Causal Inference for Spatial Treatments

Michael Pollmann*

July 7, 2022

Abstract

Many events and policies (treatments), such as opening of businesses, building of hospitals, and sources of pollution, occur at specific spatial locations, with researchers interested in their effects on nearby individuals or businesses (outcome units). However, the existing treatment effects literature primarily considers treatments that could experimentally be assigned directly at the level of the outcome units, potentially with spillover effects. I approach the *spatial treatment* setting from a similar experimental perspective: What ideal experiment would we design to estimate the causal effects of spatial treatments? This perspective motivates a comparison between individuals near realized treatment locations and individuals near counterfactual (unrealized) candidate locations, which is distinct from current empirical practice. I derive standard errors based on this design-based perspective that are straightforward to compute irrespective of spatial correlations in outcomes. Furthermore, I propose machine learning methods to find counterfactual candidate locations and show how to apply the proposed methods on observational data. I study the causal effects of grocery stores on foot traffic to nearby businesses during COVID-19 shelter-in-place policies. I find a substantial positive effect at a very short distance. Correctly accounting for possible effect "interference" between grocery stores located close to one another is of first order importance when calculating standard errors in this application.

*Department of Economics, Stanford University, 579 Jane Stanford Way, Stanford, CA, 94305. E-mail: pollmann@stanford.edu.
An earlier version of this paper can be found on arXiv: https://arxiv.org/abs/2011.00373. I am grateful to my advisor, Guido Imbens, for invaluable encouragement and guidance. I am thankful to Luis Armona, Tim Armstrong, Eric Auerbach, Ivan Canay, Caroline Hoxby, Joshua Kim, Elena Manresa, Konrad Menzel, Áureo de Paula, Daniel Pollmann, Jann Spiess, Stefan Wager, Melanie Wallskog, and Frank Wolak, as well as seminar participants at Stanford, Berkeley, and the North American and European Summer Meetings of the Econometric Society 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.

1 Introduction

How can we estimate the causal effects of spatial treatments? In the setting of this paper, a spatial treatment, such as the opening of a "million dollar plant" (Greenstone and Moretti, 2003; Greenstone et al., 2010) occurs at a spatial location, and the outcome of interest, such as earnings of workers, is measured for separate individuals who are located nearby.[1] I study an "ideal experiment" where causal effects are identified by (quasi-) random variation in the location of the spatial treatment, formalizing informal discussions about identifying variation in empirical work studying spatial treatments (for instance Linden and Rockoff, 2008). This motivates a perspective on the potential outcomes framework where the roles played by outcome and treatment units are more clearly distinct than in most existing theoretical work in causal inference. In the absence of guidance from theoretical work, most recent empirical studies using highly-detailed location data compare "treated" individuals at a given distance from treatment (on an "inner ring") to "control" individuals that are farther away but centered around the same treatment location (on an "outer ring"). However, I show that the ideal experiment does not generally ensure the validity of this control group; identification instead fundamentally relies on functional form assumptions.

In this paper, I propose (quasi-) experimental methods for spatial treatments that are motivated by an ideal experiment where the spatial locations of treatments are random. These methods are based on a simple insight: Suppose the ideal experiment randomly chooses some locations for treatment from a larger set of candidate locations. Then quasi-experimental methods should compare individuals near locations that are chosen for treatment to individuals near locations that could have been chosen but were not. For a formal characterization of estimands and estimators, I extend the potential outcomes framework for individual-level treatments to allow treatments to be randomized across space and to directly affect nearby individuals. Within this framework, I derive finite sample design-based standard errors similar to those of Neyman (1923, 1990) for randomized experiments with individual-level treatment assignment for a fixed population.

The estimators I propose differ from the approach of most recent empirical work in the choice of the control group. Suppose we want to estimate the average effect of a million dollar plant on individuals who are, say, one mile away. The method employed by most recent empirical work compares individuals on an inner ring around the million dollar plant with radius one mile to individuals on an outer ring, who are, say, five miles away from the same million dollar plant. Since many observable and unobservable characteristics correlate with distance from any one point in space (cf. Lee and Ogburn, 2021; Kelly, 2019), this comparison of inner and outer ring can be controversial: If treatment always occurs in the city center, the inner ring, or treated, individuals are urban individuals, while the outer ring, or control, individuals are suburban and rural individuals, and thus potentially fundamentally different.[2] In contrast, the estimators I propose use a control

[1] Spatial treatments are relevant for a diverse range of questions from many applied fields in economics and other social sciences. Recent studies estimating the effects of spatial treatments using individual-level outcome and location data include Stock (1989, 1991); Linden and Rockoff (2008); Currie et al. (2015); Aliprantis and Hartley (2015); Sandler (2017); Diamond and McQuade (2019); Chalfin et al. (2022); Rossin-Slater et al. (2020). Notably, Dell and Olken (2020) explicitly consider counterfactual treatment locations in a quasi-experimental setting, in an approach that is conceptually similar to the one taken in this paper. Much more existing empirical work studying spatial treatments is limited to aggregated outcome data, for instance Duflo (2001); Miguel and Kremer (2004); Cohen and Dupas (2010); Greenstone et al. (2010); Feyrer et al. (2017); Jia (2008); Keiser and Shapiro (2019). Furthermore, a recent literature has documented large geographic variation in a diverse range of outcomes (for instance Chetty et al., 2014; Chetty and Hendren, 2018; Finkelstein et al., 2016, 2021; Athey et al., 2019; Bilal, 2019), the causes and potential remedies of which, such as place-based polices, may involve spatial treatments.

[2] Researchers typically attempt to address this issue by adding a pre vs. post comparison in a difference-in-differences approach, where outcomes for urban and suburban individuals are allowed to be on different levels, but must evolve along parallel trends in the absence of treatment. Conceptually, the same issues remain. Parallel trends here is a functional form assumption that is not guaranteed to hold by experimental design, and may be more difficult to make plausible visually in the spatial treatment setting than in other settings, as discussed in Section 2.

group of individuals who are one mile away from *counterfactual candidate locations* where treatment could have occurred under the ideal experiment but did not, ensuring validity by experimental design. Greenstone and Moretti (2003), using aggregate data, choose a similar control group for counties with million dollar plants: They compare counties that "won" the bidding war for a million dollar plant to "runners-up" counties that were also very seriously considered as locations for million dollar plants, but ultimately "lost."

The "inner ring vs. outer ring" empirical strategy commonly used in current empirical research relies on two partly conflicting assumptions: First, the treatment must not affect individuals on the outer ring directly. This is most easily achieved by choosing an outer ring with large radius, such that these "control" individuals are far away. Second, the individuals on inner and outer rings must be comparable. This is most easily achieved by choosing an outer ring close to the inner ring, in conflict with the first assumption. Even when the differences in levels between inner and outer ring are differenced out with individual fixed effects in panel data, the parallel trends assumption can be particularly problematic in spatial treatment settings. Researchers oftentimes estimate the effect not just at one distance but at multiple distances, typically using the same outer ring control group. This effectively requires that individuals at all distances up to the outer ring are on the same parallel trend, with additively separable time fixed effects.

A simple example illustrates that these two assumptions are fundamental to the inner ring vs. outer ring empirical strategy, are not guaranteed to be satisfied even in a true randomized experiment, and cannot be relaxed through more flexible functional form in larger samples. Suppose that treatment only occurs in city centers. Then the assumptions may require individuals living in downtown areas to be on parallel trends (in the absence of treatment) to those on the outskirts of the city. Clearly, randomization of which cities the treatment occurs in cannot guarantee such parallel trends. Furthermore, data from additional cities does not make trends of inner cities and outskirts more likely to be parallel. Even in larger samples, one cannot choose an outer ring closer to the inner ring because such outer ring individuals continue to (potentially) be affected by the treatment themselves. These assumptions are therefore not just approximations to make finite sample analysis feasible where asymptotically an analogous nonparametric specification identifies treatment effects: Identification in this approach fundamentally rests upon the functional form assumptions even asymptotically, and even with experimental data.

Instead, I recommend estimators that are formally valid under the quasi-experimental variation in treatment location sometimes used to informally justify the assumptions of the inner ring vs. outer ring empirical strategy. The inner ring vs. outer ring empirical strategy generally yields the most credible estimates if the treatment is known not to have an effect past a short, known, distance. Then individuals on an outer ring with small radius are likely to be comparable. Sometimes, these comparisons are then argued to be justified by the fact that the exact location of the treatment was as good as random. For instance, Linden and Rockoff (2008, p. 1110), referencing Bayer et al. (2008), argue that for their treatment, sex offenders moving into neighborhoods, "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." However, as I demonstrate in Section 2, even at short distances this quasi-randomization is not sufficient for inner and outer ring comparisons: In addition to randomization, the potential outcome surface needs to be flat. The estimators proposed in this paper instead allow researchers to directly make use of such credible identifying variation in the spatial locations of treatments.

Furthermore, using this design-based perspective I derive standard errors that are both conceptually attractive and straightforward to compute irrespective of the spatial correlations in outcomes, simplifying

inference in spatial treatment settings. The interpretation of the design-based standard errors I propose is simple: If treatment had been realized in some of the counterfactual locations instead of some of the realized locations, the point estimates would be different. That is, design-based standard errors reflect the variation in estimates that arises due to the *identifying* variation in treatment locations, holding the individuals in the sample and their potential outcomes fixed. This is the same variation that is needed for internal validity of the causal effect estimates (Abadie et al., 2020). The variance formulas I derive are weighted averages of squared (aggregated) potential outcomes and are straightforward to compute with sample analogs using observed outcomes. I see the simplicity of computation as a key advantage over alternative standard errors that require correct specification and estimation of, for instance, spatial autoregressive models that are often unrelated to the research question and expertise of the researcher. In a baseline setting, the variance estimators I derive are similar to *clustering at the level of treatment assignment* (Abadie et al., 2017) or, after aggregating data, to Neyman (1923, 1990) standard errors for randomized experiments with individual-level treatment assignment for a fixed population. I also extend my results to settings where realized treatment locations close to one another cause interference. I show how to define exposure mappings (Aronow and Samii, 2017) based on restrictions on interference that are often plausible in spatial settings, such as a maximum distance past which the treatment is assumed not to affect individuals. Here, the design-based perspective allows easy computation of standard errors even when clustering based on non-overlapping, discrete, independent groups is not feasible.

I demonstrate the quasi-experimental methods I propose in an application studying the causal effects of grocery stores on foot-traffic to nearby restaurants during COVID-19 shelter-in-place policies in April of 2020. During shelter-in-places policies, consumers made few trips outside the home. Grocery stores may bring consumers to particular neighborhoods, benefiting nearby restaurants, akin to anchor stores in shopping malls. My study is informative about the extent and reach of economic externalities that businesses generate in their local neighborhoods.

I use data on the spatial locations of grocery stores and other businesses to show how to find counterfactual locations for spatial treatments in settings where these are not known by the design of an actual experiment. My approach to finding counterfactual locations borrows tools from the literature on image recognition, specifically convolutional neural networks. With some adjustments, these neural networks can use the locations of observations as well as other covariates that vary across space as inputs. Through specific choices in training, they capture the economic idea that local neighborhoods and relative spatial locations, rather than absolute latitude and longitude, matter. Intuitively, the *false positive* predictions of grocery store locations are counterfactual locations that resemble (to the algorithm) real grocery store locations. These counterfactual locations are good "control" neighborhoods that are similar to neighborhoods of actual grocery stores except for the absence of one *marginal* grocery store. I demonstrate that the neighborhoods of these counterfactual locations have a similar business composition as the neighborhoods of real grocery stores.

Using data from SafeGraph[3] for the San Francisco Bay Area, I find that grocery stores cause a substantial increase in visits to restaurants within a couple of minutes walking. The number of visits to restaurants very close to real grocery store locations is two to three times the number of visits to restaurants very close to counterfactual locations. I measure visits to a restaurant as the total number of visits that SafeGraph assigns to the restaurant based on repeated smartphone location pings at the restaurant from applications that share users' locations. However, the estimated effects are close to 0 at distances larger than 0.05 miles

[3]SafeGraph is a data company that aggregates anonymized location data from numerous applications in order to provide insights about physical places.

(approximately 80 meters or 260 feet) from the front entrance of the grocery stores, suggesting the externalities of grocery stores are specific to very local neighborhoods.

While I argue in favor of this design-based, quasi-experimental strategy in this paper, the inner ring vs. outer ring empirical strategy has its own advantages, such that both are complementary. Specifically, the comparison with an "outer ring" effectively removes time-specific noise that is shared within a larger region but distinct across regions. In contrast, the methods proposed in this paper focus mostly on eliminating confounding due to differences in the spatial neighborhoods, such as population density, of treated and control individuals. Whether spatial variation in treatment locations, temporal variation in the timing of treatments, or functional form assumptions yield the most credible estimates of causal effects depends on the particular empirical setting. Estimators that are consistent for causal effects as long as either spatial or temporal variation (or both) is as good as random, or functional form assumptions are correct, are a particularly promising topic for future research. Researchers may find studies particularly credible if several distinct identification strategies lead to similar conclusions. Doubly-robust estimators (e.g. Robins and Rotnitzky, 1995; Belloni et al., 2017), which model both the outcome (conditional expectation) and assignment process (propensity score), may offer an attractive bridge between approaches.

The potential outcomes framework I develop for spatial treatments highlights nuances in interpretation of estimated average treatment effects that have received little attention in the literature thus far. Most recent empirical work estimates the effects of spatial treatments at multiple distances. However, the average effects at different distances are not generally comparable. Since some individuals are often more likely – before realization of treatment assignment – to be close to treatment locations, their treatment effects typically get more weight in average effect estimands at shorter distances, and less weight in average effect estimands at longer distances. In other words, we cannot generally interpret effect-by-distance curves or the change in effect between distances as average within-individual effects. Even the aggregate weight placed on individuals near any one treatment location varies with distance. Both of these effects can lead to estimates of average treatment effects that *increase* in distance, even though individual-level treatment effects are *decreasing* in distance for every individual. The framework in this paper allows me to characterize estimators with alternative weights on individuals to mitigate such issues. In addition, in this framework I can show how to aggregate individual-level treatment effects to estimate the aggregate effects of treatment at a location on all nearby individuals. Such aggregate treatment effects may be relevant for cost-benefit or welfare analysis.

In recent independent work, other researchers also consider treatments that are separate from outcome units. Zigler and Papadogeorgou (2021) and Aronow et al. (2020) specifically consider the spatial treatment setting with interference. The work of Aronow et al. (2020) in particular complements the present paper by proving asymptotic normality and suggesting sampling-based inference following Conley (1999) for an estimator similar to those proposed here. Borusyak and Hull (2020) use a regression, rather than potential outcomes, framework that simplifies analysis of non-binary treatments through recentering of the realized exposure by expected (over the assignment distribution) exposure. They suggest randomization inference to test sharp null hypotheses and report Conley (1999)-type standard errors for their regression estimates. McIntosh (2008) also considers spatial treatments in a regression framework. In his setting, some individuals at the same locations as the treated individuals are assumed not to be affected by the treatment, resolving the key issue of finding the right control group and suggesting a matching (based on absolute coordinates) framework for standard errors that is not applicable when such a natural control group is unavailable. The formal results in the present paper go beyond Zigler and Papadogeorgou (2021), Aronow et al. (2020), and Borusyak and Hull (2020) by proposing design-based inference instead, which is conceptually attractive and

simplifies computation of standard errors. Compared to Borusyak and Hull (2020), the inverse probability weighting estimators and analysis in the potential outcomes framework proposed here simplify achieving desirable weighting of possibly heterogeneous treatment effects. Furthermore, I propose machine learning methods that allow finding counterfactual treatment locations in the data, a crucial step when applying any of these methods on most observational data, while Zigler and Papadogeorgou (2021), Aronow et al. (2020), and Borusyak and Hull (2020) do not address settings where these are unknown in detail.

Within the larger causal inference literature, the spatial treatment setting of this paper most closely relates to work on interference and networks. Some work in causal inference explicitly considers spatially correlated treatments (Delgado and Florax, 2015; Druckenmiller and Hsiang, 2019), but is not directly applicable when spatial treatments occur at discrete points. The literature on interference is primarily concerned with spillover, or indirect, effects of treatments assigned to individuals in violation of the stable unit treatment value assumption (Rosenbaum, 2007; Hudgens and Halloran, 2008; Tchetgen Tchetgen and VanderWeele, 2012; Aronow and Samii, 2017; Vazquez-Bare, 2017; Sävje et al., 2021; Sävje, 2021; Basse et al., 2019). In contrast, in the spatial treatment setting treatment units are separate from outcome units, requiring a different definition of potential outcomes and the stable unit treatment value assumption. The key ideas in the present paper of finding the correct control group and using design-based inference are relevant irrespective of (properly defined) interference and spillovers. For settings where interference *is* present, I build on the work on exposure mappings (Aronow and Samii, 2017) for inference that incorporates the distinct restrictions that are typically plausible for spatial treatment settings. While networks (for instance Athey et al., 2018; Basse et al., 2019, for direct applications to causal inference) and spatial distance both suggest dependence between outcomes for "nearby" units,[4] the settings where either approach is the most appropriate differ: in the treated units (nodes of the network vs. separate spatial location, see also Zigler and Papadogeorgou (2021)), in the data requirements (known network vs. observable relative locations), and in the channels "transmitting," or moderating, treatment effects (network links vs. spatial proximity). Hence, also the randomization distributions typically considered for causal inference differ between these settings.

This paper contributes to the recent literature illustrating creative uses of modern machine learning methods for economic analyses.[5]. I propose a method for finding counterfactual candidate treatment locations based on an adversarial task: Finding locations that are indistinguishable (to the algorithm) from realized treatment locations. Most closely related, Athey et al. (2021) use generative adversarial networks (Goodfellow et al., 2014) to create samples for simulation studies. Kaji et al. (2020) similarly propose "adversarial estimation" to estimate structural models using generative adversarial networks. In each of these applications, the aim is to generate *synthetic* samples which look indistinguishable from the real data. In the application of this paper, only the counterfactual candidate treatment locations are synthetic, while the outcome and covariate data around it are real. In this paper, I argue in favor of *convolutional* neural networks in particular, based on the similarity between spatial data and image data, which sparked more recent developments in this method (Krizhevsky et al., 2012). Relative spatial positions are similar to relative positions of pixels, and different covariates at each location correspond to the different color channels of images. For economic applications using satellite data (see Donaldson and Storeygard, 2016, for a review), convolutional neural networks have also shown promise (e.g. Jean et al., 2016; Engstrom et al., 2021). Convolutional neural networks are particularly attractive for spatial settings because they build on relevant economic

[4]In some cases, one may think of both network structure and relative spatial locations as different representations of the same underlying latent space (cf. Hoff et al., 2002) that would be used directly if it was observable.

[5]See Mullainathan and Spiess (2017); Glaeser et al. (2018); Athey (2018); Gentzkow et al. (2019); Athey and Imbens (2019) for recent reviews

intuition for regularization: While the geographic space might be large and high-dimensional, the immediate spatial neighborhood often matters the most, and relative distances matter similarly at different absolute locations. Through careful design decisions, the methods I propose for the spatial treatment setting retain some interpretability in addition to the good performance commonly associated with "black box" machine learning algorithms.

The framework and methods discussed in this paper may also prove useful for causal inference questions not directly related to spatial treatments. First, other non-spatial settings similarly feature "treatments" that are not directly assigned to individuals but affect them based on some measure of distance. In this paper, I briefly discuss Bartik (1991)-, or shift-share, instruments, where for instance industry-level shocks affect all cities depending on industry composition, with exposure of a city to the shocked industries as the distance measure.[6] Second, I develop an approach to finding suitable counterfactual candidate locations in observational data based on flexible machine learning methods. This approach may extend to other settings with dependency between observations where it is sometimes challenging to find (good) control observations, such as event studies and other time series settings. Third, separating treatment assignment and outcome individuals in this framework further clarifies distinctions between design-based and sampling-based inference (Abadie et al., 2020). While design-based inference captures variation in treatment locations, sampling-based inference can reflect sampling of individuals at fixed locations, within fixed regions (infill asymptotics (Cressie, 1993) in the spatial statistics literature), of a growing contiguous space (expanding domain asymptotics (Cressie, 1993)), or of independent regions (clustering). The present paper focuses on design-based inference specifically; in-depth comparisons of different modes of inference are beyond its scope.

My current analysis is limited in at least three important ways. First, I assume that outcome individuals have fixed locations. This is problematic if individuals move, or migrate, strategically in response to the treatment. In such a case, researchers can use pre-treatment locations, if such data are available, and estimate the *reduced form* effect of the treatment on those who were nearby before the treatment occurred. Second, the framework is not directly applicable to settings where researchers are interested in the causal effects of spatially correlated characteristics of places, such as in the literature on social mobility (e.g. Chetty et al., 2014). Instead, the present paper focuses on treatments that occur at discrete locations in space, rather than a general treatment intensity surface. The ideal experiment of randomizing treatment locations yields a distinct and tractable randomization distribution for the estimators studied in this paper. Third, there are no claims of efficiency in the current paper. While I attempt to offer theory and estimators for a variety of spatial treatment settings, the primary focus of this paper lies in developing a coherent conceptual framework that allows me to characterize, discuss, and exploit the ideal experiment with spatial treatments. In particular, the present paper provides no formal justification for the use of methods from the literature on sample splitting and double robustness (e.g. Chernozhukov et al., 2018). The need to consider many relative spatial locations for finding suitable counterfactual candidate locations makes this a high-dimensional estimation problem in observational settings, suggesting the importance of methods and insights from that literature.

The remainder of this paper is organized as follows. Section 2 discusses how the assumptions required for the existing inner vs. outer ring strategy differ from a (quasi-) random treatment location assumption, which can be justified by the ideal experiment underlying the estimators proposed in this paper. Section 3 develops a potential outcomes framework for spatial treatments. Section 4 contains the main results on identification, estimation, and inference under the ideal experiment. Section 5 discusses how to extend these results to

[6] Adao et al. (2019) and Borusyak et al. (2022) take a similar design-based view where the "shifts" are random. See Goldsmith-Pinkham et al. (2020) for an alternative that considers the "shares" to be random.

additional settings of empirical relevance. Section 6 shifts the focus from experimental to observational data, proposing assumptions and methods that allow researchers to emulate the ideal experiment. Section 7 shows how to apply these methods in practice. In the conclusion, I discuss fruitful directions for future research on causal inference for spatial treatments.

2 Comparison To Inner vs. Outer Ring Empirical Strategies

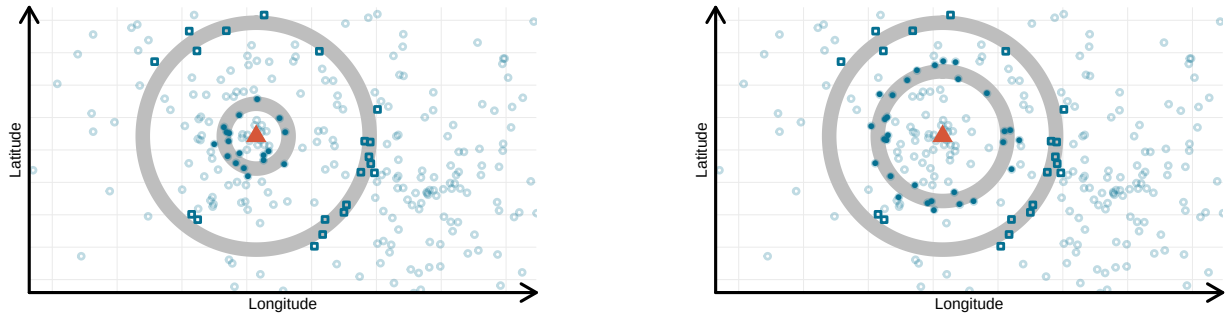
I compare the "inner vs. outer ring" empirical strategy used by most existing empirical studies of spatial treatment and the (quasi-) experimental strategy proposed in this paper.

To build intuition, consider a simplified setting of a spatial treatment occurring at a location $s \in \mathbb{R}$ in one-dimensional space. The researcher is interested in the effect the treatment at s has on outcomes Y_i of individuals i who are also located in the same one-dimensional space. Specifically, the researcher wants to estimate the effect of the treatment on individuals who are a short distance d^{short} away from treatment; that is, individuals at locations $r_i = s - d^{\text{short}}$ and $r_i = s + d^{\text{short}}$.

The "inner vs. outer ring" strategy and the experimental strategy differ in their choices for a control group. When there is treatment at location s, the researcher observes the "treated" potential outcomes for individuals at locations $s - d^{\text{short}}$ and $s + d^{\text{short}}$. The question is which other individuals have outcomes similar to the outcomes that "inner ring" individuals at $s \pm d^{\text{short}}$ would have had without treatment at location s. To make any progress on this question, one needs to assume that there are some individuals, typically individuals sufficiently far away from treatment, who are unaffected by it. Otherwise, one cannot, in general, learn about the (potential) outcome in the absence of treatment. The "inner vs. outer ring" strategy chooses "outer ring" individuals at locations $s - d^{\text{long}}$ and $s + d^{\text{long}}$ as the control group, where $d^{\text{long}} \gg d^{\text{short}}$. Figure 1 illustrates the comparison underlying this empirical strategy in two-dimensional space. The experimental strategy instead chooses individuals at locations $s' - d^{\text{short}}$ and $s' + d^{\text{short}}$ near a *counterfactual* location s' where treatment could plausibly have occurred, but (quasi-) randomly did not. Intuitively, if the treatment occurs in a business district location s such that individuals at $s \pm d^{\text{short}}$ are in the business district, then a good control group may be individuals at $s' \pm d^{\text{short}}$ in other business districts s' where treatment did not occur for exogenous reasons, while individuals at $s + d^{\text{long}}$ may be outside any business district. Formally, I justify such estimators with an ideal experiment where the locations of the treatment are randomized between some plausible *candidate* treatment locations.

For simplicity, suppose that individuals are located uniformly across space and that outcomes in the absence of treatment are quadratic in location. That is, $Y_i(0) = r_i^2$ where $Y_i(0)$ is the potential outcome of individual i at location $r_i \in \mathbb{R}$ if there is no spatial treatment nearby. Then the average outcome individuals at $s \pm d^{\text{short}}$ would have had in the absence of treatment is $\frac{1}{2}(s - d^{\text{short}})^2 + \frac{1}{2}(s + d^{\text{short}})^2 = s^2 + (d^{\text{short}})^2$ where the equal weights $1/2$ are due to the uniform distribution of individuals across space.

This example illustrates that the outer ring control group will not, in general, estimate the outcome in the absence of treatment correctly, even if the treatment location was randomized and even if the researcher has access to a large sample. Instead, the outer ring individuals have average outcomes $s^2 + (d^{\text{long}})^2 > s^2 + (d^{\text{short}})^2$. That is, no matter how the location s was chosen, whether endogenously or randomly in an experiment, the outer ring control group over-estimates the outcome in the absence of treatment by $(d^{\text{long}})^2 - (d^{\text{short}})^2$. Observing multiple separate one-dimensional spaces does not resolve this issue, irrespective of how the treatment locations are chosen in each space. It may be tempting to consider the limit of $(d^{\text{long}})^2 - (d^{\text{short}})^2$ as $d^{\text{long}} \to d^{\text{short}}$. However, when $d^{\text{long}} \approx d^{\text{short}}$, the individuals at $s \pm d^{\text{long}}$ are themselves affected by the



(a) $\hat{\tau}(d^{\text{short}})$ for short distance d^{short}

(b) $\hat{\tau}(d^{\text{med}})$ for medium distance d^{med}

Figure 1: Existing estimators focus only on locations where the treatment did occur. In this figure, the realized treatment location is shown as a filled-in triangle. The treatment group consists of individuals in an "inner ring" at a given distance of interest from treatment, here displayed as small filled-in circles. The control group consists of individuals in an "outer ring" who are farther away from realized treatment, here displayed as hollow squares. With panel data, existing estimators use pre- and post-treatment data for both groups in a difference-in-differences setup. Typically, when researchers estimate the effect at multiple distances, the same control group is used for all distances. Panel a shows the estimator $\hat{\tau}(d^{\text{short}})$ for a short distance d^{short}. Panel b shows the estimator $\hat{\tau}(d^{\text{med}})$ for a medium distance d^{med}.

treatment, and hence their outcomes may be uninformative about the outcome in the absence of treatment.

The (quasi-) experimental estimator instead uses individuals near counterfactual locations such that treatment assignment is as-good-as-random between realized and counterfactual locations. In the example of quadratic outcomes, if treatment is realized at a location s, then a counterfactual location is $s' = -s$. The average control potential outcomes of individuals at $s \pm d^{\text{short}}$ are $s^2 + (d^{\text{short}})^2$; but it is unobserved due to treatment at location s – instead, treated potential outcomes are observed. The average control potential outcomes of individuals at $s' \pm d^{\text{short}}$ are $(s')^2 + (d^{\text{short}})^2$, which is the same because $(s')^2 = (-s)^2 = s^2$. Since individuals at $s' \pm d^{\text{short}}$ could, in general, also be affected by treatment at location s, the simplest experimental estimators use multiple separate one-dimensional spaces, where treatment in one space does not affect outcomes in other spaces. Then, by randomizing which of these spaces have *any* treatment, one can use a control group comprised of individuals at $s \pm d^{\text{short}}$ and $(-s) \pm d^{\text{short}}$ from a space without treatment. I propose a machine learning algorithm to find such counterfactual locations in real applications where a (absolute or relative) location s does not directly transfer to another space, and outcomes are not quadratic so a location $-s$ may not offer a suitable counterfactual either. Furthermore, I show that even if one has data from only one space with several realized treatment locations, it is possible to estimate similar treatment effect estimands and do inference along similar lines.

Some existing work explicitly argues that the location of the treatment is as-good-as-random within some set of locations, justifying the estimators proposed in this paper. Linden and Rockoff (2008), as quoted in the introduction, argue that "the exact locations available at the time individuals seek to move into a neighborhood are arguably exogenous." Diamond and McQuade (2019, p. 1065) similarly state that "the exact geographic location of the development site within a broader neighborhood appears to be determined by idiosyncratic characteristics, such as which exact plot of land was for sale at the time." In general, the ideal experiment for spatial treatments, in analogy to more familiar individual-level treatments, randomizes the presence and absence of treatments. In this paper, I propose estimators and inference procedures for this ideal experiment, as well as methods for observational data that allow researchers to emulate the ideal experiment.

The inner vs. outer ring strategy, as demonstrated in the example above, is not justified by the same ideal experiment; instead it relies on two, partly conflicting assumptions: First, the outer ring individuals must have control potential outcomes that are comparable to the control potential outcomes of inner ring individuals. Second, the outer ring individuals must be unaffected by the treatment.

The outer ring individuals are most plausibly comparable to inner ring individuals if the outer ring has *small radius* so inner and outer individuals are close to one another. In many applications with spatial data, individuals who are close to one another in space tend to have similar outcomes. Especially when treatment locations are chosen strategically, distance from treatment is likely to correlate with other characteristics that affect outcomes. For instance, if treatments occur in city centers, inner ring individuals may be located in inner cities while outer ring individuals, if the radius of the outer ring is too large, are located on the outskirts of cities. Typically, researchers estimate the effect of the treatment at multiple distances, comparing different inner ring individuals to the same outer ring individuals (for instance, distances d^{short} and d^{med} in Figure 1). Hence, in practice they need to assume that the potential outcome surface is flat on average (or, with panel data, on parallel trends as discussed below) for the entire areas around the treatment locations up to and including the outer rings.

The outer ring individuals are most plausibly unaffected by the treatment if the outer ring has *large radius* so outer ring individuals are far away from treatment. If individuals on the outer ring were directly affected by treatment, their observed outcomes would not generally be informative about the potential outcomes in the absence of treatment for the inner ring individuals, even if these individuals are comparable in the absence of treatment. The effects of many spatial treatments plausibly decay with distance to treatment. Assuming that the treatment has no effect after a certain distance (cf. Assumption 4) then provides a reasonable approximation and may be necessary for any inference on causal effects.

The conflict between an outer ring with smaller or larger radius is inherent in this empirical strategy and cannot be alleviate through more flexible functional form or additional data. The choice of radius for the outer ring needs to carefully balance the two conflicting assumptions.

Using panel data rather than cross-sectional data does not necessarily resolve this conflict: Panel data allows potential outcomes of inner and outer ring individuals to be at different levels, but requires parallel trends. Potential outcomes may or may not follow parallel trends in the post-treatment period, even if trends appear parallel in pre-treatment periods.[7] Especially when treatment locations are chosen strategically, the (approximate) timing of treatments may also be chosen strategically even if some other factors, such as bureaucratic processes, also have some influence on the exact timing. When timing of treatment *is* exogenous, empirical strategies exploiting it can complement the strategy based on exogenous variation in treatment location proposed in this paper. In a difference-in-differences design, a control group based on the counterfactual candidate treatment locations that are the focus of this paper may offer an attractive alternative to outer ring control groups. In the earlier example, it would imply assuming parallel trends of potential outcomes in inner of similar cities, some with treatment, some without, rather than assuming parallel trends of potential outcomes of inner cities and outskirts of the cities with treatment. Which parallel trends assumption is more credible depends on the particular application, but researchers should clearly

[7]Assessing if pre-trends are parallel at many different distances from treatment may be problematic by itself (for analysis with a single treatment group, relevant for a single inner ring compared to a fixed outer ring, see for instance Manski and Pepper, 2018; Freyaldenhoven et al., 2019; Rambachan and Roth, 2019) and difficult to achieve visually in the familiar event study plots. With multiple inner rings, event study plots become three-dimensional (time, distance from treatment, outcome) and include plotting multiple planes if standard errors are to be included, where whichever plane is plotted "on top" in a two-dimensional projection would presumable obscure the other two planes. See Diamond and McQuade (2019, Figures 3, 4, 5) for an example of such a three-dimensional event study plot but without standard errors.

articulate which individuals have potential outcomes following parallel trends and why such an assumption is plausible. Exogenous variation in the locations of spatial treatments does not guarantee these parallel trends of inner and outer rings.

Exogenous variation in the locations of spatial treatments, even if created by a true randomized experiment, does not generally justify the assumptions of the inner vs. outer ring empirical strategies. In the earlier example, no matter where treatment occurred, the outer ring strategy overestimates the control potential outcomes of the inner ring individuals. Except for knife-edge cases involving a favorable combination of outcome surface curvature and the distribution of individuals across space, the outer ring individuals are not good control units for the inner ring individuals. The most plausible such edge case is one where the potential outcome surface is flat. Yet a flat potential outcome surface cannot be guaranteed by experimental design; it inherently depends on the setting and individuals and is outside an experimenter's control. To justify the comparison of inner and outer ring individuals, researchers should therefore argue that the outcome surface is flat on average (or there are no systematic trends at any distance from treatment if using panel data), rather than arguing that the location of the spatial treatments was exogenous. In such a case, researchers should justify why in their settings the particular spatial treatment of interest may cause differences in outcomes by distance from treatment, while all other features of geography, institutions, or types of individuals are either distributed uniformly across space or have effects that do *not* affect outcomes at different distances differently. Additionally, if one believes that the outcome surface is flat within some known area, there is no rationale for relying on an outer *ring*. Researchers could use all individuals who are far enough away from the treatment location to be unaffected by treatment but within the area where the outcome surface is flat.

The primary advantage of an outer ring control group under the ideal experiment is a reduction in variance. If there are substantial differences between the larger neighborhoods of different treatment locations, the variance of potential outcomes across treatment locations is large. Similar to standard randomized experiments (cf. Imbens and Rubin, 2015), this marginal variance of potential outcomes also appears in the variance of estimators of spatial treatments proposed in this paper. If individuals of an outer ring are in the same larger neighborhoods as the inner ring individuals, and have similar levels of potential outcomes, differencing outcomes of inner and outer ring individuals can reduce the variance that is due to "shocks" that are shared within the larger neighborhoods. Importantly, however, the outer ring individuals still must not be affected by the treatment themselves (otherwise one estimates an effect relative to the effect on the outer ring), and only yield identification of causal effects under appropriate functional form assumptions.

If causal effects are identified because the locations of spatial treatments are plausibly exogenous, the methods proposed in this paper allow researchers to align their empirical strategy with the source of exogenous variation and to emulate the ideal experiment. The methods proposed here can be combined with a differencing approach of inner and outer rings to reduce variance. The outer ring then plays a similar role to a pre-treatment period in a difference in differences design with random treatment assignment: Identification of causal effects is based on randomization, but the differencing with pre-treatment (outer ring) outcomes can reduce variance.

3 Setup: Spatial Treatments

In this section, I propose an extension of the potential outcomes notation (cf. Neyman, 1923, 1990; Rubin, 1974; Imbens and Rubin, 2015) that treats the level of treatment assignment as conceptually distinct from the level at which outcomes are measured. This distinction separates the intervention that is the cause of the

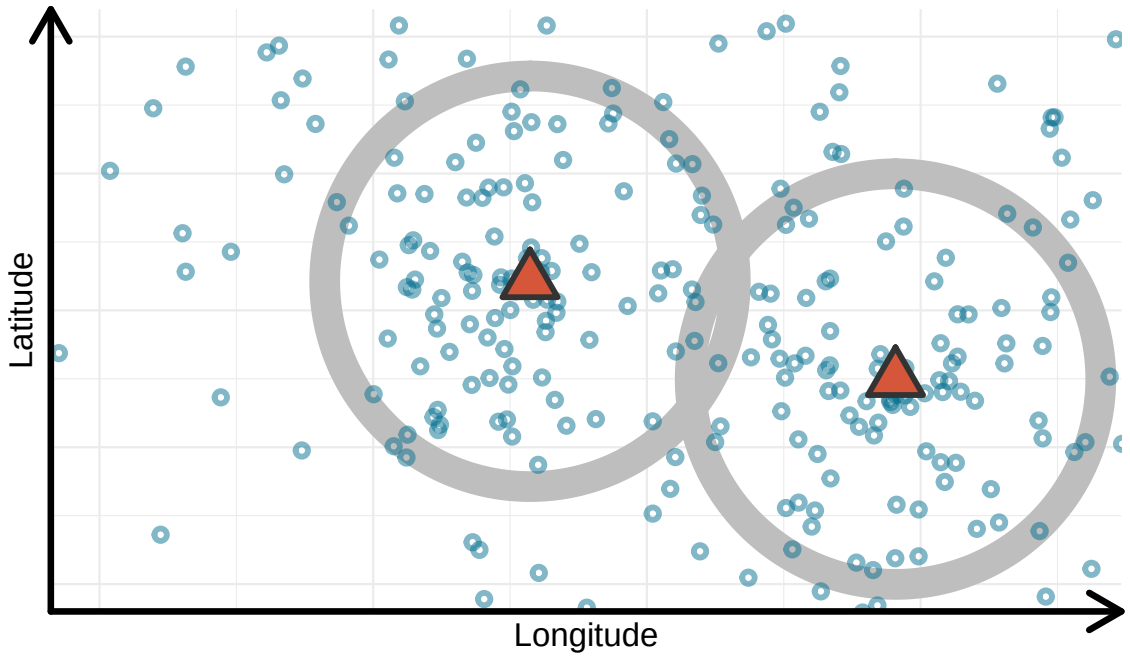


Figure 2: Illustration of the setup. While typically only relative locations matter, locations are often given by their "GPS coordinates" as latitude and longitude. In the figure, the *candidate treatment locations* where the treatment may occur are given by triangles. The small circles indicate the *locations of individuals*. The researcher typically estimates the *treatment effects*, caused by treatment at one of the candidate locations and experienced by the individuals, conditional on *distance* from treatment. When the (weighted) Euclidean distance function is used, individuals within a narrow distance bin from a candidate location are located on a ring, here displayed as an area shaded gray. If driving time is used instead to measure distance, individuals at a given distance need not be located on a circular ring. The figure shows data from a single *region*. In the baseline setting of this paper, the researcher has data from multiple such regions, with treatment realized only in some of them. If treatment is realized at multiple (both) candidate locations (triangles) within the same region, there is potential *interference* between them, complicating estimation and inference. In the baseline setting, the *probability of treatment* at locations and in regions describes a two-stage process. In the first stage, a number of regions are chosen randomly for treatment somewhere in the region. In the second stage, a single candidate location in each chosen region is chosen randomly to receive treatment.

effect from the individuals for whom the effect is measured. It allows me to formally characterize estimands of interest, and to derive estimators and their properties in the following sections.

With spatial treatments, potential outcomes of individuals are functions not of an individual-level binary or continuous treatment, but of a set of candidate treatment locations.

Treatment Locations We are interested in the effects of spatial treatments. Let $\mathbb{S}$ denote the set of *candidate treatment locations*, shown as triangles in Figure 2. The set of candidate treatment locations is finite; in the example of Figure 2 just two locations in the region shown. This reflects an inherent scarcity that is common to most applications: Only a small number of locations are ultimately realized, and most locations are infeasible, unsuitable, implausible, or exceedingly unlikely for the treatment.

In spatial settings, the candidate locations are typically given by latitude and longitude or other (potentially relative) coordinates, such that $\mathbb{S} \subset \mathbb{R}^2$. Throughout this paper, the set $\mathbb{S}$ is finite, as virtually any practical application will be based on some discretized, or rounded, locations. One can, however, take $\mathbb{S}$ as defining a

finely spaced grid over $\mathbb{R}^2$. This is convenient to conceptualize situations where treatment could be realized anywhere with some positive probability. The random variable $\xi \subset \mathbb{S}$ denotes the set of *realized treatment locations*.

Individual Locations and Potential Outcomes We measure the outcome of interest for units indexed by i. For the remainder of this paper, I will refer to these outcome units as *individuals*, but in some settings i may be a business, census tract, or similar, typically small, unit with fixed geographic location. Denote the set of all individuals by $\mathbb{I}$. Individual i has spatial location, or residence, r_i, shown as small circles in Figure 2. Throughout this paper, I assume that the locations of individuals are fixed; there is no migration. In some applications, r_i corresponds to, for instance, the workplace of individual i rather than their residence. The location of i is in the same space as the candidate treatment locations, such that typically $r_i \in \mathbb{R}^2$ are latitude and longitude. Define potential outcomes for each individual $i \in \mathbb{I}$

$$\text{Potential Outcomes:} \qquad Y_i(S) \quad \forall S \subset \mathbb{S}$$

as the outcome for individual i if treatment is realized in locations $S \subset \mathbb{S}$. To simplify notation, and consistent with standard potential outcomes notation, let the potential outcome of individual i in the absence of any realized treatment be $Y_i(0) \equiv Y_i(\emptyset)$.

Treatment Effects The treatment effects of primary interest contrast some treatment vector $S \subset \mathbb{S}$ with the absence of realized treatments, $S' = \emptyset$. Specifically, I define the effect of S on an individual $i \in \mathbb{I}$ as

$$\text{Treatment Effects:} \qquad \tau_i(S) \equiv Y_i(S) - Y_i(0) \quad \forall S \subset \mathbb{S}$$

Oftentimes, the treatment vector of interest, S, is a singleton, $S = \{s\}$ for a single candidate location $s \in \mathbb{S}$. With slight abuse of notation,[8] define

$$\text{Treatment Effects:} \qquad \tau_i(s) \equiv Y_i(s) - Y_i(0) \quad \forall s \in \mathbb{S}.$$

I define meaningful *average treatment effects* in Section 4. These average treatment effects average across both individuals i and treatment vectors S.

Distances Distances between treatment locations and individuals are central to defining interesting average treatment effects in Section 4. For instance, the researcher may estimate the average effect of a treatment *at a distance* of 1 mile. In Figure 2, the areas shaded gray highlight all locations approximately 1 mile away from any candidate treatment location. The distance between treatment location $s \in \mathbb{S}$ and individual $i \in \mathbb{I}$ is given by a distance function

$$\text{Distance Function:} \qquad d(s, r_i) \geq 0$$

Importantly, the distance between two locations must be observable (to the researcher) and must not be affected by treatment assignment, ruling out migration in response to the treatment.[9]

[8] In this paper, I consistently use lower-case letters for individual locations and upper-case letters for sets of locations. The random variable ξ of realized locations is always a set of locations. The random variable ξ_j of realized locations in region j is a particular location in the baseline setting of Section 4, and a set of locations in some of the extensions in Section 5.

[9] By fixing distances prior to treatment, one can estimate the "reduced form" effect of *assigned* distance to treatment even in the presence of migration.

The distance function is used for two purposes. First, to estimate heterogeneous average treatment effects by distance from treatment. Second, to assume distances at which treatments have no effect to limit interference and to thereby aid in estimation and inference.

When locations are given as Cartesian coordinates, we can use the Euclidean distance in $\mathbb{R}^2$:

$$\text{Euclidean Distance:} \qquad d(s, r_i) = \sqrt{(s_1 - r_{i,1})^2 + (s_2 - r_{i,2})^2}$$

When locations are given by latitude and longitude, the Great Circle distance is more accurate than a Euclidean distance with fixed weights on latitude and longitude.[10]

For some applications in social sciences, driving distances are arguably more relevant. Suppose the spatial treatment corresponds to an employer opening a new location. Then an individual's access to the treatment, and hence treatment effect, likely depends on driving time rather than straight line distance. However, computing driving times between many locations may be computationally and financially expensive. When using straight line distances instead, some interpretability, but not validity, is lost.

We can also study the effects of state-wide policies and other *clustered assignments* in this framework. In this setting, each candidate treatment location $s \in \mathbb{S}$ corresponds to one cluster, or state. The appropriate distance function for this setting is

$$\text{Cluster Membership:} \qquad d(s, r_i) = \begin{cases} 0 & \text{if } s \text{ and } r_i \text{ are in the same cluster} \\ 1 & \text{otherwise} \end{cases}$$

In the simplest case of state-wide policies, we use this distance function to estimate the treatment effect at a distance of 0. This corresponds to estimating the treatment effect of the policy by comparing individuals in treated states to individuals in untreated states where the policy could also have been implemented. We can generalize the cluster membership function to be smooth in distance to a treated state: For individuals in treated states, this distance is 0. For individuals in untreated states, the distance is smallest if they are most exposed to treated states. Exposure may measure, for instance, distance to the state border, shared media markets, number or cost of flights between airports, or the relevance of the industries of treated state s to i's occupation.[11]

Regions In many applications, it is convenient to group individuals and treatment locations into *regions*. For instance, in a sample of data from different cities, individuals and treatment locations of each city may form a separate region. When regions are not directly coded in the data, one can sometimes define regions based on geographic proximity such that treatment locations only have effects within their own regions. That is, no individual is close enough to candidate treatment locations from two or more distinct regions to be affected by both of them. Figure 2 shows data from one such region. In the baseline setting of this paper, the researcher has access to data from multiple such regions with at most one candidate location realized in each region, but this requirement is relaxed in Section 5.1.

Throughout, I denote regions by subscripts $j = 1, \ldots, J$. Let $\mathbb{S}_j \subset \mathbb{S}$ be the set of candidate treatment locations within region j. The set of realized treatment locations within region j is ξ_j. If treatment is realized

[10] For instance, moving 0.01 degrees west in New York City corresponds to a distance of about 0.52 miles. Moving 0.01 degrees west in Miami corresponds to a distance of 0.62 miles. This occurs because the distance between degrees of longitude is largest at the equator and converges to 0 at the poles, while the distance between degrees of latitude is (approximately) fixed.

[11] The example of industries is the original inspiration for Bartik (1991), or shift-share, instruments. See Adao et al. (2019); Goldsmith-Pinkham et al. (2020); Borusyak et al. (2022) for recent treatments in econometrics, and Section 5.2 for how the framework of this paper relates.

within region j, $\xi_j \neq \emptyset$, let $W_j = 1$, and otherwise, $\xi_j = \emptyset$, let $W_j = 0$. If $W_j = 1$, I say that region j "is treated" or "is a treated region." Analogously, if $W_j = 0$, I say that region j "is a control region." Let $\mathbb{I}_j \subset \mathbb{I}$ be the set of individuals i with residence in region j. The region where individual i resides is given by $j(i)$, such that $i \in \mathbb{I}_{j(i)}$.

INTERFERENCE The notation in this paper can be seen as an extension of the notation of the literature on interference (cf. Aronow and Samii, 2017). Consider first a setting with individual-level treatments. Let A_i be the treatment assigned to an individual $i = 1, \ldots, n$, and $A \in \{0,1\}^n$ be the vector stacking all of the A_i. In the absence of interference, that is, under the stable unit treatment value assumption (cf. Imbens and Rubin, 2015), the observed outcome of individual i is $Y_i = Y_i(A_i)$. With interference, the outcome of individual i may depend not only on her own treatment assignment, but also on the treatment assignment of other individuals. That is, the potential outcomes of i are functions of the entire vector A rather than only her own A_i, and her observed outcome is $Y_i = Y_i(A)$. Notationally, spatial treatments generalize this setting by allowing A to have a dimension other than n, the number of individuals. For the closest analogy, enumerate the candidate treatment locations by $k = 1, 2, \ldots, K$ where $K \equiv |\mathbb{S}|$ is the finite number of candidate treatment locations. The random variable of realized treatment locations takes on values $\xi \in \{0,1\}^K$, such that $\xi_k \equiv 1$ whenever the k-*th* candidate location is treated, and $\xi_k \equiv 0$ otherwise. The realized outcome for individual i is then $Y_i = Y_i(\xi)$, where ξ is K rather than n dimensional.

Consider the example given in Figure 2. Some individuals are at a distance of 1 mile from both treatment locations. If the treatment has an effect at that distance, the treatment states of both candidate locations jointly determine the observed outcome. The two candidate locations can interfere because conditional on the treatment state of just one of them, the outcome for some individuals still varies with the treatment assignment of the other candidate location.

The literature on interference is typically interested in answering (at least) one of two questions. First, what is the effect of changing i's treatment status, holding the treatment status of i's neighbors fixed? Second, what is the effect of changing the treatment status of i's neighbors, holding the treatment status of i fixed?

With spatial treatments, neither of these questions is of primary interest. If i is 1 mile away from a realized treatment location, then a neighbor of i, say i', is also approximately 1 mile away from the same realized treatment location. A counterfactual where i is 1 mile away from a realized treatment location, while her neighbor i' is not, is typically neither feasible nor relevant in practice. The treatment does not spill over from i to i', it affects both of them directly, such that decompositions into direct and indirect effects (cf. Hudgens and Halloran, 2008) are not well defined.[12]

Interference in the spatial treatment setting refers to multiple treatment locations affecting the same individual, rather than the treatment or effect of one individual spilling over to another individual. Formally, a treatment location s affects an individual i if for some set of treatment locations $S \subset \mathbb{S}$, the outcome of i changes when s is included or excluded: $Y_i(S \cup \{s\}) \neq Y_i(S \setminus \{s\})$. Two treatment locations $s, s' \in \mathbb{S}$ interfere with one another if there is an individual i affected by treatment at both locations, that is $\exists i \in \mathbb{I}, S, S' \subset \mathbb{S}$ with $Y_i(S \cup \{s\}) \neq Y_i(S \setminus \{s\})$ and $Y_i(S' \cup \{s'\}) \neq Y_i(S' \setminus \{s'\})$.

In spatial treatment settings, it is often natural to assume that treatment locations that are far away from an individual do not affect her. Formally, assume that whenever $d(s, r_i) > d^{\max}$ for some sufficiently

[12] If some individuals are optionally excluded from accessing a treatment location, we can think of them as two *distinct* locations s and s' sharing the same geographic location. Suppose an individual i has access to location s but not to location s'. Then $\tau_i(s)$ is the effect of an accessible treatment at that geographic location. In contrast, $\tau_i(s')$ is the effect of a spatial treatment at that geographic location to which i does not have access but some neighbors of i may have access. Loosely speaking, $\tau_i(s')$ is then the spillover, or indirect, effect on i.

large distance $d^{\max}$, $Y_i(S \cup \{s\}) = Y_i(S \setminus \{s\})$ for all $S \subset \mathbb{S}$. Assumption 1 formally states that there is no interference across regions.

Assumption 1 (No Interference Across Regions)**.** Individuals in region j are unaffected by treatment locations in regions $j' \neq j$. That is, for $i \in \mathbb{I}_j$ and $S \subset \mathbb{S}$,

$$Y_i(S) = Y_i(S')$$

whenever $S \cap \mathbb{S}_j = S' \cap \mathbb{S}_j$.

Regions are sufficiently far apart that individuals in one region are unaffected by treatment locations in another region. The results under the baseline setting of this paper, however, fundamentally rely on the absence of interference between treatment locations that are far apart, not on separate regions. Section 5.1 discusses a setting where all data available to the researcher comes from a single large contiguous region. In that setting, identification of the simple estimands of Section 4 requires additional assumptions, but the estimators I propose in Section 5.1 retain some interpretability and inference procedures are valid even if those assumptions are violated. The separate region framework, however, helps clarify key concepts by simplifying estimators, and it is applicable to a large number of empirical studies. The assumption that treatment locations only affect individuals within the same region is similar in spirit to assumptions that interference or spillovers are limited to family members, classrooms, or other subgroups in settings with individual-level treatments (e.g. Vazquez-Bare, 2017).

Treatment Probabilities The assignment mechanism (Imbens and Rubin, 2015) determines the probabilities with which treatment is realized at each of the candidate treatment locations. The marginal probability that treatment is realized at a location $s \in \mathbb{S}$ is given by $\Pr(s \in \xi)$. In the main part of the paper, I consider a two-stage assignment mechanism that imposes structure on $\Pr(s \in \xi)$ as well as on the conditional probabilities $\Pr(s \in \xi \,|\, s' \in \xi)$ and $\Pr(s \in \xi \,|\, s' \notin \xi)$. In the first stage, either a fixed number of regions is chosen to receive treatment, or assignment is through independent Bernoulli trials (coin flips). In the second stage, a single location receives treatment in each treated region. I discuss methods for some observational settings that deviate from this assignment mechanism in sections 5.1.

Suppose the randomization of treatments across regions takes the form of a completely randomized experiment with a fixed number of treated regions. Assumption 2 formalizes this design together with an assumption that each region is equally likely to be treated. Define $\pi_j \equiv \Pr(W_j = 1)$ for $j = 1, \ldots, J$ to be the probability that a region receives treatment. Note that the completely randomized design differs from experiments that are paired or stratified at the region-level. Results for stratified experiments are generally similar and can be obtained by substituting the appropriate covariances of treatment indicators in the proofs. Estimating the variance of estimators under paired designs is often difficult (e.g. Bai et al., 2019, for individual-level treatment assignment), but does not contribute conceptually to our understanding of the spatial treatment setting.

Assumption 2 (Completely Randomized Experiment)**.** Regions are chosen for treatment according to a completely randomized design (e.g. Imbens and Rubin, 2015, ch. 4.4) where each region has equal marginal probability of receiving treatment somewhere, $\pi = \pi_j$ for all regions j. That is, all assignment vectors $W \in \{0,1\}^J$ with $\sum_j W_j = J_t \equiv \pi J$ are equally likely, and assignments with $\sum_j W_j \neq \pi J$ have zero

probability:

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

As an alternative to completely randomized designs with fixed probability of treatment, I also consider designs where treatment is decided by independent coin flip for each region, potentially with different probabilities. Assumption 3 below formalizes this assumption.

Assumption 3 (Bernoulli Trial)**.** Regions are chosen for treatment according to a Bernoulli trial (e.g. Imbens and Rubin, 2015, ch. 4.3) where region j has marginal probability π_j of receiving treatment somewhere and assignment is independent across regions. That is, the probability of assignment $W \in \{0,1\}^J$ is

$$\Pr(W) = \prod_{j=1}^{J} \pi_j^{W_j} (1-\pi_j)^{1-W_j}$$

such that the number of treated regions varies.

In the main part of the paper, I consider a setting with exactly one treated location in each treated region. This restriction of the assignment mechanism rules out interference by design under the minimal assumption that treatments have no effects across regions. For each candidate treatment location in a region, $s \in \mathbb{S}_j$, define the probability of treatment conditional on the region receiving treatment as $g_j(s) \equiv \Pr(s \in \xi \,|\, W_j = 1)$. Then, by the definition of conditional probabilities, $\Pr(s \in \xi) = \Pr(s \in \xi \,|\, W_j = 1) \Pr(W_j = 1) = g_j(s)\pi_j$. The notational distinction between treatment of regions and treatment of particular locations within regions is motivated by an asymmetry in which potential outcomes are observed: In control regions, the control potential outcomes are observed for all individuals near each (counterfactual) candidate treatment location. In treated regions, in contrast, only the treated potential outcomes corresponding to one particular treatment location are observed for all individuals. This asymmetry is apparent in the estimators and variances throughout section 4.

4 Experimental Setting: Estimation and Inference

In the ideal experiment, treatment is randomized between candidate locations in a way that rules out interference by design. In this baseline setting, locations are grouped into separate regions, and some regions are randomly chosen for treatment assignment, followed by randomization of the location of treatment among candidate locations within each treated region, as outlined in Section 3. I show how to use this random variation for identification, estimation, and inference of individual-level and aggregate treatment effects.

4.1 Individual-Level Effects

Individual-level effects express the average effects of treatment locations on individuals.

4.1.1 Average Treatment Effect on the Treated

The most intuitive estimator of the average effect of a spatial treatment on nearby individuals takes the simple average of individuals near realized treatment and subtracts from it the average outcome of properly

chosen control individuals. In this section, I first show who the proper control individuals are under the ideal experiment of random variation in treatment locations. Then I present properties of this estimator and discuss its interpretation as the average treatment effect on the treated.

The average of individuals who are treated *at a distance* $d \pm h$ from a treated location is

$$\bar{Y}_t(d) = \frac{1}{\sum_{i\in\mathbb{I}} W_{j(i)} \mathbb{1}\{|d(\xi_{j(i)}, r_i) - d| \leq h\}} \sum_{i\in\mathbb{I}} W_{j(i)} \mathbb{1}\{|d(\xi_{j(i)}, r_i) - d| \leq h\} Y_i$$

where $W_{j(i)} = 1$ if and only if individual i is in a region $j(i)$ that is treated. The indicator function equals 1 if and only if the distance between individual i and the realized treatment location in her region, $\xi_{j(i)}$, is within the *distance bin* of distances between $d - h$ and $d + h$. For instance, to estimate the average outcome for individuals who are between 1 and 2 miles from treatment, calculate $\bar{Y}_t(1.5)$ with $h = 0.5$.

The choice of control individuals to compare this average of treated individuals to is less obvious. Recent empirical studies compare the treated to controls on an outer ring; that is, to individuals i' in treated regions ($W_{j(i')} = 1$) who are farther away from treatment. Effectively, this estimates the treatment effect at distance d as $\bar{Y}_t(d) - \bar{Y}_t(d')$ where $d' \gg d$. In analogy to individual-level randomized experiments, one might also consider taking the simple average of individuals in control regions, $\sum_i (1 - W_{j(i)}) Y_i / \sum_i (1 - W_{j(i)})$. While either of these strategies is valid under further assumptions or in particular settings, below I argue in favor of a different strategy that is justified by the experimental design.

One particular choice of (weighted) control average is, however, justified by the experimental design of the ideal experiment considered in this paper:

$$\bar{Y}_c(d) = \frac{\sum_{i\in\mathbb{I}} \frac{1-W_{j(i)}}{1-\pi_{j(i)}} \pi_{j(i)} \sum_{s\in\mathbb{S}_{j(i)}} g_{j(i)}(s) \mathbb{1}\{|d(s, r_i) - d| \leq h\} Y_i}{\sum_{i\in\mathbb{I}} \frac{1-W_{j(i)}}{1-\pi_{j(i)}} \pi_{j(i)} \sum_{s\in\mathbb{S}_{j(i)}} g_{j(i)}(s) \mathbb{1}\{|d(s, r_i) - d| \leq h\}}.$$

Most importantly, the estimator $\bar{Y}_c(d)$ only averages over individuals who are at approximately distance d from a counterfactual candidate location $s \in \mathbb{S}_{j(i)}$. The remaining weighting is similar to inverse probability weighting estimators of the average effect of the treatment on the treated (ATT) in settings with individual-level treatments (cf. Imbens, 2004).

To see that this particular control average $\bar{Y}_c(d)$ provides the appropriate counterfactual for the simple average of the treated $\bar{Y}_t(d)$, consider the expected value of the terms in the numerator of the latter. It is straightforward to show that

$$\begin{aligned} &E\Big(W_{j(i)} \mathbb{1}\{|d(\xi_{j(i)}, r_i) - d| \leq h\} Y_i\Big) \\ =&\pi_{j(i)} \sum_{s\in\mathbb{S}_{j(i)}} g_{j(i)}(s) \mathbb{1}\{|d(s, r_i) - d| \leq h\} Y_i(s) \end{aligned}$$

see Appendix A.1 for the details. The difference between the expression above and the terms of the estimator $\bar{Y}_c(d)$ is that the latter can only average over individuals in control regions, with $W_{j(i)} = 0$, requiring the additional inverse probability weight $\frac{\pi_{j(i)}}{1-\pi_{j(i)}}$ in $\bar{Y}_c(d)$. The estimator $\bar{Y}_c(d)$ therefore aligns, in expectation, the weights placed on each *control* potential outcome $Y_i(0)$ with those placed on the corresponding *treated* potential outcome $Y_i(s)$ by $\bar{Y}_t(d)$.

Consequently, the estimator

$$\hat{\tau}(d) = \bar{Y}_t(d) - \bar{Y}_c(d) \tag{1}$$

estimates a weighted average of the differences $Y_i(s) - Y_i(0)$, which are the individual-level treatment effects $\tau_i(s)$ defined in Section 3 above. The particular inverse probability weights make $\hat{\tau}(d)$ an estimate of the average treatment effect on the treated at a distance of $d \pm h$. Theorem 1 states approximate finite sample properties of this estimator.

Theorem 1. *Under Assumptions 1 (no interference across regions) and 2 (completely randomized design with equal treatment probabilities across regions), the estimator $\hat{\tau}(d)$ has an approximate finite sample distribution over the assignment distribution with*

(i) unbiasedness for the ATT:

$$E(\hat{\tau}(d)) \approx \frac{\sum_{j=1}^{J} \sum_{s \in \mathbb{S}_j} \pi g_j(s) \sum_{i:\ |d(s,r_i)-d| \leq h} \tau_i(s)}{\sum_{j=1}^{J} \sum_{s \in \mathbb{S}_j} \pi g_j(s) \sum_{i:\ |d(s,r_i)-d| \leq h} 1} \tag{2}$$

(ii) variance:

$$\text{var}\Big(\hat{\tau}(d)\Big) \approx \frac{J-1}{J} \frac{\tilde{V}_t^{location}(d)}{J_t} + \frac{\tilde{V}_c^{region}(d)}{J_c} + \frac{1}{J} \frac{\tilde{V}_t^{region}(d)}{J_t} - \frac{\tilde{V}_{ct}^{region}(d)}{J}$$

where

$$\tilde{V}_t^{location}(d) \equiv \frac{\sum_{j=1}^{J} \sum_{s \in \mathbb{S}_j} g_j(s) \Big(\sum_{i:\ |d(s,r_i)-d| \leq h} (Y_i(s) - \mu_t(d)) \Big)^2}{\bar{n}(d)^2 (J-1)}$$

$$\tilde{V}_c^{region}(d) \equiv \frac{\sum_{j=1}^{J} \Big(\sum_{s \in \mathbb{S}_j} g_j(s) \sum_{i:\ |d(s,r_i)-d| \leq h} (Y_i(0) - \mu_c(d)) \Big)^2}{\bar{n}(d)^2 (J-1)}$$

$$\tilde{V}_t^{region}(d) \equiv \frac{\sum_{j=1}^{J} \Big(\sum_{s \in \mathbb{S}_j} g_j(s) \sum_{i:\ |d(s,r_i)-d| \leq h} (Y_i(s) - \mu_t(d)) \Big)^2}{\bar{n}(d)^2 (J-1)}$$

$$\tilde{V}_{ct}^{region}(d) \equiv \frac{\sum_{j=1}^{J} \Big(\sum_{s \in \mathbb{S}_j} g_j(s) \sum_{i:\ |d(s,r_i)-d| \leq h} \big(Y_i(s) - Y_i(0) - (\mu_t(d) - \mu_c(d))\big) \Big)^2}{\bar{n}(d)^2 (J-1)}$$

$$\bar{n}(d) \equiv \frac{1}{J} \sum_{j=1}^{J} \sum_{s \in \mathbb{S}_j} g_j(s) \sum_{i \in \mathbb{I}_j} \mathbb{1}\{|d(s,r_i) - d| \leq h\}$$

and $\mu_t(d)$ and $\mu_c(d)$ are the average treated and control potential outcomes with the same weights as the ATT estimand given in (i). $J_t \equiv \sum_{j=1}^{J} W_j$ and $J_c \equiv J - J_t$ are the numbers of treated and control regions, which are fixed under Assumption 2.

Proof: *The theorem is a special case of Theorem 3 below with weights*

$$w_i(s) = \pi_j g_j(s) \mathbb{1}\{|d(s,r_i) - d| \leq h\}.$$

Remark 1. The approximation in Theorem 1 arises because the denominators of the estimator $\hat{\tau}(d)$ are stochastic.[13] The proof proceeds by deriving the finite sample properties of an infeasible demeaned estimator $\tilde{\tau}(d)$ with non-stochastic denominators that satisfies $\hat{\tau}(d) - \tilde{\tau}(d) = o(J^{-1/2})$, where J is the number of regions; details are given in the appendix. Even with relatively few regions, the approximation is likely to perform well

[13] An alternative approach is to normalize the weights *in expectation* rather than in the sample to derive exact results (Aronow et al., 2020). But such estimators are less attractive and rarely used in practice in the well-studied setting of individual-level treatment assignment (Imbens, 2004; Busso et al., 2014).

in practice. Similar issues arise with individual-level treatments if treatment is decided by successive coin flips, rather than by fixing the number of treated. In spatial treatment settings, however, it is rarely feasible to hold the number of individuals near treatment fixed when randomizing the assignment of treatment locations. When all candidate locations have equal numbers of individuals in the distance bin, the approximations in the theorem above hold with equality.

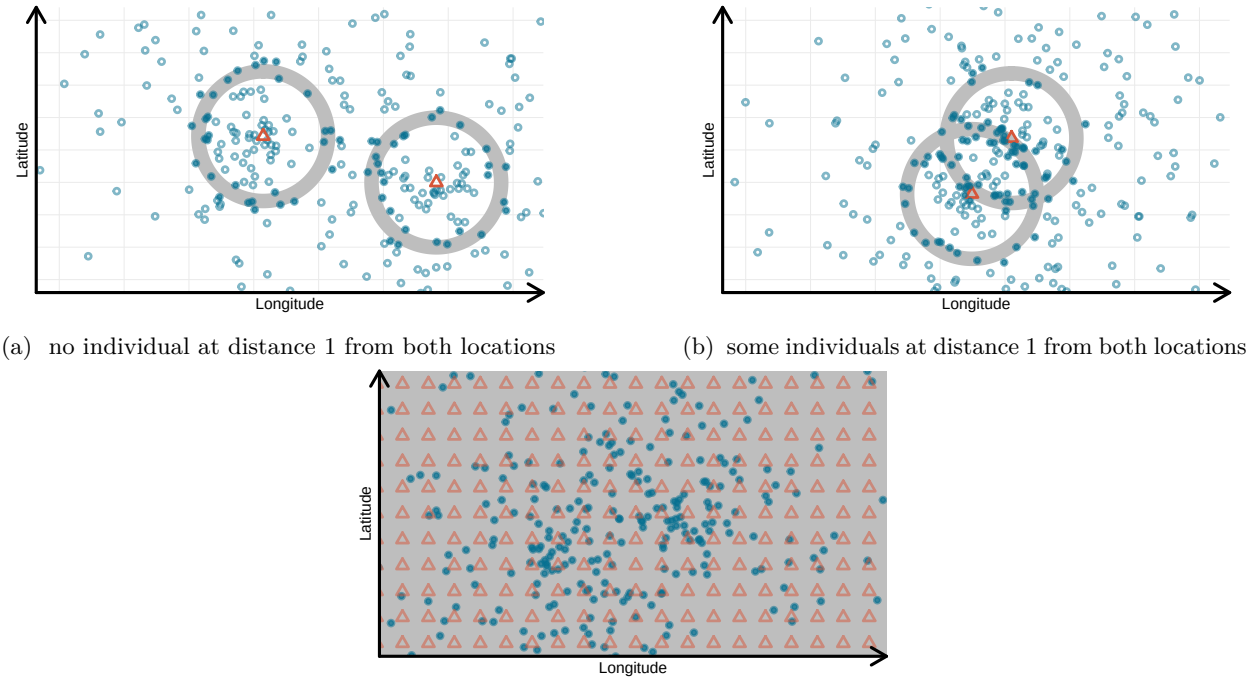
Remark 2. The expected value given by Theorem 1 is also relevant for other estimators using $\bar{Y}_t(d)$ as the mean of the treated but relying on a different control comparison group and auxiliary assumptions to justify the comparison. When researchers argue that randomization in the spatial locations for treatments allows them to estimate the treatment effect using $\bar{Y}_t(d)$ (or close analogs), they therefore implicitly estimate the average treatment effect on the treated. I believe there is value in making the estimation target explicit: As I argue in Section 4.1.2 below, the ATT as defined above does not necessarily allow the most meaningful comparisons of the effects at different distances. The estimation of the average treatment effect on the treated appears in the weights $\pi g_j(s)$ in the estimand, as well as the factors $g_j(s)$ in the variances. Since the region treatment probabilities π_j are constant under the assumptions of Theorem 1, they appear only through the number of treated and control regions, J_t and J_c.

Remark 3. The control average $\bar{Y}_c(d)$ used by the proposed estimator $\hat{\tau}(d)$ simplifies to the simple average over all individuals in control regions, $\sum_i (1 - W_{j(i)}) Y_i / \sum_i (1 - W_{j(i)})$ if each individual is equally likely to be distance d from realized treatment. This typically requires that treatment can be realized at *any* location within a region with equal probability ($g_j(s)$ is constant within j), and the probability that a region is selected for treatment (π_j) must be proportional to its area. Then the unconditional treatment probability $\pi_j g_j(s)$ is constant for all locations s, not just for a small, finite, set of candidate locations. Figure 3 illustrates and contrasts this with the more common setting where only a small number of candidate locations have positive probability of receiving treatment.

Remark 4. The variance given in the theorem is the design-based variance (Abadie et al., 2020) of the estimator. It expresses the variation in the estimate arising from assigning treatment randomly to one candidate location in a fixed number of randomly chosen regions. The individuals whose outcomes are measured are held fixed across these repeated samples; the only difference between samples lies in which potential outcome is observed for each individual. The thought experiment behind the variance above is therefore similar to performing a permutation, or placebo, test. Aronow et al. (2020) also suggest permutation tests as an alternative basis for inference in the spatial treatment setting.

Remark 5. In the variance, the terms $\tilde{V}_t^{\text{location}}(d)$, $\tilde{V}_c^{\text{region}}(d)$, $\tilde{V}_t^{\text{region}}(d)$ can loosely be interpreted as variances of potential outcomes across locations or regions, and $\tilde{V}_{ct}^{\text{region}}(d)$ can be interpreted as the variance of treatment effects across regions. $\bar{n}(d)$ is the average, across regions, expected number of treated individuals conditional on treatment somewhere in the region. If there is only one candidate treatment location per region, $\tilde{V}_t^{\text{region}}(d) = \tilde{V}_t^{\text{location}}(d)$, and combining these two terms cancels the factors involving J. The remaining three terms then resemble two marginal variances of potential outcomes and the variance of the treatment effect, as in the formula for the design-based variance in individual-level randomized experiments (cf. Imbens and Rubin, 2015, ch. 6).

Remark 6. When a region is treated, the researcher only observes the potential outcomes corresponding to treatment at one of the candidate locations in the region. In contrast, when a region is in the control group, the researcher can average over control potential outcomes for all candidate locations in the region. This distinction affects the variance of the estimator: In general, averaging within region across candidate



(a) no individual at distance 1 from both locations

(b) some individuals at distance 1 from both locations

(c) candidate treatment locations everywhere, all individuals at distance 1 from equally many locations

Figure 3: The figures show regions with individuals (small circles) and candidate treatment locations (triangles), highlighting areas that are distance 1 away from a candidate treatment location in gray. Suppose each candidate treatment location is equally likely to be realized. In panel a, all individuals who are distance 1 away from a candidate treatment location receive equal weight in the estimand $\tau^{ATT}(d)$. In estimation, if the region is in the control group, we take the simple average of outcomes of the highlighted individuals. In panel b, some individuals are distance 1 away from both candidate treatment locations, so these individuals receive greater weight in the estimand $\tau^{ATT}(d)$. In estimation, if the region is in the control group, we take the average of outcomes of the highlighted individuals, but individuals who are located in both gray rings receive twice the weight. In panel c, candidate treatment locations are *everywhere* (for illustration, only candidate treatment locations along a grid are displayed). If we assume that candidate treatment locations extend past the boundaries of the region, then all individuals in the region receive equal weight in the estimand $\tau^{ATT}(d)$. In estimation, if the region is in the control group, we take the simple average of outcomes of *all* individuals.

locations decreases the variance. Formally, $\tilde{V}_t^{\text{region}}(d) \leq \tilde{V}_t^{\text{location}}(d)$ by Jensen's inequality. That the variance formula places most weight on $\tilde{V}_t^{\text{location}}(d)$ rather than on $\tilde{V}_t^{\text{region}}(d)$ is the price to pay for only observing potential outcomes corresponding to a single candidate location in each treated region. However, if most of the variance in treated potential outcomes is across regions rather than within, $\tilde{V}_t^{\text{region}}(d) \approx \tilde{V}_t^{\text{location}}(d)$. Intuitively, if (individuals near) all candidate locations within a region are similar, little information is lost by only observing outcomes for one candidate location under treatment.

Estimation of Variance The variance in Theorem 1 depends on four variances: $\tilde{V}_t^{\text{location}}(d)$, $\tilde{V}_c^{\text{region}}(d)$, $\tilde{V}_t^{\text{region}}(d)$, and $\tilde{V}_{ct}^{\text{region}}(d)$. The first two variances are straightforward to estimate, as given below. The third variance, $\tilde{V}_t^{\text{region}}(d)$ cannot be estimated directly, but is bounded by $\tilde{V}_t^{\text{location}}(d)$. Alternatively, it can be approximated as discussed below. The fourth variance, the variance of treatment effects $\tilde{V}_{ct}^{\text{region}}(d)$, is

generally not identified,[14] but since it appears negatively in the overall variance, it can be dropped resulting in a conservative estimator of the variance (cf. Imbens and Rubin, 2015, ch. 6).

A natural estimator of $\tilde{V}_t^{\text{location}}(d)$ is

$$\hat{V}_t^{\text{location}}(d) \equiv \frac{\sum_{j=1}^{J} \sum_{s\in\mathbb{S}_j} W_j \mathbb{1}\{\xi_j = s\}\Big(\sum_{i:\ |d(s,r_i)-d|\leq h}(Y_i - \bar{Y}_t(d))\Big)^2}{(J_t-1)\Big(\frac{1}{J_t}\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\big(W_j\mathbb{1}\{\xi_j=s\}(\sum_{i:\ |d(s,r_i)-d|\leq h}1)\big)\Big)^2}$$

which takes the average squared difference from the mean over those individuals who are treated at distance d. Note that while one can calculate $\bar{n}(d)$ exactly, it is likely preferable in practice to use the average number of individuals near treated locations in the sample, which more accurately reflects the averaging in the numerator.

Similarly, a natural estimator of $\tilde{V}_c^{\text{region}}(d)$ is

$$\hat{V}_c^{\text{region}}(d) \equiv \frac{\sum_{j=1}^{J}(1-W_j)\Big(\sum_{s\in\mathbb{S}_j} g_j(s)\sum_{i:\ |d(s,r_i)-d|\leq h}(Y_i - \bar{Y}_c(d))\Big)^2}{(J_c-1)\Big(\frac{1}{J_c}\sum_{j=1}^{J}\big((1-W_j)\sum_{s\in\mathbb{S}_j} g_j(s)\sum_{i:\ |d(s,r_i)-d|\leq h}1\big)\Big)^2}.$$

Using $\hat{V}_t^{\text{location}}(d)$ as a conservative estimator of $\tilde{V}_t^{\text{region}}(d)$, the conservative estimator for the variance of the estimator $\hat{\tau}(d)$ is

$$\widehat{\text{var}}_{\text{conservative}}(\hat{\tau}(d)) \equiv \frac{\hat{V}_t^{\text{location}}(d)}{J_t} + \frac{\hat{V}_c^{\text{region}}(d)}{J_c}$$

If there is reason to believe that there is substantial variance within regions (rather than across regions), it may be preferable to approximate $\tilde{V}_t^{\text{region}}(d)$ directly rather than estimate it conservatively with $\hat{V}_t^{\text{location}}(d)$. Specifically, consider forming the estimator

$$\hat{V}_c^{\text{location}}(d) \equiv \frac{\sum_{j=1}^{J}(1-W_j)\sum_{s\in\mathbb{S}_j} g_j(s)\Big(\sum_{i:\ |d(s,r_i)-d|\leq h}(Y_i - \bar{Y}_c(d))\Big)^2}{(J_c-1)\Big(\frac{1}{J_c}\sum_{j=1}^{J}(1-W_j)\sum_{s\in\mathbb{S}_j} g_j(s)(\sum_{i:\ |d(s,r_i)-d|\leq h}1)\Big)^2}$$

which is analogous to $\hat{V}_t^{\text{location}}(d)$ but for control, rather than treated, regions. Under some assumptions, for instance constant additive or constant multiplicative treatment effects,

$$\frac{\tilde{V}_t^{\text{region}}(d)}{\tilde{V}_t^{\text{location}}(d)} = \frac{\tilde{V}_c^{\text{region}}(d)}{\tilde{V}_c^{\text{location}}(d)}$$

where $\tilde{V}_c^{\text{location}}(d)$ is the appropriate population analogue, such that a plausible estimator for $\tilde{V}_t^{\text{region}}(d)$ is

$$\hat{V}_t^{\text{region}}(d) = \hat{V}_t^{\text{location}}(d)\frac{\hat{V}_c^{\text{region}}(d)}{\hat{V}_c^{\text{location}}(d)}$$

This estimator uses that the ratio of within-region and across-region variances of treated and control potential outcomes are approximately similar in most settings where the effect of the treatment (and its

[14] For the variance of treatment effects to be zero (constant treatment effects), the distributions of treated and control must be identical up to a location shift. More generally, the covariance of treated and control potential outcomes is partially identified from the marginal variances. Heckman et al. (1997) use the Fréchet-Hoeffding inequality to form bounds on the variance of treatment effects. Aronow et al. (2014) use the same bounds to improve the Neyman (1923, 1990, cf. Imbens and Rubin (2015)) variance estimator.

heterogeneity) is small relative to other variance in outcomes. When the equality of ratios is not exact, deviations can lead to an either conservative or anti-conservative estimate of $\tilde{V}_t^{\text{region}}(d)$. In practice, even if the estimator $\hat{V}_t^{\text{region}}(d)$ is not conservative, the variance estimator

$$\widehat{\text{var}}_{\text{approx}}(\hat{\tau}(d)) \equiv \frac{J-1}{J}\frac{\hat{V}_t^{\text{location}}(d)}{J_t} + \frac{\hat{V}_c^{\text{region}}(d)}{J_c} + \frac{1}{J}\frac{\hat{V}_t^{\text{region}}(d)}{J_t}$$

likely is still conservative for $\text{var}(\hat{\tau}(d))$ by omission of the variance of treatment effects term.

Comparison of the conservative variance estimate, $\widehat{\text{var}}_{\text{conservative}}(\hat{\tau}(d))$, and the variance estimate using the approximation, $\widehat{\text{var}}_{\text{approx}}(\hat{\tau}(d))$, can serve as a plausible benchmark for the benefits any refinements of estimators of $\tilde{V}_t^{\text{region}}(d)$ can plausibly yield. In practice, since $\tilde{V}_t^{\text{region}}(d)$ receives weight $\frac{1}{J}$ relative to the other variances, the difference is likely to be small.

4.1.2 Weighting Considerations for the ATT

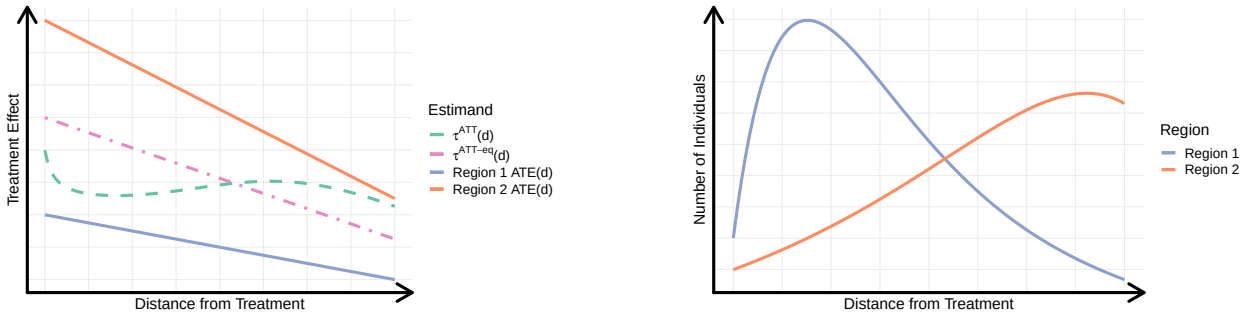
The estimand $\tau(d)$ is not generally the most informative when the researcher is interested in how the effect of the treatment changes with distance from treatment. As an alternative, I propose the estimand $\tau^{ATT-eq}(d)$ with a more attractive interpretation when comparing effects at different distances.[15] The estimand $\tau^{ATT-eq}(d)$ can additionally be interpreted as the expected average effect at distance d of a new treatment location.

Figure 4 illustrates the problem of interpreting the difference between the estimands $\tau(d)$ and $\tau(d')$ as the pattern of treatment effects across distances from treatment. Suppose the researcher is interested in comparing the average treatment effect at a short distance $d = d_{\text{short}}$ and long distance $d' = d_{\text{long}}$. Suppose further that there are two types of candidate treatment locations, each type equally likely to be realized. The first type of candidate treatment locations has many individuals located at the short distance and few individuals at the long distance. These first candidate locations all have relatively small treatment effect at both distances, but decreasing in distance from treatment. The second type of candidate treatment locations has few individuals located at the short distance, and many individuals at the long distance. These second candidate locations all have relatively large treatment effect at both distances, but also decreasing in distance from treatment.

In the example in Figure 4, the estimand $\tau(d)$ is *increasing* in distance d even though the treatment effect of any single treatment location is *decreasing* in distance from treatment. The estimand $\tau(d_{\text{short}})$ places most weight on the first type of candidate locations because most individuals at the short distance from treatment are near this type of location. In contrast, the estimand $\tau(d_{\text{long}})$ places most weight on the second type of candidate locations. Hence, $\tau(d_{\text{long}}) > \tau(d_{\text{short}})$. This inequality states that the average treatment effect at a long distance for the average individual at the long distance from a candidate treatment location is larger than the average effect at a short distance for the average individual at the short distance from candidate treatment locations. It does not imply that the average effect of any single treatment location is increasing in distance from treatment. Instead, the average individual at a long distance may simply be both a different type of individual (in terms of observables and unobservables) and also be exposed to a different treatment location on average.

An alternative estimand, $\tau^{ATT-eq}(d)$ defined below and also shown in Figure 4, avoids such issues in interpretation by placing the same aggregate weight on each candidate treatment location irrespective of the

[15] Aronow et al. (2020) focus on the estimator $\tau^{ATT-eq}(d)$, but do not discuss its interpretation and differences from alternative estimators in detail.



(a) Estimands

(b) Individuals per distance in each region

Figure 4: The estimands $\tau(d)$ and $\tau^{ATT-eq}(d)$ can be substantially different. Panel a shows the decay of average treatment effects over distance for two regions with solid lines. The dashed line shows the estimand $\tau(d)$, which weights by the relative number of individuals at distance d, as given in panel b. The dot-dashed line shows the estimand $\tau^{ATT-eq}(d)$, which weights the regions equally if both receive treatment with equal probability. Existing empirical methods implicitly target $\tau(d)$.

distance d. The estimand $\tau^{ATT-eq}(d)$ first separately averages the potential outcomes of nearby individuals for each candidate treatment location. These averages are then averaged again, with weights proportional only to the probability of treatment at the location. In contrast, the estimand $\tau(d)$ uses weights proportional to the product of the treatment probability and the number of individuals near the treatment location. Formally,

$$\begin{aligned} \tau^{ATT-eq}(d) &= \frac{\sum_{j=1}^{J} \sum_{s \in \mathbb{S}_j} \pi_j g_j(s) \bar{\tau}(s,d)}{\sum_{j=1}^{J} \sum_{s \in \mathbb{S}_j} \pi_j g_j(s)} \\ \bar{\tau}(s,d) &= \frac{\sum_{i:\ |d(s,r_i)-d| \leq h} \tau_i(s)}{\sum_{i:\ |d(s,r_i)-d| \leq h} 1} \end{aligned} \tag{3}$$

where $\bar{\tau}(s,d)$ is the average effect of a given candidate location s on individuals at distance $d \pm h$ from it. These average effects are then averaged with weights $\pi_j g_j(s)$, which do not depend on distance from treatment. Hence, the weight placed on the average effect of a given location s does not depend on the distance from treatment. To also non-parametrically control for observable differences in pre-treatment variables X_i, one can estimate $\tau^{ATT-eq}(d)$ separately using only individuals with covariate values falling into groups defined by X_i. The comparison of $\tau^{ATT-eq}(d)$ and $\tau^{ATT-eq}(d')$ then compares individuals with the same average exposure to the candidate locations and similar individual (observable) characteristics.

Holding the aggregate weight per treatment location constant across distance from treatment is attractive when the treatment effects are expected to be heterogeneous by region or location. Such heterogeneity is particularly plausible in many spatial treatment settings: Oftentimes, the exact implementation of the treatment differs substantially from location to location. For instance, the million dollar plants in the study of Greenstone and Moretti (2003) are each operated by distinct companies which may differ in their labor demand and wage setting. Hence, heterogeneous treatment effects arise not only due to differences between individuals, but also due to differences in the implementation of the treatments. Since spatial treatments are often larger, rarer, and more complex, their implementation tends to vary more than, say, the administration of a drug in medical trials to different patients, or the content of a job training program across training sites or cohorts.

Additionally, the estimand $\tau^{ATT-eq}(d)$ has an attractive interpretation as the expected effect at distance d of a new treatment location. Consider the following hierarchical model. First, when treatment is realized

at location s, its average effect at distance d is drawn as $\bar{\tau}_s(d) \sim F_d$. Second, the individual-level effect of location s on individual i is given by

$$\tau_i(s) = \tau_s(d) + \epsilon_i(s)$$

where $\epsilon_i(s)$ is a mean zero individual-specific component. Then, as the width of the distance bin, $2h$, goes to 0, the estimand $\tau^{ATT-eq}(d)$ approaches the mean of the distribution F_d. Hence, one can interpret $\tau^{ATT-eq}(d)$ as the expected average individual-level treatment effect of a new treatment location drawn in the same way as existing realized treatment locations.

I propose the following estimator to estimate $\tau^{ATT-eq}(d)$:

$$\begin{aligned}
\hat{\tau}^{ATT-eq}(d) &= \bar{Y}_t^{ATT-eq}(d) - \bar{Y}_c^{ATT-eq}(d) \\
\bar{Y}_t^{ATT-eq}(d) &= \frac{\sum_{j=1}^{J} W_j \frac{\sum_{i:\ |d(s,r_i)-d|\leq h} Y_i}{\sum_{i:\ |d(s,r_i)-d|\leq h} 1}}{\sum_{j=1}^{J} W_j} \\
\bar{Y}_c^{ATT-eq}(d) &= \frac{\sum_{j=1}^{J}(1-W_j)\sum_{s\in\mathbb{S}_j} \frac{\pi_j g_j(s)}{1-\pi_j} \frac{\sum_{i:\ |d(s,r_i)-d|\leq h} Y_i}{\sum_{i:\ |d(s,r_i)-d|\leq h} 1}}{\sum_{j=1}^{J}(1-W_j)\sum_{s\in\mathbb{S}_j} \frac{\pi_j g_j(s)}{1-\pi_j}}
\end{aligned} \tag{4}$$

Theorem 2 gives the exact properties of the finite sample distribution of $\tau^{ATT-eq}(d)$.

Theorem 2. *Under assumptions 1 (no interference across regions) and 2 (completely randomized design), the estimator $\hat{\tau}^{ATT-eq}(d)$ has an approximate finite sample distribution over the assignment distribution with*

(i) unbiasedness: $E(\hat{\tau}^{ATT-eq}(d)) = \tau^{ATT-eq}(d)$

(ii) variance:

$$\begin{aligned}
\text{var}\Big(\hat{\tau}^{ATT-eq}(d)\Big) = \frac{J-1}{J}\frac{\tilde{V}_t^{ATT-eq,location}(d)}{J_t} + \frac{\tilde{V}_c^{ATT-eq,region}(d)}{J_c} \\
+ \frac{1}{J}\frac{\tilde{V}_t^{ATT-eq,region}(d)}{J_t} - \frac{\tilde{V}_{ct}^{ATT-eq,region}(d)}{J}
\end{aligned}$$

where

$$\begin{aligned}
\tilde{V}_t^{ATT-eq,location}(d) &\equiv \frac{\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j} g_j(s)\Big(\frac{\sum_{i:\ |d(s,r_i)-d|\leq h}(Y_i(s)-\mu_t^{ATT-eq}(d))}{\sum_{i:\ |d(s,r_i)-d|\leq h} 1}\Big)^2}{J-1} \\
\tilde{V}_c^{ATT-eq,region}(d) &\equiv \frac{\sum_{j=1}^{J}\Big(\sum_{s\in\mathbb{S}_j} g_j(s)\frac{\sum_{i:\ |d(s,r_i)-d|\leq h}(Y_i(0)-\mu_c^{ATT-eq}(d))}{\sum_{i:\ |d(s,r_i)-d|\leq h} 1}\Big)^2}{J-1} \\
\tilde{V}_t^{ATT-eq,region}(d) &\equiv \frac{\sum_{j=1}^{J}\Big(\sum_{s\in\mathbb{S}_j} g_j(s)\frac{\sum_{i:\ |d(s,r_i)-d|\leq h}(Y_i(s)-\mu_t^{ATT-eq}(d))}{\sum_{i:\ |d(s,r_i)-d|\leq h} 1}\Big)^2}{J-1} \\
\tilde{V}_{ct}^{ATT-eq,region}(d) &\equiv \frac{\sum_{j=1}^{J}\Big(\sum_{s\in\mathbb{S}_j} g_j(s)\frac{\sum_{i:\ |d(s,r_i)-d|\leq h}\big(Y_i(s)-Y_i(0)-\tau^{ATT-eq}(d)\big)}{\sum_{i:\ |d(s,r_i)-d|\leq h} 1}\Big)^2}{J-1}
\end{aligned}$$

and $\mu_t^{ATT-eq}(d)$ and $\mu_c^{ATT-eq}(d)$ are the average treated and control potential outcomes with the same

weights as the ATT-eq estimand given in Equation 3; specifically potential outcome $Y_i(s)$ receives weight

$$w_i(s,d) = \pi_j g_j(s) \frac{\mathbb{1}\{|d(s,r_i) - d| \leq h\}}{\sum_{i':\ |d(s,r_{i'})-d|\leq h} 1}.$$

$J_t \equiv \sum_{j=1}^{J} W_j$ and $J_c \equiv J - J_t$ are the numbers of treated and control regions, which are fixed under Assumption 2.

Proof: *The theorem is a special case of Theorem 3 below with weight as specified in (ii).*

Remark 7. The estimator here is exactly unbiased because under a completely randomized design, the sum of weights is constant (and equal to the number of treated regions) across assignment realizations. This is different from Theorem 1 above, where the number of treated individuals varies. Here, treated individuals are averaged by candidate treatment location, and the number of treated locations is constant across assignment realizations by Assumption 2.

Remark 8. It is straightforward to estimate the variance in Theorem 2. Specifically, I propose

$$\hat{V}_t^{ATT-eq,\text{location}}(d) \equiv \frac{\sum_{j=1}^{J} \sum_{s\in\mathbb{S}_j} W_j \mathbb{1}\{\xi_j = s\} \Big(\frac{\sum_{i:\ |d(s,r_i)-d|\leq h} (Y_i(s) - \mu_t^{ATT-eq}(d))}{\sum_{i:\ |d(s,r_i)-d|\leq h} 1} \Big)^2}{J_t - 1}$$

$$\hat{V}_c^{ATT-eq,\text{region}}(d) \equiv \frac{\sum_{j=1}^{J} (1 - W_j) \Big(\sum_{s\in\mathbb{S}_j} g_j(s) \frac{\sum_{i:\ |d(s,r_i)-d|\leq h} (Y_i(0) - \mu_c^{ATT-eq}(d))}{\sum_{i:\ |d(s,r_i)-d|\leq h} 1} \Big)^2}{J_c - 1}$$

as well as bounding $\tilde{V}_t^{ATT-eq,\text{region}}(d)$ by $\tilde{V}_t^{ATT-eq,\text{location}}(d)$ (by Jensen's inequality) and dropping the variance of treatment effects term. The estimator of the variance is then

$$\widehat{\text{var}}(\hat{\tau}^{ATT-eq}(d)) = \frac{\hat{V}_t^{ATT-eq,\text{location}}(d)}{J_t} + \frac{\hat{V}_c^{ATT-eq,\text{location}}(d)}{J_c}.$$

An alternative variance estimator that approximates rather than bounds $\tilde{V}_t^{ATT-eq,\text{region}}(d)$ is also possible along the lines of the discussion of the previous section.

While Theorem 2 gives moments of the exact finite-sample distribution of $\hat{\tau}^{ATT-eq}(d)$ at a given distance d, researchers sometimes wish to compare the estimated effects at different distances d and d'. Lemma 1 provides the variance of $\hat{\tau}^{ATT-eq}(d) - \hat{\tau}^{ATT-eq}(d')$. Alternative comparisons of effects at different distances can also be based on part (iii) of Theorem 3 below.

Corollary 1. *Suppose there is at least one individual at distance d and at least one individual at distance d' for every candidate treatment location. Then, under assumptions 1 (no interference across regions) and 2 (completely randomized design), the difference $\Delta\hat{\tau}^{ATT-eq}(d,d') \equiv \hat{\tau}^{ATT-eq}(d) - \hat{\tau}^{ATT-eq}(d')$ has exact finite sample variance across the assignment distribution*

$$\text{var}\Big(\Delta\hat{\tau}^{ATT-eq}(d,d')\Big) = \frac{J-1}{J} \frac{\tilde{V}_{t,diff}^{location}(d,d')}{J_t} + \frac{\tilde{V}_{c,diff}^{region}(d,d')}{J_c} + \frac{1}{J} \frac{\tilde{V}_{t,diff}^{region}(d,d')}{J_t} - \frac{\tilde{V}_{ct,diff}^{region}(d,d')}{J}$$

where

$$\tilde{V}^{location}_{t,diff}(d,d') \equiv \frac{\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j} g_j(s)\Big(\frac{\sum_{i:\ |d(s,r_i)-d|\leq h}(Y_i(s)-\mu_t^{ATT-eq}(d))}{\sum_{i:\ |d(s,r_i)-d|\leq h}1} - \frac{\sum_{i:\ |d(s,r_i)-d'|\leq h}(Y_i(s)-\mu_t^{ATT-eq}(d'))}{\sum_{i:\ |d(s,r_i)-d'|\leq h}1}\Big)^2}{J-1}$$

$$\tilde{V}^{region}_{c,diff}(d,d') \equiv \frac{\sum_{j=1}^{J}\Big(\sum_{s\in\mathbb{S}_j} g_j(s)\big(\frac{\sum_{i:\ |d(s,r_i)-d|\leq h}(Y_i(0)-\mu_c^{ATT-eq}(d))}{\sum_{i:\ |d(s,r_i)-d|\leq h}1} - \frac{\sum_{i:\ |d(s,r_i)-d'|\leq h}(Y_i(0)-\mu_c^{ATT-eq}(d'))}{\sum_{i:\ |d(s,r_i)-d'|\leq h}1}\big)\Big)^2}{J-1}$$

$$\tilde{V}^{region}_{t,diff}(d,d') \equiv \frac{\sum_{j=1}^{J}\Big(\sum_{s\in\mathbb{S}_j} g_j(s)\big(\frac{\sum_{i:\ |d(s,r_i)-d|\leq h}(Y_i(s)-\mu_t^{ATT-eq}(d))}{\sum_{i:\ |d(s,r_i)-d|\leq h}1} - \frac{\sum_{i:\ |d(s,r_i)-d'|\leq h}(Y_i(s)-\mu_t^{ATT-eq}(d'))}{\sum_{i:\ |d(s,r_i)-d'|\leq h}1}\big)\Big)^2}{J-1}$$

$$\tilde{V}^{region}_{ct,diff}(d,d') \equiv \frac{1}{J-1}\sum_{j=1}^{J}\Big(\sum_{s\in\mathbb{S}_j} g_j(s)\big(\frac{\sum_{i:\ |d(s,r_i)-d|\leq h}\big(Y_i(s)-Y_i(0)-\tau^{ATT-eq}(d)\big)}{\sum_{i:\ |d(s,r_i)-d|\leq h}1} - \frac{\sum_{i:\ |d(s,r_i)-d'|\leq h}\big(Y_i(s)-Y_i(0)-\tau^{ATT-eq}(d')\big)}{\sum_{i:\ |d(s,r_i)-d'|\leq h}1}\big)\Big)^2$$

Proof: *For each candidate treatment location s, define an "individual" $i(s)$ with outcome*

$$Y_{\tilde{i}(s)} = \left(\frac{\sum_{i:\ |d(s,r_{i'})-d|\leq h} Y_i}{\sum_{i:\ |d(s,r_{i'})-d|\leq h}1} - \frac{\sum_{i:\ |d(s,r_{i'})-d'|\leq h} Y_i}{\sum_{i:\ |d(s,r_{i'})-d'|\leq h}1}\right).$$

Then $\Delta\hat{\tau}^{ATT-eq}(d,d')$ is the ATT with equally weighted regions of individuals $i(s)$ instead of i; that is, for $\mathbb{I}_j = \{i(s) \mid s \in \mathbb{S}_j\}$. Hence Theorem 2 can be applied to obtain the variance.

Remark 9. Applying Theorem 2 with a redefined set of individuals is possible because the estimator $\hat{\tau}^{ATT-eq}(d)$ ensures that the aggregate weight on a location is constant across distances. This allows differencing distances within candidate locations before averaging. For other estimators, the weight normalization differs by distance, such that the difference of averages does not generally equal the average difference.

Remark 10. The condition that for each candidate treatment location and either distance there is at least one individual is required both for the validity of the variance expression and for the interpretation of the difference in effects. Without the condition, the weight normalization differs between distances, such that the difference of averages does not equal the average of the difference. The condition also facilitates interpretation. Suppose the average effect at distance d averages over a treatment location s that is not averaged over at distance d'. Then a difference in estimates could either be due to an actual difference in effects at the two distances, or due to effect heterogeneity, with the inclusion or omission of treatment location s affecting the average effect. I therefore recommend restricting attention to those candidate locations that have at least one individual at both distances.

4.1.3 General Weighted Average Treatment Effects

More generally, the same ideas allow estimation of any weighted average of individual-level treatment effects that places non-zero weights only on the effects of candidate treatment locations with positive probability of realization. Write these estimands of individual-level average effects of the spatial treatment on individuals *at*

a distance d from treatment as:

$$\tau_w(d) = \frac{1}{\sum_j \sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w_i(s,d)} \sum_j \sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w_i(s,d)\tau_i(s) \tag{5}$$

where $\tau_i(s)$ is the effect of treatment at location s on individual i, and $w_i(s,d)$ are *weights specified by the researcher.* The estimand here therefore can be any weighted average of the effect of single treatments on individuals with weights w as specified by the researcher. For the average effect of the treatment at distance d, weights $w_i(s,d)$ are non-zero only when location s and individual i are (approximately) distance d apart. While I define the ATT estimands $\tau(d)$ and $\tau^{ATT-eq}(d)$ for distance bins $d \pm h$ above by using the rectangular, or uniform, kernel function $\mathbb{1}\{|d(s,r_i) - d| \leq h\}$, the weights $w_i(s,d)$ can generally use any kernel function $k\Big(\frac{d(s,r_i)-d}{h}\Big)$ in place of distance bins to estimate the effects at distance d.

The average effect of the treatment on the treated estimands in Equations 2 and 3 are special cases of the estimand in Equation 5. For the estimand corresponding to $\hat{\tau}(d)$, choose weights

$$w_i(s,d) = \pi_{j(i)} g_{j(i)}(s) \mathbb{1}\{|d(s, r_i - d| \leq h\}.$$

For the estimand corresponding to $\hat{\tau}^{ATT-eq}(d)$, choose weights

$$w_i(s,d) = \pi_{j(i)} g_{j(i)}(s) \frac{\mathbb{1}\{|d(s, r_i - d| \leq h\}}{\sum_{i'\in\mathbb{I}_{j(i)}} \mathbb{1}\{|d(s, r_{i'} - d| \leq h\}}.$$

I propose an inverse probability weighting estimator (cf. Imbens, 2004) to estimate the weighted average treatment effect in Equation 5 above. In short, the estimator is the difference between weighted average outcomes of individuals near realized treatment locations and weighted average outcomes of individuals in regions without treatments. The weights here need to account for two aspects: First, the researcher specifies the desired weights $w_i(s,d)$ in the estimand. Second, individuals near locations with high treatment probability are relatively more likely, across samples of repeated treatment assignment, to appear in the sum of "treated" individuals than in the sum of "control" individuals due to the experimental design. The estimator cancels out the probability weighting induced by averaging over individuals (not) near realized treatment for the treated (control) average.

The proposed estimator for the weighted average treatment effect in 5 is

$$\begin{aligned}
\hat{\tau}_w(d) &\equiv \frac{\sum_i \sum_{s\in\mathbb{S}_{j(i)}} \iota_i^t(s) w_i(s,d) Y_i}{\sum_i \sum_{s\in\mathbb{S}_{j(i)}} \iota_i^t(s) w_i(s,d)} - \frac{\sum_i \sum_{s\in\mathbb{S}_{j(i)}} \iota_i^c w_i(s,d) Y_i}{\sum_i \sum_{s\in\mathbb{S}_{j(i)}} \iota_i^c w_i(s,d)} \\
\iota_i^t(s) &\equiv \frac{W_{j(i)} \mathbb{1}\{\xi_{j(i)} = s\}}{\pi_{j(i)} g_{j(i)}(s)} \\
\iota_i^c &\equiv \frac{1 - W_{j(i)}}{1 - \pi_{j(i)}}
\end{aligned} \tag{6}$$

The weights $w_i(s,d)$ are fixed and specified by the researcher. The weights $\iota_i^t(s)$ and ι_i^c, in contrast, are stochastic due to their dependence on the treatment assignment random variables W and ξ. Specifically, $\iota_i^t(s) = 0$ unless treatment is realized at location s, in which case it is equal to the inverse of the probability of this event. Similarly, $\iota_i^c = 0$ unless there is no treatment in the region of individual i, in which case it is equal to the inverse of the probability of no treatment in the region. Consequently, the stochastic weights are equal to 1 in expectation. The estimator divides each term by the sum of realized weights, $\iota_i^t(s) w_i(s,d)$

and $\iota_i^c w_i(s,d)$, such that it is the difference between a convex combination of treated outcomes and a convex combination of control outcomes as long as the weights w are non-negative.

For stating the approximate finite sample distribution of the estimator $\hat{\tau}_w(d)$ in Theorem 3 below, it is helpful to define some variances and (effective) sample sizes:

$$\tilde{V}^{\text{location}}_{w,t}(d) \equiv \frac{\sum_{j=1}^J \sum_{s\in\mathbb{S}_j} \frac{1}{\pi_j g_j(s)} \Big(\sum_{i\in\mathbb{I}_j} w_i(s,d)(Y_i(s) - \mu_{w,t}(d))\Big)^2}{\frac{J-1}{J}\sum_{j=1}^J \sum_{s\in\mathbb{S}_j} \frac{1}{\pi_j g_j(s)}}$$

$$\tilde{V}^{\text{region}}_{w,c}(d) \equiv \frac{\sum_{j=1}^J \frac{1}{1-\pi_j} \Big(\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j} w_i(s,d)(Y_i(0) - \mu_{w,c}(d))\Big)^2}{\frac{J-1}{J}\sum_{j=1}^J \frac{1}{1-\pi_j}}$$

$$\tilde{V}^{\text{region}}_{w,t}(d) \equiv \frac{\sum_{j=1}^J \Big(\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j} w_i(s,d)(Y_i(s) - \mu_{w,t}(d))\Big)^2}{J-1}$$

$$\tilde{V}^{\text{region}}_{w,ct}(d) \equiv \frac{\sum_{j=1}^J \Big(\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j} w_i(s,d)\big(Y_i(s) - Y_i(0) - (\mu_{w,t}(d) - \mu_{w,c}(d))\big)\Big)^2}{J-1}$$

which closely resemble variances of marginal potential outcomes aggregated at the candidate location or region level, and a treatment effect variance. The effective sample sizes are

$$\begin{aligned}
N_w(d) &\equiv \sum_{j=1}^J \sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w_i(s,d) \\
\bar{n}_w(d) &\equiv \frac{N_w(d)}{J-1} \\
\tilde{n}_{w,t}(d) &\equiv \Big(\frac{1}{\sum_{j=1}^J \sum_{s\in\mathbb{S}_j} \frac{1}{\pi_j g_j(s)}} N_w(d)\Big)\bar{n}_w(d) \\
\tilde{n}_{w,c}(d) &\equiv \Big(\frac{1}{\sum_{j=1}^J \frac{1}{1-\pi_j}} N_w(d)\Big)\bar{n}_w(d)
\end{aligned}$$

such that $N_w(d)$ is the total weighted number of individuals near any candidate location, and $\bar{n}_w(d)$ is the average number of individuals near candidate locations per region. The expected effective number of individuals near treated and control candidate locations ($\tilde{n}_{w,t}(d)$ and $\tilde{n}_{w,c}(d)$) accounts for the non-linear effects of heterogeneous treatment probabilities and numbers of individuals per candidate location. In standard experiments, designs where some individuals are treated with high probability and others with low probability typically have larger variance than designs with a constant treatment probability (and the same average number of treated). The effective sample sizes here translate this loss in precision into the variance-equivalent loss in sample size.

Theorem 3. *Under Assumption 1 (no interference across regions) and either Assumption 2 (completely randomized design) or Assumption 3 (Bernoulli trials), the estimator $\hat{\tau}_w(d)$ has an approximate finite sample distribution over the assignment distribution with*

(i) unbiasedness: $E(\hat{\tau}_w(d)) \approx \tau_w(d)$

(ii) variance:

$$\begin{aligned}\operatorname{var}\Big(\hat{\tau}_w(d)\Big) \approx \frac{\tilde{V}^{location}_{w,t}(d)}{\tilde{n}_{w,t}(d)J} + \big(1+\frac{\mathcal{C}}{J-1}\big)\frac{\tilde{V}^{region}_{w,c}(d)}{\tilde{n}_{w,c}(d)J} \\ + \big(\frac{\mathcal{C}}{J-1}\big)\frac{\tilde{V}^{region}_{w,t}(d)}{\bar{n}_w(d)(\pi N_w(d))} - \big(1+\frac{\mathcal{C}}{J-1}\big)\frac{\tilde{V}^{region}_{w,ct}(d)}{\bar{n}_w(d)N_w(d)}\end{aligned}$$

where $\mathcal{C}$ is an indicator for a completely randomized (rather than Bernoulli) design

$$\mathcal{C} \equiv \begin{cases} 1 & \textit{under Assumption 2} \\ 0 & \textit{under Assumption 3} \end{cases}$$

and variances $\tilde{V}$ and sample sizes $\tilde{n}$, $\bar{n}$, and N are defined as above the theorem.

(iii) covariance:

$$\begin{aligned}\operatorname{cov}\Big(\tilde{\tau}_w(d), \tilde{\tau}_{w'}(d')\Big) = \frac{\tilde{V}^{location}_{w,w',t}(d,d')}{\sqrt{\tilde{n}_{w,t}(d)}\cdot\sqrt{\tilde{n}_{w',t}(d')}\cdot J} \\ + (1+\frac{\mathcal{C}}{J-1})\frac{\tilde{V}^{region}_{w,w',c}(d,d')}{\sqrt{\tilde{n}_{w,c}(d)}\cdot\sqrt{\tilde{n}_{w',c}(d')}\cdot J} \\ + \frac{\mathcal{C}}{J-1}\frac{\tilde{V}^{region}_{w,w',t}(d,d')}{\pi\cdot\sqrt{\bar{n}_w(d)}\cdot\sqrt{\bar{n}_{w'}(d')}\cdot\sqrt{N_w(d)}\cdot\sqrt{N_{w'}(d')}} \\ - (1+\frac{\mathcal{C}}{J-1})\frac{\tilde{V}^{region}_{w,w',ct}(d,d')}{\sqrt{\bar{n}_w(d)}\cdot\sqrt{\bar{n}_{w'}(d')}\cdot\sqrt{N_w(d)}\cdot\sqrt{N_{w'}(d')}}\end{aligned}$$

with $\tilde{V}^{location}_{w,w',t}(d,d')$, $\tilde{V}^{region}_{w,w',c}(d,d')$, $\tilde{V}^{region}_{w,w',t}(d,d')$, $\tilde{V}^{region}_{w,w',t}(d,d')$, and $\tilde{V}^{region}_{w,w',ct}(d,d')$ defined analogously to the variances, with products of different distances (and weights) instead of squares of single distances; see Appendix A.2.7 for explicit definitions.

Proof: *See Appendix A.2.*

Remark 11. In a completely randomized design ($\mathcal{C}=1$, Assumption 2) with only one candidate location per region and only one individual at distance d from each candidate location, the variance simplifies to the well-known variance of the difference in means for randomized experiments (cf. Imbens and Rubin, 2015, ch. 6). Specifically, $\tilde{V}^{\text{location}}_{w,t}(d)$ and $\tilde{V}^{\text{region}}_{w,c}(d)$ are the marginal variances of the treated and control, respectively; $\tilde{V}^{\text{region}}_{w,ct}(d)$ is the variance of treatment effects; $\tilde{V}^{\text{region}}_{w,t}(d) = \tilde{V}^{\text{location}}_{w,t}(d)$ and combines with the first term. The factors $(1+\frac{\mathcal{C}}{J-1})$ cancel corresponding terms in for instance $\tilde{n}_{w,t}(d)$, such that $(1+\frac{\mathcal{C}}{J-1})\frac{1}{\tilde{n}_{w,t}(d)J}$ exactly yields the inverse of the number of treated observations.

Remark 12. Compared to the variances of Theorems 1 and 2, the formula for the variance in Theorem 3 appears more involved in part because Assumption 3 (Bernoulli trials) allows the region treatment probability π_j to vary across regions. Similar to Theorems 1 and 2, a conservative estimator of the variance in Theorem 3 is possible that drops the variance of treatment effects term $\tilde{V}^{\text{region}}_{w,ct}(d)$ and bounds the variance of region-aggregate treated potential outcomes, $\tilde{V}^{\text{region}}_{w,t}(d)$. See Appendix A.2.6 for details on the variance estimator.

Remark 13. The approximate finite sample variance is *smaller* under Bernoulli trials than under a completely randomized design (most easily seen in equation A.4 in the Appendix). This is an artifact of the approximation, which does not penalize the variance as heavily when for instance few treated regions are available under an imbalanced assignment. In practice, the difference between both designs is likely negligible due to the factor

$J-1$ in the denominator of each $\mathcal{C}$, such that $\sqrt{J}\frac{\mathcal{C}}{J-1} \to 0$ and there is no difference between the two designs under standard asymptotics in the number of regions.

Remark 14. In the covariance (part (iii) of Theorem 3) above, the covariance of treatment effects at distance d and d' appears negatively. In contrast to the case of the variance, dropping this term does not unambiguously lead to a conservative estimate of the covariance. However, in a number of special cases of interest, the covariance can be combined with variances of treatment effects into the variance of a sum (or difference) of treatment effects; see Corollary 1 and Theorem 4 for examples. In other cases, the covariance can be bounded on either side by $\pm$ the square root of the product of the variances. Alternatively, in many social science applications, it may be reasonable to assume that if the (aggregated) effects of the treatment are large at one distance from treatment in a given region, then (aggregated) effects are also large at another distance from treatment in the same region. In other words, the covariance of treatment effects at different distances may often be positive, such that dropping the inestimable term yields a conservative estimator under this additional assumption.

4.2 Aggregate Effects

The aggregate effect of a single treatment on all affected individuals is of importance for cost-benefit and welfare analyses. In this section, I propose estimators of aggregate effects that build on the estimators of individual-level effects of the previous section.

In experiments with spatial treatments, there are two units of observation: outcome individuals and spatial treatments. The individual-level treatment effects of the previous section are average effects per outcome individual. The aggregate treatment effects of this section are average effects per spatial treatment.

Suppose the researcher is interested in the aggregate effect that a single treatment location has on all affected individuals. Define the estimand

$$\tau^{agg} \equiv \frac{1}{\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j} w_j(s)} \sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j} w_j(s) \sum_{i\in\mathbb{I}_j} \tau_i(s) \tag{7}$$

where, as before, $\tau_i(s) = Y_i(s) - Y_i(0)$ is the effect of treatment location s on individual i. The aggregate treatment effect sums the $\tau_i(s)$ across individuals i and averages them across candidate treatment locations s, with weights $w_j(s)$.

In this section, I focus on the average aggregate treatment effect on the treated, τ^{AATT}, which uses weights

$$w_j^{AATT}(s) \equiv \Pr(\xi_j = s | W_j = 1) \Pr(W_j = 1).$$

The estimand places larger weight on the effects of treatment locations that are more likely to be realized. The estimand τ^{AATT} therefore answers the question: What is the expected aggregate effect of a treatment location under the observed policy of assigning treatments to locations?

One can estimate the aggregate effect τ^{AATT} by aggregating outcomes at the region-level:

$$\hat{\tau}^{AATT,1} \equiv \frac{1}{\sum_{j=1}^{J} W_j} \sum_{j=1}^{J} W_j Y_j - \frac{1}{\sum_{j=1}^{J} \frac{(1-W_j)\Pr(W_j=1)}{1-\Pr(W_j=1)}} \sum_{j=1}^{J} \frac{(1-W_j)\Pr(W_j=1)}{1-\Pr(W_j=1)} Y_j$$

where $Y_j \equiv \sum_{i\in\mathbb{I}_j} Y_i$. This is the inverse probability weighting estimator of an average treatment effect on the treated, where the outcome variable of interest is the sum over the outcomes of all individuals in a

region. When there is a single candidate treatment location per region, standard results from the literature on experiments with individual-level treatments apply (cf. Imbens, 2004), with regions taking the role of individuals.

Estimators based on region-aggregate outcomes are likely to have very large variance. Each region-aggregate outcome is the sum of outcomes of individuals in the region. If there is substantial variance in the number of individuals per region and outcomes are positive, the aggregate outcome of regions with many individuals can be substantially larger than the aggregate outcome of smaller regions. For instance, suppose that the number of individuals per region is Poisson distributed with mean n, and individual-level outcomes are i.i.d. within and across regions, with mean μ and variance σ^2. Then region-aggregate outcomes have variance $n \cdot (\sigma^2 + \mu^2)$ by the law of total variance. Hence, aggregate potential outcomes have large variance, which leads to large variance of the estimator (cf. Imbens, 2004).

Variation in region sizes generates a large variance of the region-aggregate estimator $\hat{\tau}^{AATT,1}$ in two ways. First, if there is variance in the number of individuals per region, then in finite samples, some treatment assignments will be such that there are more individuals in treated regions than in control regions.[16] Suppose outcomes are positive and constant, such that all individuals have the exact same (positive) value for the outcome. Then the treatment effect estimate $\hat{\tau}^{AATT,1}$ in such a sample is positive and sensitive to the scale of the outcome value. Hence, the estimator $\hat{\tau}^{AATT,1}$ can have large variance even when there is *no* variance in potential outcomes. Second, variation in region sizes increases the variance in a sampling-of-regions thought experiment. Even if the average individual-level treatment effect was known, needing to estimate the number of times the effect is realized on average per region can create substantial variance. The design-based variances considered in this paper condition on the individuals in the sample. With known number of individuals and known individual-level average treatment effect, it is possible to form an estimator of aggregate treatment effects with a design-based variance equal to zero, in contrast to the variance results for the estimator $\hat{\tau}^{AATT,1}$ above.

I therefore recommend an estimator of average aggregate effects that reduces the variance by building on the estimators of average individual-level effects at a distance d. Let

$$\hat{\tau}^{AATT,2} \equiv \sum_{d \in \mathbb{D}} \bar{n}(d)\hat{\tau}(d) \tag{8}$$

where $\bar{n}(d)$ is the average number of individuals at distance d from candidate treatment locations:

$$\bar{n}(d) = \frac{\sum_{j=1}^{J} \sum_{s \in \mathbb{S}_j} \pi_j g_j(s) \sum_{i:\ |d(s,r_i)-d| \leq h} 1}{\sum_{j=1}^{J} \sum_{s \in \mathbb{S}_j} \pi_j g_j(s)}$$

using the same distance bins for both $\hat{\tau}(d)$ and $\bar{n}(d)$. The set of distances $\mathbb{D}$ contains the midpoints of the bins that partition the full space into distance bins. For instance, if one uses distance bins $[0, 1]$, $(1, 2]$, $\ldots$, $(9, 10]$ for a treatment that is known not to have effects past a distance of 10 miles, then $\mathbb{D} = \{0.5, 1.5, \ldots, 9.5\}$ and $h = 0.5$.

The theoretical properties of the estimator $\hat{\tau}^{AATT,2}$ follow from those of $\hat{\tau}(d)$ in Theorem 1 above, and the covariance across distances as given in part (iii) of Theorem 3.

Theorem 4. *Under Assumptions 1 (no interference across regions) and 2 (completely randomized design*

[16] Stratification in the experimental design or in analysis is an alternative solution to this problem. However, when the number of regions is small or moderate, stratification may not be practical or sufficient to resolve this issue.

with equal treatment probabilities across regions), the estimator $\hat{\tau}^{AATT,2}$ has an approximate finite sample distribution over the assignment distribution with

(i) unbiasedness: $E(\hat{\tau}^{AATT,2}) \approx \tau^{AATT}$

(ii) variance:

$$\begin{aligned}\text{var}\Big(\hat{\tau}^{AATT,2}\Big) = \sum_{d\in\mathbb{D}} \bar{n}(d)^2 \Big(\frac{J-1}{J}\frac{\tilde{V}_t^{location}(d)}{J_t} + \frac{\tilde{V}_c^{region}(d)}{J_c} + \frac{1}{J}\frac{\tilde{V}_t^{region}(d)}{J_t} - \frac{\tilde{V}_{ct}^{region}(d)}{J}\Big) \\ + 2\sum_{d\in\mathbb{D}}\sum_{d'\in\mathbb{D},d'\neq d} \bar{n}(d)\bar{n}(d')\Big(\frac{J-1}{J}\frac{\tilde{V}_t^{location}(d,d')}{J_t} + \frac{\tilde{V}_c^{region}(d,d')}{J_c} \\ + \frac{1}{J}\frac{\tilde{V}_t^{region}(d,d')}{J_t} - \frac{\tilde{V}_{ct}^{region}(d,d')}{J}\Big)\end{aligned}$$

where

$$\begin{aligned}\tilde{V}_t^{location}(d,d') \equiv \frac{1}{\bar{n}(d)\cdot\bar{n}(d')\cdot(J-1)}\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j} g_j(s)\Bigg(\Big(\sum_{i:\,|d(s,r_i)-d|\leq h}(Y_i(s)-\mu_t(d))\Big) \\ \cdot\Big(\sum_{i:\,|d(s,r_i)-d'|\leq h}(Y_i(s)-\mu_t(d'))\Big)\Bigg) \\ \tilde{V}_c^{region}(d,d') \equiv \frac{1}{\bar{n}(d)\cdot\bar{n}(d')\cdot(J-1)}\sum_{j=1}^{J}\Bigg(\Big(\sum_{s\in\mathbb{S}_j} g_j(s)\sum_{i:\,|d(s,r_i)-d|\leq h}(Y_i(0)-\mu_c(d))\Big) \\ \cdot\Big(\sum_{s\in\mathbb{S}_j} g_j(s)\sum_{i:\,|d(s,r_i)-d'|\leq h}(Y_i(0)-\mu_c(d'))\Big)\Bigg) \\ \tilde{V}_t^{region}(d,d') \equiv \frac{1}{\bar{n}(d)\cdot\bar{n}(d')\cdot(J-1)}\sum_{j=1}^{J}\Bigg(\Big(\sum_{s\in\mathbb{S}_j} g_j(s)\sum_{i:\,|d(s,r_i)-d|\leq h}(Y_i(s)-\mu_t(d))\Big) \\ \cdot\Big(\sum_{s\in\mathbb{S}_j} g_j(s)\sum_{i:\,|d(s,r_i)-d'|\leq h}(Y_i(s)-\mu_t(d'))\Big)\Bigg) \\ \tilde{V}_{ct}^{region}(d,d') \equiv \frac{1}{\bar{n}(d)\cdot\bar{n}(d')\cdot(J-1)}\sum_{j=1}^{J}\Bigg(\Big(\sum_{s\in\mathbb{S}_j} g_j(s)\sum_{i:\,|d(s,r_i)-d|\leq h}\big(Y_i(s)-Y_i(0)-(\mu_t(d)-\mu_c(d))\big)\Big) \\ \cdot\Big(\sum_{s\in\mathbb{S}_j} g_j(s)\sum_{i:\,|d(s,r_i)-d'|\leq h}\big(Y_i(s)-Y_i(0)-(\mu_t(d')-\mu_c(d'))\big)\Big)\Bigg)\end{aligned}$$

and $\bar{n}(d)$*,* $\mu_t(d)$*, and* $\mu_c(d)$ *are defined as in Theorem 1.*

Proof: *See Appendix A.3.*

Remark 15. For approximate unbiasedness, the estimator $\hat{\tau}^{AATT,2}$ must be based on $\hat{\tau}(d)$, not $\hat{\tau}^{ATT-eq}(d)$. Intuitively, when "integrating" the effect $\hat{\tau}(d)$ against the number of individuals at this distance, one needs to ensure that $\hat{\tau}(d)$ is an (approximately) unbiased estimate of the effect for these particular $\bar{n}(d)$ individuals, while the estimator $\hat{\tau}^{ATT-eq}(d)$ weights these individuals differently.

Remark 16. The variance follows from Theorems 1 and 3. Since $\hat{\tau}^{AATT,2}$ is a sum, its variance is a sum of the covariances of the terms. In the design-based perspective, the analysis is conditional on the individuals in

the sample. Hence, the (weighted) number of individuals in each bin, $\bar{n}(d)$, is fixed. The estimators $\hat{\tau}(d)$ for distances $d \in \mathbb{D}$ are therefore the only stochastic components. Section 4.1.1 proposes estimators that can be adapted straightforwardly for each of the (estimable) components in the variance of $\hat{\tau}^{AATT,2}$.

Remark 17. The (co-) variances of treatment effects, $\tilde{V}_{ct}^{\text{region}}(d)$ and $\tilde{V}_{ct}^{\text{region}}(d, d')$ are generally not identified without further assumptions, and $\tilde{V}_{ct}^{\text{region}}(d, d')$ may be negative, such that dropping this covariance does not necessarily yield a conservative expression for the variance of $\hat{\tau}^{AATT,2}$. However, as I show in Appendix A.3, dropping all the unidentified terms $\tilde{V}_{ct}^{\text{region}}(d)$ and $\tilde{V}_{ct}^{\text{region}}(d, d')$ $(\forall d, d' \in \mathbb{D})$ jointly yields a conservative estimator of the variance.

Remark 18. The optimal choice of distance bins (and bandwidths) remains an open question. If individuals are spread uniformly across space, equal-width rings with larger radii have larger area and hence contain more individuals. In practice, in densely populated areas, smaller bins may be preferable, and under suitable sequences of populations (infill asymptotics or growing number of regions), it may be possible to allow $h \to 0$ and $|\mathbb{D}| \to \infty$. Generally, in the formula above, additional distance bins decrease the (squared) weights $\bar{n}(d)$ at the cost of increasing variances $\text{var}\big(\hat{\tau}(d)\big)$.

4.3 Parametric Estimation

I discuss issues in imposing parametric assumptions on the decay of treatment effects over distance from treatment and estimation by least squares regression. First, I show how to impose a parametric model on the individual-level effects at different distances. Second, I show how to estimate aggregate effects based on such a model.

Linear parametric models for the decay of average treatment effects over distance from treatment take the form

$$\tau_w(d) = \sum_k \beta_k \tilde{\lambda}_k(d)$$

where $\tilde{\lambda}_k$ are known functions of distance, and β_k are coefficients to be estimated.

In many settings, one needs to impose a distance after which the treatment has no effect, even within region, to obtain reasonable estimates from parametric models. Assumption 4 below formalizes this assumption.

Assumption 4. The treatment has no effect after a distance $d^{\max}$. That is, for any individual $i \in \mathbb{I}$, set of treatment locations $S \subset \mathbb{S}$, and location $s \in S$ such that $d(s, r_i) > d^{\max}$,

$$\tau_i(S) = \tau_i(S \setminus \{s\}).$$

Without such a restriction, any simple functional form for $\tilde{\lambda}$ will typically offer a poor approximation for at least some distances d from treatment.

One can improve the approximation to the treatment effect at short distances by using functions that only fit the treatment effect pattern up to the maximum distance $d^{\max}$:

$$\tau_w(d) = \sum_k \beta_k \lambda_k(d) \mathbb{1}\{d \leq d^{\max}\}.$$

Relatively simple functions λ_k may well approximate the average treatment effects at distances $d \in (0, d^{\max})$. This imposes contextual knowledge that average treatment effects are negligible at large distances from treatment. It also resembles a "bet on sparsity" (Hastie et al., 2001): If treatment effects really are negligible

at distances longer than $d^{\max}$, the estimators proposed below will likely perform well. If treatment effects are not negligible even at long distances, then no (parametric) estimator will perform well.

For instance, one can impose a linear functional form on the treatment effect decay by choosing $\lambda_1(d) = 1$, $\lambda_2(d) = d$. The coefficient β_2 then measures the rate of decay, while β_1 measures the effect of the treatment on individuals right by the treatment location. A quadratic functional form is imposed by $\lambda_1(d) = 1$, $\lambda_2(d) = d$, $\lambda_3(d) = d^2$. In principle, the analysis in this section can be extended also to functional forms that are non-linear in the parameters, such as an exponential decay of treatment effects with unknown rate of decay, $\tau(d) = \exp(\beta(-d))$.

To estimate the parameters β, suppose initially that there is only a single candidate treatment location in each region. This allows the definition of the distance of individual i from the candidate treatment location uniquely as d_i, irrespective of realized treatment. Then estimate the *weighted* linear regression

$$Y_i = \sum_k \beta_k \Big(W_{j(i)} \lambda_k(d_i, x_i) \mathbb{1}\{d_i \leq d^{\max}\} \Big) + h(d_i) + \epsilon_i$$

with weights reflecting those in Equation 6, depending on the estimand. The functions $\lambda_k(d_i, x_i)$ can depend on individual characteristics x_i to allow for heterogeneity in effects, such as separate λ_k for distinct groups of individuals.

The function h models the average control potential outcomes at each distance from candidate treatment locations. For semiparametric estimation, specify the treatment effect decay (λ) parametrically, and estimate h nonparametrically, as a partially linear model (e.g. Robinson, 1988). Here, I instead focus on parametric linear estimation, which imposes known parametric functions λ and h and estimates their coefficients:

$$Y_i = \alpha_0 + \sum_k \beta_k \Big(W_{j(i)} \lambda_k(d_i, x_i) \mathbb{1}\{d_i \leq d^{\max}\} \Big) + \sum_\ell \gamma_\ell \Big(h_\ell(d_i) \mathbb{1}\{d_i \leq d^{\max}\} \Big) + \epsilon_i$$

The same caveat about setting a maximum distance applies also to h. Since there is no interest in effects at distances larger than $d^{\max}$, the constant α_0 captures the mean outcome for individuals at such larger distances.

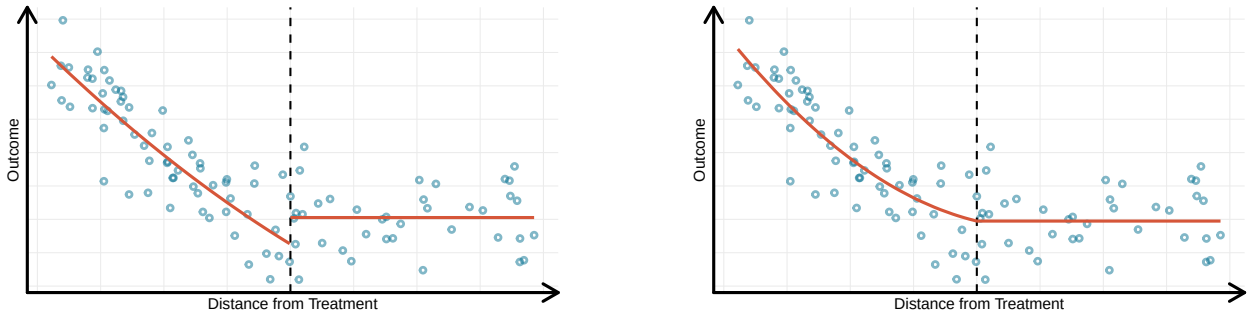
In practice, one typically not only wants to impose a zero treatment effect after distance $d^{\max}$ (Assumption 4), but a treatment effect that tends to zero continuously at $d^{\max}$.[17] To this end, estimate the linear regression with transformed covariates

$$Y_i = \alpha_0 + \sum_k \beta_k \Big(W_{j(i)} (\lambda_k(d_i, x_i) - \lambda_k(d^{\max}, x_i)) \mathbb{1}\{d_i \leq d^{\max}\} \Big) + \sum_\ell \gamma_\ell \Big(h_\ell(d_i) \mathbb{1}\{d_i \leq d^{\max}\} \Big) + \epsilon_i \quad (9)$$

which imposes the restriction $\tau(d^{\max}) = \sum_k \beta_k \lambda_k(d^{\max}) = 0$. Figure 5 illustrates what it means to impose this restriction. In panel (a), without the restriction, the estimated treatment effect will jump to 0 discontinuously at $d^{\max}$. Imposing the restriction in panel (b), the estimated treatment effect is continuous also at $d^{\max}$. The restriction generally reduces the variance of the estimator, in particular for estimating aggregate effects, as discussed below. In practice, most functional forms for λ imply not just a zero effect after distance $d^{\max}$, but also a non-zero effect at distances slightly shorter than $d^{\max}$.

The same parametric functional form can be imposed to estimate the average aggregate effects of the

[17] In principle, one could additionally impose higher order smoothness such as differentiability at $d^{\max}$. However, this generally requires more complicated functional forms λ to retain sufficient flexibility at shorter distances. In practice, this likely negates any improvements in precision.



(a) Unrestricted estimator (b) Restricted estimator

Figure 5: This figure illustrates the difference between leaving the parametric treatment effect function unrestricted (panel a) and restricting the treatment effect to be continuous at the maximum effect distance $d^{\max}$ (panel b). The figure shows a scatter plot of outcomes against distance from treatment. Both regression estimators use a quadratic in distance that is set to 0 at a distance of $d^{\max}$. The restricted estimator further restricts the regression coefficients such that the function is continuous at $d^{\max}$.

treatment. Under the parametric model, the average aggregate treatment effect on the treated is

$$\tau^{AATT} = \frac{1}{J}\sum_i \sum_k \beta_k(\lambda_k(d_i, x_i) - \lambda_k(d^{\max}, x_i))\mathbb{1}\{d_i \leq d^{\max}\}$$

Solving for β_1 and substituting the resulting expression in the regression specification above, one obtains the one-step regression specification

$$\begin{aligned} Y_i = \alpha_0 + \tau^{AATT}&\left(W_{j(i)}\frac{(\lambda_1(d_i, x_i) - \lambda_1(d^{\max}, x_i))\mathbb{1}\{d_i \leq d^{\max}\}}{\frac{1}{J}\sum_{i'}(\lambda_1(d_{i'}, x_{i'}) - \lambda_1(d^{\max}, x_{i'}))\mathbb{1}\{d_{i'} \leq d^{\max}\}}\right) \\ &+ \sum_k \beta_k \Bigg(\Big(W_{j(i)}(\lambda_k(d_i, x_i) - \lambda_k(d^{\max}, x_i))\mathbb{1}\{d_i \leq d^{\max}\}\Big) \\ &\qquad - \Big(W_{j(i)}(\lambda_k(d_i, x_i) - \lambda_k(d^{\max}, x_i))\mathbb{1}\{d_i \leq d^{\max}\} \\ &\qquad\quad \cdot \frac{\frac{1}{J}\sum_{i'}(\lambda_k(d_{i'}, x_{i'}) - \lambda_k(d^{\max}, x_{i'}))\mathbb{1}\{d_{i'} \leq d^{\max}\}}{\frac{1}{J}\sum_{i'}(\lambda_1(d_{i'}, x_{i'}) - \lambda_1(d^{\max}, x_{i'}))\mathbb{1}\{d_{i'} \leq d^{\max}\}}\Big)\Bigg) \\ &+ \sum_\ell \gamma_\ell \Big(h_\ell(d_i)\mathbb{1}\{d_i \leq d^{\max}\}\Big) + \epsilon_i \end{aligned} \tag{10}$$

where the coefficient on the first (transformed) covariate is the estimate of the average aggregate treatment effect. The transformed covariates are readily computed by realizing they are equal to the original covariates multiplied or shifted by average covariates. The average here is taken across all regions, both treated and untreated, such that this estimate has similarly attractive properties as the nonparametric estimator $\hat{\tau}^{AATT,2}$ above, in leveraging that the *number* of individuals near candidate treatment locations are available irrespective of assignment.

When there is more than one candidate treatment location per region, augment the regression approach as follows. The variable d_i is not uniquely defined, since there are multiple "distances from candidate treatment locations" for individuals. Suppose individual i in a control region ($W_{j(i)} = 0$) is 1 mile away from one candidate treatment location and 5 miles away from a different candidate treatment location. Then i should be used to estimate the control mean $h(d)$ for the two distances $d = 1$ and $d = 5$. One can therefore duplicate

observation i. Specifically, if individual i is in a region with $|\mathbb{S}_{j(i)}|$ candidate treatment locations, then include i $|\mathbb{S}_{j(i)}|$ times in the regression. Each version of i uses the distance d_i to a different candidate treatment location. This ensures $E(\epsilon_i|d_i = d, x_i = x) = 0$ and hence results in consistent parameter estimates by linear regression.

In simulations, standard errors clustered at the region level (cf. Liang and Zeger, 1986) provide a reasonable, but perhaps conservative, estimate of the variance of these estimators. One can derive formal results along the lines of Abadie et al. (2020, 2017). When there is a single candidate treatment location per region and a single distance of interest, the spatial setting considered here coincides with the setting of clustered assignment of Abadie et al. (2017), and hence their results and interpretation of Liang and Zeger (1986) clustered standard errors follow immediately. Refinements of Liang and Zeger (1986) clustered standard errors may be possible following Abadie et al. (2020) for the non-clustered setting using "attributes." In the spatial setting, such attributes are readily available in the form of the *number* of units near candidate treatment locations. Effectively, such attributes allow forming a tighter bound on the variance of treatment effects (see the discussion on variance estimation following Theorem 1 above) by exploiting heterogeneous treatment effects and appealing to the law of total variance to maintain that the estimator is still conservative for the true variance. For parametric models of the treatment effect by distance, one needs to extend the analysis of Abadie et al. (2017) to include (multiple) continuous regressors that are deterministic functions of the binary, randomly assigned, treatment. With multiple candidate treatment locations per region, one further needs to extend the binary treatment to a multi-valued (but still discrete) treatment.

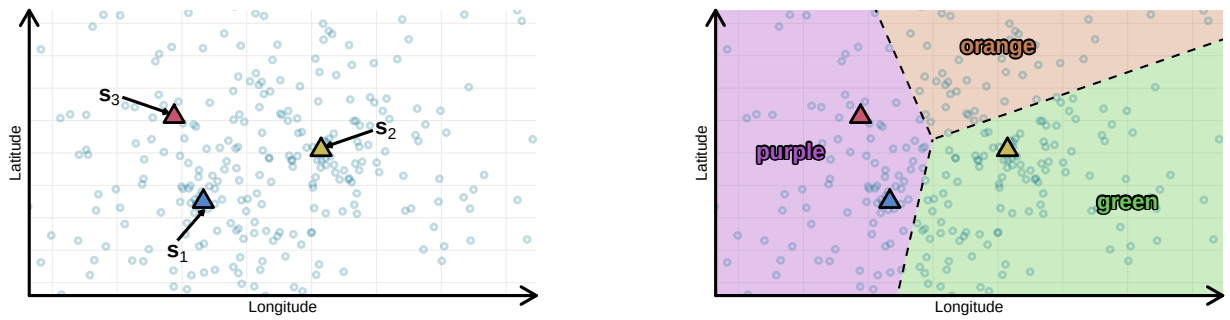
5 Extensions

5.1 Interference Between Treatment Locations

The setting discussed in Section 4 rules out interference between treatment locations by design. However, in practice treatment is sometimes realized at locations close to one another such that the same individual could be affected by more than one treatment location. The spatial treatment setting offers meaningful structure enabling identification of interesting treatment effects based on application-dependent assumptions despite this interference. I briefly outline two such assumptions, one familiar (additive separability of effects) and one more unique to the spatial treatment setting (only nearest realized treatment matters), and discuss what effects are identified under these assumptions and how to estimate them. I then turn to an even more general setup, without any true control regions, and focus on design-based inference for marginal treatment effects using exposure mappings (cf. Aronow and Samii, 2017) based on the idea that the effects of far-away treatment locations are typically negligible.

5.1.1 Assumptions Giving Structure To Interference

Suppose that if a region is randomized into the treatment group, multiple locations within the region are treated. The remaining regions remain in the control group without any treatment, such that it is straightforward to estimate average outcomes in the absence of treatment. To give an example, suppose a company operating chain stores (quasi-) randomly chooses which cities to enter, and opens multiple stores in chosen cities. Then there are multiple realized treatment locations close to one another (in the same city), but also control regions with (counterfactual) candidate treatment locations. The difficulty lies in separating the effects of multiple treatment locations. I discuss two assumptions, which may be credible in different



(a) (b)

Figure 6: An example of a region with three candidate treatment locations (panel a): s_1 (blue), s_2 (red), s_3 (yellow). Suppose exactly two of these treatment locations are realized whenever the region is treated, such that there is interference. Under the assumption that only the nearest realized treatment location matters, panel b illustrates the locations for which we can estimate effects for individuals in each area. For individuals in the orange area, we can estimate the effects of the red and yellow locations. For individuals in the green area, we can estimate the effects of the blue and yellow locations. For individuals in the purple area, we can estimate the effects of the blue and read locations.

settings, as examples for how the spatial treatment setting can provide the necessary structure.

For a simple and concrete example, suppose there are three candidate treatment locations in each region, and if treated, two of the three locations are chosen according in a completely randomized experiment.[18] Figure 6a illustrates one such region. Suppose the researcher has data from multiple regions j. Each region j has three candidate treatment locations; $s_{j,1}$, $s_{j,2}$, and $s_{j,3}$. If region j receives treatment, the assignment mechanism randomly chooses exactly two of the three candidate locations to be realized. Hence, each candidate location has marginal conditional probability of 2/3 of being realized. The set of realized treatment locations in region j, ξ_j, satisfies $\xi_j \in \big\{\emptyset, \{s_{j,1}, s_{j,2}\}, \{s_{j,1}, s_{j,3}\}, \{s_{j,2}, s_{j,3}\}\big\}$, where $\xi_j = \emptyset \iff W_j = 0$. The average effect of interest is that of implementing a single treatment location on individuals at distance d from it. The estimators of this section generalize to arbitrary numbers of candidate and realized locations, potentially differing by region.

I discuss two assumptions on interference and how to estimate effects under them in turn.

Additive Separability of Treatment Effects Assumption 5 states that the effects of all treatment locations are additively separable. Intuitively, the assumption requires that returns to additional realized treatment locations are neither diminishing nor increasing 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. For instance, the effects of toxic waste plants (cf. Currie et al., 2015) or air-polluting power plants (cf. Zigler and Papadogeorgou, 2021) on exposure to pollution are likely approximately additive.

Assumption 5. The effects of spatial treatments are additively separable. Let $S \subset \mathbb{S}$ be an arbitrary subset of the candidate treatment locations, and let $s \in S$ be an arbitrary location in this subset. Then, for all

[18] In practice, it is sometimes more plausible to assume that the assignment mechanism guarantees some minimum distance between realized treatment locations, requiring more complicated assignment mechanisms. For instance, sugar factories may be spread out such that each factory has sufficient land nearby to grow crop (Dell and Olken, 2020). It may be possible to obtain analogous results for such more complicated assignment mechanisms.

individuals $i \in \mathbb{I}$,

$$\tau_i(S) = \tau_i(S \setminus s) + \tau_i(s).$$

Additive separability as stated here neither requires different treatment locations to have the same effect, nor does it require a treatment location to have the same effect on two distinct individuals, even if they are at the same distance from the location.

Under Assumption 5, one can still identify the average treatment effects defined in Section 4. These estimands are weighted averages of individual-level treatment effects $\tau_i(s)$ of individual i and candidate treatment location s which are distance d apart. Under additive separability, Assumption 5, one of the following potential outcomes is observed

$$\begin{aligned}
Y_i(\emptyset) &= Y_i(0) \\
Y_i(\{s_{j,1}, s_{j,2}\}) &= Y_i(0) + \tau_i(s_{j,1}) + \tau_i(s_{j,2}) \\
Y_i(\{s_{j,1}, s_{j,3}\}) &= Y_i(0) + \tau_i(s_{j,1}) + \tau_i(s_{j,3}) \\
Y_i(\{s_{j,2}, s_{j,3}\}) &= Y_i(0) + \tau_i(s_{j,2}) + \tau_i(s_{j,3})
\end{aligned}$$

Hence, one can write the individual-level treatment effect of interest as a sum of observable potential outcomes as

$$\tau_i(s_{j,1}) = \frac{1}{2}\big(Y_i(\{s_{j,1}, s_{j,2}\}) + Y_i(\{s_{j,1}, s_{j,3}\}) - Y_i(\{s_{j,2}, s_{j,3}\})\big) - Y_i(\emptyset).$$

where each of the potential outcomes has positive probability of realization. One can therefore estimate $\tau_i(s_1)$ as

$$\begin{aligned}
\hat{\tau}_i^{\text{additive}}(s_{j,1}) = \frac{1}{2}\Bigg(&\frac{\mathbb{1}\{\xi_j = \{s_{j,1}, s_{j,2}\}\}}{\Pr(\xi_j = \{s_{j,1}, s_{j,2}\})} Y_i + \frac{\mathbb{1}\{\xi_j = \{s_{j,1}, s_{j,3}\}\}}{\Pr(\xi_j = \{s_{j,1}, s_{j,3}\})} Y_i - \frac{\mathbb{1}\{\xi_j = \{s_{j,2}, s_{j,3}\}\}}{\Pr(\xi_j = \{s_{j,2}, s_{j,3}\})} Y_i\Bigg) \\
&- \frac{1 - W_j}{1 - \Pr(W_j = 1)} Y_i.
\end{aligned}$$

Each term in the sum is an unbiased estimator of the corresponding potential outcome, such that $E(\hat{\tau}_i^{\text{additive}}(s_{j,1})) = \tau_i(s_{j,1})$. One can then average such estimators to estimate, for instance, the ATT estimand, $\tau(d)$. I present the generalization of the formula above with an arbitrary number of candidate treatment locations and realized treatment locations in Appendix A.4.1.

Only Nearest Realized Treatment Location Matters Assumption 6 states that if $s' \in S$ is not the nearest realized location to individual i, it does not affect her. 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 (cf. Duflo, 2001), it may be plausible to assume that only the nearest 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 (cf. Gupta et al., 2022), the appropriate assumption may depend on the type of stops that are added. 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.

Assumption 6. Only the nearest realized treatment location matters. Let $S \subset \mathbb{S}$ be an arbitrary subset of the candidate treatment locations, and $i \in \mathbb{I}$ an arbitrary individual. Whenever $s \in S$ satisfies $d(s, r_i) \leq d(s', r_i)$

for all $s' \in S$, then for $s' \in S \setminus \{s\}$

$$\tau_i(S) = \tau_i(S \setminus s').$$

The assumption also implies that if individual i is at equal distance to two treatment locations s_1 and s_2, then both have the same effect on her:

$$d(s_1, r_i) = d(s_2, r_i) \implies \tau_i(s_1) = \tau_i(s_2).$$

Under Assumption 6, only some of the average treatment effects of Section 4 are nonparametrically identified in general. Specifically, it is impossible to identify the effect of a candidate treatment location on an individual if the location is never the nearest realized location for the individual. Consider again the example of three candidate locations with two realized locations in Figure 6. The effect of location $s_{j,1}$ is unidentified for individuals nearer to both locations $s_{j,2}$ and $s_{j,3}$. Panel b of Figure 6 highlights areas in which each candidate treatment location is nearest with positive probability before realization of treatment assignment. Generally, the estimand $\tau_w(d)$ from Section 4 is identified nonparametrically under Assumption 6 if it only places weight on individual level effects $\tau_i(s)$ if s is the nearest realized location to individual i with positive probability. Formally, write this as

$$w_i(s,d) \neq 0 \implies \Pr\big(s \in \xi_j \wedge (d(s, r_i) \leq d(s', r_i) \ \ \forall s' \in \xi_j)\big) > 0$$

where the probability is taken over draws from the assignment distribution of ξ_j for fixed i, s, d, and j.

In the example illustrated in Figure 6, one can estimate $\tau_i(s_{j,1})$ for individuals i in the purple and green shaded areas. Under Assumption 6, the potential outcomes satisfy

$$Y_i(\{s_{j,1}, s_{j,2}\}) = \begin{cases} Y_i(0) + \tau_i(s_{j,1}) & \text{if } d(s_{j,1}, r_i) \leq d(s_{j,2}, r_i) \\ Y_i(0) + \tau_i(s_{j,2}) & \text{otherwise} \end{cases}$$

$$Y_i(\{s_{j,1}, s_{j,3}\}) = \begin{cases} Y_i(0) + \tau_i(s_{j,1}) & \text{if } d(s_{j,1}, r_i) \leq d(s_{j,3}, r_i) \\ Y_i(0) + \tau_i(s_{j,3}) & \text{otherwise} \end{cases}$$

An unbiased estimator of $\tau_i(s_{j,1})$ is

$$\hat{\tau}_i^{\text{nearest}}(s_{j,1}) = \begin{cases} \frac{\mathbb{1}\left\{S_j = \{s_{j,1}, s_{j,2}\}\right\}}{\Pr\left(S_j = \{s_{j,1}, s_{j,2}\}\right)} Y_i - \frac{1 - W_j}{1 - \Pr(W_j = 1)} Y_i \\ \qquad\qquad \text{if } d(s_{j,1}, r_i) \leq d(s_{j,2}, r_i) \wedge d(s_{j,1}, r_i) > d(s_{j,3}, r_i) \\ \frac{\mathbb{1}\left\{S_j = \{s_{j,1}, s_{j,3}\}\right\}}{\Pr\left(S_j = \{s_{j,1}, s_{j,3}\}\right)} Y_i - \frac{1 - W_j}{1 - \Pr(W_j = 1)} Y_i \\ \qquad\qquad \text{if } d(s_{j,1}, r_i) > d(s_{j,2}, r_i) \wedge d(s_{j,1}, r_i) \leq d(s_{j,3}, r_i) \\ \frac{\mathbb{1}\{s_{j,1} \in S_j\}}{\Pr(s_{j,1} \in S_j)} Y_i - \frac{1 - W_j}{1 - \Pr(W_j = 1)} Y_i \\ \qquad\qquad \text{if } d(s_{j,1}, r_i) \leq d(s_{j,2}, r_i) \wedge d(s_{j,1}, r_i) \leq d(s_{j,3}, r_i) \\ \text{undefined} \\ \qquad\qquad \text{otherwise} \end{cases}$$

One can then average estimates $\hat{\tau}_i^{\text{nearest}}(s)$ across individuals i and locations s to estimate average treatment effects similar to those in Section 4. However, the estimator $\hat{\tau}_i^{\text{nearest}}(s_{j,1})$ is undefined for individuals in the

orange shaded area (last line of the definition). Since location $s_{j,1}$ is never the nearest realized treatment location for these individuals, it is impossible to estimate its effect on individuals in that area. That is, only average treatment effects that place weight $w_i(s_{j,1}, d) = 0$ on individuals i in the orange area are identified nonparametrically. See Appendix A.4.2 for the estimator with arbitrary number of treatment locations per region.

5.1.2 Inference Without Control Regions

Suppose the researcher has data on a single contiguous region with individuals i, outcomes Y_i and candidate treatment locations $\mathbb{S}$.[19] The realized treatment locations are $\xi \subseteq \mathbb{S}$, with assignment to locations $s, s' \in \mathbb{S}$ independent when $s \neq s'$. Assumption 7, which is a straightforward extension of Assumption 3 above, formalizes this assignment mechanism.

Assumption 7 (Independent Treatment Assignment – Single Region)**.** Assignment of treatment to locations is independent: For $s \in \mathbb{S}$, $S \subseteq \mathbb{S}$ with $s \notin S$,

$$\Pr(s \in \xi \mid S \subseteq \xi) = \Pr(s \in \xi) \equiv \pi_s.$$

As before, the researcher is interested in the weighted average treatment effect

$$\tau_w(d) = \frac{\sum_{s\in\mathbb{S}} \sum_{i\in\mathbb{I}} w_i(s,d)(Y_i(s) - Y_i(0))}{\sum_{s\in\mathbb{S}} \sum_{i\in\mathbb{I}} w_i(s,d)}$$

with known weights $w_i(s, d)$, for instance

$$w_i(s,d) = \Pr(s \in \xi) \mathbb{1}\{|d(s, r_i) - d| \leq h\}$$

for a distance-bin estimator of the average effect of the treatment on the treated.

To estimate this average treatment effect on the treated, consider the estimator

$$\begin{aligned}\hat{\tau}(d) = &\frac{\sum_{s\in\mathbb{S}} \mathbb{1}\{s \in \xi\} \sum_{i\in\mathbb{I}} \mathbb{1}\{|d(s, r_i) - d| \leq h\} Y_i}{\sum_{s\in\mathbb{S}} \mathbb{1}\{s \in \xi\} \sum_{i\in\mathbb{I}} \mathbb{1}\{|d(s, r_i) - d| \leq h\}} \\ &- \frac{\sum_{s\in\mathbb{S}} \mathbb{1}\{s \notin \xi\} \frac{\Pr(\xi\cup\{s\})}{\Pr(\xi)} \sum_{i\in\mathbb{I}} \mathbb{1}\{|d(s, r_i) - d| \leq h\} Y_i}{\sum_{s\in\mathbb{S}} \mathbb{1}\{s \notin \xi\} \frac{\Pr(\xi\cup\{s\})}{\Pr(\xi)} \sum_{i\in\mathbb{I}} \mathbb{1}\{|d(s, r_i) - d| \leq h\}}\end{aligned} \tag{11}$$

which is the difference in average outcomes for individuals near realized candidate locations, $\mathbb{1}\{s \in \xi\} = 1$, and individuals near counterfactual candidate locations, $\mathbb{1}\{s \notin \xi\} = 1$. Note that the individuals near counterfactual candidate treatment locations may still be exposed to other *realized* treatment locations.

The estimator is distinct from common estimators for other treatment effect settings in that the same individuals may appear in both the average of the treated and in the average of the control. In Figure 7, there are three candidate treatment locations, two of which are realized (s_1 and s_2), while the last (s_3) is not realized. Individuals that are at intersections of rings of multiple treatment locations (i_2, i_3, i_4) appear multiple times in the summations in Equation 11. Individual i_2 appears twice in the sum of the treated (once for s_1 and s_2). Individual i_3 also appears twice in the sum of the treated, but additionally appears once in the sum of the control (for s_3). Individual i_4 appears once in the sum of the treated and once in the sum

[19] One can treat data from separate regions as a single contiguous region with large distances between the locations corresponding to the distinct regions.

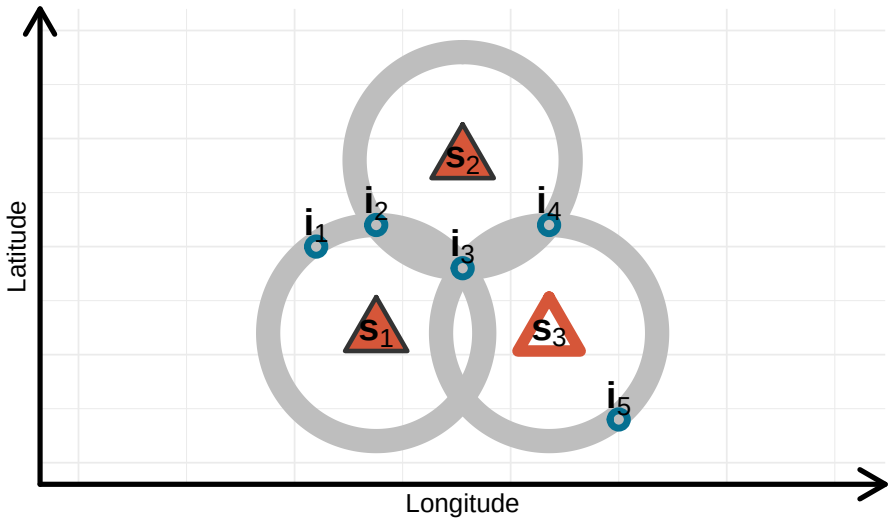


Figure 7: An example of a region with three candidate treatment locations (panel a): Treatment is realized at the locations of the solid triangles (s_1 and s_2) and not realized at the location of the hollow triangle (s_3). There are 5 individuals ($i_1, \ldots, i_5$) each at the distance of interest away from at least one of the candidate treatment locations. In the estimator given by Equation 11, individuals like i_3 and i_4 who are at the right distance from both a realized and a counterfactual treatment location appear both in the average of the treated and in the average of the control. Individuals who are at the right distance from multiple realized treatment locations, like i_2 and i_3, appear multiple times in the average of the treated. Other individuals, who are at the right distance from only one treatment location, like i_1 and i_5, may still be affected by treatments that are at a different distance. The design-based variance given in Theorem 5 accounts for these multiple occurrences of individuals and effects of treatments at other distances.

of the control. Individuals i_1 and i_5 appear only once each, in the sum of the treated and the sum of the control, respectively.

While i_1 and i_5 in Figure 7 may on first sight appear like more conventional treated and control individuals, their comparison does not necessarily yield an attractive estimate of the effect of the treatment at the distance of interest. Individual i_5 may, without other assumptions, still be affected by treatment at locations s_1 and s_2, such that i_5 is not a proper control individual. Similarly, i_1 could be affected by both s_1 and s_2 and may therefore reflect the outcome of having not just one treatment at the distance of interest, but also another treatment at a slightly longer distance. Hence, the difference between the outcome of i_1 and i_2 does not necessarily reflect the effect of one treatment at the distance of interest.

Remarkably, the estimator defined in Equation 11 yields an estimate of the marginal effect of treatment without parametric assumptions on how the effects of multiple treatments interact. The weighting of the "control" individuals by $\frac{\Pr(\xi \cup \{s\})}{\Pr(\xi)}$ results in "treated" individuals that are, on average (across individuals and the assignment distribution) exposed to one additional realized treatment location at the distance of interest, and the same number of realized treatment locations at other distances. The weights are a generalization of ATT weights, and with independent treatment assignment simplify to the familiar $\pi_s/(1-\pi_s$ weighting. If one additionally assumes that treatment effects are additive (Assumption 5), the estimand simplifies to the ATT defined in the previous section for settings without interference. Alternative assumptions on how the effects of treatments interact, such as Assumption 6, may suggest different estimators. Results for such estimators could be derived analogously. The paper focuses on the estimator in Equation 11 because additivity of effects frequently is a reasonable benchmark and the estimator retains an attractive marginal effect interpretation

even if the additivity assumption does not hold.

Theorem 5. *Suppose Assumption 7 (independent treatment assignment in a single region) holds. Then the estimator $\hat{\tau}(d)$ has an approximate finite sample distribution over the assignment distribution with*

(i) mean: $E(\hat{\tau}(d)) \approx \frac{\sum_{S\in 2^{\mathbb{S}}}\sum_{s\in S}\sum_{i\in\mathbb{I}} \Pr(\xi=S)\mathbb{1}\{|d(s,r_i)-d|\leq h\}(Y_i(S)-Y_i(S\setminus\{s\}))}{\sum_{S\in 2^{\mathbb{S}}}\sum_{s\in S}\sum_{i\in\mathbb{I}} \Pr(\xi=S)\mathbb{1}\{|d(s,r_i)-d|\leq h\}}$

(ii) variance:

$$\begin{aligned}
&\operatorname{var}(\hat{\tau}(d)) \\
&\approx \frac{\sum_{s\in\mathbb{S}}\sum_{m\in\mathbb{M}_s}\pi_s(m)(1-\pi_s(m))Y_s(m,d)^2}{(\sum_{s\in\mathbb{S}}\Pr(s\in\xi)\sum_{i\in\mathbb{I}}\mathbb{1}\{|d(s,r_i)-d|\leq h\})^2} \\
&\quad - \frac{\sum_{s\in\mathbb{S}}\sum_{m\in\mathbb{M}_s}\sum_{m'\in\mathbb{M}_s}\mathbb{1}\{m\neq m'\}\pi_s(m)\pi_s(m')Y_s(m,d)Y_s(m',d)}{(\sum_{s\in\mathbb{S}}\Pr(s\in\xi)\sum_{i\in\mathbb{I}}\mathbb{1}\{|d(s,r_i)-d|\leq h\})^2} \\
&\quad + \frac{\sum_{s\in\mathbb{S}}\sum_{s'\in\mathbb{S}}\mathbb{1}\{s\neq s'\}\sum_{m\in\mathbb{M}_s}\sum_{m'\in\mathbb{M}_{s'}}(\pi_{s,s'}(m,m')-\pi_s(m)\pi_{s'}(m'))Y_s(m,d)Y_{s'}(m',d)}{(\sum_{s\in\mathbb{S}}\Pr(s\in\xi)\sum_{i\in\mathbb{I}}\mathbb{1}\{|d(s,r_i)-d|\leq h\})^2}
\end{aligned}$$

with sets $\mathbb{M}_s$ and probabilities $\pi_s(m)$ and $\pi_{s,s'}(m,m')$ defined based on the exposure mappings described in Remark 20 below, and the exposure-specific, demeaned, potential outcomes

$$\begin{aligned}
Y_s(m,d) \equiv W_s(m)&\sum_{i\in\mathbb{I}}\mathbb{1}\{|d(s,r_i)-d|\leq h\}(Y_i(m)-\mu_t(d)) \\
&-(1-W_s(m))\frac{\pi_s}{1-\pi_s}\sum_{i\in\mathbb{I}}\mathbb{1}\{|d(s,r_i)-d|\leq h\}(Y_i(m)-\mu_c(d))
\end{aligned}$$

when exposure m determines the treatment status of candidate treatment location s, and $Y_s(m,d)=0$ otherwise. $W_s(m)$ is an indicator for location s being treated under exposure m (if uniquely determined), and $\mu_t(d)$ and $\mu_c(d)$ are the relevant means defined explicitly at the beginning of Appendix A.4.3.

If also Assumption 5 (additive separability) holds, the mean in part (i) simplifies to

(iii) mean: $E(\hat{\tau}(d)) \approx \tau(d) = \frac{\sum_{s\in\mathbb{S}}\Pr(s\in\xi)\sum_{i\in\mathbb{I}}\mathbb{1}\{|d(s,r_i)-d|\leq h\}\tau_i(s)}{\sum_{s\in\mathbb{S}}\Pr(s\in\xi)\sum_{i\in\mathbb{I}}\mathbb{1}\{|d(s,r_i)-d|\leq h\}}$

Proof: *See Appendix A.4.3.*

Remark 19. While all treatment probabilities are known to the experimenter in experimental analyses, they typically need to be estimated in observational studies. Yet researchers only observe one realization of the random variable ξ, so direct estimation of $\Pr(\xi)$ is not possible. However, under Assumption 7 of independent treatment assignment, for $S\subset\mathbb{S}$ and $s\notin S$

$$\frac{\Pr(\xi=S\cup\{s\})}{\Pr(\xi=S)} = \frac{\Pr(s\in\xi)}{1-\Pr(s\in\xi)}.$$

Hence, a practical solution is to estimate a propensity score model for $\Pr(s\in\xi)$ using $\mathbb{1}\{s\in\xi\}$ as the outcome and characteristics of the neighborhood of s, such as the number of nearby candidate treatment locations or individuals, as covariates.

Remark 20. The variance in Theorem 5 is based on exposure mappings (cf. Aronow and Samii, 2017). Exposure mappings allow defining fixed potential outcomes in settings with interference, such that each potential outcome is observed under more than one realization of the treatment assignment. Under general

interference, any candidate treatment location can affect any individual. In the spatial treatment setting, it is instead often plausible to assume that far-away treatment locations do not affect individuals. This allows breaking the full set of candidate treatment locations into (overlapping) sets of candidate treatment locations that are geographically clustered.

For spatial treatments, a simple definition of exposure mappings is based on a distance after which treatments are assumed not to have an effect. For a candidate treatment location $s \in \mathbb{S}$, the exposure mapping f_s maps a treatment assignment $S \subset \mathbb{S}$ to a member of a set $\mathbb{M}_s$. Let $\mathbb{S}_s \equiv \{s' \in \mathbb{S} : \ d(s,s') \leq 2 \cdot d^{\max}\}$ be the set of candidate treatment locations such that both s and $s' \in \mathbb{S}_s$ may have an effect on the same individual i based on Assumption 4.[20] Then $\mathbb{M}_s = 2^{\mathbb{S}_s}$ is the power set of $\mathbb{S}_s$, and exposure mapping $m \in \mathbb{M}_s$ specifies which of the candidate locations that could affect individuals near s are realized under m. That is, $f_s(S) = S \cap \mathbb{S}_s$. Then, under Assumption 4, $f_s(S) = f_s(S') \implies Y_i(S) = Y_i(S')$ for all individuals i with $d(s, r_i) \leq d^{\max}$. Hence, the exposure m uniquely determines the outcome for individuals i near location s.[21]

For these exposure mappings, the probabilities of exposure mappings m and m' being realized for candidate treatment locations s and s' are

$$\pi_s(m) \equiv \Pr(\xi \cap \mathbb{S}_s = m) = \prod_{s \in m} \pi_s \prod_{s \in \mathbb{S}_s \setminus m} (1 - \pi_s)$$

$$\pi_{s,s'}(m,m') \equiv \Pr(\xi \cap \mathbb{S}_s = m \ \wedge \xi \cap \mathbb{S}_{s'} = m'))$$

$$= \begin{cases} \prod_{\tilde{s} \in m \cup m'} \pi_{\tilde{s}} \prod_{\tilde{s} \in (\mathbb{S}_s \setminus m) \cup (\mathbb{S}_{s'} \setminus m')} (1 - \pi_{\tilde{s}}) & \text{if } \forall \tilde{s} \in \mathbb{S}_s \cap \mathbb{S}_{s'} \text{ either } \tilde{s} \in m \cap m' \text{ or } \tilde{s} \notin m \cup m' \\ 0 & \text{otherwise} \end{cases}$$

The condition in $\pi_{s,s'}(m,m')$ says that exposures m and m' agree on the treatment status for any candidate location that may affect individuals near both s and s'. The expression for $\pi_{s,s'}(m,m')$ in the first case simplifies to $\pi_s(m)$ if $s = s'$ and $m = m'$. If $s = s'$ and $m \neq m'$, then $\pi_{s,s'}(m,m') = 0$ according to the second case. Generally, $\pi_{s,s'}(m,m') = 0$ whenever exposure mappings m and m' imply contradicting assignments for treatment locations. When $\mathbb{S}_s \cap \mathbb{S}_{s'} = \emptyset$ such that s and s' share no treatment location that can affect individuals near both of them (typically because s and s' are sufficiently far apart), then $m \in \mathbb{M}_s$ and $m' \in \mathbb{M}_{s'}$ cannot contradict each other and $\pi_{s,s'}(m,m') = \pi_s(m)\pi_{s'}(m')$. When exposure mappings m and m' do not contradict each other, their joint probability is equal to the probability of the candidate locations $\tilde{s}$ in $\mathbb{S}_s \cup \mathbb{S}_{s'}$ being assigned according to m and m' (to treatment if $\tilde{s} \in m \cup m'$ and to control otherwise).

Estimation of the variance in Theorem 5 Analogous to the detailed discussion of variance estimation following Theorem 1, some of the terms in the variance in Theorem 5 are marginal variances that can be estimated, while others involve covariances, or variances of treatment effects, that cannot be estimated directly and hence must be bounded. Specifically, the first term sums over the marginal variances of "treated" and "control" potential outcomes and can be estimated from observed data as before. The second term involves distinct exposures m and m' for the same candidate location s, such that $Y_s(m,d)$ and $Y_s(m',d)$ can never

[20] Under the alternative Assumption 5, part (iii) of Theorem 5 allows estimating the average effect of the treatment by distance. A plausible choice $d^{\max}$ is therefore such that $\hat{\tau}(d) \approx 0$ for $d \geq d^{\max}$.

[21] In principle, one can define slightly finer exposure mappings at the individual-level. For an individual i who is distance d away from treatment location s, the exposure mappings defined here are conservative in the sense that they allow different exposures based on treatment at a location s' that is within $2d^{\max}$ from s, even if it is farther than $d^{\max}$ from i. While the variance expression is still correct, in estimation this leads to more conservative bounding of covariance terms corresponding to treatment states that cannot be observed jointly. The advantage of the exposure mappings defined here is the substantially reduced computational complexity achieved by aggregating data at the candidate treatment location-level, as well as simpler analytical expressions for the variance. See also Aronow and Samii (2017) for feasible approximations when the number of exposure mappings is too large for exact computation.

be observed at the same time. Rewriting the product $Y_s(m,d)Y_s(m',d)$ similarly to the variance of treatment effects in Theorem 1 suggests bounding the second term using Young's inequality for products:

$$-Y_s(m,d)Y_s(m',d) \leq |Y_s(m,d)| \cdot |Y_s(m',d)| \leq \frac{Y_s(m,d)^2}{2} + \frac{Y_s(m',d)^2}{2}$$

which allows replacing the covariance (product) by marginal variances (squares) that can be estimated. The third term includes both products of location-exposure pairs (s,m) and (s,m') that can be observed simultaneously (those with $\pi_{s,s'}(m,m') > 0$) and pairs that cannot (those with $\pi_{s,s'}(m,m') = 0$). The latter needs to be bounded using Young's inequality as well.

5.2 Non-Spatial Settings

The framework, estimators, and analysis of this paper are applicable more generally to settings where treatment assignment is separate from the units of observation, and the effect of the treatment is moderated by some observable, not necessarily geographic, distance from treatment. This setting is similar to that of Borusyak and Hull (2020), with distance from candidate treatments playing the role of "non-random exposure." Their linear regression estimators with actual treatment demeaned by expected treatment are straightforward to compute in many settings. However, in simple settings, one can show that they estimate differently weighted average treatment effects.[22] In contrast, the (inverse probability weighting) estimators proposed in this paper are only immediately applicable to treatments that are inherently binary or categorical, but their implied weighting of individual-level effects is straightforward to analyze and adjust. I give two brief examples of non-spatial settings in this section: firm entry in markets with differentiated products, and shift-share designs based on randomness of the shifts.

Firm Entry in Product Space For the first example, suppose the researcher is interested in the effects of firm entry on competition in markets with differentiated products. She has data for several markets j on prices Y_i charged by firms $i \in \mathbb{I}_j$ for products with horizontal or vertical locations r_i in characteristics space. In some markets, a new firm enters with a product with characteristics ξ_j. Here, the estimand $\tau(d)$ measures the average effect an entrant has on the price of a product at distance d in characteristics space. For short distances d, it captures competitive effects or deterrence behavior by firms selling products very similar to the product of the entrant. For longer distances d, it captures ripple effects that arise if in equilibrium firms with more different products react to the price changes of firms with products similar to the entrant's. These estimands are therefore informative about the nature of competition.

Firm entry, however, is not generally random. In the framework of this paper, the entrant's characteristics ξ_j are not a random draw from a uniform distribution. Theoretical models of competition and profits may therefore help to determine the probability of firm entry at any given point in characteristics space, conditional on the locations of existing competitors in characteristic space. For instance, expected profits of the entrant may come from a structural model based on distance to competitors in characteristics space (cf. Hotelling, 1929), perhaps calibrated to pre-treatment data. Intuitively, validity of the estimator then requires that firm entry is random conditional on the expected profitability as given by the model. The structural model provides a baseline to enhance the credibility of the quasi-experimental analysis, but does not directly restrict the estimated pattern of competition.

[22] I thank Peter Hull for sharing his derivations with me.

Quasi-Random Shifts in Shift-Share Designs For a second example, suppose the researcher is interested in the causal effects of exogenous shocks to individual industries on employment outcomes in cities based on their industry mixes (cf. Autor et al., 2013), with multiple shocks observed over time. The framework of this paper is useful in this setting if the claims of causal identification are based on randomness in which industries are shocked, rather than on randomness in industry composition. Importantly, the analysis in this paper reflects that cities with similar industry mixes are shocked similarly, in a way that is difficult to capture accurately with existing clustered standard errors. Adao et al. (2019) and Borusyak et al. (2022) develop alternative approaches based on the same idea, and show how it relates to Bartik (1991) and shift-share instruments more generally. A benefit of the framework of this paper is that its results are not specific to linear (or other) functional form and that it allows for very transparent estimation of aggregate effects.

The setting fits into the framework of this paper as follows. Data are available for time periods $j = 1, \ldots, J$. In some time periods, a single industry $\xi_j \in \mathbb{S} = \{1, \ldots, K\}$ receives an exogenous shock, potentially with different industries shocked in different time periods. Assume that the time periods are chosen such that the shock only affects outcomes within the same time period. Define the indicator $W_j = 1$ if an industry in period j is shocked, and $W_j = 0$ otherwise. The researcher observes employment outcomes Y_i for cities $i \in \mathbb{I}_j$ in time period j. City i has industry shares $r_i \in \mathbb{R}^K$, satisfying $r_{i,k} \in [0, 1]$ and $\sum_{k=1}^K r_{i,k} = 1$. Here, the distance function captures exposure to the shock. City i is heavily exposed to shocks of sector k if industry k has large share $r_{i,k}$, such that the "distance" $d(k, r_i) = 1 - r_{i,k}$ is small between industry k and city i.

The estimands $\tau(d)$ and τ^{agg} measure the effects of the exogenous industry shocks. For $d = 0$, the estimand $\tau(d)$ measures the average effect of shocking an industry on employment in cities with employment only in the shocked industry. For $d = 0.75$, the estimand measures the average effect of the exogenous shock on cities with 25% of their employment in the shocked industry. The estimand τ^{agg} measures the aggregate effect of the exogenous shock across all cities. The estimators and inference procedures of Section 4 are valid if it is random in which time periods and sectors an exogenous shock occurs. In principle, one can augment the variance calculations to allow, for instance, dependence structure in the shocked industry across time periods. The results in Section 5.1 are relevant for settings where shocks occur to multiple industries in the same time period.

6 Analysis of Observational Data

While the previous sections presume that the researcher designs the experiment for assignment of the spatial treatment, much empirical work relies on observational data. The primary challenge to observational studies in this setting is that researchers typically do not observe the exact locations of counterfactual candidate treatment locations. To emulate the analysis of the ideal experiment with observational data, researchers need to estimate candidate treatment locations and their treatment probabilities. In this section, I discuss how to address these challenges in practice with deliberate choice of machine learning methods well-suited for the spatial treatment setting. Estimation is then based on an unconfoundedness assumption stating that among individuals near candidate treatment locations, whether their treatment location is realized is as good as random, conditional on characteristics of the individuals and the neighborhood of the candidate treatment location.

6.1 Unconfoundedness Assumptions for Spatial Treatments

I focus on an unconfoundedness assumption for spatial treatments emulating an experiment with independent assignment to locations in a (possibly) single contiguous region. The setting with multiple regions is similar conceptually but may require conditioning also on features of the region. When a fixed number of locations are realized within a treated region, the setting resembles discrete choice modeling, discussed in more detail in Section 6.3 below.

Define the location-specific treatment indicator W_s to equal 1 if location s is treated, and 0 otherwise:

$$W_s \equiv \mathbb{1}\{s \in \xi\}$$

where ξ is the set of realized treatment locations.

In treatment effect settings with individual-level randomized experiments, unconfoundedness is often written as

$$W_i \perp\!\!\!\perp Y_i(0), Y_i(1) \quad | \quad X_i = x \ \ \forall x$$

which is equivalent to an assumption on densities

$$f(w, y_0, y_1 | X_i = x) = f_w(w|X_i = x) f_y(y_0, y_1 | X_i = x) \quad \forall w, y_0, y_1$$

I similarly define unconfoundedness of spatial treatments at distance d from location s as

$$W_s \perp\!\!\!\perp \big(Y_i(0), Y_i(s)\big)_{i:\ d(s,r_i)=d} \quad | \quad Z_s = z, \big(X_i\big)_{i:\ d(s,r_i)=d} = x \ \ \forall z, x \tag{12}$$

where Z_s are (neighborhood) characteristics of the candidate treatment locations $s \in \mathbb{S}$, and X_i are characteristics of the individuals $i \in \mathbb{I}$. The notation above is meant as shorthand for

$$\begin{aligned} & f(w, y_{1,0}, y_{1,1}, \ldots, y_{n_s(d),0}, y_{n_s(d),1} | Z = z, X = x) \\ = \ & f_w(w|Z = z, X = x) f_y(y_{1,0}, y_{1,1}, \ldots, y_{n_s(d),0}, y_{n_s(d),1} | Z = z, X = x) \\ \forall \ & y_{1,0}, y_{1,1}, \ldots, y_{n_s(d),0}, y_{n_s(d),1}; w \\ \wedge & \\ & \text{exchangeability of individuals} \end{aligned}$$

where $n_s(d)$ is the number of individuals at distance d from s, and argument $y_{j,w}$ for $j = 1, \ldots, n_s(d)$ and $w \in \{0, 1\}$ is the potential outcome corresponding to treatment state w of candidate location s for the j-th individual at distance d from location s. Exchangeability of individuals is defined as follows. Fix $w, y_{1,0}, y_{1,1}, \ldots, y_{n_s(d),0}, y_{n_s(d),1}$. For any permutation π of $\{1, \ldots, n_s(d)\}$, the joint density satisfies

$$\begin{aligned} & f(w, y_{1,0}, y_{1,1}, \ldots, y_{n_s(d),0}, y_{n_w(d),1} | Z = z, X = x) \\ = \ & f(w, y_{\pi_1,0}, y_{\pi_1,1}, \ldots, y_{\pi_{n_s(d)},0}, y_{\pi_{n_s(d)},1} | Z = z, X = x_\pi) \end{aligned}$$

where x_π is the similar π-permutation of x.

Remark 21. Adjusting for the true propensity score

$$e(z, x) \equiv \Pr\big(W_s = 1 \mid Z_s = z, (X_i)_{i:\ d(s,r_i)=d} = x\big)$$

rather than the full set of characteristics $\left(Z_s, (X_i)_{i:\, d(s,r_i)=d}\right)$, is sufficient for unconfoundedness in Equation 12, as in the setting with individual-level treatments (cf. Rosenbaum and Rubin, 1983). While the unconfoundedness assumption in Equation 12 is written for a specific distance d from treatment, in practice a design-based, quasi-experimental, analysis most credibly conditions the independence on $(X_i)_{i:\, d(s,r_i)=d}$ for all d of interest simultaneously. The estimated propensity score, which uses characteristics of individuals at all distances of interest, is then the same for all distances. While it may be easier to balance covariates by estimating separate propensity scores for different distances, treatment probabilities that vary by distance for the same treatment location are inconsistent with the ideal experiment underlying the quasi-experimental analysis. Furthermore, estimands such as the average effect of the treatment on the treated, in Theorems 1 and 2, become even less comparable across distances when not just the individuals, but also the probability weighting of treatment locations, differs from distance to distance.

Remark 22. The conditioning of the unconfoundedness assumption draws a visual distinction between Z_s and $(X_i)_{i:\, d(s,r_i)=d}$. This is conceptually similar to the notion that one may try to correctly model the outcome, the propensity score, or ideally both, in the literature on double robustness. Both Z_s and $(X_i)_{i:\, d(s,r_i)=d}$ "vary" at the treatment location level, so in principle the latter can be included in the former for more succinct, but less informative, notation.

Remark 23. Unconfoundedness of spatial treatments justifies comparisons of individuals near two locations s and s', one treated, the other untreated, for which $Z_s = Z_{s'}$ and $(X_i)_{i:\, d(s,r_i)=d} =^{\pi} (X_i)_{i:\, d(s',r_i)=d}$. Here, $=^{\pi}$ means that the sets of individual covariates are the same up to permutation. In practice, it is rarely feasible to find two candidate locations with equal number of individuals and equal covariates. Instead, one can assume that treatment is unconfounded conditional on, for instance, average characteristics of individuals in the neighborhoods of candidate locations. Such an assumption greatly simplifies estimation in practice.

Alternatively, when candidate treatment locations are sufficiently far apart such that interference in not an issue (based on Assumption 4),[23] one can make an individual-level unconfoundedness-type assumption for spatial treatments as a conditional mean equality,

$$E\Big(Y_i(0) \mid \min_{s\in\xi}\{d(s,r_i)\} = d, X_i = x\Big) = E\Big(Y_i(0) \mid \min_{s\in\xi}\{d(s,r_i)\} > d^{\max}, X_i = x\Big)$$

which states that, as long as one focuses on individuals with the same characteristics X_i, control potential outcomes for individuals near realized treatment are the same as control potential outcomes for individuals far away from realized treatment. To make such an assumption more plausible in a spatial context, these individual characteristics likely include characteristics of the neighborhood of i. For estimation, one then does not need to find counterfactual treatment locations, but only needs to match on, or adjust for, the individual characteristics X_i (cf. Imbens and Rubin, 2015, Parts III and IV). Such an assumption simplifies estimation, but is not justified by experimental design or arguments that the location of the treatment is as good as random, such that the design-based inference proposed in this paper does not apply.

6.2 Finding Counterfactual Candidate Treatment Locations

In this section, I outline a general strategy for finding counterfactual candidate treatment locations with observational data. These counterfactual locations for the treatment are necessary for the quasi-experimental methods I propose in this paper.

[23] With interference, a similar assumption is possible but notationally more cumbersome.

Consider first the example of Linden and Rockoff (2008) given in the introduction, where the choice of candidate locations is relatively straightforward. They argue that the exact houses where sex offenders move in are as good as random due to random availability of houses within neighborhoods. Here, the candidate treatment locations are houses in these neighborhoods. Hence, the candidate *locations* are known, but their probabilities of treatment need to be estimated. See Section 6.3 for propensity score estimation.

When there are no (or insufficiently many) known counterfactual candidate locations, however, the problem of choosing candidate locations from continuous space is hard. In principle, one could imagine estimating the probability of treatment at any location in a region conditional on all the features of the region. This is akin to estimating the spatial distribution of treatment locations $\xi_j \sim G(Z_j)$, where Z_j are the characteristics of region j, potentially relative locations of all individuals in the region as well as moments of their covariates. One could then use the estimated $\hat{G}$ to inform the treatment probabilities at each point in the region as inputs in the estimators proposed in this paper.

In practice, it is typically sufficient to find a finite number of candidate treatment locations that offer a plausible counterfactual to the realized treatment locations. Computationally, it is often impractical to use a continuous distribution G, since the weight of individual i when estimating effects at distance d would depend on the integral of the noisy $\hat{G}$ along a ring with radius d around her location, r_i, for each of the typically many individuals $i \in \mathbb{I}$. Instead, I recommend finding a finite number of candidate locations. The average across these finitely many candidate locations approximates the strategy based on the complete distribution G, setting $\hat{G}$ to exactly zero for many of the implausible locations.

I propose taking draws $\xi_j \sim G(Z_j)$ to obtain candidate treatment locations, where $G(Z_j)$ is implicitly estimated in the process of taking the draws as described below. Perhaps surprisingly, recent machine learning methods achieve good results at this task, despite the difficulty of estimating G itself. Specifically, I recommend a formulation similar to generative adversarial networks (Goodfellow et al., 2014); see Liang (2018) and Singh et al. (2018) on the relationship between generative adversarial networks and density estimation. Most closely related to this paper, Athey et al. (2019) use generative adversarial networks to draw artificial observations from the distribution that generated the (real) sample, for use in Monte Carlo simulations.

Generative adversarial methods for drawing $\xi_j \sim G(Z_j)$ are based on iteration between two steps. First, a generator generates draws $\tilde{\xi}_j \sim \tilde{G}(Z_j)$, where $\tilde{G}$ is an implicit estimate of the density maintained by the generator in the current iteration. Second, a discriminator receives as input either counterfactual locations proposed by the generator, $\tilde{\xi}_j|Z_j$, or real treatment locations, $\xi_j|Z_j$, and guesses whether its input is real. Both the generator and the discriminator are highly flexible models (typically neural networks) designed for their given tasks. The discriminator is trained by taking (stochastic) gradient descent steps in the direction that improves discrimination between real and counterfactual locations. The generator is trained by taking (stochastic) gradient descent steps in the direction that leads to fooling the discriminator into classifying counterfactual locations as real.

Effectively, the output of such models are counterfactual candidate treatment locations $\tilde{\xi}_j|Z_j$ that are indistinguishable (to the discriminator) from real treatment locations $\xi_j|Z_j$. With a sufficiently flexible discriminator, the process is therefore similar to matching.[24] If a proposed candidate location $\tilde{\xi}_j$ is noticeably different from all real treatment locations ξ, a flexible discriminator will learn to reject $\tilde{\xi}_j$. In contrast, synthetic control-type methods (Abadie et al., 2010) would average multiple candidate locations, for instance $\tilde{\xi}_a$ and $\tilde{\xi}_b$, to create a synthetic counterfactual for a real treatment location ξ_j. If $\tilde{\xi}_a$ and $\tilde{\xi}_b$ individually differ

[24]Standard matching methods, however, are unlikely to perform well due to high dimensional covariates that describe spatial data, such as relative spatial locations between many individuals as well as their characteristics.

from all real treatment locations ξ, the discriminator will reject them despite their average resembling ξ_j.

The goal therefore is to find "false positives:" Occasions when the discriminator fails to reject a counterfactual location suggested by the generator. Discriminator networks do not necessarily make binary predictions, but may give a continuous activation score that indicates how likely a location is to be real.[25] In practice, I recommend looking for locations with high activation scores, and performing an additional propensity score matching (or matching based on the activation scores of the discriminator) to the real treatment locations to select from the (potentially many) false positives one can generate. Such locations are likely to be decent matches for the real treatment locations, since they must share features of realized locations in order to achieve these high activation scores.

In the remainder of this section, I discuss how to tune generic machine learning methods to find suitable candidate treatment locations in social science applications. I recommend four high-level implementation decisions in adapting these methods. First, discretization of geographic space into a fine grid for tractability. Second, *convolutional* neural networks capture the idea that spatial neighborhoods matter in a parsimonious way. Third, incorporating the adversarial task of the discriminator into a classification task for the generator greatly simplifies training. Fourth, data augmentation (rotation, mirroring, shifting) for settings where absolute locations and orientation are irrelevant.

Discretization To tractably summarize the relative spatial locations of individuals and treatment locations, I recommend discretizing geographic space into a fine grid. Discretization provides an approximation that is particularly tractable for the convolutional neural networks recommended below. In principle, future improvements to, for instance, Capsule Neural Networks (Hinton et al., 2011) or other novel methods, may replace this as the preferred architecture and eliminate the need for discretization.

For each grid cell, one can include a count of individuals with residence in the cell, potentially separately for individuals with different values of covariates, as well as average covariate values of the individuals in the cell or other moments of their covariates. Based on the architecture of convolutional neural networks, suggested below, it is typically not necessary to also pre-compute covariates describing the neighborhood of each cell. The convolutional neural network can compute such neighborhood averages if they are helpful for predicting the outcome (here, whether a location is likely to be treated). If the grid is very fine, this discretization retains almost all meaningful information about relative locations. For instance, in the application of this paper, each grid cell has size 0.025mi $\times$ 0.025mi (approximately 40m $\times$ 40m). The discretized grid creates a three-dimensional array: The first two dimensions determine spatial location, and the third dimension enumerates the different covariates that are summarized. Rather than taking the spatial dimensions to be entire regions, I recommend using square cutouts of regions such that the probability of treatment in the approximate center of the cutout is plausibly only affected by individuals and covariates within the cutout.

Convolutional neural networks Convolutional neural networks have been particularly successful at image recognition tasks (cf. Krizhevsky et al., 2012). In image recognition tasks, the input is a 3D array: a 2D grid of pixels, with multiple layers corresponding to the RGB color channels. For spatial treatments, the input also is a 3D array: the 2D spatial grid with layers corresponding to different covariates as described above.

[25] Depending on the network architecture, the activation score can have an interpretation as the posterior (conditional) probability of the location given the covariates in a Bayesian sense. However, in practice it appears preferable to separately estimate propensity scores based on the real locations and the counterfactual locations actually chosen for the analysis. Estimating the propensity score separately allows better balancing of relevant covariates in the sample that is actually used for estimation.

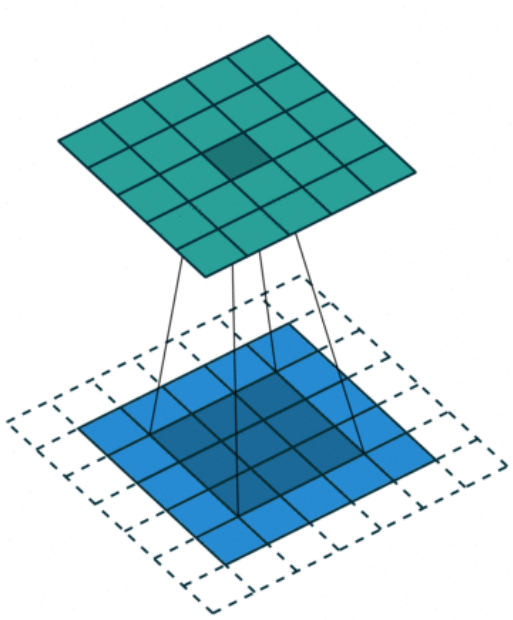


Figure 8: Convolutions in a neural network allow the prediction of a candidate location in a grid cell to depend on the characteristics of neighboring grid cells (up to a user-specified distance). These models remain parsimonious by requiring the *same* "neighborhood scan" to be performed for each grid cell.

Convolutional steps in neural networks generally retain the shape of the 2D grid, but the value of each neuron is a function of the covariates (or neurons) of the previous step not just at the same grid cell, but also the covariates (or neurons) at neighboring grid cells. Figure 8 illustrates this aspect of the convolution operation. However, the particular way in which the neighborhood of a grid cell is averaged is the same for any point in the grid. This makes convolutional layers substantially more parsimonious than fully connected layers, and allows the neural network to capture neighborhood patterns appearing in different parts of a region in a unified way.

In particular, I recommend using at least two convolutions with reasonably large spatial reach. Consider the application in this paper, where grocery stores are spatial treatments and restaurants are outcome units with foot-traffic as the outcome variable. The first convolution allows each grid cell to see what other cells are around it. In the example, the output of the first convolution for a particular grid cell may be: "There are 3 grocery stores nearby, 4 competing restaurants very close, and 10 restaurants within walking distance." The second convolution then uses the information on such neighborhoods to determine whether treatment is likely in a grid cell: "If there are many grid cells nearby (in all directions) containing restaurants or grocery stores facing much competition, this location is probably in the center of a shopping area and reasonably likely to contain another grocery store." Intuitively, the first convolution may measure what is important to the restaurants, while the second convolution translates how that is important for the treatment location choice, mirroring the two-part conditioning in the unconfoundedness assumption (Equation 12) of the previous section.

Adversarial Classification Generative adversarial networks (Goodfellow et al., 2014) are oftentimes difficult to train despite recent advances such as networks with Wasserstein-type criterion function (Arjovsky and Bottou, 2017; Arjovsky et al., 2017). The difficulty arises because the training of generator and

discriminator needs to be sufficiently balanced such that both improve. For instance, if the discriminator early on becomes (close to) perfect at discriminating between the proposals of the generator and the real treatment locations, the gradient for the generator is flat (no progress in any direction) and hence the generator fails to improve. Similarly, if the discriminator is insufficiently flexible, even poor proposals by the generator may pass, such that the false positives are not necessarily similar to the real treatment locations.

In contrast, convolutional neural networks for image classification are much easier to train, and in this case can be adapted to the same task. I therefore recommend to set up the problem of finding candidate treatment locations as a classification task. Specifically, the convolutional neural network takes a 3D input array and "classifies" it into, say, 101 categories, where categories correspond either to the $10 \times 10 = 100$ grid cells in the center of the region (cutout), or an additional "no missing treatment location" category. The distinction from other generation tasks is that here the set of possible outputs is relatively small, for instance the 101 categories described above. In contrast, in image generation there are practically infinitely many possible images that could be generated.

To retain the adversarial nature of the task, I propose simultaneously training the classification on three sets of data, and adding a final fully-connected layer. The three sets of data are as follows: First, regions with at least one real treatment location, but with one treatment location removed. The correct classification of such region data is into the category corresponding to the grid cell where the treatment location was removed. Second, regions with at least one real treatment location, but without any treatment location removed. The correct classification of such region data is into the no missing treatment location category. Third, regions without treatment locations. These are also classified as not missing any treatment location. The output of the convolutional layers is a prediction for each grid cell, of whether it is missing a treatment location. The final fully-connected layer can then combine these location-specific predictions into the categories mentioned above; one corresponding to each of the central grid cells, plus one corresponding not to a grid cell but for cases where no treatment location is missing. The neural network then balances two tasks: a generative task of picking the correct location if a treatment location is missing, predominantly performed by the convolutional layers; and a discriminatory task of deciding whether a treatment location is missing at all, predominantly performed by the final fully-connected layer. This structure retains the attractive interpretation of generative adversarial networks, but is substantially easier to train. It also resembles denoising autoencoders (cf. Vincent et al., 2008), where the removal of a real treatment location represents noise added to the input, with the autoencoder trained to remove the noise, here meaning to add the removed real treatment location. The second and third set of training examples also resemble adversarial examples and adversarial training (Biggio et al., 2013; Szegedy et al., 2013).

The setup as an adversarial task, as well as prediction of categories, additionally is beneficial because it generates draws near the local *modes* rather than the *mean* of the treatment location distribution (cf. Goodfellow, 2016; Lotter et al., 2016). Suppose regions consist of three possible locations in one-dimensional space: 1, 2, and 3. For instance, 2 may be the city center, while 1 and 3 are suburbs on either side of the city. In the data, if a region is treated, treatment always occurs in the suburbs; either at location 1 or at location 3, each with probability 0.5. However, estimating the likely location of the treatment with the familiar mean squared error loss function will estimate the mean of the treatment location distribution, predicting treatment at location 2. In contrast, the adversarial loss function as well as loss functions used for classification tasks, will predict either 1 or 3 because these are most likely to be the correct location, and there is little or no difference between *which* incorrect location or category is chosen. In general adversarial networks, one input to the network is white noise. This noise effectively chooses between the different local

modes of the distribution. In the setup as a classification task proposed here, data augmentation plays a similar role.

Data Augmentation Data augmentation (e.g. Yaeger et al., 1996; Simard et al., 2003; Cireşan et al., 2012; Krizhevsky et al., 2012) serves two closely related purposes. First, rotating, mirroring, and shifting regions, while maintaining relative distances, produces additional, albeit dependent, observations. This is helpful since training neural networks requires large numbers of training samples. Second, these transformations effectively regularize the parameters of the estimated model. One can choose transformations that induce equivariance to rotation, mirroring, and shifts as appropriate for the particular setting. For instance, in many applications in the social sciences, North-South and East-West orientation is irrelevant on a small scale; only the relative distances matter.[26] In particular, suppose an individual who visits a business to the North of her home because it is on the way to work in the North. If the whole space was rotated, the individual equally visits the same business now to the West as it is still on the way to work, now also rotated to be to the West of her home. In image classification, the use of similar data augmentation is common and often associated with greater generalizability of the learned models.

Shifting the entire grid has two further desirable effects: First, if one imposes a continuous shift of relative coordinates in combination with a fixed grid, the exact discretization becomes less relevant. The *average* (across draws from the shift distribution) distance in grid cells between two observations becomes directly proportional to their actual distance. Second, the location of an observation *within* a grid cell is no longer fixed. This is attractive because the classification is not actually informative of whether the candidate treatment location is at the center or towards the edge of a grid cell. With a continuous shift of the observations, the center of the grid cell points to different absolute locations depending on the shift. One can then average over several realizations of the shift to reduce the influence of the particular translation of grid cell to absolute location.

There are at least two notable alternatives or complements to data augmentation in the machine learning literature. First, spatial transformer networks (Jaderberg et al., 2015) attempt to estimate a rotation or other transformation that makes the subsequent classification task as easy as possible. Second, some recent work considers imposing the desired in- and equivariance properties on the convolution kernel or adding layers to the network that effectively average the appropriate kernel coefficients (Cohen and Welling, 2016; Dieleman et al., 2016; Gens and Domingos, 2014; Dudar and Semenov, 2018; Dzhezyan and Cecotti, 2019). However, data augmentation and other regularization techniques may achieve the first order gains implied by these properties (Srivastava et al., 2014; Kauderer-Abrams, 2017; Yang et al., 2019), and may hence suffice for the purpose of finding appropriate counterfactual treatment locations. One can also inspect the models to assess the implied degree of equi- and invariance (Goodfellow et al., 2009; Zeiler and Fergus, 2014; Lenc and Vedaldi, 2015).

6.3 Estimating Propensity Scores

Suppose candidate treatment locations $\mathbb{S}$ are known (in all regions), for instance as output of the convolutional neural network classification task described in the previous section. The remaining challenge in implementing the methods proposed in this paper is the estimation of the "propensity score" $\Pr(s \in \xi)$. I briefly sketch

[26] Applications in environmental economics are a notable exceptions if, for instance, wind direction is relevant. In such cases, rotation hinders the ability of the model to capture patterns due to e.g. wind consistently blowing from one direction, and may require inclusion of wind direction in estimation. The choice of appropriate data augmentation is therefore application specific.

propensity score estimation in two canonical settings: a fixed number of realized treatment locations per treated region (often just one realized location), and independent Bernoulli trials determining realization of treatment at candidate locations. When treatment probabilities are estimated rather than set by design, doubly-robust and sample-splitting approaches may reduce the impact of estimation error.

Fixed Number of Realized Treatment Locations Suppose there are a fixed number of realized treatment locations per treated region, as in Section 4. Then the problem of propensity score estimation resembles discrete choice modeling: There are $|\mathbb{S}_j|$ discrete alternatives in region j, a fixed number of which is realized. Note that for this estimation, only regions with realized treatment, $W_j = 1$, are used. Untreated regions do not contain relevant information for this task. However, the estimated discrete choice model is used to predict propensity scores for candidate locations in both treated and control regions. See, for instance, Greene (2009) for an overview of estimation methods.

It is particularly important not to treat this setting as one with independent Bernoulli trials when regions and candidate location characteristics are heterogeneous. Suppose there are two types of candidate locations and two types of regions. The first type of candidate locations is in high income neighborhoods, while the second type is in low income neighborhoods. Within a region, treatment is substantially more likely to occur at a candidate location in high income neighborhoods. However, the first type of region has many high income neighborhoods and very few low income neighborhoods, while the second type has a small number of each type of neighborhoods. Then the marginal probability of realization (unconditional on region, but conditional on neighborhood type) might be lower for candidate locations in high income neighborhoods because they, on average, face more competition for the fixed number of realized treatment locations in their regions. Hence, a researcher estimating propensity scores as if assignment was independent will systematically overestimate the probability of treatment in low income neighborhoods, and underestimate the probability of treatment in high income neighborhoods. Discrete choice modeling resolves this issue by specifically considering the choice set; that is, the particular types of candidate locations available in each region.

When data from many regions are available and regions are heterogeneous in their characteristics, one can additionally estimate the probability that a region is treated. Based on Assumption 3, one may estimate the region propensity scores π_j by (logistic) regression of the region treatment indicator W_j on region characteristics Z_j. When no separate covariate data is available, one can create region characteristics based on the number of individuals in the region, $|\mathbb{I}_j|$, the number of candidate locations in the region, $|\mathbb{S}_j|$, the area of the region (based on the locations r_i of individuals), or population density. Even when these characteristics did not directly feature in the experimental design, adjusting for such pre-treatment covariate may improve estimates in finite samples (cf. Hahn, 1998; Hirano et al., 2003; Frölich, 2004b).

Independent Bernoulli Trials When treatment assignment is independent across locations, as in Section 5.1.2 by Assumption 7, propensity score estimation for spatial treatments is similar to propensity score estimation for individual-level treatments. Logistic regression is a simple option. Each candidate treatment location $s \in \mathbb{S}$ is a separate, independent, observation. With logistic regression, regress the indicator $W_s \equiv \mathbb{1}\{s \in \xi\}$ on covariates Z_s that describe the neighborhood of candidate location s, as well as on (moments of) the characteristics of individuals near location s, $(X_i)_{i:\, d(s,r_i)=d}$ for all distances of interest d.

Using Estimated Propensity Scores In observational studies, the propensity score is typically estimated by the methods above rather than known. The design-based analysis proposed in this paper studies the behavior of estimators in the thought experiment of an experiment based on these estimated locations and

propensity scores, effectively conditioning the analysis on them. Even when the propensity score is known, there may be benefits from using estimated propensity scores for parts of the analysis as in experiments with individual-level treatments (cf. Hahn, 1998; Hirano et al., 2003; Frölich, 2004b).

To reduce the effect of estimation error from this first-stage propensity score estimation, I also report point estimates based on sample splitting and a doubly-robust moment condition (e.g. Chernozhukov et al., 2018) in the application in Section 7. When, for instance, estimated propensity scores are close to 0 or 1, the inverse propensity score weighting estimators proposed in this paper may perform poorly (cf. Frölich, 2004a; Busso et al., 2014) because small estimation errors in the propensity scores have large effects on the weights when denominators are close to zero. Note, however, that currently available inference for doubly-robust estimators is typically sampling-based and therefore does not have the same interpretation as the design-based inference of this paper. Existing inference procedures may therefore not take the spatial correlations, in exposure to treatment and in outcomes, into account correctly from a design-based perspective. Furthermore, while existing inference procedures may take into account propensity score estimation, additional work is required to also address uncertainty in the locations of the estimated counterfactual locations.[27] In practice, empirical researchers might be particularly concerned about the robustness of estimates to particular, conceptually irrelevant, design choices and random seeds used in the generation of counterfactual treatment locations. In the application, I therefore show that estimates from differently parameterized neural networks, and counterfactual treatment locations generated from different random seeds, do not produce substantively different point estimates, even though the exact counterfactual locations used may differ. Despite issues of design-based inference, estimators based on doubly-robust moments and sample splitting likely still substantively reduce the effect of error due to propensity score (and outcome model) estimation, and I report point estimates from such estimators in the application.

7 Application: Foot Traffic in Times of COVID-19

Did grocery stores bring additional foot-traffic to nearby restaurants during COVID-19 shelter-in-place policies in April of 2020? Grocery stores may have causal effects on the number of customers to nearby restaurants if they draw customers into the shopping and business area. In particular during the first few weeks of COVID-19 shelter-in-place policies, when individual mobility was greatly reduced, getting groceries may have been one of the few trips still made. If grocery store customers are more likely to stop by coffee shops or restaurants for pick-up orders right before or after getting groceries, restaurants and similar businesses may receive more foot-traffic if there is a grocery store nearby. Large department stores serving as "anchor stores" of shopping malls may play a similar role in normal times. Relatedly, Jia (2008) studies the effects of new Wal-Mart stores on existing businesses. Athey et al. (2018) study the effect of restaurant closings on nearby restaurants.

In the framework of this paper, grocery stores are the spatial treatments, restaurants are the (outcome) individuals, and foot-traffic (the number of customers) is the outcome of interest. I argue that the inner ring vs. outer ring empirical strategy used in recent studies is unattractive in this setting: Its identifying assumptions are not credible, and it requires discarding the majority of the sample for practical reasons. I show how to implement the methods proposed in this paper, and argue that the control groups these methods are based on are preferable to outer ring control groups.

[27] The difficulty of taking into account which counterfactual locations are used resembles issues in inference for matching estimators of treatment effects, where standard procedures condition on the matches and take a sampling based view of the outcomes of individuals (cf. Abadie and Imbens, 2006, 2008, 2011, 2016).

The average treatment effect of interest is identified by an ideal experiment where some grocery store locations are randomly closed during COVID-19 shelter-in-place policies. Specifically, take a restaurant i near a grocery store at location s. What is the difference between the number of customers of restaurant i during the COVID-19 shelter-in-place policies when there is a grocery store at location s, and the number of customers of the same restaurant i if there was no grocery store at location s, holding fixed the locations of all other businesses and grocery stores? In the notation of this paper, if S are the locations of other grocery stores, the treatment effect of interest is $\tau_i(s) = Y_i(S \cup \{s\}) - Y_i(S)$. This effect is distinct from fixing a spatial location near a grocery store, and considering the difference in the outcome (during COVID-19) of the restaurants that exists at this point in space when there is a grocery store nearby, and the outcome (also during COVID-19) of the, possibly different, restaurant that would have been at the same location, had there *never* been a grocery store nearby.

The effects of grocery stores on nearby restaurants are informative about several questions. Do grocery stores have (positive) externalities on other businesses? If so, should mall operators subsidize grocery stores through lower rent such that they internalize these externalities, to support other businesses in the mall? In the context of pandemics, are grocery stores likely choke points leading to bunching of customers at nearby restaurants instead of spreading out across all restaurants, increasing the risk of infections? Alternatively, grocery stores may resolve a coordination problem: Suppose that the overall reduced number of restaurant customers is insufficient to operate restaurants profitably or with reduced loss when spread across all restaurants. Grocery stores may then help to resolve a coordination problem between restaurants, by focusing potential restaurant customers on the nearby restaurants.

I use SafeGraph data on the number of customers of each business in the week starting April 13, 2020. SafeGraph (2019) describes the algorithm used for attributing visits to businesses. Generally, pick-up orders as well as outside dining are likely picked up by the algorithm as long as a customer's smartphone sends location data at the point of interest for more than one minute. For errors in attribution to matter in the application of this paper, they need to correlate with the presence or absence of nearby grocery stores. I restrict the sample to businesses in the area between San Francisco and San Jose in the San Francisco Bay Area, as highlighted in Figure 9. There are 167 grocery and convenience stores (triangles in the figure), as well as 1627 distinct restaurants (black circles) within 0.5 miles from any of the grocery stores (real or counterfactual). See Appendix B.1 for details on sample construction.

The outcome of interest is the inverse hyperbolic sine of visits to restaurants, with visits as measured by SafeGraph. For external validity, that is to interpret the percentage point effect on the number of SafeGraph-tracked customers as the overall effect, one needs to assume that SafeGraph's sample selection is orthogonal to the presence and absence of grocery stores. Otherwise, the estimates retain internal validity as the effects on the number of SafeGraph-tracked customers to these restaurants, but may not be informative about the total number of visits. The inverse hyperbolic sine allows for zero visits, and effects on it can be transformed into elasticity estimates similar to $\log(y)$ or $\log(y+1)$ specifications (see Bellemare and Wichman, 2020, for a discussion).[28]

7.1 Inner vs. Outer Ring

Figure 10 illustrates why comparisons between observations on an inner ring and observations on an outer ring around a strategically chosen location are often not attractive. Here, businesses (blue circles) on the

[28]The inverse hyperbolic sine is defined as $\text{arcsinh}(y) \equiv \ln(y + \sqrt{y^2+1})$. Hence $\text{arcsinh}(0) = 0$, $\text{arcsinh}(1) \approx 0.9$, $\text{arcsinh}(2) \approx 1.4$, and $\text{arcsinh}(y) \approx \ln(y) + 0.7$ if $y \geq 3$.

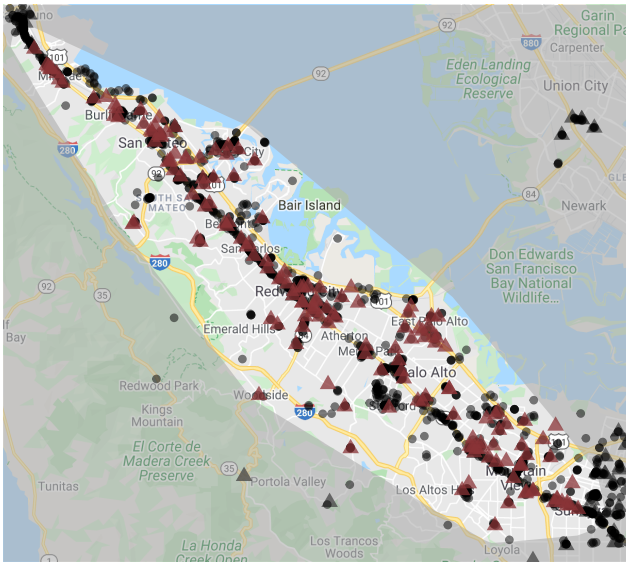


Figure 9: The sample includes businesses in the San Francisco Bay Area between San Francisco and San Jose. The locations of real grocery and convenience stores are marked by solid red triangles. Restaurants are marked by black circles. The black triangles are grocery stores outside the main study area; their location is considered fixed and restaurants near them are not part of the estimation procedure. In total, there are 167 grocery and convenience stores, as well as 1627 distinct restaurants that were open as of January 2020 within 0.5 miles from any of the grocery stores (real or counterfactual within the main study area).

inner ring are at a distance of 0.1 ± 0.025 miles from the grocery store (orange triangle), while businesses on the outer ring are at a distance of 0.25 ± 0.025 miles from the same grocery store. While inner ring businesses are part of the same strip mall, outer ring businesses are outside of the primary shopping areas. Interpreting differences in outcomes for these two groups of businesses as causal effects requires assuming that outer ring businesses are unaffected by treatment and have similar outcomes as inner ring businesses in the absence of treatment. Generally, distance from treatment often correlates with many other variables (Kelly, 2019). With small numbers of grocery stores (see below), the mode of average treatment effect estimates of an inner ring vs. outer ring empirical strategy may not be close to the true average effect, even if the locations of grocery stores were random. This arises due to spatial correlations in outcomes even in the absence of treatments (cf. Lee and Ogburn, 2021, in a network setting). See Section 2 for a concrete example suggesting that except for knife-edge cases, such estimators are only unbiased if the outcome surface is flat.

In this application, the absence of pre-trends in event study plots would not be particularly informative about the validity of the underlying assumption. With panel data, restaurants on inner and outer rings are allowed to have different average levels of customers, but trends (in the inverse hyperbolic sine) of the number of customers must be parallel. However, even if panel data suggested that trends between inner and outer ring restaurants were indeed parallel pre COVID-19, one may question whether this is informative about trends in (potential) outcomes during COVID-19 shelter-in-place policies. Given the dramatic decrease in customers for all restaurants, it is questionable that this decrease would have occurred in parallel with only a shared additive shift (in the inverse hyperbolic sine) for inner and outer ring restaurants in the absence of treatment.

Furthermore, the estimand of a difference in differences estimator in this setting is the additional effect of grocery stores on nearby restaurants during COVID-19 on top of any effects that may have already existed

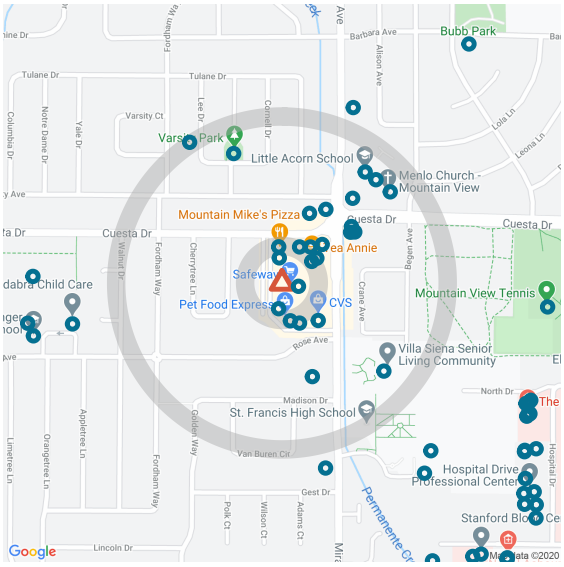


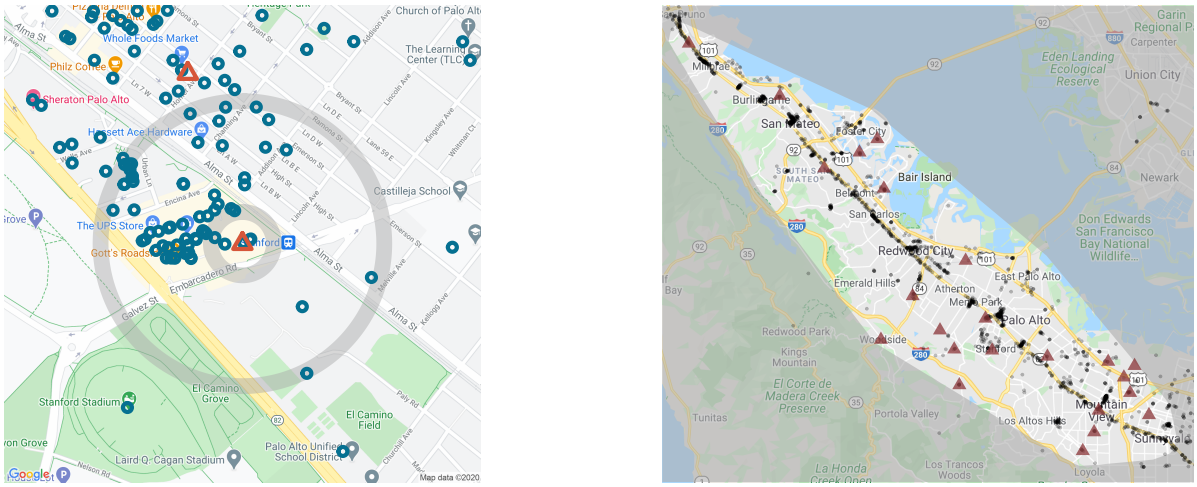
Figure 10: The comparison of businesses on an inner vs. outer ring around a particular grocery store. The grocery store is marked by an orange triangle in the center of the figure. Other businesses are small blue circles. Businesses on the gray inner ring, at a distance of 0.1 ± 0.025 miles, are primarily in strip malls, while businesses on the gray outer ring, at a distance of 0.25 ± 0.025 miles, are away from these main shopping areas.

pre COVID-19. Even if the parallel trends assumption was credible, this estimand differs from the estimand of interest described above. The difference in differences estimand can be negative even though the effect of grocery stores on nearby restaurants is positive during COVID-19 if the effect of grocery stores *pre* COVID-19 was also positive but larger in magnitude, for instance due to differences in the scale of the number of customers.

Finally, in most instances, restaurants on the outer ring around a grocery store are not actually far away from grocery stores ("untreated"), as illustrated by panel (a) of Figure 11. Here, some of the businesses on the outer ring centered around the grocery store in the center of the figure are very close to a second grocery store to the North. Applying the inner vs. outer ring estimator in this setting therefore requires restricting the sample to the neighborhoods of the few grocery stores that are sufficiently far away from other grocery stores. Specifically, to guarantee the absence of interfering grocery stores for an outer ring "no effect" distance of 0.250 miles, only grocery stores with no other grocery store within 2×0.250 miles can be used. Panel (b) of Figure 11 shows the locations of the remaining 23 grocery stores. Compared to Figure 9, these grocery stores are in more remote locations, with noticeably fewer nearby restaurants. While the average treatment effect of grocery stores in such locations may continue to be of interest, it is plausibly distinct from the treatment effect in areas with higher population or business density.

7.2 Quasi-Experimental Estimators for Spatial Treatments

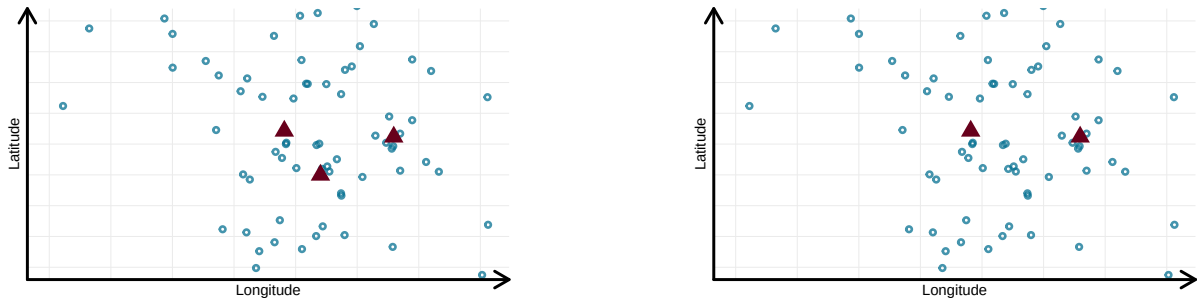
This application follows the framework of Section 5.1 for a single contiguous region with independent treatment assignment. The key idea behind identification for the proposed methods is that the location of a grocery store is as good as random between candidate locations with similar numbers and industries of nearby businesses. Figure 12 shows an example of an ideal comparison where the only difference between the (parts of the) regions is the absence of the bottom-most grocery store, and all other relative locations are the same.



(a) Example of Interference

(b) Sample of Isolated Grocery Stores

Figure 11: Panel (a) shows an example of a grocery store (triangle in the center) with a second "interfering" grocery store (triangle towards the top) nearby. Some businesses on the outer ring are close to (treated by) this second grocery store and therefore not a valid control group. Panel (b) shows that restricting the sample to the 23 (out of 167) grocery stores without interference leads to a sample selected heavily towards less business-dense areas compared to the overall sample shown in Figure 9.



(a) treated region

(b) control region

Figure 12: To estimate the effect of the bottom treatment location (orange triangle) in the region of panel (a), the task is to find another region with similar relative locations of businesses but missing the particular grocery store, as in panel (b).

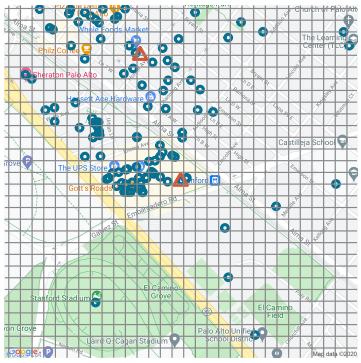
The approach I propose for observational data proceeds in two steps: First, it finds good "matches" for each grocery store; that is, locations without a grocery store that are similar in terms of the number, types, and relative locations of other businesses and grocery stores. Second, estimate treatment effects assuming the matched data resemble the ideal experiment of randomizing grocery stores between the real and counterfactual candidate treatment locations. I recommend inverse propensity score weighting estimators based on the results of sections 4 and 5.1. Conceptually similar combinations of matching or stratification and propensity score weighting or regression adjustments have been advocated for by Abadie and Imbens (2011), Imbens and Rubin (2015, ch. 17), and Kellogg et al. (2021), among others.

7.2.1 Predicting Grocery Store Locations

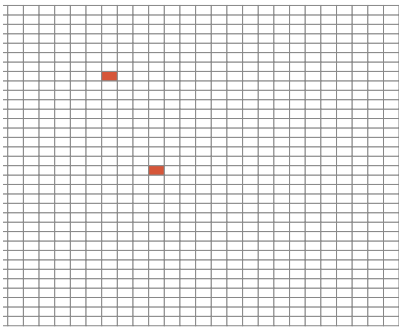
The grocery store location prediction following Section 6.2 discretizes the South Bay region into a fine grid and aggregates characteristics of businesses in each grid cell. Figure 13 illustrates the discretization for the surroundings of an example grocery store, see panel (a). For each grid cell, record the number of grocery stores as in panel (b). Other characteristics of each grid cell, for instance the number of businesses by industry are recorded in similar grids as in panel (c).

Based on this discretization, I use convolutional neural networks as described in Section 6.2 to find counterfactual candidate grocery store locations that are indistinguishable from the real grocery store locations. Details on the implementation are given in Appendix B.2. Since the method can find a very large number of counterfactual grocery store locations, I use matching (based on node activation, implied probability, and a propensity score estimate) to narrow the sample down to a smaller but more balanced sample of real and counterfactual grocery store locations. To estimate propensity scores in this setting, I assume that grocery store openings are independent decisions at each location, Assumption 7. In practice, this assumption is primarily relevant at the margin of opening (or closing) additional grocery stores relative to the existing grocery stores. Since there are neighborhoods similar in other businesses but differing in the number of grocery stores, this assumption may offer a reasonable approximation.

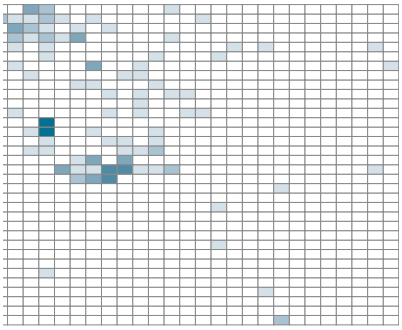
The inverse probability weighted real and counterfactual grocery store locations are similar in everything except their exposure to real grocery stores, which differs by one additional grocery store. Figure 14 shows that the exposure to treatment is as intended: Suppose we estimate the treatment effect on restaurants at a distance of 0.075 – 0.1 miles; top row, fourth panel from the left. Among these restaurants, those near real grocery store locations are exposed to the same number of grocery stores on average at all distances *except* at 0.075 – 0.1 miles as those near counterfactual locations. Hence, the estimator plausibly estimates the effect of one additional grocery store at a particular distance, holding fixed the number of grocery stores at other distances. Furthermore, the composition of nearby businesses is similar between real and counterfactual grocery store locations at any distance. Figure 15 shows that, holding fixed distance from a real or counterfactual grocery store location, the neighborhoods of "treated" restaurants (near real locations) have a similar business composition as the neighborhoods of "control" restaurants (near counterfactual locations). A comparison of restaurants on inner vs. outer rings around real grocery stores would lead to systematic differences in the industry composition of the neighborhood of treated and control restaurants. Overall, these figures lend credibility to the treatment effect estimators proposed in this paper. Treated and control restaurants are alike, except for a single additional grocery at the intended distance.



(a) fine grid overlay on map



(b) grocery stores in each cell



(c) other businesses in each cell

Figure 13: The neighborhoods of each grocery store are placed in a fine grid (0.025mi $\times$ 0.025mi) as shown in panel (a). For each cell, the number of grocery stores is shown in panel (b), and the number of other businesses in panel (c). In practice, I create multiple grids as in panel (c), each capturing the number of businesses of a different industry. Additionally, one could add similar grids containing for instance average age and 0.75 quantile of earnings of customers of businesses in each cell, or residents living in the census tract covering the cell. The prediction of grocery store locations then uses augmented data, where one grocery store is removed from the grid in panel (b), and the data from grids as in panel (c) is used together with the augmented grid of panel (b) by a convolutional neural network to predict the location of the removed grocery store.

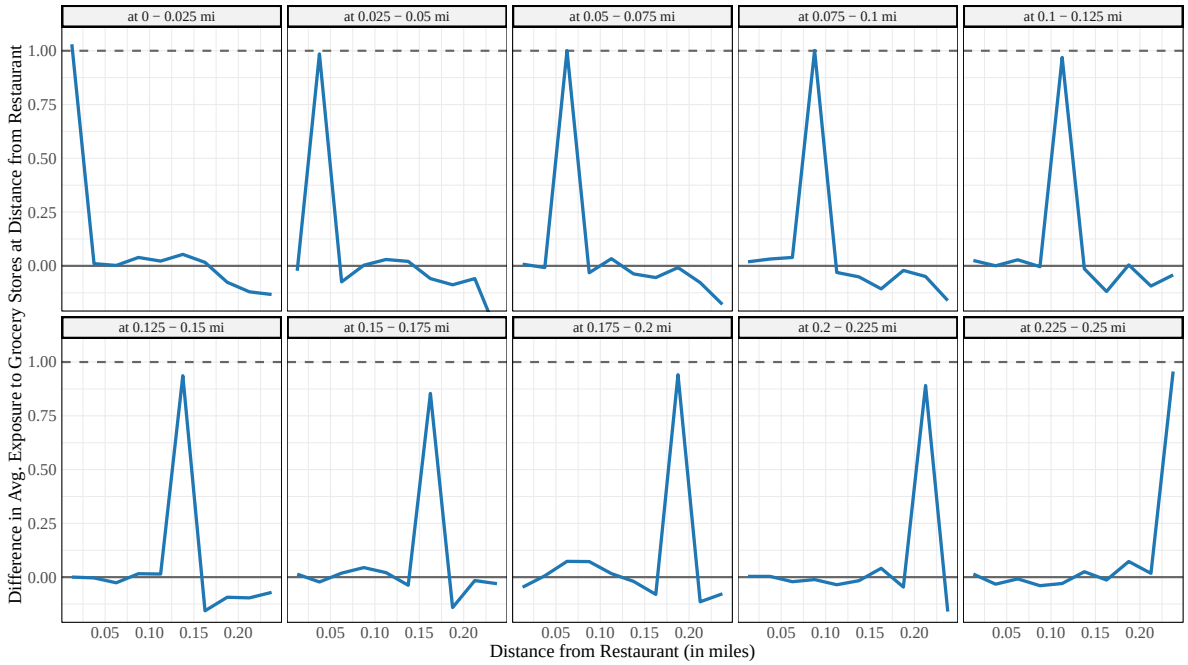


Figure 14: Each panel focuses on the restaurants used to estimate treatment effects at a specific distance. The outcome (vertical axis) shown is the difference in the average number of grocery stores by distance from the restaurant (horizontal axis) between restaurants near real and counterfactual grocery store locations. In each panel, the difference in exposure is near 0 except at the distance at which "treated" and "control" restaurants are meant to differ by one grocery store.

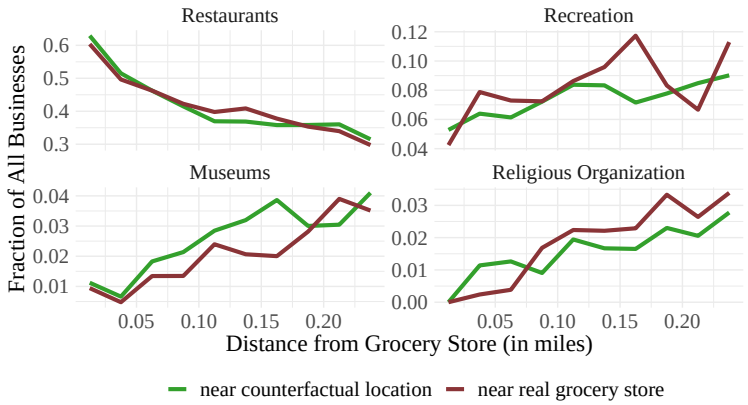


Figure 15: The composition of businesses near real and counterfactual grocery store locations is similar. The figure plots, by distance from real or counterfactual grocery store location, the fraction of businesses that fall into one of four different industries. It is encouraging that counterfactual grocery store locations mimic the business composition pattern across distance of real grocery store locations. Since the fraction of businesses that are, for instance, restaurants decreases meaningfully across distances, the neighborhoods and competition experienced by restaurants on inner vs. outer rings would differ systematically. In contrast, focusing on restaurants at a fixed distance from real vs. counterfactual locations leads to a comparison of restaurants in similar neighborhoods, as measured by business composition.

7.2.2 Estimating Treatment Effects

Given candidate treatment locations and propensity scores, I estimate treatment effects with the estimator of Section 5.1.

To interpret the estimated effect as the average effect of opening single grocery stores, rather than the marginal effect of adding a grocery store to existing exposure, one can make the additivity Assumption 5. Additivity may be plausible if each additional grocery store brings new customers into an area. During COVID-19, customers may reduce the number of different grocery stores they shop at to limit their exposure. Furthermore, there is differentiation in the grocery store market: The customers at discount grocery outlets may be distinct from the customers at Whole Foods.

Table 1 reports the estimates and standard errors for distances up to 0.25 miles, as well as the percentage effect relative to the mean number visits of the control restaurants at that distance. Figure 16 visualizes the effect by distance under the inverse hyperbolic sine transformation. Estimates for alternative transformations such as the logarithm of visits + 1 or the logarithm conditional on a strictly positive number of visits are very similar and not reported. At very short distances, restaurants have almost three times the visits (+187%) if they are near real grocery stores rather than near counterfactual grocery store locations. If the grocery store is 0.05 or more miles away, it has no more effect on the businesses.

The treatment effect estimates are relatively noisily estimated because grocery stores close to one another create (potential) interference that reduces the effective sample size. To define the exposure mappings underlying the standard errors shown in Figure 16 and reported in Table 1, I assume that grocery stores have no effect on visits to restaurants that are more than 0.15 miles away. The marginal treatment effect estimates past this distance are economically small and (under the assumption that there indeed is no effect past 0.15 miles) statistically indistinguishable from zero, suggesting the data are consistent with this assumption. The estimator uses 167 real grocery store locations as well as well as 308 counterfactual grocery store locations. If each (real as well as counterfactual) grocery store location and its nearby restaurants were in a separate, independent region such that there was no interference, standard errors would be less than half compared to the standard errors that do take into account possible interference. This implies that interference here causes the effective sample size to be only a quarter of the actual sample size. In settings with interference, there hence is a trade-off when choosing counterfactual locations: use more counterfactual locations to have more flexibility to generate (average) covariate balance and to reduce the variance of the estimate of the control mean, or use fewer counterfactual locations to reduce the extent of interference.

I also estimate the effect of grocery stores on visits to restaurants using a doubly-robust moment (e.g. Chernozhukov et al., 2018). The natural extension of the ATT-moment to the spatial treatment setting with interference is

$$\psi_{\tau(d)} = \mathbb{1}\{s \in \xi\}\Big(Y - \tau(d) - \mu\big(X, \xi \setminus \{s\}\big)\Big) - \frac{e(s)\big(1 - \mathbb{1}\{s \in \xi\}\big)\big(Y - \mu(X, \xi)\big)}{(1 - e(s))}$$

where $E(\psi_{\tau(d)})$ averages over all combinations of candidate grocery store locations s and restaurants i satisfying $d(s, r_i) \approx d$. $\mathbb{1}\{s \in \xi\}$ plays the role of the "treatment indicator." The function $\mu(X, S)$ gives the expected outcome (inverse hyperbolic sine of number of visits) for a restaurant with covariates X, including neighborhood characteristics, when there are grocery stores at locations S. For a restaurant near a real grocery store, the conditional mean function is evaluated in the absence of the nearby grocery store s, $\xi \setminus \{s\}$, with the parameter of interest, $\tau(d)$, capturing the difference between actual outcome and expected outcome in the absence of the nearby grocery store. For restaurants near a counterfactual candidate location s,

Table 1: Estimated effects on the number of visits to restaurants using different estimators. The first panel uses the inverse probability weighting estimators for spatial experiments proposed in this paper. The second panel uses a doubly-robust version of the spatial experiment estimator. For each panel, "IHS" refers to the effect in inverse hyperbolic sine unit, "level" to effects in levels, and "visits > 0" to the extensive margin of observing at least one visit to the business in the SafeGraph data. Standard errors are given in parentheses. Rows labeled "percent incr." show the percent increase (or decrease) relative to the mean visits of the control restaurants at that distance.

Distance:	0.000 mi – 0.025 mi	0.025 mi – 0.050 mi	0.050 mi – 0.075 mi	0.075 mi – 0.100 mi	0.100 mi – 0.125 mi	0.125 mi – 0.150 mi	0.150 mi – 0.175 mi	0.175 mi – 0.200 mi	0.200 mi – 0.225 mi	0.225 mi – 0.250 mi
Spatial Experiment Estimators:										
IHS:	1.05	0.41	0.25	0.09	0.16	0.01	-0.04	0.00	-0.05	0.13
	(0.48)	(0.40)	(0.42)	(0.38)	(0.43)	(0.36)	(0.35)	(0.39)	(0.42)	(0.39)
percent incr.:	187	52	29	9	17	1	-4	0	-5	14
level:	13.07	3.75	-1.35	0.68	1.84	0.83	-0.36	-0.50	-6.69	0.35
	(7.70)	(6.32)	(11.46)	(7.35)	(4.74)	(5.60)	(6.19)	(6.57)	(11.99)	(7.56)
percent incr.:	116	30	-9	5	16	7	-3	-4	-35	2
visits > 0:	0.07	0.04	0.00	-0.01	0.03	-0.02	-0.01	-0.01	0.00	0.01
	(0.07)	(0.05)	(0.05)	(0.06)	(0.07)	(0.06)	(0.06)	(0.07)	(0.07)	(0.06)
Doubly-Robust Estimators for Spatial Experiments:										
IHS:	1.16	0.45	0.27	0.03	0.12	-0.03	-0.08	0.01	-0.06	0.08
percent incr.:	221	57	31	3	13	-3	-8	1	-6	8
level:	14.97	2.86	0.28	-0.28	0.07	-0.72	-2.33	-0.64	-3.67	0.23
percent incr.:	132	23	2	-2	1	-6	-17	-5	-19	2
visits > 0:	0.06	0.04	0.01	0.00	0.03	-0.01	-0.01	-0.01	0.00	0.02

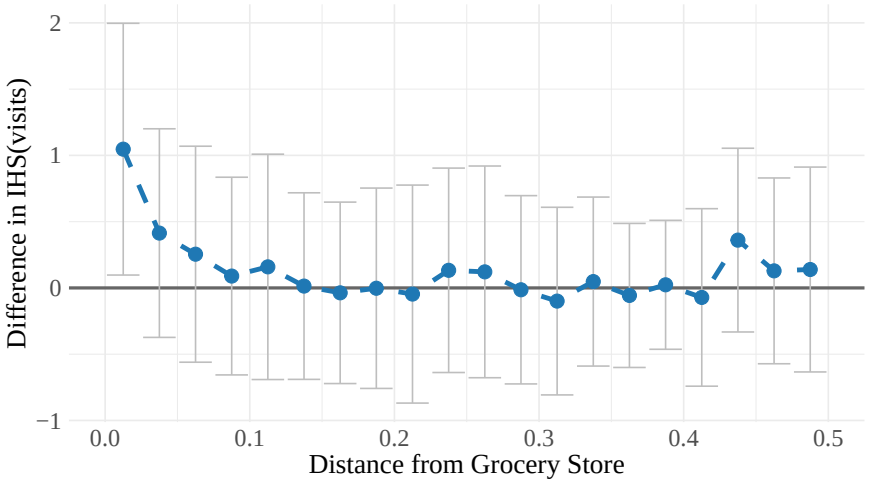


Figure 16: Treatment effect estimates by distance. The vertical axis shows the difference in the inverse probability weighted mean inverse hyperbolic sine (similar to log) of visits to restaurants near real grocery stores and restaurants near counterfactual grocery store locations. There is a substantial estimated treatment effect at very short distances of up to 0.025 miles, and no meaningful difference between treated and control businesses at larger distances.

the conditional mean function is evaluated at the background treatment exposure level S. The propensity score $e(s)$ gives the probability that there is a real grocery store at candidate location s, conditional on characteristics of the neighborhood of s. This moment function satisfies the Neyman orthogonality condition of Chernozhukov et al. (2018). Relative to the spatial experiment estimator, which treats the propensity score as known, this estimator has the advantage of reducing the impact of small errors in the estimated propensity score through orthogonalization.

Overall, the inverse propensity score weighting estimator and the doubly-robust estimator yield very similar results as shown in Table 1 above. Grocery stores have an economically large positive effect during COVID-19 shelter-in-place policies only at short distances of less than 0.025 miles. Intuitively, grocery store customers do visit nearby restaurants and coffee shops, but are unlikely to walk for more than a couple of minutes from the grocery store location.

8 Conclusion

The aim of this paper is to argue that leveraging quasi-random variation in the location of spatial treatments is both conceptually attractive and feasible in many settings in practice.

I propose a framework and experimental approach for estimating the effects of spatial treatments. This approach uses random variation in the realized locations of the spatial treatments for causal identification. I argue that an alternative empirical strategy, which compares inner and outer rings around realized treatment locations, commonly used in practice is not justified by the same random variation, but instead identifies causal effects only under partly conflicting nonparametric, or functional form assumptions.

To operationalize the (quasi-) experimental approach with observational data, I propose a machine learning method to find counterfactual locations where the treatment could have occurred but did not. The proposed method specifically leverages that neighborhood characteristics are predictive of both the location of treatments and the outcomes of individuals. Convolutional neural networks learn this rich spatial dependence

structure encoding relevant institutional features from the data. I incorporate the appealing properties of generative adversarial networks in a classification problem that leads to much simpler training in practice, similar to denoising autoencoders.

I illustrate the proposed methods in an application studying the causal effects of grocery stores on foot-traffic to nearby restaurants during COVID-19 shelter-in-place policies. I find substantial positive externalities of grocery stores on nearby restaurants: Foot-traffic to restaurants within a couple of minutes walking from real grocery stores is almost double the foot-traffic to restaurants that are in similar locations but without a nearby grocery store. This effect is very local; there is no discernible effect at larger distances, suggesting grocery stores may concentrate consumers into a relatively small number of restaurants at particularly convenient locations.

Several key questions remain for future research. In some settings, the spatial treatment is endogenous, but geographic characteristics which are continuous in space are available as plausibly exogenous instruments (cf. Feyrer et al., 2017, 2020; James and Smith, 2020). It is unclear how to construct powerful instruments from such geographic characteristics and incorporate them in the causal framework of this paper. In this paper, I also assume that there is no migration response to the treatment. To allow for migration, one could either focus on outcomes at fixed geographic locations instead of outcomes of fixed individuals or embrace a local average treatment effect (Angrist et al., 1996) interpretation with a large number of compliance types if individuals move to different distances from treatment. The analysis in this paper is focused on estimating (potentially weighted) average treatment effects. In practice, decision makers may often be more interested in the optimal location for the spatial treatment. What assumptions allow credible estimation, or prediction, of the treatment effect for locations that were not candidate locations in the randomization of the (quasi-) experiment?

References

Abadie, A., S. Athey, G. W. Imbens, and J. M. Wooldridge (2017). When should you adjust standard errors for clustering? *NBER Working Paper Series* (24003).

Abadie, A., S. Athey, G. W. Imbens, and J. M. Wooldridge (2020). Sampling-based vs. design-based uncertainty in regression analysis. *Econometrica 88*(1), 265–296.

Abadie, A., A. Diamond, and J. Hainmueller (2010). Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *Journal of the American Statistical Association 105*(490), 493–505.

Abadie, A. and G. W. Imbens (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica 74*(1), 235–267.

Abadie, A. and G. W. Imbens (2008). On the failure of the bootstrap for matching estimators. *Econometrica 76*(6), 1537–1557.

Abadie, A. and G. W. Imbens (2011). Bias-corrected matching estimators for average treatment effects. *Journal of Business & Economic Statistics 29*(1), 1–11.

Abadie, A. and G. W. Imbens (2016). Matching on the estimated propensity score. *Econometrica 84*(2), 781–807.

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

Aliprantis, D. and D. Hartley (2015). Blowing it up and knocking it down: The local and city-wide effects of demolishing high concentration public housing on crime. *Journal of Urban Economics 88*, 67–81.

Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association 91*(434), 444–455.

Arjovsky, M. and L. Bottou (2017). Towards principled methods for training generative adversarial networks. *arXiv preprint arXiv:1701.04862*.

Arjovsky, M., S. Chintala, and L. Bottou (2017). Wasserstein gan. *arXiv preprint arXiv:1701.07875*.

Aronow, P. M., D. P. Green, and D. K. K. Lee (2014). Sharp bounds on the variance in randomized experiments. *The Annals of Statistics 42*(3), 850–871.

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

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

Athey, S. (2018). The impact of machine learning on economics. In *The Economics of Artificial Intelligence: An Agenda*, Chapter 21, pp. 507–547. University of Chicago Press.

Athey, S., D. Blei, R. Donnelly, F. Ruiz, and T. Schmidt (2018). Estimating heterogeneous consumer preferences for restaurants and travel time using mobile location data. *AEA Papers and Proceedings 108*, 64–67.

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

Athey, S., B. Ferguson, M. Gentzkow, and T. Schmidt (2019). Experienced segregation. *Stanford University Working Paper*.

Athey, S. and G. W. Imbens (2019). Machine learning methods that economists should know about. *Annual Review of Economics 11*, 685–725.

Athey, S., G. W. Imbens, J. Metzger, and E. Munro (2021). Using wasserstein generative adversarial networks for the design of monte carlo simulations. *Journal of Econometrics*.

Athey, S., G. W. Imbens, J. Metzger, and E. M. Munro (2019). Using wasserstein generative adversarial networks for the design of monte carlo simulations. *NBER Working Paper Series* (26566).

Autor, D. H., D. Dorn, and G. H. Hanson (2013). The china syndrome: Local labor market effects of import competition in the united states. *American Economic Review 103*(6), 2121–68.

Bai, Y., A. Shaikh, and J. P. Romano (2019). Inference in experiments with matched pairs. *University of Chicago, Becker Friedman Institute for Economics Working Paper* (2019-63).

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

Basse, G., A. Feller, and P. Toulis (2019). Randomization tests of causal effects under interference. *Biometrika 106*(2), 487–494.

Bayer, P., S. L. Ross, and G. Topa (2008). Place of work and place of residence: Informal hiring networks and labor market outcomes. *Journal of Political Economy 116*(6), 1150–1196.

Bellemare, M. F. and C. J. Wichman (2020). Elasticities and the inverse hyperbolic sine transformation. *Oxford Bulletin of Economics and Statistics 82*(1), 50–61.

Belloni, A., V. Chernozhukov, I. Fernández-Val, and C. Hansen (2017). Program evaluation and causal inference with high-dimensional data. *Econometrica 85*(1), 233–298.

Biggio, B., I. Corona, D. Maiorca, B. Nelson, N. Šrndić, P. Laskov, G. Giacinto, and F. Roli (2013). Evasion attacks against machine learning at test time. In *Joint European conference on machine learning and knowledge discovery in databases*, Berlin, Heidelberg, pp. 387–402. Springer.

Bilal, A. (2019). The geography of unemployment.

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

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

Busso, M., J. DiNardo, and J. McCrary (2014). New evidence on the finite sample properties of propensity score reweighting and matching estimators. *Review of Economics and Statistics 96*(5), 885–897.

Chalfin, A., B. Hansen, J. Lerner, and L. Parker (2022). Reducing crime through environmental design: Evidence from a randomized experiment of street lighting in new york city. *Journal of Quantitative Criminology 38*(1), 127–157.

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

Chetty, R. and N. Hendren (2018). The impacts of neighborhoods on intergenerational mobility i: Childhood exposure effects. *The Quarterly Journal of Economics 133*(3), 1107–1162.

Chetty, R., N. Hendren, P. Kline, and E. Saez (2014). Where is the land of opportunity? the geography of intergenerational mobility in the united states. *The Quarterly Journal of Economics 129*(4), 1553–1623.

Cireşan, D., U. Meier, and J. Schmidhuber (2012). Multi-column deep neural networks for image classification. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3642–3649. IEEE.

Cohen, J. and P. Dupas (2010). Free distribution or cost-sharing? evidence from a randomized malaria prevention experiment. *The Quarterly Journal of Economics 125*(1), 1–45.

Cohen, T. S. and M. Welling (2016). Group equivariant convolutional networks. In *International conference on machine learning*, pp. 2990–2999. PMLR.

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

Cressie, N. A. C. (1993). *Statistics for Spatial Data* (Rev. ed. ed.). Wiley Series in Probability and Mathematical Statistics. Wiley.

Currie, J., L. Davis, M. Greenstone, and R. Walker (2015). Environmental health risks and housing values: evidence from 1,600 toxic plant openings and closings. *American Economic Review 105*(2), 678–709.

Delgado, M. S. and R. J. G. M. Florax (2015). Difference-in-differences techniques for spatial data: Local autocorrelation and spatial interaction. *Economics Letters 137*, 123–126.

Dell, M. and B. A. Olken (2020). The development effects of the extractive colonial economy: The dutch cultivation system in java. *The Review of Economic Studies 87*(1), 164–203.

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

Dieleman, S., J. De Fauw, and K. Kavukcuoglu (2016). Exploiting cyclic symmetry in convolutional neural networks. In *International conference on machine learning*, pp. 1889–1898. PMLR.

Donaldson, D. and A. Storeygard (2016). The view from above: Applications of satellite data in economics. *Journal of Economic Perspectives 30*(4), 171–98.

Druckenmiller, H. and S. Hsiang (2019). Accounting for unobservable heterogeneity in cross section using spatial first differences. *NBER Working Paper Series* (25177).

Dudar, V. and V. Semenov (2018). Use of symmetric kernels for convolutional neural networks. In *XVIII International Conference on Data Science and Intelligent Analysis of Information*, pp. 3–10. Springer.

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

Dzhezyan, G. and H. Cecotti (2019). Symnet: Symmetrical filters in convolutional neural networks. *arXiv preprint arXiv:1906.04252*.

Engstrom, R., J. Hersh, and D. Newhouse (2021). Poverty from space: using high-resolution satellite imagery for estimating economic well-being. *The World Bank Economic Review 0*(0), 1–31.

Feyrer, J., E. Mansur, and B. Sacerdote (2020). Geographic dispersion of economic shocks: Evidence from the fracking revolution: Reply. *American Economic Review 110*(6), 1914–1920.

Feyrer, J., E. T. Mansur, and B. Sacerdote (2017). Geographic dispersion of economic shocks: Evidence from the fracking revolution. *American Economic Review 107*(4), 1313–1334.

Finkelstein, A., M. Gentzkow, and H. Williams (2016). Sources of geographic variation in health care: Evidence from patient migration. *The Quarterly Journal of Economics 131*(4), 1681–1726.

Finkelstein, A., M. Gentzkow, and H. Williams (2021). Place-based drivers of mortality: Evidence from migration. *American Economic Review 111*(8), 2697–2735.

Freyaldenhoven, S., C. Hansen, and J. M. Shapiro (2019). Pre-event trends in the panel event-study design. *American Economic Review 109*(9), 3307–38.

Frölich, M. (2004a). Finite-sample properties of propensity-score matching and weighting estimators. *Review of Economics and Statistics 86*(1), 77–90.

Frölich, M. (2004b). A note on the role of the propensity score for estimating average treatment effects. *Econometric Reviews 23*(2), 167–174.

Gens, R. and P. M. Domingos (2014). Deep symmetry networks. In *Advances in Neural Information Processing Systems*, Volume 27, pp. 2537–2545.

Gentzkow, M., B. Kelly, and M. Taddy (2019). Text as data. *Journal of Economic Literature 57*(3), 535–74.

Glaeser, E. L., S. D. Kominers, M. Luca, and N. Naik (2018). Big data and big cities: The promises and limitations of improved measures of urban life. *Economic Inquiry 56*(1), 114–137.

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

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

Goodfellow, I., H. Lee, Q. Le, A. Saxe, and A. Ng (2009). Measuring invariances in deep networks. In *Advances in Neural Information Processing Systems*, Volume 22, pp. 646–654.

Goodfellow, I., J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio (2014). Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems*, Volume 27, pp. 2672–2680.

Greene, W. (2009). Discrete choice modeling. In T. C. Mills and K. Patterson (Eds.), *Palgrave Handbook of Econometrics*, Volume 2, Chapter 11, pp. 473–556. London, UK: Palgrave Macmillan.

Greenstone, M., R. Hornbeck, and E. Moretti (2010). Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings. *Journal of Political Economy 118*(3), 536–598.

Greenstone, M. and E. Moretti (2003). Bidding for industrial plants: Does winning a 'million dollar plant' increase welfare? *NBER Working Paper Series* (9844).

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

Hahn, J. (1998). On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica 66*(2), 315–331.

Hastie, T. J., R. J. Tibshirani, and J. H. Friedman (2001). *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer Series in Statistics. Springer.

Heckman, J. J., J. Smith, and N. Clements (1997). Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts. *The Review of Economic Studies 64*(4), 487–535.

Hinton, G. E., A. Krizhevsky, and S. D. Wang (2011). Transforming auto-encoders. In T. Honkela, W. Duch, M. Girolami, and S. Kaski (Eds.), *Artificial Neural Networks and Machine Learning – ICANN 2011*, Volume 6791 of *Lecture Notes in Computer Science*, Berlin, Heidelberg, pp. 44–51. Springer.

Hirano, K., G. W. Imbens, and G. Ridder (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica 71*(4), 1161–1189.

Hoff, P. D., A. E. Raftery, and M. S. Handcock (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association 97*(460), 1090–1098.

Hotelling, H. (1929). Stability in competition. *The Economic Journal 39*(153), 41–57.

Hudgens, M. G. and M. E. Halloran (2008). Toward causal inference with interference. *Journal of the American Statistical Association 103*(482), 832–842.

Imbens, G. W. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *The Review of Economics and Statistics 86*(1), 4–29.

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

Jaderberg, M., K. Simonyan, A. Zisserman, and K. Kavukcuoglu (2015). Spatial transformer networks. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett (Eds.), *Advances in Neural Information Processing Systems*, Volume 28, pp. 2017–2025.

James, A. G. and B. Smith (2020). Geographic dispersion of economic shocks: Evidence from the fracking revolution: Comment. *American Economic Review 110*(6), 1905–1913.

Jean, N., M. Burke, M. Xie, W. M. Davis, D. B. Lobell, and S. Ermon (2016). Combining satellite imagery and machine learning to predict poverty. *Science 353*(6301), 790–794.

Jia, P. (2008). What happens when wal-mart comes to town: An empirical analysis of the discount retailing industry. *Econometrica 76*(6), 1263–1316.

Kaji, T., E. Manresa, and G. Pouliot (2020). An adversarial approach to structural estimation. *arXiv preprint arXiv:2007.06169*.

Kauderer-Abrams, E. (2017). Quantifying translation-invariance in convolutional neural networks. *arXiv preprint arXiv:1801.01450*.

Keiser, D. A. and J. S. Shapiro (2019). Consequences of the clean water act and the demand for water quality. *The Quarterly Journal of Economics 134*(1), 349–396.

Kellogg, M., M. Mogstad, G. Pouliot, and A. Torgovitsky (2021). Combining matching and synthetic control to tradeoff biases from extrapolation and interpolation. *Journal of the American Statistical Association 116*(536), 1804–1816.

Kelly, M. (2019). The standard errors of persistence.

Krizhevsky, A., I. Sutskever, and G. E. Hinton (2012). Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems*, Volume 25, pp. 1097–1105.

Lee, Y. and E. L. Ogburn (2021). Network dependence can lead to spurious associations and invalid inference. *Journal of the American Statistical Association 116*(535), 1060–1074.

Lenc, K. and A. Vedaldi (2015). Understanding image representations by measuring their equivariance and equivalence. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 991–999.

Liang, K.-Y. and S. L. Zeger (1986). Longitudinal data analysis using generalized linear models. *Biometrika 73*(1), 13–22.

Liang, T. (2018). On how well generative adversarial networks learn densities: Nonparametric and parametric results. *arXiv preprint arXiv:1811.03179*.

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

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

Manski, C. F. and J. V. Pepper (2018). How do right-to-carry laws affect crime rates? coping with ambiguity using bounded-variation assumptions. *Review of Economics and Statistics 100*(2), 232–244.

McIntosh, C. (2008). Estimating treatment effects from spatial policy experiments: an application to ugandan microfinance. *The Review of Economics and Statistics 90*(1), 15–28.

Miguel, E. and M. Kremer (2004). Worms: identifying impacts on education and health in the presence of treatment externalities. *Econometrica 72*(1), 159–217.

Mullainathan, S. and J. Spiess (2017). Machine learning: an applied econometric approach. *Journal of Economic Perspectives 31*(2), 87–106.

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

Neyman, J. (1990). On the application of probability theory to agricultural experiments. essay on principles. section 9. *Statistical Science 5*(4), 465–472. [translated by D. M. Dabrowska and T. P. Speed].

Rambachan, A. and J. Roth (2019). An honest approach to parallel trends.

Robins, J. M. and A. Rotnitzky (1995). Semiparametric efficiency in multivariate regression models with missing data. *Journal of the American Statistical Association 90*(429), 122–129.

Robinson, P. M. (1988). Root-n-consistent semiparametric regression. *Econometrica 56*(4), 931–954.

Rosenbaum, P. R. (2007). Interference between units in randomized experiments. *Journal of the American Statistical Association 102*(477), 191–200.

Rosenbaum, P. R. and D. B. Rubin (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika 70*(1), 41–55.

Rossin-Slater, M., M. Schnell, H. Schwandt, S. Trejo, and L. Uniat (2020). Local exposure to school shootings and youth antidepressant use. *Proceedings of the National Academy of Sciences 117*(38), 23484–23489.

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

SafeGraph (2019). Determining point-of-interest visits from location data: A technical guide to visit attribution.

Sandler, D. H. (2017). Externalities of public housing: The effect of public housing demolitions on local crime. *Regional Science and Urban Economics 62*, 24–35.

Sävje, F. (2021). Causal inference with misspecified exposure mappings. *arXiv preprint arXiv:2103.06471*.

Sävje, F., P. M. Aronow, and M. G. Hudgens (2021). Average treatment effects in the presence of unknown interference. *The Annals of Statistics 49*(2), 673–701.

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

Singh, S., A. Uppal, B. Li, C.-L. Li, M. Zaheer, and B. Poczos (2018). Nonparametric density estimation under adversarial losses. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (Eds.), *Advances in Neural Information Processing Systems*, Volume 31, pp. 10225–10236.

Srivastava, N., G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov (2014). Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research 15*(1), 1929–1958.

Stock, J. H. (1989). Nonparametric policy analysis. *Journal of the American Statistical Association 84*(406), 567–575.

Stock, J. H. (1991). Nonparametric policy analysis: an application to estimating hazardous waste cleanup benefits. In W. A. Barnett, J. Powell, and G. Tauchen (Eds.), *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, Chapter 3, pp. 77–98. Cambridge University Press.

Szegedy, C., W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, and R. Fergus (2013). Intriguing properties of neural networks. *arXiv preprint arXiv:1312.6199*.

Tchetgen Tchetgen, E. J. and T. J. VanderWeele (2012). On causal inference in the presence of interference. *Statistical Methods in Medical Research 21*(1), 55–75.

Vazquez-Bare, G. (2017). Identification and estimation of spillover effects in randomized experiments. *arXiv preprint arXiv:1711.02745*.

Vincent, P., H. Larochelle, Y. Bengio, and P.-A. Manzagol (2008). Extracting and composing robust features with denoising autoencoders. In A. McCallum and S. Roweis (Eds.), *25th International Conference on Machine Learning*, pp. 1096–1103.

Yaeger, L., R. Lyon, and B. Webb (1996). Effective training of a neural network character classifier for word recognition. In *Advances in Neural Information Processing Systems*, Volume 9, pp. 807–816.

Yang, F., Z. Wang, and C. Heinze-Deml (2019). Invariance-inducing regularization using worst-case transformations suffices to boost accuracy and spatial robustness. *arXiv preprint arXiv:1906.11235*.

Zeiler, M. D. and R. Fergus (2014). Visualizing and understanding convolutional networks. In *European Conference on Computer Vision*, pp. 818–833. Springer.

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

A Theoretical Results

A.1 Expected Value of The Average Outcome of Treated Individuals

The average of treated individuals at distance $d \pm h$ from realized treatment locations is

$$\bar{Y}_t(d) = \frac{\sum_{i\in\mathbb{I}} W_{j(i)} \mathbb{1}\{|d(\xi_{j(i)}, r_i) - d| \leq h\} Y_i}{\sum_{i\in\mathbb{I}} W_{j(i)} \mathbb{1}\{|d(\xi_{j(i)}, r_i) - d| \leq h\}}.$$

Here, consider the expectation of the term in the numerator corresponding to individual i, $W_{j(i)} \mathbb{1}\{|d(\xi_{j(i)}, r_i) - d| \leq h\} Y_i$. The expected value of the term is

$$\begin{aligned}
&E\Big(W_{j(i)} \mathbb{1}\{|d(\xi_{j(i)}, r_i) - d| \leq h\} Y_i\Big) \\
&= E\Big(W_{j(i)} \mathbb{1}\{|d(\xi_{j(i)}, r_i) - d| \leq h\} Y_i(\xi_{j(i)})\Big) \\
&= E\Big(W_{j(i)} \sum_{s\in\mathbb{S}_{j(i)}} \mathbb{1}\{s = \xi_{j(i)}\} \mathbb{1}\{|d(s, r_i) - d| \leq h\} Y_i(s)\Big) \\
&= \sum_{s\in\mathbb{S}_{j(i)}} E\Big(W_{j(i)} \mathbb{1}\{s = \xi_{j(i)}\}\Big) \mathbb{1}\{|d(s, r_i) - d| \leq h\} Y_i(s) \\
&= \sum_{s\in\mathbb{S}_{j(i)}} \pi_{j(i)} g_{j(i)}(s) \mathbb{1}\{|d(s, r_i) - d| \leq h\} Y_i(s)\Big)
\end{aligned}$$

The first step uses that the realized outcome Y_i is the potential outcome corresponding to the realized treatment. The second step rewrites the potential outcome and distance bin indicator function in terms of non-stochastic candidate locations s by summing over all possible treatment locations in the region, $\sum_{s\in\mathbb{S}_{j(i)}} \mathbb{1}\{s = \xi_{j(i)}\}$. The third step moves the expectation into the summation, and the non-stochastic distance bin indicator function and potential outcome out. The final step resolves the expectation in terms of the probabilities determined by the experimental design, defined in Section 3.

A.2 Proof of Theorem 3

A.2.1 Estimator

The general estimator of interest in the setting without interference can be written as

$$\hat{\tau}_w(d) \equiv \frac{\sum_{j=1}^{J} \frac{W_j}{\pi_j} \sum_{s\in\mathbb{S}_j} \frac{\mathbb{1}\{\xi_j = s\}}{g_j(s)} \sum_{i\in\mathbb{I}_j} w_i(s, d) Y_i}{\sum_{j=1}^{J} \frac{W_j}{\pi_j} \sum_{s\in\mathbb{S}_j} \frac{\mathbb{1}\{\xi_j = s\}}{g_j(s)} \sum_{i\in\mathbb{I}_j} w_i(s, d)} - \frac{\sum_{j=1}^{J} \frac{1-W_j}{1-\pi_j} \sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w_i(s, d) Y_i}{\sum_{j=1}^{J} \frac{1-W_j}{1-\pi_j} \sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w_i(s, d)}$$

where index j denotes regions, $W_j = 1$ if region j is treated at some location, and ξ_j is the single treatment location chosen in region j (if any). The weight function $w_i(s, d)$ is chosen by the user to weight individuals

i and treatment locations s as desired and primarily place weight on pairs that are distance d apart. For instance, for the ATT estimator with distance bin, choose $w_i(s,d) = \pi_j g_j(s) \mathbb{1}\{d(s,r_i) \leq h\}$. The probabilities of treatment in regions and locations are given by $\pi_j \equiv \Pr(W_j = 1)$ and $g_j(s) \equiv \Pr(\xi_j = \{s\} | W_j = 1)$.

The first term averages over individuals at distance d from a realized treatment location. The second term averages over individuals at distance d from counterfactual (unrealized) candidate treatment locations. The estimator estimates the weighted average treatment effect:

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

with user-specified weights w.

The experiment considered here is a completely randomized experiment at the region level, where a fixed number of regions receive treatment at exactly one location each, and treatment in a region is assumed to have no effect on outcomes in other regions – regions are "far apart."

A.2.2 Approximate Estimator

The estimator $\hat{\tau}_w(d)$ is hard to analyze (in finite samples) because the denominators are random. This arises because, depending on treatment assignment, there may be more or fewer individuals near realized / counterfactual locations. The same problem exists in standard randomized experiments when the treatment is randomized by an independent coin flip for each individual, such that the number of treated varies from assignment to assignment. In that setting, we can instead analyze the experiment with number of treated fixed at the value observed in the realized sample. Conditioning on the number of realized treatment locations is not sufficient in the spatial setting because the number of *individuals* would still vary since some locations have more individuals near them than other locations. Conditioning on the number of individuals restricts the assignment distribution asymmetrically – inverting an assignment generally changes the number of individuals near treatment – such that standard estimators are no longer unbiased by design.

The theoretical analysis of $\hat{\tau}_w(d)$ therefore relies on an approximate estimator that fixes the denominators (at their expected values), and centers the numerators in a way that minimizes the difference between $\hat{\tau}_w(d)$ and its approximation.

The approximate estimator is

$$\begin{aligned} \tilde{\tau}_w(d) \equiv & \tau_w(d) + \frac{\sum_{j=1}^{J} \frac{W_j}{\pi_j} \sum_{s \in \mathbb{S}_j} \frac{\mathbb{1}\{\xi_j = s\}}{g_j(s)} \sum_{i \in \mathbb{I}_j} w_i(s,d)(Y_i - \mu_t(d))}{\sum_{j=1}^{J} \sum_{s \in \mathbb{S}_j} \sum_{i \in \mathbb{I}_j} w_i(s,d)} \\ & - \frac{\sum_{j=1}^{J} \frac{1-W_j}{1-\pi_j} \sum_{s \in \mathbb{S}_j} \sum_{i \in \mathbb{I}_j} w_i(s,d)(Y_i - \mu_c(d))}{\sum_{j=1}^{J} \sum_{s \in \mathbb{S}_j} \sum_{i \in \mathbb{I}_j} w_i(s,d)} \end{aligned} \tag{A.1}$$

where $\mu_t(d)$ and $\mu_c(d)$ are average potential outcomes:

$$\mu_{w,t}(d) \equiv \frac{\sum_{j=1}^{J} \sum_{s \in \mathbb{S}_j} \sum_{i \in \mathbb{I}_j} w_i(s,d) Y_i(s)}{\sum_{j=1}^{J} \sum_{s \in \mathbb{S}_j} \sum_{i \in \mathbb{I}_j} w_i(s,d)}$$

and

$$\mu_{w,c}(d) \equiv \frac{\sum_{j=1}^{J} \sum_{s \in \mathbb{S}_j} \sum_{i \in \mathbb{I}_j} w_i(s,d) Y_i(0)}{\sum_{j=1}^{J} \sum_{s \in \mathbb{S}_j} \sum_{i \in \mathbb{I}_j} w_i(s,d)}$$

such that

$$\tau_w(d) = \mu_{w,t}(d) - \mu_{w,c}(d)$$

A.2.3 Quality of Approximation

Here, I show that the estimators $\hat{\tau}_w(d)$ and $\tilde{\tau}_w(d)$ are *very close* in large enough samples. This motivates the use of exact finite sample results for the mean and variance of the infeasible estimator $\tilde{\tau}_w(d)$ for inference with the feasible estimator $\hat{\tau}_w(d)$. The analysis uses the mean value theorem to derive the difference $\hat{\tau}_w(d) - \tilde{\tau}_w(d)$ and argues that this difference is small in large enough samples.

As a practical matter, a sample is large enough if the number of individuals near treatment and control are close to their expected values. The approximation of $\hat{\tau}_w(d)$ by $\tilde{\tau}_w(d)$ is particular close when also the average outcomes are close to their expected values.

To simplify notation, define the following shorthands:

$$\begin{aligned}
\hat{\mu}_{w,t}(d) &= \frac{\sum_{j=1}^{J} \frac{W_j}{\pi_j} \sum_{s\in\mathbb{S}_j} \frac{\mathbb{1}\{\xi_j=s\}}{g_j(s)} \sum_{i\in\mathbb{I}_j} w_i(s,d) Y_i}{\sum_{j=1}^{J} \frac{W_j}{\pi_j} \sum_{s\in\mathbb{S}_j} \frac{\mathbb{1}\{\xi_j=s\}}{g_j(s)} \sum_{i\in\mathbb{I}_j} w_i(s,d)} \\
\hat{\mu}_{w,c}(d) &= \frac{\sum_{j=1}^{J} \frac{1-W_j}{1-\pi_j} \sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w_i(s,d) Y_i}{\sum_{j=1}^{J} \frac{1-W_j}{1-\pi_j} \sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w_i(s,d)} \\
\tilde{\mu}_{w,t}(d) &= \frac{\sum_{j=1}^{J} \frac{W_j}{\pi_j} \sum_{s\in\mathbb{S}_j} \frac{\mathbb{1}\{\xi_j=s\}}{g_j(s)} \sum_{i\in\mathbb{I}_j} w_i(s,d) Y_i}{\sum_{j=1}^{J} \sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w_i(s,d)} \\
\tilde{\mu}_{w,c}(d) &= \frac{\sum_{j=1}^{J} \frac{1-W_j}{1-\pi_j} \sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w_i(s,d) Y_i}{\sum_{j=1}^{J} \sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w_i(s,d)}
\end{aligned}$$

and

$$\begin{aligned}
\hat{p}_{w,t}(d) &= \frac{1}{J} \sum_{j=1}^{J} \frac{W_j}{\pi_j} \sum_{s\in\mathbb{S}_j} \frac{\mathbb{1}\{\xi_j=s\}}{g_j(s)} \sum_{i\in\mathbb{I}_j} w_i(s,d) \\
\hat{p}_{w,c}(d) &= \frac{1}{J} \sum_{j=1}^{J} \sum_{j=1}^{J} \frac{1-W_j}{1-\pi_j} \sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w_i(s,d) \\
p_w(d) &= \frac{1}{J} \sum_{j=1}^{J} \sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w_i(s,d)
\end{aligned}$$

where $\tilde{\mu}_{w,t}(d)$ and $\tilde{\mu}_{w,c}(d)$ replicate $\hat{\mu}_{w,t}(d)$ and $\hat{\mu}_{w,c}(d)$ but with expected values rather than sample averages in the denominators. The sample average denominators are $\hat{p}_{w,t}(d)$ and $\hat{p}_{w,c}(d)$ (scaled such that they converge under suitable conditions when J grows), and the expected value of the denominators is $p_w(d)$ (similarly scaled).

Without loss of generality, I fix the distance d and weighting w of interest and suppress the dependence on d and w in the following derivations for ease of presentation.

The feasible estimator written in terms of the shorthand notation is

$$\hat{\tau} = \hat{\mu}_t - \hat{\mu}_c = \frac{p}{\hat{p}_t}\tilde{\mu}_t - \frac{p}{\hat{p}_c}\tilde{\mu}_c$$

while the infeasible estimator is

$$\tilde{\tau} = \mu_t - \mu_c + \tilde{\mu}_t - \frac{\hat{p}_t}{p}\mu_t - \tilde{\mu}_c + \frac{\hat{p}_c}{p}\mu_c.$$

Next, define the function

$$\begin{aligned}\tilde{\Delta}(\hat{p}_t, \hat{p}_c, \tilde{\mu}_t, \tilde{\mu}_c) &\equiv \frac{p}{\hat{p}_t}\tilde{\mu}_t - \frac{p}{\hat{p}_c}\tilde{\mu}_c - (\mu_t - \mu_c + \tilde{\mu}_t - \frac{\hat{p}_t}{p}\mu_t - \tilde{\mu}_c + \frac{\hat{p}_c}{p}\mu_c) \\ &= \hat{\tau} - \tilde{\tau}\end{aligned}$$

Finally, apply the mean value theorem on $\tilde{\Delta}$ on the interval with endpoints $\hat{x} = (\hat{p}_t, \hat{p}_c, \tilde{\mu}_t, \tilde{\mu}_c)$ and $x = (p, p, \mu_t, \mu_c)$. The mean value theorem states that

$$\begin{aligned}&\tilde{\Delta}(\hat{p}_t, \hat{p}_c, \tilde{\mu}_t, \tilde{\mu}_c) - \tilde{\Delta}(p, p, \mu_t, \mu_c) \\ &= \begin{bmatrix} \hat{p}_t - p_t \\ \hat{p}_c - p_c \\ \tilde{\mu}_t - \mu_t \\ \tilde{\mu}_c - \mu_c \end{bmatrix} \cdot \begin{bmatrix} \frac{\partial \tilde{\Delta}}{\partial \hat{p}_t}(\dot{x}) \\ \frac{\partial \tilde{\Delta}}{\partial \hat{p}_c}(\dot{x}) \\ \frac{\partial \tilde{\Delta}}{\partial \tilde{\mu}_t}(\dot{x}) \\ \frac{\partial \tilde{\Delta}}{\partial \tilde{\mu}_c}(\dot{x}) \end{bmatrix}\end{aligned}$$

where $\dot{x}$ is some convex combination of $\hat{x}$ and x.

It is straightforward to see that $\tilde{\Delta}(p, p, \mu_t, \mu_c) = 0$. Hence the left-hand-side of the equality above is just $\tilde{\Delta}(\hat{p}_t, \hat{p}_c, \tilde{\mu}_t, \tilde{\mu}_c)$, such that the right-hand-side is an expression for $\hat{\tau} - \tilde{\tau}$.

Hence

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

Each of the four terms is a product with each factor close to zero under appropriate asymptotics. For instance, with independent regions and bounded outcomes and number of individuals per region, one can get $\sqrt{J}(\hat{p}_t - p) \to 0$. That is, the difference between the estimators $\hat{\tau}_w(d)$ and $\tilde{\tau}_w(d)$ is negligible under standard asymptotic frameworks. Since the difference between estimators is *very small* for large samples, exact finite sample results for $\tilde{\tau}_w(d)$ likely provide decent approximations for $\hat{\tau}_w(d)$ in smaller samples.

A.2.4 Unbiasedness of Approximate Estimator

Consider the expected value of the estimator $\tilde{\tau}_w(d)$. To show: $E(\tilde{\tau}_w(d)) = \tau_w(d)$. Since $\tau_w(d)$ is the first term of $\tilde{\tau}_w(d)$, I proceed by showing that $E(\tilde{\tau}_w(d) - \tau_w(d)) = 0$. Since the denominators are non-stochastic, it suffices to show that the expectations of the numerators are equal to zero. The "first term" and "second term" designations below therefore refer to the first and second term of $\tilde{\tau}_w(d) - \tau_w(d)$.

The expectation of the numerator of the first term is:

$$
\begin{aligned}
&E\Big(\sum_{j=1}^{J}\frac{W_j}{\pi_j}\sum_{s\in\mathbb{S}_j}\frac{\mathbb{1}\{\xi_j=s\}}{g_j(s)}\sum_{i\in\mathbb{I}_j}w_i(s,d)(Y_i-\mu_{w,t}(d))\Big)\\
&=E\Big(\sum_{j=1}^{J}\frac{W_j}{\pi_j}\sum_{s\in\mathbb{S}_j}\frac{\mathbb{1}\{\xi_j=s\}}{g_j(s)}\sum_{i\in\mathbb{I}_j}w_i(s,d)(Y_i(s)-\mu_{w,t}(d))\Big)\\
&=\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}E\Big(\frac{W_j\mathbb{1}\{\xi_j=s\}}{\pi_j g_j(s)}\Big)\sum_{i\in\mathbb{I}_j}w_i(s,d)(Y_i(s)-\mu_{w,t}(d))\\
&=\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j}w_s(s,d)(Y_i(s)-\mu_{w,t}(d))\\
&=\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j}w_i(s,d)Y_i(s)-\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j}w_i(s,d)\frac{\sum_{j'=1}^{J}\sum_{s'\in\mathbb{S}_{j'}}\sum_{i'\in\mathbb{I}_{j'}}w_{i'}(s',d)Y_{i'}(s')}{\sum_{j'=1}^{J}\sum_{s'\in\mathbb{S}_{j'}}\sum_{i'\in\mathbb{I}_{j'}}w_{i'}(s',d)}\\
&=0
\end{aligned}
$$

The first equality rewrites the observed outcome $Y_i = Y_i(\xi_{j(i)})$ in terms of potential outcome $Y_i(s) = Y_i(\xi_j)$ for $s = \xi_j$. The second equality moves all non-stochastic terms out of the expectation. The third equality rewrites the expectation of indicators as probabilities. The fourth equality distributes the difference $Y_i(s) - \mu_{w,t}(d)$ and replaces $\mu_{w,t}(d)$ by its definition. For the second term, the factor multiplying the ratio cancels with the denominator, and the numerator is equal to the first term, such that the difference is equal to zero.

Analogously, the expectation of the numerator of the second term is:

$$
\begin{aligned}
&E\Big(\sum_{j=1}^{J}\frac{1-W_j}{1-\pi_j}\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j}w_i(s,d)(Y_i-\mu_{w,c}(d))\Big)\\
&=E\Big(\sum_{j=1}^{J}\frac{1-W_j}{1-\pi_j}\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j}w_i(s,d)(Y_i(0)-\mu_{w,c}(d))\Big)\\
&=\sum_{j=1}^{J}\sum_{j=1}^{J}E\Big(\frac{1-W_j}{1-\pi_j}\Big)\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j}w_i(s,d)(Y_i(0)-\mu_{w,c}(d))\\
&=\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j}w_i(s,d)(Y_i(0)-\mu_{w,c}(d))\\
&=\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j}w_i(s,d)Y_i(0)-\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j}w_i(s,d)\frac{\sum_{j'=1}^{J}\sum_{s'\in\mathbb{S}_{j'}}\sum_{i'\in\mathbb{I}_{j'}}w_{i'}(s',d)Y_{i'}(0)}{\sum_{j'=1}^{J}\sum_{s'\in\mathbb{S}_{j'}}\sum_{i'\in\mathbb{I}_{j'}}w_{i'}(s',d)}\\
&=0
\end{aligned}
$$

Hence $E(\tilde{\tau}_w(d)) = \tau_w(d)$.

A.2.5 Variance of Approximate Estimator

The approximate estimator $\tilde{\tau}_w(d)$ in Equation A.1 is the sum of three terms. Since the first term, $\tau_w(d)$ is fixed, the variance only depends on the last two terms.

First, rewrite the variance more compactly:

$$\begin{aligned}
\text{var}\Big(\tilde{\tau}_w(d)\Big) &= \text{var}\Bigg(\frac{\sum_{j=1}^{J}\frac{W_j}{\pi_j}\sum_{s\in\mathbb{S}_j}\frac{\mathbb{1}\{\xi_j=s\}}{g_j(s)}\sum_{i\in\mathbb{I}_j}w_i(s,d)(Y_i(s)-\mu_{w,t}(d))}{\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j}w_i(s,d)} \\
&\quad - \frac{\sum_{j=1}^{J}\frac{1-W_j}{1-\pi_j}\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j}w_i(s,d)(Y_i(0)-\mu_{w,c}(d))}{\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j}w_i(s,d)}\Bigg) \\
&= \text{var}\Bigg(\frac{\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}T_j(s)\frac{1}{\pi_j g_j(s)}\sum_{i\in\mathbb{I}_j}w_i(s,d)(Y_i(s)-\mu_{w,t}(d))}{\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j}w_i(s,d)} \\
&\quad - \frac{\sum_{j=1}^{J}(1-\sum_{s\in\mathbb{S}_j}T_j(s))\frac{1}{1-\pi_j}\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j}k(\frac{d(s,r_i)-d}{h})(Y_i(0)-\mu_{w,c}(d))}{\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j}w_i(s,d)}\Bigg) \\
&= \text{var}\Bigg(\frac{\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}T_j(s)\frac{1}{\pi_j g_j(s)}\sum_{i\in\mathbb{I}_j}w_i(s,d)(Y_i(s)-\mu_{w,t}(d))}{\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j}w_i(s,d)} \\
&\quad + \frac{\sum_{j=1}^{J}(\sum_{s\in\mathbb{S}_j}T_j(s))\frac{1}{1-\pi_j}\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j}w_i(s,d)(Y_i(0)-\mu_{w,c}(d))}{\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j}w_i(s,d)}\Bigg) \\
&= \text{var}\Bigg(\frac{\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}T_j(s)\sum_{i\in\mathbb{I}_j}\Big(\frac{w_i(s,d)}{\pi_j g_j(s)}(Y_i(s)-\mu_{w,t}(d))+\sum_{s'\in\mathbb{S}_j}\frac{w_i(s',d)}{1-\pi_j}(Y_i(0)-\mu_{w,c}(d))\Big)}{\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j}w_i(s,d)}\Bigg)
\end{aligned}$$

For the first equality, I replace the observed outcome Y_i by the potential outcome corresponding to the realized treatment state, $Y_i(s)$ if $W_j = 1$ and $\mathbb{1}\{\xi_j = s\}$ and $Y_i(0)$ if $W_j = 0$. The second equality substitutes a redefined "treatment indicator"

$$T_j(s) \equiv W_j \mathbb{1}\{\xi_j = s\}$$

and $\sum_{s\in\mathbb{S}_j} T_j(s) = W_j$. For the third equality, distribute out the -1 term of $-(1-(\sum_s T_j(s)))\ldots$, which is non-stochastic and hence does not contribute to the variance, such that only $+(\sum_s T_j(s))\ldots$ remains of the second term. The fourth and final equality above distributes out the $T_j(s)$ of the second term and then combines the first and second term by factoring out $T_j(s)$.

For ease of notation, define

$$\bar{Y}^{+}_{w,j}(s,d) \equiv \frac{\sum_{i\in\mathbb{I}_j}\Big(\frac{w_i(s,d)}{\pi_j g_j(s)}(Y_i(s)-\mu_{w,t}(d))+\sum_{s'\in\mathbb{S}_j}\frac{w_i(s',d)}{1-\pi_j}(Y_i(0)-\mu_{w,c}(d))\Big)}{\sum_{j'=1}^{J}\sum_{s'\in\mathbb{S}_{j'}}\sum_{i\in\mathbb{I}_{j'}}w_i(s',d)}$$

such that

$$\text{var}\Big(\tilde{\tau}_w(d)\Big) = \text{var}\Big(\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}T_j(s)\bar{Y}^{+}_{w,j}(s,d)\Big)$$

The only stochastic terms left are the $T_j(s)$; they represent the design-based variation that is due to random treatment assignment. The average $\bar{Y}^{+}_{w,j}(s,d)$ consists only of a sum of *potential* outcomes, which are non-stochastic in the design-based perspective, in the numerator and the *expected* number of individuals near treatment, which is also non-stochastic, in the denominator.

The variance of the sum is a sum of all possible covariances:

$$\begin{aligned}
\operatorname{var}\Big(\tilde{\tau}_w(d)\Big) &= \operatorname{var}\Big(\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j} T_j(s)\bar{Y}^+_{w,j}(s,d)\Big) \\
&= \sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j} \operatorname{var}\Big(T_j(s)\Big)\bar{Y}^+_{w,j}(s,d)^2 \\
&\quad + \sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\sum_{s'\in\mathbb{S}_j} \mathbb{1}\{s\neq s'\}\operatorname{cov}\Big(T_j(s),T_j(s')\Big)\bar{Y}^+_{w,j}(s,d)\bar{Y}^+_{w,j}(s',d) \\
&\quad + \sum_{j=1}^{J}\sum_{j'=1}^{J}\sum_{s\in\mathbb{S}_j}\sum_{s'\in\mathbb{S}_{j'}} \mathbb{1}\{j\neq j'\}\operatorname{cov}\Big(T_j(s),T_{j'}(s')\Big)\bar{Y}^+_{w,j}(s,d)\bar{Y}^+_{w,j'}(s',d)
\end{aligned}$$

The mean, variances and covariances of $T_j(s)$ are determined by the experimental design. The expected value of the location-specific treatment indicator is

$$E(T_j(s)) = \pi_j g_j(s).$$

Consider

$$\operatorname{cov}(T_j(s),T_{j'}(s')) = E(T_j(s)T_{j'}(s')) - E(T_j(s))E(T_{j'}(s')).$$

$j = j'$ AND $s = s'$ Since $T_j(s)\in\{0,1\}$, $T_j(s)^2 = T_j(s)$. So

$$\operatorname{var}(T_j(s)) = \pi_j g_j(s)(1-\pi_j g_j(s))$$

$j = j'$ AND $s \neq s'$ Since at most one location per region is treated, $T_j(s)T_j(s') = 0$ if $s\neq s'$. So

$$\operatorname{cov}(T_j(s),T_j(s')) = -\pi_j^2 g_j(s)g_j(s')$$

$j \neq j'$ When j and j' are distinct regions, under Assumption 2 (completely randomized experiment):

$$\begin{aligned}
E(T_j(s)T_{j'}(s')) &= E(W_jW_{j'}\mathbb{1}\{\xi_j = s\}\mathbb{1}\{\xi_{j'} = s'\}) \\
&= \Pr(W_j=1)E(W_jW_{j'}\mathbb{1}\{\xi_j = s\}\mathbb{1}\{\xi_{j'} = s'\}|W_j=1) \\
&\quad + \Pr(W_j=0)E(W_jW_{j'}\mathbb{1}\{\xi_j = s\}\mathbb{1}\{\xi_{j'} = s'\}|W_j=0) \\
&= \pi_j g_j(s)E(W_{j'}\mathbb{1}\{\xi_{j'} = s'\}|W_j=1) \\
&= \pi_j g_j(s)\Pr(W_{j'}=1|W_j=1)E(W_{j'}\mathbb{1}\{\xi_{j'} = s'\}|W_j=1,W_{j'}=1) \\
&\quad + \pi_j g_j(s)\Pr(W_{j'}=0|W_j=1)E(W_{j'}\mathbb{1}\{\xi_{j'} = s'\}|W_j=1,W_{j'}=0) \\
&= \pi_j g_j(s)\Pr(W_{j'}=1|W_j=1)g_{j'}(s')
\end{aligned}$$

where $\Pr(W_{j'}=1|W_j=1)$ is determined by the completely randomized design. Let J_t be the (fixed) number of treated regions in a completely randomized design. Then

$$\Pr(W_{j'}=1|W_j=1) = \frac{J_t-1}{J-1}$$

since, under Assumption 2, if region j receives treatment, $J_t - 1$ of the remaining $J - 1$ regions receive treatment, each with equal probability. So

$$\begin{aligned}
&E(T_j(s)T_{j'}(s')) - E(T_j(s))E(T_{j'}(s')) \\
&= \frac{J_t}{J}\frac{J_t - 1}{J - 1} g_j(s)g_{j'}(s') - \frac{J_t}{J}\frac{J_t}{J} g_j(s)g_{j'}(s') \\
&= (\frac{J_t - 1}{J - 1} - \frac{J_t}{J})\frac{J_t}{J} g_j(s)g_{j'}(s') \\
&= \frac{J(J_t - 1) - J_t(J - 1)}{J(J - 1)}\frac{J_t}{J} g_j(s)g_{j'}(s') \\
&= -\frac{J - J_t}{J(J - 1)}\frac{J_t}{J} g_j(s)g_{j'}(s') \\
&= \frac{\pi(1 - \pi)}{J - 1} g_j(s)g_{j'}(s').
\end{aligned}$$

When j and j' are distinct regions, under Assumption 3 (Bernoulli trial):

$$\operatorname{cov}(T_j(s), T_{j'}(s')) = 0$$

To summarize, the covariances of the $T_j(s)$ are

$$\begin{aligned}
\operatorname{cov}(T_j(s), T_{j'}(s')) &= E(T_j(s)T_{j'}(s')) - E(T_j(s))E(T_{j'}(s')) \\
&= \begin{cases}
(1 - \pi_j g_j(s))\pi_j g_j(s) & \text{if } j = j' \text{ and } s = s' \\
-\pi_j^2 g_j(s)g_j(s') & \text{if } j = j' \text{ and } s \neq s' \\
-\frac{\pi(1-\pi)}{J-1} g_j(s)g_{j'}(s') & \text{if } j \neq j' \text{ under Assumption 2} \\
0 & \text{if } j \neq j' \text{ under Assumption 3}
\end{cases}
\end{aligned}$$

Hence, under either of assumptions 2 and 3,

$$\begin{aligned}
\operatorname{var}\Big(\tilde{\tau}_w(d)\Big) &= \sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}(1 - \pi_j g_j(s))\pi_j g_j(s)\bar{Y}^+_{w,j}(s,d)^2 \\
&\quad - \sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\sum_{s'\in\mathbb{S}_j}\mathbb{1}\{s \neq s'\}\pi_j^2 g_j(s)g_j(s')\bar{Y}^+_{w,j}(s,d)\bar{Y}^+_{w,j}(s',d) \\
&\quad + \sum_{j=1}^{J}\sum_{j'=1}^{J}\sum_{s\in\mathbb{S}_j}\sum_{s'\in\mathbb{S}_{j'}}\mathbb{1}\{j \neq j'\}\operatorname{cov}(T_j(s), T_{j'}(s'))\bar{Y}^+_{w,j}(s,d)\bar{Y}^+_{w,j'}(s',d)
\end{aligned}$$

The second summation is "missing" the terms where $s = s'$. Adding and subtracting

$$\pm\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\sum_{s'\in\mathbb{S}_j}\mathbb{1}\{s = s'\}\pi_j^2 g_j(s)g_j(s')\bar{Y}^+_{w,j}(s,d)\bar{Y}^+_{w,j}(s',d)$$

obtain

$$\begin{aligned}
\operatorname{var}\Big(\tilde{\tau}_w(d)\Big) &= \sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\Big((1-\pi_j g_j(s))\pi_j g_j(s) + \pi_j^2 g_j(s)^2\Big)\bar{Y}^+_{w,j}(s,d)^2 \\
&\quad - \sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\sum_{s'\in\mathbb{S}_j}\pi_j^2 g_j(s)g_j(s')\bar{Y}^+_{w,j}(s,d)\bar{Y}^+_{w,j}(s',d) \\
&\quad + \sum_{j=1}^{J}\sum_{j'=1}^{J}\sum_{s\in\mathbb{S}_j}\sum_{s'\in\mathbb{S}_{j'}}\mathbb{1}\{j\neq j'\}\operatorname{cov}(T_j(s),T_{j'}(s'))\bar{Y}^+_{w,j}(s,d)\bar{Y}^+_{w,j'}(s',d) \\
&= \sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\pi_j g_j(s)\bar{Y}^+_{w,j}(s,d)^2 - \sum_{j=1}^{J}\pi_j^2\Big(\sum_{s\in\mathbb{S}_j} g_j(s)\bar{Y}^+_{w,j}(s,d)\Big)^2 \\
&\quad + \sum_{j=1}^{J}\sum_{j'=1}^{J}\sum_{s\in\mathbb{S}_j}\sum_{s'\in\mathbb{S}_{j'}}\mathbb{1}\{j\neq j'\}\operatorname{cov}(T_j(s),T_{j'}(s'))\bar{Y}^+_{w,j}(s,d)\bar{Y}^+_{w,j'}(s',d)
\end{aligned} \tag{A.2}$$

The first equality combines the added term with the first summation and the subtracted term with the second summation. The second equality simplifies the factor of the first term, factors the second term into s and s', and notices that both summations are the same, yielding the square in the second term.

Completely Randomized Experiment Under Assumption 2 (completely randomized experiment), Equation A.2 becomes

$$\begin{aligned}
\operatorname{var}\Big(\tilde{\tau}_w(d)\Big) &= \sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\pi_j g_j(s)\bar{Y}^+_{w,j}(s,d)^2 - \sum_{j=1}^{J}\pi_j^2\Big(\sum_{s\in\mathbb{S}_j} g_j(s)\bar{Y}^+_{w,j}(s,d)\Big)^2 \\
&\quad - \sum_{j=1}^{J}\sum_{j'=1}^{J}\sum_{s\in\mathbb{S}_j}\sum_{s'\in\mathbb{S}_{j'}}\mathbb{1}\{j\neq j'\}\frac{\pi(1-\pi)}{J-1}g_j(s)g_{j'}(s')\bar{Y}^+_{w,j}(s,d)\bar{Y}^+_{w,j'}(s',d)
\end{aligned}$$

Here, the third summation is "missing" the terms where $j = j'$. Adding and subtracting

$$\pm\sum_{j=1}^{J}\sum_{j'=1}^{J}\sum_{s\in\mathbb{S}_j}\sum_{s'\in\mathbb{S}_{j'}}\mathbb{1}\{j=j'\}\frac{\pi(1-\pi)}{J-1}g_j(s)g_{j'}(s')\bar{Y}^+_{w,j}(s,d)\bar{Y}^+_{w,j'}(s',d)$$

obtain

$$\begin{aligned}
\operatorname{var}\Big(\tilde{\tau}_w(d)\Big) &= \sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\pi_j g_j(s)\bar{Y}^+_{w,j}(s,d)^2 - \sum_{j=1}^{J}\big(\pi_j^2 - \frac{\pi(1-\pi)}{J-1}\big)\Big(\sum_{s\in\mathbb{S}_j} g_j(s)\bar{Y}^+_{w,j}(s,d)\Big)^2 \\
&\quad - \sum_{j=1}^{J}\sum_{j'=1}^{J}\sum_{s\in\mathbb{S}_j}\sum_{s'\in\mathbb{S}_{j'}}\frac{\pi(1-\pi)}{J-1}g_j(s)g_{j'}(s')\bar{Y}^+_{w,j}(s,d)\bar{Y}^+_{w,j'}(s',d)
\end{aligned} \tag{A.3}$$

by combining the added $j = j'$ term into the second term and the subtracted $j = j'$ term into the third term.

Next, consider the third term in isolation.

$$\begin{aligned}
&\sum_{j=1}^{J}\sum_{j'=1}^{J}\sum_{s\in\mathbb{S}_j}\sum_{s'\in\mathbb{S}_{j'}}\frac{\pi(1-\pi)}{J-1}g_j(s)g_{j'}(s')\bar{Y}^+_{w,j}(s,d)\bar{Y}^+_{w,j'}(s',d) \\
&= \frac{\pi(1-\pi)}{J-1}\Big(\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j} g_j(s)\bar{Y}^+_{w,j}(s,d)\Big)^2
\end{aligned}$$

The term inside the square equals zero:

$$
\begin{aligned}
&\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j} g_j(s)\bar{Y}^{+}_{w,j}(s,d) \\
&= \sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j} g_j(s)\frac{\sum_{i\in\mathbb{I}_j}\Big(\frac{w_i(s,d)}{\pi_j g_j(s)}(Y_i(s)-\mu_{w,t}(d)) + \sum_{s'\in\mathbb{S}_j}\frac{w_i(s',d)}{1-\pi_j}(Y_i(0)-\mu_{w,c}(d))\Big)}{\sum_{j'=1}^{J}\sum_{s'\in\mathbb{S}_{j'}}\sum_{i\in\mathbb{I}_{j'}} w_i(s',d)} \\
&= \sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j} g_j(s)\frac{\sum_{i\in\mathbb{I}_j}\frac{w_i(s,d)}{\pi_j g_j(s)}(Y_i(s)-\mu_{w,t}(d))}{\sum_{j'=1}^{J}\sum_{s'\in\mathbb{S}_{j'}}\sum_{i\in\mathbb{I}_{j'}} w_i(s',d)} + \sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j} g_j(s)\frac{\sum_{s'\in\mathbb{S}_j}\frac{w_i(s',d)}{1-\pi_j}(Y_i(0)-\mu_{w,c}(d))}{\sum_{j'=1}^{J}\sum_{s'\in\mathbb{S}_{j'}}\sum_{i\in\mathbb{I}_{j'}} w_i(s',d)} \\
&= \frac{1}{\pi}\underbrace{\frac{\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j} w_i(s,d)(Y_i(s)-\mu_{w,t}(d))}{\sum_{j'=1}^{J}\sum_{s'\in\mathbb{S}_{j'}}\sum_{i\in\mathbb{I}_{j'}} w_i(s',d)}}_{=0} + \frac{1}{1-\pi}\underbrace{\frac{\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j} w_i(s,d)(Y_i(0)-\mu_{w,c}(d))}{\sum_{j'=1}^{J}\sum_{s'\in\mathbb{S}_{j'}}\sum_{i\in\mathbb{I}_{j'}} w_i(s',d)}}_{=0} \\
&= 0
\end{aligned}
$$

The first equality substitutes the definition of $\bar{Y}^{+}_{w,j}(s,d)$. The second equality splits the ratio into two separate sums, one of treated, the other of control potential outcomes. For the first term, the third equality factors out $\frac{1}{\pi_j} = \frac{1}{\pi}$, which is constant across regions by Assumption 2, and cancels $g_j(s)\frac{1}{g_j(s)}$. For the second term, the third equality factors out $\frac{1}{1-\pi_j} = \frac{1}{1-\pi}$, which is constant across regions by Assumption 2, and notes that the sum of conditional probabilities is equal to 1 in each region, $\sum_s g_j(s) = 1$. Both terms are equal to zero by the definitions of $\mu_{w,t}(d)$ and $\mu_{w,c}(d)$.

Hence, under Assumption 2 (completely randomized experiment), Equation A.2 becomes

$$
\mathrm{var}\Big(\tilde{\tau}_w(d)\Big) = \sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\pi_j g_j(s)\bar{Y}^{+}_{w,j}(s,d)^2 - \sum_{j=1}^{J}\big(\pi_j^2 - \frac{\pi(1-\pi)}{J-1}\big)\Big(\sum_{s\in\mathbb{S}_j} g_j(s)\bar{Y}^{+}_{w,j}(s,d)\Big)^2
$$

where $\pi_j = \pi = \frac{J_t}{J}$ by Assumption 2.

Bernoulli trial Under Assumption 3 (Bernoulli trial), Equation A.2 becomes

$$
\mathrm{var}\Big(\tilde{\tau}_w(d)\Big) = \sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\pi_j g_j(s)\bar{Y}^{+}_{w,j}(s,d)^2 - \sum_{j=1}^{J}\pi_j^2\Big(\sum_{s\in\mathbb{S}_j} g_j(s)\bar{Y}^{+}_{w,j}(s,d)\Big)^2
$$

To treat the variances under assumptions 2 and 3 jointly, define

$$
\mathcal{C} \equiv \begin{cases} 1 & \text{under Assumption 2} \\ 0 & \text{under Assumption 3} \end{cases}
$$

such that $\mathcal{C}$ is an indicator for the completely randomized design.

Then

$$
\mathrm{var}\Big(\tilde{\tau}_w(d)\Big) = \sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\pi_j g_j(s)\bar{Y}^{+}_{w,j}(s,d)^2 - \sum_{j=1}^{J}\big(\pi_j^2 - \mathcal{C}\frac{\pi(1-\pi)}{J-1}\big)\Big(\sum_{s\in\mathbb{S}_j} g_j(s)\bar{Y}^{+}_{w,j}(s,d)\Big)^2 \tag{A.4}
$$

Consider the first term,

$$
\begin{aligned}
&\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\pi_j g_j(s)\bar{Y}^{+}_{w,j}(s,d)^2 \\
&=\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\pi_j g_j(s)\left(\frac{\sum_{i\in\mathbb{I}_j}\left(\frac{w_i(s,d)}{\pi_j g_j(s)}(Y_i(s)-\mu_{w,t}(d))+\sum_{s'\in\mathbb{S}_j}\frac{w_i(s',d)}{1-\pi_j}(Y_i(0)-\mu_{w,c}(d))\right)}{\sum_{j'=1}^{J}\sum_{s'\in\mathbb{S}_{j'}}\sum_{i\in\mathbb{I}_{j'}}w_i(s',d)}\right)^2 \\
&=\frac{\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\frac{1}{\pi_j g_j(s)}\left(\sum_{i\in\mathbb{I}_j}w_i(s,d)(Y_i(s)-\mu_{w,t}(d))\right)^2}{\left(\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j}w_i(s,d)\right)^2} \\
&\quad+\frac{\sum_{j=1}^{J}\frac{\pi_j}{(1-\pi_j)^2}\left(\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j}w_i(s,d)(Y_i(0)-\mu_{w,c}(d))\right)^2}{\left(\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j}w_i(s,d)\right)^2} \\
&\quad+2\frac{\sum_{j=1}^{J}\frac{1}{1-\pi_j}\left(\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j}w_i(s,d)(Y_i(s)-\mu_{w,t}(d))\right)\left(\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j}w_i(s,d)(Y_i(0)-\mu_{w,c}(d))\right)}{\left(\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j}w_i(s,d)\right)^2}.
\end{aligned}
$$

The first equality uses the definition of $\bar{Y}^{+}_{w,j}(s,d)$, the second equality distributes the square and factors out terms involving π_j and $g_j(s)$. The results consists of three terms: a (weighted) variance of the weighted sum of treated outcomes across candidate locations, a (weighted) variance of the weighted sum of control outcomes across regions, and a cross-product of treated and control outcomes. I rewrite this cross-product in terms of the variance of candidate-location aggregate treatment effects (and marginal variances of potential outcomes) in a later step.

Similarly for the second term

$$
\begin{aligned}
&\sum_{j=1}^{J}\big(\pi_j^2-\mathcal{C}\frac{\pi(1-\pi)}{J-1}\big)\Big(\sum_{s\in\mathbb{S}_j}g_j(s)\bar{Y}^{+}_{w,j}(s,d)\Big)^2 \\
&=\sum_{j=1}^{J}\big(\pi_j^2-\mathcal{C}\frac{\pi(1-\pi)}{J-1}\big)\left(\frac{\sum_{s\in\mathbb{S}_j}g_j(s)\sum_{i\in\mathbb{I}_j}\left(\frac{w_i(s,d)}{\pi_j g_j(s)}(Y_i(s)-\mu_{w,t}(d))+\sum_{s'\in\mathbb{S}_j}\frac{w_i(s',d)}{1-\pi_j}(Y_i(0)-\mu_{w,c}(d))\right)}{\sum_{j'=1}^{J}\sum_{s'\in\mathbb{S}_{j'}}\sum_{i\in\mathbb{I}_{j'}}w_i(s',d)}\right)^2 \\
&=\sum_{j=1}^{J}\big(\pi_j^2-\mathcal{C}\frac{\pi(1-\pi)}{J-1}\big)\left(\frac{\left(\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j}\frac{w_i(s,d)}{\pi_j}(Y_i(s)-\mu_{w,t}(d))\right)+\left(\sum_{s'\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j}\frac{w_i(s',d)}{1-\pi_j}(Y_i(0)-\mu_{w,c}(d))\right)}{\sum_{j'=1}^{J}\sum_{s'\in\mathbb{S}_{j'}}\sum_{i\in\mathbb{I}_{j'}}w_i(s',d)}\right)^2 \\
&=\frac{\sum_{j=1}^{J}\frac{\pi_j^2-\mathcal{C}\frac{\pi(1-\pi)}{J-1}}{\pi_j^2}\left(\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j}w_i(s,d)(Y_i(s)-\mu_{w,t}(d))\right)^2}{\left(\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j}w_i(s,d)\right)^2} \\
&\quad+\frac{\sum_{j=1}^{J}\frac{\pi_j^2-\mathcal{C}\frac{\pi(1-\pi)}{J-1}}{(1-\pi_j)^2}\left(\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j}w_i(s,d)(Y_i(0)-\mu_{w,c}(d))\right)^2}{\left(\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j}w_i(s,d)\right)^2} \\
&\quad+2\frac{\sum_{j=1}^{J}\frac{\pi_j^2-\mathcal{C}\frac{\pi(1-\pi)}{J-1}}{\pi_j(1-\pi_j)}\left(\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j}w_i(s,d)(Y_i(s)-\mu_{w,t}(d))\right)\left(\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j}w_i(s,d)(Y_i(0)-\mu_{w,c}(d))\right)}{\left(\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j}w_i(s,d)\right)^2}
\end{aligned}
$$

where now the variance of treated potential outcomes is also only across regions, summing over all candidate

locations within each region.

Hence, the variance of the estimator is

$$
\begin{aligned}
&\operatorname{var}\Big(\tilde{\tau}_w(d)\Big) \\
&= \frac{\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\frac{1}{\pi_j g_j(s)}\Big(\sum_{i\in\mathbb{I}_j} w_i(s,d)(Y_i(s)-\mu_{w,t}(d))\Big)^2}{\Big(\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j} w_i(s,d)\Big)^2} \\
&\quad + \frac{\sum_{j=1}^{J}\frac{\pi_j}{1-\pi_j}(1+\frac{\mathcal{C}}{J-1})\Big(\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j} w_i(s,d)(Y_i(0)-\mu_{w,c}(d))\Big)^2}{\Big(\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j} w_i(s,d)\Big)^2} \\
&\quad + 2\frac{\sum_{j=1}^{J}(1+\frac{\mathcal{C}}{J-1})\Big(\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j} w_i(s,d)(Y_i(s)-\mu_{w,t}(d))\Big)\Big(\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j} w_i(s,d)(Y_i(0)-\mu_{w,c}(d))\Big)}{\Big(\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j} w_i(s,d)\Big)^2} \\
&\quad - \frac{\sum_{j=1}^{J}(1-\mathcal{C}\frac{1-\pi_j}{\pi_j(J-1)})\Big(\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j} w_i(s,d)(Y_i(s)-\mu_{w,t}(d))\Big)^2}{\Big(\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j} w_i(s,d)\Big)^2}
\end{aligned}
$$

To substitute for the cross-product term, note that

$$
\begin{aligned}
&\frac{\sum_{j=1}^{J}(1+\frac{\mathcal{C}}{J-1})\Big(\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j} w_i(s,d)\big(Y_i(s)-Y_i(0)-(\mu_{w,t}(d)-\mu_{w,c}(d))\big)\Big)^2}{\Big(\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j} w_i(s,d)\Big)^2} \\
&= \frac{\sum_{j=1}^{J}(1+\frac{\mathcal{C}}{J-1})\Big(\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j} w_i(s,d)(Y_i(s)-\mu_{w,t}(d))\Big)^2}{\Big(\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j} w_i(s,d)\Big)^2} \\
&\quad + \frac{\sum_{j=1}^{J}(1+\frac{\mathcal{C}}{J-1})\Big(\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j} w_i(s,d)(Y_i(0)-\mu_{w,c}(d))\Big)^2}{\Big(\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j} w_i(s,d)\Big)^2} \\
&\quad - 2\frac{\sum_{j=1}^{J}(1+\frac{\mathcal{C}}{J-1})\Big(\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j} w_i(s,d)(Y_i(s)-\mu_{w,t}(d))\Big)\Big(\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j} w_i(s,d)(Y_i(0)-\mu_{w,c}(d))\Big)}{\Big(\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j} w_i(s,d)\Big)^2}
\end{aligned}
$$

where the last term is the same as the cross-product of treated and control potential outcomes in the variance

of $\tilde{\tau}_w(d)$. Hence

$$
\begin{aligned}
&\operatorname{var}\Big(\tilde{\tau}_w(d)\Big) \\
&= \frac{\sum_{j=1}^{J} \sum_{s\in\mathbb{S}_j} \frac{1}{\pi_j g_j(s)} \Big(\sum_{i\in\mathbb{I}_j} w_i(s,d)(Y_i(s) - \mu_{w,t}(d))\Big)^2}{\Big(\sum_{j=1}^{J} \sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w_i(s,d)\Big)^2} \\
&\quad + \frac{\sum_{j=1}^{J} \frac{1+\frac{\mathcal{C}}{J-1}}{1-\pi_j} \Big(\sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w_i(s,d)(Y_i(0) - \mu_{w,c}(d))\Big)^2}{\Big(\sum_{j=1}^{J} \sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w_i(s,d)\Big)^2} \\
&\quad + \frac{\sum_{j=1}^{J} \frac{\mathcal{C}}{\pi_j(J-1)} \Big(\sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w_i(s,d)(Y_i(s) - \mu_{w,t}(d))\Big)^2}{\Big(\sum_{j=1}^{J} \sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w_i(s,d)\Big)^2} \\
&\quad - \frac{\sum_{j=1}^{J} (1+\frac{\mathcal{C}}{J-1}) \Big(\sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w_i(s,d)\big(Y_i(s) - Y_i(0) - (\mu_{w,t}(d) - \mu_{w,c}(d))\big)\Big)^2}{\Big(\sum_{j=1}^{J} \sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w_i(s,d)\Big)^2}
\end{aligned}
$$

To ease interpretation, define the (weighted) pseudo-variances

$$
\begin{aligned}
\tilde{V}^{\text{location}}_{w,t}(d) &\equiv \frac{\sum_{j=1}^{J} \sum_{s\in\mathbb{S}_j} \frac{1}{\pi_j g_j(s)} \Big(\sum_{i\in\mathbb{I}_j} w_i(s,d)(Y_i(s) - \mu_{w,t}(d))\Big)^2}{\frac{J-1}{J} \sum_{j=1}^{J} \sum_{s\in\mathbb{S}_j} \frac{1}{\pi_j g_j(s)}} \\
\tilde{V}^{\text{region}}_{w,c}(d) &\equiv \frac{\sum_{j=1}^{J} \frac{1}{1-\pi_j} \Big(\sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w_i(s,d)(Y_i(0) - \mu_{w,c}(d))\Big)^2}{\frac{J-1}{J} \sum_{j=1}^{J} \frac{1}{1-\pi_j}} \\
\tilde{V}^{\text{region}}_{w,t}(d) &\equiv \frac{\sum_{j=1}^{J} \Big(\sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w_i(s,d)(Y_i(s) - \mu_{w,t}(d))\Big)^2}{J-1} \\
\tilde{V}^{\text{region}}_{w,ct}(d) &\equiv \frac{\sum_{j=1}^{J} \Big(\sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w_i(s,d)\big(Y_i(s) - Y_i(0) - (\mu_{w,t}(d) - \mu_{w,c}(d))\big)\Big)^2}{J-1}
\end{aligned}
$$

which differ from actual variances in the demeaning, as well as the (approximate) sample sizes

$$
\begin{aligned}
N_w(d) &\equiv \sum_{j=1}^{J} \sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w_i(s,d) \\
\bar{n}_w(d) &\equiv \frac{N_w(d)}{J-1} \\
\tilde{n}_{w,t}(d) &\equiv \Big(\frac{1}{\sum_{j=1}^{J} \sum_{s\in\mathbb{S}_j} \frac{1}{\pi_j g_j(s)}} N_w(d)\Big) \bar{n}_w(d) \\
\tilde{n}_{w,c}(d) &\equiv \Big(\frac{1}{\sum_{j=1}^{J} \frac{1}{1-\pi_j}} N_w(d)\Big) \bar{n}_w(d)
\end{aligned}
$$

Then the variance can be written compactly as

$$\begin{aligned}\operatorname{var}\Big(\tilde{\tau}_w(d)\Big) = &\frac{\tilde{V}^{\text{location}}_{w,t}(d)}{\tilde{n}_{w,t}(d)J} + \big(1+\frac{\mathcal{C}}{J-1}\big)\frac{\tilde{V}^{\text{region}}_{w,c}(d)}{\tilde{n}_{w,c}(d)J} \\ &+ \big(\frac{\mathcal{C}}{J-1}\big)\frac{\tilde{V}^{\text{region}}_{w,t}(d)}{\bar{n}_w(d)(\pi N_w(d))} - \big(1+\frac{\mathcal{C}}{J-1}\big)\frac{\tilde{V}^{\text{region}}_{w,ct}(d)}{\bar{n}_w(d)N_w(d)}.\end{aligned} \tag{A.5}$$

A.2.6 Variance Estimation

The variance in equation A.5 consists of four terms. The first and third terms resemble a variance of outcomes of treated individuals. The second term resembles a variance of outcomes of control individuals. The fourth term resembles a variance of treatment effects. Note that dropping the unidentified variance of treatment effects (term four) unambiguously leads to a (weakly) conservative estimator of the variance.

Estimate $\tilde{V}^{\text{location}}_{w,t}(d)$ by

$$\hat{V}^{\text{location}}_{w,t}(d) \equiv \frac{\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\frac{W_j\mathbb{1}\{\xi_j=s\}}{(\pi_j g_j(s))^2}\Big(\sum_{i\in\mathbb{I}_j} w_i(s,d)(Y_i-\bar{Y}_{w,t}(d))\Big)^2}{\frac{J-1}{J}\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\frac{W_j\mathbb{1}\{\xi_j=s\}}{(\pi_j g_j(s))^2}}$$

where $\bar{Y}_{w,t}(d)$ is the weighted average outcome of the treated

$$\bar{Y}_{w,t}(d) \equiv \frac{\sum_i\sum_{s\in\mathbb{S}_{j(i)}}\iota^t_i(s)w_i(s,d)Y_i}{\sum_i\sum_{s\in\mathbb{S}_{j(i)}}\iota^t_i(s)w_i(s,d)}$$

with

$$\iota^t_i(s) \equiv \frac{W_{j(i)}\mathbb{1}\{\xi_{j(i)}=s\}}{\pi_{j(i)}g_{j(i)}(s)}$$

as in Equation 6.

Similarly, estimate $\tilde{V}^{\text{region}}_{w,c}(d)$ by

$$\hat{V}^{\text{region}}_{w,c}(d) \equiv \frac{\sum_{j=1}^{J}\frac{1-W_j}{(1-\pi_j)^2}\Big(\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j} w_i(s,d)(Y_i-\mu_{w,c}(d))\Big)^2}{\frac{J-1}{J}\sum_{j=1}^{J}\frac{1-W_j}{(1-\pi_j)^2}}$$

where $\bar{Y}_{w,t}(d)$ is the weighted average outcome of the treated

$$\bar{Y}_{w,t}(d) \equiv \frac{\sum_i\sum_{s\in\mathbb{S}_{j(i)}}\iota^c_i w_i(s,d)Y_i}{\sum_i\sum_{s\in\mathbb{S}_{j(i)}}\iota^c_i w_i(s,d)}$$

with

$$\iota^c_i(s) \equiv \frac{1-W_{j(i)}}{1-\pi_{j(i)}}$$

as in Equation 6.

The variance of treated outcomes aggregated at the region-level, $\tilde{V}^{\text{region}}_{w,t}(d)$, cannot be estimated directly because at most one candidate treatment location is realized per region, such that the covariance of treated outcomes within region, across candidate locations, is not identified.

However, $\tilde{V}_{w,t}^{\text{region}}(d)$ can be bounded from above with an estimable bound. Observe that

$$\Big(\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j} w_i(s,d)(Y_i(s)-\mu_{w,t}(d))\Big)^2 \leq |\mathbb{S}_j|\sum_{s\in\mathbb{S}_j}\Big(\sum_{i\in\mathbb{I}_j} w_i(s,d)(Y_i(s)-\mu_{w,t}(d))\Big)^2$$

by the Cauchy-Schwarz inequality. This approximation is conservative if there is more than one candidate location in a region and if the (location-aggregate) treated potential outcomes vary across candidate locations within region. To see this, note that

$$\begin{aligned}
&|\mathbb{S}_j|\sum_{s\in\mathbb{S}_j}\Big(\sum_{i\in\mathbb{I}_j} w_i(s,d)(Y_i(s)-\mu_{w,t}(d))\Big)^2 - \Big(\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j} w_i(s,d)(Y_i(s)-\mu_{w,t}(d))\Big)^2 \\
&= \frac{1}{2}\sum_{s\in\mathbb{S}_j}\sum_{s'\in\mathbb{S}_j}\Big(\big(\sum_{i\in\mathbb{I}_j} w_i(s,d)(Y_i(s)-\mu_{w,t}(d))\big) - \big(\sum_{i'\in\mathbb{I}_j} w_{i'}(s',d)(Y_{i'}(s')-\mu_{w,t}(d))\big)\Big)^2
\end{aligned}$$

which is immediate by expanding the square on the right-hand-side. If there is only a single candidate treatment location, $|\mathbb{S}_j|=1$ such that $s=s'$ on the right-hand-side and the terms inside the square are equal. Alternatively, if there is little variation in location-aggregate treated potential outcomes

$$\sum_{i\in\mathbb{I}_j} w_i(s,d)(Y_i(s)-\mu_{w,t}(d))$$

across candidate locations s within region j, the two terms inside the square are similar even when $s\neq s'$, so the difference is close to zero again.

Hence a conservative estimator for $\tilde{V}_{w,t}^{\text{region}}(d)$ is given by

$$\hat{V}_{w,t}^{\text{region}}(d) \equiv \frac{\sum_{j=1}^{J}|\mathbb{S}_j|\sum_{s\in\mathbb{S}_j}\frac{W_j\mathbb{1}\{\xi_j=s\}}{(\pi_j g_j(s))}\Big(\sum_{i\in\mathbb{I}_j} w_i(s,d)(Y_i-\bar{Y}_{w,t}(d))\Big)^2}{J-1}.$$

The (effective) sample sizes $N_w(d)$, $\bar{n}_w(d)$, $\tilde{n}_{w,t}(d)$, and $\tilde{n}_{w,c}(d)$ depend only on known weights and treatment probabilities and can hence be calculated directly.

The estimator for the variance of $\hat{\tau}_w(d)$ is then

$$\hat{V}_w(d) \equiv \frac{\hat{V}_{w,t}^{\text{location}}(d)}{\tilde{n}_{w,t}(d)J} + \big(1+\frac{\mathcal{C}}{J-1}\big)\frac{\hat{V}_{w,c}^{\text{region}}(d)}{\tilde{n}_{w,c}(d)J} + \big(\frac{\mathcal{C}}{J-1}\big)\frac{\hat{V}_{w,t}^{\text{region}}(d)}{\bar{n}_w(d)(\pi N_w(d))}. \tag{A.6}$$

A.2.7 Covariance Across Distances

The derivation of the covariance of the estimators $\tilde{\tau}_w(d)$ and $\tilde{\tau}_{w'}(d')$ (allowing for different weight functions and distance) is largely analogous to the variance of $\tilde{\tau}_w(d)$.

First, write

$$\begin{aligned}
&\operatorname{cov}\Big(\tilde{\tau}_w(d), \tilde{\tau}_{w'}(d')\Big) \\
&= \operatorname{cov}\left(\frac{\sum_{j=1}^{J} \sum_{s\in\mathbb{S}_j} T_j(s) \sum_{i\in\mathbb{I}_j} \Big(\frac{w_i(s,d)}{\pi_j g_j(s)} (Y_i(s) - \mu_{w,t}(d)) + \sum_{s'\in\mathbb{S}_j} \frac{w_i(s',d)}{1-\pi_j} (Y_i(0) - \mu_{w,c}(d)) \Big)}{\sum_{j=1}^{J} \sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w_i(s,d)}, \right. \\
&\qquad \left. \frac{\sum_{j=1}^{J} \sum_{s\in\mathbb{S}_j} T_j(s) \sum_{i\in\mathbb{I}_j} \Big(\frac{w'_i(s,d')}{\pi_j g_j(s)} (Y_i(s) - \mu_{w',t}(d')) + \sum_{s'\in\mathbb{S}_j} \frac{w'_i(s',d)}{1-\pi_j} (Y_i(0) - \mu_{w',c}(d')) \Big)}{\sum_{j=1}^{J} \sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w'_i(s,d')} \right)
\end{aligned}$$

using the definition of $T_j(s)$ from the derivation of the variance. At this point, note that the only stochastic term, $T_j(s)$, is the same as in the variance. Hence, factor out the sums of potential outcomes, $\bar{Y}^+_{w,j}(s,d)$ and $\bar{Y}^+_{w',j}(s,d')$, to obtain

$$\begin{aligned}
\operatorname{cov}\Big(\tilde{\tau}_w(d), \tilde{\tau}_{w'}(d')\Big) &= \operatorname{cov}\Big(\sum_{j=1}^{J} \sum_{s\in\mathbb{S}_j} T_j(s) \bar{Y}^+_{w,j}(s,d), \sum_{j=1}^{J} \sum_{s\in\mathbb{S}_j} T_j(s) \bar{Y}^+_{w',j}(s,d')\Big) \\
&= \sum_{j=1}^{J} \sum_{s\in\mathbb{S}_j} \operatorname{var}\Big(T_j(s)\Big) \bar{Y}^+_{w,j}(s,d) \bar{Y}^+_{w',j}(s,d') \\
&\quad + \sum_{j=1}^{J} \sum_{s\in\mathbb{S}_j} \sum_{s'\in\mathbb{S}_j} \mathbb{1}\{s \neq s'\} \operatorname{cov}\Big(T_j(s), T_j(s')\Big) \bar{Y}^+_{w,j}(s,d) \bar{Y}^+_{w',j}(s',d') \\
&\quad + \sum_{j=1}^{J} \sum_{j'=1}^{J} \sum_{s\in\mathbb{S}_j} \sum_{s'\in\mathbb{S}_{j'}} \mathbb{1}\{j \neq j'\} \operatorname{cov}\Big(T_j(s), T_{j'}(s')\Big) \bar{Y}^+_{w,j}(s,d) \bar{Y}^+_{w',j'}(s',d')
\end{aligned}$$

Substituting the covariances of $T_j(s)$ and $T_j(s')$, and adding and subtracting the terms missing from the summation,

$$\begin{aligned}
\operatorname{cov}\Big(\tilde{\tau}_w(d), \tilde{\tau}_{w'}(d')\Big) &= \sum_{j=1}^{J} \sum_{s\in\mathbb{S}_j} \pi_j g_j(s) \bar{Y}^+_{w,j}(s,d) \bar{Y}^+_{w',j}(s,d') \\
&\quad - \sum_{j=1}^{J} \pi_j^2 \Big(\sum_{s\in\mathbb{S}_j} g_j(s) \bar{Y}^+_{w,j}(s,d) \Big) \Big(\sum_{s\in\mathbb{S}_j} g_j(s) \bar{Y}^+_{w',j}(s,d') \Big) \\
&\quad + \sum_{j=1}^{J} \sum_{j'=1}^{J} \sum_{s\in\mathbb{S}_j} \sum_{s'\in\mathbb{S}_{j'}} \mathbb{1}\{j \neq j'\} \operatorname{cov}(T_j(s), T_{j'}(s')) \bar{Y}^+_{w,j}(s,d) \bar{Y}^+_{w',j'}(s',d')
\end{aligned}$$

Similarly substituting the covariances of $T_j(s)$ and $T_{j'}(s')$ where $j \neq j'$, note again that this covariance is 0 under Assumption 3. Then adding and subtracting the missing term, and noting that still $\sum_{j=1}^{J} \sum_{s\in\mathbb{S}_j} \bar{Y}^+_{w,j}(s,d) = 0$, obtain the equivalent of Equation A.4

$$\begin{aligned}
\operatorname{cov}\Big(\tilde{\tau}_w(d), \tilde{\tau}_{w'}(d')\Big) &= \sum_{j=1}^{J} \sum_{s\in\mathbb{S}_j} \pi_j g_j(s) \bar{Y}^+_{w,j}(s,d) \bar{Y}^+_{w',j}(s,d') \\
&\quad - \sum_{j=1}^{J} \Big(\pi_j^2 - \mathcal{C} \frac{\pi(1-\pi)}{J-1}\Big) \Big(\sum_{s\in\mathbb{S}_j} g_j(s) \bar{Y}^+_{w,j}(s,d) \Big) \Big(\sum_{s\in\mathbb{S}_j} g_j(s) \bar{Y}^+_{w',j}(s,d') \Big)
\end{aligned}$$

Next, substituting the definition of $\bar{Y}^+_{w,j}(s,d)$ for the first term:

$$
\begin{aligned}
&\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\pi_j g_j(s)\bar{Y}^+_{w,j}(s,d)\bar{Y}^+_{w',j}(s,d') \\
&= \frac{\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\frac{1}{\pi_j g_j(s)}\Big(\sum_{i\in\mathbb{I}_j} w_i(s,d)(Y_i(s)-\mu_{w,t}(d))\Big)\Big(\sum_{i\in\mathbb{I}_j} w'_i(s,d')(Y_i(s)-\mu_{w',t}(d'))\Big)}{\Big(\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j} w_i(s,d)\Big)\Big(\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j} w'_i(s,d')\Big)} \\
&\quad + \frac{\sum_{j=1}^{J}\frac{\pi_j}{(1-\pi_j)^2}\Big(\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j} w_i(s,d)(Y_i(0)-\mu_{w,c}(d))\Big)\Big(\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j} w'_i(s,d')(Y_i(0)-\mu_{w',c}(d'))\Big)}{\Big(\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j} w_i(s,d)\Big)\Big(\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j} w'_i(s,d')\Big)} \\
&\quad + \frac{\sum_{j=1}^{J}\frac{1}{1-\pi_j}\Big(\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j} w_i(s,d)(Y_i(s)-\mu_{w,t}(d))\Big)\Big(\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j} w'_i(s,d')(Y_i(0)-\mu_{w',c}(d'))\Big)}{\Big(\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j} w_i(s,d)\Big)\Big(\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j} w'_i(s,d')\Big)} \\
&\quad + \frac{\sum_{j=1}^{J}\frac{1}{1-\pi_j}\Big(\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j} w'_i(s,d')(Y_i(s)-\mu_{w',t}(d'))\Big)\Big(\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j} w_i(s,d)(Y_i(0)-\mu_{w,c}(d))\Big)}{\Big(\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j} w_i(s,d)\Big)\Big(\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j} w'_i(s,d')\Big)}.
\end{aligned}
$$

Similarly for the second term

$$
\begin{aligned}
&\sum_{j=1}^{J}\big(\pi_j^2-\mathcal{C}\frac{\pi(1-\pi)}{J-1}\big)\Big(\sum_{s\in\mathbb{S}_j} g_j(s)\bar{Y}^+_{w,j}(s,d)\Big)\Big(\sum_{s\in\mathbb{S}_j} g_j(s)\bar{Y}^+_{w',j}(s,d')\Big) \\
&= \frac{\sum_{j=1}^{J}\frac{\pi_j^2-\mathcal{C}\frac{\pi(1-\pi)}{J-1}}{\pi_j^2}\Big(\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j} w_i(s,d)(Y_i(s)-\mu_{w,t}(d))\Big)\Big(\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j} w'_i(s,d')(Y_i(s)-\mu_{w',t}(d'))\Big)}{\Big(\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j} w_i(s,d)\Big)\Big(\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j} w'_i(s,d')\Big)} \\
&\quad + \frac{\sum_{j=1}^{J}\frac{\pi_j^2-\mathcal{C}\frac{\pi(1-\pi)}{J-1}}{(1-\pi_j)^2}\Big(\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j} w_i(s,d)(Y_i(0)-\mu_{w,c}(d))\Big)\Big(\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j} w'_i(s,d')(Y_i(0)-\mu_{w',c}(d'))\Big)}{\Big(\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j} w_i(s,d)\Big)\Big(\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j} w'_i(s,d')\Big)} \\
&\quad + \frac{\sum_{j=1}^{J}\frac{\pi_j^2-\mathcal{C}\frac{\pi(1-\pi)}{J-1}}{\pi_j(1-\pi_j)}\Big(\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j} w_i(s,d)(Y_i(s)-\mu_{w,t}(d))\Big)\Big(\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j} w'_i(s,d')(Y_i(0)-\mu_{w',c}(d'))\Big)}{\Big(\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j} w_i(s,d)\Big)\Big(\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j} w'_i(s,d')\Big)} \\
&\quad + \frac{\sum_{j=1}^{J}\frac{\pi_j^2-\mathcal{C}\frac{\pi(1-\pi)}{J-1}}{\pi_j(1-\pi_j)}\Big(\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j} w'_i(s,d')(Y_i(s)-\mu_{w',t}(d'))\Big)\Big(\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j} w_i(s,d)(Y_i(0)-\mu_{w,c}(d))\Big)}{\Big(\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j} w_i(s,d)\Big)\Big(\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\sum_{i\in\mathbb{I}_j} w'_i(s,d')\Big)}
\end{aligned}
$$

The covariance of $\tilde{\tau}_w(d)$ and $\tilde{\tau}_{w'}(d')$ is the difference between these two terms:

$$
\begin{aligned}
&\operatorname{cov}\Big(\tilde{\tau}_w(d), \tilde{\tau}_{w'}(d')\Big) \\
&= \frac{\sum_{j=1}^{J} \sum_{s\in\mathbb{S}_j} \frac{1}{\pi_j g_j(s)} \Big(\sum_{i\in\mathbb{I}_j} w_i(s,d)(Y_i(s) - \mu_{w,t}(d))\Big)\Big(\sum_{i\in\mathbb{I}_j} w'_i(s,d')(Y_i(s) - \mu_{w',t}(d'))\Big)}{\Big(\sum_{j=1}^{J} \sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w_i(s,d)\Big)\Big(\sum_{j=1}^{J} \sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w'_i(s,d')\Big)} \\
&\quad + \frac{\sum_{j=1}^{J} \frac{\pi_j}{1-\pi_j}(1 + \frac{\mathcal{C}}{J-1})\Big(\sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w_i(s,d)(Y_i(0) - \mu_{w,c}(d))\Big)\Big(\sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w'_i(s,d')(Y_i(0) - \mu_{w',c}(d'))\Big)}{\Big(\sum_{j=1}^{J} \sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w_i(s,d)\Big)\Big(\sum_{j=1}^{J} \sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w'_i(s,d')\Big)} \\
&\quad + \frac{\sum_{j=1}^{J} (1 + \frac{\mathcal{C}}{J-1})\Big(\sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w_i(s,d)(Y_i(s) - \mu_{w,t}(d))\Big)\Big(\sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w'_i(s,d')(Y_i(0) - \mu_{w',c}(d'))\Big)}{\Big(\sum_{j=1}^{J} \sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w_i(s,d)\Big)\Big(\sum_{j=1}^{J} \sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w'_i(s,d')\Big)} \\
&\quad + \frac{\sum_{j=1}^{J} (1 + \frac{\mathcal{C}}{J-1})\Big(\sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w'_i(s,d')(Y_i(s) - \mu_{w',t}(d'))\Big)\Big(\sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w_i(s,d)(Y_i(0) - \mu_{w,c}(d))\Big)}{\Big(\sum_{j=1}^{J} \sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w_i(s,d)\Big)\Big(\sum_{j=1}^{J} \sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w'_i(s,d')\Big)} \\
&\quad - \frac{\sum_{j=1}^{J} (1 - \mathcal{C}\frac{1-\pi_j}{\pi_j(J-1)})\Big(\sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w_i(s,d)(Y_i(s) - \mu_{w,t}(d))\Big)\Big(\sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w'_i(s,d')(Y_i(s) - \mu_{w',t}(d'))\Big)}{\Big(\sum_{j=1}^{J} \sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w_i(s,d)\Big)\Big(\sum_{j=1}^{J} \sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w'_i(s,d')\Big)}
\end{aligned}
$$

There are two terms involving products of (demeaned) treated and control potential outcomes. To substitute for these terms, note that To substitute for the cross-product term, note that

$$
\begin{aligned}
&\sum_{j=1}^{J} (1 + \frac{\mathcal{C}}{J-1}) \frac{\Big(\sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w_i(s,d)\big(Y_i(s) - Y_i(0) - (\mu_{w,t}(d) - \mu_{w,c}(d))\big)\Big)}{\Big(\sum_{j=1}^{J} \sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w_i(s,d)\Big)} \\
&\qquad \cdot \frac{\Big(\sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w'_i(s,d')\big(Y_i(s) - Y_i(0) - (\mu_{w',t}(d') - \mu_{w',c}(d'))\big)\Big)}{\Big(\sum_{j=1}^{J} \sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w'_i(s,d')\Big)} \\
&= \frac{\sum_{j=1}^{J} (1 + \frac{\mathcal{C}}{J-1})\Big(\sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w_i(s,d)(Y_i(s) - \mu_{w,t}(d))\Big)\Big(\sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w'_i(s,d')(Y_i(s) - \mu_{w',t}(d'))\Big)}{\Big(\sum_{j=1}^{J} \sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w_i(s,d)\Big)\Big(\sum_{j=1}^{J} \sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w'_i(s,d')\Big)} \\
&\quad + \frac{\sum_{j=1}^{J} (1 + \frac{\mathcal{C}}{J-1})\Big(\sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w_i(s,d)(Y_i(0) - \mu_{w,c}(d))\Big)\Big(\sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w'_i(s,d')(Y_i(0) - \mu_{w',c}(d'))\Big)}{\Big(\sum_{j=1}^{J} \sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w_i(s,d)\Big)\Big(\sum_{j=1}^{J} \sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w'_i(s,d')\Big)} \\
&\quad - \frac{\sum_{j=1}^{J} (1 + \frac{\mathcal{C}}{J-1})\Big(\sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w_i(s,d)(Y_i(s) - \mu_{w,t}(d))\Big)\Big(\sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w'_i(s,d')(Y_i(0) - \mu_{w',c}(d'))\Big)}{\Big(\sum_{j=1}^{J} \sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w_i(s,d)\Big)\Big(\sum_{j=1}^{J} \sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w'_i(s,d')\Big)} \\
&\quad - \frac{\sum_{j=1}^{J} (1 + \frac{\mathcal{C}}{J-1})\Big(\sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w'_i(s,d')(Y_i(s) - \mu_{w',t}(d'))\Big)\Big(\sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w_i(s,d)(Y_i(0) - \mu_{w,c}(d))\Big)}{\Big(\sum_{j=1}^{J} \sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w_i(s,d)\Big)\Big(\sum_{j=1}^{J} \sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w'_i(s,d')\Big)}
\end{aligned}
$$

where the last two terms on the right-hand-side are the same products of treated and control potential outcomes as in the variance (but of opposite sign).

Substituting for these two product terms in the variance,

$$
\begin{aligned}
&\operatorname{cov}\Big(\tilde{\tau}_w(d), \tilde{\tau}_{w'}(d')\Big) \\
&= \frac{\sum_{j=1}^{J} \sum_{s\in\mathbb{S}_j} \frac{1}{\pi_j g_j(s)} \Big(\sum_{i\in\mathbb{I}_j} w_i(s,d)(Y_i(s) - \mu_{w,t}(d))\Big)\Big(\sum_{i\in\mathbb{I}_j} w'_i(s,d')(Y_i(s) - \mu_{w',t}(d'))\Big)}{\Big(\sum_{j=1}^{J} \sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w_i(s,d)\Big)\Big(\sum_{j=1}^{J} \sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w'_i(s,d')\Big)} \\
&\quad + \frac{\sum_{j=1}^{J} \frac{1+\frac{\mathcal{C}}{J-1}}{1-\pi_j} \Big(\sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w_i(s,d)(Y_i(0) - \mu_{w,c}(d))\Big)\Big(\sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w'_i(s,d')(Y_i(0) - \mu_{w',c}(d'))\Big)}{\Big(\sum_{j=1}^{J} \sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w_i(s,d)\Big)\Big(\sum_{j=1}^{J} \sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w'_i(s,d')\Big)} \\
&\quad + \frac{\sum_{j=1}^{J} \frac{\mathcal{C}}{\pi_j(J-1)} \Big(\sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w_i(s,d)(Y_i(s) - \mu_{w,t}(d))\Big)\Big(\sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w'_i(s,d')(Y_i(s) - \mu_{w',t}(d'))\Big)}{\Big(\sum_{j=1}^{J} \sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w_i(s,d)\Big)\Big(\sum_{j=1}^{J} \sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w'_i(s,d')\Big)} \\
&\quad - (1 + \frac{\mathcal{C}}{J-1}) \sum_{j=1}^{J} \Bigg(\frac{\Big(\sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w_i(s,d)\big(Y_i(s) - Y_i(0) - (\mu_{w,t}(d) - \mu_{w,c}(d))\big)\Big)}{\Big(\sum_{j=1}^{J} \sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w_i(s,d)\Big)} \\
&\qquad \cdot \frac{\Big(\sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w'_i(s,d')\big(Y_i(s) - Y_i(0) - (\mu_{w',t}(d') - \mu_{w',c}(d'))\big)\Big)}{\Big(\sum_{j=1}^{J} \sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w'_i(s,d')\Big)} \Bigg)
\end{aligned}
$$

To ease interpretation, define the (weighted) pseudo-covariances

$$
\begin{aligned}
\tilde{V}^{\text{location}}_{w,w',t}(d,d') &\equiv \frac{\sum_{j=1}^{J} \sum_{s\in\mathbb{S}_j} \frac{1}{\pi_j g_j(s)} \Big(\sum_{i\in\mathbb{I}_j} w_i(s,d)(Y_i(s) - \mu_{w,t}(d))\Big)\Big(\sum_{i\in\mathbb{I}_j} w'_i(s,d')(Y_i(s) - \mu_{w',t}(d'))\Big)}{\frac{J-1}{J}\sum_{j=1}^{J} \sum_{s\in\mathbb{S}_j} \frac{1}{\pi_j g_j(s)}} \\
\tilde{V}^{\text{region}}_{w,w',c}(d,d') &\equiv \frac{\sum_{j=1}^{J} \frac{1}{1-\pi_j} \Big(\sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w_i(s,d)(Y_i(0) - \mu_{w,c}(d))\Big)\Big(\sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w'_i(s,d')(Y_i(0) - \mu_{w',c}(d'))\Big)}{\frac{J-1}{J}\sum_{j=1}^{J} \frac{1}{1-\pi_j}} \\
\tilde{V}^{\text{region}}_{w,w',t}(d,d') &\equiv \frac{\sum_{j=1}^{J} \Big(\sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w_i(s,d)(Y_i(s) - \mu_{w,t}(d))\Big)\Big(\sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w'_i(s,d')(Y_i(s) - \mu_{w',t}(d'))\Big)}{J-1} \\
\tilde{V}^{\text{region}}_{w,w',ct}(d,d') &\equiv \frac{1}{J-1} \sum_{j=1}^{J} \Bigg(\Big(\sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w_i(s,d)\big(Y_i(s) - Y_i(0) - (\mu_{w,t}(d) - \mu_{w,c}(d))\big)\Big) \\
&\qquad \cdot \Big(\sum_{s\in\mathbb{S}_j} \sum_{i\in\mathbb{I}_j} w'_i(s,d')\big(Y_i(s) - Y_i(0) - (\mu_{w',t}(d') - \mu_{w',c}(d'))\big)\Big) \Bigg)
\end{aligned}
$$

Then

$$
\begin{aligned}
\operatorname{cov}\Big(\tilde{\tau}_w(d), \tilde{\tau}_{w'}(d')\Big) &= \frac{\tilde{V}^{\text{location}}_{w,w',t}(d,d')}{\sqrt{\tilde{n}_{w,t}(d)} \cdot \sqrt{\tilde{n}_{w',t}(d')} \cdot J} \\
&\quad + (1 + \frac{\mathcal{C}}{J-1}) \frac{\tilde{V}^{\text{region}}_{w,w',c}(d,d')}{\sqrt{\tilde{n}_{w,c}(d)} \cdot \sqrt{\tilde{n}_{w',c}(d')} \cdot J} \\
&\quad + \frac{\mathcal{C}}{J-1} \frac{\tilde{V}^{\text{region}}_{w,w',t}(d,d')}{\pi \cdot \sqrt{\bar{n}_w(d)} \cdot \sqrt{\bar{n}_{w'}(d')} \cdot \sqrt{N_w(d)} \cdot \sqrt{N_{w'}(d')}} \\
&\quad - (1 + \frac{\mathcal{C}}{J-1}) \frac{\tilde{V}^{\text{region}}_{w,w',ct}(d,d')}{\sqrt{\bar{n}_w(d)} \cdot \sqrt{\bar{n}_{w'}(d')} \cdot \sqrt{N_w(d)} \cdot \sqrt{N_{w'}(d')}}
\end{aligned}
$$

A.3 Proof of Theorem 4

The estimator is

$$\hat{\tau}^{AATT,2} \equiv \sum_{d\in\mathbb{D}} \bar{n}(d)\hat{\tau}(d)$$

where

$$\bar{n}(d) = \frac{\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\pi_j g_j(s)\sum_{i:\ |d(s,r_i)-d|\leq h} 1}{\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\pi_j g_j(s)}.$$

Since $\bar{n}(d)$ is non-stochastic and by Theorem 1,

$$\begin{aligned}
E(\hat{\tau}^{AATT,2}) &= \sum_{d\in\mathbb{D}} \bar{n}(d)E(\hat{\tau}(d)) \\
&\approx \sum_{d\in\mathbb{D}} \bar{n}(d)\frac{\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\pi_j g_j(s)\sum_{i:\ |d(s,r_i)-d|\leq h}\tau_i(s)}{\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\pi_j g_j(s)\sum_{i:\ |d(s,r_i)-d|\leq h}1} \\
&= \sum_{d\in\mathbb{D}} \frac{\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\pi_j g_j(s)\sum_{i:\ |d(s,r_i)-d|\leq h}\tau_i(s)}{\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\pi_j g_j(s)} \\
&= \frac{\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\pi_j g_j(s)\sum_{d\in\mathbb{D}}\sum_{i:\ |d(s,r_i)-d|\leq h}\tau_i(s)}{\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\pi_j g_j(s)} \\
&= \frac{\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\pi_j g_j(s)\sum_{i\in\mathbb{I}_j}\tau_i(s)}{\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j}\pi_j g_j(s)} \\
&= \tau^{AATT}
\end{aligned}$$

where the second-to-last equality holds because every individual is in exactly one distance bin from each candidate treatment location.

Next, consider the variance of the estimator.

$$\begin{aligned}
\operatorname{var}\big(\hat{\tau}^{AATT,2}\big) &= \sum_{d\in\mathbb{D}} \bar{n}(d)^2 \operatorname{var}\big(\hat{\tau}(d)\big) + 2\sum_{d\in\mathbb{D}}\sum_{d'\in\mathbb{D},d'\neq d} \bar{n}(d)\bar{n}(d')\operatorname{cov}\big(\hat{\tau}(d),\hat{\tau}(d')\big) \\
&\approx \sum_{d\in\mathbb{D}} \bar{n}(d)^2 \operatorname{var}\big(\tilde{\tau}(d)\big) + 2\sum_{d\in\mathbb{D}}\sum_{d'\in\mathbb{D},d'\neq d} \bar{n}(d)\bar{n}(d')\operatorname{cov}\big(\tilde{\tau}(d),\tilde{\tau}(d')\big)
\end{aligned}$$

by Theorems 1 and 3.

For the covariances across distance, simplify the the result of Theorem 3 for the ATT weights

$$w_i(s,d) = w_i'(s,d) = \pi_{j(i)}g_{j(i)}(s)\mathbb{1}\{|d(s,r_i)-d|\leq h\}$$

to obtain

$$\operatorname{cov}\Big(\tilde{\tau}(d),\tilde{\tau}(d')\Big) = \frac{J-1}{J}\frac{\tilde{V}_t^{\text{location}}(d,d')}{J_t} + \frac{\tilde{V}_c^{\text{region}}(d,d')}{J_c} + \frac{1}{J}\frac{\tilde{V}_t^{\text{region}}(d,d')}{J_t} - \frac{\tilde{V}_{ct}^{\text{region}}(d,d')}{J}$$

where

$$\tilde{V}_t^{\text{location}}(d,d') \equiv \frac{1}{\bar{n}(d)\cdot\bar{n}(d')\cdot(J-1)}\sum_{j=1}^{J}\sum_{s\in\mathbb{S}_j} g_j(s)\Bigg(\Big(\sum_{i:\ |d(s,r_i)-d|\le h}(Y_i(s)-\mu_t(d))\Big)\cdot\Big(\sum_{i:\ |d(s,r_i)-d'|\le h}(Y_i(s)-\mu_t(d'))\Big)\Bigg)$$

$$\tilde{V}_c^{\text{region}}(d,d') \equiv \frac{1}{\bar{n}(d)\cdot\bar{n}(d')\cdot(J-1)}\sum_{j=1}^{J}\Bigg(\Big(\sum_{s\in\mathbb{S}_j} g_j(s)\sum_{i:\ |d(s,r_i)-d|\le h}(Y_i(0)-\mu_c(d))\Big)\cdot\Big(\sum_{s\in\mathbb{S}_j} g_j(s)\sum_{i:\ |d(s,r_i)-d'|\le h}(Y_i(0)-\mu_c(d'))\Big)\Bigg)$$

$$\tilde{V}_t^{\text{region}}(d,d') \equiv \frac{1}{\bar{n}(d)\cdot\bar{n}(d')\cdot(J-1)}\sum_{j=1}^{J}\Bigg(\Big(\sum_{s\in\mathbb{S}_j} g_j(s)\sum_{i:\ |d(s,r_i)-d|\le h}(Y_i(s)-\mu_t(d))\Big)\cdot\Big(\sum_{s\in\mathbb{S}_j} g_j(s)\sum_{i:\ |d(s,r_i)-d'|\le h}(Y_i(s)-\mu_t(d'))\Big)\Bigg)$$

$$\tilde{V}_{ct}^{\text{region}}(d,d') \equiv \frac{1}{\bar{n}(d)\cdot\bar{n}(d')\cdot(J-1)}\sum_{j=1}^{J}\Bigg(\Big(\sum_{s\in\mathbb{S}_j} g_j(s)\sum_{i:\ |d(s,r_i)-d|\le h}\big(Y_i(s)-Y_i(0)-(\mu_t(d)-\mu_c(d))\big)\Big)\cdot\Big(\sum_{s\in\mathbb{S}_j} g_j(s)\sum_{i:\ |d(s,r_i)-d'|\le h}\big(Y_i(s)-Y_i(0)-(\mu_t(d')-\mu_c(d'))\big)\Big)\Bigg)$$

and $\bar{n}(d)$, $\mu_t(d)$, and $\mu_c(d)$ are defined as in Theorem 1.

Then

$$\begin{aligned}\text{var}\Big(\hat{\tau}^{AATT,2}\Big) = \sum_{d\in\mathbb{D}}\bar{n}(d)^2\Big(\frac{J-1}{J}\frac{\tilde{V}_t^{\text{location}}(d)}{J_t}+\frac{\tilde{V}_c^{\text{region}}(d)}{J_c}+\frac{1}{J}\frac{\tilde{V}_t^{\text{region}}(d)}{J_t}-\frac{\tilde{V}_{ct}^{\text{region}}(d)}{J}\Big)\\ +2\sum_{d\in\mathbb{D}}\sum_{d'\in\mathbb{D},d'\neq d}\bar{n}(d)\bar{n}(d')\Big(\frac{J-1}{J}\frac{\tilde{V}_t^{\text{location}}(d,d')}{J_t}+\frac{\tilde{V}_c^{\text{region}}(d,d')}{J_c}\\ +\frac{1}{J}\frac{\tilde{V}_t^{\text{region}}(d,d')}{J_t}-\frac{\tilde{V}_{ct}^{\text{region}}(d,d')}{J}\Big)\end{aligned}$$

Consider the (co-) variance of treatment effect terms:

$$\begin{aligned}&-\sum_{d\in\mathbb{D}}\bar{n}(d)^2\frac{\tilde{V}_{ct}^{\text{region}}(d)}{J}-2\sum_{d\in\mathbb{D}}\sum_{d'\in\mathbb{D},d'\neq d}\bar{n}(d)\bar{n}(d')\frac{\tilde{V}_{ct}^{\text{region}}(d,d')}{J}\\ &=-\frac{1}{J}\frac{\sum_{j=1}^{J}\Big(\sum_{d\in\mathbb{D}}\sum_{s\in\mathbb{S}_j} g_j(s)\sum_{i:\ |d(s,r_i)-d|\le h}\big(Y_i(s)-Y_i(0)-(\mu_t(d)-\mu_c(d))\big)\Big)^2}{J-1}\end{aligned}$$

which follows immediately by expanding the square and substituting the definitions of $\tilde{V}_{ct}^{\text{region}}(d)$ and $\tilde{V}_{ct}^{\text{region}}(d,d')$. Hence, dropping these terms leads to a conservative expression for the variance of the aggregate treatment effect estimator.

A.4 Nonparametric Estimators Under Assumptions on Interference

A.4.1 Additively Separable Treatment Effects

To estimate the average treatment effect on the treated, $\tau(d)$, under the assumption of additively separable treatment effects, one can use the estimator

$$\hat{\tau}^{\text{additive}}(d) = \frac{\sum_j \sum_{s_j \in \mathbb{S}_j} \sum_{i \in \mathbb{I}_j} \Pr(s \in \xi_j) \mathbb{1}\{|d(s, r_i - d| \leq h\} \hat{\tau}_i^{\text{additive}}(s)}{\sum_{j=1}^{J} \sum_{s_j \in \mathbb{S}_j} \sum_{i \in \mathbb{I}_j} \Pr(s \in \xi_j) \mathbb{1}\{|d(s, r_i - d| \leq h\}}$$

where

$$\hat{\tau}_i^{\text{additive}}(s) = \frac{\mathbb{1}\{s \in \xi_j\}}{\Pr(s \in \xi_j)} Y_i - \left(\frac{\tilde{n}_j - 1}{|\mathbb{S}_j| - 1} \cdot \frac{W_j \mathbb{1}\{s \notin \xi_j\}}{\Pr(W_j = 1) \Pr(s \notin \xi_j | W_j = 1)} \right) Y_i - \frac{1 - W_{j(i)}}{1 - \Pr(W_{j(i)} = 1)} Y_i$$

where the first term picks up all assignments with treatment at location s, the second term removes the effects of all assignments with treatment in region j but without treatment at s, and the third term estimates the control potential outcome. The estimator $\hat{\tau}_i^{\text{additive}}(s)$ here generalizes the estimator in Section 5.1 of the main text to allow for an arbitrary number of candidate locations in each region, $|\mathbb{S}_j|$. The formula is specific to completely randomized designs within treated regions with fixed number $\tilde{n}_j$ of realized locations, and equal probability for each of the $\binom{|\mathbb{S}_j|}{\tilde{n}_j}$ possible combinations of treated locations. Under this design and Assumption 5 of additively separable treatment effects, $E(\hat{\tau}_i^{\text{additive}}(s)) = \tau_i(s)$. Consequently, $\hat{\tau}^{\text{additive}}(d)$ is an unbiased estimator of the average treatment effect on the treated, $\tau(d)$.

A.4.2 Only Nearest Realized Treatment Location Matters

Under Assumption 6, it is only possible to identify the effect of a candidate treatment location s on individual i, $\tau_i(s)$, if s is the closest realized treatment location to i with positive probability. One can take

$$\hat{\tau}_w^{\text{nearest}}(d) = \frac{1}{\sum_j \sum_{s \in \mathbb{S}_j} \sum_{i \in \mathbb{I}_j} w_i(s, d)} \sum_j \sum_{s \in \mathbb{S}_j} \sum_{i \in \mathbb{I}_j} w_i(s, d) \hat{\tau}_i^{\text{nearest}}(s)$$

where

$$\hat{\tau}_i^{\text{nearest}}(s) = \frac{\mathbb{1}\{s \in S_j\}}{\Pr(s \in S_j)} \frac{\prod_{s' \in S_j} \mathbb{1}\{d(s, r_i) \leq d(s', r_i)\}}{\Pr\left(\prod_{s' \in S_j} \mathbb{1}\{d(s, r_i) \leq d(s', r_i)\} = 1 | s \in S_j\right)} Y_i - \frac{1 - W_{j(i)}}{1 - \Pr(W_{j(i)} = 1)} Y_i.$$

The estimator $\hat{\tau}_i^{\text{nearest}}(s)$ is equal to 0 if the region of individual i is treated but location s is not the closest realized treatment location to i. This happens both when s is not realized itself, and when another realized treatment location s' is closer to i. If s is the closest realized location to i, $\hat{\tau}_i^{\text{nearest}}(s)$ is equal to the outcome of i scaled by the inverse of the probability of this event. If the region j is not treated, $\hat{\tau}_i^{\text{nearest}}(s)$ is equal to the outcome of i scaled by the inverse of the probability of region j not being treated. Th estimator $\hat{\tau}_i^{\text{nearest}}(s)$ is an unbiased inverse probability weighting estimator of $\tau_i(s) \equiv Y_i(s) - Y_i(0)$ under the assumption that only the nearest realized treatment matters. Consequently, $\hat{\tau}^{\text{nearest}}(d)$ is an unbiased estimator of the weighted average treatment effect $\tau_w(d)$, as long as $\tau_w(d)$ places no weight on individual and treatment location pairs where the treatment location is never the nearest realized location to the individual.

A.4.3 Proof of Theorem 5

Approximate Estimator

As in previous proofs, I derive exact results for an approximate estimator and show that the approximate estimator is close to the estimator of interest. Define the approximate estimator for settings with interference as

$$\tilde{\tau}(d) \equiv \mu_t(d) - \mu_c(d) + \frac{\sum_{s\in\mathbb{S}} \mathbb{1}\{s \in \xi\} \sum_{i\in\mathbb{I}} \mathbb{1}\{|d(s,r_i) - d| \leq h\}(Y_i - \mu_t(d))}{\sum_{s\in\mathbb{S}} \Pr(s \in \xi) \sum_{i\in\mathbb{I}} \mathbb{1}\{|d(s,r_i) - d| \leq h\}} - \frac{\sum_{s\in\mathbb{S}} \mathbb{1}\{s \notin \xi\} \frac{\Pr(\xi\cup\{s\})}{\Pr(\xi)} \sum_{i\in\mathbb{I}} \mathbb{1}\{|d(s,r_i) - d| \leq h\}(Y_i - \mu_c(d))}{\sum_{s\in\mathbb{S}} \Pr(s \in \xi) \sum_{i\in\mathbb{I}} \mathbb{1}\{|d(s,r_i) - d| \leq h\}}$$

where

$$\mu_t(d) \equiv \frac{\sum_{S\in 2^{\mathbb{S}}} \sum_{s\in\mathbb{S}} \Pr(\xi = S) \mathbb{1}\{s \in S\} \sum_{i\in\mathbb{I}} \mathbb{1}\{|d(s,r_i) - d| \leq h\} Y_i(S)}{\sum_{s\in\mathbb{S}} \Pr(s \in \xi) \sum_{i\in\mathbb{I}} \mathbb{1}\{|d(s,r_i) - d| \leq h\}}$$

$$\mu_c(d) \equiv \frac{\sum_{S\in 2^{\mathbb{S}}} \sum_{s\in\mathbb{S}} \Pr(\xi = S\cup\{s\}) \mathbb{1}\{s \notin S\} \sum_{i\in\mathbb{I}} \mathbb{1}\{|d(s,r_i) - d| \leq h\} Y_i(S)}{\sum_{s\in\mathbb{S}} \Pr(s \in \xi) \sum_{i\in\mathbb{I}} \mathbb{1}\{|d(s,r_i) - d| \leq h\}}$$

Quality of Approximation

Here, I show that the estimators $\hat{\tau}(d)$ and $\tilde{\tau}(d)$ are *very close* in large enough samples. This motivates the use of exact finite sample results for the mean and variance of the infeasible estimator $\tilde{\tau}(d)$ for inference with the feasible estimator $\hat{\tau}(d)$. The analysis uses the mean value theorem to derive the difference $\hat{\tau}(d) - \tilde{\tau}(d)$ and argues that this difference is small in large enough samples.

As a practical matter, a sample is large enough if the number of individuals near treatment and control are close to their expected values. The approximation of $\hat{\tau}(d)$ by $\tilde{\tau}(d)$ is particular close when also the average outcomes are close to their expected values.

To simplify notation, define the following shorthands:

$$\hat{\mu}_{w,t}(d) = \frac{\sum_{s\in\mathbb{S}} \mathbb{1}\{s \in \xi\} \sum_{i\in\mathbb{I}} \mathbb{1}\{|d(s,r_i) - d| \leq h\} Y_i}{\sum_{s\in\mathbb{S}} \mathbb{1}\{s \in \xi\} \sum_{i\in\mathbb{I}} \mathbb{1}\{|d(s,r_i) - d| \leq h\}}$$

$$\hat{\mu}_{w,c}(d) = \frac{\sum_{s\in\mathbb{S}} \mathbb{1}\{s \notin \xi\} \frac{\Pr(\xi\cup\{s\})}{\Pr(\xi)} \sum_{i\in\mathbb{I}} \mathbb{1}\{|d(s,r_i) - d| \leq h\} Y_i}{\sum_{s\in\mathbb{S}} \mathbb{1}\{s \notin \xi\} \frac{\Pr(\xi\cup\{s\})}{\Pr(\xi)} \sum_{i\in\mathbb{I}} \mathbb{1}\{|d(s,r_i) - d| \leq h\}}$$

$$\tilde{\mu}_{w,t}(d) = \frac{\sum_{s\in\mathbb{S}} \mathbb{1}\{s \in \xi\} \sum_{i\in\mathbb{I}} \mathbb{1}\{|d(s,r_i) - d| \leq h\} Y_i}{\sum_{s\in\mathbb{S}} \Pr(s \in \xi) \sum_{i\in\mathbb{I}} \mathbb{1}\{|d(s,r_i) - d| \leq h\}}$$

$$\tilde{\mu}_{w,c}(d) = \frac{\sum_{s\in\mathbb{S}} \mathbb{1}\{s \notin \xi\} \frac{\Pr(\xi\cup\{s\})}{\Pr(\xi)} \sum_{i\in\mathbb{I}} \mathbb{1}\{|d(s,r_i) - d| \leq h\} Y_i}{\sum_{s\in\mathbb{S}} \Pr(s \in \xi) \sum_{i\in\mathbb{I}} \mathbb{1}\{|d(s,r_i) - d| \leq h\}}$$

and

$$\hat{p}_{w,t}(d) = \frac{1}{|\mathbb{S}|} \sum_{s\in\mathbb{S}} \mathbb{1}\{s \in \xi\} \sum_{i\in\mathbb{I}} \mathbb{1}\{|d(s,r_i) - d| \leq h\}$$

$$\hat{p}_{w,c}(d) = \frac{1}{|\mathbb{S}|} \sum_{s\in\mathbb{S}} \mathbb{1}\{s \notin \xi\} \frac{\Pr(\xi\cup\{s\})}{\Pr(\xi)} \sum_{i\in\mathbb{I}} \mathbb{1}\{|d(s,r_i) - d| \leq h\}$$

$$p_w(d) = \frac{1}{|\mathbb{S}|} \sum_{s\in\mathbb{S}} \Pr(s \in \xi) \sum_{i\in\mathbb{I}} \mathbb{1}\{|d(s,r_i) - d| \leq h\}$$

where $\tilde{\mu}_{w,t}(d)$ and $\tilde{\mu}_{w,c}(d)$ replicate $\hat{\mu}_{w,t}(d)$ and $\hat{\mu}_{w,c}(d)$ but with expected values rather than sample averages in the denominators. The sample average denominators are $\hat{p}_{w,t}(d)$ and $\hat{p}_{w,c}(d)$ (scaled such that they

converge under suitable conditions when J grows), and the expected value of the denominators is $p_w(d)$ (similarly scaled).

Without loss of generality, I fix the distance d and weighting w of interest and suppress the dependence on d and w in the following derivations for ease of presentation.

The feasible estimator written in terms of the shorthand notation is

$$\hat{\tau} = \hat{\mu}_t - \hat{\mu}_c = \frac{p}{\hat{p}_t}\tilde{\mu}_t - \frac{p}{\hat{p}_c}\tilde{\mu}_c$$

while the infeasible estimator is

$$\tilde{\tau} = \mu_t - \mu_c + \tilde{\mu}_t - \frac{\hat{p}_t}{p}\mu_t - \tilde{\mu}_c + \frac{\hat{p}_c}{p}\mu_c.$$

Next, define the function

$$\begin{aligned}\tilde{\Delta}(\hat{p}_t, \hat{p}_c, \tilde{\mu}_t, \tilde{\mu}_c) &\equiv \frac{p}{\hat{p}_t}\tilde{\mu}_t - \frac{p}{\hat{p}_c}\tilde{\mu}_c - (\mu_t - \mu_c + \tilde{\mu}_t - \frac{\hat{p}_t}{p}\mu_t - \tilde{\mu}_c + \frac{\hat{p}_c}{p}\mu_c) \\ &= \hat{\tau} - \tilde{\tau}\end{aligned}$$

Finally, apply the mean value theorem on $\tilde{\Delta}$ on the interval with endpoints $\hat{x} = (\hat{p}_t, \hat{p}_c, \tilde{\mu}_t, \tilde{\mu}_c)$ and $x = (p, p, \mu_t, \mu_c)$. The mean value theorem states that

$$\begin{aligned}&\tilde{\Delta}(\hat{p}_t, \hat{p}_c, \tilde{\mu}_t, \tilde{\mu}_c) - \tilde{\Delta}(p, p, \mu_t, \mu_c) \\ &= \begin{bmatrix} \hat{p}_t - p_t \\ \hat{p}_c - p_c \\ \tilde{\mu}_t - \mu_t \\ \tilde{\mu}_c - \mu_c \end{bmatrix} \cdot \begin{bmatrix} \frac{\partial \tilde{\Delta}}{\partial \hat{p}_t}(\dot{x}) \\ \frac{\partial \tilde{\Delta}}{\partial \hat{p}_c}(\dot{x}) \\ \frac{\partial \tilde{\Delta}}{\partial \tilde{\mu}_t}(\dot{x}) \\ \frac{\partial \tilde{\Delta}}{\partial \tilde{\mu}_c}(\dot{x}) \end{bmatrix}\end{aligned}$$

where $\dot{x}$ is some convex combination of $\hat{x}$ and x.

It is straightforward to see that $\tilde{\Delta}(p, p, \mu_t, \mu_c) = 0$. Hence the left-hand-side of the equality above is just $\tilde{\Delta}(\hat{p}_t, \hat{p}_c, \tilde{\mu}_t, \tilde{\mu}_c)$, such that the right-hand-side is an expression for $\hat{\tau} - \tilde{\tau}$.

Hence

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

Each of the four terms is a product with each factor close to zero under appropriate asymptotics. That is, the difference between the estimators $\hat{\tau}_w(d)$ and $\tilde{\tau}_w(d)$ is negligible under standard asymptotic frameworks. Since the difference between estimators is *very small* for large samples, exact finite sample results for $\tilde{\tau}_w(d)$ likely provide decent approximations for $\hat{\tau}_w(d)$ in smaller samples.

First, note that

$$
\begin{aligned}
&E\left(\frac{\sum_{s\in\mathbb{S}}\mathbb{1}\{s\in\xi\}\sum_{i\in\mathbb{I}}\mathbb{1}\{|d(s,r_i)-d|\le h\}Y_i}{\sum_{s\in\mathbb{S}}\Pr(s\in\xi)\sum_{i\in\mathbb{I}}\mathbb{1}\{|d(s,r_i)-d|\le h\}}\right)\\
&=E\left(\frac{\sum_{S\in2^{\mathbb{S}}}\sum_{s\in\mathbb{S}}\mathbb{1}\{\xi=S\}\mathbb{1}\{s\in S\}\sum_{i\in\mathbb{I}}\mathbb{1}\{|d(s,r_i)-d|\le h\}Y_i(S)}{\sum_{s\in\mathbb{S}}\Pr(s\in\xi)\sum_{i\in\mathbb{I}}\mathbb{1}\{|d(s,r_i)-d|\le h\}}\right)\\
&=\frac{\sum_{S\in2^{\mathbb{S}}}\sum_{s\in\mathbb{S}}\Pr\{\xi=S\}\mathbb{1}\{s\in S\}\sum_{i\in\mathbb{I}}\mathbb{1}\{|d(s,r_i)-d|\le h\}Y_i(S)}{\sum_{s\in\mathbb{S}}\Pr(s\in\xi)\sum_{i\in\mathbb{I}}\mathbb{1}\{|d(s,r_i)-d|\le h\}}\\
&=\mu_t(d)
\end{aligned}
$$

and similarly

$$
\begin{aligned}
&E\left(\frac{\sum_{s\in\mathbb{S}}\mathbb{1}\{s\notin\xi\}\frac{\Pr(\xi\cup\{s\})}{\Pr(\xi)}\sum_{i\in\mathbb{I}}\mathbb{1}\{|d(s,r_i)-d|\le h\}Y_i}{\sum_{s\in\mathbb{S}}\Pr(s\in\xi)\sum_{i\in\mathbb{I}}\mathbb{1}\{|d(s,r_i)-d|\le h\}}\right)\\
&=E\left(\frac{\sum_{S\in2^{\mathbb{S}}}\sum_{s\in\mathbb{S}}\mathbb{1}\{\xi=S\}\mathbb{1}\{s\notin S\}\frac{\Pr(\xi=S\cup\{s\})}{\Pr(\xi=S)}\sum_{i\in\mathbb{I}}\mathbb{1}\{|d(s,r_i)-d|\le h\}Y_i(S)}{\sum_{s\in\mathbb{S}}\Pr(s\in\xi)\sum_{i\in\mathbb{I}}\mathbb{1}\{|d(s,r_i)-d|\le h\}}\right)\\
&=\frac{\sum_{S\in2^{\mathbb{S}}}\sum_{s\in\mathbb{S}}\Pr(\xi=S\cup\{s\})\mathbb{1}\{s\notin S\}\sum_{i\in\mathbb{I}}\mathbb{1}\{|d(s,r_i)-d|\le h\}Y_i(S)}{\sum_{s\in\mathbb{S}}\Pr(s\in\xi)\sum_{i\in\mathbb{I}}\mathbb{1}\{|d(s,r_i)-d|\le h\}}\\
&=\mu_c(d)
\end{aligned}
$$

Hence $E(\tilde{\tau}(d))=\mu_t(d)-\mu_c(d)$.

Furthermore

$$
\begin{aligned}
\mu_t(d)-\mu_c(d)&=\frac{\sum_{S\in2^{\mathbb{S}}}\sum_{s\in\mathbb{S}}\Pr(\xi=S)\mathbb{1}\{s\in S\}\sum_{i\in\mathbb{I}}\mathbb{1}\{|d(s,r_i)-d|\le h\}Y_i(S)}{\sum_{s\in\mathbb{S}}\Pr(s\in\xi)\sum_{i\in\mathbb{I}}\mathbb{1}\{|d(s,r_i)-d|\le h\}}\\
&\quad-\frac{\sum_{S\in2^{\mathbb{S}}}\sum_{s\in\mathbb{S}}\Pr(\xi=S\cup\{s\})\mathbb{1}\{s\notin S\}\sum_{i\in\mathbb{I}}\mathbb{1}\{|d(s,r_i)-d|\le h\}Y_i(S)}{\sum_{s\in\mathbb{S}}\Pr(s\in\xi)\sum_{i\in\mathbb{I}}\mathbb{1}\{|d(s,r_i)-d|\le h\}}\\
&=\frac{\sum_{S\in2^{\mathbb{S}}}\sum_{s\in\mathbb{S}}\Pr(\xi=S)\mathbb{1}\{s\in S\}\sum_{i\in\mathbb{I}}\mathbb{1}\{|d(s,r_i)-d|\le h\}(Y_i(S)-Y_i(S\setminus\{s\}))}{\sum_{s\in\mathbb{S}}\Pr(s\in\xi)\sum_{i\in\mathbb{I}}\mathbb{1}\{|d(s,r_i)-d|\le h\}}
\end{aligned}
$$

by aligning the summand S of the first sum with the summand $S\setminus\{s\}$ of the second sum. This proves part (i) of Theorem 5.

Under additivity, $Y_i(S)-Y_i(S\setminus\{s\})=\tau_i(s)$ for all $S\subset\mathbb{S}$ with $s\in S$, so

$$
\begin{aligned}
\mu_t(d)-\mu_c(d)&=\frac{\sum_{S\in2^{\mathbb{S}}}\sum_{s\in\mathbb{S}}\Pr(\xi=S)\mathbb{1}\{s\in S\}\sum_{i\in\mathbb{I}}\mathbb{1}\{|d(s,r_i)-d|\le h\}\tau_i(s)}{\sum_{s\in\mathbb{S}}\Pr(s\in\xi)\sum_{i\in\mathbb{I}}\mathbb{1}\{|d(s,r_i)-d|\le h\}}\\
&=\frac{\sum_{s\in\mathbb{S}}\sum_{i\in\mathbb{I}}\mathbb{1}\{|d(s,r_i)-d|\le h\}\tau_i(s)\Big(\sum_{S\in2^{\mathbb{S}}}\Pr(\xi=S)\mathbb{1}\{s\in S\}\Big)}{\sum_{s\in\mathbb{S}}\Pr(s\in\xi)\sum_{i\in\mathbb{I}}\mathbb{1}\{|d(s,r_i)-d|\le h\}}\\
&=\frac{\sum_{s\in\mathbb{S}}\Pr(s\in\xi)\sum_{i\in\mathbb{I}}\mathbb{1}\{|d(s,r_i)-d|\le h\}\tau_i(s)}{\sum_{s\in\mathbb{S}}\Pr(s\in\xi)\sum_{i\in\mathbb{I}}\mathbb{1}\{|d(s,r_i)-d|\le h\}}
\end{aligned}
$$

proving part (iii) of Theorem 5.

Variance

Using the exposure mappings as defined in Remark 20 in the main text,[29] the variance of $\tilde{\tau}(d)$ is

$$\begin{aligned}
\operatorname{var}(\tilde{\tau}(d)) = \operatorname{var}\Bigg(& \mu_t(d) - \mu_c(d) + \frac{\sum_{s\in\mathbb{S}} \mathbb{1}\{s \in \xi\} \sum_{i\in\mathbb{I}} \mathbb{1}\{|d(s,r_i) - d| \leq h\}(Y_i - \mu_t(d))}{\sum_{s\in\mathbb{S}} \Pr(s \in \xi) \sum_{i\in\mathbb{I}} \mathbb{1}\{|d(s,r_i) - d| \leq h\}} \\
& - \frac{\sum_{s\in\mathbb{S}} \mathbb{1}\{s \notin \xi\} \frac{\Pr(\xi\cup\{s\})}{\Pr(\xi)} \sum_{i\in\mathbb{I}} \mathbb{1}\{|d(s,r_i) - d| \leq h\}(Y_i - \mu_c(d))}{\sum_{s\in\mathbb{S}} \Pr(s \in \xi) \sum_{i\in\mathbb{I}} \mathbb{1}\{|d(s,r_i) - d| \leq h\}} \Bigg) \\
= \operatorname{var}\Bigg(& \frac{\sum_{s\in\mathbb{S}} \sum_{m\in\mathbb{M}_s} \mathbb{1}\{f_s(\xi) = m\} W_s(m) \sum_{i\in\mathbb{I}} \mathbb{1}\{|d(s,r_i) - d| \leq h\}(Y_i(m) - \mu_t(d))}{\sum_{s\in\mathbb{S}} \Pr(s \in \xi) \sum_{i\in\mathbb{I}} \mathbb{1}\{|d(s,r_i) - d| \leq h\}} \\
& - \frac{\sum_{s\in\mathbb{S}} \sum_{m\in\mathbb{M}_s} \mathbb{1}\{f_s(\xi) = m\}(1 - W_s(m)) \frac{\Pr(\xi\cup\{s\})}{\Pr(\xi)} \sum_{i\in\mathbb{I}} \mathbb{1}\{|d(s,r_i) - d| \leq h\}(Y_i(m) - \mu_c(d))}{\sum_{s\in\mathbb{S}} \Pr(s \in \xi) \sum_{i\in\mathbb{I}} \mathbb{1}\{|d(s,r_i) - d| \leq h\}} \Bigg)
\end{aligned}$$

Under Assumption 7 (independent treatment assignment), $\Pr(\xi = S) = \prod_{s\in S} \pi_s \prod_{s\notin S}(1 - \pi_s)$, so

$$\begin{aligned}
\operatorname{var}(\tilde{\tau}(d)) = \operatorname{var}\Bigg(& \frac{\sum_{s\in\mathbb{S}} \sum_{m\in\mathbb{M}_s} \mathbb{1}\{f_s(\xi) = m\} W_s(m) \sum_{i\in\mathbb{I}} \mathbb{1}\{|d(s,r_i) - d| \leq h\}(Y_i(m) - \mu_t(d))}{\sum_{s\in\mathbb{S}} \Pr(s \in \xi) \sum_{i\in\mathbb{I}} \mathbb{1}\{|d(s,r_i) - d| \leq h\}} \\
& - \frac{\sum_{s\in\mathbb{S}} \sum_{m\in\mathbb{M}_s} \mathbb{1}\{f_s(\xi) = m\}(1 - W_s(m)) \frac{\pi_s}{1-\pi_s} \sum_{i\in\mathbb{I}} \mathbb{1}\{|d(s,r_i) - d| \leq h\}(Y_i(m) - \mu_c(d))}{\sum_{s\in\mathbb{S}} \Pr(s \in \xi) \sum_{i\in\mathbb{I}} \mathbb{1}\{|d(s,r_i) - d| \leq h\}} \Bigg)
\end{aligned}$$

Let $Y_s(m,d)$, the non-stochastic signed sum of potential outcomes of individuals near candidate location s under exposure m be defined as in the statement of Theorem 5.

Then

$$\begin{aligned}
& \operatorname{var}(\tilde{\tau}(d)) \\
& = \operatorname{var}\left(\frac{\sum_{s\in\mathbb{S}} \sum_{m\in\mathbb{M}_s} \mathbb{1}\{f_s(\xi) = m\} Y_s(m,d)}{\sum_{s\in\mathbb{S}} \Pr(s \in \xi) \sum_{i\in\mathbb{I}} \mathbb{1}\{|d(s,r_i) - d| \leq h\}} \right) \\
& = \sum_{s\in\mathbb{S}} \operatorname{var}\left(\frac{\sum_{m\in\mathbb{M}_s} \mathbb{1}\{f_s(\xi) = m\} Y_s(m,d)}{\sum_{s\in\mathbb{S}} \Pr(s \in \xi) \sum_{i\in\mathbb{I}} \mathbb{1}\{|d(s,r_i) - d| \leq h\}} \right) \\
& \quad + \sum_{s\in\mathbb{S}} \sum_{s'\in\mathbb{S}} \mathbb{1}\{s \neq s'\} \operatorname{cov}\Bigg(\frac{\sum_{m\in\mathbb{M}_s} \mathbb{1}\{f_s(\xi) = m\} Y_s(m,d)}{\sum_{s\in\mathbb{S}} \Pr(s \in \xi) \sum_{i\in\mathbb{I}} \mathbb{1}\{|d(s,r_i) - d| \leq h\}}, \\
& \qquad\qquad \frac{\sum_{m'\in\mathbb{M}_{s'}} \mathbb{1}\{f_{s'}(\xi) = m'\} Y_s(m,d)}{\sum_{s\in\mathbb{S}} \Pr(s \in \xi) \sum_{i\in\mathbb{I}} \mathbb{1}\{|d(s,r_i) - d| \leq h\}} \Bigg) \\
& = \frac{\sum_{s\in\mathbb{S}} \sum_{m\in\mathbb{M}_s} \operatorname{var}\Big(\mathbb{1}\{f_s(\xi) = m\}\Big) Y_s(m,d)^2}{(\sum_{s\in\mathbb{S}} \Pr(s \in \xi) \sum_{i\in\mathbb{I}} \mathbb{1}\{|d(s,r_i) - d| \leq h\})^2} \\
& \quad + \frac{\sum_{s\in\mathbb{S}} \sum_{m\in\mathbb{M}_s} \sum_{m'\in\mathbb{M}_s} \mathbb{1}\{m \neq m'\} \operatorname{cov}\Big(\mathbb{1}\{f_s(\xi) = m\}, \mathbb{1}\{f_s(\xi) = m'\}\Big) Y_s(m,d) Y_s(m',d)}{(\sum_{s\in\mathbb{S}} \Pr(s \in \xi) \sum_{i\in\mathbb{I}} \mathbb{1}\{|d(s,r_i) - d| \leq h\})^2} \\
& \quad + \frac{\sum_{s\in\mathbb{S}} \sum_{s'\in\mathbb{S}} \mathbb{1}\{s \neq s'\} \sum_{m\in\mathbb{M}_s} \sum_{m'\in\mathbb{M}_{s'}} \operatorname{cov}\Big(\mathbb{1}\{f_s(\xi) = m\}, \mathbb{1}\{f_{s'}(\xi) = m'\}\Big) Y_s(m,d) Y_{s'}(m',d)}{(\sum_{s\in\mathbb{S}} \Pr(s \in \xi) \sum_{i\in\mathbb{I}} \mathbb{1}\{|d(s,r_i) - d| \leq h\})^2}
\end{aligned}$$

[29] The derivations in the first part of this section assume that exposure mappings are correctly specified, in the sense that outcomes are fixed conditional on exposure, but do not rely on any particular definition of exposure mappings.

where the only remaining stochastic elements are indicators such as $\mathbb{1}\{f_s(\xi) = m\}$.

Let the marginal and joint probabilities of exposure be

$$\pi_s(m) = \Pr(f_s(\xi) = m)$$
$$\pi_{s,s'}(m, m') = \Pr(f_s(\xi) = m \wedge f_{s'}(\xi) = m')$$

as defined in Remark 20.

Then

$$\begin{aligned}\operatorname{var}(\tilde{\tau}(d)) = &\frac{\sum_{s\in\mathbb{S}}\sum_{m\in\mathbb{M}_s}\pi_s(m)(1-\pi_s(m))Y_s(m,d)^2}{(\sum_{s\in\mathbb{S}}\Pr(s\in\xi)\sum_{i\in\mathbb{I}}\mathbb{1}\{|d(s,r_i)-d|\le h\})^2}\\ &-\frac{\sum_{s\in\mathbb{S}}\sum_{m\in\mathbb{M}_s}\sum_{m'\in\mathbb{M}_s}\mathbb{1}\{m\neq m'\}\pi_s(m)\pi_s(m')Y_s(m,d)Y_s(m',d)}{(\sum_{s\in\mathbb{S}}\Pr(s\in\xi)\sum_{i\in\mathbb{I}}\mathbb{1}\{|d(s,r_i)-d|\le h\})^2}\\ &+\frac{\sum_{s\in\mathbb{S}}\sum_{s'\in\mathbb{S}}\mathbb{1}\{s\neq s'\}\sum_{m\in\mathbb{M}_s}\sum_{m'\in\mathbb{M}_{s'}}(\pi_{s,s'}(m,m')-\pi_s(m)\pi_{s'}(m'))Y_s(m,d)Y_{s'}(m',d)}{(\sum_{s\in\mathbb{S}}\Pr(s\in\xi)\sum_{i\in\mathbb{I}}\mathbb{1}\{|d(s,r_i)-d|\le h\})^2}\end{aligned}$$

B Details on the Empirical Application

In this appendix, I outline choices made in the processing of the data and finding counterfactual treatment locations. I also show robustness to alternative specifications and random seeds in the design and use of the convolutional neural networks.

B.1 Data Processing

I use the July 2021 release of SafeGraph's data for the year 2020. In this release of the data, SafeGraph applies its current algorithm to the data it collected in 2020, and updates its data sets attributing smartphone pings to businesses. In this paper, I focus on businesses in the San Francisco Bay Area, specifically in the Peninsula and South Bay between South San Francisco and Sunnyvale, see Figure 9 in the main text. To create the initial sample of all possibly relevant businesses for which SafeGraph has recorded data, I keep all businesses that either lie within six miles from a number of points throughout the Bay Area or have a SafeGraph determined ZIP code falling within a list of relevant ZIP codes, see Table B.1.

To define the units of interest and ensure high-quality data for this application, I take three additional steps processing the data. First, I determine the grocery and convenience stores that I consider "treatments" in this paper. Second, I manually set the location of each of these treatments to correspond to the main entrance of the store. Third, I check and de-duplicate restaurant location data to restrict to real restaurants that were likely to be open in early 2020.

Based on SafeGraph's "point of interest" data, I find 167 unique grocery and convenience store (treatment) locations that were open in 2020 in the interior of the study area. Starting from the sample defined above, I define the possible businesses of interest as those within 3 miles of Burlingame, 5 miles of Belmont, 5.5 miles of Menlo Park, or 2.95 miles of Mountain View, with the city locations as in Table B.1. Focusing on grocery stores in the interior of the study area guarantees that the full sample includes data on all businesses that are within different distances of interest from the grocery stores. To find locations consumers typically visit to purchase groceries, I start with all businesses with 4-digit NAICS code 4451 (grocery and convenience stores) assigned by SafeGraph, and then add all Costco, Target, and Walmart stores (which SafeGraph classifies as general merchandise stores, 4523), for a total of 313 stores. Of these stores, I exclude 28 stores that

Table B.1: The initial sample of all possibly relevant businesses consists of all businesses in the SafeGraph "point of interest" data with location within six miles from one of the five cities or with a zip code given in the table.

city	latitude	longitude
South San Francisco	37.653540	-122.416866
Burlingame	37.584103	-122.366083
Belmont	37.516493	-122.294191
Menlo Park	37.451967	-122.177993
Mountain View	37.389389	-122.083210
ZIP codes:		
94002, 94005, 94010, 94014, 94015, 94016, 94019, 94020, 94022, 94024, 94025, 94027, 94028, 94030, 94032, 94035, 94037, 94040, 94041, 94042, 94043, 94044, 94061, 94062, 94063, 94064, 94065, 94066, 94070, 94080, 94083, 94085, 94086, 94087, 94089, 94101, 94102, 94104, 94105, 94110, 94112, 94114, 94117, 94121, 94124, 94127, 94128, 94129, 94130, 94131, 94132, 94133, 94134, 94169, 94192, 94301, 94303, 94304, 94305, 94306, 94309, 94401, 94402, 94403, 94404, 94497, 94530, 94538, 94555, 94603, 95014, 95015, 95051, 95054, 95101, 95112		

SafeGraph determines to have closed permanently before the COVID-19 pandemic (in or before February 2020; there were no further grocery store closures until July as recorded by SafeGraph), as well as 1 store that SafeGraph determines to have opened only in November 2020. For the remaining 284 stores, I verify manually that these fit my definition of grocery or convenience store. This leads to the exclusion of 100 stores; primarily excluding convenience stores that are part of gas stations, delis, and food producers and importers / exporters that are incorrectly classified as grocery stores by SafeGraph's algorithm. I was able to confirm, based on newspaper articles, Yelp entries, and Google Street View imagery, that another 17 grocery stores were either not open in 2020 (closed before or opened after) or were duplicate entries in the data set. Hence, overall I consider 167 treatment locations; 139 of these are labeled as grocery (or general merchandise) stores by SafeGraph, with the remaining 28 labeled as convenience stores by SafeGraph.

For the 167 grocery and convenience stores in the sample, I manually determine the latitude and longitude of the main entrance, which serves two related purposes. First, the main entrance and exit is the relevant location to measure distances to or from for the purpose of trip sequencing: If a consumer considers visiting a coffee shop before or after a grocery store, the additional distance she has to travel is based on the front door of the grocery store, not a location in the interior. Second, placing the location of grocery stores at their main entrances typically reduces the differences between taking straight-line distance (as in this paper) and walking distance (likely the economically relevant distance metric) between grocery store and restaurant. When the grocery store location is instead placed in the interior of the store, restaurants that are *behind* the grocery store can appear closer than restaurants that are next door. Hence, placing the location of the grocery store at its front entrance improves the interpretability of estimates by distance. The latitude and longitude given in the SafeGraph data instead reflect "the general center of the business"[30], typically in the interior of the

[30]SafeGraph documentation, https://docs.safegraph.com/docs/core-places#section-latitude-longitude accessed on July 29, 2021.

store. I use Google Maps satellite as well as Street View imagery to locate the main entrances of all grocery stores. For about three quarters of the grocery and convenience stores, the difference in locations is less than 20 meters. The largest differences in locations (of around 70 meters) occur for a handful of particularly large Costco, Safeway, Target, and Walmart stores.

I audit the data on restaurant (outcome unit) locations in three steps. First, I de-duplicate observations by checking for similarity of business names between any two businesses with locations within 50 meters of each other according to SafeGraph data. To detect duplicates that are plausibly noticeable based on name similarity, I specifically focus my attention on businesses with high relative Levenshtein distance, which measures the minimum numbers of character edits needed to make the names of the two businesses equal, relative to the length of the longer business name. Most duplicates I detect are clear typos in the name of one of the observations, and some are abbreviations of business names that I verify to indeed describe the same business using Google Maps and Street View data. Second, I audit the SafeGraph location data by comparing the latitude and longitude in the SafeGraph "point of interest" data to the latitude and longitude obtained by running the business name and street address (also from the "point of interest" data) through Google Maps. This analysis confirms the high quality of the SafeGraph location data. Randomly inspecting the locations of a few dozen restaurant in more detail, I find that neither the SafeGraph nor the Google maps locations are systematically closer to the entrance of the restaurants. Given the much smaller size (area) of almost all restaurants compared to grocery stores, as well as the much greater number of restaurants, I do not manually record the latitudes and longitudes of their entrances. Third, I focus on businesses that were reliably being assigned visits by SafeGraph. I restrict the non-grocery store sample to businesses for which SafeGraph reported at least 7 visits in each of the four weeks starting in January 2020. This excludes businesses that were not open at the time, not properly assigned visits by SafeGraph's algorithm, or are too small to reliably measure visits for, but retains 95-97.5% of all *visits* (depending on the week) in the SafeGraph data. Importantly, each of the three steps is taken without knowledge of which businesses are, in the later analysis, considered treated or control.

B.2 Convolutional Neural Network

I use a convolutional neural network (CNN) to identify plausible counterfactual locations. First, I specify the input into the training of the CNN. Second, I describe the architecture of the CNN. Third, I use the trained CNN to predict many plausible counterfactual locations and additional matching steps to select the final counterfactual locations used in the analysis.

I project latitude and longitude of all businesses into two-dimensional Cartesian space using the NAD83 (2011) projection, EPSG:6419 California zone 3. This projection gives the location in meters East and North relative to a point near the San Francisco Bay Area. In applications where data come from different regions, different projections may be needed for different regions such that the relative distances within each region are accurate.

The CNN learns to predict treatment locations in the larger neighborhoods of prespecified locations: real grocery store locations and semi-randomly chosen locations. The semi-randomly chosen locations, together with the real grocery store locations, are meant to cover the larger neighborhoods that counterfactual locations could plausibly occur in. I start with the locations of all businesses for which the nearest grocery store is between 0.2 miles and 2 miles away. The neighborhoods of businesses even closer to a grocery store are already included in the consideration set by including that grocery store. Next, I jitter these locations by adding independent shocks from a normal distribution with mean ± 0.0004 and standard deviation 0.0001

Table B.2: Number of businesses by 4-digit NAICS code that are in the larger neighborhoods forming the input into the convolutional neural network. The number of grocery stores exceeds 167 here because additional grocery stores that are not in the interior of the main study area are included in these larger neighborhoods.

NAICS code	description	# unique businesses
7225	Restaurants and Other Eating Places	1975
7139	Other Amusement and Recreation Industries	606
7121	Museums, Historical Sites, and Similar Institutions	409
8131	Religious Organizations	324
6111	Elementary and Secondary Schools	265
6244	Child Day Care Services	264
4451	Grocery Stores	244
4471	Gasoline Stations	182
any	–	7845

to their latitudes and longitudes, where the sign of the mean is independently drawn to be $+1$ or -1 for each location and coordinate. This step ensures that the *center* of each larger neighborhood does not fall exactly onto a business because real grocery store locations never exactly coincide with the locations of other businesses. Finally, to avoid including the same neighborhoods multiple times, I detect all pairs of jittered locations that are within 100 meters of one another. I drop locations that are listed "first" (in the arbitrary order based on the row numbers of the businesses the location is based on) in any such pair. In total, this results in $1,900$ semi-random locations, in addition to the 167 real grocery store locations, that form the centers of the larger neighborhoods used by the CNN.

The CNN predictions are based on observable characteristics describing small 2D grid cells around the prespecified locations. Each grid cell covers an area of 0.025mi $\times$ 0.025mi (approximately 40m $\times$ 40m). The observable characteristics of each grid cell that I use in this application is the count of businesses by 4-digit NAICS code for the codes given in Table B.2. That is, a cell covering two gasoline stations, one car dealership, and no other businesses, will have "covariate value" 2 for the covariate indicating industry group 4471 (gasoline stations) and 3 for the covariate indicating "any" industry, with the remaining covariates at 0 because there is no separate covariate for the relatively rare car dealerships (NAICS code 4411, less than 100 in the study area). One could similarly define covariates describing the average income or age of the census tract covering the cell, or property values for properties in each grid cell.

The CNN receives as its input observations consisting of one larger neighborhood each. The covariates of the 2D grid are separate "channels" constituting a 3D tensor for each such neighborhood-observation. Each larger neighborhood consists of 50×50 grid cells. The entire neighborhood is randomly shifted, rotated, and mirrored, with the constraint that the prespecified location is placed within one of the central 10×10 grid cells (uniformly distributed), such that it has at least another 20 grid cells (0.5mi) of "padding" until the edge of the larger neighborhood.

The CNN consists of 4 sequential 2D convolutions and a final linear (fully connected) layer yielding $10 \times 10 + 1 = 101$ outputs. I use 2D instance normalization and leaky rectified linear activation for all neurons in the CNN, and replication padding to ensure the output of each convolution has the same spatial dimension as the input. The first convolution takes the 9 input channels (eight specific industries and one for any industry) and convolves it with a kernel size of 5 (considering the 5×5 grid cells centered around a given grid cell) into 18 channels. This layer can "smooth" the input such that the hard borders between grid cells due to discretization become less relevant. The increase in the number of channels allows the neural

network to learn a larger number of non-linearities. The second convolution takes the 18 channels of the previous layer and convolves them with a kernel size of 21 with a stride of 2 into 36 channels, such that each grid cell can view grid cells up to 20 cells away in any direction, but skipping every other cell for parsimony. This layer allows each grid cell to learn about its own neighborhood up to even relatively large distances (approximately $20 \times 0.025\text{mi} = 0.5\text{mi}$). The third convolution takes the 36 channels of the previous layer and convolves them with a kernel size of 5 into 36 channels, again allowing some smoothing across grid cells to counteract the skipping of every other grid cell of the previous layer. The fourth convolution takes the 36 channels of the previous layer and convolves them with a kernel size of 21 with a stride of 2 into a single channel. Intuitively, this layer forces a single prediction for each grid cell based on the large neighborhood (up to 20 cells away in any direction). The final layer linearly combines the 50×50 grid cells of the single channel of the previous layer into 101 "categories" that constitute the predictions of whether and where an additional grocery store may be located.

The 101 categories correspond to the central 10×10 grid as well as one category indicating a prediction of no additional grocery store. I train the CNN on batches consisting of 64 large neighborhood observations. Half (32) of the observations are large neighborhoods around a real grocery store, but with that grocery store removed from the input channel count of grocery stores per grid cell. The random shift and rotation in input is such that this removed grocery store could have been in any of the central 10×10 grid cells. For these observations, the prediction maximizing the cross-entropy loss is the category corresponding to the particular cell that the grocery store has been removed from. All other categories are equal to one another in terms of loss, and worse than the correct category, which trains the CNN to identify the mode, rather than average, location. A quarter (16) of the observations are large neighborhoods around a real grocery store, with no grocery store removed. The correct classification of such observations is into the category corresponding to "no missing grocery store" instead of any of the 10×10 grid cells. The last quarter (16) of the observations of each batch are large neighborhoods around the semi-random prespecified locations. Their correct classification is also the category corresponding to "no missing grocery store."

After training, I evaluate the neighborhoods of the prespecified locations for possible grocery store locations according to the CNN. In this step, I input batches consisting of 32 observation into the trained CNN. In these batches, 4 observations are large neighborhoods around real grocery stores: 2 have the grocery store removed from the input, while 2 do not have the grocery store removed. An additional 28 observations are large neighborhoods around the semi-random prespecified locations. I make predictions for $5,000$ such batches. The predictions for observations with removed grocery stores allow me to learn the prediction values (activation) of real grocery store locations. The remaining observations yield possible counterfactual locations.

I find good matches for real grocery store locations among the possible counterfactual locations in two steps. In the first step, I find separately for each real grocery store location possible counterfactual locations with similar CNN activation. Specifically, I take each prediction for a removed real grocery store separately (there are multiple such predictions for each real grocery store under different random shifts, rotation, and mirroring), and match in descending order of activation, without replacement, within the possible counterfactual locations (this excludes the prediction category for "no missing grocery store"). I repeat the same matching process (matching without replacement using the complete set of possible counterfactual locations) using relative activation within observation, corresponding to the cross-entropy loss function. Taking the union of these matches, I obtain $19,857$ possible locations that the CNN evaluated as similar to a real grocery store location under at least one shift, rotation, and mirroring. In the second step, I use propensity score matching to pick the final counterfactual locations among these $19,857$ possible locations. I estimate a propensity score model

using the real and possible counterfactual locations as observations in a logistic regression. There are three sets of regressors: the numbers of restaurants in distance bins of width 0.025 miles from the location, up to a distance of 0.25 miles; the average number of grocery stores near the restaurants in each bin broken out for each bin into similar bins of distance from the restaurant; and the total number of businesses (of any industry) in distance bins of width 0.25 miles, up to a distance of 1 mile. I match each grocery store location, without replacement, to the nearest possible counterfactual location as well as the next higher and lower counterfactual location based on the estimated propensity scores. I exclude from this list one location of any pair of possible counterfactual locations that are closer to one another than the smallest distance between any pair of real grocery store locations. I perform a final matching step (again on the estimated propensity scores) excluding these "too close" locations and keeping for each real grocery store location its closest matches and the location with the next *larger* propensity score to retain the most plausible counterfactual locations.

For the final sample of real grocery store and most plausible counterfactual locations, I estimate a propensity score to analyze the sample as a quasi-experiment conditional on these locations and propensity scores. The propensity score estimation uses the same regressors as the estimated propensity score for matching. This propensity score is only used to weight the "control" observations (restaurants near counterfactual locations) because the average treatment effect on the treated (ATT) estimator does not require reweighting of the "treated" observations (restaurants near real grocery stores). The primary purpose is to balance exposure to grocery stores appropriately between treated and control restaurants. When estimating the average effect of one marginal grocery store on restaurants at a distance d, the treated and control restaurants at that distance indeed differ on average by one grocery store at distance d, and have similar average exposure to grocery stores at other distances as Figure 14 in the main text illustrates. By selecting the counterfactual locations from the CNN predictions based on the relative locations of other businesses in the large neighborhoods, these locations and propensity score weights also balance exposure to other businesses in the neighborhood as shown in Figure 15 of the main text.