ECONOMETRICA

JOURNAL OF THE ECONOMETRIC SOCIETY

An International Society for the Advancement of Economic Theory in its Relation to Statistics and Mathematics

https://www.econometricsociety.org/

Econometrica, Vol. 94, No. 5 (September, 2026), 1529–1561

CAUSAL INFERENCE FOR SPATIAL TREATMENTS

MICHAEL POLLMANN
Department of Economics, Duke University

Econometrica, Vol. 94, No. 5 (September, 2026), 1529–1561

CAUSAL INFERENCE FOR SPATIAL TREATMENTS

MICHAEL POLLMANN
Department of Economics, Duke University

Many events and policies (treatments) occur at specific spatial locations, with researchers interested in their effects on nearby units. 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 units near realized treatment locations and units near counterfactual (unrealized) candidate locations, which differs from current empirical practice. I derive design-based standard errors that are straightforward to compute. For observational data, I propose machine learning methods to find counterfactual candidate locations when observable characteristics, rather than potential outcomes, determine treatment probabilities. To accommodate methods for high-dimensional data in the theory, I extend a double machine learning result to the design-based framework with spatial correlations. 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 large positive effect at very short distances, with no effect at larger distances.

KEYWORDS: Causal inference, spatial treatments, design-based, ideal experiment, machine learning, convolutional neural networks.

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 the setting of such studies as the "spatial treatment" setting because these "treatments" vary at the level of 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 on causal inference, each individual is thought to, in principle, be associated with a distinct treatment that generates potential outcomes, and some work considers "spillovers" and "clustered assignment" of individual-level treatments. Such a framework sufficed when treatment and outcome information was only available at aggregated levels, such as the county level. However, more recently, 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.

Michael Pollmann: 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 and three anonymous referees 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 (2021). SafeGraph is a data company that aggregates anonymized location data from numerous applications in order to provide insights about physical places.

[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 SA Table OA1 for examples of papers studying these and other spatial treatments.

 https://doi.org/10.3982/ECTA21571

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 (IPW) estimators and new expressions for their variances, which differ from commonly used existing estimators and sampling-based variances. Third, I provide theoretical results for observational data and propose using convolutional neural networks, previously used for image and satellite data, to parsimoniously yet flexibly condition on the distribution of covariates across space.

The ideal experiment for studying spatial treatments randomizes the location of the treatment among a set of plausible candidate locations. Such an ideal experiment is a formalization of settings where 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] In the ideal experiment, individuals who are located near candidate treatment locations that—by random chance alone—did not receive treatment constitute a valid control group. The ideal experiment does not generally justify comparisons of individuals near realized treatment (on an "inner ring") to individuals farther away but centered around the same treatment locations (on an "outer ring") even when embedded in a difference-in-differences analysis as is common in the empirical literature. The approach developed in this paper therefore presents an alternative identification strategy for spatial treatment settings, leveraging design-based assumptions.

I propose IPW estimators, derive their (approximate) finite population design-based unbiasedness and variance under the ideal experiment, and show asymptotic normality in a sequence of finite but growing populations. In the thought experiment underlying inference, only the realized locations of the treatment vary among a pre-defined set of plausible candidate locations. Other locations and potential outcomes are fixed. The design-based analysis of the variance, following Neyman (1923), has both conceptual and practical advantages over 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. In the design-based analysis, researchers do not need to distinguish between sample and population, which may be a difficult distinction to justify in spatial settings (Pinkse, Shen, and Slade, 2007). Estimating the design-based standard error is straightforward and does not require the correct modeling of the correlation of potential outcomes across space. Importantly, my approach and formulas generalize to settings where individuals are exposed to multiple treatments. In these settings, off-the-shelf alternatives for design-based inference, such as clustering at the level of the assignment (Abadie, Athey, Imbens, and Wooldridge, 2023), are not applicable. In SA 2, I show that different implementations of the variance estimator of Conley (1999) can have undesirable properties in a design-based framework.

I then study an observational data setting where, possibly unknown, treatment probabilities are determined by observable characteristics. Suppose locations in two neighborhoods look identical in terms of their observable pre-treatment characteristics. Then the assumption requires that the treatment is equally likely to be realized in the two locations, irrespective of potential outcomes, similar to an unconfoundedness assumption in

[2]For instance, Linden and Rockoff (2008, p. 1110) state 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." In their empirical analyses, they estimate causal effects based on a parallel trends assumption.

sampling-based analyses. I derive a novel design-based "double machine learning"-type result for spatial treatments that allows estimation of nuisance parameters using arbitrary (for instance, machine learning) methods as long as (weak) rate conditions are satisfied.

To implement flexible estimation based on this assumption on assignment, I propose using convolutional neural networks (CNNs) 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 that is lost in coarse summary statistics. However, controlling for the distribution of units or covariates across space relative to the location of estimation intrinsically is an extremely high-dimensional problem. I propose CNNs that parsimoniously condition on the distribution of covariates across space, incorporating the economic logic that typically only relative, not absolute, locations matter. 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 an increase in the number of visitors to nearby restaurants during COVID-19 shelter-in-place policies. Consumers may find it convenient to grab a coffee or meal while waiting to enter the store or before returning home. Using CNNs, 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.

A nascent methodological literature studies causal inference in spatial treatment and related settings. Zigler and Papadogeorgou (2021) define potential outcomes and estimands. Most closely related, Wang, Samii, Chang, and Aronow (2025) in contemporaneous work explore a similar experimental setting and show design-based unbiasedness, variance, and asymptotic normality for a special case of the experimental design (constant treatment probabilities) and class of estimators (simplifying weights) described in this paper under distinct structural assumptions (restricting treatment effect heterogeneity and ruling out any effect of the treatment at longer distances). In addition to arguably weaker assumptions and more general estimators and designs, I address challenges arising in observational settings. Borusyak and Hull (2023) take a design-based perspective similar to this paper but focus on regression estimators, accommodating multivalued treatments. However, the estimands of their unweighted regressions differ under treatment effect heterogeneity. These papers do not explicitly estimate design-based standard errors in their applications; instead, they report Conley (1999) standard errors. Borusyak and Hull (2023) additionally propose randomization inference with confidence intervals based on the sharp null hypothesis of constant treatment effects.

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 characteristics-determined treatment probabilities using CNNs, while prior work requires the researcher to specify it based on institutional knowledge. Similarly to Borusyak and Hull (2023), the methodological contributions of the present paper are not restricted to spatial settings. The theoretical results for observational data extend existing results on double machine learning (CCDD+, 2018) from the sampling-based framework to the design-based framework, resolving challenges that arise when observations are neither independent nor identically distributed in finite populations.

The remainder of this paper proceeds as follows. Section 2 describes the framework and notation of this paper. Section 3 contains estimation and inference results under the ideal experiment. Section 4 discusses observational data and describes the use of CNNs for finding counterfactual locations.[3] Section 5 applies the methods of this paper to study the effects of grocery stores on foot traffic to nearby restaurants. Section 6 concludes.

The Supplemental Appendix (Pollmann, 2026) contains additional simulations, theoretical derivations, and empirical analyses.

2. 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.[4] 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 individual-level treatment effect compares the outcome of i when there is treatment at location s versus no treatment at s, holding fixed treatments at other locations: $\tau_i(s|S) \equiv Y_i(S \cup \{s\}) - Y_i(S \setminus \{s\})$. $\tau_i(s|S)$ 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: $\tau_i(s) \equiv \tau_i(s|\{s\}) = Y_i(\{s\}) - Y_i(\emptyset)$. For ease of notation, define $Y_i(s) \equiv Y_i(\{s\})$ and $Y_i(0) \equiv Y_i(\emptyset)$. I state the notation and results in this paper in terms of cross-sectional data only. With panel data and staggered treatment adoption, all results remain unchanged under the same ideal experiment after subtracting the corresponding pre-treatment outcome from each (potential) outcome.

The experimental design generates randomness in where the treatment is realized. I use calligraphic letters or hats to denote random variables in contrast to regular and Greek 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 $\mathcal{S} \subset \mathbb{S}$ with probability 1.

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 (which may be asymmetric). Importantly, like "pre-treatment characteristic," the distance used must not depend on treatment assignment.

The estimand of interest is the expected (over the design distribution) average effect of the treatment on the treated (ATT) at a distance of approximately d,

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

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

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

where the weights $w_i(s, d)$ collect individuals at distance approximately d.[5] In practice, researchers often bin individuals within a bandwidth h around d together when distance is a continuous variable using weights $w_i(s, d) \equiv \mathbb{1}\{|d(s, r_i) - d| \leq h\}$. For simplicity of the results, throughout I assume that the weights are known (rather than estimated), $w_i(s, d) \geq 0$, and $w_i(s, d) = 0$ if $d(s, r_i) > D$ for some $D \in \mathbb{R}$ large enough that may depend on d. Let $\mathbb{I}_s = \{i \in \mathbb{I} : w_i(s, d) \neq 0\}$.

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:

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

This effect aggregates the marginal effects of location s given all possible background exposures $S \setminus \{s\}$ (cf. Sävje, Aronow, and Hudgens (2021)). The weights again resemble the ATT, placing more weight on assignments that are more likely to be realized.

Asymptotic results in this paper refer to sequences of finite but growing populations as in, for instance, Abadie, Athey, Imbens, and Wooldridge (2020). All randomness is due to treatment assignment. Along the sequence, populations, indexed by k, grow in the sense that $|\mathbb{S}_k| \to \infty$ and $|\mathbb{I}_k| \to \infty$ as $k \to \infty$. Additionally, for all asymptotic results, I assume outcomes and weights are bounded and weak bounds on the spatial concentration of observations hold as summarized in Assumption 8 in the Appendix, with all bounds uniform over k. Hence, the asymptotic results most plausibly approximate the finite sample behavior of the estimators below when the number of treatment locations is large and spread out in space. For readability, I suppress the dependence on the sequence index in the notation of the main part of the paper.

3. EXPERIMENTAL DATA: ESTIMATION AND INFERENCE

I discuss estimators of average treatment effects on the treated (ATT) for a setting where all data are for a single large region with multiple candidate treatment locations across which treatment is randomized independently.

ASSUMPTION 1—Independent Assignment: *Treatment is assigned to candidate locations independently, with marginal probability* $\pi_s \equiv \Pr(\mathcal{S} \ni s)$ *for location* s. *For* $S \subset \mathbb{S}$,

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

and the probability of treatment is bounded away from 0 *and* 1, $c < \pi_s < 1 - c$ *for some* $c > 0$.

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

[5] In this paper, the choice of weights, such as binning, corresponds to the desired estimand rather than a kernel used to estimate a function at a point.

errors. Even without these assumptions, the estimators estimate meaningful *marginal* effects (as defined in Equation (2)), see Theorem 1(iii), and only some of the structure is needed for the approximate variance of the estimator to remain valid, see Theorem 1(iv).

I first focus on an assumption of additive separability, also studied by Wang, Samii, Chang, and Aronow (2025), before also discussing identification and estimation under an alternative assumption.

ASSUMPTION 2—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, if, to a first approximation, the air pollution due to a power plant (Zigler and Papadogeorgou, 2021) adds pollutants into the air without affecting pollutants added by other power plants, the effects of the plants 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.

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:

$$\hat{\tau}(d) \equiv \frac{\displaystyle\sum_{s \in \mathbb{S}} \mathbb{1}\{\mathcal{S} \ni s\} \sum_{i \in \mathbb{I}} w_i(s, d) \mathcal{Y}_i}{\displaystyle\sum_{s \in \mathbb{S}} \mathbb{1}\{\mathcal{S} \ni s\} \sum_{i \in \mathbb{I}} w_i(s, d)} - \frac{\displaystyle\sum_{s \in \mathbb{S}} \frac{\mathbb{1}\{\mathcal{S} \not\ni s\}}{1 - \pi_s} \pi_s \sum_{i \in \mathbb{I}} w_i(s, d) \mathcal{Y}_i}{\displaystyle\sum_{s \in \mathbb{S}} \frac{\mathbb{1}\{\mathcal{S} \not\ni s\}}{1 - \pi_s} \pi_s \sum_{i \in \mathbb{I}} w_i(s, d)}. \tag{3}$$

To derive standard errors and show asymptotic normality, I assume that treatment effects are local: Treatments have no or relatively small effects on individuals far away from them. Let d_0 be the known distance denoting "far away," specified appropriately by the researcher depending on the treatment and outcome of interest. For individual i, define the approximate exposure mapping (Sävje, 2023) based on the set $\mathbb{M}_i \equiv 2^{\{s \in \mathbb{S}: d(s, r_i) \leq d_0\}}$. Let the random variable $\mathcal{M}_i \in \mathbb{M}_i$ be the realized approximate exposure, determining the realized treatment state of all candidate treatment locations within d_0 of i. Denote i's "potential outcome" under exposure $m \in \mathbb{M}_i$ by $Y_i(m) \equiv E(\mathcal{Y}_i | \mathcal{M}_i = m)$.

ASSUMPTION 3—Limit on Effects After Distance d_0: *Either*

(a) *the treatment has no effect at distances larger than d_0 such that for all $i \in \mathbb{I}$, $S \subset \mathbb{S}$ with $s \in S$: if $d(s, r_i) > d_0$, then $Y_i(S) = Y_i(S \setminus \{s\})$; or*

(b) *the effect of treatments at distances larger than d_0 are small relative to the variation in outcomes due to treatments at shorter distances: $\frac{\epsilon}{\sigma} = o_p(1)$, where $\epsilon \equiv \frac{1}{|\mathbb{S}|} \sum_{s \in \mathbb{S}} \sum_{i \in \mathbb{I}} \frac{w_i(s,d)}{\bar{n}(d)} (\mathbb{1}\{\mathcal{S} \ni s\} - \frac{\mathbb{1}\{\mathcal{S} \not\ni s\} \pi_s}{1 - \pi_s})(\mathcal{Y}_i - Y_i(\mathcal{M}_i))$ with $\bar{n}(d) \equiv \frac{1}{|\mathbb{S}|} \sum_{s \in \mathbb{S}} \pi_s \times \sum_{i \in \mathbb{I}} w_i(s, d)$ the average (per location) expected number of treated individuals, and σ^2 is the variance under (a) of an infeasible estimator $\tilde{\tau}(d) \approx \hat{\tau}(d)$, defined in Theorem 1.*

Assumption 3(b), which is implied by 3(a), allows treatments to affect outcomes of far-away individuals. The term ϵ captures the realized contribution to outcomes of treatment locations farther than d_0 away. By construction, $E(\epsilon) = 0$. In contrast, σ captures the variation due to treatment locations closer than d_0. For instance, individuals not affected by treatments beyond a distance of d_0 have $\mathcal{Y}_i = Y_i(\mathcal{M}_i)$ with probability 1 and therefore contribute only to σ but not to ϵ. Hence, Assumption 3(b) only requires that there is a distance d_0 capturing *most* of the effects.

The following theorem describes the (approximate) finite population properties of $\hat{\tau}(d)$. It uses an approximation to $\hat{\tau}(d)$ that replaces its stochastic denominators by their expectations and recenters its numerators appropriately. Specifically, let

$$\tilde{\tau}(d) \equiv \tau_{\text{marginal}}(d) + \frac{1}{|\mathbb{S}|} \sum_{s \in \mathbb{S}} \sum_{i \in \mathbb{I}} \frac{w_i(s, d)}{\bar{n}(d)} \left(\mathbb{1}\{\mathcal{S} \ni s\}\big(\mathcal{Y}_i - \mu_t(d)\big) - \frac{\mathbb{1}\{\mathcal{S} \not\ni s\}\pi_s}{1 - \pi_s}\big(\mathcal{Y}_i - \mu_c(d)\big) \right),$$

where $\mu_t(d)$ 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 $\tau_i(s|S)$.

THEOREM 1: *The estimator $\hat{\tau}(d)$ is similar to the infeasible estimator $\tilde{\tau}(d)$, which has analytically tractable design-based properties*:

(i) *Under regularity conditions* (*Assumption* 8 *in the Appendix*): $\hat{\tau}(d) - \tilde{\tau}(d) \to_p 0$.

(ii) *Under Assumptions* 1 *and* 2: $E(\tilde{\tau}(d)) = \tau(d)$ (*unbiasedness for ATT*).

(iii) *Under Assumption* 1: $E(\tilde{\tau}(d)) = \tau_{marginal}(d)$ (*unbiasedness for marginal ATT*).

(iv) *Under Assumptions* 1 *and* 3(*a*): *The variance of* $\tilde{\tau}(d)$ *is* $\operatorname{var}(\tilde{\tau}(d)) = \sigma^2 \equiv (\tilde{V}_t(d) + \tilde{V}_c(d) + \tilde{V}_\times(d) - \tilde{V}_{tt}(d) - \tilde{V}_{cc}(d) - \tilde{V}_{ct}(d))/|\mathbb{S}|$ *with the notation defined in Appendix* A.1.

(v) *Under Assumptions* 1, 3(*b*), *and regularity conditions* (*Assumption* 8 *in the Appendix*): $\frac{\tilde{\tau}(d) - \tau_{marginal}(d)}{\sigma} \xrightarrow{d} \mathcal{N}(0, 1)$.

REMARK 1: The variance expression contains conceptually similar terms to the finite population design-based variance of the difference-in-means estimator in standard randomized experiments with individual-level treatments. 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. 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 the finite population design-based variance of the difference-in-means estimator in standard experiments. Dropping the final three terms yields a conservative estimator of the variance because these terms are non-negative by construction.

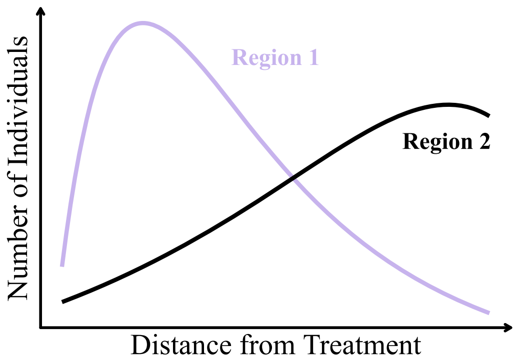
REMARK 2: As a general guideline, to meaningfully reduce the design-based variance of the estimator, one needs to expand the sampling area, rather than the number of individuals or candidate locations within a fixed area. Adding individuals while holding the sample area fixed does not generally reduce 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 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.

REMARK 3: The theorem shows attractive properties of an estimator for the effects of spatial treatments under a *design-based* framework leveraging the randomization Assumption 1 for causal identification. A popular alternative estimator, comparing individuals on an inner ring who are near the treatment to individuals on an outer ring who are farther away from treatment, typically combined with a before versus after comparison in a difference-in-differences setup, is not justified by the same design-based assumption. This lack of a natural design-based interpretation of a particular difference-in-differences estimator in spatial settings is in contrast to the standard difference-in-differences setting and estimator, where a natural form of randomization can ensure parallel trends by design.

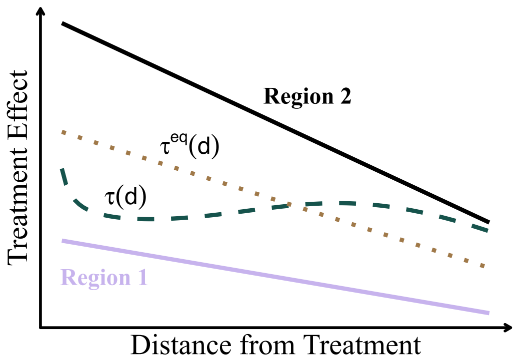
REMARK 4: In contemporaneous work, Wang, Samii, Chang, and Aronow (2025) provide a result similar to Theorem 1 but focus on a simple difference-in-means estimator that arises after averaging outcomes by candidate treatment location and assuming the probability of treatment is constant across candidate locations. The setup of the present paper nests these choices but allows me to derive more general results, covering estimands targeted by existing empirical studies of spatial treatments and accommodating heterogeneous treatment probabilities that are commonly found in the non-experimental settings considered in the following section. Expressions of the design-based variance admitting simple (conservative) estimators without restrictions on treatment effect heterogeneity were first circulated in early versions of the present paper. The asymptotic normality result of Theorem 1(v) allows treatments to have (small) effects even at large distances (Assumption 3(b)), while the earlier result of Wang, Samii, Chang, and Aronow (2025) relies on an analog of the stronger Assumption 3(a) of no effects past some distance. Due to the aggregation of outcomes by candidate treatment location in the analysis of Wang, Samii, Chang, and Aronow (2025), their assumptions are not expressed in terms of the individual-level potential outcomes and locations that are part of the primitive conceptual objects of the present paper. Hence, while closely related, the assumptions are not nested and the formulas do not coincide exactly even when specialized to the estimand they consider.

REMARK 5: When the candidate treatment locations are sufficiently far apart that individuals receiving non-zero weight are affected by at most one location, the setting corresponds to clustered assignment in randomized experiments. Intuitively, Assumption 3 allows for "overlapping clusters" where any individual may be part of an arbitrary number of clusters.

REMARK 6: The variance in Theorem 1(iv) is for $\tilde{\tau}(d)$ as an estimator for the in-sample ATT defined in Equation (2). 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.



(a) Individuals per distance in each region



(b) Estimands

FIGURE 1.—The estimands $\tau(d)$ and $\tau^{\text{eq}}(d)$ can meaningfully differ from one another. Suppose the sample area can be separated into regions such that outcomes in a region are only affected by the single candidate treatment location within the same region. Consider two types of regions, whose candidate treatment locations are equally likely to be treated and which 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 $\tau^{\text{eq}}(d)$, which weights the regions equally and decreases monotonically.

REMARK 7: 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. Second, researchers may deviate from the distance bin weight by choosing a kernel that is continuous in distance, such as triangular weights $w_i^{\text{tri}}(s, d) \equiv (1 - |d - d(s, r_i)|/h)\mathbb{1}\{|d(s, r_i) - d| \leq h\}$. 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 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.

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 the following:

ASSUMPTION 4—Only Nearest Realized Location Matters: *For all* $i \in \mathbb{I}$, $S \subset \mathbb{S}$ *with* $s \in S$, $s' \in \mathbb{S} \setminus \{s\}$: *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 (as studied by Duflo, 2001, using a difference-in-differences design), it may be plausible to assume that only the nearest built 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, Van Nieuwerburgh, and Kontokosta, 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 2: *The average effect of the treatment on the treated*, $\tau(d)$, *is nonparametrically identified if Assumptions* 1, 3(*a*), *and* 4 *are satisfied*.

The proof of Theorem 2 is constructive in that it suggests an estimator that exploits the combination of assumptions:

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

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 SA 6 for additional discussion.

4. OBSERVATIONAL DATA: THEORY AND IMPLEMENTATION

Often, researchers can only study spatial treatments in observational, rather than experimental, data. I first lay out assumptions and theory connecting "quasi-experimental" analysis using observational data to the experimental analysis outlined above. Then, I discuss the practical implementation of estimation under such assumptions with spatial data.

4.1. *Assumptions and Theory*

In Sections 4.1.1, 4.1.2, and 4.1.3, I discuss three internally coherent perspectives on "quasi-experimental" analysis using observational data that guide statistical inference. The perspectives are each founded in existing theoretical work but yield distinct prescriptions for inference and interpretation. The researcher chooses the perspective they find attractive for their application.

For any of the three perspectives, the unknown or hypothetical experiment satisfies an assumption that the probability of treatment depends solely on observable characteristics, not on potential outcomes, aligning with the concept of unconfoundedness in sampling-based analyses. Intuitively, among observationally similar locations, treatment assignment is as good as random. Let Z_s be some observable characteristics of the spatial neighborhood of $s \in \tilde{\mathbb{S}} \subset \mathbb{R}^2$, where $\tilde{\mathbb{S}}$ are the locations under consideration, and let $\mathbb{S} \equiv \{s \in \tilde{\mathbb{S}} : \pi_s > 0\}$ be the (potentially known) set of locations with positive probability of treatment. Define $\mathbb{Z} \equiv \{Z_s : s \in \mathbb{S}\}$ and $\tilde{\mathbb{Z}} \equiv \{Z_s : s \in \tilde{\mathbb{S}}\}$.

ASSUMPTION 5—Characteristics-Determined Probabilities: *Among a known set of locations* $s \in \tilde{\mathbb{S}} \subset \mathbb{R}^2$, *the treatment assignment probabilities depend solely on observed characteristics*, $\pi_s \equiv \Pr(\mathcal{S} \ni s) = p(Z_s)$, *where* Z_s *are observable characteristics of the spatial neighborhood of* s *and* $p : \tilde{\mathbb{Z}} \to [0, 1]$ *is a (possibly unknown) function*.

In spatial settings, researchers may commonly wish to invoke Assumption 5 based on a relatively large set of (or "high-dimensional") characteristics Z_s. These characteristics may include not just information about the point s itself, but also information about the surrounding neighborhood.

The different perspectives on inference outlined below each describe the (asymptotic) distribution $F_{\hat{\tau},P}$ of an estimator $\hat{\tau}$ over the treatment assignment distribution P. The perspectives differ in their choices for $\hat{\tau}$ and P.

Each of the results in this section also invokes Assumption 1 (independent assignment) to limit spatial dependence in the data. If latent treatment probabilities depend on spatially correlated characteristics, incorrectly omitting such characteristics from the conditioning set when invoking Assumption 5 can lead not only to bias but also to incorrect standard errors. Specifically, treatment assignment may behave as if it was spatially correlated rather than independent in that case.

4.1.1. *Latent Experiment Perspective*

The data are the product of a latent experiment with unknown treatment probabilities π_s. However, because treatment probabilities are deterministic functions of observable characteristics by Assumption 5, one may be able to consistently estimate the π_s under Assumption 1. The experimental estimator discussed in the previous section is then consistent by standard arguments even in observational settings. However, inference needs to account for (design-based) uncertainty in the estimated treatment probabilities: If the latent experiment had resulted in a different set of realized treatment locations, the estimated treatment probabilities could be different. Because the true experiment is unknown, Assumption 5 typically needs to hold for $\tilde{\mathbb{S}}$ the subset of $\mathbb{R}^2$ corresponding to the study area, while the unknown $\mathbb{S} \equiv \{s \in \mathbb{R}^2 : \pi_s > 0\}$ is typically finite.

I show that for a class of estimators, the uncertainty in estimated treatment probabilities does not affect the asymptotic distribution of the estimators to first order. Extending recent theoretical advances (for instance, CCDD+ (2018)) to a design-based finite population framework with spatial dependence, I study the estimator $\hat{\tau}_{\text{dml}}(d) = \hat{\tau}(d, \hat{\pi}, \hat{\mu})$ where

$$\hat{\tau}(d, p, m) = \frac{\sum_{i \in \mathbb{I}} \sum_{s \in \mathbb{S}} w_i(s, d)\left(\mathbb{1}\{\mathcal{S} \ni s\} - \dfrac{\mathbb{1}\{\mathcal{S} \not\ni s\}}{1 - p_s} p_s\right)\left(\mathcal{Y}_i - m_{s,i}(d)\right)}{\sum_{i \in \mathbb{I}} \sum_{s \in \mathbb{S}} w_i(s, d)\mathbb{1}\{\mathcal{S} \ni s\}}$$

with a choice of first step estimates for p and m, $\hat{\pi}$ and $\hat{\mu}$, based on cross-fitting described below. With (infeasible) $\pi \equiv (\pi_s)_{s \in \mathbb{S}}$ and μ such that, using covariates $Z_{s,i}$ containing Z_s,

$$\mu_{s,i}(d) \equiv \frac{\sum_{s' \in \mathbb{S}} \sum_{i' \in \mathbb{I}} w_{i'}(s', d)\mathbb{1}\{Z_{s',i'} = Z_{s,i}\}E\left(\mathcal{Y}_{i'} | \mathcal{S} \not\ni s'\right)}{\sum_{s' \in \mathbb{S}} \sum_{i' \in \mathbb{I}} w_{i'}(s', d)\mathbb{1}\{Z_{s',i'} = Z_{s,i}\}},$$

the estimator $\hat{\tau}(d, \pi, \mu)$ is similar to the estimator $\hat{\tau}(d)$ in Equation (3) with "demeaned" outcomes, and a result analogous to Theorem 1 applies.

The first step estimates $\hat{\pi}_s$ and $\hat{\mu}_{s,i}$ are based on a random split of the candidate treatment locations into F folds, with F fixed, denoted by the partition $(\mathbb{S}_f)_{f=1}^F$ of $\mathbb{S}$ with $f(s)$ the fold of s. Create estimation samples containing observations that are design-based independent of observations in a fold f as follows. For $I \subset \mathbb{I}$, let $\mathbb{S}_I \equiv \{s \in \mathbb{S} : d(s, r_i) \leq d_0 \text{ for some } i \in I\}$ be the locations that may affect individuals in I under Assumption 3(a). Let $\mathbb{I}_f \equiv \{i \in \mathbb{I} : d(s, r_i) \leq d_0 \text{ for some } s \in \mathbb{S}_f\}$ be the individuals who may be affected by treatment at locations in fold f under Assumption 3(a). Let $\tilde{\mathbb{I}}_f \equiv \bigcup_{s \in \mathbb{S}_f} \mathbb{I}_s$ be the individuals who receive weight by the estimand for being near a location in fold f. The estimate $\hat{\pi}_s$ is based only on candidate locations $s' \in \mathbb{S} \setminus (\mathbb{S}_{f(s)} \cup \mathbb{S}_{\tilde{\mathbb{I}}_{f(s)}})$ that are neither in the same fold nor affect individuals receiving weight in the estimand for being near a location in the same fold. The estimate $\hat{\mu}_{s,i}$ is based only on individuals $i' \in \mathbb{I} \setminus (\mathbb{I}_{f(s)} \cup \tilde{\mathbb{I}}_{f(s)})$ who are neither affected by locations in the same fold nor receive weight in the estimand for being near a location in the same fold. The estimation samples are weighted random samples: locations and individuals far away from other locations are more likely to be included, but sample splitting is random. Hence, it is straightforward to reweight observations inversely by their inclusion probabilities, which, intuitively, are related to the probability that all locations are not in fold $f(s)$. These sample splitting estimates ensure design-based independence of $(\hat{\pi}_s, \hat{\mu}_{s,i})$ and $(\mathbb{1}\{\mathcal{S} \ni s\}, \mathcal{Y}_i)$ under Assumptions 1 and 3(a). The individual folds cannot differ too much from the full sample, in the sense that, for each fold f, $\frac{1}{|\mathbb{S}|} \sum_{s \in \mathbb{S}_f} \sum_{i \in \mathbb{I}_s} (\tilde{\mu}_{s,i} - \mu_{s,i})^2 = O_p(|\mathbb{S}|^{-1/2})$, where $\tilde{\mu}_{s,i} = \frac{\sum_{s' \in \mathbb{S}_{f(s)}} \sum_{i' \in \mathbb{I}} w_{i'}(s') \mathbb{1}\{Z_{s',i'} = Z_{s,i}\} E(\mathcal{Y}_{i'} | \mathcal{S} \not\ni s')}{\sum_{s' \in \mathbb{S}_{f(s)}} \sum_{i' \in \mathbb{I}} w_{i'}(s') \mathbb{1}\{Z_{s',i'} = Z_{s,i}\}}$ similar to μ but averaging only within fold $f(s)$. Assumption 8(d) in the Appendix formally imposes this condition. For random sample splits, this bound is attained if the number of distinct values $Z_{s,i}$ takes on is $O(\sqrt{|\mathbb{S}|})$.

ASSUMPTION 6: *The first step estimates using sample splits described above converge sufficiently quickly such that*: (*i*) $\sqrt{|\mathbb{S}|^{-1} \sum_{s \in \mathbb{S}} \sum_{i \in \mathbb{I}_s} (\hat{\mu}_{s,i} - \mu_{s,i})^2} = o_p(|\mathbb{S}|^{-1/4})$, (*ii*) $\sqrt{|\mathbb{S}|^{-1} \sum_{s \in \mathbb{S}} (\hat{\pi}_s - \pi_s)^2} = o_p(|\mathbb{S}|^{-1/4})$, *and* (*iii*) *with probability* 1, $\hat{\pi}_s \in [0, 1-c]$ *for some* $c > 0$, *and* $\hat{\mu}_{s,i}$ *is bounded for all* s *and* i.

THEOREM 3: *Suppose Assumptions* 1, 3(*a*), 5, *and* 6, *as well as regularity conditions* (*Assumption* 8 *in the Appendix*), *hold, and* $(\hat{\tau}(d, \pi, \mu) - \tau_{marginal}(d))/\sigma \xrightarrow{d} \mathcal{N}(0, 1)$ *where* $\sigma^2 \equiv \operatorname{var}(\tilde{\tau}(d, \pi, \mu))$, *with* $\tilde{\tau}(d, p, m)$ *equal to* $\hat{\tau}(d, p, m)$ *except that the denominator is replaced by its expectation*, *is such that* $1/\sigma = O(\sqrt{|\mathbb{S}|})$. *Then*,

$$\frac{\hat{\tau}(d, \hat{\pi}, \hat{\mu}) - \tau_{marginal}(d)}{\sigma} \xrightarrow{d} \mathcal{N}(0, 1).$$

REMARK 8: The theorem describes the asymptotic distribution $F_{\hat{\tau}(d,\hat{\pi},\hat{\mu}),P}$ of the estimator $\hat{\tau}(d, \hat{\pi}, \hat{\mu})$ (including nuisance function estimation) across the treatment assignment distribution P arising from independent assignment according to the marginal probabilities π_s of the true latent experiment.

REMARK 9: The variance in the theorem does not depend on the quality of the estimators of the first step estimates $\hat{\pi}$ and $\hat{\mu}$ as long as they satisfy the rate conditions. Hence, when using this estimator of the ATT, one can effectively "ignore" the noise due to estimated treatment probabilities at the cost of also requiring an estimate of the conditional mean.

REMARK 10: The assumption on the infeasible $\hat{\tau}(d, \pi, \mu)$ typically holds based on an analog of Theorem 1.

The rate conditions in the theorem above allow for growing dimensionality of characteristics. For notational simplicity, I derive convergence rates (Theorems 4 and 5) for estimators without sample-splitting.

For a simple example of the estimator $\hat{\pi}$, let $\hat{\pi}_s = \sum_{s'\in\tilde{\mathbb{S}}} \mathbb{1}\{Z_{s'} = Z_s\}\mathbb{1}\{\mathcal{S} \ni s'\}/n_{Z_s}$, where $n_z = \sum_{s\in\tilde{\mathbb{S}}} \mathbb{1}\{Z_s = z\}$, be the average treatment status of locations with the same characteristics as s.

THEOREM 4: *Under Assumptions 1 and 5, if* $|\mathbb{Z}| = o(\sqrt{|\mathbb{S}|})$, *then* $\sqrt{|\mathbb{S}|^{-1} \sum_{s\in\mathbb{S}} (\hat{\pi}_s - \pi_s)^2} = o_p(|\mathbb{S}|^{-1/4})$, *and* $\hat{\pi}_s = 0$ *with probability* 1 *for* $s \in \tilde{\mathbb{S}} \setminus \mathbb{S}$.

REMARK 11: The particular estimator $\hat{\pi}$ above is most appropriate when the characteristics Z_s correspond to discrete types. Intuitively, one may think of the CNNs proposed below as defining locations to have the same type if their neighborhoods, discretized through a fine grid, are identical up to rotation and mirroring.

REMARK 12: The convergence rate does not depend on the cardinality of the set of characteristics $\tilde{\mathbb{Z}} \setminus \mathbb{Z}$ of locations where the unknown probability of treatment is zero. This result is useful because in spatial settings there may often be very many or even infinitely many possible locations, and hence many possible characteristics, with almost all never appearing with any treatment.

For the conditional outcome mean, consider a parametric assumption such as $\mu_{s,i} = Z_{s,i}\beta$ for L-dimensional bounded row vector $Z_{s,i}$. Under standard assumptions, one may estimate β by weighted regression of $\mathcal{Y}_i$ on $Z_{s,i}$ using each $(s, i) \in \mathbb{S} \times \mathbb{I}$ pair with $w_i(s, d) \neq 0$ and $s \notin \mathcal{S}$ as observations with weights $w_i(s, d)$. In settings where the number of covariates grows (slowly) along the sequence of finite populations, the theorem below considers the estimator including a LASSO penalty under a sparsity assumption. To define the LASSO estimator formally, define the vector $\boldsymbol{\mathcal{Y}}$ and matrix $\boldsymbol{\mathcal{Z}}$ to include all pairs $(s, i) \in \mathbb{S} \times \mathbb{I}$ with $w_i(s, d) \neq 0$, such that for row j, $\mathcal{Y}_j \equiv \sqrt{w_{i(j)}(s(j), d)}\mathbb{1}\{\mathcal{S} \not\ni s(j)\}\mathcal{Y}_{i(j)}$ and $\mathcal{Z}_j \equiv \sqrt{w_{i(j)}(s(j), d)}\mathbb{1}\{\mathcal{S} \not\ni s(j)\}Z_{s(j),i(j)}$. Both $\boldsymbol{\mathcal{Y}}$ and $\boldsymbol{\mathcal{Z}}$ are random due to their dependence on $\mathcal{S}$. Let $\dot{n}$ be the number of rows of $\boldsymbol{\mathcal{Y}}$, and λ the penalty parameter. Then

$$\hat{\beta} \in \arg\min_{\beta\in\mathbb{R}^L} \frac{1}{2\dot{n}}(\boldsymbol{\mathcal{Y}} - \boldsymbol{\mathcal{Z}}\beta)'(\boldsymbol{\mathcal{Y}} - \boldsymbol{\mathcal{Z}}\beta) + \lambda\sum_{l=1}^{L}|\beta_l|, \qquad \hat{\mu}_{s,i} = Z_{s,i}\hat{\beta}.$$

The formal result below, bounding the estimation error, uses a restriction of spatial dependence, which, effectively, is a strengthened form of Assumption 3(b). The condition groups location-individual pairs (s, i) into clusters $c(s, i) \in \mathbb{C}$ such that dependence across clusters is limited. For pair (s, i), define the approximate exposure mapping based on the set $\mathbb{M}_{s,i} \equiv 2^{\{s'\in\mathbb{S}: c(s,i)=c(s',i') \text{ for some } i'\}}$. Let the random variable $\mathcal{M}_{s,i} \in \mathbb{M}_{s,i}$ be the realized approximate exposure, determining the realized treatment state of all candidate treatment locations in the same cluster. Denote the (s,i) "potential outcome" under exposure $m \in \mathbb{M}_{s,i}$ by $Y_{s,i}(m) \equiv E(\mathcal{Y}_i|\mathcal{M}_{s,i} = m)$, and the vector of weighted potential outcomes by $\boldsymbol{Y}^{\mathcal{M}}$ with $\boldsymbol{Y}^{\mathcal{M}}_j \equiv \sqrt{w_{i(j)}(s(j),d)}\mathbb{1}\{\mathcal{S} \not\ni s(j)\}Y_{s(j),i(j)}(\mathcal{M}_{s(j),i(j)})$.

ASSUMPTION 7—Many approximate clusters: *Observations* $(s,i) \in \mathbb{S}\times\mathbb{I}$ *with* $w_i(s,d) \neq 0$ *can be grouped into approximate clusters* $c(s,i) \in \mathbb{C}$ *such that* (*i*) *the number of observations per cluster is bounded*, (*ii*) $c(s,i) = c(s,i')$ *for each* s *and any* i, i', *and* (*iii*) $\max_{l=1,\dots,L}\{|\frac{\mathcal{Z}'_{\cdot,l}(\mathcal{Y}-\boldsymbol{Y}^{\mathcal{M}})}{\dot{n}}|\} = o_p(\sqrt{\ln(L)/|\mathbb{C}|})$.

In Assumption 7(iii), $\mathcal{Y} - \boldsymbol{Y}^{\mathcal{M}}$ computes how much treatments outside an observation's cluster affect its outcome under the realized assignment. These out-of-cluster effects are multiplied with the regressors and averaged over observations. The LASSO relies on the largest covariance between regressors and residuals, $\max_l\{|\frac{\mathcal{Z}'_{\cdot,l}(\mathcal{Y}-\mathcal{Z}\beta)}{\dot{n}}|\}$, vanishing not too slowly, where $\mathcal{Z}_{\cdot,l}$ is the lth column of the regressor matrix $\mathcal{Z}$. When $\max_l\{|\frac{\mathcal{Z}'_{\cdot,l}(\boldsymbol{Y}^{\mathcal{M}}-\mathcal{Z}\beta)}{\dot{n}}|\} = O_p(\sqrt{\ln(L)/|\mathbb{C}|})$, as in Theorem 5 below, Assumption 7(iii) effectively states that variation in the residual $\mathcal{Y} - \mathcal{Z}\beta$ due to assignments outside the cluster is of smaller order of magnitude than variation due to assignments within the cluster. Assumption 7(ii) implies that the treatment indicator is independent across clusters if also Assumption 1 holds. Assumption 7(i) ensures that, for sequences of finite populations where the number of candidate locations and the number of individuals grow at the same rate, the number of clusters grows sufficiently fast.

THEOREM 5: *Suppose Assumptions* 1, 5, *and* 7, *as well as regularity conditions* (*Assumption* 8 *in the Appendix*), *hold. If the following two conditions hold*:

1. *linear conditional mean*: $\mu_{s,i} = Z_{s,i}\beta$ *for all* $(s,i) \in \mathbb{S}\times\mathbb{I}$ *with* $w_i(s,d) \neq 0$,
2. *restricted eigenvalues*: *w.p. approaching* 1 *and* $\kappa > 0$ *bounded away from zero*, $\frac{1}{\dot{n}}\|\mathcal{Z}x\|_2^2 \geq \kappa\|x\|_2^2$ *for all* $x \in \mathbb{R}^L$ *such that* $\|x_{\beta=0}\|_1 \leq 3\|x_{\beta\neq 0}\|_1$ *where* $x_{\beta=0}$ *and* $x_{\beta\neq 0}$ *are subvectors restricted to entries where corresponding entries* β *are* (*not*) *zero and* $\|a\|_p$ *denotes the* ℓ_p *norm of the vector* a,

then, *for a sequence of finite populations with* $|\mathbb{S}|$, $|\mathbb{I}|$, *and* $|\mathbb{C}|$ *growing at equal rates and* L *growing*, *there exist constants* $C, \tilde{C} \in \mathbb{R}$ *such that the LASSO with penalty parameter chosen as* $\lambda = \tilde{C}\sqrt{\ln(L)/|\mathbb{S}|}$ *yields estimation error* $\sqrt{\sum_{l=1}^{L}(\hat{\beta}_l - \beta_l)^2} \leq C\sqrt{\|\beta\|_0\ln(L)/|\mathbb{S}|}$ *with probability approaching* 1, *where* $\|\beta\|_0$ *is the number of non-zero elements in* β.

REMARK 13: Theorem 5 shows that, under the finite population design-based framework of this paper, the LASSO achieves its usual convergence rate, allowing for a growing number of regressors assuming sparsity. For prediction errors $\hat{\mu}_{s,i} - \mu_{s,i}$, if $\sum_{l=1}^{L} Z_{s,i,l}^2$ remains bounded as L grows (e.g., a growing number of types where, for a given observation, only the fixed subset of regressors corresponding to its own type are non-zero), then Theorem 5 implies Assumption 6(i) as long as $\|\beta\|_0\ln(L) = o(\sqrt{|\mathbb{S}|})$.

REMARK 14: The linear conditional mean assumption can be ensured to hold by saturating the regression in settings with discrete covariates. Sparsity allows the estimator to achieve low error even if the number of regressors grows along the sequence of finite populations. The restricted eigenvalues assumption is standard in the literature on the LASSO. When $\mathcal{Z}'\mathcal{Z}$ is invertible but all entries of β are non-zero, the assumption bounds the smallest eigenvalue away from zero such that the typical rank condition (from fixed-L asymptotics) holds even in the limit; sparsity of β weakens the assumption.

4.1.2. *Hypothetical Experiment Perspective*

The researcher identifies a particular set of counterfactual locations, determines hypothetical treatment probabilities, and analyzes the data as if these locations and probabilities were known properties of an experiment as in Section 3. The researcher uses institutional knowledge and computational techniques, such as those outlined in the following section, to identify a control group that is plausibly comparable to the observed group of treated observations. Because the hypothetical treatment probabilities π_s are determined by the researcher, they generally satisfy Assumption 5 of being based on characteristics observable to the researcher. The approach then treats the data as resulting from the corresponding experiment satisfying Assumptions 1 and 3. There is no assumption that such an experiment occurred in the past; it is solely a hypothetical device for constructing the "repeated sampling" thought experiment underlying inference. This perspective acknowledges that the thought experiment for design-based inference around in-sample causal effects is inherently divorced from any feasible or observable procedure. In contrast to repeated sampling from a larger population, which at least theoretically can be feasible in some settings, it is entirely impossible to carry out the design-based *repetition* in practice. Hence, the researcher justifies inference as reflecting one particular hypothetical thought experiment.

A convenient choice for the hypothetical experiment is the ideal experiment of Section 3 because it is feasible to study and has the attractive design-based properties discussed above. Formally, the researcher appeals to the properties stated in Theorem 1 as describing the (asymptotic) distribution $F_{\hat{\tau}(d),P}$ of the estimator $\hat{\tau}(d)$ as in Equation (3) (with π_s fixed as determined by the researcher) across the treatment assignment distribution P arising from independent assignment according to the marginal probabilities π_s determined by the researcher.

Procedurally, inference under this perspective is similar to constructing a matched sample based on estimated treatment probabilities and then taking the sample as given (for instance, Imbens and Rubin, 2015, Chapter 17.6). By never using outcome data in identifying the counterfactual locations and determining hypothetical treatment locations, this necessary first step for observational data "cannot intentionally introduce systematic biases in the subsequent analyses for causal effects on outcomes" (Imbens and Rubin, 2015, p. 374).

4.1.3. *Conditional Inference Perspective*

The researcher assumes an experiment took place, but inference is not based on the full unknown experimental assignment distribution. Instead, the researcher makes Assumption 5 with a convenient parametric functional form but an unknown parameter. By conditioning on a sufficient statistic for the unknown parameter, the inferential distribution no longer depends on the true parameter value. In other words, standard errors express the amount of variation in the estimator across assignments that lead to the same value of the sufficient statistic.

Formally, the researcher appeals to the properties stated in Theorem 1 as describing an (asymptotic) distribution $F_{\hat{\tau}(d),P}$. The estimator $\hat{\tau}(d)$ is as in Equation (3) but with estimated $\hat{\pi}_s$ given the realized assignment in place of the unknown π_s. The distribution P is the *conditional* distribution arising from restricting the true latent experiment to assignments for which $\hat{\pi}_s$ takes on the values observed in the sample and the marginal treatment probabilities of the conditional distribution coincide with $\hat{\pi}_s$. Despite the researcher acknowledging that $\hat{\pi}_s$ is estimated from the sample, the conditional inference thought experiment can hold $\hat{\pi}_s$ fixed across samples, simplifying finite sample inference as in the ideal experiment of Section 3. However, even if the true latent experiment features independent assignment, the conditional distribution may feature dependence. Nevertheless, some of the results of Theorem 1 may apply, as illustrated with a simple example below.

For the simplest example, consider an experiment according to Assumption 1 with a constant probability of treatment for locations in a known set $\mathbb{S}$, and zero probability of treatment elsewhere. The full assignment distribution depends on the unknown probability of treatment. However, conditional on the number of treated locations in the sample, the assignment distribution is known to be a uniform distribution over all assignments that hold the number of treated locations fixed. Hence, conditional inference treating the estimated (constant) treatment probabilities as fixed is valid.

The results of Theorem 1 remain useful for conditional inference in the example: The variance assuming independent assignment is conservative in expectation (by the law of total variance), but asymptotically the designs are equivalent (Hájek, 1960). The example readily extends to settings where the researcher observes several candidate locations for each value of the characteristics in Assumption 5. When many locations have distinct characteristics, but the treatment probability function takes a logistic functional form, conditional inference via permutation tests of sharp null hypotheses is possible (Rosenbaum, 1984).

4.2. *Finding Counterfactual Treatment Locations Using Convolutional Neural Networks*

I propose using CNNs to identify plausible counterfactual treatment locations that are observationally similar to realized treatment locations. The particular implementation of CNNs I advocate for is well-suited for spatial settings because it has two key features. First, it allows the prediction of counterfactual locations to depend flexibly on the spatial configuration of characteristics in the neighborhoods around the locations. Second, it is computationally feasible despite a very large number of possible locations to consider and very high-dimensional covariates.

The convolution operation, together with input data augmentation (Simard, Steinkraus, and Platt, 2003), implements the idea that the spatial distribution of characteristics relative to a location is often important for the location's plausibility as a counterfactual treatment location. In Assumption 5, characteristics Z_s of location s may contain the local values $v_{x,y}$ and relative locations of points (x, y) around s. The convolution operation f on a grid $\boldsymbol{v}$ of input values $v_{x,y}$ is

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

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), CNNs can be more parsimonious than fully connected neural networks and enforce equivariance to shift. Using multiple layers of convolutions combined with nonlinear activation functions at each grid cell allows the network to learn nonlinear relationships. Augmenting data by spatially shifting, rotating, and mirroring the input effectively imposes an equivariance with respect to these operations. Equivariance formalizes the economic logic that relative locations of spatial features and characteristics matter, rather than their absolute locations and orientations. With standard methods, such as logistic regression, it appears challenging to incorporate a similarly flexible yet equivariant relationship between the output at a point and the characteristics measured at spatial locations near it. Furthermore, the approach using CNNs is computationally feasible because the categorical predictions compute estimates for many locations (grid cells) at once, and stochastic gradient descent, used to estimate the CNN parameters, limits memory requirements and improves speed in the presence of very high-dimensional spatial data used as predictors. I describe and discuss the use of CNNs in more detail in SA 3.

I train a CNN to distinguish between realized locations of the treatment and other locations based on observable characteristics. The key insight is that spatial data can typically be "plotted on a map," and hence spatial data can resemble image data, for which CNNs have enjoyed recent popularity (Krizhevsky, Sutskever, and Hinton, 2012). The input to the CNN is a 3D tensor: a fine 2D discretization of space approximately centered around the real or possible counterfactual location, combined with a third dimension that enumerates the values of every characteristic measured for this grid cell. I design the objective function of the CNN to partly resemble a generative adversarial network (GAN, GPMX+ (2014)) while maintaining the more easily trained structure of finite categorical prediction: The output of the CNN is an "activation score" for each categorical prediction of whether ("discriminator") and in which spatial grid cell ("generator") there is a real treatment location. A GAN is attractive because its generator draws from the *modes* ("most plausible") of the treatment location distribution across space rather than estimating mean locations (Goodfellow, 2016, Lotter, Kreiman, and Cox, 2016), and the GAN implicitly maintains an internal estimate of the distribution of the treatment across space, which here resembles the treatment probabilities.

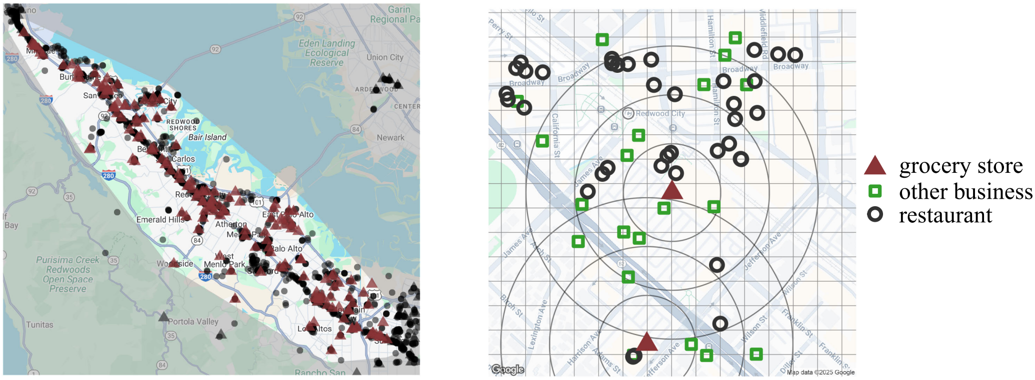
After training, the "false positives" of this algorithm (counterfactual locations drawn from the generator) are grid cells without real treatment locations that the CNN could not distinguish from real treatment locations based on observable characteristics. Under Assumption 5, observationally identical locations are a design-based valid control group. More precisely, I use matching on the activation score to create a sample consisting of the real treatment locations and those locations with the most similar activation scores, akin to propensity score matching for sample construction. I then estimate the treatment probability using only the matched sample, effectively setting the estimated treatment probabilities of all other locations to zero, and analyze the observational data following one of the perspectives above.

5. 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.[6] Here, grocery store locations are the treatment locations, and restaurants are the individuals for whom I estimate average effects by distance from treatment. For examples of existing empirical studies involving spatial treatments, see SA Table OA1.

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. For each business appearing in these data, SafeGraph records latitude and

[6]See SA 4 for the exact definitions of grocery stores and restaurants used in this analysis.



(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)). Grocery stores in study area: solid triangles (167); outside (considered fixed): black triangles. Restaurants: black circles (1627 within 0.5 miles of real or counterfactual grocery stores). Panel (b) zooms in on a location in Redwood City, also indicating locations of other businesses (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.

longitude, NAICS industry code, and the number of visitors in a given week whose smartphone location data are available to SafeGraph. SafeGraph's proprietary algorithm defines visits based on the location of the smartphone, time spent, relative locations of other businesses, and time of day. To cover the South Bay Area, I select grocery stores near Burlingame (within 3 miles of the city center), Belmont (5 miles), Menlo Park (5.5 miles), and Mountain View (2.95 miles). I manually confirm that all grocery stores used in the analysis were indeed open during the study period (the week starting April 13, 2020) and set their latitude and longitude to the front door. For restaurants, I remove duplicate observations due to typos in restaurant names and compare SafeGraph coded latitude and longitude to geocoding by Google Maps. I restrict businesses to those with at least 7 visits, as recorded by SafeGraph, in each of the first four calendar weeks of 2020 to focus on businesses that were open before the pandemic. Overall, I use 167 real and 162 counterfactual grocery store locations, with 1,894 and 1,612 instances of restaurants within 0.2 miles, respectively. These instances arise from 752 unique restaurants with a median of 8 visits in the SafeGraph data for the study period. For additional detail on sample construction, see SA 4.1.

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, Liu, and Chen, 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 Assumption 5 that assignment probabilities depend only on observable characteristics, not potential outcomes. Hence, restaurants in neighborhoods differing in their number of grocery stores, but similar in terms of all other kinds of businesses and observable characteristics, would in expectation have comparable numbers of visitors if they had similar exposure to grocery stores. The assumption may be plausible because restaurants chose their locations based on pre-pandemic potential outcomes, if at all. Restaurants that located before the pandemic are unlikely to have (accurately) predicted and sorted based on potential foot traffic patterns during shelter-in-place policies. Even pre-pandemic, grocery store locations may not have been the primary concern for restaurants, holding locations of all other businesses fixed. See SA 7 for empirical analyses assessing the identifying assumption.

For this application, I make Assumption 5 using the relative locations of businesses by industry as the observable characteristics Z_s, which form the input into the CNN and hence treatment probability estimation. Panel (b) of Figure 2 illustrates these controls by plotting as squares other businesses near a particular grocery store in the sample. I superimpose a grid with cells of size 0.025mi × 0.025mi that shows the discretization used by the CNN. I divide these other businesses into seven groups by their four-digit NAICS code, as listed in SA Table OA3, and the count of businesses by industry for each grid cell is used as a covariate. 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. When 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. The CNN passes the spatial grid of businesses by industry through four sequential 2D convolutions, calculating 16, 36, 36, and 1 weighted averages ("channels") of the neighborhood of each point based on the output of the previous layer, with one final fully connected layer. Based on a matched sample of real grocery store locations and counterfactual locations predicted by the CNN, I estimate treatment probabilities based on the number of restaurants and grocery stores by distance from each location using logistic regression. I give a complete description of the implementation of the CNN and treatment probability estimation in SA 3 and 4.

Researchers can assess the plausibility and quality of the counterfactual grocery store locations predicted by the neural network and the estimated treatment probabilities by considering summary statistics of balance and the concentration of the difference in exposure to real grocery stores. 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 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 treatment probability 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.

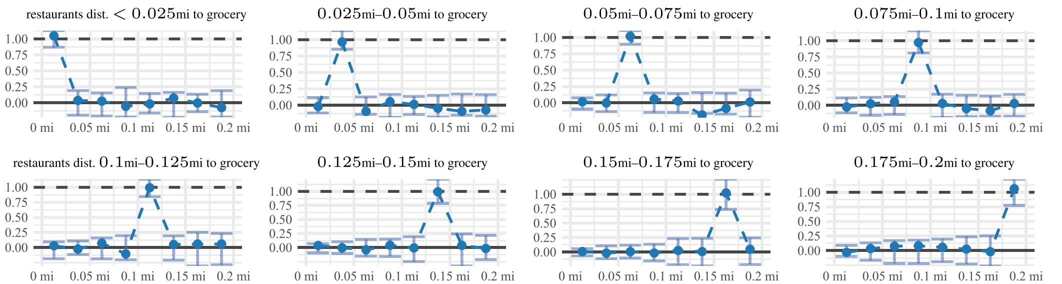


FIGURE 3.—Differential exposure of restaurants to grocery stores. Each plot shows the difference in the average number of grocery stores at multiple distances (horizontal axis) between treated and untreated restaurants. Columns show restaurants in different distance bins around candidate grocery store locations. To give a sense of statistical uncertainty in differential exposure, I take 10,000 draws from the treatment assignment distribution (given by Assumptions 1 and 5) and display error bars covering the differential exposure realized in 95% of these draws.

Figure 4 shows that other observable characteristics are balanced between the neighborhoods of treated and untreated restaurants at different distances. The top left plot focuses on restaurants within 0.025 miles of real and counterfactual grocery stores. For

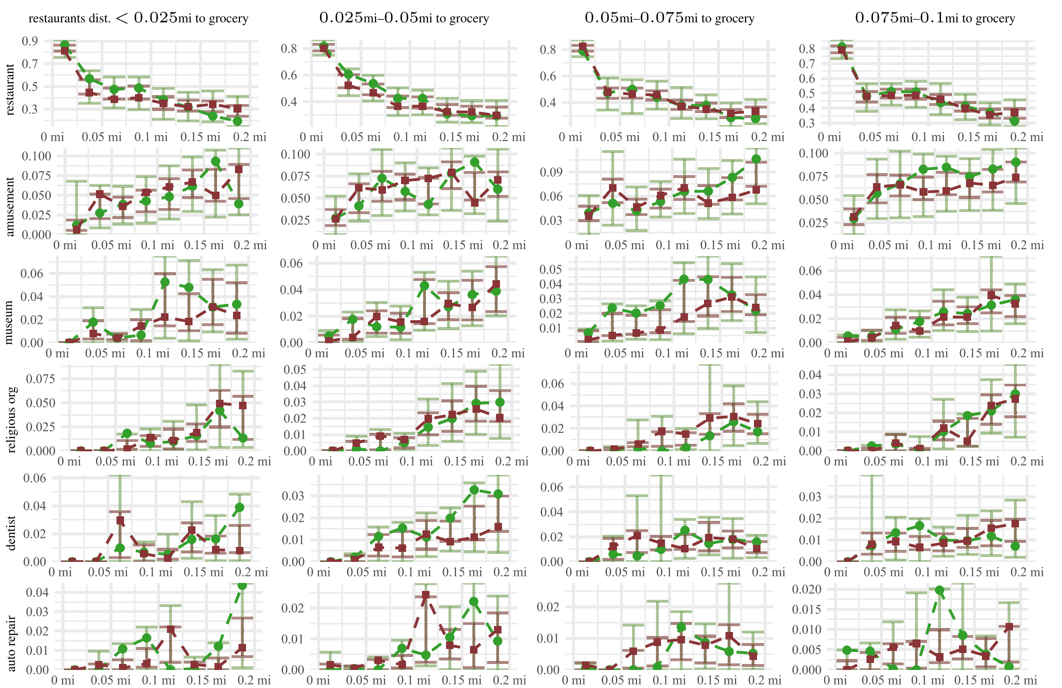


FIGURE 4.—Balance in neighborhood characteristics. Each plot shows the fraction of businesses with, in different rows, 4-digit NAICS 7225, 7139, 7121, 8131, 6212, or 8111 at multiple distances (horizontal axis) from treated (squares) and untreated (circles) restaurants. Columns show restaurants in different distance bins around candidate grocery store locations. For restaurants at longer distances from grocery stores, see SA Figure OA1. To give a sense of statistical uncertainty in characteristics balance, I take 10,000 draws from the treatment assignment distribution (given by Assumptions 1 and 5) and display error bars covering the balance in characteristics realized in 95% of these draws.

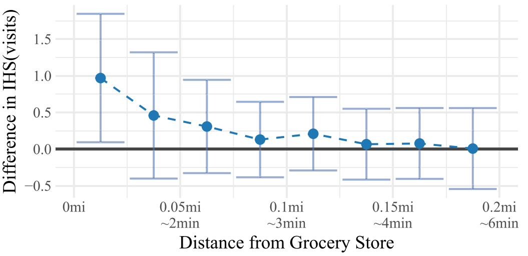


FIGURE 5.—Estimated average effect of grocery stores on restaurants at different distances. The outcome is the inverse hyperbolic sine of the number of visits as recorded by SafeGraph. Bars indicate ±1.96 standard errors.

each treated or untreated restaurant, I compute the fraction of businesses that fall into particular industries within different distance bins from the restaurant (horizontal axis). Squares show the average fraction among the treated restaurants. Circles show the ATT-weighted average fraction among the untreated restaurants. Columns focus on treated and untreated restaurants at different distances from grocery stores. Rows show the neighborhood proportions of different industries. Overall, the reasonable balance in business composition patterns across distances suggests that the neural networks succeeded in finding counterfactual locations similar to real grocery store locations. Note that restaurants, amusement, museums, and religious locations are used as predictors in the neural network, but dentists and automotive businesses are not.[7] Except for the count (not fraction) of restaurants, none of these industries are used in the treatment probability estimation, such that the balance shown in the figure is not mechanical.

Figure 5 displays estimated effects and standard errors for the estimator given in Equation (3). Standard errors indicate the variation in the estimate expected from reassigning treatment according to the fixed estimated treatment probabilities following the "hypothetical experiment perspective" of Section 4, under Assumption 1 of independent assignment and Assumption 3 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 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 2), in which case average marginal effects equal average effects.

I find large positive effects of being located very close to a grocery store, with no effect past a few minutes of walking. Table I 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

[7]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.

TABLE I

ESTIMATED EFFECTS ON THE NUMBER OF VISITS TO RESTAURANTS USING DIFFERENT ESTIMATORS.

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
IPW Estimator for Hypothetical Experiment Perspective:								
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
DML Estimator for Latent Experiment Perspective:								
IHS:	1.07	0.49	0.10	0.01	−0.02	−0.03	−0.17	−0.09
	(0.44)	(0.43)	(0.36)	(0.27)	(0.25)	(0.27)	(0.30)	(0.32)
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
	(7.40)	(7.35)	(5.49)	(5.41)	(3.54)	(4.75)	(5.27)	(5.43)
percent incr.:	171	15	−9	−3	−3	−9	−11	−21

Note: The first panel uses the IPW estimator, with inference valid for observational data under the hypothetical experiment perspective described in Section 4. The second panel uses the double machine learning (DML) estimator that may have favorable theoretical properties under the latent experiment perspective. 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.

when estimating an approximate percentage effect using inverse hyperbolic sine units[8] 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. 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. In a study of grocery store openings building on the framework of this paper, Chen, Qian, Zhang, and Zhang (2023) find a very similar pattern with estimated effects that are roughly half in magnitude of the estimates here. Larger effects in this study may be due to reduced baseline foot traffic during the COVID-19 pandemic, but overlapping confidence intervals suggest the difference may also be due to random chance.

Table I also displays estimates using the double machine learning estimator that may have favorable statistical properties under the "latent experiment perspective" of Section 4. I estimate the first step treatment probability and outcome models using cross-fitting in a post-LASSO procedure with data-driven penalty (Belloni, Chen, Chernozhukov, and Hansen, 2012). When estimating treatment probability or outcome mean for a grocery store or restaurant, I use only grocery stores or restaurants that are at least 1 mile away, satisfying design-based independence of the sample splits under the assumption that grocery stores have no effect on outcomes for restaurants at distances farther than 0.5 miles. For the treatment probability, I use the same logistic regression specification as before. For the outcome model, I use a linear model at the restaurant level with regressors for the number of grocery stores in eight distance bins of width 0.025 miles up to 0.2 miles. To obtain estimates of the conditional mean in the absence of a marginal grocery store, I use the outcome model prediction after removing a marginal grocery store from the regressors. The inferential results of Theorem 3 apply if the cross-fit estimates of treatment probabilities (including the CNN) and outcome predictions satisfy Assumption 6. The results are qualitatively and quantitatively similar to the IPW estimators discussed above.

I also implement a cross-sectional inner versus outer ring estimator. For given inner and outer ring distances, I use restaurant-grocery store pairs on both rings as observations in a regression of restaurant outcomes on an inner ring indicator and grocery store fixed effects. Note that, most commonly, the inner versus outer ring strategy is used in a difference-in-differences design. In this paper, however, the temporal differencing is not attractive for two reasons. First, because changes in foot traffic from pre-COVID-19 periods into COVID-19 periods are very large, a parallel trends assumption (in levels or percentage terms) may be questionable, and parallel trends before COVID-19 are likely uninformative about whether trends into the COVID-19 periods are parallel. Second, temporal differencing would at best yield estimates of the differential effect after versus before the COVID-19 pandemic of being near real versus counterfactual grocery store locations. In the absence of these two issues, one may wish to use *both* the estimators proposed in this paper and the inner versus outer ring estimator in a difference-in-differences design.

[8]Due to the extreme reduction in foot traffic during the COVID-19 pandemic, some businesses have zero visits recorded by SafeGraph in April 2020, such that a log transformation is not feasible. The inverse hyperbolic sine is similar to the (shifted) log at values other than zero, but the usual caveats regarding the percentage change interpretation apply (Bellemare and Wichman, 2020, Mullahy and Norton, 2023, Chen and Roth, 2024).

The inner versus outer ring design and the quasi-experimental design proposed in this paper are distinct both conceptually and in the implementation choices required. Conceptually, whether identification is most plausible based on functional form (such as the comparability of the inner and outer rings), on quasi-experimental variation in the treatment location (as in the design proposed in this paper), on both, or on neither, depends on the application. With panel data, the inner versus outer ring design requires parallel trends of outcomes on inner and outer rings, while the quasi-experimental design requires parallel trends of outcomes near realized and counterfactual treatment locations. In the inner versus outer ring design, researchers typically justify the functional form assumption for identification and describe a separate sampling scheme for inference. In the quasi-experimental design, researchers use the same design-based variation for identification and inference. The inner versus outer ring design requires the specification of an outer ring distance. If the outer ring distance is too large, the outcome individuals on the outer ring may be systematically different from individuals on the inner ring, especially if treatment locations were chosen strategically. If the outer ring distance is too small, the outcome individuals on the outer ring may be affected by the treatment such that one at best learns about the effect of being relatively closer to the treatment. Unfortunately, the data are generally uninformative about this trade-off. In contrast, the quasi-experimental design requires implementation choices for counterfactual locations, as discussed in Section 4.2. Below, I discuss an additional challenge in implementing the inner versus outer ring design when treatment locations are not far apart.

For inner versus outer ring comparisons using all restaurant-grocery store pairs, treated restaurants are not necessarily exposed to one more grocery store than control restaurants, on average. Figure 6 shows the analog to Figure 3 for the inner versus outer ring estimator. Each panel restricts the sample to restaurants that are either on the inner ring with a particular radius or on the outer ring with distance between 0.2 and 0.225 miles. Within each panel, the figure plots the coefficients of regressions of the number of grocery stores at a particular distance on an inner ring dummy and grocery store fixed effects, varying the distance along the horizontal axis. Because the outer ring restaurants, while farther away from the focal grocery store, tend to be closer to *other* grocery stores, the difference in exposure (plotted coefficient) is below 1 at the distance of interest. Hence, the inner versus outer ring estimator does not necessarily estimate the effect of one additional grocery store at the distance of interest.

To avoid exposure of outer ring restaurants to other grocery stores, one may focus on isolated grocery stores where no other grocery store is closer to the outer ring. Figure 7

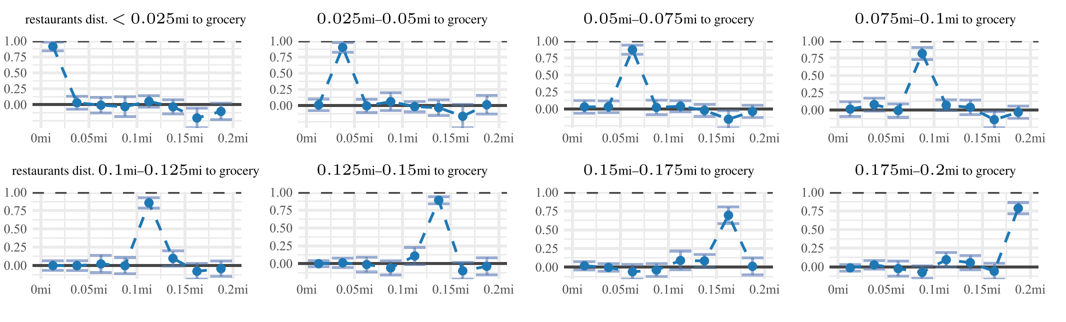


FIGURE 6.—Differential exposure of treated and control restaurants to grocery stores at different distances when using an outer ring of distance 0.2–0.225. Each panel holds fixed the inner (and outer) ring distance from a real grocery store and plots the coefficient of a regression of the number of grocery stores at a particular distance on an inner ring dummy with grocery store fixed effects.

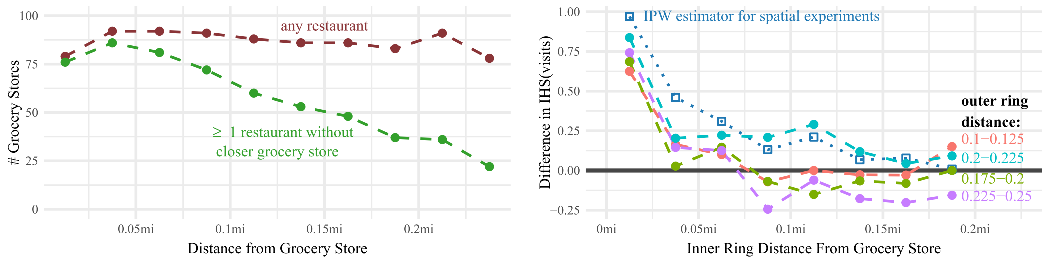


FIGURE 7.—Left: Number of grocery stores with either at least one restaurant in a given distance bin or at least one restaurant with no closer grocery store. There are few grocery stores with restaurants unexposed to other grocery stores when considering larger outer ring distances. Right: Estimated average effect of grocery stores on restaurants at different distances using control groups based on different outer ring distances as labeled on the figure, not restricting to those without closer grocery stores. The estimated effects vary noticeably depending on the choice of outer ring. When the inner and outer ring distance bins coincide, the estimates are mechanically zero. For comparison, the squares repeat the estimator proposed in this paper from Figure 5.

(left) shows the number of grocery stores that have at least one restaurant in a particular distance bin, as well as the number of grocery stores that have at least one restaurant without any closer grocery store in the distance bin. If one chooses an outer ring distance, drops all restaurants with any grocery store closer than that distance, and includes grocery store fixed effects, then the estimated treatment effect is based solely on the grocery stores constituting the lower of the two curves. Even within this subsample of grocery stores, the estimator is based on a possibly selected sample of restaurants geographically away from the more business-dense direction. Hence, when there are multiple realized treatment locations close to one another, proper differences in exposure to the treatment can be difficult to achieve with the inner versus outer ring strategy.

Furthermore, the neighborhoods of restaurants on inner rings and outer rings are not necessarily alike. Figure 4 above shows that the neighborhoods of the restaurants on an outer ring are likely noticeably different from the neighborhoods of restaurants on an inner ring. Many grocery stores are located in shopping or strip malls. The estimators proposed in this paper, intuitively, use counterfactual treatment locations that exploit differences in the number of grocery stores in such retail-heavy areas. In contrast, the inner versus outer ring strategy compares restaurants in commercial areas to restaurants in more residential or industrial areas.

Inner versus outer ring estimates are fairly sensitive to the choice of outer ring distance in this application. To estimate the effects for given inner and outer rings, restrict the sample to restaurants on these rings. Restaurants may appear multiple times if they are on the inner or outer ring around different grocery stores. Estimate the treatment effect by regressing the outcome (IHS of visits) on an inner ring dummy and grocery store fixed effects. Figure 7 (right) plots the coefficients on the inner ring dummy for regressions using different inner ring distances (horizontal axis) and a few choices for the outer ring distance (curves). Even for “adjacent” outer ring distances, for instance 0.2–0.225 miles and 0.225–0.25 miles, estimated effects can be noticeably different. Given the differences in exposure and neighborhood characteristic balance, one may wish to choose a different outer ring for each inner ring to obtain more plausible comparisons. Alternatively, functional form assumptions may allow for correcting imbalances in levels (or non-parallel trends if panel data and variation in the timing of the treatment were available). The approach proposed in this paper balances by design without manual adjustments to the control group to estimate effects at different distances. To interpret the inner versus outer ring estimates as causal effects of a marginal grocery store at a particular distance, one needs to assume that the outer ring is unaffected (as in Assumption 3), and the correct choice of this distance is required for the estimation of the effects themselves. The approach proposed in this paper uses Assumption 3 only for standard errors.

6. 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. This paper presents a design-based approach for causal identification, estimation, and inference in spatial settings. The approach differs from existing approaches in the literature to identification and estimation based on parallel trends and inference based on sampling. I argue that identification, estimation, and inference using design-based ideas are conceptually attractive, analytically tractable, and computationally feasible. The same design-based ideas do not validate the inner versus 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 similar to real treatment locations. These counterfactual locations allow the estimation of causal effects in observational data under an assumption that ensures treatment assignment is not based on potential outcomes other than through observable characteristics, similar to an unconfoundedness assumption in sampling-based analyses.

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. 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 large 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.

APPENDIX: PROOFS

ASSUMPTION 8: *For the sequence of finite populations indexed by k, there exist positive constants $c_1, \dots, c_5$ such that*

(a) *Bounded potential outcomes, weights, and concentration: For each $i \in \mathbb{I}_k$, $S \subseteq \mathbb{S}_k$, and $s \in \mathbb{S}_k$: $|Y_i(S)| < c_1$, $|w_i(s, d)| < c_2$, and $\sum_{i\in\mathbb{I}_k} w_i(s, d) \in [c_3, c_4]$.*

(b) *Minimum observation spacing: For any $i, i' \in \mathbb{I}_k$ and $s, s' \in \mathbb{S}_k$, $d(r_{k,i}, r_{k,i'}) > c_5$ if $i \neq i'$, and $d(s, s') > c_5$ if $s \neq s'$.*

(c) *Asymptotic negligibility*: $\liminf_{k\to\infty}|\mathbb{I}_k|\sigma_k^2 > 0$ *with σ_k defined in Theorem 1(iv).*

(d) *Sample splitting*: *For each fold* f, $\frac{1}{|\mathbb{S}|}\sum_{s\in\mathbb{S}_f}\sum_{i\in\mathbb{I}_s}(\tilde{\mu}_{s,i}-\mu_{s,i})^2 = O_p(|\mathbb{S}|^{-1/2})$.

Assumption 8(a) and (b) require that outcomes are bounded, and, as the population grows, the estimator places weight on a growing geographic area and hence a growing number of candidate treatment locations. Assumption 8(c) ensures that no single treatment location dominates the estimator. Assumption 8(d), needed for the double machine learning estimator, requires that the different folds each sufficiently resemble the full sample. For random sample splits, this bound is attained if the number of distinct values $Z_{s,i}$ takes on is $O(\sqrt{|\mathbb{S}|})$.

A.1. *Notation and Proof of Theorem 1*

Notation. Let the random variable $\mathcal{M}_i^m \equiv \mathbb{1}\{\mathcal{M}_i = m\}$ be the indicator for whether exposure m of individual i is realized. 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 $\pi_{i,s,i',s'}^{m,a,m',a'} \equiv \Pr(\mathcal{M}_i^m\mathcal{T}_s^a = 1 \text{ and } \mathcal{M}_{i'}^{m'}\mathcal{T}_{s'}^{a'} = 1)$, which are straightforward to compute under Assumptions 1 and 3.

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

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

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

$$\begin{aligned}\tilde{V}_{aa}(d) \equiv \frac{2}{|\mathbb{S}|}\sum_{s\in\mathbb{S}}\sum_{s'\in\mathbb{S}}\sum_{i\in\mathbb{I}}\sum_{m\in\mathbb{M}_i}\sum_{i'\in\mathbb{I}}\sum_{m'\in\mathbb{M}_i}&\mathbb{1}\{\pi_{i,i'}^{m,m'}=0\}\pi_{i,s}^{m,a}\pi_{i',s'}^{m',a}\\ &\cdot\left(\frac{\pi_s}{1-\pi_s}\frac{\pi_{s'}}{1-\pi_{s'}}\right)^{\mathbb{1}\{a=c\}}\frac{w_i(s,d)w_{i'}(s',d)}{\bar{n}(d)^2}\left(\frac{Y_i(m)+Y_{i'}(m')}{2}-\mu_a(d)\right)^2,\end{aligned}$$

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

where the fixed, computable, weights $v_{i,s}^{m,a}(d)$ are

$$\begin{aligned}v_{i,s}^{m,a}(d) \equiv \left(\frac{\pi_s}{1-\pi_s}\right)^{\mathbb{1}\{a=c\}}\Bigg(&\big(1-\pi_{i,s}^{m,a}\big)\left(\frac{\pi_s}{1-\pi_s}\right)^{\mathbb{1}\{a=c\}}\frac{w_i(s,d)}{\bar{n}(d)}\\ &+\sum_{i'\in\mathbb{I}}\sum_{m'\in\mathbb{M}_{i'}}\sum_{s'\in\mathbb{S}}\sum_{a'\in\{c,t\}}\mathbb{1}\{\pi_{i,i'}^{m,m'}=0\}\pi_{i',s'}^{m',a'}\left(\frac{\pi_{s'}}{1-\pi_{s'}}\right)^{\mathbb{1}\{a'=c\}}\frac{w_{i'}(s',d)}{\bar{n}(d)}\Bigg).\end{aligned}$$

PROOF: For part (i), 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 $\hat{\tau} - \tilde{\tau} = (\tilde{\mu}_t - \mu_t)(\frac{1}{\dot{p}_t/p} - 1) - (\tilde{\mu}_c - \mu_c)(\frac{1}{\dot{p}_c/p} - 1) + (\frac{\hat{p}_c}{p} - 1)(\frac{p^2}{\dot{p}_c^2}\dot{\mu}_c - \mu_c) - (\frac{\hat{p}_t}{p} - 1)(\frac{p^2}{\dot{p}_t^2}\dot{\mu}_t - \mu_t)$, where variables $\dot{a}$ lie between $\hat{a}$ and a for $a = \mu_t, \mu_c, p_t, p_c$. Then, by Theorem 1 of Li and Ding (2017) and the arguments given for Part (v) below, and using Slutsky's theorem and the Delta Method, each of the factors of the four products is $\sqrt{|\mathbb{S}_k|}$-asymptotically normal, implying part (i).

For parts (ii) and (iii), unbiasedness for $\tau_{\text{marginal}}(d)$ follows directly by taking expectations of the numerator of $\mathcal{D}$. Under Assumption 2, the expected value simplifies to $\tau(d)$ because $\tau_i(s|S) = \tau_i(s)$.

To characterize the variance in part (iv), note that, under Assumption 3(a), $\mathcal{Y}_i = \sum_{m\in\mathbb{M}_i} \mathcal{M}_i^m Y_i(m)$. One can rewrite $\mathcal{D}$ in terms of fixed potential outcomes by using the exposure mappings, specifically

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

with $\tilde{Y}_i^{s,a}(d,m) \equiv (-\frac{\pi_s}{1-\pi_s})^{\mathbb{1}\{a=c\}} w_i(s,d)(Y_i(m) - \mu_a(d))$. 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'. When $a \neq a'$, rewrite $(Y_i(m) - \mu_t(d))(Y_{i'}(m') - \mu_c(d)) = \frac{1}{2}((Y_i(m) - \mu_t(d))^2 + (Y_{i'}(m') - \mu_c(d))^2 - (Y_i(m) - Y_{i'}(m') - (\mu_t(d) - \mu_c(d)))^2)$ as a kind of variance of "treatment effects." When $a = a'$, these terms are multiplied by a factor of opposite sign. To obtain a formula suggesting a conservative estimator of the variance, when $a = a'$ instead rewrite

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

The remaining steps simplify the summations over such terms. I show step-by-step derivations in SA 5.

Part (v) claims that $\frac{\tilde{\tau}_k - \tau_{k,\text{marginal}}}{\sigma_k} \xrightarrow{d} \mathcal{N}(0,1)$ as $k \to \infty$. First, note that without Assumption 3(a), $\tilde{\tau}_k(d) - \tau_{k,\text{marginal}}(d) = \mathcal{D}_k = \sum_{i\in\mathbb{I}_k} \mathcal{Z}_{k,i} + \epsilon_k$, where

$$\mathcal{Z}_{k,i} \equiv \sum_{m\in\mathbb{M}_{k,i}} \sum_{s\in\mathbb{S}_k} \sum_{a\in\{c,t\}} \mathcal{M}_{k,i}^m \mathcal{T}_s^a \frac{\tilde{Y}_{k,i}^{s,a}(d,m)}{|\mathbb{S}_k|\bar{n}_k(d)}$$

and ϵ as in Assumption 3(b). Because $\epsilon_k/\sigma_k = o_p(1)$ by Assumption 3(b), it suffices to show $\sum_{i\in\mathbb{I}_k} \mathcal{Z}_{k,i}/\sigma_k \xrightarrow{d} \mathcal{N}(0,1)$. The result follows from Theorem 1 of Jenish and Prucha (2009), the conditions of which I verify below. Note that $\sum_{i\in\mathbb{I}_k} \mathcal{Z}_{k,i}/\sigma_k =$

$\sum_{i\in\mathbb{I}_k}\tilde{\mathcal{Z}}_{k,i}/\tilde{\sigma}_k$ with $\tilde{\mathcal{Z}}_{k,i}\equiv|\mathbb{I}_k|\mathcal{Z}_{k,i}$ and $\tilde{\sigma}_k\equiv|\mathbb{I}_k|\sigma_k$. By part (iv), $\text{var}(\sum_{i\in\mathbb{I}_k}\mathcal{Z}_{k,i})=\sigma_k^2$, hence $\text{var}(\sum_{i\in\mathbb{I}_k}\tilde{\mathcal{Z}}_{k,i})=|\mathbb{I}_k|^2\sigma_k^2=\tilde{\sigma}_k^2$. Assumption 1 of Jenish and Prucha (2009) is satisfied given the minimum distance between individuals. Their Assumption 2 is satisfied with their $c_{i,n}=1$ given bounded potential outcomes and weights, and the minimum distance between locations. Furthermore, given the definition of exposures depending only on locations up to a fixed distance, and 0 weight at sufficiently large distances, $\tilde{\mathcal{Z}}_{k,i}$ and $\tilde{\mathcal{Z}}_{k,j}$ are independent if $d(r_{k,i},r_{k,j})>D$ for some D, such that the random field is ϕ-mixing satisfying their Assumption 4. Finally, the asymptotic negligibility condition implies their Assumption 5. Hence, by Theorem 1 of Jenish and Prucha (2009), $\sum_{i\in\mathbb{I}_k}\mathcal{Z}_{k,i}/\sigma_k=\sum_{i\in\mathbb{I}_k}\tilde{\mathcal{Z}}_{k,i}/\tilde{\sigma}_k\xrightarrow{d}\mathcal{N}(0,1)$. *Q.E.D.*

A.2. *Proof of Theorem 2*

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.

A.3. *Proof of Theorem 3*

For ease of notation, I suppress dependence on the distance d throughout the proof.

The proof establishes that $(\hat{\tau}(\hat{\pi},\hat{\mu})-\hat{\tau}(\pi,\mu))/\sigma=o_p(1)$. The conclusion of the theorem follows because $(\hat{\tau}(\pi,\mu)-\tau_{\text{marginal}})/\sigma\xrightarrow{d}\mathcal{N}(0,1)$.

First, note that $(\frac{\hat{\pi}_s}{1-\hat{\pi}_s}-\frac{\pi_s}{1-\pi_s})^2=\frac{(\hat{\pi}_s-\pi_s)^2}{(1-\hat{\pi}_s)^2(1-\pi_s)^2}\le C_0(\hat{\pi}_s-\pi_s)^2$ for some fixed real number C_0 because $\hat{\pi}_s$ and π_s are bounded away from 1. Hence, $\sum_{s\in\mathbb{S}}(\frac{\hat{\pi}_s}{1-\hat{\pi}_s}-\frac{\pi_s}{1-\pi_s})^2\le C_0\sum_{s\in\mathbb{S}}(\hat{\pi}_s-\pi_s)^2=o_p(\sqrt{|\mathbb{S}|})$.

Note that $1/(\sum_{s\in\mathbb{S}}\sum_{i\in\mathbb{I}}w_i(s)\mathcal{T}_s^t)=O_p(|\mathbb{S}|^{-1})$ by the regularity conditions and $1/\sigma=O(\sqrt{|\mathbb{S}|})$ by assumption, so it suffices to show that the difference in numerators of $\hat{\tau}(\hat{\pi},\hat{\mu})$ and $\hat{\tau}(\pi,\mu)$ is $o_p(\sqrt{|\mathbb{S}|})$. One can write the difference in numerators as $\sum_{f=1}^F(\mathcal{R}_{1,f}+\mathcal{R}_{2,f}+\mathcal{R}_{3,f})$ where

$$\mathcal{R}_{1,f}=\sum_{s\in\mathbb{S}_f}\left(\frac{\mathcal{T}_s^t-\pi_s}{1-\pi_s}\right)\sum_{i\in\mathbb{I}}w_i(s)(\mu_{s,i}-\hat{\mu}_{s,i}),$$

$$\mathcal{R}_{2,f}=\sum_{s\in\mathbb{S}_f}\sum_{i\in\mathbb{I}}w_i(s)\mathcal{T}_s^c\left(\frac{\pi_s}{1-\pi_s}-\frac{\hat{\pi}_s}{1-\hat{\pi}_s}\right)(\mathcal{Y}_i-\mu_{s,i}),$$

$$\mathcal{R}_{3,f}=\sum_{s\in\mathbb{S}_f}\sum_{i\in\mathbb{I}}w_i(s)\mathcal{T}_s^c\left(\frac{\pi_s}{1-\pi_s}-\frac{\hat{\pi}_s}{1-\hat{\pi}_s}\right)(\mu_{s,i}-\hat{\mu}_{s,i}).$$

For $\mathcal{R}_{1,f}$, consider $E(\mathcal{R}_{1,f}^2|\mathbb{S}_f,(\hat{\mu}_s)_{s\in\mathbb{S}_f})$ which conditions on the identity of observations in fold f and the out-of-fold estimator $\hat{\mu}_s$ for all $s\in\mathbb{S}_f$, where $\hat{\mu}_s=(\hat{\mu}_{s,i})_{i\in\mathbb{I}:w_i(s)\neq 0}$. Because, by independent assignment and sample splitting, $E((\frac{\mathcal{T}_s^t-\pi_s}{1-\pi_s})(\frac{\mathcal{T}_{s'}^t-\pi_{s'}}{1-\pi_{s'}})|\mathbb{S}_f, (\hat{\mu}_s)_{s\in\mathbb{S}_f}) = 0$ for any $s \neq s'$ in fold f, and all $\hat{\mu}_{s,i}$ with $f(s) = f$ are fixed by conditioning, for fixed real numbers C_1 and C_2, $E(\mathcal{R}_{1,f}^2|\mathbb{S}_f, (\hat{\mu}_{s,i})_{s\in\mathbb{S}_f}) \leq C_1 \sum_{s\in\mathbb{S}_f}(\sum_{i\in\mathbb{I}} w_i(s)(\mu_{s,i} - \hat{\mu}_{s,i}))^2$ $\leq C_1C_2 \sum_{s\in\mathbb{S}_f}\sum_{i\in\mathbb{I}_s}(\mu_{s,i}-\hat{\mu}_{s,i})^2$, where the first inequality bounds $E((\frac{\mathcal{T}_s^t-\pi_s}{1-\pi_s})^2)$, and the second inequality uses the Cauchy–Schwarz inequality and the bound on $|w_i(s, d)|$. Hence, $\sum_{s\in\mathbb{S}_f}\sum_{i\in\mathbb{I}_s}(\mu_{s,i} - \hat{\mu}_{s,i})^2 = o_p(|\mathbb{S}|)$ implies $\mathcal{R}_{1,f} = o_p(\sqrt{|\mathbb{S}|})$ by Markov's inequality as required.

For $\mathcal{R}_{2,f}$, define $\delta_{s,i} = w_i(s)\mathcal{T}_s^c(\frac{\pi_s}{1-\pi_s} - \frac{\hat{\pi}_s}{1-\hat{\pi}_s})(\mathcal{Y}_i - \mu_{s,i})$ and $\bar{\delta}_{s,i} = E(\delta_{s,i}|\mathbb{S}_f, (\hat{p}_s)_{s\in\mathbb{S}_f})$ and $\delta_s = \sum_{i\in\mathbb{I}}\delta_{s,i}$, $\bar{\delta}_s = \sum_{i\in\mathbb{I}}\bar{\delta}_{s,i}$. Then $\mathcal{R}_{2,f} = \sum_{s\in\mathbb{S}_f}(\delta_s - \bar{\delta}_s) + \sum_{s\in\mathbb{S}_f}\bar{\delta}_s$.

For the first term, similar to $\mathcal{R}_{1,f}$, consider $E((\sum_{s\in\mathbb{S}_f}\delta_s - \bar{\delta}_s)^2|\mathbb{S}_f, (\hat{\pi}_s)_{s\in\mathbb{S}_f})$. By definition, $E(\delta_s - \bar{\delta}_s|\mathbb{S}_f, (\hat{\pi}_s)_{s\in\mathbb{S}_f}) = 0$, and $\delta_s - \bar{\delta}_s$ and $\delta_{s'} - \bar{\delta}_{s'}$ are (conditionally) independent unless there exists either an individual i' with $w_{i'}(s') \neq 0$ such that treatment at location s can affect the outcome of i, or individuals i with $w_i(s) \neq 0$ and i' with $w_{i'}(s') \neq 0$ with a common treatment location s'' that can affect the outcome of both i and i'. Under the assumptions of a minimum distance between candidate treatment locations, a maximum distance after which weights are zero, and a maximum distance after which the treatment has no effect, for any s', the number of s for which the above occurs is bounded. For these s, s', bound $(\delta_s - \bar{\delta}_s)(\delta_{s'} - \bar{\delta}_{s'}) \leq (\delta_s - \bar{\delta}_s)^2 + (\delta_{s'} - \bar{\delta}_{s'})^2$. Steps analogous to $\mathcal{R}_{1,f}$ then yield, for a fixed real number C_3, $E((\sum_{s\in\mathbb{S}_f}\delta_s - \bar{\delta}_s)^2|\mathbb{S}_f, (\hat{p}_s)_{s\in\mathbb{S}_f}) \leq C_3\sum_{s\in\mathbb{S}_f}(\frac{\pi_s}{1-\pi_s} - \frac{\hat{\pi}_s}{1-\hat{\pi}_s})^2$. So $\sum_{s\in\mathbb{S}_f}(\frac{\pi_s}{1-\pi_s} - \frac{\hat{\pi}_s}{1-\hat{\pi}_s})^2 = o_p(|\mathbb{S}|)$ implies that the first term is $o_p(\sqrt{|\mathbb{S}|})$.

For the second term, $\sum_{s\in\mathbb{S}_f}\bar{\delta}_s$, use that $\pi_s = p(Z_s)$ and $\hat{\pi}_s = \hat{\pi}_{s'}$ if $Z_s = Z_{s'}$ and $f(s) = f(s')$, to write $\sum_{s\in\mathbb{S}_f}\bar{\delta}_s = \sum_{s\in\mathbb{S}_f}\sum_{i\in\mathbb{I}} w_i(s)(1 - \pi_s)(\frac{\pi_s}{1-\pi_s} - \frac{\hat{p}_s}{1-\hat{p}_s})(\tilde{\mu}_{s,i} - \mu_{s,i})$. Then, using the Cauchy–Schwarz inequality and the bound on $|w_i(s)|$, $(\sum_{s\in\mathbb{S}_f}\bar{\delta}_s)^2 \leq C_4\sum_{s\in\mathbb{S}_f}(\frac{\pi_s}{1-\pi_s} - \frac{\hat{\pi}_s}{1-\hat{\pi}_s})^2\sum_{s\in\mathbb{S}_f}\sum_{i\in\mathbb{I}_s}(\tilde{\mu}_{s,i} - \mu_{s,i})^2$. So, $\sum_{s\in\mathbb{S}_f}(\frac{\pi_s}{1-\pi_s} - \frac{\hat{\pi}_s}{1-\hat{\pi}_s})^2 = o_p(\sqrt{|\mathbb{S}|})$ and $\sum_{s\in\mathbb{S}_f}\sum_{i\in\mathbb{I}_s}(\tilde{\mu}_{s,i} - \mu_{s,i})^2 = O_p(\sqrt{|\mathbb{S}|})$ ensure $\sum_{s\in\mathbb{S}_f}\bar{\delta}_s = o_p(\sqrt{|\mathbb{S}|})$. Hence, overall $\mathcal{R}_{2,f} = o_p(\sqrt{|\mathbb{S}|})$ by Markov's inequality.

For $\mathcal{R}_{3,f}$, applying the Cauchy–Schwarz inequality and bounding $|w_i(s)|$ yields $E(\mathcal{R}_{3,f}^2|\mathbb{S}_f, (\hat{\pi}_s, \hat{\mu}_{s,i})_{s\in\mathbb{S}_f}) \leq C_5\sum_{s\in\mathbb{S}_f}(\frac{\pi_s}{1-\pi_s} - \frac{\hat{\pi}_s}{1-\hat{\pi}_s})^2\sum_{s\in\mathbb{S}_f}\sum_{i\in\mathbb{I}_s}(\mu_{s,i} - \hat{\mu}_{s,i})^2$. Hence, $\sum_{s\in\mathbb{S}_f}(\frac{\pi_s}{1-\pi_s} - \frac{\hat{\pi}_s}{1-\hat{\pi}_s})^2 = o_p(\sqrt{|\mathbb{S}|})$ and $\sum_{s\in\mathbb{S}_f}\sum_{i\in\mathbb{I}_s}(\mu_{s,i} - \hat{\mu}_{s,i})^2 = o_p(\sqrt{|\mathbb{S}|})$ jointly imply that $\mathcal{R}_{3,f} = o_p(\sqrt{|\mathbb{S}|})$ by Markov's inequality.

A.4. *Proof of Theorem 4*

For $s \in \tilde{\mathbb{S}} \setminus \mathbb{S}$, $\pi_s = 0$ by the definition of $\mathbb{S}$, and also $\hat{\pi}_s = 0$ with probability 1 by Assumption 5. Hence, it suffices to focus on $s \in \mathbb{S}$. Under Assumption 5, one can equivalently define $n_z = \sum_{s\in\mathbb{S}}\mathbb{1}\{Z_s = z\}$ for $z \in \mathbb{Z}$ because for any $s \in \tilde{\mathbb{S}}$, if $Z_s \in \mathbb{Z}$, then $s \in \mathbb{S}$ as $p(Z_s) > 0$ by the definition of $\mathbb{Z}$. Note that $\sum_{s\in\mathbb{S}}(\hat{\pi}_s - \pi_s)^2 = \sum_{z\in\mathbb{Z}} n_z(\hat{p}(z) - p(z))^2$ where $\hat{p}(z) = \sum_{s\in\mathbb{S}}\mathbb{1}\{Z_s = z\}\mathcal{T}_s^t/n_z$. Let $\mathcal{E}_z \equiv (\hat{p}(z) - p(z))^2 - \frac{p(z)(1-p(z))}{n_z}$. Then, for $\epsilon > 0$, $\Pr(\sum_{s\in\mathbb{S}}(\hat{\pi}_s - \pi_s)^2/a > \epsilon) \leq \Pr(\sum_{z\in\mathbb{Z}} n_z\mathcal{E}_z > a\epsilon - |\mathbb{Z}|/4)$ using $\sum_{z\in\mathbb{Z}} p(z)(1 - p(z)) \leq |\mathbb{Z}|/4$.

Next, by Bernstein's inequality for sums of independent mean zero sub-Exponential random variables, $\Pr(\sum_{z\in\mathbb{Z}} n_z\mathcal{E}_z > u) \leq 2\exp(-c\min(\frac{u^2}{|\mathbb{Z}|C^2}, \frac{u}{C}))$ for universal constants c

and C: Because $\hat{p}(z)$ is the average of n_z independent bounded (binary) random variables with mean $p(z)$, there exists a fixed C such that $\hat{p}(z) - p(z)$ is sub-Gaussian with norm at most $C/\sqrt{n_z}$, and $\mathcal{E}_z$ is mean zero and sub-Exponential with norm at most C/n_z (Vershynin, 2018) for all $z \in \mathcal{Z}$. So $n_z\mathcal{E}_z$ is sub-Exponential with norm at most C. By Assumption 1, $\mathcal{E}_z$ are independent across z. Hence, Bernstein's inequality applies as claimed.

Substitute $u = \sqrt{|\mathbb{S}|}\epsilon - |\mathbb{Z}|/4$. Then, $|\mathbb{Z}| = o(\sqrt{|\mathbb{S}|})$ implies that $\min(\frac{u^2}{|\mathbb{Z}|C^2}, \frac{u}{C}) \to \infty$ as $|\mathbb{S}| \to \infty$. Hence, $\lim_{|\mathbb{S}|\to\infty} \Pr(\sum_{s\in\mathbb{S}}(\hat{\pi}_s - \pi_s)^2/\sqrt{|\mathbb{S}|} > \epsilon) = 0$ such that $\sqrt{|\mathbb{S}|^{-1}\sum_{s\in\mathbb{S}}(\hat{\pi}_s - \pi_s)^2} = o_p(|\mathbb{S}|^{-1/4})$ as stated in the theorem.

A.5. *Proof of Theorem 5*

Following Wainwright (2019, Theorem 7.13), if $\lambda \geq 2\max_l\{|\mathcal{Z}'_{\cdot,l}(\mathcal{Y} - \mathcal{Z}\beta)|\}/\dot{n}$ and the restricted eigenvalue condition holds, then $\sqrt{\sum_l(\hat{\beta}_l - \beta_l)^2} \leq \frac{3}{\kappa}\sqrt{\|\beta\|_0}\lambda$. The restricted eigenvalue condition is assumed to hold with probability approaching 1. The proof establishes that the first condition holds with high probability for the choice of λ given in the theorem, and establishes the resulting rate.

To show that the choice of λ specified in the theorem satisfies the condition above with probability approaching 1, use the approximate clustering notation and the triangle inequality to get $\max_l\{|\mathcal{Z}'_{\cdot,l}(\mathcal{Y} - \mathcal{Z}\beta)|\} \leq \max_l\{|\mathcal{Z}'_{\cdot,l}(\mathcal{Y} - \boldsymbol{Y}^{\mathcal{M}})|\} + \max_l\{|\mathcal{Z}'_{\cdot,l}(\boldsymbol{Y}^{\mathcal{M}} - \mathcal{Z}\beta)|\}$. By the approximate clustering assumption, $\max_l\{|\mathcal{Z}'_{\cdot,l}(\mathcal{Y} - \boldsymbol{Y}^{\mathcal{M}})|\}/\dot{n} = o_p(\sqrt{\ln(L)/|\mathbb{C}|})$. Below, I show that also $\max_l\{|\mathcal{Z}'_{\cdot,l}(\boldsymbol{Y}^{\mathcal{M}} - \mathcal{Z}\beta)|\}/\dot{n} \leq \sqrt{\frac{C}{|\mathbb{C}|}\ln(L)}$ with probability approaching 1 for some fixed C. Hence, there exists a fixed $\tilde{C}$ such that $\lambda = \tilde{C}\sqrt{\ln(L)/|\mathbb{C}|} \geq 2\max_l\{|\mathcal{Z}'_{\cdot,l}(\mathcal{Y} - \mathcal{Z}\beta)|\}/\dot{n}$ with probability approaching 1. Because $|\mathbb{S}|/|\mathbb{C}|$ is bounded by Assumption 7(i), one can replace $|\mathbb{C}|$ by $|\mathbb{S}|$ in the definition of λ with an updated choice of $\tilde{C}$ without affecting the validity of the claims. Substituting the choice $\lambda = \tilde{C}\sqrt{\ln(L)/|\mathbb{S}|}$ into $\sqrt{\sum_l(\hat{\beta}_l - \beta_l)^2} \leq \frac{3}{\kappa}\sqrt{\|\beta\|_0}\lambda$ yields the claim of the theorem.

To bound $|\mathcal{Z}'_{\cdot,l}(\boldsymbol{Y}^{\mathcal{M}} - \mathcal{Z}\beta)|/\dot{n}$ for fixed l, apply Hoeffding's inequality. Then, apply the union bound to bound the maximum over l. Define $\rho_{c,l} \equiv \frac{1}{\dot{n}/|\mathbb{C}|}\sum_{s\in\mathbb{S}}\sum_{i\in\mathbb{I}}\mathbb{1}\{c(s,i) = c\}w_i(s,d)\mathbb{1}\{\mathcal{S} \not\ni s\}Z_{(s,i),l}(Y_i(\mathcal{M}_{s,i}) - Z_{s,i}\beta)$ such that $\mathcal{Z}'_{\cdot,l}(\boldsymbol{Y}^{\mathcal{M}} - \mathcal{Z}\beta)/\dot{n} = \frac{1}{|\mathbb{C}|}\sum_{c\in\mathbb{C}}(\rho_{c,l} - E(\rho_{c,l}))$ because $\sum_{c\in\mathbb{C}}E(\rho_{c,l}) = 0$ by linearity. Note that the $\rho_{c,l}$ terms are independent across c because $Y_i(\mathcal{M}_{s,i})$ varies only with assignments within the cluster of (s,i) by the definition of the approximate exposures, and $\mathbb{1}\{\mathcal{S} \not\ni s\}$ is independent across clusters by Assumptions 1 and 7(ii). Furthermore, because all factors inside the summation constituting $\rho_{c,l}$ are bounded, as is the total number of observations in the cluster, $\rho_{c,l}$ is bounded. Hence, $\mathcal{Z}'_{\cdot,l}(\boldsymbol{Y}^{\mathcal{M}} - \mathcal{Z}\beta)/\dot{n}$ is the average of $|\mathbb{C}|$ independent bounded mean zero random variables. Then, by Hoeffding's inequality, for all $t > 0$, $\Pr(|\frac{1}{|\mathbb{C}|}\sum_{c\in\mathbb{C}}(\rho_{c,l} - E(\rho_{c,l}))| \geq u) \leq 2\exp(-\frac{2u^2|\mathbb{C}|}{C})$ for some $C < \infty$. Taking the union bound, $\Pr(\max_l|\frac{1}{|\mathbb{C}|}\sum_{c\in\mathbb{C}}(\rho_{c,l} - E(\rho_{c,l}))| \geq u) \leq 2\exp(-\frac{2u^2|\mathbb{C}|}{C} + \ln(L))$. So, $\Pr(\max_l\{|\mathcal{Z}'_{\cdot,l}(\boldsymbol{Y}^{\mathcal{M}} - \mathcal{Z}\beta)|\}/\dot{n} \leq \sqrt{\frac{C}{|\mathbb{C}|}\ln(L)}) \geq 1 - 2\exp(-\ln(L)) \to 1$ as $L \to \infty$.

REFERENCES

Abadie, Alberto, Susan Athey, Guido W. Imbens, and Jeffrey M. Wooldridge (2020), "Sampling-Based vs. Design-Based Uncertainty in Regression Analysis." *Econometrica*, 88 (1), 265–296. [1533]

Abadie, Alberto, Susan Athey, Guido W. Imbens, and Jeffrey M. Wooldridge (2023), "When Should You Adjust Standard Errors for Clustering?" *The Quarterly Journal of Economics*, 138 (1), 1–35. [1530]
Bellemare, Marc F. and Casey J. Wichman (2020), "Elasticities and the Inverse Hyperbolic Sine Transformation." *Oxford Bulletin of Economics and Statistics*, 82 (1), 50–61. [1551]
Belloni, Alexandre, Daniel Chen, Victor Chernozhukov, and Christian Hansen (2012), "Sparse Models and Methods for Optimal Instruments With an Application to Eminent Domain." *Econometrica*, 80 (6), 2369–2429. [1551]
Borusyak, Kirill and Peter Hull (2023), "Nonrandom Exposure to Exogenous Shocks." *Econometrica*, 91 (6), 2155–2185. [1531]
Chen, August, Franklin Qian, Qianyang Zhang, and Xiang Zhang (2023), "Identifying Agglomeration Spillovers: Evidence From Grocery Store Openings." https://papers.ssrn.com/sol3/papers.cfm?abstract_id=6609758. [1551]
Chen, Jiafeng and Jonathan Roth (2024), "Logs With Zeros? Some Problems and Solutions." *The Quarterly Journal of Economics*, 139 (2), 891–936. [1551]
[CCDD+] Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins (2018), "Double/Debiased Machine Learning for Treatment and Structural Parameters." *The Econometrics Journal*, 21 (1), C1–C68. [1531,1539]
Conley, Timothy G. (1999), "GMM Estimation With Cross Sectional Dependence." *Journal of Econometrics*, 92 (1), 1–45. [1530,1531]
Duflo, Esther (2001), "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence From an Unusual Policy Experiment." *American Economic Review*, 91 (4), 795–813. [1537]
Goodfellow, Ian (2016), "NIPS 2016 Tutorial: Generative Adversarial Networks." arXiv preprint arXiv:1701.00160. [1545]
[GPMX+] Goodfellow, Ian, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio (2014), "Generative Adversarial Nets." In *Advances in Neural Information Processing Systems*, Vol. 27, 2672–2680. [1545]
Gupta, Arpit, Stijn Van Nieuwerburgh, and Constantine E. Kontokosta (2022), "Take the Q Train: Value Capture of Public Infrastructure Projects." *Journal of Urban Economics*, 129, 103422. [1538]
Hájek, Jaroslav (1960), "Limiting Distributions in Simple Random Sampling From a Finite Population." *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, 5, 361–374. [1544]
Imbens, Guido W. and Donald B. Rubin (2015), *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press, New York, NY. [1543]
Jenish, Nazgul and Ingmar R. Prucha (2009), "Central Limit Theorems and Uniform Laws of Large Numbers for Arrays of Random Fields." *Journal of Econometrics*, 150 (1), 86–98. [1556,1557]
Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton (2012), "ImageNet Classification With Deep Convolutional Neural Networks." *in Advances in Neural Information Processing Systems*, 25, 1097–1105. [1545]
Li, Xinran and Peng 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. [1556]
Linden, Leigh and Jonah E. Rockoff (2008), "Estimates of the Impact of Crime Risk on Property Values From Megan's Laws." *American Economic Review*, 98 (3), 1103–1127. [1530]
Lotter, William, Gabriel Kreiman, and David Cox (2016), "Unsupervised Learning of Visual Structure Using Predictive Generative Networks." arXiv preprint arXiv:1511.06380. [1545]
Mullahy, John and Edward C. Norton (2023), "Why Transform Y? The Pitfalls of Transformed Regressions With a Mass at Zero." *Oxford Bulletin of Economics and Statistics*. [1551]
Neyman, Jerzey (1923), "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9." *Roczniki Nauk Rolniczych Tom X*, 1–51. [In Polish]. [1530]
Pinkse, Joris, Lihong Shen, and Margaret Slade (2007), "A Central Limit Theorem for Endogenous Locations and Complex Spatial Interactions." *Journal of Econometrics*, 140 (1), 215–225. [1530]
Pollmann, Michael (2026), "Supplement to 'Causal Inference for Spatial Treatments'." *Econometrica Supplemental Material*, 94, https://doi.org/10.3982/ECTA21571SUPP. [1532]
Rosenbaum, Paul R. (1984), "Conditional Permutation Tests and the Propensity Score in Observational Studies." *Journal of the American Statistical Association*, 79 (387), 565–574. [1544]
SafeGraph (2021), "Weekly Patterns and POI, July 2021 Release." Accessed 2021-07-22. [1529]
Sävje, Fredrik (2023), "Causal Inference With Misspecified Exposure Mappings: Separating Definitions and Assumptions." *Biometrika*. asad019. [1534]
Sävje, Fredrik, Peter M. Aronow, and Michael G. Hudgens (2021), "Average Treatment Effects in the Presence of Unknown Interference." *The Annals of Statistics*, 49 (2), 673–701. [1533]
Simard, Patrice Y., Dave Steinkraus, and John C. Platt (2003), "Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis." In *Seventh International Conference on Document Analysis and Recognition*, Vol. 3, 958–963, IEEE. [1544]

Vershynin, Roman (2018), *High-Dimensional Probability: An Introduction With Applications in Data Science*. Cambridge Series in Statistical and Probabilistic Mathematics, Vol. 47. Cambridge University Press. [1559]

Wainwright, Martin J. (2019), *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*, Vol. 48. Cambridge University Press. [1559]

Wang, Ye, Cyrus Samii, Haoge Chang, and Peter M. Aronow (2025), "Design-Based Inference for Spatial Experiments Under Unknown Interference." *The Annals of Applied Statistics*, 19 (1), 744–768. [1531,1534, 1536]

Yang, Yang, Hongbo Liu, and Xiang 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. [1546]

Zigler, Corwin M. and Georgia Papadogeorgou (2021), "Bipartite Causal Inference With Interference." *Statistical Science*, 36 (1), 109–123. [1531,1534]

Guest Editor Ulrich K. Müller handled this manuscript.

Manuscript received 30 January, 2023; final version accepted 17 April, 2026; available online 27 April, 2026.

The replication package for this paper is available at https://doi.org/10.5281/zenodo.19568729. The authors were granted an exemption to publish parts of their data because either access to these data is restricted or the authors do not have the right to republish them. However, the authors included in the package, on top of the codes and the parts of the data that are not subject to the exemption, a simulated or synthetic dataset that allows running the codes. The Journal checked the data and the codes for their ability to generate all tables and figures in the paper and approved online appendices. Whenever the available data allowed, the Journal also checked for their ability to reproduce the results. However, the synthetic/simulated data are not designed to produce the same results. Given the highly demanding nature of the algorithms, the reproducibility checks were run on a simplified version of the code, which is also available in the replication package.