

Difference-in-Differences with Geocoded Microdata

Kyle Butts*

JULY 8, 2021

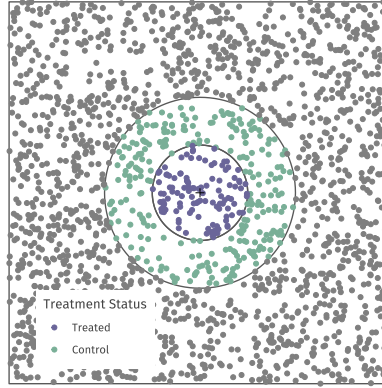
...

JEL-Classification:

Keywords:

*University of Colorado, Boulder. Email: buttskyle96@gmail.com.

Figure 1 — Rings Method



1 — Introduction

In recent years, there has been a surge of precisely geocoded microdata that has allowed researchers to answer questions about the effects of ‘treatment’ at a very granular level. When treatment is located at a specific point in space, a very common method of evaluating the effects of the treatment is to compare units within an inner-ring of a certain distance to the treatment with units just a bit further in an outer-ring. This identification strategy is illustrated in [Figure 1](#).

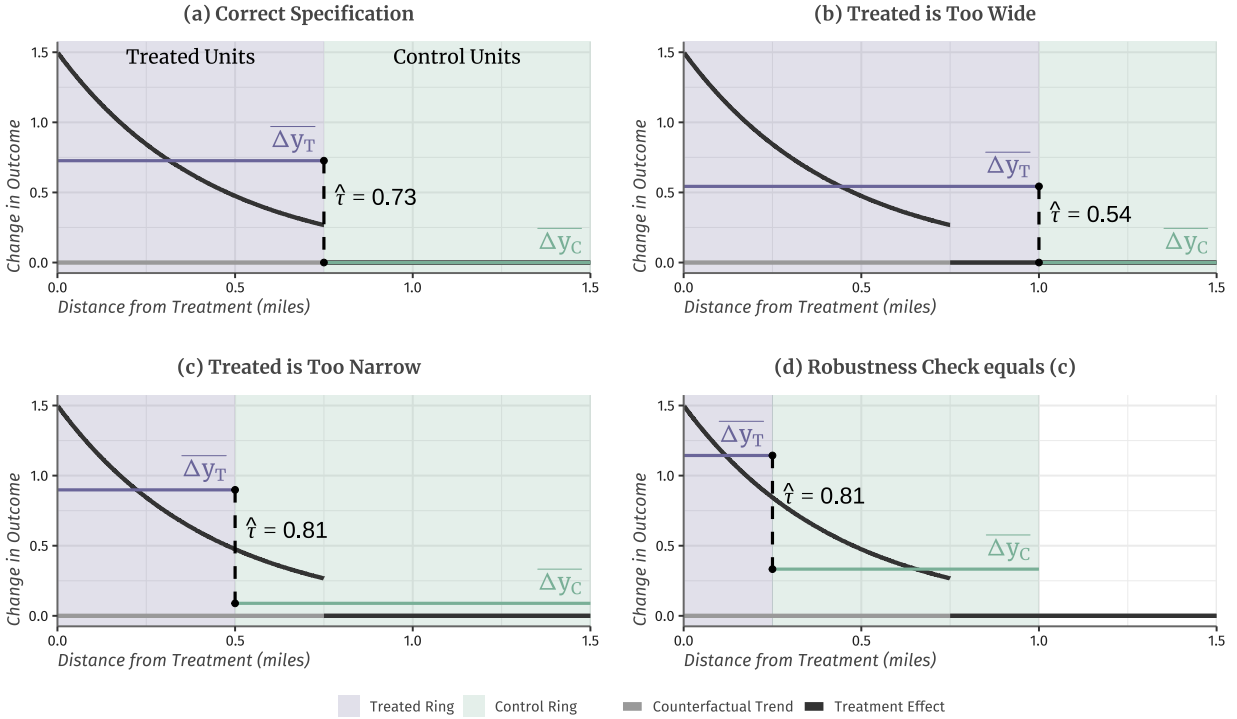
The appeal of this identification strategy, which I will label the ‘Ring method’, is that under the assumption that unobservable variables change smoothly over space, then the counterfactual outcomes will approximately be equal between both rings.

2 — Example of Problem

To illustrate the methodological difficulties in this method, I present an illustrative example. Suppose that an overgrown empty lot in a high-poverty neighborhood is cleaned up by the city and the outcome of interest is home prices. The researcher observes a panel of home sales before and after the lot is cleaned. Cleaning up the lot causes home values to go up directly nearby and as you move away from the lot, the positive treatment effect will decay to zero effect at, say, 3/4 of a mile.

[Figure 2](#) shows a plot of simulated data from this example. The black line is treatment effect at different distances from the empty lot and the grey line is the trend in home values at different distances from the empty lot. It is typically assumed that parallel trends hold at such a small radius from a location, so counterfactual trends are assumed to be constant in this example. Panel (a) of [Figure 2](#) shows the best-case

Figure 2 — Example of Problems with Ad-Hoc Ring Selection



scenario where the treated ring is correctly specified. The treatment effect estimate is the difference in the average change in outcome among the treated units and the control units. However, this singular number masks over a large amount of treatment effect heterogeneity with units very close to treatment having a treatment effect double that of $\hat{\tau}$ and units near 3/4 miles experience a treatment effect half as large as $\hat{\tau}$. For this reason, even if a researcher identifies the correct average treatment effect, they are masking a lot of heterogeneity that is potentially interesting. Therefore, I recommend non-parametrically estimating the treatment effect as a function of distance rather than using a single indicator.

However, the researcher does not typically know the distance at which treatment effects stop. Panels (b) and (c) highlights how treatment effect estimates change with a change in ring distances. Panel (b) shows when the 'treatment' ring is too wide. In this case, some of the units in the treatment ring receive no effect from treatment and therefore makes the average treatment effect among units in the treatment ring smaller. While this would be an accurate measure of the "average" treatment effect among the units in the treated ring, the estimated effect is not a measure of the average treatment effect among those actually affected by treatment that researchers aim to identify.

Panel (c) of Figure 2 shows the opposite case, where the treated ring is too narrow. In this case, there are

some units in the ‘control’ ring that experience treatment effects. Hence, the average change in outcome among the control unit is too large. This does not, though, decrease the treatment effect as you may suspect. Since the treatment effect decays with distance, the average change in outcome among the more narrow ring is larger than the correct specification. The estimated treatment effect in this case grows.

Often times, researchers try multiple sets of rings and if the estimated effect remains similar across specifications, they assume the results are ‘robust’. Panel (d) of [Figure 2](#) shows an example of why this is a problem. If Panel (c) was the researchers’ original specification and Panel (d) was run as a robustness check, then the researcher would be quite confident in their results even though the estimate is too large in both cases.

3 — Theory

A researcher observes panel data of units i at times $t = 0, 1$ located in space at point $\theta_i = (x_i, y_i)$. Treatment occurs at a location $\bar{\theta} = (\bar{x}, \bar{y})$ between periods. Therefore, units differ in their distance to treatment, defined by $\text{Dist}_i \equiv \sqrt{(x_i - \bar{x})^2 + (y_i - \bar{y})^2}$. Since treatment location is often chosen strategically, potential outcomes will have to reflect the fact that trends can change with distance to treatment. Outcomes are given by

$$Y_{it} = \tau(\text{Dist}_i)\mathbf{1}_{t=1} + \mu_i + \lambda(\text{Dist}_i)\mathbf{1}_{t=1} + \varepsilon_{it},$$

where $\tau(\text{Dist})$ is the treatment effect curve at different distances, μ_i is unit-specific time-invariant factors, and $\lambda(\text{Dist})$ is the counterfactual trend at different distances from treatment. Since we include the λ term, without loss of generality we assume that the error term ε_{it} is uncorrelated with distance to treatment. Researchers are trying to identify the average treatment effect on units experiencing treatment effects, i.e. $\bar{\tau} = \mathbb{E}[\tau(\text{Dist}) \mid \tau(\text{Dist}) > 0]$ where expectations are with respect to the distribution of distances.

Taking first differences, we see that $\Delta Y_{it} = \tau(\text{Dist}_i) + \lambda(\text{Dist}_i) + \Delta \varepsilon_{it}$. It is clear that $\tau(\text{Dist}_i)$ and $\lambda(\text{Dist}_i)$ are not separately identified unless additional assumptions are imposed. Often times researchers claim that counterfactual trends likely evolve smoothly over distance, so that $\lambda(\text{Dist}_i)$ is approximately constant within a certain distance from treatment. This is formalized in the context of our outcome model by the following assumption.

Assumption 1 (Local Parallel Trends). *For a distance \bar{d} , we say that local parallel trends hold if for all positive $d, d' \leq \bar{d}$, then $\lambda(d) = \lambda(d')$.*

This assumption requires that, in the absence of treatment, outcomes would evolve the same at every distance from treatment within a certain maximum distance. The last assumption requires that researchers correctly specify d_t to be the maximum distance that experience treatment effects.

Assumption 2 (Correct d_t). A distance d_t satisfies this assumption if for all $d \leq d_t$, $\tau(d) > 0$ and for all $d > d_t$, $\tau(d) = 0$.

The ‘ring method’ is the following procedure. Researchers select a pair of distances $d_t < d_c$ which define the “treated” and “control” groups. These groups are defined by $\mathcal{D}_t \equiv \{i : 0 < \text{Dist}_i \leq d_t\}$ and $\mathcal{D}_c \equiv \{i : d_t < \text{Dist}_i \leq d_c\}$. On the subsample of observations defined by $\mathcal{D} \equiv \mathcal{D}_t \cup \mathcal{D}_c$, they run the following specification:

$$\Delta Y_{it} = \beta_0 + \beta_1 \mathbf{1}_{i \in \mathcal{D}_t} + u_{it}. \quad (1)$$

From our potential outcome framework and standard regression results, we have the following proposition.¹

Proposition 1 (Decomposition of Ring Estimate).

(i) The estimate of β_1 in (1) has the following expectation:

$$\mathbb{E} [\hat{\beta}_1] = \underbrace{\mathbb{E} [\tau(\text{Dist}) | \mathcal{D}_t] - \mathbb{E} [\tau(\text{Dist}) | \mathcal{D}_c]}_{\text{Difference in Treatment Effect}} + \underbrace{\mathbb{E} [\lambda(\text{Dist}) | \mathcal{D}_t] - \mathbb{E} [\lambda(\text{Dist}) | \mathcal{D}_c]}_{\text{Difference in Trends}}.$$

(ii) More, if d_c satisfies Assumption 1, then

$$\mathbb{E} [\hat{\beta}_1] = \underbrace{\mathbb{E} [\tau(\text{Dist}) | \mathcal{D}_t] - \mathbb{E} [\tau(\text{Dist}) | \mathcal{D}_c]}_{\text{Difference in Treatment Effect}}.$$

(iii) If d_c satisfies Assumption 1 and d_t satisfies Assumption 2, then

$$\mathbb{E} [\hat{\beta}_1] = \bar{\tau}.$$

This proposition shows that the ‘treatment effect’ estimate is the sum of two differences. The first difference is the difference in average treatment effect among units in the treated ring and units in the control ring. If some units in the control group experience effects from treatment, the average of these effects will be subtracted from the estimate. The second difference is the difference in counterfactual trends between the treated and control rings. Since treatment can be targeted, the treated ring could be on a different trend than units further away. This difference in counterfactual trends is not separately identifiable from the difference in average treatment effects unless Assumption 1 is satisfied. This is because if d_c satisfies the assumption, then the counterfactual trend is absorbed completely by β_0 in Equation (1).

As discussed above, the decomposition in part (ii) of Proposition 1 is not necessarily unbiased estimate for $\bar{\tau}$. First, if d_t is too wide, then \mathcal{D}_t contain units that are not affected by treatment. In this case, first

1. A similar derivation is found in Sullivan (2017).

term will be smaller than $\bar{\tau}$ while the second term would be equal to zero since d_c satisfies Assumption 1. Therefore, $\hat{\beta}_1$ will be biased towards zero if d_t is too wide. Second, if d_t is too narrow then the \mathcal{D}_c will contain units that experience treatment effects. It's not in this case, though, if $\hat{\beta}_1$ will grow or shrink without knowledge of the $\tau(\text{Dist})$ curve. See the previous section for an example.

The last result of Proposition 1 shows that if d_t is correctly specified as the maximum distance that receives treatment effect, then $\hat{\beta}_1$ will be an unbiased estimate for the average treatment effect among the units affected by treatment. However, assumption 2 is a very demanding assumption and unlikely to be known by the researcher. The following section will improve estimation by allowing non-parametric identification of $\tau(\text{Dist})$. An estimate of $\tau(\text{Dist})$ can then be numerically integrated to for an estimate of $\bar{\tau}$.

References

Sullivan, Daniel M. 2017. *The True Cost of Air Pollution: Evidence from the Housing Market*. Working Paper.