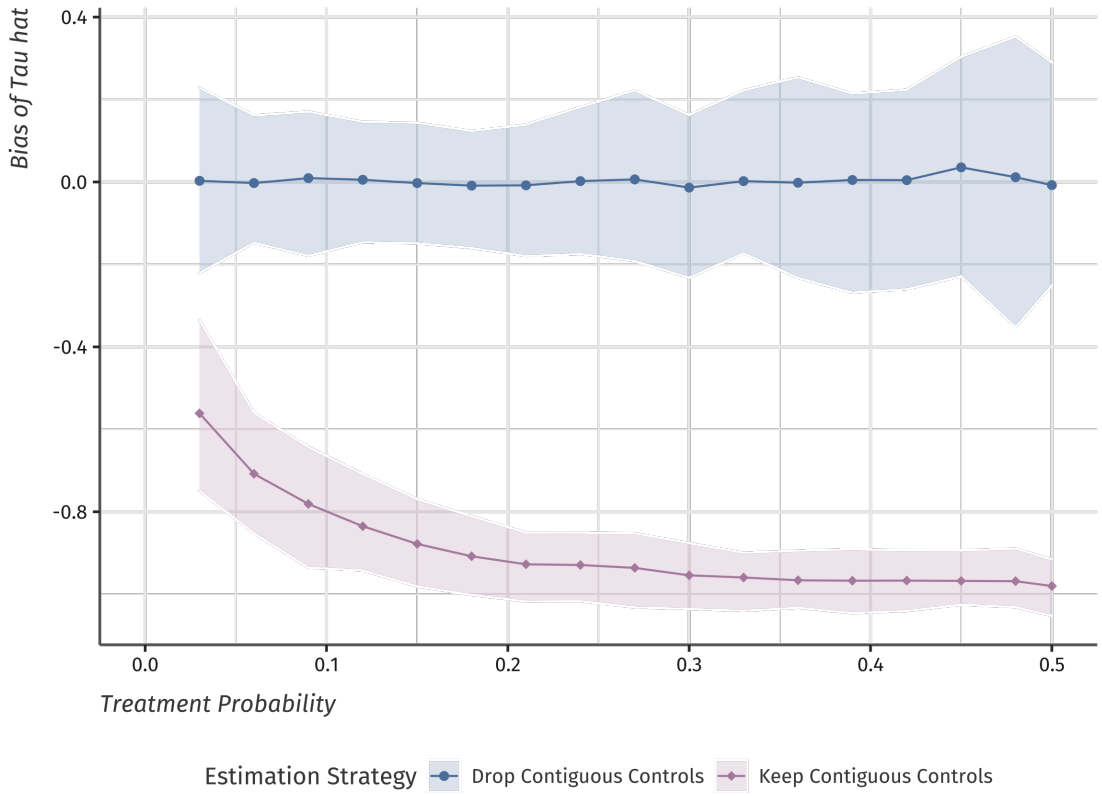
Second, the spillover effects are highly relevant for policy makers. If treatment for a given area is positive and there are positive benefits to neighboring regions, the estimated benefits will understate the true effect of treatment. First, it will not include the neighboring units' benefit in the calculation. Second, since the treatment effect will subtract the neighboring units' benefit, the neighboring units' beneift will be *double undercounted*. Similarly, the net benefits for a treated location depend on the direct effect and the spillover effect onto treated units. Therefore whether the benefits depend on the number of treated units nearby.

3 — Monte Carlo Simulations

3.1. Simulation 1

I now turn to a series of Monte Carlo Simulations to highlight the importance for controlling for spillovers in simulations. The data generating process used is defined by the exposure mapping given in Equation 2 and the data generating process given in Equa-

Figure 1 — Bias of $\hat{\tau}$ with and without Control Units with Spillovers



Notes: This figure plots the bias of $\hat{\tau}$ found from estimating Equation 1 for data generated as described in the text and Equation 4. Each point corresponds to the average bias for the given treatment probability and the band is the 95 percent empirical confidence interval over 1000 simulations. The line with diamond markers estimates with all control units. The line with circle markers removes control units that share a border with a treated county.

tion 3. A unit of observation is a US county and the periods are $t \in \{1, \ldots, 20\}$ with treatment turning on in period $t = 10$. The intercepet is $\alpha = -2$, the fixed effects are $\mu_t \sim N(t * 0.2, 0.1^2)$ and $\mu_i \sim N(6, 2^2)$, and the error term is $\epsilon \sim N(0, 2^2)$. Last, the size of treatment and spillovers are as follows: $\beta_{\text{direct}} = 2$, $\beta_{\text{spill, control}} = 1$ and $\beta_{\text{spill, treat}} = 0$.

For the first simulation, I restrict spillovers only to occur among control units, i.e. $\tau_{\text{spill, treat}} = 0$ and estimate the typical two-way fixed effects model. This data generating

process where spillovers only occur onto control units matches what has been typically assumed in the literature. I assign treatment among U.S. counties randomly with various probabilities between 3 percent and 50 percent. The DGP is therefore

$$y_{it} = -2 + \mu_t + \mu_i + 2D_{it} - (1 - D_{it})\text{Near}_{it} + \varepsilon_{it} \tag{4}$$

The size of the bias from estimating Equation 1 at different treatment probabilities are presented in Figure 1 as the line with diamond markers. As displayed in the figure, even for a low treatment probability of three percent, the bias is quite large with a 95 percent empirical confidence interval between -0.28 and -0.75. As treatment frequency increases, the bias increases as well but at a slower rate due to fewer additional control units receiving spillover units.

A common solution in the literature is to remove control units from the estimated sample that are most likely to be affected by spillovers. I do this in Figure 1 and the results are shown by the line with circle markers. Even though, I remove contiguous controls which is not hte correct measure of $h_i(\vec{D}$, it approximates it well enough such that the bias stays centered constantly around zero as most control units experiencing spillovers are removed. As the portion of control units receiving spillovers that are kept in the sapmle increases, the bias would fall between the two lines.

There is a trade-off between the bias and the variance of the estimator when using this methodolgy. As the treatment probability increases, the number of control units removed increases as well. This naturally yields a more variable estimator as seen in the wider 95 percent empirical confidence intervals.

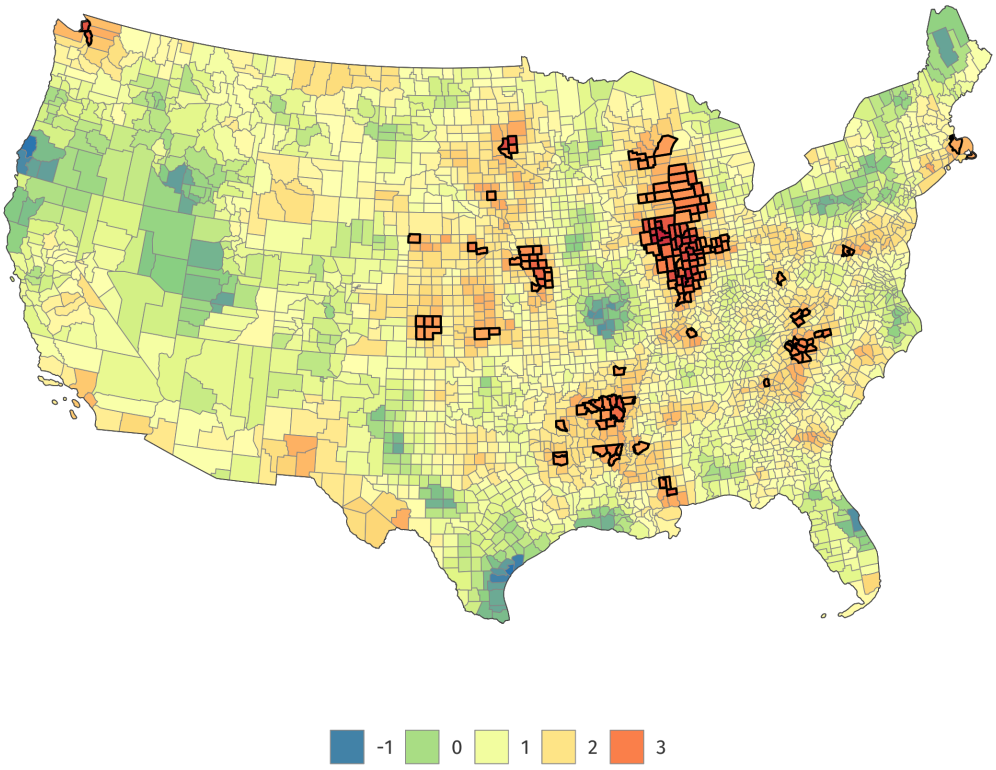
3.2. Simulation 2

For the following simulation, I will add in spillovers onto also treated units with $\beta_{\text{spill, treat}} = -0.5$. The DGP is as follows

$$y_{it} = -2 + \mu_t + \mu_i + 2D_{it} - (1 - D_{it})\text{Near}_{it} - 0.5D_{it}\text{Near}_{it} + \varepsilon_{it} \tag{5}$$

The proportion of treated units affected by the spillover depends on the spatial autocorrelation of the treatment assignment. In empirical applications, it is often the case that

Figure 2 — Example of Kriging Field

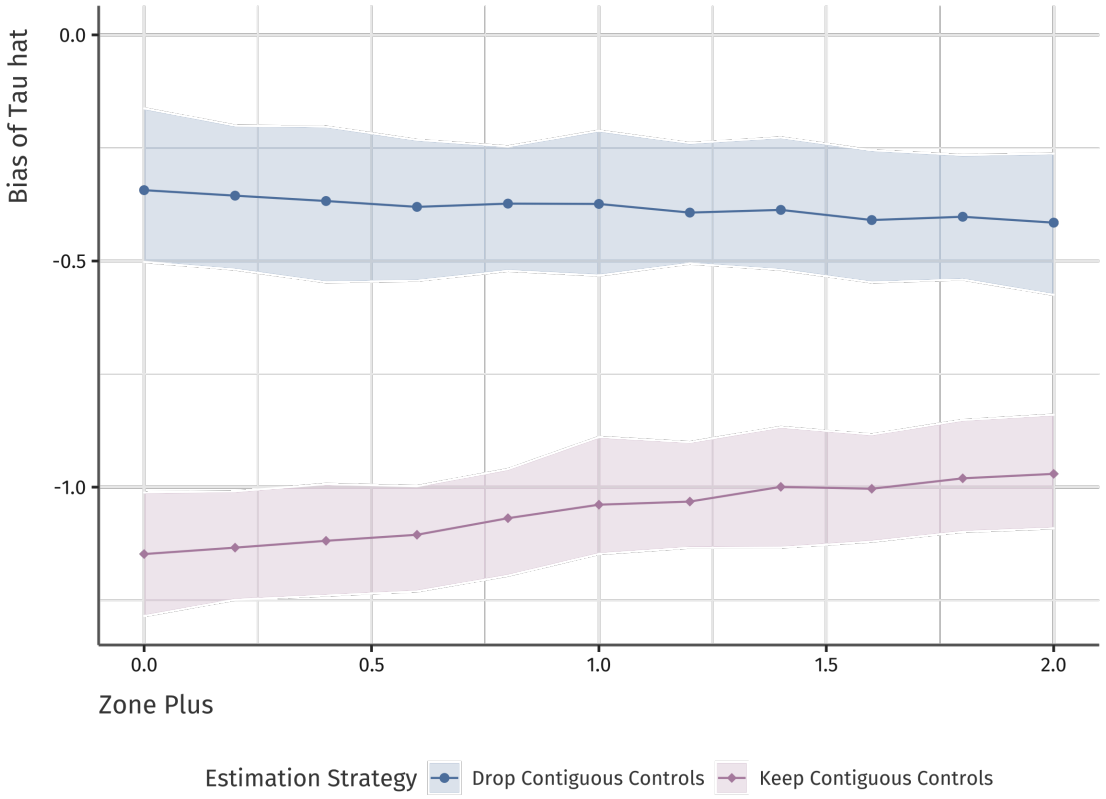


Notes: This figure plots an example of a kriging simulation using the 'gstat' package in R using an exponential variogram. The outlined counties are in the top ten percent of values for this simulation.

treatment assignment is concentrated in specific areas which yields higher proportion of treated units receiving spillover effects.

In order to model spatial autocorrelation of treatment, I turn to a method from geosciences called 'kriging'. Kriging generates a Gaussian field for a large grid of points across the entire United States where the spatial autocorrelation of the points is described by a set of parameters. One such field is depicted in Figure 2. For a given field, I find the counties in the top ten percent of values and give these counties a value of $\text{Zone}_i = 1$. Then I assign treatment with a probability that differs based on Zone. The unconditional probability is equal to 5 percent in all simulations. To do this, I use a variable 'Zone Plus'

Figure 3 — Bias of $\hat{\tau}$ at Different Levels of Spatial Autocorrelation



Notes: This figure plots the bias of $\hat{\tau}$ found from estimating Equation 1 for data generated as described in the text and Equation 5. Treatment probability is assigned according to the method described by the text and Equation 6. Each point corresponds to the average bias for the given Zone Plus and the band is the 95 percent empirical confidence interval over 1000 simulations. The line with diamond markers estimates with all control units. The line with circle markers removes control units that share a border with a treated county.

to adjust the relative probability of receiving treatment:

$$P(D_i \mid \text{Zone}_i) = (.1 + \text{Zone Plus} * \text{Zone}_i)\frac{.05}{.1 * .9 + (.1 + \text{Zone Plus}) * .1} \tag{6}$$

The second term normalizes probabilities so the unconditional probability stays at 5 percent. In our simulations, Zone Plus ranges from 0 which is the case where p = 10 percent for all counties to 2 which has $P(D_i \mid \text{Zone}_i) = 0.166 + 0.35 * \text{Zone}_i$

The results for the bias of $\hat{\tau}$ are in Figure 3. The line with diamond markers is run on the

full sample. As the level of spatial autocorrelation increases, the bias actually decreases. The reason for this is that as treatment becomes more concentrated in the 'Zones', the number of control units receiving spillovers decreases while the number of treated units receiving spillovers increases. In this particular DGP, the effect of fewer control units on the bias is larger than the bias from more treated units.

The line with circle markers in Figure 3 repeats the exercise of dropping control units that share a border with treated units. In the case where spillovers occur on treated units, all bias is not removed when dropping control units. That is because we only remove $\tau_{\text{spill,control}}$ but $\tau_{\text{spill,treated}}$ remains. To remove the second source of bias, it is necessary to acount directly for spillovers in the estimation strategy. I turn to this next.

References

Angelucci, M. and V. Di Maro (Jan. 2016). "Programme evaluation and spillover effects". In: *Journal of Development Effectiveness* 8.1, pp. 22–43. ISSN: 1943-9342, 1943-9407. DOI: 10.1080/19439342.2015.1033441.

Angrist, Joshua D. (Oct. 2014). "The perils of peer effects". In: *Labour Economics* 30, pp. 98–108. ISSN: 09275371. DOI: 10.1016/j.labeco.2014.05.008.

Basker, Emek (Feb. 2005). "Job Creation or Destruction? Labor Market Effects of Wal-Mart". In: *The Review of Economics and Statistics* 87.1, pp. 174–183.

Berg, Tobias and Daniel Streitz (2019). *Handling Spillover Effects in Empirical Research*. Working Paper, p. 59.

Berkes, Enrico and Peter Nencka (2020). *Knowledge Access: The Effects of Carnegie Libraries on Innovation*.

Clarke, Damian (2017). "Estimating Difference-in-Differences in the Presence of Spillovers". In: *Munich Personal RePEc Archive*, p. 52.

Delgado, Michael S. and Raymond J.G.M. Florax (Dec. 2015). "Difference-in-differences techniques for spatial data: Local autocorrelation and spatial interaction". In: *Economics Letters* 137, pp. 123–126. ISSN: 01651765. DOI: 10.1016/j.econlet.2015.10.035.

Halloran, M. Elizabeth and Claudio J. Struchiner (Mar. 1995). "Causal Inference in Infectious Diseases:" in: *Epidemiology* 6.2, pp. 142–151. ISSN: 1044-3983. DOI: 10.1097/00001648-199503000-00010.

Keller, Wolfgang (2002). "Geographic Localization of International Technology Diffusion". In: *The American Economic Review* 92.1, p. 52.

Rambachan, Ashesh and Jonathan Roth (2020). *An Honest Approach to Parallel Trends*. Working Paper.

Rubin, Donald B. (Sept. 1980). "Randomization Analysis of Experimental Data: The Fisher Randomization Test Comment". In: *Journal of the American Statistical Association* 75.371, p. 591. ISSN: 01621459. DOI: 10.2307/2287653.

Sävje, Fredrik, Peter M. Aronow, and Michael G. Hudgens (Oct. 2019). "Average treatment effects in the presence of unknown interference". In: *arXiv:1711.06399 [math, stat]*. arXiv: 1711.06399. URL: http://arxiv.org/abs/1711.06399.

Sobel, Michael E (Dec. 2006). "What Do Randomized Studies of Housing Mobility Demonstrate?: Causal Inference in the Face of Interference". In: *Journal of the American Statistical Association* 101.476, pp. 1398–1407. ISSN: 0162-1459, 1537-274X. DOI: 10.1198/016214506000000636.

Vazquez-Bare, Gonzalo (June 2019). "Identification and Estimation of Spillover Effects in Randomized Experiments". In: *arXiv:1711.02745 [econ]*. arXiv: 1711.02745. URL: http://arxiv.org/abs/1711.02745.

Verbitsky-Savitz, Natalya and Stephen W. Raudenbush (Jan. 2012). "Causal Inference Under Interference in Spatial Settings: A Case Study Evaluating Community Policing Program in Chicago". In: *Epidemiologic Methods* 1.1. ISSN: 2161-962X. DOI: 10.1515/2161-962X.1020. URL: https://www.degruyter.com/view/j/em.2012.1.issue-1/2161-962X.1020/2161-962X.1020.xml.

A — Proofs

Proof of Theorem 1

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