

# Difference-in-Differences with Geocoded Microdata

Kyle Butts\*

JULY 28, 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). This paper formalizes the necessary assumptions to identify the average treatment effect among the effected units and illustrates potential pitfalls 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 non-parametric estimation using partitioning-based least squares developed in [Cattaneo et al. \(2019a\)](#). This method allows for researchers to estimate how treatment effects evolve over distance. Lastly, I illustrate the advantage of this method by revisiting the effects of increased crime risk on home values studied in [Linden and Rockoff \(2008\)](#).

JEL-Classification: C13, C14, C18

Keywords: Spatial, Difference-in-Differences

---

\*University of Colorado, Boulder. Email: [kyle.butts@colorado.edu](mailto: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 some specific treatment at a very granular level. What are the effects of highly localized pollutants on child health? Does being 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 provides an improved estimator which requires fewer 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 marked by circles are considered treated due to their proximity to the treatment location, units 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 outcomes will approximately be equal between both rings. The ring method 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.

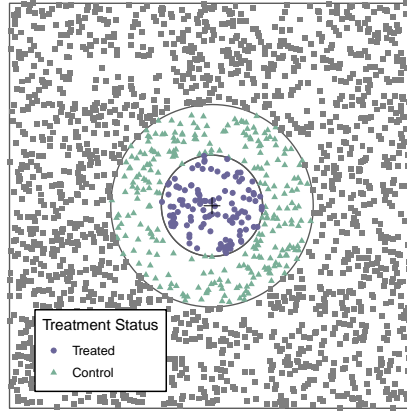
The first contribution of the paper is to fill a gap in the econometrics literature by formalizing the necessary assumptions for unbiased estimates of the average treatment effect on the affected units. The first assumption is the well understood parallel trends assumption for the treated and control units. This allows for average changes in the control units to estimate the counterfactual trend for the treated units.

The second assumption requires that the radius of the treated ring must be correctly specified as the maximum distance that experiences effects of treatment. 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 0 treatment effect of some unaffected units are averaged into the change among ‘treated’ units. Therefore results will be attenuated towards zero.<sup>1</sup>

Since this second assumption of how far treatment effects extend is difficult to know in most circumstances, I propose a method that replaces this assumption with a more mild assumption by using nonparametric partitioning-based least square regressions ([Cattaneo et al., 2019b,a](#)). My proposed methodology estimates the treatment effect curve rather than trying to estimate an average affect. There are many

1. This is equivalent to randomly assigning treated status to a portion of control units in an experiment.

**Figure 1 — Rings Method**



advantages to this approach. First, the new assumption required is much more mild as it only requires that treatment effects become zero somewhere between the distance of 0 and the control ring which does not have to be known by the researcher. Second, this 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, potentially a new bus-stop creates costs to immediate neighbors while provides benefits for homes slightly further away. Estimation of the treatment effect curve can illustrate these different effects that the “overall effect” would mask over. Lastly, this approach allows visual evidence to plausibly validate the parallel trends assumption at the heart of this method, akin to the pre-trends test in event studies.

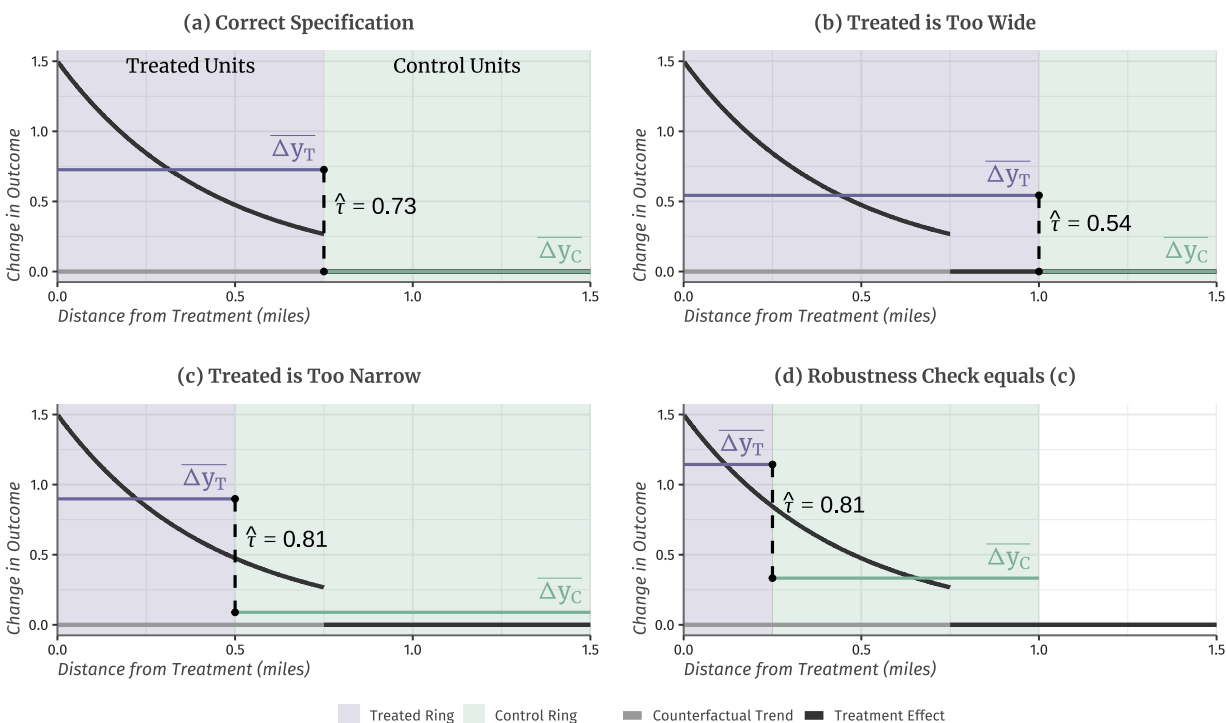
### *1.1. Relation to Literature*

## **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 trend in home values. In this simple example,

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



*Notes:* This figure shows an example of the ring method. 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.

counterfactual trends are assumed to be constant within 1.5 miles from the treated lot. 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 non-parametrically estimating the treatment effect as a function of distance rather than using a single indicator.

However, the researcher does not typically know the distance at which treatment effects stop. Panels (b) and (c) highlights how treatment effect estimates change with a change in ring distances. Panel (b) shows when the 'treatment' ring is too wide. In this case, some of the units in the treatment ring receive no effect

from treatment and therefore makes the average treatment effect among units in the treatment ring smaller. 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 you 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.

This 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 3.1 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 is a problem. If Panel (c) was the researchers’ original specification and Panel (d) was run as a robustness check, then the researcher would be quite confident in their results even though the estimate is too large in both cases.

### 3— Theory

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  $\text{Dist}_i \equiv \sqrt{(x_i - \bar{x})^2 + (y_i - \bar{y})^2}$  with a distribution function  $F$ . Since treatment location is often chosen strategically, potential outcomes will have to reflect the fact that trends can change with distance to treatment. Outcomes are given by

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

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

Taking first-differences 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. Often times re-

searchers claim that counterfactual trends likely evolve smoothly over distance, so that  $\lambda(\text{Dist}_i)$  is approximately constant within a certain distance from treatment. This is formalized in the context of our outcome model by the following assumption.

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

This assumption requires that, in the absence of treatment, outcomes would evolve the same at every distance from treatment within a certain maximum distance,  $\bar{d}$ . If this assumption 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  $\lambda$  is some constant. Therefore, the treatment effect curve  $\tau(\text{Dist}_i)$  is identifiable up to a constant under Assumption 1. To identify  $\tau(\text{Dist}_i)$  separately from the constant, researchers will often claim that treatment effects stop occurring before  $\bar{d}$  and therefore there are valid ‘control units’ in the sample. This is formalized in the following assumption.

**Assumption 2** (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)  $F(\bar{d}) - F(\bar{d}_t) > 0$*

The second part of this assumption just requires that there are observation in the control ring.

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

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

From our potential outcome framework and standard results for regressions involving only indicators, we have the following proposition.<sup>2</sup>

**Proposition 1** (Decomposition of Ring Estimate).

(i) *The estimate of  $\beta_1$  in (1) has the following expectation:*

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

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

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

2. A similar derivation of part (i) is found in Sullivan (2017).

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

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

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

Part (ii) says that if  $d_c$  satisfies Assumption 1, then the counterfactual trend 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,  $\mathbb{E} [\tau(\text{Dist}) | \mathcal{D}_t]$  will be smaller than  $\bar{\tau}$  while  $\mathbb{E} [\tau(\text{Dist}) | \mathcal{D}_c]$  would be equal to zero. Therefore,  $\hat{\beta}_1$  will be biased towards zero if  $d_t$  is too wide. Second, if  $d_t$  is too narrow then the  $\mathcal{D}_c$  will contain units that experience treatment effects. It is not clear in this case, though, if  $\hat{\beta}_1$  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 2 is a very demanding assumption and unlikely to be known by the researcher unless there are a priori theory dictating  $d_t$ .<sup>3</sup> The following section will improve estimation by allowing non-parametric identification of the entire  $\tau(\text{Dist})$  function. An estimate of  $\tau(\text{Dist})$  can then be numerically integrated to for an estimate of  $\bar{\tau}$ .

### 3.1. Estimation of $\tau(\text{Dist})$

In this section, I propose an estimation strategy that non-parametrically identifies the treatment effect curve  $\tau(\text{Dist}_i)$ . To do this, I use partitioning-based least squares estimation and inference methods developed in Cattaneo et al. (2019b,a). Partition based estimators separate the support of a covariate  $x$ , such as  $\text{Dist}$ , into a set of quantile-spaced rings (e.g. 0-25th percentiles of  $X$ , 25-50th percentiles, 50-75th, 75-100th) and estimate a polynomial within each interval. In this way, any function can be approximated nonparametrically as the number of intervals and the degree of the polynomial increase asymptotically.

3. As an example of a priori theory dictating  $d_t$ , Currie et al. (2015) uses results from chemical studies on the maximum spread of local pollutants and Marcus (2021) use the plume length of petroleum smoke.

For a given partition based on quantiles of the data  $\{\mathcal{D}_1, \dots, \mathcal{D}_k\} \equiv \mathcal{D}$ , we will estimate the conditional average of  $\Delta Y_{it}$  for each interval  $\mathcal{D}_i$ . These averages are defined as

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

This will form our estimate  $\tau(\text{Dist}_i) + \lambda$  under local parallel trends, Assumption 1. To remove  $\lambda$ , we require an Assumption like 2.

**Assumption 3** ( $d_t$  is within  $\bar{d}$ ). *A distance  $\bar{d}$  satisfies this assumption if there is a distance  $d_t$  with  $0 < d_t < \bar{d}$  such that (i) Assumption 2 holds and (ii)  $F(\bar{d}) - F(d_t) > 0$ .*

Under Assumptions 1 and 3, the mean within the last ring  $\mathcal{D}_k$  will estimate  $\lambda$  as the number of bins  $k \rightarrow \infty$ . The reason for this is simple, as  $k \rightarrow \infty$ , the last bin will have the left end-point  $d > d_t$  and therefore  $\tau(\text{Dist}) = 0$  in  $\mathcal{D}_k$ . Therefore, estimates of  $\tau(\text{Dist}_i)$  can be formed for each interval as  $\bar{\tau}_j \equiv \overline{\Delta Y}_j - \overline{\Delta Y}_k$ . The results of Cattaneo et al. (2019a) show the large-sample asymptotics of the estimates  $\bar{\tau}_j$  which allow valid inference to be performed on these estimates.

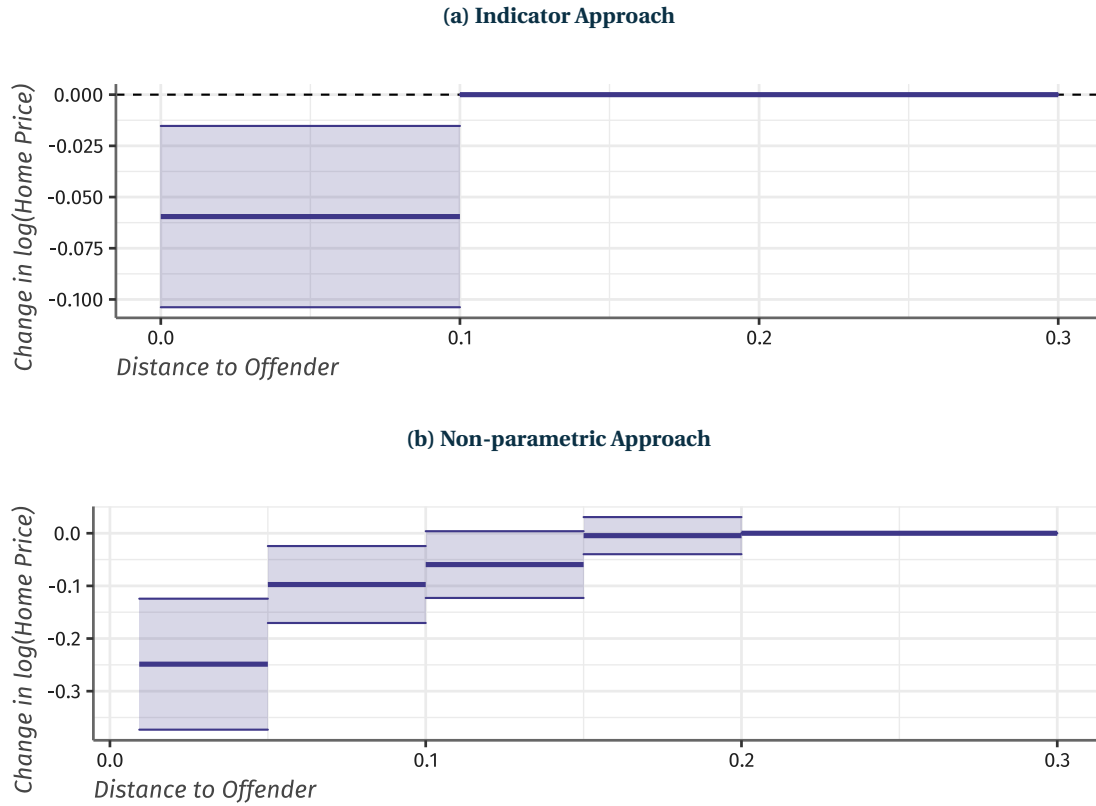
@TODO: Advantages of this method from introduction

## 4 — Application to Neighborhood Effects of Crime Risk

@TODO: Write up of Linden and Rockoff



**Figure 3 — 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. Panel (a) shows an estimate from Equation 1 with a treatment distance of 0.1 miles and a control distance of 1/3 mile. Panel (b) shows the non-parametric estimate of  $\tau(\text{Dist}_i)$  proposed in Section 3.1.

## References

- 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.
- 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](#).

**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](#).

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