Causal Inference for Spatial Treatments

Michael Pollmann*

January 24, 2023

Abstract

Many events and policies (treatments) occur at specific spatial locations, with researchers interested in their effects on nearby units of interest. I approach the *spatial treatment* setting from an experimental perspective: What ideal experiment would we design to estimate the causal effects of spatial treatments? This perspective motivates a comparison between individuals near realized treatment locations and individuals near counterfactual (unrealized) candidate locations, which differs from current empirical practice. I derive design-based standard errors that are straightforward to compute irrespective of spatial correlations in outcomes. Furthermore, I propose machine learning methods to find counterfactual candidate locations using observational data under unconfounded assignment of the treatment to locations. I apply the proposed methods to study the causal effects of grocery stores on foot traffic to nearby businesses during COVID-19 shelter-in-place policies, finding a substantial positive effect at a very short distance, with no effect at larger distances.

*Department of Economics, Duke University. E-mail: michael.pollmann@duke.edu.

I am grateful to my advisor, Guido Imbens, for invaluable encouragement and guidance. I am thankful to Luis Armona, Paul Goldsmith-Pinkham, Christian Hansen, Caroline Hoxby, Joshua Kim, Matt Masten, Áureo de Paula, Daniel Pollmann, Fredrik Sävje, Jann Spiess, Melanie Wallskog, as well as numerous seminar participants for many comments and insightful discussions. This research was supported generously by the B.F. Haley and E.S. Shaw Fellowship for Economics through a grant to the Stanford Institute for Economic Policy Research. This paper uses data from SafeGraph. SafeGraph is a data company that aggregates anonymized location data from numerous applications in order to provide insights about physical places.

1 Introduction

Many actions, events, and policies studied by economists occur at locations in space and affect (geographically) nearby units or individuals.[1] I refer to such studies' setting as the "spatial treatment" setting because the level at which these "treatments" vary are locations in space. The researcher studies the effects of such treatments on individuals who are located in the vicinity of these treatments but who are conceptually distinct units. In contrast, in most of the theoretical literature in causal inference, each individual is thought to in principle be associated with their own treatment generating potential outcomes, with some work considering "spillovers" and "clustered assignment" of individual-level treatments. Such a framework was largely sufficient when both treatment and outcome information was only available aggregated at, for instance, the county level. More recently, however, precise (geocoded) location data for treatments and individuals have become more readily available, allowing more informative analyses of the disaggregated effects of spatial treatments by distance from treatment.

This paper makes three contributions. First, I develop a framework that allows me to formalize ideal experiments and analyze questions of causal inference in spatial settings from a design-based perspective. Second, I show that this design-based perspective is tractable and useful by deriving (approximately) unbiased inverse probability weighting estimators and new expressions for their variances, which differ from commonly used existing estimators and sampling-based variances. Third, I propose using convolutional neural networks, previously used for image and satellite data, to parsimoniously condition on the distribution of covariates across space in analyses with observational data under an assumption of unconfounded treatment assignment.

These contributions yield succinct practical recommendations that answer three key methodological questions in spatial treatment settings. First, when studying the effects of a spatial treatment on individuals who are distance d away from it, who should be in the control group? Second, what is the standard error of the resulting

[1] Examples include the effects of: businesses' location decisions on local competitors, workers, or consumers; schools, hospitals, or sources of pollution on education, income, and health of nearby residents; low-income housing, local public goods, or crime risk on property values; centrally administered treatments such as deworming in schools or COVID-19 vaccination centers on treatment uptake and effectiveness. See Online Appendix Table OA1 for examples of papers studying these and other spatial treatments.

estimator under the ideal experiment? Third, how can we mimic the analysis of the ideal experiment when only observational data are available?

The ideal experiment to study the effects of a spatial treatment on nearby individuals randomizes the location of the treatment among plausible candidate locations. Such an ideal experiment is implicitly invoked when researchers argue that the location of the treatment is quasi-random, for instance, due to the exogenous (un-) availability of candidate locations at the time the treatment is implemented.[2]

I show that the control group of a commonly used estimator is not valid under this ideal experiment and discuss how to construct an alternative, valid, control group. Much of the empirical literature studying spatial treatments with observational data effectively compares individuals at the distance of interest from the treatment (on an "inner ring") to individuals farther away but centered around the same treatment location (on an "outer ring"). Yet the ideal experiment (or similar quasi-random variation in the locations of the treatment) does not directly justify this comparison: Even with random variation in treatment locations, this comparison is invalid for all but knife-edge scenarios for the surface of potential outcomes across space. Instead, the ideal experiment justifies comparing individuals on the inner ring around realized treatment locations to individuals on inner rings around those locations where the treatment could have been but is not by random chance alone.

I derive inverse probability weighting estimators and their (approximate) finite population unbiasedness and design-based variance under the ideal experiment and show asymptotic normality in a leading case. The repeated sampling thought experiment of the finite population analysis holds fixed the individuals in the population, their locations, and their potential outcomes, varying only the realized locations of the treatment among a pre-defined set of plausible candidate locations. Inverse probability weighting allows nonparametric estimation of the average effect of the treatment on individuals who are, say, distance $d \pm h$ away from candidate treatment locations, as well as other (for instance, kernel) weighted average effects.[3] The design-based analysis of the variance following Neyman (1923) has both conceptual and practical

[2]For instance, Linden and Rockoff (2008, p. 1110) argue that "the nature of the search for housing is also a largely random process at the local level. Individuals may choose neighborhoods with specific characteristics, but, within a fraction of a mile, the exact locations available at the time individuals seek to move into a neighborhood are arguably exogenous."

[3]See Online Appendix 8 for parametric estimators under correct specification of the treatment effect by distance.

advantages over other sampling-based alternatives (for instance, Conley, 1999): The design-based variance reflects the variation that the researcher exploits when claiming causality of estimated effects by appealing to "quasi-random" variation. When a researcher describes design-based variation, they do not need to distinguish between sample and population, which may be difficult to justify in spatial settings (see Pinkse et al., 2007). Estimating the design-based standard error is straightforward. The researcher does not need to correctly estimate or model the correlation of the outcome across space – a task that is often tangential to the research question. Importantly, my approach also allows me to derive standard errors for settings where individuals are exposed to multiple treatments. In these settings, off-the-shelf alternatives such as "clustering at the level of the assignment" (Abadie et al., 2023) are not applicable.

In the second half of the paper, I focus on a "spatial unconfoundedness" assumption for observational data as the analog to randomization in the ideal experiment. Suppose locations in two neighborhoods look identical in terms of their observable pre-treatment characteristics. Then the spatial unconfoundedness assumption requires that which of the two locations the treatment is realized in does not depend on the potential outcomes of individuals in the neighborhoods.

To implement flexible estimation based on spatial unconfoundedness, I propose using convolutional neural networks in a way that may be of independent interest for settings with spatial data. Researchers can plot many economic data, such as locations of businesses, property prices, school district quality, and the average income by census tract, on maps. The distribution of spatial covariates across space often encodes otherwise latent information. In contrast, coarse summary statistics such as the number of units or the average value of a covariate in a circle around a location may not fully capture the local economic environment. The spatial unconfoundedness assumption can therefore be more credible when researchers use the information contained in spatial data more fully. However, controlling for the distribution of units or covariates across space relative to the location of estimation intrinsically is an extremely high-dimensional problem. At the same time, economic and institutional knowledge often suggests equivariances: If all units and covariates in a neighborhood are shifted, their locations mirrored, or their orientation rotated along some axis equally, economically the neighborhood remains unchanged, and the location of the predicted outcome should simply be shifted, mirrored, and rotated analogously. I propose convolutional neural networks that parsimoniously condition on the distribution of

covariates across space and are automatically equivariant to shift, and I show how to build in other equivariance using data augmentation when training the network. Researchers may find convolutional neural networks useful whenever estimation is more credible conditioning on covariate values in the neighborhood around the unit of estimation, for instance, demographics of nearest and farther neighbors. In the spatial treatment setting, I use such networks to find plausible counterfactual locations of the treatment that are observationally similar to realized treatment locations.

I apply the proposed methods to study whether grocery stores caused additional visitors to nearby restaurants during COVID-19 shelter-in-place policies. During shelter-in-place policies in the San Francisco Bay Area in April 2020, residents were only allowed to make essential trips, for instance, to get groceries. As mobility was greatly reduced, restaurants may have benefited from being located near points still frequented by consumers (grocery stores): Consumers may find it convenient to grab a coffee or snack while waiting in line to get into the store or to pick up a takeout order before returning home. Using convolutional neural networks, I identify counterfactual grocery store locations that are in neighborhoods with business compositions and relative locations similar to the neighborhoods of real grocery stores.

I find that restaurants within a couple of minutes walking from real grocery stores had about twice as many visitors as restaurants at the same distance from counterfactual locations. There is no such difference in visitors at longer distances. For a causal interpretation, the unconfoundedness assumption requires that restaurants in neighborhoods with similar business composition and relative locations except differing by one grocery store do not have systematically different potential outcomes.

A nascent methodological literature studies causal inference in spatial treatment and related settings. Zigler and Papadogeorgou (2021) and, contemporaneously to the present paper, Aronow et al. (2020) set up potential outcomes frameworks and estimands for experimental settings similar to those in the present paper, and Borusyak and Hull (2020) take a similar design-based perspective but apply it to a regression framework. The approach of Borusyak and Hull (2020) has the advantage of accommodating multi-valued and other more complicated treatments, but the estimands of coefficients in their (unweighted) regressions differ under treatment effect heterogeneity. These papers do not explicitly estimate design-based standard errors in their applications, however, instead reporting Conley (1999) standard errors.

The present paper contributes to the literature by showing that design-based

inference, beyond identification, is conceptually attractive, analytically tractable, and computationally straightforward. Furthermore, I propose a data-driven method for inferring a plausible counterfactual distribution of the treatment under unconfoundedness using neural networks, while prior work requires the researcher to specify it based on institutional knowledge. Similar to Borusyak and Hull (2020), the methodological contributions of the present paper are not restricted to spatial settings. The ideas, results, and proposed methods are more generally applicable to settings where treatments are separate from the units for which outcomes are measured, rather than directly assigned to them. I discuss examples in the extensions. Spatial treatments share some resemblance to spillover effects of treatments in networks (for instance, Athey et al., 2018; Leung, 2020), and some results from one setting can be applied to special cases of the other. However, typical networks cannot be projected into low-dimensional Euclidean space, and typical settings with units in geographic space would require dense networks with weighted edges that may not be tractable analytically or may violate typical assumptions on the network structure. Consequently, the appropriate implementations of estimation under unconfoundedness differ (for instance, Leung and Loupos, 2022, in network settings).

The remainder of this paper proceeds as follows. Section 2 shows that the ideal experiment does not generally justify the inner vs. outer ring empirical strategy. Section 3 describes the framework and notation of this paper. Section 4 contains estimation and inference results under the ideal experiment. Section 5 introduces the spatial unconfoundedness assumption and describes estimation of counterfactual locations using convolutional neural networks.[4] Section 6 discusses extensions to non-spatial settings. Section 7 illustrates the use of the proposed methods with the empirical application. Section 8 concludes.

[4]A documented code tutorial implementing the approach using convolution neural networks is available at `https://github.com/michaelpollmann/spatialTreat-example`, in addition to the replication code accompanying this paper.

2 The inner vs. outer ring empirical strategy requires assumptions beyond quasi-random variation in treatment locations

Many recent empirical studies estimate the effects of spatial treatments at a distance of d by comparing the outcome of individuals d away from treatment to the outcome of individuals $D \gg d$ away from treatment.[5] Visually, the individuals d from treatment are located on an "inner ring" around the treatment, while individuals D away from treatment are located on an "outer ring." Hence, this strategy is commonly referred to as an "inner ring vs. outer ring" strategy.

I use an example to show that this estimation strategy relies on "functional form" assumptions about potential outcomes rather than "design" assumptions about the ideal experiment. For some locations of outcome units and their potential outcomes, the inner vs. outer ring estimator is inconsistent under *all* "ideal experiments" that generate random variation in treatment location. In other words, one cannot justify this strategy purely based on the ideal experiment; alternative assumptions are necessary. I present the cross-sectional example first, and then discuss the "parallel trends" assumption as one such alternative.

Some economists have the intuition that if the treatment locations are completely random (uniformly distributed and independent of potential outcomes), and one correctly specifies a distance after which the treatment has no effect, the inner vs. outer ring comparison yields a consistent estimator. The intuitive argument posits that because randomization renders all individuals equally likely to be on the inner vs. outer ring, the two groups are similar in expectation. Consequently, the estimator is unbiased, but possibly inefficient by restricting the control group to only a ring rather than all unaffected individuals. Hence, it appears worthwhile to illustrate with a simple example that this intuition is incorrect.

The setup for this section is as follows. Suppose the researcher has access to data from many cities that are sufficiently far apart for treatment in one city to not affect outcomes in another city. For simplicity, within each city consider locations in one-dimensional integer space, $\mathbb{Z}$, only. Suppose the potential outcome of individual i

[5] Typically, the comparison across individuals is combined with a comparison across time. As discussed at the end of this section, the across-time comparison conceptually does not affect identification in the ideal experiment.

in the absence of treatment is $Y_i(0) = r_i^2$ where $r_i \in \mathbb{Z}$ is the location of i. Further, suppose that exactly one individual is located at each integer. Let $d \in \mathbb{Z}_+$ be the inner ring distance, and let $D \in \mathbb{Z}_+$ with $D > d$ be an outer ring distance such that individuals D away from treatment are unaffected.

Under this setup, the inner vs. outer ring strategy yields systematically biased estimates irrespective of how treatment locations are chosen. Specifically, if treatment in a city occurs in location $S \in \mathbb{R}$, then the average (counterfactual) control potential outcome $Y_i(0)$ of the two individuals on the inner ring, at distance d, is $((S+d)^2 + (S-d)^2)/2 = S^2 + d^2$. The average control potential outcome of the two individuals on the outer ring, who are used as the control group, analogously is $S^2 + D^2$. This outcome is observed by the assumption that at distance D treatment effects are 0. Hence, the inner vs. outer ring strategy overestimates the mean of the control potential outcomes of individuals at distance d by $D^2 - d^2 > 0$, irrespective of the location of the treatment, S. That is, no matter how the location of the treatment is chosen, whether randomly or endogenously, uniformly or non-uniformly distributed, the inner vs. outer ring estimator is biased in this example even with infinite data and correct specification of a distance at which there are no effects.

This example shows that the inner vs. outer ring empirical strategy requires additional assumptions beyond the quasi-random variation induced by the ideal experiment. Random variation in the locations of treatments, by itself, cannot guarantee that the strategy yields consistent estimates. Instead, the researcher must *assume* that the outer ring control group correctly estimates the average control potential outcomes of the treated individuals. Effectively, except for knife-edge cases of offsetting biases, researchers need to assume that the outcome surface is flat: If the outcome surface is flat, any distribution of individuals yields unbiased estimates. If, in contrast, the average outcomes on inner and outer rings are not equal, only particular weighting (given by the distribution of individuals across space) of the outer ring *may* (if the inner ring average is a convex combination of outer rings values) be able to yield unbiased estimates. Hence, the validity of the estimator requires a functional form assumption that remains necessary even asymptotically. The assumption may be most plausible when $D - d$ is small such that the control group is located near the treated group. At the same time, however, D needs to be large enough for the outer ring individuals to plausibly be unaffected by the treatment unless different biases offset perfectly.

Taking empirical derivatives of the outcome with respect to distance from treatment and then integrating over distance, as in Diamond and McQuade (2019), is subject to the same conceptual issue. Consider again the example above with outcomes quadratic in location but smaller gaps between individuals such that empirical derivatives (approximately) equal the actual derivative of outcomes: $\partial Y(0)/\partial r = 2r$. Integrating these derivatives between $S+d$ and $S+D$ as well as between $S-d$ and $S-D$ yields

$$\left(\int_d^D 2(S+v)\mathrm{d}\,v + \int_{-d}^{-D} 2(S+v)\mathrm{d}\,v\right)/2 = ((D^2-d^2)+((-D)^2-(-d)^2)/2 = D^2-d^2,$$

which is the same comparison as for the inner vs. outer ring estimator. The key methodological contribution of Diamond and McQuade (2019) is to develop a computationally feasible estimator that correctly calculates the derivative and performs the integration in settings where space is multi-dimensional and the distribution of individuals is sparse and possibly not uniform. The approach still relies on a correctly specified function form (Diamond and McQuade, 2019, equation 1) that captures the conceptual essence of the inner vs. outer ring strategy.

Using the inner vs. outer ring strategy with panel data in a difference-in-differences approach, as is common in practice, does not generally resolve the conceptual issue raised in this section. The argument made about the levels of potential outcomes applies equally to the trends in potential outcomes. An absence of pre-trends may be suggestive of the required absence of differential trends also post-treatment. However, when researchers estimate the effect of the treatment at not just a single distance d but also at other distances d_2, d_3, etc., often using the same outer ring control group at distance D, assessing the absence of pre-trends (or magnitude of potential violations) becomes more challenging than in standard difference-in-differences analyses due to the large number of estimates and their correlations.

The inner vs. outer ring strategy requires functional form assumptions (such as parallel trends) rather than the variation induced by an ideal experiment of randomized treatment locations. If the researcher wishes to use this strategy, they should motivate those functional form assumptions.

If the researcher instead believes that there is quasi-random variation in the location of treatments, the estimators and theory developed in this paper are applicable. In practice, empirical studies providing evidence that distinct sources of variation, identification strategies, or assumptions, yield similar estimates may be

most convincing.

3 Setup and notation

Both individuals (outcome units) and treatments are located in a shared (geographic) space. Individuals, indexed by $i \in \mathbb{I}$, have fixed location, or residence, $r_i \in \mathbb{R}^2$ such as latitude and longitude.[6] In contrast to the standard setting of causal inference, treatments do not share the same index i with individuals. Instead, the treatment takes values $S \subset \mathbb{R}^2$ corresponding to locations in the same space as the individuals.

Each individual has a potential outcome $Y_i(S)$ for each S, and treatment effects are contrasts between different potential outcomes. The natural individual-level treatment effect compares the outcome of i when there is treatment at location s vs. no treatment at s, holding fixed treatments at other locations: $\tau_i(s \mid S) \equiv Y_i(S \cup \{s\}) - Y_i(S \setminus \{s\})$. $\tau_i(s \mid S)$ is a marginal effect with background exposure $S \setminus \{s\}$. Of particular interest is the treatment effect of s when there is no other (relevant) treatment: $\tau_i(s) \equiv \tau_i(s \mid \{s\}) = Y_i(\{s\}) - Y_i(\emptyset)$. For ease of notation, define $Y_i(s) \equiv Y_i(\{s\})$ and $Y_i(0) \equiv Y_i(\emptyset)$.

The experimental design generates randomness in where the treatment is realized. I use calligraphic letters to denote random variables in contrast to roman letters used for fixed values. The realized treatment locations are $\mathcal{S} \subset \mathbb{R}^2$, such that the observed outcome for individual i is $\mathcal{Y}_i \equiv Y_i(\mathcal{S})$. Let $\pi_s \equiv \Pr(\mathcal{S} \ni s)$ be the experimental probability of treatment at location s. The term *candidate treatment locations* ($\mathbb{S}$) refers to locations $s \in \mathbb{S} = \{s \in \mathbb{R}^2 : \pi_s > 0\}$, such that $\mathcal{S} \subset \mathbb{S}$.

I state the notation and results in this paper in terms of cross-sectional data only. If the researcher has access to panel data, all results remain unchanged under the same ideal experiment after subtracting the corresponding pre-treatment outcome from each (potential) outcome.

The researcher is interested in the average effects of treatments on individuals who are a specific *distance* away. I denote the distance between s and r_i by $d(s, r_i)$. The researcher chooses the distance function which is meaningful in their application such as "straight line distance" or driving time during rush hour. Importantly, the distance must not vary with the presence or absence of treatment; it must be a "pre-treatment characteristic."

[6] It is not essential that locations are in two-dimensional space.

The estimand of interest is the expected (over the design distribution) average effect of the treatment on the treated (ATT) at distance $d \pm h$. Researchers often bin individuals within a bandwidth h around d together when distance is a continuous variable. Distance bin weights $w_i(s,d) \equiv \mathbb{1}\{|d(s,r_i) - d| \leq h\}$ indicate if i is in the bin around s.[7] Of primary interest is a contrast between one treatment and no treatment:

$$\tau(d) \equiv \frac{\sum_{s \in \mathbb{S}} \Pr(\mathcal{S} \ni s) \sum_{i \in \mathbb{I}} w_i(s,d) \tau_i(s)}{\sum_{s \in \mathbb{S}} \Pr(\mathcal{S} \ni s) \sum_{i \in \mathbb{I}} w_i(s,d)}. \tag{1}$$

Estimating the effect of one treatment compared to no treatment is impractical in some settings because multiple treatments are observed even in small areas. Instead, the researcher may focus on an average *marginal* effect of the treatment on the treated at d:

$$\tau_{\text{marginal}}(d) \equiv \frac{\sum_{S \in 2^{\mathbb{S}}} \Pr(\mathcal{S} = S) \sum_{s \in S} \sum_{i \in \mathbb{I}} w_i(s,d) \tau_i(s \mid S)}{\sum_{S \in 2^{\mathbb{S}}} \Pr(\mathcal{S} = S) \sum_{s \in S} \sum_{i \in \mathbb{I}} w_i(s,d)}. \tag{2}$$

This effect aggregates the marginal effects of location s given all possible background exposures $S \setminus \{s\}$. The weights again resemble the ATT, placing more weight on assignments that are more likely to be realized.

4 Experimental data: estimation and inference

In this section, I discuss estimators of average treatment effects on the treated (ATT) for two settings that are particularly relevant in practice. In the first setting, the researcher has data for separate regions, defined such that treatment in one region does not affect outcomes in other regions, with at most one realized treatment location per region. In the second setting, all data are for a single large region with multiple realized treatment locations as well as unrealized, counterfactual, treatment locations.

4.1 Separate regions

Suppose the researcher collects data from separate regions (or markets) $j = 1, \ldots, J$, formalized in Assumption 1 below. This setting simplifies estimation and inference,

[7] See Online Appendix 4 for weights other than the ATT. Instead of distance bins, any other kernel weighting is possible with straightforward modifications, but bins are most common in practice. In this paper, binning corresponds to the desired estimand rather than a kernel used to estimate a function at a point. For simplicity of the results, I assume $w_i(s,d) \geq 0$ throughout.

and allows me to highlight the nature and interpretation of the theoretical results.

For ease of notation, let $\mathbb{I}_j$ and $\mathbb{S}_j$ denote the individuals and candidate treatment locations in region j, respectively, with $\{\mathbb{I}_j\}_{j=1}^J$ and $\{\mathbb{S}_j\}_{j=1}^J$ forming partitions of $\mathbb{I}$ and $\mathbb{S}$. Denote by $\mathcal{W}_j \equiv \max_{s\in\mathbb{S}_j} \mathbb{1}\{\mathcal{S} \ni s\}$ whether at least one location in region j is treated. The probability of this event is $\pi_j \equiv \Pr(\mathcal{W}_j = 1)$. The probability of treatment at location $s \in \mathbb{S}_j$ conditional on treatment somewhere in region j is $\pi_j(s) \equiv \Pr(\mathcal{S} \ni s \mid \mathcal{W}_j = 1)$. By definition, $\pi_s = \pi_j \pi_j(s)$.

Assumption 1 (Separate Regions)**.** *The data $(\mathbb{I}, \mathbb{S})$ can be partitioned into regions $\{(\mathbb{I}_j, \mathbb{S}_j)\}_{j=1}^J$, such that an individual is unaffected by treatments in other regions: For all $S \subset \mathbb{S}$, if $i \in \mathbb{I}_j$ and $s \in \mathbb{S}_{j'}$ with $j \neq j'$ then $Y_i(S) = Y_i(S \setminus \{s\})$.*

Assumption 2 (Assignment Across Regions)**.** *Treatments are assigned across regions according to a completely randomized design where each region has equal marginal probability of receiving treatment somewhere; $\pi = \pi_j$ for all regions j. That is, all assignment vectors $W \in \{0,1\}^J$ with $\sum_j W_j = J_t \equiv \pi J$ are equally likely, and assignments with $\sum_j W_j \neq \pi J$ have zero probability:*

$$\Pr(\mathcal{W} = W) = \begin{cases} \binom{J}{J_t}^{-1} & \text{if } \sum_{j=1}^J W_j = J_t \\ 0 & \text{otherwise.} \end{cases}$$

Conditional on treatment in region j, assignment to a particular location within the region is independent of assignment in other regions j'. For all $s \in \mathbb{S}_j$ and $s' \in \mathbb{S}_{j'}$ with $j \neq j'$:

$$\mathbb{1}\{\mathcal{S} \ni s\} \perp\!\!\!\perp \mathbb{1}\{\mathcal{S} \ni s'\} \mid \mathcal{W}_j = 1, \mathcal{W}_{j'} = 1.$$

Assumption 3 (One Treatment Per Region)**.** *At most one treatment is realized in each region: For $j = 1, \ldots, J$: $\sum_{s\in\mathbb{S}_j} \mathbb{1}\{\mathcal{S} \ni s\} \leq 1$ with probability 1.*

Under Assumptions 1 and 2, one can rewrite the ATT as

$$\tau(d) \equiv \frac{\sum_{j=1}^J \pi_j \sum_{s\in\mathbb{S}_j} \pi_j(s) \sum_{i\in\mathbb{I}_j} w_i(s,d)\tau_i(s)}{\sum_{j=1}^J \pi_j \sum_{s\in\mathbb{S}_j} \pi_j(s) \sum_{i\in\mathbb{I}_j} w_i(s,d)}. \tag{3}$$

I focus on inverse probability weighting estimators of the ATT that take the form of a weighted difference in means. The mean of the treated is the simple average of

individuals at the distance of interest:

$$\bar{\mathcal{Y}}_t(d) \equiv \frac{\sum_{j=1}^{J} \mathcal{W}_j \sum_{s \in \mathbb{S}_j} \mathbb{1}\{\mathcal{S} \ni s\} \sum_{i \in \mathbb{I}_j} w_i(s,d) \mathcal{Y}_i}{\sum_{j=1}^{J} \mathcal{W}_j \sum_{s \in \mathbb{S}_j} \mathbb{1}\{\mathcal{S} \ni s\} \sum_{i \in \mathbb{I}_j} w_i(s,d)}$$

while the mean of the control is based on all individuals in untreated regions who are at the distance of interest from a candidate location and weights them to match the ATT weights:

$$\bar{\mathcal{Y}}_c(d) \equiv \frac{\sum_{j=1}^{J} \frac{1-\mathcal{W}_j}{1-\pi_j} \pi_j \sum_{s \in \mathbb{S}_j} \pi_j(s) \sum_{i \in \mathbb{I}_j} w_i(s,d) \mathcal{Y}_i}{\sum_{j=1}^{J} \frac{1-\mathcal{W}_j}{1-\pi_j} \pi_j \sum_{s \in \mathbb{S}_j} \pi_j(s) \sum_{i \in \mathbb{I}_j} w_i(s,d)}$$

such that the difference in means estimator is

$$\hat{\tau}(d) \equiv \bar{\mathcal{Y}}_t(d) - \bar{\mathcal{Y}}_c(d). \tag{4}$$

The next theorem states that the estimator $\hat{\tau}(d)$ is (approximately) unbiased for the ATT and gives its (approximate) finite population variance. This variance depends on the following variances of potential outcomes and individual-level treatment effects aggregated by location or by region:

$$\begin{aligned}
\tilde{V}_t^{\text{location}}(d) &\equiv \frac{1}{J-1} \sum_{j=1}^{J} \sum_{s \in \mathbb{S}_j} \pi_j(s) \Big(\sum_{i \in \mathbb{I}_j} \frac{w_i(s,d)}{\bar{n}(d)} (Y_i(s) - \mu_t(d)) \Big)^2 \\
\tilde{V}_c^{\text{region}}(d) &\equiv \frac{1}{J-1} \sum_{j=1}^{J} \Big(\sum_{s \in \mathbb{S}_j} \pi_j(s) \sum_{i \in \mathbb{I}_j} \frac{w_i(s,d)}{\bar{n}(d)} (Y_i(0) - \mu_c(d)) \Big)^2 \\
\tilde{V}_t^{\text{region}}(d) &\equiv \frac{1}{J-1} \sum_{j=1}^{J} \Big(\sum_{s \in \mathbb{S}_j} \pi_j(s) \sum_{i \in \mathbb{I}_j} \frac{w_i(s,d)}{\bar{n}(d)} (Y_i(s) - \mu_t(d)) \Big)^2 \\
\tilde{V}_{ct}^{\text{region}}(d) &\equiv \frac{1}{J-1} \sum_{j=1}^{J} \Big(\sum_{s \in \mathbb{S}_j} \pi_j(s) \sum_{i \in \mathbb{I}_j} \frac{w_i(s,d)}{\bar{n}(d)} \big(Y_i(s) - Y_i(0) - (\mu_t(d) - \mu_c(d))\big) \Big)^2 \\
\bar{n}(d) &\equiv \frac{1}{J} \sum_{j=1}^{J} \sum_{s \in \mathbb{S}_j} \pi_j(s) \sum_{i \in \mathbb{I}_j} w_i(s,d)
\end{aligned}$$

where $\mu_t(d)$ and $\mu_c(d)$ are the average treated and control potential outcomes, respec-

tively, with the same weights as the ATT estimand given in Equation 3; they are defined explicitly in the appendix.

Theorem 1. *The estimator $\hat{\tau}(d)$ is similar to an infeasible estimator $\tilde{\tau}(d)$ (displayed in the appendix), which has analytically tractable non-asymptotic design-based properties:*

(i) Under regularity conditions (Appendix A.1), $\hat{\tau}(d) = \tilde{\tau}(d) + O_p(J^{-1})$.

Under Assumptions 1, 2, and 3:

(ii) unbiasedness: $E(\tilde{\tau}(d)) = \tau(d)$.

(iii) variance: $\text{var}\Big(\tilde{\tau}(d)\Big) = \frac{J-1}{J}\frac{\tilde{V}_t^{location}(d)}{J_t} + \frac{\tilde{V}_c^{region}(d)}{J_c} + \frac{1}{J}\frac{\tilde{V}_t^{region}(d)}{J_t} - \frac{\tilde{V}_{ct}^{region}(d)}{J}$.

Proof: *See Appendix A.1.*

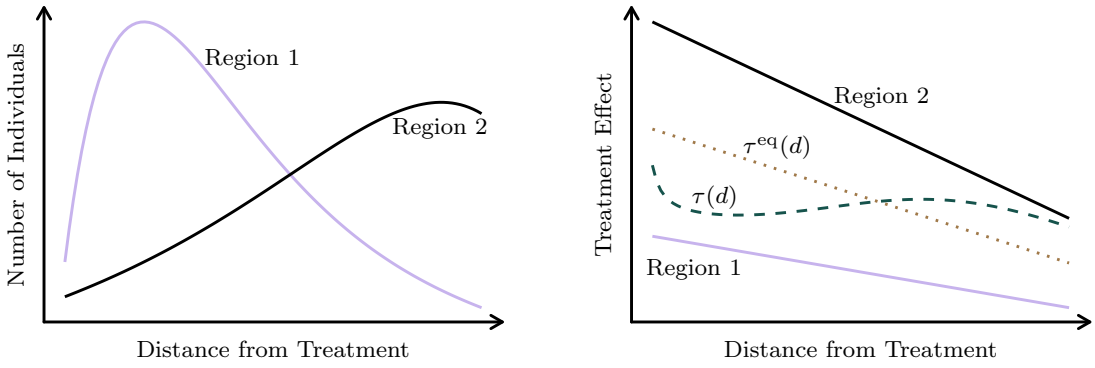
Remark 1. The variance terms $\tilde{V}_t^{\text{location}}(d)$ and $\tilde{V}_c^{\text{region}}(d)$ are straightforward to estimate with sample analogs (Online Appendix 7), and one can bound $\tilde{V}_t^{\text{region}}(d) \leq \tilde{V}_t^{\text{location}}(d)$ by Jensen's inequality. This bounding becomes negligible asymptotically due to the additional factor J^{-1} multiplying the term. $\tilde{V}_{ct}^{\text{region}}(d)$ is a variance of treatment effects that cannot be estimated consistently without strong assumptions. As is the case for the Neyman variance of the difference in means in standard randomized experiments, dropping the term yields a conservative estimator of the variance (Imbens and Rubin, 2015, ch. 6). Indeed, the variance in Theorem 1 simplifies to the familiar result if there is only one candidate location in each region and it has exactly one individual at the distance of interest. Importantly, the variance can be estimated/bounded without tuning parameters or possibly incorrect modeling of correlations between structural error terms at different distances.

Remark 2. I recommend that researchers report the square root of estimates of the variance in Theorem 1 as the (approximate) standard error of $\hat{\tau}(d)$. The approximation of $\hat{\tau}(d)$ by $\tilde{\tau}(d)$ is necessary for exact finite population results because the denominators of $\hat{\tau}(d)$ are stochastic: Depending on which candidate locations are treated, the number of individuals near treated locations may differ. $\tilde{\tau}(d)$ uses non-stochastic denominators and an appropriate re-centering of the estimator, such that the difference between the Hájek-estimator $\hat{\tau}(d)$ and the (infeasible) estimator $\tilde{\tau}(d)$ is typically much smaller than the difference with the Horvitz-Thompson estimator that only fixes the denominators. As a result, $\hat{\tau}(d)$ and $\tilde{\tau}(d)$ are very close even in small samples in simulations (Online Appendix 10).

Remark 3. The variance in Theorem 1 is for $\tilde{\tau}(d)$ as an estimator for the in-sample ATT defined in Equation 3. It relies solely on randomness due to treatment assignment, not sampling. The underlying thought experiment (repeated samples re-assign the treatment to the candidate locations) is easy to articulate and corresponds to the variation required for interpretation as a causal effect. Hence, the researcher does not need to additionally specify a hypothetical super-population and how the sample arose from it.

Remark 4. There are two variances, $\tilde{V}_t^{\text{location}}(d)$ and $\tilde{V}_t^{\text{region}}(d)$, for treated potential outcomes but only one variance, $\tilde{V}_c^{\text{region}}(d)$ for control potential outcomes. One does not need to define a variance of control potential outcomes aggregated by location (rather than region), say $\tilde{V}_c^{\text{location}}(d)$, because when a region is not treated, the researcher observes the control potential outcomes around *all* locations and can aggregate accordingly. In contrast, in treated regions, the researcher does not simultaneously observe potential outcomes for different locations being treated separately and can therefore not aggregate, such that a term involving the less aggregated $\tilde{V}_t^{\text{location}}(d)$ appears.

Remark 5. The estimator $\hat{\tau}(d)$ with distance bin weights places equal weight on all individuals at distance $d \pm h$ from a candidate treatment location (up to the ATT weights reflecting treatment probabilities). At least two alternatives may be worthwhile. First, one can apply equal weights to each treatment location, rather than individual, by taking $w_i^{\text{eq}}(s,d) \equiv \mathbb{1}\{|d(s,r_i) - d| \leq h\} / \sum_{i' \in \mathbb{I}} \mathbb{1}\{|d(s,r_{i'}) - d| \leq h\}$. These weights facilitate interpretation of the estimand across distance if there is substantial heterogeneity in population counts over distance and treatment effects by treatment location; see Figure 1 for an illustration. Inference for the estimator $\hat{\tau}^{\text{eq}}(d)$ using $w_i^{\text{eq}}(s,d)$ is straightforward based on Theorem 1 by averaging outcomes of individuals at the distance of interest by treatment location. In this case, the mean and variance results become exact for $\hat{\tau}^{\text{eq}}(d)$ because the denominator is no longer stochastic as long as each location has at least one individual at the distance of interest. Second, researchers may deviate from the distance bin weight by choosing a kernel that is continuous in distance, such as triangular weights $w_i^{tri}(s,d) \equiv (1 - |d - d(s,r_i)|/h)\mathbb{1}\{|d(s,r_i) - d| \leq h\}$ or weights that equal 1 within some bandwidth around the distance of interest and then smoothly decay to 0. In the framework of this paper, changes to the weights change both the estimator and estimand. In practice, the resulting estimator may be more robust to small errors in locations and may have



(a) Individuals per distance in each region

(b) Estimands

Figure 1: The estimands $\tau(d)$ and $\tau^{eq}(d)$ can meaningfully differ from one another. Consider two types of regions, which are equally likely to be treated and have individuals distributed across space as given in panel (a). Panel (b) shows the decay of ATTs over distance for each region as a solid line. The dashed line shows the estimand $\tau(d)$, which weights by the relative number of individuals at distance d and is increasing in distance over some range. The dotted line shows the estimand $\tau^{eq}(d)$, which weights the regions equally and decreases monotonically.

more attractive properties if one wishes to estimate the *function* $\tau(\cdot)$ in a framework where asymptotically there are individuals arbitrarily close to any distance of interest d for this estimand to be well-defined.

Remark 6. To obtain an estimate of the average *aggregate* effects of a treatment on all treated individuals (irrespective of distance), I suggest summing the estimates-by-distance $\hat{\tau}(d)$ across a partition of distance bins with weights equal to the average (by treatment location) number of individuals in the respective bin. See Online Appendix 5 for details.

The difference in means estimator $\hat{\tau}(d)$ is asymptotically normal under an appropriate sequence of growing, finite populations, by standard arguments. For brevity, the corollary below states asymptotic normality for a special case. Similar to the variance result in Theorem 1, asymptotic normality of $\hat{\tau}(d)$ around the in-sample ATT does not require assumptions about sampling or restrictions on spatial autocorrelations beyond those already introduced. Instead, the non-degenerate distribution is induced by randomness in treatment assignment.

Corollary 1. *Suppose Assumptions 1 and 2 and regularity conditions hold, and there is a single candidate location in each region. Then the estimator* $\hat{\tau}(d)$ *is asymptotically*

normal under sequences of growing but finite populations.
Proof: *See Li and Ding (2017) for conditions and Example 10 therein for the result.*

4.2 Single large region

In some settings, it is not possible to partition the data into separate regions with treatments in one region not affecting outcomes in another and some regions having no realized treatment locations. This section addresses estimation and inference in such settings.

For simplicity of the final results, I focus on designs where treatment assignment to candidate locations is independent:

Assumption 4 (Independent Assignment)**.** *Treatment is assigned to candidate locations independently, with marginal probability* $\pi_s \equiv \Pr(\mathcal{S} \ni s)$ *for location* s*. For* $S \subset \mathbb{S}$*:*

$$\Pr(\mathcal{S} = S) = \prod_{s \in S} \pi_s \prod_{s \in \mathbb{S} \setminus S} (1 - \pi_s).$$

While settings with separate regions could be analyzed as if they were far apart within a single region, this assumption on assignment differs from the combination of Assumptions 2 and 3, so the results in this section do not nest the results of the previous section.

The key idea of this section is that one can use assumptions motivated by the spatial nature of the treatments to derive estimators for treatment effects, as well as their standard errors, analogous to the previous section. Even without these assumptions, the estimators estimate meaningful *marginal* effects (as defined in Equation 2), see Theorem 2 (iii), and only some of the structure is needed for the approximate variance of the estimator to remain valid, see Theorem 2 (iv).

For simplicity, I focus primarily on the following assumption.

Assumption 5 (Additively Separable Effects)**.** *The effects of the treatment are additively separable: For all* $i \in \mathbb{I}$*,* $S \subset \mathbb{S}$ *and* $s \in S$*:*

$$Y_i(S) - Y_i(S \setminus \{s\}) = Y_i(\{s\}) - Y_i(\emptyset) \equiv \tau_i(s).$$

Intuitively, the assumption requires that returns to additional realized treatment locations are neither increasing nor decreasing in the number of realized treatment locations nearby. Additively separable treatment effects are an appropriate specification

if the effect of each treatment is independent of the realization of other treatments. Additive separability implies that one can write $Y_i(S) - Y_i(\emptyset) = \sum_{s \in S} \tau_i(s)$. For instance, the effects of air-polluting power plants (Zigler and Papadogeorgou, 2021) on exposure to pollution are likely approximately additive. The assumption does not impose homogeneity of treatment effects: It neither requires different treatment locations to have the same effect nor does it require a treatment location to have the same effect on two distinct individuals, even if they are at the same distance from the location.

The estimator based on additive separability compares individuals at the distance of interest from realized treatment locations to (properly weighted) individuals at the distance of interest from unrealized treatment locations:

$$\hat{\tau}(d) \equiv \frac{\sum_{s \in \mathbb{S}} \mathbb{1}\{\mathcal{S} \ni s\} \sum_{i \in \mathbb{I}} w_i(s,d) \mathcal{Y}_i}{\sum_{s \in \mathbb{S}} \mathbb{1}\{\mathcal{S} \ni s\} \sum_{i \in \mathbb{I}} w_i(s,d)} - \frac{\sum_{s \in \mathbb{S}} \frac{\mathbb{1}\{\mathcal{S} \not\ni s\}}{1-\pi_s} \pi_s \sum_{i \in \mathbb{I}} w_i(s,d) \mathcal{Y}_i}{\sum_{s \in \mathbb{S}} \frac{\mathbb{1}\{\mathcal{S} \not\ni s\}}{1-\pi_s} \pi_s \sum_{i \in \mathbb{I}} w_i(s,d)}. \tag{5}$$

To derive standard errors, I use an assumption that restricts, for a given treatment, which individuals can be affected by it. Specifically, the assumption states that the treatment does not affect individuals farther away than a fixed, known distance. Intuitively, the assumption generalizes the assumption of separate regions to allow *overlapping* regions. Without assumptions limiting the dependence of outcomes on treatments, one cannot estimate the variance of the estimator as otherwise only a single instance from the data generating process is observed. Assumption 6 or similar is needed to limit the dependence such that the variance can be estimated from the data. It allows defining "exposure mappings" (Aronow and Samii, 2017), which determine for a given individual which treatment configurations lead to identical (potential) outcomes.

Assumption 6 (No Effect After Distance d_0)**.** *For all $i \in \mathbb{I}$, $S \subset \mathbb{S}$ with $s \in S$: if $d(s, r_i) > d_0$, then $Y_i(S) = Y_i(S \setminus \{s\})$.*

The following theorem describes the (approximate) finite population properties of $\hat{\tau}(d)$:

Theorem 2. *The estimator $\hat{\tau}(d)$ is similar to an infeasible estimator $\tilde{\tau}(d)$ (displayed in the appendix), which has analytically tractable non-asymptotic design-based properties:*

(i) Under standard regularity conditions: $\hat{\tau}(d) - \tilde{\tau}(d) \to_p 0$.

(ii) Under Assumptions 4 and 5: $E(\tilde{\tau}(d)) = \tau(d)$ *(unbiasedness for ATT).*

(iii) Under Assumption 4: $E(\tilde{\tau}(d)) = \tau_{marginal}(d)$ *(unbiasedness for marginal ATT).*

(iv) Under Assumptions 4 and 6: The variance of $\tilde{\tau}(d)$ *is*

$$\text{var}(\tilde{\tau}(d)) = \big(\tilde{V}_t(d) + \tilde{V}_c(d) + \tilde{V}_\times(d) - \tilde{V}_{tt}(d) - \tilde{V}_{cc}(d) - \tilde{V}_{ct}(d)\big)/|\mathbb{S}|$$

with the notation defined in Appendix A.2.

Proof: *See Appendix A.2.*

Remark 7. The variance expression is similar in style to that in Theorem 1. The first two terms, $\tilde{V}_t$ and $\tilde{V}_c$, resemble variances of individual level potential outcomes corresponding to treatment and control of a candidate location at distance d. The third term, $\tilde{V}_\times(d)$ takes observable cross-products between distinct individuals, candidate treatment locations, or treatment states. The third term (jointly with the weighting inside $\tilde{V}_t$ and $\tilde{V}_c$) adjusts for the correlation in the exposure to treatment of individuals who are close to one another, as well as for individuals with positive weight $w_i(s, d)$ for multiple different locations s, some of which may be farther than d_0 away and hence not affect the outcome of the individual by Assumption 6.[8] The final three terms, $\tilde{V}_{tt}(d)$, $\tilde{V}_{cc}(d)$, and $\tilde{V}_{ct}(d)$, are averages of squares of differences in potential outcomes that cannot be observed simultaneously, and are therefore unobservable similar to the variance of treatment effects in Theorem 1. Dropping the final three terms yields a conservative estimator of the variance because these terms are non-negative by construction.

Remark 8. To meaningfully reduce the design-based variance of the estimator, one generally needs to expand the sampling area, rather than the number of individuals or candidate locations within a fixed area. Adding individuals while holding the sample area fixed does not generally reduce the variance of the estimator. The effect of increasing the number of (candidate) treatment locations within a fixed sample area on the variance of the estimator is more nuanced. However, the *estimator* of that variance will generally become more conservative because additional candidate locations increase

[8]While the true treatment effect at distances exceeding d_0 is 0 by assumption, it may nevertheless be useful to estimate the effect and compare how close it is to 0 relative to the standard error of the estimator. I discuss this analysis in Section 7.

the number of unobservable and hence inestimable treatment configurations (captured by $\tilde{V}_{cc}(d)$, $\tilde{V}_{tt}(d)$, and $\tilde{V}_{ct}(d)$) exponentially.

While the results above describe an estimator motivated by additive separability, the ideas in this paper can be used to motivate estimators and derive their properties under alternative assumptions such as:

Assumption 7 (Only Nearest Realized Location Matters)**.** *For all $i \in \mathbb{I}$, $S \subset \mathbb{S}$ with $s \in S$, $s' \in \mathbb{S}$: if $d(s, r_i) \leq d(s', r_i)$, then $Y_i(S) = Y_i(S \cup \{s'\}) = Y_i(S \setminus \{s'\})$.*

Typically, only the nearest realized treatment location matters if individuals only access, or visit, a single realized treatment location. For instance, if a developing country quasi-randomly chooses locations to construct new schools (Duflo, 2001), it may be plausible to assume that only the nearest realized school matters to an individual. For the effects of infrastructure projects, such as additional bus or subway stops, on commute times and real estate prices (Gupta et al., 2022), the appropriate assumption may depend on the type of transit stop. An additive effects specification for bus or subway stops may be a good approximation if each stop gives access to a different transit line. A specification where only the nearest stop matters may be more appropriate for stops of the same line.

Theorem 3. *The average effect of the treatment on the treated, $\tau(d)$, is identified if Assumptions 4, 6, and 7 are satisfied.*
Proof: *See Appendix A.3.*

The proof of Theorem 3 is constructive in that it suggests an estimator that exploits the combination of assumptions:

$$\begin{aligned}\hat{\tau}_{\text{nearest}}(d) \equiv & \frac{\sum_{s\in\mathbb{S}} \mathbb{1}\{\mathcal{S} \ni s\} \sum_{i\in\mathbb{I}} \frac{\mathcal{N}_i(s)}{\Pr(\mathcal{N}_i(s)=1|\mathcal{S}\ni s)} w_i(s,d)\mathcal{Y}_i}{\sum_{s\in\mathbb{S}} \mathbb{1}\{\mathcal{S} \ni s\} \sum_{i\in\mathbb{I}} \frac{\mathcal{N}_i(s)}{\Pr(\mathcal{N}_i(s)=1|\mathcal{S}\ni s)} w_i(s,d)} \\ & - \frac{\sum_{s\in\mathbb{S}} \frac{\mathbb{1}\{\mathcal{S}\not\ni s\}}{1-\pi_s}\pi_s \sum_{i\in\mathbb{I}} \frac{\mathcal{N}_i(0)}{\Pr(\mathcal{N}_i(0)=1|\mathcal{S}\not\ni s)} w_i(s,d)\mathcal{Y}_i}{\sum_{s\in\mathbb{S}} \frac{\mathbb{1}\{\mathcal{S}\not\ni s\}}{1-\pi_s}\pi_s \sum_{i\in\mathbb{I}} \frac{\mathcal{N}_i(0)}{\Pr(\mathcal{N}_i(0)=1|\mathcal{S}\not\ni s)} w_i(s,d)}\end{aligned}$$

where $\mathcal{N}_i(s)$ is an indicator for s being the nearest realized treatment location to i, and $\mathcal{N}_i(0)$ is an indicator for no treatment location within d_0 of i being realized. See Online Appendix 6 for additional discussion.

5 Observational data: unconfoundedness assumption and implementation

While the previous section analyzed stylized experiments with randomized assignment that allow the estimation of causal effects under minimal assumptions, often researchers can only study spatial treatments in observational data. In this section, I first lay out a formal assumption under which analysis using observational data can closely mirror the "ideal experiment" discussed above, and then discuss the challenging practical implementation of estimation under such an assumption with spatial data.

5.1 Unconfoundedness for spatial treatments

With observational data, the assignment of treatment to candidate locations was not randomized. Instead, the researcher views the experimental setting described above as the "ideal experiment." A close analog to true randomization is *unconfounded* treatment assignment: When comparing particular locations that are identical along observable characteristics, the treatment is not systematically assigned to locations with higher or lower potential outcomes. Propensity scores, giving the probabilities of treatment at a location given the observable characteristics, then play the roles of the probabilities such as $\Pr(\mathcal{S} \ni s)$ of the experimental assignment mechanism. The analysis proceeds *as if* the researcher had run this particular ideal experiment.

For an internally consistent ideal experiment, the unconfoundedness assumption for spatial treatments takes the treatment location as the unit of observation. It conditions jointly on characteristics of the entire spatial neighborhood and all individuals in them. Alternative individual-level unconfoundedness assumptions stating independence of an individual's potential outcomes and assignment to a nearby treatment location may not be consistent with any single experiment that assigns treatment among candidate locations. This individual-level unconfoundedness may therefore invalidate the thought experiment on which the standard errors derived in this paper are based. The formulation of unconfoundedness below therefore adapts the statements in, for instance, Rosenbaum (2002, p. 78) and Imbens and Rubin (2015, p. 259) to the spatial setting.

Assumption 8 (Spatial Unconfoundedness)**.** *Among a known set of locations* $\tilde{\mathbb{S}} \subset \mathbb{R}^2$*, treatment assignment to locations* $s \in \tilde{\mathbb{S}}$ *is unconfounded at distance* $d \pm h$*, meaning*

that

$$\pi_s \equiv \Pr(\mathcal{S} \ni s) = e\Big(Z_s,\ (X_i)_{i\in\mathbb{I}:\ |d(s,r_i)-d|\leq h}\Big)$$

where Z_s are fixed, observable characteristics of the spatial neighborhood of s, and X_i are fixed, observable characteristics of individual i. Specifically, treatment probabilities do not depend on potential outcomes, and locations with identical characteristics have equal probability of treatment as given by the propensity score function e.

The researcher specifies the neighborhood and individual characteristics (for instance, including their locations r_i relative to the location s) that one needs to condition on to satisfy unconfoundedness based on institutional knowledge relevant to the application.

In addition to unconfoundedness, a positivity (overlap) condition for the candidate locations $\mathbb{S} \subset \tilde{\mathbb{S}}$ of interest is required to ensure that for any neighborhood characteristics of treatment locations in the sample, there is some chance of observing such a neighborhood both with and without realized treatment. In practice, positivity has two implications: First, one should not condition on latitude and longitude because in any given sample one cannot find observations with the same latitude and longitude but different exposure to the spatial treatment. Second, because typically no two neighborhoods are exactly the same – among else, the unconfoundedness assumption above implicitly conditions on the number of individuals – some smoothing and equivariance are typically required. For instance, the approach recommended below can build in equivariance to shifts and rotations of space, imposing that absolute locations (latitude and longitude) and orientation (direction of North) are irrelevant. Instead, the approach only conditions on the relative locations of different units and spatial characteristics.

Turning towards estimation, some economists are concerned about the effect of using an estimated propensity score rather than the known experimental assignment probabilities. These concerns can be at least partially alleviated by using a "double robust" moment condition and sample splitting (Chernozhukov et al., 2018). The resulting estimator tends to be more robust against small estimation errors in the propensity score. Take, as the level of observation, the pair $(i, s) \in \mathbb{I} \times \mathbb{S}$ restricted to pairs with $|d(s, r_i) - d| \leq h$. Let μ_d be the expected outcome at distance d given neighborhood and individual characteristics as well as treatment status, and let e be

the propensity score. The moment condition

$$\psi_{\tau(d)}(\mathcal{Y}, \mathcal{S}, Z, X) = \mathbb{1}\{\mathcal{S} \ni s\}\Big(\mathcal{Y} - \tau(d) - \mu_d\big(Z, X, \mathcal{S} \setminus \{s\}\big)\Big) - \frac{e(Z,X)\big(1 - \mathbb{1}\{\mathcal{S} \ni s\}\big)}{1 - e(Z,X)}\big(\mathcal{Y} - \mu_d(Z, X, \mathcal{S})\big) \tag{6}$$

satisfies Neyman orthogonality (Chernozhukov et al., 2018). It corresponds closely to the IPW estimator $\hat{\tau}$ in Equation 5 under Assumption 4 of independent assignment to locations.

5.2 Finding counterfactual treatment locations using convolutional neural networks

The key challenge in implementing an approach based on the unconfoundedness Assumption 8 is to find locations without treatment that are in neighborhoods otherwise similar to the neighborhoods of locations where a treatment is observed in the data. There are two aspects to this challenge: First, standard methods for propensity score estimation, such as logistic regression, are ill-suited for conditioning on detailed (relative) locations of spatial features in the neighborhood of any given point. Second, with continuous latitude and longitude, the space of possible locations is infinite and any fine discretization may create an impractically large number of possible locations.

I use convolutional neural networks to overcome this challenge. The key insight here is that spatial data can typically be "plotted on a map." Different spatial covariates are placed at different levels of a third, non-spatial dimension. Spatial data then have the same data structure as image data, where the third non-spatial dimension corresponds to the intensity of color channels. Convolutional neural networks have enjoyed recent popularity for analyzing image data (Krizhevsky et al., 2012).

In the proposed implementation, the output of the convolutional neural network at a particular point in space is an assessment of how similar the point and its neighborhood are to the real locations of the treatment and their neighborhoods. Convolutional neural networks can computationally quickly make such assessments at many points in space. One can then match the real locations to counterfactual locations with similar assessments and proceed with propensity score estimation within this sample. Intuitively, while the neural network becomes good at distinguishing

arbitrary locations in space from real treatment locations, there will be "false positives" that it cannot distinguish from real locations. These counterfactual locations are in neighborhoods resembling the neighborhoods of real treatment locations, or else the neural network could have told them apart.

The convolution operation f on a grid $\boldsymbol{v}$ of input values $v_{x,y}$ is computed as

$$f(\boldsymbol{v})_{x,y} = \sum_{a=-k}^{k} \sum_{b=-k}^{k} \beta_{a,b} \cdot v_{x+a,y+b}$$

such that the value at grid cell (x, y) is based on input values within $x \pm k, y \pm k$ for a fixed k. The coefficients $\boldsymbol{\beta}$, which are estimated by the neural network, capture the weight placed on input values at locations relative to (x, y). By using the same $\boldsymbol{\beta}$ to compute the convolution at all points (x, y), convolutional neural networks can be dramatically more parsimonious than fully connected neural networks and enforce equivariance to shift. Using multiple layers of convolutions combined with non-linear activation functions at each grid cell allows the network to learn non-linear relationships.

In training the neural network, key implementation choices involve data augmentation (Simard et al., 2003) and the setup as a generative adversarial task (Goodfellow et al., 2014) while retaining the ease of training "image classification" algorithms. Data augmentation can effectively impose an equivariance to operations such as shift, rotation, and mirroring. Equivariance formalizes the economic logic that relative locations of spatial features and characteristics matter, rather than their absolute locations and orientations. Generative adversarial networks draw from the *modes* of the treatment distribution across space rather than estimating mean locations (Goodfellow, 2016; Lotter et al., 2016), thereby generating more realistic separate locations, and implicitly maintain an internal estimate of the distribution of the treatment across space, which here resembles the propensity score. In Online Appendix 2 and 3, I discuss the implementation both generically and for the application of this paper in more detail.

For inference, I recommend researchers report standard errors conditional on the estimated counterfactual treatment locations and propensity scores. These standard errors reflect a well-defined thought experiment of randomizing treatment assignment among a known set of locations, under which the estimator has desirable properties.

Inference results for convolutional neural networks are, to the best of my knowledge, not currently available, and the estimator selecting counterfactual locations that are most similar to realized treatment locations (akin to matching) is a highly non-smooth function of the data and neural network parameters. At the cost of a less-interpretable thought experiment, researchers may reduce the variance due to the neural network by computing estimates for different "draws" of counterfactual locations from the neural network (based on input data with slightly different shift and rotation) and averaging the estimates as well as standard errors across draws.

6 Extension to non-spatial settings

While I discuss the methods and theory in this paper in the context of spatial treatments, they are applicable more generally. There are two defining features: The treatments are separate units from the individuals whom they affect, and the distance between treatment and individuals is observed and unaffected by the treatment. For the immediate applicability of the methods and results in this paper, treatments need to be binary, but the conceptual insights apply similarly to non-binary treatments. The space treatments and individuals are located in also need not be a geographic space or two-dimensional, as long as it is observed by the researcher.

For an example of a non-spatial setting, consider the question of how a new entrant affects outcomes for existing firms selling differentiated products. In the notation of this paper, the (fixed) location of existing firm i's product in (potentially multi-dimensional) product space is given by fixed characteristics r_i, and $s \in \mathcal{S}$ is the location of an entrant. In a design-based approach to this problem, the researcher specifies alternative points ("candidate locations") in product space where new entrants would have been plausible. A structural model of potential profits at different entrant locations may determine these probabilities of entry at a given location. For quasi-random variation, the researcher postulates that idiosyncratic cost or preference shocks co-determine realized entry locations. If these shocks are independent of the potential outcomes of existing firms (conditional on the expected profit of the entrant), unconfoundedness Assumption 8 may be satisfied. The methods proposed in this paper then allow the researcher to study how entry differentially affects firms for whom entrants are or are not close (in product space) competitors.

Recent work has brought attention to some other settings that are not well-described

by individual-level treatments. Adao et al. (2019); Goldsmith-Pinkham et al. (2020); Borusyak et al. (2022) study the Bartik (1991), or shift-share, design, specifically. In the canonical example, cities i are affected by shocks to different industries $s \in \mathbb{S}$. In the context of this paper, the distance between a city and an industry is related to the industry share, which is taken as fixed, and the quasi-random variation is due to which industry experiences a shock. Borusyak and Hull (2020) study such settings within a regression framework allowing treatments to be non-binary but the interpretation of the estimated effects depends partly on correct specification of the functional form as well as treatment effect homogeneity. The present paper focuses on binary treatments in a potential outcomes framework, yielding non-parametric inverse probability weighting estimators and finite population design-based standard errors.

Some applications fitting into the framework of this paper are currently analyzed as individual-level treatments with clustered assignment. For instance, when studying the effect of state laws on the outcomes of residents, the level of the treatment does not coincide with the level of the outcome variable. For such applications, the framework of the current paper yields results identical to those for experiments with clustered assignment by imposing that state laws can only vary across states. One may argue that, if interest is in the effect of, say, a universal minimum wage, the framework of this paper is more appropriate than one with individual-level treatments. Conceptualizing potential outcomes as functions of an individual-level minimum wage may incorrectly miss spillover, or equilibrium, effects. At the individual-level, there may be a difference between changing only one's own treatment or the treatment of everyone in the state. However, in practice, this misspecification has no impact on estimation and inference as long as one only considers assignments clustered at the state-level and there are no spillovers across states. The advantage of the framework of this paper is that it very naturally allows studying for instance, "spillovers" that are mediated by distance. Distance is often left implicit in these applications. Distance may measure whether an individual resides in the state, a neighboring state, or farther away; or more generally and continuously how close the individual is to the state, for instance, geographically or by the number of flight connections.

The framework of this paper further generalizes the potential outcomes framework with interference, which itself generalizes the potential outcomes framework under the stable unit treatment value assumption (SUTVA). Specifically, outcomes for individual i under SUTVA only depend on i's own treatment, $\mathcal{Y}_i = Y_i(\mathcal{A}_i)$ with treatment $\mathcal{A}_i \in \mathbb{R}$

(Rubin, 1974, 1980). With interference, i's outcome may depend also on the treatments of other individuals, $\mathcal{Y}_i = Y_i(\mathcal{A})$ with $\mathcal{A} = [\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n] \in \mathbb{R}^n$ (Aronow and Samii, 2017). In the present paper, treatments are not directly associated with individuals. Hence, the framework of this paper generalizes the treatment to not (necessarily) be of the same dimension as the number of individuals, $\mathcal{Y}_i = Y_i(\mathcal{A})$ with $\mathcal{A} \subset \mathbb{A}$ where $\mathbb{A}$ is the set of possible treatment assignments and may differ from $\mathbb{R}^n$.

7 Application: foot traffic in times of COVID-19

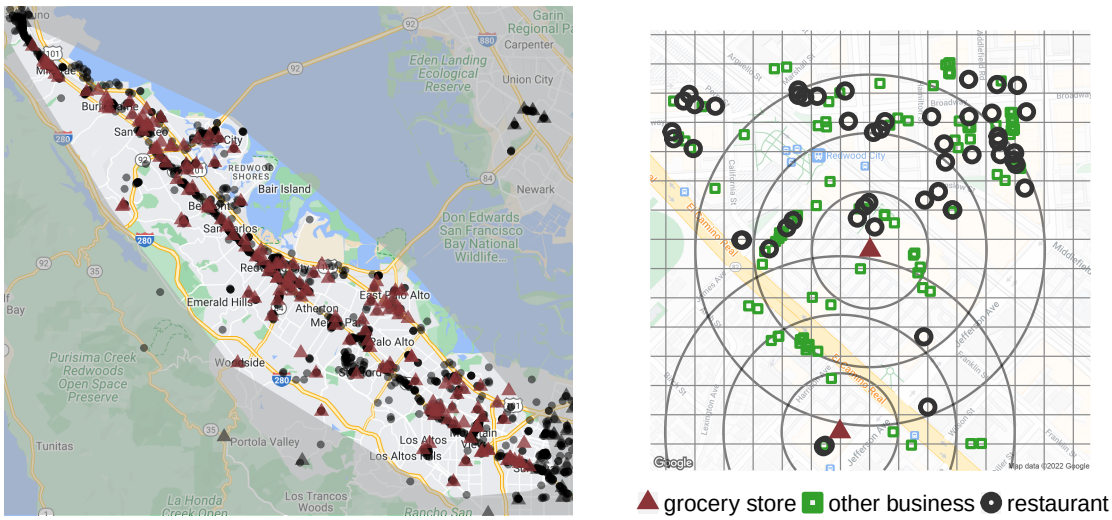
In this section, I demonstrate the use of the proposed methods to study the effect of grocery stores on the number of visitors to restaurants during COVID-19 shelter-in-place policies.[9] Here, grocery store locations are the treatment locations, and restaurants are the individuals for whom I estimate average effects by distance from treatment. I use data on the location of businesses in the San Francisco Bay Area, shown in Figure 2(a), and the number of visitors to them from SafeGraph, available to academic researchers. While this particular application has not been studied in prior work, existing empirical studies could have been replicated for this demonstration if location data was publicly available.[10]

When consumers make only essential trips, such as getting groceries, other businesses relying on foot traffic, such as restaurants, may benefit from being located nearby. Local governments in the San Francisco Bay Area urged residents to only make essential trips during shelter-in-place policies in April 2020. At the same time, other businesses such as restaurants remained open for takeout business. However, drastically reduced foot traffic and customers over time led to financial distress for many businesses (Yang et al., 2020). In such times, a location along consumers' essential trips may benefit these businesses.

The causal interpretation of the cross-sectional estimators of this paper rests on the spatial unconfoundedness assumption: Restaurants in neighborhoods differing in their number of grocery stores, but similar in terms of all other kinds of businesses and observable characteristics, would have comparable numbers of visitors if they had similar exposure to grocery stores. The key argument is that when businesses

[9] See Online Appendix 3 for the exact definitions of grocery stores and restaurants used in this analysis.

[10] See Online Appendix Table OA1 for examples of empirical studies of spatial treatments.



(a) map of sample

(b) example location

Figure 2: The sample includes businesses in the San Francisco Bay Area between San Francisco and San Jose (panel (A)). The locations of real grocery stores are marked by solid red triangles. Restaurants are marked by black circles. The black triangles are grocery stores outside the main study area; their location is considered fixed and restaurants near them are not part of the estimation procedure. In total, there are 167 grocery stores, as well as 1627 distinct restaurants that were open as of January 2020 within 0.5 miles from any of the grocery stores (real or counterfactual within the main study area). Panel (b) zooms in on a location in Redwood City, also indicating locations of other businesses (green squares), and illustrates the size of grid cells as well as circles with radii 0.05mi, 0.10mi, and 0.15mi around the two grocery store locations in the plotted area.

chose their locations before the COVID-19 pandemic, they faced many different considerations and constraints. For grocery stores, the exact locations of nearby restaurants were not the primary concern. Similarly, holding the relative locations of all other businesses fixed, neighborhoods with one more/fewer grocery store appeared similar to restaurants.

For this application, I assume unconfoundedness conditioning on the relative locations of businesses by industry. Panel (b) of Figure 2 illustrates these controls by plotting as green squares other businesses near a particular grocery store in the sample. I superimpose a grid with cells of size 0.025mi $\times$ 0.025mi that shows the discretization used by the convolutional neural network. In estimation, these other businesses are divided into seven groups by their four-digit NAICS code, as listed in Online Appendix Table OA3, and the count of businesses by industry for each grid cell is used as a covariate. In training the network, I impose continuous shifts to the grid, such that the discretization becomes less relevant, as well as rotation and mirroring to build in equivariance such that only relative locations matter. In principle, one could similarly control for any other variables that can be plotted on a map, such as average house price by grid cell or the fraction of individuals with college degrees in the census tract covering the grid cell, if such data are available and relevant for a given application. Details on the implementation of the neural network and propensity score estimation are given in Online Appendix 3.

Researchers can assess the plausibility and quality of the counterfactual grocery store locations predicted by the neural network and the estimated propensity scores by considering two notions of balance.

Figure 3 assesses whether restaurants near real grocery stores, compared to restaurants near counterfactual locations, are exposed to one additional grocery store at the distance of interest, with no differences in exposure at other distances. Each panel focuses on restaurants at a different distance from (real and counterfactual) grocery store locations. The line shows the difference in the average number of real grocery stores by distance from these restaurants. In each panel, there is little difference in exposure for restaurants near real and counterfactual restaurants, *except* at the distance for which these restaurants serve as treated and control, respectively. Hence, the estimated effect at a particular distance indeed reflects the difference between one more/fewer grocery store at that distance. Because balance in exposure to real grocery stores is essential for interpretation, I include covariates describing exposure

directly in the propensity score estimation. If there were differences in exposure at other distances, one could not interpret the estimates as the effect of adding one more grocery store. Instead, under appropriate assumptions, it may reflect the effect of shifting a grocery store from another distance to the distance of interest.

Figure 4 shows that other observable characteristics of the neighborhoods of real and counterfactual grocery store locations are similar. Each panel shows the fraction of all businesses that are classified as a particular industry, by distance from the grocery store locations. The two lines show the fraction for businesses near real and counterfactual grocery store locations, respectively. At the distances shown, the composition of businesses, as given by their industries, appears similar in the neighborhoods of either type of location. Overall, the reasonable balance suggests that the neural networks succeeded in finding counterfactual locations similar to real grocery store locations. Note that restaurants, recreation, museums, and religious locations are used as predictors in the neural network, but dentists and automotive businesses are not.[11] Except for the count (not share) of restaurants, none of these industries are used in the propensity score estimation, such that the balance shown in the figure is not mechanical.

Researchers can also informally inspect the suitability of counterfactual locations by plotting both real and counterfactual locations on a map. Systematic differences between real and counterfactual locations imply that estimated effects reflect not just differences in exposure to grocery stores, but also these other differences.

Figure 5 shows estimates and standard errors by distance from treatment for the estimator given in Equation 5. Standard errors take the counterfactual grocery store locations predicted by the neural network and the estimated propensity scores as given. Standard errors are based on Assumption 4 of independent assignment and Assumption 6 that treatments have no effect beyond a distance of $d_0 \equiv 0.075$ miles. Independent assignment may appear implausible if one believes that clustering of grocery stores close to one another is particularly likely or unlikely. In practice, I observe real grocery stores both in isolated locations and close to other grocery stores. However, if information on the covariances (joint location probabilities) was available one could impose it instead of independent assignment (zero covariance). No effect beyond 0.075 miles appears plausible given the substantively close to zero

[11] The count of dentists and automotive businesses is used by the neural network together with all "other industries" as a single covariate per grid cell.

Figure 3: Differential exposure of treated and control restaurants to grocery stores at different distances. Each panel holds fixed restaurants that are at a particular distance from real or counterfactual grocery stores.

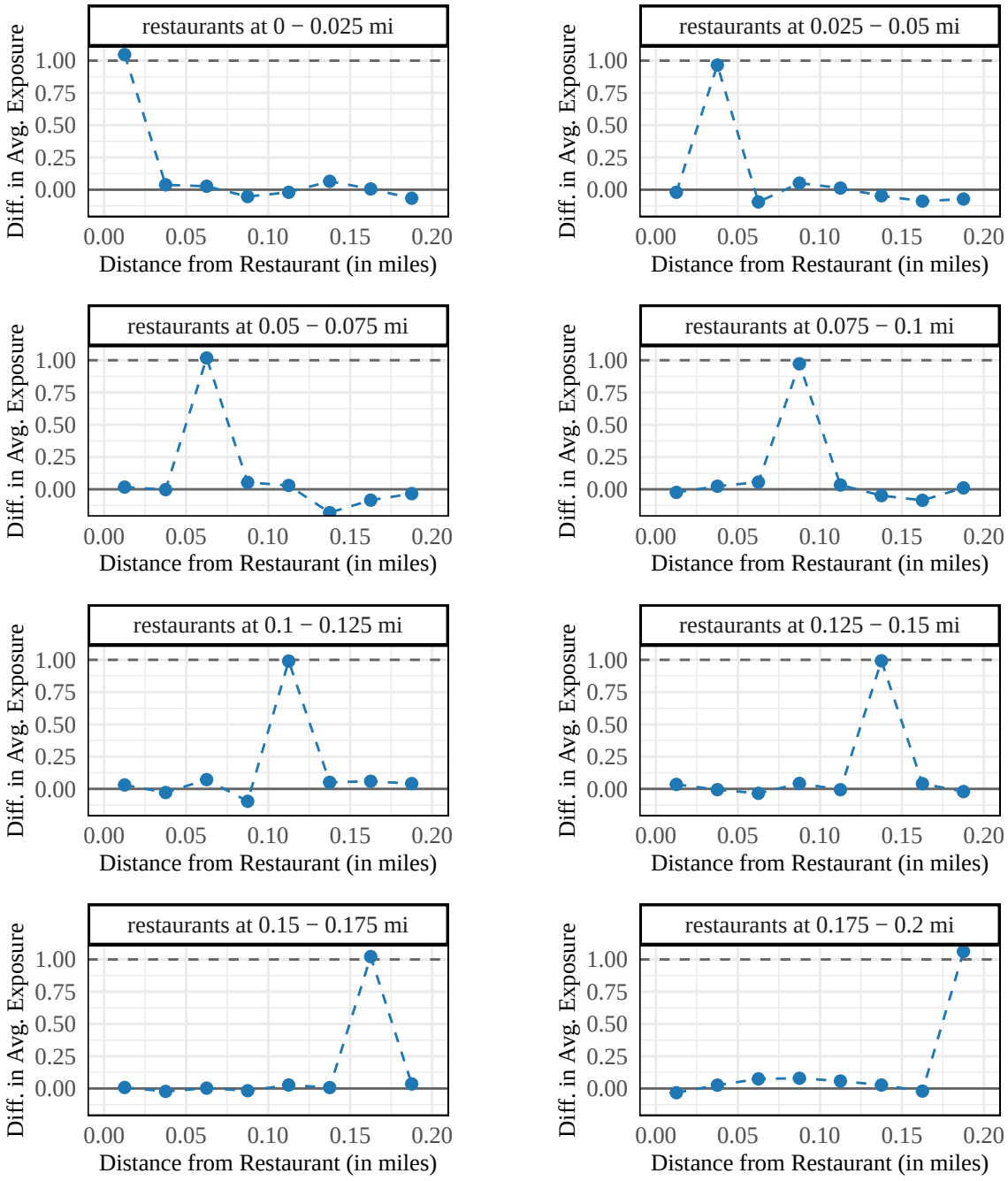


Figure 4: Industry composition of businesses near real and counterfactual grocery stores. The error bars show ±1.96 times the standard error of the difference in (weighted) means, centered around the mean of the treated.

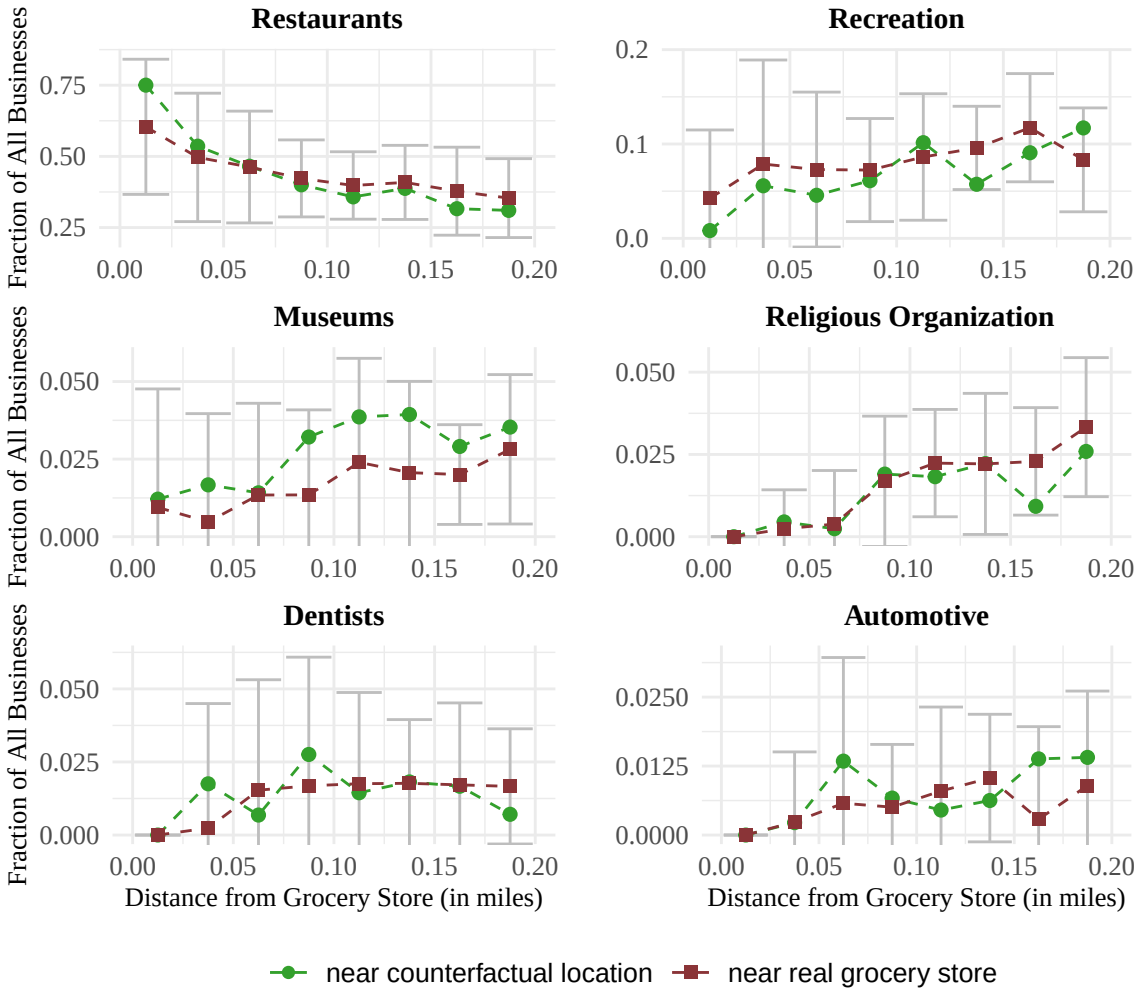
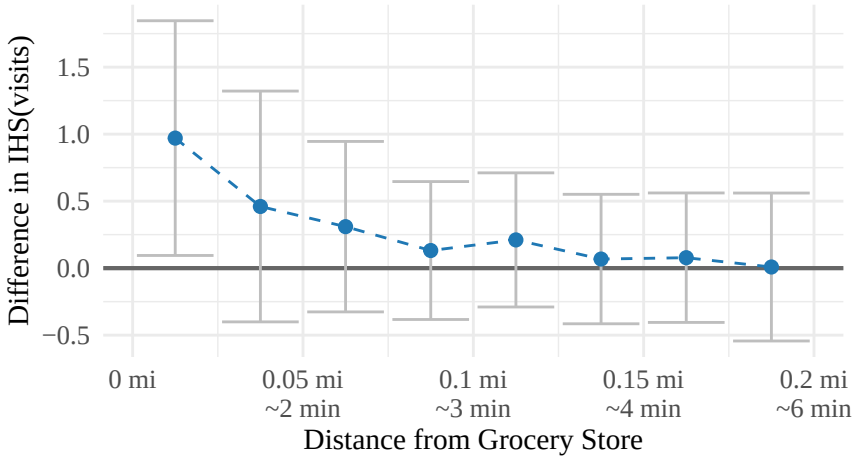


Figure 5: Estimated average effect of grocery stores on restaurants at different distances. The outcome is the inverse hyperbolic sign of the number of visits as recorded by SafeGraph. Bars indicate ±1.96 standard errors.



point estimates beyond that distance in Figure 5. Note, however, that such a figure is not proof of the sharp null hypothesis of no effect of any possible grocery store exposure beyond such distances. Without further assumptions, the figure only suggests zero *average marginal* effects. If each grocery store brings a separate set of potential customers to nearby restaurants, treatment effects may be approximately additively separable (Assumption 5), in which case average marginal effects equal average effects.

The estimates show substantial positive effects of being located very close to a grocery store, with no effect past a few minutes of walking. Table 1 shows the point estimates corresponding to Figure 5 up to a distance of 0.2 miles. For the restaurants in the closest bin of up to 0.025 miles, the average effect more than doubles the number of SafeGraph-recorded visitors both when estimating an approximate percentage effect using inverse hyperbolic sine units and when estimating effects in levels. In the second closest bin of restaurants between 0.025 and 0.05 miles from grocery stores, the estimated effects are smaller and not statistically significant at the 5% level (using a normal approximation). For any longer distance, the effects are both economically smaller and statistically insignificant. Effects close to 0 past a couple of minutes of walking may be due to either the unwillingness of consumers to walk longer distances or the lack of a need to do so because there typically is a closer alternative restaurant or coffee shop.

Table 1 also shows double robust estimates of treatment effects. For these esti-

Table 1: Estimated effects on the number of visits to restaurants using different estimators. The first panel uses the inverse probability weighting estimators for spatial experiments proposed in this paper. The second panel uses a double robust version of the spatial experiment estimator. For each panel, "IHS" refers to the effect in inverse hyperbolic sine unit and "level" to effects in levels. Standard errors are given in parentheses. Rows labeled "percent incr." show the percent increase (or decrease) relative to the mean visits of the control restaurants at that distance.

Distance:	0.000 mi – 0.025 mi	0.025 mi – 0.050 mi	0.050 mi – 0.075 mi	0.075 mi – 0.100 mi	0.100 mi – 0.125 mi	0.125 mi – 0.150 mi	0.150 mi – 0.175 mi	0.175 mi – 0.200 mi
Spatial Experiment Estimators:								
IHS:	0.97	0.46	0.31	0.13	0.21	0.07	0.08	0.01
	(0.45)	(0.44)	(0.32)	(0.26)	(0.26)	(0.25)	(0.25)	(0.28)
percent incr.:	166	59	37	14	24	7	8	1
level:	14.92	-0.42	3.78	1.53	2.52	0.43	2.93	-7.79
	(6.33)	(8.86)	(4.18)	(3.91)	(3.00)	(3.68)	(4.70)	(6.49)
percent incr.:	158	-3	36	12	23	3	28	-37
Double Robust Estimators for Spatial Experiments:								
IHS:	1.07	0.49	0.10	0.01	-0.02	-0.03	-0.17	-0.09
percent incr.:	193	63	10	1	-2	-3	-16	-9
level:	16.16	2.44	-0.98	-0.40	-0.36	-1.07	-1.19	-4.45
percent incr.:	171	15	-9	-3	-3	-9	-11	-21

mates, I estimate the propensity score and outcome models using cross-fitting: When predicting treatment status or outcome level for a grocery store or restaurant, I use only grocery stores or restaurants that are at least 1 mile away. The results are qualitatively and quantitatively similar to the inverse probability weighting estimators suggesting that "overfitting" and noise in propensity score estimation may not be of first-order concern.

8 Conclusion

The causal effects of treatments occurring at locations in space on individuals located nearby are of interest across fields of economics and social sciences. In this paper, I argue that identification, estimation, and inference using design-based ideas are conceptually attractive, analytically tractable, and computationally feasible. I show that an ideal experiment varying the location of the treatment does not validate the inner vs. outer ring empirical strategy commonly applied in empirical practice. Instead, this ideal experiment validates the comparison of individuals near realized treatment to individuals near counterfactual locations where the treatment could have happened (but did not). The finite population design-based variances derived in this paper express the variation due to the ideal experiment. Design-based inference removes the need to specify a hypothetical super-population and sampling scheme. Because counterfactual locations of treatments are typically not available in observational data, I propose a computationally feasible method using convolutional neural networks to identify locations that are observationally equivalent to real treatment locations. These counterfactual locations allow the estimation of causal effects in observational data under a "random conditional on observables" (unconfoundedness) assumption.

I demonstrate the use of these methods by studying the causal effects of grocery stores on the number of visitors to nearby restaurants during COVID-19 shelter-in-place policies. In this application, the counterfactual grocery store locations proposed by the neural network are in neighborhoods that are indeed observationally similar to the neighborhoods of real grocery stores. I estimate substantial effects for restaurants very close to a grocery store, on average more than doubling the number of visitors, as measured in data from SafeGraph. The design-based standard errors take into account the complex ways in which exposure to grocery stores (the treatment) is correlated across restaurants (outcome units) *by design* of the ideal experiment. Hence,

I find significant externalities between businesses. Such externalities may lead to socially undesirable concentrations of consumers during a pandemic, as well as spatial inequities across business owners to the extent that they are unanticipated and not internalized through, for instance, differential rent.

A Proofs

A.1 Theorem 1

Define the estimator

$$\tilde{\tau}(d) \equiv \mu_t(d) - \mu_c(d) + \frac{\sum_{j=1}^{J} \mathcal{W}_j \sum_{s\in\mathbb{S}_j} \mathbb{1}\{\mathcal{S} \ni s\} \sum_{i\in\mathbb{I}_j} w_i(s,d)(\mathcal{Y}_i - \mu_t(d))}{\sum_{j=1}^{J} \pi_j \sum_{s\in\mathbb{S}_j} \pi_j(s) \sum_{i\in\mathbb{I}_j} w_i(s,d)}$$
$$- \frac{\sum_{j=1}^{J} \frac{1-\mathcal{W}_j}{1-\pi_j} \pi_j \sum_{s\in\mathbb{S}_j} \pi_j(s) \sum_{i\in\mathbb{I}_j} w_i(s,d)(\mathcal{Y}_i - \mu_c(d))}{\sum_{j=1}^{J} \pi_j \sum_{s\in\mathbb{S}_j} \pi_j(s) \sum_{i\in\mathbb{I}_j} w_i(s,d)}$$

where

$$\mu_t(d) \equiv \frac{\sum_{j=1}^{J} \pi_j \sum_{s\in\mathbb{S}_j} \pi_j(s) \sum_{i\in\mathbb{I}_j} w_i(s,d) Y_i(s)}{\sum_{j=1}^{J} \pi_j \sum_{s\in\mathbb{S}_j} \pi_j(s) \sum_{i\in\mathbb{I}_j} w_i(s,d)}$$
$$\mu_c(d) \equiv \frac{\sum_{j=1}^{J} \pi_j \sum_{s\in\mathbb{S}_j} \pi_j(s) \sum_{i\in\mathbb{I}_j} w_i(s,d) Y_i(0)}{\sum_{j=1}^{J} \pi_j \sum_{s\in\mathbb{S}_j} \pi_j(s) \sum_{i\in\mathbb{I}_j} w_i(s,d)}.$$

For brevity, I suppress the dependence on d in the following. Let p denote the the non-stochastic denominator used in $\tilde{\tau}$; $\hat{p}_t$ and $\hat{p}_c$ the stochastic denominators of $\hat{\mu}_t = \bar{\mathcal{Y}}_t$ and $\hat{\mu}_c = \bar{\mathcal{Y}}_c$ from the main text; and $\tilde{\mu}_t = \frac{\hat{p}_t}{p}\hat{\mu}_t$ and $\tilde{\mu}_c = \frac{\hat{p}_c}{p}\hat{\mu}_c$. Then $\hat{\tau} = \frac{p}{\hat{p}_t}\tilde{\mu}_t - \frac{p}{\hat{p}_c}\tilde{\mu}_c$ and $\tilde{\tau} = \mu_t - \mu_c + \tilde{\mu}_t - \frac{\hat{p}_t}{p}\mu_t - \tilde{\mu}_c + \frac{\hat{p}_c}{p}\mu_c$.

For part (i) of Theorem 1, apply the mean value theorem to the function

$$\tilde{\Delta}(\hat{p}_t, \hat{p}_c, \tilde{\mu}_t, \tilde{\mu}_c) \equiv \frac{p}{\hat{p}_t}\tilde{\mu}_t - \frac{p}{\hat{p}_c}\tilde{\mu}_c - (\mu_t - \mu_c + \tilde{\mu}_t - \frac{\hat{p}_t}{p}\mu_t - \tilde{\mu}_c + \frac{\hat{p}_c}{p}\mu_c)$$

with endpoints $(\hat{p}_t, \hat{p}_c, \tilde{\mu}_t, \tilde{\mu}_c)$ and (p, p, μ_t, μ_c) to obtain

$$\hat{\tau} - \tilde{\tau} = (\tilde{\mu}_t - \mu_t)\Big(\frac{1}{\dot{p}_t/p} - 1\Big) - (\tilde{\mu}_c - \mu_c)\Big(\frac{1}{\dot{p}_c/p} - 1\Big)$$
$$+ (\frac{\hat{p}_c}{p} - 1)\Big(\frac{p^2}{\ddot{p}_c^2}\dot{\mu}_c - \mu_c\Big) - (\frac{\hat{p}_t}{p} - 1)\Big(\frac{p^2}{\ddot{p}_t^2}\dot{\mu}_t - \mu_t\Big)$$

where variables $\dot{a}$ lie between $\hat{a}$ and a for $a = \mu_t, \mu_c, p_t, p_c$. Consider a sequence of finite populations with a growing number of regions J where no region dominates in size: For instance, assume for all j: $\sum_{s\in\mathbb{S}_j} \pi_j(s) w_i(s,d) \in [\underline{c}, \bar{c}]$, and potential outcomes are bounded in absolute value. Then by Theorem 1 of Li and Ding (2017) and using Slutsky's Theorem and the Delta Method, each of the factors of the four products is $\sqrt{J}$-asymptotically normal, implying part (i).

For part (ii) notice that $\mathcal{W}_j \mathbb{1}\{\mathcal{S} \ni s\}\mathcal{Y}_i = \mathcal{W}_j \mathbb{1}\{\mathcal{S} \ni s\} Y_i(s)$ and $(1-\mathcal{W}_j)\mathcal{Y}_i = (1-\mathcal{W}_j)Y_i(0)$ for $i \in \mathbb{I}_j$. When taking expectations, $E(\mathcal{W}_j \mathbb{1}\{\mathcal{S} \ni s\}) = \pi_j \pi_j(s)$ and $E(1-\mathcal{W}_j) = 1-\pi_j$. The result then follows immediately by comparing the definitions because only the numerators of $\tilde{\tau}$ are stochastic and $\tau = \mu_t - \mu_c$.

I outline the key steps of the variance derivation for part (iii) here and give detailed step-by-step derivations in the proof in Online Appendix 4 for a more general case. The quantity of interest is $\text{var}(\tilde{\tau})$. Observed outcomes $\mathcal{Y}_i$ can be replaced by potential outcomes $Y_i(s)$ and $Y_i(0)$ by the same argument as for part (ii). Dropping $\mu_t - \mu_c$ and the -1 component of $-(1-\mathcal{W}_j)$, factoring out the denominator p, letting $\mathcal{T}_j(s) \equiv \mathcal{W}_j \mathbb{1}\{\mathcal{S} \ni s\}$ and using $\sum_{s\in\mathbb{S}_j} \mathcal{T}_j(s) = \mathcal{W}_j$:

$$\text{var}(\tilde{\tau}) = \text{var}\Big(\sum_{j=1}^{J} \sum_{s\in\mathbb{S}_j} \mathcal{T}_j(s)\big(Y_j^t(s) + \frac{\pi_j}{1-\pi_j} Y_j^c\big)\Big)/p^2$$

where $Y_j^t(s) \equiv \sum_{i\in\mathbb{I}_j} w_i(s,d)(Y_i(s) - \mu_t(d))$, $Y_j^c \equiv \sum_{s\in\mathbb{S}_j} \pi_j(s) \sum_{i\in\mathbb{I}_j} w_i(s,d)(Y_i(0) - \mu_c(d))$.

Because $Y_j^t(s)$ and Y_j^c are non-stochastic, the variance of the sum depends only on covariances $\text{cov}(\mathcal{T}_j(s), \mathcal{T}_{j'}(s'))$ that can be calculated based on Assumptions 2 and 3:

$$\text{cov}(\mathcal{T}_j(s), \mathcal{T}_{j'}(s')) = \begin{cases} \pi\pi_j(s)(1 - \pi\pi_j(s)) & \text{if } j = j', s = s' \\ -\pi^2 \pi_j(s)\pi_j(s') & \text{if } j = j', s \neq s' \\ -\frac{\pi(1-\pi)}{J-1} \pi_j(s)\pi_{j'}(s') & \text{if } j \neq j'. \end{cases}$$

The squared sums of potential outcomes of differing treatment states are

$$(Y_j^t(s) + \frac{\pi_j}{1-\pi_j} Y_j^c)^2 = Y_j^t(s)^2 + 2\frac{\pi_j}{1-\pi_j} Y_j^t(s) Y_j^c + \big(\frac{\pi_j}{1-\pi_j} Y_j^c\big)^2.$$

Rewrite products $Y_j^t(s)Y_j^c$ of potential outcomes of conflicting treatment states using

$$(Y_i(s)-\mu_t)(Y_i(0)-\mu_c) = \frac{1}{2}\Big((Y_i(s)-\mu_t)^2+(Y_i(0)-\mu_c)^2-(Y_i(s)-Y_i(0)-(\mu_t-\mu_c))^2\Big)$$

and similarly at different levels of aggregation. The squares of demeaned treated and control potential outcomes become the marginal variances; the final squared term becomes the variance of treatment effects. The remaining steps simplify the expression into the terms shown in the main text.

A.2 Notation and Proof of Theorem 2

Notation Define exposure mappings (Aronow and Samii, 2017) based on Assumption 6 as follows. $\mathbb{M}_i \equiv 2^{\{s\in\mathbb{S}:\, d(s,r_i)\le d_0\}}$ is the set of all possible ways in which treatment can be assigned to those locations that possibly affect i. With slight abuse of notation, denote i's potential outcome under exposure $m \in \mathbb{M}_i$ by $Y_i(m)$. Let the random variable $\mathcal{M}_m^i$ be the indicator for whether exposure m of individual i is realized. Then $\mathcal{Y}_i = \sum_{m\in\mathbb{M}_i} \mathcal{M}_m^i Y_i(m)$. Define the random variables $\mathcal{T}_s^t \equiv \mathbb{1}\{\mathcal{S} \ni s\}$ and $\mathcal{T}_s^c \equiv \mathbb{1}\{\mathcal{S} \not\ni s\}$ and probabilities $\pi_{i,s}^{m,a} \equiv \Pr(\mathcal{M}_i^m\mathcal{T}_s^a = 1)$ and $\pi_{i,s,i',s'}^{m,a,m',a'} \equiv \Pr(\mathcal{M}_m^i\mathcal{T}_s^a = 1 \text{ and } \mathcal{M}_{m'}^{i'}\mathcal{T}_{s'}^{a'} = 1)$, which are straightforward to compute under Assumptions (4) and (6).

The variance terms used in the statement of the theorem are, for $a \in \{c,t\}$,

$$\tilde{V}_a(d) \equiv \frac{1}{|\mathbb{S}|}\sum_{s\in\mathbb{S}}\sum_{i\in\mathbb{I}}\sum_{m\in\mathbb{M}_i}\pi_{i,s}^{m,a}\frac{w_i(s,d)}{\bar{n}(d)}v_{i,s}^{m,a}(d)(Y_i(m)-\mu_a(d))^2$$

$$\begin{aligned}\tilde{V}_\times(d) \equiv \frac{1}{|\mathbb{S}|}\sum_{s\in\mathbb{S}}\sum_{s'\in\mathbb{S}}\sum_{i\in\mathbb{I}}\sum_{m\in\mathbb{M}_i}\sum_{i'\in\mathbb{I}}\sum_{m'\in\mathbb{M}_i}\sum_{a\in\{c,t\}}\sum_{a'\in\{c,t\}}\Bigg(&\mathbb{1}\{i\neq i' \text{ or } s\neq s' \text{ or } a\neq a'\}\\ &\cdot\mathbb{1}\{\pi_{i,i'}^{m,m'}>0\}\big(\pi_{i,s,i',s'}^{m,a,m',a'}-\pi_{i,s}^{m,a}\pi_{i',s'}^{m',a'}\big)\Big(-\frac{\pi_s}{1-\pi_s}\Big)^{\mathbb{1}\{a=c\}}\Big(-\frac{\pi_{s'}}{1-\pi_{s'}}\Big)^{\mathbb{1}\{a'=c\}}\\ &\cdot\frac{w_i(s,d)w_{i'}(s',d)}{\bar{n}(d)^2}(Y_i(m)-\mu_a(d))(Y_{i'}(m')-\mu_{a'}(d))\Bigg)\end{aligned}$$

$$\begin{aligned}\tilde{V}_{aa}(d) \equiv \frac{2}{|\mathbb{S}|}\sum_{a\in\{c,t\}}\sum_{s\in\mathbb{S}}\sum_{s'\in\mathbb{S}}\sum_{i\in\mathbb{I}}\sum_{m\in\mathbb{M}_i}\sum_{i'\in\mathbb{I}}\sum_{m'\in\mathbb{M}_i}&\mathbb{1}\{\pi_{i,i'}^{m,m'}=0\}\pi_{i,s}^{m,a}\pi_{i',s'}^{m',a}\\ &\cdot\Big(\frac{\pi_s}{1-\pi_s}\frac{\pi_{s'}}{1-\pi_{s'}}\Big)^{\mathbb{1}\{a=c\}}\frac{w_i(s,d)w_{i'}(s',d)}{\bar{n}(d)^2}\Big(\frac{Y_i(m)+Y_{i'}(m')}{2}-\mu_a(d)\Big)^2\end{aligned}$$

$$\tilde{V}_{ct}(d) \equiv \frac{1}{|\mathbb{S}|}\sum_{s\in\mathbb{S}}\sum_{s'\in\mathbb{S}}\sum_{i\in\mathbb{I}}\sum_{m\in\mathbb{M}_i}\sum_{i'\in\mathbb{I}}\sum_{m'\in\mathbb{M}_i}\mathbb{1}\{\pi_{i,i'}^{m,m'}=0\}\pi_{i,s}^{m,t}\pi_{i',s'}^{m',c}\frac{\pi_{s'}}{1-\pi_{s'}}$$
$$\cdot\frac{w_i(s,d)w_{i'}(s',d)}{\bar{n}(d)^2}\big((Y_i(m)-Y_{i'}(m'))-(\mu_t(d)-\mu_c(d))\big)^2$$

where $\bar{n}(d)\equiv\frac{1}{\mathbb{S}}\sum_{s\in\mathbb{S}}\pi_s\sum_{i\in\mathbb{I}}w_i(s,d)$ is the average (per location) expected number of treated individuals, and the fixed, computable weights $v_{i,s}^{m,a}(d)$ are

$$v_{i,s}^{m,a}(d)\equiv\Big(\frac{\pi_s}{1-\pi_s}\Big)^{\mathbb{1}\{a=c\}}\Big((1-\pi_{i,s}^{m,a})\Big(\frac{\pi_s}{1-\pi_s}\Big)^{\mathbb{1}\{a=c\}}\frac{w_i(s,d)}{\bar{n}(d)}$$
$$+\sum_{i'\in\mathbb{I}}\sum_{m'\in\mathbb{M}_i}\sum_{s'\in\mathbb{S}}\sum_{a'\in\{c,t\}}\mathbb{1}\{\pi_{i,i'}^{m,m'}=0\}\pi_{i',s'}^{m',a'}\Big(\frac{\pi_{s'}}{1-\pi_{s'}}\Big)^{\mathbb{1}\{a'=c\}}\frac{w_{i'}(s',d)}{\bar{n}(d)}\Big).$$

Proof The proof, sketched below, is mostly analogous to the proof of Theorem 1.

Approximate $\hat{\tau}(d)$ by $\tilde{\tau}(d)\equiv\tau_{\text{marginal}}+\mathcal{D}$ where

$$\mathcal{D}\equiv\frac{1}{|\mathbb{S}|}\sum_{s\in\mathbb{S}}\Big(\mathbb{1}\{\mathcal{S}\ni s\}\sum_{i\in\mathbb{I}}\frac{w_i(s,d)}{\bar{n}(d)}(\mathcal{Y}_i-\mu_t(d))-\frac{\mathbb{1}\{\mathcal{S}\not\ni s\}}{1-\pi_s}\pi_s\sum_{i\in\mathbb{I}}\frac{w_i(s,d)}{\bar{n}(d)}(\mathcal{Y}_i-\mu_c(d))\Big)$$

and $\mu_t(d)$ and $\mu_c(d)$ are defined with the same weights as $\tau_{\text{marginal}}(d)$ (Equation 2) but with $Y_i(S)$ and $Y_i(S\setminus\{s\})$, respectively, replacing $\tau_i(s\mid S)$. The result for $\hat{\tau}-\tilde{\tau}$ in Appendix A.1 with bounded potential outcomes implies $\hat{\tau}(d)-\tilde{\tau}(d)\to_p 0$ as long as $\hat{p}_t$ and $\hat{p}_c$ converge to their expected value p. For this convergence, consider a sequence of finite populations, indexed by k, growing in the sense that $|\mathbb{I}_k|\to\infty$ and $|\mathbb{S}_k|\to\infty$. Assume, for all s, $\pi_s\in(\eta,1-\eta)$ with $\eta>0$ and that no treatment location dominates in size asymptotically (for instance, for all $s\in\mathbb{S}_k$: $\sum_{i\in\mathbb{I}_k}w_i(s,d)\in[\underline{c},\bar{c}]$). Then $\hat{p}_t$ and $\hat{p}_c$ converge in probability to p by the law of large numbers for independent ($\mathbb{1}\{\mathcal{S}\ni s\}$ are independent across s by Assumption 4) non-identically distributed random variables, establishing part (i).

Unbiasedness for $\tau_{\text{marginal}}(d)$ follows directly by taking expectations of the numerator of $\mathcal{D}$. Under Assumption 5, the expected value simplifies to $\tau(d)$ because $\tau_i(s\mid S)=\tau_i(s)$.

To characterize the variance, one can rewrite $\mathcal{D}$ in terms of fixed potential outcomes

by using the exposure mappings, specifically

$$\begin{aligned}&\sum_{s\in\mathbb{S}}\mathbb{1}\{\mathcal{S}\ni s\}\sum_{i\in\mathbb{I}}w_i(s,d)(\mathcal{Y}_i-\mu_t(d))-\sum_{s\in\mathbb{S}}\mathbb{1}\{\mathcal{S}\not\ni s\}\frac{\pi_s}{1-\pi_s}\sum_{i\in\mathbb{I}}w_i(s,d)(\mathcal{Y}_i-\mu_c(d))\\=&\sum_{i\in\mathbb{I}}\sum_{m\in\mathbb{M}_i}\sum_{s\in\mathbb{S}}\sum_{a\in\{c,t\}}\mathcal{M}_i^m\mathcal{T}_s^a\tilde{Y}_i^{s,a}(d,m)\end{aligned}$$

with $\tilde{Y}_i^{s,a}(d,m)\equiv\left(-\frac{\pi_s}{1-\pi_s}\right)^{\mathbb{1}\{a=c\}}w_i(s,d)(Y_i(m)-\mu_a(d))$. Importantly, only $\mathcal{M}_i^m\mathcal{T}_{s,a}$ is stochastic in the expression above. Hence, the variance depends on covariances

$$\operatorname{cov}(\mathcal{M}_i^m\mathcal{T}_s^a,\mathcal{M}_{i'}^{m'}\mathcal{T}_{s'}^{a'})=\pi_{i,s,i',s'}^{m,a,m',a'}-\pi_{i,s}^{m,a}\pi_{i',s'}^{m',a'}.$$

Where $\pi_{i,s,i',s'}^{m,a,m',a'}=0$ such that m and m' cannot be observed simultaneously, rewrite terms $\sum_s\sum_{s'}(Y_i(m)-\mu_a(d))(Y_i(m')-\mu_{a'}(d))$ with $a,a'\in\{t,c\}$ depending on whether the current locations s,s' are treated or control under exposures m,m', respectively. When $a=a'$, these terms are multiplied by a factor of opposite sign compared to the terms appearing in the proof of Theorem 1. To obtain a formula suggesting a conservative estimator of the variance, when $a=a'$ instead rewrite

$$\begin{aligned}&(Y_i(m)-\mu_a(d))(Y_{i'}(m')-\mu_a(d))\\=&\frac{1}{2}\Big((Y_i(m)+Y_{i'}(m')-2\mu_a(d)))^2-(Y_i(m)-\mu_a(d))^2-(Y_{i'}(m')-\mu_a(d))^2\Big).\end{aligned}$$

The remaining steps simplify the summations over such terms. Step-by-step derivations are available in Online Appendix 9.

A.3 Proof of Theorem 3

Independent assignment implies that for each individual i and candidate treatment location s, there is a positive probability the location is the nearest realized treatment location. The assumption that only the nearest realized location matters implies that in this case $Y_i(s)$ is observed, rather than $Y_i(s\cup S)$ for some set of other locations S farther from i than s. The control potential outcome $Y_i(0)$ is observed when no treatment location within distance d_0 is treated, which occurs with positive probability under independent assignment.

References

Abadie, A., S. Athey, G. W. Imbens, and J. M. Wooldridge (2023). When Should You Adjust Standard Errors for Clustering? *The Quarterly Journal of Economics 138*(1), 1–35. 4

Adao, R., M. Kolesár, and E. Morales (2019). Shift-share designs: Theory and inference. *The Quarterly Journal of Economics 134*(4), 1949–2010. 26

Aronow, P. M. and C. Samii (2017). Estimating average causal effects under general interference, with application to a social network experiment. *The Annals of Applied Statistics 11*(4), 1912–1947. 18, 27, 38

Aronow, P. M., C. Samii, and Y. Wang (2020). Design-based inference for spatial experiments with interference. *arXiv preprint arXiv:2010.13599*. 5

Athey, S., D. Eckles, and G. W. Imbens (2018). Exact p-values for network interference. *Journal of the American Statistical Association 113*(521), 230–240. 6

Bartik, T. J. (1991). *Who benefits from state and local economic development policies?* Kalamazoo, MI: W.E. Upjohn Institute for Employment Research. 26

Borusyak, K. and P. Hull (2020). Non-random exposure to exogenous shocks: Theory and applications. *NBER Working Paper Series* (27845). 5, 6, 26

Borusyak, K., P. Hull, and X. Jaravel (2022). Quasi-experimental shift-share research designs. *The Review of Economic Studies 89*(1), 181–213. 26

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal 21*(1), C1–C68. 22, 23

Conley, T. G. (1999). Gmm estimation with cross sectional dependence. *Journal of Econometrics 92*(1), 1–45. 4, 5

Diamond, R. and T. McQuade (2019). Who wants affordable housing in their backyard? an equilibrium analysis of low-income property development. *Journal of Political Economy 127*(3), 1063–1117. 9

Duflo, E. (2001). Schooling and labor market consequences of school construction in indonesia: Evidence from an unusual policy experiment. *American Economic Review 91*(4), 795–813. 20

Goldsmith-Pinkham, P., I. Sorkin, and H. Swift (2020). Bartik instruments: What, when, why, and how. *American Economic Review 110*(8), 2586–2624. 26

Goodfellow, I. (2016). Nips 2016 tutorial: Generative adversarial networks. *arXiv preprint arXiv:1701.00160*. 24

Goodfellow, I., J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio (2014). Generative adversarial nets. In *Advances in Neural Information Processing Systems*, Volume 27, pp. 2672–2680. 24

Gupta, A., S. Van Nieuwerburgh, and C. E. Kontokosta (2022). Take the q train: Value capture of public infrastructure projects. *Journal of Urban Economics 129*, 103422. 20

Imbens, G. W. and D. B. Rubin (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction.* New York, NY: Cambridge University Press. 14, 21

Krizhevsky, A., I. Sutskever, and G. E. Hinton (2012). Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, Volume 25, pp. 1097–1105. 23

Leung, M. P. (2020). Treatment and spillover effects under network interference. *The Review of Economics and Statistics 102*(2), 1–13. 6

Leung, M. P. and P. Loupos (2022). Unconfoundedness with network interference. *arXiv preprint arXiv:2211.07823*. 6

Li, X. and P. Ding (2017). General forms of finite population central limit theorems with applications to causal inference. *Journal of the American Statistical Association 112*(520), 1759–1769. 17, 37

Linden, L. and J. E. Rockoff (2008). Estimates of the impact of crime risk on property values from Megan's laws. *American Economic Review 98*(3), 1103–1127. 3

Lotter, W., G. Kreiman, and D. Cox (2016). Unsupervised learning of visual structure using predictive generative networks. *arXiv preprint arXiv:1511.06380*. 24

Neyman, J. (1923). On the application of probability theory to agricultural experiments. essay on principles. section 9. *Roczniki Nauk Rolniczych Tom X*, 1–51. [in Polish]. 3

Pinkse, J., L. Shen, and M. Slade (2007). A central limit theorem for endogenous locations and complex spatial interactions. *Journal of Econometrics 140*(1), 215–225. 4

Rosenbaum, P. R. (2002). *Observational Studies*. New York, NY: Springer. 21

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology 66*(5), 688–701. 27

Rubin, D. B. (1980). Discussion of "randomization analysis of experimental data in the fisher randomization test" by d. basu. *Journal of the American Statistical Association 75*, 591–593. 27

Simard, P. Y., D. Steinkraus, and J. C. Platt (2003). Best practices for convolutional neural networks applied to visual document analysis. In *Seventh International Conference on Document Analysis and Recognition*, Volume 3, pp. 958–963. IEEE. 24

Yang, Y., H. Liu, and X. Chen (2020). Covid-19 and restaurant demand: early effects of the pandemic and stay-at-home orders. *International Journal of Contemporary Hospitality Management 32*(12), 3809–3824. 27

Zigler, C. M. and G. Papadogeorgou (2021). Bipartite causal inference with interference. *Statistical Science 36*(1), 109–123. 5, 18