Geographic Boundaries as Regression Discontinuities

Luke J. Keele

*Associate Professor, Department of Political Science, 211 Pond Lab,
Penn State University, University Park, PA 16802
e-mail: ljk20@psu.edu*

Rocío Titiunik

*Assistant Professor, Department of Political Science, University of Michigan,
5700 Haven Hall, 505 South State Street, Ann Arbor, MI 48109-1045
e-mail: titiunik@umich.edu (corresponding author)*

Edited by Betsy Sinclair

Political scientists often turn to natural experiments to draw causal inferences with observational data. Recently, the regression discontinuity design (RD) has become a popular type of natural experiment due to its relatively weak assumptions. We study a special type of regression discontinuity design where the discontinuity in treatment assignment is geographic. In this design, which we call the Geographic Regression Discontinuity (GRD) design, a geographic or administrative boundary splits units into treated and control areas, and analysts make the case that the division into treated and control areas occurs in an as-if random fashion. We show how this design is equivalent to a standard RD with two running variables, but we also clarify several methodological differences that arise in geographical contexts. We also offer a method for estimation of geographically located treatment effects that can also be used to validate the identification assumptions using observable pretreatment characteristics. We illustrate our methodological framework with a re-examination of the effects of political advertisements on voter turnout during a presidential campaign, exploiting the exogenous variation in the volume of presidential ads that is created by media market boundaries.

1 Introduction

Selection and endogeneity are often key threats to inference in the social sciences. Recently, analysts have turned to natural experiments and quasi-experimental methods as one way to overcome these obstacles in observational studies—studies where the assignment of the treatment of interest is not under the researcher's control. Among these techniques, the regression discontinuity (RD) design has been revived with great fanfare. Lee and Lemieux (2010, 282) summarize the promise that surrounds this design, attributing the recent wave of RD studies to "the belief that the RD design is not 'just another' evaluation strategy and that causal inferences from RD designs are potentially more credible than those from typical 'natural experiment' strategies." Moreover, analysts have used RD designs to recover experimental benchmarks, which have only bolstered their credibility (Cook, Shadish, and Wong 2008; Green et al. 2009). The use of RD designs has exploded recently.

Lee and Lemieux (2010) count 78 applications of RD designs in economics, and the design is spreading quickly in political science (see, e.g., Broockman 2009; Eggers and Hainmueller 2009; Caughey and Sekhon 2011; Gerber, Kessler, and Meredith 2011; Gerber and Hopkins 2011; Trounstine 2011; Eggers et al. 2014).

In this article, we study what we call the Geographic Regression Discontinuity (GRD) design, a design in which a geographic or administrative boundary splits units into treated and control areas and analysts make the case that the division into treated and control areas occurs in an as-if random fashion. One of the earliest and most famous examples of exploiting geographic variation to estimate causal effects is the study by Card and Krueger (1994), who estimated the effect of increasing the minimum wage on employment by comparing fast-food restaurants in New Jersey (where the minimum wage was increased) to restaurants in adjacent eastern Pennsylvania. In political science, political boundaries are often associated with variation in key treatments such as national or state institutions. For example, Posner (2004) used the colonial border between Zambia and Malawi, which was drawn by the British South African Company and split two different ethnic groups, to study the political salience of cultural cleavages. Research designs based on geographic discontinuities are an increasingly popular type of natural experiment in political science, and have been recently used to study a variety of topics, including nation building, governance and ethnic relations in Africa (Asiwaju 1985; Laitin 1986; Miles and Rochefort 1991; Miles 1994; Miguel 2004; Posner 2004; Berger 2009), media effects in Europe and the United States (Huber and Arceneaux 2007; Kern and Hainmueller 2008; Krasno and Green 2008), local policies in U.S. cities (Gerber, Kessler, and Meredith 2011), and mobilization and polarization in the American electorate (Middleton and Green 2008; Nall 2012).

In this article, we clarify the methodological difficulties and opportunities that may arise in geographic applications of the RD framework. We first show that GRD designs lead to identification of the local treatment effect at the boundary under a two-dimensional continuity assumption that generalizes the seminal identification assumption in Hahn et al. (2001). In this regard, the GRD design behaves as any other standard RD design with two scores or running variables. However, applying the two-dimensional continuity assumption to geography produces some subtle but important differences. We highlight three in particular. First, analysts who study GRDs encounter *compound* treatments—multiple treatments that affect the outcome of interest simultaneously—more frequently than those who study nongeographic RD designs. Second, in a GRD design, different measures of distance from the cutoffs may require different identification assumptions and affect fundamentally the interpretation of the results. Third, spatial variation in treatment effects can be mapped to specific locations, which can be used to detect geographic areas where the identification assumptions are more (or less) likely to hold. In other words, unlike nongeographic two-dimensional RDs, individual points on the boundary have a clear interpretation.

Moreover, any method of inference applied to data from a GRD design must account for possible spatial correlation. To that end, we propose using nonparametric estimation methods that are standard practice in the analysis of classical RD designs (see, e.g., Imbens and Lemieux 2008) to GRD designs. Finally, we devote special attention to the practical applicability of GRD designs, in light of the specially demanding restrictions that are required when the RD assumptions are applied to a geographic context. We argue that the continuity assumptions needed for identification will hold less often when applied to geography, because when discontinuities are geographic, agents may sort very precisely around the boundaries and undermine the validity of the design. To that end, we provide researchers with practical advice on how to judge the plausibility of the key identifying assumptions. More generally, we argue that considerable substantive knowledge is needed to credibly exploit geographic boundaries as RD designs.

We illustrate our methodological framework and practical guidelines with an empirical application that replicates the research design in Huber and Arceneaux (2007) and Krasno and Green (2008). Following these previous studies, we use the exogenous variation in the volume of TV ads that is created by media market boundaries to understand whether campaign ads increase voter turnout. Using individual-level voter turnout data, we replicate the previous finding that political ads seem to have no effect on turnout. Applying our framework to this empirical application allows us to highlight and address important features of GRD designs, including that media market

boundaries tend to be identical to county boundaries and that the required assumptions are most likely to hold in non-battleground states.

The article is organized as follows. In the next section, we introduce the details of our motivating empirical application. In Section 3, we discuss the related literature on multidimensional RDs, formally state the identification assumptions, and discuss estimation strategies. In Section 4, we discuss the specific issues that arise in geographic applications of the two-dimensional RD design. We present our estimation framework in Section 5, and the empirical results obtained from applying our GRD framework to the media-markets example in Section 6. In Section 7, we discuss recommendations for practice and conclude. We collect additional discussions, derivations, and analyses in the Appendix.

2   A Motivating Example

Presidential campaigns spend millions of dollars on television advertisements to both persuade voters to choose their candidate and motivate them to turn out to vote on election day. Such spending on TV ads, however, is not spread evenly across the United States. Candidates spend heavily in so-called battleground states, states where the outcome of the presidential election is expected to be close, while spending little or nothing in states that clearly favor one candidate. Campaigns generally buy television advertisements by designated market areas (DMAs), also known as "media markets." DMAs are designated by Nielsen Media Research for the purposes of measuring television ratings. Each DMA is an "exclusive geographic area of counties in which the home market television stations hold a dominance of total hours viewed."[1] A single DMA may include counties in more than one state, since residents in a county can watch TV stations located in a neighboring state's metropolitan area. For example, the Chicago DMA includes not only counties in Illinois but also counties in Indiana. There are a total of 210 DMAs in the United States, typically named after the city (or cities) where the most-viewed TV stations are located.

One prominent example of within-state media-market variation, and one that we reanalyze below, is in New Jersey. Southern New Jersey has been designated as part of the Philadelphia DMA or media market. Northern New Jersey, however, is designated as part of the New York City DMA. This variation leads to very different experiences for residents of New Jersey during presidential campaigns. For example, during the 2008 presidential campaign, we calculated that residents of New Jersey who were in the Philadelphia media market were exposed to an average of 177 presidential campaign ads per day from September 1 until election day. By contrast, residents of New Jersey in the New York media market saw no presidential ads at all during the same time period (Goldstein, Franz, and Ridout 2008).

Huber and Arceneaux (2007) and Krasno and Green (2008) exploited this within-state variation in media markets to study whether presidential campaign advertisement affects turnout and political attitudes. Both studies compare voters in adjacent counties in the same state that are in different media markets, where one media market had a high volume of ads and the other media market had few or no ads. In both studies, the authors find little evidence that being exposed to presidential campaign ads during the 2000 election increases voter turnout. While these studies do not use detailed geographic data, the assumption in both cases is that voters who live near a media market boundary are as-if randomized to presidential TV ads exposure. In the following sections, we use this empirical application to illustrate our methodological framework and present general features of the GRD design that are likely to be encountered by practitioners in different subfields. But before turning to practical issues, we discuss the connections between our approach and other studies on RD with multiple scores and, also outline our methodological framework formally.

---

[1]See Nielsen Media's research glossary of terms at http://www.nielsenmedia.com/glossary.

3 Geographic Regression Discontinuity Design

3.1 Setup and Notation

In a regression discontinuity design, assignment of a binary treatment, T, is a function of a known covariate, S, usually referred to as the *forcing variable* or *score*. In the sharp RD design, treatment assignment is a deterministic function of the score, where all units with scores less than a known cutoff are assigned to the control condition ($T=0$) and all units with scores above the cutoff are assigned to the treatment condition ($T=1$). The crucial aspect of the design is that the probability of receiving treatment jumps discontinuously at the known cutoff, while all other factors related to the outcome vary smoothly.[2]

In a geographic-based RD design, we compare units in a *treated area* to units in a *control area*, which we denote by $\mathcal{A}^t$ and $\mathcal{A}^c$, respectively. We adopt the potential outcomes framework and assume that unit or individual i has two potential outcomes, Y_{i1} and Y_{i0}, which correspond to levels of treatment $T_i = 1$ and $T_i = 0$, respectively.[3] In this context, $T_i = 1$ denotes that unit i is within $\mathcal{A}^t$ and $T_i = 0$ denotes that i is within $\mathcal{A}^c$.[4] In our empirical application, $\mathcal{A}^t$ is an area in the state of New Jersey that is within the Philadelphia media market, while $\mathcal{A}^c$ is a region in New Jersey that is within the New York media market. Thus, in our example, $T_i = 1$ when individual i resides in the Philadelphia media market, and $T_i = 0$ when he resides in the New York media market. We are interested in the effect of treatment for unit i, $\tau_i = Y_{i1} - Y_{i0}$, where in our application the potential outcomes are binary and they represent the decision to turn out to vote (or not). The observed outcome is $Y_i = T_i Y_{i1} + (1 - T_i) Y_{i0}$, and the fundamental problem of causal inference is that we cannot observe both Y_{i1} and Y_{i0} simultaneously for any given unit, which implies that we cannot recover the individual effect τ_i. However, under certain assumptions, we will be able to learn about local averages of τ_i.

3.2 Identification

We now discuss the assumptions needed to identify treatment effects in the GRD design. Informally, a parameter is said to be identifiable if changing the value of the true parameter that generated the data implies a different distribution of the observed data (see Matzkin 2007, § 3.1). Identification for the GRD design holds under a central continuity assumption, also considered by Imbens and Zajonc (2011) for the general case of RD designs with multiple forcing variables. An example of an RD with two forcing variables is students taking two exams (say, language and mathematics) and a rule that allows students to graduate when their test scores on each exam exceed a particular threshold. As we formally outline below, the central identification assumption for this two-dimensional RD design is equivalent to the identification assumption needed for GRD designs. In this regard, GRD designs are a special case of an RD design with two arbitrary scores and cutoffs. In a GRD design, coordinate systems like latitude and longitude are analogous to the scores on the math and language exams.

We now state the two-dimensional continuity assumption for the GRD. We exploit the spatial proximity to the border between $\mathcal{A}^c$ and $\mathcal{A}^t$, and the fact that the treatment jumps discontinuously along this boundary. To illustrate, in our empirical example we concentrate on areas on either side of the boundary that separates the Philadelphia and New York media markets, where the volume of ads changes discontinuously from very high to zero. We define a score that uniquely represents unit

---

[2]See Imbens and Lemieux (2008) and Lee and Lemieux (2010) for comprehensive reviews of RD designs.

[3]Throughout, we assume that the potential outcomes of one unit do not depend on the treatment of other units, sometimes called SUTVA, the Stable Unit Treatment Value Assumption (Cox 1958; Rubin 1986). But we discuss the issue of interference in Appendix A and the related issue of compound treatments in Section 4.

[4]Note that we are defining our unit of observation as individuals within geographic areas, which implies that the underlying manipulation we are considering is one where individuals are assigned to treated or control areas. An alternative would be to consider a cluster assignment whereby the geographic areas themselves are assigned to treatment or control. We do not pursue this alternative here, but discuss it more fully in Appendix B.

i's geographic location, and allows us to compute i's distance to any point on the border. We use vectors, in bold, to simplify the notation. The geographic location of individual i is given by two coordinates such as latitude and longitude, $(S_{i1}, S_{i2}) = \mathbf{S}_i$. We call the set that collects the locations of all boundary points $\mathcal{B}$, and denote a single point on the boundary by $\mathbf{b}$, with $\mathbf{b} = (S_1, S_2) \in \mathcal{B}$. Thus, $\mathcal{A}^t$ and $\mathcal{A}^c$ are the sets that collect, respectively, the locations that receive treatment and control. The treatment assignment is a deterministic function of the score $\mathbf{S}_i$, and can be written as $T_i = T(\mathbf{S}_i)$, with $T(\mathbf{s}) = 1$ for $\mathbf{s} \in A^t$ and with $T(\mathbf{s}) = 0$ for $\mathbf{s} \in A^c$. This assignment has a discontinuity at the known boundary $\mathcal{B}$. In addition, we assume throughout that the density of $\mathbf{S}_i$, $f(\mathbf{s})$, is positive in a neighborhood of the boundary $\mathcal{B}$—an assumption that, as we illustrate below, can be restrictive in geographic contexts. This setup, together with the identification assumption below, formally constitute the Geographic Regression Discontinuity design.

Assumption 1 (Continuity in two-dimensional score). *The conditional regression functions are continuous in* $\mathbf{s}$ *at all points* $\mathbf{b}$ *on the boundary:*

$$\lim_{\mathbf{s} \to \mathbf{b}} E\{Y_{i0}|\mathbf{S}_i = \mathbf{s}\} = E\{Y_{i0}|\mathbf{S}_i = \mathbf{b}\}$$

$$\lim_{\mathbf{s} \to \mathbf{b}} E\{Y_{i1}|\mathbf{S}_i = \mathbf{s}\} = E\{Y_{i1}|\mathbf{S}_i = \mathbf{b}\},$$

for all $\mathbf{b} \in \mathcal{B}$.

This assumption requires that the average potential outcomes under treatment and control be continuous at all points on the boundary. In the context of our example, this means that the average potential turnout that would be observed under low (high) advertisement very near point $\mathbf{b}$ on the boundary between the New York and Philadelphia media markets is very similar to the average potential turnout that would be observed under low (high) advertisement exactly at this boundary point, regardless of the direction in which we approach the boundary. In other words, the difference in the average potential turnout with low/high advertisement between the boundary point $\mathbf{b}$ and a point very close to it can be made arbitrarily small by getting increasingly closer to $\mathbf{b}$.

Note that the probability of treatment jumps discontinuously along an infinite collection of points—the collection of all points $\mathbf{b} \in \mathcal{B}$. This implies that the parameter identified under this assumption is infinite-dimensional, as it is a curve on a plane. In other words, since the cutoff is not a point but a *boundary*, the GRD design will identify the treatment effect at *each* of the boundary points. This result is summarized in the following proposition, where superscripts t and c are used to denote locations in the treated and control areas, respectively, so that, for example, $\mathbf{s}^c \in \mathcal{A}_c$ and $\mathbf{s}^t \in \mathcal{A}_t$.

Proposition 1 (Geographic Treatment Effect Curve). If $\Pr(T_i = 1) = 1$ for all i such that $\mathbf{s}_i \in \mathcal{A}^t$ and $\Pr(T_i = 0) = 1$ for all i such that $\mathbf{s}_i \in \mathcal{A}^c$ (the discontinuity is sharp), and Assumption 1 holds, then

$$\tau(\mathbf{b}) \equiv E\{Y_{i1} - Y_{i0}|\mathbf{S}_i = \mathbf{b}\}$$

$$= \lim_{\mathbf{s}^t \to \mathbf{b}} E\{Y_i|\mathbf{S}_i = \mathbf{s}^t\} - \lim_{\mathbf{s}^c \to \mathbf{b}} E\{Y_i|\mathbf{S}_i = \mathbf{s}^c\} \quad \text{for all } \mathbf{b} \in \mathcal{B}.$$

(The proof follows Hahn, Todd, and van der Klaauw (2001); see Appendix C for details.) In other words, under the GRD assumptions, we can identify one (possibly different) treatment effect $\tau(\mathbf{b})$ for every point $\mathbf{b}$ on the boundary, defining a *treatment effect curve*.[5] In addition to the parameter $\tau(\mathbf{b}) = E\{Y_{i1} - Y_{i0}|\mathbf{S}_i = \mathbf{b}\}$, researchers may be interested in the average of effect across all boundary points, $\tau = E\{Y_{i1} - Y_{i0}|\mathbf{S}_i \in \mathcal{B}\}$. This average parameter could be obtained by

---

[5]In Appendix D, we provide alternative versions of Assumption 1 and Proposition 1 that collapse the two-dimensional score into a scalar measure of geographic distance.

integrating the $\tau(\mathbf{b})$ effects over the entire boundary. As discussed by Imbens and Zajonc (2011), we can write the average effect τ as

$$\tau = \int_{\mathbf{s}\in\mathcal{B}} \tau(\mathbf{s})f(\mathbf{s}|\mathbf{S}\in\mathcal{B})\,d\mathbf{s} = \frac{\int_{\mathbf{s}\in\mathcal{B}}\tau(\mathbf{s})f(\mathbf{s})\,d\mathbf{s}}{\int_{\mathbf{s}\in\mathcal{B}}f(\mathbf{s})\,d\mathbf{s}},$$

which can be easily recovered once the local effects $\tau(\mathbf{b})$ and the density $f(\mathbf{b})$ are estimated at multiple boundary points. This estimate can serve as a useful summary if researchers wish to report an overall effect instead of, or in addition to, geolocated local estimates.

3.3   *Related Literature on RD with Multiple Forcing Variables*

There are a number of recent studies that discuss RD designs with multiple forcing variables that are closely related to our project. Papay, Willett, and Murnane (2011) were among the first to consider the generalization of the RD design to multiple forcing variables. However, this study does not discuss formal identification results, and considers cases where multiple scores assign individuals to a range of different treatment conditions, as opposed to a single treatment as we discuss here.

Dell (2010) uses a geographic discontinuity to study the effect of a forced labor mining system in Peru and Bolivia (the *mita*) on household consumption and other measures. Her approach is related to ours in that it incorporates geographic coordinates into the analysis, but our approach differs from hers in important ways. First, Dell (2010) discusses identification informally, and focuses mostly on estimation issues. In contrast, we focus on formal identification issues and highlight specific threats that tend to arise in geographic-based designs. Second, her estimation strategy uses a two-dimensional score (latitude and longitude) that is the same for all units contained in the same cluster: the specifications include a cubic polynomial in the latitude and longitude of each observation's district capital, but the unit of observation is the individual or household, with many individual observations contained in each district. In contrast, our approach focuses on cases where the individual observations are directly georeferenced and individual geographic locations are directly incorporated into the analysis. Moreover, with clustered geographic coordinates one cannot define the individual-level treatment assignment as a deterministic function of this score, as in the traditional sharp RD, since there are no individual-level scores available, only cluster-level scores. Another distinction is on the parameter of interest; while Dell (2010) captures geographic treatment effect heterogeneity with boundary segment fixed effects, we focus on identification and estimation of treatment effects at every boundary point, which captures geographic heterogeneity in a more general way, as these effects for specific boundary points can always be integrated to obtain the effect for any larger segment.

Finally, in an independent study, Imbens and Zajonc (2011) develop identification and estimation results for the general case of the RD design with multiple forcing variables. This study is similar to ours in that it discusses identification formally, and it also proposes local regression estimation. But unlike these authors, who focus on general treatments based on multidimensional scores, we focus specifically on geographic treatments and the particular inferential obstacles and opportunities that arise in geographic applications.

We have pointed out the equivalency between the GRD design and the nongeographic, two-dimensional RD design. But when it comes to defining the treatment assigned, calculating the distance to the cutoffs, interpreting treatment effects, and evaluating the plausibility of the identification assumptions, GRD designs lead to very specific issues that do not often arise in nongeographic contexts. We now turn to these differences.

4   Particularities of Geographic Regression Discontinuities

We highlight three areas in which the GRD design differs from the two-dimensional non-geographic RD design: the possibility of compound treatments, the role of distance to the boundary, and geographic treatment heterogeneity.

4.1 *Compound Treatments*

In GRD designs, we often are confronted with compound treatments, that is, a situation in which two or more treatments that affect the outcome of interest occur simultaneously. This poses a serious challenge if the researcher is interested in only one of those treatments since, absent any restrictions or assumptions, it will not be possible to separate the effect of the treatment of interest on the outcome from the effect of all other simultaneous "irrelevant" treatments. In geographic applications, compound treatments typically arise when two or more geographically defined borders are located at the sample place. Although this phenomenon can also occur in standard RD designs (e.g., multiple policies may change when a person reaches 65 years of age), it is much more common in geographic RD designs, since the discontinuity of interest is typically the boundary of some administrative unit (a county, an electoral district, a school district, etc.), and these boundaries often perfectly overlap with each other—for example, county boundaries often coincide exactly with school district boundaries.

As illustrated by our empirical application, compound treatments may pose a serious challenge even when the boundary of interest is seemingly unrelated to administrative boundaries. Figure 1 contains a map of media markets in 2008 overlaid onto a map of U.S. counties. Close examination of the map reveals that, with only a few exceptions, media market boundaries exactly follow county boundaries. That is, moving from one media market to another implies moving from one county to another. In many places, U.S. House and other legislative districts follow county boundaries as well. As a result, county, media market and legislative district boundaries may all overlap in the same location.

Why does this matter? The outcome in our application is voter turnout. In many states, counties are important units in terms of electoral administration, as county election officials are often allowed to decide the number of polling stations, set precinct boundaries, and enforce voter iden-tification laws. It is quite possible that turnout might differ at a county boundary due to the specific features of the counties' electoral administration. If media-market and county boundaries are identical, one will not be able to isolate the two, possibly different, effects. Alternatively, if a congressional district border is also shared, a competitive House race might also confound the media-market effect on turnout.

When analysts face compound treatments, an additional assumption will be needed for identi-fication (see VanderWeele 2009; Hernán and VanderWeele 2011).[6] We can state this assumption formally as follows. We assume there are K binary treatments that occur simultaneously, which we denote as $T_{ij}, j = 1, 2, \ldots, K$, for each individual i, and $T_{ij} = \{0, 1\}$. We assume that only the kth treatment, T_{ik}, is of interest. In the most general case, potential outcomes will be affected by each of these simultaneously occurring treatments, and isolating the effect of T_{ik} will not be possible. We generalize our potential outcomes notation to illustrate this general case, and let $Y_{i\mathbf{T}_i}$ be the potential outcome of individual i with $\mathbf{T}_i = (T_{i1}, T_{i2}, \ldots, T_{ik} \ldots, T_{iK})'$ a K-dimensional vector. This general notation allows all K versions of treatment to affect the potential outcomes of individual i.

When the boundary of interest is simultaneously the boundary of multiple institutional, admin-istrative. or political units and we wish to make inferences about the effect of only one of these treatments, it may be appropriate to assume that the treatment of interest is the only treatment that affects potential outcomes:

Assumption 2 (Compound Treatment Irrelevance). *Assume the treatment of interest is the kth treat-ment. For each i and for all possible pairs of treatment vectors $\mathbf{T}_i$ and $\mathbf{T}'_i$, $Y_{i\mathbf{T}_i} = Y_{i\mathbf{T}'_i}$ if $T_{ik} = T'_{ik}$.*

When the Compound Treatment Irrelevance assumption holds, the potential outcomes are only a function of the treatment of interest, so $Y_{i\mathbf{T}_i} = Y_{iT_{ik}}$ and we can denote potential outcomes simply as Y_{i1} and Y_{i0} corresponding, respectively, to $T_{ik} = 1$ and $T_{ik} = 0$. In our application, there are at least two simultaneous treatments. The first, which we denote T_{i1}, is the county treatment, which

---

[6]These authors use the term *compound treatment* to describe a situation in which there are multiple versions of the same treatment. We use the term slightly differently, to refer to situations in which there are entirely different treatments and all of them simultaneously affect the potential outcomes.
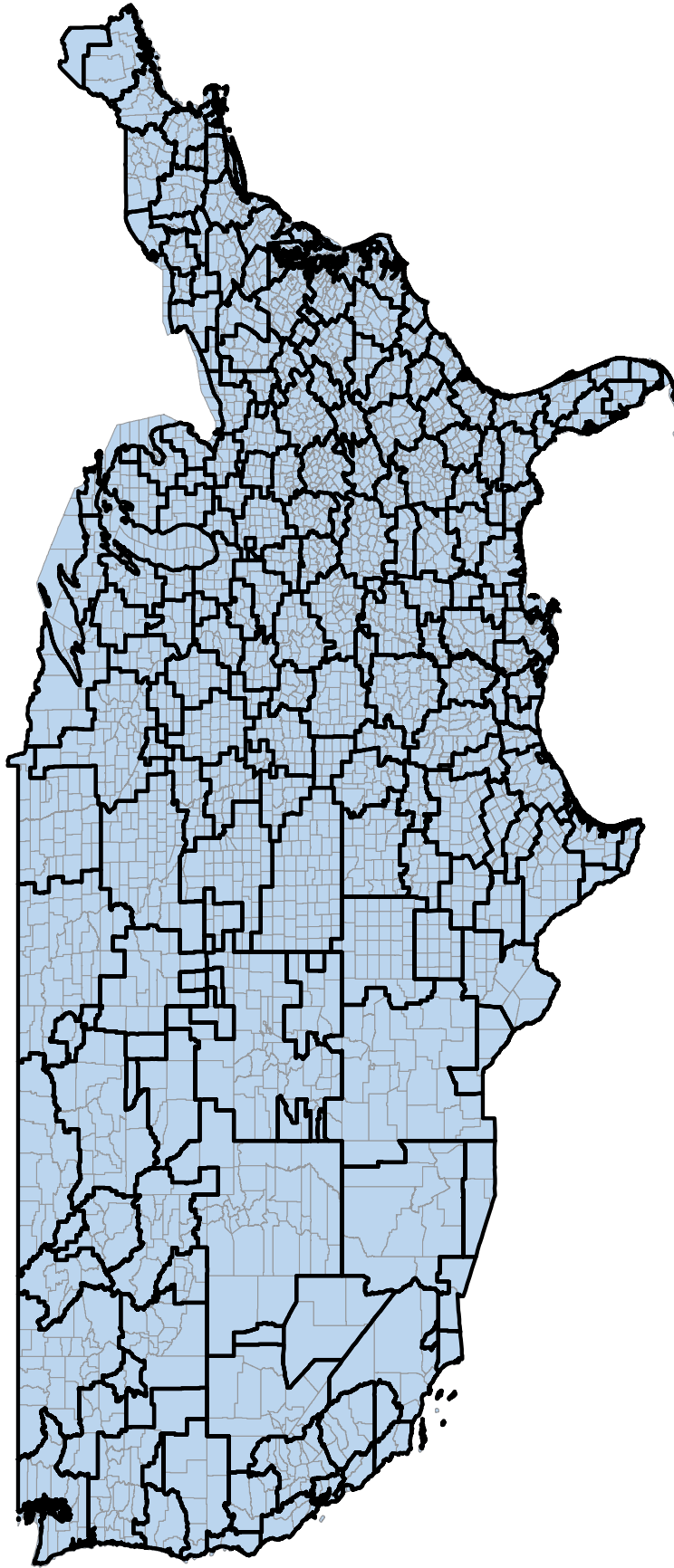
Fig. 1 Map of U.S. counties overlaid with designated media market boundaries creating compound treatments.
Note: Gray lines indicate county borders, and black lines indicate media market boundaries. Media market boundaries follow county borders except in rare instances.

is all the ways in which county-level electoral administration may affect voter turnout. But there is a second treatment, which we denote T_{i2}, that is the media market, and captures all the ways in which being exposed to high presidential TV advertising may affect voter turnout.

When the two boundaries overlap exactly, one alternative to make inferences about the effect of T_{i2} alone is to assume that $Y_{i(T_{i1}, T_{i2})'} = Y_{iT_{i2}}$. In our example, this implies assuming that there is no separate county effect on turnout, so that the county treatment can be exactly reduced to the media-market treatment.[7] Another alternative is to define the estimand as a compound treatment effect that includes both a media-market effect and a county effect, but for the purposes of our example this is unsatisfactory, because our substantive interest is on media-market effects isolated from county effects.

In some applications where units can be observed before and after the treatment of interest occurs and all the "irrelevant" treatments occur in both periods, the differences between treated and control areas in the first period could be used to infer the effect of the irrelevant treatments on the outcome. Then, under appropriate assumptions, the irrelevant effect observed in the first period could be subtracted from the overall effect observed in the second period—when the treatment of interest occurs simultaneously with all the irrelevant treatments—to isolate the effect of the treatment of interest on the outcome. Developing the type of assumptions under which this "differences-in-differences" strategy is valid is beyond the scope of this article, but we note that in general this strategy will require conditions that specify the ways in which the treatment of interest and the irrelevant treatments affect the potential outcomes over time.

The ideal scenario is one where the Compound Treatment Irrelevance assumption can be avoided altogether. In the media-market application, we found one place where this holds: a small area of Northern California along Lake Tahoe, where voters are part of the Reno, Nevada, DMA. Citizens in this region of California get the full volume of presidential ads from Nevada (a battleground state), while other California residents do not. In this case, the Reno media market does not follow a county boundary and instead splits El Dorado County in California into two media markets: the Reno DMA to the east, and the Sacramento DMA to the west. Figure 2 displays this area of California.

Since the boundary between the Reno and Sacramento media markets splits El Dorado County, the GRD estimate will isolate the media-market effect from any county effect. The dots in Fig. 2 represent the locations of households with at least one registered voter. Unfortunately, as the figure shows, the density of the data close to the media-market border is very low: there are almost no observations close to the boundary, which violates one of the needed assumptions for implementation of a GRD design. In addition, rudimentary comparisons between the observations on either side of the boundary show that those who live along Lake Tahoe are wealthier than those who live outside the Reno DMA in El Dorado County. The sales price of homes in the Reno media market were on average nearly $200,000 higher than for homes in El Dorado County outside the Reno DMA. Incomes were also substantially higher for residents in the Reno media market. Thus, we observe a correlation between income and the higher volume of presidential campaign ads, which will confound the effect of ads on voter turnout. This is not surprising, given that the treated and control groups on either side of the border are very far from each other.

Thus, despite the fact that we can isolate the precise media-market effect, we do not pursue the analysis in this area and focus on New Jersey instead. Nonetheless, this area in California is useful to illustrate a case where the compound treatment irrelevance assumption is not needed. Unfortunately, avoiding this extra assumption comes at too high a price in this example, since there is simply no data density close to the border.

The presence of compound treatments is very common in GRD designs. When compound treatments occur, analysts will need to either find areas where a single treatment can be isolated

---

[7]Of course, the assumption $Y_{i,(T_i(1), T_i(2))} = Y_{i,T_i(1)}$ also satisfies Assumption 2, since it allows us to reduce the media-market treatment to the county treatment. However, since our substantive interest is on media-market effects, this version of Assumption 2 is not helpful.
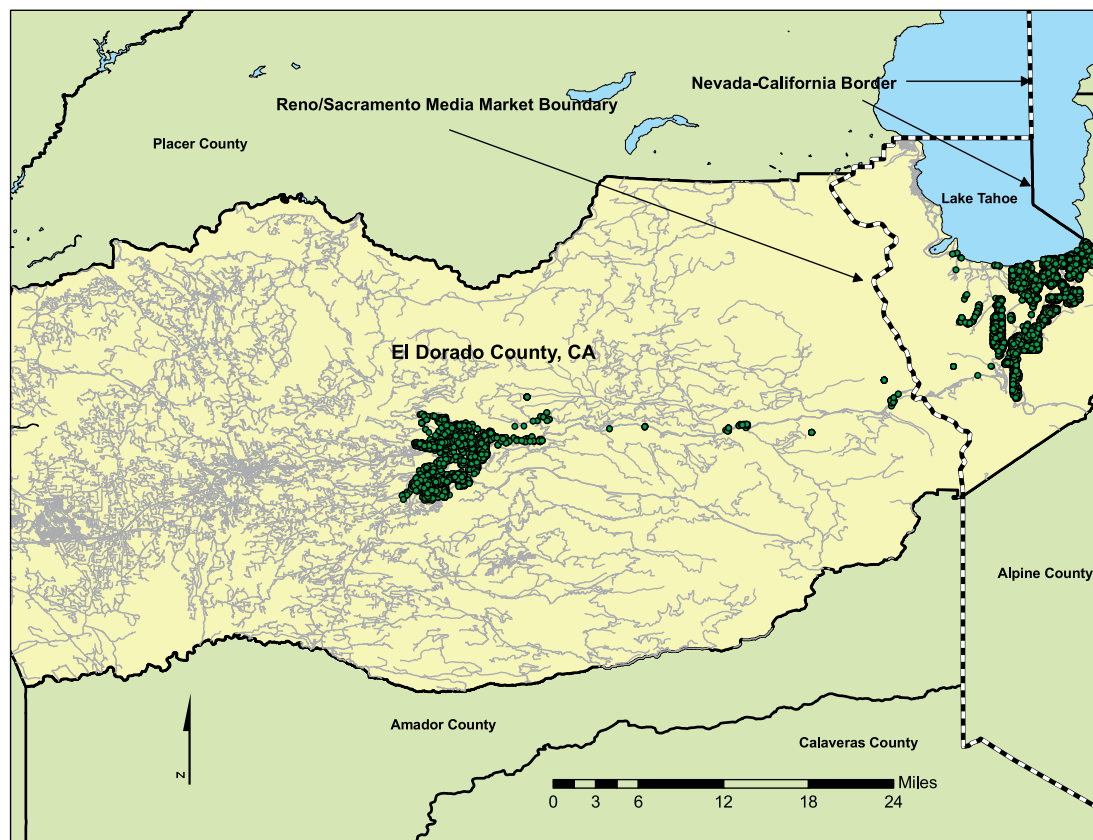
Fig. 2 Boundary between Reno and Sacramento media markets. The boundary between the Reno, NV, media market (located east of the boundary) and the Sacramento, CA, media market (located west of the boundary) splits California's El Dorado County. As a result, the media-market treatment is isolated from the "county treatment." The 2008 volume of ads was high in the Reno media market and low in the Sacramento media market. Dots represent the location of households with registered voters.

or make a case that the assumption of compound treatment irrelevance holds. As we discuss in detail below, we are unable to avoid this assumption in the area of New Jersey that we study.[8]

4.2 *Naive Distance*

Often in the GRD design, the score S is defined as the shortest (i.e., perpendicular) distance to the boundary, and units that are close to the boundary in terms of this distance but on opposite sides of it are taken as valid counterfactuals for each other. In a GRD design of this type, the analyst compares all units that are within a fixed distance from the border; see Black (1999) for an example. Here, individual i has distance $S_i = d$ if the distance from i's location to the point on the boundary that is closest to i is equal to d. While this concept of distance is well defined in the context of geography, it makes little sense in the context of the two-dimensional RD based on test scores.

---

[8]The Compound Treatment Irrelevance assumption is closely connected to the exclusion restriction in instrumental variables contexts, where researchers must assume that the instrument has an effect on the outcome only through the treatment of interest but does not affect the outcome directly. Any GRD design is ultimately focused on the effect of some treatment that occurs in the geographic unit, not on the effect of the geographic unit itself. In this sense, the geographic units in the analysis (media markets, in our example) may be seen as analogous to an instrument, and Assumption 2 may be seen as analogous to an exclusion restriction that requires that the only feature of the geographic units that affects the outcome is the presence or absence of the treatment of interest (political TV ads).
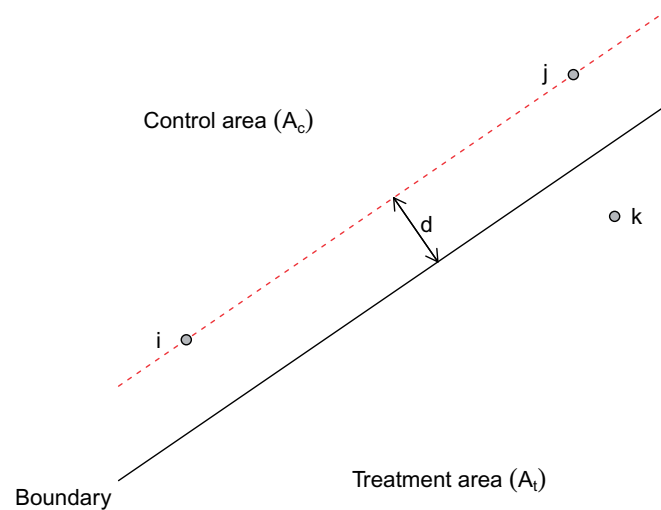
Fig. 3 Failure of one-dimensional distance to single out individual boundary points.

In the two-dimensional non-geographic RD, we cannot sensibly define a shortest distance to the "boundary" between treated and control, primarily because this boundary does not in any sense correspond to a pre-existing feature of the world such as a county, district, DMA, etc. Thus, the possibility of computing this shortest distance is another key distinction between the two designs.

Moreover, using the perpendicular distance to the boundary as the score may mask important heterogeneity and may not allow researchers to fully evaluate the plausibility of the identification assumptions in the GRD design. The problem is that this measure of distance ignores the spatial nature of geographic locations. We refer to this distance as "naive," to distinguish it from the two-dimensional distance we introduce below, which we term "geographic." As illustrated in Fig. 3, the shortest distance from individual i's location to the boundary does not determine the exact location of i in the map, since two individuals i and j in different locations can both have $S_i = S_j = d$. That is, this naive distance does not account for distance *along* the border. As one can see in Fig. 3, a naive implementation of the RD design along a geographic boundary that does not take into account both dimensions would treat individuals i and j in the control area as equally distant from individual k in the treatment area, when in fact j is much closer to k than i. This problem will be exacerbated when the boundary is longer; in Fig. 3, as the boundary becomes longer, the distance between control unit i and treated unit k can be made arbitrarily large even as $S_i = d$ remains constant, by moving i along the dotted line.

Naturally, a naive strategy cannot recover $\tau(\mathbf{b})$ for any $\mathbf{b} \in \mathcal{B}$, except in special cases such as when treatment effects are constant at all boundary points, i.e., $\tau(\mathbf{b}^p) = \tau(\mathbf{b}^q)$ for all $\mathbf{b}^p \in \mathcal{B}$, $\mathbf{b}^q \in \mathcal{B}$.[9] In contrast, no information is lost when geography is fully exploited in a GRD approach: once the local effects $\tau(\mathbf{b})$ are estimated for all $\mathbf{b} \in \mathcal{B}$, researchers can compute any average effect of interest. Moreover, a GRD approach allows researchers to detect possible discontinuities in predetermined covariates on different points on the boundary, while these local discontinuities might be masked in a naive approach.

4.3 *Spatial Treatment Effects*

We note one final important difference between GRD designs and the two-dimensional nongeographic RD design. As we noted in Section 3, the effect identified is not a point estimate but a line

---

[9]One avenue for future research is to explore the conditions under which a naive design leads to a consistent estimator of the overall average effect τ, which will depend on how this design is implemented—for example, simple means on either side of the boundary within a narrow band versus local-polynomial on one-dimensional distance normalizing the cutoffs.

of treatment effects along the border that separates the treated and control areas. In the GRD design, this leads to estimated effects that are spatially located, and these treatment effects can be in principle heterogeneous. In the standard two-dimensional RD design, this heterogeneity may be difficult to interpret, but in the GRD design, we can map this heterogeneity to specific geographic locations to observe whether the treatment effect varies along the geographic border of interest. In other words, a GRD can uncover interesting patterns of geographic treatment effect heterogeneity that may have, for example, important policy implications. Analysts should either specify whether they can articulate a pattern in the treatment effects or treat such heterogeneity as an exploratory analysis. In the next section, we develop an estimator that is faithful to the spatial nature of the GRD design.

5 Estimation in the GRD Design

We now provide an estimation framework that is well suited to the features of the GRD design, and can be used to both estimate treatment effects and assess the plausibility of the continuity assumptions. We generalize the local polynomial regression estimator commonly used for estimation in one-dimensional RD designs (see Hahn, Todd, and van der Klaauw [2001]; Porter [2003]; and Imbens and Lemieux [2008] for an overview). Our goal is to estimate a conditional expectation of the outcomes as a function of the distance to the boundary. This estimate, however, needs to be faithful to local spatial variation around the discontinuity of interest. In Appendix E, we discuss how our method relates to methods in statistical geography, such as geographically weighted regression and the analysis of spatial autocorrelation.

First, we define $\mu(x) = E(Y|X = x)$ as the regression function of the observed outcome of interest Y on some univariate X. Assuming that the first $p + 1$ derivatives of $\mu(X)$ at the point $X = x_0$ exist, we can approximate $\mu(x)$ in a neighborhood of x_0 by a Taylor expansion:

$$\mu(x) \approx \mu(x_0) + \mu^1(x_0)(x - x_0) + \frac{\mu^2(x_0)}{2}(x - x_0)^2 + \ldots + \frac{\mu^p(x_0)}{p!}(x - x_0)^p,$$

where $\mu(x_0), \mu^1(x_0), \mu^2(x_0), \ldots, \mu(x_0)^p$ denote the first $(p + 1)^{th}$ derivatives of $\mu(x_0)$.

In local regression estimation, this polynomial is fitted locally, minimizing a weighted sum of squared residuals. The estimated coefficients $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_p)'$ are defined as

$$\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{N} \left\{ Y_i - \sum_{j=0}^{p} \beta_j (X_i - x_0)^j \right\}^2 w_i,$$

with weights $w_i = \frac{1}{h} K(\frac{X_i - x_0}{h})$ for a given kernel function $K(\cdot)$ and bandwidth h. This yields $\hat{\mu}_j(x_0) = j! \hat{\beta}_j$ as an estimator of $\mu(x_0)^j$ for $j = 0, 1, \ldots, p$. In particular, an estimator for the conditional expectation of Y given $X = x_0$ is given by $\hat{\mu}(x_0) = \hat{\beta}_0$. See Fan and Gijbels (1996) for an extensive discussion of local polynomial estimation.

Using this local polynomial estimator, we borrow the basic estimation approach from RD designs, which involves estimating the left and right limits of $\mu(c)$, denoted $\mu^l(c)$ and $\mu^r(c)$, respectively, with a local polynomial of degree one, where c is a known cutoff in the score. The estimation of $\mu^l(c)$ uses only observations to the left of c and, similarly, estimation of $\mu^r(c)$ uses only observations to the right of the cutoff. For given weights w_i and a scalar score S_i, this involves computing the weighted regression of the observed outcome Y_i on a constant and $S_i - c$; the estimated effect is then $\hat{\tau} = \widehat{\mu^r(c)} - \widehat{\mu^l(c)}$.

We modify this standard estimation approach in several ways. For a given point $\mathbf{b}$ on the boundary, we calculate a measure such as the Euclidean distance, which can accommodate multiple dimensions, between the location $\mathbf{S}_i$ of unit i and $\mathbf{b}$. For every unit i in the sample, this distance is defined as $f_\mathbf{b}(\mathbf{S}_i)$. Letting

$$\mu(\mathbf{b})^c \equiv \lim_{\mathbf{s}^c \to \mathbf{b}} E\{Y_{i0}|f_\mathbf{b}(\mathbf{S}_i) = f_\mathbf{b}(\mathbf{s}^c)\}$$

$$\mu(\mathbf{b})^t \equiv \lim_{\mathbf{s}^t \to \mathbf{b}} E\{Y_{i1}|f_\mathbf{b}(\mathbf{S}_i) = f_\mathbf{b}(\mathbf{s}^t)\},$$

we estimate these functions by local linear regression. To do so, we solve

$$(\hat{\alpha}_{\mathbf{b}}^c, \hat{\beta}_{\mathbf{b}}^c) = \arg\min_{\alpha_{\mathbf{b}}^c, \beta_{\mathbf{b}}^c} \sum_{i \in A^c} \left\{ Y_i - \alpha_{\mathbf{b}}^c - \beta_{\mathbf{b}}^c (f_{\mathbf{b}}(\mathbf{S}_i) - f_{\mathbf{b}}(\mathbf{b})) \right\}^2 w_{i\mathbf{b}}$$

$$(\hat{\alpha}_{\mathbf{b}}^t, \hat{\beta}_{\mathbf{b}}^t) = \arg\min_{\alpha_{\mathbf{b}}^t, \beta_{\mathbf{b}}^t} \sum_{i \in A^t} \left\{ Y_i - \alpha_{\mathbf{b}}^t - \beta_{\mathbf{b}}^t (f_{\mathbf{b}}(\mathbf{S}_i) - f_{\mathbf{b}}(\mathbf{b})) \right\}^2 w_{i\mathbf{b}},$$

where

$$w_{i\mathbf{b}} = \frac{1}{h_{\mathbf{b}}} K\left(\frac{f_{\mathbf{b}}(\mathbf{S}_i) - f_{\mathbf{b}}(\mathbf{b})}{h_{\mathbf{b}}} \right)$$

are a set of spatial weights where $K(\cdot)$ represents a kernel weighting function and $h_{\mathbf{b}}$ is the bandwidth at the boundary point $\mathbf{b}$. Since $f_{\mathbf{b}}(\mathbf{b}) = 0$, all formulas above simplify immediately. Given these solutions, the GRD effect is estimated as

$$\hat{\tau}(\mathbf{b}) = \widehat{\mu^t(\mathbf{b})} - \widehat{\mu^c(\mathbf{b})} = \hat{\alpha}_{\mathbf{b}}^t - \hat{\alpha}_{\mathbf{b}}^c. \tag{1}$$

Conventional confidence intervals are typically based on the standard asymptotic distribution of the least squares estimator and robust standard errors (see, e.g., Imbens and Lemieux 2008). These confidence intervals, however, ignore the asymptotic bias of the nonparametric local polynomial estimator, a simplification that is justified only if the bandwidth in the estimator is chosen to be small enough (i.e., if one does *undersmoothing*). The common bandwidth selection methods include cross-validation and mean-squared error (MSE) minimization (see, e.g., Imbens and Kalyanaraman 2012), which in general lead to bandwidth choices that are too large for conventional confidence intervals to be valid (Calonico, Cattaneo, and Titiunik 2014b). In order to obtain valid inferences, researchers may select a smaller bandwidth to undersmooth, a procedure that is simple but often ad hoc, as there are no general rules about how much undersmoothing should be done. An automatic, data-driven alternative is to estimate the asymptotic bias ignored by conventional inference, and correct the standard errors appropriately to produce robust confidence intervals that are valid even for large bandwidths, including those selected by MSE minimization (Calonico, Cattaneo, and Titiunik 2014b). In our application below, we sometimes found that the MSE optimal bandwidths tended to be large, particularly in sparse areas around a boundary point. We thus make inferences in two ways: (i) we select a fixed bandwidth, smaller than the MSE bandwidth, which justifies using conventional confidence intervals; and (ii) we employ robust confidence intervals with MSE-optimal bandwidth. For implementation, we use the rdrobust software.[10]

In practice, since the boundary $\mathcal{B}$ is an infinite collection of points, we selected a grid of G points along the boundary for estimation, $\mathbf{b}^1, \mathbf{b}^2, \ldots, \mathbf{b}^G$. Under this grid of points, we defined a series of treatment effects $\tau(\mathbf{b}^g)$ for $g = 1, 2, \ldots, G$. Our estimation procedure thus produces a collection of G treatment effects that can vary along the boundary that separates the treatment and control areas, and in fact leads to a treatment effect curve, where each effect can then be mapped in its specific location, $\mathbf{b}^g$. In addition, if researchers are interested in the average effect along the entire boundary defined above, τ, this effect can be estimated as $\hat{\tau} = \sum_{g=1}^G \hat{\tau}(\mathbf{b}^g)\hat{f}(\mathbf{b}^g) / \sum_{g=1}^G \hat{f}(\mathbf{b}^g)$, and its standard error can be obtained using either resampling methods or the delta method—see Imbens and Zajonc (2011) for details.

Since our method produces G different estimated treated-control differences, researchers may face a multiple testing problem when G is large. Even if the true effect is zero, we will expect to reject the null hypothesis $\alpha \times G$ times if we use an α-level test. This multiple testing problem poses no

---

[10]Software available at https://sites.google.com/a/umich.edu/rdrobust. See Calonico, Cattaneo, and Titiunik (2014a) for details on the STATA implementation, and Calonico, Cattaneo, and Titiunik (2014c) for details on the R implementation.

difficulties since there are well-known solutions such as the Bonferroni correction or false discovery rates (Benjamini and Hochberg 1995; Anderson 2008).

5.1 *Falsification Tests*

Before presenting the results from our application, we briefly discuss how to perform falsification tests in the GRD design. It is now established practice in the analysis of RD designs to provide empirical evidence that speaks to the credibility of the continuity assumption needed for identification. Different types of evidence can be presented, but the most common is a series of tests on *predetermined* or *pretreatment* covariates—covariates that are determined before the treatment occurs and for which the treatment effect is zero by construction. We review two different but related falsification tests for GRD designs.

First, predetermined covariates can be treated as outcomes using the local linear estimator outlined above. Since the effect of the treatment is zero by construction, the estimated effects should be statistically indistinguishable from zero, ideally with short confidence intervals. In our example, we would hope to find that the estimated effect of presidential campaigning on housing prices and other covariates at each boundary point is indistinguishable from zero. This same procedure can also be used to plot these placebo treatment effects on a map for an assessment of geographical treatment heterogeneity. Ideally, the researcher will choose to study treatment effects only at those points where the placebo analysis indicates that pretreatment covariates are indistinguishable across treatment and control areas.

A second common type of falsification test based on pretreatment covariates is a series of "balance tests," tests that investigate whether the mean (or other feature of the distribution) of the covariates is statistically indistinguishable between treated and control units near the cutoff. Strictly speaking, covariate balance in a small neighborhood around the geographic boundary is expected under a local randomization assumption, but not necessarily under a continuity assumption. See Cattaneo, Frandsen, and Titiunik (2014) for a discussion of the differences between local randomization and continuity in the general RD design, and Keele and Titiunik (2014b) for a discussion in geographic contexts.

Assessing covariate balance in a GRD design, however, requires a different strategy since covariate balance is a spatial construct. Under a naive approach, balance testing is straightforward since we can simply compare averages based on equal distances around the discontinuity. That is, we could look at the average difference in housing prices for all houses within 100 m from the media-market border. We could also increase the distance around the border to understand if balance changes as a function of distance, repeating the balance test for all houses within 200 m, 300 m, etc. All things being equal, the credibility of the design would be enhanced if balance improves as we get closer to the boundary.

The key difficulty with this geographically naive form of balance testing is that balance might change as we move along the border. While balance may hold along one section of the border, it may not hold along another section. Such geographic heterogeneity will be missed in a geographically naive balance test. To that end, we developed a method for testing spatial balance that compares nearest geographic neighbors. As we show later, balance tests based on spatial proximity can differ substantially from balance tests based on naive distance.

We developed the following algorithm to assess spatial covariate balance:

- For treated unit i, calculate the geographic distance between it and all control units.
- Match unit i to the nearest control unit (or set of control units) in terms of this geographic distance.
- Break ties randomly, so that each treated unit i is matched to a single control unit.
- Repeat for all treated units.
- Apply standard balance tests such as KS tests or t-tests to the spatially matched data.

A few issues about this form of balance test are worth noting. First, while this balance test uses a matching algorithm, matching is applied only to distance, so the end result is a set of spatially

proximate pairs. Second, one can also combine this spatial balance test with pruning of observations based on overall distance to the border. For example, we might apply the spatial balance test only to voters that are within 200 m of the border. Third, this balance test, like all balance tests, is not a strict hypothesis test based on statistical significance. As with any falsification test, we prefer not just statistically insignificant estimates, but small differences that translate into both small point estimates and large *p*-values.

All in all, whether falsification tests based on covariates lend credibility to the design will depend on whether the covariates selected are closely connected to the outcome and the treatment of interest, an issue that will depend on substantive knowledge behind each particular application. Moreover, falsification tests based on balance should be interpreted carefully, as there could simultaneously be evidence of continuity at the border and imbalance near the border for these covariates. In these situations, researchers should pay attention to the implementation of the estimation strategy, as the credibility of the results will hinge on the credibility of the extrapolation to the boundary point.

6 Estimating the Effect of Campaign Ads on Voter Turnout

We now apply our GRD framework to the research design developed by Huber and Arceneaux (2007) and Krasno and Green (2008) and examine the effect of political television advertisements on turnout during the 2008 presidential election, using the variation in the volume of ads created by different media markets. We first describe the data we used for the analysis, and then the geographic analytic tools that are essential to study this question with a GRD design. We then explore the issue of compound treatments in this application, and demonstrate how to make the assumption more plausible. Next, we explore whether pretreatment covariates (variables that are determined before the volume of ads) differ significantly near the media-market border where the discontinuity in advertisement volume occurs. This is equivalent to the requirement in the standard RD design that pretreatment covariates be similar or "balanced" close to the cutoff, just as one would expect in a randomized experiment (see Lee 2008). The final part of our empirical analysis focuses on estimating and mapping the turnout effects of presidential campaign advertisement.

6.1 Data

Our analysis is greatly aided by the form of our data, in particular the fact that it is individual-level data with enough information to geo-locate all units. We rely on two data sources.[11] Our main source is the New Jersey voter file. We purchased a version of this file from the Catalist corporation. This dataset has measures of party registration, gender, and age directly from the voter file, and imputed values of education, income, poverty status, and employment status.[12] Most importantly, the voter file also contains the address of each voter, which allows us to find each voter's geographic location and avoid the use of naive distances. Our second data source is property sales records. Data from housing sales have a number of advantages. First, according to hedonic pricing theory, housing prices should reflect a wide variety of neighborhood characteristics, including school quality (Sheppard 1999; Malpezzi 2002). Second, these data are not aggregated, which allows us to precisely estimate how they vary around the boundary of interest. We acquired records for all houses sold in the appropriate zip codes in New Jersey from January 2006 to November 2008. In this time period, nearly 3000 homes were sold in this area—although we used only the 1800 house sales inside one specific school district; see below. The housing sales data allow us to conduct a fine-grained analysis of the sales price differential along the boundary of interest.

We did not use Census data. Census units such as census block groups typically contain between 600 and 3000 people. Given the large size of block groups, it is often difficult to tell whether there

---

[11]Replication files are available at Political Analysis Dataverse. Data citation: Keele, Luke; Titiunik, Rocio, 2014. Replication data for: Geographic Boundaries as Regression Discontinuities, http://dx.doi.org/10.7910/DVN/26453 IQSS Dataverse Network [Distributor] V1 [Version].

[12]See Ansolabehere and Hersh (2012) for details about the features of voter files compiled by Catalist.

is meaningful spatial variation in block-group-level measures as one approaches the boundary of interest. In addition, past turnout behavior might seem to be an important pretreatment covariate to either condition on or use as a placebo outcome in our application. However, past turnout is affected by similar ad differentials in past elections. As such, past turnout cannot be considered a pretreatment covariate.

6.2 *Geographic Analysis*

We use Geographic Information Systems (GIS) software to process the data before the final statistical analysis. Indeed, we believe that without GIS analysis the GRD design is significantly weakened. GIS software allows analysts to more fully exploit geography and spatial proximity. Here, we outline the geographic analysis we performed to implement the GRD design in New Jersey.[13]

We can use certain GIS techniques to avoid the biases caused by aggregate data and the modifiable areal unit problem (MAUP) (Openshaw 1984). MAUP refers to the fact that areal units such as census tracts or blocks are often relatively arbitrary with respect to the spatial variation of the units measured. Thus, aggregate measures do not accurately reflect individual-level phenomena unless those phenomena are spatially constant with respect to the areal unit. It is well understood that MAUP introduces bias into estimates based on aggregate units (Openshaw 1984; Fotheringham and Wong 1991; Reardon and O'Sullivan 2004; Wong 2008; Cohn and Jackman 2011). One solution to MAUP, made possible by recent advances in GIS, is to eschew aggregate data and give each individual a unique neighborhood that is centered on the individual and includes the area around that specific point. Such a measure varies with the location of the individual, which allows the measure to vary spatially and accurately reflect the composition of the area around the individual. In our case, since the voter file contains information about the geographic location of individual voters, we can create and analyze spatial measures that are not subject to the bias caused by the MAUP.

Next, we detail how we used GIS techniques. First, we geocoded both the voter file and data on housing sales. Geocoding is the process of converting addresses into a coordinate system, typically latitude and longitude.[14] Geocoding allows us to know the distance between voters and the media market boundary that forms the discontinuity of interest, and to develop a score that reflects the two-dimensional geographic space.[15] We also used GIS software for two other tasks. First, we created what is called a buffer around the media market boundary. The buffer is a spatial object that records which voters fall within a specified distance of a geographic boundary. We used a buffer to identify which voters are within 100, 200, 300, 400, 500, 600, 700, 800, 900, and 1000 m from the border on either side. Second, we used GIS to obtain a grid of points on the media market boundary for the calculation of treatment effects. We did this by dividing the boundary into points defined by latitude and longitude, spaced at selected intervals.

---

[13]We performed all the geographic analysis in ArcgGIS 9.3.

[14]Geocoding involves taking formatted addresses and comparing them to a known database of addresses and street locations with an assigned geographic reference such as latitude and longitude.

[15]One might imagine that a simple application of the Euclidean distance with the points defined by latitude and longitude would be sufficient for calculation of the score in the GRD design. This would be appropriate if voters resided on a plane, but the Earth is a sphere. Naive Euclidean distances calculated between geographic locations can severely overestimate the distance (Banerjee 2005). There are two standard alternatives to the naive Euclidean distance: the geodetic and chordal distance. We used the chordal distance, which is a rescaling of the Euclidean distance and is very close to the geodetic distance for locations that are less than 2000 km apart. The additional advantage of the chordal distance is that it allows for valid calculations of spatial correlations, which the geodetic does not allow for (Banerjee 2005).

6.3   *Compound Treatment Reduction*

Before calculating each unit's distance to each of the boundary points, we discuss how to minimize the compound treatment assumption in this application. Figure 4 contains a map of the state of New Jersey along with the location of the boundary between the New York and Philadelphia media markets. We could calculate distances between voters and points along the entire media market boundary and compare voters who are near each other along this border. However, it is first important to reduce the incidence of compound treatments as much as possible. To do that, we examined the boundaries of four different administrative units: U.S. congressional districts, state senate districts, state house districts, and school districts. We found that for many parts of the media market boundary, the boundaries of at least one of these units overlapped perfectly with the media market boundary. In other words, in various boundary segments, not only did the media market change at the boundary, but so did the school and/or the legislative districts. This is not entirely surprising since, as we discussed above, the media market boundary in this area (and in most of the United States) follows county boundaries.

The overlap between media and county boundaries means that we cannot escape the problem of compound treatments entirely, but we can minimize it by restricting our analysis to those segments
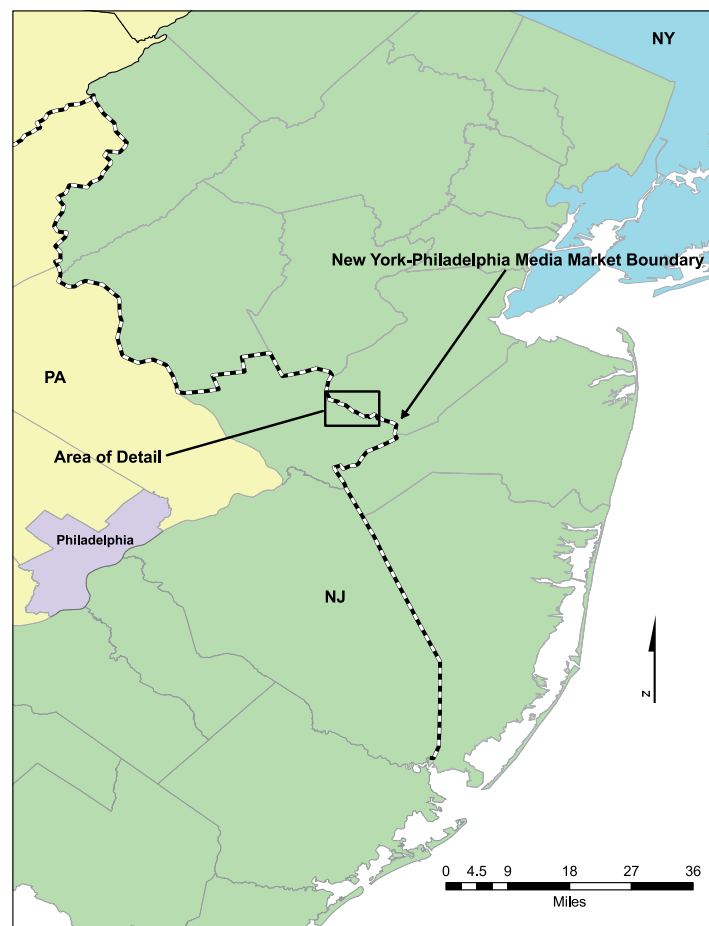
Fig. 4   Boundary between Philadelphia and New York City media markets. The dashed line represents the boundary between the Philadelphia, PA, media market (located southwest of the boundary) and the New York City, NY, media market (located northeast of the boundary), which divides the state of New Jersey. Area of detail is where legislative districts and school district are constant on both sides of media market boundary. Figure 5 contains a detailed map of this area.

Fig. 5 Detail of the boundary between Philadelphia and New York City media markets. Area marked with gray hash lines indicates the West Windsor-Plainsboro school district, which straddles the media market boundary. Empirical analysis is confined to the West Windsor-Plainsboro school di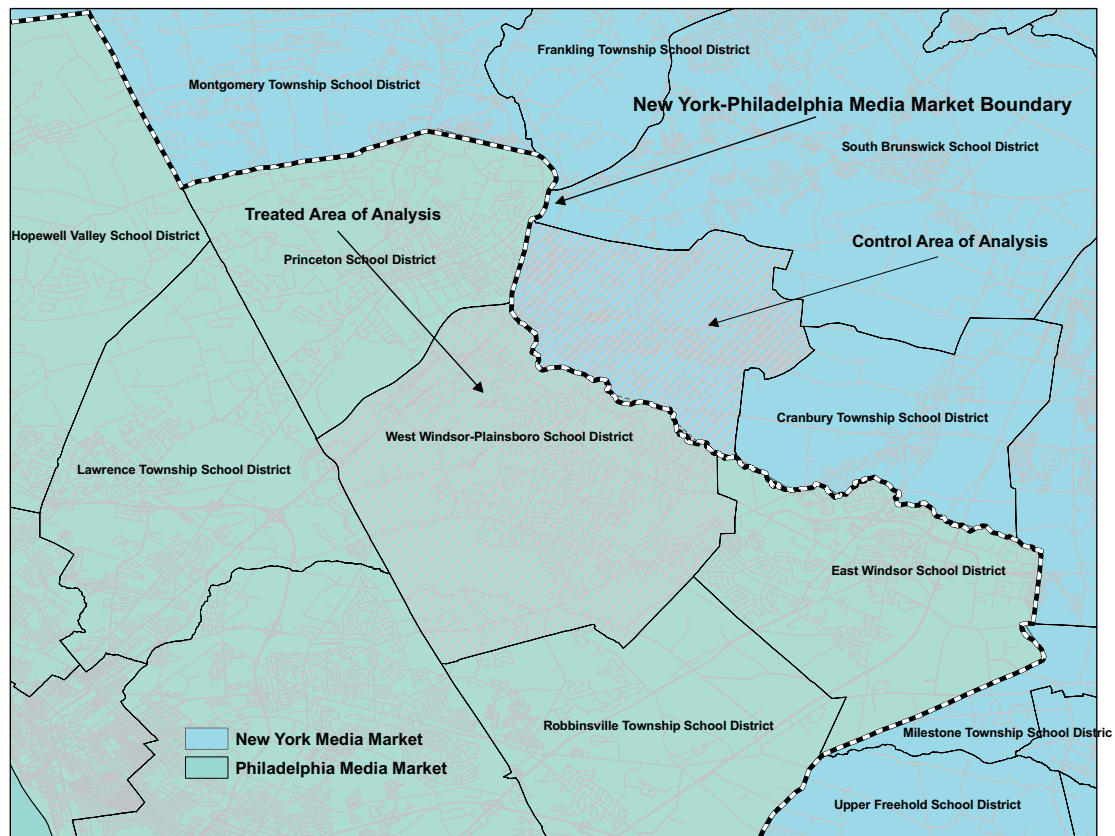strict only, where legislative districts are also the same on both sides of the border. Treated area is southwest of media market boundary, inside Philadelphia media market, where volume of political ads is high; control area is northeast of media market boundary, inside New York City media market, where volume of ads is zero.

along the media market boundary where voters are in identical legislative districts and school districts. Despite the length of the media market boundary, we found only one short segment along the border where both legislative district or school district boundaries did not also follow the county–media market boundary. The area of detail in Fig. 4 marks this boundary segment. Figure 5 contains a detailed map of this area. The area marked with gray hash lines is West Windor-Plainsboro school district, which is split in two by the county–media market boundary. This school district also lies within a single U.S. House, state house, and state senate district. We restrict our analysis to residents in this school district since we can more plausibly assume that areas on either side of this segment of the media market border are comparable—though we test this assumption empirically below. Thus, while we cannot avoid compound treatments in the application, by finding an area where school and legislative districts are constant, we hope to increase the plausibility of the compound treatment irrelevance assumption.

It is also worth noting that by holding units such as legislative and school districts constant in order to reduce compound treatments, we are making our analysis conditional on distance. That is, by removing compound treatments, we are already restricting our analysis to areas within a specific distance of the border. We suspect that, in many applications of the GRD design, addressing the compound treatment problem will prompt researchers to indirectly condition on distance to the boundary and will be a useful first step to identify areas that are similar along the discontinuity of interest.

Table 1 Covariate balance across New York and Philadelphia media markets as a function of distance

		Raw Comparison	*1000m Buffer[a]*	*500m Buffer*	*200m Buffer*	*100m Buffer*
18 covariates						
Naive	χ^2 Statistic (voters)[b]	18,921	3175	1094	217	49
	Tr Sample size	15,000	1880	731	170	36
	Co Sample size	9460	1339	367	49	15
Geographic	χ^2 Statistic (voters)[b]	–	2660	732	96	29
	Tr Sample size	–	1339	367	49	15
	Co Sample size	–	1339	367	49	15
Housing Price						
Naive	Price Sq. Ft. Diff (houses)[c]	$22	$2.95	$2.93	–	–
	Tr Sample size	1024	86	27	–	–
	Co Sample size	7743	77	14	–	–
Geographic	Price Sq. Ft. Diff (houses)[c]	–	$3.54	$2.82	–	–
	Tr Sample size	–	86	27	–	–
	Co Sample size	–	86	27	–	–

Note. [a]A buffer is a specified distance around the media market boundary. For example, with a 500 m buffer all voters who live more than 500 m from the city limit are removed from the analysis before matching on geographic distance occurs. [b]The metric is the χ^2 test statistic from a global balance test applied to 18 pretreatment covariates; the analysis is performed at the voter level (each individual observation used is a voter). [c]Price Sq. Ft. Diff is the absolute value of the difference in the average house price per square foot; the analysis is performed at the house level (each individual observation used is a house). Rows labeled *Naive* show the unadjusted mean difference between treatment and control areas included in the buffer. Rows labeled *Geographic* show the mean difference between treatment and control areas included in the buffer after nearest-neighbor matching on chordal (spatial) distance alone.

6.4 Results

Despite the differences between the GRD design and nongeographic RDs, the analysis of both designs should start in an identical fashion. Under both designs, researchers should provide evidence about the plausibility of the identification assumption invoked using falsification tests. We now report whether covariates are similar on either side of the media market boundary segment we selected for analysis using a balance analysis. Given that we have already conditioned on geographic proximity by selecting a segment of the border where legislative and school districts overlap, we expect that covariates should already be comparable.

We check balance on 18 covariates in the voter file and for house prices. Because we analyze a large number of covariates in the voter file, instead of reporting individual-level balance statistics, we use a global measure of balance developed by Hansen and Bowers (2008). For housing prices, we report price per square foot to avoid capturing differences in house sizes. We compare differences in pretreatment covariates in a number of different ways. We begin with a raw comparison between treated and control areas for all voters and house sales in the West Windsor-Plainsboro school district. We then wish to understand whether balance improves depending on whether we use naive distance versus a measure of distance that accounts for spatial locations.

Table 1 contains the results from the balance analysis. We first discuss the results for housing prices. The basic treated-control difference in housing prices indicates a $22 per square foot price difference. As expected, given that we have focused on a relatively homogeneous area, this difference is not large.[16] We then restricted the balance analysis to sales that were within 1000 m of the media market boundary. Now the difference is a mere $3 a square foot. For housing prices, it did not matter whether we conditioned on naive distance or not; in each case we find that the house price differential along the boundary is quite small. Given that house prices should be correlated with school quality and other neighborhood characteristics, the balance we observe helps validate the GRD design.

---

[16]The average in the New York media market area is $211, and the average in the Philadelphia media market is $233. The standard deviation for the entire area is $37. So this $22 difference is less than a standard deviation.

Next, we examine balance in the voter file covariates. Here, we find that using naive distance versus geographic distance matters. While balance improves as we near the border, the imbalance as measured by the global χ^2 statistic is nearly half when we use geographic distance as opposed to naive distance. While the global measure does little to convey individual characteristics, close examination reveals that registered voters in the treated and control areas are predominantly white and have similar levels of education and income. Our analysis reveals that voters along this segment of the media market boundary are well balanced in terms of pretreatment covariates, which lends credibility to the design. We note, however, that geographic distance does not remove all the imbalances in observed covariates. This is a common phenomenon in geographic applications, where conditioning on distance often improves balance but does not eliminate imbalances entirely—see Keele et al. (2014) for a matching estimator that conditions on covariates and distance simultaneously and allows for a sensitivity analysis, and Keele and Titiunik (2014b) for an example with conditioning based on a regression model.

We have seen thus far that we can improve balance considerably by analyzing units that are spatially proximate. While those results are encouraging, they require us to average along the entire segment of the media market boundary that we have selected, which is approximately 7 km long. However, since spatial heterogeneity may occur along these 7 km, the design (i.e., its identification assumptions) might be more credible in some parts of the border than others. And even if the design is equally credible along the entire segment of the border we analyze, the effects of presidential advertisement on turnout may vary along the border—that is, we might have geographically heterogeneous treatment effects. We should note, however, that we do not have any strong reason to suspect heterogeneity. As such, the analysis is exploratory.

We now turn to the results of our local regression estimator introduced in Section 5. We use this estimation framework to first assess the validity of the design at various points along the border, and then to estimate the treatment effect of interest and probe its heterogeneity. For this, we select three different points along the border. The two extreme points are chosen so that they split the boundary in three equal segments, each of which is roughly 2.3 km long. The third point is the midpoint between the two end points. The distance between each of the extreme points and the middle point is therefore approximately 1.15 km. For this reason, whenever data density permits, we include results based on a fixed bandwidth of 1.1 km, which ensures that each observation is used for estimation in exactly one boundary point.

First, to asses the validity of our design, we apply the local linear estimator to covariates other than the outcome to identify segments along the border where the assumption of continuity of potential outcomes (Assumption 1) seems most plausible. We report the results for house prices, a covariate that we think is extremely important because it captures both individual-level and neighborhood-level characteristics that are correlated with turnout—but we also report results for other covariates such as age and party registration in the Appendix. Table 2 reports the media market effects on housing prices, estimated in the three different points using local linear regression. As before, treatment is defined as being in the Philadelphia media market and control as being in the New York media market. The Estimate column contains the difference between the limit of the expected outcome for treated and control units at the boundary point, as defined in Equation 1, estimated with a fixed bandwidth of 1.7 km (data density was too low to consider a 1.1 km bandwidth). We report the two types of confidence intervals described in Section 5. In the Conventional Inference columns, we calculate conventional 95% confidence intervals for the fixed bandwidth of 1.7 km. Since this fixed bandwidth is smaller than the MSE-optimal bandwidth at every point, these conventional confidence intervals are undersmoothed and are therefore valid. In the Robust Inference columns, we report the robust 95% confidence interval calculated with a MSE-optimal bandwidth at each point. The point estimates in Table 2 range from negative to positive across the three boundary points, but in all cases the confidence intervals include zero. Thus, we find no evidence of a significant discontinuity in housing prices at any of the boundary points we consider. We note, however, that the confidence intervals are wide, which is likely because the number of house transactions that occur near the border is relatively small. Indeed, the data density at the house level is much lower than at the voter level, which we analyze next. This illustrates an obstacle that is commonly encountered in RD designs: the number of observations

Table 2 Placebo effect of presidential ads on housing price per sq ft (in dollars)

Boundary Point	Estimate	Conventional Inference				Robust Inference			
		95% CI	h	NTr	NCo	95% CI	h	NTr	Nco
1	10.02	[−106.13, 32.49]	1.70	73	54	[−37.06, 55.47]	3.29	264	242
2	−41.99	[−112.85, 16.40]	1.70	48	85	[−136.95, 36.61]	1.90	56	93
3	−10.47	[−68.76, 52.99]	1.70	40	26	[−122.11, 63.27]	1.74	42	28

Note. Results estimated with local linear regression with triangular kernel weights on each observation's chordal distance to the point of estimation. Estimate indicates the point estimate (difference in price per square foot) using the MSE-optimal bandwidth reported under Robust Inference; NTr and NCo indicate the effective sample size used for estimation in the treated and control areas, respectively; h indicates the bandwidth used (in km). All results estimated with package rdrobust (Calonico, Cattaneo, and Titiunik 2014a, 2014c). In conventional inference, bandwidth is chosen manually. In robust inference, all bandwidths are chosen with the CCT MSE minimization method described in Calonico, Cattaneo, and Titiunik (2014b). MSE-optimal pilot bandwidths used in robust confidence intervals are 4.10, 2.65, and 2.61 in boundary points 1, 2, and 3, respectively.

Table 3 Effect of presidential ads on 2008 turnout

Boundary Point	Estimate	Conventional Inference				Robust Inference			
		95% CI	h	NTr	NCo	95% CI	h	NTr	Nco
1	0.043	[−0.66,0.79]	1.10	233	574	[−0.03, 0.17]	5.59	11729	8983
2	−0.002	[−0.29,0.31]	1.10	401	342	[−0.20, 0.04]	2.89	3994	3222
3	0.040	[−0.37,0.47]	1.10	324	326	[−0.08, 0.24]	4.10	7507	7438

Note. Results estimated with local linear regression with triangular kernel weights on each observation's chordal distance to the point of estimation. Estimate indicates the point estimate (difference turnout proportion across treated and control areas) using the MSE-optimal bandwidth reported under Robust Inference; NTr and NCo indicate the effective sample size used for estimation in the treated and control areas, respectively; h indicates the bandwidth used (in km). All results estimated with package rdrobust (Calonico, Cattaneo, and Titiunik 2014a, 2014c). In conventional inference, bandwidth is chosen manually. In robust inference, all bandwidths are chosen with the CCT MSE minimization method described in Calonico, Cattaneo, and Titiunik (2014b). MSE-optimal pilot bandwidths used in robust confidence intervals are 5.06, 4.07, and 3.78 in boundary points 1, 2, and 3, respectively.

near the cutoff is typically small and leads potentially to low statistical power to reject the null hypothesis. This challenge may be exacerbated in a geographic RD design, where researchers require data density in two dimensions as opposed to one.

Second, in Table 3, we report the effects of presidential ads on voter turnout in 2008 for the same three boundary points. The columns in this table are analogous to those in Table 2. The results are fairly consistent across the three boundary points considered. The point estimates in Points 1 and 3 are remarkably similar to each other, at roughly 4 percentage points, while the point estimate in Point 2 is considerably closer to zero, at −0.2 percentage points. In all cases the (undersmoothed) 95% conventional confidence intervals are roughly centered around zero (all associated *p*-values are above 0.80). The robust 95% confidence intervals include zero in every boundary point as well, although they are no longer centered around zero.[17] Taken together, these results suggest that presidential ads have little effect on voter turnout during the 2008 presidential campaign, consistent with the national-level findings in Huber and Arceneaux (2007) and Krasno and Green (2008).

In Fig. 6, we plot the location of these treatment effects on a map. On the map, the reader can see the points along the border where we estimated the media market effect on turnout. The estimates at these points are limits of conditional expectations, estimated with spatial weights that give stronger weight to voters near the boundary and less weight to voters farther from the boundary. Given the relative homogeneous nature of this area, we do not expect much heterogeneity.

The possibility of plotting both actual and placebo effects on a map is one of the distinctive and, we believe, extremely useful features of the GRD design. This ability to plot provides researchers

---

[17]The robust *p*-values in Points 1, 2, and 3 are, respectively, 0.19, 0.21, and 0.30.
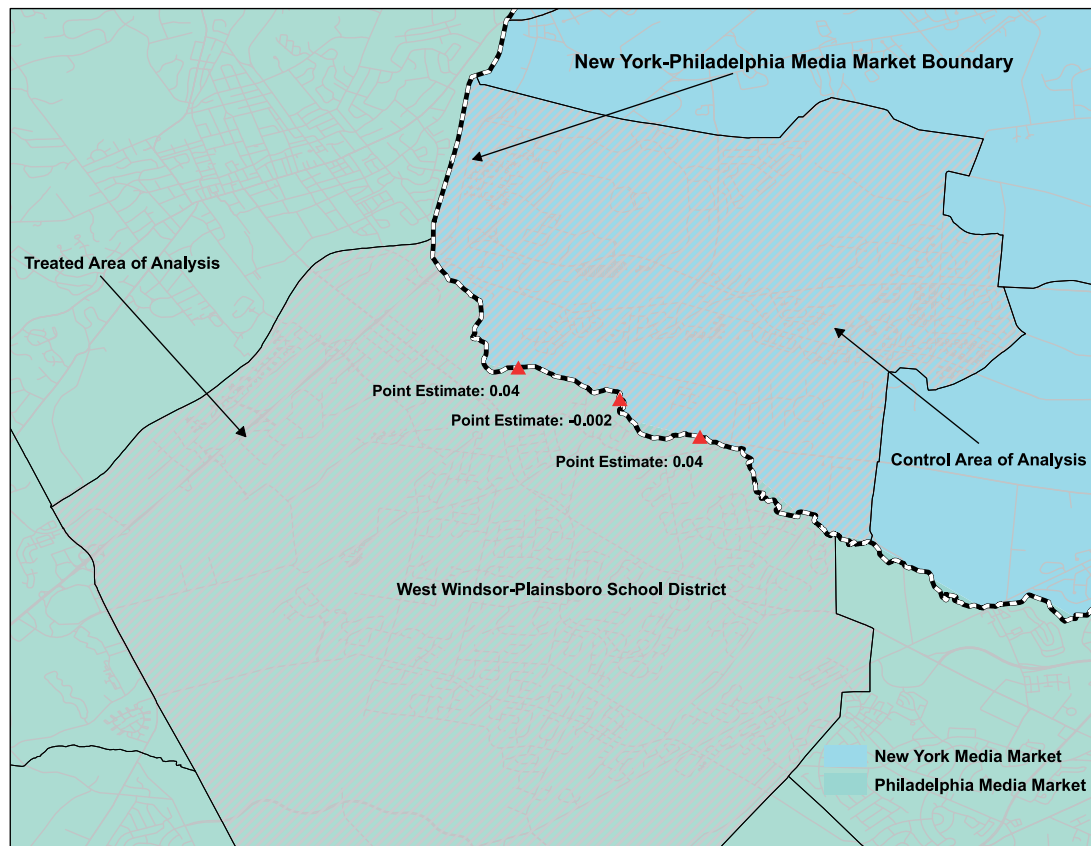
Fig. 6 Geographically located estimated advertisement effects on 2008 voter turnout. Treatment effects estimated at three different points along the boundary between the Philadelphia, PA, media market (located southwest of the boundary) and New York City, NY, media market (located northeast of the boundary) in the state of New Jersey (see Fig. 4). Results estimated with local linear regression with triangular kernel weights on each observation's chordal distance to the point of estimation and bandwidth fixed at 1.1 km. Area marked is the West Windsor-Plainsboro school district, which straddles the media market boundary. Empirical analysis is confined to the West Windsor-Plainsboro school district only, where legislative districts are also the same on both sides of the border. Treated area is southwest of media market boundary, inside Philadelphia media market, where volume of political ads is high; control area is northeast of media market boundary, inside New York City media market, where volume of ads is zero.

with a summary of the geographic heterogeneity in both the treatment effects and the plausibility of the continuity conditions that are needed for the GRD design to yield valid inferences.

7 Recommendations for practice and concluding remarks

We now summarize our recommendations for analysts who wish to use geographic discontinuities to estimate treatment effects, and offer some concluding remarks about the promise and limitations of the methodological framework used here.

- **Data.** Much of the analysis we presented depended on having data that can be georeferenced. Without information about geographic locations, boundaries cannot be fully exploited as discontinuities. Thus, researchers need to collect geographic data (addresses, latitude and longitude, or other geographic information that can be used for geocoding) along with more traditional covariates. Moreover, qualitative research on the history of the border and conditions around it will often prove useful to justify various assumptions.

- **Falsification tests.** Following standard practice in non-spatial RD designs, the credibility of the design can be enhanced by providing evidence that pretreatment covariates become more and more similar as the distance to the border decreases. Identification for the GRD design requires that people cannot precisely sort around the boundary in a way that makes potential outcomes discontinuous. In many GRD designs, we expect that people will be able to sort very precisely around the boundary of interest. For example, features such as the quality of schools and the price of housing may vary discontinuously at the border of interest. Thus, researchers should at least rule out nonzero treatment effects on predetermined covariates, which can be easily implemented using covariates as outcomes in the estimation for each boundary point.

- **Isolating the treatment.** As discussed extensively above, compound treatments are common in GRD designs. In political science applications, a first step is to restrict the analysis to areas around the border where other important geographically defined institutional units are kept constant on either side of border. This includes units such as legislative districts, school districts, counties, cities, municipalities, states, and countries. When not all the relevant units are kept constant on either side of the border, researchers should evaluate whether a plausible Compound Treatment Irrelevance assumption can be made. Placebo outcomes will be helpful to assess whether this assumption is plausible. For example, researchers could show that the outcome of interest was similar among units on both sides of the border *before* the treatment was introduced.

- **Appropriate analysis.** The statistical analysis for the best GRD designs should be relatively simple. Analysts should assess balance on pretreatment covariates for groups on either side of the border. If covariates are balanced, a simple comparison of the outcomes for those on either side of the border will be adequate if the border segment is short and there is little concern about heterogeneity. On the other hand, if compound treatments are present or covariates are imbalanced, more complex analysis will be required. In these cases, placebo maps or tables showing geographically located effects of the treatment on predetermined covariates at various points along the border will be important to identify segments of the border where the GRD's continuity assumptions are most plausible. The estimation framework we proposed in Section 5 is well suited for this purpose.

All in all, we believe that geographic discontinuities are a promising form of natural experiment. The biggest challenge is that often agents are able to sort very precisely around the boundary that forms the discontinuity in the design. Compound treatments will tend to increase sorting, as there will be a larger number of reasons for agents to sort at the boundary. Moreover, an inability to avoid compound treatments can make defining the exact estimand difficult. As such, while the GRD design is a particular case of the two-dimensional RD design, geography creates a number of complications that are not necessarily common in nongeographic designs. We have provided a framework that we believe will allow analysts to address these issues. The implementation of the GRD design requires careful attention to both the statistical analysis needed to justify the continuity assumption at the center of the design and the geographic analysis needed to fully assess and exploit the discontinuity.

Appendix A Interference

In the estimation of treatment effects in randomized experiments or observational studies, we must assume that there is no interference between units, that is, that the outcome of one unit is not affected by the treatment status of other units. Rubin (1986) called the assumption of no interference the "stable unit-treatment value assumption," or SUTVA.[18] In the classic RD, SUTVA

---

[18]SUTVA also includes an assumption that there is only one version of the treatment or, if there are multiple versions of the treatment, that the effects of these different treatment versions are the same.

violations via interference are typically of little concern. For example, there is no reason to suspect that the outcomes for a student just under a scholarship threshold will be affected by the fact that a student just over the same threshold receives the treatment in the form of the scholarship. For a GRD design, however, SUTVA violations might be a concern. In our application, we would worry that a voter exposed to presidential ads may urge a voter not exposed to ads to vote. In general, we doubt such SUTVA violations occur, since it is unlikely that contagion occurs outside households. In our cases, households are close but not so close that we suspect easy spillovers, since the area is suburban and the density of housing is not high. Recent work designed to detect treatment spillovers in the context of turnout finds little evidence of contagion (Sinclair, McConnell, and Green 2012). That is, we doubt that citizens in urban areas actively discuss politics with their neighbors to the extent that they encourage increased participation. Moreover, a SUTVA violation of this kind would tend to bias the effect toward zero, so any positive effects should be conservative estimates. However, we need not ignore SUTVA violations of this type. One simple way to account for a possible SUTVA violation is to estimate treatment effects but exclude voters that might be nearly adjacent (see Keele and Titiunik 2014a).

Appendix B Inference in natural experiments

One question that arises in the context of geographic natural experiments (and may arise in other types of natural experiments) is whether treatment assignment occurs at the individual level or at a more aggregated level of geography. With geographic natural experiments, one could argue that the assignment mechanism operates not at the individual level but at an aggregated geographic level. For example, Card and Krueger (1994) estimated the effect of increasing the minimum wage on employment by comparing fast-food restaurants in New Jersey (where the minimum wage was increased) to restaurants in adjacent eastern Pennsylvania. One could argue that treatment occurs at the state level, and as such this is an experiment with two observations. Card and Krueger, however, analyze their data at a disaggregated level. In our application, it could be argued that the assignment mechanism occurs at the media market level or instead that voters choose to live one side of the media market boundary by chance, which would lead to an individual-level assignment mechanism. If the former assignment mechanism holds, then we have two observations as well: the treated media market and the control media market. The difficulty is that in most natural experiments, the analyst does not explicitly control the assignment mechanism, which means the answer to the level of treatment assignment remains ambiguous.

Statistical inference in settings where either random sampling or random assignment of treatment has not explicitly occurred must invariably rely on concepts such as super populations or hypothetical treatment assignment mechanisms. In natural experiments, one might assume the assignment mechanism is known and use randomization inference (Rosenbaum 2002). In our case, we assume that individual-level assignment occurs around the media market boundary, which serves as the discontinuity. As such, the assignment mechanism is one that focuses on uncertainty about where one lives, and therefore it is an individual-level assignment mechanism. Some would argue that in these cases one should report estimates with no standard errors, since we are working with population data (Berk 2006). Others might argue that only inference based on Bayesian principles is sound (Gill 2002). These are questions about the philosophy of statistical inference that cannot be resolved here or perhaps anywhere. Such questions and debates are endemic to any situation where the analyst does not control either sampling or assignment through a random mechanism.

Appendix C Proof of proposition 1

We write the observed outcome as $Y_i = Y_{i0} + T_i(Y_{i1} - Y_{i0})$ and generalize the proof in Hahn et al. (2001). We let superscripts t and c denote locations that are in the treated and control areas, respectively, $\mathbf{s}^c \in \mathcal{A}_c$ and $\mathbf{s}^t \in \mathcal{A}_t$. Assuming that $\Pr(T_i = 1) = 1$ for all i such that $\mathbf{S}_i \in \mathcal{A}^t$ and $\Pr(T_i = 0) = 1$ for all i such that $\mathbf{S}_i \in \mathcal{A}^c$ $T_i = 1$ for all $i \in \mathcal{A}^t$ (the discontinuity is sharp), that the

density of $\mathbf{S}_i$ is positive in a neighborhood around $\mathbf{b}$, and Assumption 1, we have:

$$E\{Y_i|\mathbf{S}_i = \mathbf{s}^t\} - E\{Y_i|\mathbf{S}_i = \mathbf{s}^c\} = E\{Y_{i0}|\mathbf{S}_i = \mathbf{s}^t\} - E\{Y_{i0}|\mathbf{S}_i = \mathbf{s}^c\}$$

$$+ E\{T_i(Y_{i1} - Y_{i0})|\mathbf{S}_i = \mathbf{s}^t\} - E\{T_i(Y_{i1} - Y_{i0})|\mathbf{S}_i = \mathbf{s}^c\}$$

$$= E\{Y_{i0}|\mathbf{S}_i = \mathbf{s}^t\} - E\{Y_{i0}|\mathbf{S}_i = \mathbf{s}^c\}$$

$$+ 1 \cdot E\{(Y_{i1} - Y_{i0})|\mathbf{S}_i = \mathbf{s}^t\} - 0 \cdot E\{(Y_{i1} - Y_{i0})|\mathbf{S}_i = \mathbf{s}^c\}.$$

Taking limits,

$$\lim_{\mathbf{s}^t \to \mathbf{b}} E\{Y_i|\mathbf{S}_i = \mathbf{s}^t\} - \lim_{\mathbf{s}^c \to \mathbf{b}} E\{Y_i|\mathbf{S}_i = \mathbf{s}^c\} = \lim_{\mathbf{s}^t \to \mathbf{b}} E\{Y_{i0}|\mathbf{S}_i = \mathbf{s}^t\} - \lim_{\mathbf{s}^c \to \mathbf{b}} E\{Y_{i0}|\mathbf{S}_i = \mathbf{s}^c\}$$

$$+ \lim_{\mathbf{s}^t \to \mathbf{b}} E\{Y_{i1} - Y_{i0}|\mathbf{S}_i = \mathbf{s}^t\}$$

$$= E\{Y_{i0}|\mathbf{S}_i = \mathbf{b}\} - E\{Y_{i0}|\mathbf{S}_i = \mathbf{b}\} + E\{Y_{i1} - Y_{i0}|\mathbf{S}_i = \mathbf{b}\}$$

$$= E\{Y_{i1} - Y_{i0}|\mathbf{S}_i = \mathbf{b}\}$$

$$\equiv \tau(\mathbf{b}).$$

Appendix D Alternative statement of identification assumption

It is useful to consider a special case of Assumption 1 that will be useful in applications. For a point $\mathbf{b}$ on the boundary, we define a function $f_{\mathbf{b}}(\mathbf{S}) : \mathbb{R}^2 \to \mathbb{R}$, that represents the distance between any point $\mathbf{S} = (S_1, S_2)$ on the map and the boundary point $\mathbf{b} = (b_1, b_2)$. Particular forms for $f(\cdot)$ include Euclidean and chordal distances and are discussed in Section 6.2.

For every boundary point $\mathbf{b}$, this allows us to collapse the underlying two-dimensional score of unit i into a single-dimensional score, $f_{\mathbf{b}}(\mathbf{S}_i)$, which is the two-dimensional distance between unit i's location on the map and the location of boundary point $\mathbf{b}$. Note that the estimand remains a plane since for every unit i, there will be as many values $f_{\mathbf{b}}(\mathbf{S}_i)$ as points $\mathbf{b}$ on the boundary. As we discuss below, since the number of boundary points is infinite, in practice we consider only a grid of points along the boundary, so that $f_{\mathbf{b}}(\mathbf{S}_i)$ needs to be calculated only for points on the grid.

Assumption 3 (Continuity in a scalar function of two-dimensional score). *The conditional regression functions are continuous in $f_{\mathbf{b}}(\mathbf{b})$ at all points $\mathbf{b}$ on the boundary:*

$$\lim_{\mathbf{s} \to \mathbf{b}} E\{Y_{i0}|f_{\mathbf{b}}(\mathbf{S}_i) = f_{\mathbf{b}}(\mathbf{s})\} = E\{Y_{i0}|f_{\mathbf{b}}(\mathbf{S}_i) = f_{\mathbf{b}}(\mathbf{b})\}$$

$$\lim_{\mathbf{s} \to \mathbf{b}} E\{Y_{i1}|f_{\mathbf{b}}(\mathbf{S}_i) = f_{\mathbf{b}}(\mathbf{s})\} = E\{Y_{i1}|f_{\mathbf{b}}(\mathbf{S}_i) = f_{\mathbf{b}}(\mathbf{b})\}.$$

Since $f_{\mathbf{b}}(\mathbf{b}) = 0$, Assumption 3 can be recast as

$$\lim_{\mathbf{s} \to \mathbf{b}} E\{Y_{i0}|f_{\mathbf{b}}(\mathbf{S}_i) = f_{\mathbf{b}}(\mathbf{s})\} = E\{Y_{i0}|f_{\mathbf{b}}(\mathbf{S}_i) = 0\}$$

$$\lim_{\mathbf{s} \to \mathbf{b}} E\{Y_{i1}|f_{\mathbf{b}}(\mathbf{S}_i) = f_{\mathbf{b}}(\mathbf{s})\} = E\{Y_{i1}|f_{\mathbf{b}}(\mathbf{S}_i) = 0\}.$$

Under Assumption 3, unit i's score, although a scalar, is a function of i's two-dimensional location on the map, and the scalar score $f_{\mathbf{b}}(\mathbf{S}_i)$ of unit i is defined relative to the boundary point $\mathbf{b}$. Therefore, although this score is unidimensional for a fixed boundary point, it will vary as different points $\mathbf{b}$ along the boundary are considered.

Under Assumption 3, the treatment effect identified by the GRD design at the boundary point $\mathbf{b}$ is summarized in the following proposition:

Proposition 2 (Geographic treatment effect curve with scalar function of two-dimensional score). *Assuming that $\Pr(T_i = 1) = 1$ for all i such that $\mathbf{S}_i \in \mathcal{A}^t$ and $\Pr(T_i = 0) = 1$ for all i such that $\mathbf{S}_i \in \mathcal{A}^c$ (the discontinuity is sharp), that the density of $\mathbf{S}_i$ is positive in a neighborhood around zero, and*

Assumption 3, we have:

$$\tau(\mathbf{b}) = \lim_{\mathbf{s}^t \to \mathbf{b}} E\big\{Y_i | f_{\mathbf{b}}(\mathbf{S}_i) = f_{\mathbf{b}}(\mathbf{s}^t)\big\} - \lim_{\mathbf{s}^c \to \mathbf{b}} E\big\{Y_i | f_{\mathbf{b}}(\mathbf{S}_i) = f_{\mathbf{b}}(\mathbf{s}^c)\big\}$$

$$= E\big\{Y_{i1} - Y_{i0} | f_{\mathbf{b}}(\mathbf{S}_i) = 0\big\},$$

where $\mathbf{s}^c \in \mathcal{A}_c$ *and* $\mathbf{s}^t \in \mathcal{A}_t$.

Appendix E Connections to geographic weighted regression

We proposed a framework for estimation in the GRD design that uses local linear regression to estimate the treatment effect at a given boundary point, using the two-dimensional distance of each data point to the boundary point as the score or running variable. In this framework, the treatment effect at each boundary point is estimated giving the highest weight to data points that are closest to the boundary point. This method is closely related to geographically weighted regression (GWR) (Fotheringham, Brundson, and Charlton 1998, 2002), a technique commonly employed in geography and spatial econometrics. In GWR, an outcome is modeled as a function of one or more covariates, and the coefficients on these covariates are allowed to vary at different geographic locations, giving higher weight in the estimation to observations near the location of estimation. GWR is typically used to gauge how the relationship between the dependent and independent variables changes through space. The focus of a GRD design, in contrast, is about estimating the change in the average outcome that occurs exactly at a boundary point when a treatment is active on one side of the boundary but inactive on the other. Because the goal is to estimate the change in the conditional expectation of the outcome at the boundary point, GRD designs estimate the model separately on either side of the border. Also, unlike typical GWR analysis, GRD designs do not include covariates on the right-hand side, because the identification assumption is typically about continuity of the regression function conditional on the score alone. Thus, typically in GRD designs there is no interest in describing the relationship between outcome and covariates generally through space; interest lies only on the boundary points, and the relationship of interest is usually between the outcome and the score alone.

Nonetheless, our treatment of borders and the discontinuity in treatment assignment they generate is related to the treatment of borders in geography. For example, Tam Cho and Nicely (2008) study whether counties that are near and on opposite sides of state borders share political similarities. Analyzing all counties in the United States, the authors find that there is positive spatial autocorrelation in political tendencies (as measured by presidential vote) between counties on a state border and adjacent counties within the same state. But the spatial autocorrelation is no longer significant when the authors analyze the relationship between border counties and their adjacent neighbors in a different state, suggesting that political tendencies change abruptly at state borders. Applying this type of analysis to GRD designs would be fruitful and could provide researchers with one additional tool to gauge the plausibility of the identification assumptions. In particular, an ideal GRD would be one where there is high and positive spatial autocorrelation in a neighborhood of the border between units on both sides of the border, indicating that the two groups compared in the GRD design are indeed similar. Since the treatment would of course be a difference between the adjacent areas, and that difference will itself affect the spatial auto-correlation measure, one would need to look at pre-treatment spatial autocorrelations, in the same spirit as one looks at balance statistics on pre-treatment covariates.

Appendix F Additional empirical results

In this section, we report additional empirical results for important predetermined covariates. Table A1, which is analogous to Table 2 in the main body of the article, shows local linear regression results for several pretreatment covariates available in the voter file. These covariates are age and female, black, Hispanic, and democratic registration indicators at the individual level.

Table A1 Placebo effect of presidential ads on pre-treatment covariates

Boundary Point	Variable	Estimate	Conventional Inference				Robust Inference			
			95% CI	h	NTr	NCo	95% CI	h	NTr	Nco
1	Age	2.790	[−6.753,57.194]	1.10	233	574	[−14.564,6.594]	2.28	2226	1406
2	Age	7.497	[5.459,27.043]	1.10	401	342	[5.263,13.176]	3.58	5960	4832
3	Age	1.424	[−15.256,19.049]	1.10	324	326	[−0.934,12.111]	3.14	4405	4841
1	Black	−0.007	[−0.101,0.066]	1.10	233	574	[−0.112,0.037]	1.93	1693	1097
2	Black	−0.008	[−0.191,0.027]	1.10	401	342	[−0.076,0.040]	2.68	3064	2795
3	Black	−0.033	[−0.094,0.033]	1.10	324	326	[−0.056,0.003]	1.37	589	337
1	Hisp	−0.001	[−0.206,0.105]	1.10	233	574	[−0.075,0.025]	3.48	5421	3457
2	Hisp	−0.005	[−0.352,0.003]	1.10	401	342	[−0.088,0.023]	3.32	5240	3936
3	Hisp	−0.004	[−0.106,0.075]	1.10	324	326	[−0.054,0.040]	3.12	4359	4798
1	Dem	0.063	[−0.350,1.292]	1.10	233	574	[0.016,0.188]	6.36	13261	9460
2	Dem	−0.054	[−0.399,0.279]	1.10	401	342	[−0.270,0.003]	3.02	4345	3454
3	Dem	−0.106	[−0.505,0.132]	1.10	324	326	[−0.298,−0.097]	3.92	6745	7181
1	Female	0.052	[−0.908,0.808]	1.10	233	574	[0.027,0.200]	5.30	11053	8466
2	Female	0.049	[−0.393,0.294]	1.10	401	342	[−0.062,0.190]	4.04	7183	7370
3	Female	−0.002	[−0.482,0.517]	1.10	324	326	[−0.091,0.137]	6.31	14372	9460

Note. Results estimated with local linear regression with triangular kernel weights on each observation's chordal distance to the point of estimation. Estimate indicates the point estimate (difference in pre-treatment covariate across treated and control areas) using the MSE-optimal bandwidth reported under Robust Inference; NTr and NCo indicate the effective sample size used for estimation in the treated and control areas, respectively; h indicates the bandwidth used (in km). All results estimated with package rdrobust (Calonico, Cattaneo, and Titiunik 2014a, 2014c). In conventional inference, bandwidth is chosen manually. In robust inference, all bandwidths are chosen with the CCT MSE minimization method. MSE-optimal pilot bandwidths used in robust confidence intervals are 2.63, 4.07, 3.28, 2.88, 3.15, 1.11, 4.05, 3.66, 3.34, 6.19, 3.95, 5.28, 5.87, 4.16, and 5.39 for every boundary point and variable in the order reported in the table, respectively. Last four covariates reported are indicators equal to one if individual is black, Hispanic, registered Democrat, or female.

References

Anderson, Michael L. 2008. Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, Perry preschool, and early training projects. *Journal of the American Statistical Association* 103(484):1481–95.

Ansolabehere, Stephen D., and Eitan Hersh. 2012. What big data reveal about survey misreporting and the real electorate. *Political Analysis* 20(4):437–59.

Asiwaju, A.I. 1985. *Partitioned Africa: Ethnic relations an Africa's international boundaries, 1884–1984*. London: C. Hurst.

Banerjee, Sudipto. 2005. On geodetic distance computations in spatial modeling. *Biometrics* 61(2):617–25.

Benjamini, Yoav, and Yosef Hochberg. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B* 57(1):289–300.

Berger, Daniel. 2009. Taxes, institutions and local governance: Evidence from a natural experiment in colonial Nigeria. Unpublished manuscript.

Berk, Richard A. 2006. *Regression analysis: A constructive critique*. Thousand Oaks, CA: Sage Publications.

Black, Sandra E. 1999. Do better schools matter? Parental valuation of elementary education. *Quarterly Journal of Economics* 114(2):577–99.

Broockman, David E. 2009. Do congressional candidates have reverse coattails? Evidence from a regression discontinuity design. *Political Analysis* 17(4):418–34.

Calonico, Sebastian, Matias D. Cattaneo, and Rocío Titiunik. 2014a. Robust data-driven inference in the regression-discontinuity design. *Stata Journal*.

———. Forthcoming 2014b. Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*. https://www.econometricsociety.org/publications/econometrica/journal-materials/forthcoming-papers.

———. 2014c. rdrobust: An R package for robust inference in regression-discontinuity designs. Working paper.

Card, David, and Alan B. Krueger. 1994. Minimum wages and employment: A case study of the fast-food industry in New Jersey and Pennsylvania. *American Economic Review* 84(4):772–93.

Cattaneo, Matias D., Brigham Frandsen, and Rocío Titiunik. 2014. Randomization inference in the regression discontinuity design: An application to party advantages in the U.S. Senate. *Journal of Causal Inference*. DOI: 10.1515/jci-2013-0010.

Caughey, Devin, and Jasjeet S. Sekhon. 2011. Elections and the regression discontinuity design: Lessons from close U.S. House races, 1942–2008. *Political Analysis* 19(4):385–408.

Cohn, Molly J., and Saul P. Jackman. 2011. A comparison of aspatial and spatial measures of segregation. *Transactions in GIS* 14(1):47–66.

Cook, Thomas D., William R. Shadish, and Vivian C. Wong. 2008. Three conditions under which experiments and observational studies produce comparable causal estimates: New findings from within-study comparisons. *Journal of Policy Analysis and Management* 27(4):724–50.

Cox, David R. 1958. *Planning of experiments*. New York: Wiley.

Dell, Melissa. 2010. The persistent effects of Peru's mining Mita. *Ecometrica* 78(6):1863–903.

Eggers, Andrew C., and Jens Hainmueller. 2009. MPs for sale? Returns to office in postwar British politics. *American Political Science Review* 103(4):513–33.

Eggers, Andrew C., Anthony Fowler, Jens Hainmueller, Andrew B. Hall, and James M. Snyder. 2014. On the validity of the regression discontinuity design for estimating electoral effects: New evidence from over 40,000 close races. *American Journal of Political Science*, doi: 10.1111/ajps.12127.

Fan, J., and I. Gijbels 1996. *Local polynomial modelling and its applications*: Monographs on statistics and applied probability, Vol 66. Boca Raton, FL: Chapman & Hall/CRC.

Fotheringham, A. Stewart, Chris Brundson, and Martin Charlton. 1998. Geographically weighted regression: A natural evolution of the expansion method for spatial data analysis. *Environment and Planning* 30:1905–27.

———. 2002. *Geographically weighted regression: The analysis of spatially varying relationships*. Hoboken, NJ: Wiley and Sons.

Fotheringham, A. Stewart, and D. Wong. 1991. The modifiable areal unit problem in multivariate statistical analysis. *Environment and Planning* 23(4):1025–44.

Gerber, Alan S., Daniel P. Kessler, and Marc Meredith. 2011. The persuasive effects of direct mail: A regression-discontinuity-based approach. *Journal of Politics* 73(1):140–55.

Gerber, Elisabeth R., and Daniel J. Hopkins. 2011. When mayors matter: Estimating the impact of mayoral partisanship on city policy. *American Journal of Political Science* 55(2):326–39.

Gill, Jeff. 2002. *Bayesian methods: A social and behavioral sciences approach*. Boca Raton, FL: Chapman & Hall/CRC.

Goldstein, Kenneth, Michael Franz, and Travis Ridout. 2008. Political advertising in 2008. Combined file [dataset], final release.

Green, Donald P., Terence Y. Leong, Holger Kern, Alan S. Gerber, and Christopher W. Larimer. 2009. Testing the accuracy of regression discontinuity analysis using experimental benchmarks. *Political Analysis* 17(4): 400–17.

Hahn, Jinyong, Petra Todd, and Wilbert van der Klaauw. 2001. Identification and estimation of treatments effects with a regression-discontinuity design. *Econometrica* 69(1):201–09.

Hansen, Ben B., and Jake Bowers. 2008. Covariate balance in simple, stratified, and clustered comparative studies. *Statistical Science* 23(2):219–36.

Hernán, Miguel A., and Tyler J. VanderWeele. 2011. Compound treatments and transportability of causal inference. *Epidemiology* 22(3):368–77.

Huber, Gregory A., and Kevin Arceneaux. 2007. Identifying the persuasive effects of presidential advertising. *American Journal of Political Science* 51(4):957–77.

Imbens, Guido W., and Karthik Kalyanaraman. 2012. Optimal bandwidth choice for the regression discontinuity estimator. *Review of Economic Studies* 79(3):933–59.

Imbens, Guido W., and Thomas Lemieux. 2008. Regression discontinuity designs: A guide to practice. *Journal of Econometrics* 142(2):615–35.

Imbens, Guido W., and Tristan Zajonc. 2011. *Regression discontinuity design with multiple forcing variables*. Working Paper.

Keele, Luke J., Rocío Titiunik, and José Zubizarreta. 2014. Enhancing a geographic regression discontinuity design through matching to estimate the effect of ballot initiatives on voter turnout. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. doi: 10.1111/rssa.12056.

Keele, Luke J., and Rocío Titiunik. 2014a. *Bounding treatment effects under interference in geographic natural experiments: An application to all-mail voting in Colorado*. Working Paper.

———. Forthcoming 2014b. Natural experiments based on geography. *Political Science Research and Methods*. Working Paper.

Kern, Holger L., and Jens Hainmueller. 2008. Opium for the masses: How foreign media can stabilize authoritarian regimes. *Political Analysis* 17(2):377–99.

Krasno, Jonathan S., and Donald P. Green. 2008. Do televised presidential ads increase voter turnout? Evidence from a natural experiment. *Journal of Politics* 70(1):245–61.

Laitin, David D. 1986. *Hegemony and culture: Politics and religious change among the Yoruba*. Chicago: University of Chicago Press.

Lee, David S. 2008. Randomized experiments from non-random selection in U.S. House elections. *Journal of Econometrics* 142(2):675–97.

Lee, David S., and Thomas Lemieux. 2010. Regression discontiuity designs in economics. *Journal of Economic Literature* 48(2):281–355.

Malpezzi, Stephen. 2002. Hedonic pricing models and house price indexes: A select review. In *Housing economics and public policy: Essays in honour of Duncan Maclennan*, eds. Kenneth Gibb and Anthony O'Sullivan, 67–89. Oxford: Blackwell Publishing.

Matzkin, Rosa L. 2007. Nonparametric identification. *Handbook of Econometrics* 6:5307–68.

Middleton, Joel A., and Donald P. Green. 2008. Do community-based voter mobilization campaigns work even in battleground states? Evaluating the effectiveness of Move On's 2004 outreach campaign. *Quarterly Journal of Political Science* 3(1):63–82.

Miguel, Edward. 2004. Tribe or nation? Nation building and public goods in Kenya versus Tanzania. *World Politics* 56(3):327–62.

Miles, William F. S. 1994. *Hausaland divided: Colonialism and independence in Nigeria and Niger*. Ithaca, NY: Cornell University Press.

Miles, William F. S., and David Rochefort. 1991. Nationalism versus ethnic identity in sub-Saharan Africa. *American Political Science Review* 85(2):393–403.

Nall, Clayton. 2012. The road to division: Interstate highways and geographic polarization. Unpublished manuscript.

Openshaw, S. 1984. *The modifiable areal unit problem*. Norwich, CT: Geo Books.

Papay, John P., John B. Willett, and Richard J. Murnane. 2011. Extending the regression-discontinuity approach to multiple assignment variables. *Journal of Econometrics* 161(2):203–07.

Porter, Jack. 2003. Estimation in the regression discontinuity model. Unpublished manuscript.

Posner, Daniel N. 2004. The political salience of cultural difference: Why Chewas and Tumbukas are allies in Zambia and adversaries in Malawi. *American Political Science Review* 98(4):529–45.

Reardon, Sean F., and John O'Sullivan. 2004. Measures of spatial segregation. *Sociological Methodology* 34(1):121–62.

Rosenbaum, Paul R. 2002. *Observational studies*. 2nd ed. New York: Springer.

Rubin, Donald B. 1986. Which ifs have causal answers. *Journal of the American Statistical Association* 81(396):961–62.

Sheppard, Stephen. 1999. Hedonic analysis of housing markets. In *Applied urban economics: Handbook of regional and urban economics*, volume 3, chapter 41, eds. Paul Cheshire and Edwin S. Mills, 1595–635. Amsterdam, The Netherlands: Elsevier.

Sinclair, Betsy, Margaret McConnell, and Donald P. Green. 2012. Detecting spillover in social networks: Design and analysis of multilevel experiments. *American Journal of Political Science* 56(4):1055–69.

Tam Cho, Wendy K., and Erinn P. Nicely. 2008. Geographic proximity versus institutions: Evaluating borders as real political boundaries. *American Politics Research* 36(6):803–23.

Trounstine, Jessica. 2011. Evidence of a local incumbency advantage. *Legislative Studies Quarterly* 36(2):255–80.

VanderWeele, Tyler J. 2009. Concerning the consistency assumption in causal inference. *Epidemiology* 20(6):880–83.

Wong, D. 2008. A local multidimensional approach to evaluate changes in segregation. *Urban Geography* 29(3):455–72.