Differences-in-Differences with Spatial Spillover

Kyle Butts

October 14, 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, differences-in-differences produces biased estimates for the average treatment effect. In this paper, I identify two sources of bias. 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 show when the biases are particularly large. Further, I show that 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 economic often considers treatment assigned by geographic boundaries such as cities, counties, and states. The effects of treatment does not typically stay within these boundaries whether it be from people crossing borders or from general equilibrium mechanisms affecting neighboring areas. In causal inference, 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 a unit's outcome.[1]

Despite the problem of spatial spillovers being common across many settings, Table 1 of Berg and Streitz (2019) document that little empirical analysis takes the problem seriously. In a survey of 8 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 8 include spillovers 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 differences-in-differences. 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 differences-in-differences, the spillover onto the 'close' control units is averaged into the untreated change in outcomes. Therefore the spillover is subtracted from the estimated treatment effect and the sign of the bias is the opposite sign of the spillover. For example, if a factory opening in a given county benefits both that

---

[1] Close can refer to many different things, e.g. geographic distance, node distance in graph theory, 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.

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 in the counterfactual.

Second, changes in treated units' outcomes reflect the effect of their own treatment status and the effect from the treatment status of "close" units. Therefore 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 a researcher expects that spillovers are quite large in magnitude or spread far over distance, then the bias will generally be large. My paper provides an explicit form for this bias and therefore researchers can give bounds for the bias under assumptions about the upper bound of the spillover effect and on number of units affected. In the paper, I use Monte Carlo simulations in order to quantify the magnitude of bias in various spillover settings.

In Berg and Streitz (2019), they find that 6 of the 8 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.

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. Between-group spillovers is this case where groups are overlapping which allows for more general forms of spillover. For example, a counties' economic outcomes depend on the economic outcomes of nearby counties (even across state borders).

One set of literature discusses 'partial intereference' where spillovers are assumed to only occur within a group and not across groups.[3] This setting compares "control" units in a partially treated group with contrl 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.

The potential outcome framework used in my paper extends the model in Vazquez-Bare (2019) which assumes that potential outcomes are a function of both own treatment-status D_i and a function of the vector of treatment assignments $h_i(\vec{D})$ (this is referred to as an exposure mapping). Adding an exposure mapping to the potential outcomes framework allows to explicitly model SUTVA violations. In Vazquez-Bare (2019), he models within-group spillovers such that their exposure mapping is only a function of unit i's group's treatment vector. Using this framework, he shows that a difference in means comparison between treated and control units is a biased estimated for the treatment effect. My paper considers the case where treatment spillover is not limited to within-group spillovers and considers estimation of treatment effects by differences-in-differences.

Sävje, Aronow, and Hudgens (2019) consider a similar potential outcome frame-

---

[3]Halloran and Struchiner (1995) considers community-vaccine rates in epidemiology; 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.

4

work 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 diff-in-diff averages over both individual heterogeneity and, central to their contribution, over heterogeneity in the exposure mapping, $h_i(\vec{D})$. This paper does not include spillovers in the estimation of ATE because researchers often wish to understand the effects seperately.

The paper closest to mine is Delgado and Florax (2015) where they consider spillover only on to control units and identify a bias that can come from estimating a diff-in-diff 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. My 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.

A similar paper, Clarke (2017), finds an explicit form for bias when estimating a diff-in-diff model with spillover onto control units. However, they model a specific kind of spillover that only allows spillovers onto control units and assumes spillovers are non-additive. 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. ...

2. Potential Outcomes Framework

In this paper, I am assuming there is a time t_0 where treatment turns on for all treated units and remains on afterwards. The potential outcome framework used here extends Vazquez-Bare (2019) model which defines potential outcomes for unit i at time t as a function of own treatment-status D_i and of a function of the vector of treatment assignments $h_i(\vec{D})$ where $\vec{D} \in \{0,1\}^n$. This function $h_i(\vec{D})$ is referred to as an 'exposure mapping' and in this paper will be assumed to be a non-negative scalar.[5] The function $h_i(\vec{D})$ measures the exposure to spatial spillover and no exposure to spatial spillovers is when $h_i(\vec{D}) = 0$.

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

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

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. Last, is the spatial decay function which depends on the decay parameter α. In this case, spillover intensity is the sum of across all treated observations' decay term, i.e.

$$h_i(\vec{D}) = \sum_{j \neq i} D_j \exp^{-\alpha d_i(j)}$$

The exposure mapping is part of the parameterization of spillovers, but the functional form of the potential outcome still would need to be specified.

---

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

2.1 Spatial Spillovers

With a new definition of potential outcomes, let me formalize what is meant by "spatial spillovers". I will define 'spillover onto control units' as:

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

Then, we will take the average spillover onto control units which averages over potential heterogeneity in spillovers and over heterogeneity in exposure intensity $h_i(\vec{D})$:

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

These spillover effects depend on the average of exposure mapping. For example, assuming the potential outcomes is additively linear in D_i and $h_i(\vec{D})$, then an individual control unit's spillover effect is $\beta_{\text{spillover, control}} \, h_i(\vec{D})$. The average spillover onto control unit would therefore be $\tau_{\text{spillover, control}} = \beta_{\text{spillover, control}} * \mathbb{E}_i \left[h_i(\vec{D}) \right]$, i.e. the average over all control units exposure mapping.

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

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

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

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

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

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

where the expectation is over individuals and their exposures. I prefer the former because it allows for seperation between the direct effect of treatment and the spillover effects themselves.[6]

2.2 Bias in Differences-in-Difference Estimation

In this section, I identify the two sources of bias in differences-in-differences estimation. For exposition, I will refer to two periods, $t = 0$ and $t = 1$, but this can be replaced by averages across Y in the pre-period and the post-period respectively. First, I present the equivalent to the parallel counterfactual trends assumption in the context of the new potential outcome framework.

Assumption 1 (Parallel Counterfactual Trends).

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

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 typical Parallel Counterfactual Trends because typically SUTVA assumes that every unit has zero exposure.

When researchers run the canonical differences-in-differences regression

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

$\hat{\tau}$ 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.

Theorem 1. *If Assumption 1 holds,*

$$\mathbb{E}[\hat{\tau}] = \mathbb{E}\left[Y_{i1} - Y_{i0} \mid D_i = 1\right] - \mathbb{E}\left[Y_{i1} - Y_{i0} \mid D_i = 0\right]$$

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

---

[6]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.

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. Second, the change in outcomes among control units combines the parallel counterfactual trend with the average spillover onto control units. Since τhat is found by subtracting this second difference, we subtract the average spillover onto the control.

2.3 Example Parameterization of Bias

3.   Monte Carlo Simulations

I now turn to a series of Monte Carlo Simulations to highlight the importance for controlling for spillovers in simulations.

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.

Berg, Tobias and Daniel Streitz (2019). "Handling Spillover Effects in Empirical Research". In: p. 59.

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.

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

$\mathbb{E}_i\left[\hat{\tau}\right] = \mathbb{E}_i\left[Y_{i1} - Y_{i0} \mid D_i = 1\right] - \mathbb{E}_i\left[Y_{i1} - Y_{i0} \mid D_i = 0\right]$

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

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

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

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

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

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

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

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

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

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

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

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