

Difference-in-Differences with Geocoded Microdata

Kyle Butts^{*}

October 12, 2021

This paper formalizes a common approach for estimating effects of treatment at a specific location using geocoded microdata. This estimator compares units immediately next to treatment (an inner-ring) to units just slightly further away (an outer-ring). I introduce intuitive assumptions needed to identify the average treatment effect among the affected units and illustrates pitfalls that occur when these assumptions fail. Since one of these assumptions requires knowledge of exactly how far treatment effects are experienced, I propose a new method that relaxes this assumption and allows for nonparametric estimation using partitioning-based least squares developed in [Cattaneo et al. \(2019a,a\)](#). Since treatment effects typically decay/change over distance, this estimator improves analysis by estimating a *treatment effect curve* as a function of distance from treatment. This is contrast to the traditional method which, *at best*, identifies the average effect of treatment. To illustrate the advantages of this method, I show that Linden and Rockoff (2008) under estimate the effects of increased crime risk on home values closest to the treatment and overestimate how far the effects extend by selecting a treatment ring that is too wide.

JEL-Classification: C13, C14, C18

Keywords: Spatial Econometrics, Difference-in-Differences, Nonparametric Estimation

^{*}University of Colorado, Boulder. Email: kyle.butts@colorado.edu.

1 – Introduction

The rise of microdata with precisely geocoded locations has allowed researchers to begin answering questions about the effects of spatially-targeted treatments at a very granular level. What are the effects of local pollutants on child health? Does living within walking distance to a new bus stop improve labor market outcomes? How far do neighborhood shocks such as foreclosures or new construction spread? When treatment is located at a specific point in space, a standard method of evaluating the effects of the treatment is to compare units that are close to treatment to those slightly further away – what I will label the ‘ring method’. This paper fills a gap in the literature by formalizing the assumptions required for identification, highlighting potential pitfalls to the currently used estimator, and proposes an improved estimator which relaxes these assumptions.

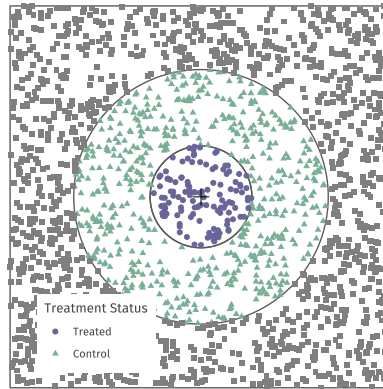
The ring method is illustrated in [Figure 1](#). The center of the figure marks an x which represents the location of treatment, e.g. a foreclosed home. Units within the inner circle, marked by dots, are considered treated due to their proximity to the treatment location; units between the inner and outer circles, marked in triangles, are considered control units; and then the remaining units are removed from the sample. The appeal of this identification strategy is that since the treated and control units are all very close in physical location, e.g. having access to the same labor market and consumptive amenities, the counterfactual untreated outcomes will approximately be equal between both rings. The ring estimate for the treatment effect compares average changes in outcomes between units in the inner ‘treated’ ring and the outer ‘control’ ring to form an estimate for the treatment effect, i.e. a difference-in-differences estimator.

My first contribution is to fill a gap in the econometrics literature by formalizing the necessary assumptions for unbiased estimates of the average treatment effect on the affected units.¹ The first assumption is the well understood parallel trends assumption for the treated and control units. This requires that the *average* change in (counterfactual) untreated outcomes in the treated ring is equal to the *average* change in the control ring. This allows the control units to estimate the counterfactual trend for the treated units.

The second assumption requires the researcher to correctly identify how far treatment effects are experienced (the inner ring). This is a very strict assumption that when not satisfied, causes biased estimates of the treatment effect. If the treated ring is too narrow, then units in the control ring experience effects of treatment and the change among ‘control’ units would no longer identify the counterfactual trend. On the other hand, if the treated ring is too wide, then the zero treatment effect of some unaffected units are averaged into the change among ‘treated’ units. Therefore results will

1. This generalizes the treatment effect on the treated in the case where treatment isn’t assigned to specific units.

Figure 1 – Rings Method



be attenuated towards zero in this case.

Since researchers often do not know how far treatment effects extend in most circumstances, I propose an estimator that replaces the second assumption with a more mild assumption by using a nonparametric partitioning-based least square estimator (Cattaneo et al., 2019a,b). My proposed methodology estimates the treatment effect curve as a function of distance by using many rings rather than trying to estimate the average treatment effect with one inner ring. This method requires that treatment effects become zero somewhere between the distance of 0 and the control ring *without* the need to specify the exact distance. It however requires a stronger assumption that the counterfactual trend is constant across distance.² This new assumption is more strict than parallel trends holding *on average*. However, researchers motivate the identification strategy by saying within a small distance from treatment that units are subject to a common set of shocks which implies the more strict assumption. While this assumption is not directly testable, the estimator creates a set of point estimates of treatment effects that can be used to visually inspect the plausibility of the assumption. If after some distance, treatment effects become centered at zero, this is suggestive evidence of common trends holding, akin to the pre-trends test in event study regressions.

The nonparametric approach allows researcher to get a more complete picture of how the intervention affects units at various distance rather than estimating an “overall effect”. For example, the construction of a new bus-stop potentially creates net costs to immediate neighbors while providing net benefits for homes slightly further away. Estimation of the treatment effect curve can illustrate these different effects that the “overall effect” would mask over. In this case, the average effect could be zero even though most units experience non-zero effects.

2. Note that the original requirement is that the *average* change in each ring is equal but allows variation across distance.

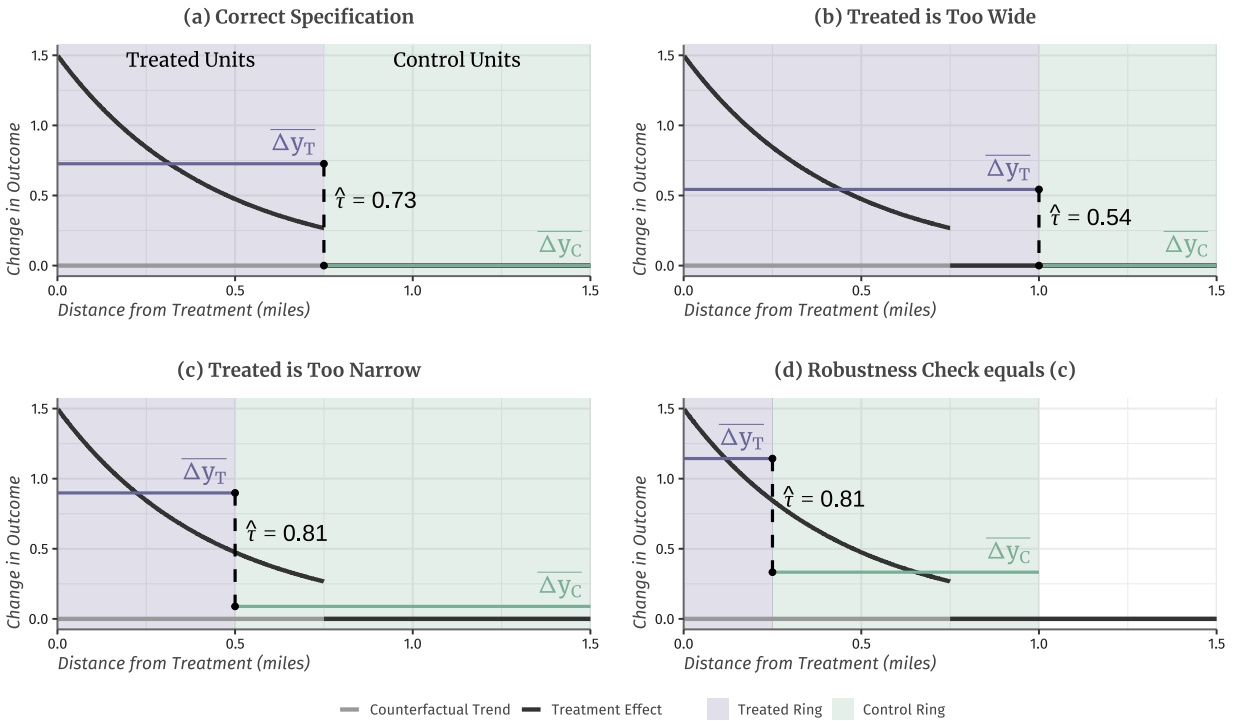
1.1. *Relation to Literature*

This paper relates to a few papers that address difficulties with using the rings method for causal effect estimation. In the online appendix, [Gerardi et al. \(2015\)](#) discuss the problem that if the treated ring is defined too narrowly, then control units will be affected by treatment causing a biased estimate of the counterfactual trend. [Sullivan \(2017\)](#) discuss the problem more formally and derive that the bias will be the difference in treatment effects experienced by the ‘treated’ ring and the ‘control’ ring. My paper nests the results of [Sullivan \(2017\)](#) by including the additional source of bias that can result from a violation of parallel trends. Other researchers have recognized that estimating a single average treatment effect is less informative than a treatment effect curve and use multiple rings to estimate treatment effects at different distances (e.g. [Alexander et al. \(2019\)](#)). However, this approach selects multiple rings in an ad-hoc manner, still requires treatment effects become zero after the outer-most treatment ring, and is prone to problems of specification searching. This paper’s proposed estimator selects the number and location of rings in a data-driven way and does not require correct specification of where treatment effects become zero.

[Diamond and McQuade \(2019\)](#) propose a nonparametric estimator aimed at estimating a treatment effect surface. They use two-dimensions (latitude/longitude) to better approximate a smooth change in counterfactual outcomes (e.g. north-west and south-east from treatment might have different treatment effects). My method, instead, uses a singular measure of distance which pools units at similar distance but different direction from treatment and therefore delivers more precise estimates. However, the treatment effect estimate may mask heterogeneity of effects at different directions. If a researcher has a reason to suspect significant heterogeneity, then their proposed estimator will make a better fit.

This paper also contributes to a small literature on difference-in-differences estimators from a spatial lens ([Butts, 2021](#); [Clarke, 2017](#); [Berg and Streitz, 2019](#); [Verbitsky-Savitz and Raudenbush, 2012](#); [Delgado and Florax, 2015](#)). These papers concern the case where treatment is well defined by administrative boundaries but spillovers cause similar problems of defining who is ‘treated’ and at what level of exposure. [Butts \(2021\)](#); [Clarke \(2017\)](#) recommend a method of using many rings to estimate treatment effects similar to the proposed nonparametric estimator, but did not specify how to choose the number and width of the rings. Since this paper focuses on more local shocks where constant parallel trends are plausible, I am able to provide a data-driven approach to choosing rings.

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



Notes: This figure shows an example of the ring method. For each panel, the inner ring marks units considered 'treated', the outer ring marks units considered 'control' units, and the rest of the observations are removed from the sample. Then average changes in outcomes are compared between the treated and the control units to form a treatment effect estimate.

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. Since treatment is targeted to the high poverty neighborhood, comparisons with other neighborhoods in the cities could be biased if the neighborhood home prices are on different trends. Hence, the researcher wants to look only at the homes in the immediate neighborhood.

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 underlying (constant) counterfactual change in home prices, normalized to 0. Panel (a) of Figure 2 shows the best-case scenario where

the treated ring is correctly specified. The two horizontal lines show the average change in outcome in the treated ring and the control ring. The treatment effect estimate, $\hat{\tau}$, is the difference between these two averages. 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, later in this paper I recommend nonparametrically estimating the treatment effect curve as a function of distance rather than using average effect.

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. Therefore when the treated ring is too large, the estimated treatment effect is too small. 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 one may suspect. Since the treatment effect decays with distance, the average change in outcome among the more narrow ‘treatment’ ring is larger than the correct specification. The estimated treatment effect in this case grows, but it is not clear more generally whether the treatment effect will increase or decrease.³ From these three examples, it’s clear that the estimation strategy requires researchers to know the exact distance at which treatment effects become zero. Since this is a very demanding assumption, I propose an improved estimator in section 4 that relaxes this assumption.

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 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. Now, I turn to econometric theory in order to formalize the intuition developed in this section.

3. This primarily depends on the curvature of the treatment effect curve.

3 – Theory

Now, I develop econometric theory to formalize the intuition developed in the previous section. A researcher observes panel data of a random sample 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 $\text{Dist}_i \equiv d(\theta_i, \bar{\theta})$ for some distance metric d (e.g. Euclidean distance) with a distribution function F . Outcomes are given by

$$Y_{it} = \mu_i + \tau_i \mathbf{1}_{t=1} + \lambda_i \mathbf{1}_{t=1} + u_{it}, \quad (1)$$

where μ_i is unit-specific time-invariant factors, λ_i is the change in outcomes due to non-treatment shocks in period 1, τ_i is unit i 's treatment effect. Both λ and τ can be split into a systematic function of distance $z(\text{Dist}_i)$ and an idiosyncratic term $\tilde{z}_i \equiv z_i - z(\text{Dist}_i)$ with z being τ and λ . $\tau(d)$ is the average effect of treatment at a given distance and $\lambda(d)$ summarizes how covariates and shocks change over distance. Therefore, we could rewrite our model as

$$Y_{it} = \mu_i + \tau(\text{Dist}_i) \mathbf{1}_{t=1} + \lambda(\text{Dist}_i) \mathbf{1}_{t=1} + \varepsilon_{it}, \quad (2)$$

where $\varepsilon = u_{it} + \tilde{\tau}_i + \tilde{\lambda}_i$ which is uncorrelated with distance to treatment. Researchers are trying to identify the average treatment effect on units experiencing treatment effects, i.e. $\bar{\tau} = \mathbb{E}[\tau_i \mid \tau(\text{Dist}_i) > 0]$.

Assumption 1 (Random Sampling). The observed data consists of $\{Y_{i1}, Y_{i0}, \text{Dist}_i\}$ which is independent and identically distributed.

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

Assumption 2 (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, \bar{d} . This assumption requires that treatment location can't be targeted based on trends *within a small-area/neighborhood*. Note that

Local Parallel Trends implies the standard assumption that parallel trends holds *on average* between the treated and control rings:

Assumption 3 (Average Parallel Trends). For a pair of distances d_t and d_c , we say that ‘average parallel trends’ hold if $\mathbb{E}[\lambda_d \mid 0 \leq d \leq d_t] = \mathbb{E}[\lambda_d \mid d_t < d \leq d_c]$.

If **Local Parallel Trends** holds for some d_c , then our first-difference equation can be simplified to $\Delta Y_{it} = \tau(\text{Dist}_i) + \lambda + \Delta \varepsilon_{it}$ where λ is some constant for units in the subsample $\mathcal{D} \equiv \{i : \text{Dist}_i \leq d_c\}$. Therefore, the treatment effect curve $\tau(\text{Dist}_i)$ is identifiable up to a constant under Assumption 2. To identify $\tau(\text{Dist}_i)$ separately from the constant, researchers will often claim that treatment effects stop occurring before some distance $d_t < d_c$. This is formalized in the following assumption.

Assumption 4 (Correct d_t). A distance d_t satisfies this assumption if (i) for all $d \leq d_t$, $\tau(d) > 0$ and for all $d > d_t$, $\tau(d) = 0$ and (ii) $F(d_c) - F(d_t) > 0$.

With this assumption, the first difference equation simplifies to $\Delta Y_{it} = \lambda + \Delta \varepsilon_{it}$ for units with $d_t < \text{Dist}_i < d_c$. These units therefore identify λ . 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 $\mathcal{D}_t \equiv \{i : 0 \leq \text{Dist}_i \leq d_t\}$ 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 estimate the following regression:

$$\Delta Y_{it} = \beta_0 + \beta_1 \mathbf{1}_{i \in \mathcal{D}_t} + u_{it}. \quad (3)$$

From standard results for regressions involving only indicators, $\hat{\beta}_1$ is the difference-in-differences estimator with the following expectation:

$$\mathbb{E}[\hat{\beta}_1] = \mathbb{E}[\Delta Y_{it} \mid \mathcal{D}_t] - \mathbb{E}[\Delta Y_{it} \mid \mathcal{D}_c].$$

This estimate is decomposed in the following proposition.⁴

Proposition 1 (Decomposition of Ring Estimate). Given that units follow model (2),

(i) The estimate of β_1 in (5) has the following expectation:

$$\begin{aligned} \mathbb{E}[\hat{\beta}_1] &= \mathbb{E}[\Delta Y_{it} \mid \mathcal{D}_t] - \mathbb{E}[\Delta Y_{it} \mid \mathcal{D}_c] \\ &= \underbrace{\mathbb{E}[\tau(\text{Dist}) \mid \mathcal{D}_t] - \mathbb{E}[\tau(\text{Dist}) \mid \mathcal{D}_c]}_{\text{Difference in Treatment Effect}} + \underbrace{\mathbb{E}[\lambda(\text{Dist}) \mid \mathcal{D}_t] - \mathbb{E}[\lambda(\text{Dist}) \mid \mathcal{D}_c]}_{\text{Difference in Trends}}. \end{aligned}$$

4. A similar derivation of part (i) is found in [Sullivan \(2017\)](#) but does not include difference in parallel trends.

(ii) If d_c satisfies **Local Parallel Trends** or, more weakly, if d_t and d_c satisfy **Average Parallel Trends**, then

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

(iii) If d_c satisfies **Local Parallel Trends** and d_t satisfies Assumption 8, then

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

Part (i) of this proposition shows that the 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. The second difference is the difference in counterfactual trends between the treated and control rings. This presents two possible problems. If some units in the control group experience effects from treatment, the average of these effects will be subtracted from the estimate. Second, since treatment can be targeted, the treated ring could be on a different trend than units further away and hence control units do not serve as a good counterfactual for treated units.

Part (ii) says that if d_c satisfies **Local Parallel Trends**, then the difference in trends from part (i) is equal to 0. 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, $\hat{\beta}_1$ will be biased towards zero from the inclusion of unaffected units from d_t being too wide. Second, if d_t is *too narrow* then the \mathcal{D}_c will contain units that experience treatment effects. It is not clear in this case, though, whether $\hat{\beta}_1$ will grow or shrink without knowledge of the $\tau(\text{Dist})$ curve, but typically $\hat{\beta}_1$ will not be an unbiased estimate for $\bar{\tau}$. See the previous section for an example.

Part (iii) 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 8 is a very demanding assumption and unlikely to be known by the researcher unless there are *a priori* theory dictating d_t .⁵ The following section will improve estimation by allowing consistent nonparametric estimation of the entire $\tau(\text{Dist})$ function. An estimate of $\tau(\text{Dist})$ can then be numerically integrated to for an estimate of $\bar{\tau}$.

4 – Nonparametric Estimation of the Treatment Effect Curve

In this section, I propose an estimation strategy that nonparametrically identifies the treatment effect curve $\tau(\text{Dist}_i)$ using partitioning-based least squares estimation and inference methods developed

5. As an example, Currie et al. (2015) uses results from scientific research on the maximum spread of local pollutants and Marcus (2021) use the plume length of petroleum smoke.

in Cattaneo et al. (2019a,b). Partition-based estimators separate the support of a covariate, Dist_i , into a set of quantile-spaced intervals (e.g. 0-25th percentiles of Dist_i , 25-50th, 50-75th, and 75-100th). Then the conditional $\mathbb{E}[Y_i | \text{Dist}_i]$ is estimated separately within each interval as a k -degree polynomial of the covariate x_i .

For a given d_c , we will form a partition of our sample $\mathcal{D} = \{i : \text{Dist}_i \leq d_c\}$ into L intervals based on quantiles of the distance variable. Denote a given quantile as $\mathcal{D}_j \equiv \{i : F_n^{-1}(\frac{j-1}{L}) \leq \text{Dist}_i < F_n^{-1}(\frac{j}{L})\}$ where F_n is the empirical distribution of Dist . Let $\{\mathcal{D}_1, \dots, \mathcal{D}_L\}$ be the collection of the L intervals. This paper will impose $k = 0$ which will predict ΔY_{it} with a constant within each interval.⁶

These averages are defined as

$$\overline{\Delta Y}_j \equiv \frac{1}{n_j} \sum_{i \in \mathcal{D}_j} \Delta Y_{it},$$

where the number of units in bin \mathcal{D}_j is $n_j \approx n/L$. Our estimator for $\mathbb{E}[\Delta Y_{it} | \text{Dist}_i]$ is then given by

$$\widehat{\Delta Y}_{it} = \sum_{j=1}^L \mathbf{1}_{i \in \mathcal{D}_j} \overline{\Delta Y}_j$$

As the number of intervals approach infinity, this estimate will approach $\mathbb{E}[\Delta Y_{it} | \text{Dist} = d]$ in a mean-squared error sense. Under **Local Parallel Trends**, $\mathbb{E}[\Delta Y_{it} | \text{Dist} = d] \equiv \mathbb{E}[\tau(\text{Dist}) | \text{Dist} = d] + \lambda$. To remove λ , we require a less-strict version of assumption 8.

Assumption 5 (d_t is within d_c). A distance d_c satisfies this assumption if there exists a distance d_t with $0 < d_t < d_c$ such that (i) Assumption 8 holds and (ii) $F(d_c) - F(d_t) > 0$.

If a distance d_c satisfies **Local Parallel Trends** and (5), the mean within the last ring \mathcal{D}_L will estimate λ as the number of bins $L \rightarrow \infty$. The reason for this is simple, as $L \rightarrow \infty$, the last bin will have the left end-point $> d_t$ and therefore $\tau(\text{Dist}) = 0$ in \mathcal{D}_L . Under local parallel trends, the last ring will therefore estimate λ . Therefore, estimates of $\tau(\text{Dist}_i)$ can be formed for each interval as $\hat{\tau}_j \equiv \overline{\Delta Y}_j - \overline{\Delta Y}_L$.

Proposition 2 (Consistency of Nonparametric Estimator). Given that units follow model (2) and d_c satisfies **Local Parallel Trends** and assumption (5), as n and $L \rightarrow \infty$

$$\hat{\tau} \equiv \sum_{i=1}^L \hat{\tau}_j \mathbf{1}_{i \in \mathcal{D}_j} \rightarrow^{unif} \tau(\text{Dist})$$

where d_j corresponds to the $F^{-1}(D_j)$.

6. Approximation can be made arbitrarily close to the true conditional expectation function by *either* increasing the number of intervals or by increasing the polynomial order to infinity, so setting $k = 0$ does not impose any cost.

As discussed in Section 2, specifying d_t correctly is important to identify the average treatment effect among the affected in the parametric estimator. The nonparametric estimator only requires that treatment effects become zero before d_c , i.e. that such a d_t exists. However, the estimator would no longer identify the treatment effect curve under the milder **Average Parallel Trends** assumption. Therefore, a researcher should justify explicitly the assumption that, within the d_c ring, every unit is subject to the same trend. This is most likely to be satisfied on a very local level and not very plausible in the case of larger units, e.g. counties.

The nonparametric approach allows estimation of the treatment effect curve whereas the indicator approach, *at best*, can only estimate an *average* effect among units experiencing effects. The treatment effect curve allows researcher to understand differences in treatment effect across distance. For example, typically one would assume treatment effects shrink over distance and evidence of this from the nonparametric approach can strengthen a causal claim. In some cases, such as a negative hyper-local shock and a positive local shock (e.g. a local bus-stop), the treatment effect can even change sign across distances. In this case, the average effect could be near zero even though there are significant effects occurring.

Plotting estimates $\hat{\tau}_j$ can provide visual evidence for the underlying **Local Parallel Trends** assumption. Typically, treatment effect will stop being experienced far enough away from d_c that some estimates of $\hat{\tau}_j$ with j 'close to' L will provide informal tests for parallel trends holding. **Figure 4** provide an example where plotting of $\hat{\tau}_j$ provide strong evidence in support of local parallel trends as it appears that after some distance, average effects are consistently centered around zero. This is not a formal test as it could be the case that the true treatment effect curve, $\tau(\text{Dist})$ is perfectly cancelling out with the counterfactual trends curve $\lambda(\text{Dist})$ producing near zero estimates, but this is a knife's edge case.

The above proposition shows that the series estimator will consistently estimate the treatment effect curve, $\tau(\text{Dist})$ as the number of bins L and the number of observations n both go to infinity. In finite-samples though, we will have a fixed L and hence a fixed set of treatment effect estimates $\{\tau_1, \dots, \tau_L\}$ with $\tau_L \equiv 0$ by definition. The estimates $\hat{\tau}_j$ are approximately equal to $\mathbb{E}[\tau(\text{Dist}) \mid \text{Dist} \in \mathcal{D}_j]$ or the average treatment effect within the interval \mathcal{D}_j .

The choice of L in finite samples is not entirely clear. **Cattaneo et al. (2019a)** derive the IMSE-optimal choice of L which is a completely data-driven choice. The optimal L is driven by two competing terms in the IMSE formula. On the one hand, as L increases, the conditional expectation function is allowed to vary more across values of Dist and hence bias of the estimator decreases. However, larger values of L increase the variance of the estimator. Balancing this trade-off depends on the shape and curvature of $\tau(\text{Dist})$. The resulting choice of L^* and the use of quantiles of the data allows

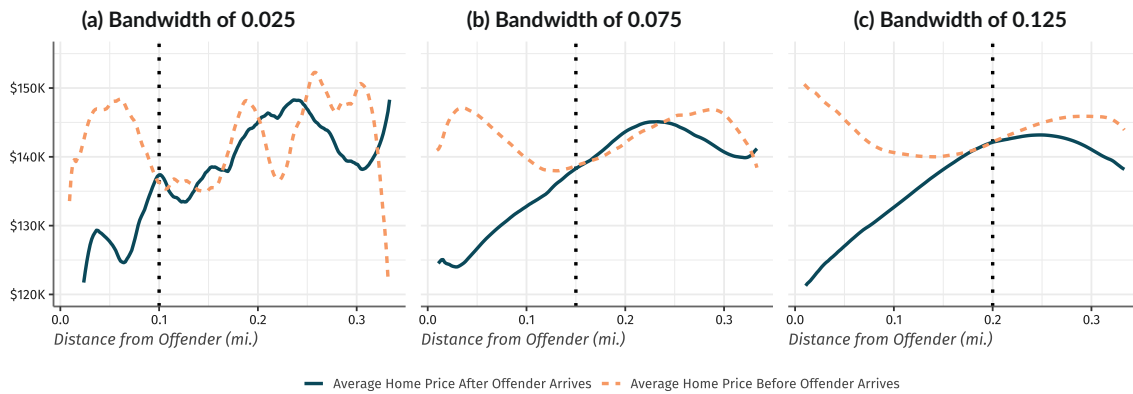
a completely data-driven choice of the number of rings and their endpoints which allows for estimation in a principled and objective way. This principled estimator removes researcher-incentives to search across choices of rings to provide the best evidence.

For a given L^* , Cattaneo et al. (2019a) show the large-sample asymptotics of the estimates $\overline{\Delta Y}_j$ and provide robust standard errors for the conditional means that account for the additional randomness due to quantile estimation. Since our estimator is a difference in means, standard errors on our estimate $\hat{\tau}_j$ are given by $\sqrt{\sigma_j^2 + \sigma_L^2}$, where σ_j is the standard error recommended by Cattaneo et al. (2019a). These standard errors are produced by the Stata/R package `binsreg`. Inference can be done by using the estimated t-stat with the standard normal distribution.

5 – Application to Neighborhood Effects of Crime Risk

To highlight the advantages of my proposed estimator, I revisit the analysis of Linden and Rockoff (2008). This paper analyzes the effect of a sex offender moving to a neighborhood on home prices. This paper uses the ring method with treated homes being defined as being within $1/10^{th}$ of the sex offender's home and the control units being between $1/10^{th}$ and $1/3^{rd}$ of a mile from the home. The authors make a case for the ring method by arguing that *within a neighborhood*, Local Parallel Trends holds since they are looking in such a narrow area and purchasing a home is difficult to be precisely located with concurrent hyper-local shocks.

Figure 3 – Price Gradient of Distance from Offender



Notes: This figure plots estimates of home prices in the year before and the year after the arrival of a sex offender estimated using a Local Polynomial Kernel Density estimation with an Epanechnikov kernel. Panel (b) recreates Figure 2 from Linden and Rockoff (2008) and the other panels change the bandwidth.

As for the choice of the treatment ring, there is little *a priori* reasons to know how far the effects of sex offender arrival will extend in the neighborhood. The authors provide graphical evidence

of nonparametric estimates of the conditional mean home price at different distances in the year before and the year after the arrival of a sex offender. The published plot can be seen in Panel (b) of [Figure 3](#). They ‘eyeball’ the point at which the two estimates are approximately equal to decide how far treatment effects extend. However, this approach is less precise than it may seem. Panels (a) and (c) show that changing the bandwidth for the kernel density estimator will produce very different guesses at how far treatment effects extend. My proposed estimator works in a data-driven way that does not require these ad-hoc decisions.

The standard rings approach is equivalent to my proposed method with two rings: \mathcal{D}_1 being the treated homes between 0 and 0.1 miles away and \mathcal{D}_2 being the control homes between 0.1 and 0.3 miles away. The average change among \mathcal{D}_2 estimates the counterfactual trend and the average change among \mathcal{D}_1 minus the estimated counterfactual trend serves as the treatment effect. Panel (a) of [Figure 4](#) shows the basic results of their difference-in-differences analysis which plots estimates $\hat{\tau}_j$ for $j = 1, 2$. On average, homes between 0 and 0.1 miles decline in value by about 7.5% after the arrival of a sex offender. As an *assumption* of the rings method, homes between 0.1 and 0.3 miles away are not affected by a sex offender arrival. The choice of 0.1 miles is an untestable assumption and as seen above the evidence provided is highly dependent on the choice of bandwidth parameter. My proposed estimator does not require a specific choice for a ‘treated’ area.

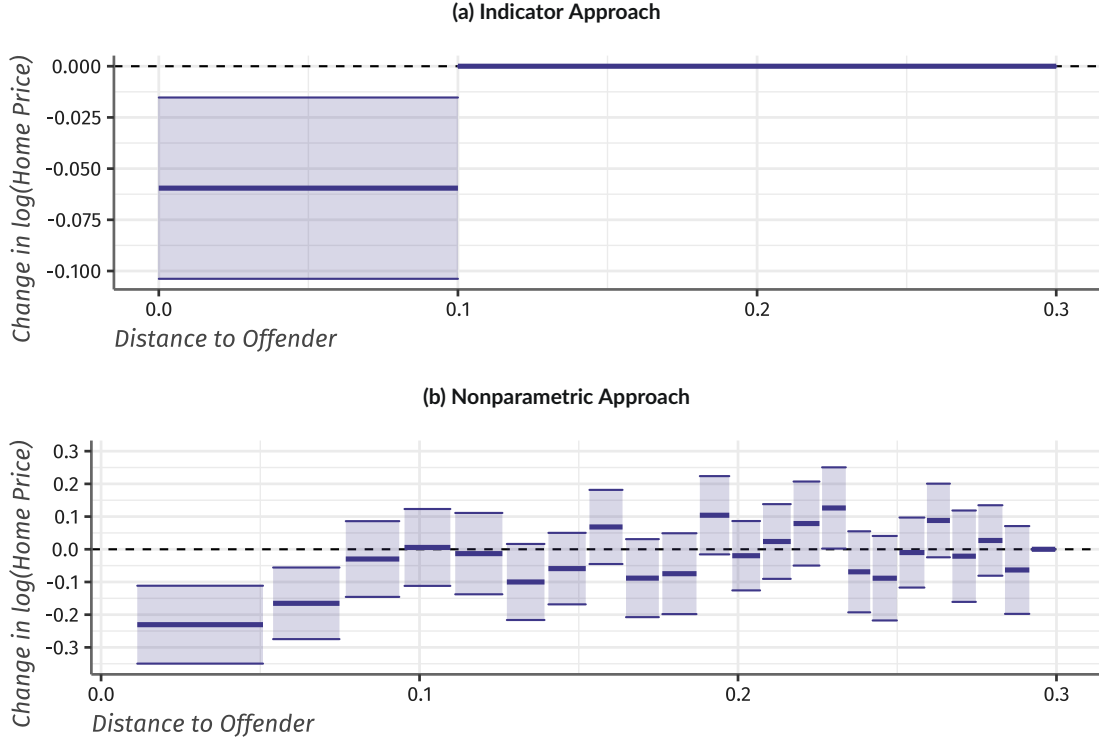
[Linden and Rockoff \(2008\)](#) only have access to a non-panel sample of home sales, so identification requires another assumption for identification, namely that the composition of homes at a given distance does not change over time. Further, since we can no longer form first-differences of the outcome variable, separate nonparametric estimators must be estimated before and after treatment and subtracted from one another. Details of this theory are in the Online Appendix

Panel (b) of [Figure 4](#) applies the nonparametric approach described in [Section 4](#). Two differences in results occur. First, homes in the two closest rings i.e. within a few hundred feet, are most affected by sex-offender arrival with an estimated decline of home value of around 20%. homes a bit further away but still within in Linden and Rockoff’s ‘treated’ sample do not experience statistically significant treatment effects. As discussed above, Linden and Rockoff’s estimate of $\bar{\tau}$ is attenuated towards zero because of the inclusion of homes with little to no treatment effects, leading them to understate the effect of arrival on home prices. The nonparametric approach improves on answering this question by providing a more complete picture of the treatment effect curve. The magnitude of treatment effects decrease over distance, providing additional evidence that the arrival causes a drop in home prices.⁷

The second advantage of this approach is that the produced figure provides an informal test of

7. This is similar to estimating a dose-response function as evidence supporting a causal mechanism.

Figure 4 – Effects of Offender Arrival on Home Prices (Linden and Rockoff, 2008)



Notes: This figure plots the estimated change in home prices after the arrival of a registered sex offender as a function of distance from offender. Each line plots $\hat{\tau}_j = \overline{\Delta Y_j} - \overline{\Delta Y_l}$ with associated standard errors. Panel (a) shows an estimate from Equation 5 with a treatment distance of $1/10^{th}$ miles and a control distance of $1/3^{rd}$ mile. Panel (b) shows the nonparametric estimate of $\tau(\text{Dist}_i)$ proposed in Section 4.

the local parallel trends assumption. After 0.1 miles, the estimated treatment effect curve becomes centered at zero consistently. This implies that units within each ring have the same estimated trend as the outer most ring, providing suggestive evidence that homes in this neighborhood are subject to the same trends.

6 – Discussion

This article formalizes a common applied identification strategy that has a strong intuitive appeal. When treatment effects of shocks are experienced in only part of an area that would otherwise be on a common neighborhood-trend, difference-in-differences comparisons within a neighborhood can identify treatment effects. However, this paper shows that the typical *estimator* for treatment effects requires a very strong assumption and returns only an average treatment effect among affected units when this assumption holds.

This article then proposes an improved estimator that relies on nonparametric series estimators. The nonparametric estimator allows for estimation of the treatment effect at different distances from treatment, similar to a dose-response function, which can allow better understanding of *who* is experiencing effects and how this changes across 'exposure' to a shock. More, in some cases it can provide explanation for null results. For example, if a bus station creates negative externalities for apartments that border the station but positive externalities for apartments within walking distance, the average effect could be zero. However, nonparametric estimation would reveal the two effects separately.

References

- Alexander, Diane, Janet Currie, and Molly Schnell. 2019. "Check up before you check out: Retail clinics and emergency room use." *Journal of Public Economics* 178 104050. [10.1016/j.jpubeco.2019.104050](https://doi.org/10.1016/j.jpubeco.2019.104050).
- Berg, Tobias, and Daniel Streitz. 2019. *Handling Spillover Effects in Empirical Research*. 59.
- Butts, Kyle. 2021. "Difference-in-Differences Estimation with Spatial Spillovers."
- Cattaneo, Matias D., Richard K. Crump, Max H. Farrell, and Yingjie Feng. 2019a. "On Binscatter." *arXiv:1902.09608 [econ, stat]*, <http://arxiv.org/abs/1902.09608>, arXiv: 1902.09608.
- Cattaneo, Matias D., Max H. Farrell, and Yingjie Feng. 2019b. "Large Sample Properties of Partitioning-Based Series Estimators." *arXiv:1804.04916 [econ, math, stat]*, <http://arxiv.org/abs/1804.04916>, arXiv: 1804.04916.
- Clarke, Damian. 2017. "Estimating Difference-in-Differences in the Presence of Spillovers." *Munich Personal RePEc Archive* 52.
- Currie, Janet, Lucas Davis, Michael Greenstone, and Reed Walker. 2015. "Environmental Health Risks and Housing Values: Evidence from 1,600 Toxic Plant Openings and Closings." *American Economic Review* 105 (2): 678–709. [10.1257/aer.20121656](https://doi.org/10.1257/aer.20121656).
- Delgado, Michael S., and Raymond J.G.M. Florax. 2015. "Difference-in-differences techniques for spatial data: Local autocorrelation and spatial interaction." *Economics Letters* 137 123–126. [10.1016/j.econlet.2015.10.035](https://doi.org/10.1016/j.econlet.2015.10.035).
- Diamond, Rebecca, and Tim McQuade. 2019. "Who Wants Affordable Housing in Their Backyard? An Equilibrium Analysis of Low-Income Property Development." *journal of political economy* 55.
- Gerardi, Kristopher, Eric Rosenblatt, Paul S. Willen, and Vincent Yao. 2015. "Foreclosure externalities: New evidence." *Journal of Urban Economics* 87 42–56. [10.1016/j.jue.2015.02.004](https://doi.org/10.1016/j.jue.2015.02.004).
- Linden, Leigh, and Jonah E Rockoff. 2008. "Estimates of the Impact of Crime Risk on Property Values from Megan's Laws." *American Economic Review* 98 (3): 1103–1127. [10.1257/aer.98.3.1103](https://doi.org/10.1257/aer.98.3.1103).
- Marcus, Michelle. 2021. "Going Beneath the Surface: Petroleum Pollution, Regulation, and Health." *American Economic Journal: Applied Economics* 13 (1): 72–104. [10.1257/app.20190130](https://doi.org/10.1257/app.20190130).

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

Verbitsky-Savitz, Natalya, and Stephen W. Raudenbush. 2012. “Causal Inference Under Interference in Spatial Settings: A Case Study Evaluating Community Policing Program in Chicago.” *Epidemiologic Methods* 1 (1): . [10.1515/2161-962X.1020](#).

A – Proofs

A.1. Proof of Proposition 1

Proof. Note using our model (2), we have

$$\begin{aligned}\mathbb{E}[\hat{\beta}_1] &= \mathbb{E}[\Delta Y_{it} \mid \mathcal{D}_t] - \mathbb{E}[\Delta Y_{it} \mid \mathcal{D}_c] \\ &= \mathbb{E}[\tau(\text{Dist}_i) + \lambda(\text{Dist}_i) + \Delta \varepsilon_i \mid \mathcal{D}_t] - \mathbb{E}[\tau(\text{Dist}_i) + \lambda(\text{Dist}_i) + \Delta \varepsilon_i \mid \mathcal{D}_c] \\ &= \mathbb{E}[\tau(\text{Dist}_i) \mid \mathcal{D}_t] - \mathbb{E}[\tau(\text{Dist}_i) \mid \mathcal{D}_c] + \mathbb{E}[\lambda(\text{Dist}_i) \mid \mathcal{D}_t] - \mathbb{E}[\lambda(\text{Dist}_i) \mid \mathcal{D}_c] + \mathbb{E}[\Delta \varepsilon_i \mid \mathcal{D}_t] - \mathbb{E}[\Delta \varepsilon_i \mid \mathcal{D}_c].\end{aligned}$$

By construction, $\Delta \varepsilon_i$ is uncorrelated with distance, so the final two terms in the sum is zero giving us result (i). Result (ii) comes from the fact that within d_c , $\lambda(\text{Dist}_i) = \lambda$. Result (iii) comes from the fact that if d_t is the correct cutoff $\mathbb{E}[\tau(\text{Dist}_i) \mid \mathcal{D}_c] = 0$. \square

A.2. Proof of Proposition 2

Proof. Note that $L \rightarrow \infty$ implies $d_t \leq F_n^{-1}(\frac{L-1}{L})$ by assumption (5). This implies $\overline{\Delta Y}_L \xrightarrow{p} \lambda$ as $n \rightarrow \infty$ by assumption (5).

From assumption (2) and from our model (2), we have

$$\begin{aligned}\hat{\tau}_j &= \overline{\Delta Y}_j - \overline{\Delta Y}_L \\ &\xrightarrow{p} \mathbb{E}[\tau(\text{Dist}) \mid \text{Dist} \in \mathcal{D}_j] + \lambda - \lambda \\ &= \mathbb{E}[\tau(\text{Dist}) \mid \text{Dist} \in \mathcal{D}_j]\end{aligned}$$

As $L \rightarrow \infty$ and $n \rightarrow \infty$, we have that \mathcal{D}_j approaches a set containing a singular point, say d_j . Therefore

$$\hat{\tau}_j \xrightarrow{p} \mathbb{E}[\tau(\text{Dist}) \mid \text{Dist} = d_j]$$

The sum of $\hat{\tau}_j$ therefore approach the conditional expectation function of $\tau(\text{Dist})$ pointwisely.

See [Cattaneo et al. \(2019b\)](#) for proof of uniform convergence and underlying smoothness conditions for nonparametric consistency. \square

B – Cross-Sectional Data

In the case of cross sectional data, we have individuals i that appear in the data in period $t(i) \in \{0, 1\}$. However, since we no longer are able to observe units in both periods, we are not able to take first differences. Our model therefore will have a λ term for both periods. Therefore, λ includes the average of μ_i , covariates, and period shocks at a given distance.

$$Y_i = \tau(\text{Dist}_i) \mathbf{1}_{t(i)=1} + \lambda_{t(i)}(\text{Dist}_i) + \nu_{it}. \quad (4)$$

The parallel trends assumption must be modified now in the case of cross sections:

Assumption 6 (Local Parallel Trends (RC)). For a distance \bar{d} , we say that ‘local parallel trends’ hold if for all positive $d, d' \leq \bar{d}$, then $\lambda_1(d) - \lambda_0(d) = \lambda_1(d') - \lambda_0(d')$.

Assumption 7 (Average Parallel Trends (RC)). For a pair of distances d_t and d_c , we say that ‘average parallel trends’ hold if $\mathbb{E}[\lambda_1(d) \mid 0 \leq d \leq d_t] - \mathbb{E}[\lambda_0(d) \mid 0 \leq d \leq d_t] = \mathbb{E}[\lambda_1(d) \mid d_t < d \leq d_c] - \mathbb{E}[\lambda_0(d) \mid d_t < d \leq d_c]$.

The parallel trends assumption is a bit more complicated now and is theoretically more strict. **Local Parallel Trends (RC)** still require changes in outcomes over time for a given unit i must be constant across distance (or on average in the case of **Average Parallel Trends (RC)**). However since the composition of units can change over time, this also requires that the average of individual fixed effects must be constant across time. This is well understood in the hedonic pricing literature that the composition of homes being sold can not change over time for identification (e.g. [Linden and Rockoff \(2008\)](#)).

For completeness, I rewrite the other necessary assumption

Assumption 8 (Correct d_t). A distance d_t satisfies this assumption if (i) for all $d \leq d_t$, $\tau(d) > 0$ and for all $d > d_t$, $\tau(d) = 0$ and (ii) $F(d_c) - F(d_t) > 0$.

The ring estimate in the case of cross-sections is given by:

$$\Delta Y_i = \beta_0 + \beta_1 \mathbf{1}_{i \in \mathcal{D}_c} \mathbf{1}_{t(i)=1} + \beta_2 \mathbf{1}_{i \in \mathcal{D}_t} \mathbf{1}_{t(i)=0} + \beta_4 \mathbf{1}_{i \in \mathcal{D}_t} \mathbf{1}_{t(i)=1} + u_{it}. \quad (5)$$

Proposition 3 (Decomposition of Ring Estimate (RC)). Given that units follow model (4),

(i) The estimate of β_4 in (5) has the following expectation:

$$\begin{aligned}
\mathbb{E} [\hat{\beta}_4] &= \mathbb{E} [\Delta Y_{it} \mid \mathcal{D}_t] - \mathbb{E} [\Delta Y_{it} \mid \mathcal{D}_c] \\
&= \underbrace{\mathbb{E} [\tau(\mathbf{Dist}) \mid \mathcal{D}_t] - \mathbb{E} [\tau(\mathbf{Dist}) \mid \mathcal{D}_c]}_{\text{Difference in Treatment Effect}} \\
&\quad + (\mathbb{E} [\lambda_1(\mathbf{Dist}) \mid \mathcal{D}_t, t(i) = 1] - \mathbb{E} [\lambda_0(\mathbf{Dist}) \mid \mathcal{D}_t, t(i) = 0]) \\
&\quad - (\mathbb{E} [\lambda_1(\mathbf{Dist}) \mid \mathcal{D}_c, t(i) = 1] - \mathbb{E} [\lambda_0(\mathbf{Dist}) \mid \mathcal{D}_c, t(i) = 0])
\end{aligned}$$

(ii) If d_c satisfies **Local Parallel Trends (RC)** or, more weakly, if d_t and d_c satisfy **Average Parallel Trends (RC)**, then

$$\mathbb{E} [\hat{\beta}_4] = \underbrace{\mathbb{E} [\tau(\mathbf{Dist}) \mid \mathcal{D}_t] - \mathbb{E} [\tau(\mathbf{Dist}) \mid \mathcal{D}_c]}_{\text{Difference in Treatment Effect}}.$$

(iii) If d_c satisfies **Local Parallel Trends (RC)** and d_t satisfies Assumption 8, then

$$\mathbb{E} [\hat{\beta}_4] = \bar{\tau}.$$

Proof. With some algebraic manipulation, we can rewrite our difference-in-differences estimator as

$$\begin{aligned}
&(\mathbb{E} [Y_i \mid \mathcal{D}_t, t(i) = 1] - \mathbb{E} [Y_i \mid \mathcal{D}_t, t(i) = 0]) - (\mathbb{E} [Y_i \mid \mathcal{D}_c, t(i) = 1] - \mathbb{E} [Y_i \mid \mathcal{D}_c, t(i) = 0]) \\
&= (\mathbb{E} [\tau(\mathbf{Dist}_i) + \lambda_1(\mathbf{Dist}_i) + \nu_{i1} \mid \mathcal{D}_t, t(i) = 1] - \mathbb{E} [\lambda_0(\mathbf{Dist}_i) + \nu_{i0} \mid \mathcal{D}_t, t(i) = 0]) \\
&\quad - (\mathbb{E} [\tau(\mathbf{Dist}_i) + \lambda_1(\mathbf{Dist}_i) + \nu_{i1} \mid \mathcal{D}_c, t(i) = 1] - \mathbb{E} [\lambda_0(\mathbf{Dist}_i) + \nu_{i0} \mid \mathcal{D}_c, t(i) = 0]) \\
&= \mathbb{E} [\tau(\mathbf{Dist}_i) \mid \mathcal{D}_t, t(i) = 1] - \mathbb{E} [\tau(\mathbf{Dist}_i) \mid \mathcal{D}_c, t(i) = 0] + \\
&\quad (\mathbb{E} [\lambda_1(\mathbf{Dist}_i) \mid \mathcal{D}_t, t(i) = 1] - \mathbb{E} [\lambda_0(\mathbf{Dist}_i) \mid \mathcal{D}_t, t(i) = 0]) \\
&\quad - (\mathbb{E} [\lambda_1(\mathbf{Dist}_i) \mid \mathcal{D}_c, t(i) = 1] - \mathbb{E} [\lambda_0(\mathbf{Dist}_i) \mid \mathcal{D}_c, t(i) = 0]),
\end{aligned}$$

where the terms consisting of ν_{it} cancel out as they are uncorrelated with distance. Propositions (ii) and (iii) follow the same arguments as in the panel case. \square

Part (i) of this theorem shows that under no parallel trends assumption, the ‘Difference in Trends’ term becomes the change in λ_t for the treated ring minus the change in for the control ring. As discussed above, this change in lambdas can be do to period specific shocks or changes in the composition of units observed in each period.

B.1. Nonparametric Estimation

Since we can no longer perform a single nonparametric regression on first differences in the context of cross-sections, our nonparametric estimator must be adjusted. The modified procedure will fit a nonparametric estimate of $\mathbb{E}[Y_i | \text{Dist}_i, t]$ separately for $t = 0$ and $t = 1$ with a restriction that the bin intervals $\{\mathcal{D}_1, \dots, \mathcal{D}_L\}$ must be the same in both samples.⁸ Then, for each distance bin we calculate an estimate of $\bar{Y}_{j,t}$ which corresponds to the sample average of observations in period t in bin \mathcal{D}_j .

Then estimates of τ_j can be formed as

$$\hat{\tau}_j = [\bar{Y}_{j,1} - \bar{Y}_{j,0}] - [\bar{Y}_{L,1} - \bar{Y}_{L,0}],$$

where, as before, the change in trends in the last ring serve as an estimate for the counterfactual trend. Under **Local Parallel Trends (RC)**, estimates of $\hat{\tau}_j$ are consistent for $\mathbb{E}[\tau(\text{Dist}_i) | i \in \mathcal{D}_j]$ and the treatment effect curve converges uniformly to the treatment effect curve $\tau(d)$.

Standard errors are formed similarly as before, but is the difference of four means so they can be formed as $\sqrt{\sigma_{j,1}^2 + \sigma_{j,0}^2 + \sigma_{L,1}^2 + \sigma_{L,0}^2}$. These individual estimates and standard errors can be produced by the Stata/R package `binsreg`.

8. The number of intervals are decided based on a different IMSE condition described in Cattaneo et al. (2019a) and the quantiles are calculated using the distribution of distances in both periods.