

Difference-in-Differences with Geocoded Microdata^{*}

Kyle Butts[†]

March 20, 2022

I formalize a commonly-used estimator for the effects of spatially-targeted treatment with geocoded microdata. This estimator compares units immediately next to treatment to units slightly further away. I introduce intuitive identifying assumptions for the average treatment effect among affected units and illustrate problems 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 allows for nonparametric estimation following methods introduced in Cattaneo et al. (2019b). Since treatment effects can change with distance, the proposed estimator improves estimation by estimating a *treatment effect curve*.

JEL-Classification: C13, C14, C18

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

^{*}I am grateful to Max Farrell, Taylor Jaworski, Damian Clarke, Alexander Bentz, James Flynn, Brach Champion, and Hannah Denker for the helpful insights.

[†]University of Colorado, Boulder. Corresponding Author. Email: kyle.butts@colorado.edu.

1 – Introduction

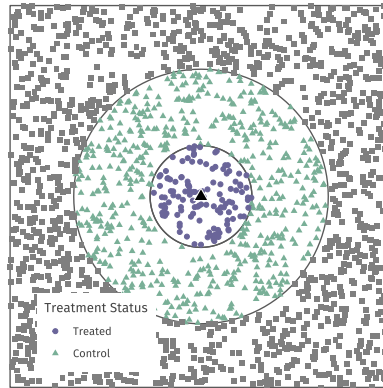
The rise of geocoded microdata has allowed researchers to begin answering questions about the effects of spatially-targeted treatments at a very granular level. How do local pollutants affect child health?¹ Does living within walking distance to a new bus stop improve labor market outcomes?² How far do neighborhood shocks, such as foreclosures or new construction spread?³ A standard method of evaluating the effects of the treatment is to compare changes in outcomes for units that are close to treatment to those slightly further away – what I label the ‘ring method’ as illustrated in [Figure 1](#). This paper formalizes the assumptions required for identification in the ring method, highlighting potential pitfalls of the currently-used estimator, and proposes an improved estimator.

First, I formalize the necessary assumptions for unbiased estimates of the average treatment effect on the affected units.⁴ 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. The rings method is typically motivated by the fact 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, shocks over time should be common across units in the neighborhood.⁵

Instead of *fully* leveraging the common neighborhood trends assumption that motivates the ring method, researchers rely on a second assumption which requires

1. See, e.g., [Currie et al. \(2015\)](#) and [Marcus \(2021\)](#).
2. See, e.g., [Gibbons and Machin \(2005\)](#) and [Billings \(2011\)](#).
3. See, e.g., [Asquith et al. \(2021\)](#), [Cui and Walsh \(2015\)](#), [Gerardi et al. \(2015\)](#) and [Campbell et al. \(2011\)](#).
4. This generalizes the treatment effect on the treated in the case where treatment is not assigned to specific units.
5. For example, [Asquith et al. \(2021\)](#) write "The idea is that within a small area, developers have few sites that are available and properly zoned, leading to hyper-local variation in the location of new construction that is not related to future price changes."

Figure 1 — Rings Method



Notes: This figure illustrates the ring method. The center triangle represents the location of treatment, e.g. a foreclosed home. Units within the inner circle, marked by dots, are considered 'treated'; 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 ring estimate compares average changes in outcomes between the inner 'treated' ring and the outer 'control' ring.

correct identification of 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 be attenuated towards zero.

I propose an alternative estimator that fully leverages the 'common neighborhood trends' assumption 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 outer ring *without* the need to specify the *exact* distance at which it does. To do this, the estimator leverages the stronger 'common neighborhood trends' assumption that

the counterfactual trend is constant across distance. This new assumption is more strict in that the standard method only requires that parallel trends holds *on average* in each ring. However, researchers typically 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, my proposed estimator creates a set of point estimates of treatment effects that can be used to visually inspect the plausability of the assumption. If after some distance, treatment effects become centered at zero, this suggests that common trends hold, akin to the pre-trends test in event study regressions.

The nonparametric estimator provides many benefits. First, the estimator allows the researcher to get a more complete picture of how the intervention affects units at various distances rather than estimating an “overall effect”. For example, 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 a single ring may average out to near-zero effects. Additionally, the proposed nonparametric estimator selects the rings in a *completely data-driven and optimal manner* removing incentives for specification searching or ‘pre-testing’ the data to determine the ‘correct’ rings (McCloskey and Michailat 2021, Andrews and Kasy 2019).

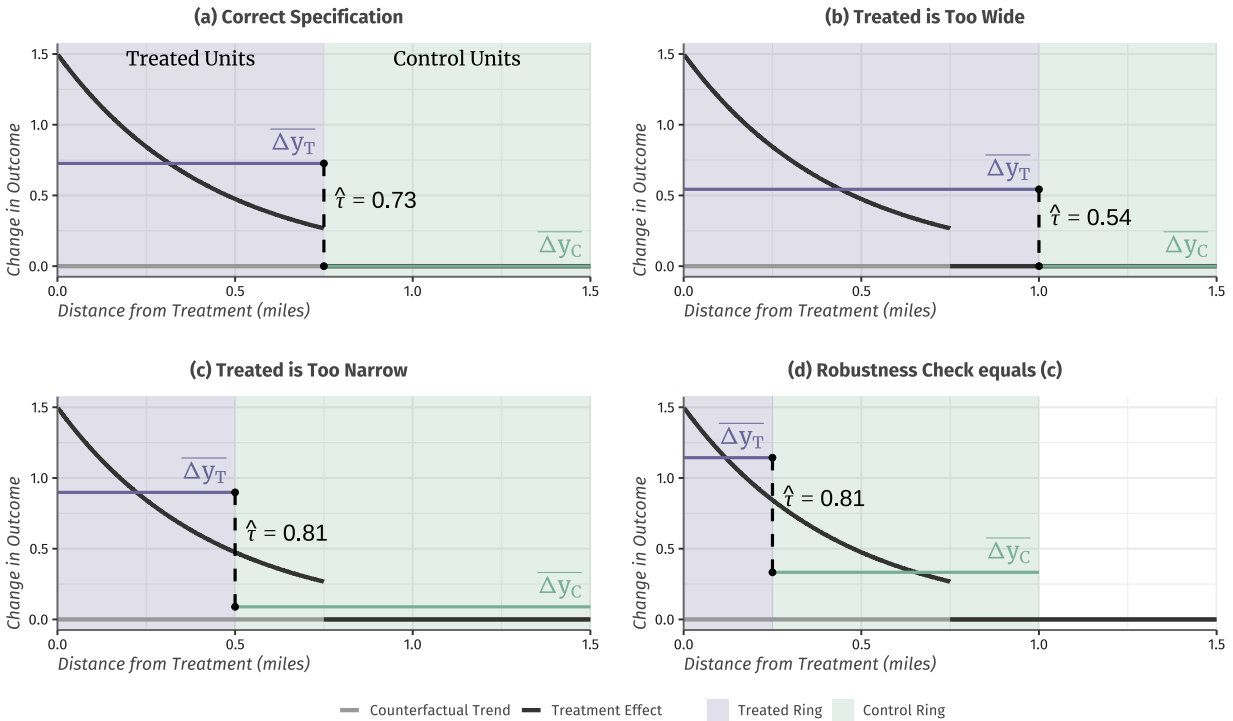
This paper contributes to the literature in many ways. First, I contribute to a literature that discuss identification using the rings method. Sullivan (2017) and Gerardi et al. (2015) discuss the identification assumption of identifying the maximum treatment distance. My paper formalizes all the necessary assumptions for identification in the ring method. Other researchers have recognized that estimating a single average treatment effect is less informative than a treatment effect curve, instead using multiple rings to estimate treatment effects at different distances (e.g. Alexander et al. (2019), Casey et al. (2018), Di Tella and Schargrodsky (2004)). However, this approach selects multiple rings in an ad-hoc manner, requires treatment effects to

become zero after the outer-most treatment ring, and is prone to problems of specification searching. The current study'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 an alternative nonparametric estimator aimed at estimating a treatment effect surface. Similar to this paper, they leverage the "common local trends" assumption to estimate a *two-dimensional* treatment effect curve. Their procedure allows "location" fixed effects to differ along two-dimensions, latitude and longitude (e.g. the mean home price is higher in the north-west than the south-east) which in cross-sectional data can absorb more variation in the data yielding smaller standard errors than my method, though both are consistent. In the case of panel data, this does not matter since first-differencing remove this systematic variation. To produce noise estimates/standard errors, their method requires integrating over the 2D curve for arbitrarily chosen rings. In cases where researchers chose these aggregation ranges based on the results of the estimation procedure, standard inference tools are not valid even when bootstrapping the standard errors (Leeb and Pötscher 2005). As discussed above, my proposed estimator provides valid inference after selection of aggregation regions and chooses these aggregation regions in a data-driven manner.

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 address instances where treatment is well defined by administrative boundaries but spillovers cause problems of defining who is 'treated' and at what level of exposure. This paper contributes to this literature by considering the setting where treatment occurs at a point in space. Butts (2021) and Clarke (2017) both recommend a method of using many rings to estimate a treatment effect curve, but are not able to provide a data-driven way to select the rings. Since this paper focuses on local shocks where

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



Notes: This figure shows an example of the difficulties in estimation of treatment effects via the ring method.

constant parallel trends are plausible, I am able to provide a data-driven approach to choosing rings.

2 – Example of Problem

To illustrate the methodological difficulties in this method, 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. 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 and our 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 experiencing treatment effects as little as half and as large as double the magnitude that of $\hat{\tau}$. 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. 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. 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 and the counterfactual trend estimate is biased upward. The treatment effect estimate grows since the new ‘treated’ group is closer and hence experiences larger treatment effects. Panel (d) of [Figure 2](#) shows an example of why using different rings as a ‘robustness check’ is a problematic practice. Using Panel (d) as a robustness check for the researchers’ original specification of Panel (c), a researcher would be quite confident in their results even though the estimate is too large in both cases.

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 by leveraging the constant common trends assumption.

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} . To clarify the assumption, it is helpful to think of ways that it can fail. First, if treatment location is targeted based on trends *within a small-area/neighborhood*, then trends would not be constant within the control ring. Second, if units sort either towards or away from treatment in a way that is systematically correlated with the outcome variable, then the compositional change can cause a violation in trends over time. 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.⁶

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

(i) The estimate of β_1 in (3) has the following expectation:

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

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

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

(iii) If d_c satisfies **Local Parallel Trends** and d_t satisfies Assumption 4, 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

6. A similar derivation of part (i) is found in Sullivan (2017) but does not include difference in parallel trends.

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 4 is a very demanding assumption and unlikely to be known by the researcher unless there are *a priori* theory dictating d_t .⁷ 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 $\bar{\tau}$.

7. 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.

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 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 Dist_i .

For a given outer-ring distance, d_c , our sample is given by $\mathcal{D} = \{i : \text{Dist}_i \leq d_c\}$. We will split the sample into L intervals based on quantiles of the distance variable. For a given $j \in \{1, \dots, L\}$, I denote the j^{th} interval 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 use a 0-degree polynomial for each interval, predicting ΔY_{it} with a constant within each interval.⁸ 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 $\mathbb{E}[\Delta Y_{it} | \text{Dist} = d]$ uniformly. 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 4.

Assumption 5 (d_t is within d_c). A distance d_c satisfies this assumption if there exists a

8. 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 asymptotic cost.

distance d_t with $0 < d_t < d_c$ such that (i) Assumption 4 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}_k 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 a 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 $\hat{\tau}_j \equiv \overline{\Delta Y}_j - \overline{\Delta Y}_L$. These $\hat{\tau}_j$ can be plotted over their intervals to provide a graphical estimate of the treatment effect curve (see below for an example).

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 $L \rightarrow \infty$

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

where d_j corresponds to the $F^{-1}(D_j)$.⁹

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 comes at a cost, namely it 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 which allows researcher to understand differences in treatment effect across distance whereas the indicator approach, *at best*, can only estimate an *average* effect

9. See Appendix A for the proof.

among units experiencing effects. 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 optimal choice of L which is a completely data-driven choice. The optimal L is driven by two competing terms in the mean-squared error 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 $\tau(\text{Dist})$ which is estimated by the data. The

resulting choice of L^* determines the number of bins and the rings are determined by quantiles of the data. Since the rings are themselves estimated, Cattaneo et al. (2019a) provide inference that is valid after the ‘first-stage’ estimation of quantiles.

This data-driven choice allows for estimation in a principled and objective way that lets the data ‘speak for itself’. Further, this method removes incentives from ‘pre-testing’ the choice of rings to find a significant result since the estimates are determined completely by the data ¹⁰

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. There may be concern that the standard errors need to adjust for spatial correlation. However, this is not the case under assumption (2) as this implies the error term is uncorrelated with distance.

Remark 1 (Overall Average Treatment Effect). A researcher may be tempted to ‘pool’ together the significant rings to estimate an overall average treatment effect among the affected. Since the estimates are quantiles in the data, a simple average of the significant $\hat{\tau}_j$ ’s can produce a back of the envelope treatment effect. However, inference on this estimand is not possible, since, in repeated sampling the number of significant $\hat{\tau}_j$ can change and this additional source of ‘model selection’ makes inference a very difficult problem (Leeb and Pötscher 2005). A potential solution would be to use cross-validation where half the data would determine the ‘inner ring’ and then the second half of the data would estimate the overall average treatment effect,

10. See, for example, discussions in Andrews and Kasy (2019) and McCloskey and Michailat (2021) about researcher degrees of freedom.

though potentially requiring a large sample size.

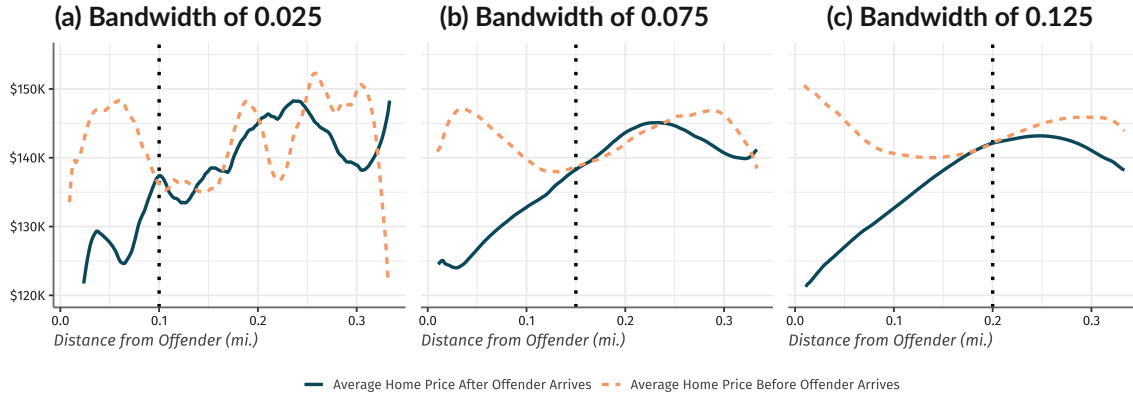
Remark 2 (Covariates). A researcher may often be interested The partitioning based series-estimator can allow for covariates to be included in the model with valid inference, so the estimation can easily be done with the Stata/R package `binsreg`. However, including covariates changes the necessary common trends assumption to have to hold conditional on the vector of covariates X (See Sant’Anna and Zhao (2020) for modern discussion of conditional parallel trends). In this paper, this estimation strategy would require neighborhood shocks to be common across values of X_i , i.e. shocks are experienced equally regardless of characteristics X .

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. This application uses cross-sectional data which similar identification and estimation results are presented in the Online Appendix.

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

Figure 3 – Price Gradient of Distance from Offender

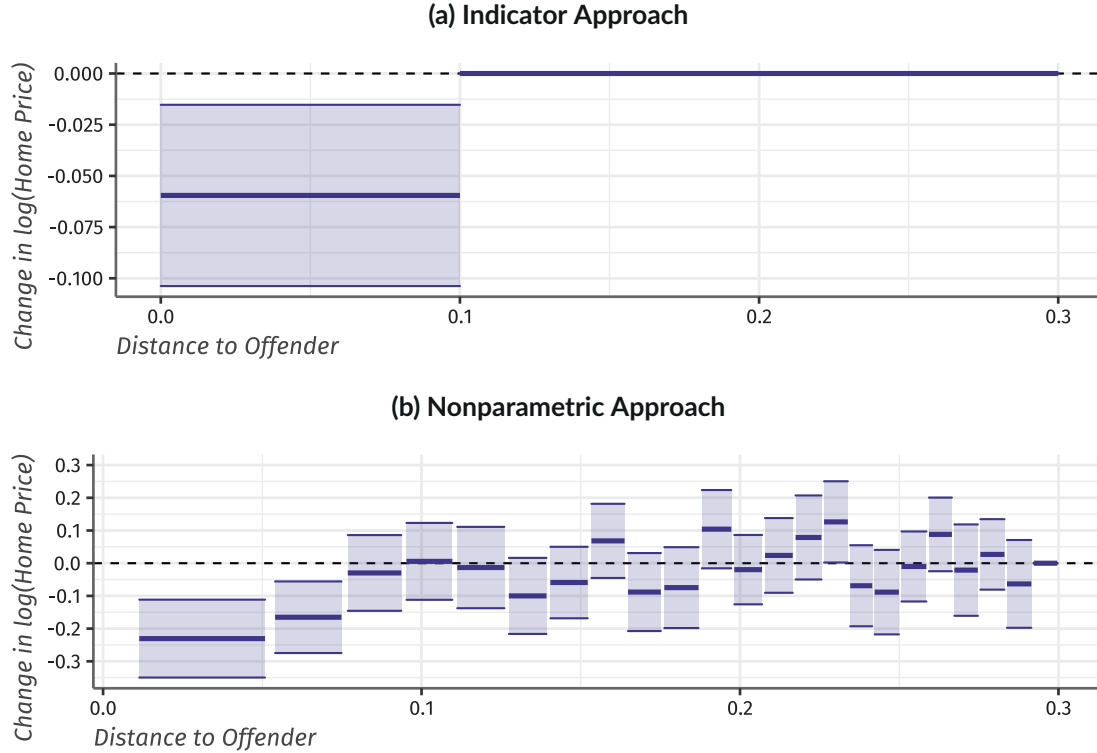


Notes: This figure plots estimates of home prices in the year before and the year after the arrival of a sex offender estimated using a Local Polynomial Kernel Density estimation with an Epanechnikov kernel. Panel (b) recreates Figure 2 from [Linden and Rockoff \(2008\)](#) and the other panels change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

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

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 3 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.

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.¹¹

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

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 se-

11. This is similar to estimating a dose-response function as evidence supporting a causal mechanism. The results of this paper are similar to the results of Callaway et al. (2021) with continuous treatment. In their framework, *Local Parallel Trends* is analogous to their assumption of common trends at different doses of a continuous treatment. In this setting, I am able to provide an estimator for the treatment effect curve, the average level effect in their terminology, by relying on an assumption that the treatment effect curve $\tau(\text{Dist})$ is homogeneous across units.

ries 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).
- Andrews, Isaiah, and Maximilian Kasy. 2019. "Identification of and correction for publication bias." *American Economic Review* 109 32.
- Asquith, Brian J., Evan Mast, and Davin Reed. 2021. "Local Effects of Large New Apartment Buildings in Low-Income Areas." *The Review of Economics and Statistics* 1–46. [10.1162/rest_a_01055](https://doi.org/10.1162/rest_a_01055).
- Berg, Tobias, and Daniel Streitz. 2019. *Handling Spillover Effects in Empirical Research*. Working Paper, 59.
- Billings, Stephen B. 2011. "Estimating the value of a new transit option." *Regional Science and Urban Economics* 41 (6): 525–536. [10.1016/j.regsciurbeco.2011.03.013](https://doi.org/10.1016/j.regsciurbeco.2011.03.013).
- Butts, Kyle. 2021. "Difference-in-Differences Estimation with Spatial Spillovers."
- Callaway, Brantly, Andrew Goodman-Bacon, and Pedro H. C. Sant'Anna. 2021. "Difference-in-Differences with a Continuous Treatment." *arXiv:2107.02637 [econ]*, <http://arxiv.org/abs/2107.02637>, arXiv: 2107.02637.
- Campbell, John Y, Stefano Giglio, and Parag Pathak. 2011. "Forced Sales and House Prices." *American Economic Review* 101 (5): 2108–2131. [10.1257/aer.101.5.2108](https://doi.org/10.1257/aer.101.5.2108).
- Casey, Marcus, Jeffrey C. Schiman, and Maciej Wachala. 2018. "Local Violence, Academic Performance, and School Accountability." *AEA Papers and Proceedings* 108 213–216. [10.1257/pandp.20181109](https://doi.org/10.1257/pandp.20181109).

- 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.
- Cui, Lin, and Randall Walsh.** 2015. "Foreclosure, vacancy and crime." *Journal of Urban Economics* 87 72–84. [10.1016/j.jue.2015.01.001](https://doi.org/10.1016/j.jue.2015.01.001).
- 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).
- Di Tella, Rafael, and Ernesto Schargrodsky.** 2004. "Do Police Reduce Crime? Estimates Using the Allocation of Police Forces After a Terrorist Attack." *American Economic Review* 94 (1): 115–133. [10.1257/000282804322970733](https://doi.org/10.1257/000282804322970733).
- 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. "Fore-

- closure externalities: New evidence." *Journal of Urban Economics* 87 42–56. [10.1016/j.jue.2015.02.004](#).
- Gibbons, Stephen, and Stephen Machin.** 2005. "Valuing rail access using transport innovations." *Journal of Urban Economics* 57 (1): 148–169. [10.1016/j.jue.2004.10.002](#).
- Leeb, Hannes, and Benedikt M. Pötscher.** 2005. "Model Selection and Inference: Facts and Fiction." *Econometric Theory* 21 (01): . [10.1017/S0266466605050036](#).
- 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](#).
- 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](#).
- McCloskey, Adam, and Pascal Michailat.** 2021. "Incentive-Compatible Critical Values." *Working Paper* 44.
- Sant'Anna, Pedro H.C., and Jun Zhao.** 2020. "Doubly robust difference-in-differences estimators." *Journal of Econometrics* 219 (1): 101–122. [10.1016/j.jeconom.2020.06.003](#).
- 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

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

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 $\overline{\Delta Y}_L \rightarrow^p \lambda$ as $n \rightarrow \infty$ by assumption (5).

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

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

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

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

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 – Monte Carlo Simulations

In the following section, I describe a set of Monte Carlo simulations that compare the standard rings method with the proposed nonparametric method. I generate a number of units on the unit circle for two periods using the following data-generating process proposed in [Diamond and McQuade \(2019\)](#):

$$p_{it} = 1 + \tau(\text{Dist}_i)\mathbf{1}_{t=1} + \beta_{\text{Lat}}\text{Lat}_i + \lambda\mathbf{1}_{t=1} + \beta_{\text{Lon}}\text{Lon}_i + \varepsilon_{it}, \quad (4)$$

where $(\text{Lat}_i, \text{Lon}_i)$ is units' x and y coordinates on the unit circle, $\beta_{\text{Lat}} \sim N(0, 0.036)$ and $\beta_{\text{Lon}} \sim N(0, 0.036)$ determine how the price levels evolve over the unit circle, $\lambda \sim N(0, 0.025)$ is the constant common trend, $\varepsilon_{it} \sim N(0, 0.036)$ are idiosyncratic errors (uncorrelated with distance).

The key component of my simulations is the treatment effect curve, $\tau(\text{Dist}_i)$, which I vary across simulations:

$$\begin{aligned} \tau_1(\text{Dist}) &= 0.15 * \mathbf{1}_{\text{Dist} < 0.4} \\ \tau_2(\text{Dist}) &= (0.5 * (0.8 - \text{Dist})^2) * \mathbf{1}_{\text{Dist} < 0.8} \\ \tau_3(\text{Dist}) &= (-0.15 + 1.2875 * \text{Dist} - 1.375 * \text{Dist}^2) * \mathbf{1}_{\text{Dist} < 0.8} \\ \tau_4(\text{Dist}) &= (0.5 * (0.8 - \text{Dist})^2) * \mathbf{1}_{\text{Dist} < 0.25} \end{aligned}$$

First, to compare the rings method and the nonparametric under favorable conditions for the ring method, I use τ_1 which assumes the treatment effect curve is flat for a fixed distance which aligns with the ring method. Second, I use the treatment

effect curve proposed by [Diamond and McQuade \(2019\)](#) which is a positive effect that declines to 0 by $\text{Dist} = 0.8$. This treatment effect curve is difficult to estimate since it evolves smoothly over space. However, the estimate of a ring might be close to the average treatment effect since the curve is positive everywhere. Third, I use τ_3 which starts with a negative effect which becomes positive around $\text{Dist} = 0.3$ and then becomes 0 by $\text{Dist} = 0.8$. This treatment effect curve is the most difficult for the rings method since the average treatment effect is near zero even though there are significant positive and negative effects. Last, I modify τ_2 to cutoff at $\text{Dist} = 0.25$, so that only units *really* close to treatment are affected by treatment. This is potentially difficult for the rings estimator which will average over many non-affected units.

For each treatment effect curve, I generate data according to (4) and estimate the treatment effect curve in three ways. I use that standard rings method with two rings $\{(0, 0.6], (0.6, 1]\}$, with three rings $\{(0, 0.3], (0.3, 0.6], (0.6, 1]\}$, and the nonparametric ring method proposed in section 4. Note that these simulations are performed with the correct maximum treatment distance satisfied, which in practice is not known. In cases where the true treatment effect distance is not known, this would introduce additional bias in the treatment effect curve estimate.

I compare each estimator on their ability to estimate the true treatment effect curve. To do so, for each point in the data, I compute the predicted treatment effect $\hat{\tau}(\text{Dist}_i)$ and compare it to the true treatment effect $\tau(\text{Dist}_i)$. I'll label this quantity, $v_i \equiv \hat{\tau}(\text{Dist}_i) - \tau(\text{Dist}_i)$, as the prediction error. To summarize the prediction error, I calculate the mean-squared prediction error, $\hat{E}(v_i^2)$, and the average absolute prediction error, $\hat{E}(|v_i|)$, where the average is taken over the sample of observations with positive treatment effects. These numbers are a combination (i) how well the predicted treatment effect curve is approximated (bias of estimator) and (ii) the estimator's noise in repeated sampling (variance of estimator). Additionally, since it is often desirable to accurately estimate the largest treatment effects, I report the bias in predicted treatment effect at 0.4 miles following [Diamond and McQuade \(2019\)](#).

For each metric, I divide by the nonparametric estimator's value, so that values > 1 perform worse than the nonparametric method and values < 1 perform better.

The results are presented in [Table 1](#). For τ_1 , the 2 ring and 3 ring method performed better than the nonparametric method across metrics and across sample sizes. This is due to (i) the large rings accurately approximate the treatment effect curve and (ii) the sampling variation of the nonparametric estimator. For τ_2 and τ_3 , the rings estimators perform better at small sample sizes, but by 5000 observations, the nonparametric method performs better in all 3 metrics which makes sense given that nonparametric estimators perform better with larger amounts of data. Last, for τ_4 , the nonparametric estimator performs better than the rings method even at small sample sizes. This is because the rings method is able to better tease out the large treatment effect *very close* to treatment than the rings method. Overall, the results of [Table 1](#) show that the rings method performs better in most settings with at least a few thousand observations except for in the most-favorable case where the treatment effect curve is approximately flat.

Table 1 – Monte Carlo Simulations

n	$\hat{E}(v_i^2)$		$\hat{E}(v_i)$		Bias at 0.4 miles	
	2 Rings	3 Rings	2 Rings	3 Rings	2 Rings	3 Rings
τ_1						
250	0.33	0.46	0.58	0.73	0.59	0.87
500	0.26	0.36	0.52	0.66	0.54	0.79
1000	0.21	0.29	0.47	0.58	0.49	0.69
5000	0.12	0.16	0.36	0.45	0.38	0.55
τ_2						
250	0.81	0.56	1.13	0.81	1.68	0.91
500	1.05	0.55	1.32	0.84	2.06	1.06
1000	1.47	0.61	1.59	0.93	2.54	1.29
5000	3.81	1.15	2.48	1.32	4.32	2.22
τ_3						
250	0.62	0.72	0.85	0.89	1.06	0.98
500	0.72	0.78	0.93	0.94	1.21	1.09
1000	0.91	0.95	1.06	1.04	1.45	1.31
5000	2.18	2.17	1.61	1.54	2.49	2.24
τ_4						
250	1.57	0.93	1.34	1.00	1.37	1.01
500	2.17	1.16	1.55	1.14	1.60	1.18
1000	3.63	1.85	1.95	1.44	2.06	1.52
5000	6.96	3.37	3.64	2.67	5.76	4.24

Notes. The table shows the results from 2,000 Monte Carlo simulations. For each treatment effect curve τ_i , I generate data following (4) and estimate the treatment effect curve using the standard ring methods with 2 or 3 rings and the nonparametric method. Columns (2)-(3) report the mean-squared prediction error, (4)-(5) report the mean absolute prediction error, and (6)-(7) report the prediction error at 0.4 miles. All results are normalized by the nonparametric value, so that values > 1 performed worse than the nonparametric method and vice-versa.